



Universidad del Desarrollo
Facultad de Ingeniería

**IMPACTO DE LAS RESTRICCIONES SANITARIAS DEL COVID EN LA COMISIÓN
DE DELITOS**

Estudio de los efectos colaterales del Covid-19 en la evolución de delitos en la región
metropolitana de Chile en los periodos 2018 a 2022.

POR: BERNABÉ ESCOBAR ARRIAGADA Y OMAR VELOSO OYARCE.

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

PROFESOR GUÍA:

Sr., MAURICIO HERRERA MARIN

Diciembre 2022

LAS CONDES

AGRADECIMIENTO

A nuestras familias con mucho cariño

*A través de los diversos azares, a través de
todas las vicisitudes, cubrimos nuestro
camino...*

Eneida

Índice

3.	RESUMEN	4
1.	INTRODUCCIÓN.....	5
2.	TRABAJO RELACIONADO	6
3.	HIPÓTESIS Y OBJETIVOS.....	8
3.1.	HIPÓTESIS	8
3.2.	OBJETIVOS	8
4.	DATOS Y METODOLOGÍA.....	10
4.1.	DATOS	10
4.2.	METODOLOGÍA	11
5.	RESULTADOS	15
5.1.	PRECEDENTES	15
5.2.	EDA: VARIACIÓN DE DELITOS EN EL TIEMPO - AGLOMERADO RM	16
5.3.	EDA: VARIACIÓN DE UN DELITO Y UNA COMUNA EN PARTICULAR	19
5.4.	EDA: CORRELACIÓN EN LA REGIÓN METROPOLITANA POR TIPO DE CRIMEN	27
5.5.	MODELO NO SUPERVISADO: HIERARCHICAL CLUSTERING	31
5.6.	TEST ESTADÍSTICO: T-STUDENT Y SHAPIRO-WILKS	35
5.7.	REGRESIÓN LINEAL	40
5.8.	MODELO CAUSAL: DoWhy	41
5.9.	MODELO SUPERVISADO: PROPHET	42
5.10.	MODELO SUPERVISADO: ARIMA Y SARIMAX	48
6.	CONCLUSIONES Y TRABAJO FUTURO	52
6.1.	CONCLUSIONES	52
6.2.	TRABAJO FUTURO	53
7.	REFERENCIAS.....	55

Resumen

En el mundo actual, y particularmente en Chile, hemos sido testigos de dos fenómenos que abundan en el contenido en redes sociales y medios de comunicación, como son la pandemia y la delincuencia. En particular, la pandemia, declarada en marzo del año 2020 sorprendió a todas las generaciones, una enfermedad desconocida y de gran propagación llegó para quedarse y detener, momentáneamente, el flujo constante de vida cotidiana.

Ahora, la delincuencia con el pasar de los años transitó de “lanzazos” a “portonazos” o “encerronas”, en octubre del año 2019, se realizaron diversas denuncias por hurto, desorden público o violencia con el inmobiliario público. Una vez llegada la pandemia y con ella las restricciones de movilidad hubo un cambio en cuanto a la actividad y comportamiento delictual, aumentando delitos al interior del hogar.

A la fecha, las medidas de restricción de movilidad se han ido mutando, volviendo a patrones similares a un periodo prepandemia. En el presente trabajo, se expone cómo se ha comportado la delincuencia durante pre, durante y post pandemia dentro de la región metropolitana, buscando una correlación entre restricciones de movilidad y delincuencia.

Se procedió a ranquear y segmentar a cada comuna según su comportamiento a la hora de hablar de criminología y propagación del covid-19. Se proyectó mediante modelos supervisados para poder predecir el estado de la comuna a futuro.

1. Introducción

La pandemia del Covid-19 y las restricciones posteriores en Chile representan una serie de "experimentos naturales" en el cual cambios afectaron a la población en todas sus rutinas, interacciones sociales, y el uso del espacio público, y así también la base de nuestro estudio, a saber, la delincuencia.

Como primer objetivo queremos analizar *la evolución de la pandemia* a través del tiempo en la región metropolitana entre los años 2018 a 2022 y la influencia de la multiplicidad de factores y causas en la propagación del virus. Como objetivo secundario estudiar el impacto de las restricciones del covid-19 sobre *la ocurrencia y comisión de diversos delitos* en la región metropolitana.

En general, por parte de entidades públicas, se realizan investigaciones en cuanto a la causalidad y consecuencias de distintos fenómenos sociales, pero de forma independiente no buscando una correlación entre distintos tipos de eventos que por su naturaleza podrían parecer diferentes o no relacionados.

Finalmente, la pandemia Covid-19 tanto en Chile, como en el mundo, ha ido siendo controlada con inmunización, por efecto rebaño e incluso por aumento en la oferta hospitalaria, donde ha sido capaces de responder a los rebrotes que han surgido, a pesar de ello, nada garantiza que pueda haber una próxima enfermedad o variante, por ende, se vuelve importante entender los fenómenos sociales que implica una pandemia.

2. Trabajo Relacionado

El presente trabajo considera como referencias principalmente tres estudios relacionados al análisis de Covid-19, donde la particularidad de cada país juega un rol fundamental, además de que en Sudamérica las olas de contagios se producían con un desfase en comparación a Europa. Adicionalmente, en Chile hubo un gran porcentaje de habitantes con inmunización por vacunas recibidas con respecto a los vecinos latinoamericanos.

En cuanto a estudios de covid-19 en Chile, (Goic, Bozanic-Leal, Badal, y Basso, 2020) estudia la predicción de ocupación de camas en las Unidades de Cuidados Intensivos (UCI) en tiempos de crisis. Si bien, el data set de uso varía dado que ellos estudian productos de datos basados en el uso de camas UCI y Muertes diarias provocadas por Covid-10, la metodología de interés se basa en la puesta en marcha de modelos predictivos autorregresivos como ARIMAX y TBTS, y Redes neuronales como MLPR, GMDH y ELM.

Por otro lado, en el desarrollo de forecasting comparativo (Saavedra G. 2022), tenemos un análisis del comportamiento de la pandemia más desarrollado por el tiempo transcurrido. En dicho análisis se ven variables como *Personas vacunadas*, *Personas ancianas*, *Muertes diarias*, *Uso de Camas UCI*, Países, entre otros. Además, este análisis usa *SimpleRNN* de la librería *Keras*, y a partir de allí se centra en grupos, United Kingdom, United States, Chile y a nivel Mundial.

Finalmente, diversos estudios se han realizado sobre el Covid-19, pero no así de los efectos que ha causado en distintos ámbitos de estudios, entre ellos encontramos el crimen. En el estudio titulado "*Measuring the impact of the COVID-19 lockdown on crime in a medium-sized city in China*" (Chen P., Kurland J., Piquero A. & Borrion H. 2021) encontramos un análisis sobre las

consecuencias del confinamiento en pandemia sobre los crímenes en China. En el presente estudio se utilizó métodos de Regresión discontinua en el tiempo para ver los efectos del confinamiento, como principales outputs se tienen los efectos inmediatos y a largo plazo según los tipos de crímenes, como ciber fraude, robo de autos, robo de bienes, entre otros.

3. Hipótesis y Objetivos

3.1. hipótesis

Nuestra hipótesis versa sobre las restricciones sanitarias de Covid-19 en la región metropolitana, y cómo impactaron en la forma y comisión de ciertos delitos. De manera tal, que disminuyeron aquellos ilícitos que se cometen regularmente en el exterior de los hogares, sin perjuicio del incremento de manera preocupante de aquellos asociados a la permanencia en los mismos.

Queremos entender qué delitos cayeron y cuales aumentaron en medio de la pandemia en la región metropolitana. Si existe una correlación evidente por tipos de delitos y casos covid-19 en las comunas de la región metropolitana y por último cuáles fueron los efectos de la vacunación en la comisión de delitos en la región metropolitana.

3.2. Objetivos

Los objetivos nacen junto a un problema como que la pandemia del Covid-19 y las restricciones posteriores en Chile representan una serie de "experimentos naturales" en el que los cambios afectaron a la población en todas sus rutinas, interacciones sociales, y el uso del espacio público, y así también la delincuencia.

Teniendo así dos objetivos:

1. Observar la evolución de la pandemia a través del tiempo en la RM y la influencia de la multiplicidad de factores y causas en la propagación del virus.
2. Comprender el impacto de la pandemia sobre la ocurrencia y comisión de diversos delitos en la región metropolitana.

4. Datos y Metodología

Para la prueba o refutación de cualquier hipótesis, se debe conseguir los datos adecuados, poder limpiarlos, procesarlos y descubrir patrones que resulten de interés para el proyecto en sí.

4.1. Datos

En cuanto a los sets de datos tenemos los siguientes:

Bases de crímenes:

Se ingresa al portal del *Centro de estudios y análisis del delito*, se selecciona una por una cada comuna de la región metropolitana como unidad territorial cada una de las tipologías de crimen en un periodo determinado de tiempo.

Se descarga en formato .XLSX y posteriormente se carga, transforma y usa para un análisis en un data set único.

COVID - 19:

A partir de la información disponible en el *Producto 1 estandarizado* del ministerio de salud en su GitHub oficial. En dicho data set se encuentra la cantidad de casos por comuna y por región, esta cantidad de casos tienen una característica incremental. Cada cierto periodo de tiempo el ministerio realizaba una corrección de datos, pudiendo incrementar o no la sumatoria.

4.2. Metodología

Una vez definida la hipótesis, objetivos y alcance del proyecto se inicia con la descarga de información de crímenes, dado que tiene una mayor complejidad que en su manera de obtención (Se tuvo que descargar por cada comuna de la región metropolitana, y la única forma disponible era mediante Excel), una vez que se tiene el conjunto de data sets de crímenes por comuna condensados en una carpeta se incorporan al notebook, se anexan para así tenerlos todos juntos en un solo dataframe, a lo cual posteriormente se le agrega el código de comuna como una nueva columna.

Una vez se tiene el dataframe de crímenes se procede a seleccionar solo tres grupos de delitos según la relevancia en la discusión ciudadana, como son *Delitos de mayor connotación social*, *Abusos sexuales y otros delitos sexuales* y *Violencia intrafamiliar*.

Con el fin de realizar un Análisis Exploratorio de los Datos (EDA) se toma el dataframe de delitos, se agrupa por fecha y se visualizan las series de tiempo que, en la sección de resultados se podrá observar. Posteriormente se cruza con información del periodo de restricciones sanitarias para poder observar el comportamiento de los crímenes con respecto al covid-19, esto se puede realizar de manera agregada por región, provincia o ciudad.

A su vez, se puede sumar y agrupar por comunas en espacios temporales para poder ranquear e inferir cuál tiene más casos de un determinado grupo de crímenes.

Una vez realizada la exploración de datos al data set de crímenes, se trae la base de covid desde el GitHub del Ministerio de Salud y se realiza la limpieza y preparación de datos para

poder unir a la base de crímenes, por ende, deben estar unidas por un mismo ID, en un periodo de tiempo igual.

Finalmente, y con el fin de poder realizar una comparación entre comunas con niveles de población considerablemente distintos, se obtiene la tasa de cada uno de los crímenes y de los casos de covid que se tenían en el dataframe, por ende, se puede realizar comparaciones, correlaciones y mapeo viendo a las comunas de forma homogénea.

Una vez finalizado el proceso EDA, se procede a realizar modelos tanto supervisados como no supervisados. En el caso de los no supervisados, deseamos ver la segmentación de comunas desde dos frentes, los crímenes y el covid. Para darle una mayor validez se compara con el ranking del total de crímenes realizados.

Limpieza de datos como input de los modelos:

Respecto a los datos de covid-19 utilizamos los datos del ministerio de ciencia, tecnología, conocimiento e innovación. En especial el producto 1: El cual contiene los casos totales por comuna incremental. Filtramos los datos por la región metropolitana, y nos damos cuenta que no contamos con todos los días en la serie temporal, para esto unimos fechas completas creadas antes con los datos que si tenemos, luego buscamos dejar la data con casos covid de manera diaria, rellenamos NaNs con valores interpolados para los días que el data set no contenía datos, los casos que sigan teniendo NaNs que son los primeros de cada comuna les asignamos ceros, continuando dividimos los casos interpolados con los datos de la población de cada comuna para tener una medida con un peso más representativo a cada grupo comunal, también completamos todo el resto

de columnas con el método fillna(bfill). Con estos pasos relevantes podemos seleccionar una comuna de estudio para efectos de mejor análisis.

