



**Universidad del Desarrollo**  
Facultad de Ingeniería

IMPACTO DE LAS VARIABLES  
SOCIOECONÓMICAS Y DEMOGRÁFICAS EN LA  
PROPAGACIÓN DEL SARS-COV-2 EN LA RM DE CHILE,  
CON MODELOS RANDOM FOREST REGRESSION Y SARIMAX.

POR:

ANDRES ALEXIS LEIVA HERNÁNDEZ

FRANCISCO JAVIER SÁNCHEZ LEÓN

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:

SR. MAURICIO HERRERA.

12 2023

SANTIAGO.

## DEDICATORIAS

Quiero dedicar este logro a mi novia Alejandra Silva. Agradezco tu paciencia, comprensión y aliento constante, que han sido el motor que impulsa cada logro alcanzado. Este trabajo no solo es el resultado de esfuerzos individuales, sino también el fruto de una relación de confianza y apoyo mutuo.

Andrés Leiva.

Quiero dar las gracias a mi pareja por acompañarme en este gran paso en mi carrera profesional, agradezco tu paciencia en comprender que para lograr los objetivos se debió invertir mucho tiempo el cual pude haber dedicado el compartir contigo las metas que tenemos juntos son y serán siempre mi motivación para alcanzar este logro académico.

Gracias por compartir este viaje conmigo, por celebrar cada pequeño triunfo como tuyo.

Con todo mi amor y gratitud,

Francisco Sánchez

## AGRADECIMIENTOS

A mi compañero de tesis, Andrés Leiva.

Quiero expresar mi profundo agradecimiento por tu dedicación en este proyecto final, tu compromiso con este proyecto ha sido invaluable.

Aprecio la paciencia que has demostrado en los momentos difíciles, tu disposición constante para compartir ideas brillantes y tu habilidad para trabajar en equipo.

Espero que este sea solo el comienzo de muchos éxitos que tengas en tu carrera profesional.

Con gratitud,

Francisco Sánchez

Quiero agradecerle a mi compañero de tesis, Francisco Sánchez.

Trabajar contigo en este proyecto fue una experiencia invaluable. Aprecio sinceramente tu dedicación, colaboración y la manera en que enfrentamos juntos los desafíos. Tu aporte ha sido fundamental para alcanzar estos resultados.

Gracias por compartir este viaje académico y hacerlo más llevadero. La sinergia entre nosotros hizo que este proceso fuera más enriquecedor y, sin duda, contribuyó al éxito de nuestra tesis.

Con gratitud,

Andrés Leiva.

## TABLA DE CONTENIDO

<b>RESUMEN</b>	<b>1</b>
<b>1. INTRODUCCIÓN</b>	<b>2</b>
<b>2. TRABAJO RELACIONADO</b>	<b>4</b>
<b>3. HIPÓTESIS Y OBJETIVOS</b>	<b>6</b>
<b>4. DATOS Y METODOLOGÍA</b>	<b>7</b>
4.1. DATOS	7
4.2. METODOLOGÍA	9
<b>5. RESULTADOS</b>	<b>11</b>
5.1. ANÁLISIS EXPLORATORIO	11
4.2. Modelo	23
<b>6. CONCLUSIÓN Y TRABAJO FUTURO</b>	<b>28</b>
<b>BIBLIOGRAFÍA</b>	<b>29</b>

## Resumen

El análisis exploratorio de datos de este estudio reveló una propagación inicial del COVID-19 desde el sector oriente de Santiago, caracterizado por una mejor condición económica con respecto al resto de la RM. A medida que avanzaba el tiempo, la enfermedad se expandió hacia áreas más densamente pobladas y finalmente se concentró en la zona sur, asociada a un menor ingreso per cápita. Por otro lado, la movilidad, inicialmente entre la zona sur y el centro, evolucionó hacia desplazamientos entre comunas oriente y centro. La zona central mostró mayor hacinamiento y densidad poblacional, aunque esta última se ve influenciada por áreas periféricas con baja urbanización.

Al implementar un modelo Random Forest Regression se utilizaron las variables predictoras Hacinamiento, Ingreso Per Capita, Densidad Poblacional e Índice de Movilidad, en donde se identificó que el hacinamiento fue crucial en el primer periodo (entre marzo y mayo del año 2020), pero la capacidad predictiva fue limitada ( $R^2 = 0,57$ ), esto significa que las variables exógenas del modelo no son capaces de explicar el comportamiento de la tasa de contagios. En el pick de casos (alrededor de junio y julio del año 2020), el hacinamiento y el ingreso per cápita fueron determinantes, mejorando significativamente la capacidad predictiva ( $R^2 = 0,88$ ) y representando entre ambas variables casi un 80% el comportamiento de la tasa de contagios. En el resto del año, los ingresos ganaron relevancia en la predicción ( $R^2 = 0,98$ ), teniendo un porcentaje de representatividad cercano al 42% por sí solo. En estos 3 periodos, el Índice de Movilidad no resultó ser determinístico en la predicción.

El modelo SARIMAX, inicialmente impreciso con respecto a la amplitud de la curva pero que seguía las tendencias en cambios, este mostró mejoras al considerar dos periodos separados, uno antes del pick de casos contagiados y otro después. Con esto mejoró considerablemente la predicción, destacando que la movilidad, si bien no explicaba la mayoría de los casos por sí sola, fue un factor importante en la propagación del virus.

## **1. Introducción**

La irrupción del COVID-19 a nivel mundial marcó un hito crítico en la historia contemporánea, desencadenando una serie de eventos sin precedentes que afectaron todos los aspectos de la vida cotidiana. La respuesta a esta pandemia se caracterizó por la implementación de diversas medidas gubernamentales, siendo una de las más prominentes la reducción de la movilidad como estrategia para frenar la propagación del virus. En el caso específico de Chile, estas medidas fueron adoptadas con celeridad, buscando proteger la salud de la población y mitigar el impacto en el sistema de salud.

Sin embargo, la efectividad de estas medidas ha sido objeto de debate y análisis. La presente tesis surge en un contexto donde se plantea la hipótesis de que las estrategias gubernamentales, centradas principalmente en la reducción de la movilidad, pueden haber pasado por alto variables socioeconómicas y demográficas cruciales en la propagación del virus. Este supuesto se fundamenta en la idea de que la comprensión completa de los factores que influyen en la transmisión del COVID-19 va más allá de la movilidad y requiere un análisis integral que abarque dimensiones como el hacinamiento, la densidad poblacional y los ingresos.

Chile, en su diversidad geográfica y demográfica, proporciona un escenario idóneo para explorar estas relaciones complejas. Este estudio se concentra en la Región Metropolitana, específicamente en Santiago, donde la interconexión entre la movilidad urbana, las condiciones socioeconómicas y la propagación del virus ofrece un terreno fértil para la investigación. A través de un análisis detallado de datos epidemiológicos y socioeconómicos, este proyecto busca arrojar luz sobre cómo estas variables interactúan y contribuyen a la dinámica de la pandemia.

