



Universidad del Desarrollo
Facultad de Ingeniería

ESTUDIO DE LA RELACIÓN ENTRE CASOS POLICIALES Y PERSONAS
EN HORARIO HÁBIL EN LAS COMUNAS DE CHILE

POR: SALOMÓN DENNIS NILO DURÁN

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:

Loreto Bravo

Leonardo Ferres

Enero 2023

SANTIAGO

A mis padres por su apoyo incondicional.

TABLA DE CONTENIDO

RESUMEN.....	1
1. INTRODUCCIÓN	2
2. TRABAJO RELACIONADO	4
3. HIPÓTESIS Y OBJETIVOS.....	5
4. DATOS Y METODOLOGÍA	6
4.1. DATOS	6
<i>Datos de dispositivos móviles</i>	<i>6</i>
<i>Datos de Casos Policiales</i>	<i>7</i>
<i>Shapefiles - Biblioteca del Congreso Nacional de Chile.....</i>	<i>8</i>
4.2. METODOLOGÍA.....	8
4.2.1 <i>Recolección de datos.....</i>	<i>8</i>
4.2.2 <i>Obtención de cantidades.....</i>	<i>9</i>
4.2.3 <i>Recuento de datos de personas en horario laboral por comuna</i>	<i>10</i>
4.3.4. <i>Serie temporal de frecuencia de delitos por comuna.....</i>	<i>11</i>
4.4.4. <i>Comparación de resultados.</i>	<i>11</i>
5. RESULTADOS	13
5.1. ANÁLISIS EXPLORATORIO	13
5.2. COMPARACIÓN DE RESULTADOS	14
5.2.1 <i>Coficiente de Pearson.....</i>	<i>15</i>
5.2.2 <i>Coficiente de correlación de Spearman.....</i>	<i>17</i>
5.2.3 <i>Regresión lineal sobre datos de 2020 para explicar datos de 2021 y 2022.</i>	<i>18</i>
5.3 HALLAZGOS ADICIONALES	21
6. CONCLUSIONES	25

6.2 CONTINUACIÓN DEL ESTUDIO.....	25
6.2 LIMITACIONES	26
BIBLIOGRAFÍA	27

Resumen

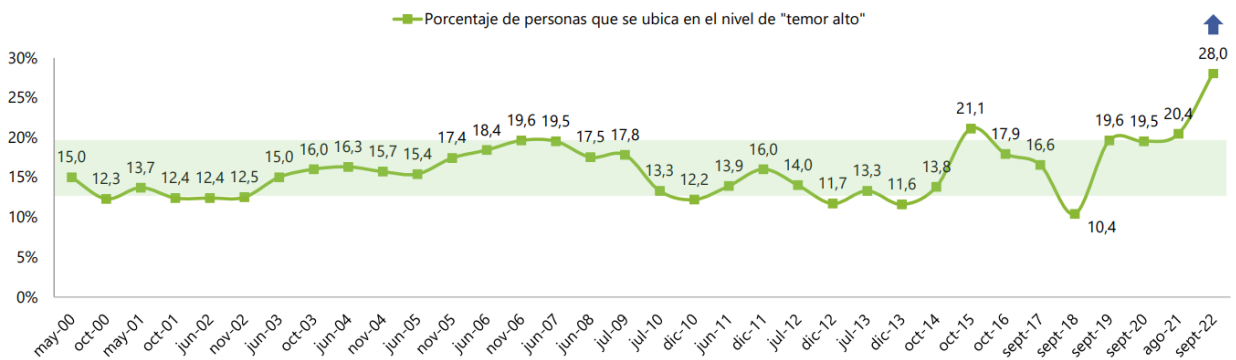
Según el Banco Mundial (Ayres, 1998) la delincuencia tiene un impacto negativo para un buen clima de inversión, contribuyen a la reducción de la inversión en infraestructura y otros. Asimismo, el crimen tiene un impacto muy negativo para los gobiernos, pues produce que deban invertir en programas costosos contra la delincuencia, empujando a que la ciudadanía invierta en seguridad privada, haciendo que la imagen del gobierno se vea mermada.

Es de preocupación colectiva el combate contra la delincuencia para un estado de bienestar común, este proyecto nos ayuda a entender cómo se comporta la delincuencia a lo largo de Chile. Usaremos datos de dispositivos móviles y su geolocalización en cada comuna los cuales compararemos contra la cantidad de casos policiales ocurridos en cada unidad geográfica. Este análisis se hizo con datos de los años 2020, 2021 y 2022.

El proyecto concluyó que sí existe una relación directa entre cantidad de crímenes y cantidad de personas y también cómo es dicha relación, así poder realizar distintos análisis comunales y construir semáforos de tendencia al alza o baja de delitos.

1. Introducción

Durante el último tiempo, ha existido un significativo aumento en la percepción del temor, el índice del temor alcanzó un máximo histórico llegando al 28% (Índice Paz Ciudadana, pág. 23)



Podrían ser muchas las posibles razones del por qué el aumento anterior. Para poder tener una mayor comprensión sobre esta alza, analizaremos la cantidad de personas en horario hábil dentro de cada una de las comuna de Chile, utilizaremos sólo los datos de personas en horario hábil, ya que durante 2019 como muestra la ilustración 1 (INE, 2019, pág. 35) se observa que los tres tramos con mayor cantidad de delitos están entre la banda horaria desde las 8 hasta las 20 horas, y para obtener una mayor precisión, solo nos quedaremos con las personas que trabajan en cada comuna, quitando así las personas que viven en dichas comunas, usaremos un algoritmo opuesto al de *home detection* TC-WK-19-7 (Detección del hogar de un individuo) (Pappalardo, Ferres, Sacasa, & et al, 2021),

en consecuencia, utilizaremos el criterio TC-WK-7-19 para obtener personas que trabajan dentro de cada comuna, usando la banda horaria desde las 7 a las 19 horas.