A continuación, el ranking de casos covid acumulados por comuna al 12.12.2022: Puente alto y Santiago tienen por lejos los primeros lugares, pero esto también debido a la gran población de estas comunas:

Imagen 1: ranking de casos covid acumulados por comuna

Comuna	Fecha	Region	Codigo region	Codigo comuna	Poblacion	Casos confirmados	Casos confirmados interpolados	Casos diarios interpolados	tasa_casos_poblacion	
0	Puente Alto	2022-12-12 00:00:00	Metropolitana	13.0	13201.0	645909.0	162230.0	162230.0	1069.333333	165.554797
1	Santiago	2022-12-12 00:00:00	Metropolitana	13.0	13101.0	503147.0	138501.0	138501.0	1368.500000	271.988107
2	Maipu	2022-12-12 00:00:00	Metropolitana	13.0	13119.0	578605.0	128868.0	128868.0	1012.000000	174.903432
3	La Florida	2022-12-12 00:00:00	Metropolitana	13.0	13110.0	402433.0	106229.0	106229.0	714.500000	177.545082
4	Las Condes	2022-12-12 00:00:00	Metropolitana	13.0	13114.0	330759.0	84213.0	84213.0	729.500000	220.553333

Fuente: Elaboración propia

Luego pensamos en que debíamos abrir esto y agregar los días faltantes al data set y rellenar estos días con valores interpolados, y así llegamos a un ranking de casos covid diarios interpolados por comuna al 12.12.2022:

Imagen 2: ranking de casos covid interpolados por comuna

Comuna	Fecha	Region	Codigo region	Codigo comuna	Poblacion	Casos confirmados	Casos confirmados interpolados	Casos diarios interpolados	tasa_casos_poblacion	
0	Santiago	2022-12-12 00:00:00	Metropolitana	13.0	13101.0	503147.0	138501.0	138501.0	1368.500000	271.988107
1	Desconocido Metropolitana	2022-07-18 00:00:00	Metropolitana	13.0	0.0	0.0	24806.0	24806.0	1075.333333	0.000000
2	Puente Alto	2022-12-12 00:00:00	Metropolitana	13.0	13201.0	645909.0	162230.0	162230.0	1069.333333	165.554797
3	Maipu	2022-12-12 00:00:00	Metropolitana	13.0	13119.0	578605.0	128868.0	128868.0	1012.000000	174.903432
4	Las Condes	2022-12-12 00:00:00	Metropolitana	13.0	13114.0	330759.0	84213.0	84213.0	729.500000	220.553333

Fuente: Elaboración propia

Aquí podemos observar que existe un gran volumen de casos diarios para las comunas de Santiago, puente alto y Maipú debido nuevamente a la alta población de estas comunas, sin embargo, también aparece un gran número de casos diarios desconocidos, investigando un poco más en este detalle, se puede indicar que son casos reportados en base a la región de notificación o de toma de muestra y no necesariamente son personas que residen en ella.

Luego queríamos entender cómo se comportan los casos covid, si aplicáramos la variable de población de cada comuna y así tener un orden de magnitud más homogéneo, ya que las comunas que tienen mayor población siempre aparecerán en los primeros lugares y queríamos validar que esto fuera así o no, si consideramos esta variable cada 100.000 habitantes, con esto llegamos al ranking de casos covid diarios interpolados considerando, la tasa de población por 100.000 habitantes por comuna al 12.12.2022: y aquí realmente podemos validar que comunas más pequeñas tales como Alhué y san José de Maipo, se vieron afectadas significativamente por los casos covid.

Imagen 3: Dataframe de tasas covid diarias

	Comuna	Fecha	Region	Codigo region	Codigo comuna	Poblacion	Casos confirmados	Casos confirmados interpolados	Casos diarios interpolados	tasa_casos_poblacion
0	Alhue	2022-12-12 00:00:00	Metropolitana	13.0	13502.0	7405.0	2742.0	2742.0	22.666667	306.099482
1	Santiago	2022-12-12 00:00:00	Metropolitana	13.0	13101.0	503147.0	138501.0	138501.0	1368.500000	271.988107
2	San Miguel	2022-12-12 00:00:00	Metropolitana	13.0	13130.0	133059.0	43063.0	43063.0	333.250000	250.452807
3	San Jose de Maipo	2022-12-12 00:00:00	Metropolitana	13.0	13203.0	18644.0	4791.0	4791.0	46.666667	250.303940
4	La Granja	2022-12-12 00:00:00	Metropolitana	13.0	13111.0	122557.0	30584.0	30584.0	283.500000	231.320936

Fuente: Elaboración propia

5. Resultados

5.1. Precedentes

Como precedente tenemos que las fechas de restricciones de movilidad en **Santiago Centro** son:

- La OMS declaró el brote como una pandemia global el 11 de marzo de 2020.
- Inicio restricción en la Comuna de Santiago: 26/03/2020.
- Fase 1: Santiago, Puente alto, Buin, San Bernardo, Padre Hurtado, Lampa y Colina entraron en cuarentena total el 15/05/2020.
- Salida Fase 1 Santiago: 17/08/2020 Santiago sale de cuarentena total.
- Retrocede Fase 1 Santiago: 20/03/2021 Santiago ingresa nuevamente a cuarentena total hasta el 13/05/2021
- Última restricción de movilidad 05/06/2021 a 01/07/2021

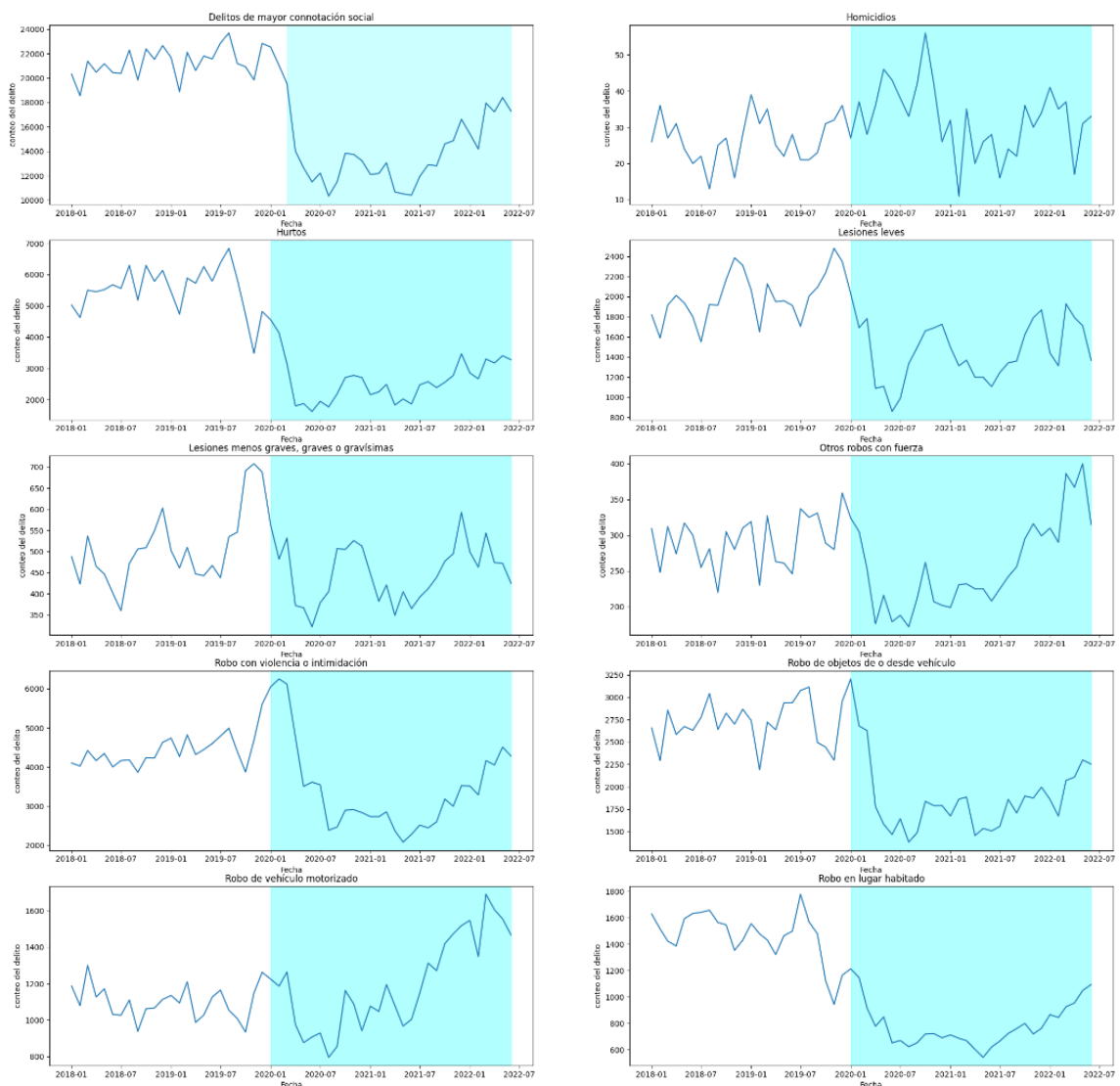
Con respecto a las fechas de restricciones de movilidad en **Puente Alto**:

- Inicio restricción en la Comuna: 09/04/2020 hasta la Salida: 28/09/2020
- Retrocede Fase 1: 25/03/2021 hasta el 29/06/2021

5.2. EDA: Variación de delitos en el tiempo - aglomerado RM

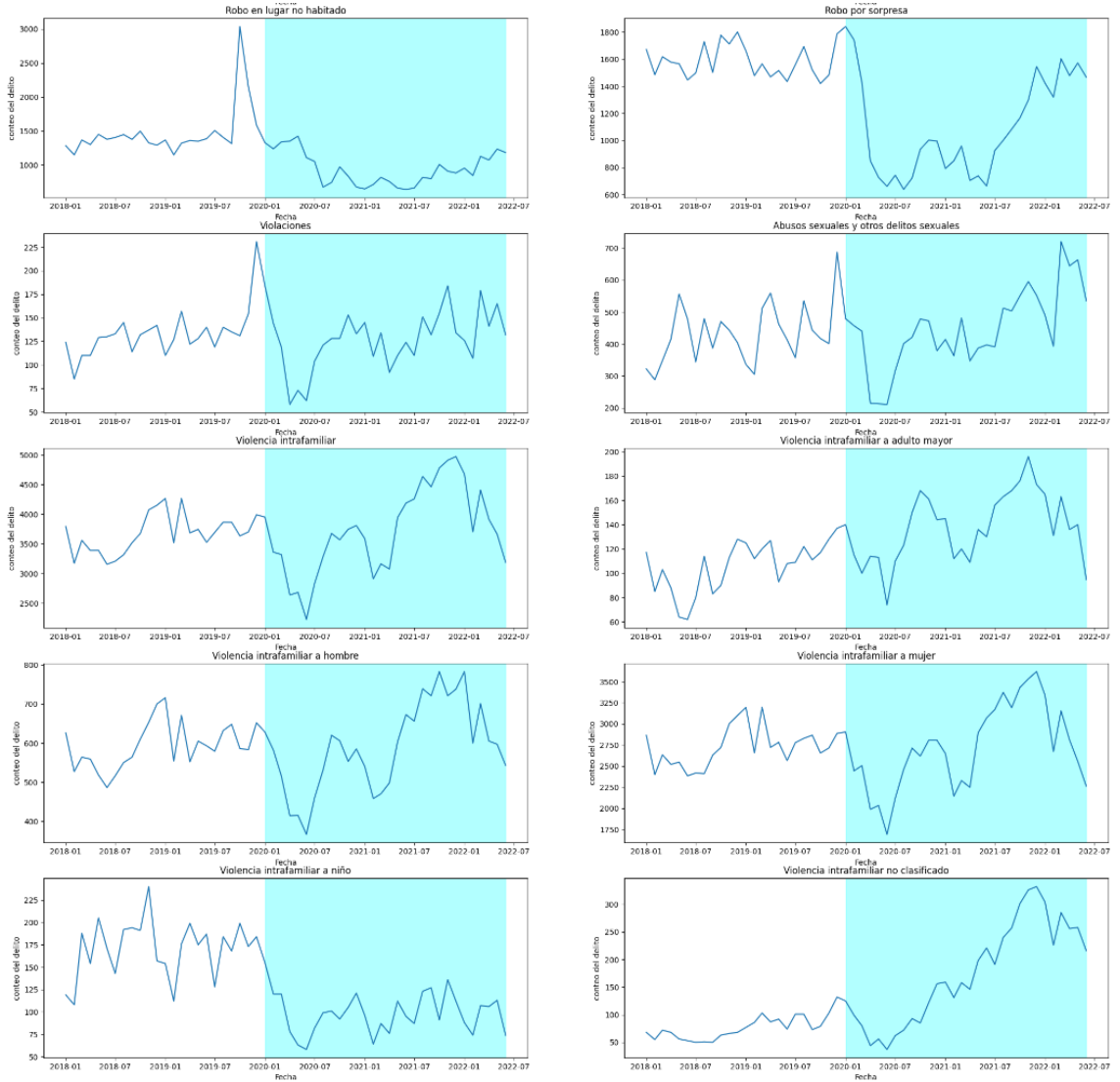
En la *Imagen 4* e *Imagen 5* se logra apreciar en color celeste el periodo durante y post pandemia.

Imagen 4: Serie temporal de delitos



Fuente: Elaboración propia

Imagen 5: Serie temporal de delitos, parte dos



Fuente: Elaboración propia

Cuando queremos visualizar por ejemplo a nivel comunal, las denuncias de “Robo con Violencia” para entender cómo fluctúan en los años 2018 a 2022 y ver si existe cierta tendencia, observamos lo siguiente:

Imagen 6: Serie temporal de Robo con violencia 6 comunas



Fuente: Elaboración propia

Se puede identificar que existen diferencias importantes entre Santiago y Las condes en cuanto a casos denunciados, las magnitudes de Santiago fluctúan entre los 400 a 600 casos mayormente, sin embargo, las condes siempre en menos de 100. Aquí podemos también mencionar que existe una estrecha relación entre lo que invierten de su presupuesto ambas comunas, las condes es la comuna en todo Chile que más invierte en seguridad, en cambio Santiago la sexta con datos al 2018¹³, por otro lado también se puede observar un quiebre en la serie temporal en todas estas comunas para marzo 2020 cuando se comienzan a producir las restricciones de movilidad, se observa una caída en este delito en particular, en las condes es más evidente debido a un mayor cumplimiento de las restricciones sanitarias, Santiago y puente alto al inicio de la pandemia tenían mayor cantidad de personas multadas por no cumplir las restricciones de movilidad, como fue expuesto en un reportaje del diario La tercera¹⁴.