El análisis exploratorio de datos (EDA) revela patrones intrigantes, desde el inicio del virus en los sectores económicamente más acomodados hasta su propagación en áreas densamente pobladas y, finalmente, su concentración en zonas con mayores índices de pobreza. El contexto de movilidad también ha evolucionado, transformando la dinámica del virus a lo largo del tiempo. La movilidad inicial entre zonas específicas ha dado paso a patrones diferentes que podrían tener implicaciones significativas en la propagación del virus.

Este estudio utiliza modelos analíticos avanzados, como Random Forest Regression y SARIMAX, para profundizar en la influencia relativa de variables clave en diferentes periodos de la pandemia. El análisis se enfoca en el hacinamiento, la densidad poblacional y los ingresos, examinando cómo estas variables se relacionan con la propagación del COVID-19. A través de este enfoque integral, se busca no solo entender el pasado, sino también aportar a la formulación de estrategias más efectivas para el manejo de futuras crisis sanitarias.

En resumen, esta investigación se adentra en la complejidad de la dinámica de la pandemia en Santiago, Chile, explorando más allá de la movilidad para comprender las variables socioeconómicas y demográficas que han moldeado la propagación del COVID-19. Los resultados de este estudio no solo tendrán implicaciones para el abordaje de la pandemia en Chile sino también para la formulación de políticas de salud pública a nivel global.

## 2. Trabajo Relacionado

Durante la pandemia en la Región Metropolitana, el fenómeno del hacinamiento en viviendas ha sido una preocupación significativa.

En algunas comunas del Gran Santiago, la pobreza multidimensional, caracterizada por elementos como trabajo informal, malnutrición y baja escolaridad, se asocia positiva y significativamente con mayores niveles de contagios por COVID-19. El hacinamiento emerge como una variable clave, correlacionando positivamente tanto con la tasa de contagios como con la de fallecidos, según un análisis estadístico. Se exploran cinco dimensiones de pobreza multidimensional (educación, salud, vivienda, trabajo y seguridad social, entorno y redes), así como cinco indicadores adicionales relacionados con la vulnerabilidad al virus (escolaridad, malnutrición, seguridad social, hacinamiento y estado de la vivienda). Los mapas clasifican las tasas de pobreza y contagios en tres grupos: bajas, medias y altas. (CIPER (2012), “*Hacinamiento: la variable clave en la propagación del Covid-19 en el Gran Santiago*” (artículo))

El hacinamiento, que se define como la presencia de un número excesivo de personas en relación con el espacio disponible en una vivienda, puede tener importantes implicaciones en la propagación de enfermedades infecciosas, incluyendo el COVID-19. En este contexto, varios factores contribuyen a la variable de hacinamiento en esta región durante la crisis sanitaria.

Uno de los principales factores es la prevalencia de viviendas multifamiliares. En las áreas urbanas, especialmente en regiones metropolitanas, es común encontrar edificios de departamentos o complejos residenciales que albergan a múltiples familias en unidades de vivienda relativamente pequeñas. Esta concentración de personas en un espacio limitado puede aumentar las posibilidades de propagación del virus, ya que el distanciamiento social se vuelve más difícil de lograr en entornos con alta densidad de población.

Las condiciones socioeconómicas también desempeñan un papel crucial en el problema del hacinamiento. Aquellas personas y familias con recursos limitados pueden tener dificultades para acceder a viviendas más grandes o para mudarse a lugares con condiciones de vida más espaciales. La falta de recursos económicos puede limitar las

opciones de vivienda, lo que resulta en una mayor concentración de personas en un mismo espacio, incrementando así el riesgo de transmisión de enfermedades.

También podemos apreciar según diversos estudios como estas variables son parte importante en la propagación del covid 19 en el gran santiago.

El informe destaca que diversas características comunales influyen en la efectividad de las cuarentenas durante la pandemia. Se compararon la densidad poblacional, los ingresos, la informalidad laboral, el nivel de trabajadores no calificados y el uso de transporte público en diferentes comunas. Aquellas con alta densidad, bajos ingresos, alto trabajo no calificado y mayor uso de transporte público muestran una correlación con la ineficacia de las cuarentenas. La combinación de estos factores hace que el confinamiento sea menos efectivo en estas áreas. (ministerio de Ciencia, Tecnología, Conocimiento e Innovación y CASEN (2020), “*Encuesta de Caracterización Socioeconómica Nacional*” (encuesta)).

Artículo (Epidemic prediction of COVID-19” )

Se destaca como gracias a los datos recopilados de diferentes fuentes de datos se pudo implementar un modelo dinámico de transmisión del COVID-19 y un modelo estadístico basado en el análisis de series temporales por lo cual basándose en esto podemos ver que es factible implementar un modelo de series de tiempo, en el cual se compara los efectos de predicción de estos modelos matemáticos en la propagación de la epidemia de COVID-19. Debido a que el brote de datos existentes no es un conjunto de datos de muestra relativamente grande, en esta etapa de la propagación del COVID-19,

también podemos ver cómo al igual que Chile el mundo tuvo problemas para poder contener el covid-19 debido que es un virus de muy rápida propagación en el artículo Epidemic prediction of COVID-19”

El gobierno chino de varios niveles ha intervenido de diversas maneras para controlar la epidemia. Según los resultados del análisis del modelo, se cree que las medidas de intervención de emergencia adoptadas en las etapas iniciales de la epidemia, como el bloqueo de Wuhan, la restricción del flujo de personas en la provincia de Hubei y el aumento del apoyo a Wuhan, tuvieron un efecto crucial para frenar la propagación original de la epidemia. Aumentar la inversión en diversos recursos médicos para garantizar que los pacientes sospechosos puedan ser diagnosticados y tratados de manera oportuna es un método muy efectivo de prevención y tratamiento. Según los resultados del análisis de sensibilidad, creemos que mejorar el tratamiento de los

## Artículo El nivel socioeconómico determina la incidencia de COVID-19 y la mortalidad relacionada en Santiago de Chile

indica que Santiago está altamente segregada con distintas zonas de riqueza y privación. Este escenario ofrece una ventana a cómo los factores sociales impulsan la pandemia del coronavirus 2 del síndrome respiratorio agudo severo (SARS-CoV-2) en una sociedad económicamente vulnerable con altos niveles de desigualdad de ingresos. Se analizaron la incidencia y mortalidad atribuidas al SARS-CoV-2 para comprender las variaciones espaciales en la carga de la enfermedad. Las tasas de letalidad por infección fueron más altas en los municipios de menores ingresos debido a las comorbilidades y la falta de acceso a la atención médica. Las disparidades entre los municipios en cuanto a la calidad de su sistema de prestación de servicios de salud se hicieron evidentes en los retrasos en las pruebas y en la capacidad. Estos indicadores explican gran parte de la variación en el subregistro y las muertes por COVID-19.

