



ANÁLISIS COMPARATIVO SOBRE INDICADORES DE MOVILIDAD PARA LA
ZONA CENTRAL DE CHILE DURANTE LA PANDEMIA
Exploración de diferencias entre grupos socioeconómicos y propuesta de modelo predictivo
(SVR) para el índice de movilidad interno

POR: DAVID FERNANDO ORELLANA MONTECINO

&

IGNACIO ANDRES FUENTES NAVARRO

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. LEO FERRES

NOVIEMBRE 2023

SANTIAGO

David Orellana:

Dedico este trabajo a mi familia, quienes me han apoyado de manera incondicional en las decisiones que he tomado. Al amor de mi vida, Carolina, quien con su amor y apoyo me han hecho llegar más lejos de lo que hubiese podido imaginar.

Ignacio Fuentes:

Dedico este trabajo a mi esposa, Paulina, cuyo apoyo constante y paciencia han sido fundamentales a lo largo de este viaje académico. También quiero dedicar estas páginas a mi hija, Martina, quien se ha unido a nuestra vida hace muy poco tiempo, llenándola de aún más motivación y alegría

AGRADECIMIENTOS

David Orellana:

Agradezco a los docentes del programa por permitirme seguir avanzando en mi desarrollo personal. A nuestro tutor Sr Leo Ferres, por su amabilidad para guiarnos en este trabajo y a mi compañero Ignacio por su buena energía y entrega.

Ignacio Fuentes:

Expreso mi profundo agradecimiento al Sr. Leo Ferres, nuestro tutor, por su constante apoyo fundamental en la realización de este trabajo. También quiero reconocer a la Universidad del Desarrollo por ofrecer un programa académico de alta calidad y a mi compañero David, cuyo esfuerzo y dedicación han sido ejemplares.

TABLA DE CONTENIDO

RESUMEN.....	1
1. INTRODUCCIÓN	2
2. TRABAJO RELACIONADO	3
3. HIPÓTESIS Y OBJETIVOS.....	4
4. DATOS Y METODOLOGÍA.....	5
4.1. DATOS	5
4.2. METODOLOGÍA	7
5. RESULTADOS.....	10
5.1. RESULTADOS PRELIMINARES	10
5.2. EXPLORACIÓN DE DATOS ENTRE CUARTILES 1 Y 4	12
5.3. RESULTADOS Y ANÁLISIS ENTRE CUARTILES 1 Y 4.....	13
5.4. RESULTADOS ENTRE CUARTILES 1 Y 4 EN FIN DE SEMANA	18
5.5. RESULTADOS Y ANÁLISIS INTRA CUARTILES 1 Y 4	22
5.6. MOVILIDAD DURANTE DIFERENTES FASES	26
5.7. RESULTADOS DEL MODELO DE REGRESIÓN DE VECTORES DE SOPORTE (SVR)	27
6. CONCLUSIONES Y TRABAJO FUTURO	30
BIBLIOGRAFÍA	33

Resumen

La presente investigación intenta aportar al entendimiento de la inercia de movilidad de la población, especialmente ante situaciones de alta complejidad como lo fue la pandemia de COVID – 19 y las medidas no farmacológicas que se aplicaron en Chile para restringir el libre desplazamiento. Para este análisis, se utilizaron datos detallados de los Índices de Movilidad (IM) contruidos por Instituto Data Science de la UDD, y facilitados al Ministerio de Ciencia, los cuales se encuentran desglosados en índices internos (referentes a viajes dentro de la comuna) e índices externos (considerando desplazamientos entre comunas), así como la sumatoria de ambos como un indicador de movilidad total (IM).

Uno de los resultados más significativos de esta investigación es la correlación negativa entre el índice de Movilidad interno y los niveles socioeconómicos de las comunas. Este hallazgo sugiere que factores como el ingreso y la situación económica local desempeñan un papel crucial en la configuración de los patrones de movilidad. Esta comprensión pudiese ser fundamental a la hora de diseñar estrategias de manejo de la movilidad que sean efectivas, equitativas y sensibles a las disparidades socioeconómicas.

También, se desarrolla un modelo predictivo utilizando Regresión de Vectores de Soporte (SVR), que pretende ser una herramienta eficaz para predecir el índice de movilidad interno (IM_interno). El modelo se entrena con datos de comunas en la zona central de Chile, utilizando variables clave como la densidad poblacional (población/superficie en Km²) y el porcentaje de pobreza. El modelo destaca la importancia de considerar estas variables en la formulación de políticas de gestión de la movilidad.

En conclusión, este estudio intenta ofrecer una visión más completa y contextualizada de la movilidad durante los tiempos de cuarentena, integrando datos cuantitativos y comprensión cualitativa de los factores subyacentes, especialmente el porcentaje de población en situación de pobreza.

1. Introducción

La pandemia del COVID-19, que ocurre a principios de 2020, desencadenó una serie de cambios significativos en todo el mundo, afectando no solo la salud pública sino también la dinámica social y económica. Entre los aspectos más impactantes se encuentra la movilidad de la población, un indicador clave que refleja la adaptabilidad de las comunidades frente a las restricciones impuestas para contener la propagación del virus.

Chile, como muchos otros países, se vio inmerso en la complejidad de gestionar la movilidad de su población durante este periodo. La Universidad del Desarrollo, en colaboración con diversas entidades, ha recopilado y procesado datos detallados de la movilidad de los residentes chilenos, utilizando información anónima de las posiciones de los teléfonos celulares. Estos datos no solo ofrecen una idea clara de la movilidad en el país, sino que también sirven como base para explorar patrones, desafíos y posibles soluciones en el contexto de una crisis sanitaria global.

El trabajo desarrollado a continuación propone un análisis de los indicadores de movilidad en Chile durante el año 2020, con un enfoque particular en comunas de la zona central del país. Se busca vislumbrar patrones de movilidad asociados a variables socioeconómicas y demográficas. De igual manera, comprender cómo las comunidades se han adaptado a las restricciones, ya sea mediante cambios en los patrones de viaje o la adopción de medidas preventivas como el plan “Paso a Paso”, proporciona información valiosa para diseñar estrategias de gestión de crisis y planificación futura.

Se busca proponer un modelo predictivo que estime el índice de Movilidad Interno (IM_interno) basándose en variables de mayor correlación y testarlo con datos al azar. Este modelo no solo serviría como herramienta de análisis, sino también como un indicador potencial para futuras crisis.

2. Trabajo Relacionado

El estudio presentado en este informe se beneficia y se nutre de investigaciones relevantes en campos afines, cada una proporcionando una perspectiva única. Tres de los textos científicos citados en este trabajo fueron desarrollados por nuestro tutor, Sr. Leo Ferres. Sin duda, cada uno de los textos citados fueron relevantes para llevar a cabo lo presentado en los apartados siguientes.

En el trabajo de "A dataset to assess mobility changes in Chile following local quarantines", se examina el impacto de las cuarentenas localizadas en Chile durante la pandemia de COVID-19. La implementación de medidas de cuarentena a nivel local, en lugar de a nivel nacional, plantea la importancia de entender cómo estas estrategias afectan la movilidad. Este estudio, basado en datos de Telefónica Chile, establece un precedente fundamental para evaluar cambios en los patrones de movilidad y entender su eficacia.

En el artículo que se centra en la "Evaluation of home detection algorithms on mobile phone data using individual-level ground truth", se aborda el desafío de inferir la ubicación del hogar de los usuarios de teléfonos móviles. Lo que adoptamos como punto central de este artículo, es la necesidad de contar con datos cada vez más precisos y ajustados, que permitan establecer patrones de movilidad y por consiguiente políticas eficientes en términos de transporte.

El estudio sobre "Estimating the effect of social inequalities on the mitigation of COVID-19 across communities in Santiago de Chile" proporciona una visión detallada de la propagación del virus en Santiago de Chile. La correlación entre las intervenciones no farmacéuticas (NPIs) y la movilidad, especialmente en áreas más desarrolladas, plantea preguntas valiosas sobre cómo los factores socioeconómicos influyen en la dinámica de la pandemia. Estos hallazgos sirven como una guía esencial para entender las disparidades en la Zona Central de Chile y siembran muchas de las ideas que desarrollamos a lo largo del informe.

El artículo sobre "Mobility in pandemic times: Exploring changes and long-term effects of COVID-19 on urban mobility behavior" ofrece una perspectiva longitudinal sobre la transformación del comportamiento de movilidad en Berlín. Aunque la ubicación geográfica difiere, las tendencias generales en la adaptación a lo largo del tiempo son fundamentales para contextualizar nuestra investigación.

En síntesis, estos trabajos relacionados no solo fortalecen la base teórica de nuestro estudio, sino que también ofrecen enfoques y metodologías que podemos adaptar a nuestro contexto específico e intereses. Establecen un marco integral que enriquece nuestra comprensión de los indicadores de movilidad durante la pandemia de COVID-19.

Cada uno de estos artículos ha sido citado correctamente en el apartado de bibliografías.

3. Hipótesis y Objetivos

El objetivo principal de este Capstone Project es realizar un análisis comparativo de los indicadores de movilidad desarrollados durante la época de pandemia en Chile, específicamente para el año 2020. Adicionalmente, se propone un modelo que permita estimar el índice de Movilidad Interno en base a variables relevantes, como lo son la demografía, densidad y factores económicos de la población.

HIPÓTESIS

- No hay diferencias significativas en los indicadores de movilidad entre distintos grupos socioeconómicos.
- No hay diferencias significativas en los indicadores de movilidad dentro del mismo grupo socioeconómicos.
- No hay diferencias significativas en los indicadores de movilidad entre distintos grupos socioeconómicos para los días de fin de semana.
- No existe una relación significativa entre el índice de movilidad interno promedio de las comunas de la zona central de Chile durante la pandemia y variables socioeconómicas como el porcentaje de pobreza observado por comuna y la densidad poblacional por Km², por lo tanto, no es posible predecir dicho índice.

4. Datos y Metodología

4.1. Datos

La principal fuente de información fue obtenida desde el “DP33”, que está compuesto por los Índices de Movilidad (IM) a nivel comunal para Chile, el cual incluye viajes dentro de la comuna (IM_interno) y viajes desde una comuna a otra (IM_externo), así como también, un tercer indicador (IM) que es la suma de estos últimos dos índices¹. Para poder comparar los indicadores, estos están normalizado por el número de habitantes de la comuna.

En detalle, la información está compuesta por las columnas de: Región, Código Región, Comuna, Código comuna, Población, Superficie_km2 y Fecha, así como sus respectivos valores para cada Índice de Movilidad (Figura N°1). Este dataset tiene en total de 95.976 registros de filas y 10 columnas, comprendidas entre el 26 de febrero de 2020 y el 30 de noviembre del 2020. Esto abarca un periodo anterior a las medidas sanitarias no farmacológicas tomadas por la autoridad de Chile, ya que las primeras restricciones correspondieron al día 16 de marzo, con el cierre de algunos colegios, mientras que los primeros periodos de cuarentena no se establecieron hasta el 18 de marzo en adelante.

	Region	Codigo region	Comuna	Codigo comuna	Superficie_km2	Poblacion	IM_interno	IM_externo	IM	Fecha
0	Coquimbo	4.0	La Serena	4101	1901.22	249656.0	7.842179	2.378377	10.220555	2020-02-26
1	Coquimbo	4.0	La Serena	4101	1901.22	249656.0	7.886014	2.345445	10.231459	2020-02-27
2	Coquimbo	4.0	La Serena	4101	1901.22	249656.0	8.238552	2.406954	10.645506	2020-02-28
3	Coquimbo	4.0	La Serena	4101	1901.22	249656.0	7.236426	2.242430	9.478856	2020-02-29
4	Coquimbo	4.0	La Serena	4101	1901.22	249656.0	5.994744	1.981816	7.976560	2020-03-01

Figura N°1. Muestra del dataset denominado “DP33”.

Adicionalmente, y como parte necesaria para la estructuración de la información, se utiliza el “DP74”², que contiene las etapas del Plan Paso a Paso, disgregado a nivel comunal desde su implementación. La información corresponde a las publicaciones en el diario Oficial proporcionadas por el Ministerio de Salud. Los registros comienzan desde el día 28 de julio hasta el 11 de julio del 2022, completando un total de 390 filas y 838 columnas.