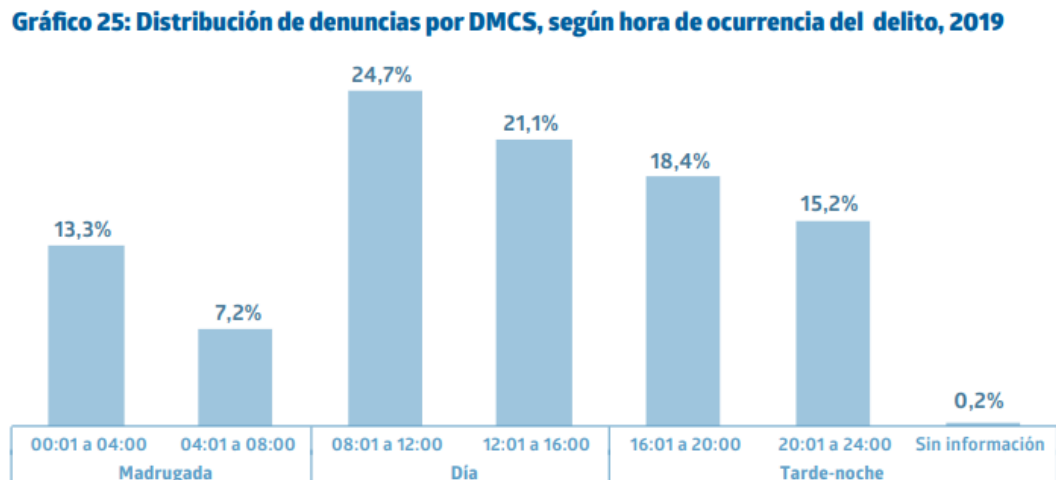


Ilustración 1 - Distribución de denuncias por DMCS, según hora de ocurrencia del delito 2019. (INE, 2019)

Esta cantidad de personas en horario hábil en cada comuna debiese estar fuertemente ligada a la de casos policiales DMCS (Delitos de mayor connotación social) (ChileAtiende, 2022). Es así como surge la siguiente hipótesis, el número de personas en horario hábil dentro de las comunas de Chile es directamente proporcional a la de casos policiales registrados durante 2020, usaremos datos de una empresa de telefonía móvil y de Carabineros de Chile para poder entender esta relación.

2. Trabajo Relacionado

Dentro del mundo de la criminalística, han existido distintos análisis, pero pocos con respecto a los últimos años en Chile. En el ámbito internacional James Nolan (Nolan, 2004) postula que la relación entre la cantidad de delitos y la población de una jurisdicción ha sido bastante establecida, no obstante, la relación entre tasa de delitos y tamaño de población es menos clara. Cuando se compara la relación entre cantidad de población y crímenes puede haber muchos factores que alteren dicha relación, pues algunas características de la población pueden ser significativas, como los cambios en el tamaño, la densidad de la población o su composición demográfica y socioeconómica.

Dentro del contexto nacional, J. Ñuñez et al. (Ñuñez, Rivera, Villavicencio, & Molina, 2003) postulan que la densidad poblacional está positivamente asociada con algunos delitos, además que algunas variables socioeconómicas y demográficas regionales pueden estar asociadas a la prevalencia de algunos delitos. Como por ejemplo, el porcentaje de población entre 18 y 40 años, índice de masculinidad, índice de ruralidad, el porcentaje de la población regional que vive en un gran centro urbano (Gran Ciudad).

En este trabajo tomaremos en consideración una variable que no había sido tomada en cuenta, la población presente solo en horarios hábiles en cada unidad geográfica, así quedarnos solo con las personas que trabajan en cada lugar. Evaluaremos qué tan directa es su relación contra la cantidad de crímenes en cada unidad geográfica.

3. Hipótesis y objetivos

“La cantidad de personas en horario hábil dentro de las comunas de Chile es directamente proporcional a la de casos policiales registrados”, es en base a esta hipótesis que se trabaja el presente proyecto.

El principal objetivo es demostrar la directa relación entre la cantidad de personas que están en una determinada comuna en horario hábil contra la cantidad de casos policiales de cada una de las comunas de Chile, para obtener la cantidad de personas en horario hábil usaremos el criterio TC-WK-7-19, tomaremos como muestra los datos de las personas y sus dispositivos móviles que ocurrieron dentro de la semana, es decir, de lunes a viernes desde las 7 de la mañana a las 19 horas de tarde. Para la información necesaria de casos policiales se usarán los datos de carabineros de Chile, que los proporciona CEAD, el Centro de Estudios y Análisis del Delito.

El primer objetivo es poder entender cuántas personas hay en horario hábil en cada comuna de Chile.

Una vez realizado el paso anterior, lo siguiente es poder comparar las cantidades anteriormente señaladas con la cantidad de casos policiales que se definen como *“aquellos delitos conocidos por la policía y que consideran detenciones por flagrancia y denuncias formales realizadas por la ciudadanía en Carabineros de Chile o la PDI”* (CEAD, 2022).

4. Datos y Metodología

4.1. Datos

Datos de dispositivos móviles

La primera fuente de datos hace relación a los datos de una empresa de telefonía móvil, en formato XDRs (eXtended Detail Records), constan de 4 días distintos de información de 3 años diferentes (2020, 2021 y 2022). El foco inicial e hipótesis del proyecto es comparar la relación cantidad de personas y crímenes en 2020, pero se procesarán adicionalmente los datos de los años 2021 y 2022 para poder hacer un análisis más robusto que sirva de comparativo entre los resultados obtenidos entre los distintos años. Para comenzar, es necesario organizar la información de forma correcta, ordenada y consistente, se generarán tres carpetas de datos organizadas en distintos repositorios.