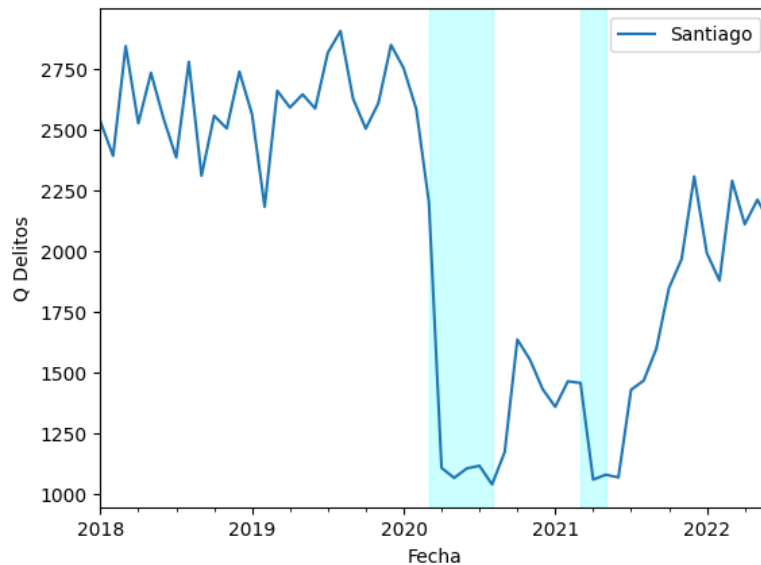
Como principales insights de estas series de tiempo tenemos:

- En general, la mayoría de los crímenes se han visto reducidos desde marzo de 2020.
- Los crímenes de tipo Violencia Intrafamiliar y Homicidio han sido la excepción, durante la pandemia han ido creciendo. Esto puede relacionarse con los niveles de estrés y hacinamiento vividos en la pandemia

5.3. EDA: Variación de un delito y una comuna en particular

Para ver un caso en particular se eligió un delito de mayor connotación social y restricciones de movilidad en la comuna de Santiago centro expuestos en la imagen 6, además se agrega en celeste los períodos con restricciones de movilidad estrictas, donde existían cuarentenas obligatorias.

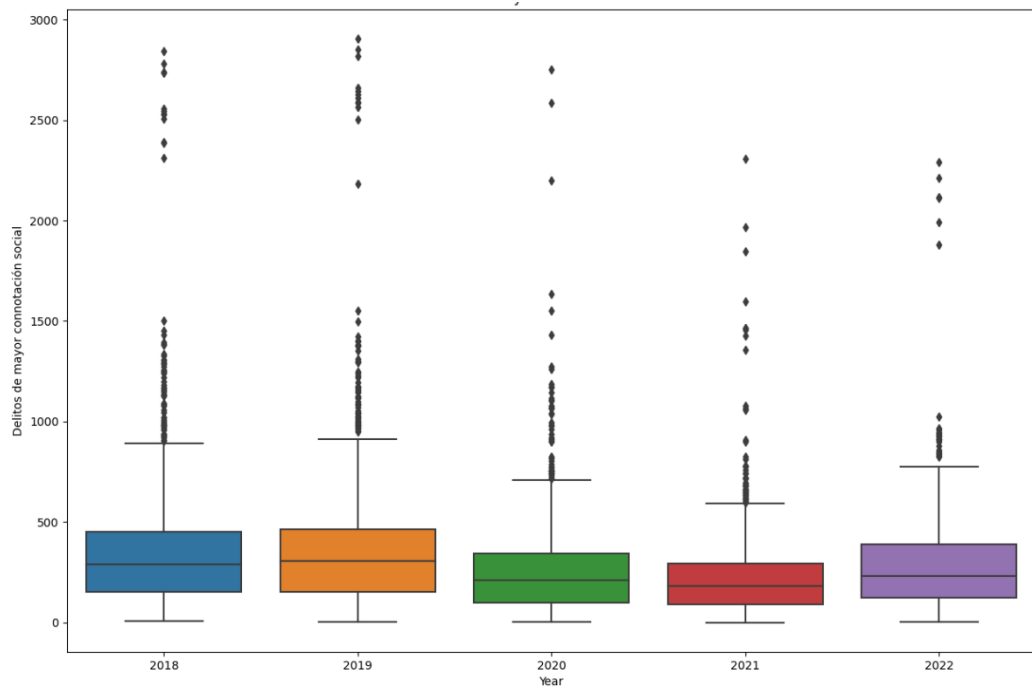
Imagen 7: Delitos de mayor connotación social en Santiago



Fuente: Elaboración propia

Respecto a los outliers de los Delitos de mayor connotación social anualizados podemos visualizar los siguientes movimientos, una caída de la media importante el 2020 y 2021 debido a las restricciones de movilidad y el efecto en este grupo de delitos:

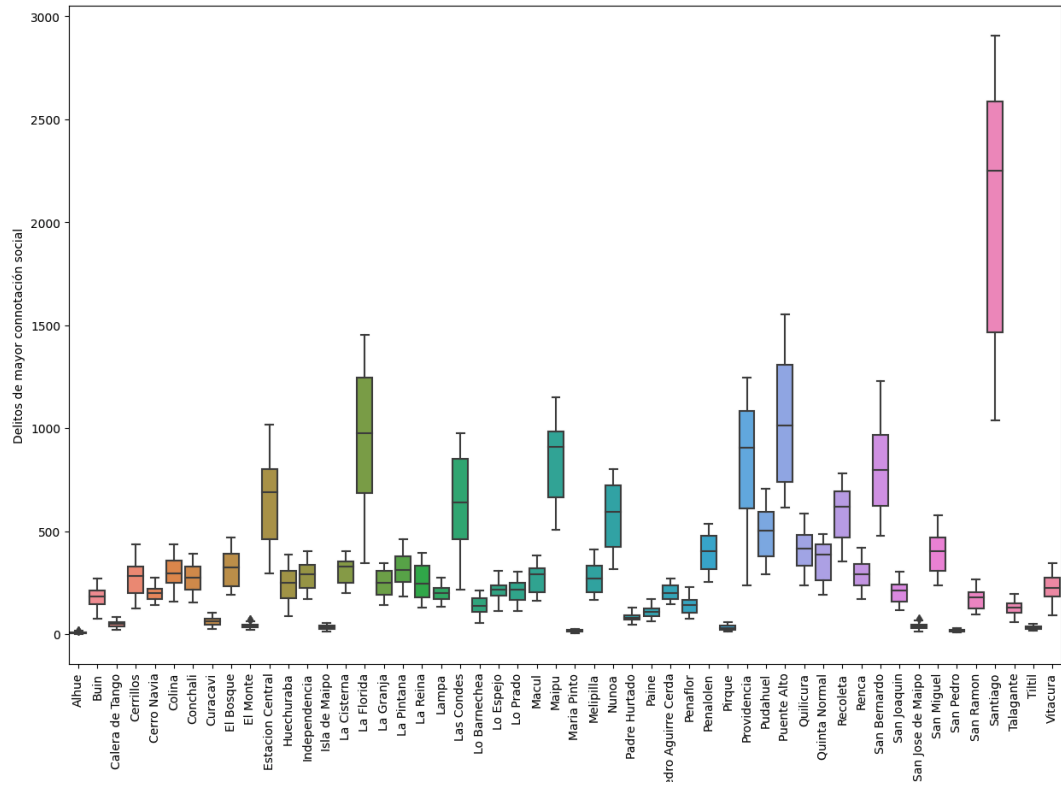
Imagen 8: Outliers delitos de mayor connotación social - anuales



Fuente: Elaboración propia

Por otro lado, tenemos como comunas relevantes en este grupo de delitos a Santiago, Puente Alto, La Florida, Providencia, San Bernardo, y Maipú.

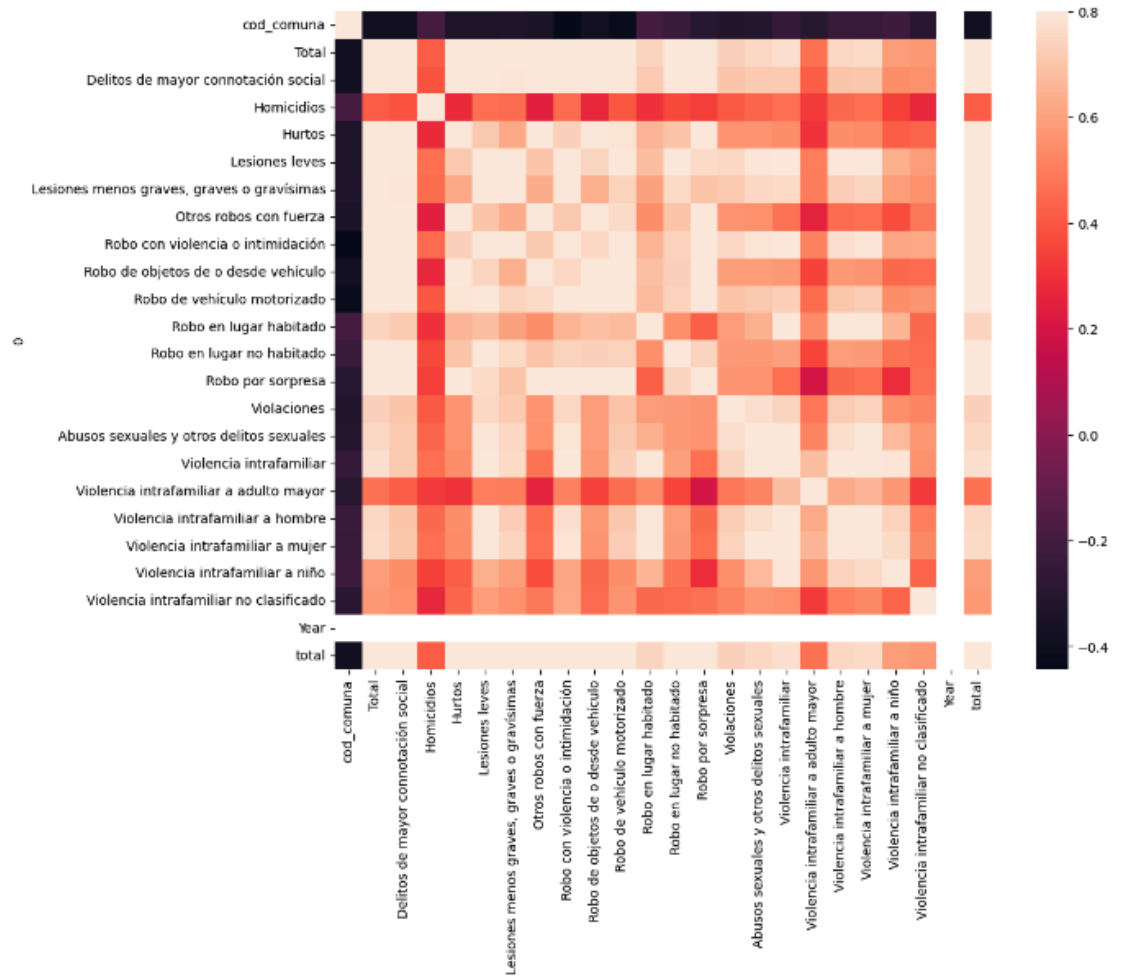
Imagen 9: Outliers delitos de mayor connotación social -comunal



Fuente: Elaboración propia

En cuanto a la correlaciones entre delitos se observa en la siguiente imagen:

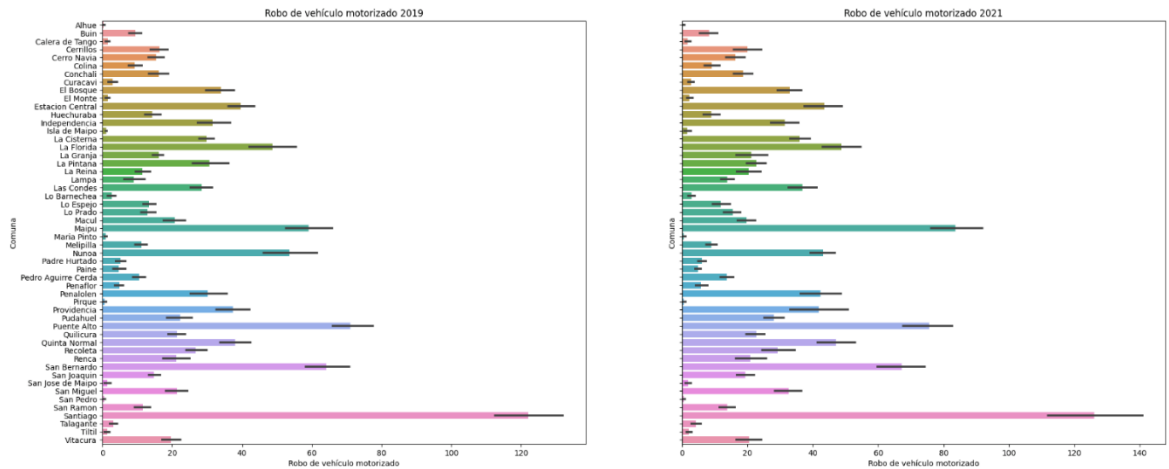
Imagen 10: HeatMap de correlaciones entre tipos de delitos



Fuente: Elaboración propia

Ahora, al estudiar el delito de “Robo de vehículo motorizado” entre el 2019 al 2021 se puede observar un aumento importante en la comuna de Maipú y Puente Alto.