¹ Data obtenida desde el DP33, por el Ministerio de Ciencias de Chile.

² Data obtenida desde el DP74, por el Ministerio de Ciencias de Chile.

codigo_region	region_residencia	codigo_comuna	comuna_residencia	zona	2020-07-28	2020-07-29	2020-07-30	2020-07-31	2020-08-01	...	2022-10-29	2022-10-30	2022-10-31	2022-11-01
0	15	Arica y Parinacota	15101	Arica Rural	3	3	3	3	3	...	Escenario de Apertura	Escenario de Apertura	Escenario de Apertura	Escenario de Apertura
1	15	Arica y Parinacota	15101	Arica Urbana	1	1	1	1	1	...	Escenario de Apertura	Escenario de Apertura	Escenario de Apertura	Escenario de Apertura
2	15	Arica y Parinacota	15102	Camarones Total	3	3	3	3	3	...	Escenario de Apertura	Escenario de Apertura	Escenario de Apertura	Escenario de Apertura
3	15	Arica y Parinacota	15202	General Lagos Total	3	3	3	3	3	...	Escenario de Apertura	Escenario de Apertura	Escenario de Apertura	Escenario de Apertura
4	15	Arica y Parinacota	15201	Putre Total	3	3	3	3	3	...	Escenario de Apertura	Escenario de Apertura	Escenario de Apertura	Escenario de Apertura

Figura N°2: DP74, Fases del plan Paso a Paso, desgregado a nivel de día y comuna.

Para complementar la información proporcionada por el Producto 74, se integra también de DP29, que contiene el registro de cuarentenas totales, para abarcar los periodos desde el 18 de marzo al 27 de julio del año 2020. Este dataframe, está compuesto por 187 líneas, con 12 columnas de atributos³.

ID	Nombre	Estado	Alcance	Fecha de Inicio	Fecha de Término	Código CUT Comuna	Detalle	Superficie en m ²	Perímetro en m	Region	n_REGION
0 1	Arica	Histórica	Área Urbana Completa	2020-04-16 22:00:00	2020-05-15 22:00:00	15101	Corresponde al Radio Urbano de la Ciudad de Ar...	28050418.80	48722.30001	15	R15
1 2	Santiago Norte	Histórica	Sector Específico	2020-04-13 05:00:00	2020-05-05 22:00:00	13101	Corresponde a la extensión de la cuarentena en...	19939218.57	21599.35134	13	R13
2 3	Puente Alto Poniente	Histórica	Sector Específico	2020-04-09 22:00:00	2020-05-08 22:00:00	13201	Corresponde al sector delimitado por la Av. Co...	42364268.21	30428.05152	13	R13
3 4	El Bosque	Histórica	Comuna completa	2020-04-16 22:00:00	2020-09-21 05:00:00	13105	Aplicada a la totalidad de la comuna. Se incor...	20622713.23	18788.87075	13	R13
4 5	San Bernardo Nororiental	Histórica	Sector Específico	2020-04-16 22:00:00	2020-05-08 22:00:00	13401	Corresponde al sector delimitado por los limit...	16326216.83	24929.14572	13	R13

Figura N°3: DP29, registro de cuarentenas totales.

Finalmente, también se utiliza un set de datos proporcionado por el Ministerio de Desarrollo Social llamado “Estimaciones de Pobreza Comunal⁴”. Esta información, de carácter oficial, está construida a través de una combinación de encuestas y registros administrativos y aporta la variable socioeconómica desglosada a nivel comunal para el país durante el año 2020. Se utilizan los valores relacionados a la cantidad de personas en situación de pobreza por comuna y su porcentaje para cada comuna. Este dataframe contiene en total 349 filas de registros y 10 columnas.

³ Data obtenida desde el DP29, por el Ministerio de Ciencias de Chile

⁴ Observatorio del Ministerio de Desarrollo Social, en su apartado “Pobreza Comunal 2020.

	Código	Región	Nombre comuna	Número de personas según proyecciones de población (*)	Número de personas en situación de pobreza por ingresos (**)	Porcentaje de personas en situación de pobreza por ingresos 2020	Límite inferior	Límite superior	Presencia de la comuna en la muestra Casen	Tipo de estimación SAE (***)
0	1101	I de Tarapacá	Iquique	224971.0	27709.108893	0.123167	0.102136	0.144199	Si	Directa y Sintética (Fay-Herriot)
1	1107	I de Tarapacá	Alto Hospicio	131699.0	21207.973305	0.161034	0.138915	0.183153	Si	Directa y Sintética (Fay-Herriot)
2	1401	I de Tarapacá	Pozo Almonte	17580.0	2752.085009	0.156546	0.119535	0.193558	Si	Directa y Sintética (Fay-Herriot)
3	1402	I de Tarapacá	Camíña	1382.0	315.572403	0.228345	0.167431	0.289259	Si	Sintética
4	1403	I de Tarapacá	Colchane	1592.0	337.637330	0.212084	0.138230	0.285938	No	Sintética

Figura N°4. Datos en “Estimaciones de Pobreza Comunal para el año 2020”.

4.2. Metodología

La metodología para este proyecto consiste en realizar un trabajo de depuración sobre los distintos dataframe descritos como fuentes de información, esto incluye el proceso de:

- Carga de la información, exploración de su composición y resumen estadísticos, tipo de datos almacenados y cantidad de valores nulos para cada uno.
- Limpieza y preprocesamiento de los datos: El proceso incluye la eliminación de valores nulos según corresponda, y ajuste de los tipos de dato para su tratamiento posterior.
- Generación de nuevas columnas de datos que se consideran relevantes para este estudio, como lo es, la identificación de días de semana y fin de semana, la densidad para cada comuna, asignación de comunas con alto o bajo porcentaje de pobreza, entre otros.
- Unificación de las diversas fuentes de información obtenida de los diversos dataframe, donde se extrajo las variables de interés adicionales a los índices de movilidad, como lo es: cantidad de personas en situación de pobreza por comuna, porcentaje de pobreza para igual segmentación, fases del Plan Paso a Paso y periodo de cuarentenas totales, previo a la marcha de dicho plan.
- Visualización de datos preliminar y su exploración de las relaciones tanto para datos cuantitativos como cualitativos.

Posteriormente, y una vez que se obtiene el dataframe unificado de trabajo, este fue segmentado para poder acotar el área de estudio. El enfoque está centrado sobre la zona

Centra de Chile continental, por lo que se opta por filtrar la información en base a las regiones del país que componen este segmento, en este caso: la región Metropolitana de Santiago, Biobío, Valparaíso, Libertador General Bernardo O'Higgins, Ñuble y Maule, excluyendo las comunas que no corresponden a esta división, como lo es, por ejemplo, las comunas fuera del Chile continental. De igual manera, se ha definido demarcar el estudio solo sobre las restricciones de movilidad que corresponden a la Fase 0 (extraído desde las cuarentenas totales), Fase 1 y Fase 2 de la implementación del Paso a Paso.

Adicionalmente, y utilizando la información obtenida desde el Ministerio de Desarrollo Social en base a sus resultados expuestos en el dataframe de “Estimaciones de Pobreza Comunal para el año 2022”, se realiza la segmentación en base a “Cuartiles” para el indicador de “Porcentaje de personas en situación de pobreza”. La implementación de esta variable proporciona la capacidad de realizar análisis sobre 2 conjuntos de comunas que poseen características socioeconómicas disimiles, facilitando la extracción de valor de los datos en búsqueda de comparativas o diferencias más robustas y acotadas.

Para la formulación de los resultados en cuanto a la inercia de movilidad, el dataframe resultante fue sometido a diversos ejercicios los cuales abarcaron:

- Realización de pruebas de correlación sobre las variables cuantitativas que lo componen.
- Regresiones (lineales u otras) y pruebas estadísticas para establecer diferencia significativa sobre los datos, tanto para los índices de movilidad total (IM), interno (IM_interno) y también externo (IM_externo), esto, aplicado entre los Cuartil 1 (menor porcentaje de pobreza) y el Cuartil 4 (mayor porcentaje de pobreza), así como también intra-Cuartil (IM_interno vs IM_externo), y las diferencias para la segmentación de “Días de semana” vs “Días de fin de semana”.
- Visualización del comportamiento temporal de los índices de movilidad a lo largo del periodo de estudio, y su comportamiento para cada segmentación de “fase” y su posible efecto sobre la movilidad como medida de restricción no farmacológica.

Finalmente, se desarrolla un modelo que pretende comprender la relación entre el índice de movilidad interna (IM_interno) de habitantes en comunas de la zona central de Chile y factores socioeconómicos clave, como el porcentaje de pobreza y la densidad poblacional.

Para esto el modelo busca analizar el Índices de pobreza y densidad poblacional y su impacto sobre la movilidad interna utilizando regresión de vectores de soporte (SVR). Esta técnica de aprendizaje automático es utilizada para abordar problemas de regresión, donde el objetivo es prever un valor numérico en lugar de clasificar en categorías, además de proveer fundamentos más robustos para manejar no linealidades y adaptarse a patrones más complejos sin requerir transformaciones explícitas de las variables. Adicionalmente, resistente a valores atípicos debido al uso de vectores de soporte y la posibilidad de usar distintas funciones kernel, siendo el kernel radial (RBF) el más común y el utilizado en esta investigación. Este modelo es escalable, mediante validación cruzada, y es posible evaluar cómo se desempeña el modelo con nuevos datos. Siendo esta última característica, crucial a la hora de determinar su utilidad práctica.

5. Resultados

5.1. Resultados preliminares

El desarrollo de la limpieza, la transformación, la segmentación y todo el preprocesamiento detallado en la sección anterior, da como resultado que la información se acotó desde un total de más de 90 mil registros, a poco más de 11 mil, lo cual se ve reflejado en el resumen de la estructuración final del dataframe, disponible en la siguiente Figura N°5:

```

RangeIndex: 11520 entries, 0 to 11519
Data columns (total 18 columns):
#   Column              Non-Null Count  Dtype  
---  -
0   Region              11520 non-null  object  
1   Codigo region       11520 non-null  int32   
2   Comuna              11520 non-null  object  
3   Codigo comuna       11520 non-null  int32   
4   Superficie_km2      11520 non-null  float64  
5   Poblacion           11520 non-null  float64  
6   IM_interno          11520 non-null  float64  
7   IM_externo          11520 non-null  float64  
8   IM                  11520 non-null  float64  
9   Fecha               11520 non-null  datetime64[ns]
10  Año_Mes             11520 non-null  period[M]
11  Día                 11520 non-null  int32   
12  Weekend             11520 non-null  bool    
13  Personas en pobreza  11520 non-null  float64  
14  Porcentaje de pobreza 11520 non-null  float64  
15  Fase                11520 non-null  int32   
16  Densidad            11520 non-null  float64  
17  Cuartil             11520 non-null  category
dtypes: bool(1), category(1), datetime64[ns](1), float64(8), int32(4), object(2), period[M](1)

```

Figura N°5. Resultado del dataframe tras el pre-procesamiento.

Para profundizar en la exploración de la distribución y concentración de las variables cuantitativas, así como también sobre la presencia de outliers, se construye de manera visual diferentes gráficos de caja, donde se busca detectar posibles puntos de interés o valores inusuales en cada variable:

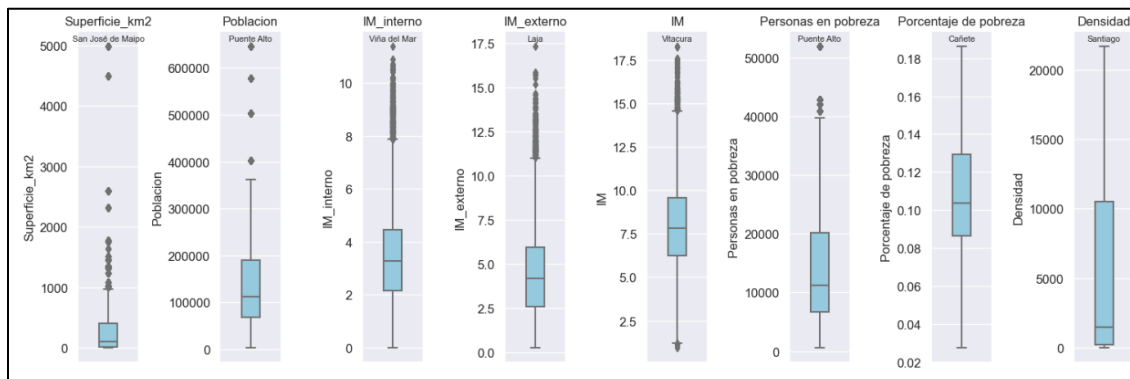


Figura N°6. Visualización de Boxplot para las variables cuantitativas.