El primer paso es la descarga de información desde el origen de datos de la compañía de telefonía móvil, estos archivos se encuentran en formato columnar (.parquet). Los datos poseen grandes volúmenes de información, siendo cada uno de ellos aproximadamente de 500MB.

Todos los archivos serán llevados a la carpeta *source* del proyecto mediante código Python que copia los archivos a dicho directorio.

La estructura de los archivos es la siguiente:

Device	Datetime	Lat	Lon
Dispositivo móvil encriptado.	Fecha donde se hizo la conexión a la antena	Coordenadas (Latitud donde se encuentra la antena)	Coordenadas (Longitud donde se encuentra la antena)

Datos de Casos Policiales

La segunda fuente de información son los datos policiales del Centro de Estudios y Análisis del Delito, se definen como *“aquellos delitos conocidos por la policía y que consideran detenciones por flagrancia y denuncias formales realizadas por la ciudadanía en Carabineros de Chile o la PDI”* (CEAD, 2022)

Casos policiales considerados para calcular la frecuencia total por comuna:

- Robos con violencia o intimidación.
- Robo por sorpresa
- Robo con fuerza, incluye:
 - Robo de vehículos motorizados.
 - Robo de Accesorios u objetos de o desde vehículo.
 - Robo de Lugar habitado o destinado a la habitación.
 - Robo de Lugar no habitado.
 - Otros robos con fuerza.
- Hurtos.
- Lesiones.

- Lesiones leves.
- Lesiones graves.
- Violación.
- Homicidio.

Los datos anteriormente mencionados son públicos y se pueden descargar en la página de CEAD (CEAD, 2022).

Shapefiles - Biblioteca del Congreso Nacional de Chile

Estos datos contienen los límites, formas y estructuras de las distintas comunas de Chile y servirán para cruzar la información de los dispositivos móviles con la información de las comunas, así determinando en qué comuna se encontraba cada dispositivo por hora descrita. Estos datos se pueden encontrar directamente por varias formas de agrupación, en este caso usaremos comuna (BCN).

4.2. Metodología

Para poder alcanzar el objetivo final se definieron las siguientes etapas:

4.2.1 Recolección de datos.

Esta etapa comprende el copiado de archivos a los repositorios del proyecto, para poder procesar la información más fácilmente dentro del entorno del proyecto.

Lo primero es añadir información adicional a los archivos XDRs de dispositivos móviles, pues necesitamos saber en qué comuna de Chile está ubicado cada dispositivo móvil en el momento determinado que nos dictan los datos de latitud, longitud y fecha.

Se dividirá el conjunto de datos en 3 partes, es decir, cada archivo pasará a ser 3 archivos, cada un año con información de un año distinto, a cada uno de dichos archivos se les agregará columna “comuna”.

Filtros en el uso de datos

Para cada archivo es necesario sólo usar los datos *TC-WK-7-19*, el foco de estos filtros es poder entender cuántas personas estuvieron en las comunas en cuestión en horario comercial (Personas laborando en dichas comunas):

TC: Time Constraints Criterion (Restricciones de tiempo)

WK: Sólo usando días de la semana laborales de lunes a viernes.

7-19: Daytime period - En el rango horario de 7 horas de la mañana a 19 horas de la tarde.

Marzo: Solo usaremos marzo para los datos de XDRs, de modo de poder compararlo con los datos de marzo de casos policiales.

Cuando ya tenemos los datos almacenados por año y adicionalmente con la información de qué comuna es cada registro, es posible proceder al siguiente paso.

4.2.2 Obtención de cantidades

Ya dada esta etapa tenemos la información almacenada en archivos parquet para cada uno de los años de datos (2020, 2021, 2022). Es necesario agrupar por comuna y entender cuántos dispositivos móviles hubo en cada comuna en los años mencionados anteriormente.

Para ello es necesario leer cada uno de los 500 archivos parquet para cada año (1500 archivos en total) y realizar contadores de cuántas veces estuvo cada dispositivo móvil en cada comuna.

Según cada uno de los dispositivos, la comuna que tenga mayor cantidad de registros relacionados será la ganadora, como se muestra en la tabla 1, con lo anterior ya tendremos la relación año-dispositivo-comuna.

Dispositivo	Año	Comuna	Cantidad de ocurrencia	Comuna Seleccionada
Dispositivo1	2020	Las Condes	4	Sí (Dato Final)
Dispositivo1	2020	Santiago	1	No
Dispositivo1	2020	Providencia	1	No
Dispositivo1	2021	Santiago	5	Sí (Dato Final)
Dispositivo1	2021	Las Condes	2	No
Dispositivo1	2021	Providencia	2	No
Dispositivo1	2022	Las Condes	8	Sí (Dato Final)
Dispositivo1	2022	Ñuñoa	4	No
Dispositivo1	2022	Independencia	3	No
Dispositivo1	2022	Providencia	2	No

Tabla 1 – Ejemplo de datos procesados

4.2.3. Recuento de datos de personas en horario laboral por comuna

Luego de la etapa anterior, agruparemos los datos y realizaremos un conjunto de datos que posea cuántas personas estuvieron en cada comuna, año y mes (Personas en horario hábil).

Esta serie consta de datos de **marzo de 2020, marzo 2021 y marzo del 2022** tal como lo muestra la tabla 2.

Dispositivo	Año	Comuna
Dispositivo 1	2020	Las Condes
Dispositivo 2	2020	Maipú
Dispositivo 3	2020	Arica
Dispositivo 1	2021	Santiago
Dispositivo 2	2021	Ñuñoa
Dispositivo 1	2022	Las Condes

Tabla 2

4.3.4. Serie temporal de frecuencia de delitos por comuna

Es necesario hacer la misma serie temporal del ítem anterior, pero para los datos de casos policiales, considerando algunos delitos para marzo 2020, marzo 2021 y marzo 2022.