### **3. Hipótesis y Objetivos**

La hipótesis subyacente en este proyecto de tesis sostiene que las medidas implementadas en Chile podrían no haber sido completamente efectivas debido a su enfoque exclusivo en la movilidad, descuidando factores críticos como el hacinamiento, la densidad poblacional y el ingresos per cápita por comuna. Se postula que estas variables desempeñan un papel crucial en la dinámica de la enfermedad, impactando directamente en las tasas de contagio y propagación del virus.

El hacinamiento, por ejemplo, se presenta como un factor de riesgo, ya que las familias con condiciones de vida más apretadas podrían experimentar mayores dificultades para practicar el distanciamiento social o el aislamiento en caso de exposición al virus. Por otro lado, la densidad poblacional, especialmente en áreas urbanas como Estación Central y Santiago Centro, podría amplificar la probabilidad de contagio debido a la concentración de personas y la presencia de edificaciones de gran altura.

La variable de ingreso per cápita también emerge como un elemento clave, ya que las familias con bajos ingresos suelen estar vinculadas a trabajos presenciales, donde el teletrabajo no es una opción viable. Esto podría exponer a estos grupos a un mayor riesgo de infección, destacando la necesidad de considerar aspectos socioeconómicos en las estrategias de mitigación.

En este contexto, el presente estudio busca analizar de manera integral la relación entre estas variables y la propagación del COVID-19 en Chile, proporcionando una evaluación crítica de la efectividad de las medidas implementadas y ofreciendo insights que podrían contribuir a estrategias más robustas y equitativas para el manejo de futuras crisis sanitarias.

#### **Objetivo principal:**

Evaluar la capacidad predictiva exclusiva de la variable Índice de Movilidad para anticipar el comportamiento del virus SARS-CoV-2 (COVID-19).

**Objetivos secundarios:**

1. Analizar el impacto de factores socioeconómicos en la propagación del COVID-19, focalizando la atención en las siguientes variables:

- Hacinamiento.
- Densidad Poblacional.
- Ingreso Per Cápita.

2. Explorar la relación entre variables demográficas y la expansión del virus SARS-CoV-2 (COVID-19), priorizando el examen de las siguientes variables:

- Hacinamiento.
- Densidad Poblacional.
- Ingreso Per Cápita.

Estos objetivos se desarrollarán mediante un análisis estadístico riguroso y la aplicación de métodos predictivos para determinar la viabilidad de la variable Índice de Movilidad como indicador exclusivo de comportamiento del COVID-19, así como para comprender el impacto de las variables socioeconómicas y demográficas mencionadas en la propagación del virus.

## 4. Datos y Metodología

### 4.1. Datos

Nuestro capstone Project comenzó con la definición cuidadosa de datos de entrada con el fin de poder construir un conjunto de datos sólido. Esta selección minuciosa es determinante para garantizar la robustez de nuestro modelo.

Exploramos la diversidad de estos conjuntos de datos, cada uno aportando una perspectiva única y esencial para la realización de nuestro análisis. Este proceso de construcción del conjunto de datos no solo fue técnico, sino también estratégico, enfocándonos en la calidad de los datos que obtuvimos.

los cuales se obtuvieron desde el Ministerio de Ciencia, Tecnología, Conocimiento e Innovación.

Nombre y estructura de los dataset ocupados:

- Cuarentenas-Activas.csv
  - ID: Identificador de cada comuna.
  - Nombre: Nombre de cada comuna.
  - Estado: Estado en que se encuentra la comuna.
  - Alcance: Indica qué porcentaje de la comuna está en cuarentena.
  - Fecha de Inicio: Inicio de la cuarentena.
  - Fecha de término: Fin de la cuarentena.
  - Código CUT Comuna: Corresponde al código único territorial.
  - Detalle: Detalle de la cuarentena.
  - Superficie en m2: Superficie en metros cuadrados de la comuna.
  - Perímetro en m: Perímetro de la comuna expresadas en metro.
  
- Cuarentenas-Históricas.csv
  - ID: Identificador de cada comuna.
  - Nombre: Nombre de cada comuna.
  - Estado: Estado en que se encuentra la comuna.
  - Alcance: Indica qué porcentaje de la comuna está en cuarentena.
  - Fecha de Inicio: Inicio de la cuarentena.
  - Fecha de término: Fin de la cuarentena.

- Código CUT Comuna: Corresponde al código único territorial.
- Detalle: Detalle de la cuarentena.
- Superficie en m2: Superficie en metros cuadrados de la comuna.
- Perímetro en m: Perímetro de la comuna expresadas en metro.
  
- Cuarentenas-Totales.csv
  - ID: Identificador de cada comuna.
  - Nombre: Nombre de cada comuna.
  - Estado: Estado en que se encuentra la comuna.
  - Alcance: Indica qué porcentaje de la comuna está en cuarentena.
  - Fecha de Inicio: Inicio de la cuarentena.
  - Fecha de término: Fin de la cuarentena.
  - Código CUT Comuna: Corresponde al código único territorial.
  - Detalle: Detalle de la cuarentena.
  - Superficie en m2: Superficie en metros cuadrados de la comuna.
  - Perímetro en m: Perímetro de la comuna expresadas en metro.
  - Región: Número de la región.
  - Número de región: Código de la región.
  
- Ingreso per cápita.csv
  - Comuna: Nombre de la comuna.
  - Ingreso: Corresponde al ingreso per cápita por comuna, este valor está expresado en pesos.
  
- paso\_a\_paso.csv
  - Codigo\_region: Corresponde al código de la región.
  - Region\_residencia: Corresponde a la región de Chile.
  - Codigo\_comuna: código de la comuna.
  - Comuna\_residencia: comuna de la región.
  - Zona: indica si es zona rural, urbana o total.
  - 28-07-2020: fecha que indica en qué paso está la comuna .
  - 29-07-2020: fecha que indica en qué paso está la comuna
  
- Índice de movilidad.csv
  - codigo\_comuna : código de la comuna.
  - comuna: Nombre de la comuna
  - IM: Corresponde al índice de movilidad, estas fueron obtenidas desde las antenas de telecomunicaciones de la empresa telefónica.

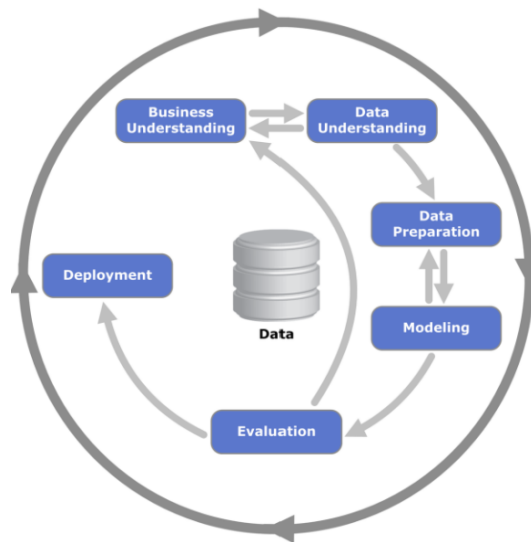
## 4.2. Metodología

La metodología que se implementara en el proyecto se llama CRISP-DM la cual tiene una serie de etapas que se deben implementar de forma correcta para tener éxito en la implementación de un modelo de data science

Podemos destacar que CRISP-DM no es un proceso lineal estricto; las fases pueden superponerse y retroceder en función de las necesidades específicas del proyecto. Además, la comunicación continua con los interesados y la adaptación a cambios en los objetivos comerciales son fundamentales para el éxito del proyecto.