Imagen11: Delitos de robo de vehículos por comuna, variación 2019-2021

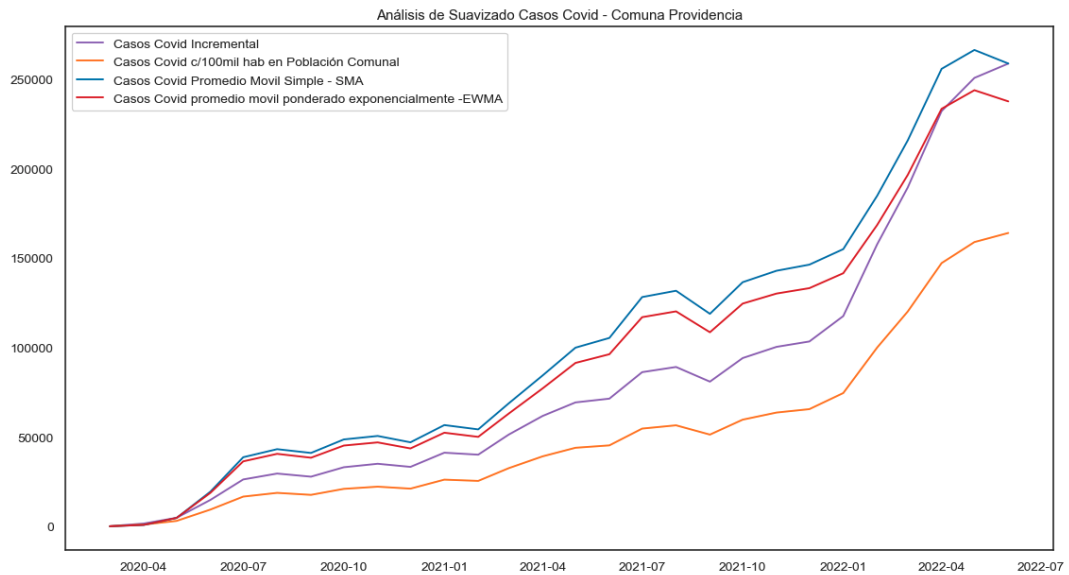


Fuente: Elaboración propia

Por otro lado, en una comuna con mayor población, más alejada del centro, con un mayor nivel de hacinamiento, se puede observar que los crímenes de violencia intrafamiliar crecen en un periodo determinado de restricción de movilidad.

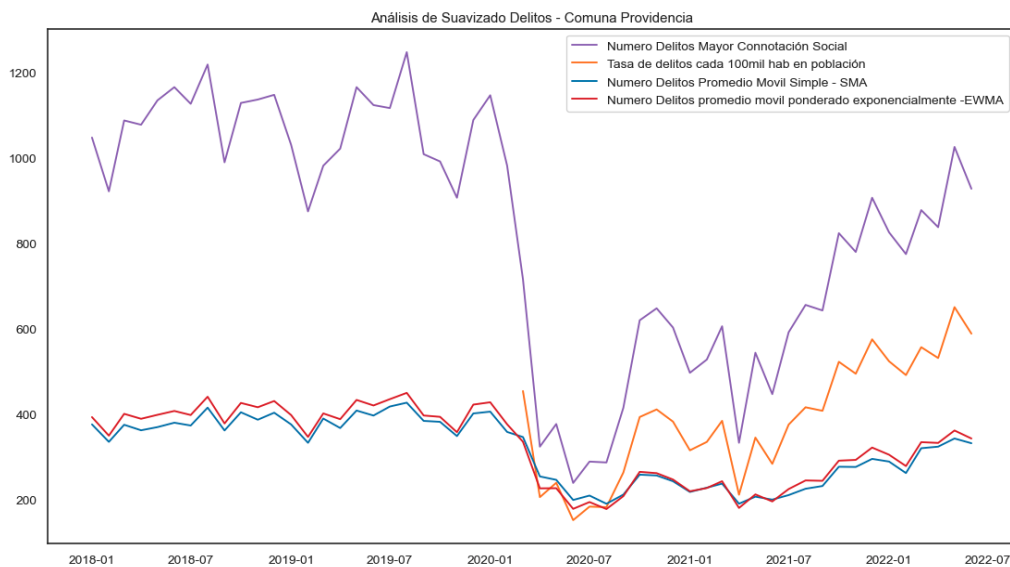
A continuación, veremos un análisis de suavizado tanto para casos covid entre el periodo marzo 2020 a junio 2022 y delitos de mayor connotación social entre enero 2018 a junio 2022, ambos en la comuna de Providencia. El suavizado de datos temporales a menudo revela tendencias o ciclos a largo plazo mientras suaviza el ruido y las fluctuaciones a corto plazo.

Cuando observamos el gráfico de casos covid en la comuna de providencia, se evidencia que la serie temporal de tasa de casos covid cada 100 mil habitantes de la comuna es la más suavizada, pero se mantiene en línea bastante con los casos covid incrementales.



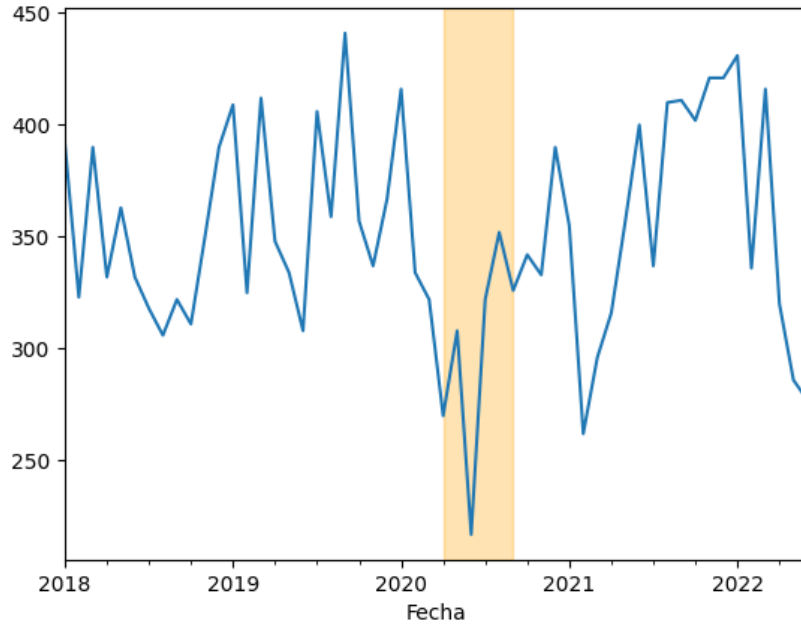
Fuente: Elaboración propia

En el siguiente gráfico se observa la caída abrupta de los delitos de mayor connotación social con la entrada de las restricciones sanitarias en abril de 2020 y también en abril 2021 se observa otra caída importante debido a la cuarenta del 27 de marzo 2021 al 29 de abril 2021:



Fuente: Elaboración propia

Imagen 12: Delitos de mayor connotación social en Santiago



Fuente: Elaboración propia

Como tal, Santiago centro es la comuna que tiene una mayor cantidad de crímenes, pero dada su alta cantidad de habitantes se ve superado por la comuna de Providencia en cuanto a tasa de crímenes. En la siguiente tabla se resume los crímenes de providencia en 2019, 2021 y la desviación porcentual respectiva:

Tabla 1: Comparativo de crímenes en Providencia

Comuna	2019	2021	Variación
Violencia intrafamiliar no clasificado	13	24	85%
Abusos sexuales y otros delitos sexuales	81	96	19%
Violencia intrafamiliar a adulto mayor	15	17	13%
Robo de vehículo motorizado	449	502	12%
Violencia intrafamiliar a mujer	316	304	-4%
Violencia intrafamiliar	450	420	-7%
Violaciones	29	27	-7%
Robo por sorpresa	1022	951	-7%
Otros robos con fuerza	260	240	-8%
Violencia intrafamiliar a hombre	81	70	-14%
Robo de objetos de o desde vehículo	1674	1077	-36%
Delitos de mayor connotación social	12550	7345	-41%
Robo con violencia o intimidación	1305	741	-43%
Robo en lugar no habitado	846	452	-47%
Lesiones leves	562	299	-47%
Homicidios	4	2	-50%
Hurtos	5679	2828	-50%
Lesiones menos graves, graves o gravísimas	167	68	-59%
Robo en lugar habitado	553	158	-71%
Violencia intrafamiliar a niño	25	5	-80%
total	25528	15468	-39%

Fuente: Elaboración propia

Como principal conclusión del cuadro comparativo para dos años en particular para la comuna de Providencia se tiene que delitos que usualmente se realizan en la vía pública como Hurto y Delitos de mayor connotación social se ven drásticamente disminuidos, tanto porcentual como unitariamente.

5.4. EDA: Correlación en la Región metropolitana por tipo de crimen

Comenzamos realizando la importación de los datos de covid-19 desde el producto 1 del GitHub del Ministerio de Salud, los datos están con una frecuencia diaria, y por otro lado los datos de delitos mensuales, por esta razón llevaremos los datos a frecuencia mensual, una vez realizado esto unimos los data set de delitos y los de covid con un merge por los códigos de comuna. Luego limpiamos la data, y realizamos una función llamada: “Correlacion_Comunas” en la cual hacemos un ciclo For para considerar cada comuna y todos los delitos analizados en este proyecto inyectando las correlaciones, por supuesto considerando la variable “casos covid”. Posteriormente creamos variables para cada uno de los delitos y agrupamos las correlaciones.

En el siguiente paso es obtener los mapas vectoriales de la región metropolitana, de la siguiente página logramos obtener todos los archivos requeridos para lograr trabajar el shapefile¹⁵. A continuación, seleccionamos la región metropolitana dentro del archivo y agregamos los datos de las correlaciones de delitos con los casos covid, población, y casos per cápita cada 100.000 habitantes, todo en un solo dataframe.

Finalmente realizamos variadas iteraciones para probar parámetros para la visualización de estos datos (usamos la librería matplotlib y sus diferentes), creamos la infografía y la grilla, para finalmente presentar los siguientes mapas de casos covid cada 100.000 habitantes y correlaciones por comuna en la región metropolitana: Por ejemplo, arriba a la izquierda se puede observar que existen dos leyendas una para las denuncias de los delitos y otra para los casos covid.

1.- En el mapa de “Total Casos Covid cada 100M Hab”, tenemos los casos covid per cápita cada 100.000 habitantes y realizando un ranking observamos que lo lideran las comunas de: Renca, La Granja, La Pintana, y Lo Prado:

Comuna	objectid	shape_leng	dis_elec	cir_sena	cod_comuna	codregion	st_area_sh	st_length_	Delitos de mayor connotación social	Abusos sexuales y otros delitos sexuales	Violencia intrafamiliar	Violencia intrafamiliar a adulto mayor	Violencia intrafamiliar a hombre	Violencia intrafamiliar a mujer	Violencia intrafamiliar a niño	Violencia intrafamiliar no clasificado	Poblacion		
0	Renca	329	28974.956711	9	7	13128	13	3.407127e+07	34707.518216	0.102795	...	0.572384	0.322852	0.070425	0.110190	0.231606	0.315745	0.458349	160847
1	La Granja	312	14001.424778	10	7	13111	13	1.451988e+07	16893.836165	0.151013	...	0.324123	0.428035	0.456885	0.005701	0.446512	0.394035	0.019510	122557
2	La Pintana	313	26302.790697	12	7	13112	13	4.406310e+07	31608.304419	0.122874	...	0.484705	0.427075	-0.076644	0.336715	0.412074	0.281674	0.384069	189335
3	Lo Prado	318	10795.433316	9	7	13117	13	9.412959e+06	12827.159439	0.483294	...	0.362566	0.621080	0.663108	0.440876	0.504967	0.092299	0.368918	104403
4	San Joaquín	330	13987.326781	10	7	13129	13	1.424341e+07	16812.167615	0.185795	...	0.332508	0.518234	0.234686	0.492045	0.450495	0.242576	0.197668	103485

Fuente: Elaboración propia

(Nota: Se pueden observar con un color más morado estas comunas).

2.- Al observar el segundo mapa titulado “Grupo delitos de mayor connotación social” se observan las mayores correlaciones entre este grupo de delitos y el covid, en las siguientes comunas en orden descendente: Providencia, Talagante, Santiago, Las Condes y Buin de esto da cuenta el color más rojizo señalado en leyenda de “Denuncias”.

Comuna	objectid	shape_leng	dis_elec	cir_sena	cod_comuna	codregion	st_area_sh	st_length_	Delitos de mayor connotación social	Abusos sexuales y otros delitos sexuales	Violencia intrafamiliar	Violencia intrafamiliar a adulto mayor	Violencia intrafamiliar a hombre	Violencia intrafamiliar a mujer	Violencia intrafamiliar a niño	Violencia intrafamiliar no clasificado	Poblacion		
0	Providencia	324	16072.408276	10	7	13123	13	2.073083e+07	19223.740933	0.778026	...	0.408149	0.583225	-0.079121	0.488399	0.464292	0.419639	0.574967	157749
1	Talagante	339	69373.905485	14	7	13601	13	1.854226e+08	83423.753031	0.775659	...	0.405199	0.565509	0.586071	0.390180	0.494436	0.120252	0.081899	81838
2	Santiago	333	24644.018619	10	7	13101	13	3.337678e+07	29171.609500	0.692283	...	0.843722	0.618361	-0.127904	0.486492	0.511333	0.318448	0.704649	503147
3	Las Condes	315	51230.563241	11	7	13114	13	1.416658e+08	61618.090148	0.658015	...	0.599338	0.185404	0.184460	-0.018105	0.116910	0.108619	0.480645	330759
4	Buín	294	89486.542349	14	7	13402	13	3.150330e+08	107037.540755	0.651109	...	0.426268	0.333657	0.329842	0.215372	0.239946	-0.060253	0.264196	109641

Fuente: Elaboración propia

3.- El siguiente mapa titulado “Homicidios” muestra una correlación mayor, en orden descendente para las siguientes comunas: Quinta normal, Isla de Maipo, Buin, Santiago, Peñalolén, que son las coloreadas más rojizas en el mapa.

