



Universidad del Desarrollo
Facultad de Ingeniería

**PATRONES DE MOVILIDAD ENTRE LOS PASOS 1 Y 2 DEL PLAN PASO A
PASO DURANTE LA PANDEMIA DEL COVID-19 EN CHILE**

POR: KEVIN DIEGO GALLARDO TORO

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. Leo Ferres

Noviembre 2023

VALPARAÍSO

A mi Madre, quien sin su apoyo no estaría aquí;
a mi pareja, quien estuvo a mi lado durante el
desarrollo de este proyecto; a mis hermanos y a
mi abuela, que son pilar fundamental en mi
vida.

AGRADECIMIENTO

En primer lugar, agradezco a la Universidad de Desarrollo por mantener un nivel espectacular en sus programas de postgrado. Agradezco también al profesor Dr. Leo Ferres por su apoyo en la realización de este proyecto y a todos los demás profesores y compañeros que participaron en este programa de Magíster.

TABLA DE CONTENIDO

RESUMEN.....	1
1. INTRODUCCIÓN	2
2. TRABAJO RELACIONADO	4
3. HIPÓTESIS Y OBJETIVOS.....	5
4. DATOS Y METODOLOGÍA	6
4.1. DATOS	6
4.2. METODOLOGÍA.....	11
5. RESULTADOS	12
6. CONCLUSIÓN Y TRABAJO FUTURO.....	46
BIBLIOGRAFÍA	48

Resumen

La implementación de cierres parciales durante la pandemia del Covid-19 fue, es y probablemente seguirá siendo tema de investigación y de discusión para muchos. Chile, al igual que otros países en el mundo, estableció este tipo de medidas y en Julio de 2020, para hacer más claras las etapas para la población, creo el Plan Paso a Paso. Este plan consistía en 5 Pasos, que iban desde el cierre total a aperturas casi totales de la movilidad, permitiendo la obtención de permisos para movilizarse durante las etapas de mayor restricción. En este proyecto, analizaremos de manera gráfica las diferencias de movilidad entre los pasos 1 y 2 del Plan Paso a Paso, utilizando como principal indicador un índice de movilidad creado a partir de datos de interacción de teléfonos móviles con antenas (específicamente eXtended Detail Records o XDRs). Además, veremos la relación de la movilidad en Chile con factores económicos y medioambientales a través de correlaciones a nivel nacional y regional, junto con importancia de estas variables según modelos de regresión.

1. Introducción

El virus SARS-CoV-2 causó un antes y después para muchos, un patógeno que hizo que una tremenda parte del mundo se paralizara^[11] y comenzara a tomarse en serio el usar mascarillas, lavarse las manos y mantener distancia social.^[1] Por supuesto estas no fueron las únicas medidas “No Farmacológicas” implementadas, en Chile el uso de termómetros y alcohol gel en prácticamente todos los lugares públicos junto con las cuarentenas, fueron algunas de las medidas más cuestionadas.^[2]

El 3 de Marzo del 2020 se confirmó el primer caso de Covid-19 en Chile, la enfermedad producida por el coronavirus había definitivamente llegado al país. Luego de este lamentable acontecimiento, no pasó mucho tiempo hasta que las primeras cuarentenas fueran comunicadas a la población. El gobierno declaró el estado de catástrofe, uno de los cuatro estados de excepción constitucional^[6], el 18 de Marzo de 2020; el 25 de Marzo se declararon las primeras cuarentenas en la región metropolitana y desde ahí continuaron distintas medidas de confinamiento.^{[3][4][5]}

El 19 de Julio del 2020, para hacer más clara la comunicación de las medidas de confinamiento en el país, se estableció el Plan Paso a Paso^[7], un plan consistente de 5 etapas:



Imagen 1 - Etapas del Plan Paso a Paso. [7]

En este proyecto nos concentramos en las dos primeras, siendo estas las más restrictivas en cuanto a disminuir la movilidad de las personas se refiere. El Paso 1, tal como dice la descripción en la Imagen 1, establece una movilidad limitada para disminuir al mínimo la interacción y propagación del virus, esto significaba que las personas no podían moverse libremente ningún día de la semana, excepto por permisos muy limitados. ^[40]

El Paso 2 por su parte “Disminuye el grado de confinamiento”, esto significó que las personas podían moverse casi libremente durante los días de la semana, pero las restricciones se mantenían en un grado elevado para los fines de semana y festivos.

Es importante mencionar que el movimiento desde comunas en Pasos superiores hacia comunas de Pasos inferiores estaba también muy limitado, estableciéndose cordones sanitarios en varios puntos de alto tráfico. Además, durante la noche existía un toque de queda que se mantuvo, con varias modificaciones, hasta Septiembre del 2021. ^[8]

En este proyecto inspeccionaremos de manera gráfica como fue la movilidad de los Chilenos durante la pandemia del covid-19, con especial énfasis en la movilidad de las comunas en Pasos 1 y 2; utilizando como ejemplo principal la región Metropolitana, la cual es conocida por sus separaciones icónicas entre “Ricos y Pobres”, “Sobre y bajo plaza Italia” u “Oriente, centro y periferia”, que hacen referencia a las diferencias sociales y económicas existentes en las distintas zonas de la capital. ^{[9][10][12]}

Además, revisaremos si fueron relevantes aspectos socioeconómicos y ambientales en la movilidad, apoyándonos de correlaciones e importancia de variables entregadas por modelos de regresión tipo Random Forest, Gradient Boosting y Arboles de decisión.

2. Trabajo Relacionado

Distintas disciplinas han estudiado la movilidad, relacionándola a factores geográficos y climáticos ^{[14][15]}, a segregación territorial y residencial ^{[10][12][16]}, además de otros factores sociales y económicos, como la distancia a puntos neurálgicos, trabajos y disponibilidad de transporte ^{[17][18][31]}.

También es fácil encontrarse con estudios de la movilidad durante o posterior a importantes pandemias, que toman en cuenta todo lo mencionado anteriormente. Considerando como el movimiento de las personas afectó a la propagación de las enfermedades, como ciertas medidas de restricción de movimiento y contacto ayudaron al control de estas; y como las distintas realidades de las personas influyeron en como vivieron o viven la lucha contra estos patógenos. Ejemplos tales como el dengue ^[19], el virus de la influenza H1N1 ^{[20][21]} y por supuesto el Covid-19 ^{[22][23][24]}.

Fundamental para este proyecto fue la creación del dataset para medir cambios en la movilidad siguiendo cuarentenas locales ^[25] (nombre original: “A dataset to assess mobility changes in Chile following local quarantines”), trabajo realizado por Luca Pappalardo, Giuliano Cornacchia, Victor Navarro, Loreto Bravo y Leo Ferres, donde se creó el índice de movilidad utilizado en este proyecto. Desde ese estudio ya podemos notar como cambió la movilidad cuando se implementaron distintas medidas de restricción en Chile.

Integrantes del equipo antes mencionado tienen también otros estudios donde revisan la movilidad de las personas utilizando datos de telefonía celular, explicando en ellos distintos detalles de cómo se pueden aprovechar estos datos para identificar lugares de trabajo y de vivienda. Explicando también como se dio origen al índice de movilidad, las ventajas del uso de eXtended Detail Records (XDRs) en comparación a Call Detail Records (CDRs) o Control Plane Records (CPRs); y según esto último, como se relaciona la movilidad con la interacción de dispositivos celulares y antenas de telefonía.^{[26][27]}

3. Hipótesis y Objetivos

La principal hipótesis de nuestro proyecto es que factores socioeconómicos y ambientales afectaron a la movilidad durante los Pasos 1 y 2 del Plan Paso a Paso, de tal manera que comunas con bajos índices de pobreza, bajos índices de hacinamiento o con altos ingresos municipales (por mencionar algunos factores) se diferencian de aquellas con una situación opuesta.

Objetivo general:

- Establecer si existe una diferencia en el índice de movilidad entre los Pasos 1 y 2, teniendo en cuenta las variaciones entre días de semana y fines de semana, considerando factores socioeconómicos y ambientales.

Objetivos específicos:

- Analizar las diferencias en movilidad durante los Pasos 1 y 2.
- Analizar la correlación de variables socioeconómicas y ambientales con el índice de movilidad (interno y externo)
- Analizar la importancia de las variables socioeconómicas y ambientales según modelos de regresión.

4. Datos y Metodología

4.1. Datos

4.1.1 Índice de Movilidad

El principal set de datos que será utilizado en este proyecto es el índice de movilidad por comuna, elaborado a partir de datos de 31 mil millones (31 Billion según el artículo original) ^[25] de eXtended Detail Records (XDRs) correspondientes a la navegación de 5 millones de usuarios, en un período entre el 26 de Febrero y el 30 de Noviembre del 2020 (el rango creció un poco en comparación al artículo original del Dr. Leo Ferres que menciona que los datos llegaban hasta el 20 de Septiembre del 2020).

Para entender la creación del dataset de movilidad, primero revisemos los diferentes datos de navegación que manejan las empresas de telecomunicaciones, siendo estos Call Detail Records (CDRs), eXtended Detail Records (XDRs) and Control Plane Records (CPRs). ^[27]

- **CDRs:** Contiene detalles de llamadas realizadas por un usuario a otro (se guardan ambos identificadores), manteniendo registro de cuánto dura la llamada, la posición de la antena de origen, la posición de la antena que recibe la llamada y cuando ocurrió esta interacción. Se crea un registro por llamada, dejando estos datos como los más escasos en cantidad de conexión a antenas.
- **XDRs:** Estos registros se crean según la descarga y subida de datos de un usuario cada 15 minutos, si el usuario no utilizó una antena de telefonía para descargar datos o si la información no supera cierto límite de MBs en ciertos períodos de tiempo, entonces no se escribe un registro de XDR. Un registro de XDR contiene la posición de la antena involucrada en la descarga o subida de datos, cuantos datos fueron descargados en kilobytes, una estampa de tiempo y el identificador del usuario.
- **CPRs:** Estos registros se crean por cada interacción o “evento” que ocurra en la red de una antena, entre los eventos nombrados en el artículo se encuentran: “handovers” (lo entiendo como el cambio forzado de antena en un dispositivo) y “(re-)activación” de un teléfono. Debido a su naturaleza, este tipo de registros son almacenados en mucha mayor cantidad en comparación a los dos antes mencionados. En este caso se guarda la ubicación de la antena, el identificador del usuario, el evento y la estampa de tiempo.

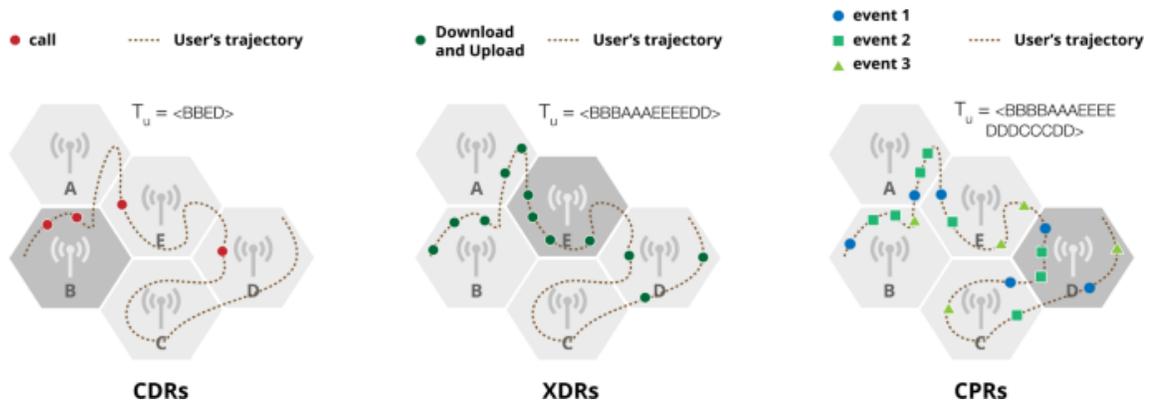


Imagen 2 - Diferencia entre CDR XDR y CPR en términos de conexiones a las antenas. Cada hexágono representa una antena, la línea punteada indica recorrido real de un usuario y los puntos de colores indican distintos tipos de conexión hacia las antenas. [27]

El dataset de movilidad considera movimientos intra-comunales e inter-comunales por día. Cada vez que se registra un XDR en dos antenas distintas dentro de una misma comuna, se considera que el usuario realizó un movimiento intra-comunal; mientras que, si un mismo usuario genera un XDR en dos antenas de comunas distintas, se considera inter-comunal. De esta manera se da origen a tres indicadores:

- IM_{int} o índice de movilidad interna: número de movimientos intra-comunales para ese día.
- IM_{ext} o índice de movilidad externa: número de movimientos inter-comunales para ese día.
- $IM = IM_{int} + IM_{ext}$ (Índice de movilidad)

** Todos normalizados por el número de habitantes de cada comuna. [25]

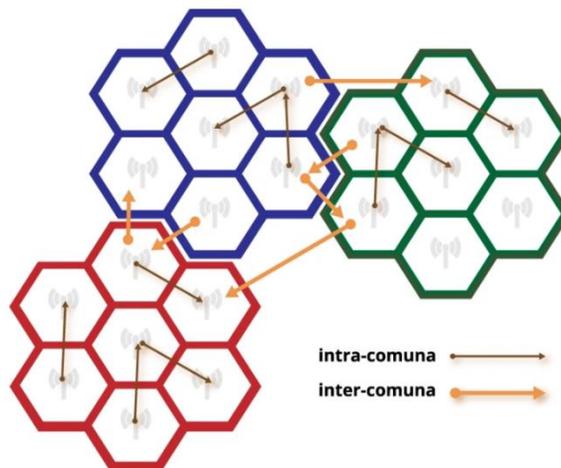


Imagen 3 - Movimientos intra-comunales e inter-comunales, hexágonos del mismo color indican antenas de una misma comuna. [25]

Los datos del índice de movilidad, junto con otros datos relacionados al Covid-19, fueron puesto a disposición por el Ministerio de Ciencia y por el Dr. Leo Ferres en dos repositorios distintos, siendo el más actualizado el del Ministerio de Ciencia (considerar que los datos de movilidad son expuestos como Producto 33).^{[28][29]}

Columnas en el dataset:

- Región
- Código región
- Comuna
- Código comuna
- Superficie km2
- Población
- IM interno
- IM externo
- IM

4.1.2 Etapas del Plan Paso a Paso

Este dataset fue construido a partir de datos expuestos por el Ministerio de Salud al curar información publicada por el Diario Oficial de Chile.^[29] Contiene registros de las distintas etapas de restricción por las que pasó Chile, por comuna, entre las fechas 28 de Julio del 2020 y 7 de Noviembre del 2022. Para este proyecto, considerando las fechas del dataset de movilidad, solo son necesarios los datos hasta el día 30 de Noviembre del 2020. Para este período en específico, se encuentran registros para comunas en los Pasos 1 a 4 del plan Paso a Paso, lo cual es suficiente si consideramos que solo nos centraremos en comunas en los Pasos 1 y 2. (Estos registros fueron publicados como Producto 74 en los repositorios)

Los Pasos no siempre fueron iguales para zonas urbanas y rurales, para algunas comunas y en algunos días encontraremos dos registros distintos que muestran esta diferencia. Sin embargo, para simplificar el análisis nos hemos quedado solo con el registro indicado como “Zona Urbana”, descartando las zonas rurales.

Columnas en el dataset:

- Código región
- Región residencia
- Código comuna
- Comuna residencia
- Zona
- Fecha
- Paso

4.1.3 Perfiles Comunales de la Biblioteca del Congreso Nacional

Desde el 2011 hasta la fecha, el Sistema Integrado de Información Territorial (SIIT) de la Biblioteca del Congreso Nacional (BCN) ha puesto a disposición distintos Perfiles Comunales con el objetivo de comparar unidades territoriales con características similares. Para resumir de manera simple y clara la información, el SIIT construye mapas interactivos donde exponen esta información, permitiendo descargar la base de datos utilizada. Esta base de datos contiene variables Demográficas, Sociales, de Salud, Educativas, Económicas, Municipales, de Seguridad y de Vivienda. ^[30]

Para este proyecto se ha realizado una selección de variables que se consideran representativas de la condición socioeconómica de las distintas comunas:

- Variables Sociales (Ambas basadas en la encuesta CASEN y Censo del 2017)
 - o Tasa pobreza por ingresos
 - o Tasa pobreza multidimensional
 - o Porcentaje población inmigrantes internacionales
- Variables Económicas (Ambas basadas en el Censo del 2017)
 - o Porcentaje de ocupados en rama de actividad económica Primaria
 - o Porcentaje de ocupados en rama de actividad económica Terciaria
- Variables Municipales (Datos del 2019, el documento no especifica fuente para estos datos, aunque se nombran informes municipales, se asume que los datos provienen de ahí)
 - o Disponibilidad presupuestaria municipal por Habitante (M\$)
- Variables de Vivienda (Datos del año 2017 del Observatorio Urbano del Ministerio de Vivienda y Urbanismo)
 - o Porcentaje de Viviendas con hacinamiento

Junto con lo anteriormente mencionado, la base de datos también contiene las columnas de identificación de las comunas, que nos permitirán cruzar la información con los otros datasets (Nombre de la comuna e ID de la comuna).

4.1.4 Datos ambientales

Sabemos que la contaminación del aire está intrínsecamente relacionada a la movilidad, al ser el transporte, tanto personal como público, uno de los principales (sino el principal) sector contaminante en muchas partes del mundo. ^{[32][33]} Considerando esto, recurrimos nuevamente al repositorio del Ministerio de Ciencias ^[29], esta vez para obtener el “Producto 43 - Datos de calidad del aire por hora”. Cada archivo publicado contiene registros de la emisión de algún contaminante por hora durante un año, junto con información de la estación que tomó la medición, siendo la fuente de estos datos lo entregado por el Sistema de Información Nacional de Calidad del Aire (SINCA).

Para este proyecto, tomando en cuenta la disponibilidad de registros y lo leído en artículos relacionados, se han seleccionado los registros de Material Particulado Respirable de menos de 10 µm de tamaño (MP10) y Material Particulado Fino de menos de 2.5 µm de tamaño (MP2.5), ambos del año 2020 y medidos en microgramos por metro cúbico normalizado (µg/m³N).

Por último, como los datos están publicados por hora, se calculó el promedio de las mediciones por día para así coincidir con el dataset principal de movilidad, dejando así una medición por comuna por día.

Cada archivo publicado contiene las siguientes columnas:

- Nombre de estación
- Región
- Código región
- Comuna
- Código comuna
- Coordenadas UTM Este (Ubicación de la estación)
- Coordenadas UTM Norte (Ubicación de la estación)
- Fecha
- Contaminante (En este caso MP 10 y MP 2.5)

4.1.5 Mapas vectoriales

Estos datos nos permitirán construir los mapas coropléticos al proveer de distintos archivos con extensiones compatibles con programas de Sistemas de Información Geográfico. Disponibles en el sitio de la BCN ^[34] nos encontramos con divisiones por comuna, región y provincia (entre otros datasets interesantes).