De manera general, se aprecia una mayor dispersión en variables como: la superficie de cada comuna, la cantidad de personas en situación de pobreza o la población, que responde a un comportamiento más bien esperable debido a la alta heterogeneidad del país, no obstante, también se evidencia una cantidad de valores anómalos en los índices de movilidad.

Para entender de mejor manera la interacción entre las variables cuantitativas, se realiza una matriz de correlación sobre el dataset y los resultados se aprecian en la Figura N°7.

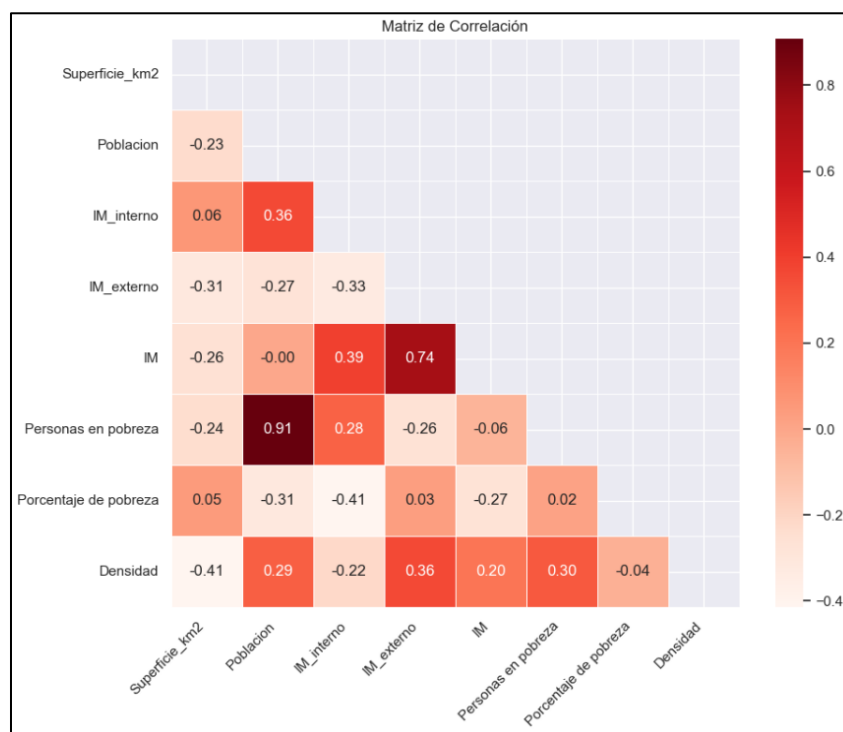


Figura N°7. Visualización de la matriz de correlación sobre las variables cuantitativas.

De manera preliminar, se observa una correlación fuerte y positiva entre las variables de “Personas en Pobreza” y “Población”. Esto debido a la alta densidad que presentan algunas comunas de la zona de estudio, y que alberga problemas conocidos como el deficit habitacional o el hacinamiento, entre otros problemas. Por otra parte, también se evidencia una correlación negativa, aunque de menor magnitud, entre los valores de “Porcentaje de pobreza” y el “Índice de Movilidad Interna”, en contraste al “Índice de movilidad externo”. Esto resulta interesante sobre los objetivos del estudio, por lo que se hará una profundización de como la segmentación realizada para “Porcentaje de pobreza” pudiese tener injerencia sobre la inercia de movilidad.

5.2. Exploración de datos entre Cuartiles 1 y 4

Durante la preparación del set de datos, se utiliza la información proveniente desde el Ministerio de Desarrollo social para realizar una segmentación en base al “Porcentaje de Pobreza” que presenta cada comuna. Se generaron 4 cuartiles, los cuales están descritos, como Cuartil 1, el de menor porcentaje de pobreza, y en su contra posición, el Cuartil 4, el cual presenta para el año 2020 el mayor porcentaje. Esto proporciona una distribución equitativa de comunas para el análisis, dejando una segmentación final con igual número para ambos grupos, como muestra Figura N°8:

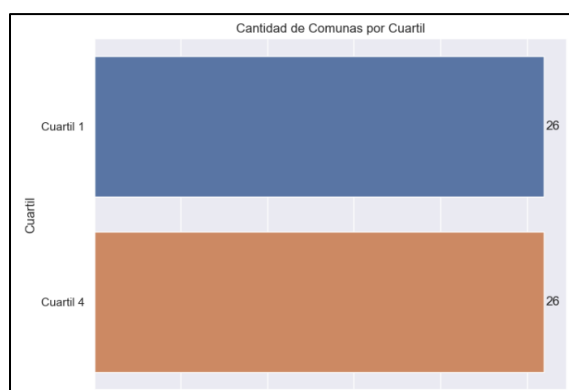


Figura N°8. Cantidad de comunas para los Cuartiles 1 y 4.

Por otra parte, al indagar en mayor profundidad en la cantidad de datos que fueron segmentados bajo estos cuartiles, y a su vez, también por día de semana, es posible ver una frecuencia de registros asimétrica, como lo muestra la siguiente Figura N°9:



Figura N°9. Frecuencia de registros segmentados por Cuartil y Día de semana/ Fin de semana.

Si la mirada está centrada en el aporte de cada Fase de restricción de movilidad al dataset, se observa que el 38.6% de los datos corresponden a la denominada “Fase 0”, mientras que el 34% a Fase 2 y el 27.1% restante de los datos a Fase 1.

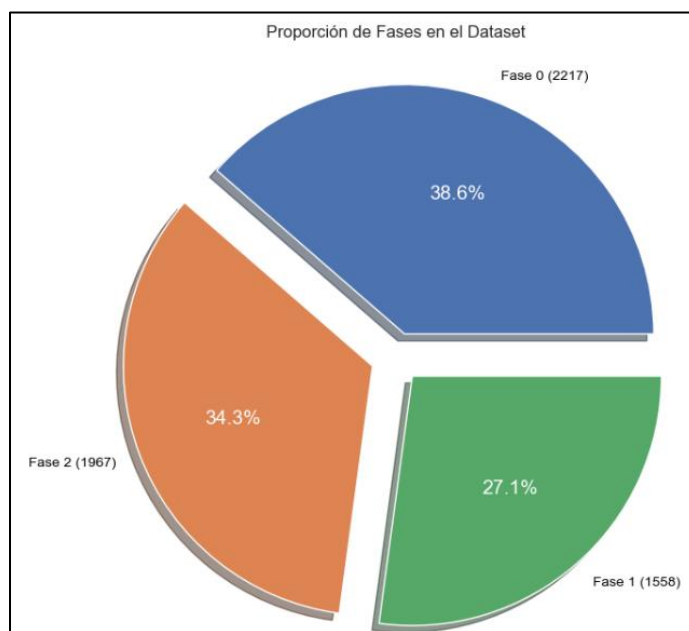


Figura N°10. Frecuencia de registros segmentados por Fase de restricción de movilidad.

5.3. Resultados y análisis entre Cuartiles 1 y 4

La matriz de correlación para nuestro set de datos evidencia una correlación negativa importante entre el porcentaje de pobreza y el “índice de movilidad interno”. Para poder indagar en mayor profundidad, en como estaría relacionado el Porcentaje de Pobreza sobre los diferentes índices de movilidad, se construyen diferentes gráficas para entender el comportamiento a lo largo del tiempo para cada índice de movilidad y su diferenciación para el Cuartil 1 y el Cuartil 4.

El comportamiento a lo largo del tiempo para el IM total (IM) se representa en la Figura N°11. A primera vista, se distingue que presentan conductas más bien heterogéneas, en el cual, para el primer Cuartil, tiende a mantener su valor después de la aplicación del plan Paso a Paso en contra punto a la respuesta en el Cuartil 4.

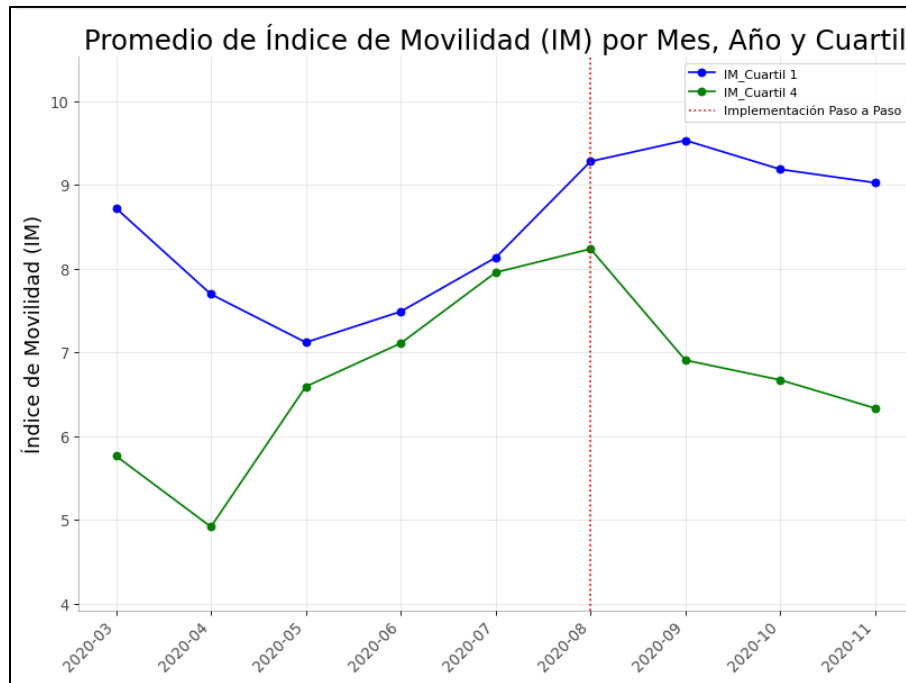


Figura N°11. Promedio de índice de Movilidad (IM) a lo largo del tiempo para ambos cuartiles.

Si la mirada se centra tan solo en el IM interno, se aprecia que ambas líneas tienden a presentar más bien un comportamiento paralelo, aunque nuevamente, tras la aplicación del Plan Paso a Paso, el cuartil 1 se contrae en contra posición a lo que se ve para el Cuartil 4, que tiende ligeramente al alza.

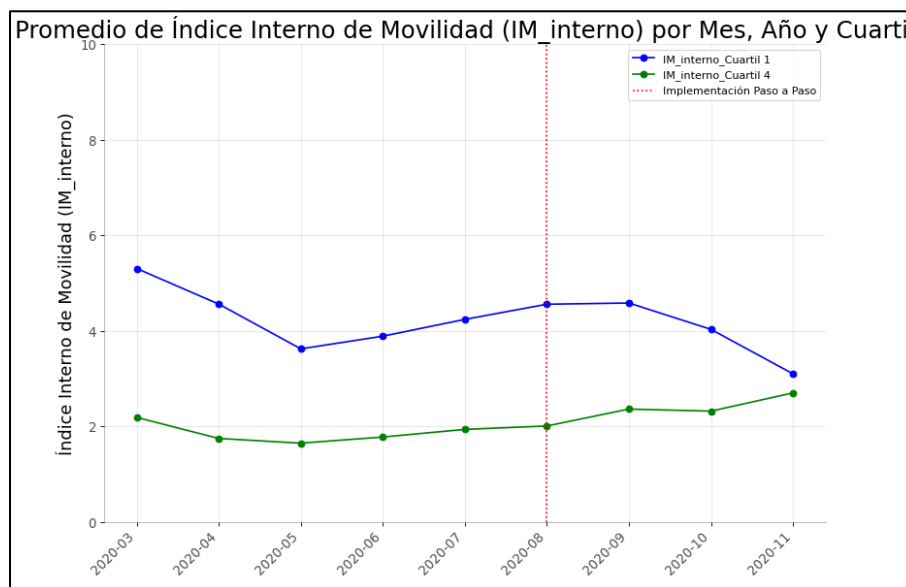


Figura N°12. Promedio de índice de Movilidad Interno a lo largo del tiempo para ambos cuartiles.