Comuna	objectid	shape_leng	dis_elec	cir_sena	cod_comuna	codregion	st_area_sh	st_length	Delitos de mayor connotación social	Abusos sexuales y otros delitos sexuales	Violencia intrafamiliar	Violencia intrafamiliar a adulto mayor	Violencia intrafamiliar a hombre	Violencia intrafamiliar a mujer	Violencia intrafamiliar a niño	Violencia intrafamiliar no clasificado	Poblacion	
0	Quinta Normal	327	17674.576790	9	7	13126	13	1.708430e+07	21186.419393	0.524182	0.439963	0.633048	0.321473	0.488113	0.572786	0.202238	0.627012	136368
1	Isla de Maipo	336	81883.461053	14	7	13603	13	2.710438e+08	98708.092703	0.643613	0.411785	0.591368	0.286165	0.279664	0.548792	0.181335	0.328549	40171
2	Buin	294	89486.542349	14	7	13402	13	3.150330e+08	107037.540755	0.651109	0.426268	0.333657	0.328842	0.215372	0.239946	-0.060253	0.264196	109641
3	Santiago	333	24644.018619	10	7	13101	13	3.337678e+07	29171.609500	0.692283	0.843722	0.618361	-0.127904	0.486492	0.511333	0.318448	0.704649	503147
4	Peñalolén	323	36618.971695	11	7	13122	13	7.709271e+07	43954.336968	0.396718	0.475833	0.564585	0.393742	0.621246	0.471308	0.013563	0.356631	266798

Fuente: Elaboración propia

Con estas ideas en mente, se pueden enfocar mejoras en las políticas públicas, a fin de planificar para estas comunas en caso de restricciones sanitarias futuras u otros casos de restricciones de movilidad, el apoyo para la prevención o denuncia oportuna de casos de riesgo.

4.- El mapa de “Robo de vehículo motorizado” muestra mayor presencia y correlación de en las comunas de: Providencia, Santiago, Buin, Paine, y Ñuñoa:

Comuna	objectid	shape_leng	dis_elec	cir_sena	cod_comuna	codregion	st_area_sh	st_length	Delitos de mayor connotación social	Abusos sexuales y otros delitos sexuales	Violencia intrafamiliar	Violencia intrafamiliar a adulto mayor	Violencia intrafamiliar a hombre	Violencia intrafamiliar a mujer	Violencia intrafamiliar a niño	Violencia intrafamiliar no clasificado	Poblacion	
0	Puente Alto	292	48961.654840	12	7	13201	13	1.269446e+08	58874.136921	0.324210	0.642108	0.316694	0.107347	0.119399	0.257413	0.296731	0.734600	645909
1	Talagante	339	69373.905485	14	7	13601	13	1.854226e+08	83423.753031	0.775659	0.405199	0.565509	0.586071	0.390180	0.494436	0.120252	0.081899	81838
2	Pedro Aguirre Cerda	322	13054.563524	13	7	13121	13	1.225808e+07	15005.049440	0.119000	0.381737	0.329410	-0.112716	0.134754	0.364179	0.087994	0.356899	107803
3	Maipú	320	61424.982526	8	7	13119	13	1.990876e+08	73734.913427	0.366856	0.621120	0.370795	0.101017	0.366267	0.365908	-0.090557	0.281164	578605
4	Estación Central	18258	13054.563524	8	7	13106	13	2.122054e+07	22973.040873	0.386782	0.598423	0.698410	0.482904	0.468855	0.605808	0.027563	0.828069	206792

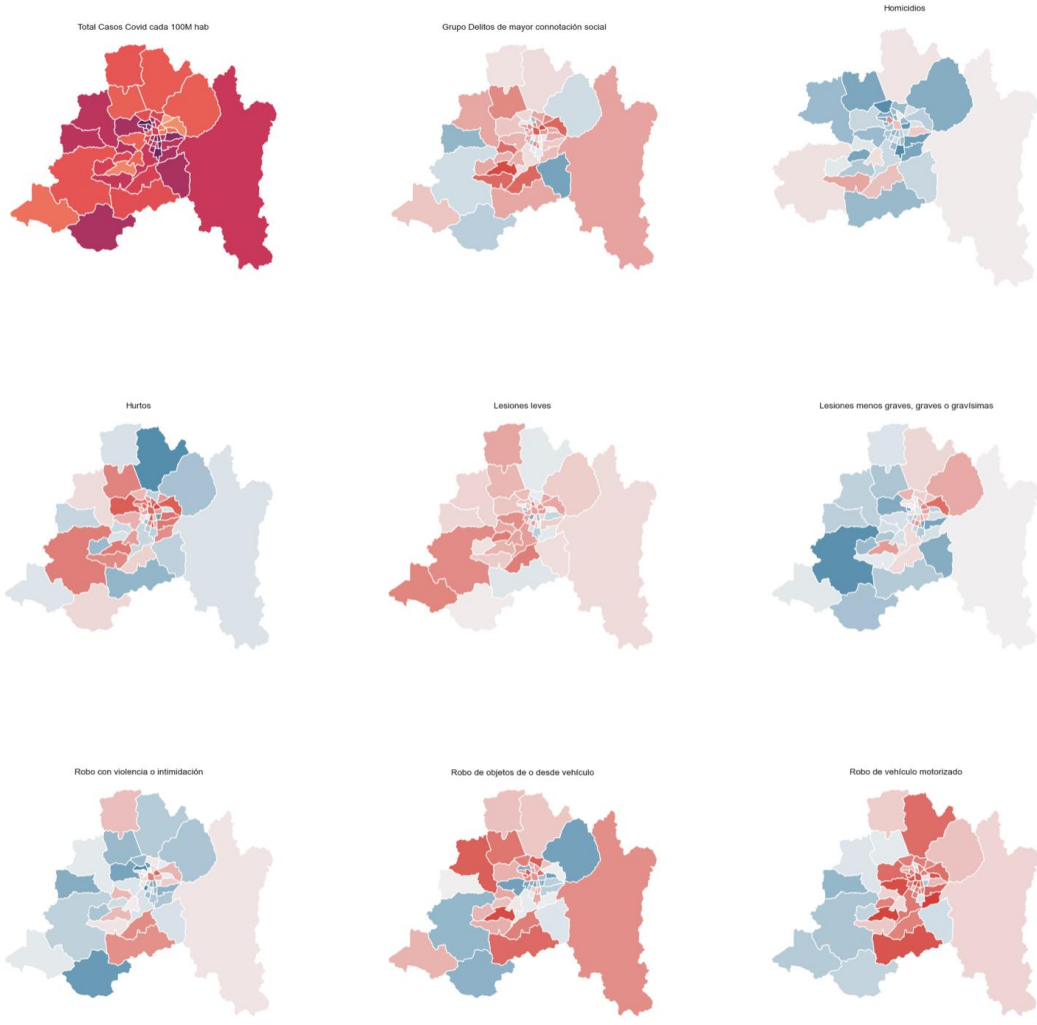
Fuente: Elaboración propia

Imagen 13: Correlación por tasas – Mapa Región metropolitana

Impacto de los casos Covid según tipo de delito por comuna en la RM



En los siguientes mapas podemos observar los casos covid cada 100.000 habitantes en la Región Metropolitana (primer mapa de izquierda a derecha), y en el resto de mapas se observan las correlaciones de los delitos respectivos con los casos covid por comuna.



Fuente: Elaboración propia

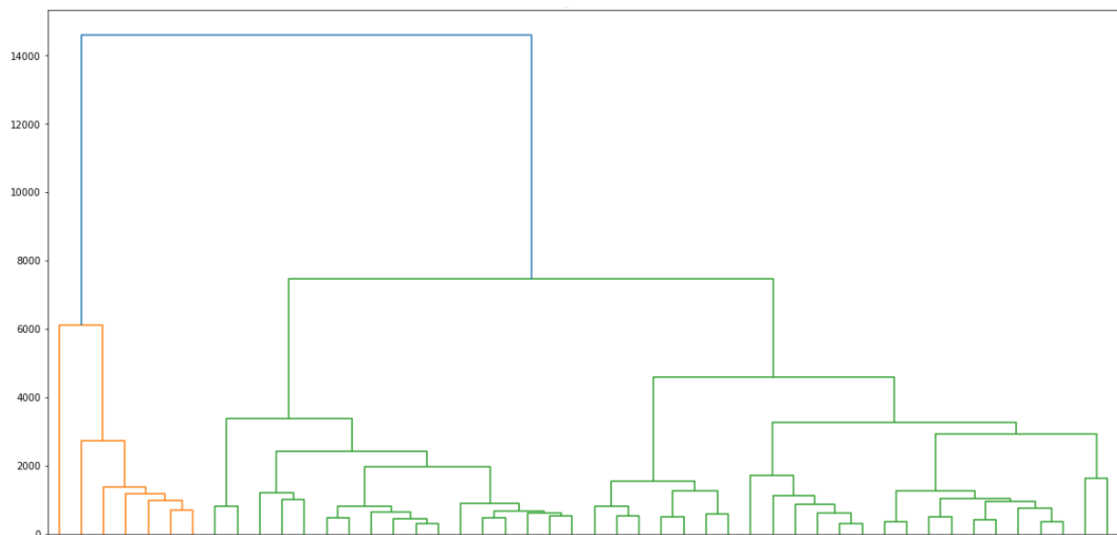
5.5. Modelo no supervisado: Hierarchical Clustering

Se realiza un dendrograma para las tasas de delitos, en esta ocasión se excluyó la columna de Tasa Total, dado que es creada en base a una operación aritmética simple de los delitos.

En este modelo se usó la librería Sklearn, con un método *Ward* o *método Ward de varianza mínima*, este método realiza un clustering en base a un criterio de elección del par de cluster a mezclar según el valor óptimo de una función objetivo.

En el eje horizontal de la imagen 9 se observan las comunas, mientras que en el eje vertical las distancias entre tasas.

Imagen 14: dendrograma delitos 2020-2021



Fuente: Elaboración propia

A simple vista se observan tres grupos principalmente, es importante en términos de gestión del crimen, dado que entrega un segmento de comunas donde se podrían realizar acciones para combatir la delincuencia. En cuanto a los grupos, con el método *cut_tree* de la librería *Scipy* tenemos los siguientes grupos:

Tabla 2: Agrupación de comunas obtenidas por el método cut tree

Grupo 1	Grupo 2	Grupo 3
La Cisterna	Macul	San Pedro
Estación Central	La Florida	Curacaví
Recoleta	San Ramon	San José de Maipo
Cerrillos	San Joaquín	Calera de Tango
San Miguel	San Bernardo	Cerro Navia
Santiago	Renca	Colina
Providencia	Quinta Normal	Talagante
	Pudahuel	Quilicura
	Pedro Aguirre Cerda	Puente Alto
	Lo Prado	Pirque
	Lo Espejo	Isla de Maipo
	La Reina	El Bosque
	La Granja	Paine
	Vitacura	Padre Hurtado
	Independencia	Melipilla
	Conchalí	María Pinto
	Huechuraba	Maipú
		Buin
		El Monte
		Lo Barnechea
		Las Condes
		Lampa
		La Pintana
		Alhué

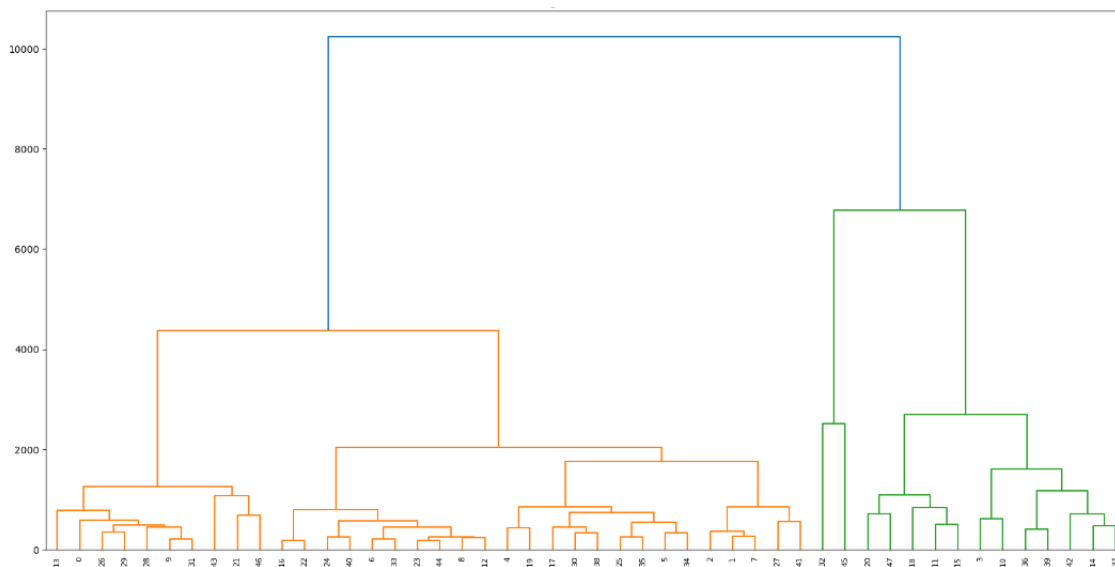
Fuente: Elaboración propia

Tabla 3: Tasas totales de crímenes, Top 7

Comuna	Sumatoria Tasa total de crímenes
Providencia	22.805
Santiago	19.479
La Cisterna	17.303
Recoleta	16.983
Cerrillos	16.713
San Miguel	15.832
Estación Central	15.413

