



Universidad del Desarrollo  
Facultad de Ingeniería

# ANÁLISIS DE MOVILIDAD HUMANA PARA EL AÑO 2020 DURANTE LAS CUARENTENAS EN CHILE

*Modelamiento considerando las comunas de la region Metropolitana.*

POR: PATRICIO RAMÍREZ VEAS Y PABLO ELGUETA BARRIOS

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

## TABLA DE CONTENIDO

<b>RESUMEN.....</b>	<b>1</b>
<b>1. INTRODUCCIÓN.....</b>	<b>3</b>
<b>2. TRABAJO RELACIONADO.....</b>	<b>4</b>
<b>3. HIPÓTESIS Y OBJETIVOS.....</b>	<b>10</b>
<b>4. DATOS Y METODOLOGÍA.....</b>	<b>11</b>
4.1. DATOS.....	11
4.2. METODOLOGÍA.....	14
<b>5. RESULTADOS.....</b>	<b>23</b>
<b>6. CONCLUSIONES.....</b>	<b>36</b>
<b>BIBLIOGRAFÍA.....</b>	<b>38</b>

## **Resumen**

El 11 de marzo del 2020, se declaró la pandemia en Chile debido al virus Sars-CoV-2, o mejor conocido como COVID-19. Este anuncio trajo varios cambios dentro del país por los últimos tres años, uno de esos cambios es la movilidad que tuvo el país desde el 2020 hasta el 2022, en donde debido a la pandemia el país instauró un plan paso a paso en donde cada comuna tenía sus restricciones a la hora de poder salir a la calle. Debido a esto, la gente en Chile comenzó a salir menos, quedándose en sus casas haciendo cuarentena.

Durante la pandemia, se pudo hacer un rastreo del movimiento de las personas gracias a las antenas telefónicas repartidas por todo el país, en donde cada una podía rastrear cuando un teléfono entraba y salía de su señal. Gracias a esto, se puede ver a través de todo el país como la pandemia afectó la movilidad de los chilenos desde el 2020 hasta la actualidad, se puede ver como cada comuna reaccionó a este suceso y como cada una fue evolucionando durante el plan Paso a Paso del gobierno de Chile hasta su punto final en donde llegó al paso de apertura el 1 de octubre del 2022 hasta el día de hoy.

Los datos que se han recopilado durante el 2020 del índice de movilidad de las personas fueron hechos accesibles para poder hacer un análisis futuro acerca de los efectos de este virus y las futuras consecuencias que podría llegar a tener si es que ocurriese otra pandemia similar.

Utilizando los datos recopilados, se generó un análisis y exploración de datos el cual cuenta un poco la historia durante la pandemia en el 2020, cuenta cosas como las comunas que más movimiento interno y/o externo tuvieron, además del total de la movilidad llamada el índice de movilidad. Además, se pueden buscar explicaciones a las del por qué el movimiento de las comunas si es que se toman en cuenta los datos socioeconómicos de ella, además de los ambientales, en donde las comunas con mejor estrato socioeconómico tuvieron menos movilidad en contraste a las que más lo tuvieron en donde en ellas si se observó una mayor tasa de movilidad en comparación a la media.

Junto con los datos de movilidad, también estaban accesibles los datos del plan Paso a Paso del gobierno de Chile, en donde se podía ver para cada comuna en esa semana en qué fase se encontraban. Este data set fue esencial para este trabajo dado que ayudo mucho en el trabajo del análisis de datos, además del trabajo del modelo de machine learning.

Finalmente, se hizo un modelo que fue capaz de predecir un muy buen nivel de precisión los índices de movilidad de las regiones de Chile externas a la región metropolitana. Específicamente fueron utilizadas todas las comunas de la region metropolitana, y se utilizaron los datos de su índice de movilidad para los días del 2020 desde que comenzó el plan Paso a Paso, en donde las fases para cada día cumplieron un rol muy importante.

# 1. Introducción

El siguiente informe muestra a continuación presenta el desarrollo de investigación, análisis y resultados respecto al estudio de datos de movilidad durante los periodos de restricciones de movilidad durante el desarrollo de la pandemia COVID-19 durante el año 2020.

El presente análisis se divide en dos partes. La primera se encuentra orientada a un análisis exploratorio de datos de movilidad a nivel de la Región Metropolitana que permita identificar patrones o correlaciones respecto al impacto de las cuarentenas en la movilidad en las diferentes comunas, buscando identificar patrones de comportamiento en las reducciones o incrementos de desplazamientos.

El segundo análisis se centra en el desarrollo de un modelo de machine learning que permitan predecir el impacto que tendría una cuarentena en la movilidad de una determinada comuna. El modelo de Machine Learning debe ser capaz de predecir el comportamiento que tendrían las comunas de la Región Metropolitana como respuesta a un determinado nivel o fase de restricción de movilidad.

La base de datos estudiadas corresponde a datos estructurados de los índices de movilidad comunales durante el periodo de febrero a noviembre de 2020. Estos datos tienen su origen en el resultado obtenido del procesamiento de los registros de la compañía Movistar correspondientes a las conexiones de dispositivos móviles a las antenas de servicio. Este tipo de registros, individualizados por usuario y horarios de conexión, permite estimar los desplazamientos de los usuarios identificando los cambios de conexión a una antena diferente.

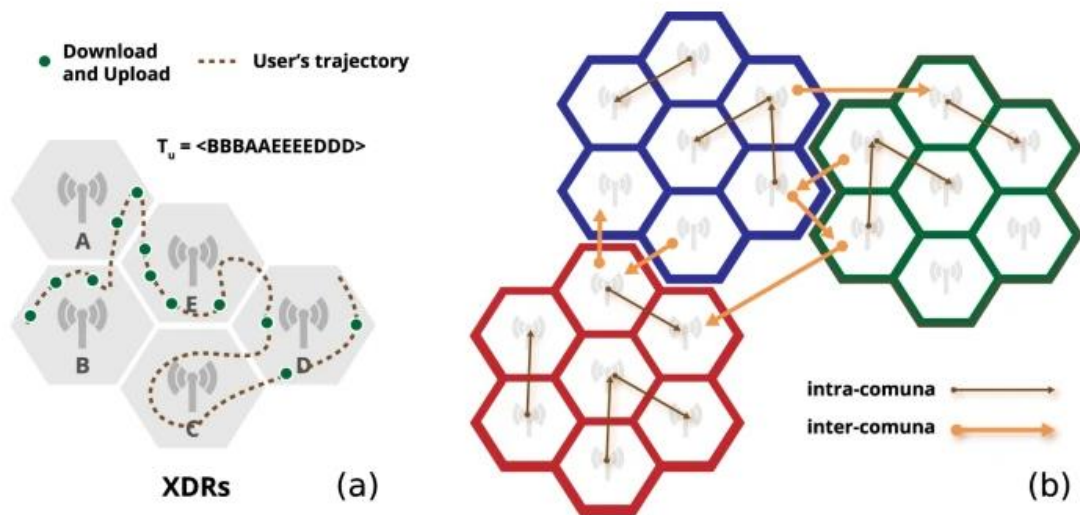
## **2. Trabajo Relacionado**

Se han hecho varias investigaciones con respecto a datos telefónicos y los efectos que ellos pueden tener.

Una investigación muy importante para comenzar todo es la cual crea el dataset para ver los cambios de movilidad en Chile para las cuarentenas locales. Esta investigación cuenta con 31 billón de datos XDRs o Registros de Detalle Extendidos, y 5.4 millón de usuarios desde el 26 de febrero hasta el 20 de septiembre. Con estos datos, los investigadores crearon tres nuevos índices los cuales nos dicen cuál fue el grado de movilidad de una comuna.

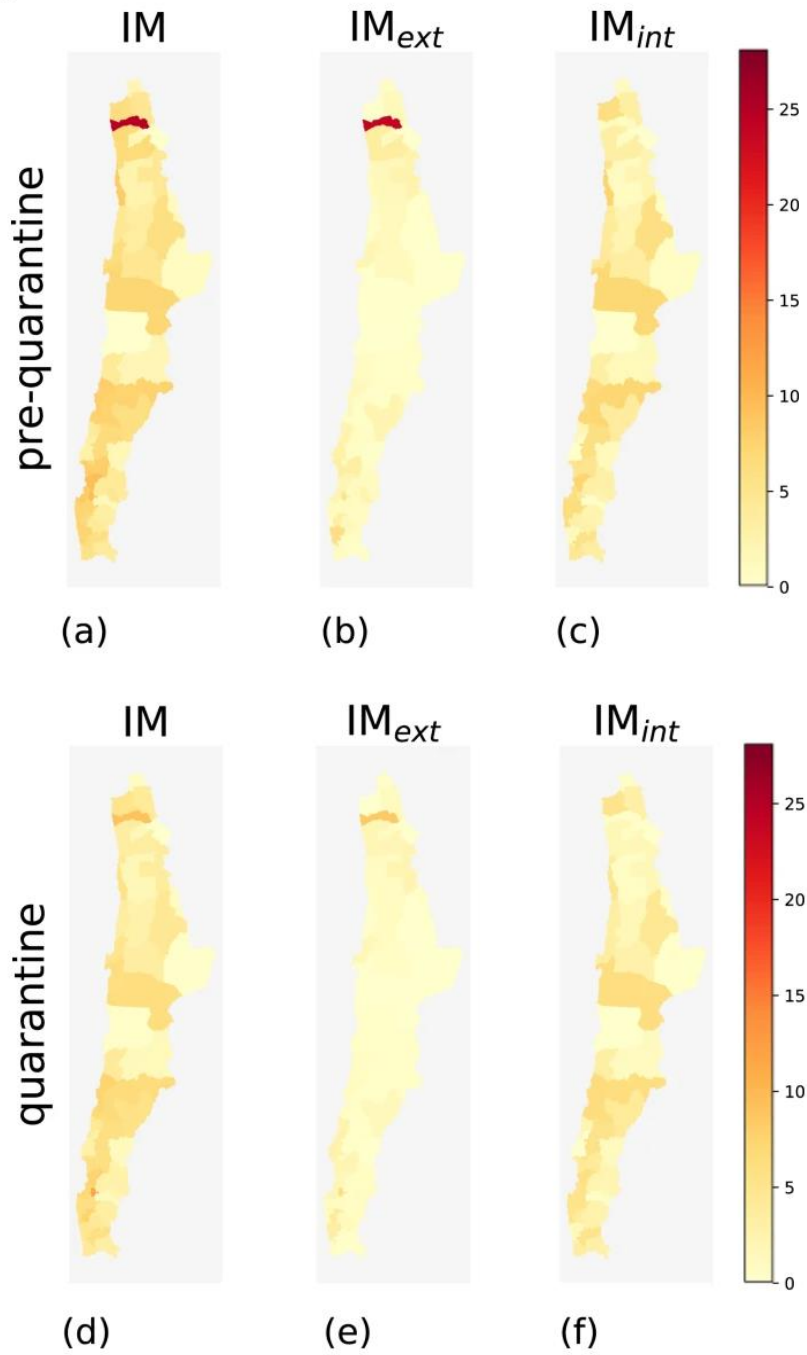
Utilizando los XDRs, se puede detectar el uso de un teléfono móvil a través de cualquier descarga o uso de páginas http, con lo cual se puede ver a qué antena se conecta a qué hora en específico. Con esta información, si después se toma la siguiente antena a la que el usuario se conectó, se puede con un grado de precisión decir si la persona siguió o no en la misma comuna, por lo que podemos decir que esa persona se movió por dentro o hacia afuera de la comuna. Gracias a esta información, los investigadores crearon el índice de movilidad interno y externo, en donde el interno simplemente nos dice la movilidad interna de la comuna (que se obtiene por la cantidad de conexiones dividida por la cantidad de habitantes en la comuna), la externa que nos dice la movilidad externa de la comuna (calculada de igual manera que la interna, pero hacia afuera de la comuna), y finalmente el índice de movilidad que es la suma entre la interna y la externa.

