



Universidad del Desarrollo
Facultad de Ingeniería

EVOLUCIÓN DE LA PANDEMIA DE COVID-19 EN LA REGIÓN
METROPOLITANA: UN ESTUDIO DE ESTADOS OCULTOS USANDO
HIDDEN MARKOV MODEL Y SU RELACIÓN CON ÍNDICES DE MOVILIDAD Y
FACTORES SOCIODEMOGRÁFICOS

POR: FERNANDO LAGOS ALVARADO Y CONSTANZA NEIRA URRUTIA

Proyecto de grado presentado a la Facultad de Ingeniería de la Universidad
del Desarrollo para optar al grado académico de Magíster en Data Science

PROFESOR GUÍA: DR. MAURICIO HERRERA MARÍN

Diciembre, 2024

SANTIAGO

Dedicatoria Fernando Lagos:

A mi esposa, por su amor diario, paciencia infinita y por ser siempre mi mayor fuente de inspiración. A mi madre y a mi abuela, por ser mis primeras maestras y enseñarme con amor y sabiduría.

Dedicatoria Constanza Neira:

A mi mayor tesoro, mi Juli. Para que siempre recuerdes que los sueños se alcanzan con esfuerzo, dedicación y fe. Nunca dejes de perseguirlos.

AGRADECIMIENTOS

Agradecimientos Fernando Lagos

Quiero expresar mi más profundo agradecimiento a todos los que permitieron el desarrollo de este camino. En primer lugar, quiero agradecer a mi esposa, cuya paciencia, confianza y amor incondicional me permitieron dedicar el tiempo y la energía necesarios para desarrollar este Capstone y culminar el Magíster. Sin su constante apoyo, no habría tenido la fortaleza ni el ánimo para enfrentar los desafíos que surgieron día a día.

A mi madre y abuela, les agradezco por inculcarme valores y virtudes que me han guiado a ser una persona de bien, así como por el amor y respaldo que me han brindado a lo largo de mi vida. Han sido mi fuente constante de motivación para esforzarme y ser cada día mejor.

A mi compañera Constanza, le agradezco profundamente por su dedicación, esfuerzo y compromiso en cada etapa de este proyecto. Su entrega y sacrificio fueron esenciales para alcanzar nuestras metas y superar los retos que enfrentamos juntos.

A nuestro profesor guía, expreso mi gratitud por su confianza, colaboración y apoyo constante. Su guía nos motivó a dar lo mejor de nosotros y a superar cada desafío con determinación.

Finalmente, gracias a Dios por las experiencias y el camino que me ha otorgado en esta vida. Estas vivencias me han moldeado y dado la oportunidad de ser quien soy hoy, impulsándome a seguir aprendiendo y a contribuir con un granito de arena para construir un mundo mejor.

Agradecimientos Constanza Neira

Quiero agradecer a todos quienes fueron parte de este proceso. En primer lugar a Fernando, mi compañero de tesis, por su dedicación, compromiso y trabajo en equipo a lo largo de este proyecto. Juntos enfrentamos varios desafíos, demostrando que el esfuerzo compartido nos llevó a alcanzar este importante logro. Tu apañe fue fundamental en este camino y siempre estaré agradecida por eso.

A nuestro profesor guía, Mauricio Herrera, por su orientación, paciencia y sabiduría durante todo el proceso. Gracias por su dedicación y por creer en nuestro potencial.

A mi familia, por ser mi refugio incondicional y por brindarme siempre la fortaleza para seguir adelante. A mi pareja, por caminar a mi lado en cada paso de este camino, compartiendo mis alegrías y aliviando mis cargas en los momentos difíciles. A mi hija, quien con su sonrisa y ternura iluminó mis días y me recordó la importancia de seguir adelante con determinación y esperanza.

A mis amigos, quienes con sus palabras de aliento y compañía me acompañaron en este viaje, demostrando que su apoyo es un pilar esencial en mi vida. A mis suegros, por su generosidad y cariño, quienes siempre me ofrecieron apoyo y confianza.

Finalmente, agradezco a Dios, por guiarme en este trayecto, darme fuerzas en los momentos más desafiantes y por las bendiciones que me permitieron llegar hasta aquí. Sin su luz y la presencia de todas estas personas maravillosas en mi vida, este logro no habría sido posible. A todos ellos, les dedico este trabajo con el corazón lleno de gratitud y amor.

TABLA DE CONTENIDO

ÍNDICE DE TABLAS.....	1
ÍNDICE DE FIGURAS	3
RESUMEN.....	8
INTRODUCCIÓN.....	10
PREGUNTA DE INVESTIGACIÓN	12
MARCO REFERENCIAL	13
1. Pandemia del Covid-19: Una Perspectiva General	13
2. Factores Sociodemográficos y su Relación con la Propagación del Covid-19 .	16
3. Movilidad Interna y Externa como Factor Crítico en la Propagación del Virus..	22
4. Análisis y Modelamiento de Series de Tiempo en el Estudio de Pandemias....	29
5. Modelamiento Predictivo con Hidden Markov Model	37
HIPÓTESIS	42
OBJETIVOS	43
Objetivo General	43
Objetivos Específicos	43
DATOS Y METODOLOGÍA.....	44
1. Datos	44
2. Metodología.....	47
2.1. Exploración de los Datos	47
2.2. Predicción Baseline.....	48
2.3. Modelado de Estados Ocultos	50
2.4. Análisis de Secuencias de Estados	53
2.5. Caracterización de Estados Ocultos	54
2.6. Análisis de Probabilidades de Transición	55

2.7.	Modelos Lineales Mixtos	57
2.8.	Análisis Complementarios.....	62
RESULTADOS		63
1.	Análisis Descriptivo Exploratorio	63
2.	Predicción Baseline	103
2.1.	Baseline con Shallow Classifiers	103
2.2.	Baseline con DeepAR.....	119
3.	Estados Ocultos con Hidden Markov Model.....	125
3.1.	Determinación de Estados Ocultos de Hidden Markov Model.....	125
3.2.	Caracterización de los Estados Ocultos de Hidden Markov Model	130
3.3.	Correlación entre Casos Diarios Covid-19 y Movilidad Interna por cada Estado Oculto de Hidden Markov Model.....	142
3.4.	Correlación entre Casos Diarios Covid-19 y Movilidad Externa por cada Estado Oculto de Hidden Markov Model.....	152
4.	Transiciones entre Estados Ocultos de Hidden Markov Model.....	161
4.1.	Correlaciones entre Índices de Movilidad y Probabilidades de Transición entre Estados Ocultos	162
4.2.	Correlaciones entre Covariables y Probabilidades de Transición	170
5.	Linear Mixed Models para Transiciones entre Estados Ocultos de Hidden Markov Model.....	174
5.1.	Linear Mixed Models para la Mantenición en Estado de Contagio Leve: Transición 1 → 1	177
5.2.	Linear Mixed Models para la Mantenición en Estado de Contagio Moderado: Transición 2 → 2.....	184
5.3.	Linear Mixed Models para la Mantenición en Estado de Contagio Severo: Transición 3 → 3	191
5.4.	Linear Mixed Models para la Transición de Estado de Contagio Leve a Moderado: Transición 1 → 2.....	198
5.5.	Linear Mixed Models para la Transición de Estado de Contagio Leve a Severo: Transición 1 → 3.....	205
5.6.	Linear Mixed Models para la Transición de Estado de Contagio Moderado a Severo: Transición 2 → 3.....	211

CONCLUSIONES.....	218
TRABAJOS FUTUROS.....	223
LIMITACIONES.....	225
REFERENCIAS.....	227

ÍNDICE DE TABLAS

Tabla 1. Comparación de Métricas MAE y RMSE entre Modelos Support Vector Machine, Random Forest y XGBoost.....	104
Tabla 2. Comparación de Métricas MAE y RMSE entre Modelos Support Vector Machine, Random Forest y XGBoost utilizando Ventana Deslizante de 7 días.	105
Tabla 3. Comparación de Métricas entre Modelos Support Vector Machine, Random Forest y XGBoost Post Validación Cruzada y Ajuste de Hiperparámetros.	112
Tabla 4. Comparación de Métricas de Resultados Globales del modelo DeepAR.	120
Tabla 5. Comparación de Métricas de Resultados del modelo DeepAR en Vitacura.	121
Tabla 6. Comparación de Métricas de Resultados del modelo DeepAR en La Pintana. ...	123
Tabla 7. Evaluación de métricas de ajuste en función del número de Estados Ocultos en el modelo HMM.	128
Tabla 8. Estadísticos Descriptivos de Casos Diarios y Tasas Acumuladas de Covid-19 para cada Estado.....	131
Tabla 9. Estadísticos Descriptivos de Índices de Movilidad Interna y Externa para cada Estado.	132
Tabla 10. Estadísticos Descriptivos de Variables Sociodemográficas para cada Estado..	135
Tabla 11. Correlaciones entre el Índice de Movilidad Interna y las Probabilidades de Transición.	164
Tabla 12. Correlaciones entre el Índice de Movilidad Externa y las Probabilidades de Transición.	168
Tabla 13. Resultados de Linear Mixed Models para la Transición 1 → 1.	179
Tabla 14. Coeficientes de Determinación Marginal y Condicional para la Transición 1 → 1.	180
Tabla 15. Comunas con Residuales Extremos en el Modelo C para Mantenición en Estado 1.	184
Tabla 16. Resultados de Linear Mixed Models para la Transición 2 → 2.	186

Tabla 17. Coeficientes de Determinación Marginal y Condicional para la Transición 2 → 2.	187
Tabla 18. Comunas con Residuales Extremos en el Modelo D para Mantenición en Estado 2.	190
Tabla 19. Resultados de Linear Mixed Models para la Transición 3 → 3.	192
Tabla 20. Coeficientes de Determinación Marginal y Condicional para la Transición 3 → 3.	194
Tabla 21. Comunas con Residuales Extremos en el Modelo D para Mantenición en Estado 3.	197
Tabla 22. Resultados de Linear Mixed Models para la Transición 1 → 2.	200
Tabla 23. Coeficientes de Determinación Marginal y Condicional para la Transición 1 → 2.	201
Tabla 24. Comunas con Residuales Extremos en el Modelo D para la Transición 1 → 2.	204
Tabla 25. Resultados de Linear Mixed Models para la Transición 1 → 3.	206
Tabla 26. Coeficientes de Determinación Marginal y Condicional para la Transición 1 → 3.	207
Tabla 27. Comunas con Residuales Extremos en el Modelo D para la Transición 1 → 3.	211
Tabla 28. Resultados de Linear Mixed Models para la Transición 2 → 3.	213
Tabla 29. Coeficientes de Determinación Marginal y Condicional para la Transición 2 → 3.	214
Tabla 30. Comunas con Residuales Extremos en el Modelo D para la Transición 2 → 3.	217

ÍNDICE DE FIGURAS

Figura 1. Evolución de la Tasa de Casos Acumulados de Covid-19 durante el año 2020...	64
Figura 2. Evolución de la Tasa de Casos Acumulados de Covid-19 por Comuna durante el año 2020.....	66
Figura 3. Distribución de la Tasas de Casos Acumulados de Covid-19 por Comuna durante el año 2020.....	67
Figura 4. Evolución Mensual de la Tasa de Casos Acumulados de Covid-19 por Comuna durante el año 2020.....	69
Figura 5. Cantidad de Inmigrantes por cada Comuna de la Región Metropolitana.....	70
Figura 6. Promedio de Personas por Hogar en las Comunas de la Región Metropolitana..	71
Figura 7. Total de Hogares con Hacinamiento por cada Comuna de la Región Metropolitana.	72
Figura 8. Ingreso Per Cápita por Comuna en la Región Metropolitana.....	73
Figura 9. Índice de Precariedad de la Vivienda en distintas comunas de la Región Metropolitana.....	74
Figura 10. Índice de Prioridad Social por Comuna en la Región Metropolitana.....	75
Figura 11. Distribución de las Comunas de la Región Metropolitana según Índice de Prioridad Social.....	76
Figura 12. Porcentaje de Personas con Educación Superior por Comuna.....	77
Figura 13. Porcentaje de Personas por Comuna que poseen Postgrado.....	78
Figura 14. Distribución del Índice de Movilidad Interna por Comuna durante el año 2020..	80
Figura 15. Índice de Movilidad Externo por Comuna durante el año 2020.	81
Figura 16. Evolución de los Índices de Movilidad Interno y Externo durante el año 2020. ...	83
<i>Figura 17. Visualización del Flujo de Viaje Origen – Destino entre las distintas Comunas de la Región Metropolitana.....</i>	<i>84</i>
Figura 18. Visualización del Flujo de Viajes hacia las 10 Comunas con más Movimiento de Destino.	85
Figura 19. Visualización del Flujo de Viajes Salientes desde las 10 Comunas con más Movimiento de Origen a sus respectivos Destinos.....	86

Figura 20. Visualización del Flujo de Viajes Pre - Pandemia durante Horario Punta AM entre 07:00 y 08:59 horas.....	88
Figura 21. Visualización del Flujo de Viajes Pre - Pandemia durante Horario Punta PM entre 18:00 y 19:59 horas.....	89
Figura 22. Visualización del Flujo de Viajes a Inicios de Pandemia durante Horario Punta AM entre 07:00 y 08:59 horas.....	90
Figura 23. Visualización del Flujo de Viajes a Inicios de Pandemia durante Horario Punta PM entre 18:00 y 19:59 horas.....	92
Figura 24. Visualización del Flujo de Viajes en Pandemia durante Horario Punta AM entre 07:00 y 08:59 horas.....	93
Figura 25. Visualización del Flujo de Viajes en Pandemia durante Horario Punta PM entre 18:00 y 19:59 horas.....	94
Figura 26. Movilidad Interna entre Comunas de la Región Metropolitana Previo, al Inicio y Durante la Pandemia.....	96
Figura 27. Evolución de Tasas Acumuladas e Índices de Movilidad durante Pandemia el año 2020.....	98
Figura 28. Evolución de las Tasas de Contagio y Fases del Plan Paso a Paso por Comuna el año 2020.....	99
Figura 29. Correlación entre tasa de casos acumulados, índices de movilidad y variables sociodemográficas.....	101
Figura 30. Series de Tasa de Casos Diarios Acumulados por Comuna.....	103
Figura 31. Comparación de MAE y RMSE entre Modelos Support Vector Machine, Random Forest y XGBoost.....	105
Figura 32. Visualización de la comparación de Métricas MAE y RMSE entre Modelos Support Vector Machine, Random Forest y XGBoost utilizando Ventanas de Tiempo.....	106
Figura 33. Comportamiento Real de la Tasa de Casos Diarios por Comuna versus Predicciones realizadas con Random Forest utilizando Ventana Deslizante.....	108
Figura 34. Comportamiento Real de la Tasa de Casos Diarios por Comuna versus Predicciones realizadas con Support Vector Machine utilizando Ventana Deslizante.....	109

Figura 35. Comportamiento Real de la Tasa de Casos Diarios por Comuna versus Predicciones realizadas con XGBoost utilizando Ventana Deslizante.	110
Figura 36. Visualización de la Comparación de Métricas entre Modelos Support Vector Machine, Random Forest y XGBoost Post Validación Cruzada y Ajuste de Hiperparámetros.	112
Figura 37. Visualización de las Predicciones Random Forest versus Valores reales por Comuna Posterior a los Ajustes.	114
Figura 38. Visualización de las Predicciones Support Vector Machine versus Valores reales por Comuna Posterior a los Ajustes.	115
Figura 39. Visualización de las Predicciones XGBoost versus Valores reales por Comuna Posterior a los Ajustes.	116
Figura 40. Curvas de Aprendizaje para Random Forest.....	117
Figura 41. Curvas de Aprendizaje para Support Vector Machine.	118
Figura 42. Curvas de Aprendizaje para XGBoost.	119
Figura 43. Predicción DeepAR vs Contagios Diarios de Vitacura, ventana 07 días.	121
Figura 44. Predicción DeepAR vs Contagios Diarios de Vitacura, ventana 15 días.	122
Figura 45. Predicción DeepAR vs Contagios Diarios de Vitacura, ventana 30 días.	122
Figura 46. Predicción DeepAR vs Contagios Diarios de La Pintana, ventana 07 días.	123
Figura 47. Predicción DeepAR vs Contagios Diarios de La Pintana, ventana 15 días.	124
Figura 48. Predicción DeepAR vs Contagios Diarios de La Pintana, ventana 30 días.	124
Figura 49. Comparación de Métricas para Selección del Número de Estados Ocultos.....	127
Figura 50. Serie de Viterbi de Estados Ocultos por Comuna y Tiempo.	130
Figura 51. Resumen de estadísticos descriptivos de todas las variables por cada Estado Oculto.	137
Figura 52. Distribución de la Cantidad de Días por Estado Oculto para cada Comuna.....	139
Figura 53. Resumen de estadísticos descriptivos de todas las variables por cada estado considerando cálculos ponderados por cantidad de días en cada estado oculto.	141
Figura 54. Correlación entre Casos Diarios Covid-19 y Movilidad Interna en Estado 1 de Contagio Leve.	144

Figura 55. Correlación entre Casos Diarios Covid-19 y Movilidad Interna en Estado 2 de Contagio Moderado.	146
Figura 56. Correlación entre Casos Diarios Covid-19 y Movilidad Interna en Estado 3 de Contagio Severo.	148
Figura 57. Comparación de la Correlación entre Casos Diarios y Movilidad Interna previo y posterior a Periodo de Contagio Severo.	152
Figura 58. Correlación entre Casos Diarios Covid-19 y Movilidad Externa en Estado 1 de Contagio Leve.	154
Figura 59. Correlación entre Casos Diarios Covid-19 y Movilidad Externa en Estado 2 de Contagio Moderado.	156
Figura 60. Correlación entre Casos Diarios Covid-19 y Movilidad Externa en Estado 3 de Contagio Severo.	158
Figura 61. Comparación de la Correlación entre Casos Diarios y Movilidad Externa previo y posterior al Periodo de Contagio Severo.	160
Figura 62. Relación entre el Índice de Movilidad Interna y las Probabilidades de Transición.	165
Figura 63. Relación entre el Índice de Movilidad Interna y las Probabilidades de Transición por Comuna.	166
Figura 64. Relación entre el Índice de Movilidad Externa y las Probabilidades de Transición.	169
Figura 65. Relación entre el Índice de Movilidad Externa y las Probabilidades de Transición por Comuna.	170
Figura 66. Correlación entre Covariables y las Probabilidades de Transición.	174
Figura 67. Resultados de Elastic Net para evaluación selección de variables relevantes.	176
Figura 68. Efectos Aleatorios por Comuna del Modelo C para Mantenimiento en Estado 1.	181
Figura 69. Efectos Aleatorios por Tiempo del Modelo C para Mantenimiento en Estado 1.	182
Figura 70. Análisis de Residuales de las Comunas a lo largo del Tiempo Modelo C para Mantenimiento en Estado 1.	183
Figura 71. Efectos Aleatorios por Comuna del Modelo D para Mantenimiento en Estado 2.	188
Figura 72. Efectos Aleatorios por Tiempo del Modelo D para Mantenimiento en Estado 2.	189

Figura 73. Análisis de Residuales de las Comunas a lo largo del Tiempo Modelo D para Mantención en Estado 2.....	190
Figura 74. Efectos Aleatorios por Comuna del Modelo D para Mantención en Estado 3. .	195
Figura 75. Efectos Aleatorios por Tiempo del Modelo D para Mantención en Estado 3.	196
Figura 76. Análisis de Residuales de las Comunas a lo largo del Tiempo Modelo D para Mantención en Estado 3.....	197
Figura 77. Efectos Aleatorios por Comuna del Modelo D para la Transición 1 → 2.	202
Figura 78. Efectos Aleatorios por Tiempo del Modelo D para la Transición 1 → 2.	203
Figura 79. Análisis de Residuales de las Comunas a lo largo del Tiempo Modelo D para la Transición 1 → 2.....	204
Figura 80. Efectos Aleatorios por Comuna del Modelo D para la Transición 1 → 3.	208
Figura 81. Efectos Aleatorios por Tiempo del Modelo D para la Transición 1 → 3.	209
Figura 82. Análisis de Residuales de las Comunas a lo largo del Tiempo Modelo D para la Transición 1 → 3.....	210
Figura 83. Efectos Aleatorios por Comuna del Modelo D para la Transición 2 → 3.	215
Figura 84. Efectos Aleatorios por Tiempo del Modelo D para la Transición 2 → 3.	216
Figura 85. Análisis de Residuales de las Comunas a lo largo del Tiempo Modelo D para la Transición 2 → 3.....	217

RESUMEN

Este estudio analizó la evolución de los contagios de Covid-19 en las comunas de la Región Metropolitana de Chile, empleando modelos Hidden Markov Models no homogéneos y Linear Mixed Models para identificar estados latentes y modelar las transiciones entre ellos. El objetivo principal fue identificar patrones clave que explicaran la propagación del virus y evaluar el impacto de factores sociodemográficos y de movilidad en las dinámicas de contagio, proporcionando información útil para la toma de decisiones en salud pública.

A través del análisis, se identificaron tres estados latentes asociados a la intensidad de los contagios, definidos como estados leve, moderado y severo. Las transiciones entre estos estados fueron modeladas incorporando índices de movilidad interna y externa, variables sociodemográficas como el hacinamiento, el ingreso per cápita y el nivel educativo, así como efectos aleatorios por comuna y tiempo para capturar diferencias espaciales y temporales.

Los resultados evidenciaron que la movilidad interna y externa desempeñaron un papel determinante en las probabilidades de transición entre estados, destacándose como factores clave en la propagación del virus. Asimismo, las características sociodemográficas demostraron ser esenciales para explicar la heterogeneidad en las dinámicas comunales, mostrando cómo variables como el hacinamiento, ingreso per cápita y el nivel educativo influyen en la evolución de los contagios. Además, los patrones específicos por comuna y temporalidad

revelaron diferencias significativas en las probabilidades de transición, subrayando la necesidad de intervenciones localizadas.

En conjunto, estos hallazgos resaltan la necesidad de segmentar estrategias de salud pública basándose en las características específicas de las comunas y de ajustar las intervenciones según las dinámicas temporales y locales. El uso de este enfoque, que combina análisis contextualizado y modelado avanzado, podría optimizar los resultados en futuras emergencias sanitarias, mejorando la eficiencia y efectividad de las medidas de contención.

Palabras Clave: Covid-19, Hidden Markov Model, Linear Mixed Models, Factores Sociodemográficos, Movilidad Interna y Externa.

INTRODUCCIÓN

La pandemia del Covid-19 ha sido uno de los hitos más disruptivos del siglo XXI, impactando no solo la salud de las personas, sino también la estructura económica y social de las naciones, afectando en mayor medida a los más pobres y vulnerables (1). En Chile, la Región Metropolitana de Santiago, que concentra una alta densidad poblacional y una marcada segregación socioespacial, se vio particularmente impactada por la propagación del virus (2), convirtiéndose rápidamente en el epicentro de los casos confirmados de Covid-19 en el país.

Si bien la evolución del Covid-19 ha sido objeto de numerosos estudios a nivel global, aún existe una necesidad de comprender cómo las dinámicas urbanas y los factores sociodemográficos específicos de cada comuna contribuyeron a la expansión de la pandemia en el contexto local.

Durante los primeros meses de la pandemia, la capital chilena implementó una serie de medidas que incluyeron cuarentenas dinámicas, seguidas del plan “Paso a Paso”, el cual establecía diferentes fases de restricción según la situación epidemiológica de cada comuna. Sin embargo, a pesar de estos esfuerzos, la propagación del virus mostró diferencias notables entre comunas, lo que sugiere que las condiciones sociodemográficas y los patrones de movilidad jugaron un papel importante en la efectividad de las políticas de contención.

Comunas con menores ingresos presentaron dificultades para reducir la movilidad y mostraron una menor efectividad en el control de la pandemia, en

contraste con las comunas de mayores ingresos, donde las medidas fueron más exitosas. Este fenómeno subraya la necesidad de estudiar en profundidad cómo los factores estructurales, como la movilidad urbana y la inequidad socioeconómica, influyen en la propagación del virus y en la efectividad de las intervenciones sanitarias.

El análisis de la pandemia en áreas urbanas complejas, como la Región Metropolitana (RM), requiere la integración de múltiples variables que van más allá de los contagios diarios, considerando factores como el índice de movilidad interna y externa, la densidad de población, el hacinamiento, el ingreso per cápita de las comunas, entre otros, los cuales jugaron un papel esencial en la evolución de los contagios (3). Además, la implementación de políticas de cuarentenas dinámicas, que restringieron de manera desigual la movilidad entre comunas, generó dinámicas específicas que no han sido completamente exploradas.

En este trabajo, se propone la aplicación de Hidden Markov Model (HMM) con el fin identificar y modelar los estados ocultos de la evolución de la pandemia de Covid-19 en las comunas de la Región Metropolitana de Santiago, para analizar las diferencias en la evolución de estos estados según las características sociodemográficas y de movilidad de las comunas, lo que permitirá comprender cómo evolucionaron las etapas de la pandemia a lo largo del tiempo.

PREGUNTA DE INVESTIGACIÓN

¿Cómo se pueden identificar y caracterizar los estados ocultos de la evolución de la pandemia de Covid-19 en las comunas de la Región Metropolitana (RM) mediante Hidden Markov Model (HMM) no Homogéneo, y cómo influyen los índices de movilidad y variables sociodemográficas en las probabilidades de transición entre estos estados?

MARCO REFERENCIAL

1. Pandemia del Covid-19: Una Perspectiva General

El Covid-19, causado por el virus SARS-CoV-2, fue identificado por primera vez en diciembre de 2019 en la ciudad de Wuhan, China (4), y debido a su alta propagación, el 11 de marzo de 2020, la Organización Mundial de la Salud (OMS) declaró al Covid-19 como una pandemia mundial, lo que desencadenó una serie de respuestas a nivel internacional para contener la propagación del virus (5).

Esta enfermedad viral, que comenzó como un brote localizado, rápidamente se convirtió en una pandemia de escala global, afectando a millones de personas en todos los continentes, situándose como la tercera causa de muerte a nivel mundial en el año 2020 y la segunda causa en el año 2021 (6).

La rápida propagación del virus, la falta inicial de tratamientos efectivos y de una vacuna, y las graves complicaciones en los pacientes más vulnerables, como los adultos mayores y aquellos con comorbilidades, contribuyeron a la sobrecarga de los sistemas de salud en todo el mundo (7). Esto generó una crisis sanitaria sin precedentes, acompañada de impactos económicos, sociales y psicológicos que se han extendido más allá del ámbito sanitario (8)(9)(10).

Además de los desafíos médicos, la pandemia del Covid-19 puso de relieve las desigualdades estructurales y las brechas socioeconómicas en diversas sociedades (11). Las respuestas de los gobiernos, como el confinamiento, el

cierre de fronteras, la cuarentena y la restricción de movilidad, variaron ampliamente en su efectividad dependiendo de factores como la infraestructura sanitaria, la gobernanza y la capacidad de hacer cumplir las políticas públicas.

El desarrollo y la distribución de vacunas a finales de 2020 y durante 2021 marcaron un punto de inflexión en la lucha contra la pandemia. Sin embargo, la aparición de variantes del virus, demostraron la continua amenaza de nuevas olas de contagio, subrayando la importancia de mantener medidas de vigilancia epidemiológica, prevención, y estrategias de vacunación masiva (12).

Desde una perspectiva global, el Covid-19 no solo ha sido un desafío médico, sino también un fenómeno multifacético que ha reconfigurado la forma en que los gobiernos y las sociedades enfrentan emergencias sanitarias. La pandemia ha revelado la necesidad de fortalecer los sistemas de salud pública, mejorar la colaboración internacional, y desarrollar modelos predictivos que permitan anticipar y mitigar el impacto de futuras pandemias (13).

Las áreas urbanas fueron particularmente vulnerables a la rápida propagación del Covid-19 debido a la alta densidad poblacional, la movilidad constante de sus habitantes y las complejas interacciones socioeconómicas que caracterizan estos entornos, donde la transmisión del virus se vio facilitada por la cercanía física entre los habitantes, la dependencia del transporte público y la concentración de actividades laborales y recreativas en espacios cerrados.

Además, la infraestructura sanitaria de muchas ciudades metropolitanas se vio rápidamente sobrepasada debido a la gran cantidad de casos graves que requerían hospitalización, sumado a la limitación de recursos como camas de cuidados intensivos, respiradores y personal médico especializado (14)(15). Las grandes ciudades fueron también el punto de entrada y expansión inicial del virus en muchos países, debido a su rol como hubs de transporte internacional, con aeropuertos y terminales de transporte que sirvieron de puerta de entrada para viajeros infectados.

El Covid-19 exacerbó las desigualdades preexistentes en las áreas metropolitanas, donde los barrios más pobres, a menudo con mayor densidad de población y condiciones de vivienda precarias, enfrentaron tasas más altas de contagio y mortalidad. En estos entornos, el acceso limitado a servicios de salud, el hacinamiento y la imposibilidad de implementar medidas de distanciamiento social de manera efectiva contribuyeron a la rápida propagación del virus (16).

Se han estudiado distintos factores que influyeron en la propagación del virus sobre todo en ciudades densamente pobladas, dentro de los cuales se encuentran la densidad poblacional, la movilidad interna y externa, factores sociodemográficos, las condiciones socioeconómicas, la infraestructura de salud y las políticas públicas como restricciones de movilidad, confinamiento, cordón sanitario, entre otras.

2. Factores Sociodemográficos y su Relación con la Propagación del Covid-19

La pandemia del Covid-19 no afectó a todas las personas por igual. Diversos factores sociológicos y demográficos influyeron de manera significativa en la propagación del virus y en la gravedad de sus efectos en distintas comunidades. Factores como el nivel socioeconómico, el acceso a servicios de salud, la calidad de la vivienda, la densidad poblacional y las características específicas de cada región jugaron un papel fundamental en la distribución desigual de las tasas de infección y mortalidad.

Los estudios realizados a lo largo de la pandemia mostraron que las comunidades más vulnerables desde el punto de vista socioeconómico y aquellas que viven en condiciones de hacinamiento fueron las más afectadas por la pandemia (17). En las grandes áreas urbanas, donde las disparidades sociales y económicas son más marcadas, estas diferencias se hicieron aún más evidentes, con tasas de contagio y mortalidad significativamente más altas en barrios de bajos recursos en comparación con áreas más acomodadas.

Las poblaciones con menor acceso a servicios de salud, menor estabilidad económica y condiciones de vivienda más precarias enfrentaron una mayor exposición al virus, una mayor prevalencia de comorbilidades, y un acceso limitado a cuidados médicos, lo que contribuyó a una mayor vulnerabilidad frente al Covid-19, además, las políticas de confinamiento y distanciamiento social

fueron más difíciles de implementar en las comunidades con menores recursos, lo que exacerbó aún más las desigualdades (18)(19).

En el caso específico de la Región Metropolitana de Santiago (Chile), las comunas más vulnerables social y económicamente, como La Pintana y Lo Espejo, registraron algunas de las tasas de contagio y mortalidad más altas del país, en contraste con las comunas con mayor nivel socioeconómico, como Vitacura y Las Condes, lograron mantener tasas más bajas de contagio (20). Esto se debe, en parte, a factores como el hacinamiento, la falta de acceso a servicios de salud de calidad, y la imposibilidad de cumplir con medidas preventivas como el aislamiento social.

Las desigualdades socioeconómicas desempeñaron un papel crucial en la manera en que el Covid-19 afectó a las comunidades urbanas. Desde el acceso a servicios de salud hasta la capacidad de implementar medidas de distanciamiento social, las personas de menor nivel socioeconómico se vieron más expuestas a los efectos de la pandemia y sufrieron mayores tasas de contagio y mortalidad. Las áreas con alta pobreza, empleo informal y falta de acceso a servicios esenciales enfrentaron una propagación del virus más rápida y tasas de recuperación más bajas (21).

El nivel socioeconómico de una comunidad afecta directamente su capacidad para seguir las recomendaciones sanitarias y mitigar el riesgo de contagio. Las personas de ingresos más bajos, que suelen trabajar en sectores esenciales o

informales, no pudieron permitirse el lujo de trabajar desde casa, lo que aumentó su exposición al virus en comparación con aquellos que pudieron realizar teletrabajo (22)(23). Además, estas personas, al depender en mayor medida del transporte público, estuvieron más expuestas a ambientes propensos a la transmisión del virus.

El acceso a servicios de salud de calidad es uno de los principales determinantes del bienestar durante una pandemia. Las comunidades de bajos ingresos, tanto en países desarrollados como en desarrollo, tienen menos acceso a infraestructura hospitalaria y servicios médicos adecuados, lo que afectó su capacidad para recibir atención temprana y tratamiento adecuado durante el brote de Covid-19.

En América Latina, un informe de CEPAL (2020) destacó que las disparidades en el acceso a servicios de salud exacerbaron las desigualdades en los resultados de salud durante la pandemia. Países como Brasil y Perú, donde las desigualdades en salud ya eran marcadas antes de la pandemia, registraron altas tasas de mortalidad en las comunidades más pobres debido a la saturación de los hospitales públicos y la falta de recursos médicos en las zonas más afectadas.

La Organización Panamericana de la Salud (OPS) refirió que la saturación de los hospitales públicos y la falta de recursos en las comunas de bajos ingresos contribuyeron significativamente a las altas tasas de mortalidad, no solo por

Covid-19, sino a nivel general, debido a la limitación de acceso para atención de todo tipo de pacientes.

El hacinamiento y las condiciones de vivienda precarias fueron factores decisivos en la propagación del Covid-19 en muchas áreas urbanas, especialmente en comunidades de bajos ingresos (16). El hacinamiento se refiere a la alta densidad de personas que habitan en espacios reducidos, lo que hace que sea casi imposible cumplir con las medidas de distanciamiento social o aislar a las personas contagiadas de manera efectiva. En muchas ciudades, los hogares de bajos recursos, que a menudo consisten en varias personas compartiendo pequeños espacios, fueron los más afectados por estas limitaciones.

El hacinamiento crea un entorno ideal para la propagación de enfermedades infecciosas, ya que el contacto cercano y constante entre las personas facilita la transmisión del virus. En el caso del Covid-19, donde las medidas de distanciamiento social y aislamiento de personas infectadas fueron cruciales para controlar la propagación, los hogares hacinados no pudieron cumplir adecuadamente con estas recomendaciones.

Varios estudios han evidenciado que las áreas con mayor hacinamiento experimentaron tasas de contagio más altas debido a la imposibilidad de las personas de aislarse adecuadamente dentro de sus hogares (24)(25)(26). En estos espacios, el virus se propagó rápidamente entre los miembros del hogar y luego hacia la comunidad circundante. El estudio concluyó que el hacinamiento

fue un factor determinante en la velocidad de propagación del Covid-19 en áreas urbanas densamente pobladas.

Las viviendas precarias, que a menudo carecen de servicios básicos como acceso a agua potable, ventilación adecuada o sistemas de saneamiento, también contribuyeron a la propagación del Covid-19. En muchas zonas marginalizadas, la falta de recursos y la infraestructura deficiente limitaron la capacidad de los residentes para seguir las medidas preventivas recomendadas, como el lavado frecuente de manos y la ventilación de espacios cerrados.

En América Latina, donde una proporción significativa de la población vive en condiciones de precariedad, las tasas de contagio fueron particularmente altas en los barrios marginales.

En la Región Metropolitana de Santiago, el hacinamiento fue un factor clave en la propagación del Covid-19 en comunas como La Pintana, Puente Alto y El Bosque, donde las tasas de contagio fueron significativamente más altas en comparación con comunas de mayor ingreso. En estas áreas, es común que varias generaciones de una familia vivan en una misma casa, compartiendo espacios reducidos sin posibilidad de aislar a los miembros del hogar que resultan contagiados.

La densidad poblacional y el ingreso per cápita son dos factores clave que influyen en las tasas de mortalidad y morbilidad durante una pandemia como la del Covid-19. Mientras que la densidad poblacional determina la cantidad de

contactos entre las personas y, por lo tanto, las posibilidades de transmisión del virus, el ingreso per cápita afecta la capacidad de los individuos y las comunidades para acceder a recursos médicos y seguir las recomendaciones de salud pública.

La densidad poblacional es un factor que, históricamente, ha estado vinculado a la propagación de enfermedades infecciosas. En las ciudades densamente pobladas, el contacto entre personas es inevitable, ya sea en el transporte público, en el trabajo o en espacios públicos, lo que aumenta las oportunidades de transmisión del virus. Durante la pandemia del Covid-19, las áreas con mayor densidad poblacional registraron tasas más altas de contagio y, en muchos casos, de mortalidad.

El ingreso per cápita también tuvo un impacto significativo en las tasas de mortalidad y morbilidad por Covid-19. Las comunidades de mayores ingresos generalmente tienen mejor acceso a servicios de salud de calidad, viviendas más espaciales donde es posible el aislamiento, y la capacidad económica para cumplir con las medidas preventivas, como el teletrabajo o la adquisición de suministros de protección personal. En contraste, las personas de menores ingresos suelen tener trabajos que requieren presencia física, viven en condiciones de hacinamiento, y enfrentan barreras para acceder a atención médica o realizarse pruebas diagnósticas.

En el contexto de la Región Metropolitana de Santiago, la combinación de alta densidad poblacional y bajos ingresos exacerbó los impactos de la pandemia en las comunas más vulnerables. Puente Alto, La Florida y El Bosque, comunas densamente pobladas y con menores ingresos, registraron algunas de las tasas más altas de contagio y mortalidad en la región. Estas comunas, además de enfrentar hacinamiento, contaban con menor acceso a servicios de salud privados y recursos médicos, lo que agravó la situación durante los peaks de la pandemia.

3. Movilidad Interna y Externa como Factor Crítico en la Propagación del Virus

La movilidad urbana fue uno de los factores más importantes en la propagación del Covid-19 en las áreas metropolitanas. El desplazamiento diario de personas, tanto dentro de las ciudades como entre regiones, facilitó la transmisión del virus, especialmente en lugares con alta densidad poblacional y una fuerte dependencia del transporte público. La capacidad del virus para transmitirse a través del contacto cercano, tanto en espacios abiertos como cerrados, hizo que la movilidad urbana se convirtiera en un desafío clave para los gobiernos al intentar controlar los brotes.

En muchas ciudades del mundo, las restricciones de movilidad, como el confinamiento y la reducción de la actividad económica, lograron disminuir temporalmente las tasas de contagio (27). Sin embargo, la necesidad de

mantener ciertas actividades esenciales y la dependencia de los trabajadores esenciales del transporte público significó que la movilidad nunca pudo reducirse completamente. Este equilibrio entre las restricciones de movilidad y la necesidad de mantener el funcionamiento de las ciudades fue un factor central en la estrategia para contener la pandemia.

El papel de la movilidad urbana en la transmisión del Covid-19 fue particularmente evidente en las grandes áreas metropolitanas, donde el flujo constante de personas a través de medios de transporte masivo y espacios públicos generó oportunidades continuas para el contagio. Ciudades como Nueva York, Londres y Santiago de Chile experimentaron aumentos rápidos en las tasas de infección durante las primeras olas de la pandemia, en parte debido a la incapacidad de detener completamente la movilidad interna y externa.

El análisis de datos de movilidad urbana se convirtió en una herramienta fundamental para comprender y predecir la propagación del Covid-19. Las tecnologías de seguimiento de movimientos, como los informes de movilidad de y los datos de transporte público, permitieron a los investigadores y a las autoridades sanitarias observar los patrones de desplazamiento en tiempo real y relacionarlos con las tasas de contagio (28). Esto facilitó la implementación de políticas de salud pública más precisas, como cuarentenas selectivas, y ayudó a identificar focos de transmisión en áreas con alta densidad de desplazamientos.

La movilidad interna y externa referente al desplazamiento de personas en una ciudad o área metropolitana fueron factores determinantes en la propagación del Covid-19. Estas formas de movilidad facilitaron la transmisión del virus en las primeras etapas de la pandemia, permitiendo que los brotes localizados se expandieran rápidamente. La alta interconexión entre regiones urbanas, donde las personas se desplazan diariamente para trabajar, estudiar o realizar otras actividades, aceleró la transmisión del virus a lo largo de diferentes territorios.

La evidencia indica que la movilidad en las áreas metropolitanas fue uno de los principales factores que contribuyó a la propagación del Covid-19, especialmente en las ciudades con una alta dependencia del transporte público y una alta densidad de población. Las personas que realizan desplazamientos diarios para trabajar, comprar bienes esenciales o acceder a servicios de salud tienen más oportunidades de entrar en contacto con personas infectadas. En las ciudades donde una gran parte de la población depende del transporte público, como trenes y autobuses, el riesgo de exposición al virus fue considerablemente mayor.

A lo largo de la pandemia, los gobiernos de todo el mundo implementaron diversas políticas para restringir la movilidad con el fin de reducir la propagación del virus. Estas restricciones incluían confinamientos domiciliarios, cierres de fronteras, limitaciones en el transporte público y la suspensión de vuelos internacionales. Aunque estas medidas lograron reducir temporalmente las tasas de contagio, su efectividad dependió en gran medida del nivel de cumplimiento

por parte de la población y de la capacidad de los gobiernos para hacerlas cumplir.

Al respecto, varios estudios han demostrado que las restricciones de movilidad durante las primeras etapas de la pandemia fueron eficaces para reducir la tasa de reproducción del virus y que el hecho de que se haya limitado la movilidad interna y externa permitió controlar el brote en las áreas más afectadas, sin embargo, los estudios también advirtieron que las restricciones por sí solas no eran suficientes y debían combinarse con otras políticas de salud pública (28)(29)(30).

Otra de las variables consideradas para evaluar el impacto de la movilidad sobre la propagación del Covid-19 corresponde al transporte público, el cual fue considerado uno de los principales puntos críticos para la transmisión del Covid-19 en áreas urbanas, debido a las características propias de estos sistemas: espacios cerrados, gran densidad de personas y alta rotación de pasajeros. En ciudades donde el transporte público es esencial para la movilidad diaria de millones de personas, su uso continuado durante la pandemia facilitó la propagación del virus, especialmente en las horas pico. El contacto cercano entre los pasajeros, la dificultad para mantener el distanciamiento social y la dependencia de este servicio por parte de trabajadores esenciales aumentaron el riesgo de contagio.

El transporte público, como autobuses, trenes y metros, presenta varias condiciones que aumentan la probabilidad de transmisión de enfermedades respiratorias como el Covid-19. Los pasajeros suelen estar en contacto cercano durante los desplazamientos, y muchas superficies compartidas, como pasamanos, asientos y puertas, pueden convertirse en puntos de contagio. Además, la ventilación limitada en los vagones y vehículos facilita la propagación de partículas virales en el aire, lo que incrementa el riesgo de transmisión entre los pasajeros.

Al respecto, un estudio evidenció que las áreas de la ciudad donde los residentes dependían más del transporte público para sus desplazamientos diarios tenían tasas de contagio significativamente más altas, ya que las personas que utilizan el transporte público de manera regular enfrentan una exposición prolongada al virus en espacios cerrados, lo que contribuyó al rápido aumento de casos en las primeras fases de la pandemia (31).

En Santiago de Chile, el sistema de transporte público, que incluye buses y el metro, jugó un papel importante en la transmisión del virus en comunas periféricas con alta densidad poblacional. La alta dependencia de los residentes de estas comunas en el transporte público, combinada con la dificultad para implementar el distanciamiento social en las estaciones y vehículos, facilitó la propagación del Covid-19. Un informe del Ministerio de Transportes y Telecomunicaciones de Chile (2020) destacó que, a pesar de los esfuerzos por

reducir la capacidad de los vehículos y aumentar las frecuencias de limpieza, el riesgo de contagio siguió siendo elevado debido a la gran cantidad de personas que no podían evitar usar el transporte público.

Durante la pandemia del Covid-19, el análisis de datos de movilidad se convirtió en una herramienta esencial para entender y predecir los patrones de propagación del virus. Los datos de movilidad, obtenidos a través de diversas fuentes como aplicaciones móviles, redes sociales, dispositivos GPS y operadores de transporte público, proporcionaron a los investigadores y responsables de políticas una visión en tiempo real sobre cómo se desplazaban las personas durante la pandemia y qué impacto tenía esto en la transmisión del virus. Estos datos fueron fundamentales para crear modelos predictivos que ayudaron a guiar las decisiones de salud pública y a implementar medidas de contención más efectivas.

Un estudio de Kraemer y cols. (2020), en el que se utilizaron datos de movilidad para modelar la propagación del Covid-19 en China durante las primeras fases del brote, evidenció que los patrones de movilidad tenían una correlación directa con la tasa de reproducción del virus (R_0), y que las restricciones de movilidad implementadas en ciudades como Wuhan fueron efectivas para reducir significativamente la propagación del virus (27).

Los modelos predictivos son herramientas matemáticas y computacionales que permiten simular cómo se propagarán las enfermedades en función de diversos

factores, como el comportamiento de la población, la movilidad y las intervenciones de salud pública. Durante la pandemia del Covid-19, los datos de movilidad se integraron en estos modelos para predecir cómo se movería el virus en diferentes regiones y qué impacto tendrían las políticas de confinamiento y distanciamiento social.

Un ejemplo destacado de estos modelos es el Modelo SIR (Susceptible-Infected-Recovered), que se utiliza para simular la dinámica de una epidemia. Al combinar este modelo con datos de movilidad, los investigadores pudieron predecir cómo los cambios en la movilidad, como la reducción del uso de transporte público o el cierre de fronteras, afectarían la propagación del Covid-19. Este tipo de análisis permitió a los gobiernos ajustar sus políticas en tiempo real, dependiendo de los flujos de movilidad y el nivel de exposición de diferentes áreas.

A la vez, un estudio realizado en Italia por Pepe y cols. (2020), publicado en Nature, utilizó datos de movilidad y un modelo epidemiológico para simular la propagación del Covid-19 en las diferentes regiones del país. Los datos de movilidad mostraron que, a medida que las restricciones se relajaban, los desplazamientos entre las ciudades aumentaban, lo que resultó en un aumento en los contagios. Este tipo de modelado ayudó a las autoridades italianas a implementar cuarentenas regionales y a ajustar las medidas de distanciamiento social de manera más precisa (32).

4. Análisis y Modelamiento de Series de Tiempo en el Estudio de Pandemias

Las series de tiempo son conjuntos de datos recolectados en intervalos regulares de tiempo, lo que las convierte en un formato ideal para analizar la evolución temporal de fenómenos dinámicos como la propagación de enfermedades infecciosas, es por esto que su análisis se ha convertido en una herramienta esencial en el estudio de brotes epidémicos y pandemias.

Durante la pandemia del Covid-19, el análisis de series de tiempo permitió a los investigadores monitorear la dinámica del virus a lo largo del tiempo, identificar patrones de transmisión y pronosticar futuros brotes en función de los datos pasados. Además, las técnicas avanzadas de descomposición de series de tiempo permitieron a los expertos distinguir entre diferentes componentes del fenómeno estudiado, como la tendencia a largo plazo, la estacionalidad y el ruido aleatorio, análisis que son clave para poder anticipar los efectos de la relajación de restricciones, cambios estacionales y comportamientos sociales en la evolución de la pandemia.

Estos modelos se basan en datos históricos de casos y muertes para prever la trayectoria de una enfermedad en el corto o mediano plazo, permitiendo modelar el crecimiento y la disminución de los casos durante un brote epidémico. Es por esto que los investigadores utilizan estos modelos para entender la dinámica de

transmisión del virus, y así poder ajustar las políticas públicas en función de los escenarios proyectados.

Un ejemplo importante es el uso de modelos ARIMA y SARIMA, que han sido aplicados en la predicción de brotes de enfermedades infecciosas, incluyendo el Covid-19, ya que permiten capturar tanto las tendencias a largo plazo como los patrones estacionales, lo que es crucial en el contexto de una pandemia prolongada.

Al respecto, un estudio de Benvenuto y cols. (2020), utilizó un modelo ARIMA para predecir el número de casos de Covid-19, demostrando un alto grado de precisión en las predicciones a corto plazo, ayudando a los gobiernos a tomar decisiones informadas sobre restricciones y medidas de contención (33).

Otro de los principales usos de las series de tiempo en epidemiología es la detección de peaks y olas de contagio, lo cual se vio reflejado a lo largo de la pandemia Covid-19, donde los datos de series de tiempo fueron utilizados para identificar aumentos repentinos en los contagios y prever la llegada de nuevas olas de casos. Estos análisis permitieron a los gobiernos y sistemas de salud anticiparse a los momentos de mayor demanda de atención médica y optimizar la asignación de recursos, como camas de hospital y ventiladores.

Las series de tiempo también han sido utilizadas para evaluar el impacto de las intervenciones de salud pública, como los confinamientos, la campaña de vacunación y el distanciamiento social, ya que, al analizar cómo evolucionan los

casos antes y después de una intervención, los investigadores pueden medir su efectividad en la reducción de la transmisión del virus.

Esto se puede evidenciar en un estudio de Flaxman y cols. (2020), publicado en Nature, en el cual se utilizaron modelos de series de tiempo para evaluar el impacto de las medidas de confinamiento implementadas en 11 países europeos durante la primera ola del Covid-19, y los resultados mostraron que las intervenciones redujeron significativamente la tasa de transmisión del virus (R_0) y evitaron millones de muertes (34).

A la vez, este tipo de análisis también ha sido clave para identificar patrones estacionales en la propagación de enfermedades infecciosas, demostrando por ejemplo, que enfermedades respiratorias como la gripe y el Covid-19, tienden a mostrar fluctuaciones estacionales, con aumentos durante los meses más fríos del año. Los modelos estacionales, como el SARIMA, han sido utilizados para capturar estos patrones y prever cuándo es más probable que se produzcan brotes (35).

Uno de los patrones más evidentes en la propagación del Covid-19 fue el crecimiento exponencial de los casos, lo cual es una característica típica de las primeras etapas de un brote o epidemia, donde a medida que el virus se propaga sin restricciones, el número de casos se duplica en periodos cortos de tiempo, lo que da lugar a un aumento abrupto de contagios. El análisis de series de tiempo

permite identificar este tipo de crecimiento en sus primeras fases y tomar medidas tempranas para contener la propagación.

Durante la primera ola del Covid-19 en muchos países, los datos de series de tiempo mostraron una clara tendencia de crecimiento exponencial. A medida que el Covid-19 avanzaba, muchas regiones experimentaron olas recurrentes de contagio, un patrón en el que los casos aumentan y disminuyen a lo largo del tiempo. Estas olas suelen estar asociadas con factores como la relajación de las medidas de distanciamiento social, la aparición de nuevas variantes del virus o la estacionalidad. En este contexto, el análisis de series de tiempo permitió a los investigadores identificar estos ciclos y prever el surgimiento de nuevas olas de contagio, lo que ayudó a ajustar las intervenciones de salud pública.

Desde el inicio de la pandemia, se han planteado diversas metodologías para modelar el comportamiento del Covid-19. Entre los enfoques más comunes se encuentran los modelos econométricos basados en series de tiempo, los modelos Susceptible-Exposed-Infectious-Removed (SEIR) y técnicas de aprendizaje automático. También se ha sugerido el uso de modelos Autoregresivos Integrados de Medias Móviles (ARIMA) para predecir la evolución del número de contagios por Covid-19. Por ejemplo, Ibrahim y cols. (2020) (36) propusieron un modelo del tipo ARIMA de orden (1,1,0) para predecir la propagación del virus en Nigeria.

De forma similar, Talkhi y cols. (2021) (37) realizaron un análisis comparativo de varias técnicas de series de tiempo, concluyendo que el modelo más adecuado para predecir los casos confirmados en Irán fue el Perceptrón Multicapa (MLP), mientras que el modelo Holt-Winter fue más efectivo para prever los casos de fallecimientos por Covid-19.

Por su parte, Yonar y cols. (2020) (38) evaluaron la predicción del número de casos de Covid-19 utilizando dos enfoques: ARIMA y métodos de suavizado exponencial, concluyendo que, aunque no existe un modelo único para todos los países en estudio, ambos métodos resultan efectivos para describir las curvas de propagación del virus.

Como alternativa a los enfoques econométricos tradicionales, se han explorado métodos de aprendizaje automático (AA). Al respecto, Ghafouri-Fard y cols. (2021) (39) realizaron una exhaustiva revisión de trabajos que se enfocaron en la predicción de la tendencia de propagación del Covid-19. En esta revisión, destacaron el beneficio de usar el modelo Long Short Term Memory (LSTM) propuesto por Hochreiter y cols. (1997), el cual pertenece a la familia de las redes neuronales recurrentes.

Paralelamente, se ha evidenciado que el uso de SARIMAX puede ser de gran utilidad en este tipo de contextos. Al respecto, en el artículo "Predicting Time Series with SARIMAX" (40), se explica cómo utilizar el modelo SARIMAX para la predicción de series temporales, donde se destaca que esta extensión del

modelo ARIMA permite incorporar variables exógenas para predecir más efectivamente cuando existen factores externos que influyen en la serie temporal, haciéndolo más robusto que ARIMA en escenarios donde los cambios en la serie dependen de variables adicionales a su propia estructura histórica. Para utilizar SARIMAX, es necesario que la serie sea estacionaria, es decir, que no presente tendencias a lo largo del tiempo, por lo que se deben aplicar técnicas de diferenciación que eliminan las tendencias y patrones estacionales, lo que permite trabajar con datos más estables. Además, este modelo permite capturar patrones estacionales, lo que es clave en series temporales que repiten patrones en intervalos regulares, como días o meses.

En el artículo de Pelin Okutan (41), se realiza una comparativa entre varios modelos de predicción de series temporales, como SARIMAX, RNN, LSTM, Prophet y Transformer, donde en el análisis se enfatiza cómo SARIMAX puede capturar estacionalidades y factores externos de manera efectiva, lo que es beneficioso en contextos donde hay influencia de múltiples variables, sin embargo, se señala que este modelo tiene limitaciones cuando los datos tienen patrones más complejos o cuando se requiere manejar relaciones de larga duración.

Por otro lado, los modelos basados en redes neuronales, como RNN y LSTM, son apreciados por su capacidad para captar relaciones temporales de largo plazo, en especial el modelo LSTM, el cual destaca por su capacidad para

manejar dependencias complejas a lo largo del tiempo, sin embargo, este tipo de modelos requieren grandes cantidades de datos y un alto poder computacional, lo que puede complicar su uso en entornos con recursos limitados.

En este sentido, Prophet se presenta como una alternativa más sencilla para la predicción de series temporales con tendencia y estacionalidad marcada, siendo valorado por su facilidad de uso y su capacidad de ajustarse automáticamente a estos patrones sin requerir tanto esfuerzo en ajuste de parámetros. Aunque no alcanza la precisión de otros modelos más complejos como LSTM o Transformer, es recomendado en casos donde se necesita una implementación rápida y efectiva.

Finalmente, en este artículo Okutan resalta el uso de Transformers, los cuales originalmente fueron diseñados para el procesamiento del lenguaje natural, pero han demostrado ser muy eficaces en la predicción de series temporales. Estos modelos son capaces de manejar grandes volúmenes de datos y relaciones complejas sin necesidad de procesarlos de manera secuencial, lo que los convierte en una poderosa herramienta en escenarios donde los datos son voluminosos o intrincados. No obstante, su uso implica una curva de aprendizaje más pronunciada y una implementación técnica más compleja que otros enfoques.

Otro de los modelos que ha mostrado buenos resultados en el análisis de series de tiempo corresponde a DeepAR. Al respecto, un artículo llamado "DeepAR:

Probabilistic Forecasting with Autoregressive Recurrent Networks" (42), explica que en lugar de ajustar un modelo separado para cada serie temporal, DeepAR utiliza una arquitectura de red neuronal recurrente (RNN) para entrenar un modelo global que maneja escalas variadas y patrones de velocidad mediante el uso de una distribución probabilística (como la gaussiana o binomial negativa). Esto permite generar predicciones probabilísticas, mejorando la precisión frente a enfoques tradicionales de predicción para un solo artículo. La metodología de DeepAR es adecuada para problemas donde las series temporales están relacionadas y se necesita generar predicciones con incertidumbre cuantificable.

El modelo DeepAR se basa en la creación de predicciones paso a paso, donde se utiliza la última observación como entrada para predecir el siguiente paso de la serie temporal, generando una distribución de probabilidad para cada predicción. Este enfoque permite realizar múltiples predicciones (con la técnica Monte Carlo), lo que facilita el cálculo de intervalos de confianza y la evaluación de la incertidumbre en las predicciones.

Además de su capacidad para trabajar con covariables, DeepAR puede utilizar características adicionales, como el tiempo del año o promociones en el contexto de previsión de demanda. Estas características son valiosas cuando los valores de las covariables están disponibles tanto para el conjunto de entrenamiento como para el rango de predicción, mejorando la capacidad del modelo para hacer predicciones más precisas en función de la información adicional proporcionada.

El artículo "DeepAR Forecasting Algorithm" de Cory Maklin (43), detalla el uso del modelo DeepAR, una red recurrente basada en LSTM, para realizar predicciones probabilísticas en series temporales a gran escala. Este enfoque resulta particularmente útil en escenarios donde es necesario manejar múltiples series temporales simultáneamente, como en la predicción de la demanda de productos en diversas tiendas o ubicaciones.

Una de las ventajas clave de DeepAR es su capacidad para capturar relaciones complejas entre distintas series temporales, lo que puede mejorar el rendimiento frente a métodos más tradicionales como ARIMA o ETS. Al entrenarse en todas las series temporales del conjunto de datos de forma simultánea, el modelo es capaz de aprender patrones grupales dependientes. Otro aspecto importante es cómo DeepAR maneja problemas comunes en las series temporales, como la presencia de valores atípicos y valores faltantes, que son frecuentes en datos de demanda o epidemiológicos.

5. Modelamiento Predictivo con Hidden Markov Model

En el contexto de la pandemia del Covid-19 y otros brotes epidémicos, las técnicas avanzadas de modelamiento predictivo han cobrado una gran relevancia para anticipar la evolución de los contagios, prever fases latentes y optimizar la toma de decisiones en salud pública. A medida que las técnicas tradicionales de análisis y modelamiento, como los modelos de series de tiempo, se combinan con enfoques más complejos, los investigadores han podido crear modelos más

precisos que capturan mejor las dinámicas complejas de la transmisión de enfermedades infecciosas.