La base de datos a usar son los Casos Policiales del Centro de Estudios y Análisis del Delito (CEAD, 2022).

Dentro de los archivos que comparte CEAD, se encuentran por tasa y por frecuencia, para este análisis, usaremos los datos crudos en frecuencia, y solo usando marzo para poder comparar las cantidades con la cantidad de dispositivos móviles.

4.4.4. Comparación de resultados.

Para el análisis y pruebas finales, es necesario comparar ambos conjuntos de datos obtenidos para el **año 2020** y entender si existe alguna correlación entre ellas (Cantidad de personas en horario hábil y casos policiales), utilizando los siguientes métodos:

Coefficientes de correlación de Pearson y Spearman

Dos de los métodos más usados en la industria para entender si existe alguna correlación entre dos variables son Pearson y Spearman.

Con esto podemos entender si la relación es:

1. Correlación directa lineal (Pearson).
2. Correlación monótona, que puede ser no necesariamente lineal, sino que exponencial. (Spearman).

Este paso se realiza para entender si existe una correlación positiva débil (Coeficiente Pearson mayor a 0,2), correlación positiva fuerte (Coeficiente Pearson mayor a 0,8) o quizá una correlación positiva perfecta (Coeficiente positiva perfecta).

Regresión lineal simple

Otra de las aristas posibles y corroborar de que la tendencia y relación de la hipótesis es sostenible en el tiempo es comparar los datos con 2021 y 2022.

Para ello se estimarán los casos policiales para 2020, 2021 y 2022 con ayuda de una regresión lineal simple, donde la variable dependiente serán los Casos Policiales, y la variable independiente serán la cantidad de personas en la comuna en horario hábil.

Aplicando los métodos comparativos a los años 2021 y 2022

Serán los datos de 2021 y 2022 corroboradores de la hipótesis, para poder tener mayor comparación entre variables durante el tiempo y a través de cada una de las comunas de

Chile. Por consecuente se agregará a todos los análisis anteriores los datos de 2021 y 2022 de referencia.

5. Resultados

5.1. Análisis exploratorio

Uno de los primeros desafíos fue trabajar con datos georreferenciados a gran escala pues los datos totales a usar son 1.816.961.236 registros. Estos datos se dividen en 3 años diferentes, siendo 2022 el de mayor volumen de datos.

Las fechas a utilizar de los archivos proporcionados fueron:

2020: 2020-03-02, 2020-03-03, 2020-03-04, 2020-03-05

2021: 2021-03-01, 2021-03-02, 2021-03-03, 2021-03-04

2022: 2022-03-01, 2022-03-02, 2022-03-03

A cada registro hubo que agregarle la información de la comuna en la cual se encontraba cada registro, proceso que demoró aproximadamente unas 24 horas.

Luego de procesar la información y filtrar solo los datos de Dispositivos móviles únicos por comunas por año quedó un total de 19.346.439 dispositivos así quedando con un 1,06% de los datos, de esta muestra se excluyeron los dispositivos móviles que estuvieron en más de una comuna en el mismo año.

El resultado final fue un archivo con 344 comunas, para 3 años dando un total de 1032 registros.

El ejemplo de la salida final se ve a continuación en la tabla 3 – Salida de datos final.

Comunas	Date	CasosPoliciales	Personas
Arica	Mar-2020	620	41894
Camarones	Mar-2020	1	867
General Lagos	Mar-2020	1	52
Putre	Mar-2020	3	957
Alto Hospicio	Mar-2020	303	19592

Tabla 3 - Salida de datos final

5.2. Comparación de resultados

Los primeros resultados ya arrojan una dependencia entre las variables, como se puede ver en el gráfico de dispersión en la Ilustración 2.

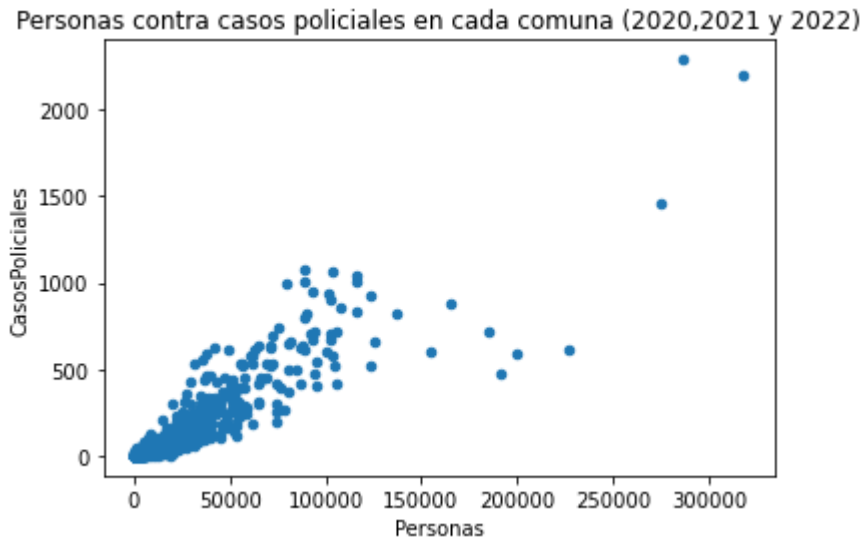


Ilustración 2 - Personas contra casos policiales en cada comuna de Chile (2020,2021 & 2022)

La tendencia anterior también se puede apreciar en los distintos años como se puede ver en la Ilustración 3.