Fig. 1



En esta figura se aprecian tres comunas denotadas por el color rojo azul y verde, en donde se puede ver las conexiones dentro de las comunas cómo las conexiones entre comunas. Además, se aprecia también el uso de XDRs para poder detectar la conexión a una antena y el movimiento que hace el usuario durante el día conectándose a varias antenas por el resto del día.

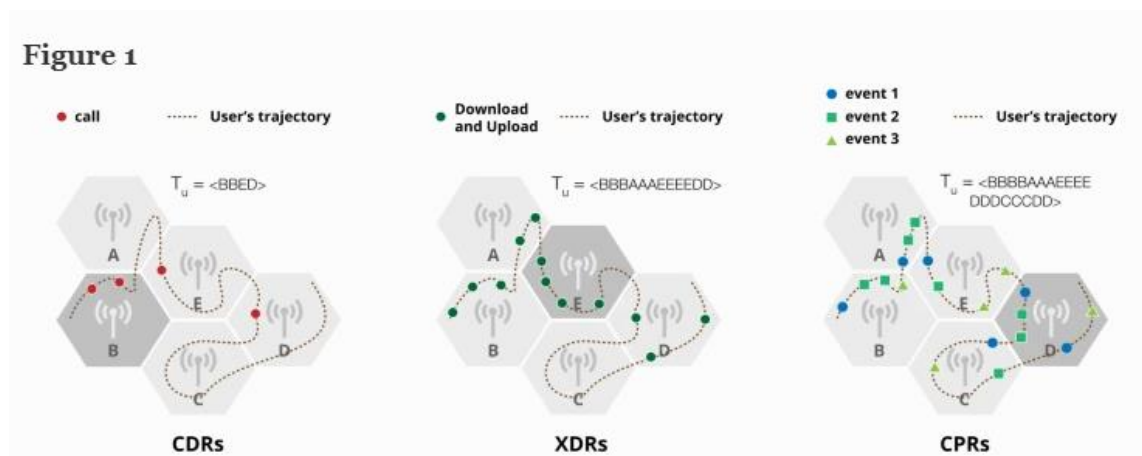
Fig. 2



Choropleth maps of IM, IM<sub>int</sub> and IM<sub>ext</sub> for the comunas in northern Chile for the pre-quarantine (first row) and the quarantine (second row) periods.

Los investigadores a través de los datos llegaron a la lógica conclusión de que antes de la cuarentena y después de la cuarentena se aprecia una gran discrepancia entre los índices de movilidad, en donde antes de la cuarentena existía un nivel más grande de movilidad.

Otra investigación la cual intenta evaluar los algoritmos de detección de hogar utilizando datos móviles, hace la comparación entre los Registros de Detalle de Llamadas (CDRs), Registros de Detalle Extendidos (XDRs) y Registros de Plano de Control (CPRs). Entre estos tres, detalla que los CDRs son simplemente las llamadas que un usuario recibe, los XDRs, las descargas que se hace en el teléfono móvil de un usuario, y finalmente los CPRs son los eventos en la red en la cual un usuario toma parte cómo por ejemplo conectarse a una nueva antena o conectarse a un nuevo dispositivo ya sea wifi o bluetooth.

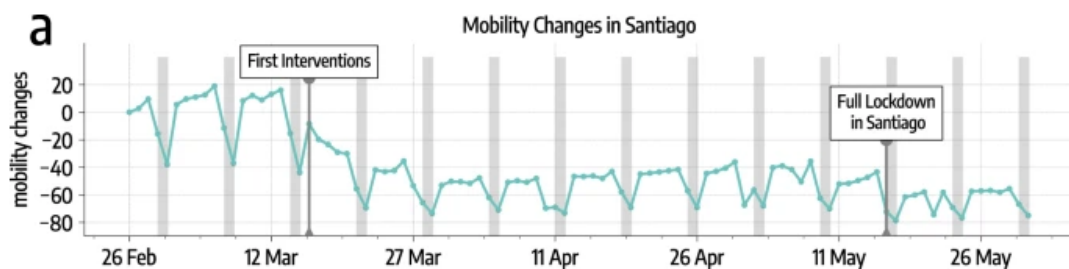


En esta figura se ve en resumidas cuentas cómo se comportan los tres tipos de registros por separado. Se puede apreciar que de izquierda a derecha la cantidad de datos recopilados va en aumento, siendo el CDRs el menor y el CPRs el más alto. En esta investigación se trataron 19,234 CDRs, 43,607 XDRs y 772,871 CPRs. Finalmente,

concluyeron que los CPRs y los XDRs son los mejores datos para poder predecir la vivienda de un usuario.

En otra investigación, se utilizaron los datos móviles para estimar los efectos en la mitigación del COVID-19 a través de las distintas comunas dependiendo de su estado socioeconómico. En esta investigación, se utilizaron datos anonimizados de 1.4 millón de usuarios de la empresa de Telefónica. Se estudiaron los efectos del plan del gobierno de Chile para combatir el rápido crecimiento de casos de COVID-19 en la población, desde el comienzo de ésta en marzo del 2020, incluyendo eventos tales como el cierre total de colegios, y el plan Paso a Paso colocado por el gobierno de Chile.

**Fig. 1: Mobility and contacts changes in Santiago.**



Tal cómo se puede apreciar en este gráfico, los cambios de movilidad cambiaron drásticamente durante las primeras intervenciones del estado y durante el cierre total de la región. Detectaron que los actos del gobierno de Chile ayudaron mucho a la mitigación del virus y lograron prevenir un 34.7% más de muertes.

Lamentablemente, en la región Metropolitana existe una gran diferencia entre las comunas en términos socio-económicos, en donde hay comunas las cuales tienen un ingreso mucho mayor a otras, y en esta investigación se concluyó que mientras las intervenciones hechas por el gobierno fueron exitosas, estas fueron principalmente exitosas en las comunas más

ricas y con una mayor calidad de vida, ya que estas no tuvieron grandes problemas para poder hacer uso del distanciamiento social y quedarse en sus casas. Pero, a diferencia de las comunas con más dinero, las comunas más pobres no tuvieron el mismo efecto, ya que, aunque el movimiento en general para estas comunas se redujo, no se redujo tanto en las comunas con una menor calidad de vida.

### **3. Hipótesis y Objetivos**

Hipótesis:

A través de los datos obtenidos en la region metropolitana, se puede generar un modelo el cual pueda correctamente predecir los índices de movilidad con un grado de precisión aceptable en las demás comunas de características urbanas o rurales cercanas a las principales ciudades de Chile.

Objetivo General: Generar un modelo el cual pueda predecir los índices de movilidad en las comunas de características urbanas o rurales cercanas a las principales ciudades de Chile durante la pandemia del 2020.

Objetivos Específicos:

- Realizar un análisis y exploratorio sobre el conjunto de datos buscando características o patrones que permitan identificar similitudes en los comportamientos de movilidad como respuesta a los diferentes grados de restricción de movilidad.
- Generar un esquema para los conjuntos de datos que permitan ser utilizados para entrenar un modelo de machine learning. Que esto considere identificar las variables significativas para así evitar un sobreajuste en el modelo o bien posibles distorsiones.
- Complementar el conjunto de datos de índice de movilidad utilizando los datos correspondientes al plan paso a paso del gobierno de Chile. Para así, mejorar el modelo final e identificar como estos se relacionan entre sí.
- Identificar por lo menos dos modelos de machine learning se ajusten adecuadamente para la predicción de los índices de movilidad ante variaciones en las restricciones de movilidad.

## 4. Datos y Metodología

### 4.1. Datos

Los datos utilizados para el análisis de movilidad en la Región Metropolitana corresponden al resultado obtenidos por el Instituto de Data Science de la Universidad del Desarrollo en el cual se obtuvieron 3 series de tiempo correspondientes a los Índices de Movilidad a nivel nacional en Chile durante el periodo entre febrero de 2020 y diciembre de 2020. A su vez, el análisis desarrollado por la Universidad del Desarrollo se basó en los Datos proporcionados por la Mesa de Datos Covid-19 liderada por el Ministerio de Ciencia, Tecnología, Conocimiento e Innovación.

<https://raw.githubusercontent.com/MinCiencia/Datos-COVID19/master/output/producto33/IndiceDeMovilidad.csv>

Los archivos contienen las columnas 'Región', 'Código Región', 'Comuna', 'Código comuna', 'Población', 'Superficie\_km2' y múltiples columnas correspondientes a '[fecha]'. Estas últimas columnas, '[fecha]', contienen el valor para el índice de movilidad en la fecha indicada.

	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

Los índices de movilidad correspondientes a 3 series de tiempo y se definen como:

- **IM Interno:** Representa la evolución del índice de movilidad interno a la comuna es una medida de los viajes que ocurren al interior de dicha unidad administrativa.
- **IM Externo:** Representa la evolución del índice de movilidad externo la comuna es una medida tanto de los viajes que tienen origen al interior de la comuna y destino al exterior de la comuna, como de los viajes que tienen origen al exterior de la comuna y destino al interior de la comuna.
- **IM:** Representa la evolución del índice de movilidad, que es la suma de estos últimos dos índices.

Expuesto lo anterior, es importante decir que, los datos se encuentran estructurados de tal forma que se han separado en 2 escalas de movilidad humana. En un primer nivel se han registrado los desplazamientos a nivel interno de una comuna, es decir identificando cambios de conexión a antena, pero dentro de los límites territoriales de una misma comuna. Es importante mencionar que, dentro de las estructuras de administración política territorial de Chile, la comuna corresponde a la estructura de división más pequeña.

El segundo nivel en que se encuentran agrupados los datos corresponde a los desplazamientos externos a la comuna, es decir, a aquellos usuarios que han cambiado de antena de conexión a datos móviles a una fuera de la misma administración comunal de la antena previa. En este punto, es importante destacar que debido a que los datos no se encuentran individualizados por usuario, no es posible conocer los orígenes y destinos exactos de los desplazamientos realizados por los usuarios, sino que solo saber la cantidad de personas que entraron o salieron a una determinada comuna durante el transcurso de un determinado día. De este modo, para la obtención de las series de tiempo, se consideró un viaje cómo el paso de una antena a otra, es decir para dos comunas A y B, tenemos que

el número de viajes de A hacia B queda estimado como la suma de los viajes entre antenas que se encuentran dentro de A y antenas que se encuentran dentro de B.