Para relevar la diferencia entre las magnitudes visualizadas en la gráfica anterior, se realiza una regresión en base al “Porcentaje de pobreza” y el índice de movilidad interno, añadiendo la segmentación para las comunas pertenecientes a cada Cuartil.

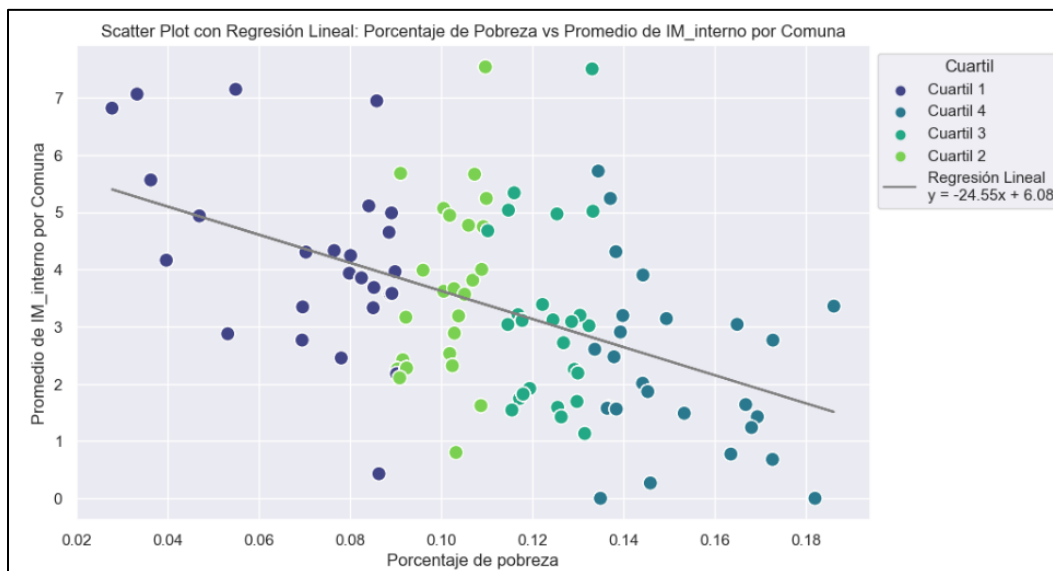


Figura N°13. Scatter Plot para Porcentaje de pobreza y el índice de movilidad interno por comuna.

Profundizando en el estudio de los datos, se construye un boxplot para dar otra mirada a los valores del índice de movilidad interno por comuna. La figura N°14 representa de mejor manera las diferencias que presentarían ambos cuartiles.

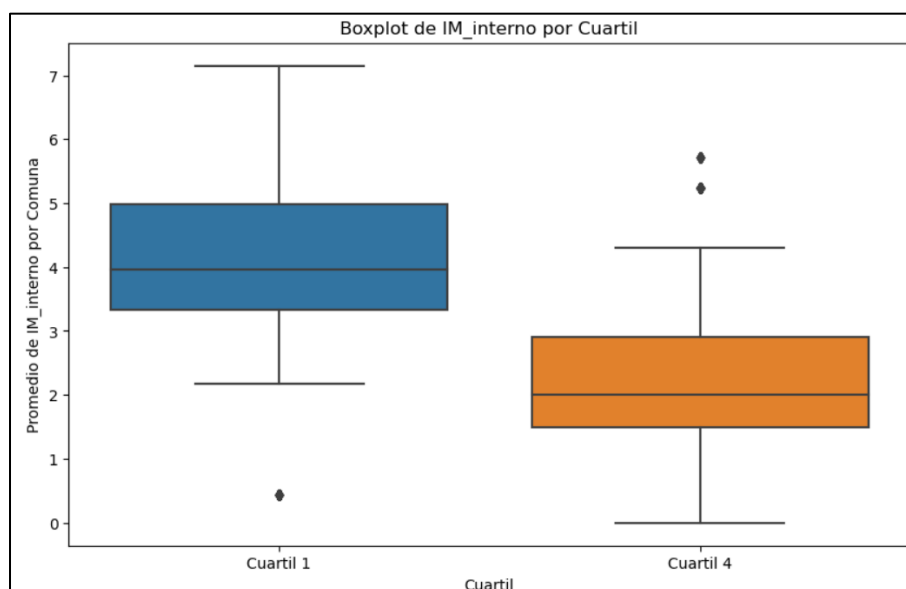


Figura N°14. Boxplot para el IM interno diferenciado por Cuartil.

Adicionalmente, se utiliza un histograma para visualizar la distribución de estos valores. La gráfica refleja la falta de normalidad de los datos, representados en la Figura N°15. Para dar sustento a una diferencia entre ambos cuartiles, teniendo presente este antecedente, se utilizará una prueba estadística no paramétrica como lo es “U de Mann-Whitney”. El resultado obtenido para el índice de movilidad interno revela que efectivamente existe una diferencia significativa para un valor $\alpha = 0.05$.

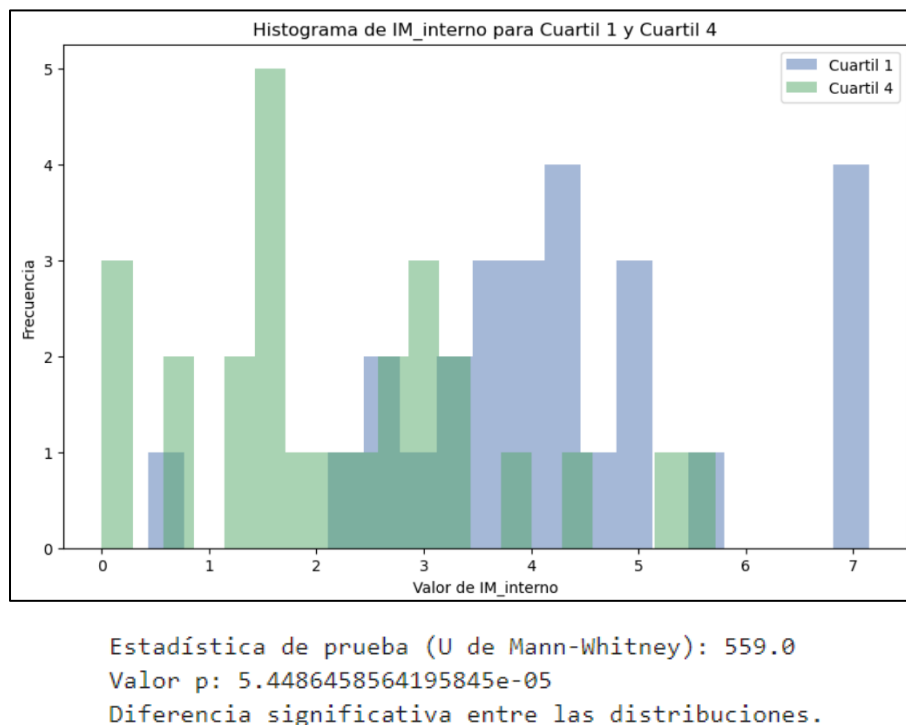


Figura N°15. Histograma para IM Interno en cada Cuartil. Valores y resultado de la prueba estadística.

En cuanto a la comparativa de cuartiles utilizando el “Índice de Movilidad Externo”, el comportamiento a lo largo del tiempo es bastante disímil al anterior, tal como se aprecia en la Figura N°16. El Cuartil 4 presenta un mayor valor promedio mensual durante los primeros meses del confinamiento, pero nuevamente la tendencia cambia tras la aplicación del Plan Paso a Paso a finales del mes de julio de 2020.

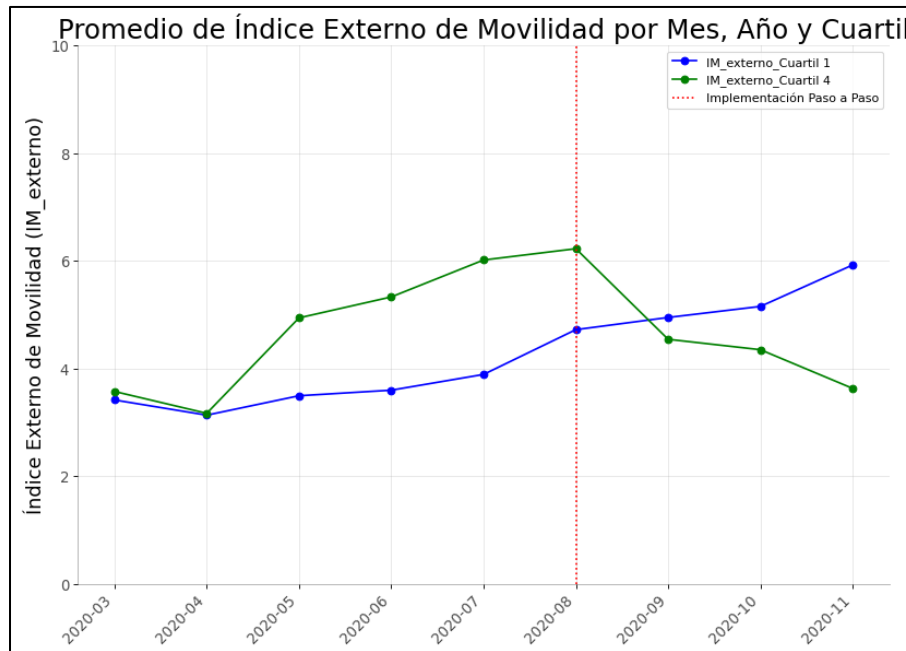


Figura N°16. Valores para IM_externo segmentado por Cuartil.

Para el caso de la gráfica de boxplot, los valores promedio de las comunas muestran otra perspectiva sobre los datos y difiere de lo analizado anteriormente.

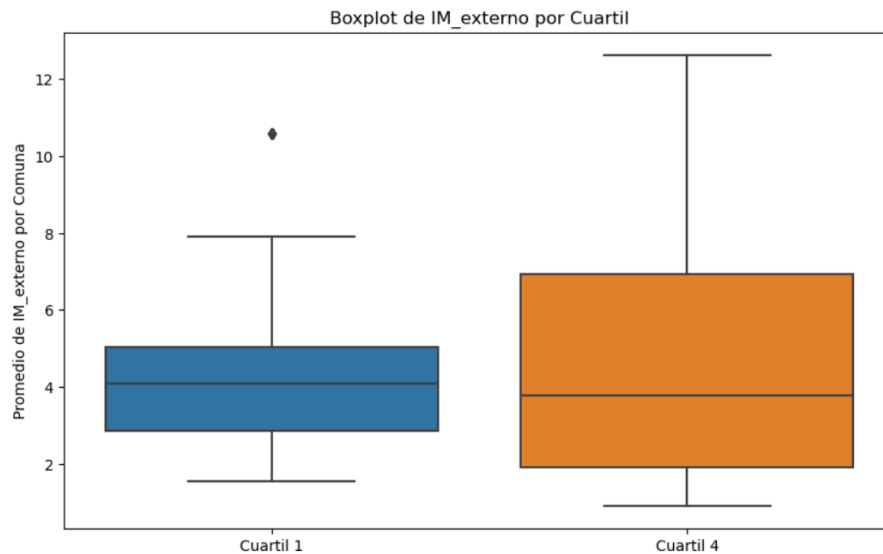
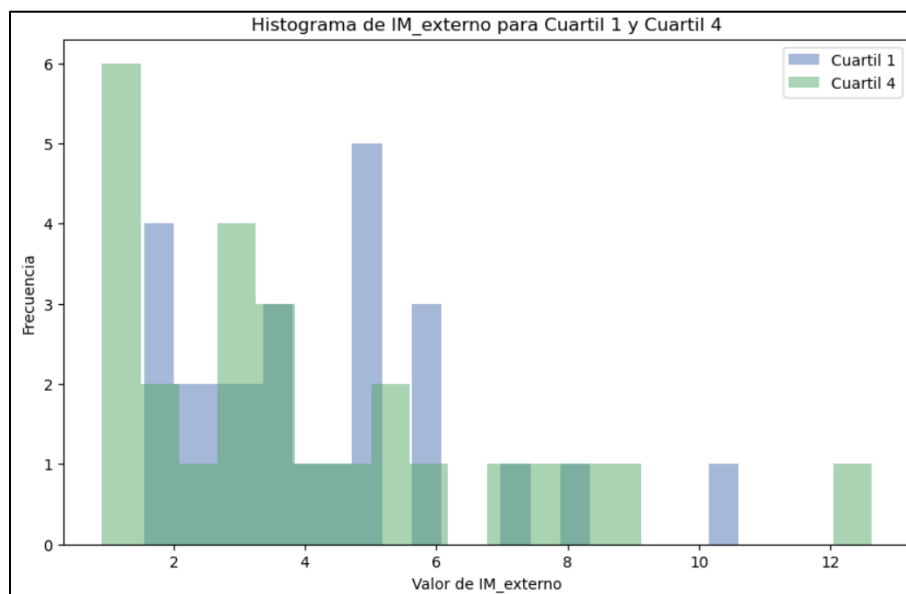


Figura N°17. Boxplot para los valores de IM_externo segmentado por Cuartil.