Estas técnicas avanzadas permiten identificar fases latentes en las pandemias, analizar patrones no lineales en los sistemas complejos y simular dinámicas basadas en ecuaciones diferenciales que representan la transmisión de virus en la población, por lo que se han aplicado en una variedad de situaciones durante la pandemia del Covid-19, ayudando a modelar brotes, fases de transmisión y prever la evolución del virus con mayor precisión.

Una de estas técnicas corresponde a los Modelos Ocultos de Markov (HMM, por sus siglas en inglés), los cuales son una técnica avanzada de modelamiento estadístico utilizada para representar sistemas que evolucionan con el tiempo y cuyas fases o estados no son directamente observables (ocultos), pero pueden inferirse a partir de datos observados. En el contexto de las epidemias, los HMM permiten modelar el desarrollo de una enfermedad infecciosa y predecir fases latentes de propagación, como el estado de transmisión del virus en la población o la aparición de nuevas olas de contagio.

Este tipo de modelos ha sido ampliamente utilizado en el estudio de enfermedades infecciosas, ya que ofrece un marco flexible para modelar los procesos epidémicos que siguen dinámicas estocásticas, es decir, que están influenciados por probabilidades y comportamientos inciertos (44).

Estos modelos son una técnica probabilística utilizada para modelar sistemas en los que el proceso observado depende de una secuencia oculta de estados, donde las observaciones visibles son el resultado de estados ocultos no observables directamente, pero que influyen en cómo se generan esas observaciones. El modelo asume que el sistema transita entre estos estados ocultos siguiendo ciertas probabilidades de transición, además, cada estado oculto genera observaciones según probabilidades de emisión específicas.

Aunque los "estados" del sistema, como las fases de crecimiento de la pandemia o los cambios en el comportamiento de los indicadores sociales, no son observables directamente, el modelo puede inferir estos estados ocultos a partir de las observaciones visibles, como la pendiente o aceleración suavizadas de las tasas de contagio.

Además, los HMM pueden abordar tres problemas clave, el problema de evaluación, calcula la probabilidad de que una secuencia observada haya sido generada por un modelo en particular, lo cual es útil para validar si un modelo es adecuado para describir los patrones observados; el problema de decodificación, busca la secuencia más probable de estados ocultos que podrían haber generado las observaciones visibles, permitiendo identificar fases ocultas o transiciones importantes en la evolución de la pandemia o de los factores socioeconómicos; y el problema de aprendizaje, ajusta las probabilidades de

transición y emisión para optimizar el modelo en función de los datos observados, lo que permite mejorar las predicciones (45).

En el artículo de Dhiman (46), se presenta el uso de los modelos Hidden Markov Model (HMM) para la predicción de series temporales, destacando su relevancia en contextos donde existen cambios de régimen o comportamientos volátiles, como en la predicción de ventas de productos con baja rotación. Aunque tradicionalmente los HMM se han utilizado para comprender la estructura interna de los datos secuenciales, el artículo muestra cómo también pueden ser una herramienta poderosa para realizar predicciones a futuro.

Un aspecto clave es la capacidad de los HMM para capturar dinámicas no lineales, lo que los hace adecuados para problemas en los que se otorga mayor importancia al estado actual, además permiten calcular la probabilidad de transición entre estados ocultos, lo que proporciona una capa adicional de interpretación sobre los cambios en el sistema.

En una investigación titulada “Application of Hidden Markov Models to Analyze, Group and Visualize Spatio-Temporal Covid-19 Data” (47) se evidenció que HMM permite modelar la incertidumbre en la dinámica de propagación del virus y las observaciones de los casos, proporcionando una herramienta efectiva para visualizar la propagación del Covid-19 por regiones.

En este estudio se emplearon secuencias de observación de vectores, como infecciones y muertes diarias, para caracterizar la gravedad del Covid-19

mediante modelos HMM. Los modelos HMM, aprendidos a partir de datos de varios países, incluidos Estados Unidos, varios países europeos y naciones con estrictas políticas de control, se utilizaron para estudiar la distribución de la propagación del Covid-19 en pequeñas regiones geográficas.

Los resultados obtenidos fueron consistentes con las observaciones en estos países. Además, se compararon los resultados de modelos multinacionales con aquellos obtenidos a partir de modelos de una sola nación, así como las secuencias de observación de vectores con secuencias de observación única, como infecciones o muertes diarias. Este análisis demostró la eficacia de utilizar datos multinacionales y vectores de observación en el conjunto de entrenamiento, lo que aportó nuevos conocimientos que no podrían haberse revelado mediante otros métodos o la simple inspección de datos brutos.

Este enfoque también permite hacer predicciones más precisas al integrar diversas modalidades de observación, como infecciones y muertes diarias, mejorando el rendimiento en comparación con otras estrategias de fusión de datos que se aplican después de la predicción. Dado lo anterior, los modelos HMM y las inferencias que se derivan de ellos pueden ser útiles tanto como herramienta de visualización como para proporcionar un sistema proactivo de apoyo a la toma de decisiones para los responsables de políticas públicas.

HIPÓTESIS

- La evolución de la pandemia en las comunas de la Región Metropolitana presenta tres estados ocultos principales (contagio leve, moderado y severo), los cuales pueden ser identificados mediante el uso de Hidden Markov Model (HMM).
- Los índices de movilidad interna y externa y las características sociodemográficas de las comunas tienen un efecto significativo en las probabilidades de transición entre los estados ocultos.

OBJETIVOS

Objetivo General

- Determinar los estados ocultos de la evolución de los contagios de Covid-19 en las comunas de la Región Metropolitana de Santiago y su relación con índices de movilidad y variables sociodemográficas utilizando Hidden Markov Model no Homogéneo.

Objetivos Específicos

- Aplicar Hidden Markov Model no homogéneo para identificar los estados ocultos asociados a los niveles de contagio de Covid-19.
- Describir y caracterizar los estados ocultos identificados en relación con los casos diarios de Covid-19, los índices de movilidad interna y externa, y las características sociodemográficas de las comunas.
- Estimar las probabilidades de transición entre los distintos estados ocultos y analizar su comportamiento en base a los índices de movilidad y las características sociodemográficas.
- Implementar modelos lineales mixtos para cuantificar el efecto de las variables de movilidad y las características sociodemográficas sobre las probabilidades de transición entre estados, considerando efectos aleatorios por comuna y tiempo.

DATOS Y METODOLOGÍA

1. Datos

Los datos utilizados corresponden a las tasas de casos acumulados y casos diarios de contagios por comuna, la movilidad interna y externa, las fases de restricción y las características sociodemográficas de 34 comunas de la Región Metropolitana en el primer año de la pandemia. Las variables que se incluyeron fueron las siguientes:

- 'IM_interno': Es el índice de movilidad interno de cada comuna, diseñado para medir el número de desplazamientos realizados al interior de los límites comunales. Este índice fue desarrollado por el Instituto de Data Science (IDS) de la Universidad del Desarrollo (UDD), en colaboración con Telefónica Chile y con el apoyo de Cisco. Su cálculo se basa en el análisis de datos de telecomunicaciones y desplazamientos, utilizando información anonimizada proveniente de las conexiones de los dispositivos móviles entre antenas.
- 'IM_externo': Representa un indicador de movilidad externa que mide el número de desplazamientos realizados fuera de los límites de cada comuna. Al igual que índice 'IM_interno', este fue desarrollado por el Instituto de Data Science de la Universidad del Desarrollo, en colaboración con Telefónica Chile y con el apoyo de Cisco.
- 'inc_pc_comuna': Ingreso per cápita de la comuna.

- 'Personas_hogar_promedio': Promedio de personas por hogar.
- 'inmigrantes': Número de inmigrantes en la comuna.
- 'total_hog_hacinados': Total de hogares con hacinamiento.
- 'indic_precariedad_vivienda': Índice de precariedad de la vivienda en la comuna.
- 'codigo': Código de la comuna.
- 'Comuna': Nombre de la comuna.
- 'fecha': Fecha del registro.
- 'dias': Días correlativo desde el inicio de los registros a analizar (considerado desde el 30 de marzo de 2020 al 30 de octubre de 2020).
- 'TasaCasosAcumulados': Tasa de contagio por cada 100.000 habitantes.
- 'CasosDiarios': Número de casos diarios, calculados como la diferencia entre los contagios acumulados de un día y el día anterior.
- 'Aceleracion': Velocidad de cambios en los casos diarios, obtenido de la segunda derivada de la Tasa de Casos Acumulados.
- 'IPS_2020': Índice de Prioridad Social por comuna. Indicador compuesto elaborado por el Ministerio de Desarrollo Social y Familia que integra aspectos relevantes del desarrollo social comunal, como ingresos, educación y salud.
- 'Prioridad_2020': Categorización del Índice de Prioridad Social del 2020 por comuna (Alta Prioridad, Media Alta Prioridad, Media Baja Prioridad, Baja Prioridad y Sin Prioridad).

- 'por_educ_sup': Porcentaje de población con educación superior, incluyendo niveles de educación profesional y postgrado. Esta información fue obtenida de las bases de datos del Censo 2017, proporcionadas por el Instituto Nacional de Estadísticas (INE).
- 'porc_postgrado': Porcentaje de la población que ha completado estudios de postgrado, basado en los datos recopilados del Censo 2017, proporcionados por el Instituto Nacional de Estadísticas (INE).
- 'paso': Corresponde a las etapas definidas del Plan Paso a Paso 2020, implementado por el Ministerio de Salud. Este plan estableció distintas fases (Cuarentena, Transición, Preparación y Apertura) para regular las medidas sanitarias en cada comuna. Según la fase asignada, se determinaban restricciones específicas, como los niveles de movilidad permitidos, el alcance de la población en cuarentena y las actividades autorizadas. Estas etapas sirvieron como un marco normativo para gestionar la respuesta sanitaria frente a la pandemia, adaptándose a la situación epidemiológica de cada territorio.
- 'inicio_cuarentena': Fecha en la que inició la cuarentena, de acuerdo con el Plan Paso a Paso 2020.
- 'termino_cuarentena': Fecha en la que terminó la cuarentena, de acuerdo con el Plan Paso a Paso 2020.

Por otro lado, se incorporaron datos de matrices origen-destino de viajes en transporte público, correspondientes a los sistemas de buses de Transantiago y

Metro, proporcionados por el Directorio de Transporte Público Metropolitano (DTPM). Estos datos fueron generados a partir de las bases de información de la tarjeta “Bip!” y los GPS de los buses, utilizando metodologías desarrolladas en colaboración entre la Universidad de Chile y el DTPM. Los datos disponibles permitieron analizar periodos representativos en distintos momentos de interés: prepandemia (5, 6, 7, 8, 10, 11 y 12 de agosto de 2019), al inicio de la pandemia (7, 8, 9, 10, 11, 12 y 13 de marzo de 2020) y durante la pandemia (9, 10, 11, 12, 13, 14 y 15 de noviembre de 2020).

2. Metodología

2.1. Exploración de los Datos

Se analizaron las series temporales de la tasa de casos acumulados y contagios diarios. Los casos diarios se calcularon como la diferencia entre los contagios acumulados de un día y el día anterior. Ambas métricas se suavizaron utilizando medias móviles de 7 días para reducir el ruido diario causado por variaciones en los reportes de casos. Lo anterior, permitió capturar no solo el crecimiento en los contagios, sino también los cambios dinámicos en su velocidad a lo largo del tiempo.

Posteriormente se evaluó la relación entre las tasas de casos acumulados y casos diarios con índices de movilidad internos y externos y factores sociodemográficos de las comunas mediante análisis de estadística descriptiva y

visualizaciones como gráficos de dispersión, heatmap de correlación de Pearson y mapas de visualización espacial para analizar cómo estas variables variaban entre comunas, proporcionando evidencia inicial sobre cómo las condiciones estructurales y socioeconómicas impactaron la propagación del virus.

Asimismo, se realizaron grafos para representar el flujo de viajes en la Región Metropolitana entre comunas en base a los datos de matriz Origen-Destino del DTMP, con el propósito de visualizar la dinámica de viajes, donde los nodos representan las comunas, el grosor de las aristas el tiempo de viaje promedio y el color (de azul a rojo) representa el número de viajes realizado.

2.2. Predicción Baseline

Se establecieron modelos baseline para predecir la tasa de contagios diarios suavizada en cada comuna, utilizando enfoques de Shallow Classifiers (Random Forest, XGBoost y Support Vector Machine) y un modelo basado en series temporales (DeepAR).

En cuanto a las covariables utilizadas en el entrenamiento, se incluyeron movilidad interna y externa, índice de prioridad social (IPS 2020), porcentaje de educación superior, y otros indicadores sociodemográficos. Luego, se descartaron variables con alta colinealidad usando el Variance Inflation Factor (VIF).

Se utilizó una proporción de entrenamiento y prueba del 80/20, preservando la secuencia temporal y se generaron ventanas deslizantes de tamaño 7, 15 y 30 días para modelar las dependencias temporales.

a) Shallow Classifiers:

- Random Forest: Captura relaciones no lineales mediante un ensamblado de árboles de decisión.
- XGBoost: Modelo de boosting que ajusta iterativamente árboles para reducir errores residuales.
- Support Vector Machine (SVM): Modelo que busca encontrar un hiperplano óptimo que separe los datos en un espacio de alta dimensionalidad.

b) DeepAR

- Modelo basado en redes neuronales recurrentes diseñado para series temporales, integra covariables exógenas como movilidad y características sociodemográficas para capturar patrones complejos y relaciones temporales.

Finalmente, se comparó la precisión de cada modelo mediante Errores Absolutos y Cuadrático Medio (MAE y RMSE).

2.3. Modelado de Estados Ocultos

Se aplicó Hidden Markov Model (HMM) no Homogéneo, el cual se basa en un enfoque probabilístico diseñado para modelar series temporales mediante una secuencia de estados no observables (estados ocultos), generando datos observables a partir de estos, siendo ideal para describir fenómenos dinámicos en los que los patrones latentes cambian a lo largo del tiempo, proporcionando una representación más precisa de su evolución.

La estructura principal de un HMM incluye:

- a) Estados Ocultos (S_t): Representan categorías o niveles no observables directamente en los datos. En este caso, representan distintas intensidades de contagio.

$S_t \in \{1, 2, 3, \dots, K\}$, donde K es el número de estados ocultos.

- b) Probabilidades de Transición ($P(S_t | S_{t-1})$): Describen la probabilidad de pasar de un estado oculto (S_{t-1}) a otro (S_t) en el tiempo t entre periodos consecutivos. Estas probabilidades definen la dinámica del sistema subyacente, en el caso de un modelo no homogéneo, dependen de covariables X_t :

$$P(S_t | S_{t-1}, X_t) = \pi_{ij}(X_t), \text{ para } i, j \in \{1, 2, 3, \dots, K\}$$

Aquí, $\pi_{ij}(X_t)$ describe cómo las covariables (como movilidad interna y externa) afectan a la probabilidad de transición entre los estados i y j .

- c) Distribución de Emisión $P(O_t | S_t)$: Define la relación entre los estados ocultos y las observaciones O_t . Esta distribución modela cómo los estados generan las observaciones, en este caso, las tasas de contagios diarios. Si asumimos una distribución normal para las emisiones:

$$O_t | S_t = k \sim \mathcal{N}(u_t, \sigma_k^2), \text{ para } k \in \{1, 2, 3, \dots, K\}$$

Se utilizó un HMM porque permite representar explícitamente la evolución de los contagios como una transición entre estados latentes, capturando los patrones subyacentes en la pandemia, donde los comportamientos de contagio son altamente dinámicos.

En este caso se optó por un modelo no homogéneo debido a que las probabilidades de transición entre niveles de contagio no son constantes y están influenciadas por factores externos, como la movilidad y características sociodemográficas (a diferencia del homogéneo, donde las probabilidades de transición son constantes en el tiempo). El modelo no homogéneo ajusta las probabilidades de transición en función de estas covariables, ofreciendo una comprensión de los mecanismos que impulsan la evolución del contagio.

El modelo fue implementado utilizando la biblioteca depmixS4 en R, debido a su capacidad para especificar probabilidades de transición dependientes de covariables, además usa métodos avanzados como el algoritmo Expectation-Maximization (EM), para estimar los parámetros que maximizan la probabilidad de los datos observados ($P(O|\theta)$, donde θ son los parámetros del modelo). Se

utilizaron como covariables de transición la movilidad interna y externa, y se establecieron como emisiones los casos de contagios diarios con media móvil de 7 días.

Una vez implementado el modelo, con el fin de determinar el número adecuado de estados ocultos (K), se evaluaron varios modelos con un rango de 3 a 10 estados ocultos, priorizando la interpretabilidad y simplicidad de los resultados, para capturar información relevante a lo largo de la evolución de la pandemia.

Los criterios utilizados en la evaluación del número de estados ocultos fueron los siguientes:

- a) Log-Likelihood (ℓ): Mide el ajuste del modelo a los datos observados, con valores más altos de (ℓ) indicando un mejor ajuste.

$$\ell = \sum_{t=1}^T \log P(O_t | \theta)$$

- b) Akaike Information Criterion (AIC): Penaliza la complejidad del modelo para evitar el sobreajuste.

$$AIC = -2\ell + 2p, \text{ donde } p \text{ es el número de parámetros del modelo.}$$

- c) Bayesian Information Criterion (BIC): Similar al AIC, pero con una penalización más fuerte para modelos complejos.

$$BIC = -2\ell + p \log(N), \text{ donde } N \text{ es el tamaño de la muestra.}$$

2.4. Análisis de Secuencias de Estados

El algoritmo de Viterbi es una técnica de programación dinámica que permite encontrar la secuencia más probable de estados ocultos $(S_1, S_2, S_3, \dots, S_T)$ dado los datos observados $(O_1, O_2, O_3, \dots, O_T)$ y un modelo Hidden Markov Model con parámetros conocidos, permitiendo en este caso, inferir los estados ocultos que mejor explican la evolución de los contagios en cada comuna a lo largo del tiempo.

Dado un conjunto de probabilidades de transición $P(S_t | S_{t-1})$, de emisión $P(O_t | S_t)$ y un modelo inicial $P(S_1)$, el algoritmo maximiza la probabilidad conjunta:

$$P(S_1, S_2, \dots, S_T | O_1, O_2, \dots, O_T)$$

Esto se logra mediante un proceso iterativo que calcula las probabilidades acumuladas de las rutas más probables hasta cada estado en cada paso del tiempo.

Después de calcular las probabilidades acumuladas para todos los estados en todos los momentos (t), el algoritmo realiza un proceso llamado “Backtracking”, retrocediendo para reconstruir la secuencia óptima de estados ocultos.

En este estudio, el algoritmo de Viterbi fue implementado para cada comuna, obteniendo la secuencia más probable de estados a lo largo del periodo de estudio. Esto permitió identificar momentos críticos en la evolución de los

contagios y observar cómo la pandemia evolucionó en diferentes comunas y en distintos momentos, ofreciendo una visión clara de las fases críticas de contagio.

2.5. Caracterización de Estados Ocultos

Posterior al ajuste del modelo Hidden Markov Model y a la identificación de las secuencias más probables de estados a través del algoritmo de Viterbi, se procedió a la caracterización de los estados ocultos. Este análisis se enfocó en describir y comparar las condiciones que definen cada estado en términos de variables sociodemográficas, movilidad y epidemiológicas. Lo que permitió entender cómo las comunas experimentaron la pandemia de manera diferenciada y cómo sus características influyeron la propagación del virus.

El enfoque metodológico adoptado para la caracterización de los estados ocultos fue el siguiente:

- a) Análisis descriptivo por estado oculto: Se realizó un análisis descriptivo de las variables relacionadas con el comportamiento de los contagios, índices de movilidad y características sociodemográficas, incluyendo medidas de tendencia central y dispersión para cada uno de los estados ocultos.
- b) Análisis de distribución de la cantidad de días por estado para cada comuna: Debido a que las comunas transitaron por los tres estados ocultos de forma heterogénea, se llevó a cabo un análisis de la distribución temporal, contabilizando la cantidad de días que cada comuna permaneció

en los estados 1, 2 y 3, facilitando la identificación de variaciones temporales y espaciales.

- c) Relación entre movilidad y contagios: Se analizaron las correlaciones entre los casos diarios de COVID-19 y los índices de movilidad interna y externa en cada estado oculto. Además, se incluyó información contextual, como índice de precariedad de la vivienda e ingreso per cápita, permitiendo identificar patrones diferenciados entre comunas.
- d) Relación entre características sociodemográficas y estados ocultos: Para cada estado, se evaluaron asociaciones con variables sociodemográficas.
- e) Asignación de denominaciones a los estados ocultos: Basado en los resultados obtenidos, los estados ocultos fueron clasificados de manera interpretativa para facilitar su comprensión, reflejando un gradiente progresivo en la intensidad de los contagios.

2.6. Análisis de Probabilidades de Transición

El análisis de las probabilidades de transición entre estados ocultos, obtenido a partir del modelo Hidden Markov Model No Homogéneo, permitió identificar dinámicas relevantes en el comportamiento de la pandemia y explorar cómo variables explicativas como movilidad interna y externa o características sociodemográficas influenciaron los cambios entre estados.

El enfoque metodológico utilizado fue el siguiente:

- a) Definición y Obtención de Probabilidades de Transición: Se analizaron las nueve transiciones posibles entre los tres estados ocultos, clasificándolos en tres categorías: Permanencia en un mismo estado ($1 \rightarrow 1$, $2 \rightarrow 2$, $3 \rightarrow 3$), Transiciones hacia estados de mayor severidad ($1 \rightarrow 2$, $1 \rightarrow 3$, $2 \rightarrow 3$) y Transiciones hacia estados de menor severidad ($3 \rightarrow 2$, $3 \rightarrow 1$, $2 \rightarrow 1$). Luego a partir del modelo HMM ajustado, se extrajeron las probabilidades de transición para cada intervalo temporal en cada comuna, reflejando la probabilidad condicional de que una comuna transite de un estado oculto a otro en un periodo dado.
- b) Análisis de correlaciones con variables dinámicas: Se examinó la relación entre las probabilidades de transición y las variables de movilidad interna y externa, para evaluar cómo la actividad dentro y hacia otras comunas afecta la dinámica de contagio influyendo en el avance o retroceso entre estados ocultos.
- c) Análisis de correlaciones con covariables sociodemográficas: Las probabilidades de transición se correlacionaron con covariables estructurales, es decir, características comunales como ingreso per cápita, promedio de personas por hogar, porcentaje de educación superior o índice de prioridad social. Dado que estas covariables no varían en el tiempo de estudio, las correlaciones reflejan diferencias estructurales entre comunas en lugar de dinámicas temporales, por lo tanto, estas covariables

se analizaron como factores contextuales que condicionan las transiciones.

2.7. Modelos Lineales Mixtos

En este estudio, se aplicaron Modelos Lineales Mixtos (LMM) para modelar las probabilidades de transición entre estados ocultos de contagio identificados previamente a través del modelo Hidden Markov Model (HMM) no homogéneo. Los LMM permiten capturar tanto el efecto promedio de variables explicativas (efectos fijos) como las variaciones específicas atribuibles a unidades jerárquicas, como comunas y periodos temporales (efectos aleatorios). Esto resulta esencial para abordar la naturaleza espaciotemporal de los datos analizados.

2.7.1. Selección de Covariables mediante Elastic Net

Se utilizó el método de Elastic Net para realizar la selección de covariables relevantes. Elastic Net es un modelo de regresión regularizado que combina penalizaciones ℓ_1 (Lasso) y ℓ_2 (Ridge), lo que permite manejar la multicolinealidad entre variables y seleccionar aquellas con mayor impacto explicativo en la variable objetivo. Las covariables iniciales consideradas fueron los índices de movilidad interno y externo, y las variables sociodemográficas: ingreso per capita, promedio de personas por hogar, inmigrantes, número total de hogares

hacinados, indicador de precariedad de la vivienda, porcentaje de educación superior e índice de prioridad social (IPS 2020).

La métrica de desempeño utilizada como criterio de evaluación fue el Error Cuadrático Medio (RMSE), y las covariables seleccionadas mediante Elastic Net se utilizaron como covariables para los Modelos Lineales Mixtos (LMM), asegurando que solo las variables con relevancia significativa sean incluidas, mejorando la robustez de los resultados.

2.7.2. Especificaciones de los Modelos Lineales Mixtos

Un modelo lineal mixto se compone de efectos fijos y efectos aleatorios. Los efectos fijos representan el impacto promedio de las covariables sobre la variable de respuesta (probabilidad de transición), por ejemplo, la movilidad interna o externa, el ingreso per cápita y las características sociodemográficas (hacinamiento, nivel de educación, entre otros). Los efectos aleatorios capturan la variabilidad específica no explicada entre subunidades. En este caso, se incluyeron efectos aleatorios para comunas (para reflejar las diferencias geográficas en las probabilidades de transición) y días (para modelar las variaciones temporales, incluyendo efectos específicos asociados a días laborales, fines de semana y feriados).

La ecuación general de un LMM utilizado en este estudio puede expresarse como:

$$P(y_{ij}) = \beta_0 + \beta_1 X_{ij} + u_j + u_t + \epsilon_{ij}$$

Donde:

- $P(y_{ij})$ es la probabilidad estimada de transición para la observación i en la comuna j y el día t .
- $\beta_0 + \beta_1 X_{ij}$ representa los efectos fijos, donde X_{ij} incluye variables como movilidad y características socioeconómicas.
- u_j y u_t son los efectos aleatorios asociados a la comuna y al tiempo, respectivamente.
- ϵ_{ij} es el término de error residual.

2.7.3. Modelamiento de las Probabilidades de Transición mediante Linear Mixed Models

Se construyeron cuatro variantes de LMM para modelar cada una de las posibles transiciones entre estados ocultos. La elección de estos modelos buscó descomponer la variabilidad en componentes espaciales y temporales, permitiendo un análisis más detallado. La descripción de los modelos se detalla a continuación.

- Modelo A: Incluye efectos aleatorios por cada comuna para capturar la variabilidad geográfica en la transición.

$$P(\text{Transición}) \sim (1|\text{Comuna})$$

- Modelo B: Incorpora efectos aleatorios por comuna y día, permitiendo modelar tanto la variación geográfica como la temporal.

$$P(\text{Transición}) \sim (1|\text{Comuna}) + (1|\text{Tiempo})$$

- Modelo C: Adiciona efectos fijos relacionados con indicadores de movilidad (movilidad interna y externa) a los efectos aleatorios, permitiendo evaluar el impacto de la movilidad en las probabilidades de transición.

$$P(\text{Transición}) \sim IM_{\text{Externo}} + IM_{\text{Interno}} + (1|\text{Comuna}) + (1|\text{Tiempo})$$

- Modelo D: Integra efectos fijos relacionados con la movilidad e indicadores sociodemográficos (como el promedio de personas por hogar, el ingreso per cápita y el nivel educativo) junto con efectos aleatorios, ofreciendo una visión más completa de las dinámicas espacio temporales.

$$P(\text{Transición}) \sim IM_{\text{Externo}} + IM_{\text{Interno}} + \text{Covariables sociodemográficas}_{\text{ElasticNet}} + (1|\text{Comuna}) + (1|\text{Tiempo})$$

Los LMM se utilizaron para analizar las probabilidades de transición entre estados ocultos, desglosados en:

- Mantenimiento de estados: Como el estado leve (1→1) o severo (3→3).

- Transiciones entre estados: Por ejemplo, leve a moderado (1→2), leve a severo (1→3) y moderado a severo (2→3).

Los efectos aleatorios por comuna y día capturan diferencias estructurales en los datos no explicados completamente por las covariables. Esto incluye diferencias socioeconómicas entre comunas y patrones específicos asociados a días de la semana, fines de semana y días festivos. Por otra parte, al incluir efectos fijos, los LMM permiten identificar y cuantificar el impacto de variables explicativas como la movilidad y características sociodemográficas, en las probabilidades de transición.

Para comparar los modelos y determinar qué modelo (A, B, C y D) ofrecía el mejor ajuste para cada transición, se evaluaron métricas como el AIC y el BIC. Esto permitió identificar la configuración más adecuada para capturar la dinámica de los contagios en cada caso.

Posteriormente, para los modelos C y D se realizó análisis de R^2 marginal y condicional, con el fin de determinar la varianza explicada únicamente por los efectos fijos, y la varianza explicada por los efectos fijos y aleatorios.

El R^2 marginal representa la proporción de la varianza explicada únicamente por los efectos fijos del modelo, por ende, en el contexto de estudio, este análisis permite evaluar qué tan bien las covariables explicativas (como movilidad y características sociodemográficas) capturan las probabilidades de transición entre estados.

Por otra parte, el R^2 condicional representa la proporción de la varianza explicada por el modelo completo, incluyendo tanto los efectos fijos como los efectos aleatorios, lo que proporciona una visión integral de cómo las variaciones espaciotemporales (capturadas por los efectos aleatorios) contribuyen al ajuste del modelo.

2.8. Análisis Complementarios

Para los modelos que obtuvieron mejores resultados para cada transición, se realizó análisis de efectos aleatorios y análisis de residuales.

Mediante el análisis de efectos aleatorios, se evaluó cómo las diferencias entre comunas y días contribuyen a la variabilidad de las probabilidades de transición. Además, se analizaron los interceptos específicos por comuna y sus posibles patrones espaciales y, por otra parte, los efectos aleatorios por día permitieron identificar variaciones temporales, como el impacto de días festivos o fines de semana en las transiciones.

Finalmente, se analizaron los residuales del modelo, para verificar si las variaciones estaban adecuadamente capturadas por los efectos aleatorios.

Este análisis fue relevante para verificar la validez del modelo y para identificar posibles áreas de mejora o factores no considerados explícitamente en el modelo.

RESULTADOS

Este estudio se enfoca en el análisis de los datos de la pandemia por Covid-19 en la Región Metropolitana durante el primer año de la pandemia, desde marzo a octubre del año 2020. Los datos abarcan una variedad de aspectos clave relacionados con la movilidad, los casos de Covid-19, las condiciones de vivienda y factores sociodemográficos. Entre las variables principales analizadas se encuentran los casos diarios y tasas diarias acumuladas, los índices de movilidad interna y externa, además de indicadores sociodemográficos como el indicador de precariedad de vivienda, ingreso per cápita, personas hogar promedio, total de hogares con hacinamiento, cantidad de inmigrantes, porcentaje de población con educación superior, y el índice de prioridad social (IPS) de cada comuna.

1. Análisis Descriptivo Exploratorio

El análisis descriptivo exploratorio (EDA) fue la primera etapa del estudio y permitió obtener una comprensión inicial de la distribución, tendencias y correlaciones entre las diferentes variables, lo que resulta fundamental para identificar patrones importantes, como las diferencias entre comunas en términos de contagios y movilidad, así como los factores sociales y económicos que pueden influir en la propagación del virus.

Con respecto a la evolución de las tasas de casos acumulados de Covid-10 por cada 100.000 habitantes en la Región Metropolitana durante el año 2020 (Figura

1), se puede observar una tendencia creciente constante desde el inicio de la pandemia en marzo, con un ascenso más pronunciado durante los meses de mayo y junio, lo que coincide con el período de mayor propagación del virus, en el que se implementaron medidas de cuarentena más estrictas en diversas comunas.

A partir de julio, se aprecia una disminución en la pendiente de la curva, lo que sugiere un control relativo de la tasa de contagios debido, probablemente, al efecto acumulado de las políticas de confinamiento, la reducción de la movilidad y la adopción de medidas preventivas por parte de la población. Luego se observa que los contagios continúan incrementando pero a un ritmo más lento, evidenciando el comportamiento natural de las pandemias.

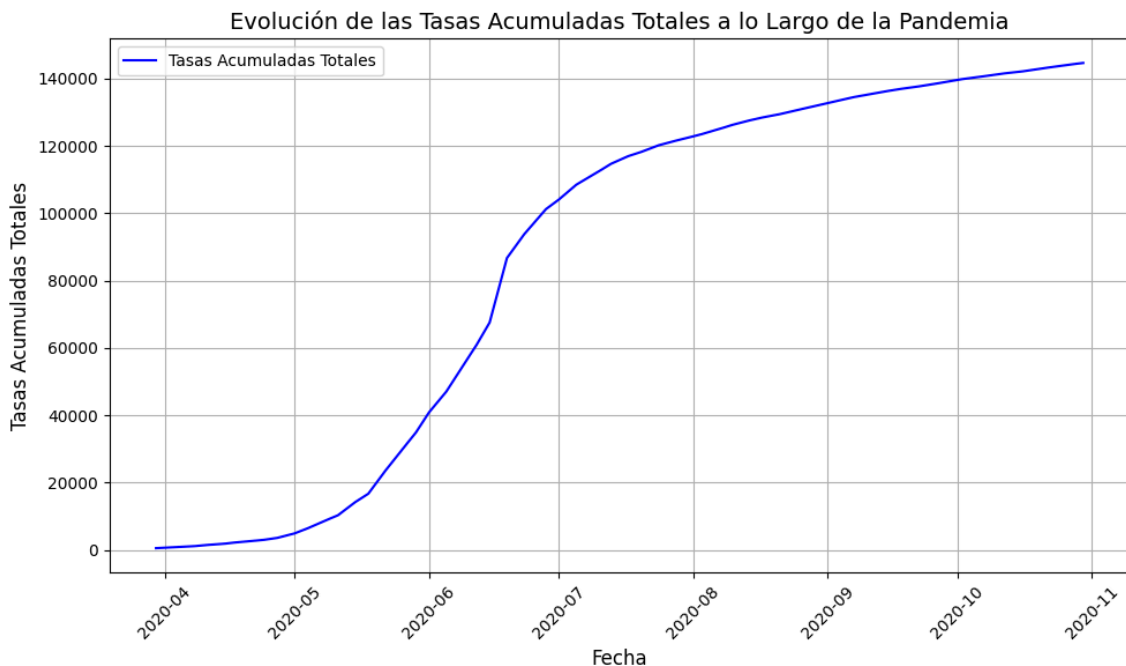


Figura 1. Evolución de la Tasa de Casos Acumulados de Covid-19 durante el año 2020.

En cuanto a la evolución de la tasa de contagios acumulados por cada 100.000 habitantes en 34 comunas centrales de la Región Metropolitana (Figura 2), se puede observar que la pandemia afectó de manera desigual a las distintas comunas. Durante los primeros meses, se observa un ascenso acelerado en las tasas de contagio, especialmente a partir de mayo, periodo en que la RM experimentó un rápido incremento de casos, lo que llevó a la implementación de medidas de cuarentena más estrictas en varias comunas. Aunque todas las comunas presentaron un patrón de crecimiento general en las tasas de contagio, la intensidad y velocidad de propagación fue diferente entre ellas.

Comunas como La Granja, San Ramón, La Pintana y Renca mostraron un crecimiento más acelerado de la tasa de contagios acumulados, mientras que comunas como Vitacura, Lo Barnechea y Las Condes mostraron un crecimiento más lento y constante en comparación con las otras comunas del sector sur de la región.

A partir de julio, se aprecia una desaceleración en el crecimiento de las tasas de contagio en la mayoría de las comunas, lo que sugiere que las medidas de control comenzaron a ser efectivas en la contención de la pandemia.

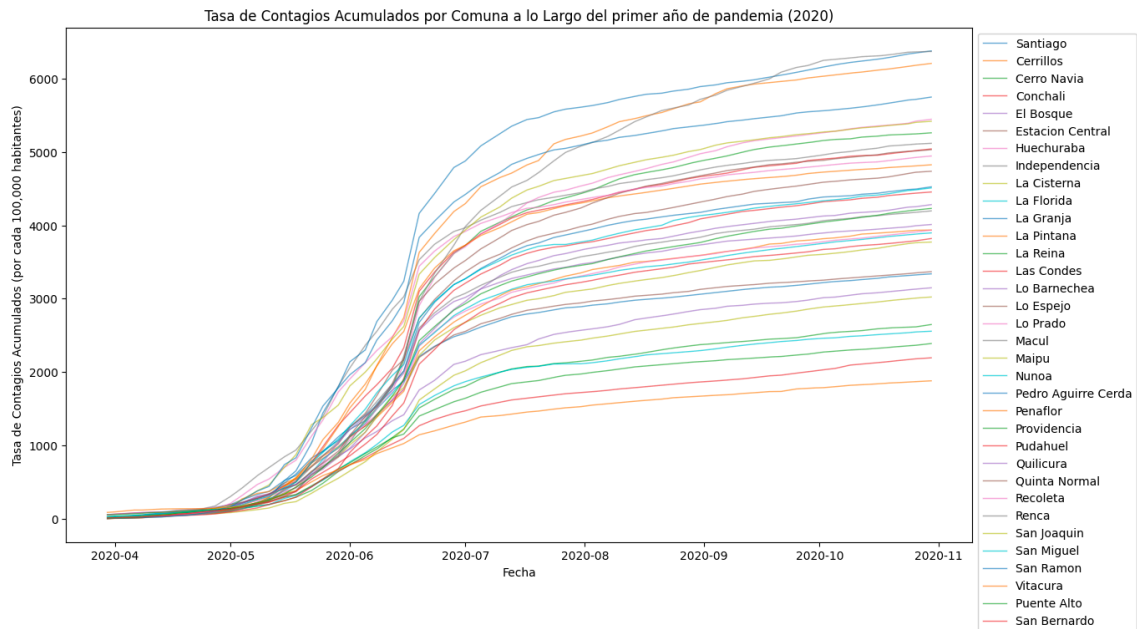


Figura 2. Evolución de la Tasa de Casos Acumulados de Covid-19 por Comuna durante el año 2020.

Al analizar la distribución de la tasa de casos acumulada por comuna (Figura 3), se observan las diferencias en la propagación del Covid-19 entre las distintas comunas. Las comunas con mayor tasa acumulada de contagios fueron La Granja, San Ramón, La Pintana, y Renca, todas ubicadas en el sector sur de la Región Metropolitana, las cuales superaron los 5.000 a 6.000 casos acumulados por cada 100.000 habitantes, evidenciando un mayor impacto de la pandemia en estas zonas, lo que podría estar relacionado con factores como mayor densidad poblacional, condiciones de hacinamiento y menor acceso a recursos de salud, lo que contribuyó a la rápida propagación del virus.

En contraste, comunas como Vitacura, Las Condes y Providencia presentaron las tasas más bajas, con valores que no superaron los 3.000 casos acumulados

por cada 100.000 habitantes. Estas comunas tienen un mayor nivel socioeconómico, menor densidad poblacional y mejores condiciones de vivienda, lo que pudo haber favorecido un mejor control de la propagación del virus.

Con respecto a la dispersión de las tasas, se observa que existe cierta variabilidad en la tasa de contagios acumulados incluso dentro de comunas que podrían parecer similares en términos de características sociodemográficas.

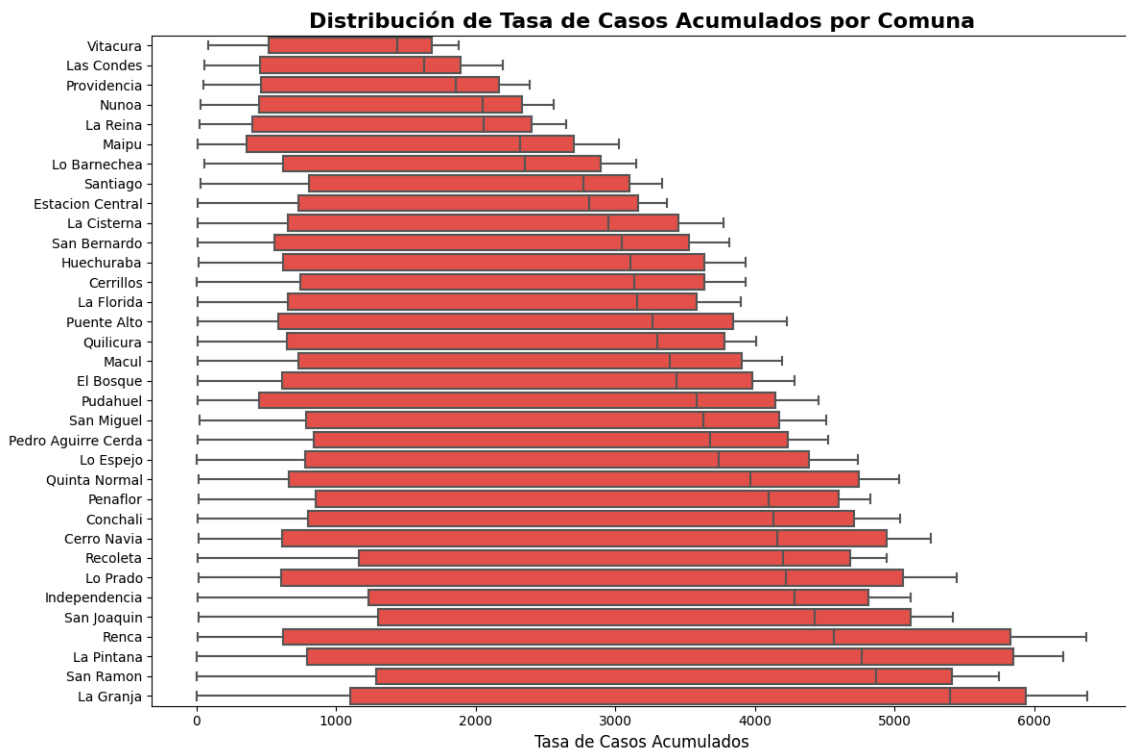


Figura 3. Distribución de la Tasas de Casos Acumulados de Covid-19 por Comuna durante el año 2020.

Al visualizar la evolución mensual de las tasas de casos acumulados por cada comuna a través de un heat map (Figura 4), se puede observar que hay un aumento notable de las tasas durante los meses de junio y julio, lo que es

esperable debido al comportamiento normal de una pandemia pero también se puede haber visto exacerbado producto del clima, ya que en los meses de invierno las enfermedades respiratorias suelen mostrar un comportamiento similar.

Al igual que lo descrito anteriormente, aquí también se puede observar que comunas como La Granja, La Pintana, y Renca mostraron los niveles más altos de contagio durante estos meses (rojos más oscuros), versus comunas como Vitacura, Las Condes, y Lo Barnechea (tonos más claros), que presentaron tasas relativamente más bajas en comparación con comunas del sector sur.

A partir de agosto y septiembre, se observa una disminución en la intensidad de los contagios en muchas comunas, lo que coincide con la relajación de las medidas de cuarentena y la llegada de climas más cálidos. Sin embargo, algunas comunas como San Ramón y La Pintana siguen mostrando altas tasas de contagio acumulado, lo que indica que la recuperación no fue uniforme en todas las áreas.

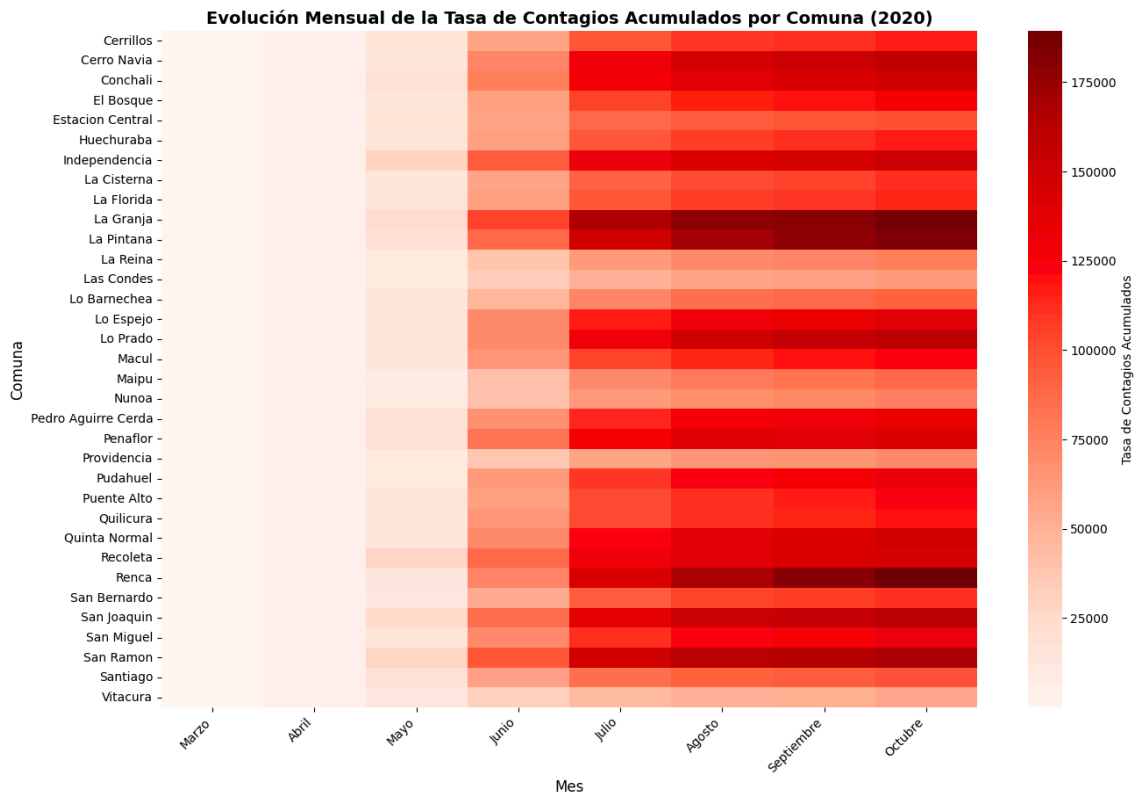


Figura 4. Evolución Mensual de la Tasa de Casos Acumulados de Covid-19 por Comuna durante el año 2020.

Con respecto al total de inmigrantes en las 34 comunas analizadas (Figura 5), que corresponde a 428.737 personas, un 25,8% reside en la Comuna de Santiago (101.389 personas), seguido de un 7,1% en Las Condes, un 7% en Independencia, un 5,4% en Recoleta y un 5,2% en Estación Central, el resto de las comunas fluctúan con porcentajes entre 0,5% y 3,6%, siendo las comunas de La Pintana y San Ramón las que presentan los valores más bajos.

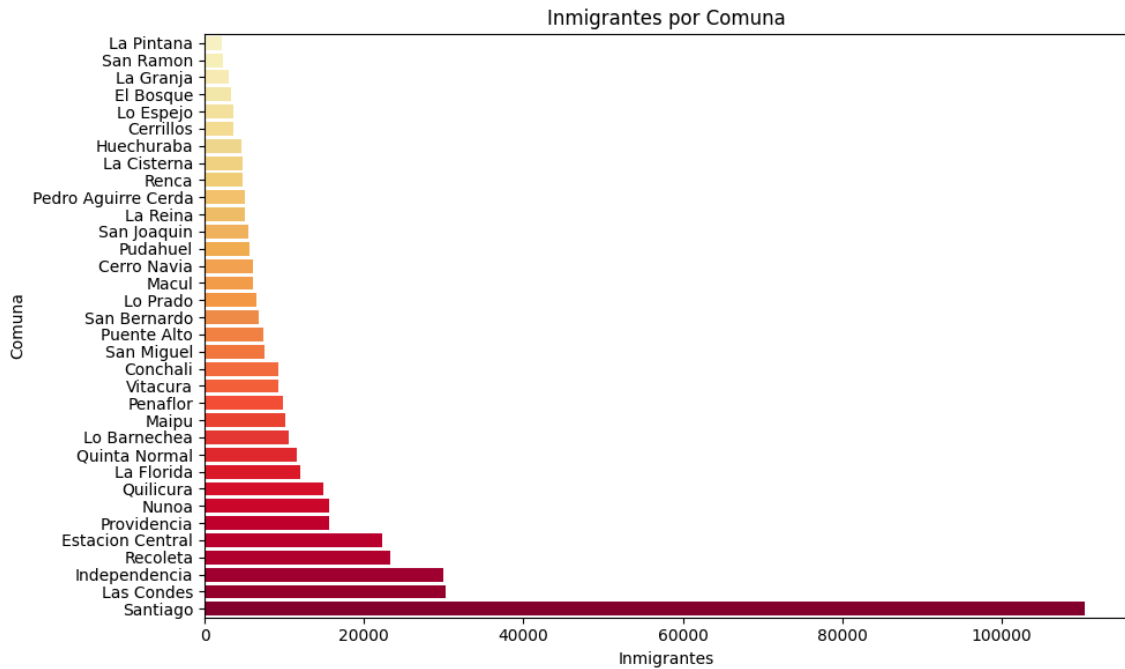


Figura 5. Cantidad de Inmigrantes por cada Comuna de la Región Metropolitana.

En cuanto al promedio de personas por hogar en las distintas comunas analizadas (Figura 6), se puede evidenciar que las comunas que poseen el promedio más bajo corresponden a Providencia y Santiago, con 1,7 y 1,9 personas por hogar respectivamente. Al contrario, las comunas que presentan un promedio mayor corresponden a Lo Barnechea, con 3,6 personas por hogar, seguido de Huechuraba, Lo Espejo, La Pintana, Quilicura, Cerro Navia, Peñaflor y Renca, todas con 3,4 personas por hogar. El resto de las comunas se encuentran con valores promedio entre 2,2 y 3,3 personas por hogar.

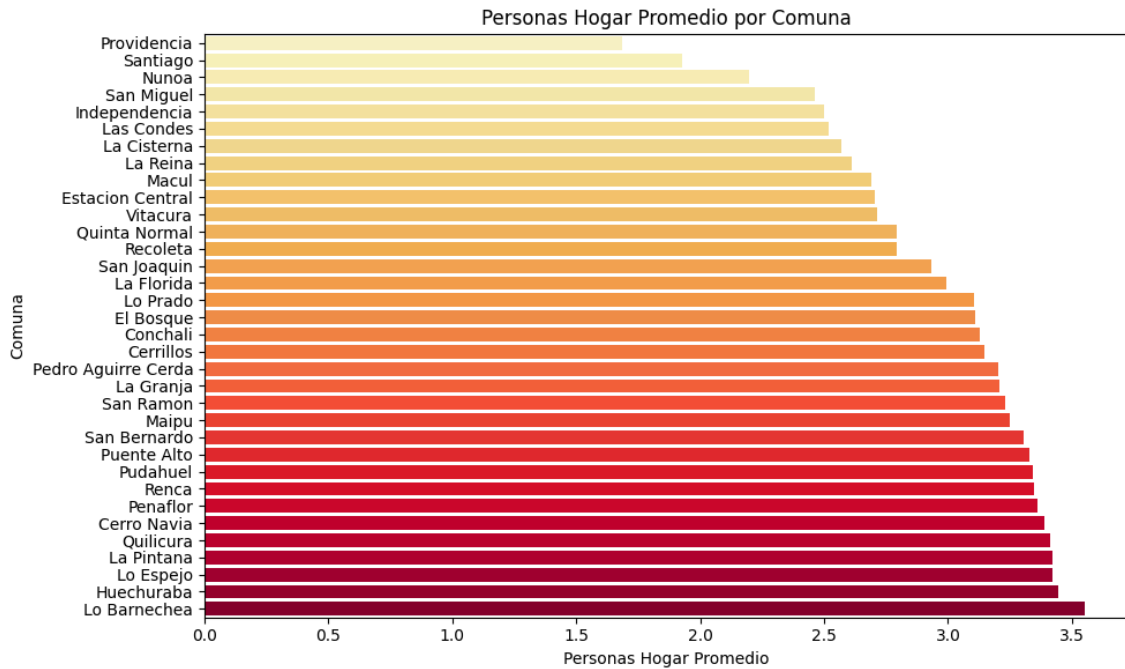


Figura 6. Promedio de Personas por Hogar en las Comunas de la Región Metropolitana.

Por otra parte, al analizar la cantidad de hogares por comuna que presentaron hacinamiento (Figura 7), se evidenció que las comunas que presentaron una mayor cantidad de hogares con hacinamiento fueron Puente Alto (376.908), San Bernardo (308.396), Maipú (265.835) y La Pintana (204.312), mientras que las comunas que presentaron valores menores fueron Vitacura (4.770), Providencia (15.738) y Lo Barnechea (24.911). El resto de las comunas se encuentra con valores entre 33.000 y 190.000 hogares hacinados.

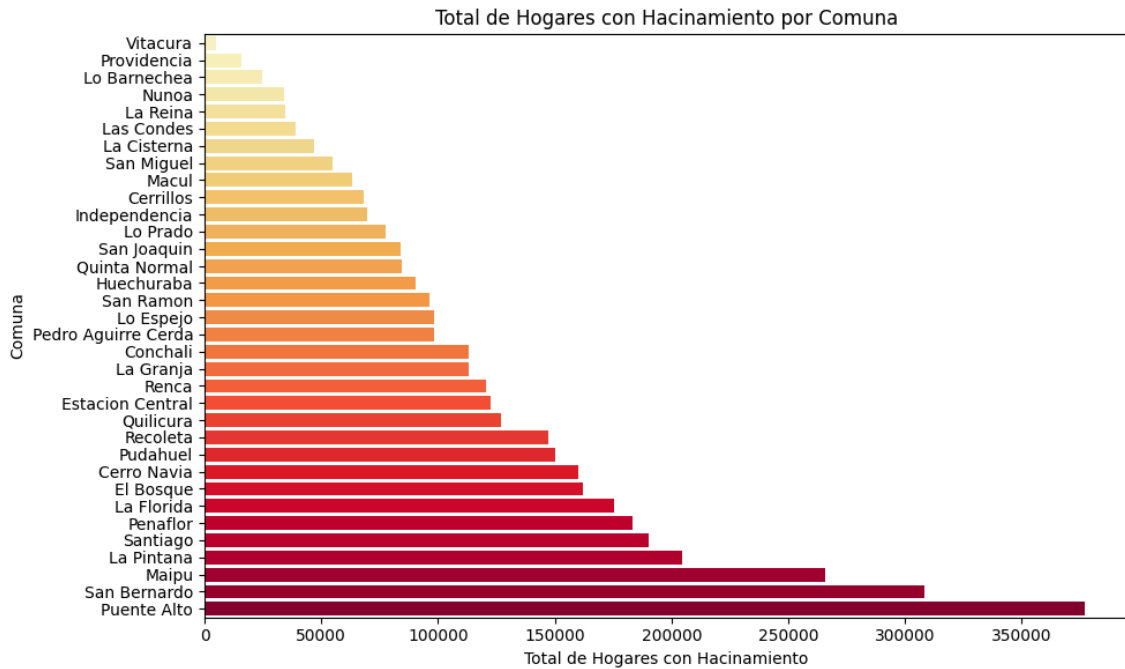


Figura 7. Total de Hogares con Hacinamiento por cada Comuna de la Región Metropolitana.

Al analizar el ingreso per cápita por comuna (Figura 8), se evidenció que las comunas que presentaron mayor ingreso fueron Vitacura con 1.659.485 pesos, seguido de Providencia con 1.300.158 pesos y Las Condes con 1.280.186 pesos. Luego vienen las comunas de La Reina, Ñuñoa, Lo Barnechea y Santiago, con valores entre 581.000 y 917.000 pesos. El resto de las comunas posee ingresos per cápita menores a 500.000 pesos, siendo La Pintana, Pedro Aguirre Cerda, San Bernardo y Lo Espejo, las que presentaron los valores más bajos, entre 211.800 y 249.700 pesos.

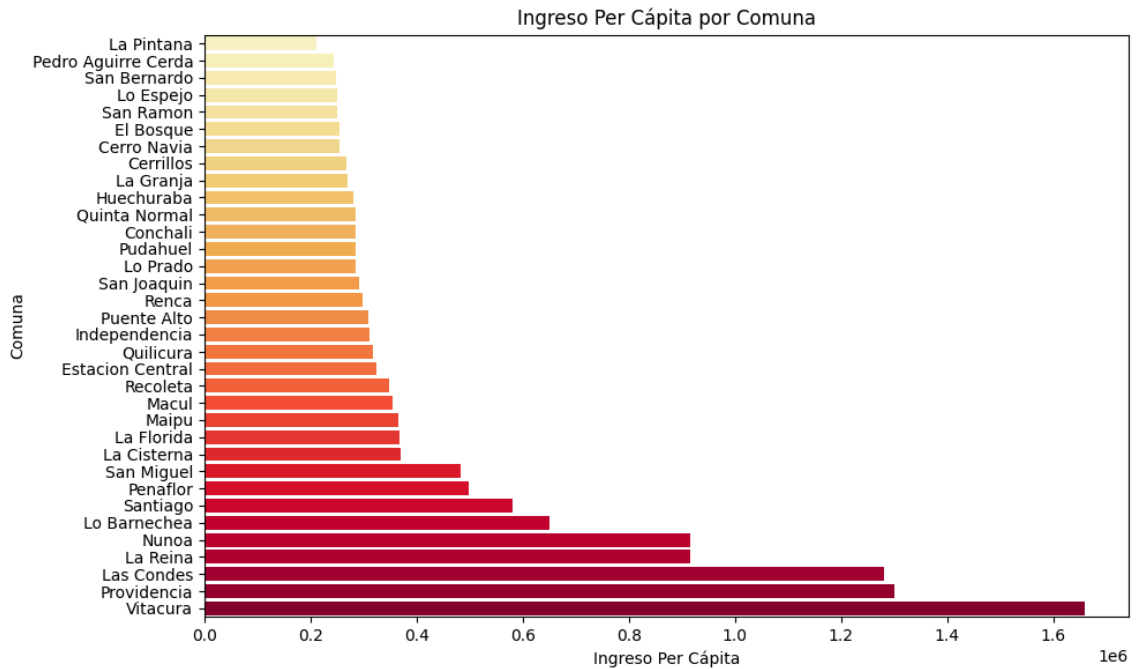


Figura 8. Ingreso Per Cápita por Comuna en la Región Metropolitana.

En cuanto al índice de precariedad por vivienda se observó que las comunas que presentaron valores más bajos fueron Vitacura (0,087), Providencia (0,210), Ñuñoa (0,268) y Las Condes (0,427); por el contrario, las comunas con valores de precariedad más alto fueron La Granja (5,048), San Ramón (4,939), Huechuraba (4,737) y Cerro Navia (4,679) (Figura 9).