A modo de datos complementarios, se ha utilizado el conjunto de datos correspondiente al registro de cambio en las condiciones de restricciones de movilidad durante el periodo de cuarentena COVID-19 en Chile. Esta política pública que estableció criterios y grados de restricción se catalogó como “Plan Paso a Paso” y buscaba aplicar restricciones de movilidad graduales acordes al nivel de contagios y variaciones en dichas tasas, entre otros factores.

codigo_region	region_residencia	codigo_comuna	comuna_residencia	zona	2020-07-28	2020-07-29	2020-07-30	2020-07-31	2020-08-01	2020-08-02
0	15	Arica y Parinacota	15101	Arica Rural	3	3	3	3	3	3
1	15	Arica y Parinacota	15101	Arica Urbana	1	1	1	1	1	1
2	15	Arica y Parinacota	15102	Camarones Total	3	3	3	3	3	3
3	15	Arica y Parinacota	15202	General Lagos Total	3	3	3	3	3	3
4	15	Arica y Parinacota	15201	Putre Total	3	3	3	3	3	3

Finalmente, es importante mencionar que los datos se encuentran agrupados de forma diaria y por comuna, lo cual para fines de análisis permite siempre resguardar la privacidad de las personas, trabajando con datos anonimizados.

## 4.2. Metodología

### Exploración de Datos

Tal como se mencionó inicialmente, el primer desafío a afrontar correspondió a la exploración de datos de los Índices de Movilidad. Para esto fue necesario realizar un ordenamiento de los datos con una estructura que permitiera analizarlos de forma clara y eficiente.

### Ordenamiento

Dado que el conjunto de datos original se encontraba individualizados por el registro diario de los Índices de Movilidad para cada comuna, donde cada fecha constituía una fila diferente para cada comuna, el primer paso fue reorganizar cada uno de los grados de Índices de Movilidad en diferentes conjuntos de datos, y agrupando los registros por comuna, dejando una fila única por comuna y múltiples para cada una de las fechas. Esto debido a que el objetivo de primer análisis exploratorio era visualizar gráficamente las tendencias de las variaciones en la movilidad poniendo foco en el comportamiento por comuna.

	Region	Codigo region	Comuna	Codigo comuna	Superficie_km2	Poblacion	tipo_dia	Fecha	IM
18080	Metropolitana de Santiago	13.0	Alhué	13502	844.43	7405.0	Día laboral	2020-03-02	0.174494
18081	Metropolitana de Santiago	13.0	Alhué	13502	844.43	7405.0	Día laboral	2020-03-09	0.150289
18082	Metropolitana de Santiago	13.0	Alhué	13502	844.43	7405.0	Día laboral	2020-03-16	0.122832
18083	Metropolitana de Santiago	13.0	Alhué	13502	844.43	7405.0	Día laboral	2020-03-23	0.102601
18084	Metropolitana de Santiago	13.0	Alhué	13502	844.43	7405.0	Día laboral	2020-03-30	0.078613

Llegado a este punto, se realizó la primera gran determinación respecto a este estudio, que fue la de acotar el análisis y posterior modelamiento a la Región Metropolitana, esto debido a que permitía reducir la cantidad de comunas a analizar, pero que a su vez permitía trabajar con un amplio espectro de comunas con diferentes características, algunas rurales

y otras urbanas, unas con alta concentración de ingresos y otras con mayor índice de pobreza, u otras con características de comuna dormitorio o bien con alta actividad comercial o industrial.

En este punto, fue posible visualizar también que la variación de movilidad entre los días hábiles y fines de semana o festivos era demasiado significativa y que ya desde un punto de vista del análisis generaba un ruido, sin aportar demasiada información. Dado esto, los promedios semanales de movilidad se estimaron utilizando solo los días de lunes a viernes.

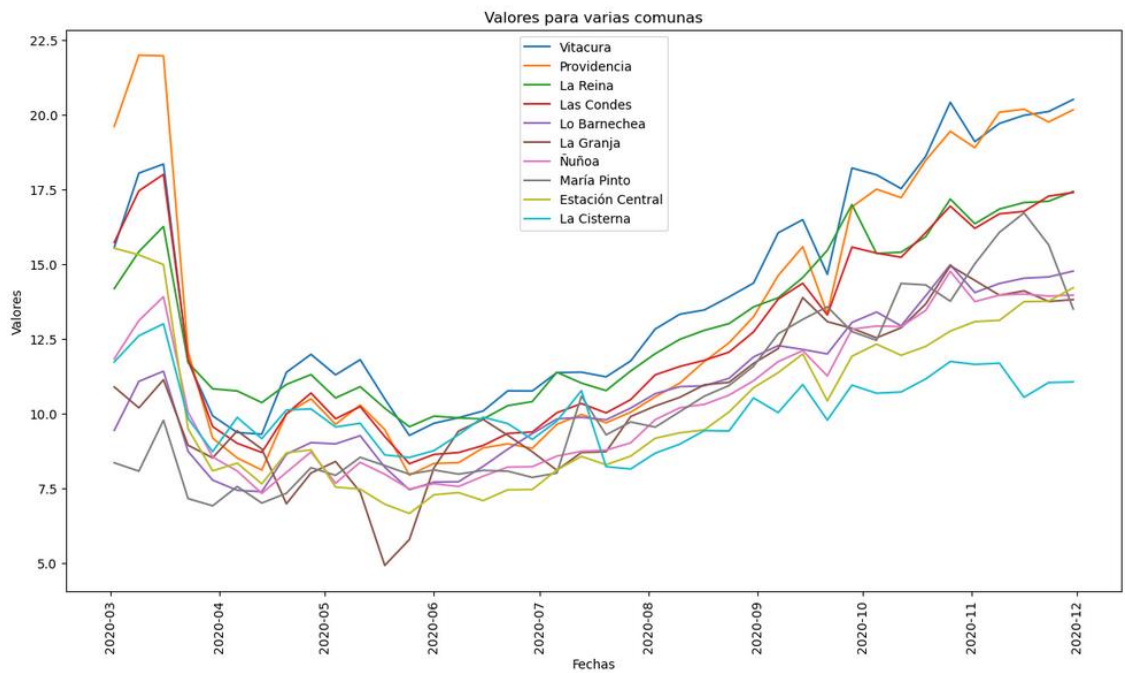
Fecha	2020-03-02	2020-03-09	2020-03-16	2020-03-23	2020-03-30	2020-04-06	2020-04-13	2020-04-20	2020-04-27	2020-05-04	...	2020-09-28	2020-10-05	202
<b>Alhué</b>	0.165643	0.147760	0.124639	0.077673	0.058815	0.062211	0.043550	0.044188	0.065406	0.058053	...	0.080672	0.098459	0
<b>Buín</b>	8.028068	7.759716	7.731720	5.617206	5.464093	6.091349	5.342207	6.154463	6.125326	6.370804	...	7.521142	7.804496	7
<b>Calera de Tango</b>	8.956020	9.160844	9.830933	7.678451	7.887560	9.130335	8.751247	8.952455	8.903095	8.068459	...	9.154609	9.989340	11
<b>Cerrillos</b>	12.487010	12.040197	12.848607	8.157036	7.702067	8.222445	6.489877	8.409857	7.953976	8.726952	...	11.418604	11.443908	11
<b>Cerro Navia</b>	8.214965	8.469082	9.584786	7.187308	6.837717	7.267777	7.541194	7.543055	7.831555	6.717183	...	11.929106	11.682784	11

Dado que el conjunto de datos obtenido al filtrar las comunas de la región metropolitana corresponde a los registros de 52 comunas, se hizo necesario aplicar un criterio que permitiera identificar aquellas comunas que presentaran condiciones que las hicieran interesantes para analizar. Es por esto que el siguiente paso fue estimar los estadísticos básicos tales como los valores máximos, mínimos y el promedio de movilidad.

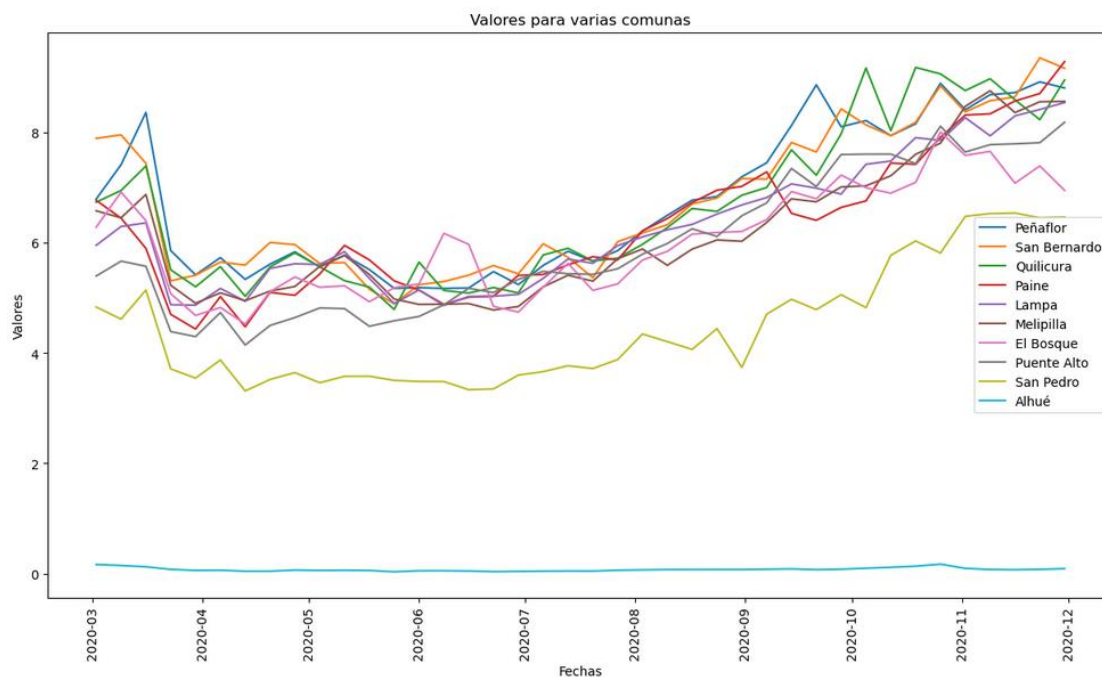
Fecha	mean	min	max	range
<b>Vitacura</b>	14.069556	9.272029	20.513021	11.240991
<b>Providencia</b>	13.330940	7.941017	21.994958	14.053941
<b>La Reina</b>	13.032795	9.563166	17.434681	7.871514
<b>Las Condes</b>	12.473722	8.324781	18.005964	9.681184
<b>Lo Barnechea</b>	10.724647	7.389177	14.984135	7.594958

Otro índice calculado y llamativo para analizar es el rango de variación de movilidad, es decir la diferencia entre el valor máximo y mínimo. Esto debido a que dicho valor

representa una primera aproximación para ver qué tan susceptibles es cada comuna a variaciones en sus índices de movilidad, siempre teniendo en consideración que este valor no se encontraría normalizado y tendría tendencia a ser proporcional a la media de movilidad.

