



Universidad del Desarrollo
Facultad de Ingeniería

**ESTUDIO SOBRE LA RELACIÓN ENTRE ACTIVIDAD COMERCIAL Y DELITOS
DE MAYOR CONNOTACIÓN SOCIAL**

Aplicado a transacciones en comercios físicos por manzana censal.

POR: NACHBAUER RAMOS, FERNANDO Y VEGA GAJARDO, JAVIER

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

PROFESOR GUÍA:

Dra. Daniela Opitz

Noviembre 2023

SANTIAGO

A nuestras familias, amigos y seres queridos.

A mi hija; el motor de mi vida.

“Si pones agua en una botella se convierte en la botella. Si la pones en una tetera se convierte en la tetera. El agua puede fluir o puede golpear. Yo soy agua, sé como el agua mi amigo”

Bruce Lee.

AGRADECIMIENTO

En esta etapa tan importante del desarrollo profesional, no hay que olvidar a las grandes personas que fueron parte del camino. En primer lugar, y de forma especial, el más sincero agradecimiento es para mi compañero de trabajo Javier Zacarías Vega Guajardo, quien ha tenido el liderazgo y el compromiso incansable para enfrentar los desafíos aportando siempre de forma significativa al logro en este trabajo. Gracias a su colaboración y constante entrega, es posible contar con un trabajo que nos orgullece como equipo.

Por otro lado, quiero destacar el apoyo incondicional de nuestras familias que de forma directa e indirectamente aportaron para que hoy podamos superar esta etapa de crecimiento profesional.

Primero agradecer especialmente a mi padre Ramón Nachbauer Contreras por creer en mí y apoyarme no solo en lo moral si no que económicamente para ser hoy un profesional de excelencia. En segundo lugar, con mucho cariño y gratitud, brindando su colaboración en momentos difíciles donde había que cuidar a la hija recién nacida, mientras el foco era sacar adelante cada asignatura; gracias, Jocelyn Fernández por darme la tranquilidad necesaria siendo una excelente madre.

A su vez recordar con gratitud a los compañeros de trabajos especialmente a David Orellana e Ignacio Fuentes quienes fueron parte importante como equipo de trabajo en nuestro paso por el magister, entregando su voluntad y buena disposición para concretar cada desafío académico. A nuestros profesores por su dedicación y compromiso en la formación de los futuros científicos de datos del país y a nuestra profesora guía por su apreciada orientación en este proceso de finalización ¡Muchas gracias!

Fernando Iván Nachbauer Ramos.

AGRADECIMIENTO

Un profundo agradecimiento a mis padres, Daniza y Francisco a quienes les debo todo y son los responsables de todos los logros y éxitos que me convierten en el hombre que soy.

A mis hermanas Francesca y Maylin, quienes son mis mejores amigas acompañándome incansablemente en éxitos y fracasos.

A mi ahijada Antonia, que estas palabras sirvan de inspiración para todo el futuro brillante que la vida te depara.

A mi novia Succelle, quien me apoyó en cada paso de este camino, en el día a día que a veces es tan duro y me animó a siempre creer en mí.

A mi compañero Fernando, por toda la disposición, actitud positiva, conocimiento y confianza entregadas durante el transcurso de los cursos y este proyecto.

A los profesores, por su entrega y de dedicación al transmitir con pasión el conocimiento. A nuestra profesora guía, por acompañarnos y confiar en el trabajo realizado durante este proyecto.

Finalmente, a Dios, a quien en mi profunda fe agradezco todo lo bueno, todo lo malo, todo lo pasado y todo lo que vendrá.

Javier Zacarías Vega Gajardo.

TABLA DE CONTENIDO

RESUMEN.....	1
1. INTRODUCCIÓN	2
2. TRABAJO RELACIONADO	3
3. HIPÓTESIS Y OBJETIVOS.....	5
3.1. HIPÓTESIS.....	5
3.2. OBJETIVOS	5
4. DATOS Y METODOLOGÍA.....	6
4.1. DATOS	6
4.1.1. <i>Fuentes de Datos</i>	6
4.1.2. <i>Preparación, Limpieza y Enriquecimiento de Datos</i>	7
4.2. METODOLOGÍA.....	10
4.2.1. <i>Análisis exploratorio (EDA)</i>	10
4.2.2. <i>Validación Estadística de las relaciones entre variables</i>	11
4.2.3. <i>Visualización Temporal y Correlación Móvil</i>	12
4.2.4. <i>Análisis de sensibilidad con XGBoost</i>	12
5. RESULTADOS	13
5.1. EXPLORACIÓN Y ANÁLISIS GENERAL DE CORRELACIONES.....	13
5.2. CORRELACIÓN MÓVIL.....	27
5.3. ANÁLISIS DE SENSIBILIDAD CON XGBOOST.....	28
5.4. ANÁLISIS FOCALIZADO EN LAS PRINCIPALES COMUNAS.....	31
5.4.1. <i>Selección de Comunas con Mayor Actividad Comercial</i>	31
5.4.2. <i>Correlación en Comunas Seleccionadas</i>	32
5.4.3. <i>Análisis de sensibilidad para comunas TOP actividad comercial</i>	32

6.	CONCLUSIONES	34
6.1.	CONCLUSIONES PRINCIPALES	34
6.2.	TRABAJO FUTURO.....	35
6.3.	CONTRIBUCIONES Y APLICACIONES.....	36
7.	BIBLIOGRAFÍA.....	37

Resumen

Este estudio explora la relación entre la actividad comercial física y los índices de Delitos de Mayor Connotación Social (DMCS) en las comunas de Chile, un tema de alto interés en el contexto actual del país donde la sensación de inseguridad va en crecimiento (Sanhueza, 2023) acompañada de un escenario general de estancamiento de la actividad económica (Vergara, 2023).