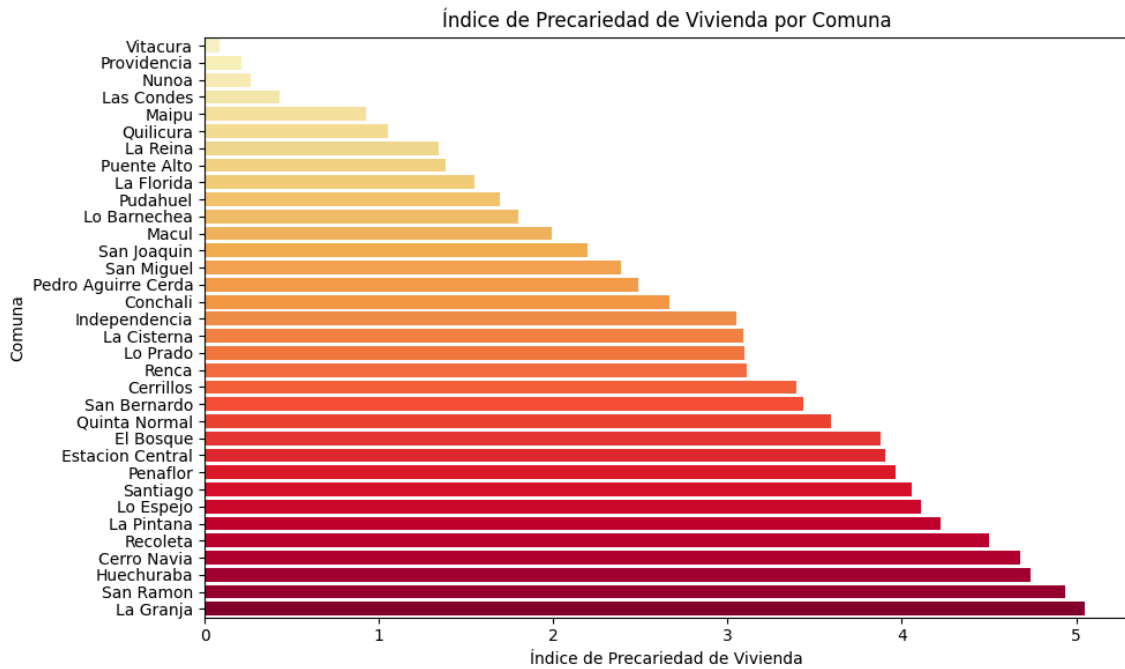


Figura 9. Índice de Precariedad de la Vivienda en distintas comunas de la Región Metropolitana.

Con respecto al índice de prioridad social por comuna, se puede observar que las comunas que presentaron menores porcentajes de prioridad social fueron Vitacura (7,9%), Las Condes (11,6%) y Providencia (24,9%) y las que presentaron mayores porcentajes de prioridad social fueron La Pintana (89,3%), Lo Espejo (88,8%), Cerro Navia (85,9%), San Ramón (83,5%) y El Bosque (81%). Al categorizar la prioridad social considerando las categorías de sin prioridad, baja, media baja, media alta y alta, se observa más detenidamente el comportamiento por comuna (Figura 10 y 11).

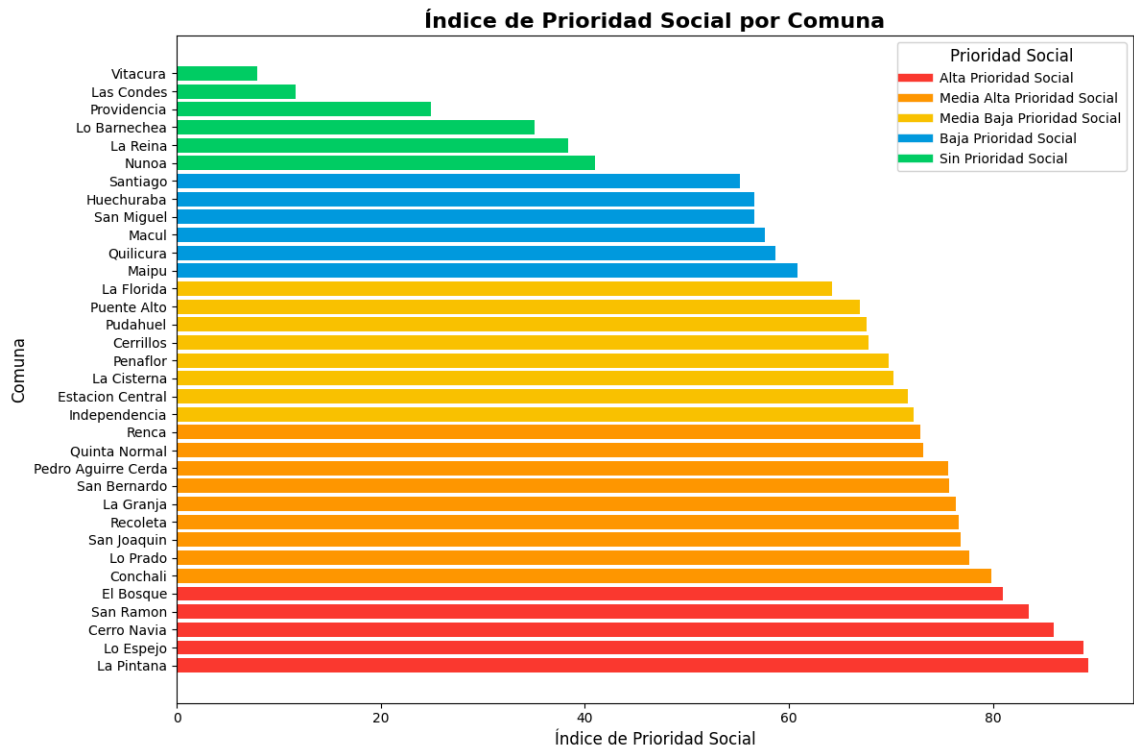


Figura 10. Índice de Prioridad Social por Comuna en la Región Metropolitana.

Mapa de Prioridad Social en la Región Metropolitana (2020)

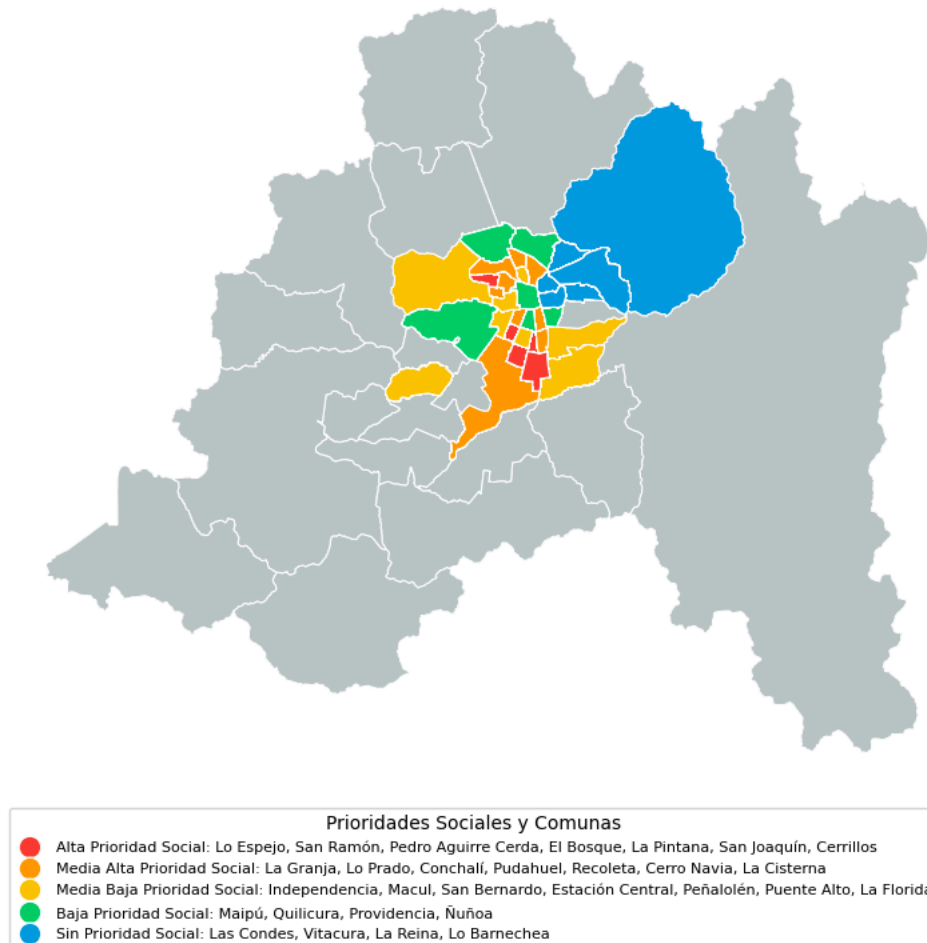


Figura 11. Distribución de las Comunas de la Región Metropolitana según Índice de Prioridad Social.

Otra de las variables sociodemográficas que se analizaron corresponde al nivel educacional, considerando el porcentaje de personas con educación superior por cada comuna incluyendo a personas profesionales, magíster y/o doctorados (Figura 12), donde se observó que las comunas que presentaron un mayor porcentaje de personas con educación superior fueron Providencia con 59,7%,

Las Condes con 54% y Vitacura con 53,2%; a estas le siguen las comunas de Ñuñoa, La Reina, Santiago y Lo Barnechea con valores entre 35% y 48%.

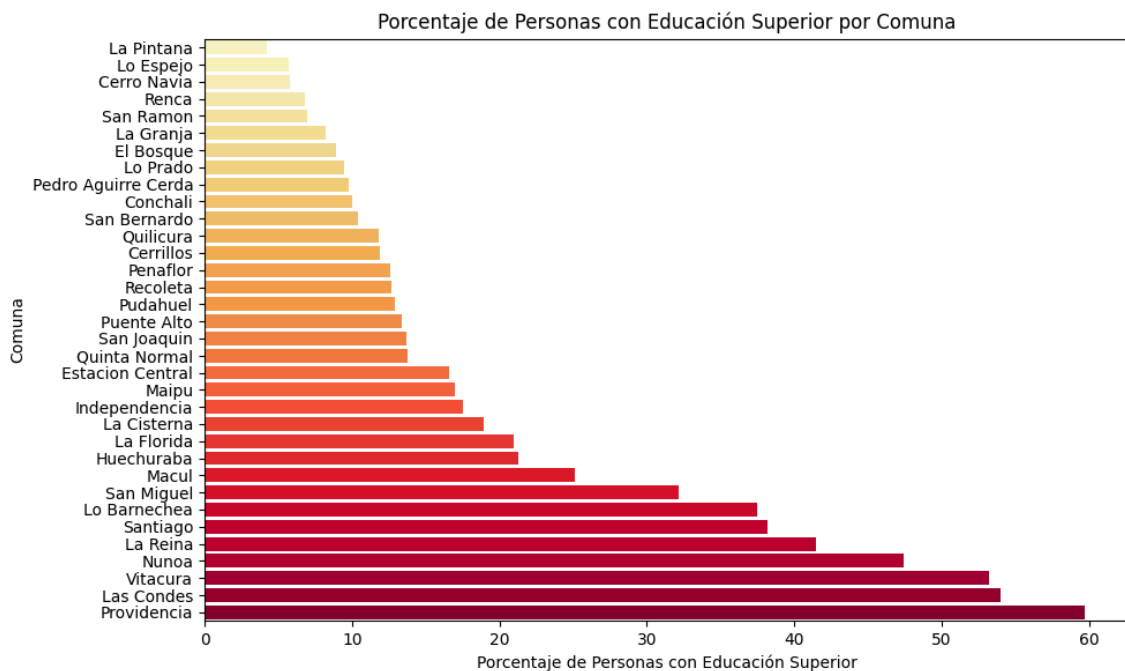


Figura 12. Porcentaje de Personas con Educación Superior por Comuna.

Por otra parte, dentro de las comunas que poseen un menor porcentaje de personas con educación superior se encuentran La Pintana con 4,2%, Lo espejo con 5,7%, Cerro Navia con 5,8% y Renca con 6,8%. En cuanto al porcentaje de personas con postgrado (magíster o doctorado) (Figura 13), se evidenció un comportamiento similar a los porcentajes de personas profesionales, donde nuevamente lideraron Providencia (12,7%), Vitacura (11,5%) y Las Condes (10,6%), mientras que las comunas con valores más bajos fueron La Pintana, Lo Espejo y Cerro Navia, todas con un 0,1% de personas con postgrado.

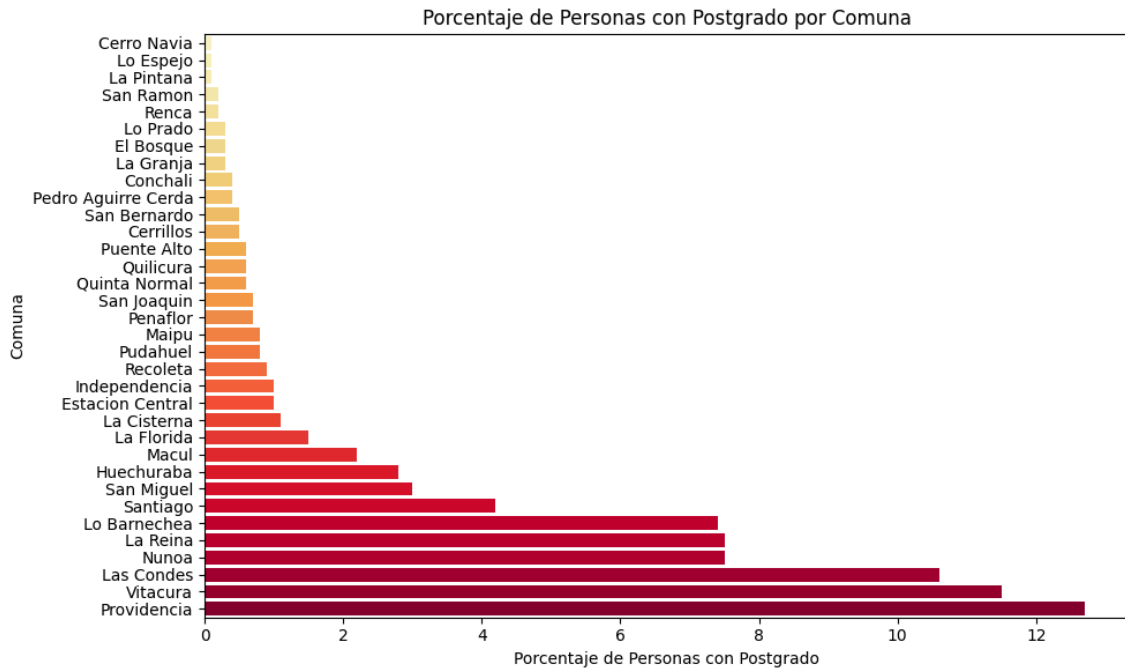


Figura 13. Porcentaje de Personas por Comuna que poseen Postgrado.

Con respecto a las variables relacionadas con la movilidad, se analizó el índice de movilidad interno correspondiente a la movilidad dentro de la comuna y el índice de movilidad externo correspondiente a las salidas y entradas en una comuna. Ambos índices fueron medidos mediante el uso de celulares y registros de Telefónica Móvil, por lo tanto, se consideraron como proxys de la movilidad de las personas en la RM.

Al visualizar la Figura 14 se puede observar que las comunas de Vitacura, Las Condes, Lo Barnechea y Providencia, presentaron los índices de movilidad interna más altos, mientras que San Ramón, La Pintana y Lo Prado fueron las comunas que presentaron los índices más bajos. En cuanto a los valores

específicos, la comuna que tuvo una mayor movilidad interna promedio fue Vitacura, con un índice de $7,9 \pm 1,9$, seguida de Las Condes con $7,7 \pm 1,9$ y Lo Barnechea con $7,5 \pm 1,5$. Las comunas que presentaron menor movilidad interna fueron San Ramón, con un índice promedio de $1,1 \pm 0,3$, La Pintana con $1,4 \pm 0,3$ y El Bosque con $1,6 \pm 0,3$.

También se observa la presencia de outliers en comunas como Quilicura, Recoleta, Lo Espejo, Huechuraba y La Florida, entre otras, lo que indica que, en ciertos días, estas comunas experimentaron índices de movilidad que se escaparon del comportamiento normal evidenciado durante el periodo de estudio. Comunas como Providencia, Santiago, Las Condes y Vitacura, presentaron una mayor variabilidad, lo que refleja que el índice de movilidad interna varió considerablemente a lo largo del tiempo.

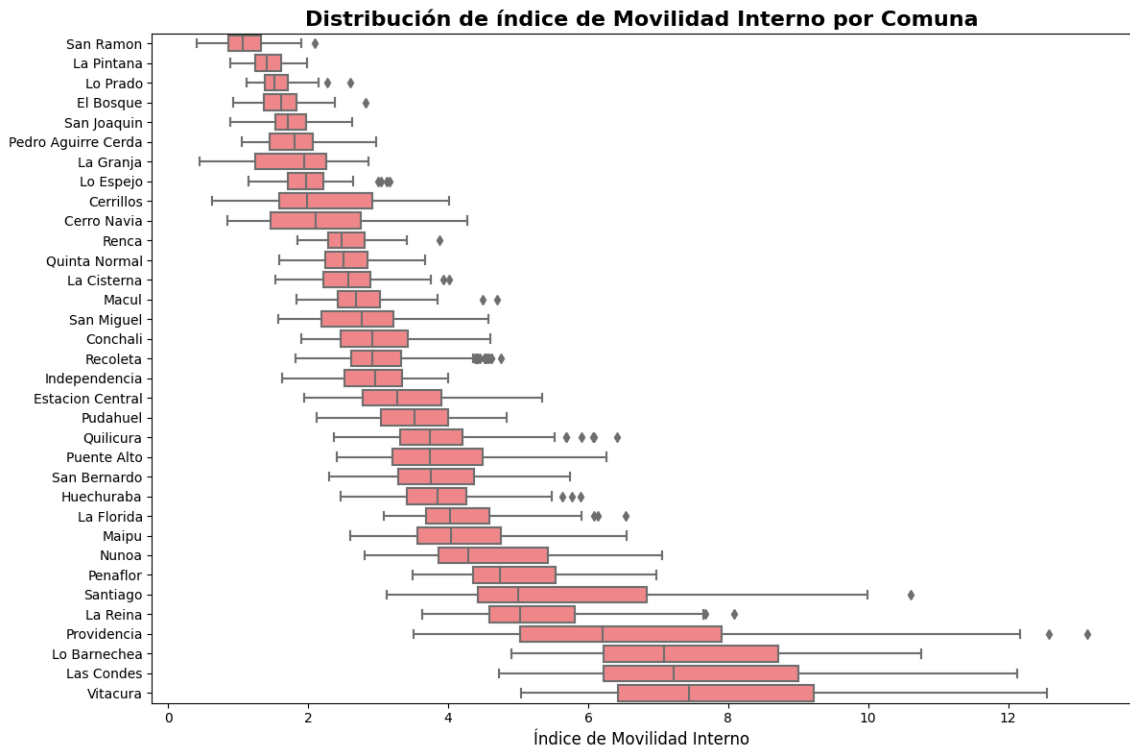


Figura 14. Distribución del Índice de Movilidad Interna por Comuna durante el año 2020.

Con respecto a la movilidad externa, las comunas que presentaron valores más altos fueron La Granja con un índice promedio de $8,7 \pm 2,3$, Cerrillos con $8,2 \pm 1,9$, La Cisterna con $7,7 \pm 1,4$ y Lo Espejo con $7,5 \pm 1,5$, y las que presentaron valores promedio más bajos fueron Puente Alto con $2,1 \pm 0,4$, Maipú con $2,7 \pm 0,8$ y San Bernardo con $2,9 \pm 0,8$. Las comunas que presentaron una mayor presencia de outliers fueron San Joaquín, Recoleta y Maipú, y en las que se observó mayor variabilidad en el comportamiento de los datos fueron La Granja, Vitacura, Providencia y San Joaquín (Figura 15).

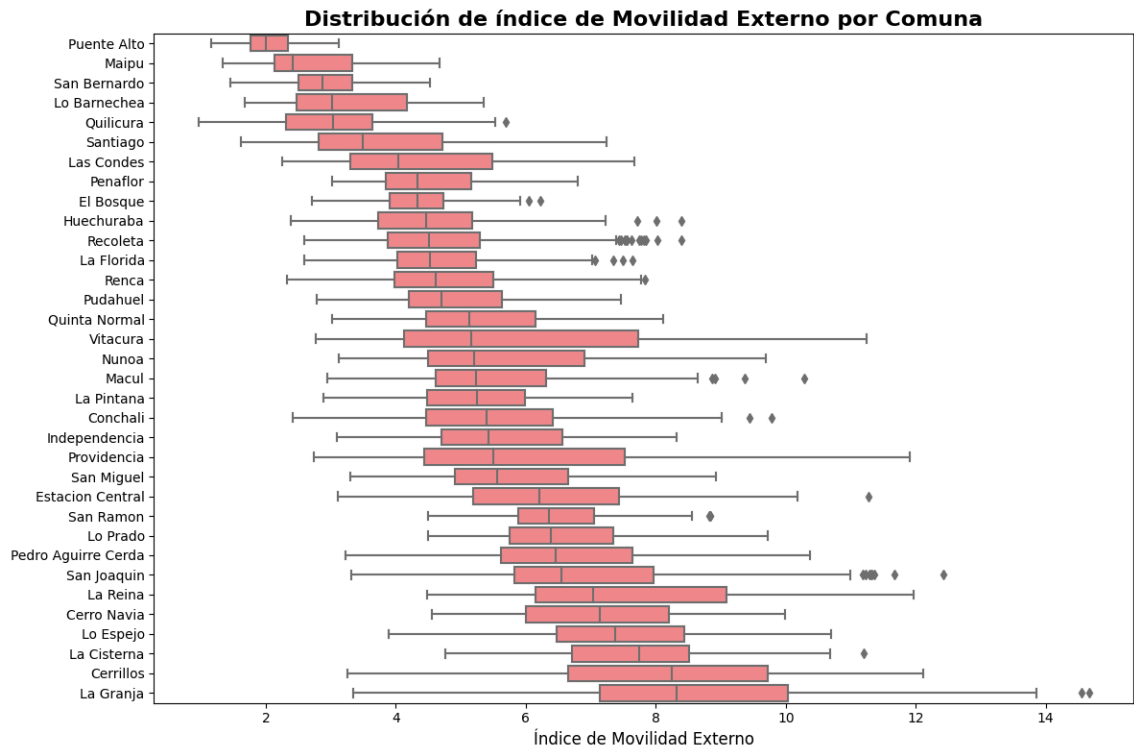


Figura 15. Índice de Movilidad Externo por Comuna durante el año 2020.

Por otra parte, al analizar la evolución del índice de movilidad interno y externo para todas las comunas en conjunto durante el año 2020, se puede observar que ambos presentan una tendencia a la alza constante, especialmente a partir de mediados de mayo. El índice de movilidad externo presenta picos pronunciados durante ciertos días a la semana, con unas bajas abruptas entre picos, evidenciando un comportamiento cíclico semanal, lo que se podría explicar debido a que de lunes a viernes hay mayor movilidad producto del traslado hacia los trabajos, en cambio durante los fines de semana las personas tendían a quedarse en sus casas, producto de la preocupación y de las medidas de restricción implementadas. A la vez, durante la primera mitad de mayo se observa

una caída notoria en la movilidad, lo que se puede deber a las medidas de restricción que fueron implementadas en dicho periodo.

Al comparar el comportamiento de movilidad entre ambos índices, se puede observar que el índice de movilidad interno presentó un crecimiento más moderado en comparación al índice externo, donde igual se observa aumento sostenido a partir de mediados de mayo, pero con una menor variabilidad a lo largo del tiempo y durante cada mes. Esto sugiere que, si bien las personas aumentaron la movilidad interna a lo largo del tiempo, los desplazamientos internos se mantuvieron más constante, posiblemente debido a la permanencia de restricciones en los viajes intercomunales. Paralelamente, para ambos índices la diferencia o variabilidad entre el comportamiento de los días de semana versus fines de semana se va acentuando a medida que pasa el tiempo, sobre todo desde agosto en adelante.

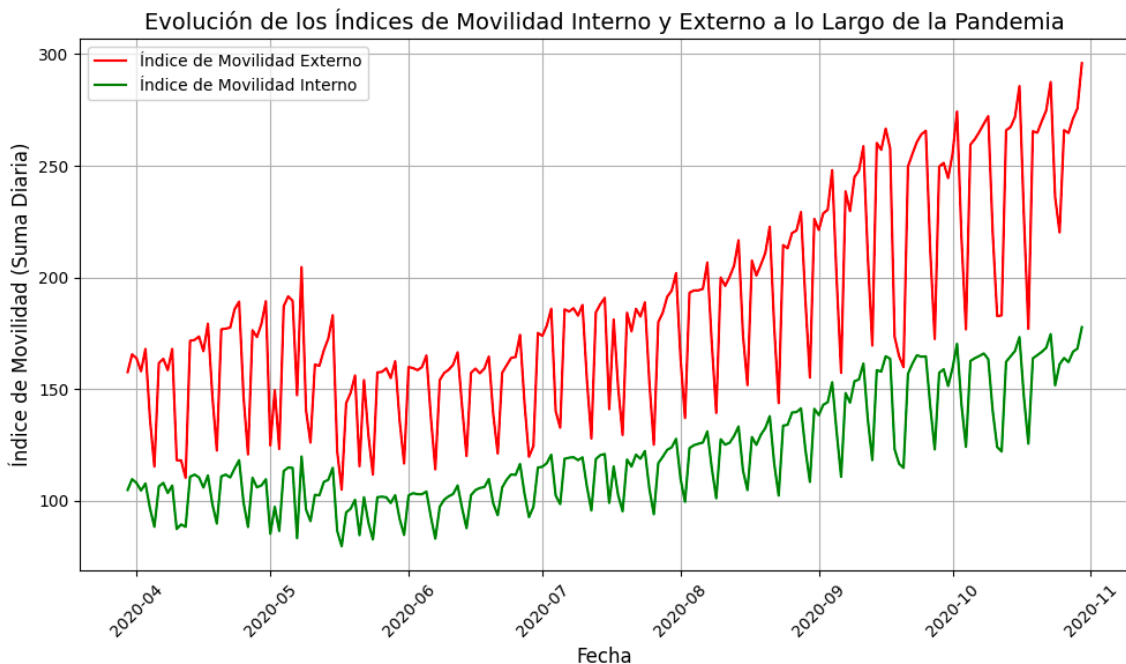


Figura 16. Evolución de los Índices de Movilidad Interno y Externo durante el año 2020.

Para analizar el comportamiento de la movilidad también se utilizó el número de viajes diarios entre comunas, a partir de los datos de transporte públicos disponibles, los cuales abarcaron desde marzo 2020 hasta mediados de mayo 2020. La visualización se presenta a través de grafos, en los cuales el color de las aristas representan la cantidad de viajes (origen y destino).

Al observar el flujo de viajes origen – destino entre comunas de la RM (Figura 17), los resultados reflejan que las comunas céntricas como Santiago, Providencia, Las Condes y Estación Central funcionan como hubs de transporte, es decir, son las comunas donde se realizan la mayor cantidad de viajes, lo que sugiere que la dinámica de viajes en la RM se produce mayoritariamente entre

estas comunas nodos. Por otro lado, comunas más periféricas tienen menos participación en el flujo de viajes, lo que puede sugerir que son zonas con menor actividad económica o que están más aisladas en términos de movilidad.

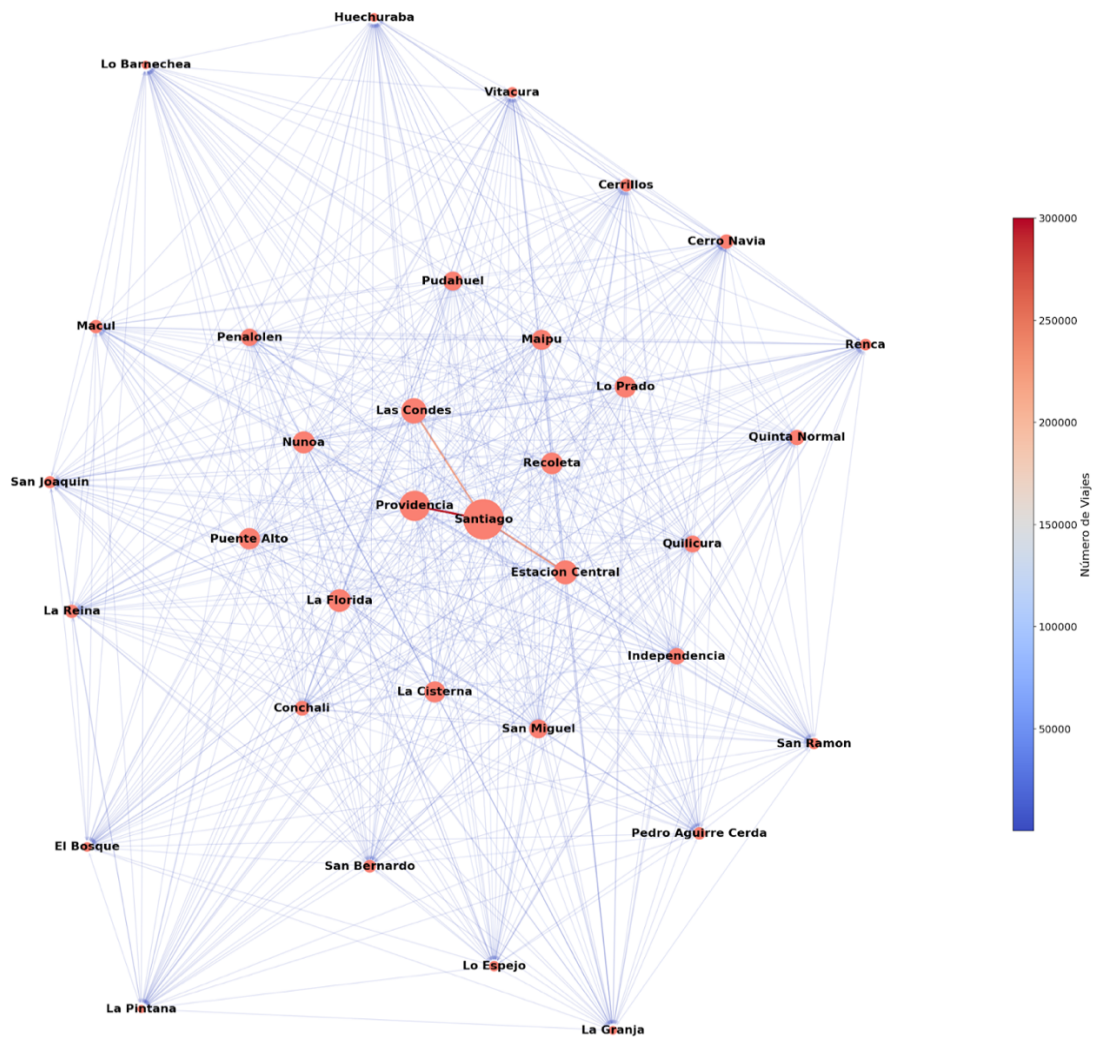


Figura 17. Visualización del Flujo de Viaje Origen – Destino entre las distintas Comunas de la Región Metropolitana.

Al analizar el flujo de viajes hacia las 10 comunas con más movimientos de destino (Figura 18), se puede deducir que el desplazamiento desde las comunas

de nodos pequeños es equitativo hacia las comunas hubs, destacándose Santiago y Providencia como comunas principales, seguidas de comunas como Ñuñoa, Puente Alto, Estación Central y La Florida, entre otras. Esto permite inferir que estas comunas también presentan características propias asociadas a la funcionalidad, como la presencia de comercio, estaciones de trenes y estaciones de buses, lo que las convierte en nodos importantes debido al gran flujo de personas a estos lugares.

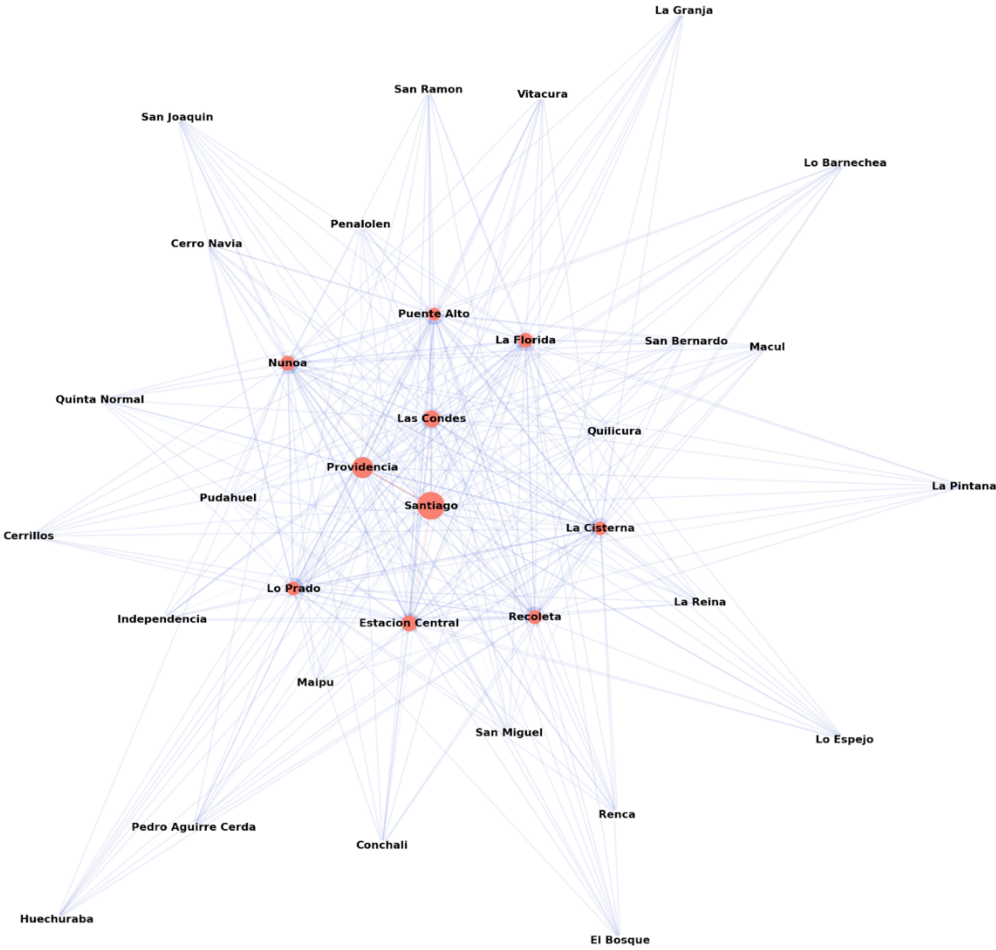


Figura 18. Visualización del Flujo de Viajes hacia las 10 Comunas con más Movimiento de Destino.

Por otra parte, al observar el flujo de viajes desde las 10 comunas con más movimientos de origen hacia sus respectivos destinos (Figura 19), que podría corresponder al viaje de regreso desde las comunas de alta actividad laboral y comercial hacia comunas de tipo dormitorio, se puede evidenciar la densidad de conexiones hacia comunas centrales y densamente pobladas como Puente Alto, La Florida, Estación Central y La Cisterna, entre otras, resaltando su rol en la red de movilidad, pudiendo ser consideradas como hubs de transporte.

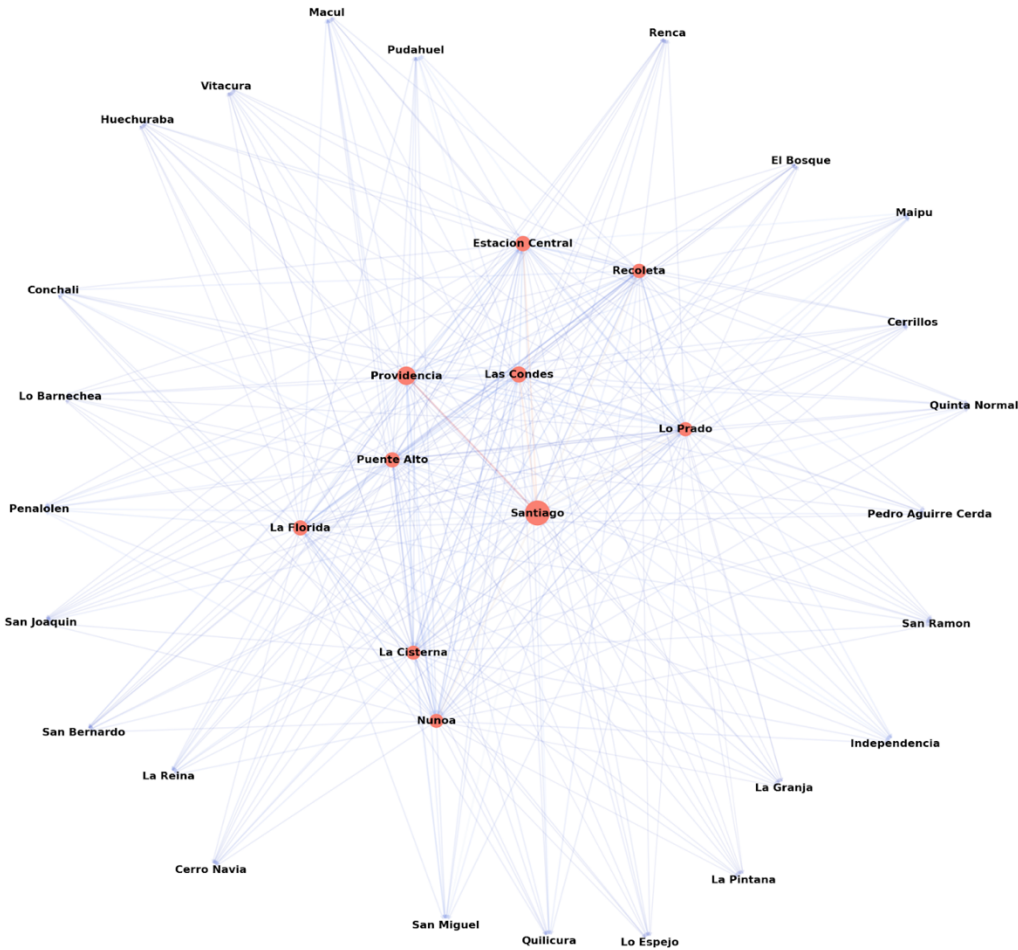


Figura 19. Visualización del Flujo de Viajes Salientes desde las 10 Comunas con más Movimiento de Origen a sus respectivos Destinos.

Paralelamente, para visualizar los cambios en la movilidad producto de la pandemia, se analizó la variación del flujo de viajes en horario punta AM y PM para distintos periodos de tiempo, considerando un periodo de comportamiento normal previo a la pandemia (agosto 2019), el periodo de inicio de la pandemia (marzo 2020) y un periodo posterior al establecimiento (noviembre 2020). Esto fue representado a través de grafos, donde el tamaño de los nodos es proporcional al total de entradas o salidas según su clasificación y el grosor de las aristas representa la cantidad de tiempo total de viaje acumulado entre comunas.

Al analizar el flujo de viajes en la RM durante el horario punta AM, entre las 07:00 y 08:59 horas abarcando desde el 5 hasta el 12 de agosto de 2019 (periodo pre – pandemia) (Figura 20), se puede observar que las comunas con mayor relación de destinos corresponden Santiago, Providencia y Las Condes, y las comunas con mayor relación de orígenes son Puente Alto, seguido de Maipú y La Florida. Estación Central, San Miguel e Independencia corresponden a las zonas de paso o mixtas que presentan mayor movimiento.

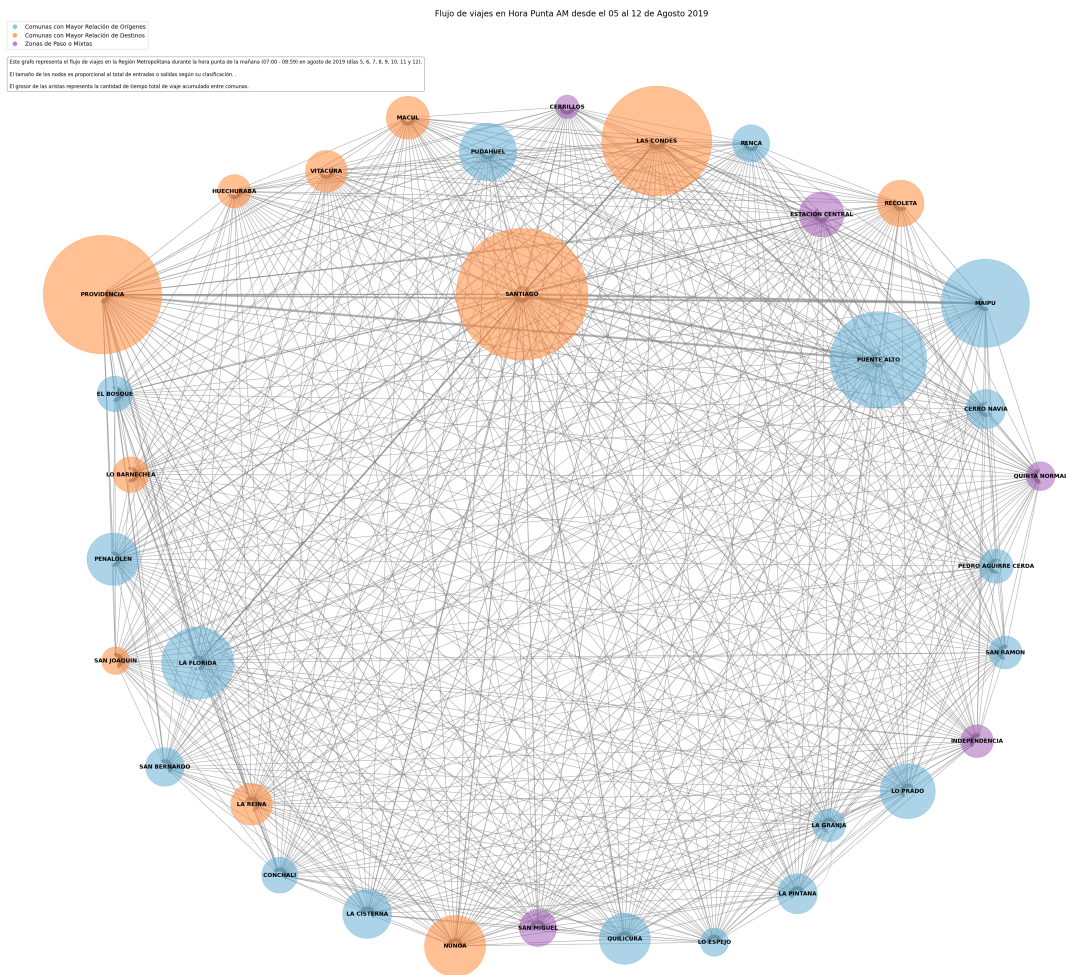


Figura 20. Visualización del Flujo de Viajes Pre - Pandemia durante Horario Punta AM entre 07:00 y 08:59 horas.

En cuanto al flujo de viajes durante el horario punta PM, entre las 18:00 y 19:59 horas de los mismos días (Figura 21), las comunas con mayor relación de destinos fueron Puente Alto, La Florida, Estación Central, Maipú, y las comunas con mayor relación de orígenes fueron Las Condes, Providencia, Santiago, Vitacura, seguido de Lo Barnechea y Huechuraba. Además, se observa que las comunas con mayores tiempos de traslado están interconectadas principalmente

entre sí, destacándose las conexiones entre Las Condes, Providencia, Santiago, Maipú y La Florida.

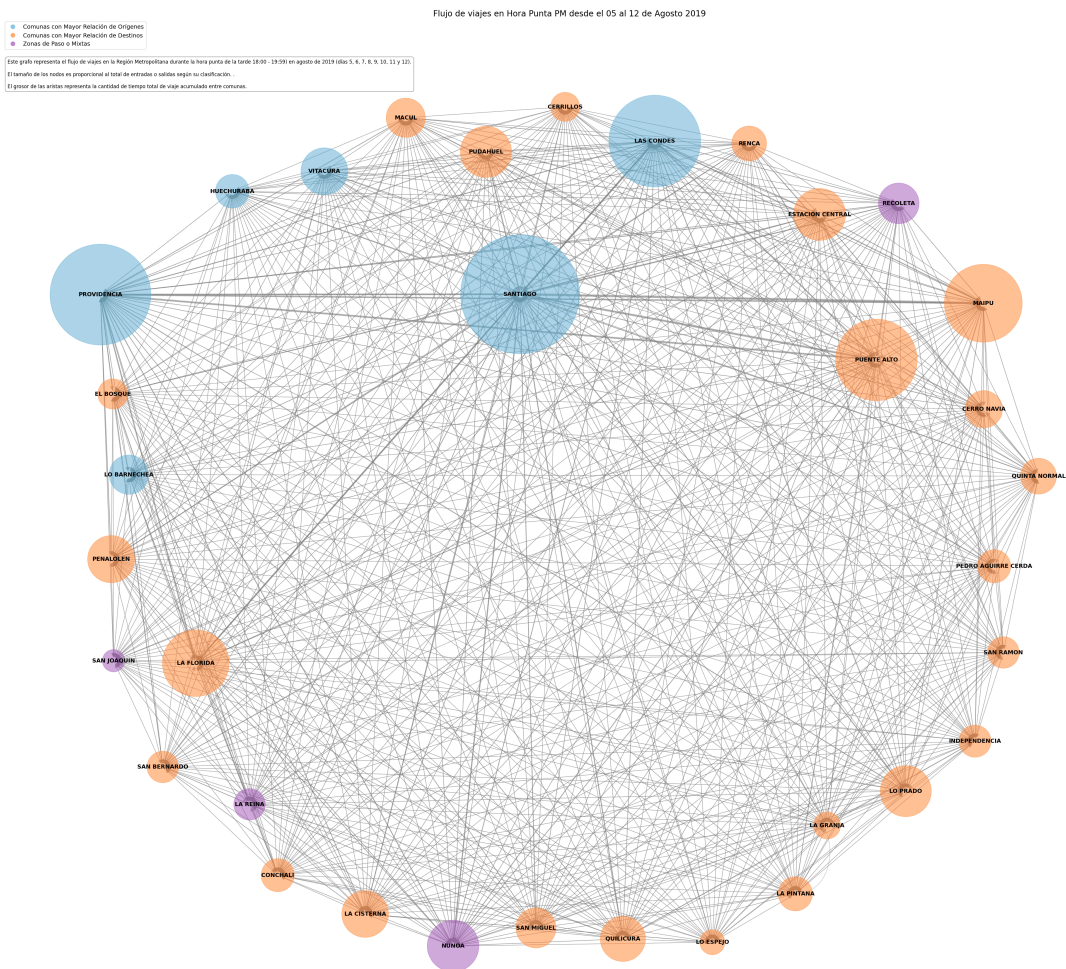


Figura 21. Visualización del Flujo de Viajes Pre - Pandemia durante Horario Punta PM entre 18:00 y 19:59 horas.

Con respecto al flujo de viajes durante el horario punta AM en el periodo de inicio de la pandemia (desde el 7 al 13 de marzo) (Figura 22), se puede observar que el tamaño de los nodos es menor en comparación al grafo de movilidad pre –

pandemia (Figura 20), lo que evidencia que el número de movimientos se vio disminuido, pero se mantuvieron las comunas con mayor relación de destinos y las comunas con mayor relación de orígenes. A la vez, las comunas con mayores tiempos de traslado están interconectadas principalmente entre ellas, destacándose las conexiones entre Las Condes, Providencia, Santiago, Maipú y La Florida.

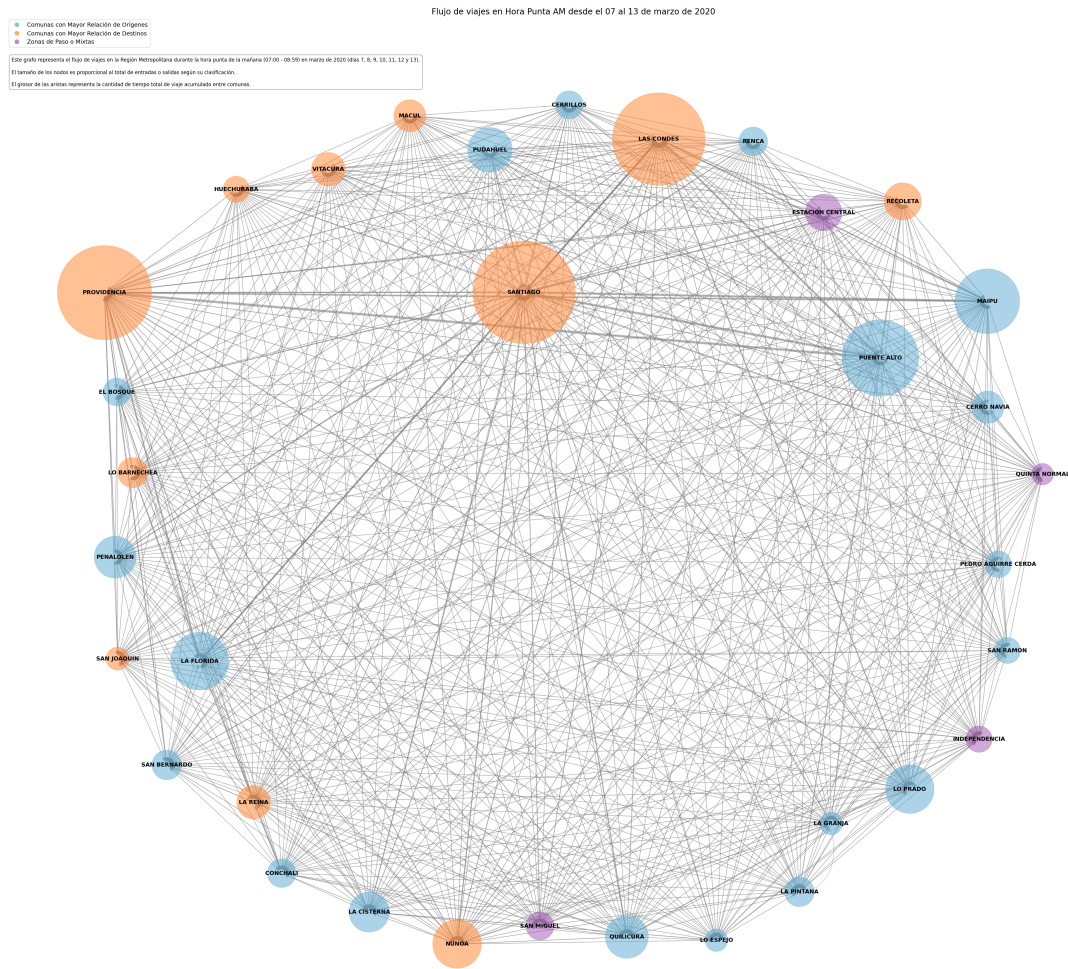


Figura 22. Visualización del Flujo de Viajes a Inicios de Pandemia durante Horario Punta AM entre 07:00 y 08:59 horas.

En el caso de los viajes en hora punta PM para el mismo periodo de inicio de pandemia (Figura 23), se puede observar que los nodos del grafo también disminuyeron su tamaño en comparación al grafo del periodo pre – pandemia (Figura 21), reflejando que la cantidad de viajes se redujo considerablemente. Asimismo, las comunas con mayores tiempos de traslado también se encuentran interconectadas principalmente entre sí, destacándose las conexiones entre Las Condes, Providencia, Santiago, Maipú y La Florida.

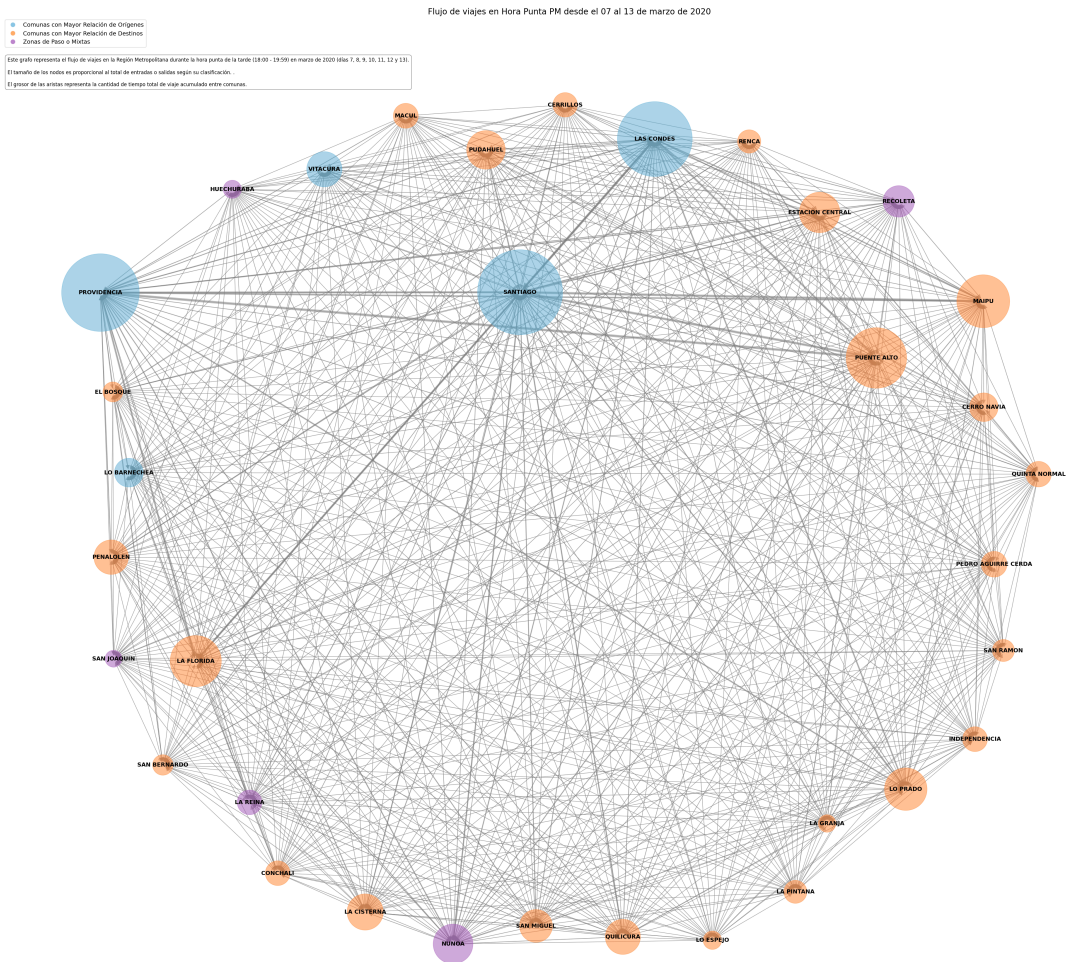


Figura 23. Visualización del Flujo de Viajes a Inicios de Pandemia durante Horario Punta PM entre 18:00 y 19:59 horas.

En cuanto al flujo de viajes durante el horario punta AM en el periodo de pandemia ya establecida (desde el 9 al 15 de noviembre de 2020) (Figura 24), se puede observar una disminución del tamaño de los nodos en comparación al periodo previo pandemia y al periodo de inicio de pandemia (Figura 20 y 22 respectivamente), reflejando una mayor reducción en la cantidad de viajes, sin

embargo, se mantuvo la dinámica respecto a las comunas con mayores tiempos de traslado, estando interconectadas entre sí.

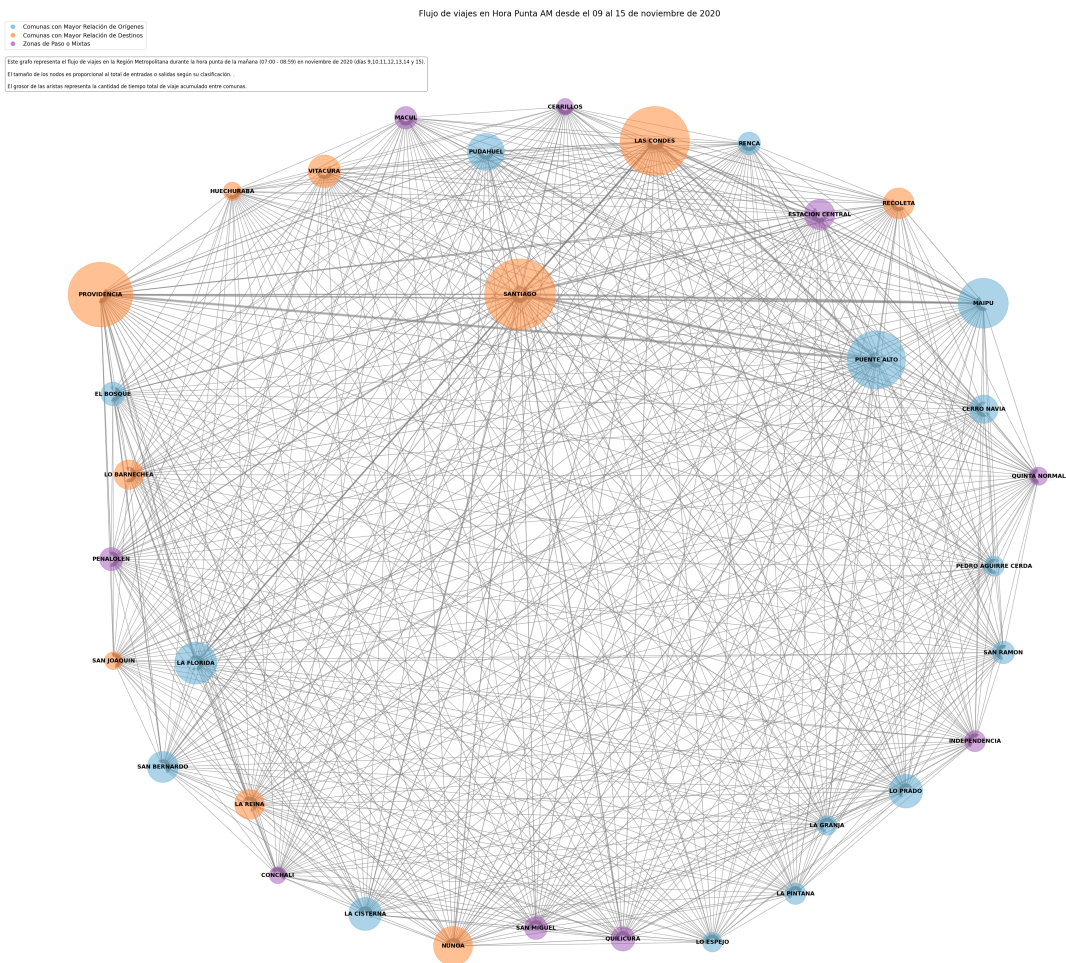


Figura 24. Visualización del Flujo de Viajes en Pandemia durante Horario Punta AM entre 07:00 y 08:59 horas.