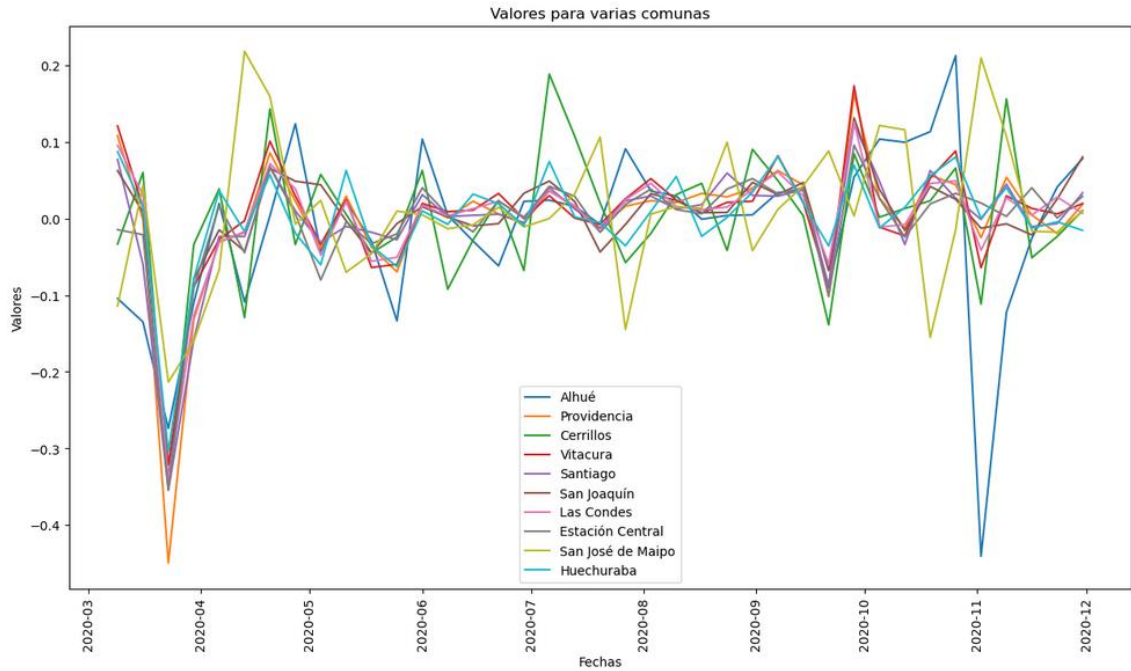


El gráfico anterior, representa a las comunas con mayor movilidad media. En esta primera parte de la exploración de Datos, es posible identificar que entre las comunas con mayor promedio de índice de movilidad se encuentran aquellas del sector oriente de Santiago (Las Conde, Providencia, Lo Barnechea, Vitacura, Ñuñoa y La Reina).

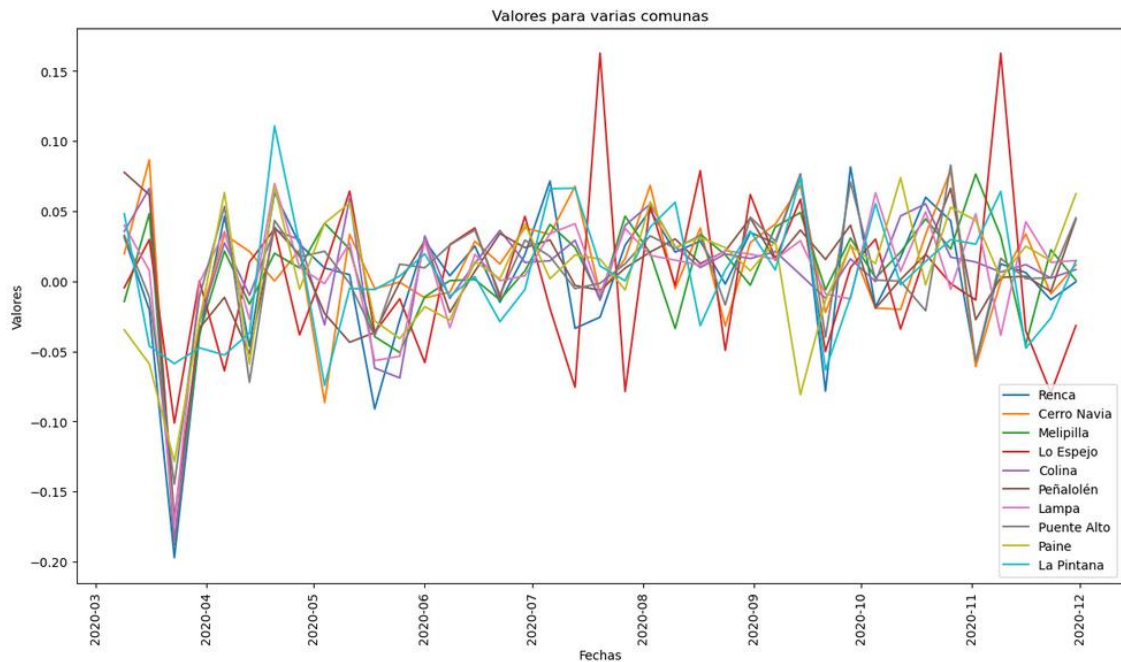


Por otro lado, el gráfico anterior, muestra que entre aquellas con menor movilidad de encuentran comunas de sectores más rurales (San Pedro, Lampa, Melipilla, Alhué y Paine) o comunas de gran número de habitantes, pero orientadas a barrios residenciales (Peñafior, San Bernardo, Quilicura, El Bosque y Puente Alto).

Si embargo, para poder visualizar de mejor manera el comportamiento en las variaciones de movilidad, se procedió a realizar un análisis del rango de variación de movilidad de las comunas, a fin de identificar aquellas que pudieran tener mayor o menor respuestas ante la imposición de restricciones de desplazamiento. Para esto fue necesario normalizar los valores de acuerdo a su valor máximo por comuna, ya que por lógica aquellas comunas que presentan valores más altos de movilidad tendrían mayores probabilidades de obtener valores de variación de movilidad más alta, mientras que, al normalizar los valores, es posible tener un criterio de evaluación del comportamiento de las variaciones de movilidad, independiente del nivel de movilidad que tuviera normalmente cada comuna.



Nuevamente, en este análisis, es posible apreciar que las comunas de Providencia, Vitacura y Las Condes, correspondientes al sector Oriente de Santiago y que se encontraban entre aquellas con mayores índices de movilidad se encuentran entre las comunas con mayor impacto en la movilidad debido a las restricciones de pandemia COVID-19.



Por otro lado, las comunas de Paine, Lampa, Melipilla y Puente Alto, que se encontraban entre aquellas con menores índices de movilidad, vuelven a figurar entre aquellas con menor variación de movilidad producto de las restricciones.

Como punto aparte se encuentra Alhué, comuna que, si bien se encuentra entre aquellas con menor índice de movilidad, figura entre aquellas con mayor impacto en su índice producto de las restricciones, con una significativa baja de movilidad durante el mes de noviembre de 2020.

Para la generación del modelo de Machine Learning, luego de realizar el análisis y estructuración de los datos contenidos en el *Dataset Movilidad*, se debió realizar la preparación de datos y estructuración del segundo conjunto de datos corresponde al *Dataset Paso a Paso*, que contiene el registro de las fases de restricciones definidas para cada día entre el periodo de cuarentena COVID-19 entre 2020 y principio de 2023.

También fue necesario estructurar ambos conjuntos de datos dentro de un mismo período de tiempo y unificarlos en un único dataset, para luego poder generar un conjunto de datos

con la estructura necesaria para los algoritmos de Machine Learning que permitan establecer correlaciones entre las restricciones, características comunales y tendencias en la Movilidad, así como también la correlación entre la Movilidad Interna de cada comuna y su Movilidad Total.

En este caso los datos se encontraban organizados por comuna, con registros de los correspondientes estados de fase por día, donde cada día correspondía a una columna. Producto de esto, los datos se reorganizan dejando cada registro de estado de fase como una fila independiente para cada comuna, es decir, se generó una columna denominada "Fecha" y otra "Valor", en la que se individualizaran los datos mencionados anteriormente.

	<b>Comuna</b>	<b>Fecha</b>	<b>Valor</b>
<b>87</b>	Alhué	2020-07-28	3
<b>477</b>	Alhué	2020-07-29	3
<b>867</b>	Alhué	2020-07-30	3

En este punto, se unificaron ambos conjuntos de datos, para posteriormente filtrarlos entre la fecha de julio y noviembre de 2020, que corresponde al periodo de tiempo común para ambos conjuntos de datos. Luego se identificaron los días hábiles y se eliminaron los fines de semana, a fin de evitar distorsiones en las predicciones de movilidad ante un cambio de fase. Esto se debe a que los fines de semana presentan normalmente una caída en la movilidad y esto no contribuye al entrenamiento de modelo, dado que interfieren en la tendencia de movilidad para días hábiles.

	Comuna	Fecha	Valor	Region	Codigo region	Codigo comuna	Superficie_km2	Poblacion	IM_interno	IM_externo	IM	tipo_dia
153	Alhué	2020-07-28	3	Metropolitana de Santiago	13.0	13502	844.43	7405.0	0.0	0.099440	0.099440	Día laboral
154	Alhué	2020-07-29	3	Metropolitana de Santiago	13.0	13502	844.43	7405.0	0.0	0.093838	0.093838	Día laboral
155	Alhué	2020-07-30	3	Metropolitana de Santiago	13.0	13502	844.43	7405.0	0.0	0.092437	0.092437	Día laboral

Luego se aplicó la estructura definida para ingresar los datos al modelo de Machine Learning, por ejemplo, normalizando valores numéricos o generando dummies para las variables categóricas.

Punto aparte sería lo correspondiente a la agregación de columnas correspondientes a las movilidades y fases de días previos, es importante mencionar que la cantidad de días previos con que se trabaje implica eliminar la misma cantidad de filas de registros por comuna, ya que las primeras filas de cada comuna no podrán contener registros de estados en días previos.

	Comuna	Fecha	Region	Codigo region	Codigo comuna	Superficie_km2	Poblacion	tipo_dia	Valor2	Valor1	Valor
155	Alhué	2020-07-30	Metropolitana de Santiago	13.0	13502	0.16843	0.0	Día laboral	3	3	3
156	Alhué	2020-07-31	Metropolitana de Santiago	13.0	13502	0.16843	0.0	Día laboral	3	3	3
159	Alhué	2020-08-03	Metropolitana de Santiago	13.0	13502	0.16843	0.0	Día laboral	3	3	3

De este modo el esquema del conjunto de datos queda de la siguiente forma:

- 'Comuna': Corresponde a la variable que agrupa los registros y que es el foco de la predicción, ya que se desea predecir los valores de Movilidad por comuna, en respuesta a un cambio de estado de fase en el plan Paso a Paso.
- 'Fecha': Corresponden a los días con registros entre el periodo de julio y noviembre de 2020.
- 'Región': Corresponde al nivel de organización territorial que agrupa a las comunas.
- 'Código región': Identificador numérico para las regiones.

- 'Código comuna': Identificador numérico para las comunas.
- 'Superficie\_km2': Superficie territorial de la comuna.
- 'Población': Población de la comuna.
- 'tipo\_dia': Identificador de día Laborar/Descanso
- 'Valor2': Identificador de Fase Paso a Paso 2 días antes.
- 'Valor1': Identificador de Fase Paso a Paso 1 días antes.
- 'Valor': Identificador de Fase Paso a Paso definido para dicho día.
- 'IM\_interno2': Identificador de IM interno 2 días antes.
- 'IM\_interno1': Identificador de IM interno 1 días antes.
- 'IM\_interno': Identificador de IM interno definido para dicho día.
- 'IM\_externo2': Identificador de IM externo 2 días antes.
- 'IM\_externo1': Identificador de IM externo 1 días antes.
- 'IM\_externo': Identificador de IM externo definido para dicho día.
- 'IM': Variable objetivo que se desea predecir.