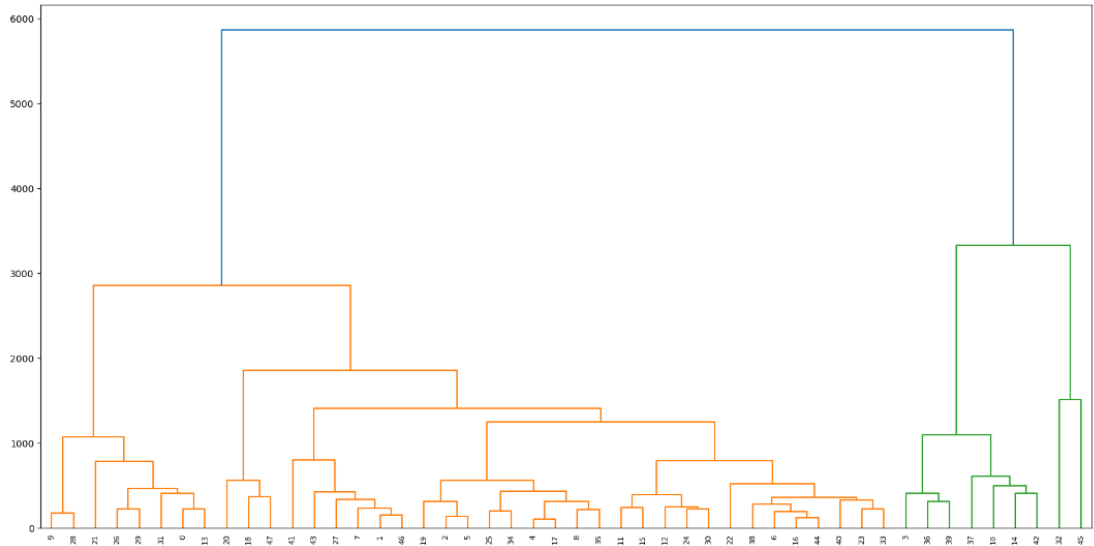
Fuente: Elaboración propia

Imagen 15: dendrograma delitos 2019



Fuente: Elaboración propia

Imagen 16: dendrograma delitos 2021



Fuente: Elaboración propia

Como principales insights del modelo no supervisado tenemos:

- Destacan tres segmentos de delitos.
- Dentro del primer grupo de segmento, tenemos a Providencia encabezando el ranking, lo que resulta poco lógico en primera instancia, pero entendiendo que su población con residencia en la comuna es baja, la tasa se vuelve mayor que la comuna con mayor cantidad de crímenes como Santiago Centro.
- El primer segmento Expuesto en la Tabla 1, concuerda con el top 7 comunas con una mayor cantidad total de crímenes por su respectiva población
- Ha habido un cambio en los Cluster de las diversas comunas, como ejemplo de ello tenemos a Providencia.

5.6. Test estadístico: T-Student y Shapiro-Wilks

Para llegar a concluir si existe o no una diferencia significativa en la cantidad de crímenes entre los años 2019 y 2021 se usó la librería *researchpy*, donde las variables claves del análisis fueron tres: *Tasas de delitos de mayor connotación social*, *Suma de tasa de delito* y *Tasa de robo de vehículos motorizados*.

Las comunas en cuestión para análisis determinados fueron Santiago, Providencia y Puente Alto. Y finalmente, se realiza un Test Shapiro-Wilk para comprobar normalidad

Santiago:

Imagen 17: t-test tasa delitos de mayor connotación social

(Variable	N	Mean	SD	SE	95% Conf.	Interval
0	2019	12.0	522.378814	37.186732	10.734885	498.751492	546.006137
1	2021	12.0	299.630790	75.521769	21.801257	251.646548	347.615033
2	combined	24.0	411.004802	127.799161	26.086895	357.039949	464.969655,
Independent t-test results							
0	Difference (2019 - 2021) =			222.7480			
1	Degrees of freedom =			22.0000			
2	t =			9.1663			
3	Two side test p value =			0.0000			
4	Difference < 0 p value =			1.0000			
5	Difference > 0 p value =			0.0000			
6	Cohen's d =			3.7421			
7	Hedge's g =			3.6131			
8	Glass's delta1 =			5.9900			
9	Point-Biserial r =			0.8902)			

Fuente: Elaboración propia

Test Shapiro con valor-p= 0,077 >0,05

Imagen 18: t-test tasa de Robo de vehículos motorizados

(Variable	N	Mean	SD	SE	95% Conf.	Interval
0	2019	12.0	24.230825	3.718697	1.073495	21.868077	26.593572
1	2021	12.0	25.042383	5.360534	1.547453	21.636463	28.448304
2	combined	24.0	24.636604	4.530842	0.924854	22.723397	26.549811,
Independent t-test results							
0	Difference (2019 - 2021) =		-0.8116				
1	Degrees of freedom =		22.0000				
2	t =		-0.4309				
3	Two side test p value =		0.6707				
4	Difference < 0 p value =		0.3354				
5	Difference > 0 p value =		0.6646				
6	Cohen's d =		-0.1759				
7	Hedge's g =		-0.1699				
8	Glass's delta1 =		-0.2182				
9	Point-Biserial r =		-0.0915)				

Fuente: Elaboración propia

Test Shapiro con valor-p= 0,077 >0,05

Imagen 19: t-test tasa del total de delitos

(Variable	N	Mean	SD	SE	95% Conf.	\
0	2019	12.0	1112.762937	74.285656	21.444422	1065.564083	
1	2021	12.0	676.210597	162.268770	46.842959	573.109940	
2	combined	24.0	894.486767	254.849684	52.020974	786.873184	
Interval							
0	1159.961791						
1	779.311255						
2	1002.100350 ,						
Independent t-test results							
0	Difference (2019 - 2021) =		436.5523				
1	Degrees of freedom =		22.0000				
2	t =		8.4737				
3	Two side test p value =		0.0000				
4	Difference < 0 p value =		1.0000				
5	Difference > 0 p value =		0.0000				
6	Cohen's d =		3.4594				
7	Hedge's g =		3.3401				
8	Glass's delta1 =		5.8767				
9	Point-Biserial r =		0.8749)				

Fuente: Elaboración propia

Test Shapiro con valor-p= 0,086 >0,05

Providencia:

Imagen 20: t-test tasa del total de delitos

(Variable	N	Mean	SD	SE	95% Conf.	\
0	2019	12.0	1348.555828	134.578506	38.849468	1263.048725	
1	2021	12.0	817.120869	217.422852	62.764571	678.976980	
2	combined	24.0	1082.838349	323.954068	66.126847	946.044543	
Interval							
0	1434.062932						
1	955.264759						
2	1219.632154 ,						
Independent t-test results							
0	Difference (2019 - 2021) = 531.4350						
1	Degrees of freedom = 22.0000						
2	t = 7.1995						
3	Two side test p value = 0.0000						
4	Difference < 0 p value = 1.0000						
5	Difference > 0 p value = 0.0000						
6	Cohen's d = 2.9392						
7	Hedge's g = 2.8378						
8	Glass's delta1 = 3.9489						
9	Point-Biserial r = 0.8379)						

Fuente: Elaboración propia

Test Shapiro con valor-p= 0,29 >0,05

Imagen 21: t-test robo de vehículo motorizado

(Variable	N	Mean	SD	SE	95% Conf.	Interval
0	2019	12.0	23.719115	5.608124	1.618926	20.155883	27.282347
1	2021	12.0	26.518921	10.447277	3.015869	19.881038	33.156805
2	combined	24.0	25.119018	8.323870	1.699103	21.604156	28.633880,
Independent t-test results							
0	Difference (2019 - 2021) = -2.7998						
1	Degrees of freedom = 22.0000						
2	t = -0.8180						
3	Two side test p value = 0.4222						
4	Difference < 0 p value = 0.2111						
5	Difference > 0 p value = 0.7889						
6	Cohen's d = -0.3339						
7	Hedge's g = -0.3224						
8	Glass's delta1 = -0.4992						
9	Point-Biserial r = -0.1718)						

Fuente: Elaboración propia

Test Shapiro con valor-p= 0,56 >0,05

Maipú:

Imagen 22: t-test robo de vehículo motorizado

(Variable	N	Mean	SD	SE	95% Conf.	Interval
0	2019	12.0	10.211342	2.065371	0.596221	8.899068	11.523616
1	2021	12.0	14.445664	2.568194	0.741374	12.813911	16.077416
2	combined	24.0	12.328503	3.141947	0.641347	11.001775	13.655230,
Independent t-test results							
0	Difference (2019 - 2021) =		-4.2343				
1	Degrees of freedom =		22.0000				
2	t =		-4.4507				
3	Two side test p value =		0.0002				
4	Difference < 0 p value =		0.0001				
5	Difference > 0 p value =		0.9999				
6	Cohen's d =		-1.8170				
7	Hedge's g =		-1.7544				
8	Glass's delta1 =		-2.0502				
9	Point-Biserial r =		-0.6883)				

Fuente: Elaboración propia

Test Shapiro con valor-p= 0,67 >0,05

Imagen 23: t-test total de delitos

(Variable	N	Mean	SD	SE	95% Conf.	Interval
0	2019	12.0	425.722211	29.377973	8.480690	407.056338	444.388085
1	2021	12.0	301.140963	37.693864	10.881281	277.191424	325.090501
2	combined	24.0	363.431587	71.701599	14.636028	333.154657	393.708517,
Independent t-test results							
0	Difference (2019 - 2021) =		124.5812				
1	Degrees of freedom =		22.0000				
2	t =		9.0304				
3	Two side test p value =		0.0000				
4	Difference < 0 p value =		1.0000				
5	Difference > 0 p value =		0.0000				
6	Cohen's d =		3.6866				
7	Hedge's g =		3.5595				
8	Glass's delta1 =		4.2406				
9	Point-Biserial r =		0.8874)				

Fuente: Elaboración propia

Test Shapiro con valor-p= 0,09 >0,05

Como principales insights de los test y pruebas de normalidad:

- Todos los dataframes obtuvieron pruebas de normalidad no significativas, con un valor-p mayor a 0,05, por ende, tienen una distribución normal y los análisis de la prueba t-student válidos.
- En las tres comunas muestreadas existe una diferencia significativa entre las medias de las tasas del total de crímenes.
- En Santiago y Providencia no existe evidencia para rechazar la hipótesis de no significancia de la diferencia de medias en la tasa de robo de vehículos motorizados.
- En la comuna de Maipú sí existe evidencia suficiente y podemos concluir que hay una diferencia significativa en las medias de las tasas de robo de vehículos motorizados.
- En la comuna de Santiago existe una diferencia significativa entre las medias de las tasas de delitos de mayor connotación social, es decir, ha cambiado la media entre el año 2019 y el año 2022.

5.7. Regresión Lineal

Para la búsqueda de causalidad, se recurrió a la librería Sklearn, utilizaron el modelo de Regresión Lineal, para ello se definieron las siguientes variables, siendo Y la variable dependiente:

Y	X
T_Covid	T_Delitos de mayor connotación social
	T_Homicidios
	T_Hurtos
	T_Lesiones leves
	T_Lesiones menos graves, graves o gravísimas
	T_Otros robos con fuerza
	T_Robo con violencia o intimidación
	T_Robo de objetos de o desde vehículo
	T_Robo de vehículo motorizado
	T_Robo en lugar habitado
	T_Robo en lugar no habitado
	T_Robo por sorpresa
	T_Violaciones
	T_Abusos sexuales y otros delitos sexuales
	T_Violencia intrafamiliar
	T_Violencia intrafamiliar a adulto mayor
	T_Violencia intrafamiliar a hombre
	T_Violencia intrafamiliar a mujer
	T_Violencia intrafamiliar a niño
	T_Violencia intrafamiliar no clasificado

Ahora, para probar la precisión del modelo, se separó el data set en una proporción 80-20 tal que se pueda entrenar y por separado testear. Del data set completo, se agrupó por fecha cada una de las comunas, obteniendo los siguientes resultado, donde se destaca el R^2 :

Imagen 24: Performance del modelo de regresión lineal

```
The model performance for training set
-----
RMSE is 4.722083606231969
R2 score is 0.9586715916001916

The model performance for testing set
-----
RMSE is 15.585519442297144
R2 score is 0.6162184277471097
```

Fuente: Elaboración propia

5.8. Modelo causal: DoWhy

Para este caso se usó el mismo dataset de la regresión lineal, donde, dependiendo de la tasa total de casos covid en la región metropolitana, pueda explicar la causalidad de la tasa total de crímenes en la misma zona.

Para este modelo se usó la librería DoWhy, creando el modelo y sus variables de `Treatmente='T_Covid'` y `Outcome='T_TOTAL'`, se identifica el efecto causal y se entrega el ATE.

Imagen 25: Modelo de causalidad, librería DoWhy

```
# pip install dowhy

from dowhy import CausalModel

model=CausalModel(data = df_r12,treatment='T_Covid',outcome='T_TOTAL')

identified_estimand = model.identify_effect(proceed_when_unidentifiable=True)

causal_estimate = model.estimate_effect(identified_estimand,
method_name="backdoor.linear_regression")

print(causal_estimate)

*** Causal Estimate ***

## Identified estimand
Estimand type: EstimandType.NONPARAMETRIC_ATE

### Estimand : 1
Estimand name: backdoor
Estimand expression:
      d
----- (E[T_Covid])
d[T_TOTAL]
Estimand assumption 1, Unconfoundedness: If  $U \perp \{T\_TOTAL\}$  and  $U \perp T\_Covid$  then  $P(T\_Covid|T\_TOTAL,,U) = P(T\_Covid|T\_TOTAL,)$ 

## Realized estimand
b: T_Covid~T_TOTAL
Target units: ate

## Estimate
Mean value: 82.93847536242902
```

Fuente: Elaboración propia

Como principales insights del modelo de causalidad:

- Identificamos el tratamiento como el covid y el efecto como la tasa total del delitos, se le declara al modelo que realice los cálculos y que identifique las operaciones que se deben realizar dentro del modelo para resolver el problema.
- Podemos concluir que , aproximadamente, el 83% de la tasa de delitos está explicada en la tasa de covid dentro de la región metropolitana para el periodo marzo 2020 en adelante.