En cuanto a la distribución, representada en el histograma de la Figura N°18 también se aprecian cambios. No obstante, la distribución de los datos sigue sin presentar normalidad, por lo que se aplica nuevamente la prueba estadística U de Mann-Whitney. El resultado

obtenido para el índice de movilidad externo en este caso establece que efectivamente no existe una diferencia significativa para un valor $\alpha = 0.05$.



Estadística de prueba (U de Mann-Whitney): 384.0
 Valor p: 0.4050115949637756
 No hay evidencia suficiente para afirmar una diferencia significativa entre las distribuciones.

Figura N°18. Histograma para IM externo en cada Cuartil. Valores y resultado de la prueba estadística.

5.4. Resultados entre Cuartiles 1 y 4 en Fin de semana

Un procedimiento equivalente se realiza sobre el set de datos, pero esta vez segmentado particularmente para los días solamente de “Fin de semana”. El comportamiento del Índice de Movilidad interno mantiene un comportamiento similar, tal como lo refleja la Figura N°19:

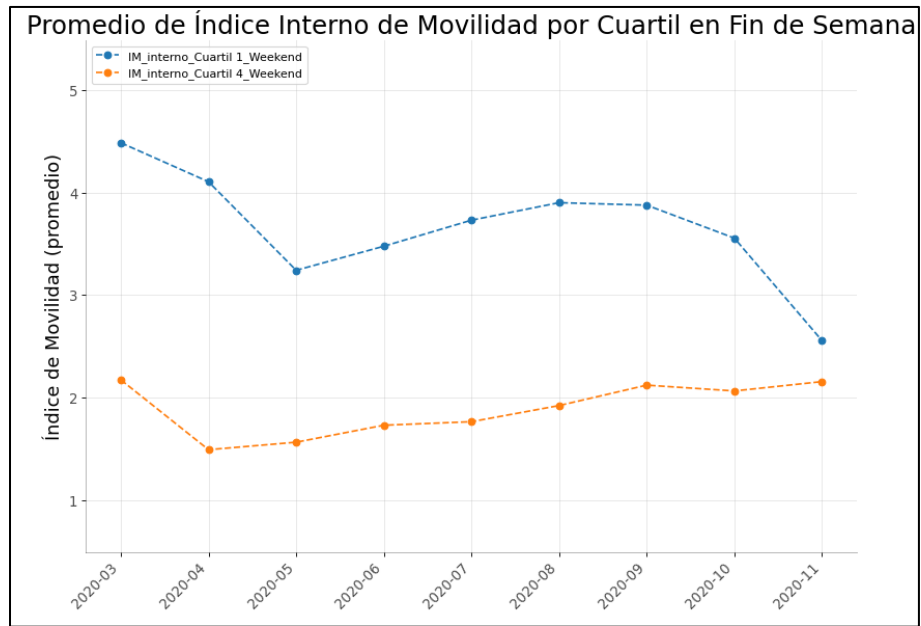


Figura N°19. Gráfico de líneas para valores de IM_interno segmentado por Cuartil y Fin de semana.

De manera análoga, se construye la gráfica de boxplot, así como también el histograma de frecuencias para esta nueva segmentación de la información. Al igual que en el análisis anterior, se aplica el estadístico U de Mann-Whitney para reflejar la existencia de diferencias significativas, el cual es nuevamente positivo teniendo un umbral de alpha 0.05.

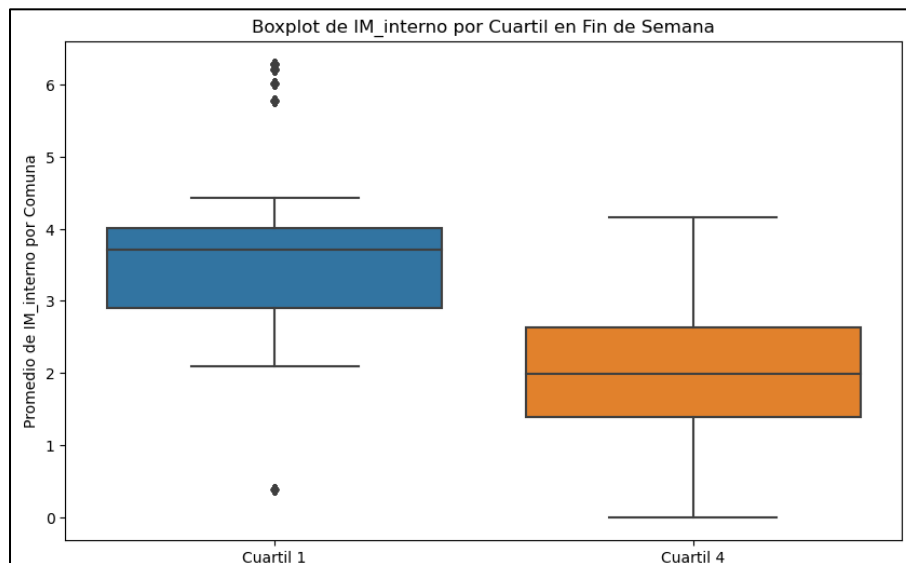
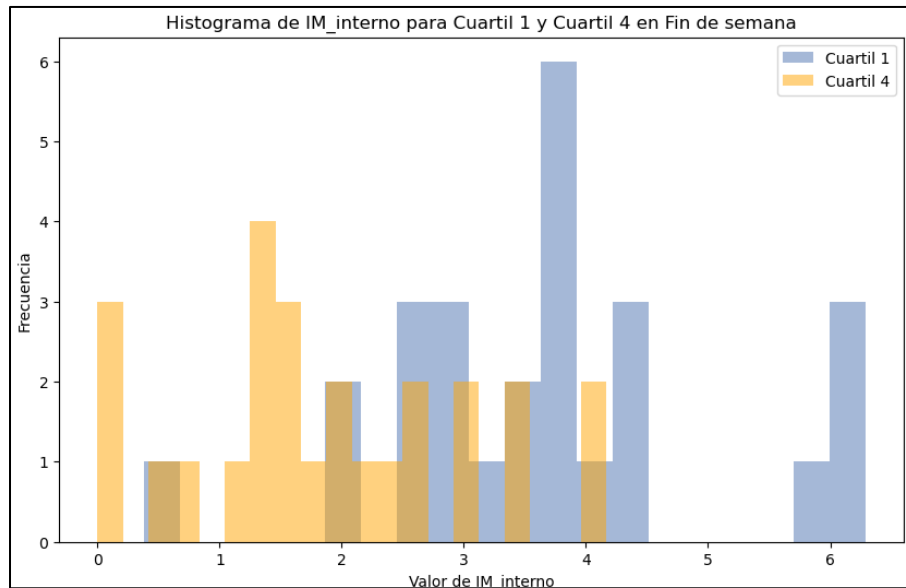


Figura N°20. Boxplot para los valores de IM_interno segmentado por Cuartil y fin de semana.



Estadística de prueba (U de Mann-Whitney): 573.0
 Valor p: 1.7729381809115864e-05
 Diferencia significativa entre las distribuciones.

Figura N°21. Histograma para IM interno en cada Cuartil y fin de semana.

Para el caso del análisis del “Índice de Movilidad Externo” bajo el filtro de días de “Fin de semana”, no se observan cambios en relación con el análisis anterior (todos los días de la semana). Esto a pesar de que la gráfica temporal de las medias para este indicador se devela más pronunciados durante algunos meses de confinamiento.

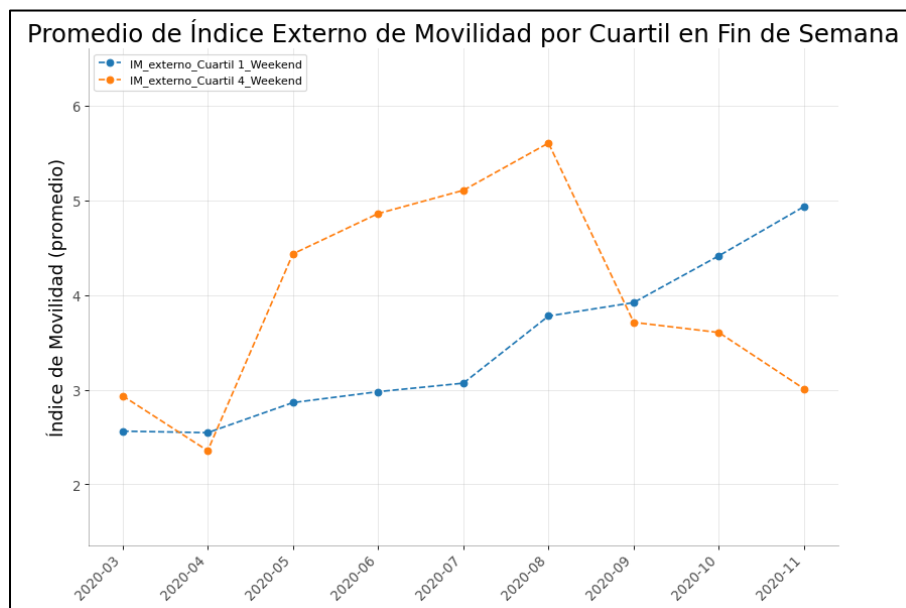


Figura N°22. Gráfico de líneas para valores de IM_externo segmentado por Cuartil y Fin de semana.