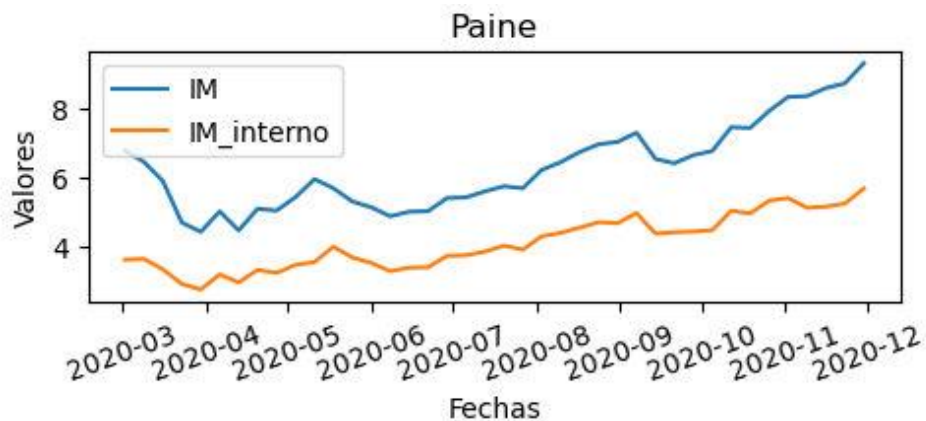
Finalmente, el conjunto de datos es utilizado para entrenar los modelos de predicción mediante los algoritmos de Regresión Lineal y XGBoost

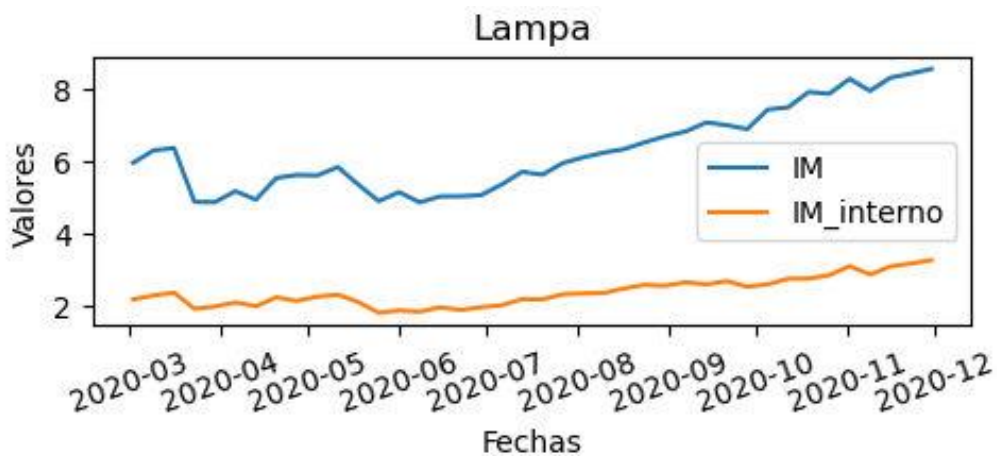
## 5. Resultados

Habiendo procesado los datos a fin de llegar a un resultado que permitiera realizar un análisis respecto al comportamiento de las movilidades durante la pandemia COVID-19, se buscó visualizar las movilidades de las comunas identificadas previamente a dos niveles de escalas de movilidad, la primera a nivel interno de la comuna y a un segundo nivel de movilidad de desplazamientos de entrada o salida en cada comuna. Para esto, se representaron gráficamente que impacto tenían las restricciones en la movilidad interna de cada comuna y como esto afecta la movilidad total de dicha comuna.

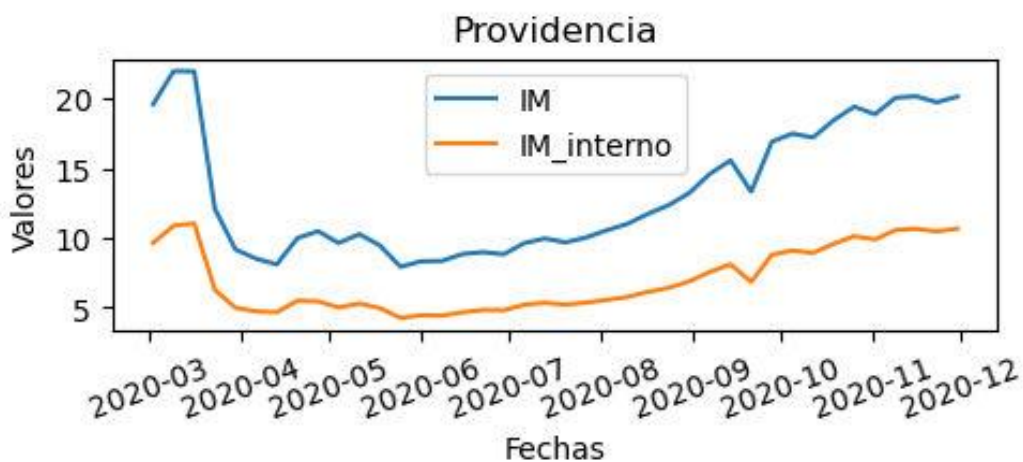
Las comunas a analizadas fueron:

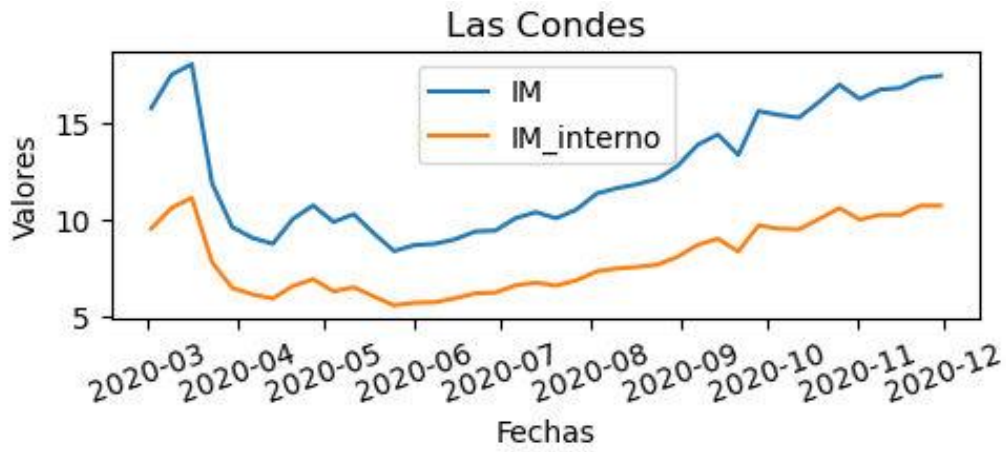
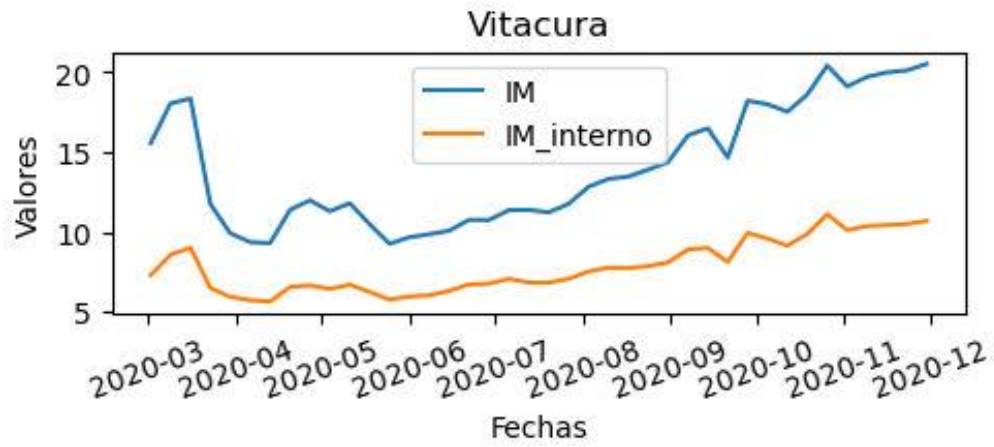
- Providencia
- Vitacura
- Las Condes
- Paine
- Lampa
- Melipilla
- Puente Alto



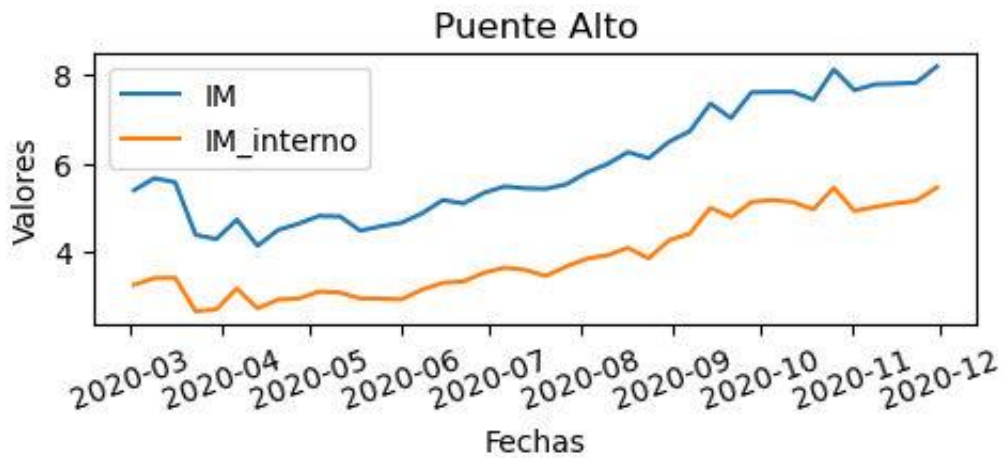
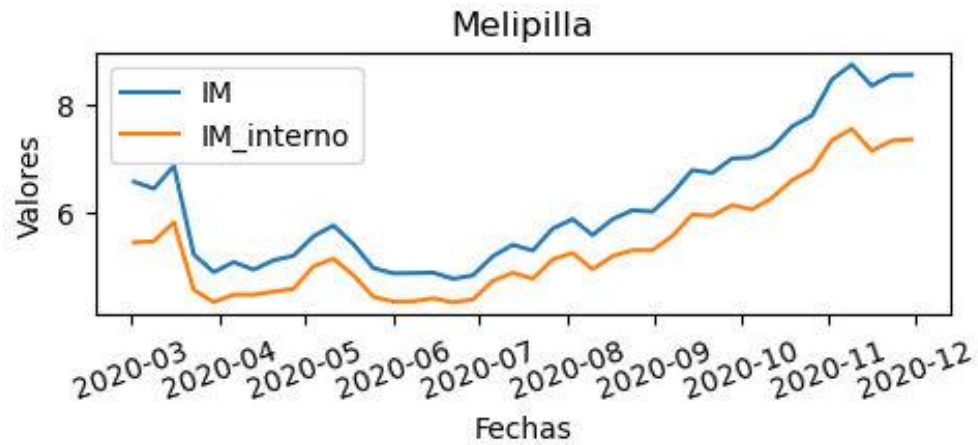


Tal como es posible apreciar en los gráficos de más abajo, las comunas del sector oriente de la Región Metropolitana Providencia, Vitacura y Las Condes, presentan un una diferencia respecto a las variaciones del Índice de Movilidad Interna y el Índice de Movilidad, donde si bien en general ambos decrecen o se incrementan en los mismo periodos, los incrementos o descensos en el Índice de Movilidad son de una proporción mayor al del Índice de Movilidad Interna, donde en definitiva el Índice de Movilidad se ve fuertemente afectado por las variaciones del Índice de Movilidad Externa que sería la diferencia entre el Índice de Movilidad menos el Índice de Movilidad Interna.