Por otra parte, al analizar el flujo de viajes en horario punta PM durante este último periodo (Figura 25), también se observa una reducción considerable en la cantidad de viajes en comparación con los periodos anteriores (Figura 21 y 23).

Además, la comuna de Santiago cambió su dinámica, convirtiéndose en una comuna crítica desde el punto de vista del traslado de las personas, la relación entre personas que salen de esta comuna es similar al de las que entran, pudiendo ser un efecto de las medidas de restricción de movilidad en ciertas comunas que cumplen un rol importante en la interconexión entre comunas.

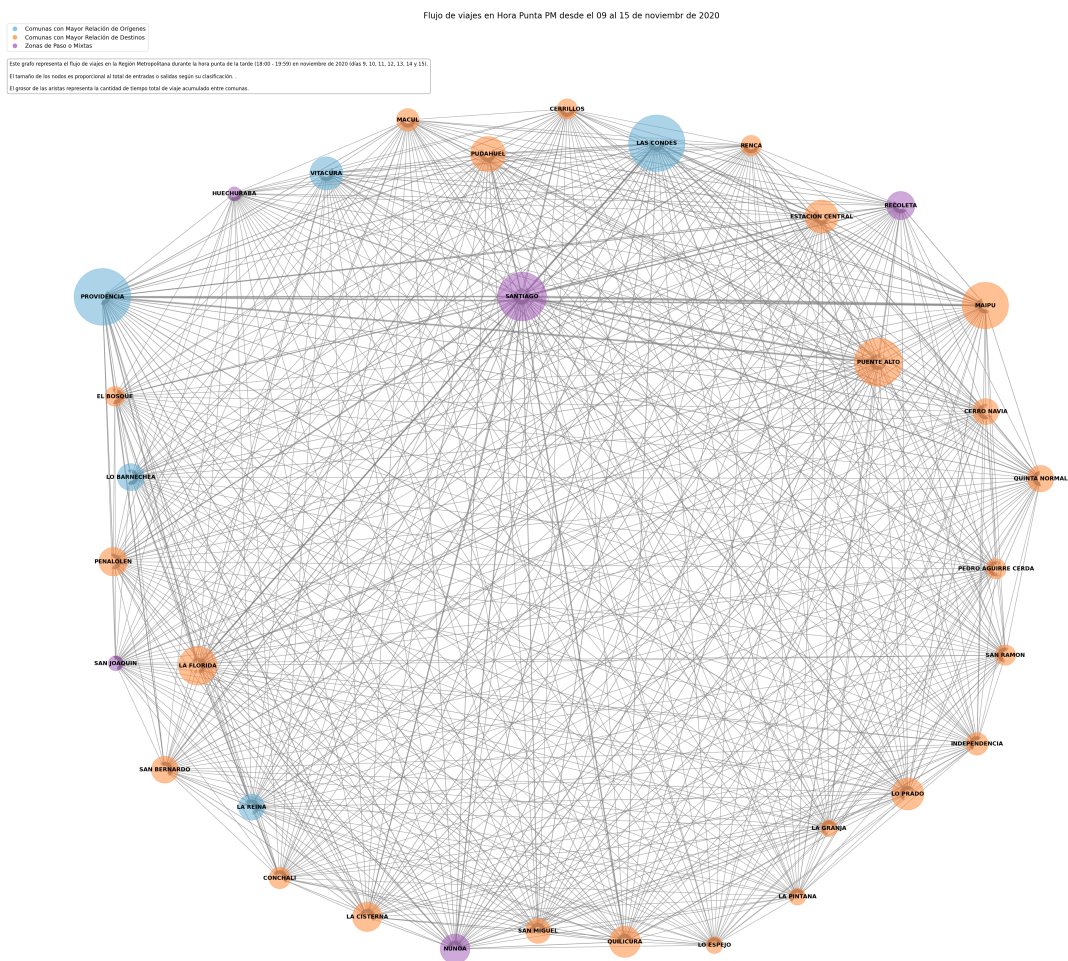


Figura 25. Visualización del Flujo de Viajes en Pandemia durante Horario Punta PM entre 18:00 y 19:59 horas.

A continuación se presenta el total de viajes internos por comuna, comparando los tres periodos clave: agosto de 2019 (pre – pandemia), marzo de 2020 (inicios de la pandemia) y noviembre de 2020 (pandemia ya instaurada) (Figura 26).

Los resultados evidencian una disminución progresiva en la movilidad interna entre estos tres periodos, destacando el elevado nivel de movimientos internos en la comuna de Santiago, lo cual es consistente con las dinámicas observadas en los grafos previos sobre los flujos de viajes, consolidando a Santiago como un nodo primordial en la red de movilidad urbana.

Por otro lado, la comuna de Maipú resalta por ser la única en registrar un aumento en la cantidad de viajes internos durante la pandemia (noviembre de 2020) en comparación con los inicios de esta (marzo de 2020). Esto pone en manifiesto las dinámicas particulares de movilidad de cada comuna, evidenciando que las respuestas ante la pandemia varían significativamente según el contexto local.

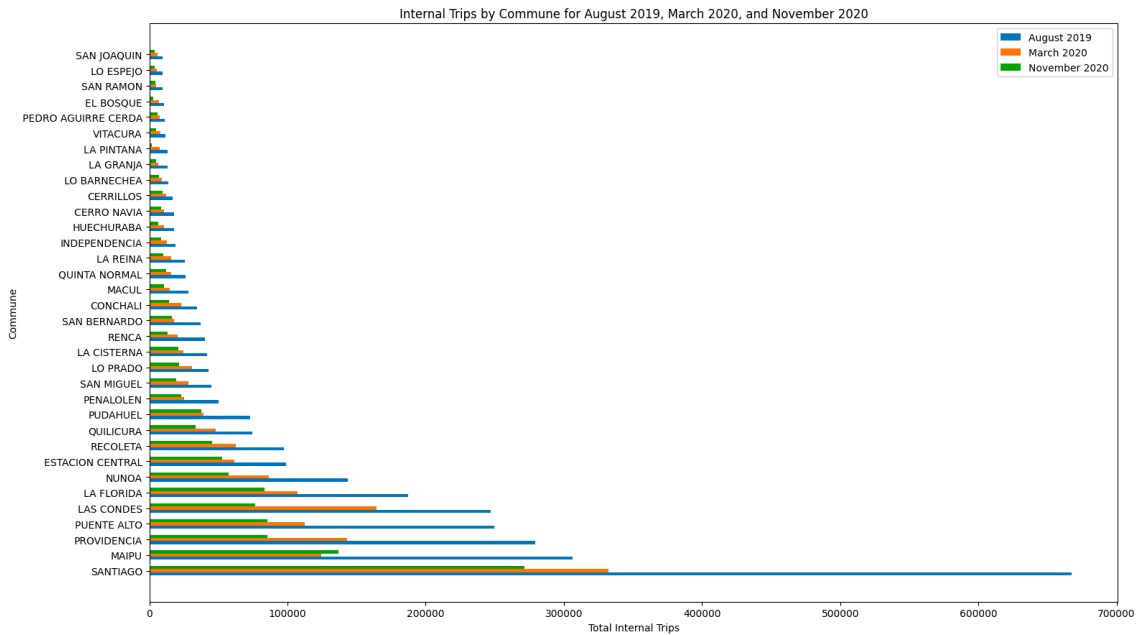


Figura 26. Movilidad Interna entre Comunas de la Región Metropolitana Previo, al Inicio y Durante la Pandemia.

Al analizar el comportamiento de la tasa de casos acumulado en paralelo a la evolución de los índices de movilidad a lo largo del periodo de estudio, se puede observar cómo la movilidad se fue adaptando y evolucionando frente a las distintas etapas de la pandemia y medidas sanitarias implementadas (Figura 27).

Los índices de movilidad externo (línea roja discontinua) e interno (línea verde discontinua), presentan valores bajos durante los primeros meses de pandemia, principalmente hasta fines de junio, tiempo en que la mayoría de las comunas se encontraba en cuarentena, y entre finales de julio e inicios de agosto estos comienzan a aumentar progresivamente, lo que coincide con el periodo en que las primeras comunas pasan de cuarentena a transición. A la vez, se observa un comportamiento estacional marcado, donde se repite un patrón semanal

caracterizado por mayor movimiento de lunes a viernes y una baja brusca los fines de semana.

Si bien, lo descrito anteriormente refleja el comportamiento global de la movilidad durante dicho periodo, hay algunos episodios que destacan, como lo es la caída brusca en el índice de movilidad externo que se observa a mediados de mayo, cuando comenzó la fase exponencial de contagios.

Por otro lado, el índice de movilidad interna, aunque con valores más bajos que el externo, presentó un comportamiento consistente con una tendencia de crecimiento moderado a lo largo del tiempo. Esto sugiere que las actividades dentro de las comunas continuaron desarrollándose, aunque en menor medida, posiblemente como resultado de la priorización de desplazamientos locales frente a las restricciones de movilidad intercomunal.

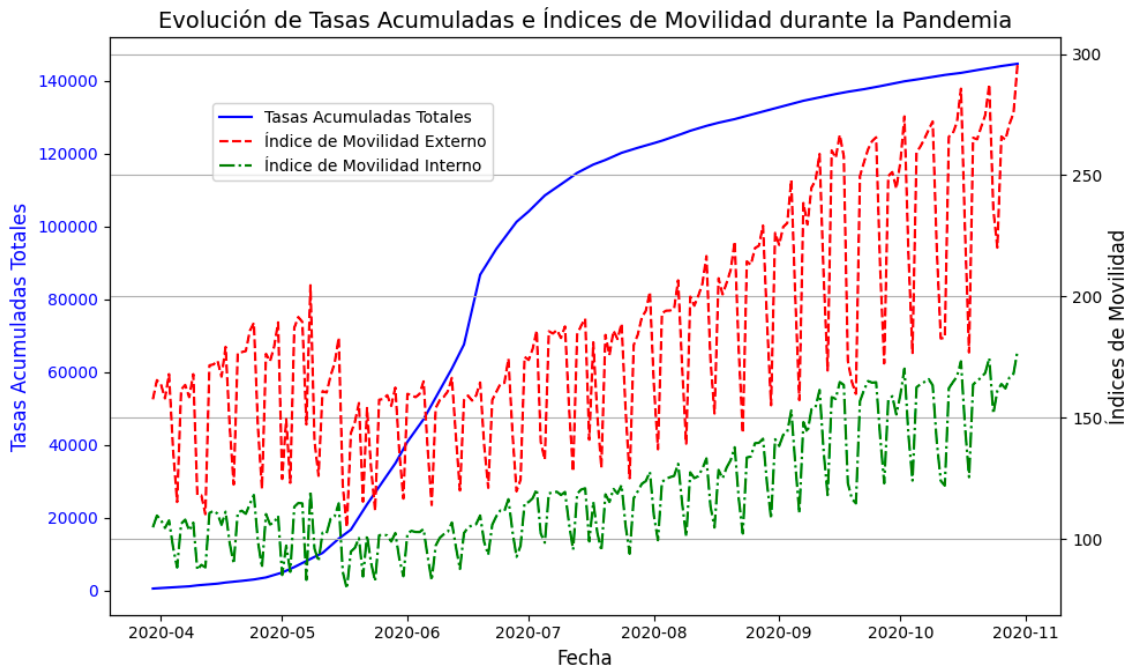


Figura 27. Evolución de Tasas Acumuladas e Índices de Movilidad durante Pandemia el año 2020.

A continuación se analiza la evolución de las tasas acumuladas de casos diarios de Covid-19 por comuna, en paralelo con las fases del Plan Paso a Paso implementado en Chile, donde cada barra representa una comuna, y el gradiente de colores ilustra las tasas de contagio acumuladas, mientras que las franjas horizontales indican las distintas fases del plan en las que se encontraba cada comuna en un momento determinado (Figura 28).

En este gráfico se evidencia la complejidad de la propagación del COVID-19 en un área metropolitana heterogénea como Santiago. Comunas densamente pobladas y con mayores índices de vulnerabilidad socioeconómica, presentaron mayores tasas acumuladas de contagio, lo que las mantuvo por más tiempo en

fases restrictivas del Plan Paso a Paso. En contraste, comunas con mejor infraestructura y menor densidad poblacional, como Vitacura y Las Condes, avanzaron más rápidamente hacia fases de apertura, reflejando una mejor contención del virus, quedando en manifiesto la importancia de adaptar las medidas sanitarias a las características particulares de cada comuna para mejorar la gestión de futuras crisis sanitarias.

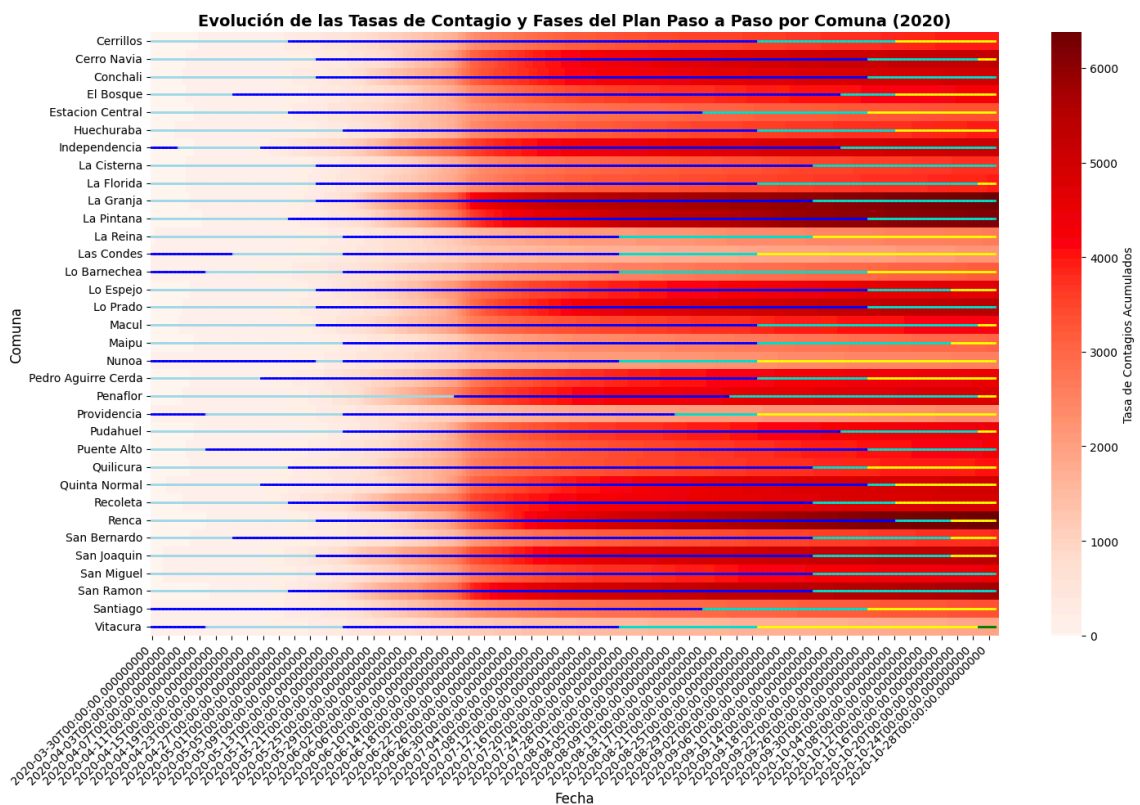


Figura 28. Evolución de las Tasas de Contagio y Fases del Plan Paso a Paso por Comuna el año 2020.

Por otra parte, con el fin de identificar de forma exploratoria las relaciones lineales entre las variables sociodemográficas, índices de movilidad y tasa de casos

acumulados de Covid-19, se realizó una matriz de correlación visualizada a través de un heatmap, la cual evidencia patrones clave para comprender las posibles interacciones entre los factores analizados (Figura 29).

La tasa de casos acumulados mostró una correlación positiva moderada con el índice de movilidad externo (0,40), indicando que comunas con mayor movilidad externa presentan mayores tasas de contagio acumulado. En contraste, la correlación entre la tasa de casos acumulados y el índice de movilidad interno fue prácticamente nula (-0,09), sugiriendo que la movilidad dentro de las comunas no estuvo directamente asociada con el aumento acumulado de casos.

Al analizar las relaciones entre las variables sociodemográficas y los casos de Covid-19, se observó que el ingreso per cápita presentó una correlación negativa con la tasa de casos acumulados (-0,29), lo que podría indicar que las comunas con mayores ingresos lograron contener mejor el avance del virus. En contraste, el índice de precariedad de vivienda mostró una correlación positiva con la tasa de casos acumulados (0,27), reflejando que las condiciones habitacionales desfavorables pudieron contribuir al aumento del contagio.

Respecto a las variables sociodemográficas y su relación con la movilidad, el ingreso per cápita mostró una fuerte correlación positiva con el índice de movilidad interno (0,74), evidenciando que las comunas con mayores ingresos presentan mayores niveles de desplazamiento interno. De manera similar, se detectó una correlación positiva entre el porcentaje de educación superior y el

índice de movilidad interno (0,74), lo que sugiere que las comunas con mayor nivel educativo tienen patrones más marcados de movilidad dentro de sus límites.



Figura 29. Correlación entre tasa de casos acumulados, índices de movilidad y variables sociodemográficas.

Finalmente, se analizó la evolución de las tasas de contagio de COVID-19 abarcando tres niveles de análisis: casos acumulados, casos diarios y la aceleración de los contagios, diferenciados por comuna (Figura 30). Estas series permiten identificar patrones y dinámicas en la propagación del virus, desde distintos puntos de vista, por ejemplo, la tasa de contagios acumulados, que muestra el crecimiento progresivo del número total de contagios en cada comuna,

muestra que Santiago, Puente Alto y Maipú presentan tasas de contagio acumulado más altas, alcanzando valores superiores a 5.000 casos hacia finales del 2020. Aunque todas las comunas siguen una tendencia ascendente, algunas como Vitacura y Lo Barnechea muestran un crecimiento más lento y tasas acumuladas considerablemente menores, probablemente debido a diferencias en densidad poblacional, nivel socioeconómico y acceso a medidas preventivas.

Por otra parte, en el gráfico central de Tasas de Contagios Diarios se observa un peak pronunciado de contagios entre mayo y junio, consistente con el período de mayor propagación en la Región Metropolitana, después las tasas diarias comienzan a disminuir, reflejando los efectos de las medidas sanitarias. Además, se puede observar que comunas como Puente Alto y Maipú presentan peaks más altos indicando una mayor velocidad de propagación en estos territorios densamente poblados.

El gráfico inferior de aceleración de la tasa de contagios permite evaluar los cambios en la velocidad de contagio, mostrando dónde se acelera o desacelera la propagación del virus. Las curvas muestran un peak positivo de aceleración durante mayo y junio, seguido por una desaceleración abrupta, lo que coincide con la implementación de restricciones más estrictas como cuarentenas y cierre de actividades no esenciales. Posteriormente, hacia agosto y septiembre de 2020, la aceleración se estabiliza en valores cercanos a cero, lo que sugiere un control más efectivo de la propagación del virus.

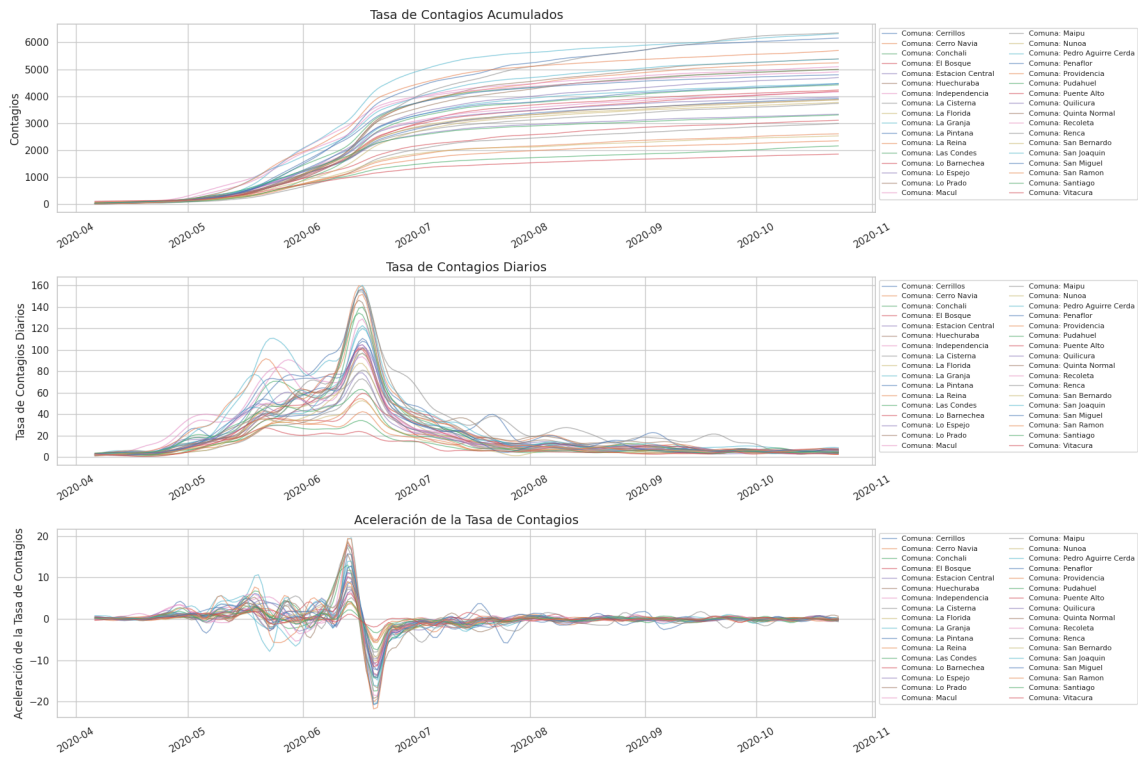


Figura 30. Series de Tasa de Casos Diarios Acumulados por Comuna.

2. Predicción Baseline

2.1. Baseline con Shallow Classifiers

Tras analizar el comportamiento de las variables mediante el análisis exploratorio de datos, se procedió a establecer una línea base para la predicción del número de casos diarios de contagio en cada comuna. Para ello, se ajustaron clasificadores y regresores simples (Shallow Classifiers) a las series de tiempo, con el objetivo de evaluar modelos iniciales que sirvieran como referencia para comparaciones posteriores, dentro de los cuales se consideraron Support Vector Machine (SVM), Random Forest (RF) y XGBoost (XGB), debido a su capacidad

de capturar relaciones no lineales y patrones en los datos. A continuación, se presentan los resultados iniciales de estos modelos, así como los obtenidos tras implementar una ventana móvil de 7 días, y realizar ajustes de hiperparámetros y aplicar validación cruzada para optimizar su rendimiento.

Los resultados iniciales indicaron que Random Forest mostró el menor Error Absoluto Medio (MAE) (13.272) y Raíz del Error Cuadrático Medio (RMSE) (21.932), con un desempeño aceptable para una línea base; XGBoost obtuvo un MAE ligeramente superior (13.984) y RMSE (23.272), destacándose por su capacidad para capturar patrones más complejos, en cambio SVM presentó los mayores errores MAE (16.618) y RMSE (25.568) reflejando una menor capacidad para modelar la variabilidad temporal (Tabla 1) (Figura 30).

Tabla 1. Comparación de Métricas MAE y RMSE entre Modelos Support Vector Machine, Random Forest y XGBoost.

Modelos	MAE	RMSE
Random Forest	13,272	21,932
Support Vector Machine	16,618	25,568
XGBoost	13,984	23,272

MAE: Error Absoluto Medio; RMSE: Raíz del Error Cuadrático Medio

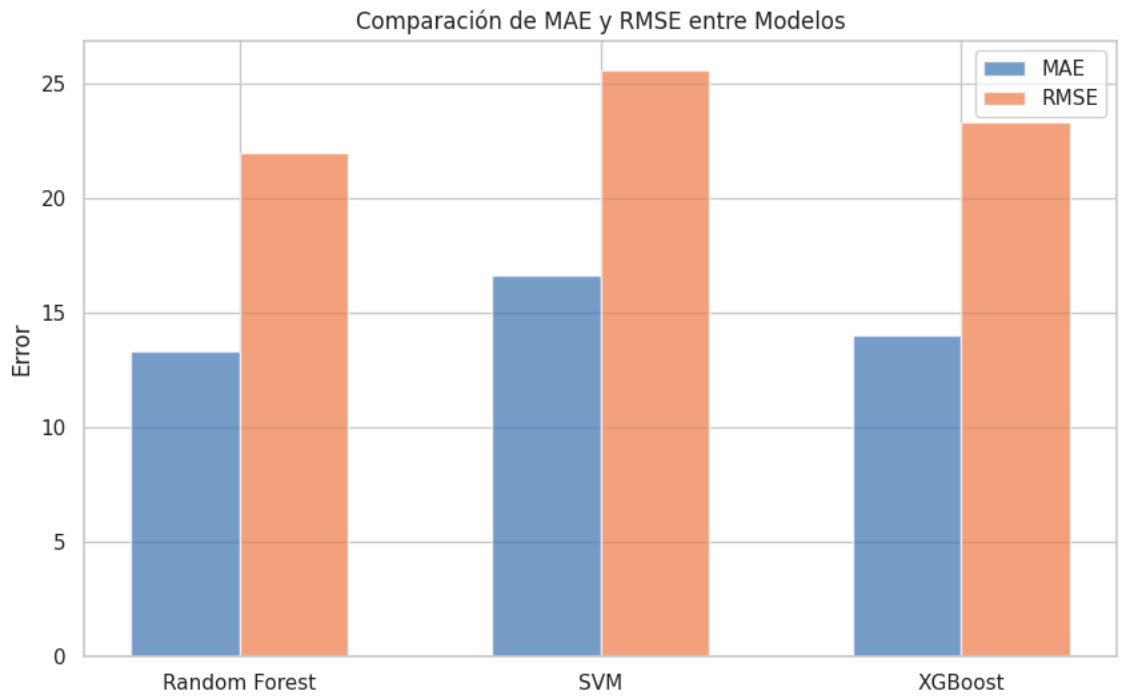


Figura 31. Comparación de MAE y RMSE entre Modelos Support Vector Machine, Random Forest y XGBoost.

Posteriormente, se incorporó una ventana deslizante de tamaño 7 para mejorar la capacidad predictiva, reduciendo significativamente los errores, donde se observa que XGBoost mejoró su rendimiento con un MAE de 1.613 y RMSE de 3.549 (Tabla 2) (Figura 32).

Tabla 2. Comparación de Métricas MAE y RMSE entre Modelos Support Vector Machine, Random Forest y XGBoost utilizando Ventana Deslizante de 7 días.

Modelos	MAE	RMSE
Random Forest	2,392	5,749
Support Vector Machine	8,005	17,088
XGBoost	1,613	3,549

MAE: Error Absoluto Medio; RMSE: Raíz del Error Cuadrático Medio

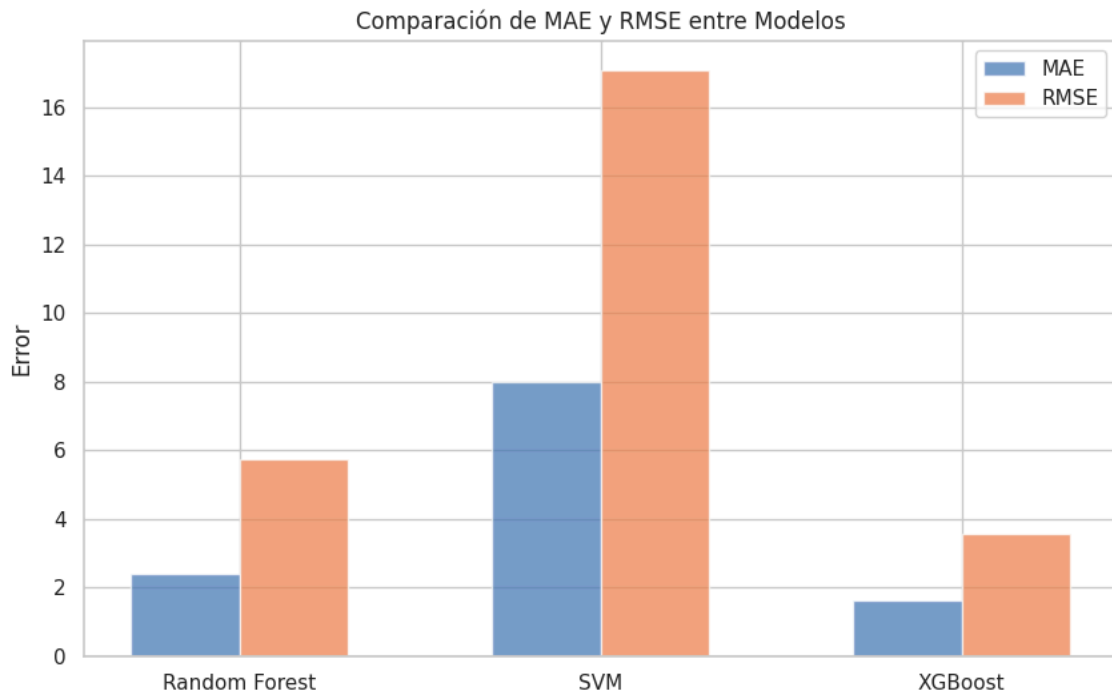


Figura 32. Visualización de la comparación de Métricas MAE y RMSE entre Modelos Support Vector Machine, Random Forest y XGBoost utilizando Ventanas de Tiempo.

Pese a obtener mejores métricas de MAE y RMSE en comparación a los resultados iniciales, en las figuras 33, 34 y 35 se puede observar que los modelos Shallow Classifiers han sobreajustado los datos en algunas comunas, debido a las características inherentes de los modelos, que tienden a aprender patrones específicos del conjunto de entrenamiento. En este caso, puede que Random Forest se haya sobreajustado debido a la alta profundidad de los árboles, capturando detalles muy específicos del conjunto de datos.

En el caso de SVM, cuando se utiliza un kernel no lineal como función de base radial (RBF) e hiperparámetros como C muy elevados, el modelo busca ajustar

incluso las pequeñas variaciones perdiendo capacidad de generalización, y en el caso de XGBoost, como es un modelo basado en boosting, se adapta iterativamente para reducir errores residuales, lo que también puede derivar a ajustes excesivos. Por lo tanto, los próximos pasos fueron implementar validación cruzada k-fold para evaluar la estabilidad de los modelos y reducir el riesgo de sobreajuste. Asimismo, se ajustaron hiperparámetros como la profundidad de los árboles (Random Forest y XGBoost) o los hiperparámetros C y γ en SVM.

Predicciones Random Forest vs Valores Reales por Comuna

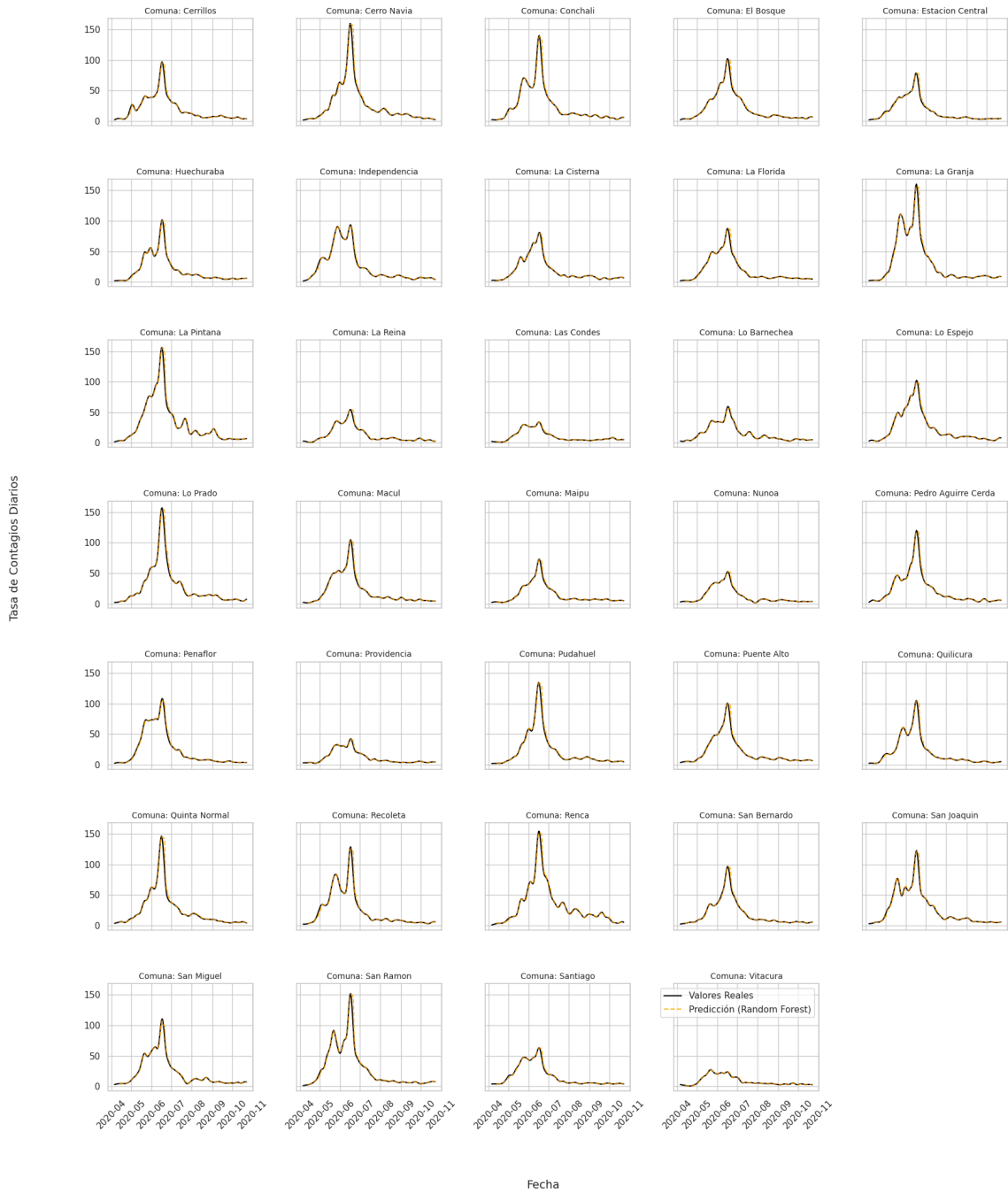


Figura 33. Comportamiento Real de la Tasa de Casos Diarios por Comuna versus Predicciones realizadas con Random Forest utilizando Ventana Deslizante.

Predicciones SVM vs Valores Reales por Comuna

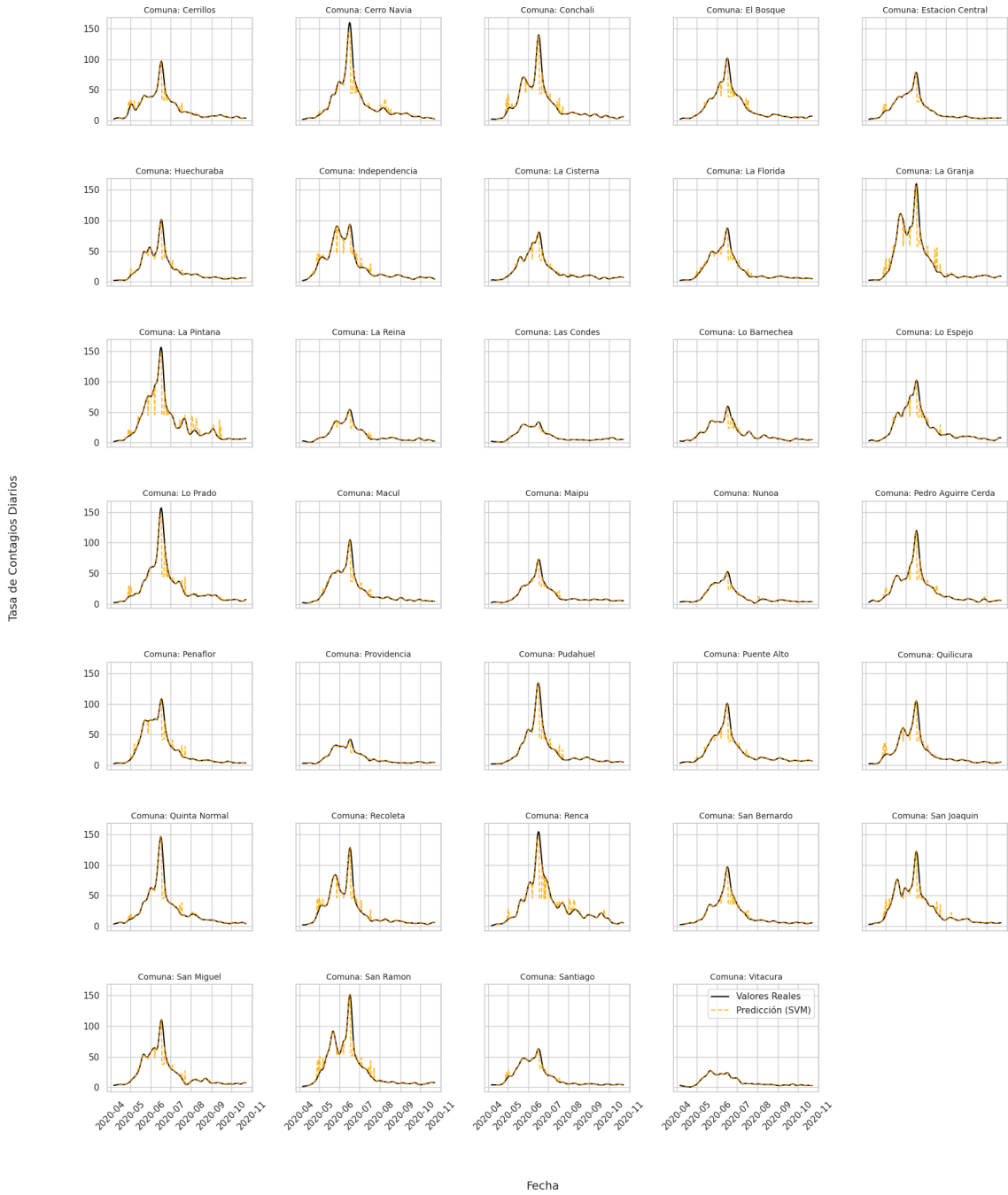


Figura 34. Comportamiento Real de la Tasa de Casos Diarios por Comuna versus Predicciones realizadas con Support Vector Machine utilizando Ventana Deslizante.

Predicciones XGBoost vs Valores Reales por Comuna

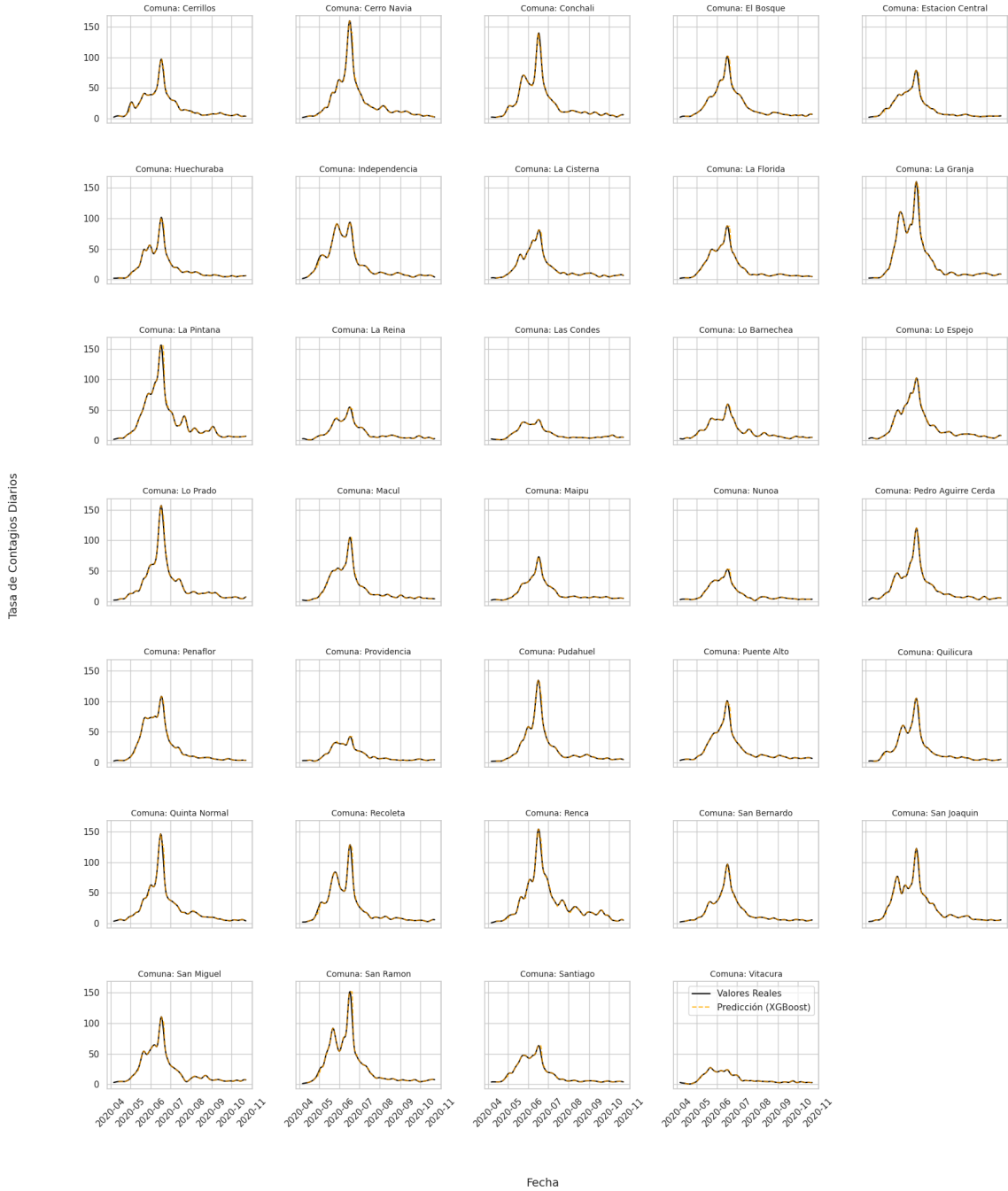


Figura 35. Comportamiento Real de la Tasa de Casos Diarios por Comuna versus Predicciones realizadas con XGBoost utilizando Ventana Deslizante.

Posteriormente, se realizaron ajustes de hiperparámetros de los modelos y validación cruzada con k-fold (Tabla 3) (Figura 36). Al analizar los resultados de métricas se evidenció que Random Forest fue el mejor modelo ya que tuvo un menor MAE (0,61) y RMSE (0,94), lo que indica que realiza predicciones más precisas y estables. Los hiperparámetros óptimos fueron $\text{max_depth}=5$, $\text{n_estimators}=50$, al limitar la profundidad del árbol se reduce el riesgo de sobreajuste.

Por otra parte, XGBoost arrojó un error ligeramente mayor que Random Forest (MAE de 4,30 y RMSE de 5,00). Los hiperparámetros óptimos fueron $\text{learning_rate}=0,01$, $\text{max_depth}=10$, $\text{n_estimators}=50$. En este modelo, la profundidad de los árboles y el bajo learning rate ayudan a captar patrones sin sobreajustarse.

Finalmente, el modelo SVM presentó el desempeño más bajo con un MAE de 4,80 y un RMSE de 7,80, además de una desviación estándar más alta, indicando inestabilidad en las predicciones. Los hiperparámetros óptimos fueron $C=1$, $\gamma=1$ y $\text{kernel}='rbf'$. A pesar de los ajustes, SVM muestra limitaciones significativas para capturar las fluctuaciones rápidas de los datos, siendo menos adecuado para modelar series temporales complejas.

Tabla 3. Comparación de Métricas entre Modelos Support Vector Machine, Random Forest y XGBoost Post Validación Cruzada y Ajuste de Hiperparámetros.

Modelos	MAE	RMSE	Std MAE
Random Forest	0,613	0,939	0,166
Support Vector Machine	4,804	7,798	0,887
XGBoost	4,306	5,002	0,158

MAE: Error Absoluto Medio; RMSE: Raíz del Error Cuadrático Medio; Std: Desviación Estándar.

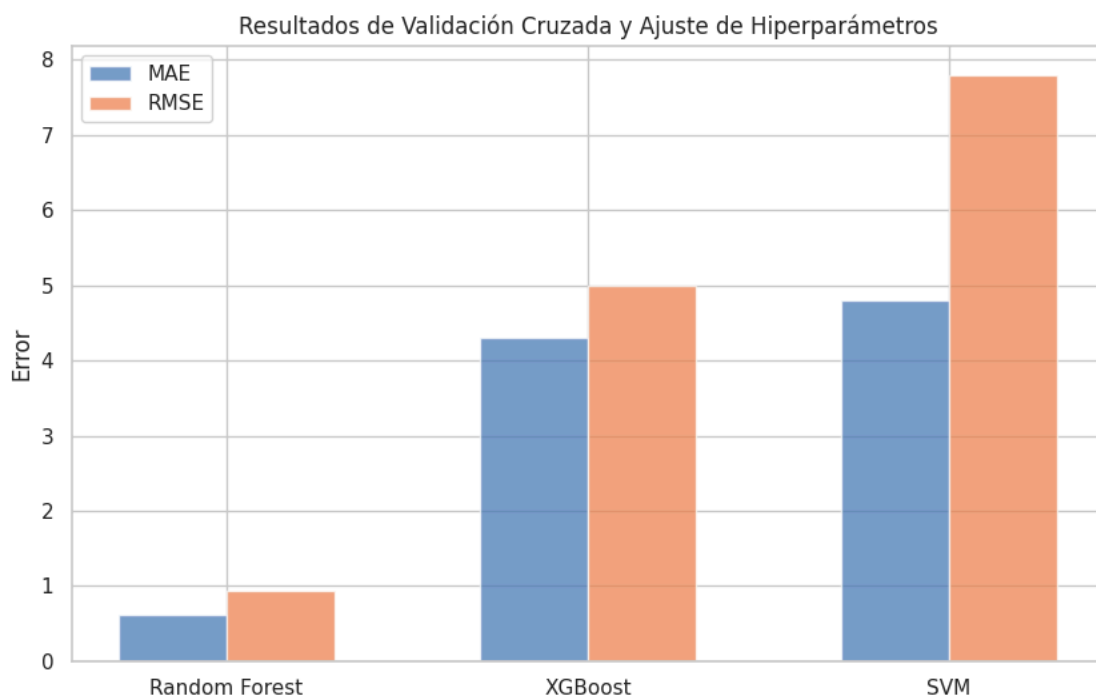


Figura 36. Visualización de la Comparación de Métricas entre Modelos Support Vector Machine, Random Forest y XGBoost Post Validación Cruzada y Ajuste de Hiperparámetros.

Con respecto a la predicción de la serie temporal, se pudo observar que Random Forest captura mejor la tendencia de los datos reales (Figura 37), ajustándose bien a los peaks y a la mayoría de los cambios significativos en las tasas de contagio. Sin embargo, es posible que en algunas comunas se vea un leve

sobreajuste, ya que Random Forest puede aprender patrones locales de los datos que no necesariamente generalizan bien.

En el caso de Support Vector Machine las predicciones fueron más suaves y lineales, lo que indica que no está capturando completamente las fluctuaciones o peaks presentes en los datos reales (Figura 38). Este comportamiento puede ser indicativo de que los hiperparámetros utilizados requieren un ajuste más exhaustivo, o que el modelo no es ideal para este tipo de datos con alta variabilidad.

Por otra parte, XGBoost también muestra un ajuste razonable, pero en algunos casos subestima o sobreestima los peaks, dependiendo de la comuna (Figura 39). Dado que este modelo es altamente flexible, es posible que los resultados se beneficien de una mayor regularización para evitar el sobreajuste.

Dado lo anterior, Random Forest parece ser el modelo más consistente para estas series temporales en términos de ajuste, SVM puede no ser adecuado para este tipo de problema debido a la complejidad de los datos, y XGBoost tiene potencial, pero requiere más ajustes para mejorar la precisión y evitar errores sistemáticos.

Predicciones Random Forest vs Valores Reales por Comuna

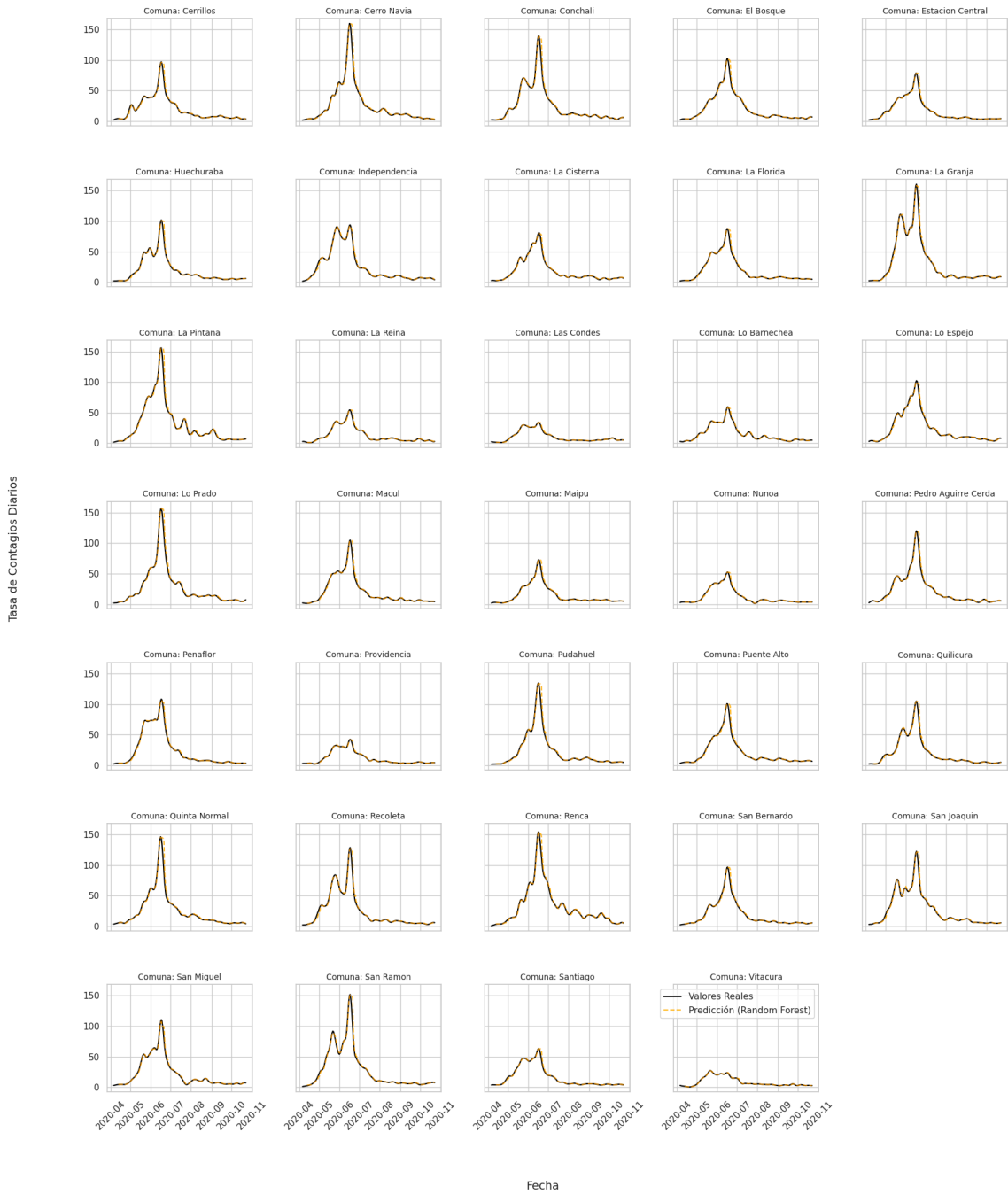


Figura 37. Visualización de las Predicciones Random Forest versus Valores reales por Comuna Posterior a los Ajustes.

Predicciones SVM vs Valores Reales por Comuna

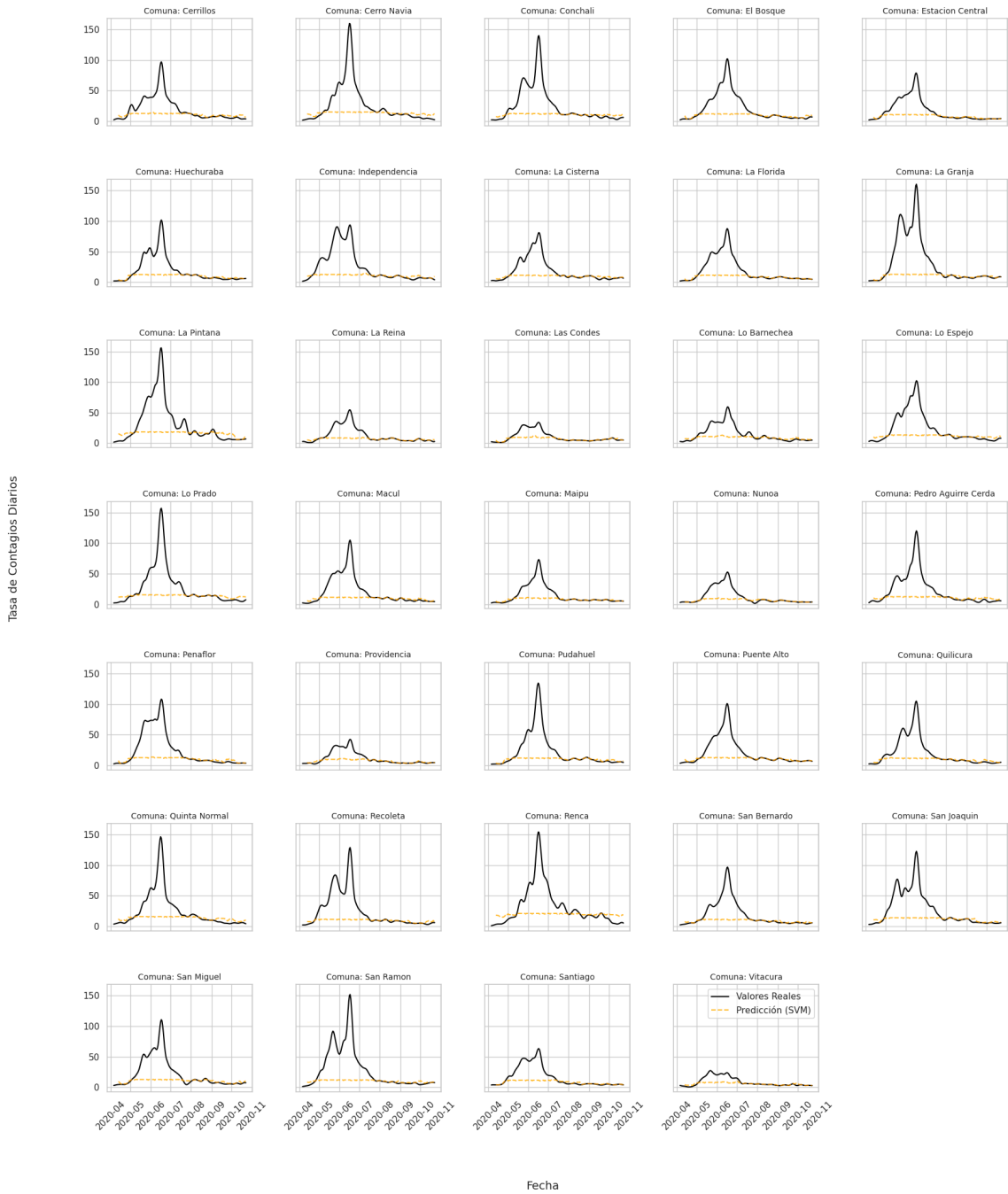


Figura 38. Visualización de las Predicciones Support Vector Machine versus Valores reales por Comuna Posterior a los Ajustes.