4.2. Metodología

Primero, en la exploración de los datos, intentaremos entender la movilidad en general durante el año 2020. Para eso hemos segmentado los registros de movilidad en 3 períodos:

- Anterior a las restricciones
 - Período entre el 26 de Febrero y el 18 de Marzo
- Restricciones severas antes del Paso a Paso
 - Período entre el 18 de Marzo y el 18 de Julio
- Post implementación del Plan Paso a Paso
 - Período posterior al 19 de Julio

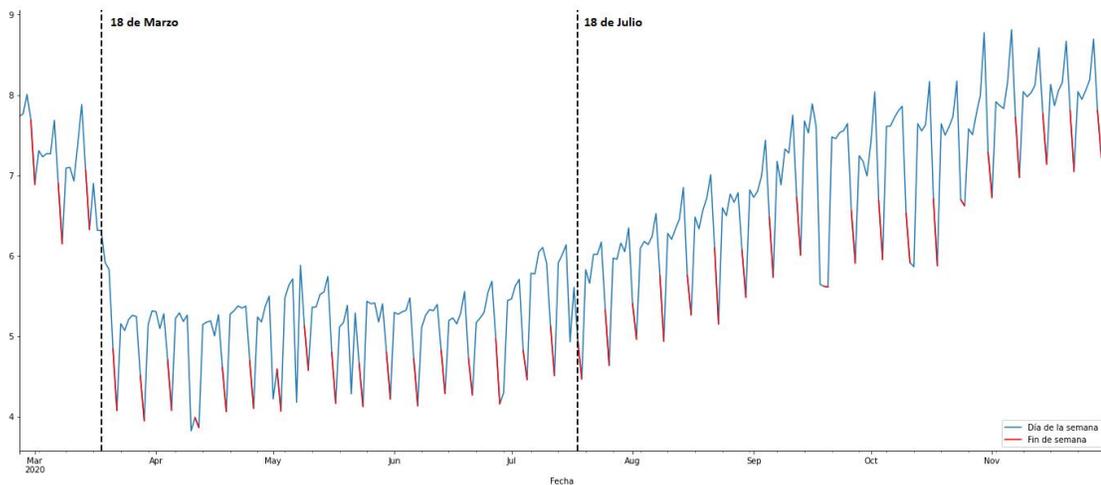


Imagen 4 - Índice de Movilidad promedio y división de períodos. Líneas azules representan variaciones entre días de semana. Líneas rojas representan variaciones en fines de semana (Sábados y Domingos).

Revisaremos la correlación de Pearson entre las variables numéricas que hemos seleccionado y revisaremos algunas de las más destacadas de manera gráfica, de tal manera de entender cómo se diferencia la dispersión entre todas las regiones.

También revisaremos otras visualizaciones que nos ayuden a entender las diferencias entre estos tres períodos seleccionados.

Luego de revisar el panorama general, nos concentraremos en las Fases 1 y 2, buscando si existen diferencias relevantes en la Movilidad. Para esto también nos apoyaremos de visualizaciones, junto con comparaciones entre ambas Fases y el período de mayor restricción (previo al Paso a Paso).

Finalizaremos nuestro análisis revisando la importancia de las variables seleccionadas según modelos de regresión. Los modelos de regresión seleccionados fueron Gradient Boosting, Random Forest y Decision Trees ya que son modelos que en general no se ven afectados por las relaciones lineales o no lineales de las features y por lo mismo menos condiciones deben cumplirse con nuestros datos. Estos modelos han sido usados para selección de características en varias oportunidades, lo que habla de lo confiable que pueden ser sus coeficientes de importancia. [35][36][37]

5. Resultados

5.1. Análisis exploratorio

Más allá de lo pudimos leer en artículos relacionados, lo primero que nos permite darnos cuenta de la relevancia de ciertas variables socioeconómicas y ambientales en la movilidad, es la correlación.

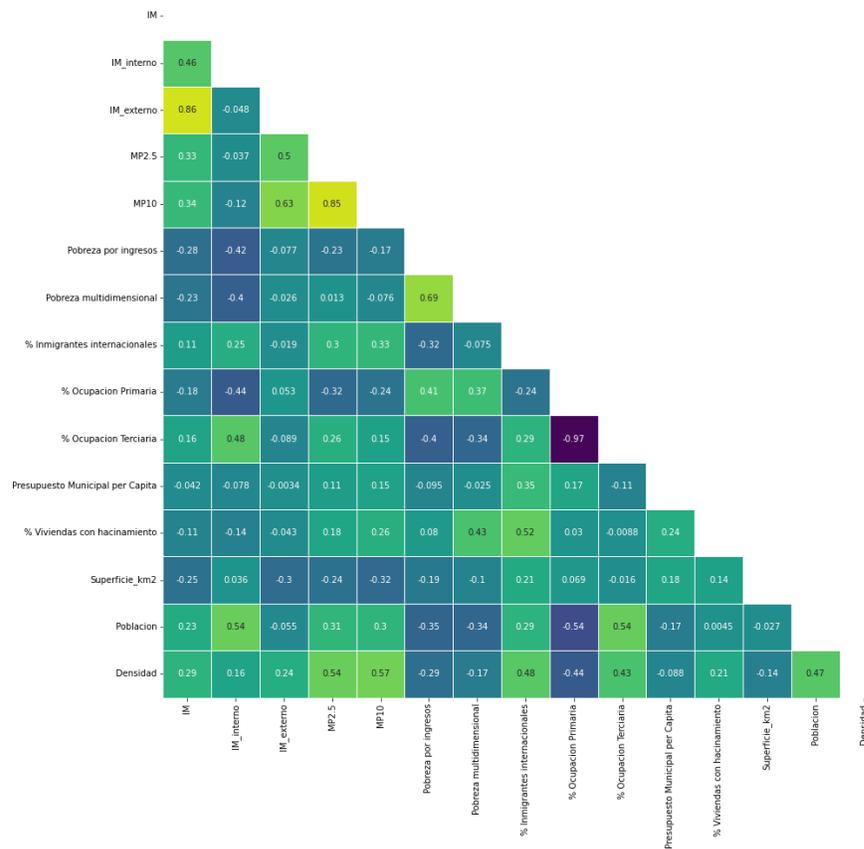


Imagen 5 - Correlación en período previo al 18 de Marzo (sin restricciones)

En este primer acercamiento a los datos ya podemos ver que a lo menos el Índice de Movilidad interno presenta altas correlaciones con variables socioeconómicas, tales como las tasas de pobreza, las tasas de ocupación, aunque también se ve una fuerte correlación con la población (que a su vez también tiene alta correlación con variables sociales).

Por su parte, el IM externo presenta más correlación con variables ambientales. Cabe destacar que este método excluye cualquier valor nulo para el cálculo de la correlación y que se están considerando datos para todo el país.

Pasando al período donde las restricciones fueron más severas, no vemos grandes cambios en la correlación general, aunque si se nota una cierta disminución en varias variables, especialmente en variables ambientales que presentaban alta correlación con el IM externo.

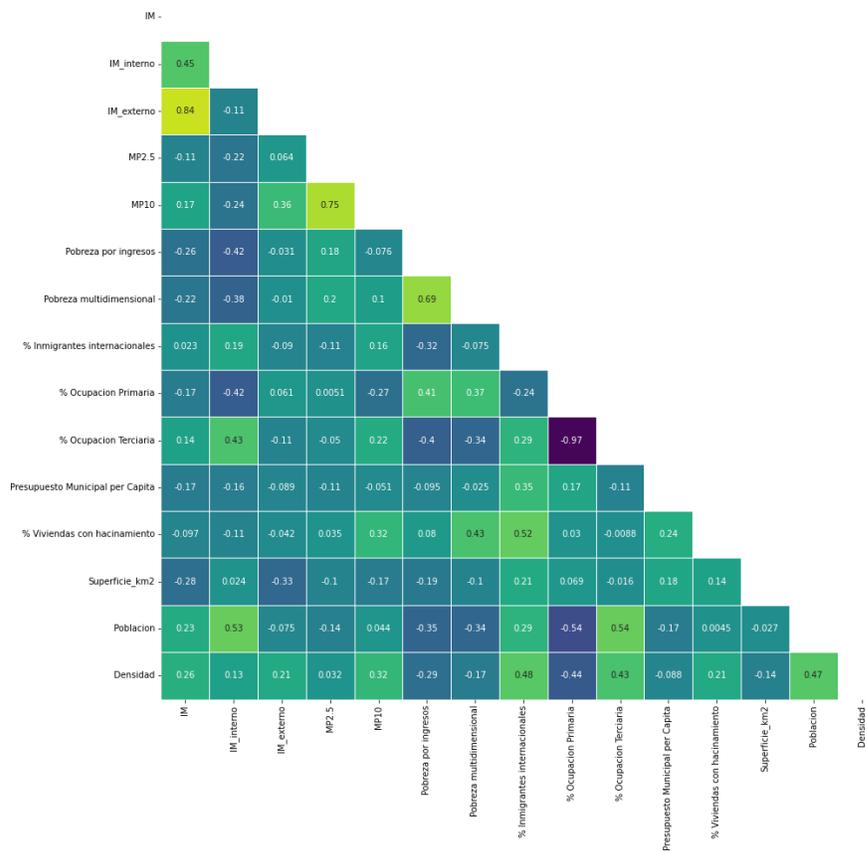


Imagen 6 - Correlación en período entre el 18 de Marzo y 18 de Julio (medidas más severas)

Finalmente, en el período donde se desarrolla el plan Paso a Paso, ciertos valores recuperan algo de fuerza, pero se mantiene la baja correlación en variables ambientales.

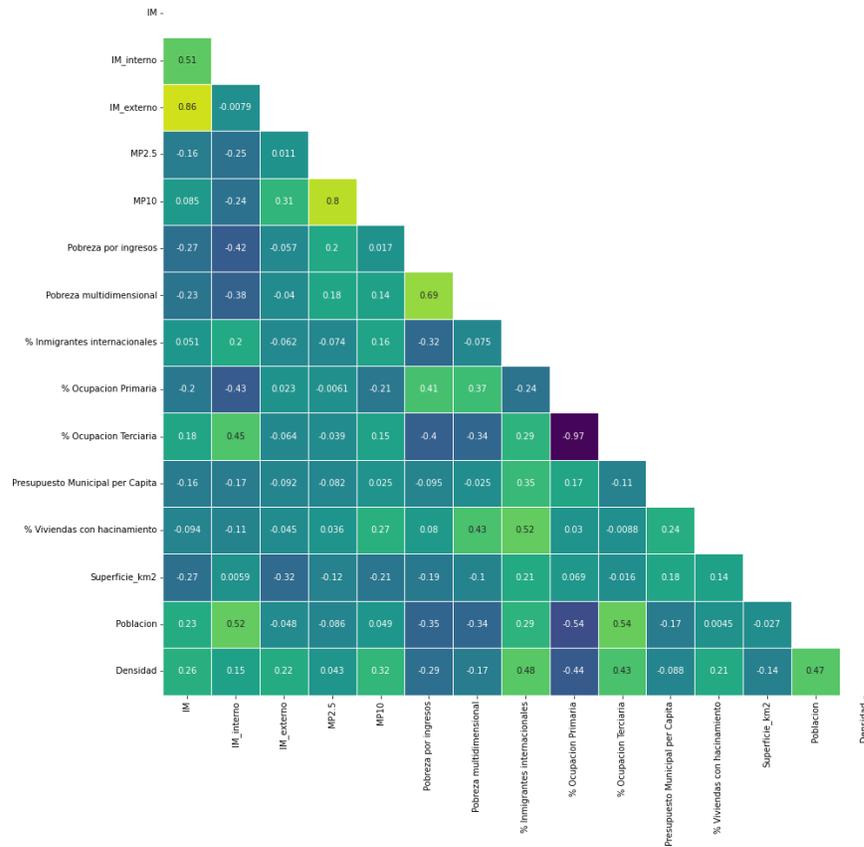


Imagen 7 - Correlación en período posterior al 18 de Julio (desarrollo del Plan Paso a Paso)

Revisar por región podría resultar incluso más interesante que la visión global, un caso a destacar es el de, por ejemplo, la región metropolitana, conocida por ser muy desigual entre sus comunas, donde el sector oriente es destacado por su diferencia con el centro y la periferia. En la Imagen 8 podemos observar cómo en esta región se notan las diferencias sociales solo mirando las diferentes correlaciones. Además, más abajo en el mapa vemos como la movilidad se destaca en el sector Oriente, reconocido por sus comunas con altos ingresos y donde se centra mucha de la fuerza laboral del área de servicios (actividad terciaria).

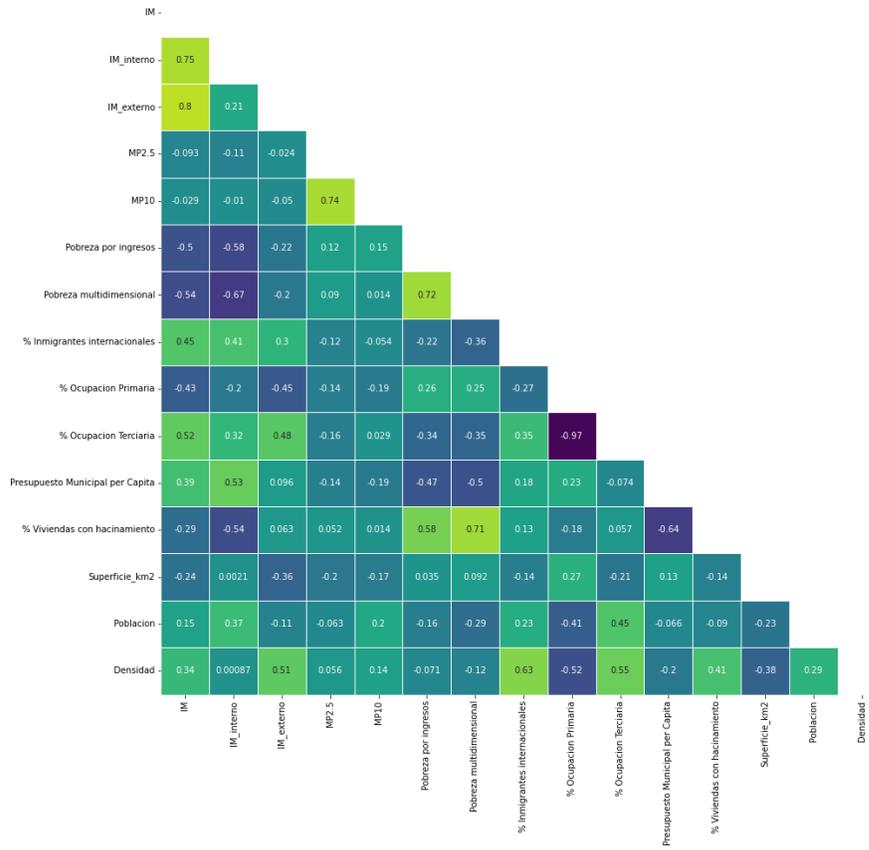


Imagen 8 - Correlación en la región metropolitana período antes del 18 de Marzo

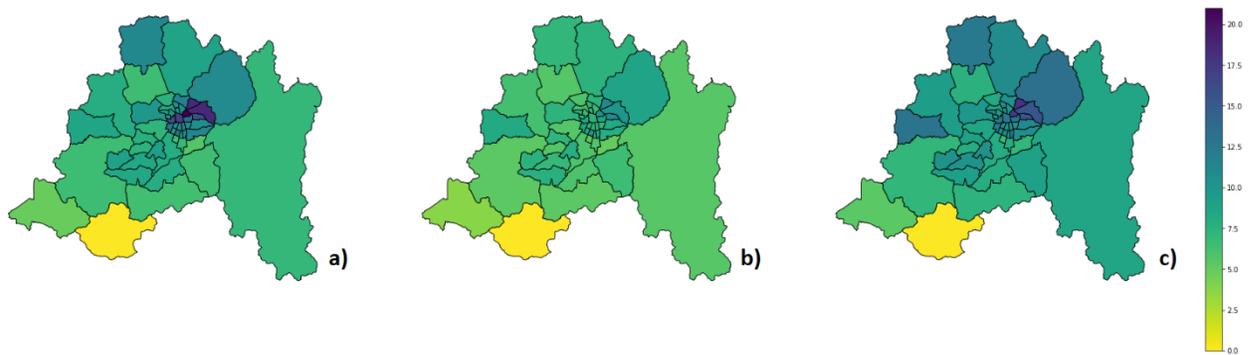


Imagen 9 - Índice de Movilidad en la región metropolitana en los tres períodos. a) Antes del 18 de Marzo. b) Entre el 18 de Marzo y el 18 de Julio. c) Post 18 de Julio

5.1.1 Viendo el panorama nacional

Chile es bastante conocido por su centralismo, donde la mayor población, densidad poblacional y recursos económicos se encuentran en la región metropolitana y sus alrededores. Regiones donde se ubican los principales poderes de gobierno, como lo son la casa de gobierno, el congreso y el senado. [38]

Tengamos en cuenta esto al mirar lo siguientes gráficos, ya que notaremos como muchas métricas importantes se destacan en la zona central.

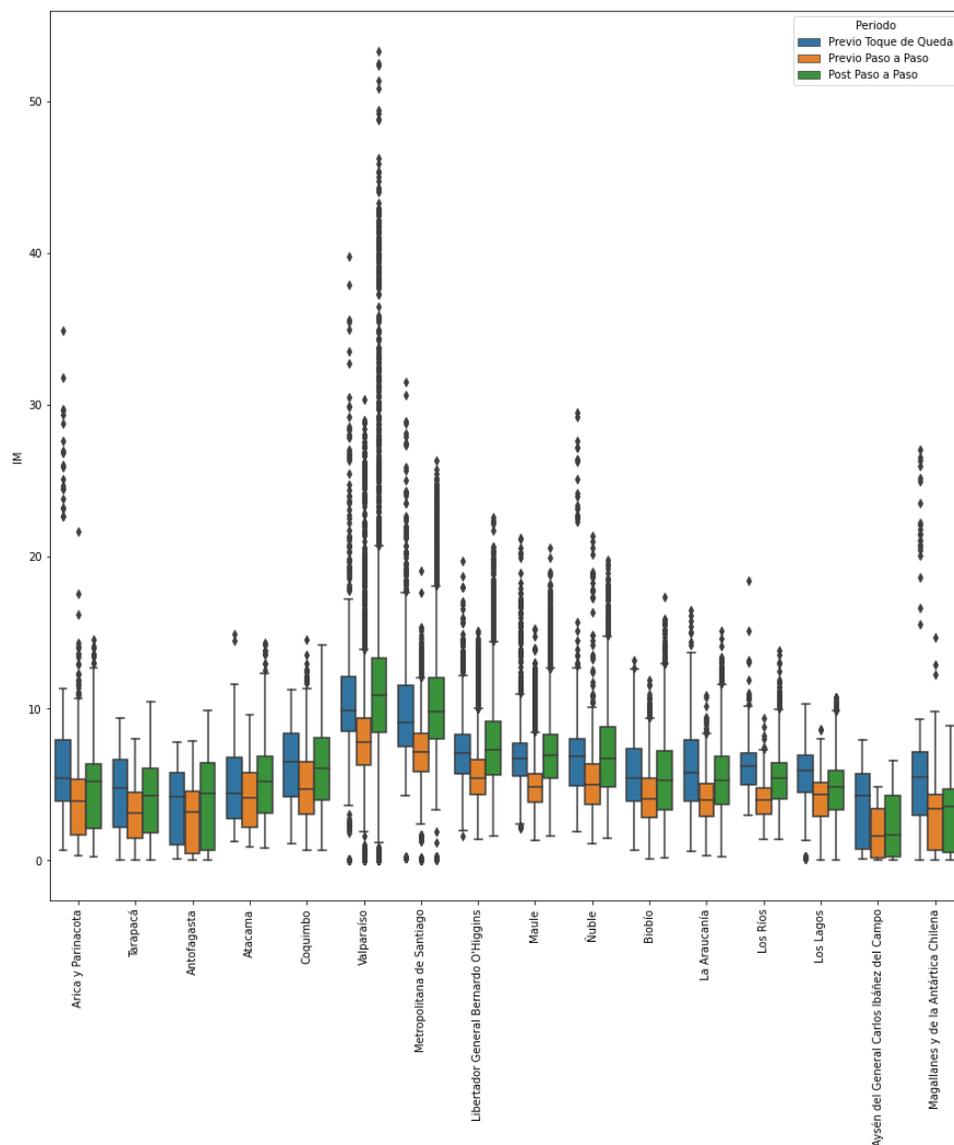


Imagen 10 - Índice de movilidad en las distintas regiones y dividido en los tres periodos. Se destaca el aumento de movilidad en la zona central, especialmente en las regiones de Valparaíso y Metropolitana.