Por otro lado, comunas Melipilla y Puente Alto, presentan variaciones con proporciones similares tanto para el Índice de Movilidad como para el Índice de Movilidad Interna, lo que hace suponer que, en estos casos, la Movilidad Externa de estas comunas no se ve afectadas por las cuarentenas.

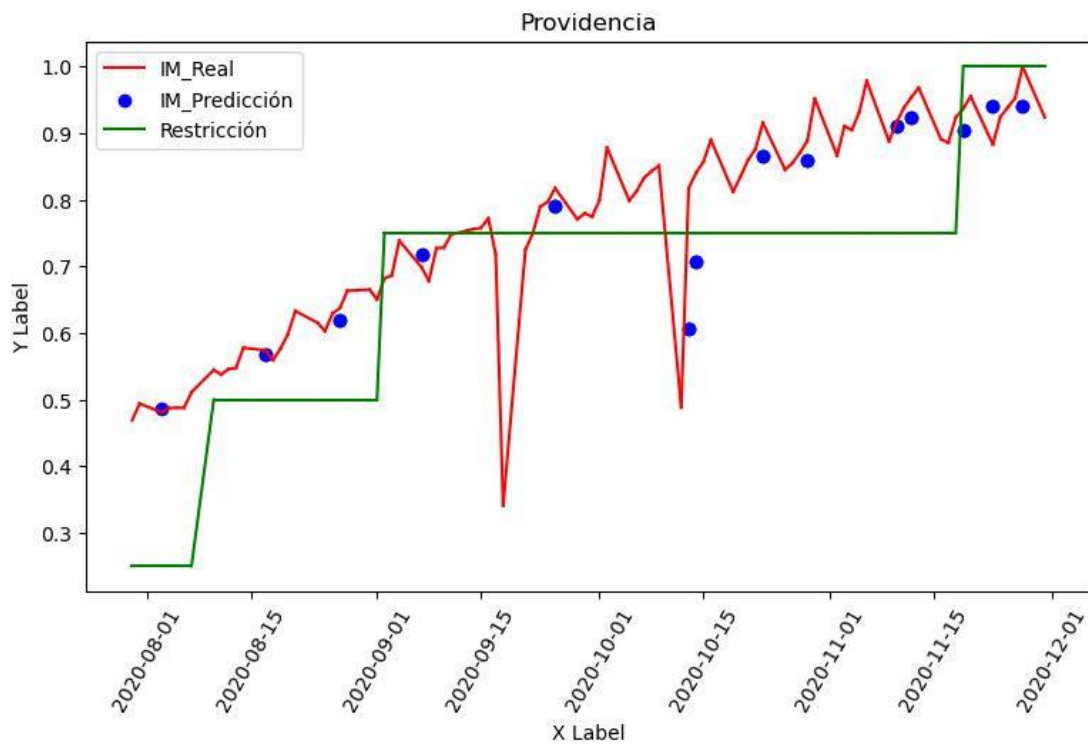


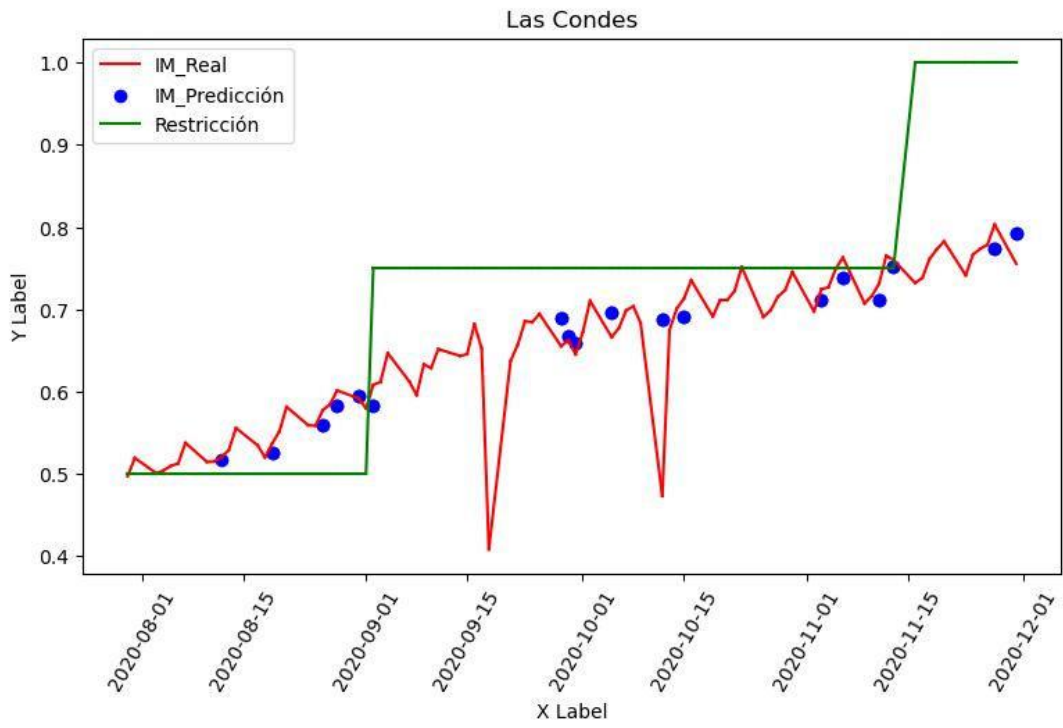
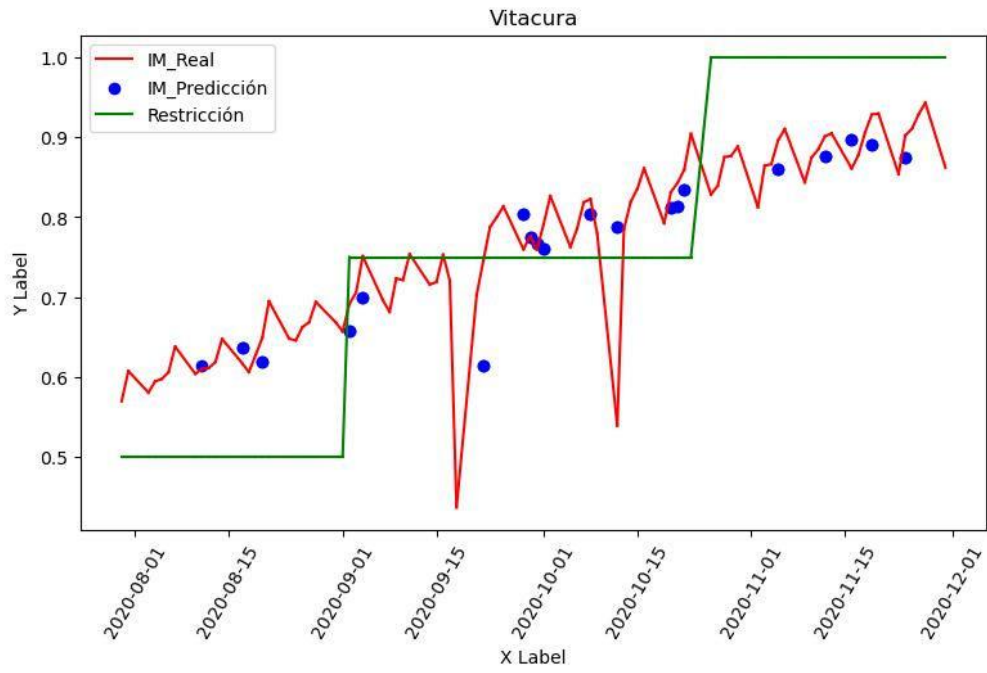
Adicionalmente, se identifica que la magnitud de la respuesta a un cambio de fase se encuentra fuertemente relacionado con el nivel del índice de movilidad existente en cada comuna, donde aquellas que presentan mayor movilidad, también son más propensas a que dichos cambios sean más significativos ante un cambio de fase, por lo cual un modelo predictivo debería considerar las condiciones de movilidad y restricción en los días previos a dicho cambio de fase, normalizando los valores a fin de que estas respuestas puedan ser comparable entre las diferentes comunas independiente de la magnitud del promedio de su índice de movilidad.

Respecto a los resultados obtenidos del modelamiento y predicción de índices de movilidad utilizando modelos de Machine Learning, fue posible obtener un modelo con

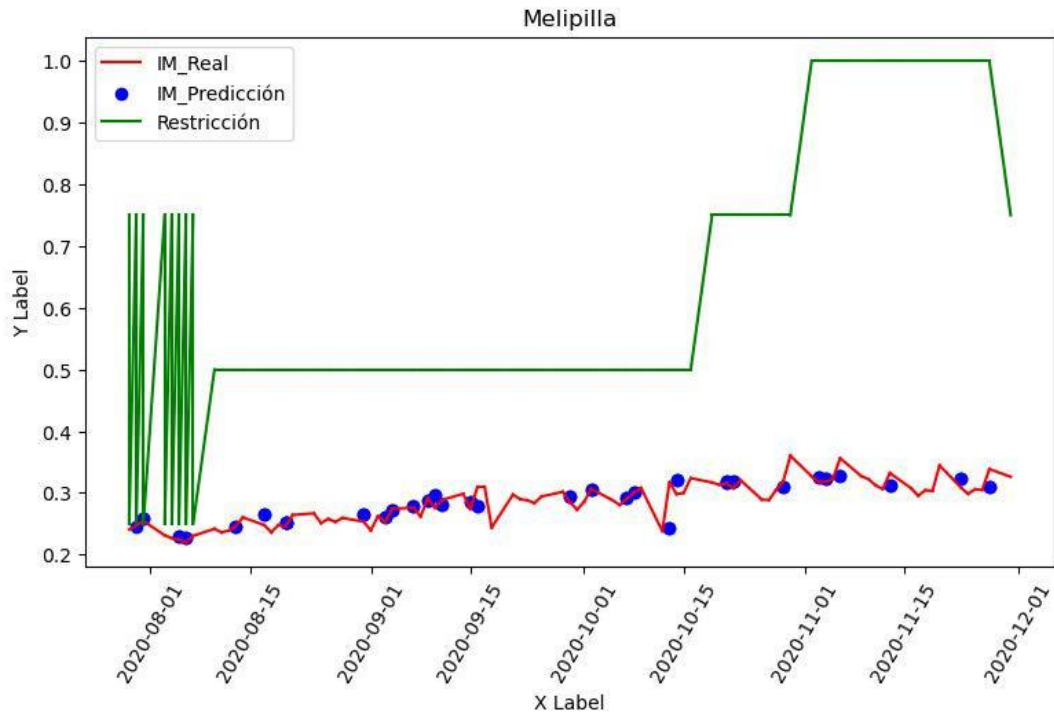
error cuadrático medio (RMSE) con valores inferiores a 0,05, lo cual permite predecir con bastante exactitud el comportamiento de las comunas de la Región Metropolitana.