**Figura 1**

Metodología CRISP-DM



Según esta metodología, el análisis se abordará de la siguiente forma:

### **1. Recopilación de Datos**

Se recopilaron datos detallados de las variables mencionadas para varias ubicaciones geográficas en las fechas de interés. Los datos de casos de COVID-19 fueron utilizados como variable objetivo, mientras que el "Índice de Movilidad", "Hacinamiento", "Población", "Densidad Poblacional" e "Ingreso Per Cápita" se consideraron como variables predictoras.

### **2. Análisis Exploratorio de Datos**

Se realizó un análisis exploratorio para comprender la distribución, tendencias y posibles correlaciones entre las variables. Se utilizaron gráficos y estadísticas descriptivas para visualizar la información.

### **3. Modelado de datos**

#### **Random Forest Regression**

Se implementó un modelo Random Forest Regression para calcular el peso de cada variable en la cantidad de casos de COVID-19 en tres fechas específicas. Este enfoque permitirá identificar la importancia relativa de cada variable en el contexto de las fechas seleccionadas.

#### **Modelado con SARIMAX**

Se aplicó un modelo SARIMAX (Seasonal Autoregressive Integrated Moving Average) para realizar predicciones a lo largo del tiempo. El "Índice de Movilidad" se consideró como la variable principal en este modelo, ya que se busca comprender y prever su impacto en la evolución de los casos de COVID-19.

### **4. Evaluación de Modelos**

Se llevaron a cabo evaluaciones detalladas de ambos modelos, considerando métricas como el error cuadrático medio (MSE), el coeficiente de determinación ( $R^2$ ) y otras

métricas relevantes. Esto permitirá validar la efectividad de los modelos en la explicación de la variabilidad de los casos de COVID-19.

## 5. Interpretación de Resultados

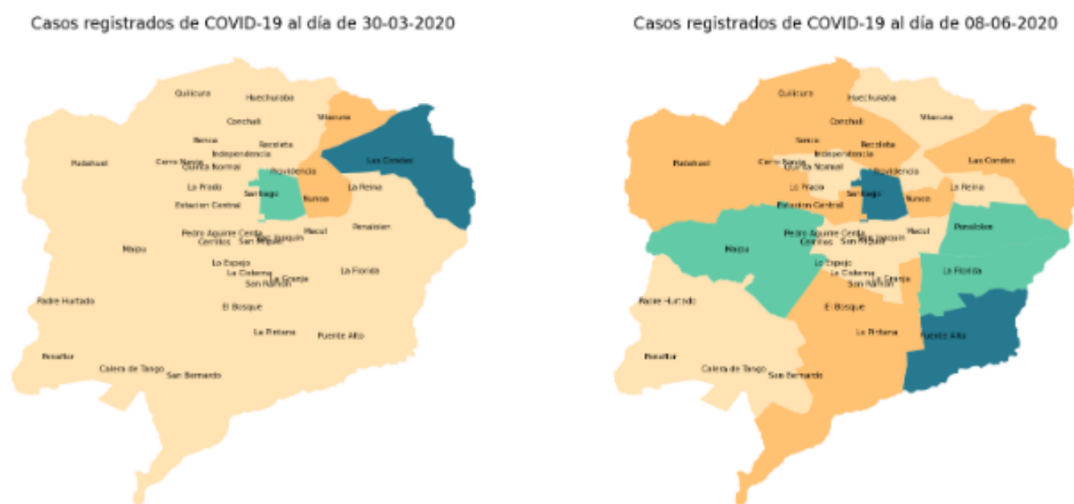
Los resultados obtenidos de los modelos Random Forest Regression y SARIMAX se interpretaron en conjunto para comprender el impacto relativo de las variables analizadas en la propagación del COVID-19. Se prestará especial atención a las tendencias identificadas y a las predicciones a largo plazo basadas en el "Índice de Movilidad".

## 5. Resultados.

### 5.1. Análisis Exploratorio

Figura 2

Cantidad de casos confirmados de COVID-19 (30-03-2020 -08-06-2020)



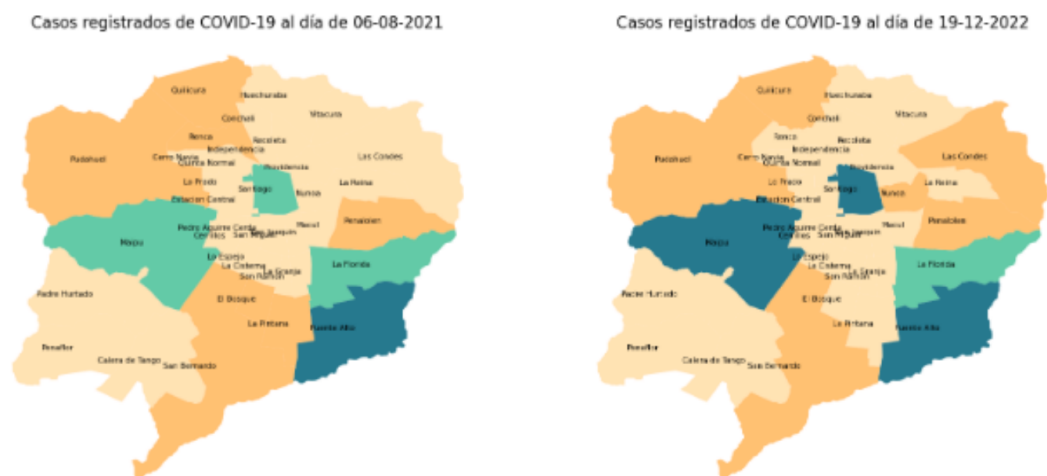
En la figura 2 al lado izquierdo se puede apreciar que el 30 de marzo de 2020 marcó el inicio de la pandemia en la Región Metropolitana, dando lugar a la propagación inicial en el sector oriente de dicha región, particularmente en la comuna de Las Condes, este fenómeno se atribuye a que esta zona alberga a la población más adinerada de la capital, lo que implica que sus residentes tenían mayores recursos y facilidades para viajar a otros países.

Esta capacidad de movilización internacional se convirtió en un factor de riesgo, ya que, según varios informes, muchos de estos destinos ya estaban afectados por la presencia del virus SARS-CoV-2. En consecuencia, al regresar a Chile, aquellos que habían viajado a estas regiones regresaban contagiados, desencadenando así la introducción y propagación del virus en la comunidad local.

En la imagen 2 al lado derecho podemos ver que ya transcurren unos meses desde que se inició la pandemia por lo cual ya no solo se concentra la pandemia en la comuna de Las Condes sino que ahora ya podemos observar que casi toda la Región Metropolitana tiene presente el Covid-19 pero ahora la mayor concentración está en la comuna de Puente Alto.