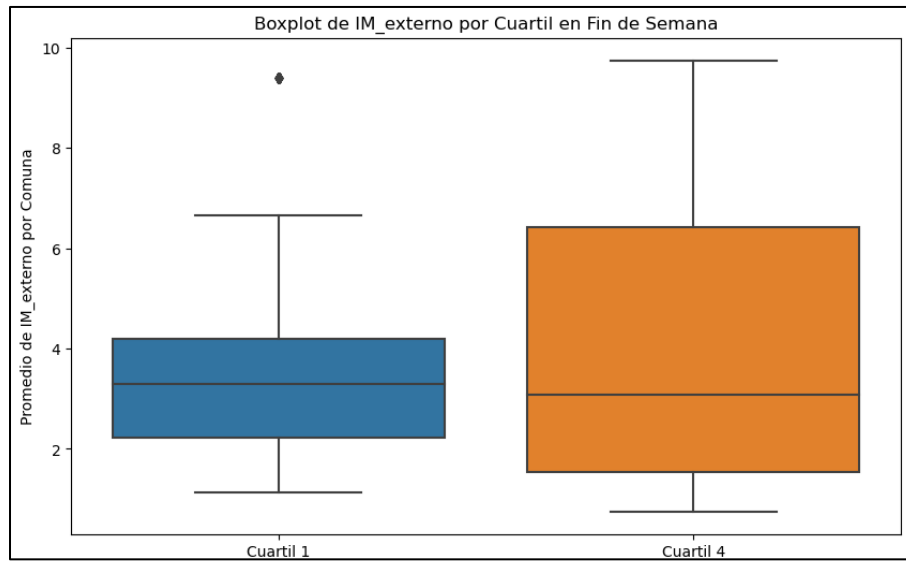
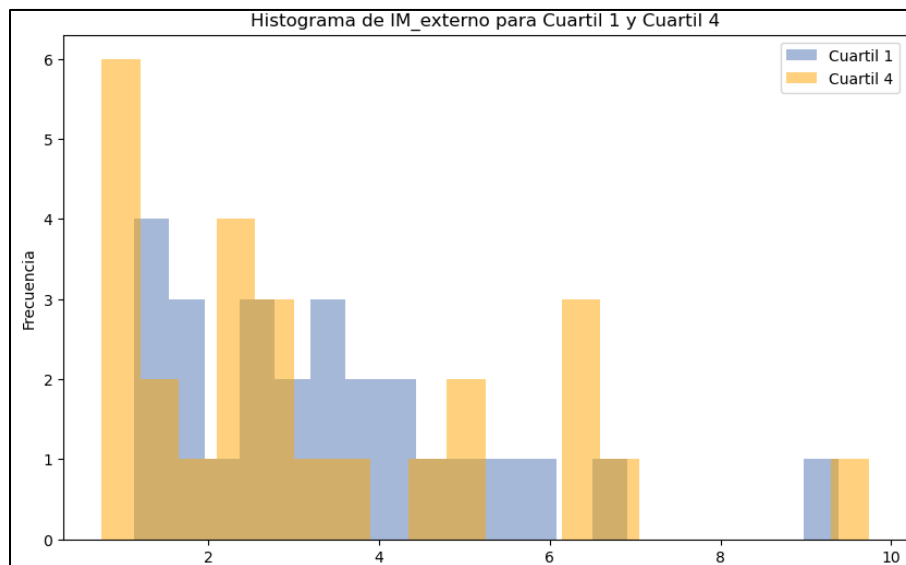


Figura N°23. Boxplot para valores de IM_externo segmentado por Cuartil y Fin de semana.

Nuevamente, y debido a la falta de normalidad en la distribución de las frecuencias, que se repite el mismo estimador estadístico, con umbral alpha de 0.05, el cual entrega como resultado que no hay evidencia suficiente para desechar la hipótesis nula.



Estadística de prueba (U de Mann-Whitney): 375.0
 Valor p: 0.5041377728786078
 No hay evidencia suficiente para afirmar una diferencia significativa entre las distribuciones.

Figura N°24. Histograma para IM interno en cada Cuartil y fin de semana.

5.5. Resultados y análisis intra Cuartiles 1 y 4

Para complementar el análisis sobre los indicadores de movilidad, es que también se analiza el comportamiento de los índices tanto interno como externo dentro de cada cuartil. Para el primer caso, que corresponde a las comunas asignadas como Cuartil 1, la gráfica de valores a lo largo del tiempo se aprecia bastante pareja, como se ve en la Figura N°25, así como también en su boxplot correspondiente.

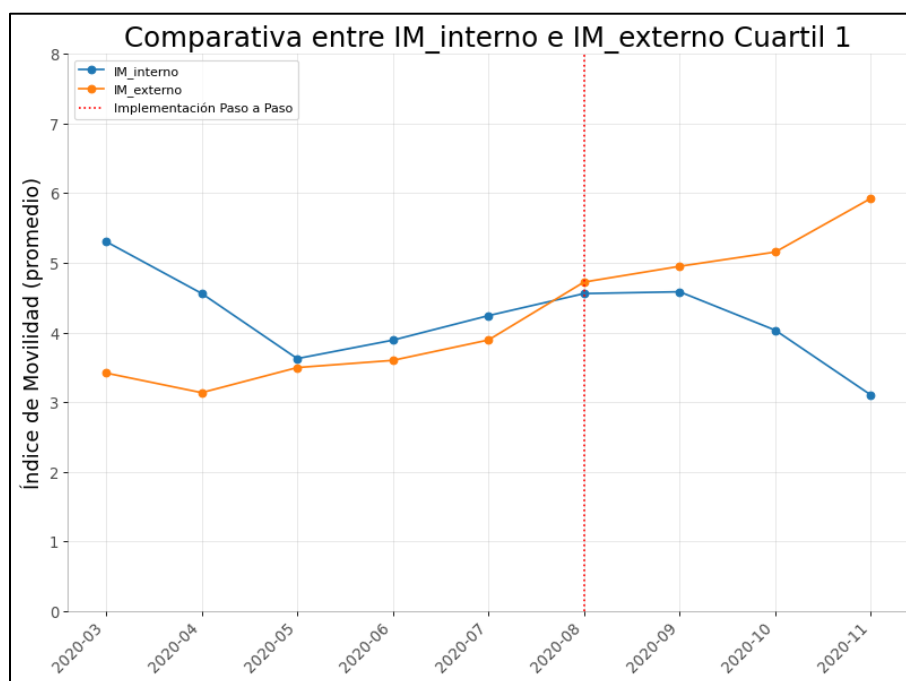


Figura N°25. Gráfico de líneas para valores de IM_externo e interno segmentado por Cuartil 1.

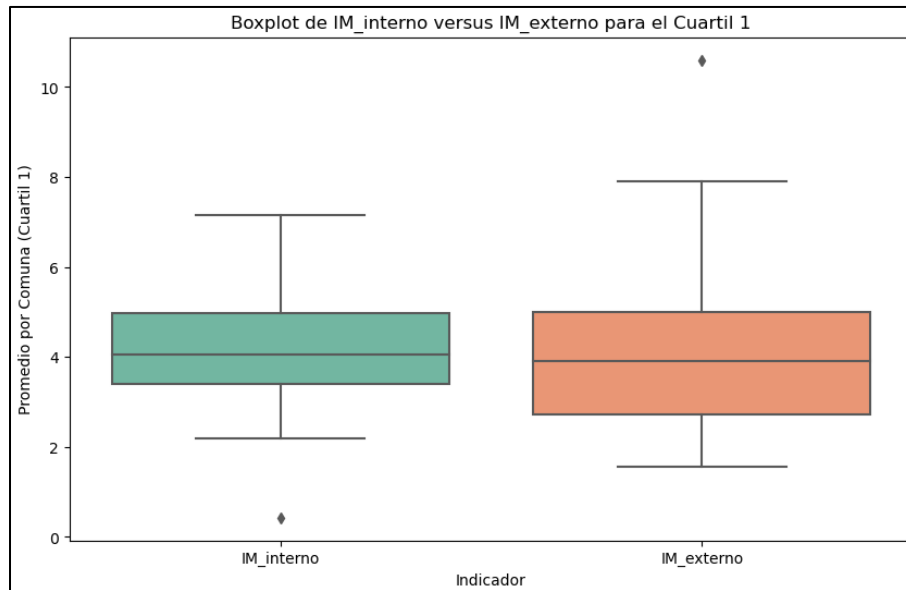
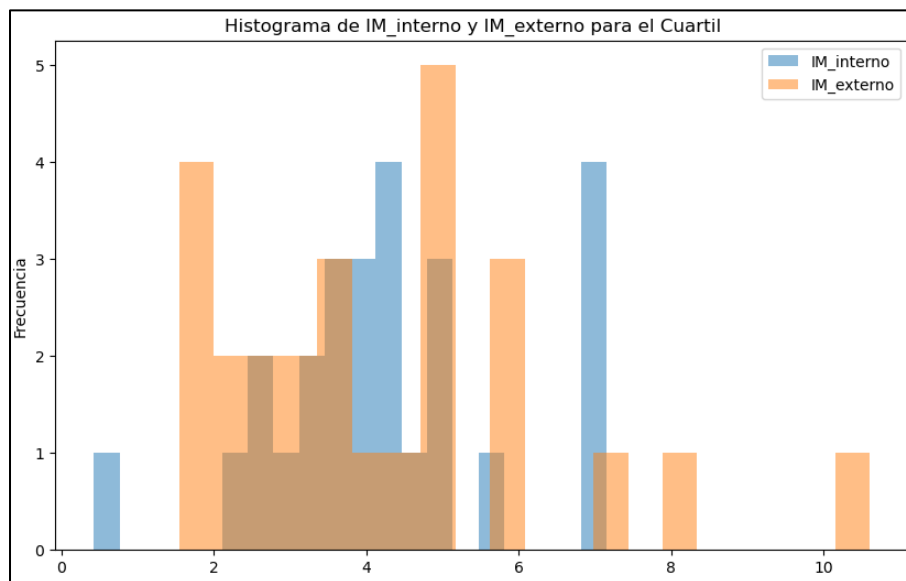


Figura N°26. Boxplot para valores de IM_externo e interno segmentado por Cuartil 1.

La distribución de la frecuencia reflejada en el histograma vuelve a estar sin normalidad, por lo que se aplica la prueba U de Mann-Whitney con límite alpha 0.05, entregando como resultado que no hay evidencia suficiente para afirmar que existe una diferencia significativa.



Estadística de prueba (U de Mann-Whitney) para Cuartil 4: 364.0
 Valor p para Cuartil 4: 0.6407285860495562
 No hay evidencia suficiente para afirmar una diferencia significativa entre las distribuciones de IM_interno e IM_externo.

Figura N°27. Histograma para IM interno y externo para el Cuartil 1.

Para el caso del Cuartil 4, la gráfica temporal de los indicadores se ve bastante distante a lo largo de la escala de tiempo para los valores agrupados por mes. De igual forma para este caso, el resumen entregado por el boxplot se ve más diferenciado a simple vista, mientras que el histograma, fuerza nuevamente a aplicar igual estadístico, revelando en esta oportunidad que hay información suficiente para desechar la hipótesis nula, por lo que habría diferencias significativas entre los dos indicadores.

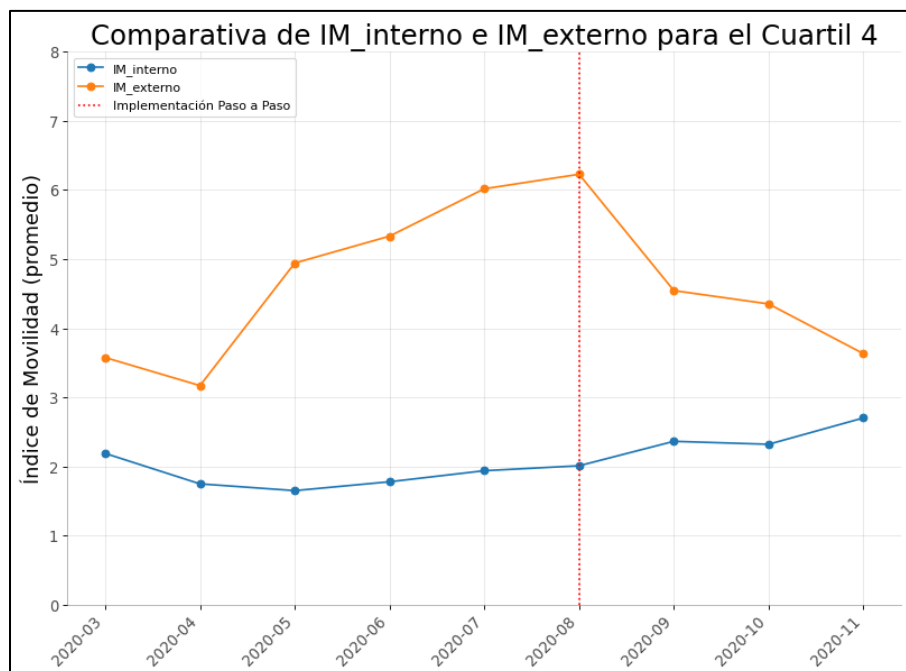


Figura N°28. Gráfico de líneas para IM interno y externo para el Cuartil 4.