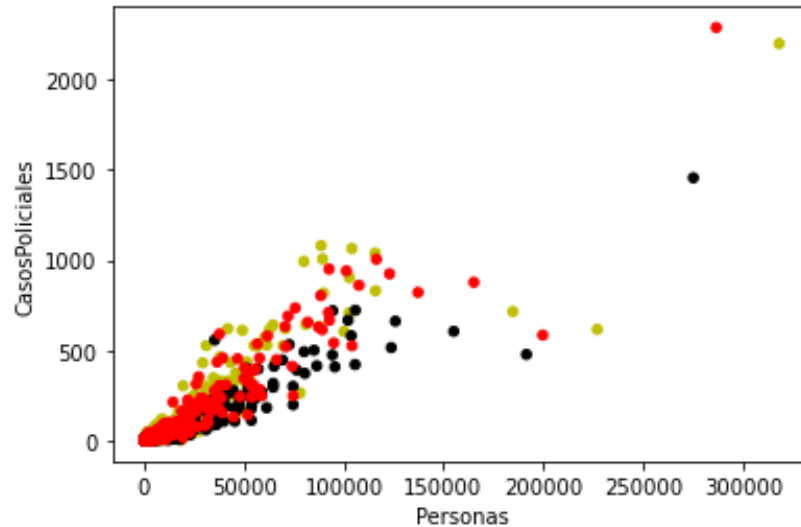


Ilustración 3 - Personas contra casos policiales en cada comuna de Chile por año (2020,2021 & 2022)

5.2.1 Coeficiente de Pearson

“Dadas dos variables, la correlación permite hacer estimaciones del valor de una de ellas conociendo el valor de la otra variable.

Los coeficientes de correlación son medidas que indican la situación relativa de los mismos sucesos respecto a las dos variables, es decir, son la expresión numérica que nos indica el grado de relación existente entre las 2 variables y en qué medida se relacionan. Son números que varían entre los límites +1 y -1. Su magnitud indica el grado de asociación entre las variables; el valor $r = 0$ indica que no existe relación entre las variables; los valores 1 son indicadores de una correlación perfecta positiva (al crecer o decrecer X, crece o decrece Y) o negativa (Al crecer o decrecer X, decrece o crece Y)”. (Orlando & Suárez, 2011).

Luego de procesar los datos para los años 2020, 2021 y 2022 se obtuvo lo que se puede apreciar en la Tabla 4.

Año	Coefficiente de correlación de Pearson	Significancia
2020	0.9082	$1.449717 \times 10^{-129}$
2021	0.9382	$2.125731 \times 10^{-157}$
2022	0.9378	$6.302255 \times 10^{-157}$

Tabla 4 - Coeficiente de correlación de Pearson

Según la Tabla 5, se puede apreciar que todos los valores obtenidos para 2020, 2021 y 2022 tienen una correlación positiva muy alta. Se puede decir entonces que, a mayor cantidad de personas en horario hábil en una determinada comuna, mayor será la cantidad de casos policiales.

Valor	Significado
-1	Correlación negativa grande y perfecta
-0,9 a -0,99	Correlación negativa muy alta
-0,7 a -0,89	Correlación negativa alta
-0,4 a -0,69	Correlación negativa moderada
-0,2 a -0,39	Correlación negativa baja
-0,01 a -0,19	Correlación negativa muy baja
0	Correlación nula
0,01 a 0,19	Correlación positiva muy baja
0,2 a 0,39	Correlación positiva baja
0,4 a 0,69	Correlación positiva moderada
0,7 a 0,89	Correlación positiva alta
0,9 a 0,99	Correlación positiva muy alta
1	Correlación positiva grande y perfecta

Tabla 5 - Escala de coeficiente de correlación de Pearson

5.2.2 Coeficiente de correlación de Spearman

“El coeficiente de Spearman no es una medida de la relación lineal entre dos variables. Evalúa qué tan bien una función monótona arbitraria puede describir una relación entre dos variables, sin hacer suposiciones sobre la distribución de frecuencias de las variables. A diferencia del coeficiente de correlación producto-momento de Pearson, no requiere la suposición de que la relación entre las variables es lineal, ni requiere que las variables sean medidas en escalas de intervalo; puede ser usado para variables medidas a nivel ordinal” (Hauke, 2011).

Para los datos del conjunto de datos obtenidos anteriormente, los resultados se pueden ver en la tabla 6 del coeficiente de correlación de Spearman.

Año	Coeficiente de correlación de Spearman
2020	0.889304
2021	0.891549
2022	0.907309

Tabla 6 - Coeficiente de correlación de Spearman

Lo anterior nos confirma que la relación entre las personas en horario hábil de las comunas de Chile es directamente proporcional a la cantidad de casos policiales y aumenta en forma **monotónica y lineal**.

5.2.3 Regresión lineal sobre datos de 2020 para explicar datos de 2021 y 2022.

¿Es posible explicar los valores de 2021 y 2022 con 2020? La respuesta es que tienen un comportamiento muy similar estos tres años, pues tienen correlaciones muy similares, siendo los años 2021 y 2022 con una correlación muy fuerte entre variables.

Al realizar una recta de regresión utilizando sólo el 70% de los datos del 2020, se puede ver una recta como la de la Ilustración 4 - Regresión lineal simple con datos de 2020.