Predicciones XGBoost vs Valores Reales por Comuna

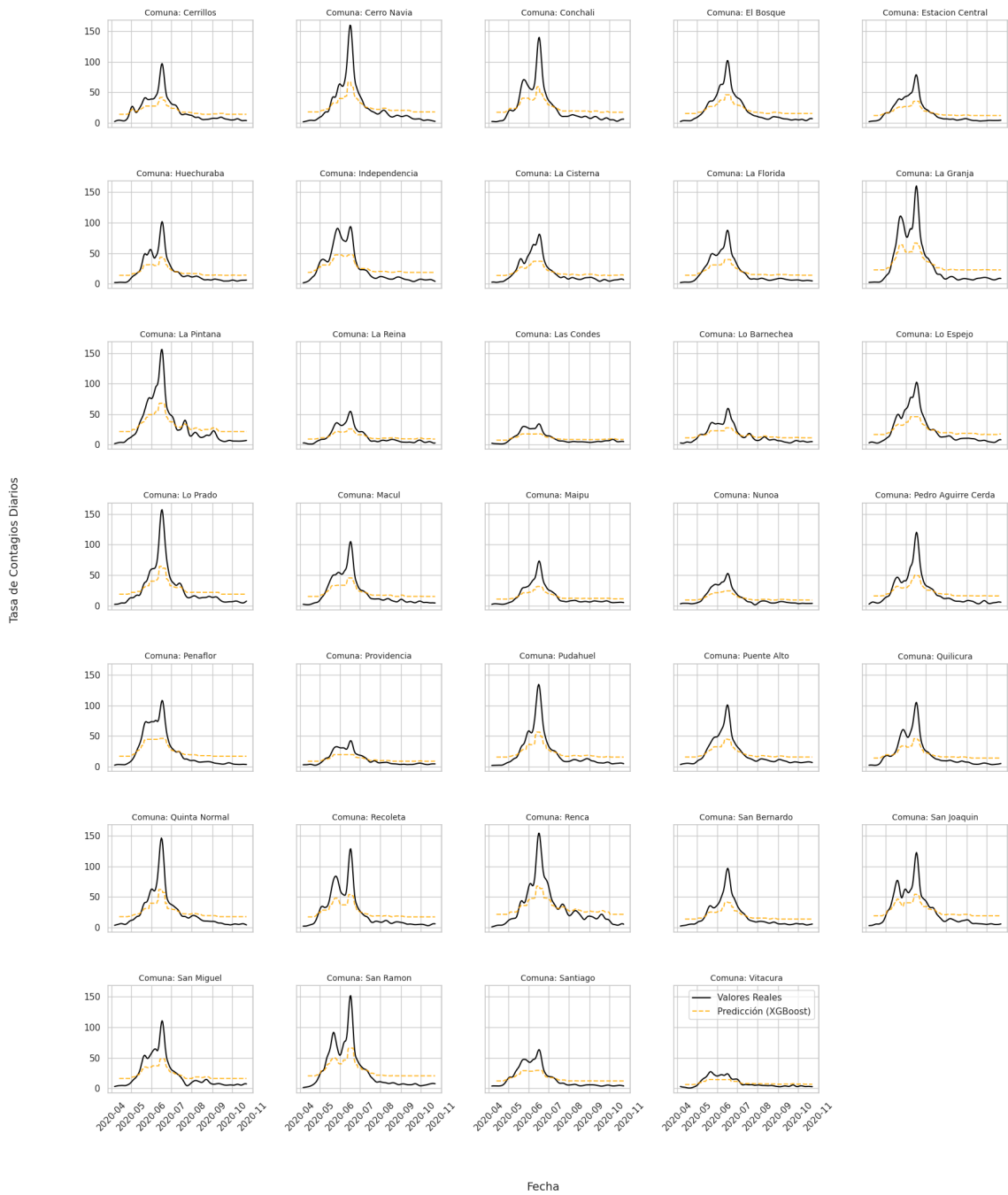


Figura 39. Visualización de las Predicciones XGBoost versus Valores reales por Comuna Posterior a los Ajustes.

Al analizar las curvas de aprendizajes al utilizar Random Forest, se pudo observar que los errores de entrenamiento y prueba convergieron a medida que aumentaba el tamaño del conjunto de entrenamiento, lo que sugiere que el modelo generalizó bien, con baja varianza y sesgo (Figura 40).

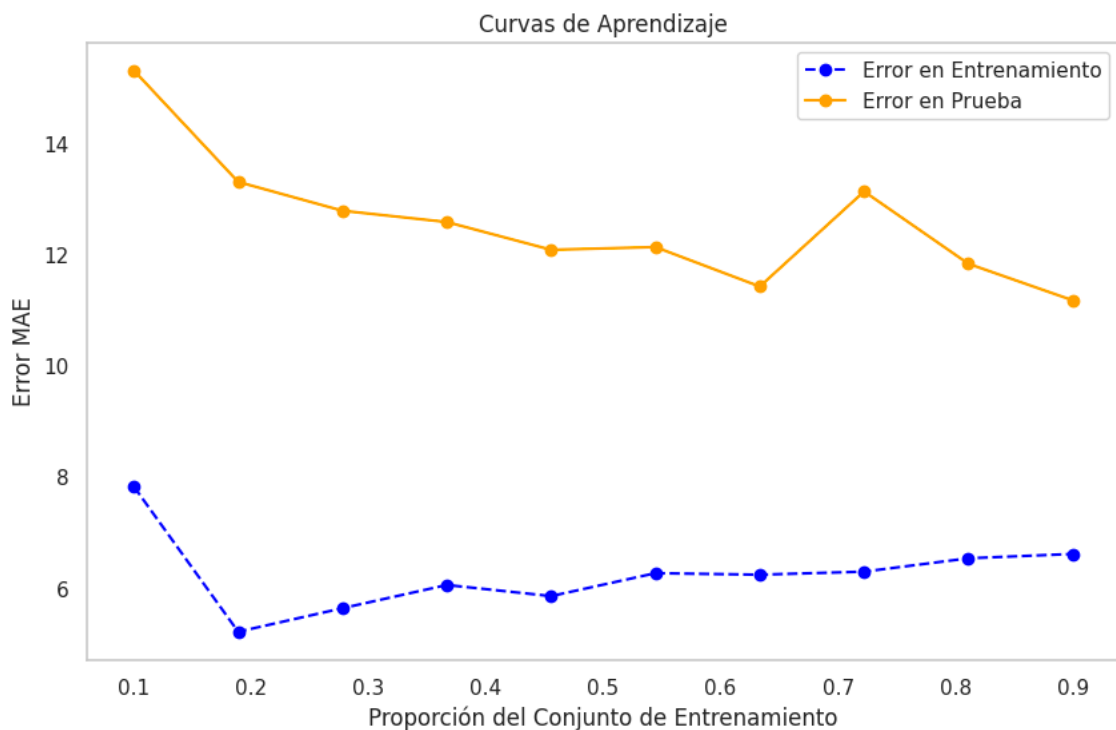


Figura 40. Curvas de Aprendizaje para Random Forest

En cuanto a las curvas de SVM se pudo observar una gran brecha entre los errores de entrenamiento y prueba, especialmente a medida que aumentaba el tamaño del conjunto de entrenamiento, lo que indica sobreajuste; este modelo funcionó bien con los datos de entrenamiento, pero tuvo un rendimiento deficiente en datos no vistos (Figura 41).

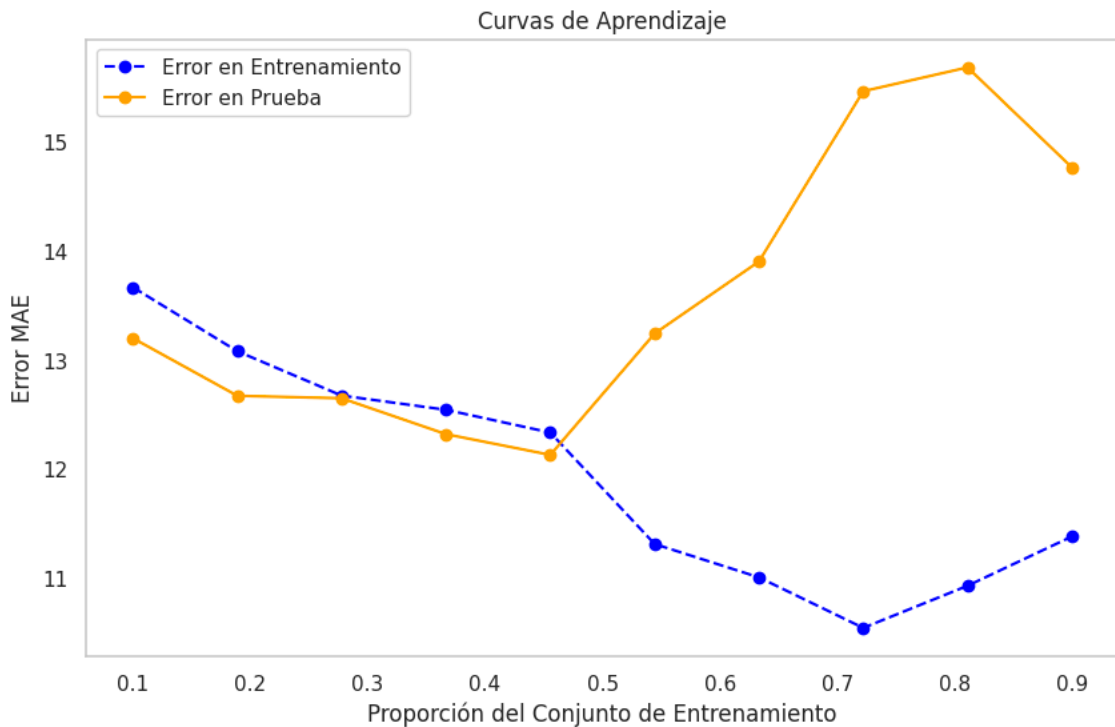


Figura 41. Curvas de Aprendizaje para Support Vector Machine.

Por último, en el caso de XGBoost, los errores de los datos de entrenamiento y prueba mostraron una convergencia moderada, pero aún se observó una brecha (Figura 42). Este modelo parece equilibrar mejor el sesgo y la varianza en comparación con SVM, pero todavía persiste cierto sobreajuste.

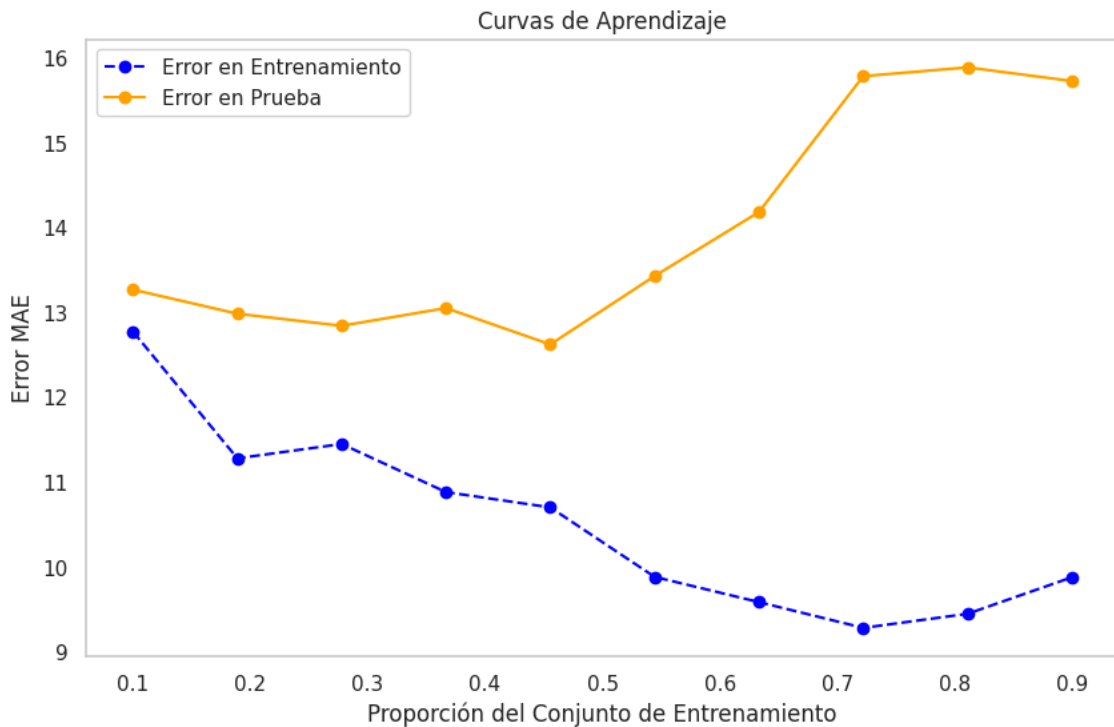


Figura 42. Curvas de Aprendizaje para XGBoost.

En base a los ajustes realizados y mejoras obtenidas en los errores de predicción, las limitaciones inherentes a los Shallow Classifiers dificultan la captura de patrones temporales complejos presentes en los datos analizados. Por lo tanto, el siguiente paso lógico a analizar fue la implementación de modelos más avanzados, como DeepAR.

2.2. Baseline con DeepAR

Con el fin de incorporar la información temporal en las series desde una perspectiva de predicción probabilística, se utilizaron modelos DeepAR para predecir las tasas de contagios diarios de forma univariada en cada comuna,

incorporando como covariables los índices de movilidad interno y externo, el índice de prioridad social (IPS 2020) y porcentaje de educación superior. Cada modelo se configuró con los siguientes hiperparámetros: frecuencia temporal diaria, se probaron ventanas de 7, 15 y 30 días; se utilizó el optimizador Adam, con un learning rate de $1e-4$, un total de 20 epochs y la opción de hibridación activada para mejorar la eficiencia.

Tras los resultados obtenidos se evidenció que el aumento del tamaño de la ventana afectó al MAE, mientras que el RMSE permaneció constante, por lo tanto, para balances entre precisión y longitud del horizonte, se deben considerar ventanas más pequeñas para menor error, sin embargo, el modelo siguió siendo estable con ventanas largas (Tabla 4).

Tabla 4. Comparación de Métricas de Resultados Globales del modelo DeepAR.

Días de Ventana	MAE Promedio	RMSE Promedio
07 días	8,75	2,68
15 días	9,24	2,67
30 días	10,25	2,68

MAE: Error Absoluto Medio; RMSE: Raíz del Error Cuadrático Medio.

Posteriormente, para evaluar el desempeño local de los modelos, se decidió realizar un análisis con dos comunas específicas (Vitacura y La Pintana).

En Vitacura el modelo mostró un rendimiento consistente a lo largo de los distintos escenarios, capturando la baja variabilidad de los datos en esta comuna, donde los valores reales y predichos presentaron una correlación estrecha,

especialmente con la ventana de 30 días, la cual obtuvo el menor MAE (1,07) y RMSE (1,04) (Tabla 5) (Figuras 43, 44 y 45).

Tabla 5. Comparación de Métricas de Resultados del modelo DeepAR en Vitacura.

Días de Ventana	MAE	RMSE
07 días	2,99	1,73
15 días	10,34	3,22
30 días	1,07	1,04

MAE: Error Absoluto Medio; RMSE: Raíz del Error Cuadrático Medio.

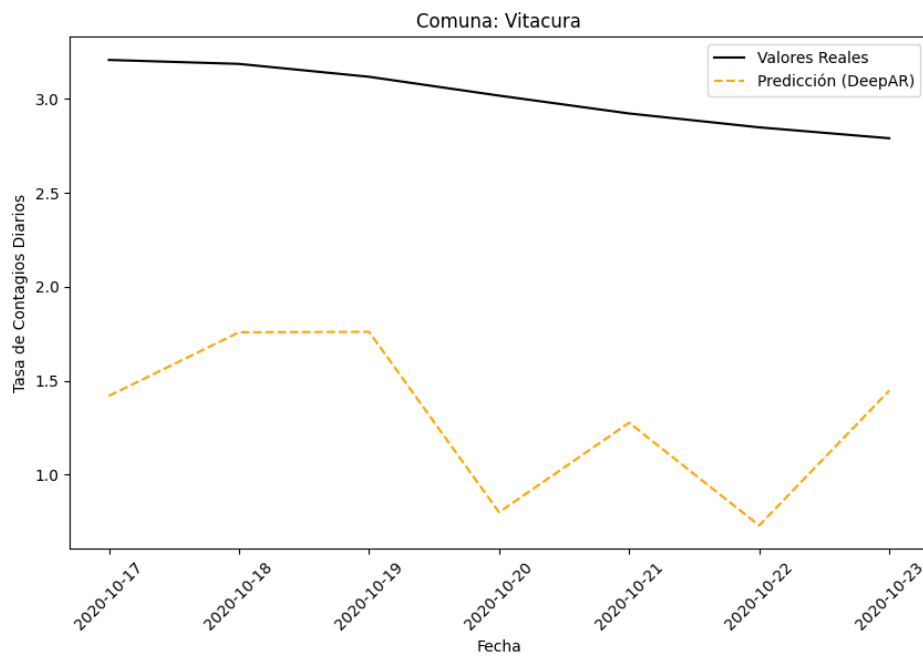


Figura 43. Predicción DeepAR vs Contagios Diarios de Vitacura, ventana 07 días.

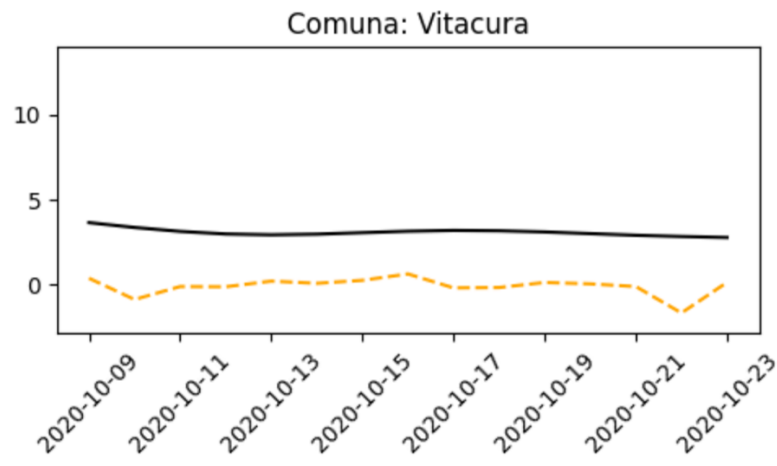


Figura 44. Predicción DeepAR vs Contagios Diarios de Vitacura, ventana 15 días.

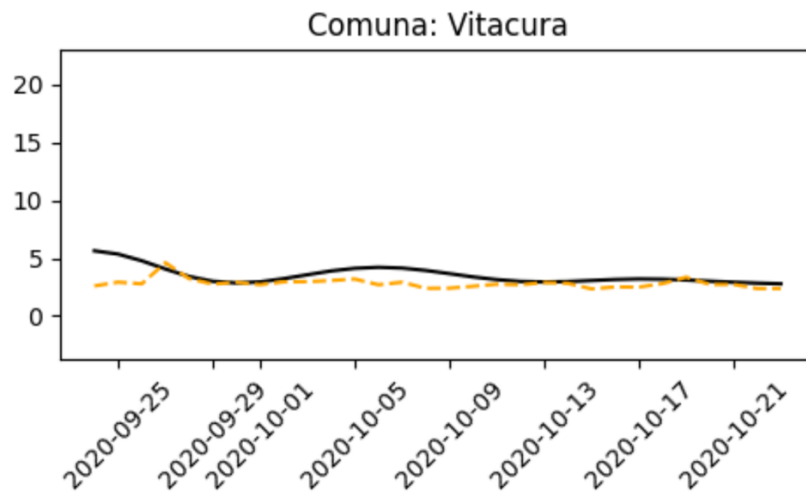


Figura 45. Predicción DeepAR vs Contagios Diarios de Vitacura, ventana 30 días.

Con respecto al comportamiento en La Pintana, el modelo tuvo mayor dificultad para ajustarse debido a la alta variabilidad de datos, especialmente durante los peaks de contagio (Tabla 6) (Figura 46, 47 y 48). El modelo con ventana de 15

días presentó un balance entre precisión y estabilidad, obteniendo un MAE de 6,69 y un RMSE de 2,59.

Tabla 6. Comparación de Métricas de Resultados del modelo DeepAR en La Pintana.

Días de Ventana	MAE	RMSE
07 días	9,95	3,15
15 días	6,69	2,59
30 días	7,98	2,82

MAE: Error Absoluto Medio; RMSE: Raíz del Error Cuadrático Medio.

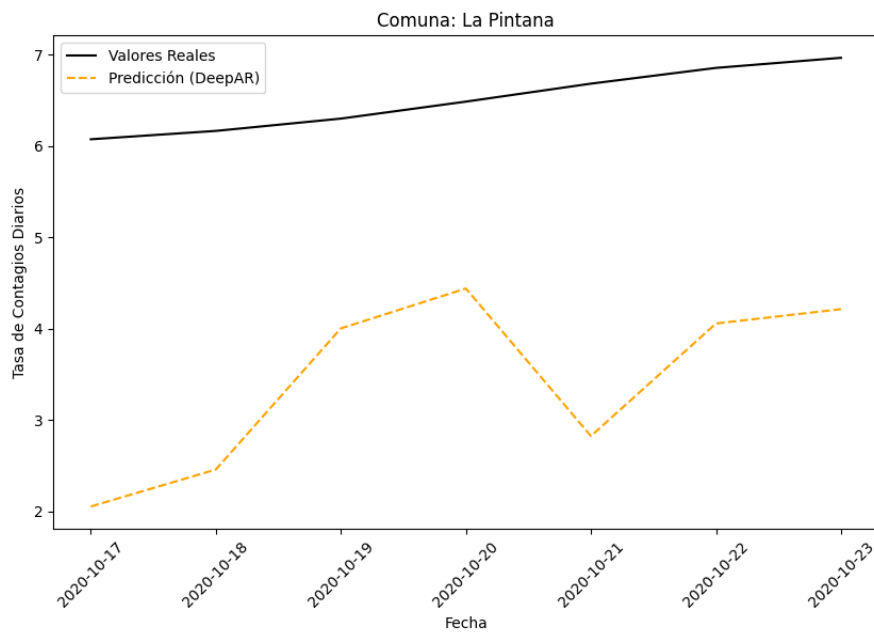


Figura 46. Predicción DeepAR vs Contagios Diarios de La Pintana, ventana 07 días.

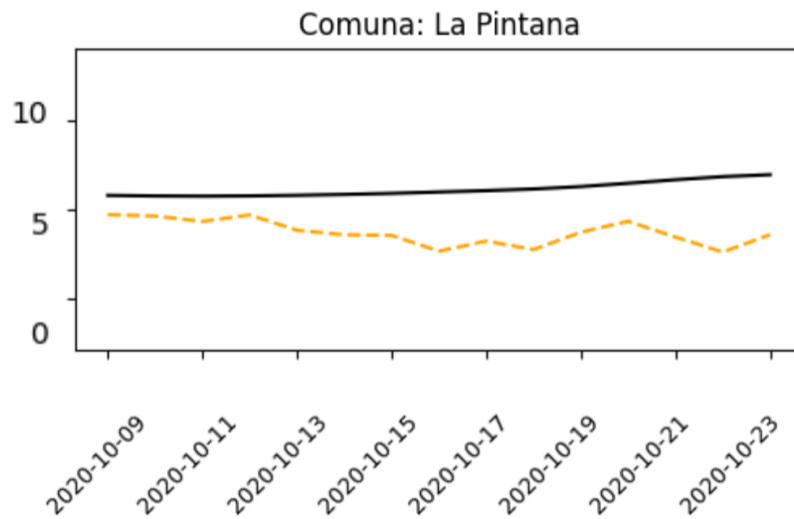


Figura 47. Predicción DeepAR vs Contagios Diarios de La Pintana, ventana 15 días.

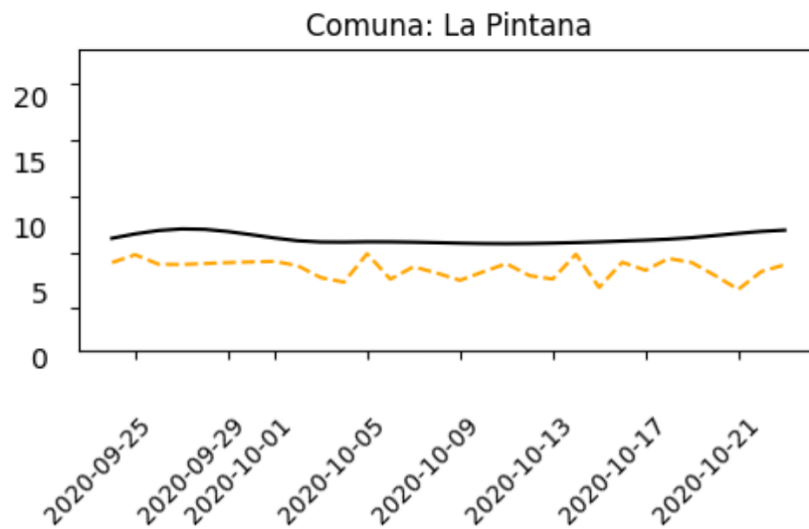


Figura 48. Predicción DeepAR vs Contagios Diarios de La Pintana, ventana 30 días.

Los modelos DeepAR demostraron ser efectivos para capturar tendencias temporales y realizar predicciones en comunas con diferentes niveles de variabilidad en los datos. Sin embargo, su principal desventaja radica en que

están diseñados para series univariadas, por lo que se debe tratar las series de contagios de cada comuna como independientes y realizar predicciones separadas para cada una.

A pesar de la precisión predictiva obtenida, estos modelos carecen de una perspectiva secuencial explícita para capturar transiciones entre estados epidémicos y no permiten capturar el dinamismo ni la interdependencia entre distintas comunas.

Este vacío metodológico, en conjunto con el foco principal de la presente investigación, que busca comprender cómo ocurrió la propagación del virus entre las diferentes comunas, motivó al uso de Hidden Markov Model (HMM), ya que esta técnica permite modelar la evolución temporal del fenómeno y analizar las interacciones subyacentes entre las comunas, ofreciendo una comprensión más profunda de las dinámicas de contagio.

3. Estados Ocultos con Hidden Markov Model

3.1. Determinación de Estados Ocultos de Hidden Markov Model

Para modelar la dinámica de contagio por Covid-19 a nivel comunal, se utilizó un modelo de Hidden Markov Models (HMM) no homogéneo con tres estados ocultos. Inicialmente, el modelo no incluyó una definición previa de los estados, sino que estos fueron interpretados y nombrados tras observar los resultados

obtenidos y analizar el comportamiento de las probabilidades de transición y las distribuciones asociadas a cada estado.

El modelo fue configurado utilizando la biblioteca depmixS4 de R, con los datos de casos diarios ajustados por media móvil de siete días (MM7). La respuesta principal fue definida como el número de casos diarios ajustados, mientras que las probabilidades de transición se modelaron como dependientes de variables explicativas seleccionadas, incluyendo la movilidad interna y externa, así como la identificación de cada comuna.

3.1.1. Evaluación del Número de Estados Ocultos

Para determinar el número adecuado de estados ocultos en el modelo Hidden Markov Model (HMM) No Homogéneo, se realizó un análisis exhaustivo basado en los siguientes criterios de selección:

- Log-Likelihood: Mide qué tan bien el modelo ajusta los datos observados, con valores más altos indicando un mejor ajuste.
- Akaike Information Criterion (AIC): Introduce una penalización por la complejidad del modelo para evitar el sobreajuste.
- Bayesian Information Criterion (BIC): Similar al AIC, pero con una penalización más fuerte, lo que favorece modelos más simples.

Se evaluaron modelos con un rango de estados ocultos (entre 3 y 10 estados), priorizando la interpretabilidad y simplicidad de los resultados, para capturar información relevante a lo largo de la evolución de la pandemia.

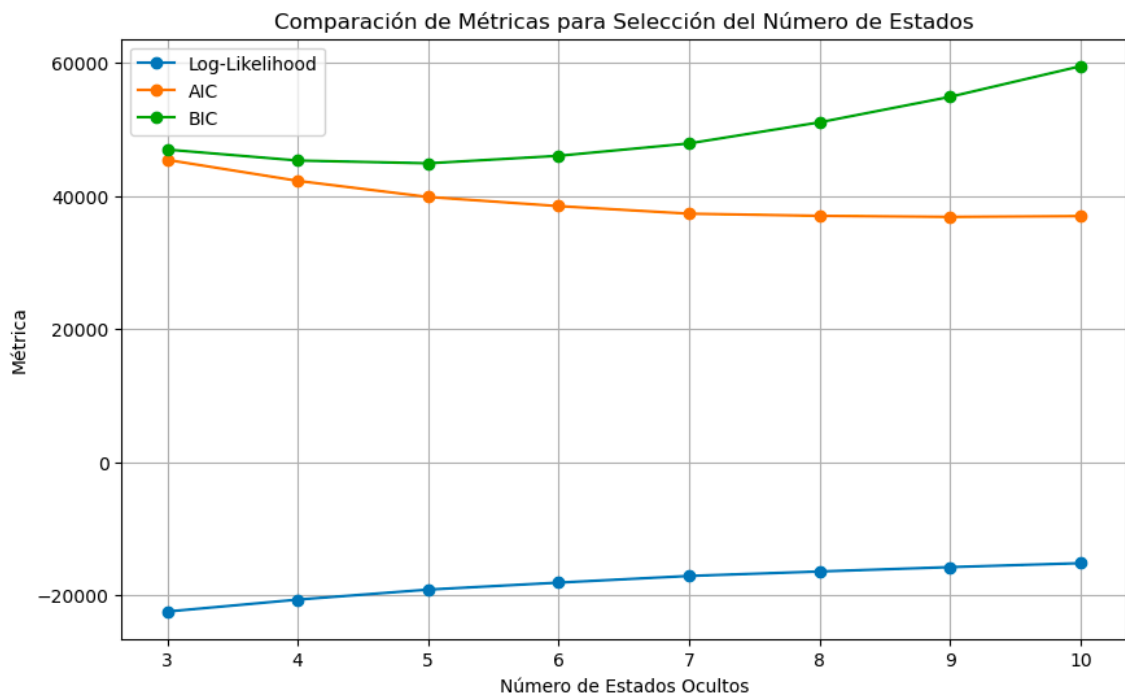


Figura 49. Comparación de Métricas para Selección del Número de Estados Ocultos.

Aunque el BIC fue menor para el modelo con 5 estados, la diferencia con respecto al modelo con 3 estados no fue significativa en términos prácticos, 44.906 versus 46.967 respectivamente. La reducción del BIC en aproximadamente 2.000 puntos no compensa el aumento en la complejidad del modelo. A la vez, un modelo con menos estados es más fácil de interpretar y ajustar a otro conjunto de datos similares, lo que facilita la interpretabilidad.

Al evaluar la convergencia y estabilidad, se observó que las iteraciones necesarias para la convergencia en el modelo con 3 estados fue de 33 iteraciones, con un log-likelihood de -22.487, un AIC de 45.423 y un BIC de 46.967 (Tabla 7), evidenciando una convergencia estable, menor número de iteraciones y más robustez, lo que hace que el modelo sea menos propenso a problemas de sobreajuste. A la vez, las métricas evaluadas no mostraron grandes mejoras al agregar estados adicionales. Por estas razones, se seleccionaron 3 estados ocultos con el fin de equilibrar simplicidad e interpretabilidad.

Tabla 7. Evaluación de métricas de ajuste en función del número de Estados Ocultos en el modelo HMM.

Número de Estados	Log Likelihood	AIC	BIC
3	-22.487,2	45.422,5	46.967,4
4	-20.691,8	42.269,6	45.325,0
5	-19.187,9	39.843,7	44.906,1
6	-18.137,7	38.469,4	46.035,4
7	-17.133,6	37.331,2	47.897,4
8	-16.464,4	37.006,7	51.069,7
9	-16.015,3	37.266,6	55.322,9
10	-15.220,6	36.979,2	59.525,5

AIC: Akaike Information Criterion; BIC: Bayesian Information Criterion

El ajuste del modelo se llevó a cabo mediante un algoritmo de máxima verosimilitud, logrando estimar tanto las probabilidades de transición como las distribuciones asociadas a los estados ocultos. Una vez ajustado, se extrajo la matriz de transición que describe las probabilidades de pasar de un estado oculto a otro, condicionadas por las covariables incluidas en el modelo. Además, se

generó la secuencia de Viterbi, la cual permite identificar la trayectoria más probable de estados ocultos a lo largo del periodo de análisis para cada comuna.

La interpretación de los estados se realizó en esta etapa, asignando los nombres de leve, moderado y severo a los estados 1, 2 y 3, respectivamente, en base a los valores observados y los patrones temporales y espaciales de las comunas.

Posteriormente, se utilizó la secuencia de Viterbi para construir un heatmap con el fin de visualizar la evolución temporal de los estados más probables en cada comuna (Figura 50). En este gráfico, cada celda representa el estado más probable de una comuna en un día específico, donde el tono más claro corresponde al estado 1 (contagio leve), el tono intermedio al estado 2 (contagio moderado) y el tono más oscuro corresponde al estado 3 (contagio severo).

Al visualizar los resultados mediante el heatmap, se logra evidenciar patrones claros en la evolución de los estados de contagio, donde comunas como Lo Espejo, Cerro Navia y La Pintana muestran periodos prolongados en el estado severo, reflejando altos niveles de contagio sostenido, mientras que comunas como Las Condes, Lo Barnechea y Vitacura exhiben mayor predominancia del estado leve, indicando una menor intensidad de contagio en comparación con otras comunas.

Los resultados obtenidos tras la aplicación de HMM no homogéneo, permitieron caracterizar la dinámica temporal y espacial de los estados de contagio en

diferentes comunas, dando paso a la asignación de los nombres para cada estado, lo que aporta claridad interpretativa y facilita una comprensión más profunda de las trayectorias de contagio.

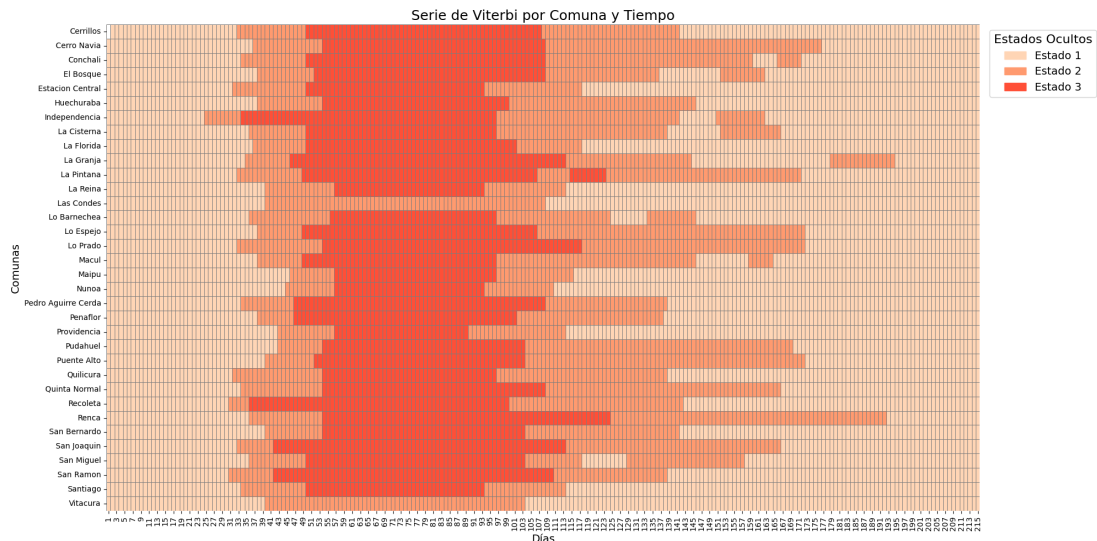


Figura 50. Serie de Viterbi de Estados Ocultos por Comuna y Tiempo.

3.2. Caracterización de los Estados Ocultos de Hidden Markov Model

Posterior al ajuste del modelo y la obtención de la secuencia de Viterbi, se procedió a realizar una caracterización de los tres estados identificado, basada en variables clave relacionadas con el comportamiento de los casos de Covid-19, índices de movilidad y factores sociodemográficos.

Un aspecto importante a considerar, es que la mayoría de las comunas transitaron por los tres estados a lo largo del periodo analizado, a excepción de Las Condes y Vitacura, las cuales no presentaron estado 3 (contagio severo), lo que podría explicar la variación en los resultados observados.

Respecto al comportamiento de los casos de Covid-19, se observan diferencias claras entre los tres estados analizados (Tabla 8). Para los casos diarios brutos, en el primer estado los valores presentan una media de $5,2 \pm 2,7$, en el segundo estado, estos valores aumentan a una media de $15,4 \pm 8,1$, y en el tercer estado se registra un incremento considerable, alcanzando una media de $55,4 \pm 35,4$ casos diarios. Las tasas acumuladas de casos también muestran un comportamiento creciente a lo largo de los tres estados, con medias de 2.511, 2.968 y 2.180, respectivamente, y un rango amplio que refleja la variabilidad entre comunas.

Tabla 8. Estadísticos Descriptivos de Casos Diarios y Tasas Acumuladas de Covid-19 para cada Estado.

	Estado	Mínimo	Máximo	Mediana	Media	DE
Casos Diarios	1	0,00	22,62	4,90	5,17	2,69
	2	1,99	58,75	13,37	15,35	8,09
	3	8,16	231,73	45,84	55,41	35,37
Casos Diarios MM7	1	0,00	10,56	4,99	4,94	2,10
	2	5,56	35,50	13,58	15,09	5,48
	3	16,67	174,18	47,61	55,97	27,83
Tasa de Casos Acumulados	1	0,00	6377,44	2816,80	2511,29	1984,16
	2	123,89	6273,04	3461,96	2968,18	1846,08
	3	316,72	5467,66	1911,99	2179,68	1247,93

MM7: Media Móvil 7 Días; DE: Desviación Estándar.

En relación con los índices de movilidad interna y externa, se observa una tendencia decreciente a medida que se avanza entre los estados (Tabla 9). La movilidad interna en el primer estado presenta una media de $4,2 \pm 2,4$, mientras que en el segundo estado la media disminuye a $3,2 \pm 1,6$ y en el tercer estado

alcanza su valor más bajo, con una media de $2,6 \pm 1,2$. Por otra parte, la movilidad externa sigue un patrón similar, comenzando con una media de $6 \pm 2,3$ en el primer estado, disminuyendo a $5,3 \pm 2$ en el segundo y alcanzando $4,7 \pm 1,6$ en el tercer estado. Estos resultados evidencian una posible relación entre la reducción de la movilidad y el comportamiento de los casos de Covid-19 a lo largo del tiempo.

Tabla 9. Estadísticos Descriptivos de Índices de Movilidad Interna y Externa para cada Estado.

	Estado	Mínimo	Máximo	Mediana	Media	DE
IM Interna	1	0,54	13,12	3,72	4,19	2,38
	2	0,45	9,08	2,90	3,19	1,57
	3	0,40	7,51	2,41	2,62	1,17
IM Externa	1	1,12	14,66	5,80	5,97	2,34
	2	0,97	13,19	5,19	5,25	2,01
	3	1,03	10,40	4,61	4,67	1,59

IM: Índice de Movilidad; DE: Desviación Estándar.

Al analizar las características sociodemográficas para cada estado se observan patrones importantes que permiten describir los estados observados, teniendo en consideración de que en el estado 3 no están incluidas las comunas de Las Condes y Vitacura, lo que también podría influir en los resultados obtenidos (Tabla 10).

En relación al ingreso per cápita comunal, se observa una tendencia decreciente a medida que se avanza de un estado a otro. En el estado 1, donde se encuentran comunas como Las Condes y Vitacura, se alcanza la media más alta con

516.508, mientras que en el estado 2 disminuye a 436.048 y en el estado 3, donde estas comunas no están presentes, desciende a 365.185. La ausencia de comunas con altos ingresos en el estado 3 también podría explicar en parte este comportamiento, indicando que las comunas más vulnerables económicamente tienen mayor probabilidad de encontrarse en estados más críticos.

Por otra parte, el número total de hogares hacinados presenta una tendencia al alza conforme aumenta la severidad del estado. En el estado 1, el promedio fue de 114.134, seguido por 118.240 en el estado 2 y 125.100 en el estado 3. La mayor concentración de hogares hacinados en el estado 3 podría estar vinculada a las condiciones de vulnerabilidad habitacional, lo cual se ve reforzado por la exclusión de comunas con mejores indicadores, como Las Condes y Vitacura.

En cuanto a la variable inmigrantes, se observa una disminución progresiva en los valores medios entre los estados. El estado 1 presentó un promedio de 14.024, mientras que en el estado 2 descendió a 11.163, y en el estado 3 se mantuvo en 11.298. La ausencia de algunas comunas en el estado 3 podría contribuir a que los valores sean comparativamente similares a los del estado 2, sugiriendo que esta variable no varía de manera sustancial entre los estados más severos.

Con respecto al índice de precariedad de la vivienda se observa un aumento progresivo en los valores medios a medida que se avanza en la severidad de los

estados. En el estado 1, la media fue de 2,6, la cual ascendió a 2,8 en el estado 2 y alcanzó su valor máximo en el estado 3 con una media de 3,1.

Por su parte, el porcentaje de educación superior mantuvo una tendencia decreciente entre estados. En el estado 1, se registra un promedio de 23, en el estado 2 disminuye a 18,8, y en el estado 3 desciende a 16,3, lo que también podría estar asociado a la ausencia de comunas con mayores niveles educativos en el estado 3, las cuales se caracterizan por tener porcentajes de educación superior significativamente más altos.

A la vez, el índice de prioridad social, que integra componentes de vulnerabilidad económica y social, presentó un incremento en los valores medios a medida que aumenta la severidad del estado, con una media de 60,3 en el estado 1, 65,2 en el estado 2 y 69,5 en el estado 3. La ausencia de comunas menos vulnerables en el estado 3 contribuye a reforzar este patrón, donde predominan comunas con mayores niveles de vulnerabilidad social.

Finalmente, el promedio de personas por hogar muestra un leve aumento entre estados. En el estado 1, el valor medio es de 2,9, en el estado 2 asciende a 3, y en el estado 3 se mantiene en 3. La presencia de valores similares en este caso se podría explicar debido a que el promedio de personas por hogar no refleja hacinamiento, debido a que pueden haber varias personas en un hogar de gran tamaño, por lo tanto, este indicador no rescata componentes de vulnerabilidad social.

Tabla 10. Estadísticos Descriptivos de Variables Sociodemográficas para cada Estado.

	Estado	Mínimo	Máximo	Mediana	Media	DE
Índice Per Cápita Comuna	1	211826,83	1659485,00	324016,69	516508,37	386949,42
	2	211826,83	1659485,00	297526,34	436047,89	339480,50
	3	211826,83	1300158,10	290623,38	365184,64	200545,52
Total de Hogares Hacinados	1	4770,00	376908,00	98088,00	114133,94	83118,87
	2	4770,00	376908,00	98088,00	118239,89	82529,94
	3	15738,00	376908,00	113165,00	125100,13	76426,05
Inmigrantes	1	2127,00	110389,00	9219,00	14024,41	20342,76
	2	2127,00	110389,00	6520,00	11163,19	15366,16
	3	2127,00	110389,00	6108,00	11298,01	17401,35
Índice de Precariedad de la Vivienda	1	0,09	5,05	2,67	2,56	1,53
	2	0,09	5,05	3,09	2,81	1,40
	3	0,21	5,05	3,11	3,05	1,29
Porcentaje de Educación Superior	1	4,20	59,70	16,60	23,04	16,54
	2	4,20	59,70	12,90	18,75	14,68
	3	4,20	59,70	12,70	16,31	11,80
Índice de Prioridad Social	1	7,94	89,29	67,64	60,28	21,63
	2	7,94	89,29	71,67	65,19	20,29
	3	24,91	89,29	72,85	69,51	13,96
Personas Hogar Promedio	1	1,68	3,55	3,10	2,91	0,48
	2	1,68	3,55	3,13	3,01	0,42
	3	1,68	3,55	3,15	3,02	0,42

DE: Desviación Estándar.

Pese a las limitaciones que tiene el hecho de considerar el cálculo de estadígrafos de variables sociodemográficas para cada estado, considerando que la mayoría de las comunas pasan por los 3 estados, las características observadas igual permiten distinguir a modo general diferencias importantes entre los estados, donde se observa que los estados 2 y 3 que presentan un mayor contagio, están marcados por menores ingresos per cápita, mayores niveles de hacinamiento, mayor precariedad de la vivienda y menores

porcentajes de educación superior, además de un incremento en los niveles de prioridad social. Sin embargo, la ausencia de comunas como Las Condes y Vitacura en el estado 3 también juega un papel relevante, influyendo en las diferencias observadas, ya que estas comunas tienden a presentar indicadores más favorables en comparación con las demás.

La caracterización de los 3 estados ocultos considerando la distribución de los casos y tasas acumuladas de Covid-19, los índices de movilidad interna y externa y las variables sociodemográficas, se puede visualizar a continuación en la Figura 51.

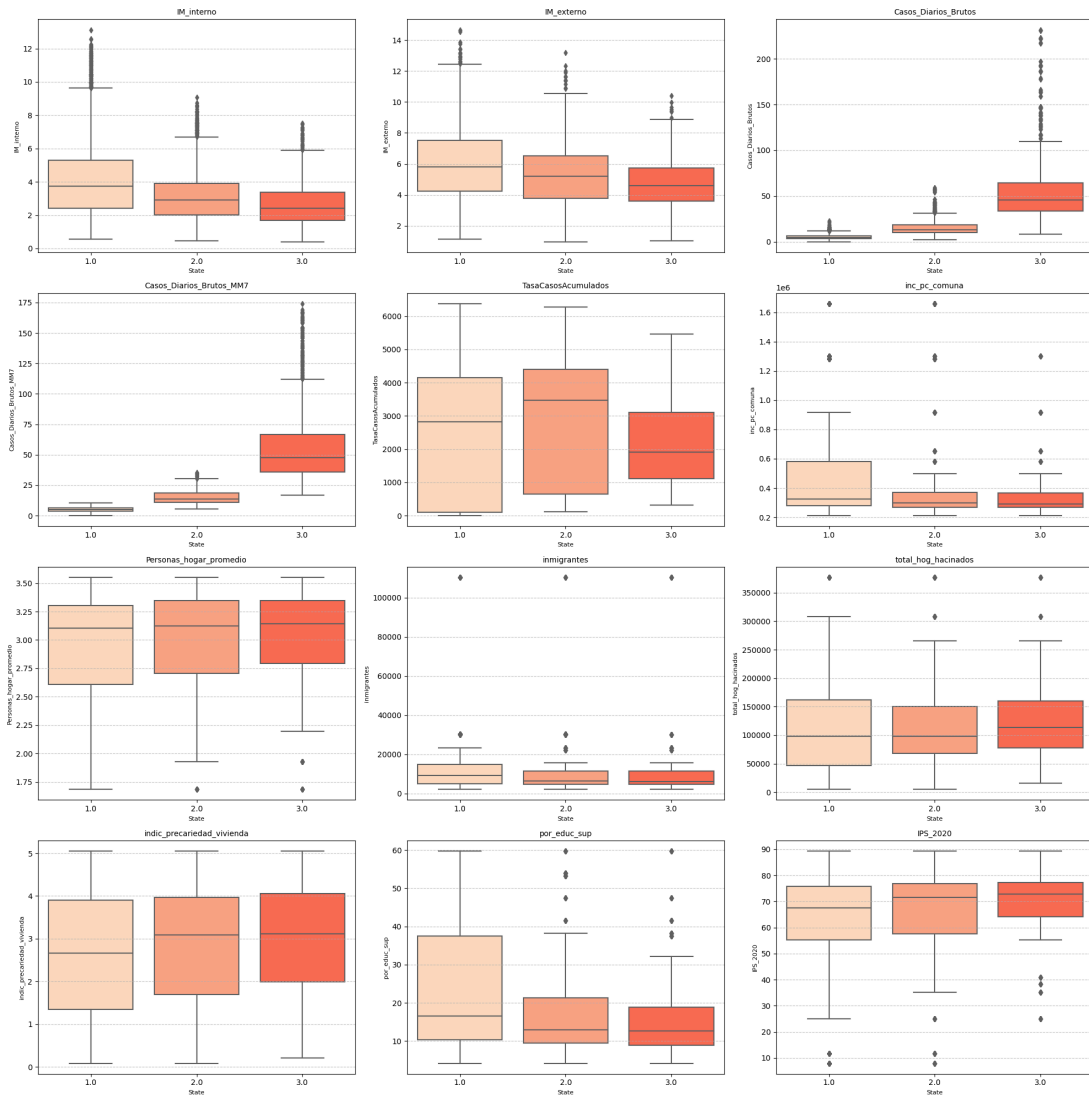


Figura 51. Resumen de estadísticos descriptivos de todas las variables por cada Estado Oculto.

Debido a que casi todas las comunas pasan por los tres estados, pero la cantidad de días que permanecen en cada uno varía considerablemente, se decidió analizar la distribución temporal de los estados ocultos para cada comuna, con el fin de aproximarse un poco más al comportamiento de los estados (Figura 52).

El análisis de la cantidad de días que cada comuna permaneció en los distintos estados permitió identificar variaciones importantes en la distribución temporal de los estados ocultos, ya que, aunque casi todas las comunas atraviesan los tres estados, la cantidad de días en que permanecen en cada uno de ellos es heterogénea y muestra patrones distintivos entre las comunas.

En el estado 1, que corresponde a la menor intensidad de contagio, se observa que la mayoría de las comunas pasan una cantidad considerable de días en este estado, destacando comunas como Las Condes, Lo Barnechea y Vitacura, que presentan los valores más altos, superando los 140 días. Esto es consistente con sus mejores indicadores socioeconómicos y su menor exposición a situaciones de vulnerabilidad, lo que podría estar asociado a una transición más lenta hacia estados de mayor severidad. Otras comunas, como Estación Central, La Cisterna y La Florida, también tienen una cantidad considerable de días en este estado, aunque en menor proporción.

Por otro lado, en el estado 2, que corresponde a una intensidad intermedia, la distribución de días fue más equilibrada. Comunas como Puente Alto, Cerro Navia y Renca muestran una mayor cantidad de días en este estado, superando en algunos casos los 80 días.

Finalmente, en el estado 3, que representa la mayor intensidad, la cantidad de días varía significativamente entre comunas, destacando comunas como La Pintana, La Granja, Renca y San Joaquín, las cuales permanecieron entre 60 y

70 días en este estado. Estas comunas, que se caracterizan por mayores niveles de vulnerabilidad, evidencian una permanencia prolongada en el estado más crítico. A la vez, tal como se mencionó anteriormente, las comunas de Las Condes y Vitacura no registran días en el estado 3, lo cual repercute directamente en los resultados y refuerza las diferencias observadas en las variables sociodemográficas y de comportamiento de los casos.

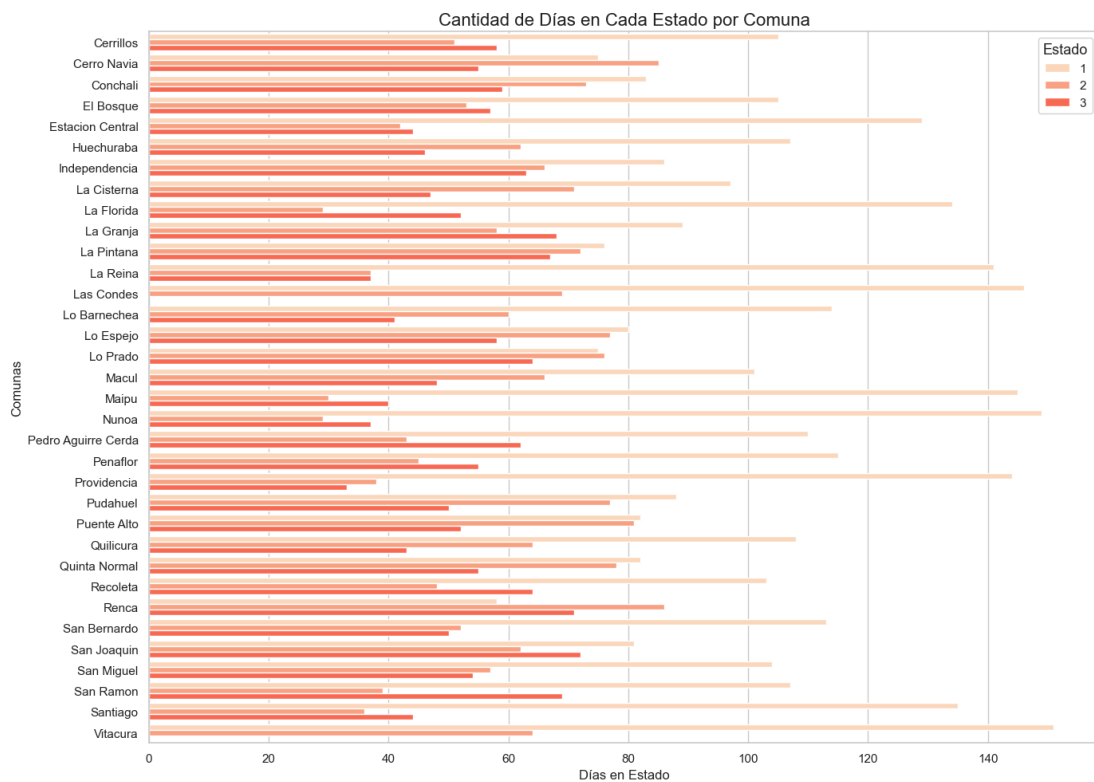


Figura 52. Distribución de la Cantidad de Días por Estado Oculto para cada Comuna.

Para tener una mayor aproximación al comportamiento de las variables en cada estado, se realizó nuevamente el cálculo de los estadígrafos pero considerando la asignación de peso distinto según la cantidad de días que estuvo cada comuna

en un estado. Al analizar visualmente las distribuciones de las variables considerando esta ponderación, se puede observar un comportamiento similar al descrito anteriormente utilizando los datos reales, sin embargo, existen unas variaciones sutiles en el comportamiento de algunas variables (Figura 53).

En el caso de los casos diarios y casos diarios MM7, el comportamiento general entre los estados se mantiene, donde a medida que se avanza del estado 1 al estado 3, se observa un aumento en los valores de casos, no obstante, en los datos ponderados, los rangos se ajustan ligeramente, especialmente en el estado 3, donde los valores parecen menos dispersos. Este ajuste podría explicarse por la distribución desigual de días en estado severo entre las comunas.

En cuanto a los índices de movilidad interna y externa, la tendencia descendente entre estados se mantiene tras el cálculo ponderado, pero se observa una ligera variación en las medianas de movilidad interna y externa en el estado 1 y 3.

Las variables sociodemográficas, como el promedio de personas por hogar y el porcentaje de educación superior, también muestran ligeros ajustes en los valores ponderados. El promedio de personas por hogar se mantiene más alto en el estado 3, mientras que el porcentaje de educación superior disminuye en ese estado.

Finalmente, en el caso del índice de prioridad social y la precariedad de la vivienda, los resultados ponderados confirman que en el estado 3 se concentran

las comunas con mayores niveles de vulnerabilidad, aunque los valores de dispersión son menores en comparación con los datos originales, lo que sugiere que las comunas que acumularon más días en el estado severo comparten características socioeconómicas similares, lo que contribuye a la persistencia del estado de mayor contagio.

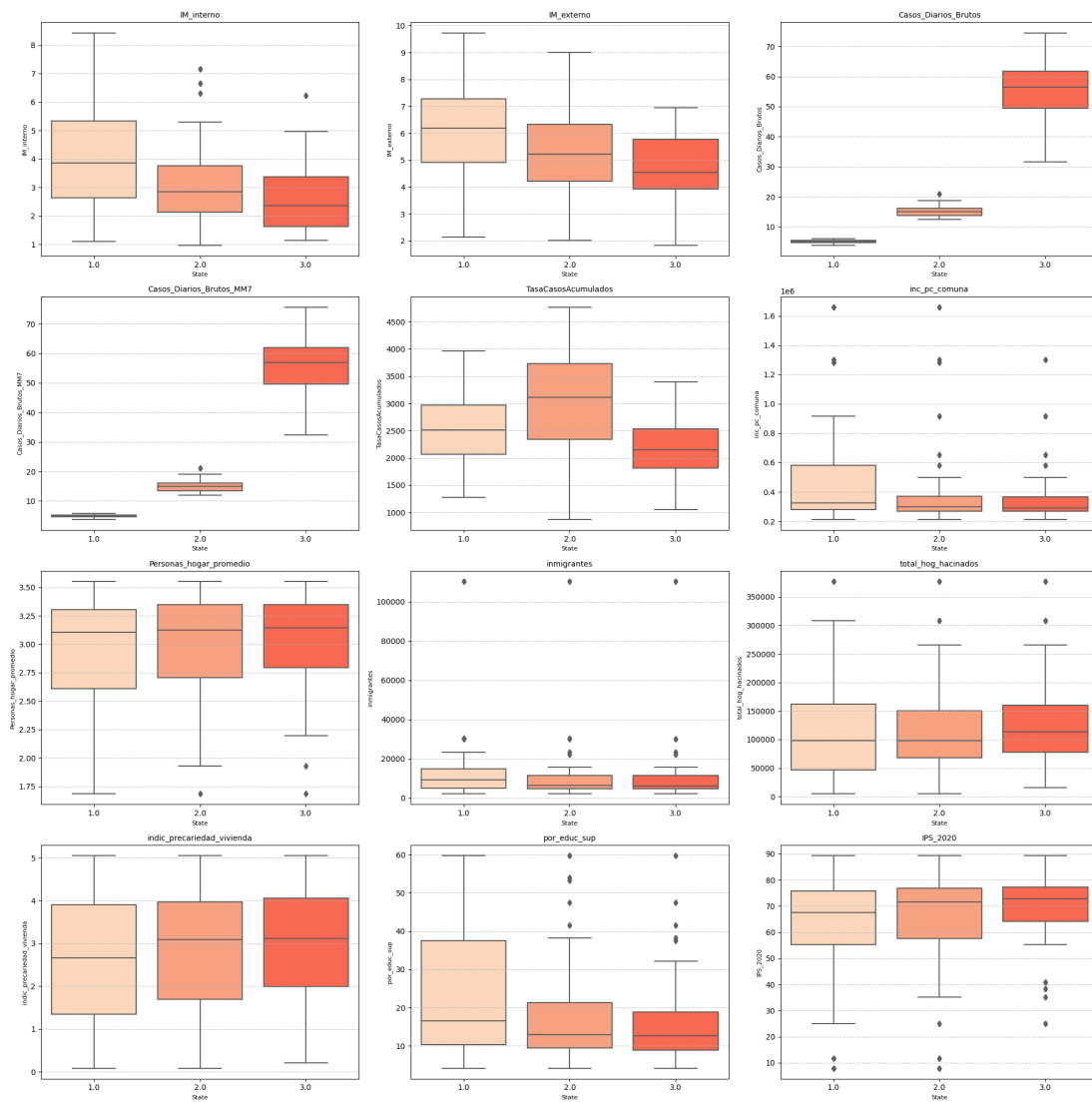


Figura 53. Resumen de estadísticos descriptivos de todas las variables por cada estado considerando cálculos ponderados por cantidad de días en cada estado oculto.

3.3. Correlación entre Casos Diarios Covid-19 y Movilidad Interna por cada Estado Oculto de Hidden Markov Model

Para analizar la correlación entre los casos diarios de Covid-19 y los índices de movilidad interna, se realizó una regresión lineal para cada estado, la cual fue visualizada mediante gráficos de dispersión, donde el tamaño de cada burbuja refleja el ingreso per cápita por comuna y el color representa el índice de precariedad de la vivienda de la comuna.