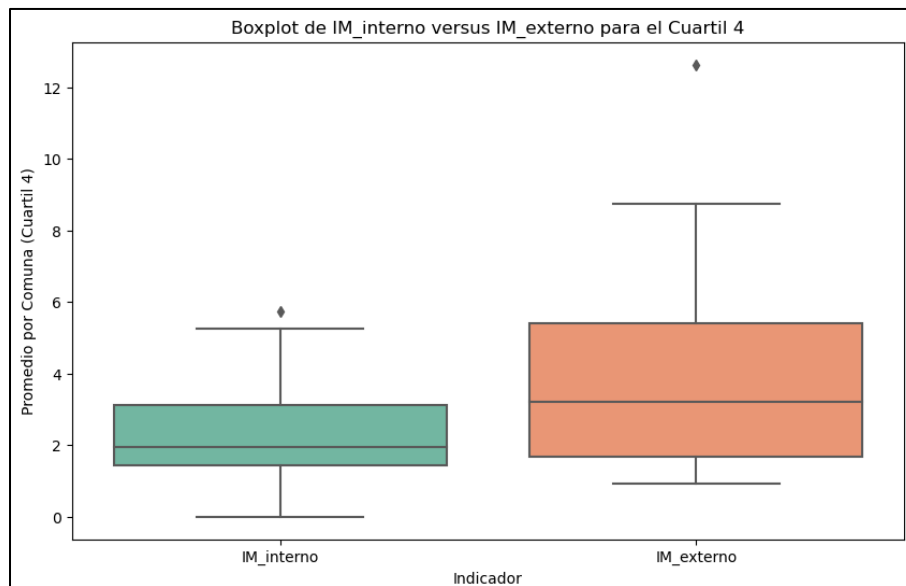
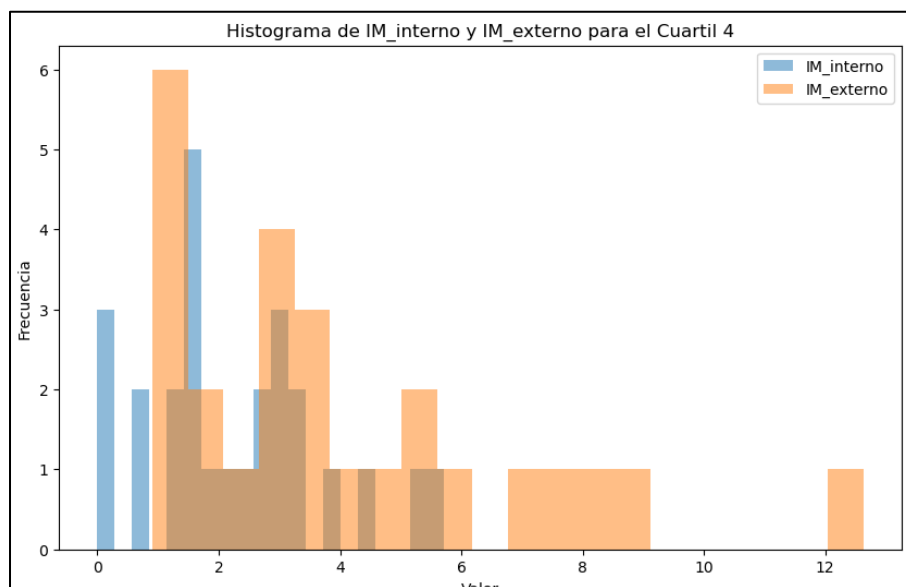


Figura N°29. Boxplot para valores de IM_externo e interno segmentado por Cuartil 4.



Estadística de prueba (U de Mann-Whitney) para Cuartil 4: 218.0
 Valor p para Cuartil 4: 0.028739681602974555
 Diferencia significativa entre las distribuciones de IM_interno e IM_externo.

Figura N°30. Histograma para IM interno y externo para el Cuartil 4.

Esto revela una diferencia entre los dos grupos analizados, ya que las comunas que presentan un mayor porcentaje de pobreza presentan un índice de movilidad externo que es significativamente diferente a su índice de movilidad interno.

5.6. Movilidad durante diferentes fases

De manera complementaria se presenta la evolución de los índices de movilidad separado por cada Fase del Plan Paso a Paso. Para el IM total, en primera instancia se aprecia el cambio de comportamiento en la movilidad debido a la instauración de las Fases, aumentando de manera notoria para los valores promedio de la Fase 2. No obstante, tanto las Fase 1 como la Fase 2 presentan una caída casi sostenida hasta final de año.

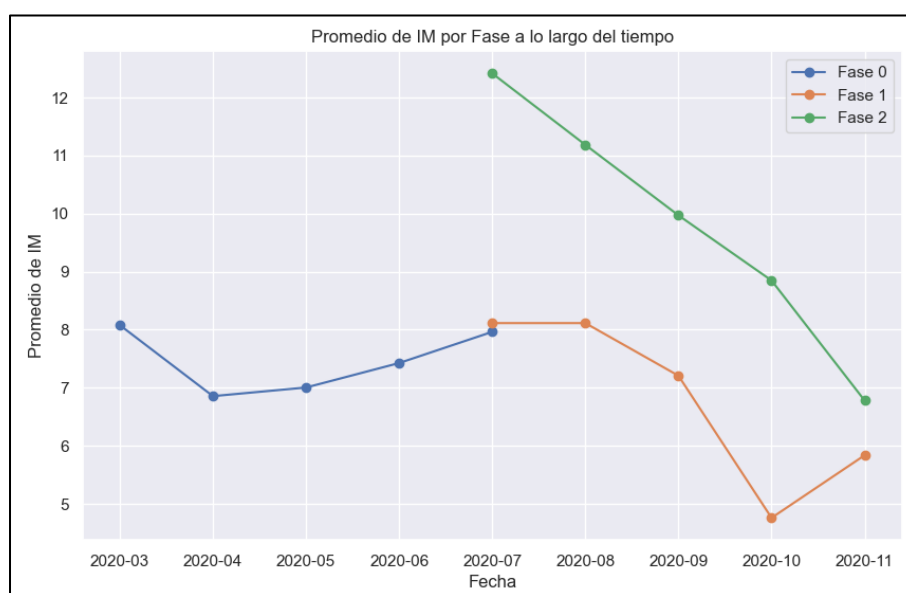


Figura N°31. Evolución temporal del IM total separado por fase.

Si se busca ampliar la mirada sobre el impacto de las Fases del Plan Paso a Paso dentro de los dos cuartiles de estudio, se puede apreciar en la Figura N°32 que el comportamiento a rasgos generales es bastante ambiguo, aunque presiona sobre el indicador de IM total sostenidamente tras su implementación. No obstante, se observan también valores menores para el Cuartil 4, al menos durante la Fase 2 del plan.

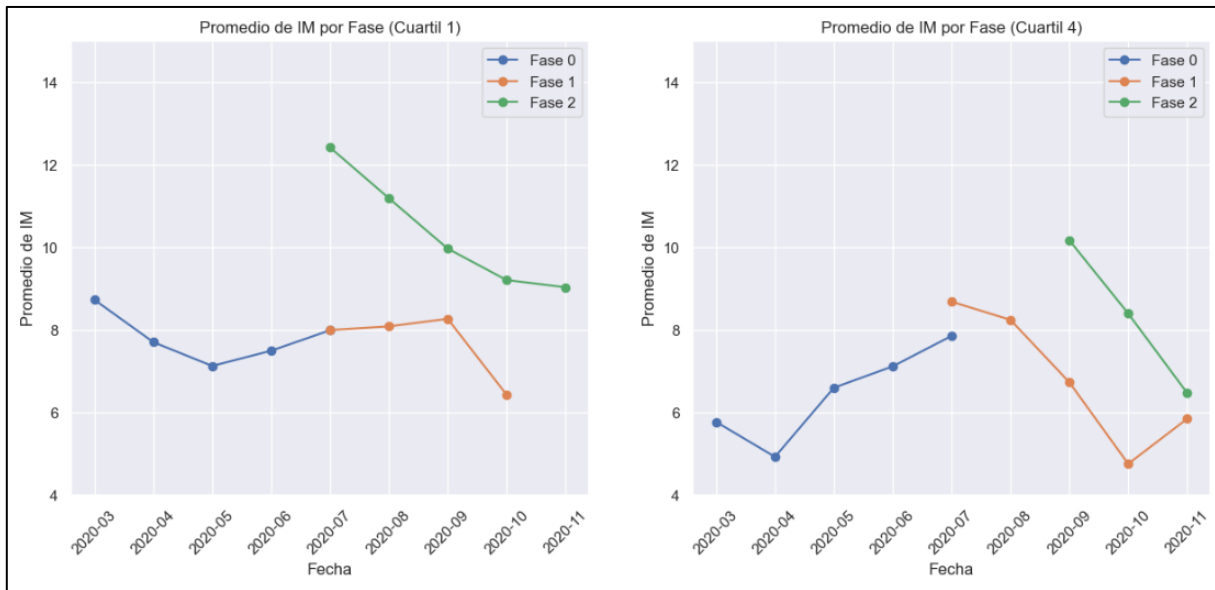


Figura N°32. Evolución temporal del IM total separado por fase y Cuartil.

5.7. Resultados del modelo de Regresión de Vectores de Soporte (SVR)

Para la creación del modelo, se utilizará el porcentaje de pobreza y la densidad poblacional como variables independientes, y el IM_interno como la variable dependiente. La elección de estas variables radica en su relevancia socioeconómica y su potencial impacto en la movilidad interna, respaldado por nuestro análisis de correlaciones.

En cuanto a los preparativos del modelo, se utiliza StandardScaler para eliminar el sesgo en la escala en la que se encuentran los datos, muy importante para la naturaleza del modelo elegido. Estandarizar, ayuda a que los algoritmos de optimización converjan más rápidamente. Muchos algoritmos, incluyendo SVR, pueden beneficiarse de datos estandarizados, ya que esto ayuda a evitar que ciertas características tengan un peso dominante debido a su escala.

Al entrenar el algoritmo con el 80% de datos, se obtienen resultados moderados que permiten avanzar y responder la hipótesis:

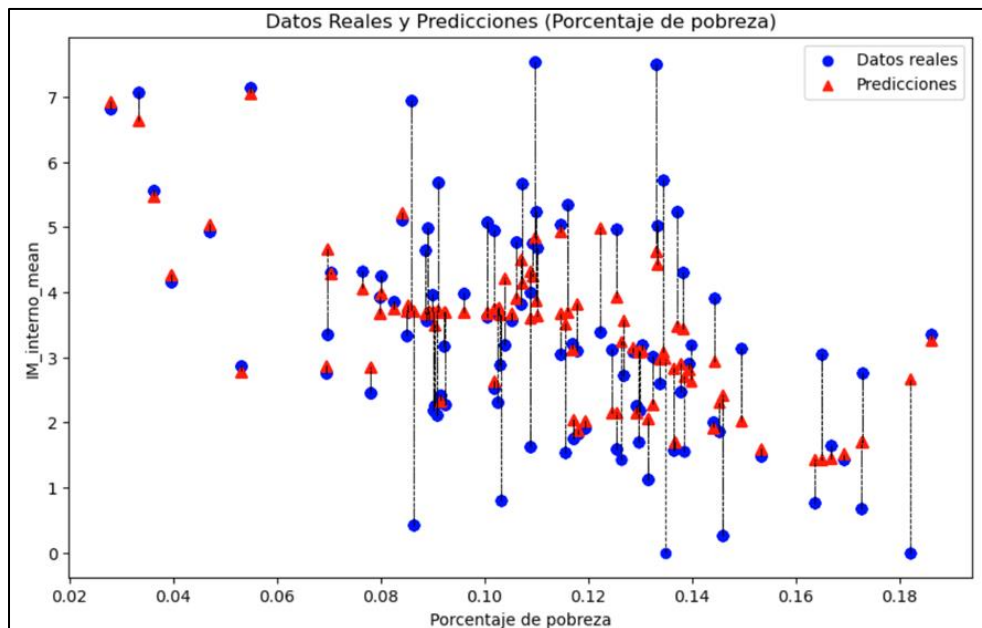


Figura N°33. Datos reales y predicción realizadas por el modelo SVR.