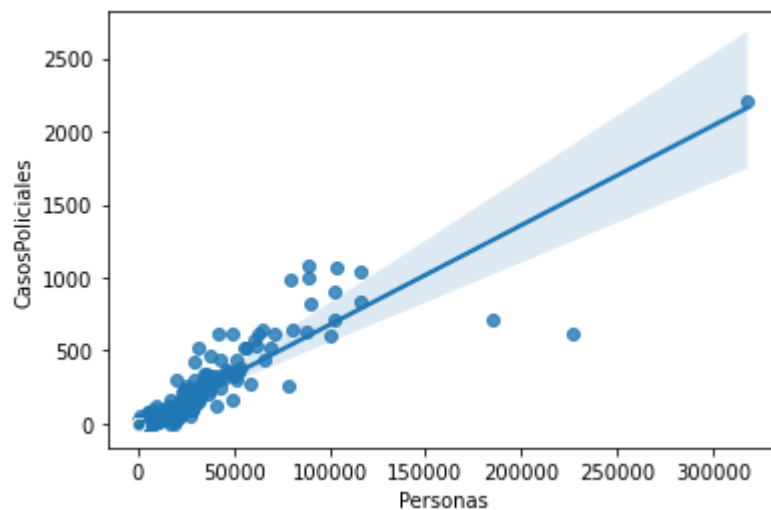


Ilustración 4 - Regresión lineal simple con datos de 2020

Este resultado se obtuvo con un coeficiente de determinación R^2 para los datos de *testing* (30% de los datos) igual a 0.9342. Con estos datos se obtuvo los siguientes valores para la función de regresión:

- Coeficiente de la variable explicativa (Personas TC-WK-7-19) = 0.00627513
- Intercepto con eje y (Casos Policiales) = 5.64405954

Con los datos anteriores podemos concluir que la función resultante es:

$$\text{CasosPoliciales} = 5.64405954 + 0.00627513 \times \text{CantidadPersonas} \text{ (TC-WK-7-19)}$$

Esta es la fórmula que se puede aplicar a los datos de 2021 y 2022, pero primero, la proyección aplicada a los datos de 2021 se vería como lo muestra la Ilustración 5 -

Estimación Casos Policiales.

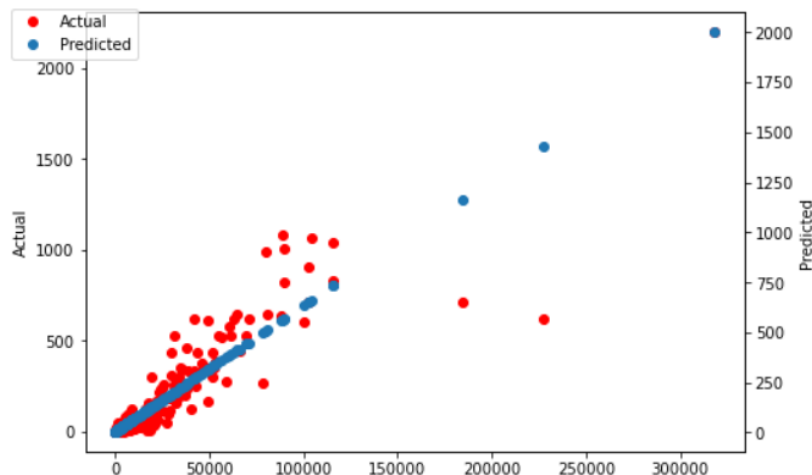


Ilustración 5 - Estimación Casos Policiales

El error cuadrático medio (MSE – Mean Squared Error) obtenido para la muestra anterior resultó en 9616.

Aplicando predicciones para 2021 y 2022 con datos del 2020 se pudo obtener errores mucho menores, indicando que los valores de los años 2021 y 2022 se ajustan mejor a la función de regresión como se puede ver en la Ilustración 6 - Comparativo datos predichos vs actual 2021 y 2022.

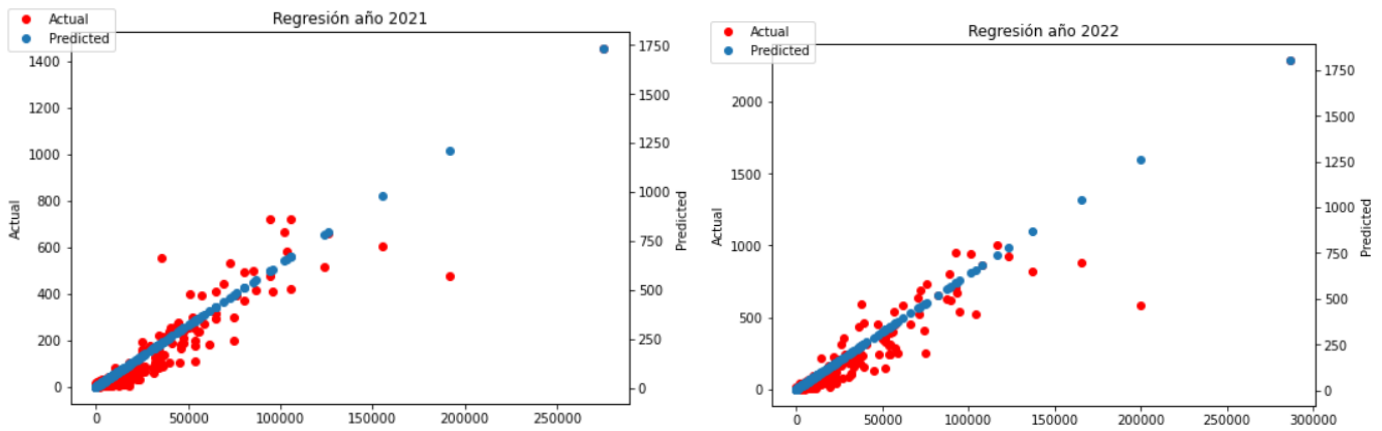


Ilustración 6 - Comparativo datos predichos vs actual 2021 y 2022

Ya con lo anterior podemos llegar a concluir lo siguiente con respecto a errores del modelo:

Año	Mean Squared Error
2020	9616.76
2021	6509.30
2022	6634.77

Adicionalmente al ejercicio anterior, se realizó una segunda regresión, esta vez utilizando el 70% de todos los datos, no solo usando 2020 sino que también 2021 y 2022. Con esta nueva regresión los errores del modelo fueron bastante menores, dando así lo siguiente:

Año	Mean Squared Error
2020	98.73

2021	76.36
2022	79.68

5.3 Hallazgos adicionales

Al realizar el entendimiento de cuánto se alejan los datos reales a la predicción, fue necesario calcular los residuos de la predicción, esto quiere decir, la resta entre los datos predichos menos la cantidad actual de casos policiales en cada comuna, al llevar esto a un histograma como el de la Ilustración 7 - Histograma diferencias de casos reales vs predicción, se pueden observar algunos valores que se alejan bastante de la distribución como lo es la comuna de Las Condes de la Región Metropolitana, donde se pueden ver hasta -1016 casos menos de lo dictado por la predicción.