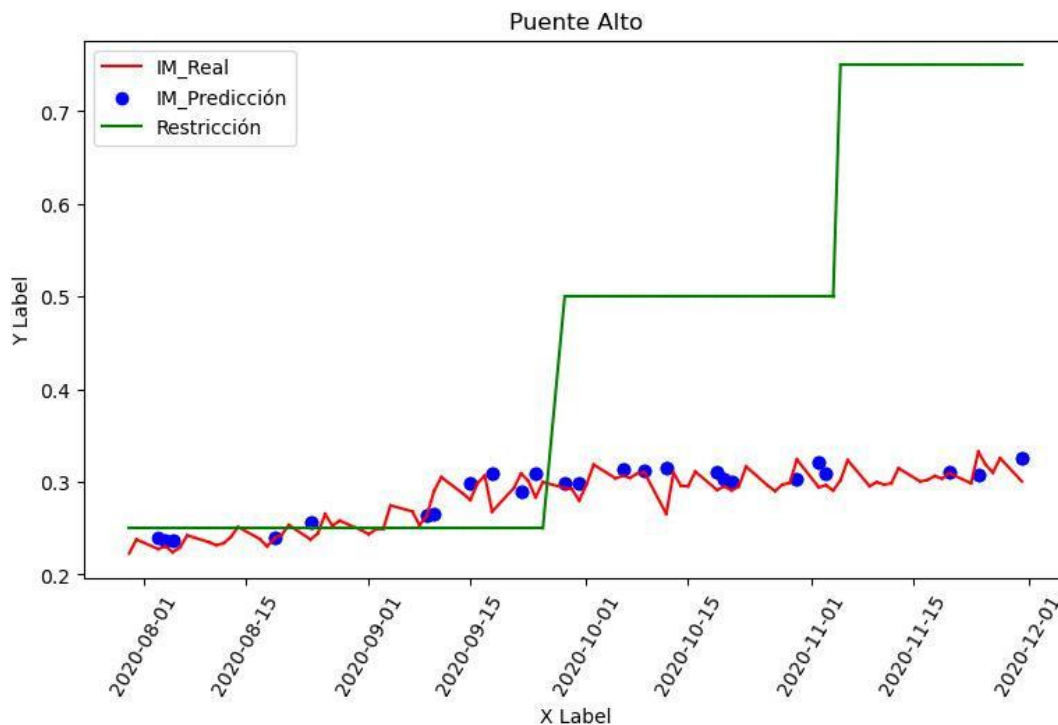
En este punto se presentan los gráficos de las movilidades y de las predicciones realizadas para las comunas de la Región Metropolitana. Para el análisis solo se utilizarán los valores correspondientes a la Regresión Lineal, ya que ambos modelos poseen errores cuadráticos medios similares.





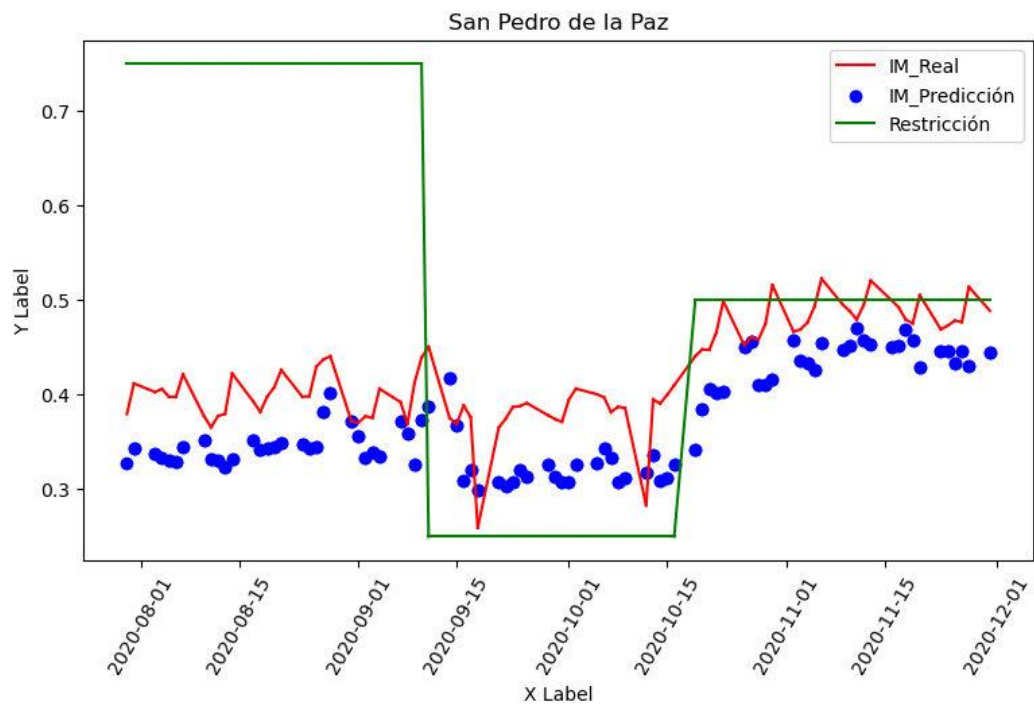
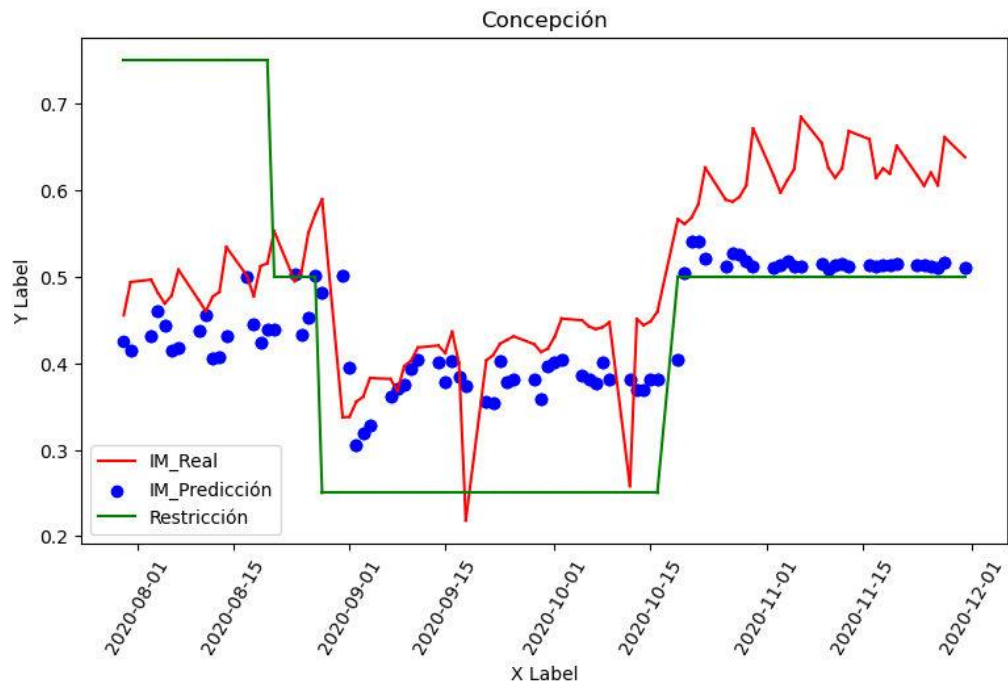
En estos gráficos es posible apreciar que las predicciones coinciden efectivamente con los valores de movilidad para las diferentes fechas usadas para el testeo del modelo obtenido, en este caso de Regresión Lineal. Esto aplica tanto para las comunas del sector oriente graficadas más arriba como Providencia, Vitacura y Las Condes, así como también para aquellas de características más periféricas graficadas más abajo tales como Melipilla y Puente Alto. De este modo es posible apreciar de que el modelo de regresión lineal es capaz de predecir adecuadamente los índices de movilidad para diferentes comunas independientemente de que posean diferentes características.

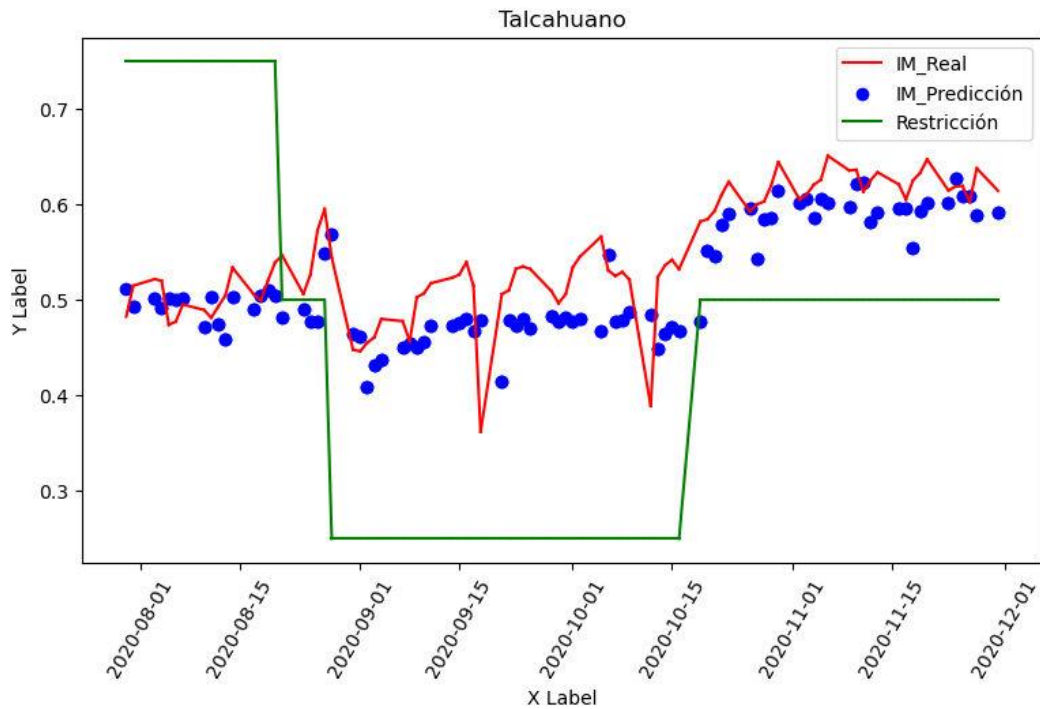




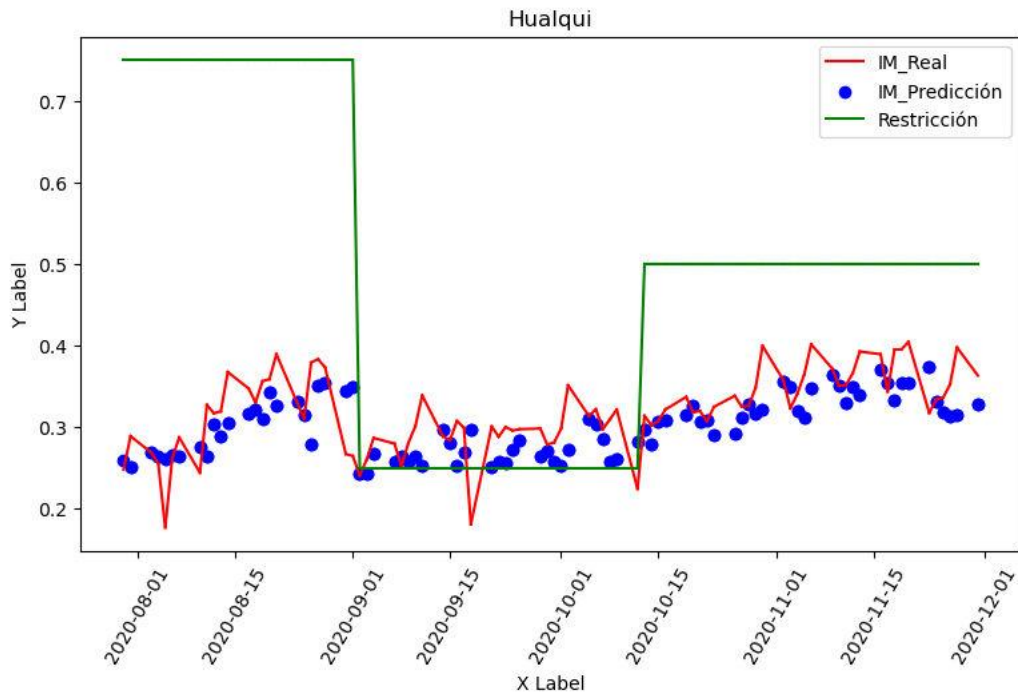
Adicionalmente, se realizaron pruebas de validación del modelo utilizando datos correspondientes a las comunas de la Región del Bio Bío, ya que estos datos no fueron incluidos en el entrenamiento del modelo y que eventualmente dichas comunas pudieran poseer características diferentes a las de la Región Metropolitana, por ejemplo, en cuanto a la relación entre superficie y población.