3.3.1. Correlación entre Casos Diarios Covid-19 y Movilidad Interna en Estado 1: Contagio Leve

Al visualizar la correlación entre los casos diarios de Covid-19 y el índice de movilidad interna en el estado 1 (contagio leve), ambas variables normalizadas, se observa una tendencia negativa moderada, lo que sugiere que a medida que aumenta la movilidad interna, los casos diarios tienden a disminuir (Figura 54). Esta relación está representada por la recta de ajuste lineal en rojo, cuya pendiente es negativa, con un valor de R^2 de 0,17, lo que indica que el modelo lineal explica aproximadamente un 17% de la variabilidad en los casos diarios.

En cuanto a las características de las comunas, se observa que comunas con mayores ingresos, como Vitacura, Providencia y Las Condes, tienden a ubicarse hacia la derecha del gráfico, es decir, presentan mayores niveles de movilidad interna, sin embargo, estas mismas comunas reportan valores bajos de casos

diarios, lo que podría estar asociado con un menor índice de precariedad en la vivienda, como se refleja en los puntos de colores más oscuros en la escala y con un mayor ingreso per cápita que se evidencia en el tamaño de los puntos.

Esto contrasta con comunas como San Ramón y La Granja, que muestran índices de precariedad de la vivienda más elevados, representados por los colores más claros, menor ingreso per cápita, evidenciado por el tamaño de los puntos, y niveles de movilidad interna cercanos a cero, pero con valores más altos de casos diarios.

Por otra parte, la mayoría de las comunas tienden a concentrarse entre niveles bajos y moderados de movilidad interna, con una dispersión considerable en los casos diarios. Por ejemplo, comunas como Renca y Cerro Navia se encuentran en posiciones con baja movilidad y bajos casos, mientras que comunas como Puente Alto y La Florida exhiben niveles más altos de casos a pesar de una movilidad intermedia.

En resumen, la relación entre la movilidad interna y los casos de Covid-19 en este estado sugiere una correlación inversa moderada, donde las comunas con mayores niveles de movilidad, asociadas en general a mayores ingresos y menor precariedad en la vivienda, reportan menores valores de casos diarios. No obstante, existen excepciones notables que podrían explicarse por otros factores contextuales, como la densidad poblacional y las condiciones habitacionales específicas.

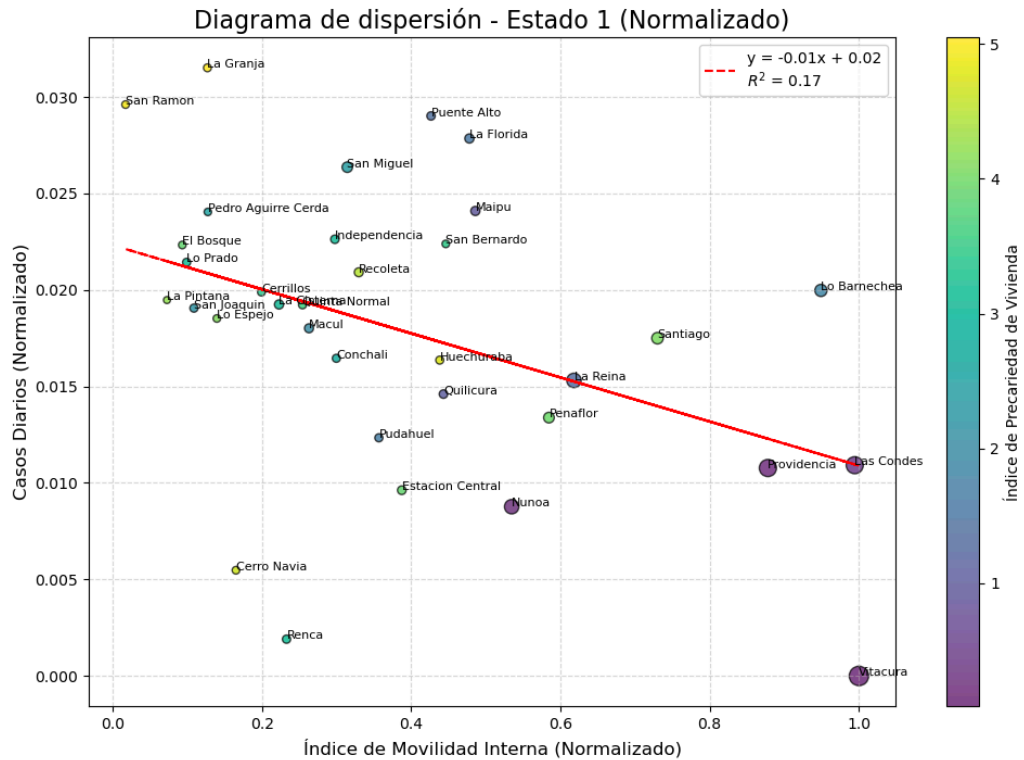


Figura 54. Correlación entre Casos Diarios Covid-19 y Movilidad Interna en Estado 1 de Contagio Leve.

3.3.2. Correlación entre Casos Diarios Covid-19 y Movilidad Interna en Estado 2: Contagio Moderado

En el estado 2, que corresponde a un nivel moderado de contagio, se observa una relación positiva entre los casos diarios y el índice de movilidad interna, donde a medida que aumenta la movilidad interna, los casos diarios también tienden a incrementarse, lo cual está representado por una recta de ajuste lineal con pendiente positiva (Figura 55). El valor de R^2 es de 0,20, lo que indica que aproximadamente un 20% de la variabilidad en los casos diarios puede ser explicada por los niveles de movilidad interna.

Al analizar las comunas, se aprecia una distribución más dispersa de los puntos en comparación con el estado anterior, lo que refleja una mayor variabilidad entre los casos diarios y la movilidad. Comunas como Vitacura y Las Condes se posicionan hacia la derecha del gráfico, mostrando niveles más altos de movilidad interna, así como valores más elevados de casos diarios normalizados en comparación con otras comunas.

Por otro lado, comunas como Renca, Santiago y Maipú también muestran niveles intermedios a altos de casos diarios, a pesar de que sus niveles de movilidad interna varían considerablemente.

En contraste, comunas con menor movilidad interna, como San Ramón, El Bosque y San Joaquín, registran casos diarios relativamente bajos, lo que sugiere una posible asociación inversa en ciertos contextos, aunque no tan pronunciada como en el estado 1. En estas comunas, el índice de precariedad de la vivienda es más elevado, lo cual se refleja en los colores más claros de las burbujas.

En general, la tendencia positiva entre los casos diarios y la movilidad interna en este estado podría indicar que el incremento de la movilidad está asociado a una mayor propagación del virus en niveles moderados de contagio. Sin embargo, esta relación no es uniforme entre todas las comunas, ya que factores como el ingreso per cápita y la precariedad de la vivienda parecen desempeñar un papel importante en la distribución de los casos.

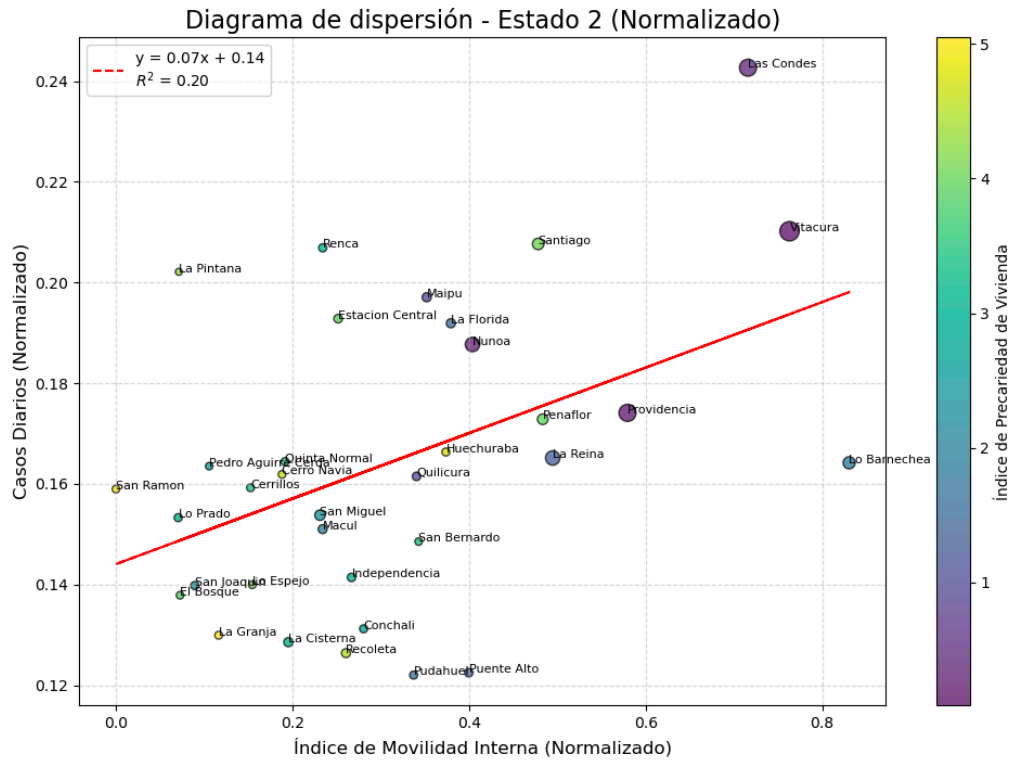


Figura 55. Correlación entre Casos Diarios Covid-19 y Movilidad Interna en Estado 2 de Contagio Moderado.

3.3.3. Correlación entre Casos Diarios Covid-19 y Movilidad Interna en Estado 3: Contagio Severo

En el estado 3, correspondiente al estado de contagio severo, se observa una relación negativa entre los casos diarios y el índice de movilidad interna, donde a medida que el índice de movilidad interna aumenta, los casos diarios tienden a disminuir de manera pronunciada, como lo refleja la pendiente negativa de la recta ajustada (Figura 56). Esta relación es más marcada en comparación con los estados anteriores, con un coeficiente de determinación mayor, lo que indica

una mejor explicación de la variabilidad de los casos diarios a partir de la movilidad interna.

Las comunas que presentan los mayores valores de casos diarios se concentran en la parte superior izquierda, donde la movilidad interna es más baja. Esto incluye comunas como La Granja, San Ramón, La Pintana y Cerro Navia. Estas comunas también tienden a mostrar índices de precariedad de vivienda más elevados (color puntos) e ingresos per cápita más bajos (tamaño puntos), lo que sugiere una mayor vulnerabilidad en estas áreas.

A medida que la movilidad interna aumenta, las comunas tienden a situarse hacia la parte inferior derecha del gráfico, donde los casos diarios son considerablemente menores. Comunas como Providencia, Ñuñoa, y Barnechea destacan en este extremo, caracterizándose por niveles más altos de movilidad interna y, simultáneamente, índices de precariedad de vivienda más bajos.

En cuanto al tamaño de los puntos, que representa el ingreso per cápita comuna, se mantiene una tendencia similar a los estados anteriores. Comunas con menor movilidad y mayores casos diarios, como La Granja y San Ramón, tienden a tener un menor índice per cápita. Por otro lado, comunas como Barnechea y Providencia, que exhiben niveles más altos de movilidad interna y menores casos diarios, presentan puntos de mayor tamaño, reflejando su mayor ingreso per cápita.

En resumen, en el estado de contagios severos, la relación negativa entre la movilidad interna y los casos diarios es más pronunciada, destacando que las comunas con mayores niveles de precariedad y menor movilidad concentran los valores más altos de contagios. Al mismo tiempo, las comunas con mayor movilidad y mejores condiciones sociodemográficas tienden a tener una menor incidencia de casos en este estado.

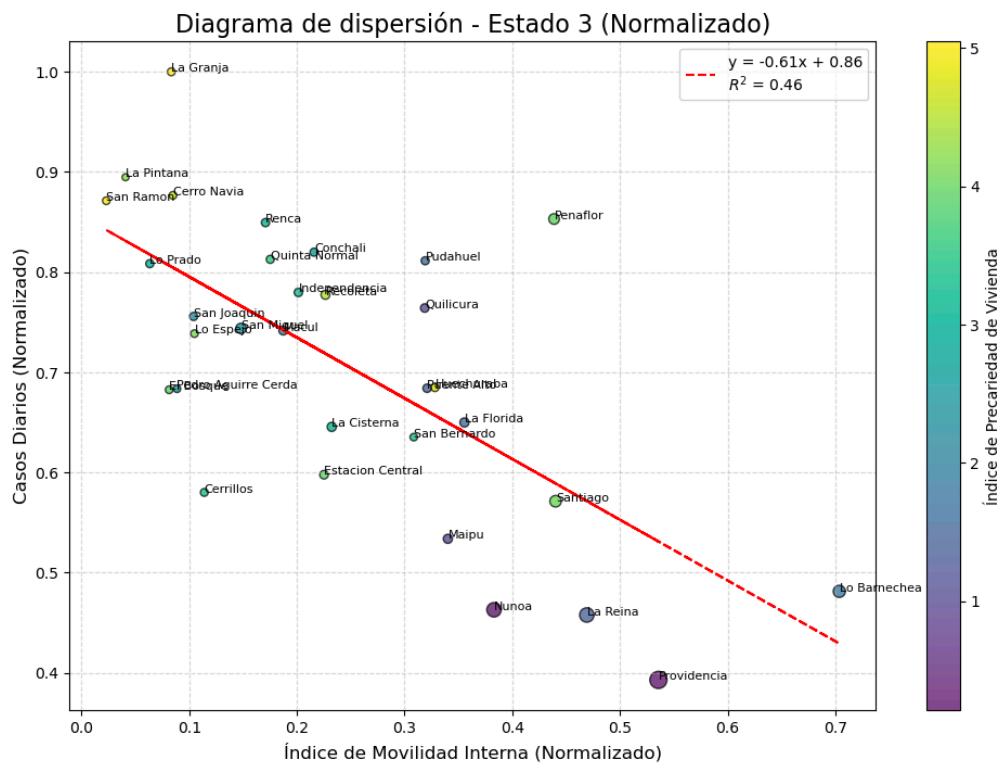


Figura 56. Correlación entre Casos Diarios Covid-19 y Movilidad Interna en Estado 3 de Contagio Severo.

3.3.4. Comparación de la Correlación entre Casos Diarios Covid-19 y Movilidad Interna previo y posterior a Periodo de Contagio Severo

Para poder obtener un mayor acercamiento del comportamiento de las correlaciones, se decidió analizar la correlación entre las mismas variables pero abarcando dos periodos distintos, el periodo previo al estado 3, considerando esto como proxys de la primera fase de la pandemia correspondiente a la fase de crecimiento moderado, y luego el periodo posterior al estado 3, que sería el proxys de la fase posterior al crecimiento exponencial, donde se empieza a estabilizar el aumento de casos (Figura 57).

Al realizar en análisis considerando la data particionada, se observó un cambio en el comportamiento de la correlación entre casos y movilidad interna para el estado 1 y 2, siendo más evidente la diferencia en el estado 2.

En el estado 1 de contagio leve, el análisis del periodo previo al estado 3 de contagio severo, muestra una correlación negativa débil entre las variables, con un valor de R^2 de 0,11, lo que sugiere que, antes de alcanzar el peak de contagios, a medida que aumentaba la movilidad interna, los casos diarios tendían a disminuir levemente, aunque la relación es débil y con alta dispersión. En este gráfico, las comunas con mayores valores de precariedad de la vivienda, representadas en tonos más claros, se concentran mayormente en niveles intermedios de movilidad interna, mientras que comunas con menores índices de

precariedad, como Las Condes y Vitacura, se ubican hacia los extremos con menor cantidad de casos diarios.

Posterior al estado 3, la relación se vuelve más marcada, con un R^2 de 0,18, lo que sugiere una correlación negativa más clara entre las variables, sin embargo sigue siendo débil. En esta etapa, la tendencia indica que un aumento en la movilidad interna está asociado a una leve disminución en los casos diarios, aunque sigue existiendo una alta variabilidad en los datos. En este segundo periodo, comunas como Providencia y Ñuñoa mantienen niveles bajos de casos, mientras que comunas con índices más altos de precariedad, como La Granja y San Ramón, tienden a concentrarse en niveles superiores de casos diarios.

Por otro lado, en el estado 2 de contagio moderado, el periodo previo al estado 3 muestra una correlación negativa muy débil, con un R^2 de apenas 0,04, lo que indica que no existe relación entre la movilidad interna y los casos diarios de Covid-19 antes del peak de contagios. A la vez, las comunas presentan una alta dispersión, con valores de casos diarios relativamente altos en comparación con el estado 1. Comunas como San Joaquín, La Granja y Recoleta, con mayores niveles de precariedad, se sitúan en la parte superior del gráfico, mostrando casos elevados independientemente de la movilidad interna.

En contraste, el análisis del periodo posterior al estado 3 revela una inversión en la relación, con una correlación positiva y un R^2 de 0,14, lo que sugiere que a medida que la movilidad interna aumenta, también lo hacen los casos diarios, aunque de manera leve.

La división de los datos en periodos previos y posteriores al estado 3 permite evidenciar un cambio en la relación entre la movilidad interna y los casos diarios. Mientras que antes del peak de contagios predominan correlaciones negativas débiles o casi inexistentes, posterior al peak la relación tiende a ser más marcada, ya sea negativamente en el estado 1 o positivamente en el estado 2. Esto sugiere que, luego de la etapa de contagios severos, la movilidad interna comienza a desempeñar un rol más relevante en la dinámica de los contagios, aunque de manera diferenciada según el estado en que se encuentren las comunas.

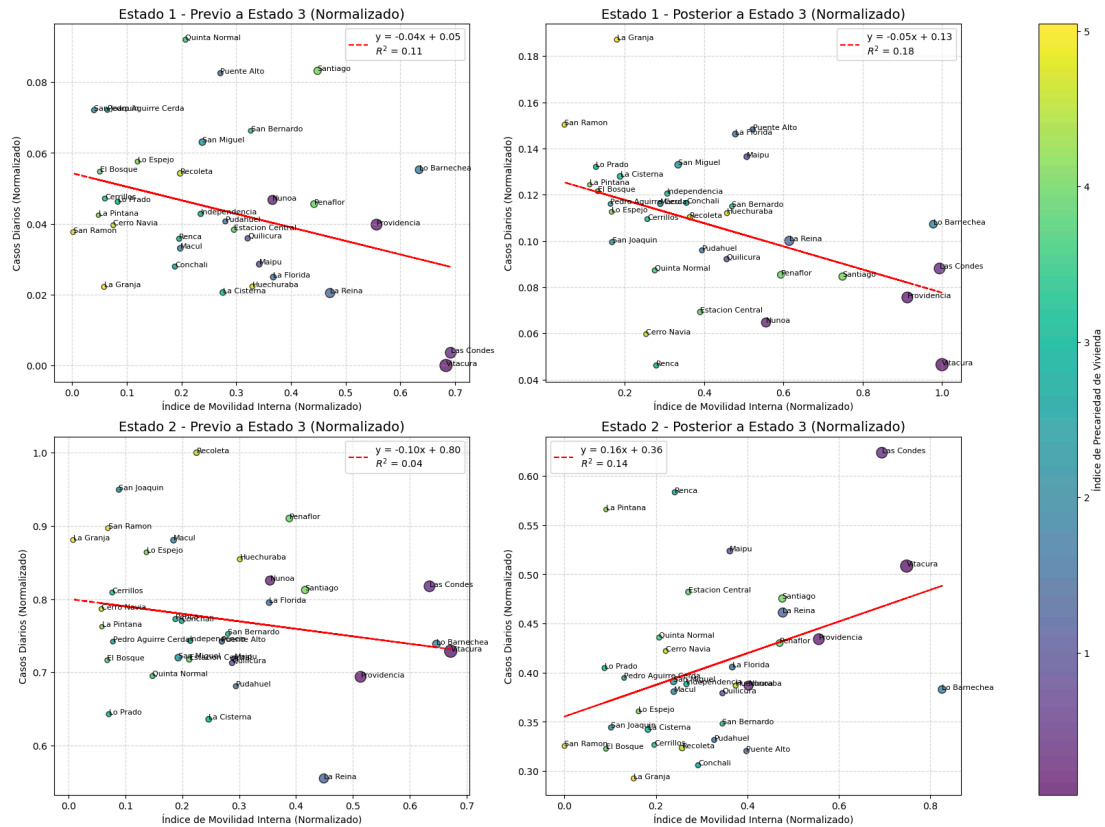


Figura 57. Comparación de la Correlación entre Casos Diarios y Movilidad Interna previo y posterior a Periodo de Contagio Severo.

3.4. Correlación entre Casos Diarios Covid-19 y Movilidad Externa por cada Estado Oculto de Hidden Markov Model

Para el análisis de la correlación entre los casos diarios de Covid-19 y los índices de movilidad externa, también se realizó una regresión lineal para cada estado, la cual fue visualizada mediante gráficos de dispersión, considerando como complemento las mismas variables que fueron el ingreso per cápita (tamaño de las burbujas) y el índice de precariedad de la vivienda (color).

3.4.1. Correlación entre Casos Diarios Covid-19 y Movilidad Externa en Estado 1: Contagio Leve

En el estado 1, caracterizado como estado de contagio leve, el análisis de la relación entre los casos diarios y el índice de movilidad externa muestra una correlación prácticamente nula, con un valor de R^2 igual a cero (Figura 58). En términos de distribución de las comunas, los puntos se dispersan de manera amplia a lo largo del eje de movilidad externa, sin una clara concentración en torno a valores específicos.

Al observar el color de las burbujas, se puede notar que comunas con mayor precariedad, como La Granja, San Ramón y algunas otras al extremo superior derecho del gráfico, presentan valores de movilidad externa relativamente bajos o moderados, pero exhiben niveles ligeramente más elevados de casos diarios. Por el contrario, comunas con menores índices de precariedad, como Vitacura y Las Condes, se encuentran en la parte inferior del gráfico con valores de movilidad externa más altos, pero con niveles de casos diarios más bajos.

El tamaño de las burbujas, que representa el ingreso per cápita por comuna, permite identificar que se mantiene el comportamiento observado anteriormente, donde las comunas con ingresos más altos como Vitacura, Providencia y Las Condes, presentan generalmente menores niveles de casos diarios, aunque exhiben cierta variabilidad en la movilidad externa. En contraste, comunas con

burbujas más pequeñas, correspondientes a menores niveles de ingresos, muestran una distribución más dispersa en términos de movilidad y casos diarios.

Estos resultados evidencian que para el estado 1 de contagio leve, no se observan tendencias significativas entre los casos diarios normalizados y la movilidad externa, lo cual sugiere que, en este nivel de contagio, la movilidad hacia fuera de las comunas no influye de manera relevante en la cantidad de casos. Este resultado es coherente con la baja intensidad de los contagios en esta etapa y la falta de un patrón claro de comportamiento en las comunas estudiadas.

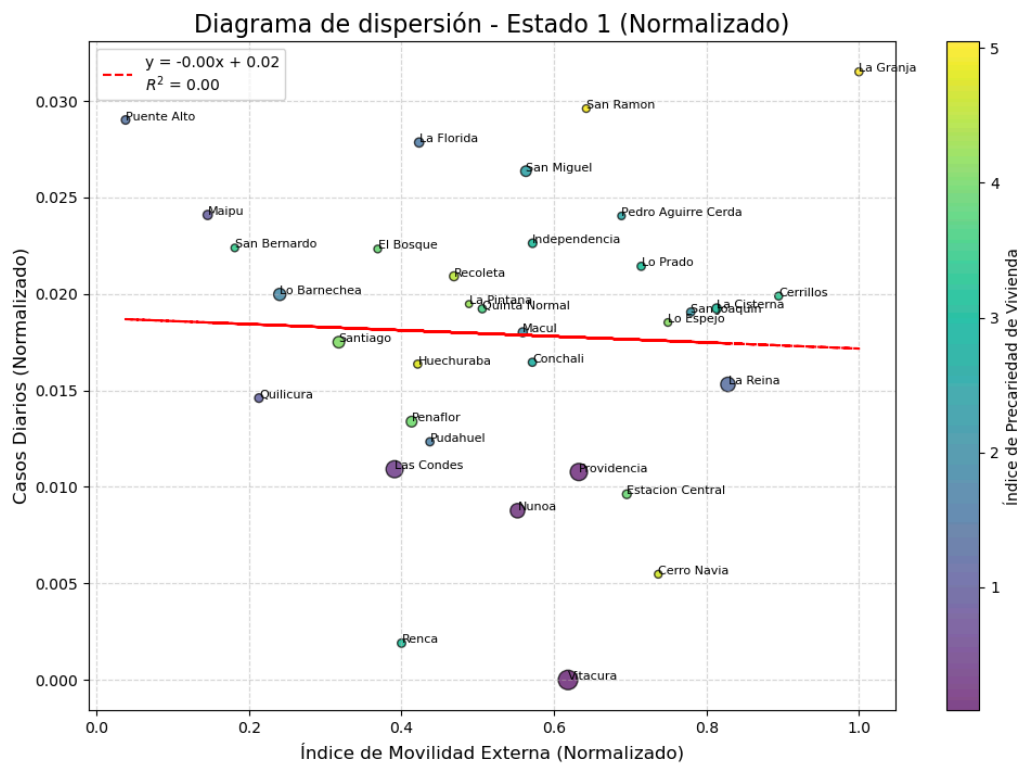


Figura 58. Correlación entre Casos Diarios Covid-19 y Movilidad Externa en Estado 1 de Contagio Leve.

3.4.2. Correlación entre Casos Diarios Covid-19 y Movilidad Externa en Estado 2: Contagio Moderado

Al analizar la correlación entre los casos diarios y el índice de movilidad externa en el estado 2 correspondiente a periodo de contagio moderado, se puede observar una correlación negativa y débil, con un valor R^2 de 0,11, sugiriendo que a medida que aumenta la movilidad externa, los casos tienden a disminuir, sin embargo, la correlación es débil y presenta una alta dispersión de los puntos (Figura 59).

En este caso, la distribución de las comunas evidencia una mayor variabilidad en comparación con el estado 1, donde comunas como Las Condes y Vitacura, ubicadas en el cuadrante superior del gráfico, presentan altos niveles de casos diarios a pesar de tener valores intermedios o bajos de movilidad externa. Por el contrario, comunas como Lo Espejo, La Cisterna y La Granja, con índices más altos de precariedad de la vivienda, se ubican en la parte inferior derecha del gráfico, mostrando menores casos diarios pero con niveles más elevados de movilidad externa.

En síntesis, en el estado 2 de contagio moderado, se observa una relación débilmente negativa entre la movilidad externa y los casos diarios, sugiriendo que, aunque la movilidad hacia fuera de las comunas podría estar asociada a una leve reducción en los contagios, dicha relación no es lo suficientemente fuerte para ser concluyente. A la vez, la dispersión de los datos y la heterogeneidad

entre las comunas reflejan que otros factores, como la precariedad de la vivienda y los niveles de ingreso, podrían influir en esta dinámica.

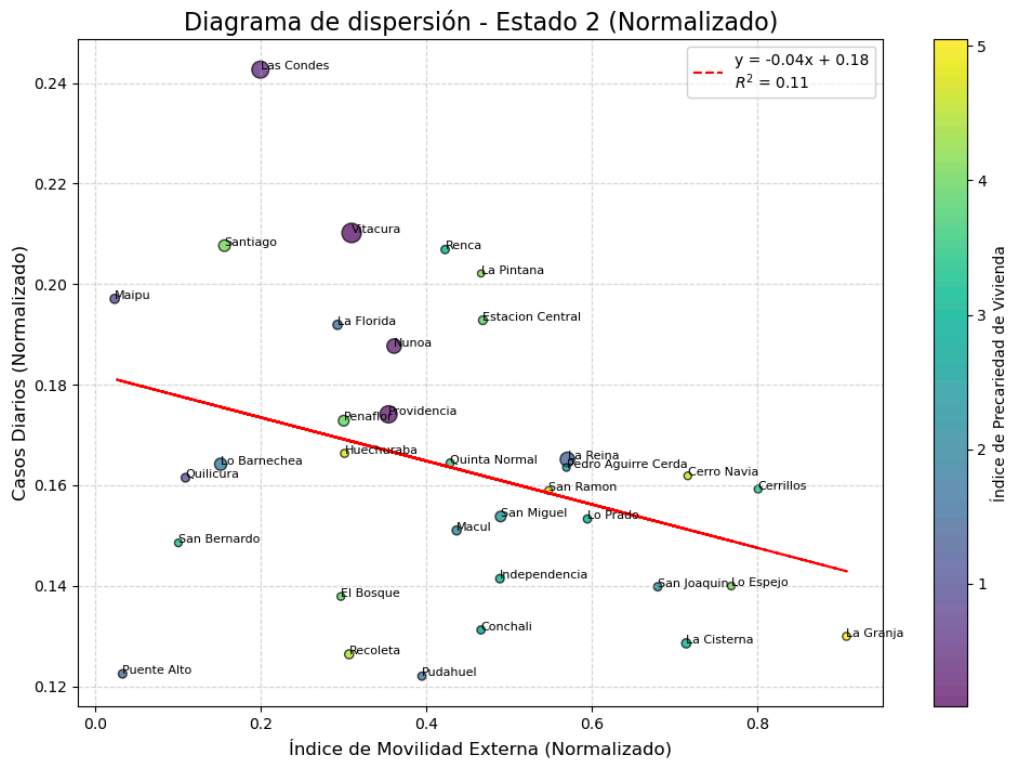


Figura 59. Correlación entre Casos Diarios Covid-19 y Movilidad Externa en Estado 2 de Contagio Moderado.

3.4.3. Correlación entre Casos Diarios Covid-19 y Movilidad Externa en Estado 3: Contagio Severo

En el estado 3, correspondiente a contagio severo, a diferencia de los estados anteriores, se observa una correlación positiva, pero muy débil, con un valor de R^2 de 0,07, lo que sugiere que la movilidad externa tiene una relación directa mínima con los casos diarios en este estado, donde a medida que aumenta la

movilidad externa, los casos diarios tienden a incrementarse levemente, aunque la dispersión de los datos es alta y no se logra establecer una relación clara (Figura 60).

En la distribución de las comunas, se destacan comportamientos heterogéneos. Comunas como La Granja y Cerro Navia, ubicadas en la parte superior izquierda del gráfico, presentan los valores más altos de casos diarios, y al mismo tiempo presentan niveles altos de movilidad externa. Estas comunas también están caracterizadas por índices de precariedad de la vivienda más elevados, como lo indica la coloración en tonos más claros del gráfico. Por el contrario, comunas como Providencia, Ñuñoa y La Reina, ubicadas en la parte inferior del gráfico, presentan menores casos diarios junto con niveles intermedios o bajos de movilidad externa, además, estas comunas muestran índices de precariedad de la vivienda más bajos, reflejados en los tonos oscuros de las burbujas.

En cuanto al tamaño de las burbujas, que representa el ingreso per cápita comuna, se aprecia que comunas como Providencia y Ñuñoa, con mayores ingresos, tienden a registrar menos casos en este estado severo, mientras que comunas con menores ingresos, como La Pintana, Renca y San Ramón, muestran altos niveles de casos diarios, a pesar de su variabilidad en la movilidad externa.

El análisis de este gráfico sugiere que, en el estado 3 de contagio severo, la relación entre la movilidad externa y los casos diarios es prácticamente

inexistente, con un R^2 muy bajo. La heterogeneidad en los datos refleja que, en este nivel de gravedad, otros factores adicionales, como las condiciones socioeconómicas y los índices de precariedad de la vivienda, podrían estar jugando un rol más significativo en la dinámica de los contagios. Esto es especialmente evidente en comunas con altos índices de precariedad, que tienden a registrar mayores niveles de contagio independientemente de los patrones de movilidad.

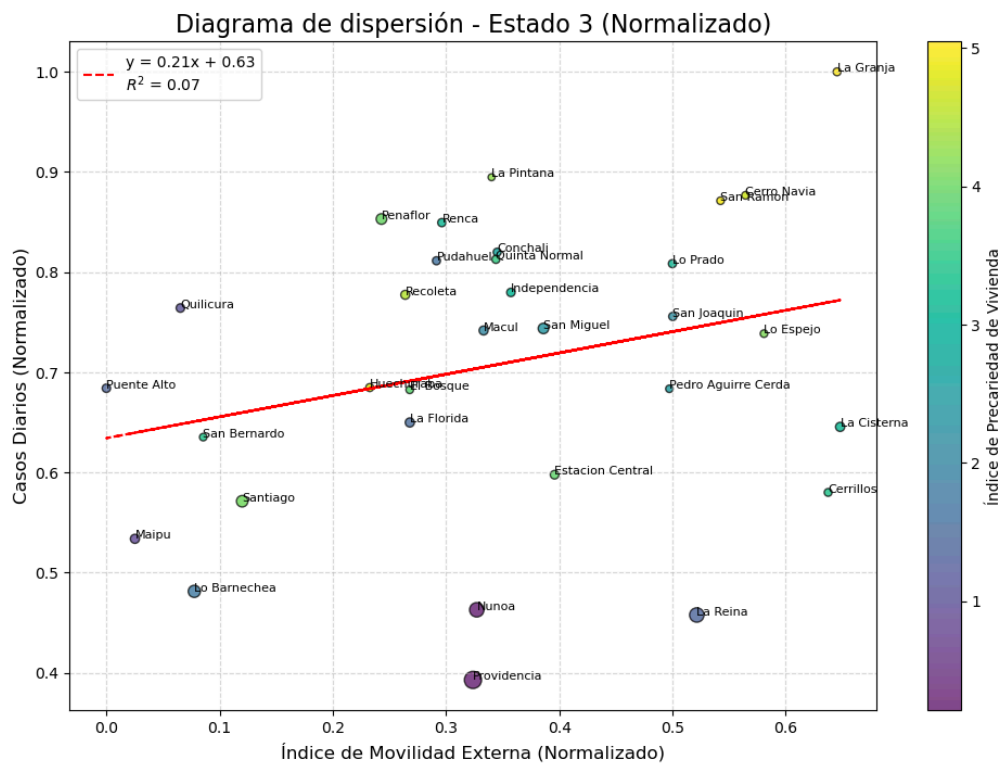


Figura 60. Correlación entre Casos Diarios Covid-19 y Movilidad Externa en Estado 3 de Contagio Severo.

3.4.4. Comparación de la Correlación entre Casos Diarios Covid-19 y Movilidad Externa previo y posterior a Periodo de Contagio Severo

Al analizar las correlaciones entre los casos diarios y el índice de movilidad externa, dividiendo los datos en periodo previo y posterior al estado 3, también se pueden observar algunas diferencias con respecto al análisis del periodo completo (Figura 61).

En el estado 1, el periodo previo al estado 3 presenta una correlación negativa casi nula, con un R^2 de 0,05, lo que indica que no existe una relación significativa entre las variables en esta etapa. A la vez, la distribución de los puntos muestra una gran dispersión, con comunas como Las Condes y Vitacura situadas en niveles bajos tanto de movilidad externa como de casos diarios, mientras que otras comunas con índices más altos de precariedad de la vivienda, como Puente Alto y Santiago, tienden a concentrarse en valores más altos de casos diarios.

En el periodo posterior al estado 3, la correlación se vuelve prácticamente inexistente, con un R^2 de 0,00, lo que sugiere que la movilidad externa no tuvo un efecto relevante en los casos diarios en esta etapa. Las comunas mantienen una distribución similar, con altos niveles de dispersión y sin una tendencia clara entre las variables.

En el estado 2, el periodo previo al estado 3 se observa nuevamente una correlación débil, pero positiva, con un R^2 de 0,01, por lo que la movilidad externa no juega un rol relevante en la variación de los casos diarios.

Por otra parte, en el periodo posterior al estado 3, la relación se invierte, mostrando una correlación negativa leve, con un R^2 de 0,07, lo que sugiere que, después del peak de contagios, la movilidad externa tiende a asociarse con una ligera disminución en los casos diarios, aunque la relación sigue siendo débil.

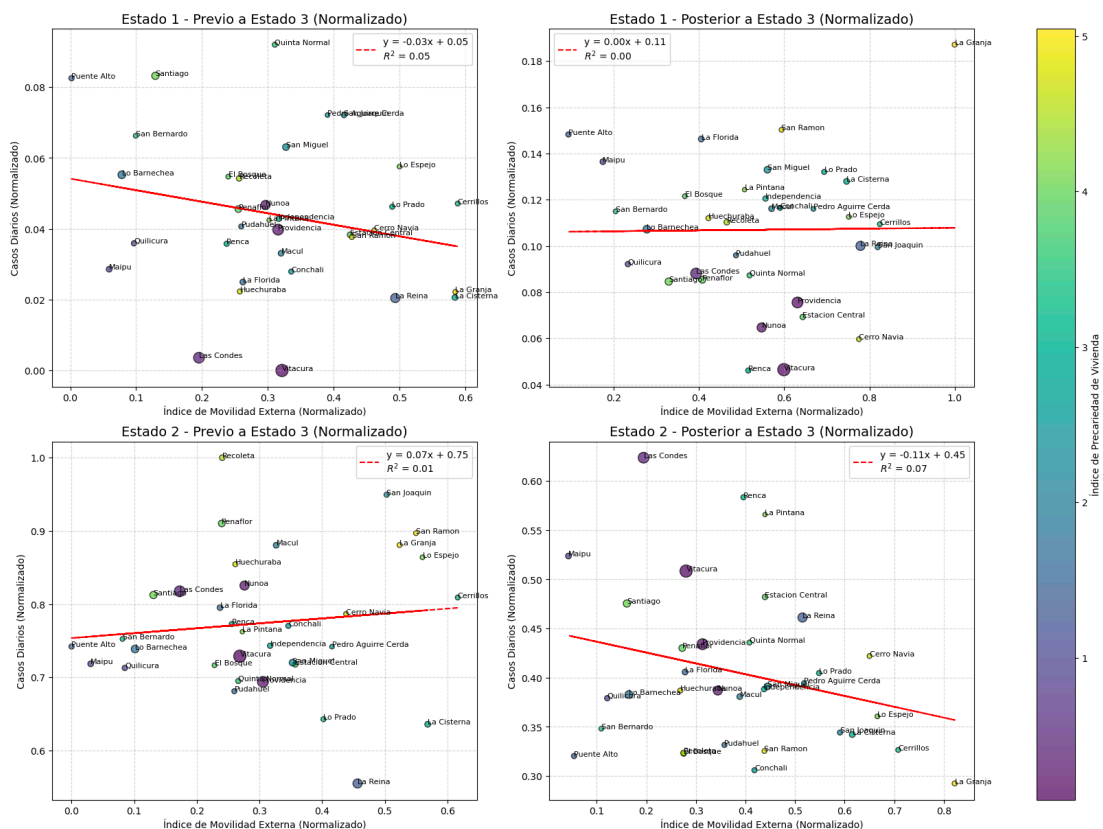


Figura 61. Comparación de la Correlación entre Casos Diarios y Movilidad Externa previo y posterior al Periodo de Contagio Severo.

4. Transiciones entre Estados Ocultos de Hidden Markov Model

Dado el comportamiento identificado en la caracterización de los estados ocultos, resulta fundamental analizar las transiciones entre estos estados para comprender la dinámica de cambio en la situación epidemiológica y su relación con las variables de movilidad y las características sociodemográficas. Las transiciones representan los movimientos de las comunas entre los distintos estados a lo largo del tiempo, permitiendo identificar los factores que podrían haber facilitado o inhibido dichos cambios. En particular, se consideran las nueve transiciones posibles entre los tres estados: la permanencia en un mismo estado, así como los movimientos hacia estados de mayor o menor severidad.

Para abordar este análisis, se examina la relación entre la probabilidad de cada transición y las variables de movilidad, tanto interna como externa. Además, se exploran las correlaciones entre las covariables sociodemográficas y las probabilidades de transición a través de un heatmap, permitiendo identificar patrones relevantes en la dinámica de cambio entre estados. Esta exploración busca determinar qué factores podrían explicar la permanencia o el paso de un estado de contagio leve a moderado o severo, o bien la recuperación hacia estados menos críticos.

4.1. Correlaciones entre Índices de Movilidad y Probabilidades de Transición entre Estados Ocultos

Para interpretar los resultados de correlación entre las probabilidades de transición y los índices de movilidad interna y externa, se presenta un análisis detallado para cada una de las nueve transiciones observadas en el modelo de estados ocultos. La correlación entre estas variables permite identificar la relación existente entre los cambios de estado y la movilidad, tanto interna como externa, lo que resulta clave para entender el comportamiento de la dinámica pandémica.

4.1.1. Correlaciones entre el Índice de Movilidad Interna y las Probabilidades de Transición entre Estados Ocultos

En el análisis de la relación entre el índice de movilidad interna y las probabilidades de transición, se observaron comportamientos variados según cada tipo de transición, lo que sugiere diferencias en la influencia de la movilidad interna en los cambios de estado (Tabla 11) (Figura 62 y 63).

En las transiciones $1 \rightarrow 1$, $1 \rightarrow 2$ y $1 \rightarrow 3$, correspondientes al estado 1, se observa una correlación positiva significativa, con valores de 0,28, 0,58 y 0,70, respectivamente. Esto indica que, dentro del estado de contagio leve, a medida que aumenta la movilidad interna, existe una mayor probabilidad de permanecer en el mismo estado ($1 \rightarrow 1$) o de avanzar hacia estados de mayor severidad ($1 \rightarrow 2$ y $1 \rightarrow 3$). El valor particularmente alto de la correlación en la transición $1 \rightarrow 3$ (0,70)

sugiere que la movilidad interna tiene un rol importante en el avance hacia el estado de mayor severidad, lo que podría explicarse por una mayor exposición al contagio en situaciones de mayor interacción dentro de las comunas.

Por otro lado, en las transiciones asociadas al estado 2, como $2 \rightarrow 1$, $2 \rightarrow 2$ y $2 \rightarrow 3$, los resultados muestran patrones opuestos. La correlación entre movilidad interna y la probabilidad de permanecer en el mismo estado ($2 \rightarrow 2$) es negativa y significativa (-0,54), indicando que a medida que aumenta la movilidad interna, disminuye la probabilidad de quedarse en el estado de contagio moderado. Este comportamiento también se replica de forma más moderada en la transición $2 \rightarrow 1$, con una correlación negativa de -0,28, lo que sugiere que el incremento en movilidad podría estar asociado a un retroceso al estado leve. Sin embargo, la transición $2 \rightarrow 3$, que representa el avance hacia el estado severo, muestra una correlación positiva muy baja (0,03), aunque sigue siendo estadísticamente significativa, lo cual indica que la movilidad interna no tiene un rol predominante en esta transición.

En las transiciones $3 \rightarrow 1$, $3 \rightarrow 2$ y $3 \rightarrow 3$, que involucran el estado de contagio severo, se evidencia una relación inversa entre la movilidad interna y las probabilidades de transición. Las correlaciones de -0,11 para la transición $3 \rightarrow 1$ y -0,23 para la transición $3 \rightarrow 2$ reflejan que un aumento en la movilidad interna se asocia a una menor probabilidad de retroceder a estados menos severos.

Finalmente, la transición 3→3, que representa la permanencia en el estado severo, presenta una correlación negativa más alta, de -0,55, lo que sugiere que la movilidad interna elevada podría estar contribuyendo a una menor probabilidad de que la comuna salga del estado de contagio severo.

En conjunto, estos resultados evidencian que la movilidad interna juega un rol relevante en las transiciones entre estados, principalmente en las etapas iniciales (estado 1), donde la correlación es positiva y significativa para avanzar hacia estados más graves. En cambio, en los estados 2 y 3, predominan las correlaciones negativas, lo que sugiere que una mayor movilidad podría dificultar el retroceso a estados menos severos o la permanencia en estados intermedios. Este patrón destaca la complejidad de la relación entre movilidad interna y la dinámica de los estados pandémicos, subrayando la necesidad de considerar políticas de restricción de movilidad especialmente en etapas tempranas para prevenir avances hacia estados más críticos.

Tabla 11. Correlaciones entre el Índice de Movilidad Interna y las Probabilidades de Transición.

Transición	Correlación	p-value
Transición 1 → 1	0,280	0,000
Transición 1 → 2	0,586	0,000
Transición 1 → 3	0,702	0,000
Transición 2 → 1	-0,280	0,000
Transición 2 → 2	-0,538	0,000
Transición 2 → 3	0,037	0,001
Transición 3 → 1	-0,115	0,000
Transición 3 → 2	-0,237	0,000
Transición 3 → 3	-0,554	0,000

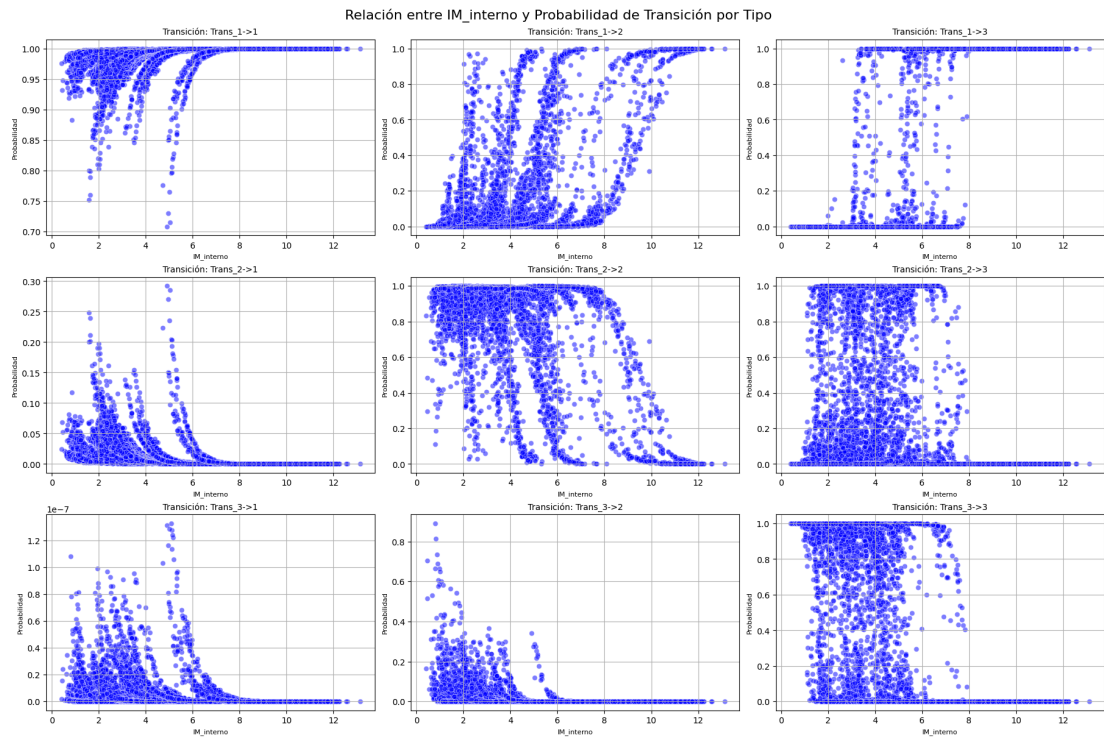


Figura 62. Relación entre el Índice de Movilidad Interna y las Probabilidades de Transición.

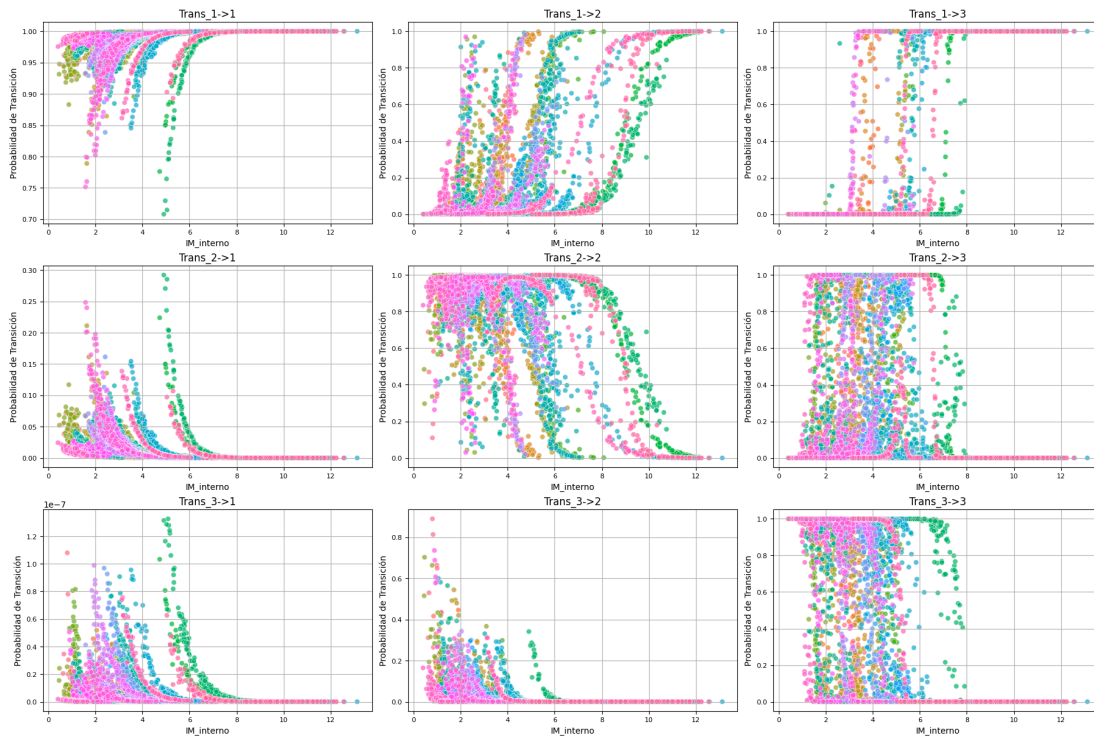


Figura 63. Relación entre el Índice de Movilidad Interna y las Probabilidades de Transición por Comuna.

4.1.2. Correlaciones entre el Índice de Movilidad Externa y las Probabilidades de Transición entre Estados Ocultos

Al analizar la relación entre el índice de movilidad externa y las probabilidades de transición, se observaron comportamientos variados según cada tipo de transición, lo que sugiere diferencias en la influencia de la movilidad externa en los cambios de estado (Tabla 12) (Figura 64 y 65).

En las transiciones dentro del mismo estado, es decir, $1 \rightarrow 1$, $2 \rightarrow 2$ y $3 \rightarrow 3$, se identificaron correlaciones negativas significativas, con valores de $-0,27$, $-0,42$ y $-0,46$, respectivamente. Estas correlaciones indican que a medida que aumenta

la movilidad externa, la probabilidad de permanecer en el mismo estado tiende a disminuir, siendo este efecto más pronunciado en el estado 3, lo que podría interpretarse como una mayor sensibilidad del estado severo ante niveles elevados de movilidad externa.

Por otro lado, en las transiciones hacia estados más graves, como $1 \rightarrow 2$ y $1 \rightarrow 3$, las correlaciones fueron positivas y significativas, con valores de 0,48 y 0,27, respectivamente. Esto sugiere que un incremento en la movilidad externa está asociado con una mayor probabilidad de pasar a estados de mayor severidad. La transición de $1 \rightarrow 2$, correspondiente al cambio de estado leve a moderado, mostró la correlación positiva más alta, lo cual podría reflejar el impacto de la movilidad externa en la propagación de los contagios durante fases iniciales de la pandemia.

En las transiciones hacia estados menos severos, como $2 \rightarrow 1$ y $3 \rightarrow 1$, se encontraron correlaciones negativas significativas, con valores de -0,27 y -0,50, respectivamente. Estas asociaciones indican que una menor movilidad externa favorece la probabilidad de transición hacia estados menos severos, especialmente desde el estado 3, donde la relación es más fuerte.

Finalmente, en las transiciones intermedias, como $2 \rightarrow 3$ y $3 \rightarrow 2$, se evidenciaron relaciones más débiles pero aún significativas. En la transición de $2 \rightarrow 3$, la correlación fue positiva (0,31), mientras que en la transición de $3 \rightarrow 2$ fue negativa (-0,28). Esto sugiere que, aunque la movilidad externa puede facilitar el avance

a estados más severos, su reducción también podría favorecer la transición hacia estados menos graves.

Estos resultados muestran que la movilidad externa tiene una relación clara con las probabilidades de transición, particularmente en las transiciones hacia estados más graves y en la permanencia en el estado severo.

Tabla 12. Correlaciones entre el Índice de Movilidad Externa y las Probabilidades de Transición.

Transición	Correlación	p-value
Transición 1 → 1	0,273	0,000
Transición 1 → 2	0,485	0,000
Transición 1 → 3	0,274	0,000
Transición 2 → 1	-0,273	0,000
Transición 2 → 2	-0,426	0,000
Transición 2 → 3	0,306	0,000
Transición 3 → 1	-0,500	0,000
Transición 3 → 2	-0,280	0,000
Transición 3 → 3	-0,461	0,000

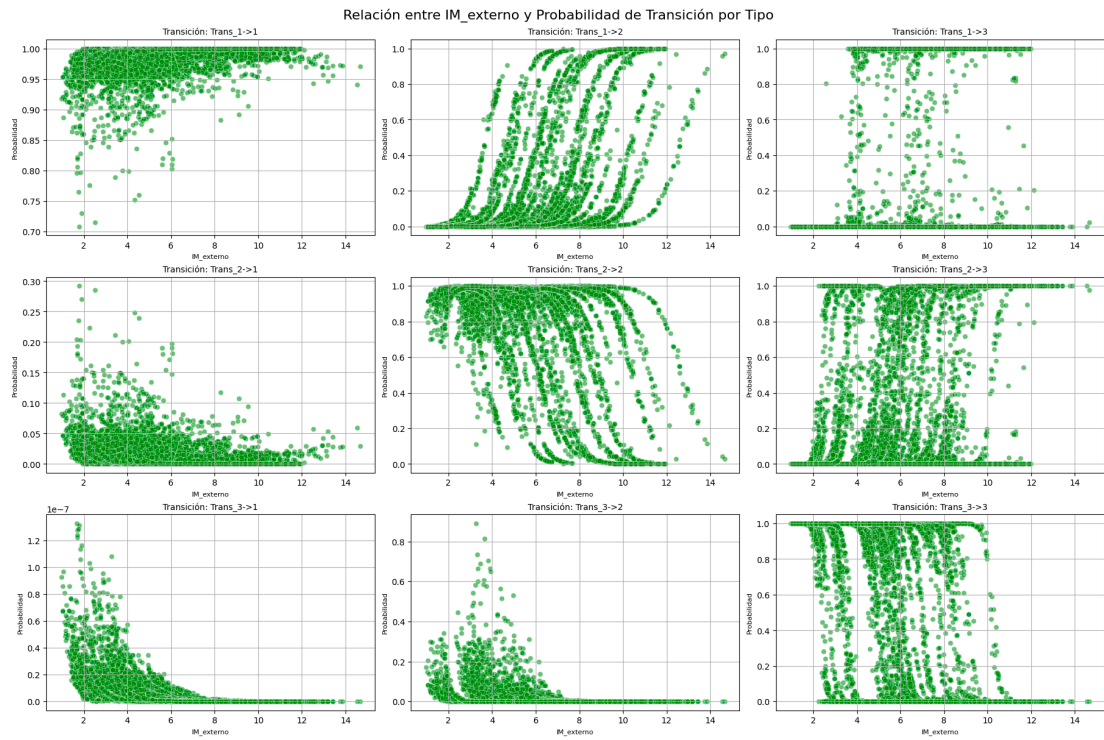


Figura 64. Relación entre el Índice de Movilidad Externa y las Probabilidades de Transición.

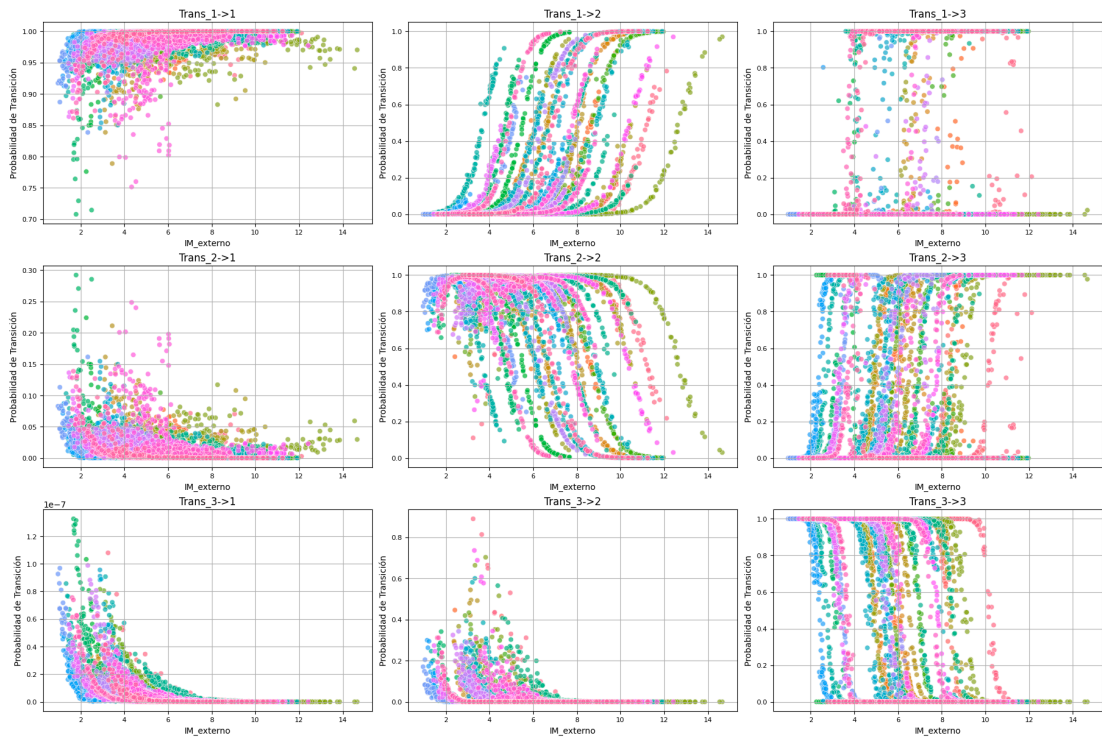


Figura 65. Relación entre el Índice de Movilidad Externa y las Probabilidades de Transición por Comuna.