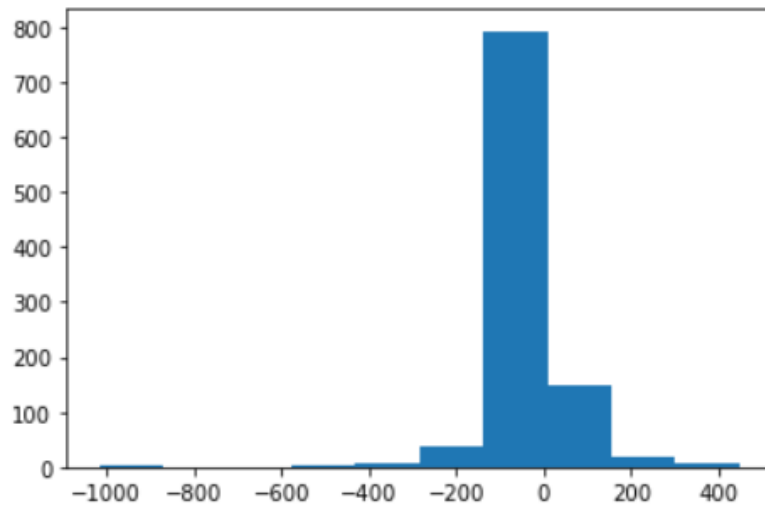


Ilustración 7 - Histograma diferencias de casos reales vs predicción

A continuación, en la tabla 7, se muestran algunos de los datos a modo de ejemplo:

Comuna	Fecha	Personas	Casos Policiales	Predicción	Diferencia
Las Condes	mar-20	227.026	616	1.632	-1.016
Las Condes	mar-21	191.556	477	1.376	-899
Las Condes	mar-22	199.727	584	1.435	-851
Providencia	mar-20	184.741	715	1.326	-611
Santiago	mar-21	275.150	1456	1.980	-524
Providencia	mar-21	155.128	605	1.112	-507
Viña del Mar	mar-21	123.816	515	886	-371
Ñuñoa	mar-21	105.718	421	755	-334
Vitacura	mar-21	74.667	198	531	-333
Providencia	mar-22	165.203	877	1.185	-308
Vitacura	mar-20	78.149	264	556	-292

Tabla 7 – Ejemplos de diferencias de casos reales vs predicción

De estos ejemplos, es la comuna de Las Condes la que más se aleja a los valores de predicción en los 3 años, acumulando 2.766 casos menos que lo que nos dice la predicción, las diferencias anuales para la comuna se pueden ver en la Ilustración 8- Comparativo de casos policiales actuales vs predicción.

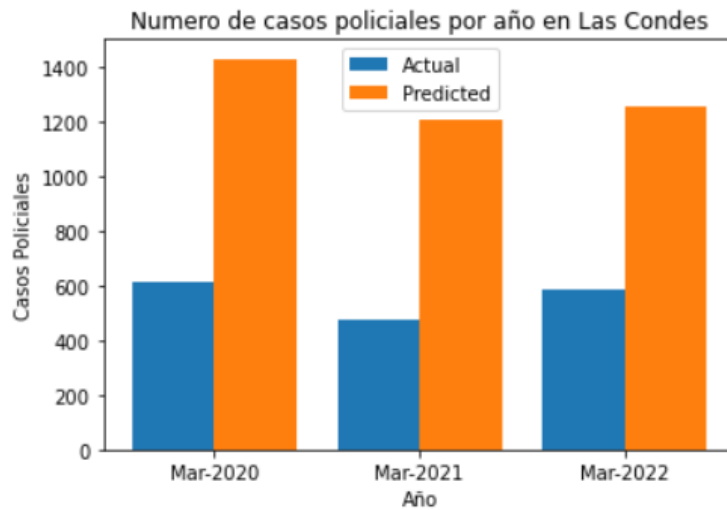


Ilustración 8- Comparativo de casos policiales actuales vs predicción Las Condes

La comuna que continúa en el listado es la comuna de Providencia que acumula 1.426 casos menos de diferencia contra lo que nos dicta la predicción. Esto se puede ver en la Ilustración 9 - Comparativo de casos policiales actuales vs predicción Providencia.

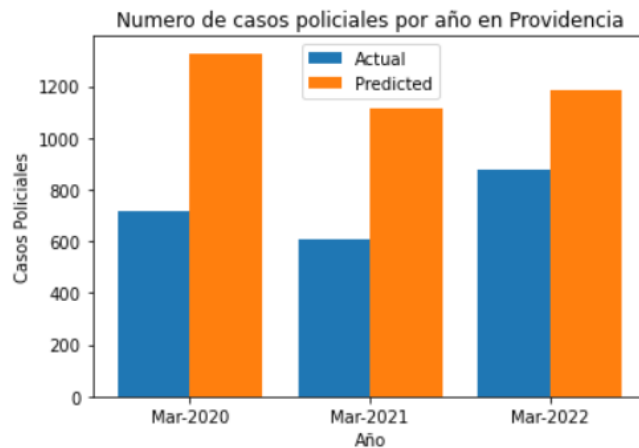


Ilustración 9 - Comparativo de casos policiales actuales vs predicción Providencia