**Figura 3**

**Cantidad de casos confirmados de COVID-19 (06-08-2021 -19-12-2022)**



En la figura 3 la figura que está al lado izquierdo, se aprecia que la incidencia de contagios se mantiene constante en cuanto a la ubicación geográfica y a las características específicas. Sin embargo, es relevante destacar que, con la implementación del plan paso a paso, se observa una mejora significativa en la situación de la comuna de Santiago en comparación con la imagen anterior.

El plan paso a paso, diseñado para gestionar la reapertura gradual y segura de actividades en el contexto de la pandemia, parece haber tenido un impacto positivo en la comuna de Santiago. Esta mejora podría ser atribuible a la efectividad de las medidas implementadas en el marco del plan, como el control de la movilidad, restricciones en la capacidad de establecimientos, campañas de vacunación, y otras acciones preventivas.

En la figura 3 la figura que está al lado derecha, el 19 de diciembre de 2022, se aprecia claramente cómo la Región Metropolitana experimenta gradualmente un aumento en la propagación e infección por COVID-19. Este fenómeno se manifiesta especialmente en las tres comunas más grandes, tanto en términos de extensión territorial como de población.

La situación resaltada en la imagen sugiere un posible deterioro en la situación epidemiológica de la región y enfatiza la importancia de analizar las tendencias a lo largo del tiempo. El aumento de la propagación e infección podría deberse a varios factores, como la movilidad de la población, la variabilidad en la efectividad de las medidas de control, la aparición de nuevas variantes del virus, entre otros.

**Figura 4**

**Evolución de casos COVID-19 v/s índice de movilidad.**



En la figura 4, se puede apreciar la evolución durante un período específico, desde el 30 de marzo de 2020 hasta el 30 de noviembre de 2020, en relación con el índice de movilidad (representado en verde claro) y la tasa de contagio (en verde oscuro).

En las etapas iniciales, se observa que el índice de movilidad era normal y la tasa de contagio era baja. Esto sugiere que, al principio, la movilidad de la población era típica y que la propagación del virus era limitada, ya que la mayoría de las personas aún no estaban contagiadas.

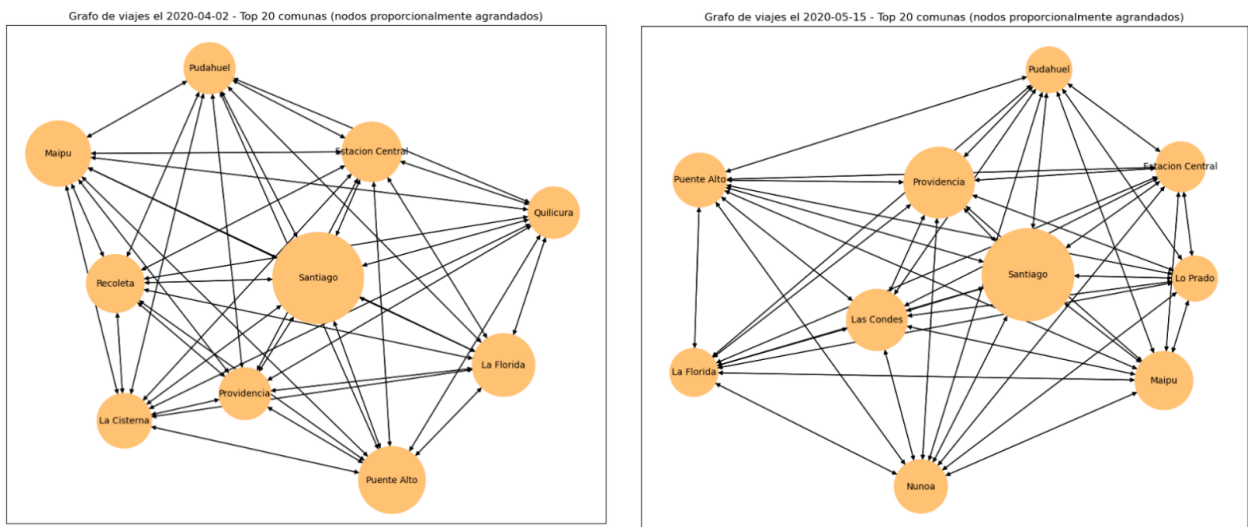
En una segunda fase, se evidencia un cambio significativo, marcado por un aumento abrupto en la tasa de contagio. Este salto podría indicar un momento crítico en el cual la propagación del virus se aceleró considerablemente, posiblemente debido a factores como la falta de medidas de contención efectivas, mayor interacción social o la introducción de nuevas cepas del virus.

Finalmente, en la tercera fase, correspondiente a finales de 2020, se observa que tanto la tasa de contagio como el índice de movilidad son elevados. Esto podría explicarse por la liberación progresiva de comunas del plan paso a paso, lo que lleva a un aumento en la movilidad de las personas y, consecuentemente, a un aumento en los casos de

COVID-19. La necesidad de las personas de regresar a sus lugares de trabajo u otras actividades podría haber contribuido a este aumento en la movilidad y, por ende, en la propagación del virus.

**Figura 5**

### Viajes origen-destino(transporte público)



En la figura 5, se observa el comportamiento de las personas en dos períodos de tiempo en relación con los viajes de origen y destino, específicamente en los 20 comunas con mayor viajes. En este análisis, es evidente que hay una comuna común en ambos periodos, que es Santiago. A partir de esta observación, podemos inferir que Santiago es el lugar donde la mayor cantidad de personas tiene sus lugares de trabajo. en la se destaca que otras dos comunas, Las Condes y Providencia, comparten características similares en términos de viajes de origen y destino. Este fenómeno puede atribuirse a un cambio significativo que ocurrió después del estallido social, donde un porcentaje considerable de empresas optó por trasladar sus oficinas a estas dos comunas mencionadas anteriormente.