5.9. Modelo supervisado: Prophet

Prophet es un procedimiento para pronosticar datos de series temporales basado en un modelo aditivo en el que las tendencias no lineales se ajustan a la estacionalidad anual,

semanal y diaria, además de los efectos de las vacaciones. Funciona mejor con series temporales que tienen fuertes efectos estacionales y varias temporadas de datos históricos. Prophet es resistente a los datos faltantes y los cambios en la tendencia, y por lo general maneja bien los valores atípicos.

A nuestro data set filtrando por la comuna de Santiago, que contiene la serie temporal sin datos faltantes (ya que agregamos los días faltantes e interpolamos recordemos) más la tasa de casos por población le cambiamos los títulos de columnas a “ds” y “y” para que prophet funcione correctamente, creamos y ajustamos el modelo y predecimos como objetivo 20 días, calculamos los pronósticos y guardamos en una variable forecast:

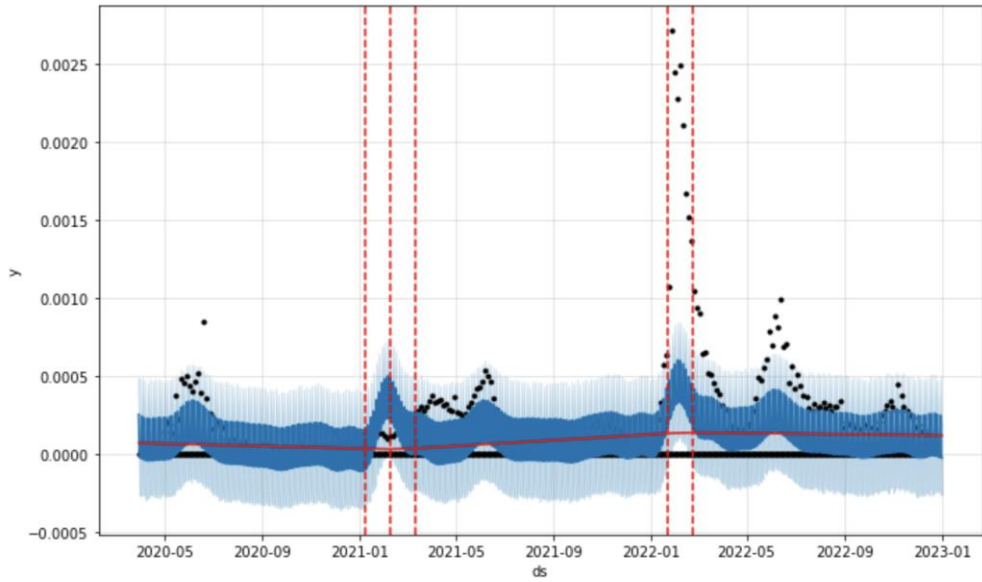
Imagen 26: Dataframe obtenido por Prophet

```
forecast[['ds', 'yhat', 'yhat_lower', 'yhat_upper']].tail(12)
```

	ds	yhat	yhat_lower	yhat_upper
996	2022-12-21	-1.071962	-23.152097	20.747623
997	2022-12-22	-1.291837	-25.954003	23.086720
998	2022-12-23	25.951482	2.604907	51.666645
999	2022-12-24	-1.444286	-24.927678	23.357843
1000	2022-12-25	-1.276046	-24.915857	20.913470
1001	2022-12-26	25.175043	1.974332	48.896440
1002	2022-12-27	-1.703179	-24.829718	22.021328
1003	2022-12-28	-1.962341	-24.545418	20.657895
1004	2022-12-29	-2.342763	-25.566683	20.199358
1005	2022-12-30	24.780212	2.307491	48.489709
1006	2022-12-31	-2.686528	-25.351293	19.253804
1007	2023-01-01	-2.531404	-24.781044	22.237960

Fuente: Elaboración propia

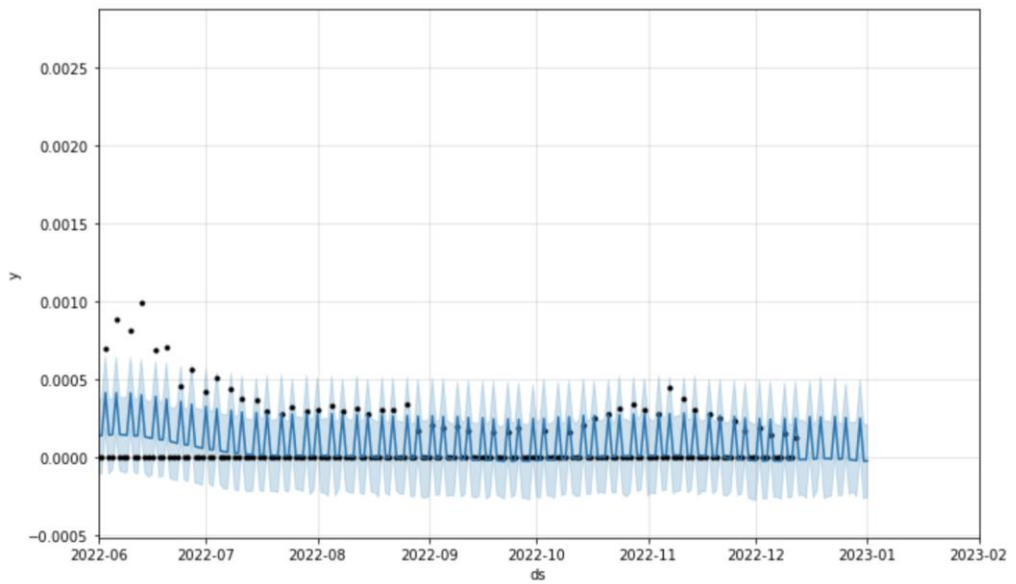
Imagen 27: Forecast obtenido por Prophet



Fuente: Elaboración propia

Realizando un acercamiento de junio 2022 a enero 2023:

Imagen 28: Proyección hasta enero 2023



Fuente: Elaboración propia

Imagen 29: Forecast obtenido por Prophet



Fuente: Elaboración propia

Prophet incluye funcionalidad para la validación cruzada de series de tiempo para medir el error de pronóstico utilizando datos históricos. Esto se hace seleccionando puntos de corte en el historial y, para cada uno de ellos, ajustando el modelo usando datos solo hasta ese punto de corte. Luego podemos comparar los valores pronosticados con los valores reales.

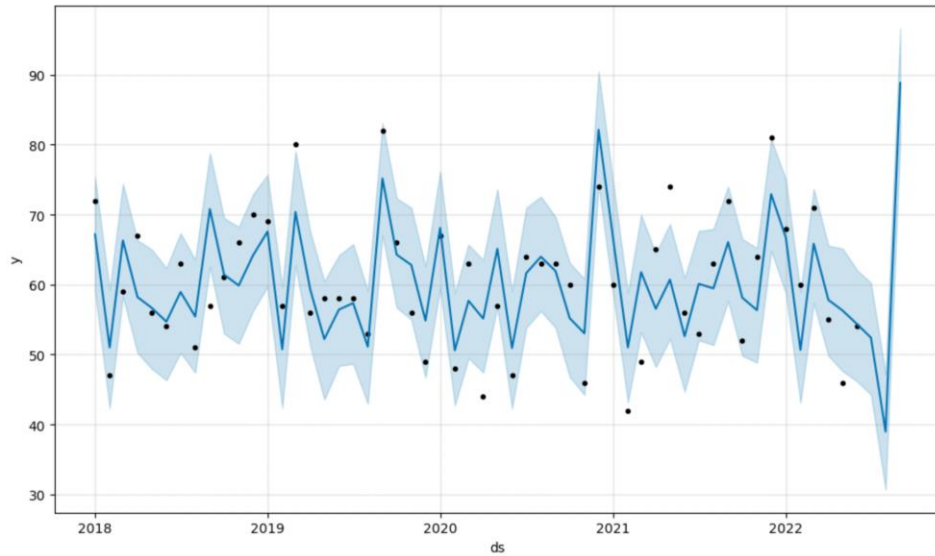
Realizando cross-validation aplicamos un año de entrenamiento fijo, luego hicimos predicciones cada 100 días de 100 días de ventana de predicción sucesivamente hasta que los datos se agotaron, así y con el output llegamos a la conclusión de que el modelo más eficiente es el aditivo con parámetros por default.

Imagen 30: Resultados Cross-validation

```
{'changeoint_prior_scale': 0.001,  
'seasonality_prior_scale': 10.0,  
'seasonality_mode': 'additive'}
```

Fuente: Elaboración propia

Finalmente, y posterior a esto realizamos la predicción del delito de 'Violencia intrafamiliar' en la comuna de puente alto (debido a que tenía la correlación más alta), esto considerando el dataset mensual desde enero 2018 a junio 2022, encontrando los siguientes resultados: Un error RMSE de 1 delito con un pronóstico a 60 días lo que equivale a un MAPE de 1,9% lo que refleja uno de los mejores resultados identificados hasta el momento:

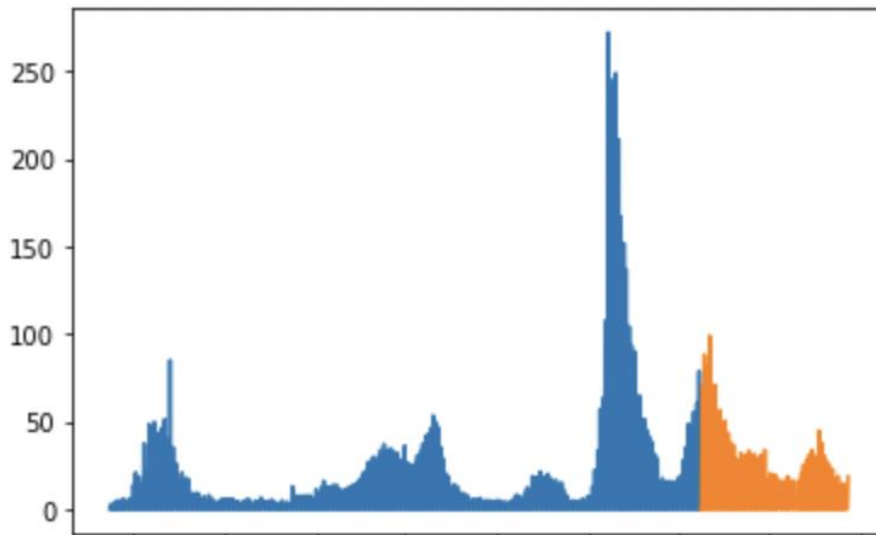


	horizon	mse	rmse	mae	mape	mdape	smape	coverage
0	13 days	564.043511	23.749600	23.749600	0.409476	0.409476	0.339888	0.0
1	17 days	68.786162	8.293742	8.293742	0.148103	0.148103	0.159947	0.0
2	19 days	29.370715	5.419476	5.419476	0.095079	0.095079	0.090764	1.0
3	23 days	13.292100	3.645833	3.645833	0.079257	0.079257	0.076236	1.0
4	24 days	364.566818	19.093633	19.093633	0.258022	0.258022	0.296240	0.0
5	28 days	141.531448	11.896699	11.896699	0.185886	0.185886	0.204933	0.0
6	29 days	196.765921	14.027328	14.027328	0.304942	0.304942	0.264598	0.0
7	44 days	221.206688	14.873019	14.873019	0.256431	0.256431	0.294145	0.0
8	47 days	730.622103	27.030022	27.030022	0.551633	0.551633	0.432377	0.0
9	50 days	271.719526	16.483917	16.483917	0.350722	0.350722	0.298395	0.0
10	53 days	39.688439	6.299876	6.299876	0.085133	0.085133	0.081658	1.0
11	55 days	57.878662	7.607803	7.607803	0.135854	0.135854	0.145754	0.0
12	58 days	159.263610	12.619969	12.619969	0.155802	0.155802	0.168965	0.0
13	60 days	1.061773	1.030424	1.030424	0.019082	0.019082	0.018902	1.0

5.10. Modelo supervisado: ARIMA Y SARIMAX

Aquí realizamos un modelo autorregresivo de medias móviles, separamos la data de test y train:

Imagen 31: Resultado modelo Autorregresivo, separación de data set



Fuente: Elaboración propia

Luego establecemos los parámetros: afectamos un poco de estacionalidad, 50 iteraciones como máximo, Alpha 5%, entre otros.

Los resultados del resumen se explican de la siguiente forma:

Orden no estacional (4,1,4) = "d" 1 integración explica la nueva serie temporal que son las diferencias de tasas de casos covid, Orden estacional (4,0,0,5) = "4"retraso de un día, "0" sin integración. "0"retraso de medias móviles, "5" días atrás.