Al observar el gráfico donde se comparan los datos reales con la predicción del modelo, se valida la hipótesis de la correlación negativa entre estas dos variables y se aprecia el rendimiento moderado del modelo desarrollado.

El error cuadrático medio (MSE) obtenido por el modelo es de 1.2082, y funciona como soporte a lo mencionado anteriormente, ya que medimos la discrepancia entre las predicciones y los valores reales. Otro indicador que refleja el rendimiento del modelo es el R^2 , que tiene un valor de 0.5475, donde se logra explicar el 54.75% de la variabilidad en el IM_Interno.

Con el objetivo de testear el modelo, se genera un ejemplo de predicción con datos al azar ingresados por teclado. En este caso, se ingresa un porcentaje de pobreza de 15%, una densidad de 10,000 personas por Km², dando como resultado un índice de movilidad de interno de 1.667. El intervalo de confianza proporciona una estimación de la incertidumbre asociada a esta predicción, mostrando que el valor predicho podría estar en el rango (-0.4892, 3.8236). Este intervalo es útil para comprender la variabilidad potencial en las predicciones del modelo.

De esta manera, los resultados obtenidos a través de la predicción se alinean con lo expresado por los datos originales, tal como se evidencia en las Figura N°34 y Figura N°35.

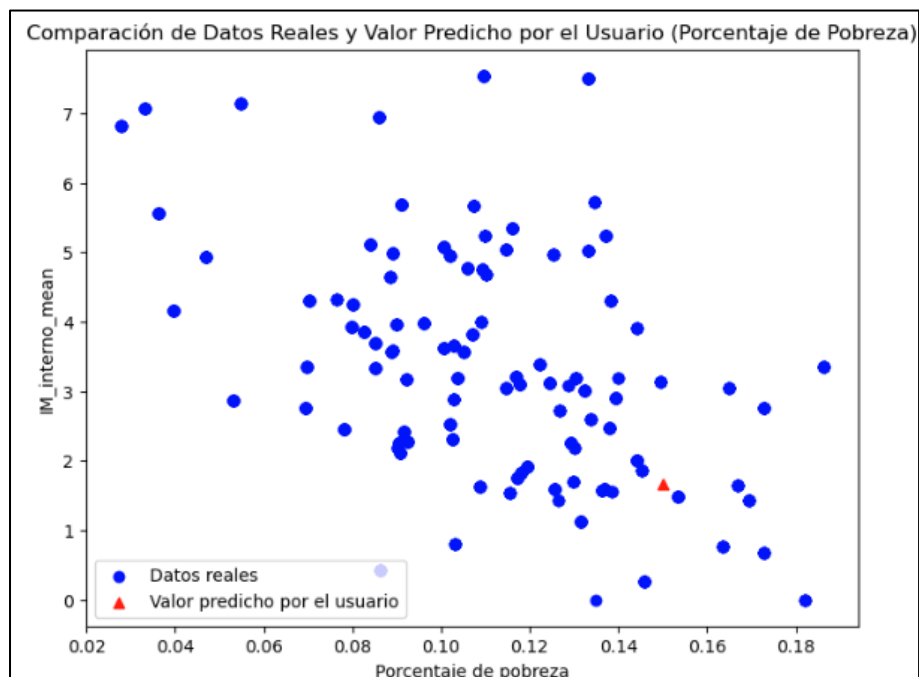


Figura N°34. Valor predicho por el usuario en relación al Porcentaje de Pobreza.

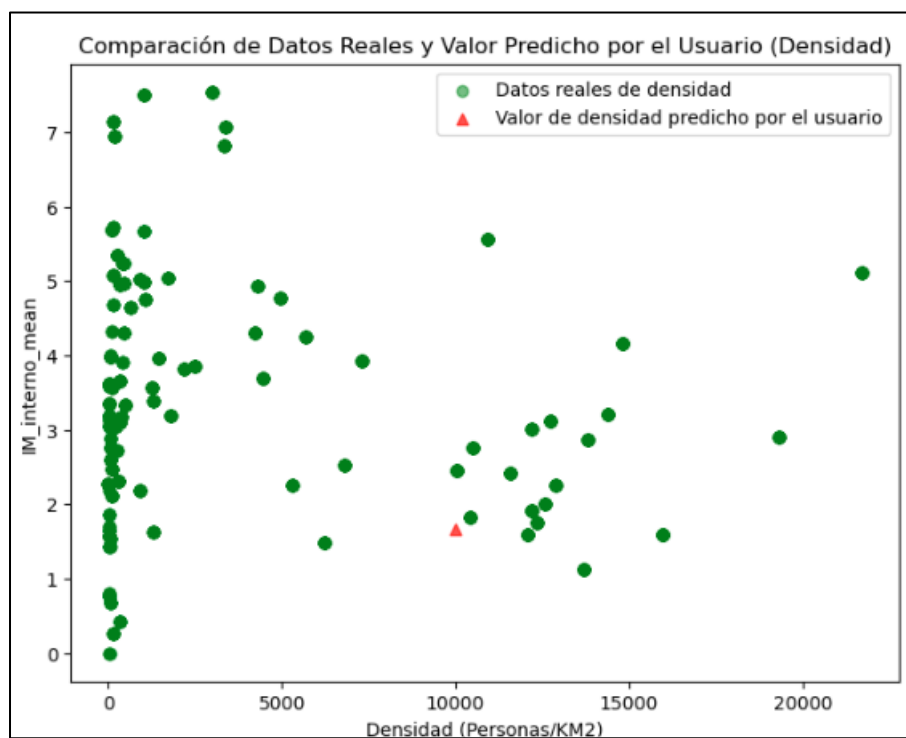


Figura N°35. Valor predicho por el usuario en relación a la densidad.

6. Conclusiones y trabajo futuro

Sin dudas que el impacto de la pandemia en Chile aún presenta muchas áreas por investigar. En esta oportunidad, y a través de este trabajo, se ha tratado de vislumbrar sus consecuencias con un foco particular como lo son las diferencias socioeconómicas, la densidad de la población y su impacto sobre la movilidad de las personas.

Los resultados obtenidos en la comparativa sobre los indicadores de movilidad, a pesar de que son bastante escuetos, abren una puerta al entendimiento de cómo reaccionan diferentes grupos de la población ante medidas tan drásticas como una cuarentena.

Por ejemplo, no deja de ser llamativo que zonas que tienen un mayor porcentaje de pobreza dentro de su población, requieran o se hayan movilizadado en mayor medida entre diferentes comunas, más que dentro de la de su residencia, o por qué el grupo identificado con una menor pobreza, que intuitivamente tiene una mayor posibilidad de adaptación, haya mostrado una mayor movilidad interna comparada al Cuartil 4. Así mismo, se evidencia que el comportamiento de ambos grupos tampoco difirió durante los días de fin de semana, aunque sus dinámicas sociales, por default sean muy disimiles.

La mayor limitante detectada durante la evaluación de los índices de movilidad viene explicada por la baja posibilidad de utilizar estadísticos de mayor robustez o paramétricos que provean un mayor soporte. A pesar de que se utilizaron técnicas que permitieron evidenciar ciertas diferencias significativas, la alta heterogeneidad de los datos, aún más durante las diversas etapas del Plan Paso a Paso, obliga a reducir el área de estudio para poder acotar y mejorar estos indicadores.

Sin dudas, que, como perspectiva de futuro, sería bueno analizar el comportamiento de los índices de movilidad posterior a los estados de excepción y restricciones de movilidad, para corroborar que estos patrones se mantienen o han diferido hacia una etapa prepandemia. Logrando extraer más aún información de las consecuencias en la limitación de la movilidad y las medidas restrictivas o no farmacológicas.

Debiese ser imperante para los tomadores de decisiones a nivel regional, comunal o hasta nacional, entender y estudiar en profundidad este tipo de dinámicas, ya que este tipo de conocimiento permite aplicar políticas públicas más efectivas o eficaces ante escenarios adversos.

Por otra parte, la construcción de modelos predictivos nos permite ir un paso adelante ante las catástrofes, permitiendo salvar vidas o impactos negativos en la población. Este estudio, que se centró en analizar la relación entre variables socioeconómicas, específicamente el porcentaje de pobreza y la densidad poblacional, con el índice de movilidad interna (IM_interno) puede abrir una puerta en esa dirección.

El modelo de regresión de vectores de soporte (SVR) entrenado con el 80% de los datos proporcionó resultados moderados. La validación de la hipótesis de correlación negativa entre las variables independientes y el IM_interno se respalda al observar la comparación entre datos reales y predicciones en el gráfico. El error cuadrático medio (MSE) de 1.2082 indica una discrepancia moderada entre las predicciones y los valores reales. Además, el coeficiente de determinación (R^2) de 0.5475 sugiere que el modelo explica el 54.75% de la variabilidad en el IM_interno.

Para evaluar la robustez del modelo, se realizó una prueba utilizando datos al azar ingresados por teclado. Los resultados de esta prueba, que incluyen un intervalo de confianza, respaldan la capacidad del modelo para proporcionar predicciones coherentes incluso en condiciones no observadas durante el entrenamiento.

A pesar de los logros, hay aspectos que podrían abordarse en investigaciones futuras. Por ejemplo, la inclusión de más variables podría mejorar la precisión del modelo. Explorar la dinámica temporal de las variables y su impacto en el IM_interno podría proporcionar una comprensión más profunda de los patrones de movilidad. Además, la evaluación de la robustez del modelo podría extenderse mediante pruebas con conjuntos de datos adicionales,

lo que ayudaría a determinar la generalización del modelo a diferentes contextos geográficos o temporales

Las limitaciones del estudio incluyen la dependencia de los datos disponibles y la posible simplificación de la complejidad del fenómeno de la movilidad interna. Además, aunque se utilizó StandardScaler para abordar el sesgo en la escala de datos, podrían explorarse otras técnicas de normalización y su impacto en el rendimiento del modelo.

Una limitación evidente es la dependencia de los datos disponibles, lo que podría introducir sesgos y limitar la generalización del modelo a diferentes contextos. La interpretación de la correlación negativa entre las variables independientes y el IM_interno también puede ser influenciada por la calidad y representatividad de los datos utilizados.

Aunque se utilizó StandardScaler para estandarizar los datos y facilitar la convergencia de los algoritmos de optimización, existen otras técnicas que podrían explorarse para abordar la escala de datos de manera más efectiva.

La capacidad del modelo para manejar situaciones extremas o cambios abruptos en las variables también podría ser una limitación. Además, la interpretación de la correlación como causalidad debe abordarse con precaución, ya que otros factores no considerados en este estudio podrían influir en la movilidad interna.

En resumen, a pesar de los resultados positivos, es esencial abordar estas limitaciones y considerar el trabajo futuro para mejorar la robustez y aplicabilidad del modelo en entornos más amplios.

Bibliografía

- Kellermann, R., Sivizaca Conde, D., Rößler, D., Kliewer, N., & Dienela, H.-L. (2022). Mobility in pandemic times: Exploring changes and long-term effects of COVID-19 on urban mobility behavior. *Transportation Research Interdisciplinary Perspectives*, 15, 100668. <https://doi.org/10.1016/j.trip.2022.100668>
- Gozzi, N., Tizzoni, M., Chinazzi, M., Ferres, L., Vespignani, A., & Perra, N. (2021). Estimating the effect of social inequalities on the mitigation of COVID-19 across communities in Santiago de Chile. *Nature Communications*, 12, 2429. <https://doi.org/10.1038/s41467-021-22687-w>
- Pappalardo, L., Cornacchia, G., Navarro, V., Bravo, L., & Ferres, L. (2023). A dataset to assess mobility changes in Chile following local quarantines. *Scientific Data*, 10, 6. <https://doi.org/10.1038/s41597-022-01256-1>
- Pappalardo, L., Ferres, L., Sacasa, M., Cattuto, C., & Bravo, L. (2021). Evaluation of home detection algorithms on mobile phone data using individual-level ground truth. *EPJ Data Science*, 10, 29. <https://doi.org/10.1140/epjds/s13688-021-00298-4>