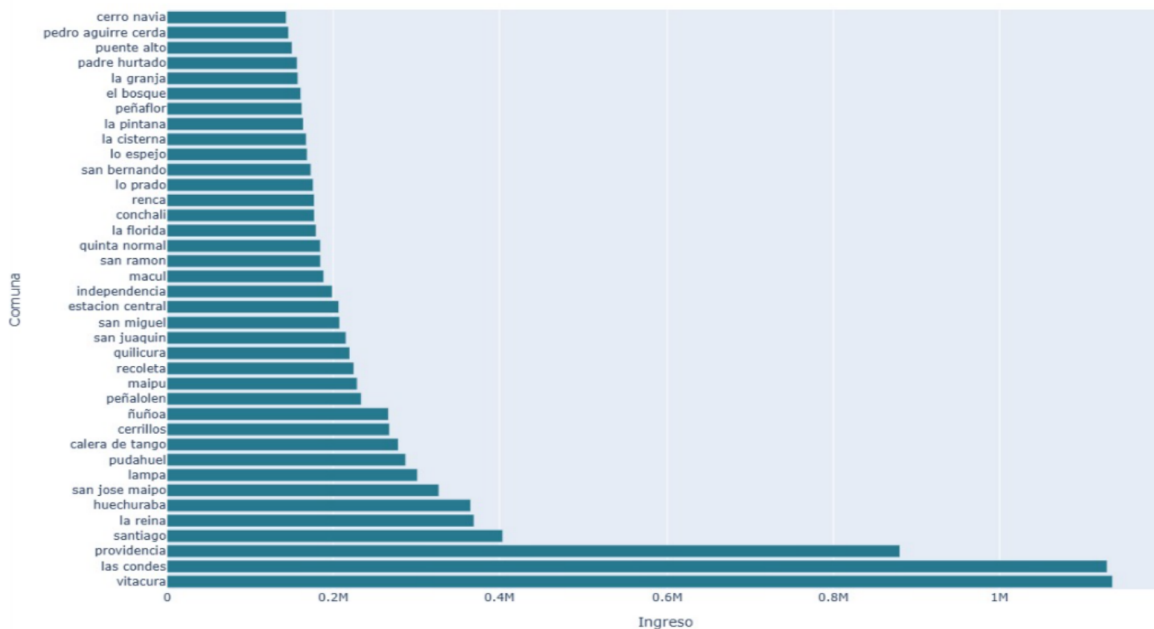
Este cambio en la ubicación de las oficinas podría deberse a diversas razones, como la búsqueda de un entorno más seguro, una mejor infraestructura, o la concentración de actividades comerciales y empresariales en esas áreas específicas. En resumen, el análisis de los patrones de viaje destaca la importancia de Santiago, Las Condes

Providencia como centros neurálgicos de actividad laboral, lo que puede influir significativamente en los flujos de movilidad y en la dinámica socioeconómica de la región.

Como conclusión podemos decir que en ambas imágenes los nodos más grandes representan que en esas comunas existen más viajes, esto se puede deber que los centros de trabajos se centran en estas 3 comunas como lo son : Santiago, Providencia, las condes.

**Figura 6**

**Ingreso Per capital.**



En la figura 6, se muestra la distribución del ingreso per cápita en las diversas comunas de la Región Metropolitana. Esta variable se revela como una de las consideraciones clave al analizar la propagación del COVID-19. Esto se debe a que las comunas con ingresos más altos corresponden, en su mayoría, a aquellas ubicadas en el sector oriente de la capital.

Esta relación entre el nivel de ingreso y la propagación del virus puede explicarse por diversos factores. En las comunas de ingresos más altos, las personas suelen tener mejores condiciones de vida, lo que incluye la capacidad de desplazarse cómodamente a supermercados, centros comerciales, estaciones de metro y servicios de atención médica.

Estas condiciones favorables podrían aumentar las interacciones sociales y la movilidad, lo que, a su vez, podría contribuir a una mayor propagación del virus.

Además, las personas con ingresos más altos pueden tener la flexibilidad de trabajar desde casa o de acceder a servicios de salud privados, lo que podría influir en su exposición y en la velocidad de propagación del virus en comparación con aquellas comunas con ingresos más bajos.

**Figura 7**

**Plan paso a paso.**



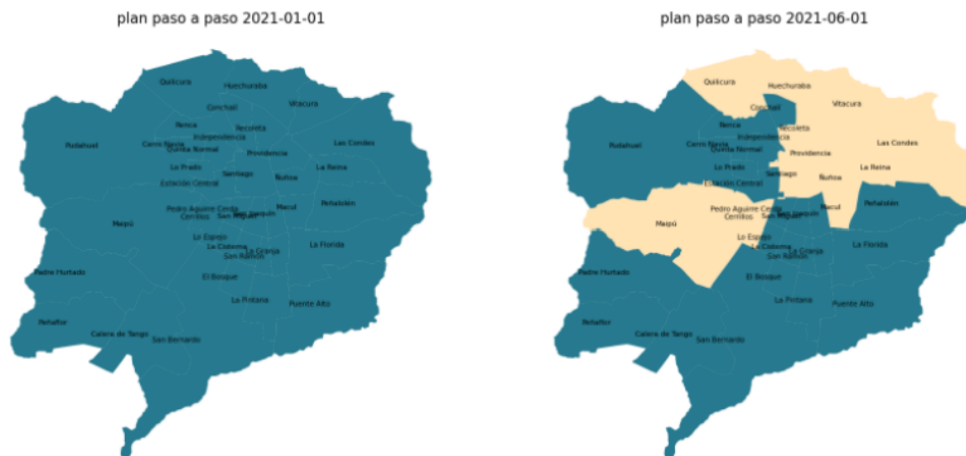
En la figura 7, se ilustra la evolución temporal de la aplicación del plan paso a paso. Inicialmente, se observa que durante una etapa inicial, toda la Región Metropolitana se encuentra bajo medidas de cuarentena. Sin embargo, se destaca que algunas comunas del sector oriente de la capital, a diferencia de otras áreas, experimentan cierta flexibilidad en las restricciones.

A medida que avanzamos hacia septiembre de 2020, se inicia un proceso de liberación progresiva de comunas en la Región Metropolitana. Esta fase se caracteriza por la implementación de permisos que permiten la salida de las personas durante determinadas horas del día para realizar actividades específicas. Este enfoque busca equilibrar la necesidad de reactivar la actividad económica con la importancia de controlar la propagación del virus.

Este cambio en las restricciones indica un intento de adaptarse a la evolución de la situación epidemiológica y económica. La estrategia de liberación gradual está vinculada a la implementación de medidas más focalizadas, donde se busca permitir ciertas actividades esenciales y limitar la movilidad, especialmente en áreas de mayor riesgo.

**Figura 8**

**Plan paso a paso.**



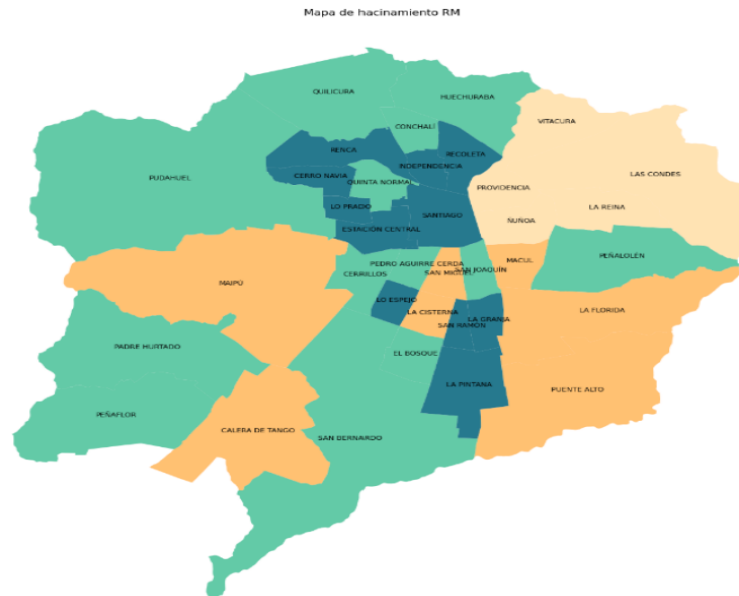
En la figura 8 al lado izquierdo se puede observar la llamada "segunda ola", caracterizada por un constante aumento en el número de casos. Ante esta situación, se evidencia la necesidad de imponer nuevamente medidas estrictas de cuarentena en la Región Metropolitana para contener la propagación del virus y evitar la saturación de los sistemas de salud.