El índice de AIC es Akaike Information Criterion y sirve para elegir un modelo entre un conjunto de posibles modelos. El índice calcula la distancia de Kullback - Leibler entre el

modelo y la serie. Una forma de interpretar el índice es: buscamos el modelo con el menor índice porque este es el más simple que se ajusta a los datos. Nuestro AIC es de: 6.132

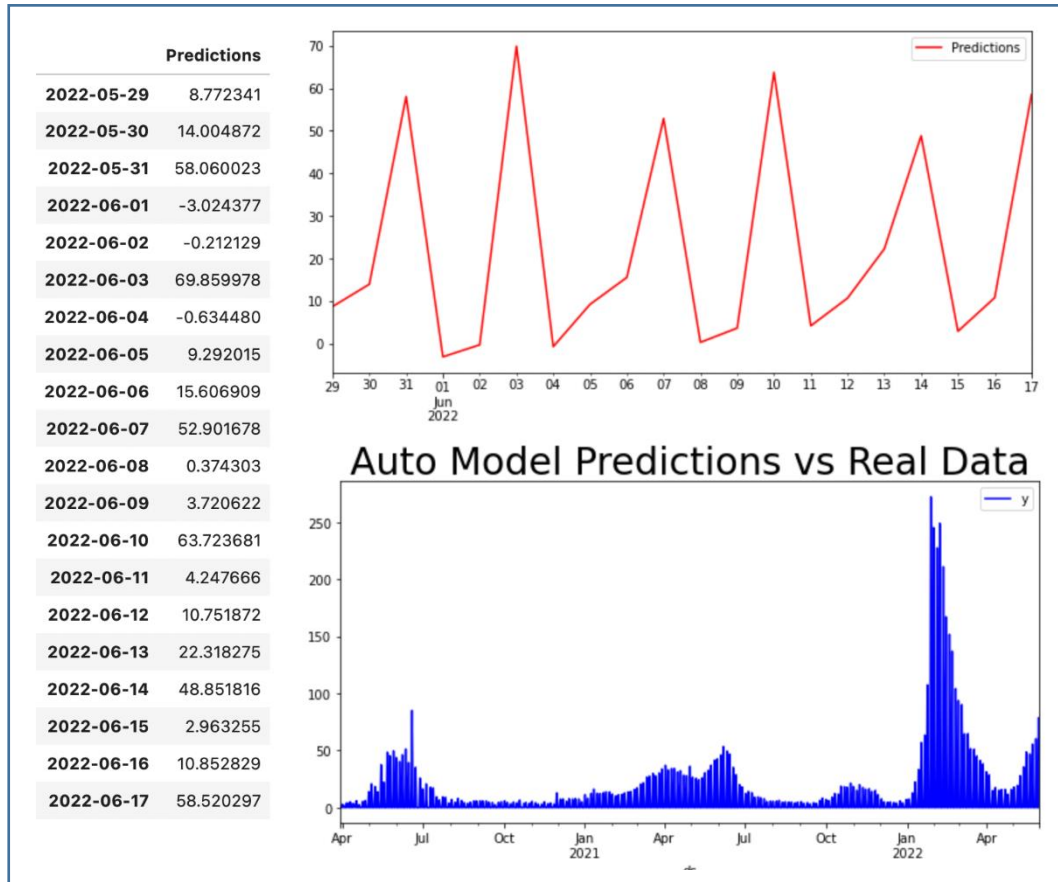
Imagen 32: Resultados modelo ARIMA

Dep. Variable:	y	No. Observations:	793			
Model:	SARIMAX(4, 1, 4)x(4, 0, [], 5)	Log Likelihood	-3052.819			
Date:	Sun, 18 Dec 2022	AIC	6131.637			
Time:	15:21:26	BIC	6192.407			
Sample:	0	HQIC	6154.993			
	- 793					
Covariance Type:	opg					
	coef	std err	z	P> z 	[0.025	0.975]
ar.L1	-2.2153	0.008	-263.034	0.000	-2.232	-2.199
ar.L2	-2.7210	0.016	-175.121	0.000	-2.751	-2.691
ar.L3	-2.1347	0.015	-145.240	0.000	-2.163	-2.106
ar.L4	-0.9281	0.007	-125.724	0.000	-0.943	-0.914
ma.L1	1.1671	0.019	62.135	0.000	1.130	1.204
ma.L2	0.5188	0.042	12.408	0.000	0.437	0.601
ma.L3	-0.0915	0.038	-2.427	0.015	-0.165	-0.018
ma.L4	-0.2645	0.017	-15.525	0.000	-0.298	-0.231
ar.S.L5	-0.2800	0.026	-10.653	0.000	-0.332	-0.228
ar.S.L10	-0.1939	0.018	-10.784	0.000	-0.229	-0.159
ar.S.L15	0.0987	0.039	2.551	0.011	0.023	0.175
ar.S.L20	0.2014	0.025	7.992	0.000	0.152	0.251
sigma2	129.9631	1.568	82.880	0.000	126.890	133.036
Ljung-Box (Q):	250.18	Jarque-Bera (JB):	553556.13			
Prob(Q):	0.00	Prob(JB):	0.00			

Fuente: Elaboración propia

Finalmente obtenemos las siguientes predicciones:

Imagen 33: Predicciones modelo SARIMAX



Fuente: Elaboración propia

Al haber ejecutado primero el auto-arima nos permitió obtener los parámetros del orden estacional y no estacional para aplicar SARIMAX:

Imagen 34: Resultados modelo SARIMAX

SARIMAX Results						
Dep. Variable:	y			No. Observations:	793	
Model:	SARIMAX(4, 1, 4)x(4, 1, 4, 5)			Log Likelihood	-3093.867	
Date:	Sun, 18 Dec 2022			AIC	6221.734	
Time:	20:56:07			BIC	6301.094	
Sample:	03-30-2020			HQIC	6252.244	
	- 05-31-2022					
Covariance Type:	opg					
	coef	std err	z	P> z 	[0.025	0.975]
ar.L1	-2.2367	0.007	-310.558	0.000	-2.251	-2.223
ar.L2	-2.7950	0.012	-226.357	0.000	-2.819	-2.771
ar.L3	-2.2290	0.009	-254.296	0.000	-2.246	-2.212
ar.L4	-0.9840	0.005	-197.037	0.000	-0.994	-0.974
ma.L1	1.1936	0.019	64.032	0.000	1.157	1.230
ma.L2	0.5093	0.028	17.934	0.000	0.454	0.565
ma.L3	-0.1605	0.023	-6.951	0.000	-0.206	-0.115
ma.L4	-0.2812	0.017	-16.455	0.000	-0.315	-0.248
ar.S.L5	-2.7001	1.010	-2.674	0.007	-4.679	-0.721
ar.S.L10	-2.8950	1.760	-1.645	0.100	-6.345	0.555
ar.S.L15	-1.4395	1.223	-1.177	0.239	-3.836	0.957
ar.S.L20	-0.2517	0.265	-0.951	0.342	-0.770	0.267
ma.S.L5	1.5084	0.737	2.046	0.041	0.063	2.953
ma.S.L10	-0.1824	1.223	-0.149	0.881	-2.579	2.214
ma.S.L15	-1.5185	2.004	-0.758	0.449	-5.447	2.410
ma.S.L20	-0.8075	1.233	-0.655	0.513	-3.225	1.610
sigma2	149.7334	91.507	1.636	0.102	-29.616	329.083
Ljung-Box (Q):	689.91	Jarque-Bera (JB):	289694.67			
Prob(Q):	0.00	Prob(JB):	0.00			
Heteroskedasticity (H):	5.49	Skew:	4.87			
Prob(H) (two-sided):	0.00	Kurtosis:	96.49			

Fuente: Elaboración propia

6. Conclusiones y Trabajo futuro

6.1. Conclusiones

En cuanto a las conclusiones que uno puede obtener del proceso EDA, se observa a las comunas de Puente Alto o Santiago que mantienen mayores tasas de crimen, sin embargo, la realidad es que encontramos a una comuna del sector oriente como Providencia que tiene una mayor tasa de crimen agrupado en el total de delitos y considerando el factor de habitantes por comuna, lo cual es bastante interesante descubrir.

La pandemia afectó la tasa de delitos que se realizan en el exterior como hurtos, homicidios o lesiones reduciéndose, mientras que aumentó todos los que se refieren a crímenes dentro del hogar como violencia intrafamiliar o abusos sexuales. Particularmente en periodos de restricciones de movilidad la mayoría de los crímenes se vieron drásticamente disminuidos, en especial al inicio de la pandemia.

A lo largo de este proyecto, hemos profundizado en series temporales, y utilizado diversos modelos para estimar los delitos y casos de covid-19 y el impacto eventual que esto podría ocasionar en las restricciones de movilidad.

Al utilizar prophet el proceso de predicción se torna bastante amigable, su versatilidad permite pronosticar un gran número y una variedad de problemas de series temporales. Utilizamos como métricas, el RMSE, el error absoluto medio (MAE) y el MAPE, para medir el error final del modelo porcentualmente sobre el periodo seleccionado. Lo valioso de esta métrica es que da una medición del error absoluto porcentual, es decir respecto del dato real. Esto se debe a que el valor está en la misma escala que el objetivo para el cual está

prediciendo. Obtenemos con Prophet unos resultados muy buenos, por un lado, si queremos predecir delitos a 60 días el error porcentual absoluto medio es de 1,9% lo cual equivale a un delito de violencia intrafamiliar de error. A medida que vamos aumentando los días a predecir el error aumentará en proporción.

A nivel de segmentación de comunas de la región Metropolitana, hemos llegado a la conclusión que ha habido cambios, dado que diversos crímenes han cambiado a lo largo de la pandemia comparado con el año 2019, previo a la pandemia.

Además de analizar correlaciones, se realizaron modelos de causalidad y regresión lineal con las librerías dowhy y Sklearn, respectivamente. Dentro de la causalidad, hemos concluido que la tasa de covid produce un efecto en la tasa total de crímenes. Finalmente, en el modelo de regresión lineal se aglomeraron las diversas comunas, se amplía como set de variables a cada una de las tasas de crímenes mientras se elimina la tasa total, dado que proviene de la suma de cada una de las anteriores, mientras que se selecciona como la tasa de covid como la variable dependiente, obteniendo así un coeficiente de determinación $R^2 = 61,6\%$.

6.2. Trabajo futuro

En cuanto a trabajo futuro, se proponen dos ideas, primero por profundidad del tema realizar muchas más pruebas de causalidad, no solo considerando como variables delitos, restricción de movilidad y casos covid, por ejemplo, existen variables como hacimiento, nivel socioeconómico, que sería interesante evaluar dentro del modelo de causalidad, segundo por un tema de plazo, creando modificaciones mayores a los dataset utilizados para el presente trabajo.

Con respecto a resultados concretos, se puede resumir el trabajo futuro en los siguientes dos puntos:

- Mayores test de agregación de variables para ver si es robusto o no la causalidad de mayores restricciones de movilidad menores delitos en el espacio público. Siempre buscando tener significancia estadística.
- La creación de un solo dataset que combine los delitos y el covid de forma más frecuente a una mensualidad, por ejemplo, semanal o diario. Luego crear un modelo GLM con distribución Poisson, donde se pueda explicar y predecir una cantidad de casos confirmados de Covid-19 según la cantidad o tasa de crímenes por comuna.

7. Referencias

1. Paper: Covid-19: Short-term forecast of ICU beds in times of crisis. URL: <https://repositorio.uchile.cl/bitstream/handle/2250/182097/COVID-19%20%281%29.pdf?sequence=1&isAllowed=y> Autores: (Goic, Bozanic-Leal, Badal, y Basso). Departamento de ingeniería industrial universidad de chile. Instituto de sistemas complejos de ingeniería (ISCI) y departamento de ingeniería civil Universidad de Chile.
2. Código proyecto Kaggle: Best COVID-19 Forecasting in US, UK and Chile! Autor: George. URL: Saavedra. <https://www.kaggle.com/code/georgesaavedra/best-covid-19-forecasting-in-us-uk-and-chile>
3. Measuring the impact of the COVID-19 lockdown on crime in a medium-sized city in China: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8577180/>
4. COVID-19 and Crime: Effects of Stay-at-Home Orders on Domestic Violence: <https://www.journals.uchicago.edu/doi/full/10.1086/713787>
5. Impact Report: COVID-19 and Crime 21.05.2021: <https://counciloncj.org/impact-report-covid-19-and-crime/>
6. Promising Practices from Law Enforcement's COVID-19 Response: https://www.rand.org/pubs/research_briefs/RBA108-1.html
7. Crime Rates in a Pandemic: the Largest Criminological Experiment in History: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7297511/>
8. How the Criminal Justice System's COVID-19 Response Has Provided Valuable Lessons for Broader Reform: https://www.rand.org/pubs/research_briefs/RBA108-6.html
9. <https://pazciudadana.cl/noticias/daniel-johnson-disminucion-delitos/> Fecha: 12-05-2020
10. “Análisis comparativo de dos técnicas de predicción de datos de covid’19 en chile utilizando herramientas de análisis de datos R y RapidMiner” URL: http://repobib.ubiobio.cl/jspui/bitstream/123456789/3727/1/Silva_Castro_Paula.pdf Autor: Paula Silva Castro.
11. “estadísticas para la vigilancia epidemiológica del covid-19 en chile” URL: <http://repositorio.udec.cl/bitstream/11594/6283/1/TESIS%20ESTADISTICAS%20PARA%20LA%20VIGILANCIA%20EPIDEMIOLOGICA%20.Image.Marked.pdf> Autor: Nixon Andrés Jerez Lillo
12. Producto 1 GitHub Minsal: https://raw.githubusercontent.com/MinCiencia/Datos-COVID19/master/output/producto1/Covid-19_std.csv
13. <https://www.df.cl/economia-y-politica/macro/las-condes-es-la-comuna-que-mas-gasta-en-seguridad-y-antofagasta-lidera>
14. <https://www.latercera.com/nacional/noticia/santiago-puente-alto-y-temuco-las-comunas-con-mas-infracciones-por-no-cumplir-la-cuarentena/TO3DDTBHJFDCNAAWSA7DXEC3Y/>

15. https://www.bcn.cl/siit/mapas_vectoriales/index.html
16. <https://facebook.github.io/prophet/>