Otro aspecto importante es que, si bien el Gran Concepción concentra varias comunas en un contexto urbano, la mayoría de las otras comunas no poseen vínculos tan fuertes entre sí, es decir o corresponden a una localidad rural, o bien a una "ciudad-comuna" con una única administración territorial.

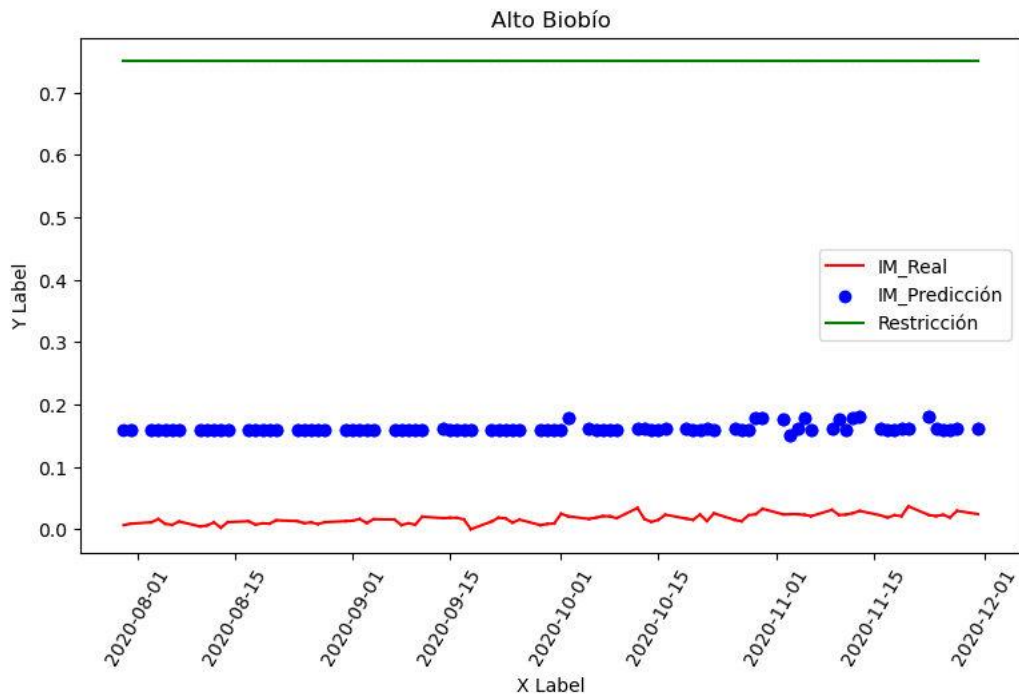
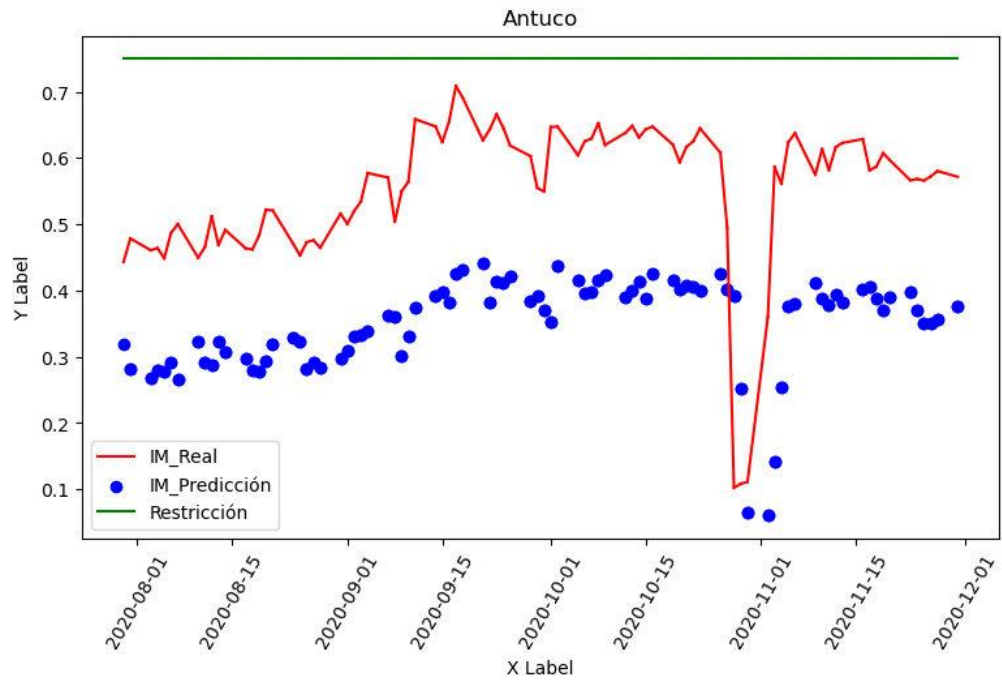




Al observar los gráficos correspondientes a Concepción, San Pedro de la Paz y Talcahuano, es posible apreciar que existe un alto grado de acierto en las predicciones de índice de movilidad para estas comunas. Esto eventualmente podría explicarse debido a que estas comunas son parte del entorno urbano del Gran Concepción, donde eventualmente podrían existir características similares a las de las presentes en las comunas de la Región Metropolitana.



Respecto a Hualqui, es posible apreciar que también existe un alto grado de precisión de las predicciones. Ahora bien, es importante tener presente que esta comuna, si bien es rural, pertenece al entorno urbano del Gran Concepción, pudiendo hacerse cierta analogía con comunas como Melipilla o Lampa para Santiago.



Sin embargo, tal como se esperaba, efectivamente para comunas como Antuco o Alto Bio Bío el escenario es diferente. Si bien es posible apreciar que el modelo es capaz de predecir el comportamiento de la comuna en cuanto a variaciones en los índices de movilidad, existe evidentemente un desfase constante entre los valores predichos y los valores registrados para dichas comunas. Es importante mencionar, que tal como se esperaba, efectivamente comunas como Alto Bio Bío o Antuco, poseen características completamente diferentes a cualquiera de las utilizadas para entrenar el modelo, corresponden a localidades muy rurales, alejadas de entornos urbanos, en sectores cordilleranos y con bastante presencia de comunidades Mapuches.

## 6. Conclusiones

Tal como es posible apreciar en los gráficos, las comunas del sector oriente de la Región Metropolitana Providencia, Vitacura y Las Condes, presentan una diferencia respecto a las variaciones del índice de movilidad interna y el índice de movilidad. Esto nos hace tener una primera apreciación de que la movilidad de las comunas durante las cuarentenas puede encontrarse fuertemente vinculada a condiciones socioeconómicas o bien al nivel de actividad económica que se registran en las diferentes comunas.

Por otro lado, comunas Melipilla y Puente Alto, presentan variaciones con proporciones similares tanto para el índice de movilidad como para el Índice de movilidad interna, lo que hace suponer que en estos casos, la Movilidad Externa de estas comunas no se ve afectadas por las cuarentenas, lo que por ejemplo podría deberse a factores laborales, donde las personas residentes en dichas comunas poseen trabajos o condiciones socioeconómicas en las cuales deben continuar desplazándose de forma normal a sus trabajos, que por ejemplo no pueden realizarse de manera telemática.

De los dos puntos anteriores, se concluye que las diferentes comunas poseen diferentes tipos de correlaciones entre el índice de movilidad externa y el índice de Movilidad, por lo cual se puede intentar generar un modelo que prediga la respuesta los patrones de respuestas de una comuna ante las diferentes fases de cuarentena en utilizando las correlaciones entre las linealidades para los índices de movilidad interna y externa, como las correspondientes respuestas a los cambios de fase de cuarentena.

Respecto al modelo realizado mediante los algoritmos de Machine Learning, tanto de XGBoost como de Regresión Lineal, fue posible apreciar un alto grado de precisión en las predicciones con un Error Cuadrático Medio por debajo de 0,05. Esto, se traduce en que las variables y la metodología de esquematización de los datos previo al entrenamiento del modelo fue el adecuado, permitiendo que el modelo fuera capaz de

predecir la respuesta que tendrían las comunas de la Región Metropolitana ante un eventual cambio en las restricciones de movilidad.

Otra conclusión importante es que, al realizar la validación del modelo, se utilizaron los datos correspondientes a las comunas de la Región del Bio-Bío, fue que el modelo se ajusta muy bien a comunas de características urbanas o rurales cercanas a un entorno urbano, como es el caso de la Región Metropolitana, sin embargo, no posee buenas respuestas para comunas con características de entornos rurales retirados como es el caso de las comunas en sectores más cordilleranos de la región del Bio-Bío, tomando en cuenta que la principal concentraciones urbanas corresponden al Gran Concepción ubicado en un sector costero.

De los dos puntos anteriores también se puede establecer que, conociendo algunas características de una determinada comuna tal como su población o extensión geográfica, los algoritmos de Machine Learning son capaces de establecer relaciones entre el comportamiento de las comunas sin necesidad de ingresar una determinada variable en forma directa, por ejemplo, como la ruralidad o urbanidad de una determinada comuna.

Finalmente, concluir que los resultados obtenidos nos permiten suponer como posible trabajo a futuro y extensión de este trabajo realizado, el complementar los datos con algún otro factor, tal como el promedio de ingresos de los habitantes de las comunas, podría mejorar aún más el modelo de predicción y a su vez permitir afrontar mejor diferentes escenarios ante otras localidades con características diferentes a las comunas utilizadas para entrenar el modelo.

## Bibliografía

1. 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(1). <https://doi.org/10.1038/s41597-022-01893-3>
2. 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(1). <https://doi.org/10.1140/epjds/s13688-021-00284-9>
3. 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(1). <https://doi.org/10.1038/s41467-021-22601-6>
4. Leoferres. (s. f.). *GitHub - leoferres/Datos-COVID19: en formato estándar*. GitHub. <https://github.com/leoferres/Datos-COVID19>