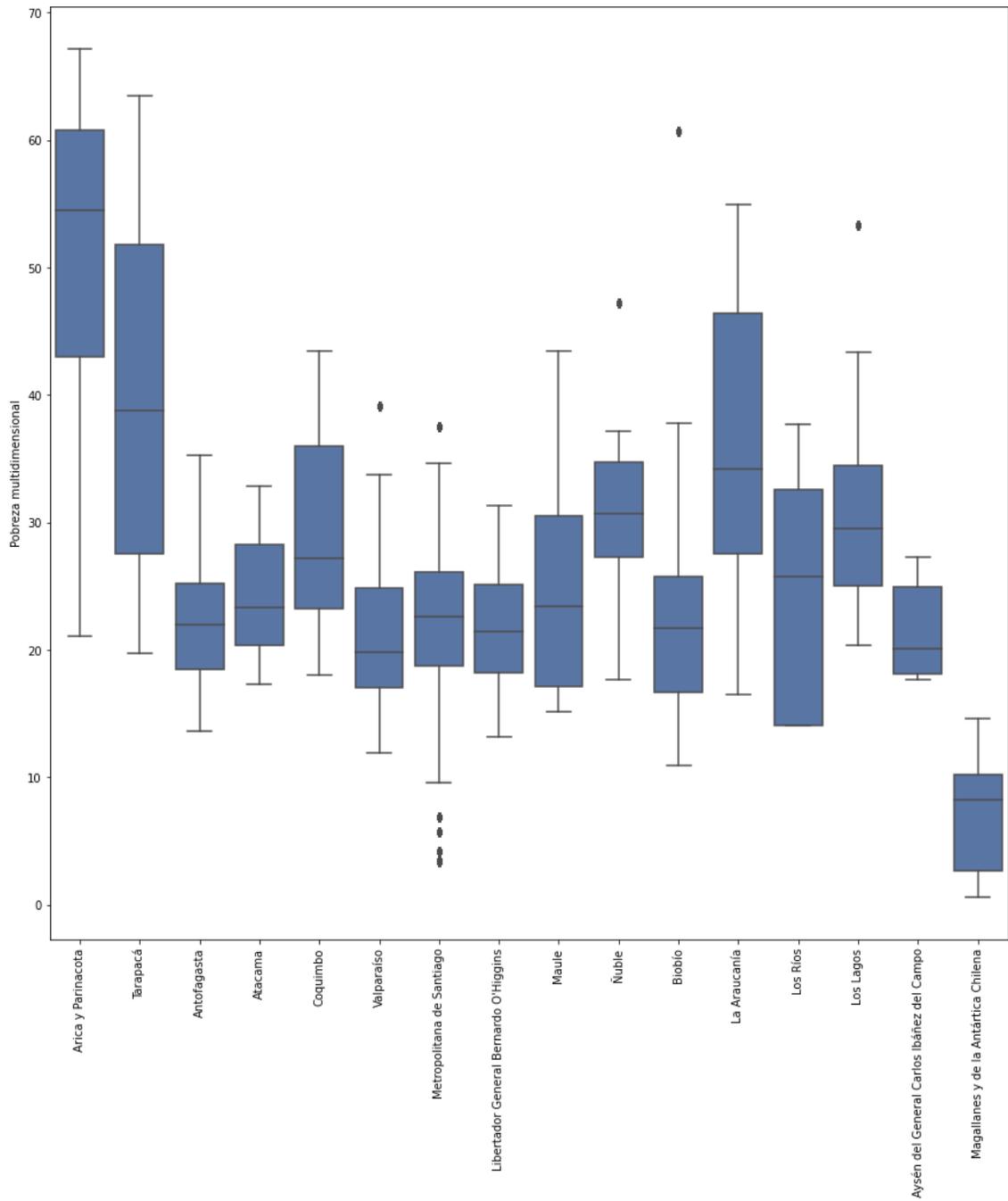


Imagen 11 - Tasa de pobreza multidimensional en las distintas regiones

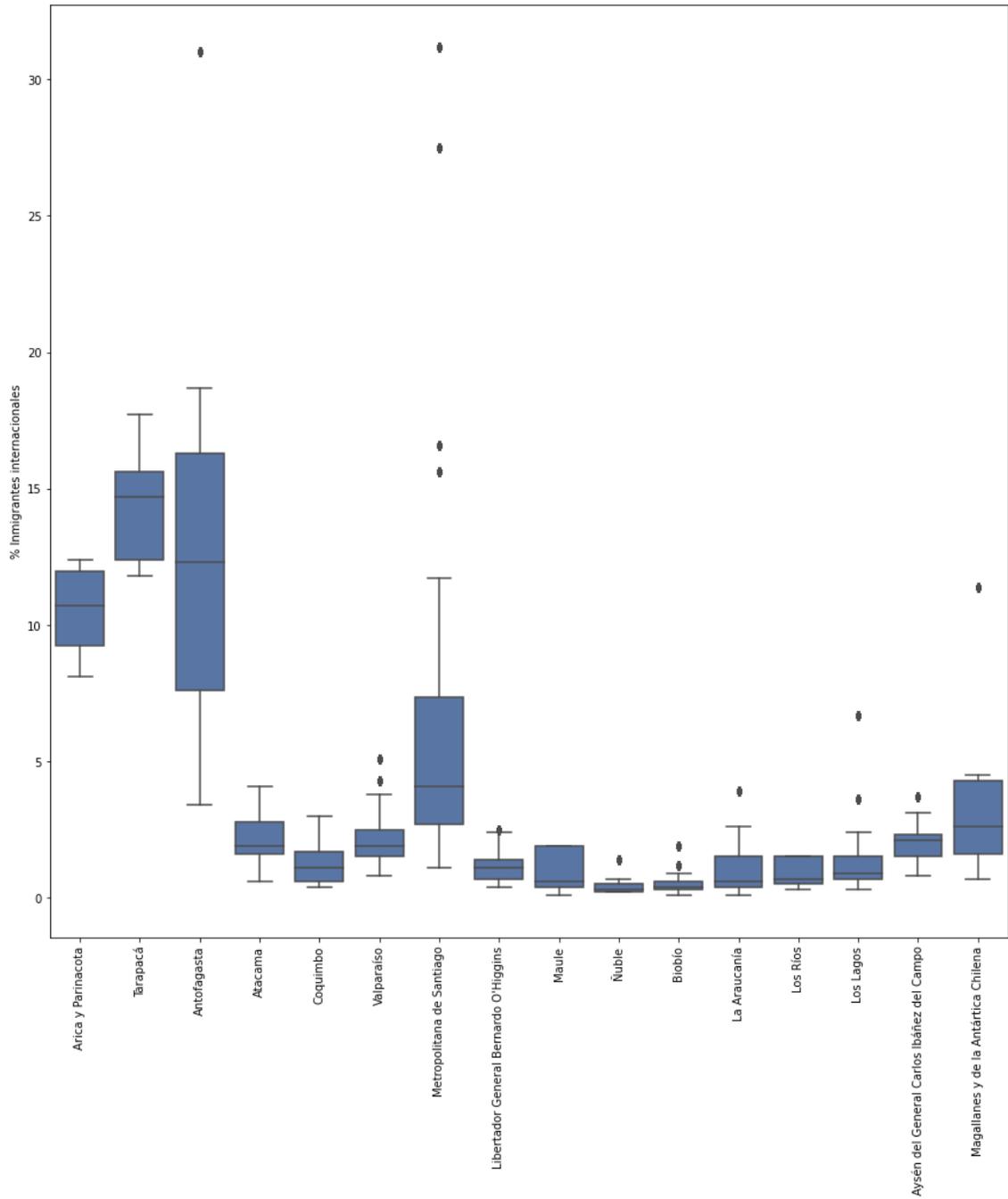


Imagen 12 - Porcentaje de inmigrantes internacionales en las distintas regiones. Se destaca la región metropolitana y la zona Norte. Ambas zonas con alta inmigración irregular.

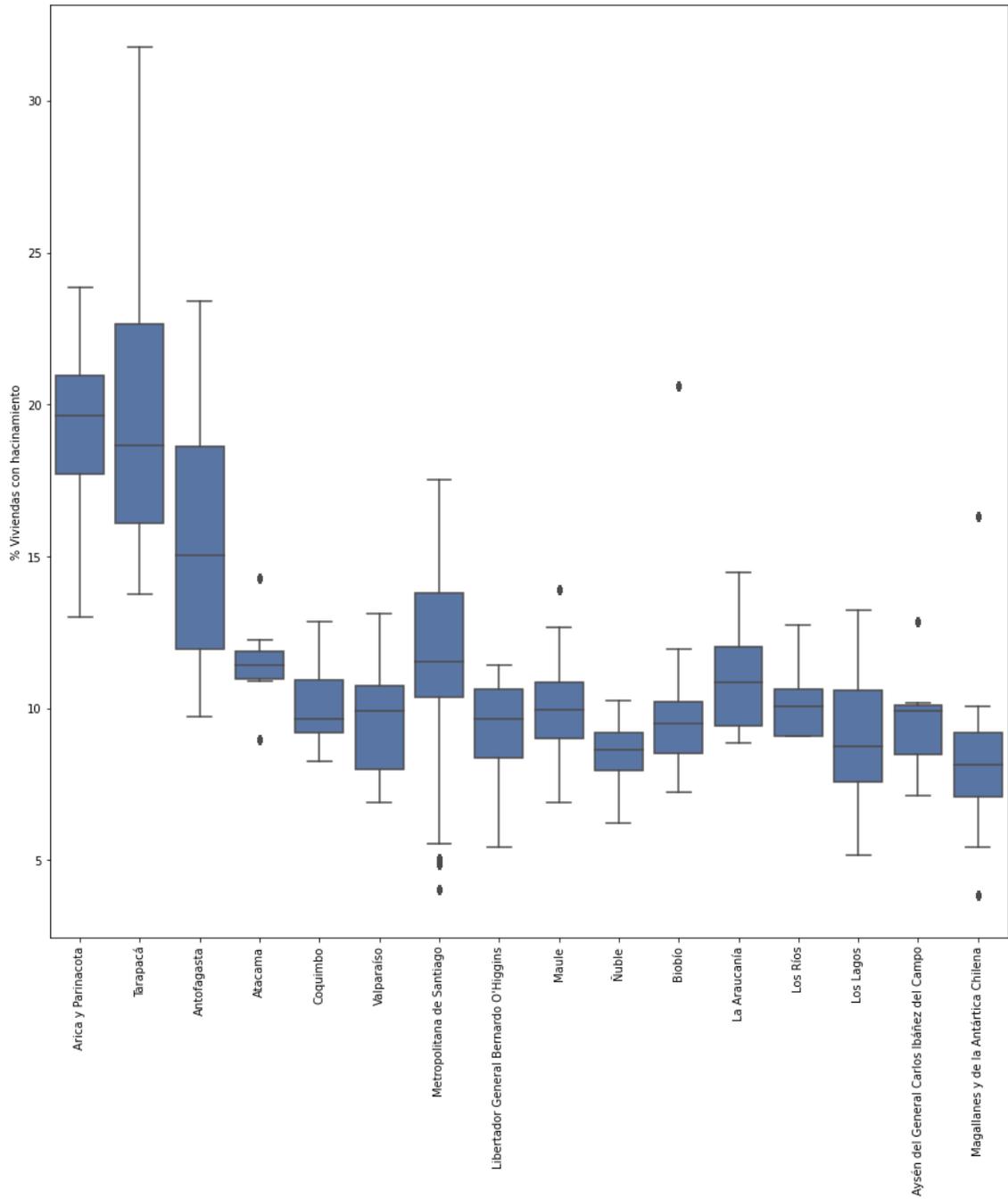


Imagen 13 - Porcentaje de viviendas con hacinamiento en las distintas regiones. Se destaca la región metropolitana y la zona Norte. Ambas zonas con alta inmigración irregular.

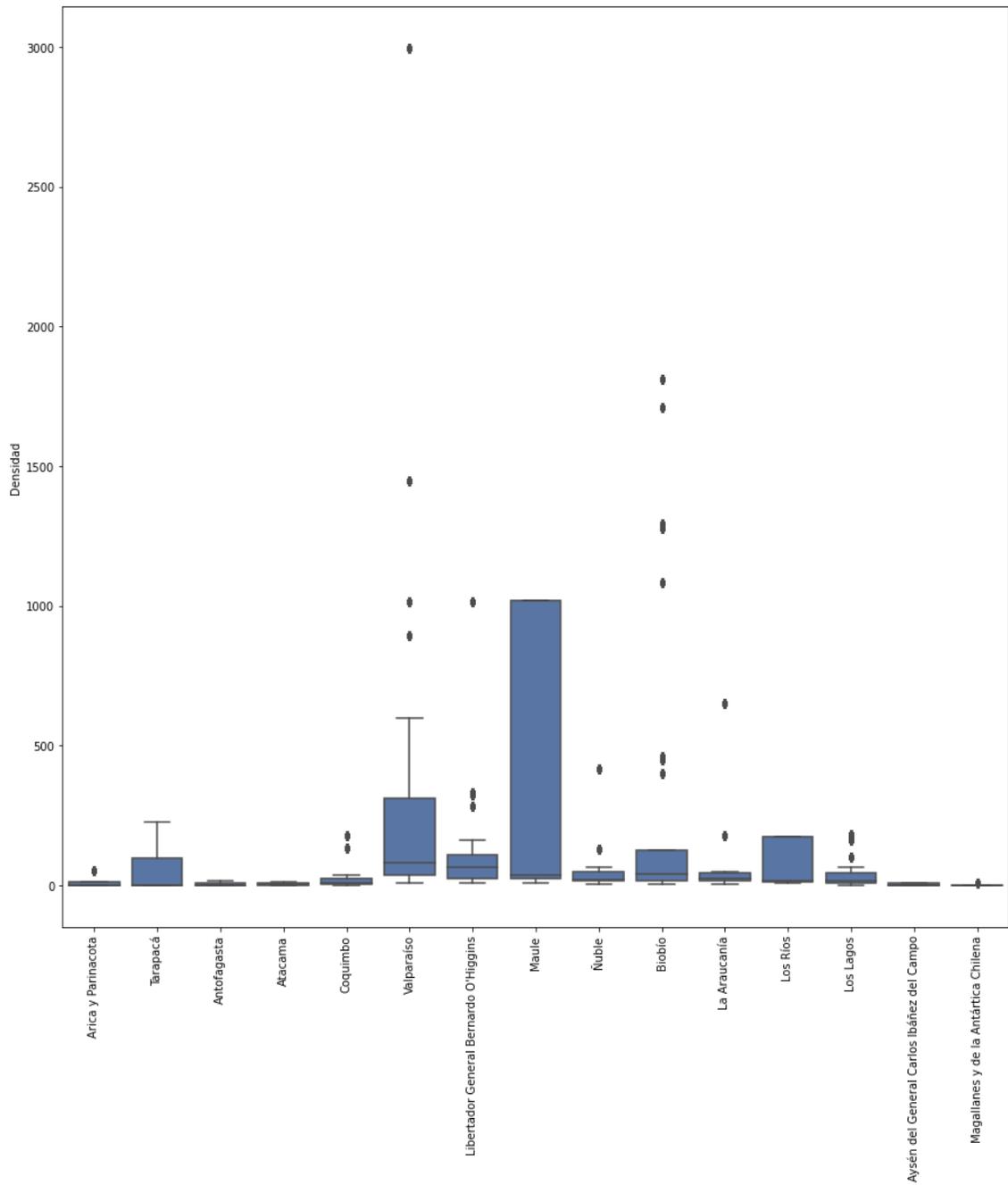


Imagen 14 - Densidad poblacional en las distintas regiones. Se omitió la región metropolitana porque deja pequeña a las demás regiones.

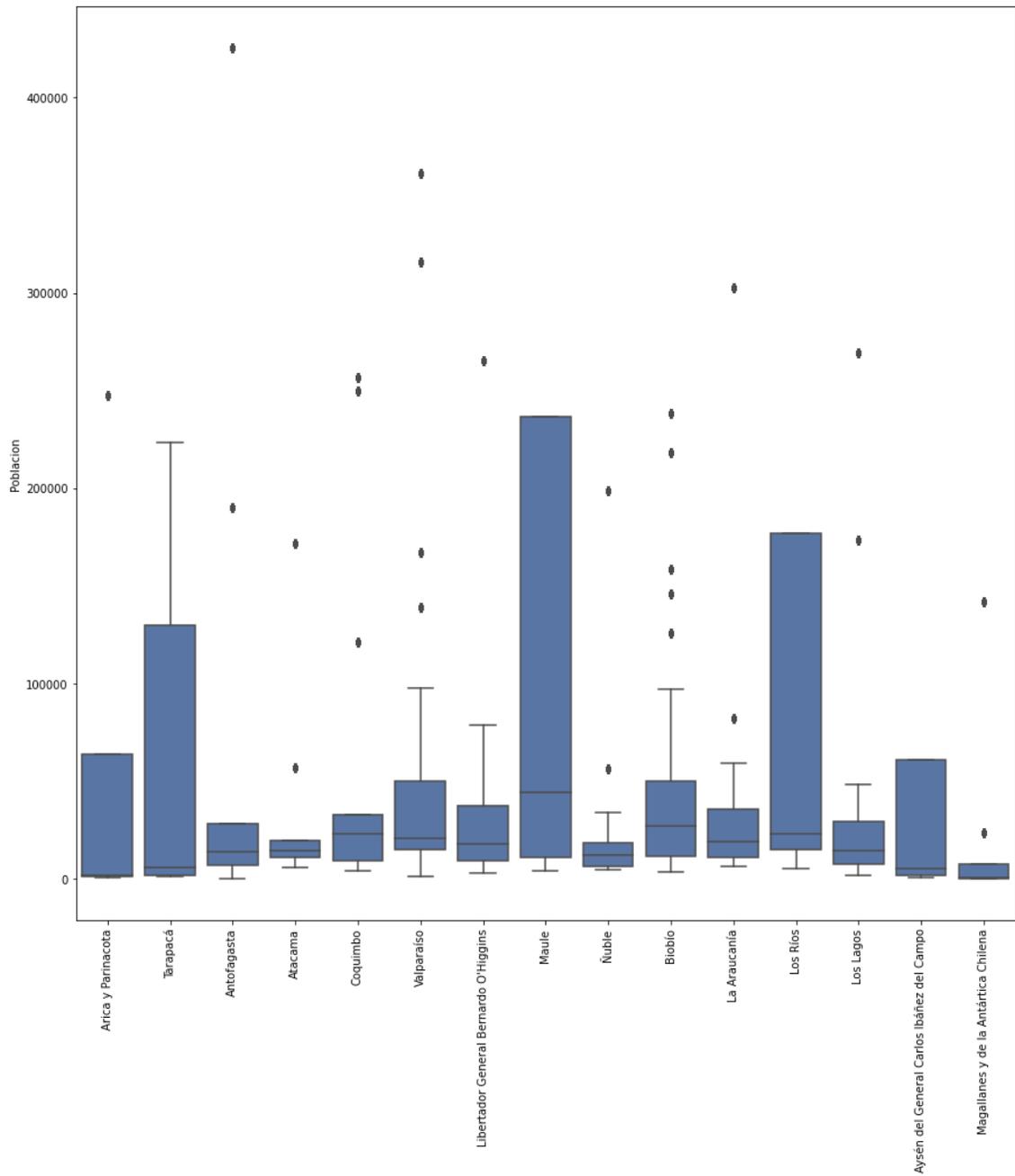


Imagen 15 – Población de las distintas regiones. Se omitió la región metropolitana porque deja pequeña a las demás regiones.