Como una de las comunas que tiene el caso contrario, más casos actualmente de los que dicta la predicción está Puente Alto con un total de 628 casos más de lo que

debiese, la evolución anual se puede ver en la Ilustración 10 - Comparativo de casos policiales actuales vs predicción Puente Alto

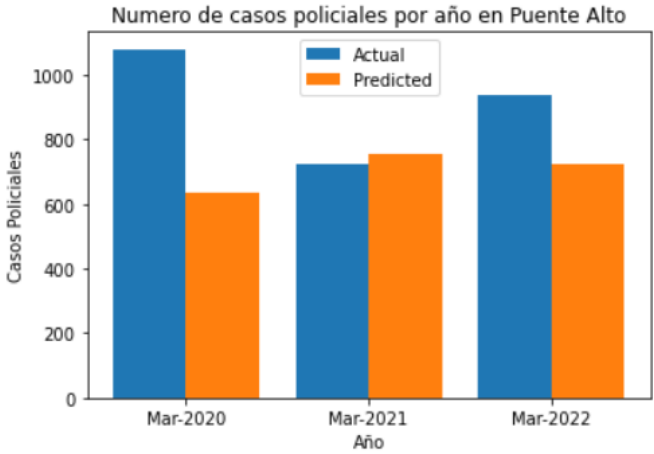


Ilustración 10 - Comparativo de casos policiales actuales vs predicción Puente Alto

6. Conclusiones

Podemos comprobar que sí existe una relación fuerte, directa, lineal y monótona entre las variables cantidad de personas en horario hábil con la cantidad de delitos policiales en las distintas comunas de Chile durante 2020, 2021 y 2022 esta relación se vuelve aún más marcada en los años 2021 y 2022 donde las correlaciones entre los datos son incluso mayores. Los datos se han comportado de manera muy similar en los últimos tres años, pues tienen indicadores muy similares, muchas de las comunas analizadas se comportan con tendencias parecidas durante estos tres años.

6.2 Continuación del estudio

Con este estudio se abre un abanico importante de futuros análisis para entender dicha relación y qué variables hacen que suban o bajen los casos policiales dentro de cada comuna. Algunas de las interrogantes que surgieron durante el desarrollo son las siguientes:

1. ¿Es posible predecir qué comunas están en un aumento de casos policiales o lo tendrán en el futuro?

En este punto es necesario contar con mayor información de años anteriores y de más meses de información para poder analizar tendencias de alza o caída en los indicadores de delincuencia y poder realizar series temporales y estimaciones. La idea sería poder actuar de manera oportuna y corregir posibles tendencias al alza de casos policiales.

2. ¿Es el origen de las personas un factor determinante en el aumento de casos policiales?

Uno de los análisis interesantes es, si dentro de cada comuna existe alguna relación entre las personas que se mudaron a esa comuna y el aumento de casos en los últimos tres años.

3. ¿Por qué algunas comunas se alejan tanto de la recta de regresión?

Dentro de los análisis que se abren también está el poder agregar más variables explicativas al predictor de los datos, por ejemplo, sumar información de cantidad de personal de seguridad en cada comuna, planes comunales de seguridad, movilidad dentro de la comuna, etc. Para así poder lograr quizá una predicción mucho más exacta de los valores.

6.2 Limitaciones

A la hora de abordar cualquier análisis con datos, es necesario tener la mayor cantidad posible de datos históricos para poder hacer buenos modelos de datos. El estudio realizado tuvo la limitante de no poder ver más meses de información y solo ver marzo por los datos disponibles hasta el momento.

Bibliografía

- Ayres, R. L. (1998). *Crime and violence as development issues in Latin America and the Caribbean*. Obtenido de <https://documents1.worldbank.org/curated/en/385501468742891106/pdf/multi-page.pdf>
- BCN, B. d. (s.f.). Shapefiles Comunas de Chile. Obtenido de <https://www.bcn.cl/obtienearchivo?id=repositorio/10221/10396/2/Comunas.zip>
- CEAD. (2022). *CEAD*. Obtenido de <http://cead.spd.gov.cl/centro-de-documentacion/?r=1>
- ChileAtiende. (2022). *Estadísticas de delitos de mayor connotación social, violencia intrafamiliar y ley de drogas*.
- Fundación Paz Ciudadana. (2022). *Índice Paz Ciudadana*. Santiago, Chile.
- Hauke, J. &. (2011). *Comparison of values of Pearson's and Spearman's correlation coefficients on the same sets of data*. *Quaestiones geographicae*.
- INE, I. N. (2019). *Estadísticas Policiales*.
- Jorge, D. (2014). *Correlación Lineal*.
- Nolan, J. J. (2004). *Establishing the statistical relationship between population size and UCR crime rate: Its impact and implications*. *Journal of Criminal Justice*.
Obtenido de <https://www.sciencedirect.com/science/article/pii/S0047235204000832>

Ñuñez, J., Rivera, J., Villavicencio, X., & Molina, O. (2003). *DETERMINANTES SOCIOECONOMICOS Y DEMOGRAFICOS DEL CRIMEN EN CHILE*. Santiago. Obtenido de <https://econ.uchile.cl/uploads/publicacion/001898cb-526c-41c7-a359-1844809fa80c.pdf>

Orlando, M., & Suárez, I. (2011). *Coficiente de correlación de Karl Pearson*. Obtenido de <http://repositorio.utn.edu.ec/handle/123456789/766>

Pappalardo, L., Ferres, L., Sacasa, M., & et al. (2021). *Evaluation of home detection algorithms on mobile phone data using individual-level ground truth*. EPJ Data Sci. Obtenido de <https://doi.org/10.1140/epjds/s13688-021-00284-9>