4.2. Correlaciones entre Covariables y Probabilidades de Transición

El análisis de correlación entre las covariables y las probabilidades de transición permite identificar relaciones específicas que influyen en los cambios de estado. En primer lugar, el índice de movilidad interna presenta correlaciones positivas y significativas con las probabilidades de transición hacia estados más graves, destacándose la transición de estado de contagio leve a severo (1→3) con una correlación de 0,64, seguida de la transición de estado leve a moderado (1→2) con 0,52. En contraste, la movilidad interna muestra una relación negativa con las probabilidades de transición, lo cual se observa más marcado en la transición

de estado de contagio moderado a leve ($2 \rightarrow 1$) (-0,48) y en la mantención en estado de contagio severo ($3 \rightarrow 3$) (-0,51), lo que sugiere que a mayores valores de movilidad interna se reduce la probabilidad de permanecer en estados severos o retroceder a estados menos críticos.

Por otro lado, el índice de movilidad externa presenta correlaciones más moderadas en comparación con la movilidad interna. Las transiciones $1 \rightarrow 2$ (0,38) y $2 \rightarrow 3$ (0,24) muestran asociaciones positivas con la movilidad externa, aunque con menor magnitud.

Al analizar los resultados de correlación entre las covariables sociodemográficas y las probabilidades de transición, es importante tener en consideración que las características sociodemográficas, como el índice de precariedad de la vivienda, el ingreso per cápita y otras variables relacionadas, son valores fijos por comuna y no varían en el tiempo. Esto significa que estas covariables no capturan las dinámicas temporales observadas en el comportamiento de la pandemia, a diferencia de otras variables como los índices de movilidad, que son medidos diariamente y reflejan cambios a lo largo del tiempo.

Este aspecto tiene implicaciones importantes para la interpretación de las correlaciones. En el caso del índice de precariedad de la vivienda, por ejemplo, las correlaciones negativas con las transiciones $1 \rightarrow 2$ (-0,20) y $1 \rightarrow 3$ (-0,32) indican que, en comunas con mayores niveles de precariedad, la probabilidad de

pasar a estados de contagio moderado o severo es menor. A primera vista, esto parece un comportamiento opuesto a lo esperado, ya que en contextos de mayor precariedad se esperaría una mayor vulnerabilidad frente a la pandemia. Sin embargo, al tratarse de valores fijos por comuna, estas correlaciones reflejan diferencias estructurales entre comunas, más que dinámicas temporales en la transición de estados. En este sentido, la ausencia de algunas comunas con niveles socioeconómicos más altos en el estado severo (como Las Condes y Vitacura) contribuye a este resultado, ya que estas comunas no alcanzaron el estado 3 y presentan, en general, menores niveles de precariedad.

De manera similar, el ingreso per cápita comunal muestra una correlación positiva con la transición 1→3 (0,53), lo que sugiere que comunas con mayores niveles de ingreso tienen una mayor probabilidad de transitar directamente al estado severo. Este resultado puede explicarse porque las comunas de mayores ingresos, al contar con mejores sistemas de diagnóstico y monitoreo, pueden haber detectado de manera más rápida y precisa el aumento de casos graves, facilitando la observación de transiciones hacia estados más críticos. Por otro lado, las comunas con menores ingresos y mayores niveles de precariedad pudieron haber tenido menor capacidad de detección temprana, lo que retrasa o subestima las transiciones hacia el estado severo.

Es relevante considerar que, al ser valores fijos, las covariables sociodemográficas no permiten capturar las dinámicas temporales propias de la

pandemia, como el aumento exponencial de contagios o las fluctuaciones en la movilidad. Por esta razón, su correlación con las probabilidades de transición debe interpretarse desde una perspectiva estructural, donde estas variables actúan como factores contextuales que condicionan las diferencias entre comunas. Por ejemplo, comunas con altos niveles de precariedad habitacional pueden haber tenido dinámicas de contagio más prolongadas y menos detectables, lo que reduce la probabilidad de observar transiciones rápidas hacia estados graves.

En contraste, variables como los índices de movilidad, que varían diariamente, reflejan cambios más dinámicos y específicos en el tiempo, lo que puede explicar en mayor medida las transiciones entre estados durante el desarrollo de la pandemia. Esto refuerza la idea de que las características sociodemográficas proporcionan un marco estructural, mientras que las variables dinámicas, como la movilidad, capturan las fluctuaciones temporales que influyen directamente en la evolución de los casos y las transiciones entre estados.

El comportamiento aparentemente opuesto de algunas covariables sociodemográficas, como el índice de precariedad de la vivienda, podría explicarse por la distribución desigual de las comunas entre los estados y por la naturaleza fija de estas variables. Al no variar en el tiempo, estas covariables reflejan diferencias estructurales entre comunas, condicionadas además por la

ausencia de algunas comunas con características particulares en estados más críticos.

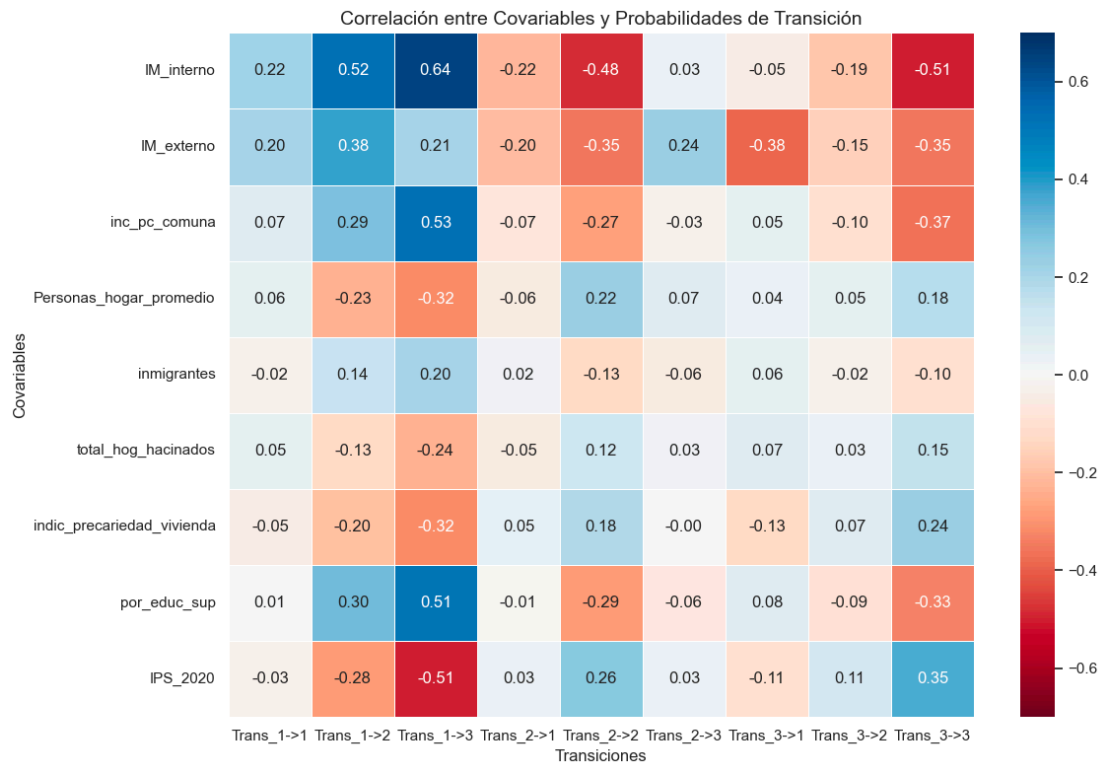


Figura 66. Correlación entre Covariables y las Probabilidades de Transición.

5. Linear Mixed Models para Transiciones entre Estados Ocultos de Hidden Markov Model

Previo al modelamiento de los Linear Mixed Models, se utilizó Elastic Net para seleccionar las covariables más relevantes a ser consideradas. La selección inicial incluyó todas las covariables disponibles: índice de movilidad interna, índice de movilidad externa, ingreso per cápita comunal, promedio de personas

por hogar, porcentaje de inmigrantes, número total de hogares hacinados, indicador de precariedad de la vivienda, porcentaje de educación superior y el índice de prioridad social (IPS 2020). Sin embargo, se observó que al incluir el IPS 2020 en el modelo, muchas covariables clave fueron excluidas debido a la naturaleza compuesta de este indicador, ya que integra información económica, educacional y de vulnerabilidad. Esto dificultaba la interpretación separada de los efectos específicos de las dimensiones económicas, educacionales y de vulnerabilidad.

Con el objetivo de facilitar la interpretación de los resultados y analizar de forma independiente el impacto de cada dimensión, se realizó una nueva selección de variables excluyendo el IPS 2020. La nueva configuración consideró las siguientes covariables: índice de movilidad interna, índice de movilidad externa, ingreso per cápita comunal, promedio de personas por hogar, porcentaje de inmigrantes, número total de hogares hacinados, indicador de precariedad de la vivienda y porcentaje de educación superior (Figura 67).

El modelo Elastic Net ajustado sin incluir el IPS 2020 arrojó un error cuadrático medio de 23,37, lo que refleja un ajuste razonable a los datos. Los coeficientes estimados para las covariables se muestran a continuación, donde los valores negativos indican efectos reductores sobre la variable objetivo, mientras que los valores positivos indican efectos incrementales:

- Índice de movilidad interna (IM_interno): -5,0591
- Índice de movilidad externa (IM_externo): -6,3160
- Ingreso per cápita comunal (inc_pc_comuna): 1,8741
- Promedio de personas por hogar: -0,3373
- Porcentaje de inmigrantes: -0,5658
- Número total de hogares hacinados: -2,7817
- Indicador de precariedad de la vivienda: 2,4493
- Porcentaje de educación superior: -2,3354

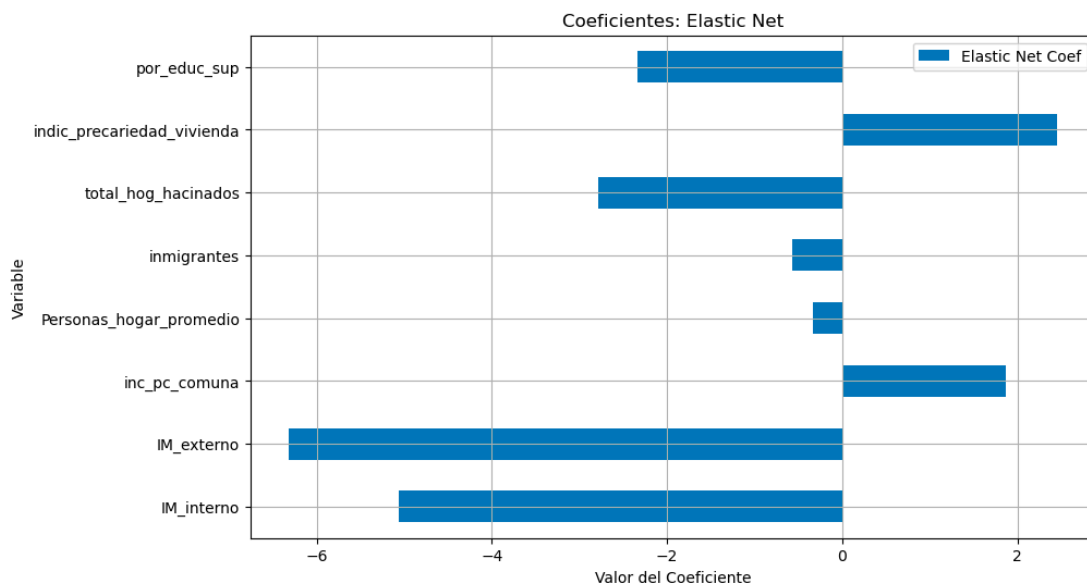


Figura 67. Resultados de Elastic Net para evaluación selección de variables relevantes.

Los resultados muestran que tanto el índice de movilidad interna como el externo tienen un efecto negativo considerable, sugiriendo que una mayor movilidad en estas dimensiones se asocia con una disminución en el número de casos diarios brutos ajustados por media móvil de siete días. Por otro lado, variables como el

ingreso per cápita comunal y el indicador de precariedad de la vivienda presentan efectos positivos, mientras que el porcentaje de educación superior, el promedio de personas por hogar y el número total de hogares hacinados se asocian con efectos negativos.

Estos hallazgos proporcionaron una base sólida para seleccionar las covariables incluidas en los modelos de análisis posteriores, permitiendo una mejor interpretación de los factores asociados con las transiciones entre los diferentes niveles de contagio.

5.1. Linear Mixed Models para la Mantención en Estado de Contagio

Leve: Transición 1 → 1

En la transición 1→1, correspondiente a la probabilidad de mantención en estado de contagios leves, se evaluaron cuatro modelos (A, B, C y D) considerando distintos niveles de complejidad y variables explicativas (Tabla 13).

El modelo A, que incluye únicamente un efecto aleatorio por comuna, presentó un log-likelihood de 17.830 y un AIC de -35.654, con una varianza residual estimada de 0,0004 y una varianza aleatoria por comuna de 0,00008. El modelo B, que incorpora además un efecto aleatorio por día, mostró una mejora en el ajuste con un log-likelihood de 18.859 y un AIC de -37.711, así como una varianza residual de 0,0003, una varianza aleatoria por comuna de 0,00008 y una varianza aleatoria por día de 0,00013.

Por su parte, el modelo C, que incluye los índices de movilidad interna y externa como efectos fijos, evidenció un mejor desempeño, con un log-likelihood de 19.738 y un AIC de -39.464. Este modelo mostró una varianza residual de 0,0002, una varianza aleatoria por comuna de 0,00083 y una varianza aleatoria por día de 0,00063. En este caso, se observó que un aumento de una unidad en el índice de movilidad interna incrementa la probabilidad de mantención en el estado leve en un 1,4 % ($p < 0,001$), mientras que un aumento de una unidad en el índice de movilidad externa disminuye dicha probabilidad en un 0,3% ($p < 0,001$), siendo ambos resultados estadísticamente significativos.

Finalmente, el modelo D, que incorpora tanto los índices de movilidad como variables sociodemográficas (promedio de personas por hogar, porcentaje de educación superior, precariedad de la vivienda, hacinamiento, cantidad de inmigrantes e ingreso per cápita), presentó un log-likelihood de 19.697 y un AIC de -39.371. La varianza residual estimada fue de 0,0002, mientras que la varianza aleatoria por comuna y por día fueron de 0,00024 y 0,00006, respectivamente.

En este modelo, se identificó que un aumento de una persona en el promedio de personas por hogar reduce la probabilidad de mantención en el estado leve en un 3,5% ($p = 0,002$). De manera similar, un aumento de un punto porcentual en el porcentaje de personas con educación superior disminuye esta probabilidad en un 0,2% ($p = 0,001$). Por otro lado, aunque la cantidad de inmigrantes también fue estadísticamente significativa ($p = 0,043$), su efecto práctico fue casi nulo, con

un estimate de 0,000. En el caso de la precariedad de la vivienda, hogares hacinados e ingreso per cápita, no se evidenciaron efectos significativos sobre la probabilidad de transición.

Tabla 13. Resultados de Linear Mixed Models para la Transición 1 → 1.

Modelo	Log-Likelihood	AIC	Varianza Residual	Varianza Aleatoria Comuna	Varianza Aleatoria Día	Efectos Fijos	Estimate	p-value
A	17830	-35654	0,0004	0,00008				
B	18859	-37711	0,0003	0,00008	0,00013			
C	19738	-39464	0,0002	0,00083	0,00063	IM Interno	0,014	0,000*
						IM Externo	-0,003	0,000*
D	19697	-39371	0,0002	0,00024	0,00006	IM Interno	0,014	0,000*
						IM Externo	-0,003	0,000*
						Personas Hogar \bar{x}	-0,035	0,002*
						Educación Superior %	-0,002	0,001*
						Precariedad Vivienda	-0,002	0,533
						Hogares Hacinados	0,000	0,115
						Cantidad Inmigrantes	0,000	0,043*
Ingreso Per Cápita	0,000	0,657						

Posteriormente, se analizaron los coeficientes de determinación marginal (R^2 marginal) y condicional (R^2 condicional) de los modelos C y D, los cuales permiten evaluar el ajuste y la variabilidad explicada por los efectos fijos y aleatorios (Tabla 14).

Para el modelo C, el R^2 marginal fue de 0,287, lo que indica que aproximadamente un 28,7% de la variabilidad en la probabilidad de mantenerse en el estado de contagios leves se explica exclusivamente por los efectos fijos incluidos en el modelo. Por otro lado, el R^2 condicional fue de 0,849, lo que

sugiere que al incluir los efectos aleatorios, el modelo es capaz de explicar un 84,9% de la variabilidad total, lo que refleja la relevancia de los efectos aleatorios asociados a la comuna y al día para capturar patrones específicos no considerados por las covariables sociodemográficas y de movilidad.

Tabla 14. Coeficientes de Determinación Marginal y Condicional para la Transición 1 → 1.

Transición	Modelo	R ² Marginal	R ² Condicional
1 → 1	C	0,287	0,849
	D	0,696	0,866

En base a los resultados obtenidos se seleccionó el modelo C como mejor opción para modelar esta transición debido a su balance entre simplicidad y capacidad explicativa. Aunque el R² marginal (0,287) fue menor en comparación con el modelo D (0,696), el R² condicional de C (0,849) demuestra que el modelo captura adecuadamente la variabilidad total al incluir los efectos aleatorios, por lo tanto, se puede decir que es suficientemente robusto para explicar los datos.

5.1.1. Efectos Aleatorios por Comuna y Tiempo: Transición 1 → 1

El análisis de efectos aleatorios por comuna (Figura 68) muestra que comunas como Las Condes, Lo Barnechea, Providencia y Vitacura presentan interceptos ajustados negativos. Esto indica que, a pesar de iniciar con una alta probabilidad de mantenerse en el estado 1 (leve), estas comunas tenían una mayor predisposición a transitar hacia el estado 2 (moderado) y estado 3 (severo).

Este comportamiento podría explicarse por el hecho de que estas comunas fueron las primeras en reportar casos de Covid-19 en Chile posiblemente debido al mayor acceso a viajes internacionales, lo que incrementa la probabilidad de exposición al virus. Además, se caracterizan por una alta movilidad interna y externa, derivada de su importante rol económico y laboral dentro de la RM.

Por otro lado, comunas como San Ramón, Lo Espejo y La Pintana tienen interceptos ajustados positivos, lo que indica una mayor probabilidad de permanecer en el estado 1 durante el período analizado. Esto se podría deber a la existencia de subregistros de pesquisa de casos debido al menor acceso a pruebas diagnósticas y seguimiento médico, lo que reduciría la identificación de casos, aumentando artificialmente la permanencia en este estado.

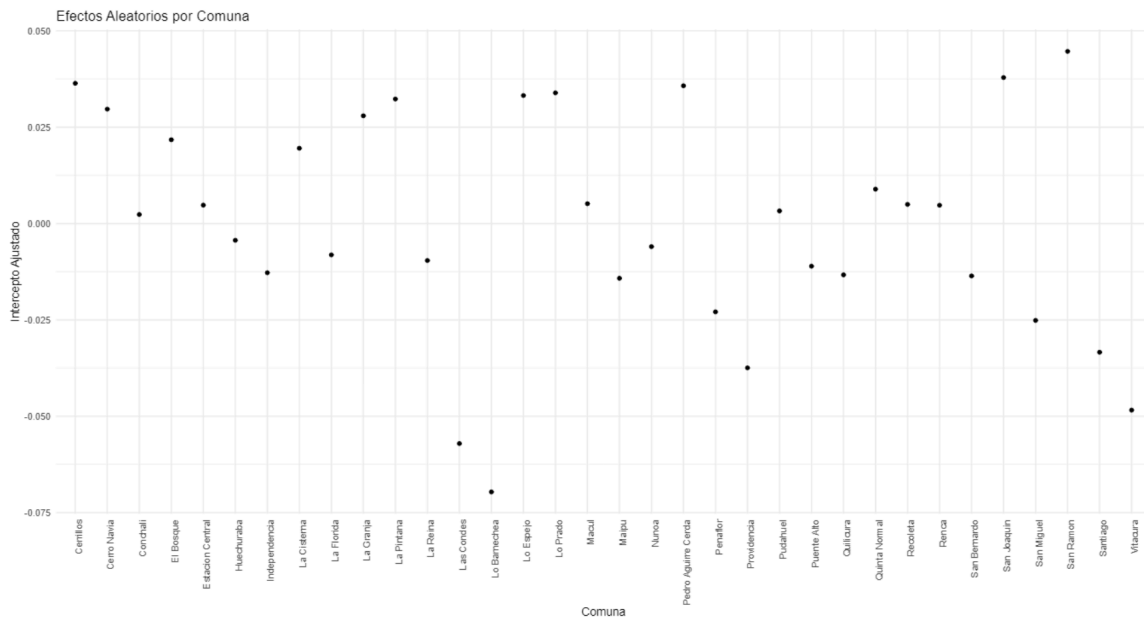


Figura 68. Efectos Aleatorios por Comuna del Modelo C para Mantención en Estado 1.

Al analizar los efectos aleatorios por tiempo (Figura 69), se observaron diferencias significativas según el tipo de día. En azul se muestran los efectos correspondientes a días de semana, mientras que en rojo se observan los días festivos y fin de semana.

Durante los días laborales, se identifican interceptos ajustados más negativos, lo que indica una mayor probabilidad de transición hacia el estado 2 (moderado) o 3 (severo). En cambio, los fines de semana y festivos presentan interceptos más cercanos a cero o positivos, lo que sugiere una menor probabilidad de transición a estados más críticos.

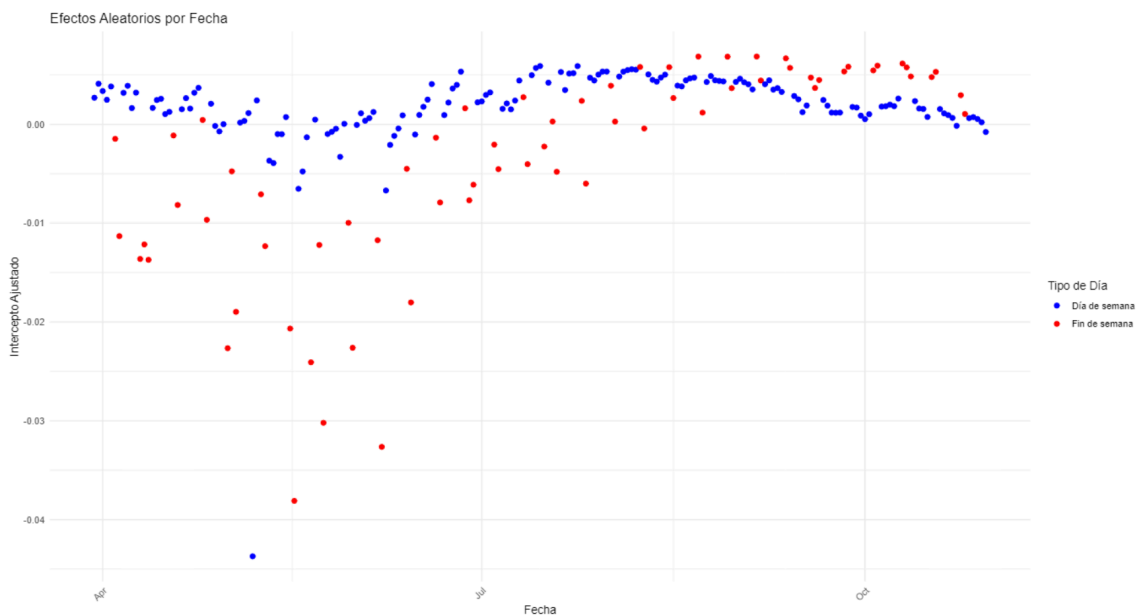


Figura 69. Efectos Aleatorios por Tiempo del Modelo C para Mantenición en Estado 1.

5.1.2. Análisis de Residuales: Transición 1 → 1

El análisis de residuales permite identificar variaciones no explicadas por el modelo en la mantención 1→1 (estado de contagio leve). En el gráfico del análisis de residuales (Figura 70) se observa que los residuales alcanzan su mayor dispersión al inicio del periodo analizado, indicando que algunas comunas presentan variaciones considerables en las predicciones del modelo de mantención en el estado leve, resaltando la importancia de explorar posibles sesgos o variables omitidas que puedan estar influyendo en estos resultados.



Figura 70. Análisis de Residuales de las Comunas a lo largo del Tiempo Modelo C para Mantención en Estado 1.

A continuación, se presentan las comunas con residuales atípicos, donde se incluyeron las comunas con mayores y menores valores residuales (Tabla 15), lo

que indica que existen otros factores que influyen en la probabilidad de transición que no fueron considerados en este modelo.

Tabla 15. Comunas con Residuales Extremos en el Modelo C para Mantención en Estado 1.

Comuna	Rango de Residuales	DE
Lo Barnechea	-0,197 a 0,031	0,03
San Miguel	-0,178 a 0,039	0,04
Las Condes	-0,138 a 0,024	0,02
La Cisterna	-0,074 a 0,051	0,01
Providencia	-0,069 a 0,047	0,02
San Ramón	-0,009 a 0,045	0,01

DE: Desviación Estándar

5.2. Linear Mixed Models para la Mantención en Estado de Contagio

Moderado: Transición 2 → 2

En la transición 2→2, correspondiente a la probabilidad de mantención en el estado de contagio moderado, se evaluaron los mismos 4 modelos (Tabla 16). El modelo A, que incluye únicamente un efecto aleatorio por comuna, presentó un log-likelihood de 269 y un AIC de -533, con una varianza residual de 0,05349 y una varianza aleatoria por comuna de 0,00728.

El modelo B, que incorpora un efecto aleatorio por día además del efecto aleatorio por comuna, mostró una mejora considerable en el ajuste, con un log-likelihood de 2.936 y un AIC de -5.865. Este modelo presentó una varianza residual de 0,02296, una varianza aleatoria por comuna de 0,00742 y una varianza aleatoria por día de 0,03053.

Por su parte, en el modelo C, que incluye los índices de movilidad interna y externa como efectos fijos, mejoró aún más el ajuste, con un log-likelihood de 4.670 y un AIC de -9.328. En este modelo, la varianza residual fue de 0,01447, mientras que la varianza aleatoria por comuna y por día fueron de 0,01874 y 0,00907, respectivamente. En cuanto a los efectos fijos, se observó que un aumento de una unidad en el índice de movilidad interna disminuye la probabilidad de mantención en el estado moderado en un 11,3% ($p < 0,001$), mientras que un aumento de una unidad en el índice de movilidad externa reduce esta probabilidad en un 5,2% ($p < 0,001$).

Finalmente, el modelo D, que incorpora además de los índices de movilidad variables sociodemográficas como efectos fijos, presentó un log-likelihood de 4.637 y un AIC de -9.251. La varianza residual fue de 0,01447, mientras que las varianzas aleatorias por comuna y por día fueron de 0,00619 y 0,00906, respectivamente.

En este modelo, un aumento de una unidad en el índice de movilidad reduce la probabilidad de mantención en el estado moderado en un 11,5% ($p < 0,001$), mientras que un aumento de una unidad en el índice de movilidad externa disminuye esta probabilidad en un 5,1% ($p < 0,001$). Además, un incremento de una unidad en el promedio de personas por hogar aumenta la probabilidad de mantención en el estado moderado en un 17,7% ($p = 0,002$).

Aunque el porcentaje de personas con educación superior no fue estadísticamente significativo ($p = 0,139$), se observó una tendencia positiva, con un efecto estimado del 0,5%. Por otro lado, las variables de precariedad en la vivienda, hogares hacinados, cantidad de inmigrantes e ingreso per cápita fueron significativas, pero sus estimates (0,000) sugieren efectos prácticamente imperceptibles.

Tabla 16. Resultados de Linear Mixed Models para la Transición 2 → 2.

Modelo	Log-Likelihood	AIC	Varianza Residual	Varianza Aleatoria Comuna	Varianza Aleatoria Día	Efectos Fijos	Estimate	p-value
A	269	-533	0,05349	0,00728				
B	2936	-5865	0,02296	0,00742	0,03053			
C	4670	-9328	0,01447	0,01874	0,00907	IM Interno	-0,113	0,000*
						IM Externo	-0,052	0,000*
D	4637	-9251	0,01447	0,00619	0,00906	IM Interno	-0,115	0,000*
						IM Externo	-0,051	0,000*
						Personas Hogar \bar{x}	0,177	0,002*
						Educación Superior (%)	0,005	0,139
						Precariedad Vivienda	0,010	0,514
						Hogares Hacinados	0,000	0,234
						Cantidad Inmigrantes	0,000	0,227
Ingreso Per Cápita	0,000	0,036*						

En cuanto al ajuste y la capacidad explicativa de los modelos evaluados para la transición 2→2 (Tabla 17), el modelo C presentó un R^2 marginal de 0,435, lo que indica que el 43,5% de la variabilidad en la probabilidad de mantención en el estado de contagio moderado puede ser explicado únicamente por los efectos fijos incluidos en el modelo, que fueron los índices de movilidad interna y externa.

Por otra parte, el R^2 condicional alcanzó un valor de 0,807, sugiriendo que el modelo completo, al considerar tanto los efectos fijos como los aleatorios (comuna y día), explica el 80,7% de la variabilidad total, lo que destaca la importancia de los efectos aleatorios en este contexto, ya que capturan heterogeneidades adicionales no consideradas por las variables explicativas.

Con respecto al modelo D, el R^2 marginal fue de 0,643, lo que refleja que el 64,3% de la variabilidad puede atribuirse a los efectos fijos incluidos en el modelo, que fueron los índices de movilidad y las variables sociodemográficas, lo que deja en evidencia el aporte significativo de las variables adicionales en la explicación de la probabilidad de mantención en el estado moderado. El R^2 condicional, por su parte, fue de 0,827, lo que implica que al incluir los efectos aleatorios, el modelo explica el 82,7% de la variabilidad total.

Tabla 17. Coeficientes de Determinación Marginal y Condicional para la Transición 2 → 2.

Transición	Modelo	R^2 Marginal	R^2 Condicional
2 → 2	C	0,435	0,807
	D	0,643	0,827

En base a los resultados obtenidos se seleccionó el modelo D como mejor opción para modelar esta transición debido a que presentó mejores coeficientes de determinación.

5.2.1. Efectos Aleatorios por Comuna y Tiempo: Transición 2 → 2

El análisis de efectos aleatorios por comuna muestra que comunas como El Bosque, Ñuñoa, La Pintana y San Ramón presentaron menor probabilidad de permanecer en el estado 2 (moderado), mientras que comunas como La Cisterna, Pudahuel, Cerro Navia y Providencia presentaron mayor probabilidad de permanecer en el estado moderado por más tiempo (Figura 71).

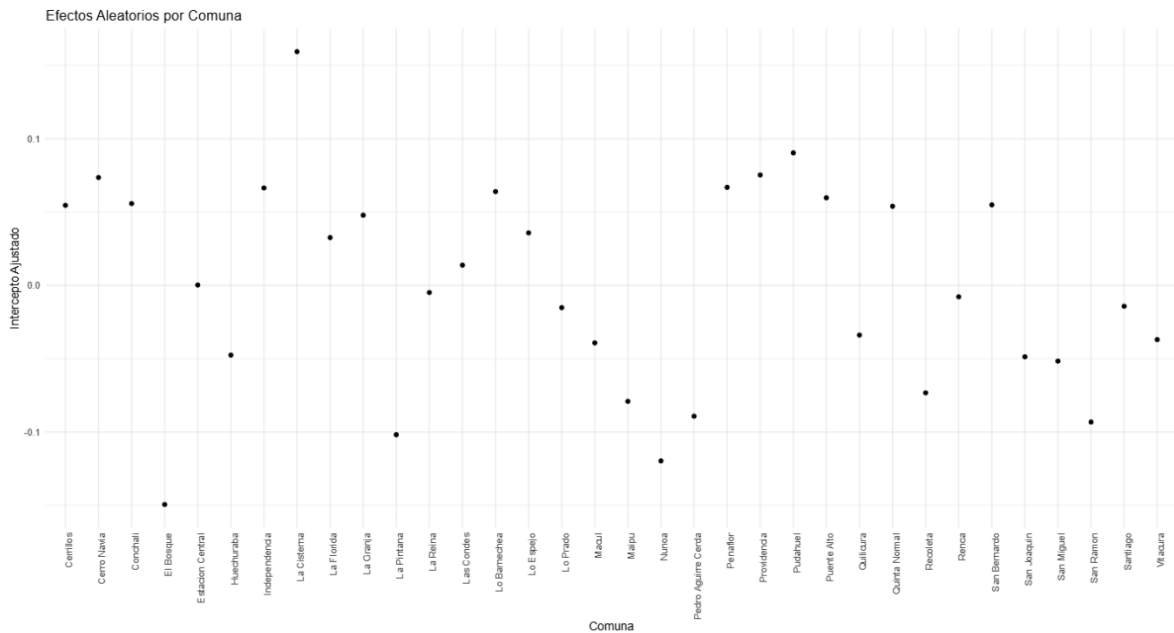


Figura 71. Efectos Aleatorios por Comuna del Modelo D para Mantenición en Estado 2.

En cuanto al efecto aleatorio por tiempo (Figura 72), se observa que en los días de semana, los interceptos ajustados positivos predominan hasta mediados de julio, disminuyendo hacia septiembre, cuando son superados por los interceptos de los fines de semana y festivos, que muestran mayor variabilidad hacia finales

del período analizado. Esto se podría deber a los cambios en las dinámicas de movilidad o a la efectividad de las medidas de restricción aplicadas.

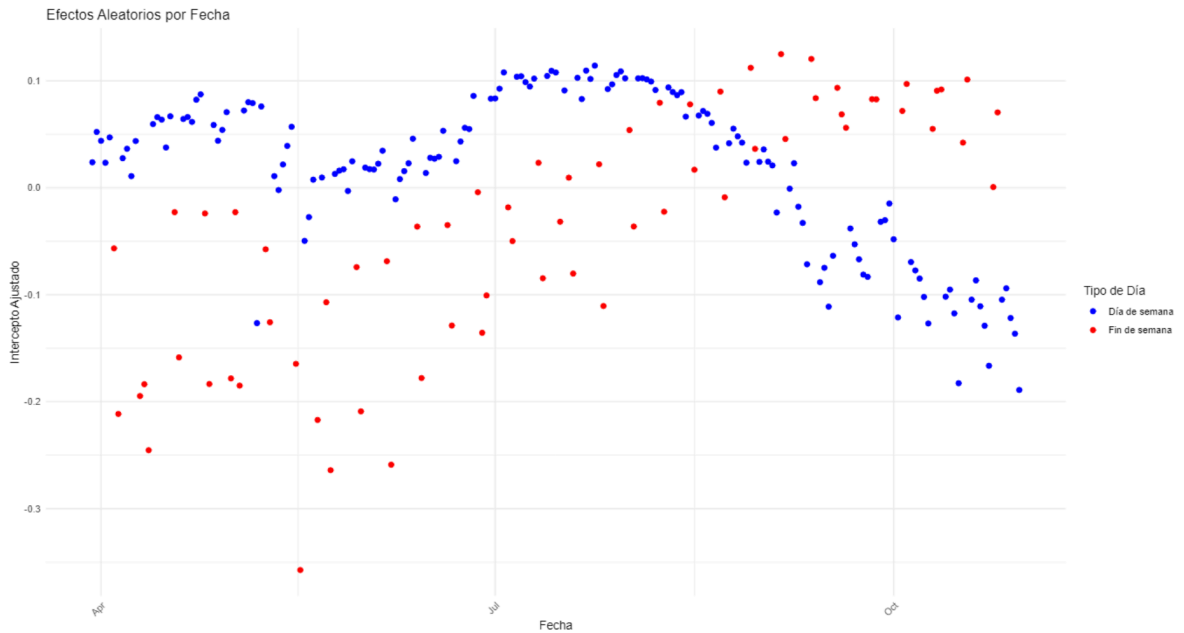


Figura 72. Efectos Aleatorios por Tiempo del Modelo D para Mantenición en Estado 2.

5.2.2. Análisis de Residuales: Transición 2 → 2

En cuanto al análisis de residuales del modelo D para la mantención en el estado moderado, se puede observar que existe una tendencia general en residuales 0, sin embargo, algunas comunas presentan una mayor variabilidad al inicio de la pandemia y después de septiembre (Figura 73).

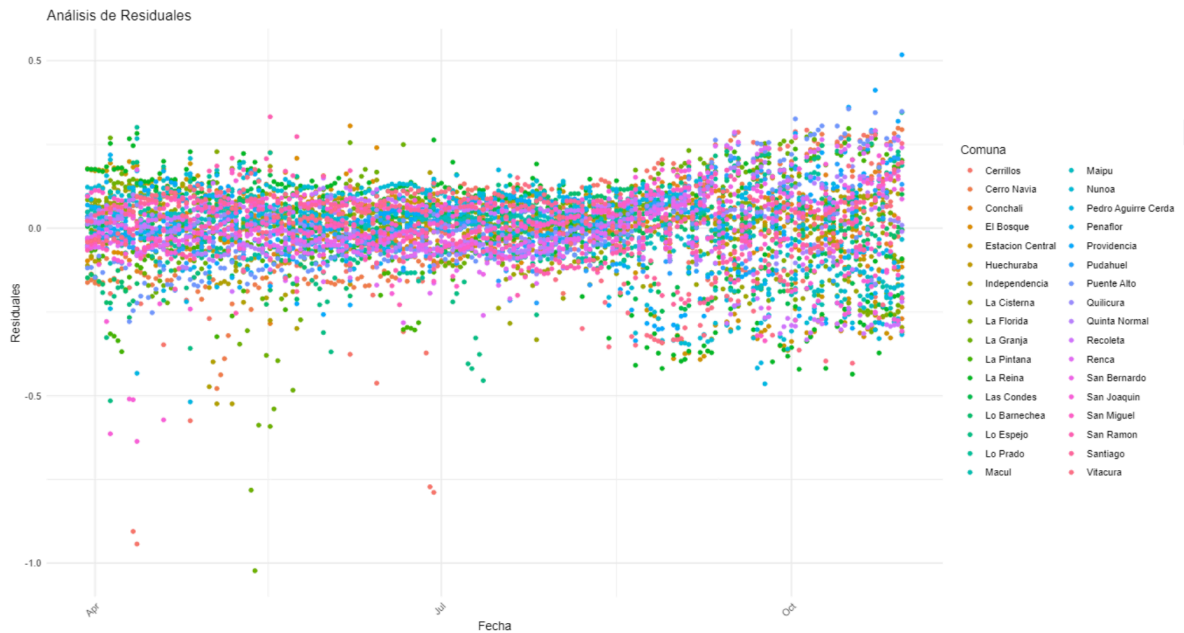


Figura 73. Análisis de Residuales de las Comunas a lo largo del Tiempo Modelo D para Mantención en Estado 2.

A continuación, se presentan las comunas con residuales atípicos, donde se incluyeron las comunas con mayores y menores valores residuales (Tabla 18), lo que indica que existen otros factores que influyen en la probabilidad de transición que no fueron considerados en este modelo.

Tabla 18. Comunas con Residuales Extremos en el Modelo D para Mantención en Estado 2.

Comuna	Rango de Residuales	DE
La Granja	-1,022 a -0,281	0,18
Cerrillos	-0,942 a -0,292	0,17
San Joaquín	-0,636 a 0,216	0,14
Providencia	-0,335 a 0,517	0,13
Puente Alto	-0,278 a 0,356	0,13
Lo Espejo	-0,515 a 0,345	0,14

DE: Desviación Estándar

5.3. Linear Mixed Models para la Mantención en Estado de Contagio

Severo: Transición 3 → 3

En la transición 3→3, correspondiente a la mantención en estado severo (Tabla 19), el modelo A, presentó un log-likelihood de -3.823, un AIC de 7.652, con una varianza residual de 0,16361 y una varianza aleatoria por comuna de 0,03418, mientras que el modelo B, que incorpora un efecto aleatorio por día además del efecto aleatorio por comuna, mostró una mejora significativa en el ajuste, con un log-likelihood de 569 y un AIC de -1.130.

Por su parte, el modelo C, que incluye los índices de movilidad interna y externa como efectos fijos, evidenció un mejor ajuste, con un log-likelihood de 727 y un AIC de -1.443. En este modelo, la varianza residual fue de 0,04125, mientras que la varianza aleatoria por comuna y por día fueron de 0,08219 y 0,07257, respectivamente. Entre los efectos fijos, se observó que un aumento de una unidad en el índice de movilidad interna incrementa la probabilidad de mantención en el estado severo en un 3,2% ($p < 0,001$), mientras que un aumento de una unidad en el índice de movilidad externa reduce esta probabilidad en un 8% ($p < 0,001$).

Finalmente, el modelo D, que incluye tanto los índices de movilidad como variables sociodemográficas como efectos fijos, presentó un log-likelihood de 707 y un AIC de -1.391. La varianza residual fue de 0,04124, mientras que la varianza aleatoria por comuna y por día fueron de 0,01450 y 0,07322, respectivamente.

En este modelo se evidenció que un aumento de una unidad en el índice de movilidad interna incrementa la probabilidad de mantención en el estado severo en un 3,4% ($p < 0,001$), mientras que un aumento de una unidad en el índice de movilidad externa reduce esta probabilidad en un 8% ($p < 0,001$). Además, un aumento de una unidad en el promedio de personas por hogar reduce la probabilidad en un 33,3% ($p < 0,001$), y un incremento en el porcentaje de educación superior disminuye esta probabilidad en un 1% ($p = 0,050$). Por otro lado, aunque algunas otras variables sociodemográficas también presentaron resultados estadísticamente significativos (total de hogares hacinados, cantidad de inmigrantes e ingreso per cápita), los efectos fueron prácticamente imperceptibles.

Tabla 19. Resultados de Linear Mixed Models para la Transición 3 → 3.

Modelo	Log-Likelihood	AIC	Varianza Residual	Varianza Aleatoria Comuna	Varianza Aleatoria Día	Efectos Fijos	Estimate	p-value
A	-3823	7652	0,16361	0,03418				
B	569	-1130	0,04276	0,03475	0,12085			
C	727	-1443	0,04125	0,08219	0,07257	IM Interno	0,032	0,000*
						IM Externo	-0,080	0,000*
D	707	-1391	0,04124	0,01450	0,07322	IM Interno	0,034	0,000*
						IM Externo	-0,080	0,000*
						Personas Hogar \bar{x}	-0,333	0,000*
						Educación Superior (%)	-0,010	0,050*
						Precariedad Vivienda	0,038	0,099
						Hogares Hacinados	0,000	0,008*
						Cantidad Inmigrantes	0,000	0,003*
Ingreso Per Cápita	0,000	0,014*						

En cuanto al ajuste del modelo y la proporción de variabilidad explicada en la transición 3→3, los coeficientes de determinación marginal y condicional mostraron diferencias importantes entre los modelos evaluados (Tabla 20).

En el modelo C, el R^2 marginal fue de 0,056, lo que indica que solo el 5,6% de la variabilidad en la probabilidad de mantención en el estado severo puede ser explicado por los índices de movilidad interna y externa. Sin embargo, el R^2 condicional alcanzó un valor de 0,801, sugiriendo que el modelo completo, al incluir tanto los efectos fijos como los efectos aleatorios (comuna y día), explica el 80,1% de la variabilidad total, lo que refleja la relevancia predominante de los efectos aleatorios en la explicación de la probabilidad de mantención en el estado severo para este modelo, dado el bajo aporte explicativo de los efectos fijos.

Por otro lado, el modelo D presentó un R^2 marginal mayor, alcanzando un valor de 0,393, lo que indica que el 39,3% de la variabilidad puede ser atribuida a los índices de movilidad y variables sociodemográficas. Este incremento en el R^2 marginal resalta la importancia de incorporar las variables sociodemográficas para mejorar la capacidad explicativa del modelo. El R^2 condicional para este modelo fue de 0,806, indicando que el modelo completo, al incluir los efectos aleatorios, explica el 80,6% de la variabilidad total, por lo que los efectos aleatorios continúan siendo relevantes.

Tabla 20. Coeficientes de Determinación Marginal y Condicional para la Transición 3 → 3.

Transición	Modelo	R ² Marginal	R ² Condicional
3 → 3	C	0,056	0,801
	D	0,393	0,806

En base a los resultados obtenidos se seleccionó el modelo D como mejor opción para modelar esta transición debido a que presentó mejores coeficientes de determinación.

5.3.1. Efectos Aleatorios por Comuna y Tiempo: Transición 3 → 3

Al analizar los efectos aleatorios por comuna (Figura 74), se identificaron mayores interceptos ajustados positivos en las comunas de Cerrillos, La Reina, Cerro Navia y Pedro Aguirre Cerda, reflejando condiciones que favorecen la persistencia en un estado severo. Por otro lado, las comunas con mayores interceptos negativos fueron Estación Central, Las Condes, Quinta Normal y Conchalí, indicando menos probabilidades de mantenerse en este estado.

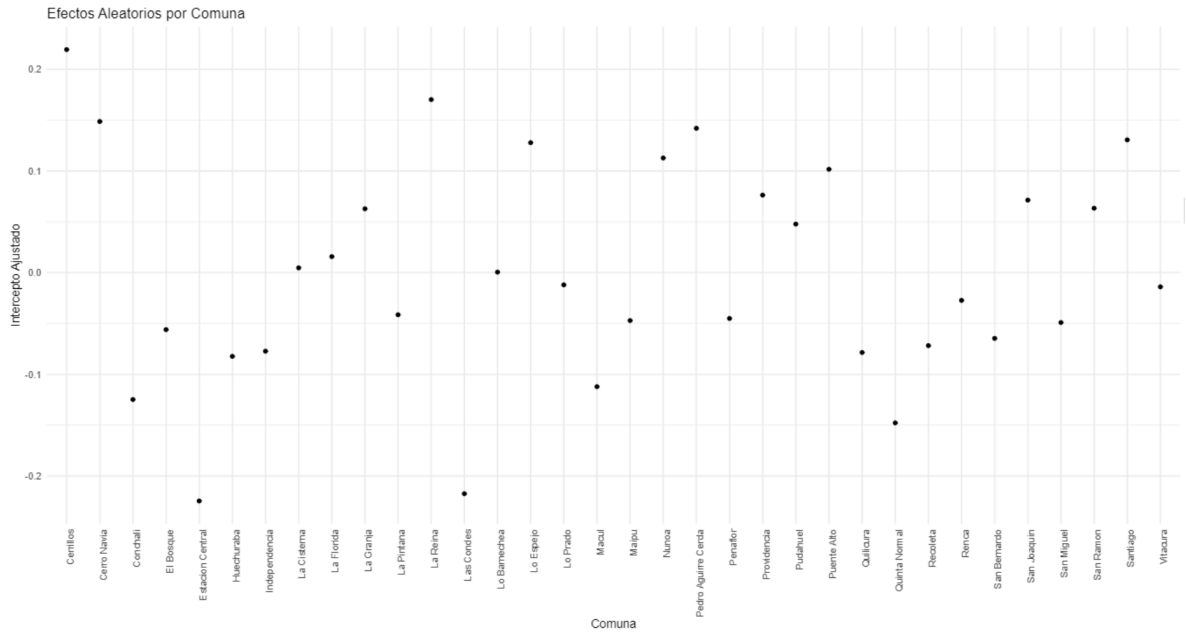


Figura 74. Efectos Aleatorios por Comuna del Modelo D para Mantenición en Estado 3.

En cuanto los efectos aleatorios por tiempo (Figura 75), se puede observar que hasta el mes de julio el intercepto ajustado se mantuvo constante, tanto para días de semana, festivos o fines de semana (alrededor de 0,25). Posteriormente, se produjo una caída abrupta en los interceptos ajustados para días de semana, mientras que los interceptos en fines de semana mantuvieron un patrón similar hasta finales de septiembre, con una disminución progresiva hasta fin de año.

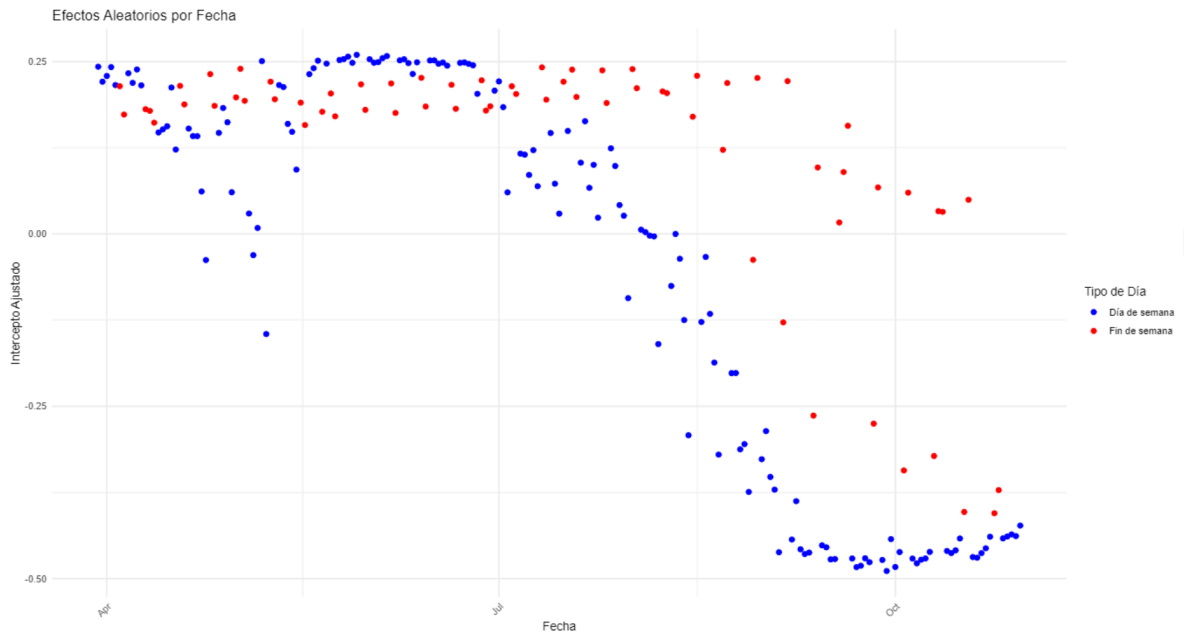


Figura 75. Efectos Aleatorios por Tiempo del Modelo D para Mantenición en Estado 3.

5.3.2. Análisis de Residuales: Transición 3 → 3

Con respecto al análisis de residuales del modelo D para la mantención en el estado de contagio severo, se puede observar una alta variabilidad entre comunas, con una mayor dispersión en el segundo semestre (Figura 76). Las comunas con rangos de residuales extremos y mayor desviación estándar fueron La Pintana, Lo Prado y Vitacura, mostrando un comportamiento heterogéneo que podría asociarse a diferencias estructurales y socioeconómicas no consideradas en este modelo.

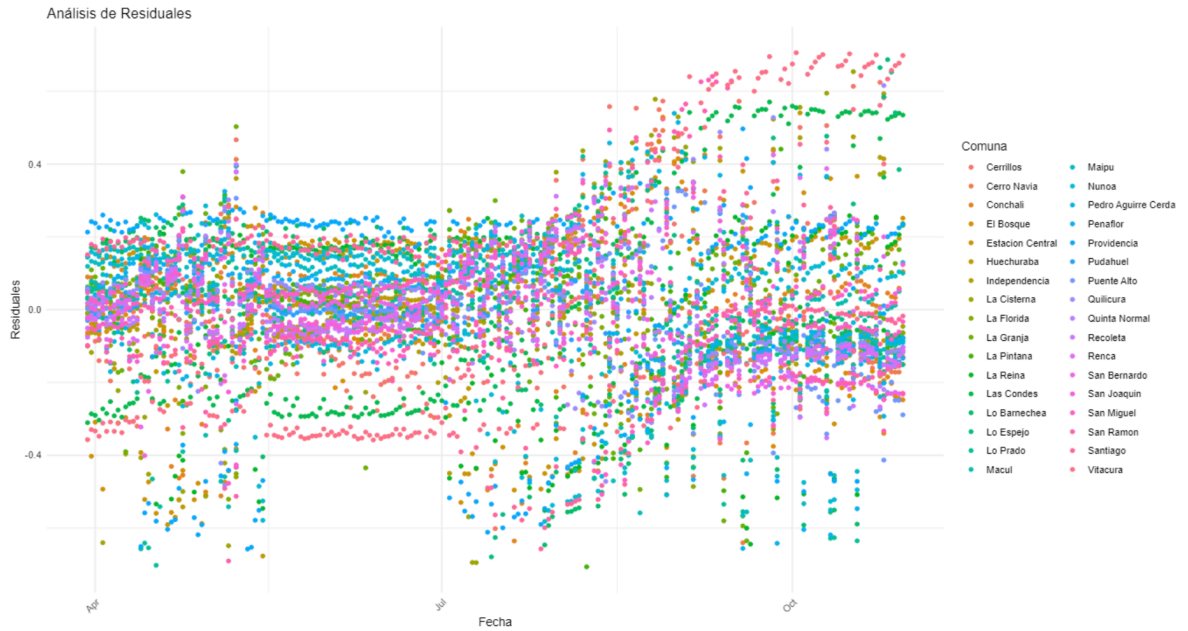


Figura 76. Análisis de Residuales de las Comunas a lo largo del Tiempo Modelo D para Mantención en Estado 3.

A continuación se presentan las comunas con residuales atípicos, donde se incluyeron las comunas con mayores y menores valores residuales (Tabla 21), lo que indica que existen otros factores que influyen en la probabilidad de transición que no fueron considerados en este modelo.

Tabla 21. Comunas con Residuales Extremos en el Modelo D para Mantención en Estado 3.

Comuna	Rango de Residuales	DE
La Pintana	-0,706 a 0,440	0,15
Lo Prado	-0,702 a 0,420	0,19
La Cisterna	-0,695 a 0,653	0,22
Vitacura	-0,358 a 0,705	0,38
Lo Espejo	-0,679 a 0,686	0,19

DE: Desviación Estándar

5.4. Linear Mixed Models para la Transición de Estado de Contagio

Leve a Moderado: Transición 1 → 2

En la transición 1→2, que corresponde al cambio de estado de contagio leve a moderado, los cuatro modelos presentaron diferentes resultados en términos de bondad de ajuste y significancia estadística de las variables incluidas (Tabla 22).

El modelo A, que corresponde al modelo más simple, presentó un log-likelihood de 218 y un AIC de -429, con una varianza residual de 0,05420 y una varianza aleatoria atribuible a la comuna de 0,008. Paralelamente, se evidenció que el modelo B presentó una mejora en el ajuste, con un log-likelihood de 3.324 y un AIC de -6.640, además de una reducción en la varianza residual a 0,02100, con una varianza aleatoria atribuible a la comuna de 0,00820 y una varianza aleatoria atribuible al día de 0,03370.

Por otra parte, el modelo C, con log-likelihood de 5.708 y AIC de -11.405, presentó una varianza residual de 0,01097, una varianza aleatoria asociada a la comuna de 0,01736 y una varianza aleatoria asociada al día de 0,00496. Entre los efectos fijos, se observó que el aumento de una unidad en el índice de movilidad interno incrementa la probabilidad de transición al estado de contagio moderado en un 11%, mientras que un aumento en el índice de movilidad externo incrementa esta probabilidad en un 7%, siendo ambos resultados estadísticamente significativos ($p < 0,001$).

Finalmente, el modelo D, que incluyó las variables sociodemográficas además de los índices de movilidad, presentó un log-likelihood de 5.673 y un AIC de -11.322, con una varianza residual de 0,01098, una varianza aleatoria asociada a la comuna de 0,00691 y una varianza aleatoria atribuible al día de 0,00496. Por su parte, los índices de movilidad interna y externa presentaron casi los mismos efectos que en el modelo C, incrementando la probabilidad de transición en un 11,1% y un 7%, respectivamente ($p < 0,001$).

Adicionalmente, se evidenció que el incremento en el promedio de personas por hogar disminuye la probabilidad de transición en un 12,8% ($p = 0,026$). La cantidad de hogares hacinados y el ingreso per cápita también presentaron resultados estadísticamente significativos, sin embargo, los estimate fueron casi nulos. El porcentaje de personas con educación superior, índice de precariedad de la vivienda y la cantidad de inmigrantes, no mostraron efectos estadísticamente significativos.

Tabla 22. Resultados de Linear Mixed Models para la Transición 1 → 2.

Modelo	Log-Likelihood	AIC	Varianza Residual	Varianza Aleatoria Comuna	Varianza Aleatoria Día	Efectos Fijos	Estimate	p-value
A	218	-429	0,05420	0,00800				
B	3324	-6640	0,02100	0,00820	0,03370			
C	5708	-11405	0,01097	0,01736	0,00496	IM Interno	0,110	0,000*
						IM Externo	0,070	0,000*
D	5673	-11322	0,01098	0,00691	0,00496	IM Interno	0,111	0,000*
						IM Externo	0,070	0,000*
						Personas Hogar \bar{x}	-0,128	0,026*
						Educación Superior (%)	-0,002	0,454
						Precariedad Vivienda	-0,017	0,266
						Hogares Hacinados	0,000	0,035*
						Cantidad Inmigrantes	0,000	0,565
Ingreso Per Cápita	0,000	0,030*						

En cuanto a la capacidad explicativa de los modelos en la transición de estado leve a moderado (Tabla 23), el modelo C mostró un R^2 marginal de 0,507, lo que indica que el 50,7% de la variabilidad en la probabilidad de transición de estado leve a moderado puede ser explicado por los índices de movilidad interna y externa. Por otro lado, el R^2 condicional fue de 0,838, lo que refleja que, al considerar los efectos aleatorios asociados a la comuna y al día, el modelo completo explica el 83,8% de la variabilidad total, por lo que los efectos aleatorios continúan desempeñando un papel importante en esta transición, no obstante, los efectos fijos explican una proporción considerable de la variabilidad.