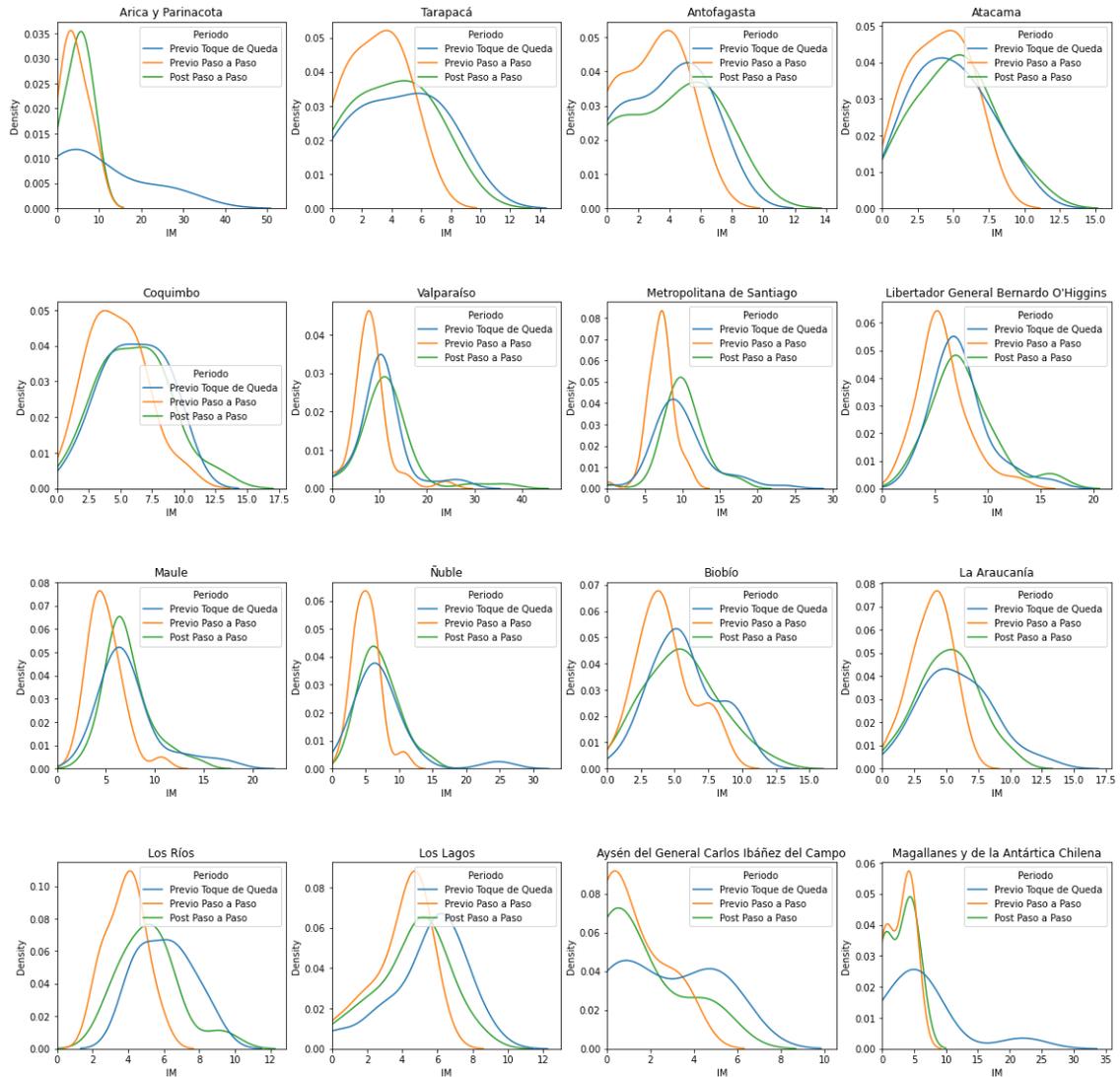


Imagen 16 - Variación Índice de movilidad en los distintos períodos por región. Se puede destacar como la línea naranja se mantiene hacia el lado izquierdo, lo cual condice con tener índices de movilidad más bajos durante el período más restrictivo. Además, la mayoría de los casos las otras dos líneas son muy similares, dando a entender que el plan Paso a Paso ayudó a volver a la normalidad.

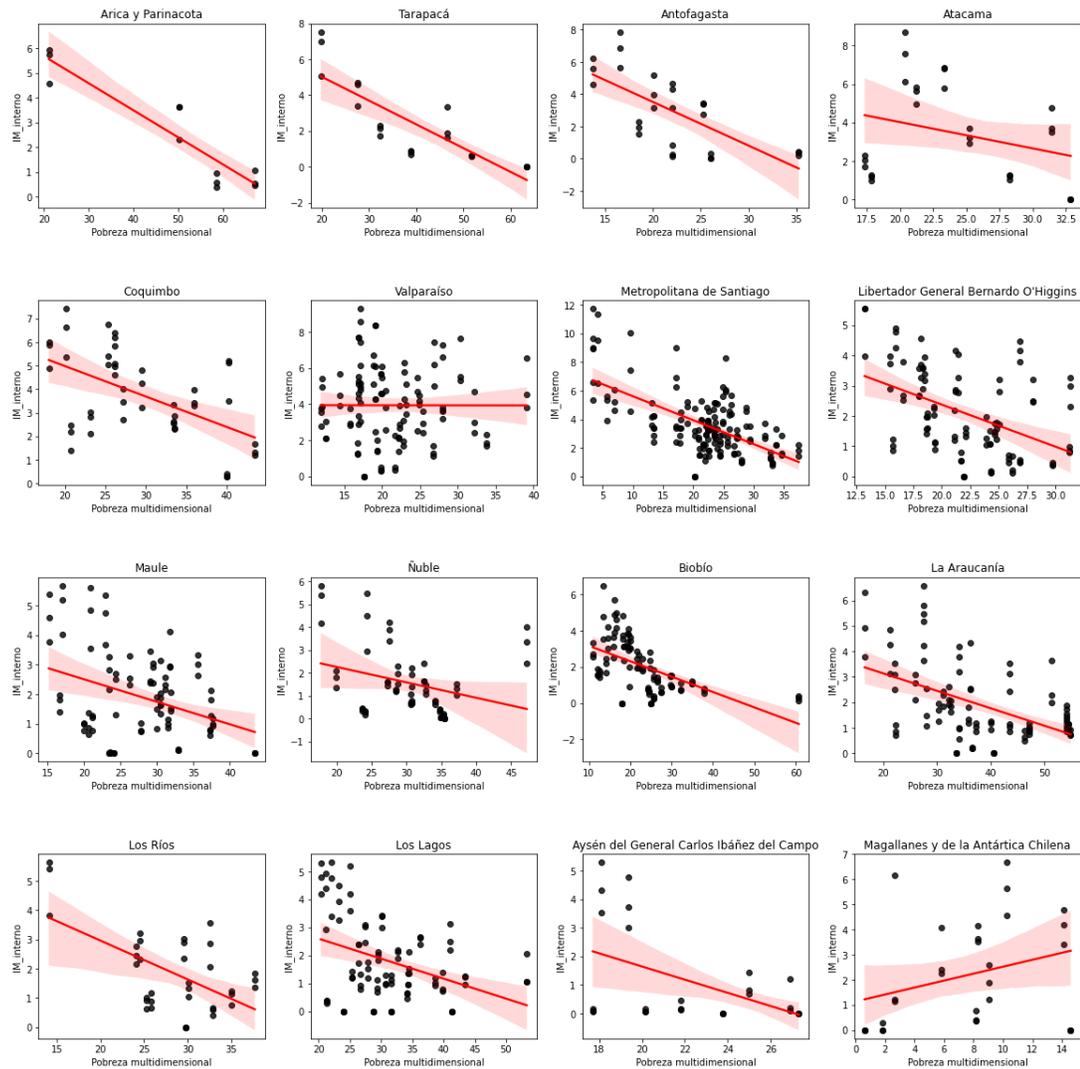


Imagen 17 - Correlación destacada, IM interno con tasa de pobreza. Se puede notar como en casi todas las regiones (excepto por Magallanes y Valparaíso) la correlación entre ambas variables es notable. Cada punto representa el promedio durante uno de los tres períodos del IM interno y la tasa de pobreza multidimensional de una comuna. Rango rojo representa el intervalo de confianza del 95%.

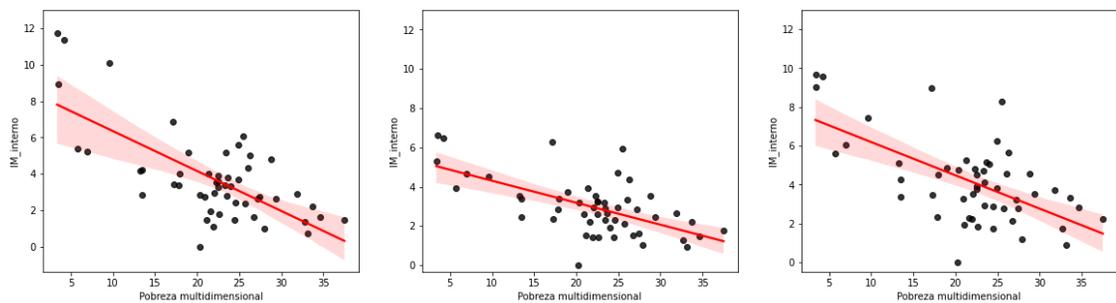


Imagen 18 - Variación en los tres períodos definidos para la región metropolitana.

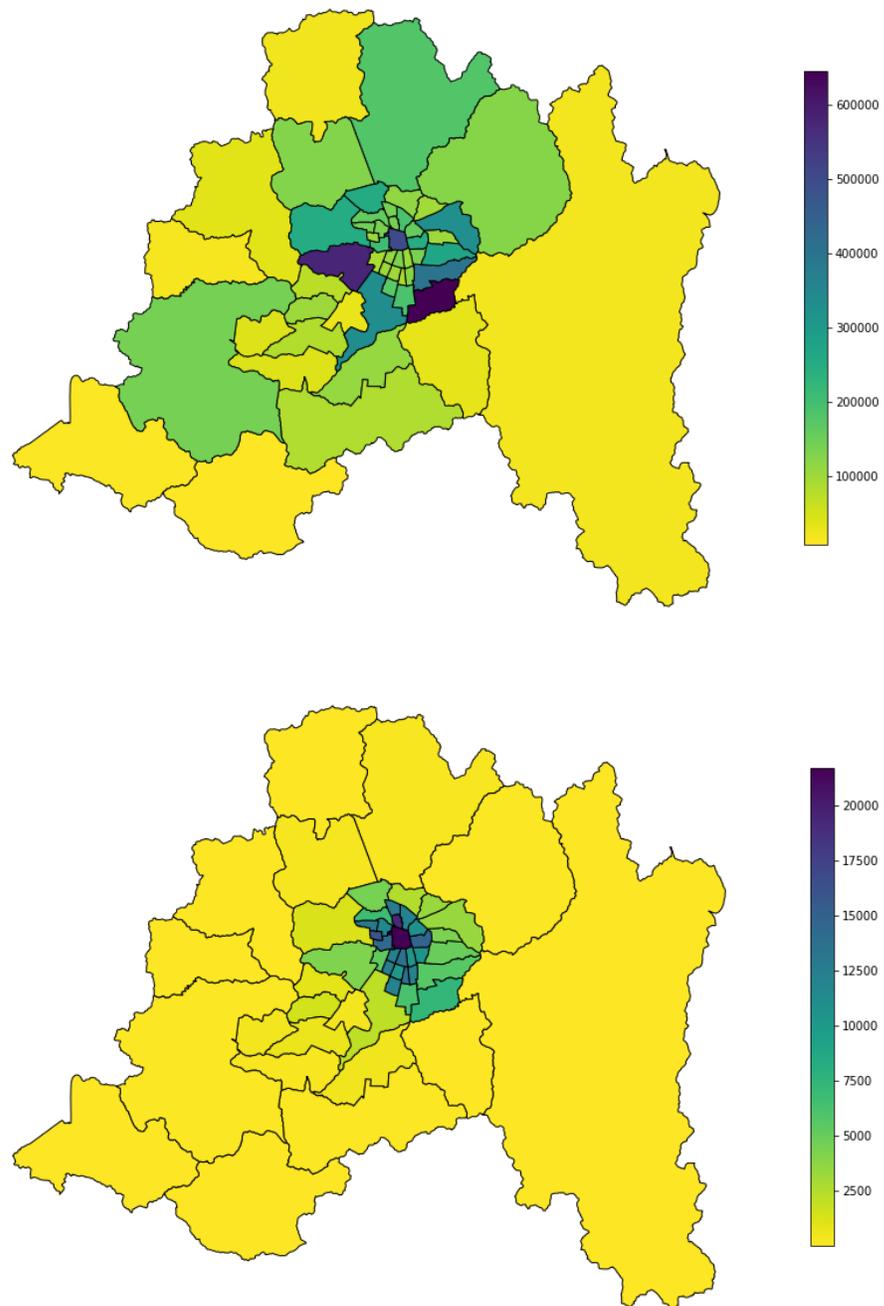


Imagen 19 - Arriba: Número de habitantes en la región metropolitana. Abajo: Densidad poblacional. Es destacable como incluso en la capital se nota la centralización, esto en general se explica por temas de tiempo y distancias de viaje, considerando que muchos puestos de trabajo se concentran en el centro y oriente.

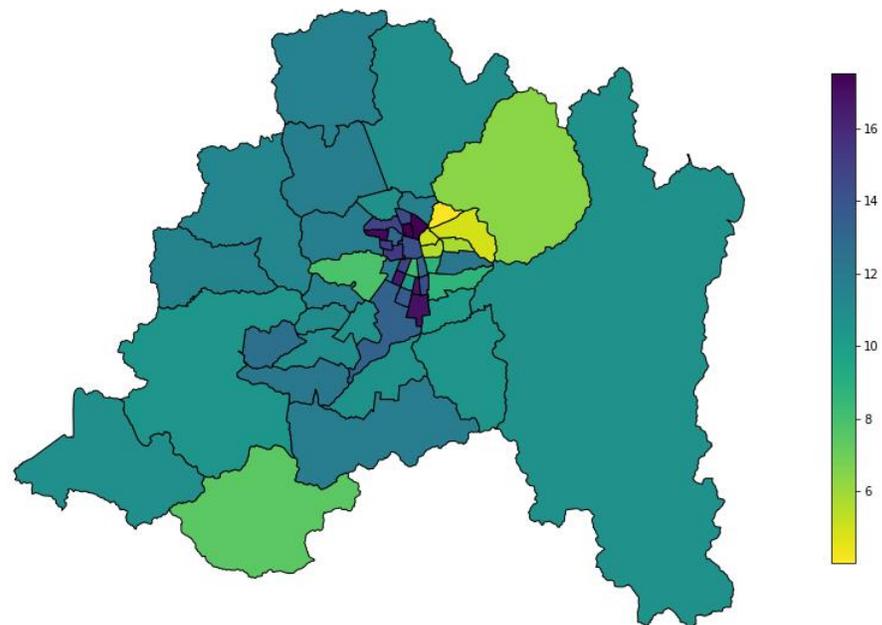
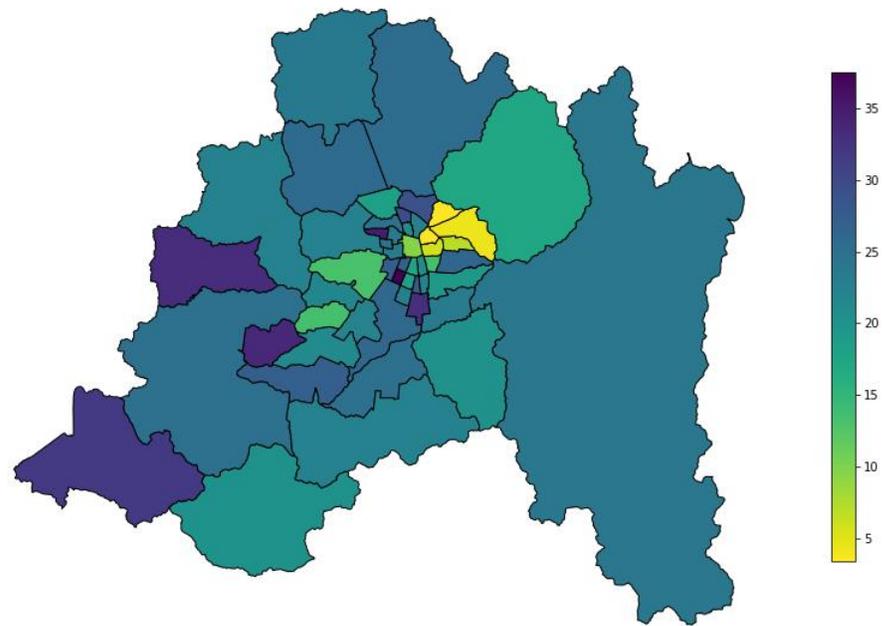


Imagen 20 - Arriba pobreza multidimensional y abajo porcentaje de viviendas con hacinamiento. En ambos casos más oscuro significa mayor tasa y por lo tanto mayor precariedad social.

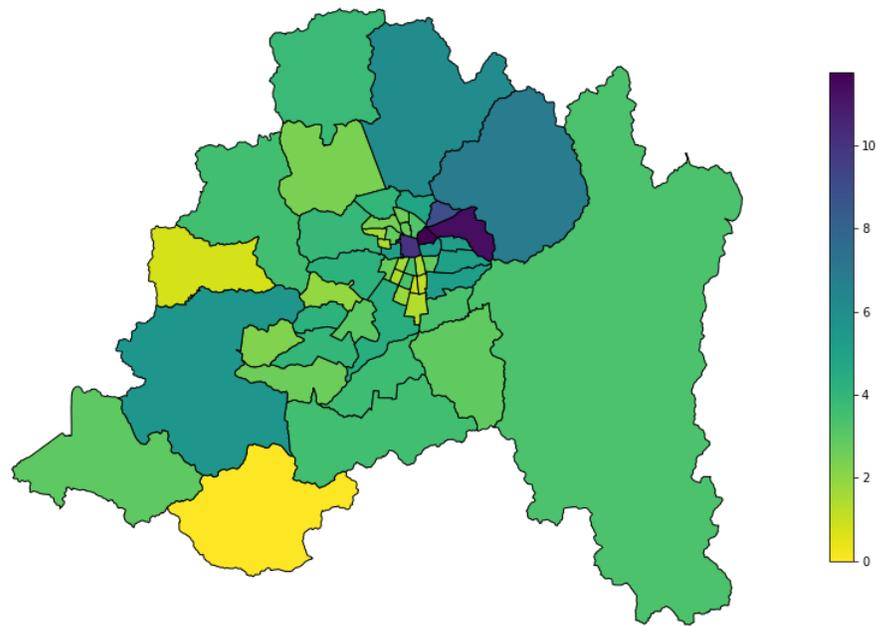


Imagen 21 - Índice de movilidad interna en la región metropolitana durante el período previo a las restricciones. Vemos como se destacan las comunas del sector oriente y Santiago en el centro.