Posteriormente, en junio de 2021, después de haber logrado controlar la segunda ola, se observa una transición hacia una fase de mayor flexibilización y apertura. Durante este periodo, se comienza gradualmente a liberar y permitir actividades en algunas comunas de la región. Este cambio puede deberse a la mejora en las condiciones epidemiológicas, el aumento en la disponibilidad de pruebas y la implementación exitosa de estrategias de mitigación.

Es importante destacar que la decisión de liberar y ser más permisivos en algunas comunas implica un equilibrio delicado entre la reactivación económica y la protección de la salud pública.

**Figura 9**

**Hacinamiento.**



En la figura 9 podemos observar ciertas categorías de hacinamiento que establecen diferentes niveles de densidad de personas en un dormitorio, ofreciendo una medida de la calidad de vida y las condiciones de vivienda en una determinada área. Estas categorías son definidas de la siguiente manera:

1. Sin hacinamiento: Menos de 2,5 personas por dormitorio. Esta categoría indica que la densidad de personas en los dormitorios es relativamente baja, sugiriendo una situación de vivienda menos congestionada.



La "densidad poblacional" se refiere a la cantidad de personas que viven por metro cuadrado. Cuando se observan los valores en la tabla, se evidencia que estas tres comunas tienen una mayor concentración de habitantes en comparación con las demás.

Esta conclusión sugiere que estas áreas experimentan una mayor aglomeración de personas por unidad de superficie. Las razones detrás de esta alta densidad pueden variar y podrían incluir aspectos como la presencia de infraestructuras urbanas importantes, centros comerciales, zonas industriales, o la ubicación céntrica y accesible que atrae a una mayor población.

## **5.2. Modelo**

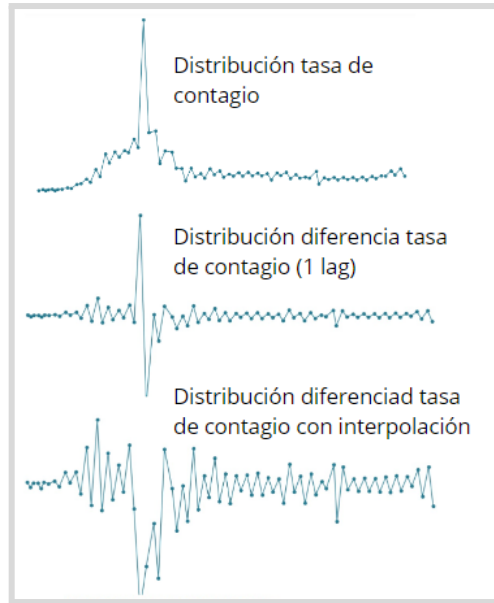
### **Resultados del Modelo SARIMAX**

Se implementó un modelo SARIMAX utilizando la tasa de contagio como variable endógena y el índice de movilidad como variable exógena.

Para que nuestro modelo SARIMAX funcione adecuadamente, es imperativo que nuestra serie principal sea estacionaria, es decir, que mantenga su media y varianza constantes a lo largo del tiempo. Al analizar la distribución inicial, observamos que no cumple con este requisito. Sin embargo, mediante la aplicación de un método de diferenciación, logramos homogeneizar la varianza y la media, a excepción del pico de contagios. Este inconveniente se solventó mediante una imputación por interpolación, como se observa en la figura 11.

**Figura 11**

**Distribución tasa de contagios**



Al realizar la prueba de estacionariedad KPSS a la última distribución, observamos que el estadístico de prueba supera el valor crítico, lo que lleva al rechazo de la hipótesis nula. Este resultado sugiere que la serie ahora exhibe estacionariedad, lo cual es fundamental para la adecuada aplicación del modelo SARIMAX.

```
Estadística KPSS: 0.10459353647950516
P-valor: 0.1
Lags utilizados: 4
Valores críticos: {'10%': 0.347, '5%': 0.463, '2.5%': 0.574, '1%': 0.739}
La serie es estacionaria (no podemos rechazar la hipótesis nula de estacionariedad).
```

## Autocorrelación

Se llevó a cabo un análisis de autocorrelación con el objetivo de evaluar la relación entre la variable y sus valores en períodos anteriores. Los resultados obtenidos se detallan a continuación:

En el gráfico de la función de autocorrelación (ACF), de la figura 12, se aprecia una disminución en los rezagos, lo que sugiere la posible presencia del factor autoregresivo (AR). Por otro lado, en el gráfico de la función de autocorrelación parcial (PACF), se observa que los rezagos se sitúan por encima del nivel de confianza, indicando la existencia de un componente de media móvil y una correlación significativa en dichos rezagos.

Figura 12

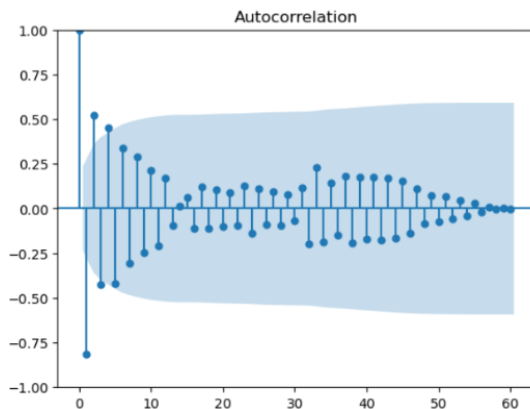
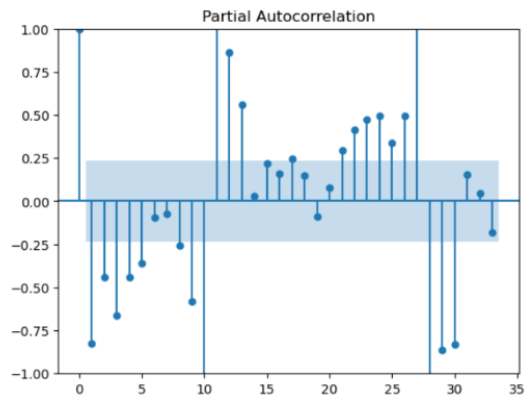