En el caso del modelo D, el R^2 marginal incrementó a 0,655, lo que indica que el 65,5% de la variabilidad en la probabilidad de transición puede ser atribuida a los índices de movilidad y variables sociodemográficas, dejando en evidencia la

importancia de incluir variables sociodemográficas para mejorar la capacidad explicativa del modelo. Por otra parte, el R^2 condicional fue de 0,839, muy similar al observado en el modelo C, lo que sugiere que los efectos aleatorios mantienen una contribución estable al ajuste del modelo.

Tabla 23. Coeficientes de Determinación Marginal y Condicional para la Transición 1 → 2.

Transición	Modelo	R^2 Marginal	R^2 Condicional
1 → 2	C	0,507	0,838
	D	0,655	0,839

5.4.1. Efectos Aleatorios por Comuna y Tiempo: Transición 1 → 2

Al analizar el efecto aleatorio por comuna en la transición del estado leve a moderado (1→2), se puede observar diferencias significativas entre las comunas (Figura 77). Comunas como El Bosque, seguido de Ñuñoa, San Ramón, Recoleta y Maipú, presentan los interceptos ajustados más altos, lo que indica una mayor predisposición a transitar desde el estado leve a moderado, mientras que comunas como La Cisterna, Cerrillos, Pudahuel y Cerro Navia muestran interceptos ajustados negativos, sugiriendo que estas áreas presentaron menor probabilidad de transitar rápidamente al estado 2.

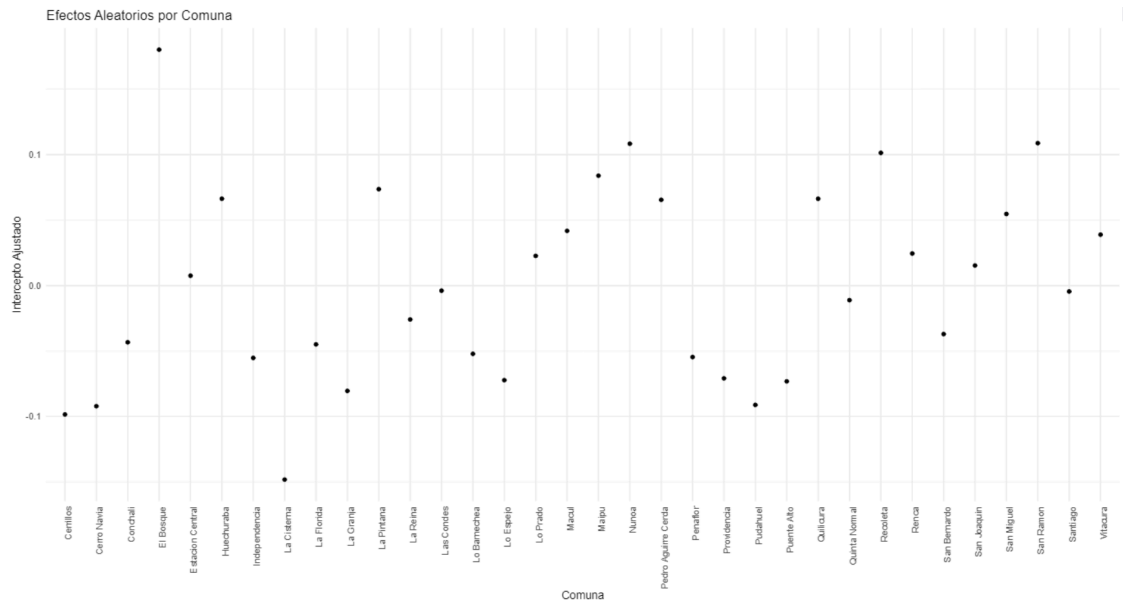


Figura 77. Efectos Aleatorios por Comuna del Modelo D para la Transición 1 → 2.

Con respecto al análisis temporal se observa una variación en los interceptos ajustados a lo largo del periodo (Figura 78). Al inicio de la pandemia, las transiciones de leve a moderado fueron más frecuentes durante los días festivos o fines de semana. Sin embargo, a partir de julio de 2020, se observa un cambio en la dinámica; los interceptos ajustados comenzaron a disminuir en los días festivos, mientras que los días de semana toman protagonismo en las transiciones desde septiembre en adelante.

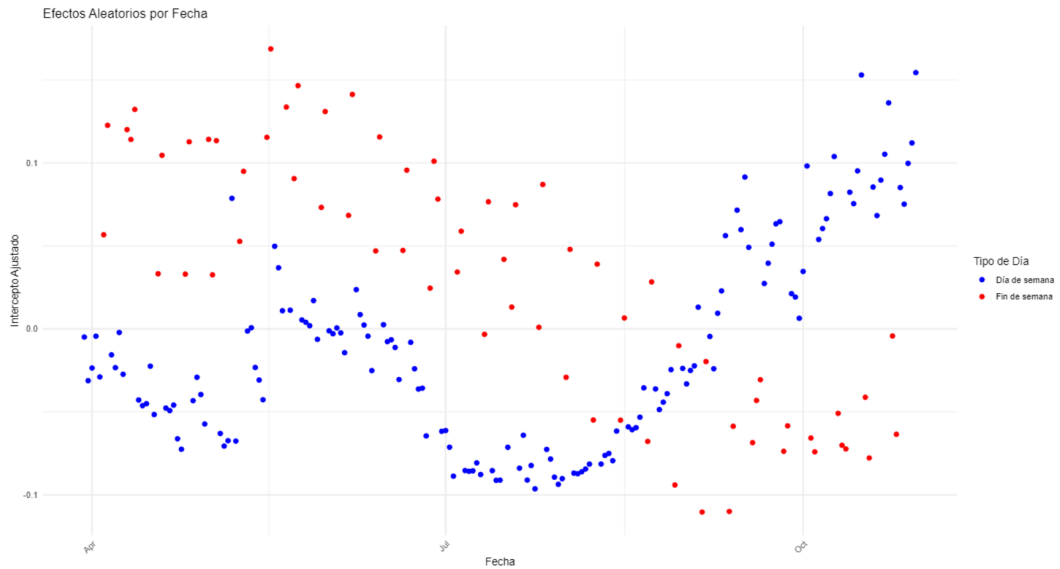


Figura 78. Efectos Aleatorios por Tiempo del Modelo D para la Transición 1 → 2.

5.4.2. Análisis de Residuales: Transición 1 → 2

Con respecto al análisis de residuales para la transición del estado leve a moderado, se observa que al inicio del periodo las comunas presentan una variabilidad baja, con valores cercanos a cero, consistente con una progresión gradual al estado moderado, sin embargo, desde agosto, los residuales comienzan a dispersarse, reflejando un aumento en la variabilidad y en las diferencias entre comunas en la probabilidad de transitar desde el estado leve a moderado (Figura 79).



Figura 79. Análisis de Residuales de las Comunas a lo largo del Tiempo Modelo D para la Transición 1 → 2.

A continuación, se presentan comunas con residuales atípicos, abarcando las comunas con mayores y menores valores residuales (Tabla 24). Las comunas con mayor rango de residuales fueron Providencia y Vitacura, ambas con baja variabilidad.

Tabla 24. Comunas con Residuales Extremos en el Modelo D para la Transición 1 → 2.

Comuna	Rango de Residuales	DE
Providencia	-0,552 a 0,313	0,13
Vitacura	-0,326 a 0,363	0,13
Puente Alto	-0,313 a 0,167	0,11
Pedro Aguirre Cerda	-0,148 a 0,443	0,09
La Granja	-0,291 a 0,428	0,14
La Reina	-0,200 a 0,407	0,17

DE: Desviación Estándar

5.5. Linear Mixed Models para la Transición de Estado de Contagio Leve a Severo: Transición 1 → 3

En la transición 1→3, que corresponde al paso de estado de contagio leve a severo (Tabla 25), el modelo A presentó un log-likelihood de -929 y un AIC de 1.863, con una varianza residual de 0,07382 y una varianza aleatoria asociada a la comuna de 0,03624. El modelo B, con un log-likelihood de -86 y un AIC de 180, mostró una varianza residual de 0,05428, mientras que las varianzas aleatorias para comuna y día fueron de 0,03633 y 0,01954, respectivamente.

Por su parte, el modelo C, que agrega como efectos fijos los índices de movilidad, presentó un log-likelihood de 1.359 y un AIC de -2.707, con una varianza residual de 0,03902, y varianzas asociadas a la comuna y al día de 0,03072 y 0,00031, respectivamente. Al analizar los efectos fijos, se observa que un aumento de una unidad en el índice de movilidad interna incrementa la probabilidad de transición al estado severo en un 17,8%, mientras que un aumento en el índice de movilidad externa incrementa esta probabilidad en un 1,2%. Ambos efectos fueron estadísticamente significativos ($p < 0,001$).

Finalmente, el modelo D, que incluyó tanto los índices de movilidad como las variables sociodemográficas, presentó un log-likelihood de 1.326 y un AIC de -2.629, con una varianza residual de 0,03901, una varianza aleatoria asociada a la comuna de 0,01203 y una varianza aleatoria atribuible al día de 0,00031. Los índices de movilidad interna y externa mantuvieron efectos significativos,

incrementando la probabilidad de transición en un 18,1% y un 1%, respectivamente ($p < 0,001$).

Con respecto al efecto de las variables sociodemográficas, se evidenció que un aumento en el promedio de personas por hogar reduce la probabilidad de transición al estado severo en un 30,9% ($p < 0,001$) y que un incremento en el porcentaje de educación superior disminuye levemente la probabilidad de transición en un 1,3% ($p = 0,006$). El resto de las variables no presentaron resultados estadísticamente significativos.

Tabla 25. Resultados de Linear Mixed Models para la Transición 1 → 3.

Modelo	Log-Likelihood	AIC	Varianza Residual	Varianza Aleatoria Comuna	Varianza Aleatoria Día	Efectos Fijos	Estimate	p-value
A	-929	1863	0,07382	0,03624				
B	-86	180	0,05428	0,03633	0,01954			
C	1359	-2707	0,03902	0,03072	0,00031	IM Interno	0,178	0,000*
						IM Externo	0,012	0,000*
D	1326	-2629	0,03901	0,01203	0,00031	IM Interno	0,181	0,000*
						IM Externo	0,010	0,000*
						Personas Hogar \bar{x}	-0,309	0,000*
						Educación Superior (%)	-0,013	0,006*
						Precariedad Vivienda	0,026	0,216
						Hogares Hacinados	0,000	0,137
						Cantidad Inmigrantes	0,000	0,051
Ingreso Per Cápita	0,000	0,720						

Al analizar los coeficientes de determinación (Tabla 26), el modelo C presentó un R^2 marginal de 0,505, lo que indica que los índices de movilidad explican el 50,5% de la variabilidad en la probabilidad de transición de estado de contagio leve a

moderado. Paralelamente, el R^2 condicional, fue de 0,724, por lo tanto, al incluir los efectos aleatorios asociados a la comuna y al día, el modelo alcanza a explicar el 72,4% de la variabilidad total.

Por otra parte, el modelo D mostró un incremento importante en el R^2 marginal, alcanzando un valor de 0,725, lo que indica que el 72,5% de la variabilidad en la probabilidad de transición es explicada por los índices de movilidad y variables sociodemográficas. En el caso del R^2 condicional, este fue levemente mayor, con un valor de 0,791, lo que refleja que los efectos aleatorios influyen en la explicabilidad del modelo, pero en menor medida, ya que el mayor porcentaje de la variabilidad es explicado por los efectos fijos.

Tabla 26. Coeficientes de Determinación Marginal y Condicional para la Transición 1 → 3.

Transición	Modelo	R^2 Marginal	R^2 Condicional
1 → 3	C	0,505	0,724
	D	0,725	0,791

Estos resultados evidencian que el modelo D, al integrar información adicional de las variables sociodemográficas, ofrece una explicación más completa y ajustada de los factores que influyen en la transición del estado leve al severo, por lo que se seleccionó como mejor opción para modelar esta transición.

5.5.1. Efectos Aleatorios por Comuna y Tiempo: Transición 1 → 3

Al analizar los interceptos ajustados por comuna (Figura 80), se puede observar que comunas como San Miguel, La Pintana, San Ramón, Ñuñoa y El Bosque presentan valores significativamente positivos, sugiriendo que estas comunas tienen una mayor probabilidad de transitar directamente del estado 1 de contagio leve, al estado 3 de contagio severo. Por otra parte, comunas como Recoleta, La Cisterna, Independencia, Peñaflor, San Bernardo o Quinta Normal presentan interceptos ajustados más bajos, lo que indica una menor probabilidad de realizar esta transición directa.

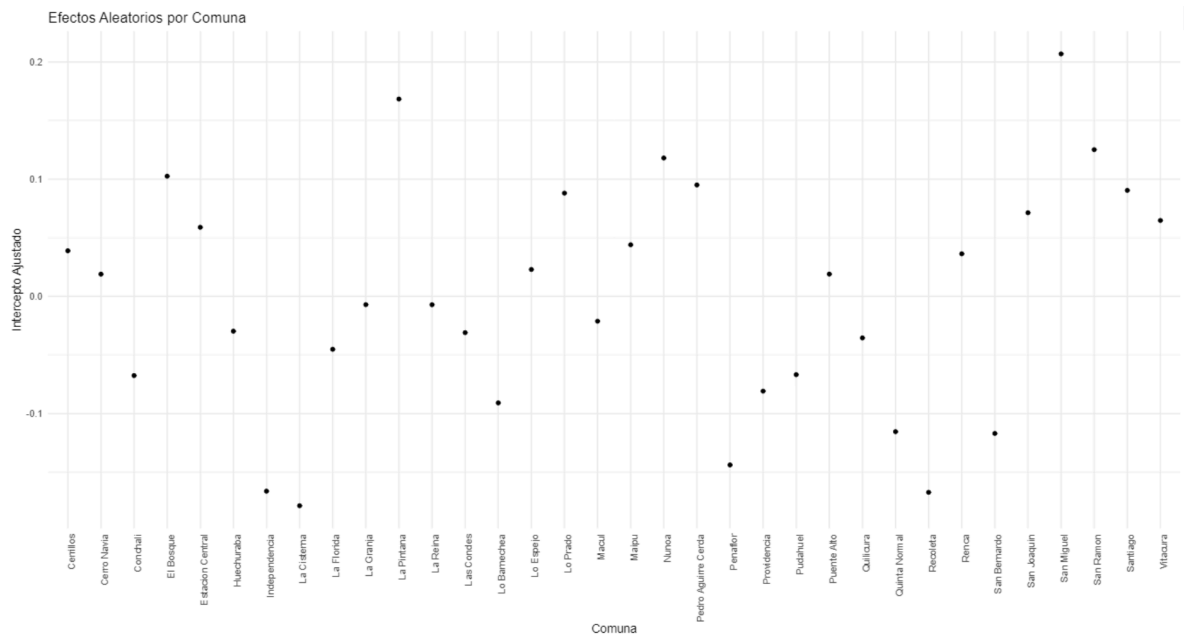


Figura 80. Efectos Aleatorios por Comuna del Modelo D para la Transición 1 → 3.

En cuanto al efecto aleatorio por tiempo (Figura 81), se puede observar que al inicio del periodo, los días de semana presentan interceptos ajustados negativos,

mientras que los fines de semana y festivos tienen valores positivos. Esto refleja una menor probabilidad de transición de estado leve a severo durante los días laborales. A partir de abril, los interceptos ajustados de ambos tipos de días disminuyen, aunque los fines de semana mantienen valores mayores y desde julio en adelante los días de semana muestran un incremento sostenido en los interceptos, mientras que en los fines de semana tiende a disminuir.

Paralelamente, en septiembre se exhibe una mayor dispersión, con interceptos entre -0,01 y 0,01, sugiriendo dinámicas más heterogéneas en la transición desde el estado de contagio leve al estado de contagio severo (1→3).

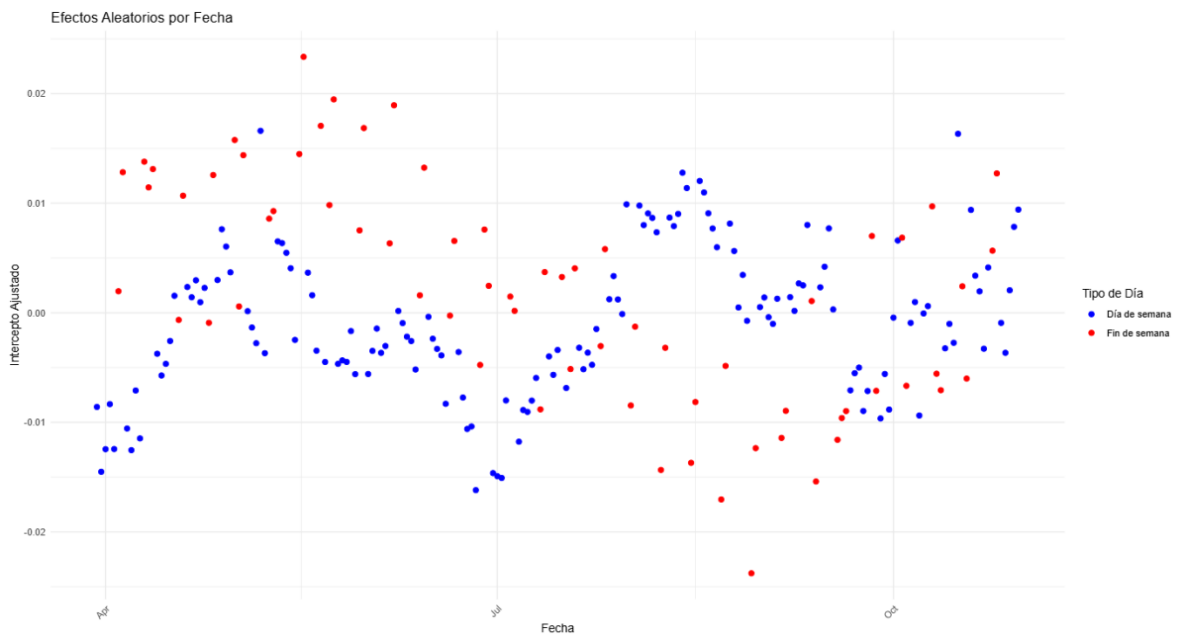


Figura 81. Efectos Aleatorios por Tiempo del Modelo D para la Transición 1 → 3.

5.5.2. Análisis de Residuales: Transición 1 → 3

Para la transición del estado leve a severo, el análisis de residuales muestra una tendencia central con valores predominantemente cercanos a cero, sin embargo, hay comunas específicas en las que se observa una mayor dispersión y variabilidad, como Providencia, Vitacura y Lo Barnechea, las cuales presentan sobreestimaciones y subestimaciones en la probabilidad de transitar al estado severo, reflejando una mayor incertidumbre en su comportamiento (Figura 82).

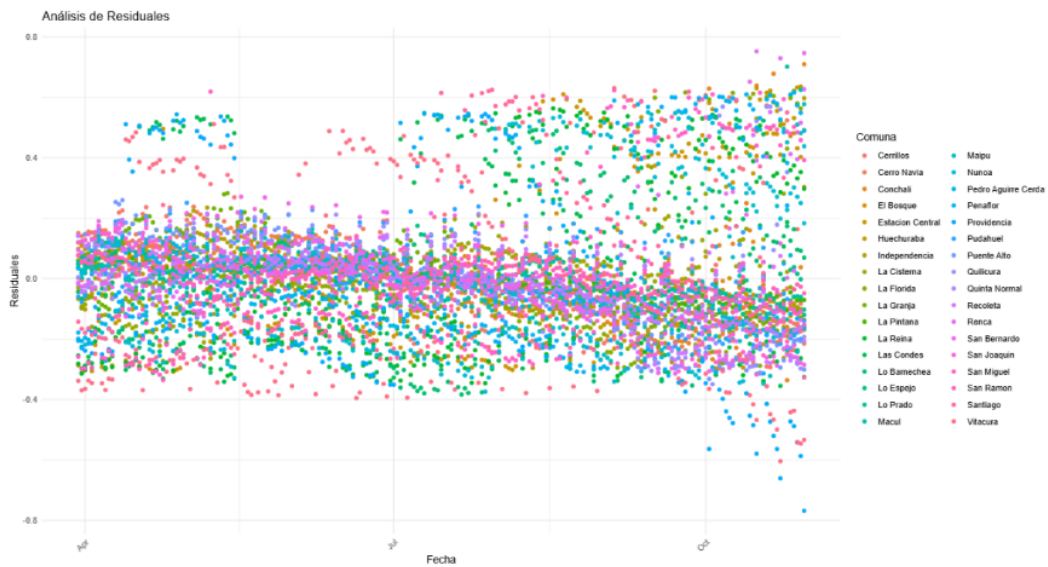


Figura 82. Análisis de Residuales de las Comunas a lo largo del Tiempo Modelo D para la Transición 1 → 3.

A continuación se presentan las comunas con mayor rango de residuales y desviación estándar (Tabla 27), las cuales destacan la heterogeneidad existente en las dinámicas locales, lo que podría estar relacionado con otros factores sociodemográficos y contextuales no considerados en este modelo.

Tabla 27. Comunas con Residuales Extremos en el Modelo D para la Transición 1 → 3.

Comuna	Rango de Residuales	DE
Providencia	-0,768 a 0,557	0,33
Vitacura	-0,604 a 0,527	0,33
Lo Barnechea	-0,386 a 0,531	0,26
Renca	-0,191 a 0,752	0,15
Conchalí	-0,336 a 0,709	0,15
Lo Prado	-0,140 a 0,701	0,06

DE: Desviación Estándar

5.6. Linear Mixed Models para la Transición de Estado de Contagio Moderado a Severo: Transición 2 → 3

En la transición 2→3, de estado moderado a severo (Tabla 28), el Modelo A, presentó un log-likelihood de -2.987 y un AIC de 5.980, con una varianza residual de 0,13084 y una varianza aleatoria asociada a la comuna de 0,00823. El Modelo B, presentó un log-likelihood de -1.899 y un AIC de 3.807, con una varianza residual de 0,08929, y varianzas aleatorias asociadas a la comuna y al día de 0,00842 y 0,04154, respectivamente.

El Modelo C, que incorporó los índices de movilidad interna y externa, mostró una mejora considerable en el ajuste, con un log-likelihood de -1.221 y un AIC de 2.454, con una varianza residual de 0,07147, y varianzas aleatorias asociadas a la comuna y al día de 0,20568 y 0,06711, respectivamente. En este modelo se observa que el aumento en una unidad en el índice de movilidad interna disminuye la probabilidad de transición en un 21,5%, mientras que el aumento en

el índice de movilidad externa incrementa la probabilidad en un 7,5%, siendo ambos efectos estadísticamente significativos ($p < 0,05$).

Por último, el Modelo D, que incluyó tanto los índices de movilidad como variables sociodemográficas, presentó un log-likelihood de -1.238 y un AIC de 2.500, con una varianza residual de 0,07145, una varianza aleatoria asociada a la comuna de 0,03507, y una varianza aleatoria atribuible al día de 0,06760.

En cuanto a los efectos fijos, se observa que un aumento de una unidad en el índice de movilidad interna disminuye la probabilidad de transición de estado moderado a severo en un 21,7%, mientras que un aumento de una unidad en el índice de movilidad externa incrementa la probabilidad en un 7,5%, siendo ambos efectos estadísticamente significativos ($p < 0,05$).

Con respecto a las variables sociodemográficas, se evidenció que el aumento en el promedio de personas por hogar incrementa la probabilidad de transición en un 65,5% ($p < 0,05$). A la vez, el porcentaje de población con educación superior también mostró un efecto positivo y significativo, aumentando la probabilidad de transición en un 2,3% ($p < 0,05$). La cantidad de hogares hacinados y de inmigrantes, también presentaron resultados estadísticamente significativos, sin embargo, el efecto observado en ambas fue mínimo (estimate 0,000). Finalmente, en el caso del ingreso per cápita y la precariedad de la vivienda, los resultados no fueron estadísticamente significativos.

Tabla 28. Resultados de Linear Mixed Models para la Transición 2 → 3.

Modelo	Log-Likelihood	AIC	Varianza Residual	Varianza Aleatoria Comuna	Varianza Aleatoria Día	Efectos Fijos	Estimate	p-value
A	-2987	5980	0,13084	0,00823				
B	-1899	3807	0,08929	0,00842	0,04154			
C	-1221	2454	0,07147	0,20568	0,06711	IM Interno	-0,215	0,000*
						IM Externo	0,075	0,000*
D	-1238	2500	0,07145	0,03507	0,06760	IM Interno	-0,217	0,000*
						IM Externo	0,075	0,000*
						Personas Hogar \bar{x}	0,655	0,000*
						Educación Superior (%)	0,023	0,004*
						Precariedad Vivienda	-0,066	0,067
						Hogares Hacinados	0,000	0,008*
						Cantidad Inmigrantes	0,000	0,002*
						Ingreso Per Cápita	0,000	0,163

En cuanto al análisis de los coeficientes de determinación para esta transición (Tabla 29), el modelo C presentó un R^2 marginal de 0,245, lo que indica que los índices de movilidad explican el 24,5% de la variabilidad en la probabilidad de transición, sin embargo, al considerar los efectos aleatorios asociados a la comuna y al día, el R^2 condicional aumentó a 0,843, lo que refleja que la mayor parte de la varianza es explicada por los efectos aleatorios.

Por otro lado, el modelo D, que incluye tanto los índices de movilidad como las variables sociodemográficas, mostró un incremento notable en el R^2 marginal, alcanzando un valor de 0,692, lo que sugiere que los efectos fijos, en este caso, explican el 69,2% de la variabilidad en la probabilidad de transición al estado severo. Al incorporar los efectos aleatorios, el R^2 condicional del modelo D

alcanzó un valor de 0,874, lo que demuestra que, aunque los efectos fijos explican una proporción considerable de la variabilidad, la inclusión de los efectos aleatorios mejora el ajuste general del modelo.

Tabla 29. Coeficientes de Determinación Marginal y Condicional para la Transición 2 → 3.

Transición	Modelo	R ² Marginal	R ² Condicional
2 → 3	C	0,245	0,843
	D	0,692	0,874

En base a los resultados obtenidos se seleccionó el modelo D como mejor opción para modelar esta transición debido a que presentó mejores coeficientes de determinación.

5.6.1. Efectos Aleatorios por Comuna y Tiempo: Transición 2 → 3

Con respecto al efecto aleatorio por comuna (Figura 83), las localidades que presentaron mayores interceptos ajustados fueron Independencia, Quinta Normal, Recoleta y Las Condes, las cuales reflejan una mayor probabilidad de transición del estado moderado a severo. Por otra parte, Cerrillos, Ñuñoa, Pedro Aguirre Cerda y Santiago mostraron interceptos ajustados negativos, indicando menor probabilidad de transición del estado de contagio moderado a severo (2→3).

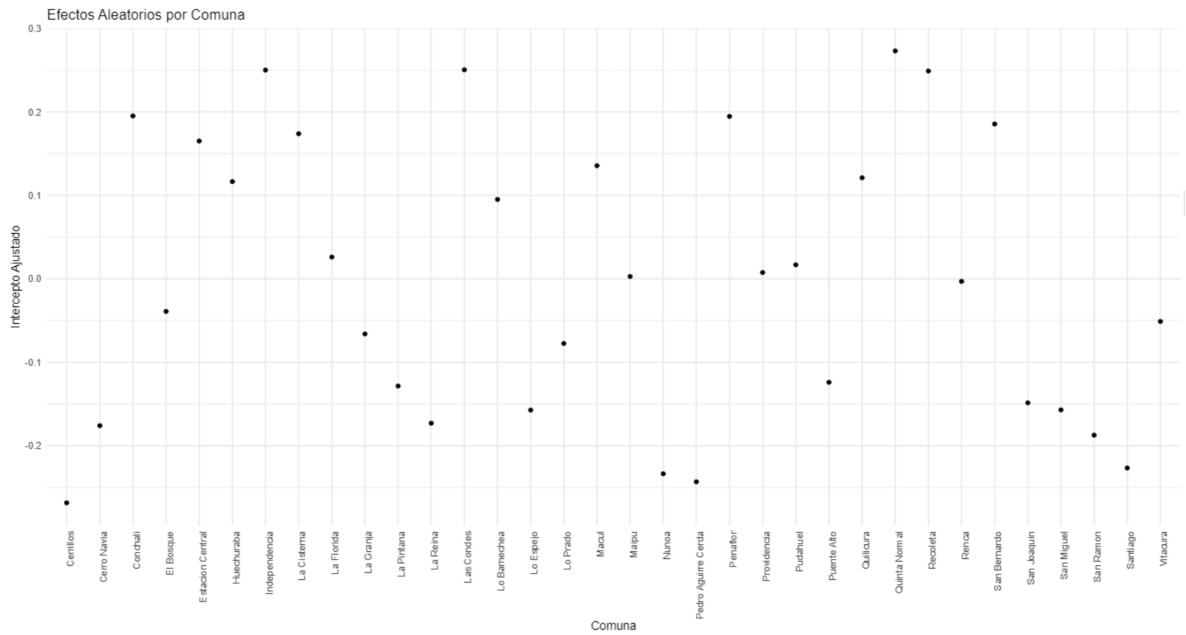


Figura 83. Efectos Aleatorios por Comuna del Modelo D para la Transición 2 → 3.

En cuanto a los efectos aleatorios por día, se puede observar que hasta el mes de julio, los días de semana, fines de semana y festivos presentan una tendencia general de intercepto ajustado negativo, sin embargo, después de julio, se observa un incremento, con valores positivos para ambos tipos de días (Figura 84).

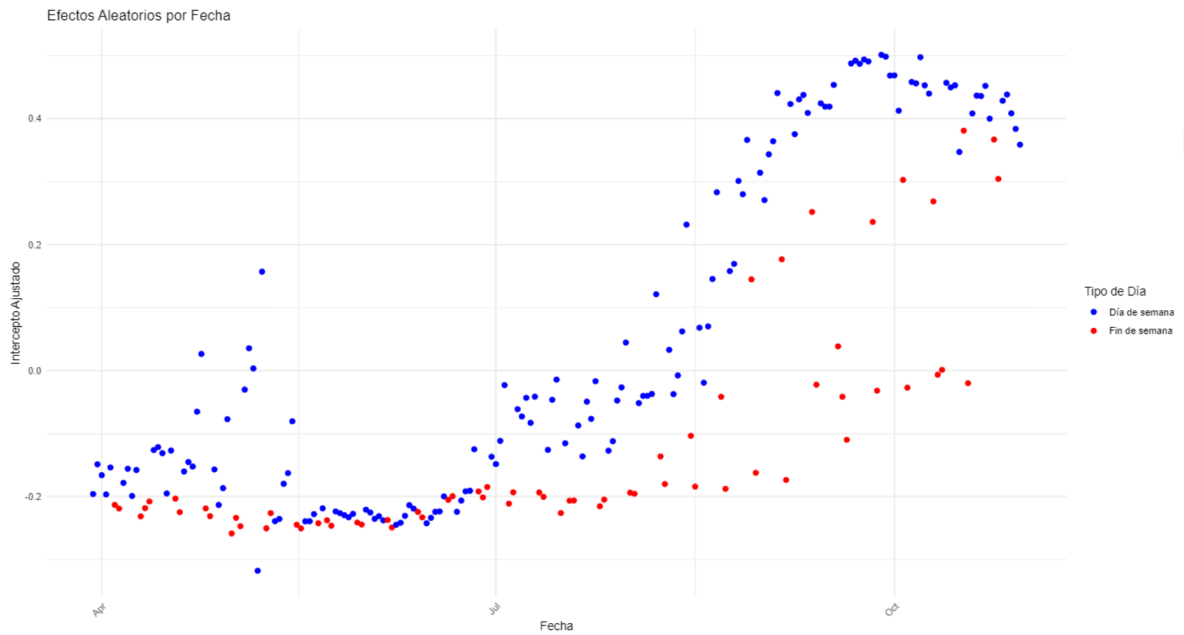


Figura 84. Efectos Aleatorios por Tiempo del Modelo D para la Transición 2 → 3.

5.6.2. Análisis de Residuales: Transición 2 → 3

Al analizar los residuales para la transición de estado moderado a severo, se observa una alta variabilidad a lo largo del tiempo (Figura 85). Durante el primer semestre se puede observar que existió un grupo mayor con residuales cercanos a cero y otro pequeño grupo con residuales extremos sostenido a lo largo del tiempo, sin embargo, pasado el mes de julio se evidencia una alta variabilidad en los residuales a modo general, sin un patrón significativo.

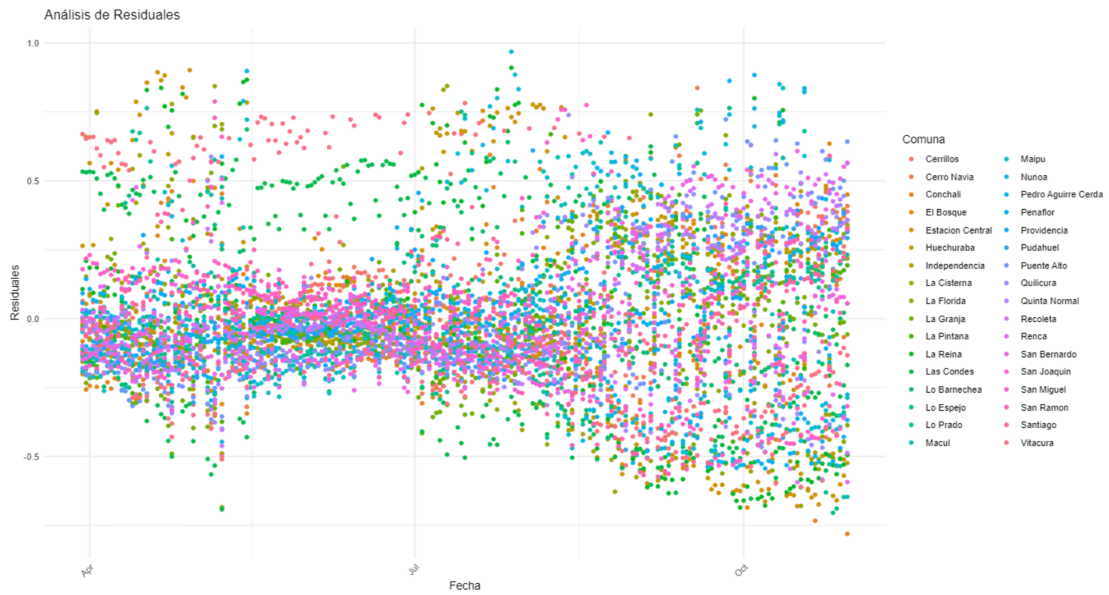


Figura 85. Análisis de Residuales de las Comunas a lo largo del Tiempo Modelo D para la Transición 2 → 3.

A continuación, se presentan las comunas con residuales atípicos, donde se incluyeron las que presentaron mayores y menores valores residuales (Tabla 30).

Tabla 30. Comunas con Residuales Extremos en el Modelo D para la Transición 2 → 3.

Comuna	Rango de Residuales	DE
Conchalí	-0,781 a 0,717	0,25
Lo Espejo	-0,704 a 0,750	0,21
Las Condes	-0,692 a 0,575	0,44
Ñuñoa	-0,538 a 0,967	0,33
La Reina	-0,685 a 0,909	0,41
Estación Central	-0,674 a 0,900	0,43

DE: Desviación Estándar

CONCLUSIONES

La pandemia por Covid-19, declarada por la Organización Mundial de la Salud en marzo de 2020, se convirtió en una de las mayores crisis sanitarias y sociales del último siglo, afectando a millones de personas en todo el mundo. Su rápida propagación y alto impacto en los sistemas de salud, las economías locales y las dinámicas sociales, reforzó la existencia de importantes desigualdades y desafíos en las capacidades de respuesta de las regiones y países.

En contextos urbanos densamente poblados, como la Región Metropolitana de Chile, la alta concentración de personas, las marcadas diferencias socioeconómicas y las dinámicas de movilidad, fueron factores clave que contribuyeron a la heterogeneidad en la evolución de la pandemia, evidenciando la necesidad de herramientas analíticas que permitan entender estos fenómenos de manera más detallada y precisa.

Dada la magnitud de esta problemática y sus implicancias a nivel sanitario, social y económico, resulta fundamental profundizar en el estudio de los factores que determinan la evolución de la pandemia y las diferencias en su impacto entre territorios. A la vez, la alta heterogeneidad en las dinámicas de contagio, influenciada por características como la movilidad y las condiciones sociodemográficas, subraya la necesidad de emplear enfoques analíticos robustos que permitan capturar estos patrones complejos, es por esto que se decidió utilizar modelos avanzados como los Hidden Markov Models no

homogéneos, capaces de identificar patrones latentes en las transiciones entre niveles de contagio y de modelar la influencia de variables contextuales en estas dinámicas.

La utilización de Hidden Markov Model no Homogéneo para el análisis de la evolución de la pandemia Covid-19, permitió identificar y caracterizar tres estados ocultos, los cuales fueron asociados al nivel de contagio: leve, moderado y severo. Además, este modelo permitió obtener las probabilidades de transición entre estos estados, proporcionando información clave sobre los cambios dinámicos en los niveles de contagio.

A partir de estas probabilidades, se decidió utilizar Linear Mixed Models para profundizar en el análisis, considerando como variables explicativas los índices de movilidad interna y externa y las características sociodemográficas de cada comuna, permitiendo el modelamiento de las probabilidades de transición como dependientes de dichas variables explicativas, enriqueciendo la comprensión de los factores que influyen en las dinámicas de contagio.

Este enfoque permitió validar la primera hipótesis planteada al confirmar la existencia de estos estados ocultos y su relación con las transiciones entre niveles de contagio en un contexto dinámico y heterogéneo. Asimismo, la creación del Hidden Markov Model permitió corroborar la segunda hipótesis, al evidenciar que los índices de movilidad interna y externa, junto con las

características sociodemográficas de las comunas, tienen un efecto significativo en las probabilidades de transición entre los estados.

Dentro de los principales resultados obtenidos se evidenció que el aumento en la movilidad interna incrementó las probabilidades de transición hacia estados más graves, siendo pronunciado en la transición de leve a severo, con un aumento del 18,1%. Por otro lado, la movilidad externa presentó un comportamiento complejo, mientras que en algunas transiciones, como de estado leve a moderado, y de moderado a severo, incrementó las probabilidades en un 7% y 7,5% respectivamente, también se observó que una mayor movilidad externa redujo la duración cuando se encontraba en estado severo en un 8%. Por lo tanto, dependiendo del estado y la transición, la movilidad externa puede actuar como un factor de dispersión o disminución del contagio, en vez de agravarlo directamente.

Respecto a las variables sociodemográficas, se puede destacar que, el promedio de personas por hogar tuvo efectos diferenciados, por una parte, redujo la probabilidad de transición del estado leve a severo en un 30,9%, pero incrementó las probabilidades de transición de moderado a severo en un 65,5%, reflejando la importancia del hacinamiento en la propagación y agravamiento del virus. Paralelamente, el porcentaje de educación superior demostró un efecto mitigador en las transiciones hacia estados severos, con una reducción del 1,3% en la probabilidad de transición de leve a severo.

Por otra parte, para evaluar de manera robusta los efectos observados, se implementaron modelos lineales mixtos que incorporaron tanto efectos fijos como aleatorios. El coeficiente de determinación marginal reflejó la contribución de las variables de movilidad y sociodemográficas, destacando su relevancia como predictores de las probabilidades de transición. Paralelamente, el coeficiente de determinación condicional evidenció que gran parte de la variabilidad observada podía explicarse por diferencias estructurales entre comunas y por las dinámicas temporales, enfatizando la importancia de considerar el contexto espacial y temporal en el modelado de los contagios por Covid-19.

A la vez, el análisis de efectos aleatorios por comuna evidenció patrones contradictorios en la evolución del contagio, donde comunas como Las Condes, Lo Barnechea, Providencia y Vitacura indicaron una mayor predisposición a transitar hacia estados más graves (moderado y severo), sin embargo, esto podría explicarse por su exposición inicial al virus, dado su mayor acceso a viajes internacionales y su alta movilidad interna y externa, derivada de su rol económico en la Región Metropolitana. Por otra parte, comunas como San Ramón, Lo Espejo, Pedro Aguirre Cerda y La Pintana mostraron una mayor probabilidad de mantenerse en el estado leve, no obstante, este comportamiento podría estar asociado a subregistros en la detección de casos, derivados de limitaciones en el acceso a pruebas diagnósticas y seguimiento médico.

Adicionalmente, el análisis de residuales reveló que algunas comunas presentaron variaciones considerables en las predicciones del modelo, sobre todo en la mantención del estado leve, lo que podría estar relacionado a variables omitidas que influyen en la dinámica del contagio, como el cumplimiento de las medidas sanitarias. Por lo tanto, este hallazgo podría dar pie a continuar explorando estos factores para mejorar la precisión y validez de los modelos.

Finalmente, este estudio ofrece una herramienta analítica robusta para comprender la dinámica de la pandemia en contextos urbanos densamente poblados. Los resultados ofrecen información relevante para diseñar políticas sanitarias más focalizadas, implementando medidas adaptadas a las características propias de cada comuna, como restricciones de movilidad o campañas de concientización en áreas de mayor hacinamiento, reforzando la importancia de combinar este tipo de análisis con estrategias de intervención contextualizadas para enfrentar desafíos sanitarios futuros en regiones metropolitanas.

TRABAJOS FUTUROS

El presente estudio ha demostrado que los Hidden Markov Models (HMM) no homogéneos son herramientas valiosas para identificar y modelar dinámicas latentes en la propagación del COVID-19. Sin embargo, hay varias áreas en las que se pueden realizar mejoras y extensiones en trabajos futuros.

Una dirección importante es la incorporación de datos temporales extendidos que incluyan fases posteriores de la pandemia, lo cual permitiría analizar los efectos de las variantes del virus, las campañas de vacunación y los cambios en las políticas sanitarias, ofreciendo una visión más completa y dinámica de la evolución de la pandemia. Este enfoque podría contribuir a entender cómo estos factores influenciaron las transiciones entre estados de contagio a lo largo del tiempo.

Otro aspecto clave a considerar es la evaluación de la efectividad de las políticas implementadas, como el plan “Paso a Paso”. A través de pruebas de hipótesis, se podría determinar la eficacia de estas medidas sanitarias en la contención de los contagios y en la reducción de la probabilidad de transitar hacia estados más severos. Complementando este análisis, el uso de variables adicionales como la ocupación hospitalaria, la cobertura de vacunación, las mutaciones del virus y el impacto socioeconómico de las medidas sanitarias, podría mejorar significativamente la comprensión de las transiciones entre estados y enriquecer los modelos existentes con una perspectiva multivariada más avanzada.

Una extensión natural de este trabajo sería abordar la predicción de contagios. Con un modelamiento ajustado correctamente desde el punto de vista explicativo, los estados ocultos identificados podrían utilizarse como base para construir modelos predictivos que anticipen los casos diarios en función de las condiciones actuales y futuras. Esto podría lograrse integrando técnicas de series temporales avanzadas, como SARIMA o LSTM, para modelar tendencias dinámicas con mayor precisión. Asimismo, explorar la aplicabilidad de este enfoque en otros contextos epidemiológicos, como áreas metropolitanas con características diferentes o enfermedades infecciosas con patrones similares, permitiría evaluar su generalización y utilidad en escenarios diversos.

LIMITACIONES

A pesar de los hallazgos relevantes, este estudio presenta varias limitaciones que deben ser consideradas. En primer lugar, los datos analizados se limitan al primer año de la pandemia, excluyendo fases posteriores donde factores como variantes más contagiosas y estrategias de vacunación jugaron un papel crucial, lo que restringe la capacidad del modelo para capturar la evolución completa de la pandemia. Además, el análisis se centró exclusivamente en 34 comunas de la Región Metropolitana, dejando fuera a comunas rurales o áreas con dinámicas poblacionales y sanitarias diferentes. Esto podría limitar la generalización de los resultados a otros contextos.

Otro aspecto a considerar es la falta de interacción espacial explícita en el modelamiento. Aunque los efectos aleatorios por comuna capturan variabilidad local, no se analizaron las interacciones espaciales entre comunas adyacentes, que podrían haber ofrecido información valiosa sobre cómo las conexiones geográficas y de movilidad afectan las transiciones entre estados. Asimismo, los modelos utilizados simplifican ciertas dinámicas complejas del fenómeno, omitiendo variables contextuales relevantes como las dinámicas laborales específicas, las políticas comunales diferenciadas y la percepción social del riesgo. Esto puede limitar la capacidad del modelo para reflejar completamente los factores que impulsan la propagación del virus.

Finalmente, la calidad de los datos utilizados representa una limitación inherente. Los índices de movilidad y las características sociodemográficas podrían estar afectados por sesgos o errores de medición, lo que puede influir en la precisión de las probabilidades de transición calculadas. Además, dado que el enfoque de este estudio fue principalmente explicativo, los modelos no fueron diseñados para realizar predicciones explícitas de casos diarios. Esto limita su aplicabilidad directa en la planificación de políticas preventivas a futuro y subraya la necesidad de desarrollar herramientas predictivas basadas en esta metodología como un siguiente paso.

Pese a estas limitaciones, el presente estudio establece una base metodológica sólida que puede ser mejorada y ampliada en futuros trabajos. La incorporación de datos más amplios y variados, el análisis de la efectividad de políticas específicas y el desarrollo de enfoques predictivos permitirán no solo comprender mejor la evolución de pandemias, sino también diseñar estrategias más efectivas para su control y mitigación en diferentes contextos.

REFERENCIAS

1. Blogs del Banco Mundial [Internet]. [citado 21 de octubre de 2024]. Resumen anual 2020: El impacto de la COVID-19 (coronavirus) en 12 gráficos. Disponible en: <https://blogs.worldbank.org/es/voices/resumen-anual-2020-el-impacto-de-la-covid-19-coronavirus-en-12-graficos>
2. Perucich FV. Ciudad y COVID-19: Desigualdad socio espacial y vulnerabilidad [Internet]. CIPER Chile. 2020 [citado 21 de octubre de 2024]. Disponible en: <https://www.ciperchile.cl/2020/03/25/ciudad-y-covid-19-desigualdad-socio-espacial-y-vulnerabilidad/>
3. Herrera M, Godoy-Faúndez A. Exploring the Roles of Local Mobility Patterns, Socioeconomic Conditions, and Lockdown Policies in Shaping the Patterns of COVID-19 Spread. *Future Internet*. mayo de 2021;13(5):112.
4. Nuevo coronavirus 2019 [Internet]. [citado 23 de octubre de 2024]. Disponible en: <https://www.who.int/es/emergencias/diseases/novel-coronavirus-2019>
5. COVID-19: cronología de la actuación de la OMS [Internet]. [citado 23 de octubre de 2024]. Disponible en: <https://www.who.int/es/news/item/27-04-2020-who-timeline---covid-19>
6. La COVID-19 ha acabado con una década de avances en esperanza de vida a nivel mundial [Internet]. [citado 23 de octubre de 2024]. Disponible en: <https://www.who.int/es/news/item/24-05-2024-covid-19-eliminated-a-decade-of-progress-in-global-level-of-life-expectancy>
7. Caribe CE para AL y el. La prolongación de la crisis sanitaria y su impacto en la salud, la economía y el desarrollo social [Internet]. Comisión Económica para América Latina y el Caribe; 2021 [citado 23 de octubre de 2024]. Disponible en: <https://www.cepal.org/es/publicaciones/47301-la-prolongacion-la-tesis-sanitaria-su-impacto-la-salud-la-economia-desarrollo>
8. CEPAL. El impacto social de la pandemia en América Latina. Panorama Social de América Latina [Internet]. 2021;(8). Disponible en: https://www.cepal.org/sites/default/files/publication/files/46740/S2000182_es.pdf
9. Nations U. United Nations. United Nations; [citado 23 de octubre de 2024]. Estado de la salud mental tras la pandemia del COVID-19 y progreso de la Iniciativa Especial para la Salud Mental (2019-2023) de la OMS | Naciones Unidas. Disponible en: <https://www.un.org/es/cr%C3%B3nica-onu/estado-de-la-salud-mental-tras-la-pandemia-del-covid-19-y-progreso-de-la-iniciativa>
10. Banco Mundial. World Bank. [citado 23 de octubre de 2024]. Los impactos económicos de la crisis de la COVID19. Disponible en:

<https://www.bancomundial.org/es/publication/wdr2022/brief/chapter-1-introduction-the-economic-impacts-of-the-covid-19-crisis>

11. Caribe CE para AL y el. Pandemia provoca aumento en los niveles de pobreza sin precedentes en las últimas décadas e impacta fuertemente en la desigualdad y el empleo [Internet]. Comisión Económica para América Latina y el Caribe; [citado 23 de octubre de 2024]. Disponible en: <https://www.cepal.org/es/comunicados/pandemia-provoca-aumento-niveles-pobreza-sin-precedentes-ultimas-decadas-impacta>

12. Maza Malave M, Celis LG, Maza Malave M, Celis LG. Mutaciones del virus SARS-CoV2 y el impacto de la enfermedad COVID-19 en el sistema de salud. *Revista Salud Uninorte*. diciembre de 2021;37(3):534-8.

13. Arancibia D, Araos R, Basso L, Bravo L, Guridi J, Hartel S, et al. Una mirada a la cultura de datos en Chile: Desafíos y lecciones de la pandemia por Covid-19 [Internet]. Ministerio de Ciencias, Tecnología, Conocimiento e Innovación; 2020. Disponible en: https://www.minciencia.gob.cl/legacy-files/whitepaper_datos_covid19.pdf

14. Chero-Córdova LF, Yabar-Torres G, Chero-Córdova LF, Yabar-Torres G. Vulnerabilidad de la infraestructura sanitaria en relación a la Covid-19 en el distrito de Castilla-Piura 2021- 2022. *Revista de la Facultad de Medicina Humana*. julio de 2023;23(3):35-45.

15. Medscape [Internet]. [citado 23 de octubre de 2024]. Efectos de la saturación hospitalaria en la mortalidad por COVID-19 grave. Disponible en: <https://espanol.medscape.com/verarticulo/5906609>

16. Herath S, Mansour A, Bentley R. Urban density, household overcrowding and the spread of COVID-19 in Australian cities. *Health & Place*. 1 de septiembre de 2024;89:103298.

17. Caribe CE para AL y el. Mortalidad por COVID-19 y las desigualdades por nivel socioeconómico y por territorio [Internet]. Comisión Económica para América Latina y el Caribe; [citado 23 de octubre de 2024]. Disponible en: <https://www.cepal.org/es/enfoques/mortalidad-covid-19-desigualdades-nivel-socioeconomico-territorio>

18. Alkire S, Nogales R, Oldiges C. Multidimensional Poverty and COVID-19 Risk Factors: A Rapid Overview of Interlinked Deprivations across 5.8 Billion People | OPHI. OPHI Briefing 53 [Internet]. 2020 [citado 23 de octubre de 2024]; Disponible en: <https://ophi.org.uk/publications/B-53a>

19. Gallardo PS. COVID-19: la vulnerabilidad en el ojo del huracán. *Enfermería Clínica*. 23 de mayo de 2020;31:S2.

20. Mena GE, Martínez PP, Mahmud AS, Marquet PA, Buckee CO, Santillana

- M. Socioeconomic status determines COVID-19 incidence and related mortality in Santiago, Chile. *Science*. 28 de mayo de 2021;372(6545):eabg5298.
21. Antoñanzas Serrano A, Gimeno Feliu LA, Antoñanzas Serrano A, Gimeno Feliu LA. Los determinantes sociales de la salud y su influencia en la incidencia de la COVID-19. Una revisión narrativa. *Revista Clínica de Medicina de Familia*. 2022;15(1):12-9.
22. Sobre vulnerabilidad y teletrabajo durante la pandemia [Internet]. Facultad de Economía y Negocios. 2020 [citado 23 de octubre de 2024]. Disponible en: <https://fen.uahurtado.cl/2020/articulos/sobre-vulnerabilidad-y-teletrabajo-durante-la-pandemia/>
23. Guntin R. Trabajo a Distancia y con Contacto en Uruguay. 2020; Disponible en: https://www.rguntin.com/research/employment_u ru_covid_new.pdf
24. Almagro M, Coven J, Gupta A, Orane-Hutchinson A. Disparities in COVID-19 risk exposure: Evidence from geolocation data. *Regional Science and Urban Economics*. 1 de septiembre de 2023;102:103933.
25. Kamis C, Stolte A, West JS, Fishman SH, Brown T, Brown T, et al. Overcrowding and COVID-19 mortality across U.S. counties: Are disparities growing over time? *SSM - Population Health*. 1 de septiembre de 2021;15:100845.
26. Russette H, Graham J, Holden Z, Semmens EO, Williams E, Landguth EL. Greenspace exposure and COVID-19 mortality in the United States: January–July 2020. *Environmental Research*. 28 de abril de 2021;198:111195.
27. Kraemer MUG, Yang CH, Gutierrez B, Wu CH, Klein B, Pigott DM, et al. The effect of human mobility and control measures on the COVID-19 epidemic in China. *Science*. mayo de 2020;368(6490):493-7.
28. Badr HS, Du H, Marshall M, Dong E, Squire MM, Gardner LM. Association between mobility patterns and COVID-19 transmission in the USA: a mathematical modelling study. *The Lancet Infectious Diseases*. 1 de noviembre de 2020;20(11):1247-54.
29. Lami F, Khaleel HA, Khader YS. Mobility indicators and COVID-19 growth ratio in Iraq: a correlation study. *Journal of Public Health (Oxford, England)*. 27 de septiembre de 2021;43(Suppl 3):iii29.
30. Gao S, Rao J, Kang Y, Liang Y, Kruse J, Dopfer D, et al. Association of Mobile Phone Location Data Indications of Travel and Stay-at-Home Mandates With COVID-19 Infection Rates in the US. *JAMA Netw Open*. 1 de septiembre de 2020;3(9):e2020485.
31. Tapiador L, Gomez J, Vassallo JM. Exploring the relationship between

public transport use and COVID-19 infection: A survey data analysis in Madrid Region. *Sustainable Cities and Society*. 1 de mayo de 2024;104:105279.

32. Pepe E, Bajardi P, Gauvin L, Privitera F, Lake B, Cattuto C, et al. COVID-19 outbreak response, a dataset to assess mobility changes in Italy following national lockdown. *Sci Data*. 8 de julio de 2020;7(1):230.

33. Benvenuto D, Giovanetti M, Vassallo L, Angeletti S, Ciccozzi M. Application of the ARIMA model on the COVID-2019 epidemic dataset. *Data in Brief*. 26 de febrero de 2020;29:105340.

34. Flaxman S, Mishra S, Gandy A, Unwin HJT, Mellan TA, Coupland H, et al. Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature*. agosto de 2020;584(7820):257-61.

35. Yuan H, Yeung A, Yang W. Interactions among common non-SARS-CoV-2 respiratory viruses and influence of the COVID-19 pandemic on their circulation in New York City. *Influenza and Other Respiratory Viruses*. 12 de marzo de 2022;16(4):653.

36. Ibrahim RR, Oladipo HO. Forecasting the spread of COVID-19 in Nigeria using Box-Jenkins Modeling Procedur [Internet]. 2020 [citado 24 de octubre de 2024]. Disponible en: <https://europepmc.org/article/PPR/PPR159867>

37. Talkhi N, Akhavan Fatemi N, Ataei Z, Jabbari Nooghabi M. Modeling and forecasting number of confirmed and death caused COVID-19 in IRAN: A comparison of time series forecasting methods. *Biomedical Signal Processing and Control*. 1 de abril de 2021;66:102494.

38. Yonar H, Yonar A, Tekindal MA, Tekindal M. Modeling and Forecasting for the number of cases of the COVID-19 pandemic with the Curve Estimation Models, the Box-Jenkins and Exponential Smoothing Methods. *Eurasian Journal of Medicine and Oncology*. 2020;4(2):160-5.

39. Ghafouri-Fard S, Mohammad-Rahimi H, Motie P, Minabi MAS, Taheri M, Nateghinia S. Application of machine learning in the prediction of COVID-19 daily new cases: A scoping review. *Heliyon*. 1 de octubre de 2021;7(10):e08143.

40. Andrenacci G. Predicting Time-Series with SARIMAX [Internet]. *Data Bistrot*. 2024 [citado 25 de octubre de 2024]. Disponible en: <https://medium.com/data-bistrot/predicting-time-series-with-sarimax-f805f6608cd2>

41. Okutan P. Time Series Forecasting: A Comparative Analysis of SARIMAX, RNN, LSTM, Prophet, and Transformer... [Internet]. *Medium*. 2024 [citado 25 de octubre de 2024]. Disponible en: <https://medium.com/@pelinokutan/time-series-forecasting-a-comparative-analysis-of-sarimax-rnn-lstm-prophet-and-transformer-46c85d960b4b>

42. G K. DeepAR: Probabilistic Forecasting with Autoregressive Recurrent Networks [Internet]. Medium. 2022 [citado 25 de octubre de 2024]. Disponible en: <https://kshavg.medium.com/deepar-probabilistic-forecasting-with-autoregressive-recurrent-networks-fa65dddae1f5>
43. Maklin C. DeepAR Forecasting Algorithm [Internet]. Medium. 2022 [citado 25 de octubre de 2024]. Disponible en: <https://medium.com/@corymaklin/deepar-forecasting-algorithm-6555efa63444>
44. Zhou X, Li Q, Zhu Z, Zhao H, Tang H, Feng Y. Monitoring Epidemic Alert Levels by Analyzing Internet Search Volume. *IEEE Transactions on Biomedical Engineering*. febrero de 2013;60(2):446-52.
45. Bento C. Medium. 2023 [citado 25 de octubre de 2024]. Hidden Markov Models Explained with a Real Life Example and Python code. Disponible en: <https://towardsdatascience.com/hidden-markov-models-explained-with-a-real-life-example-and-python-code-2df2a7956d65>
46. DHIMAN A. Medium. 2023 [citado 25 de octubre de 2024]. Time Series forecasting in Python with Hidden Markov Models. Disponible en: <https://medium.datadriveninvestor.com/time-series-forecasting-in-python-with-hidden-markov-models-dac146592237>
47. Zhou S, Braca P, Marano S, Willett P, Millefiori LM, Gaglione D, et al. Application of Hidden Markov Models to Analyze, Group and Visualize Spatio-Temporal COVID-19 Data. *IEEE Access*. 2021;9:134384-401.