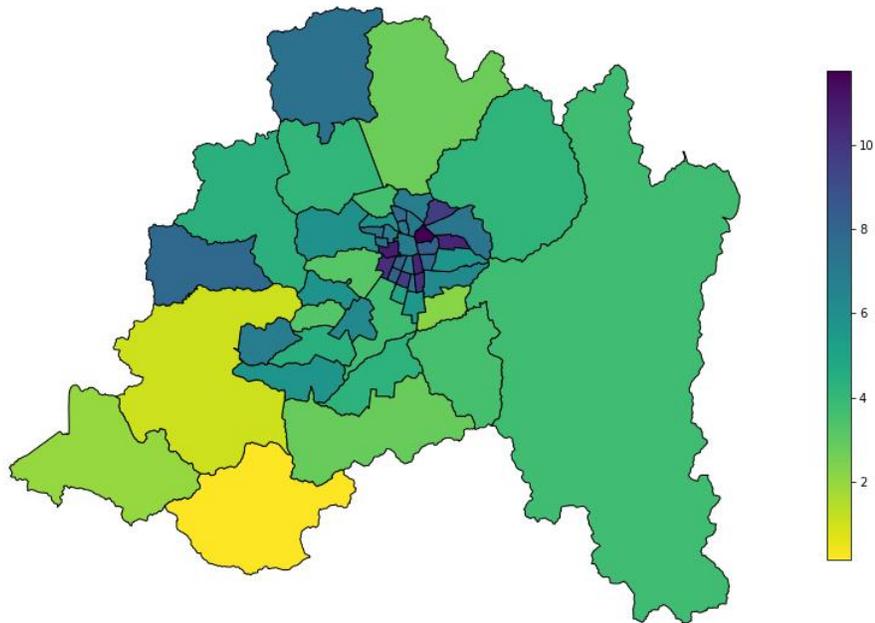


Imagen 22- Índice de movilidad externa en la región metropolitana durante el período previo a las restricciones. En este caso se destacan las comunas alrededor de Santiago centro y algunas comunas fronterizas.

5.2. Comparación Pasos 1 y 2 del plan Paso a Paso

En esta sección nos centraremos en la comparación entre las Fases 1 y 2 del Plan Paso a Paso. Buscando diferencias en la movilidad, especialmente durante los fines de semana, aunque no dejaremos fuera los días de la semana. Aprovecharemos también de comparar con el período de mayor restricción, ya que estos dos pasos buscaron ser lo más parecido a las severas restricciones impuestas durante las cuarentenas de Marzo hasta Julio.

Es importante mencionar o recordar que durante el Plan Paso a Paso, durante todas sus Fases, era posible obtener permisos para circular.^[40] Los cuales fueron falsificados por algunas personas e incluso vendidos.^{[39][41]} Por lo que, legal o ilegalmente, se podía circular más los meses posteriores a Julio del 2020 y por lo tanto deberíamos notar un aumento en el índice de movilidad.

Comencemos nuevamente con la correlación:

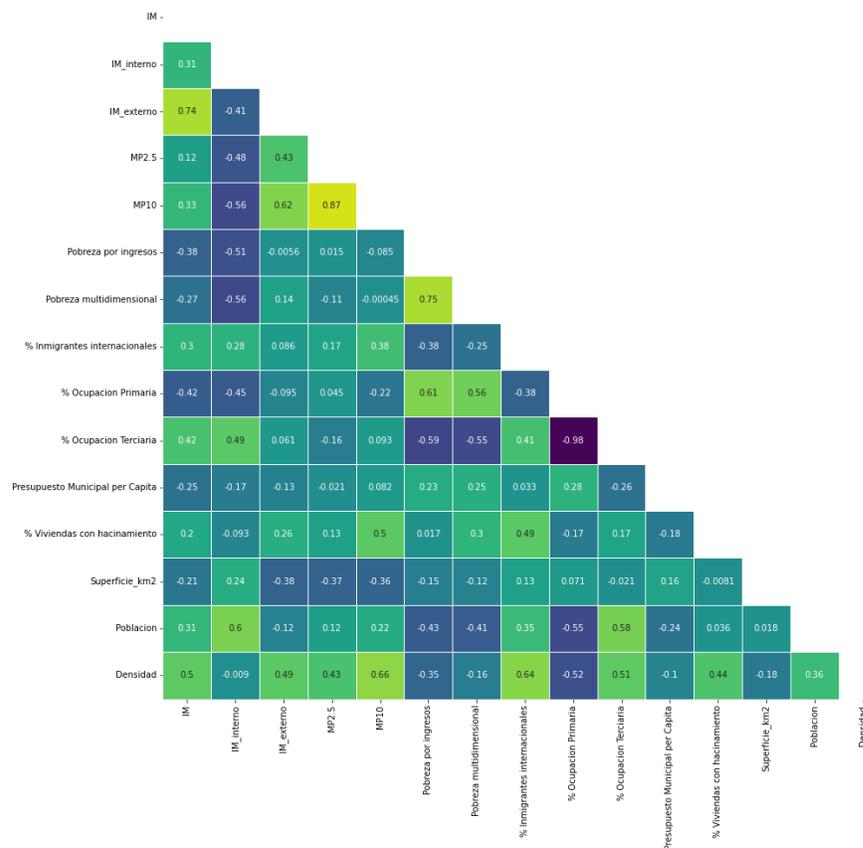


Imagen 23 - Correlación de variables numéricas en comunas durante el Paso 1

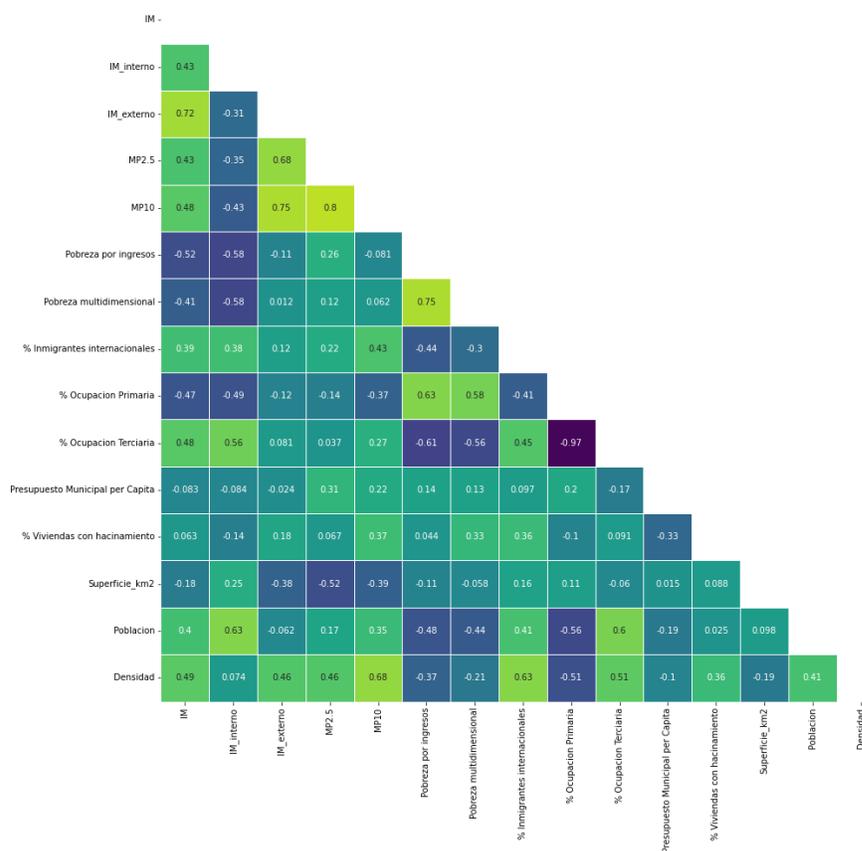


Imagen 24 - Correlación de variables numéricas en comunas durante el Paso 2

Al parecer, en las comunas en Paso 2, las variables socioeconómicas tuvieron una injerencia más fuerte sobre la movilidad, lo cual no es tan extraño si consideramos que durante el Paso 2 las comunas podían funcionar casi normalmente durante los días de semana y con restricciones durante los fines de semana. Siendo los días de semana donde más movilidad existe en general, ya que las personas deben movilizarse a sus lugares de trabajo, dentro o fuera de sus mismas comunas.

Por esto último también es probable que se noten más las diferencias sociales, ya que las personas se movilizan hacia o dentro de comunas con altos índices económicos o con altas tasas de ocupación terciaria.

Comparemos en términos numéricos la diferencia de movilidad entre estas tres distintas etapas:

	Pre Paso a Paso	Paso 1	Paso 2
IM	5,37	7,18 ^{+34%}	8,67 ^{+21%}
IM Interno	2,18	3,54 ^{+62%}	4,04 ^{+14%}
IM Externo	3,19	3,64 ^{+14%}	4,63 ^{+27%}

Tabla 1 - Índices de movilidad promedio considerando solo los días de semana. Variaciones porcentuales aproximadas, calculadas en función del valor en la columna de la izquierda. Pre Paso a Paso considera datos entre el 18 de Marzo y el 18 de Julio.

	Pre Paso a Paso	Paso 1	Paso 2
IM	4,51	5,90 ^{+31%}	6,46 ^{+9%}
IM Interno	1,89	2,98 ^{+58%}	3,04 ^{+2%}
IM Externo	2,61	2,93 ^{+12%}	3,42 ^{+17%}

Tabla 2 - Índices de movilidad promedio considerando solo los fines de semana. Variaciones porcentuales aproximadas, calculadas en función del valor en la columna de la izquierda. Pre Paso a Paso considera datos entre el 18 de Marzo y el 18 de Julio.

En la tabla 1 se pueden notar grandes diferencias, a decir verdad, esperaba que las diferencias entre el Paso 1 y el “Pre Paso a Paso” fueran más pequeñas, pero eso da muestra de las libertades que se comenzaron a dar posterior a Julio de 2020.

Otro punto importante por destacar es la pequeña diferencia entre el Paso 1 y el Paso 2 durante los fines de semana, pero no debemos olvidar que son promedios generales, veamos ahora como se distribuye esta diferencia en las regiones de manera gráfica.

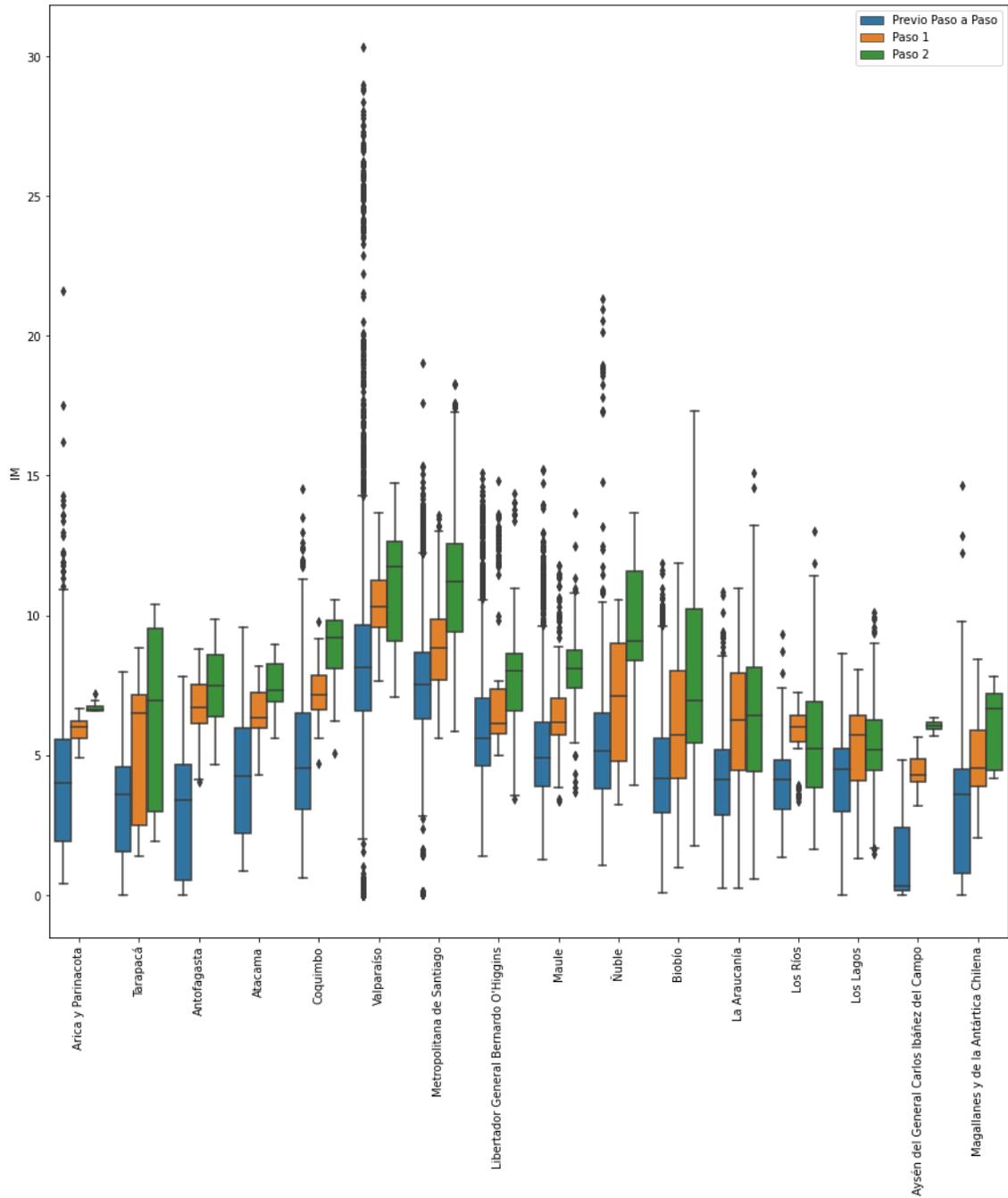


Imagen 25 – Diferencia en Índice de movilidad promedio en cada una de las regiones en días de semana. Se puede notar en muchas regiones como se cumple la diferencia vista en la Tabla 1, es decir una gran diferencia entre el período previo al Paso a Paso con el Paso 1 y una diferencia más pequeña entre el Paso 1 y Paso 2. Se sigue viendo como los mayores índices de movilidad se encuentran en la zona central.

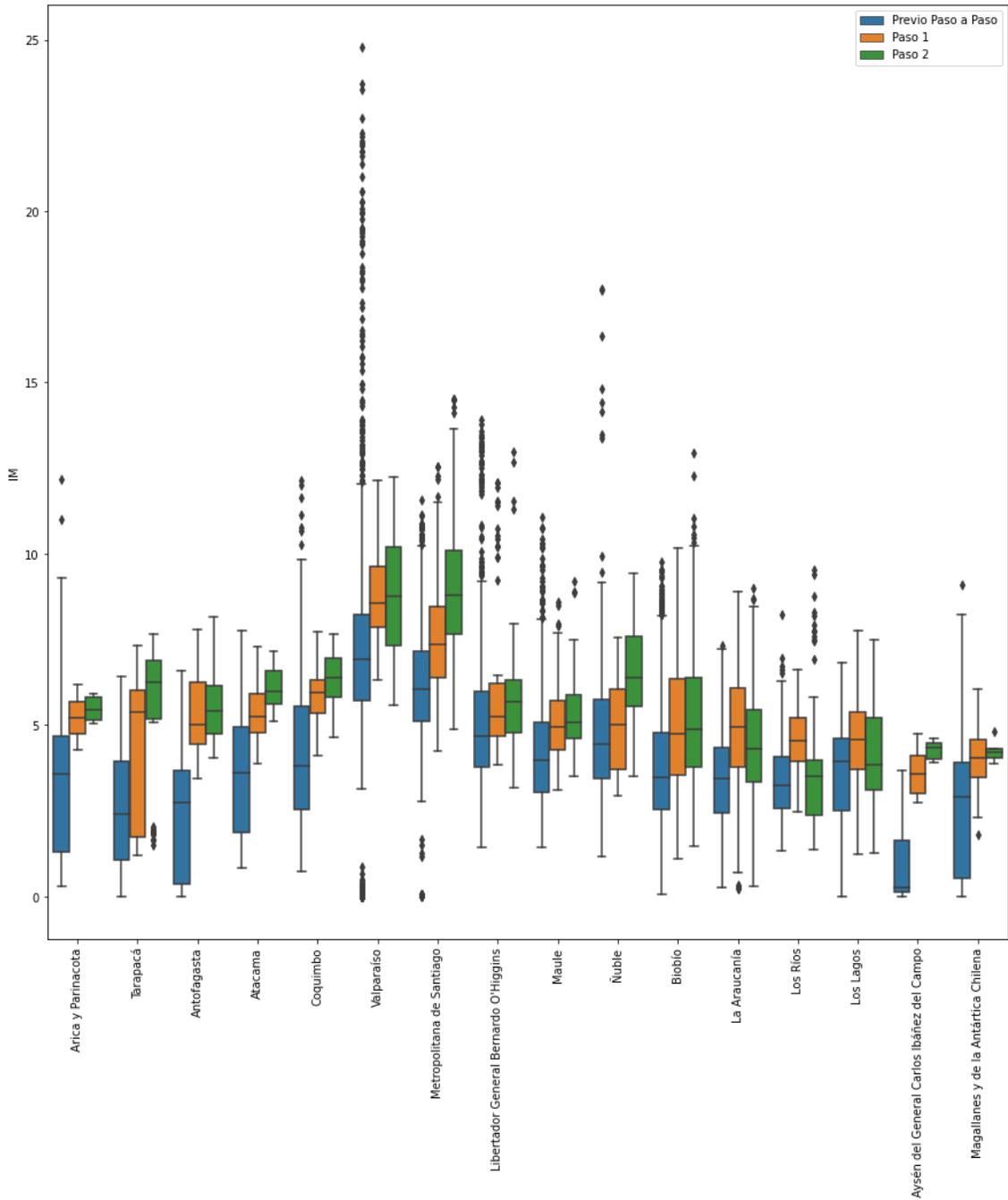


Imagen 26 - Diferencia en Índice de movilidad promedio en cada una de las regiones en fines de semana. La historia en esta imagen es similar a lo que vimos antes en la Imagen 25, alta movilidad en el centro y diferencias concordantes con lo visto en la Tabla 2. Llama la atención ver Regiones al sur de la capital de Chile con índices de movilidad más bajos en Paso 2 en comparación al Paso 1; como La Araucanía, Los Ríos y Los Lagos.

También podemos observar cómo se comportó el Índice de movilidad durante los últimos meses de los cuales tenemos registro. Es importante entender que la cantidad de comunas en los distintos pasos también fue cambiando a medida que pasó el tiempo, junto con a que región pertenecen, debemos tenerlo en cuenta al analizar los promedios de movilidad graficados en la Imagen 27.

Si obviamos las diferencias de comunas que componen cada día para cada Paso, seguimos observando una diferencia que se condice con lo anteriormente descrito; diferencias más grandes durante la semana y no tan grandes durante los fines de semana.