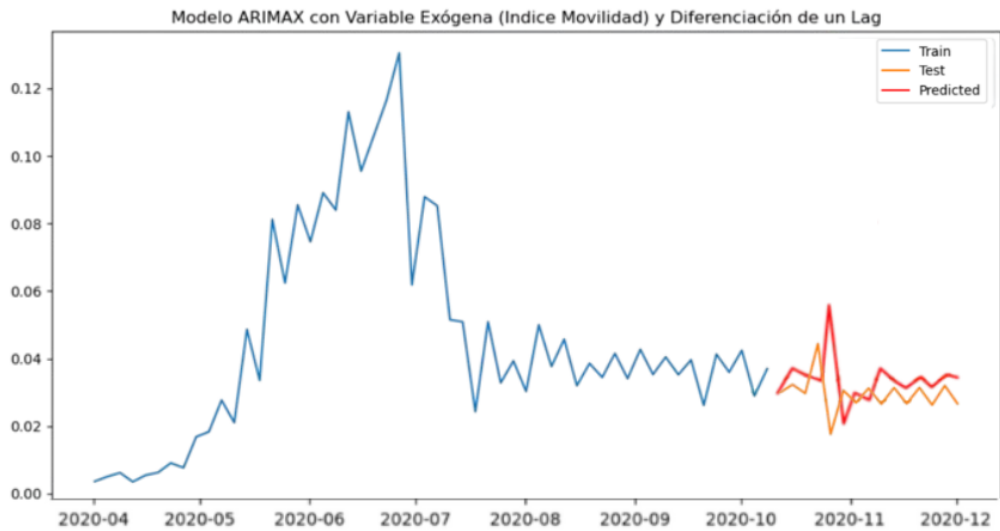
Figura 13



Al considerar estos criterios, se activó el modelo respectivo. Al revisar las predicciones de la figura 13, se aprecia que, aunque se adaptan a las fluctuaciones de la pendiente, se evidencia un desfase y una falta de precisión. Estos hallazgos sugieren la importancia de incorporar variables adicionales que puedan ofrecer una explicación más integral del modelo.

Figura 13

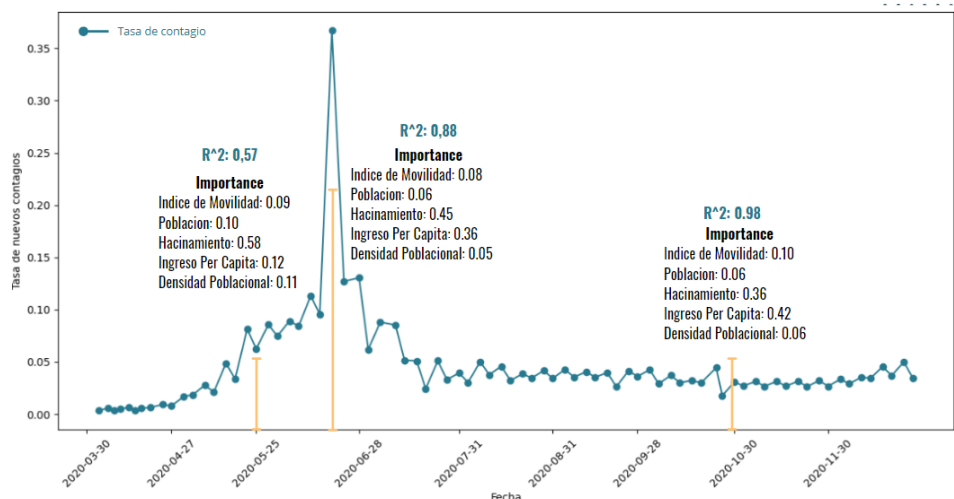
### Resultados Modelo SARIMAX



### Resultados del Modelo Random Forest Regression en 3 Periodos

El modelo Random Forest Regression se aplicó en tres periodos definidos, lo que proporcionó una visión detallada de la compleja relación entre las variables y la propagación del virus.

Figura 14



En el primer periodo, la movilidad emergió como un factor clave en la propagación inicial del COVID-19, respaldando la hipótesis inicial. Sin embargo, este impacto se vio influenciado por otras variables socioeconómicas, lo que sugiere una red de factores interconectados.

Durante el segundo periodo, el hacinamiento surgió como un factor crítico, especialmente en momentos de alta incidencia. La variabilidad en la importancia relativa de las variables destacó la evolución dinámica de la pandemia y la necesidad de adaptar las estrategias de intervención a las condiciones específicas de cada fase.

En el tercer periodo, se observaron cambios significativos en la importancia relativa de las variables, reflejando la continua evolución de la dinámica del virus. La mejora en la precisión de las predicciones al dividir los datos en periodos subraya la influencia temporal en la propagación del COVID-19.

## **6. Conclusión y Trabajo Futuro**

Los resultados respaldan parcialmente la hipótesis planteada, demostrando que la movilidad inicialmente fue un factor influyente, pero su contribución se vio matizada al considerar otras variables socioeconómicas. El hacinamiento emergió como un factor crítico en la propagación del virus, especialmente en momentos de alta incidencia. La importancia relativa de las variables cambió con el tiempo, reflejando la dinámica evolutiva de la pandemia.

El modelo Random Forest Regression resaltó la complejidad de la relación entre las variables y la propagación del COVID-19, subrayando la necesidad de considerar múltiples factores para obtener predicciones más precisas. La mejora en la predicción al dividir los datos en periodos sugiere la influencia temporal en la dinámica del virus.

En resumen, la gestión efectiva de la pandemia requiere estrategias más allá de la reducción de movilidad, incorporando medidas específicas según las condiciones socioeconómicas y demográficas de cada área. Estos hallazgos tienen implicaciones significativas para el diseño de políticas de salud pública y la preparación para futuras crisis sanitarias, destacando la importancia de enfoques integrales y adaptativos.

Como trabajo futuro queda pendiente la incorporación de las variables propias del virus, como cantidad de contagios por persona o rangos etarios a los que afecta más. Con estos datos se podría mejorar aún más la estimación de la propagación de un virus.

## Bibliografía

- Factores socioeconómicos y perceptuales que influyeron en el teletrabajo durante la pandemia COVID-19 en Santiago, Chile.  
[Factores socioeconómicos y perceptuales que influyeron en el teletrabajo durante la pandemia COVID-19 en Santiago, Chile | Estudios de Transporte.](#)
- La economía chilena frente a la pandemia del COVID-19: fortalezas, desafíos y riesgos  
[83f103c6-53c9-4c96-9190-7b0314a4574d \(bcentral.cl\).](#)
- Hacinamiento: la variable clave en la propagación del Covid-19 en el Gran Santiago  
[Hacinamiento: la variable clave en la propagación del Covid-19 en el Gran Santiago - CIPER Chile.](#)
- Conjunto de dataset del Ministerio de Ciencia, Tecnología e Innovación:  
[https://github.com/MinCiencia/Datos-COVID19/tree/master/input.](https://github.com/MinCiencia/Datos-COVID19/tree/master/input)
- El nivel socioeconómico determina la incidencia de COVID-19 y la mortalidad relacionada en Santiago de Chile  
[El nivel socioeconómico determina la incidencia de COVID-19 y la mortalidad relacionada en Santiago de Chile | Ciencia \(science.org\)](#)