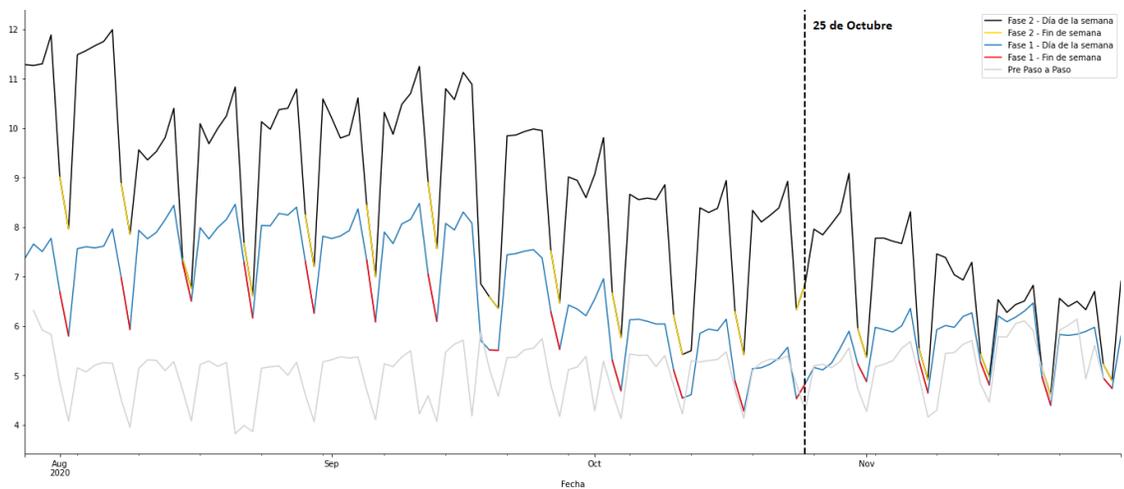


Imagen 27 - Índices de movilidad promedio en los últimos meses de 2020. Se puede observar que durante el Domingo 25 de Octubre aumentó la movilidad promedio, lo cual coincide con el plebiscito para Aprobar o Rechazar la propuesta de constitución Chilena, donde existió una participación histórica ^[42] Línea gris representa el período previo al Paso a Paso, se aplicó un offset para poder mostrarlo en el gráfico y que fuera comparable la diferencia con los dos Pasos.

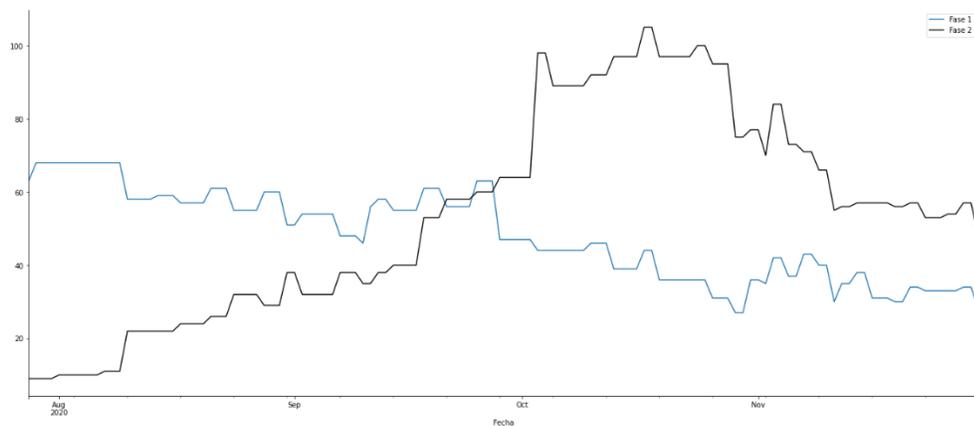


Imagen 28 - Cantidad de comunas en los Pasos 1 y 2 durante los últimos meses de 2020

Revisitemos algunas correlaciones destacables para ver gráficamente como se diferencian entre estos tres períodos.

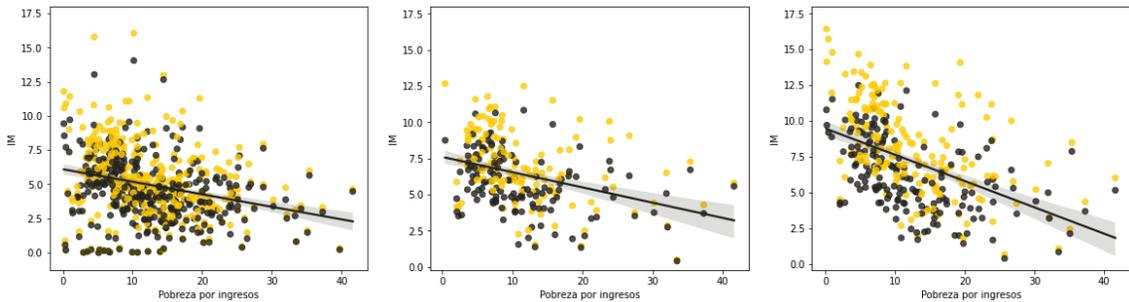


Imagen 29 - Relación entre Pobreza por ingresos e Índice de Movilidad. Los puntos negros indican promedio de movilidad en fines de semana, puntos amarillos indican promedio de movilidad en días de semana y línea negra representa la regresión ajustada a todos los datos, se muestra también intervalo de confianza al 95% de la regresión. Gráfico de la izquierda contiene registros entre el 18 de Marzo y el 18 de Julio; gráfico del medio contiene registros de Paso 1 y finalmente el gráfico de la derecha contiene registros del Paso 2. Se puede notar como en Fase 2 la movilidad aumenta en lugares con tasas de pobreza más bajas, especialmente en días de semana.

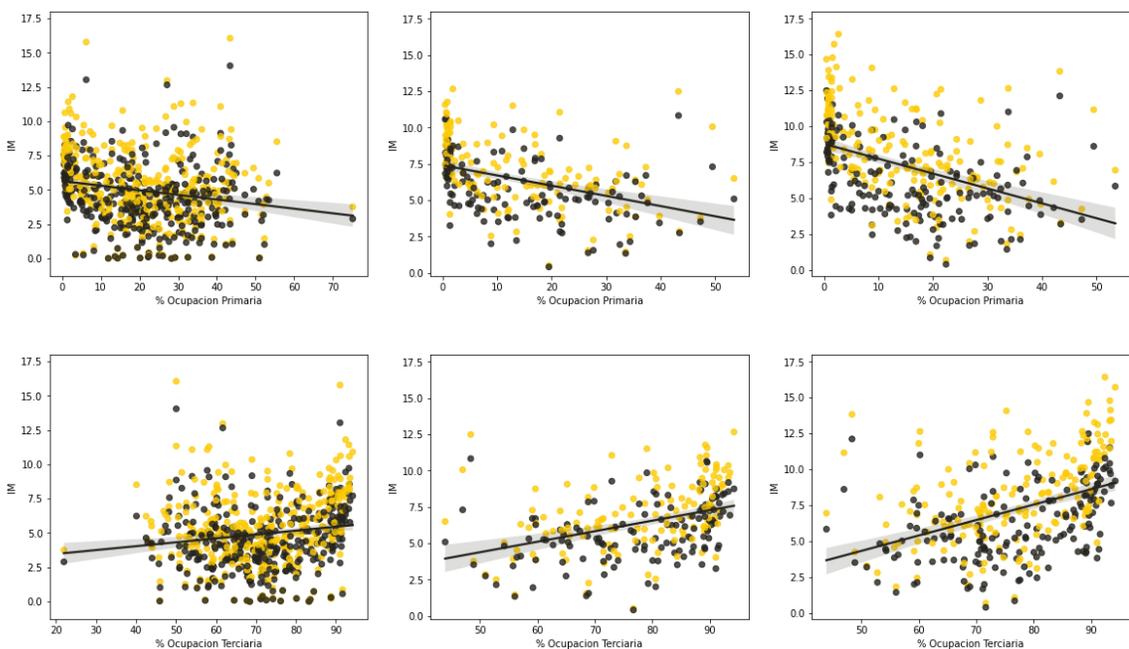


Imagen 30 - Relación entre porcentaje de Ocupación Primaria y Terciaria con el Índice de Movilidad. Misma representación de colores y orden que en la Imagen 29. Se puede observar cómo son casi un espejo un área de ocupación en comparación a la otra, algo esperable teniendo en cuenta la correlación casi 100% negativa entre ambas variables.

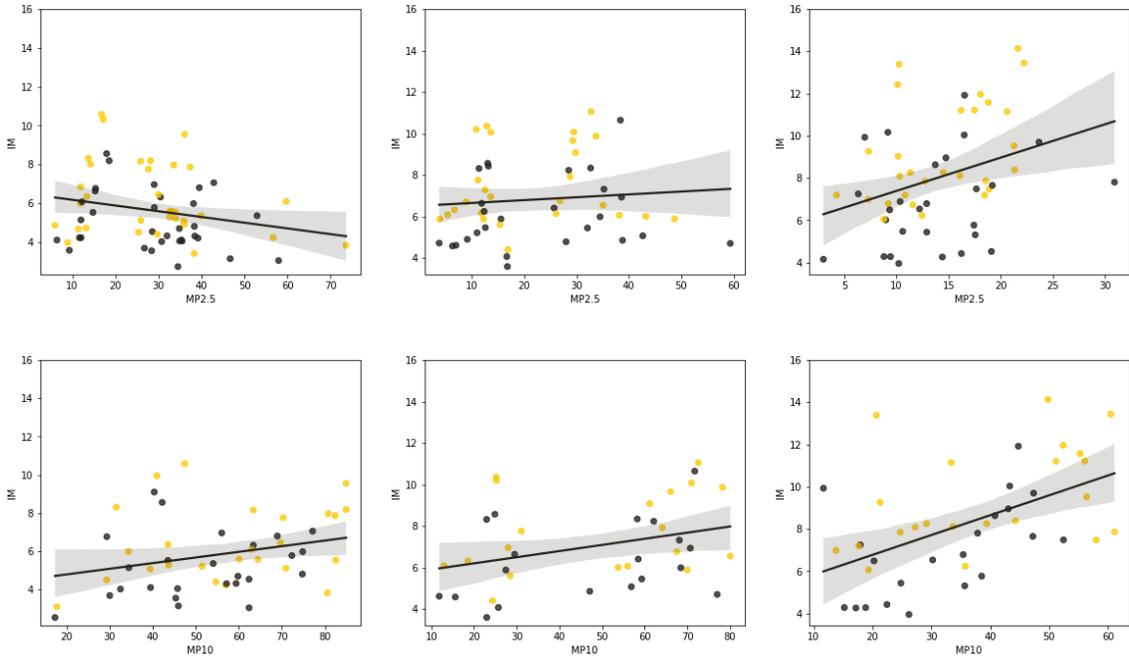


Imagen 31 - Relación de factores ambientales con el Índice de movilidad. Misma representación de colores y orden que en la Imagen 29. En general observamos un comportamiento esperable, considerando que las dos variables medioambientales están altamente correlacionadas. Sorprende lo graficado en los cuadros de más a la izquierda, los cuales representan el período entre el 18 de Marzo y el 18 de Julio, y aunque en ninguno de los dos casos la correlación contra el Índice de movilidad es muy alta (esto coincide con lo visto anteriormente en la Imagen 6) se nota una diferencia en la dispersión de ambos gráficos, estando el MP2.5 más agrupado hacia la izquierda mientras que para el MP10 no existe un grupo tan definido.

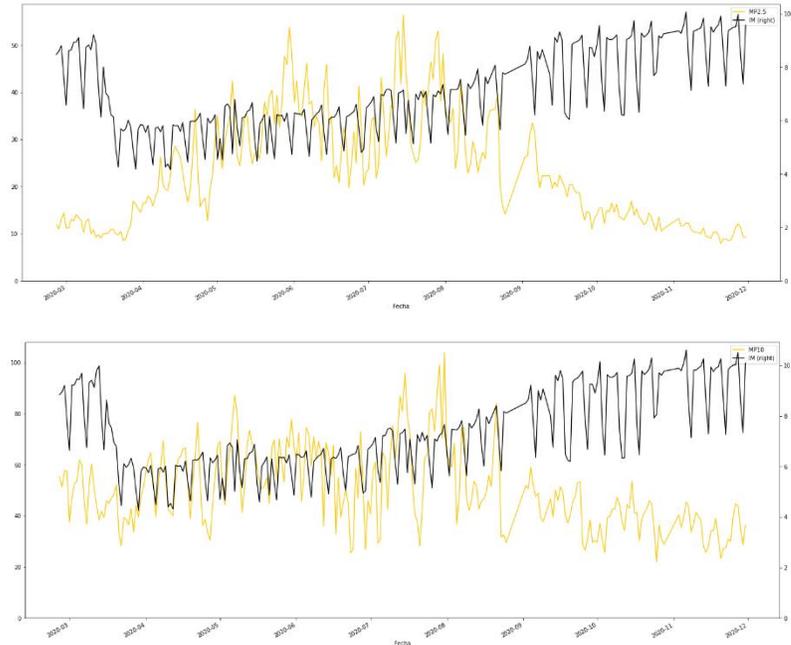


Imagen 32 - Tendencia (valores absolutos) del Índice de movilidad y las variables de calidad del aire. Se ve un aumento claro de los contaminantes en el período invernal. Solo se incluyen las comunas con mediciones de contaminantes válidas.

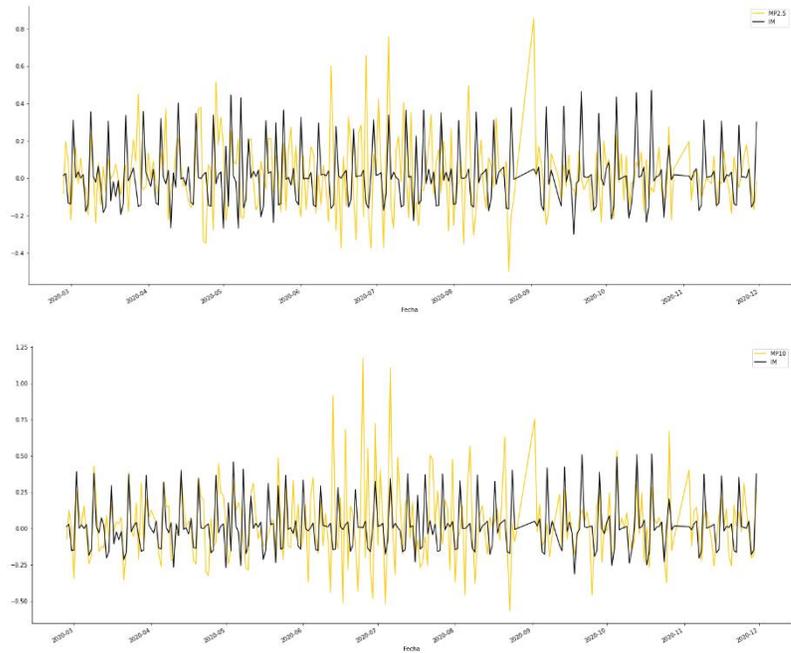


Imagen 33 - Diferencia porcentual (valores entre 0 y 1) del promedio de las variables ambientales y el promedio del Índice de movilidad. La diferencia es calculada en cada variable por separado, teniendo en cuenta el valor del día anterior. Solo se incluyen las comunas con mediciones de contaminantes válidas.

Por último, ya que lo hicimos antes también y para confirmar si se mantiene la tendencia hacia “más movilidad en Paso 1 que en Pre Paso a Paso” y “más movilidad en Paso 2 que en Paso 1”, revisemos como se comportó la región metropolitana a través de mapas coropléticos.

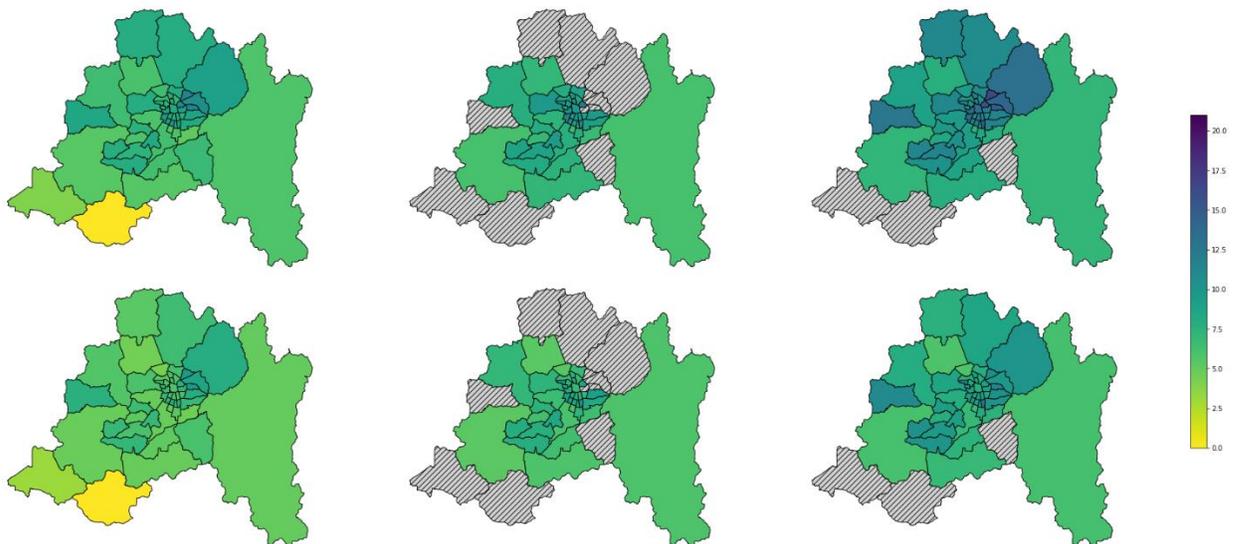


Imagen 34 - Índice de movilidad en la región metropolitana. Fila de arriba contiene registros para los días de semana, mientras que la fila de abajo representa la movilidad durante los fines de semana. Columna de más a la izquierda tiene registros del período comprendido entre el 18 de Marzo y el 18 de Julio del 2020; columna del medio contiene los registros para el Paso 1 y más a la derecha se encuentran los registros para el Paso 2. Colores oscuros representan una movilidad más alta y zonas grises muestran comunas de las cuales no hay registros para el Paso que corresponde.

En esta oportunidad las diferencias entre el Paso 1 y el período previo al Paso a Paso son un poco más difíciles de ver, especialmente si consideramos que las comunas que generalmente concentran la movilidad no presentan datos (Sector oriente y comunas fronterizas como Tiltil y María Pinto). Sin embargo, es innegable que los colores se notan un par de tonos más oscuros.

Ya si miramos las comunas durante el Paso 2, ubicadas más a la derecha, es más clara la diferencia (o quizá debería decir más oscura). Vuelven a tomar protagonismo las comunas del sector Oriente y el centro de la capital. Aunque esta vez no se alejan tanto del resto, viéndose casi del mismo color que sus vecinos del poniente, tanto en días hábiles (semana) como fines de semana.

5.3. Importancia de las variables según Modelos de regresión

Quizá alguien podría estar pensando “por fin se acabaron los gráficos y viene el Machine Learning”, pero nada más lejos de la realidad. Para estas pruebas se han implementado modelos de regresión de manera simple, sin un Feature Engineering espectacular y sin Hyperparameter Tunning.

Que no sea el objetivo de este proyecto no quiere decir que las técnicas de optimización no ayuden a la selección de variables según su importancia, se pueden encontrar muchos artículos al respecto, donde se busca mejorar la selección de variables para disminuir la cantidad de columnas con las que, por ejemplo, se va a entrenar esos modelos, consiguiendo reducciones bastante sorprendentes. ^{[36][37][43]}

El objetivo de estas pruebas no fue obtener el mejor modelo, sino intentar entender, a ojos de estos modelos supervisados de aprendizaje automático, cuales variables son más importantes y por lo tanto las que mejor podrían explicar a la variable objetivo (Índices de movilidad). Pensando en esto, los modelos que menos requisitos necesitan para ser entrenados son aquellos basados en árboles, siendo los más comunes Decision Tree, Random Forest y Gradient Boosting. ^[35]

Es importante considerar que se hicieron pruebas con distintos tipos de escalamiento, normalización, distintas variables y otros modelos, los cuales no se seleccionaron como finalistas por distintos motivos (MSE alto, poco explicables, etc.) y no se presentan en los resultados. Por poner un ejemplo, no se seleccionaron resultados donde se probó el método Normalizer, ya que no fue posible entender cómo se relaciona el cambio en importancia con la transformación lineal realizada por este método, quedando ese punto como algo a investigar en el futuro.

Por último, es relevante mencionar que no se están usando métodos de imputación de datos, si una fila tiene cualquier valor nulo, se elimina completa. Esto es especialmente relevante respecto de las variables de Material Particulado (MP), ya que presentan una mayoría nula.

5.3.1 Período Previo al Paso a Paso (más restricciones)

Importancia de variables para el período previo al plan Paso a Paso comprendido entre el 18 de Marzo y el 18 de Julio. Se utilizaron todas las variables numéricas disponibles, por lo que la cantidad de registros se ve reducida. Se debe considerar que se está mostrando el mejor modelo para cada variable objetivo, según el R^2 .

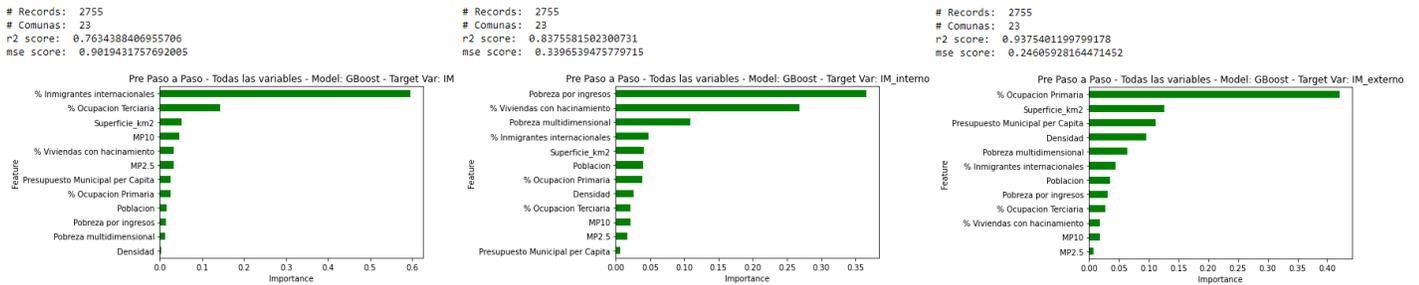


Imagen 35 – Importancia de las variables en período previo al Paso a Paso. Se destacan muchas variables sociales, especialmente para la IM interno. Para IM e IM externo, se destacan las tasas de Ocupación principalmente. Y, por su parte, las variables ambientales no destacaron en prácticamente ningún caso.

La siguiente prueba es revisar la importancia sin las variables de calidad del aire, de tal manera de utilizar todos los registros disponibles para el período.

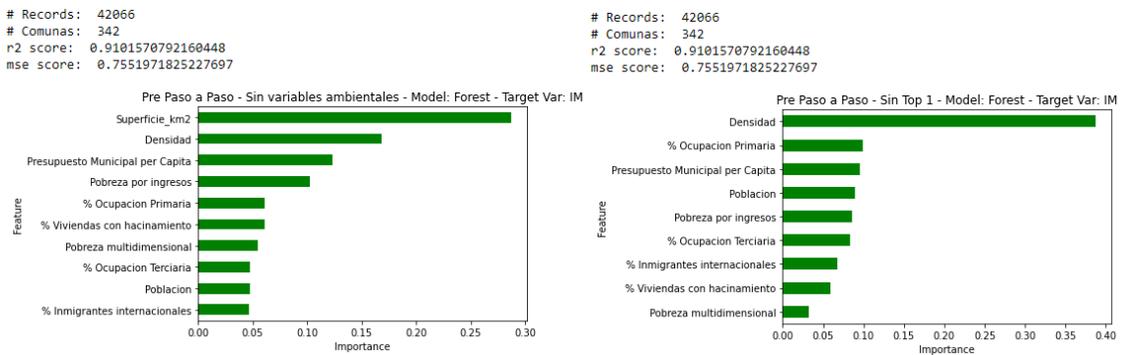
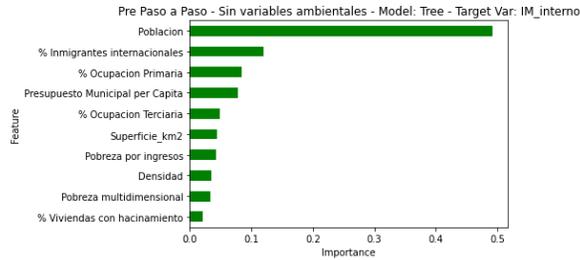


Imagen 36 - En ambos casos vemos que la Densidad o la Superficie son importantes para calcular el Índice de movilidad. En la imagen de la derecha vemos cómo cambian las variables más importantes si quitamos la que fue reconocida como la más importante al lado izquierdo, destacándose más la importancia de la tasa de ocupación primaria.

Records : 42066
 # Comunas : 342
 r2 score: 0.9442016616182418
 mse score: 0.1364651811175846



Records : 42066
 # Comunas : 342
 r2 score: 0.9442016616182418
 mse score: 0.1364651811175846

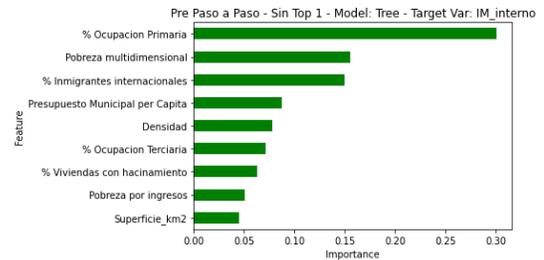
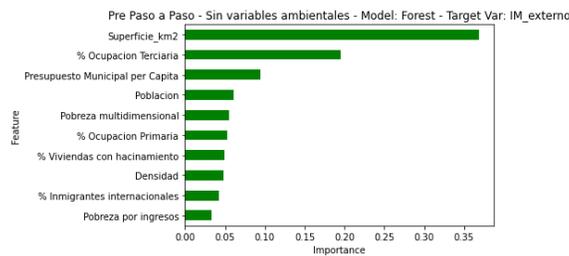


Imagen 37 - Mismo ejercicio que en la imagen anterior, pero esta vez con el IM interno como variable objetivo. En este caso destaca Población como variable más importante, seguida por porcentaje de inmigrantes internacionales, ocupación primaria y pobreza multidimensional.

Records : 42066
 # Comunas : 342
 r2 score: 0.9390462801128038
 mse score: 0.4180218727861839



Records : 42066
 # Comunas : 342
 r2 score: 0.9390462801128038
 mse score: 0.4180218727861839

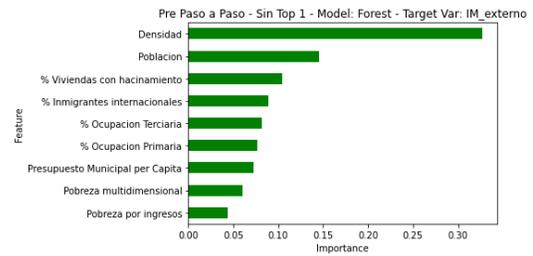


Imagen 38 - Igual que en el caso anterior, pero con IM externo. En este caso ninguna variable social destacó por sobre la densidad, la superficie y la población.

Pareciera ser que durante este período de mayor restricción fue poca la injerencia de las variables sociales, en mínimas instancias se destacaron por sobre las variables demográficas. Y eso no solo ocurrió con los mejores modelos, en general, luego de varias ejecuciones con distintos parámetros, el resultado era igual o muy similar. De todas maneras, en este período esto se podría considerar esperable, ya que podemos pensar que todos estábamos encerrados bajo las mismas condiciones, siendo así las variables sociales serían bastante irrelevantes, esto es algo que tendremos que validar al comparar con los otros períodos.

5.3.2 Comparación Paso 1 y 2 durante los días de semana

Al igual que en la sección anterior, primero veamos el mejor modelo para cada variable objetivo considerando las variables ambientales, recordemos que esto afecta mucho a la cantidad de registros.

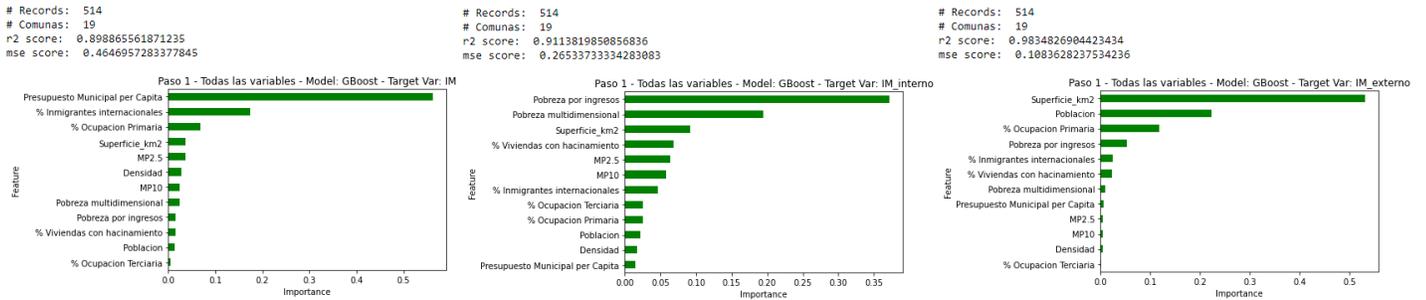


Imagen 39 – Importancia de variables en días de semana para comunas en Paso 1. Innegablemente se destacaron variables socioeconómicas, especialmente en el IM interno e IM (a secas). Las variables ambientales siguen sin destacarse mucho en comparación a las variables demográficas o socioeconómicas. También llama la atención, teniendo en cuenta lo protagonista de las variables sociales para los otros índices, que para el IM externo la superficie y la población sean tan protagonistas, tomando más del 70% de la importancia.

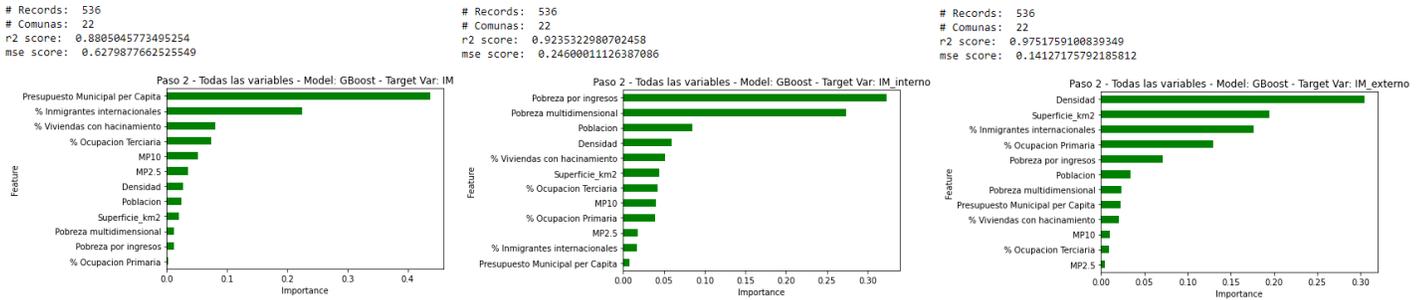


Imagen 40 - Importancia de variables en días de semana para comunas en Paso 2. Vemos que se mantiene la tendencia anterior, donde el Presupuesto Municipal es lo más relevante para el IM, las tasas de pobreza lo son para el IM interno y las variables demográficas son más importantes en el IM externo. Algo que llama la atención es la escalada del porcentaje de Viviendas con hacinamiento en la importancia del modelo para el IM, además de que las variables ambientales y sociales ocupan ahora los 6 primeros lugares, en vez de solo 4 de esos 6 en el Paso 1 (con el mismo modelo Gradient Boosting).

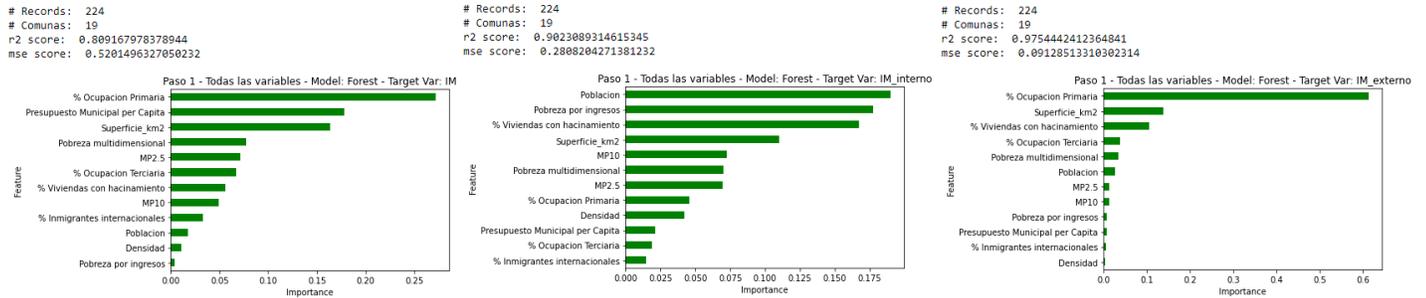


Imagen 41 - Importancia de variables en Fines de semana para comunas en Paso 1. Se destacan bastante las variables socioeconómicas, no así las ambientales. Ocupación primaria aparece como la más relevante para el IM e IM externo, subiendo en importancia en comparación a lo que vimos en el análisis de los días de semana. Vemos un cambio también en el ranking del IM interno, destacándose más la población en este caso, aunque también son más homogéneas las proporciones entre todas las variables.

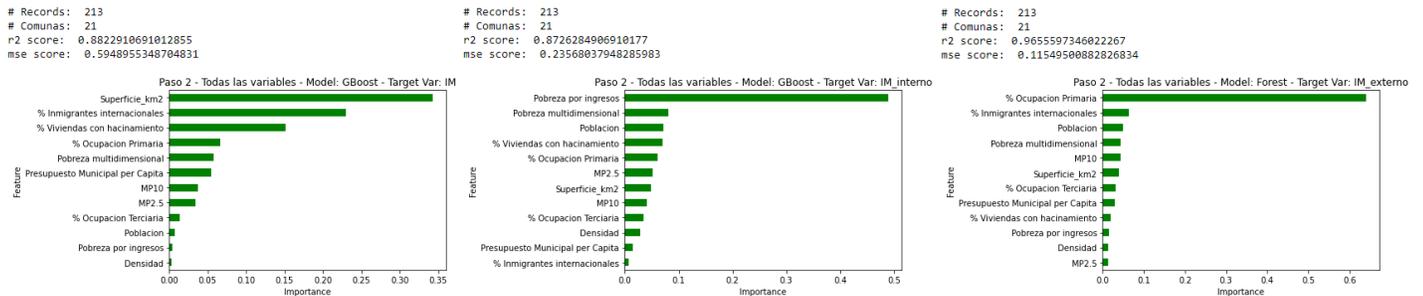


Imagen 42 - Importancia de variables en Fines de semana para comunas en Paso 2. Vuelven a destacar variables socioeconómicas en todos los Índices de movilidad. Todos con casos muy particulares, donde para el IM interno y externo las variables de tasa de pobreza y ocupación primaria (respectivamente) se toman más del 50% de las proporciones; mientras que en el IM se destaca la superficie, pero bajo ella solo encontramos variables socioeconómicas.

Para continuar esta comparación cambiaremos la estrategia anterior, de tal manera que mostraremos el mejor modelo sin variables ambientales de cada Índice en cada Paso solo una vez. Si en algún momento sacamos la mejor variable en una ejecución para comparar resultados, se comentará.

Records: 3905
 # Comunas: 132
 r2 score: 0.916206359411478
 mse score: 0.48711276325936487

Records: 4059
 # Comunas: 177
 r2 score: 0.932426421400282
 mse score: 0.7057367586345558

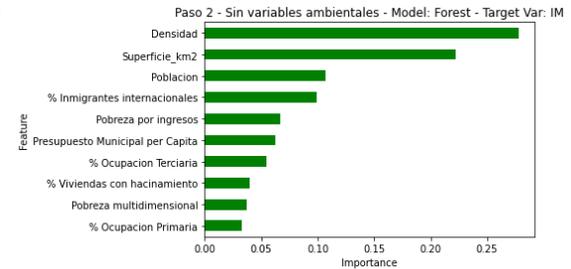
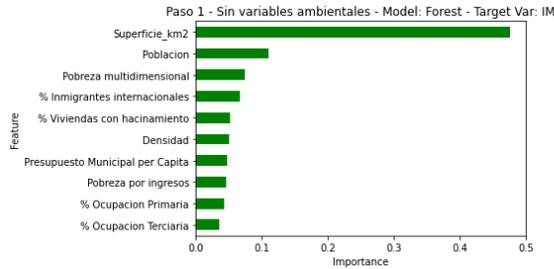


Imagen 43 - Importancia del modelo para el Índice de Movilidad en días de semana. A la izquierda el Paso 1 y a la derecha el Paso 2. No se incluyen las variables de material particulado en el cálculo. Vemos que ninguna variable socioeconómica destaca por sobre las variables demográficas.

Records: 3905
 # Comunas: 132
 r2 score: 0.9555082040821165
 mse score: 0.16682356775128707

Records: 4059
 # Comunas: 177
 r2 score: 0.9736180810731109
 mse score: 0.15009697278050352

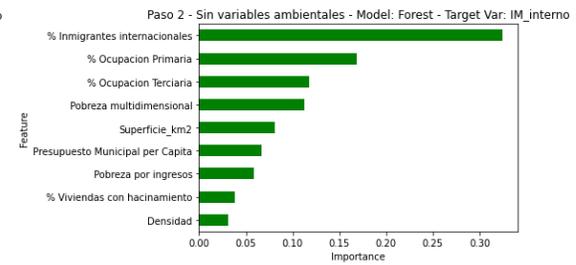
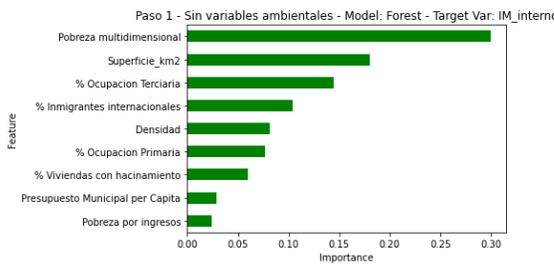


Imagen 44 - Importancia del modelo para el Índice de Movilidad interna en días de semana. A la izquierda el Paso 1 y a la derecha el Paso 2. No se incluyen las variables de material particulado ni la población en el cálculo. En este caso la población se destacó por mucho sobre las variables socioeconómicas, al quitarla se ve que ni la densidad ni la superficie toman protagonismo, esto puede tener sentido con las correlaciones vistas en la Imagen 23 e Imagen 24, donde se ve que la población tiene una alta correlación con las variables sociales.

Records: 3905
 # Comunas: 132
 r2 score: 0.9759055730276694
 mse score: 0.17251788922674144

Records: 4059
 # Comunas: 177
 r2 score: 0.9674970188990095
 mse score: 0.31482885236531943

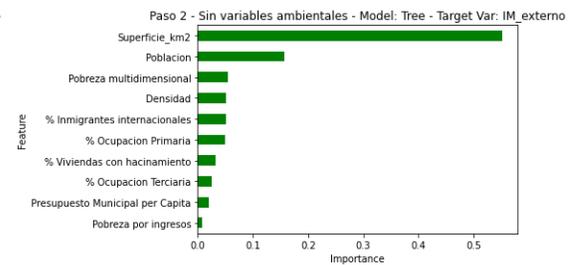
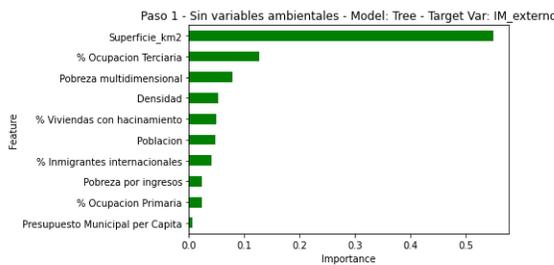
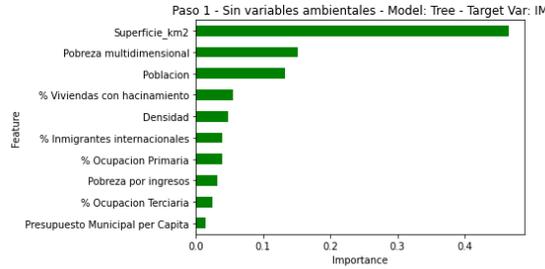


Imagen 45 - Importancia del modelo para el Índice de Movilidad externa en días de semana. A la izquierda el Paso 1 y a la derecha el Paso 2. No se incluyen las variables de material particulado en el cálculo. En este caso, si quitamos la superficie, Densidad y Población toman protagonismo, dejando por detrás a las variables socioeconómicas.

Records: 1666
 # Comunas: 132
 r2 score: 0.86647083424206
 mse score: 0.6134440470847312



Records: 1698
 # Comunas: 172
 r2 score: 0.8993388153455146
 mse score: 0.7575350027596607

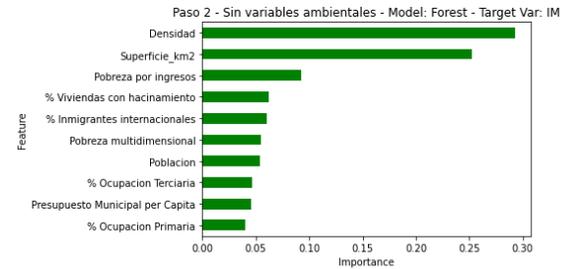
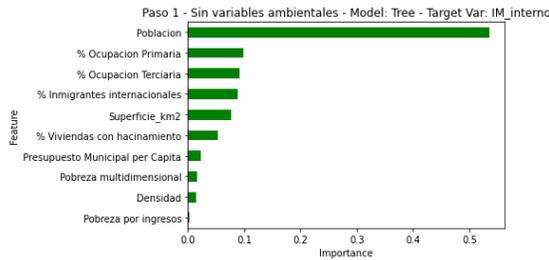


Imagen 46 - Importancia del modelo para el Índice de Movilidad en fines de semana. A la izquierda el Paso 1 y a la derecha el Paso 2. No se incluyen las variables de material particulado en el cálculo. Vemos nuevamente que ninguna variable socioeconómica destaca por sobre las variables demográficas.

Records: 1666
 # Comunas: 132
 r2 score: 0.934360648152857
 mse score: 0.16398939236004997



Records: 1698
 # Comunas: 172
 r2 score: 0.9530493538658658
 mse score: 0.14135174252353047

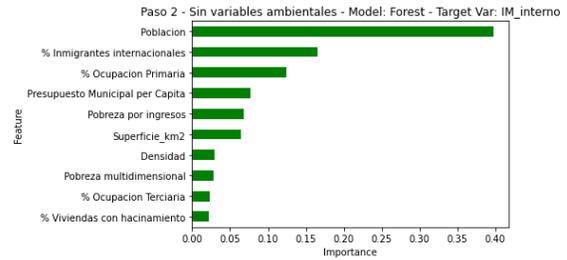
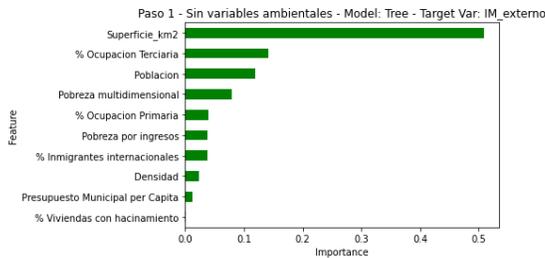


Imagen 47 - Importancia del modelo para el Índice de Movilidad interna en fines de semana. A la izquierda el Paso 1 y a la derecha el Paso 2. No se incluyen las variables de material particulado en el cálculo. Vemos el mismo caso que en la Imagen 44, en este caso se decidió mantener la Población para ilustrar la diferencia.

Records: 1666
 # Comunas: 132
 r2 score: 0.9477918184005559
 mse score: 0.2697907050118217



Records: 1698
 # Comunas: 172
 r2 score: 0.9239775264228269
 mse score: 0.542467344593092

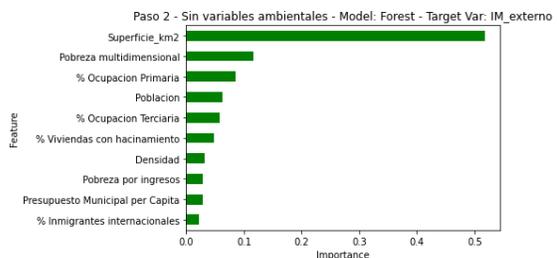


Imagen 48 - Importancia del modelo para el Índice de Movilidad externa en fines de semana. A la izquierda el Paso 1 y a la derecha el Paso 2. No se incluyen las variables de material particulado en el cálculo. Y volvemos a ver un comportamiento similar a lo que veíamos en los días de semana, aunque es necesario mencionar que, al igual que antes, se destacan, bajo la variable de superficie, las tasas de ocupación y pobreza.

5.3.3 Período previo a las restricciones

Ya finalizando y para usarlo de referencia, veamos la importancia durante el período previo a las restricciones, es decir, antes del 18 de Marzo. En este caso no incluiremos tanto detalle como antes, dejando de lado las variables ambientales que, como vimos antes, no toman protagonismo en casi ningún caso. Tampoco haremos una diferencia entre Fin de semana y días hábiles.

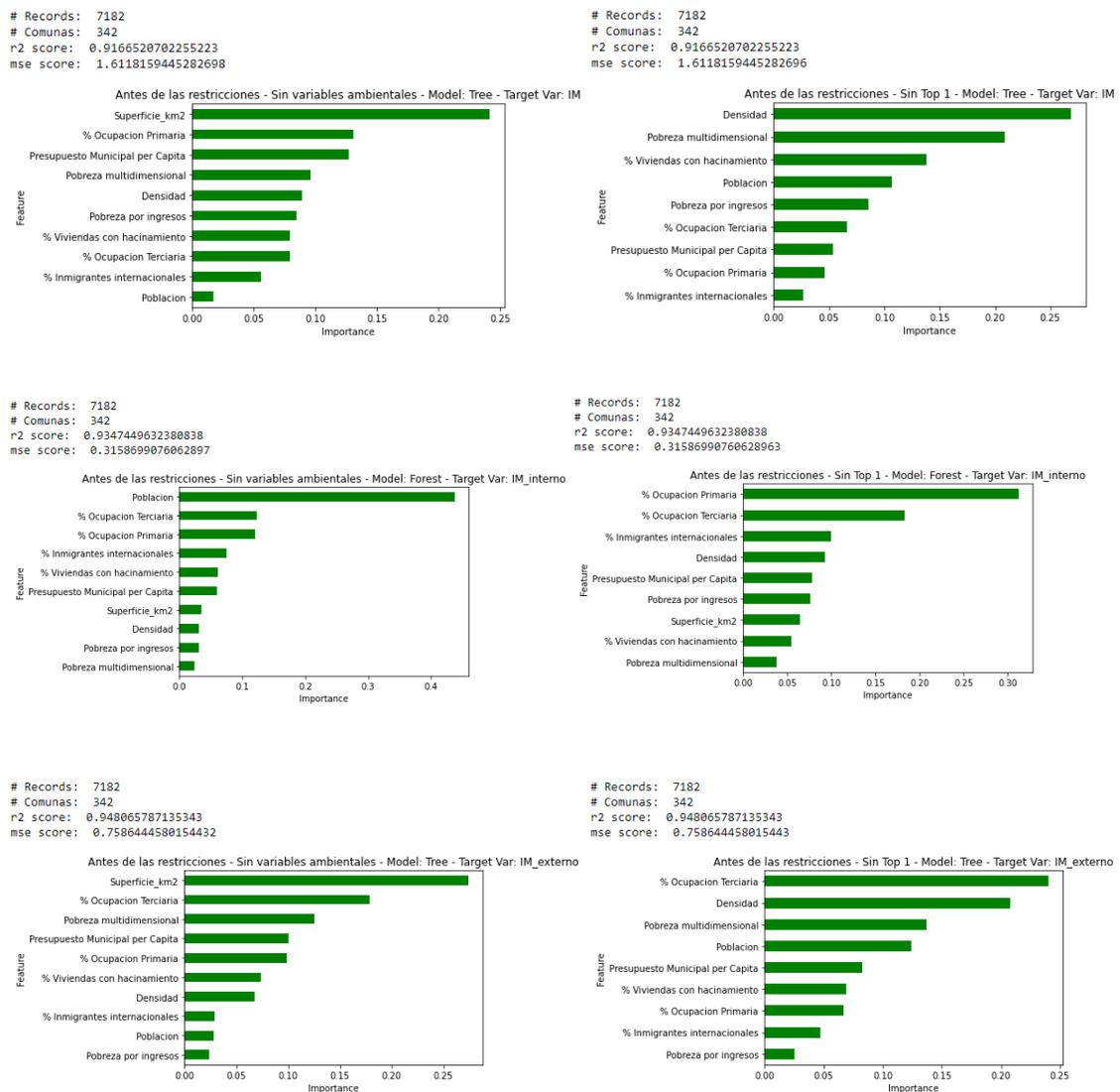


Imagen 49 - Importancia de variables en el período previo a las restricciones. No se incluyen las variables de material particulado en el cálculo y en los gráficos de la izquierda no se incluye la variable que en el lado izquierdo haya salido como primera en importancia. Llama la atención el protagonismo de las variables de ocupación en casi todos los gráficos, aunque no siempre en primer lugar.

Región	Top 2 Variables más importantes - Random Forest - IM
1	Población y Superficie
2	Población y Densidad
3	Población y Densidad
4	Densidad y Pobreza multidimensional
5	Población y Densidad
6	Hacinamiento y Pobreza multidimensional
7	Superficie y Ocupación primaria
8	Ocupación primaria y Hacinamiento
9	Densidad y Presupuesto municipal
10	Densidad y Pobreza multidimensional
11	Tasa de inmigrantes y Hacinamiento
12	Hacinamiento y Superficie
13	Ocupación terciaria y Pobreza multidimensional
14	Población y Presupuesto municipal
15	Ocupación primaria y Densidad
16	Superficie y Pobreza por ingresos

Tabla 3 - Variables más importantes por región en el período previo a las restricciones. Modelo Random Forest e Índice de Movilidad como variable objetivo.

6. Conclusiones

Es innegable que existe una relación entre las variables socioeconómicas y la movilidad, lo mismo ocurre con las variables ambientales. Lo vimos en las comparaciones realizadas para los índices de movilidad y también en artículos relacionados, donde se compara la movilidad desde distintos puntos de vista con las variables que en este proyecto revisamos.

Si hablamos de variables económicas, el caso más notorio de esta relación se da en la región metropolitana. Gracias a las diferencias sociales muy marcadas que se presentan entre el sector Oriente y el centro de la capital versus la periferia y el poniente.

En esta región somos capaces de notar gráficamente como la movilidad interna se concentra en las zonas de mayor ingreso. Estas zonas son también donde se concentra la mayor cantidad de trabajadores terciarios, ya que muchas empresas han posicionado sus sedes ahí. Mientras que la movilidad externa se concentra en comunas de Paso (como las vecinas de Santiago centro) o fronterizas. Y esto se repite en otras regiones, donde pudimos ver que la importancia de variables socioeconómicas toma protagonismo por sobre otras.

Por lo que muy probablemente seríamos capaces de predecir la movilidad según variables socioeconómicas. Aunque deberíamos apuntar a mejorar nuestros modelos de regresión, además de conseguir más variables (o robustecer las que tenemos) para describir de mejor manera las condiciones socioeconómicas de las regiones, comunas o individuos.

Ahora, **sobre las variables ambientales**, aquí tenemos dos problemas; el primero, no es claro que la movilidad sea afectada por las condiciones ambientales, sino que podría ser más bien al revés; y el segundo, nos falta información ambiental más precisa para otras comunas o, nuevamente, robustecer la que existe de tal manera de extrapolar la contaminación de otras comunas.

Sobre el primer problema tenemos lo encontrado en algunos artículos, además del sentido común. La contaminación aumentaría por la movilidad ya que las personas usan más medios de transporte contaminantes al desplazarse. Además, la polución aumenta en invierno por el uso de métodos de calefacción contaminantes, mientras que la movilidad podría disminuir teniendo en cuenta que las personas saldrían a la calle lo menos posible para evitar el frío o la lluvia. Aunque para esto no tenemos pruebas en este informe.

Sobre el segundo problema, la información ambiental es difícil de conseguir y extrapolarla requeriría de mucho trabajo de validación, además necesitaríamos datos de movilidad de igual manera. Esto puede quedar como trabajo futuro, aunque ojalá no en el contexto de otra pandemia.

Por último, y respondiendo a nuestro objetivo principal, somos capaces de notar diferencias en la movilidad durante los Pasos 1 y 2. Esta diferencia es mucho más clara en los días de semana, de manera esperable por supuesto, ya que el Paso 2 entregaba muchas más libertades durante los días de trabajo.

En fines de semana esta diferencia es mucho menos notable, siendo casi idénticos, en algunos casos, los niveles de movilidad entre ambos Pasos, incluso las variables de importancia son técnicamente las mismas.

Además, intentar dilucidar si las variables socioeconómicas tienen alguna incidencia en las pequeñas diferencias es complejo (dejaremos fuera de esta discusión a las variables ambientales por razones explicadas en los párrafos anteriores), no existen indicadores que nos den de manera convincente que la movilidad cambió por alguna de las variables utilizadas en este proyecto.

Es importante mencionar que existen diferencias notorias entre las regiones, por lo que podría ser más efectivo hacer análisis más exhaustivos en cada una de ellas o agrupando por comunas con características similares (podría apoyarse esto último de modelos de segmentación o clustering como k-means).

Se entiende la importancia del análisis de Movilidad, tanto en contexto de emergencias de salud, como en también en planificación de transporte, infraestructura, territorio, entre otros. Se pueden aprovechar las nuevas tecnologías para mejorar este tipo de investigaciones, aprovechando las grandes capacidades de cómputo que ofrece la nube, junto con modelos entrenados con volúmenes de datos y servidores impensados para equipos de investigación pequeños.

Bibliografía

- [1] Lovo J. ¿Covid-19: son las medidas no farmacológicas la respuesta? *Mul Med.* 2021
- [2] Artículo en *France24*. Polémica en Chile por la efectividad de las nuevas medidas contra el Covid-19. 2020
- [3] Artículo Universidad de Chile. COVID-19 en Chile: a 365 días del primer caso en el país. 2021
- [4] Artículo *pauta.cl*. Un año de pandemia en Chile. 2021
- [5] Artículo *minsal.cl*. Gobierno anunció cuarentena total para siete comunas de la RM. 2020
- [6] Archivo *senado.cl*. Estado de Excepción Constitucional. Sin fecha
- [7] Archivo *minsal.cl*. Conoce el Plan Paso a Paso. 2020
- [8] Artículo *radio.uchile.cl*. Se acaba el Toque de Queda: Gobierno confirma que no renovará estado de excepción. 2021
- [9] Artículo *latercera.cl*. Historia de dos ciudades. 2015
- [10] Felipe Link et al. Segregación, estructura y composición social del territorio metropolitano en Santiago de Chile. Complejidades metodológicas en el análisis de la diferenciación social en el espacio. 2015
- [11] Artículo *euronews.com*. Coronavirus: Half of humanity now on lockdown as 90 countries call for confinement. 2020
- [12] Ernesto López. Impacto del crecimiento del gran Santiago en el deterioro funcional de sus espacios pericentrales. 2005
- [13] Claudio A. Agostini et al. Segregación residencial de ingresos en el Gran Santiago, 1992-2002: una estimación robusta. 2016
- [14] Marian Rybansky et al. The Impact of Terrain on Cross-Country Mobility – Geographic Factors and their Characteristics. 2014
- [15] Lars Böcker et al. Weather and daily mobility in international perspective: A cross-comparison of Dutch, Norwegian and Swedish city regions. 2019
- [16] Paola Jirón M. et al. Exclusion and spatial inequality: an analysis from a daily mobility perspective. 2010
- [17] Daniel Moreno et al. Desigualdades urbanas: costos y tiempos de viaje en el Área Metropolitana de Santiago. 2021

- [18] Thiago Hérick de Sá et al. Impact of travel mode shift and trip distance on active and non-active transportation in the São Paulo Metropolitan Area in Brazil. 2015
- [19] Amy Wesolowski et al. Impact of human mobility on the emergence of dengue epidemics in Pakistan. 2015
- [20] Choong-Ki Lee et al. The impact of non-pharmaceutical interventions for 2009 H1N1 influenza on travel intentions: A model of goal-directed behavior. 2012
- [21] Paolo Bajardi et al. Human Mobility Networks, Travel Restrictions, and the Global Spread of 2009 H1N1 Pandemic. 2011
- [22] André de Palma et al. An overview of effects of COVID-19 on mobility and lifestyle: 18 months since the outbreak. 2022
- [23] Marco Dueñas et al. Changes in mobility and socioeconomic conditions during the COVID-19 outbreak. 2021
- [24] Nicolò Gozzi et al. Estimating the effect of social inequalities on the mitigation of COVID-19 across communities in Santiago de Chile. 2021
- [25] Luca Pappalardo et al. A dataset to assess mobility changes in Chile following local quarantines. 2023
- [26] Leo Ferres et al. Measuring Levels of Activity in a Changing City. Sin fecha
- [27] Leo Ferres et al. Evaluation of home detection algorithms on mobile phone data using individual-level ground truth. 2021
- [28] Repositorio <https://github.com/leoferres/Datos-COVID19>
- [29] Repositorio <https://github.com/MinCiencia/Datos-COVID19>
- [30] Artículo Biblioteca del Congreso Nacional de Chile. SIIT de la BCN libera nuevos "Perfiles Comunales". 2023
- [31] Santi Phithakkitnukoon et al. Socio-Geography of Human Mobility: A Study Using Longitudinal Mobile Phone Data. 2012
- [32] Guillermo Sepúlveda. Impacto de la restricción vehicular sobre la calidad del aire: lecciones para Santiago de Chile. 2017
- [33] Beatriz Pineda et al. Aspectos relevantes de la movilidad y su relación con el medio ambiente en el Valle de Aburrá: una revisión. 2017
- [34] Archivo bcn.cl. Mapas vectoriales. Sin fecha
- [35] Artículo Harpreet Sahota. Can't Decide Between a Linear Regression or a Random Forest? Here, Let Me Help. 2022
- [36] Ahmad Alsaahaf et al. A framework for feature selection through boosting. 2022

- [37] Feng Pan et al. Feature Selection for Ranking using Boosted Trees. 2009
- [38] Egon Montecinos. Antecedentes sobre la relación histórica centralismo y descentralización en Chile. 2005
- [39] Artículo soychile.cl. Dos personas fueron formalizadas por falsificar permiso temporal en Mejillones. 2020
- [40] Archivo cdn.digital.gob.cl. Instructivo para permisos de desplazamiento. 2020.
- [41] Artículo diarioconcepcion.cl. PDI de Concepción detuvo a sujeto que vendía permisos falsos para transitar en cuarentena. 2020
- [42] Artículo Biblioteca del Congreso Nacional de Chile. Plebiscito logra la mayor participación electoral en la era del voto voluntario. 2020
- [43] Miron B. Kursa and Witold R. Rudnicki. The All Relevant Feature Selection using Random Forest. 2011