Enfocándose en el período 2021-2022, la investigación utiliza un conjunto de datos que combina transacciones financieras proporcionadas por Transbank, información demográfica del censo de 2017, y estadísticas oficiales de DMCS obtenidos desde la biblioteca del congreso nacional.

Los resultados muestran una relación compleja y dinámica entre la actividad comercial y los índices DMCS. Mediante el uso de técnicas de Machine Learning y ciencia de datos, se logra establecer una relación entre la actividad económica de las comunas con el índice de DMCS, fundamentado en distintas correlaciones estadísticas lineales, pero principalmente por los resultados de los análisis de sensibilidad, donde la incorporación de variables de actividad económica mejora en proporciones considerables la performance de los modelos de predicción del índice DMCS en comparación a aquellos que no las consideran.

Este estudio aporta a la comprensión de las interacciones entre la economía y la seguridad pública en el contexto urbano chileno, ofreciendo insights que podrían ser valiosos para la formulación de políticas y estrategias de desarrollo urbano. Las conclusiones subrayan la importancia de considerar las dinámicas económicas locales, por ejemplo, en la planificación de la seguridad y la prevención del delito.

1. Introducción

En la era actual, caracterizada por una rápida urbanización y una búsqueda incesante del crecimiento económico, la relación entre la actividad comercial en áreas urbanas con distintos indicadores de seguridad pública adquiere una importancia significativa. En Chile, como en muchos otros países, las dinámicas urbanas y económicas están intrínsecamente ligadas a patrones de criminalidad y seguridad, sobre todo cuando hablamos del comercio físico.

Según el informe publicado “Claves Ipsos” durante 2023, considerando un periodo de 12 meses, un 85% de las personas encuestadas declaró que los problemas de seguridad pública han aumentado, mientras que un 12% dijo que se han mantenido igual y sólo un 3% cree que han disminuido.

De los principales problemas de seguridad que aquejan a los chilenos, un 57% mencionó la categoría “asaltos, robos, portonazos y delincuencia urbana” como la más importante, aumentando 15 puntos porcentuales en relación con 2022.

En este sentido, la comprensión de cómo la actividad comercial física se relaciona con los índices de delitos nos podrían otorgar datos de alto valor para, por ejemplo, el desarrollo de políticas públicas efectivas, la planificación urbana, y la promoción de entornos comerciales seguros y prósperos.

2. Trabajo Relacionado

Los argumentos teóricos y empíricos sugieren que el temor a ser víctima de un delito provoca que los consumidores, trabajadores y emprendedores modifiquen sus actividades cotidianas (Wilcox, Land, & Hunt, 2017; Hamermesh, 1999) El crimen y los cambios en el comportamiento resultantes aumentan el costo de hacer negocios en una localidad y, en última instancia, afectan la trayectoria de desarrollo de todo el vecindario (Greenbaum, 2004).

Es importante atender la complejidad de la relación entre crimen y crecimiento económico. Por ejemplo, en Estados Unidos la relación entre ingreso y crimen sigue un patrón de Curva de Kuznets en forma de U invertida desde 1970: el crimen aumenta con el ingreso hasta alcanzar un nivel máximo, a partir del cual empieza a decrecer (Zuñiga Sergio, 2015). Esto sugiere que la relación entre ingreso y crimen no es lineal, y que otros factores pueden influir en cómo el crimen se relaciona con el crecimiento económico en diferentes etapas de desarrollo de una sociedad.

Varias teorías pueden racionalizar este fenómeno, incluyendo aumentos en las oportunidades criminales, acceso a ingresos disponibles para actividades que complementan el crimen y cambios en la población. Sin embargo, no está claro en qué medida cada una de estas teorías explica este enigma, especialmente dado que los cambios en el crimen suelen observarse a nivel agregado (Brittany, 2019).

En el marco de la actualidad, recientes investigaciones han comenzado a utilizar datos de tarjetas y otras trazas digitales como medios innovadores para estudiar distintos fenómenos sociales. Por ejemplo, algunos enfoques han utilizado datos de telefonía móvil para explorar la movilidad humana y transacciones con tarjetas de crédito para perfilar comportamientos socio económicos de la población a gran escala (Di Clemente, y otros, 2018), proporcionando una nueva perspectiva sobre cómo las distintas aristas de la denominada huella digital pueden aplicarse a estudios de índole social.

Otras investigaciones recientes analizan cómo las trazas digitales podrían ayudar a comprender la dinámica de la criminalidad, al vincular datos de registros de teléfonos móviles con características sociales y adquisitivas como la independencia económica, escolaridad y segregación ocupacional de las mujeres (Contreras, y otros, 2023).

Con este estudio, entonces, se busca comenzar a entender como el crimen y la actividad económica (en este caso dada por actividad comercial en locales físicos) se relacionan en nuestro contexto local, aprovechando los datos de transacciones físicas en locales comerciales con tarjetas de crédito/débito, abordando las complejidades que esto representa y con enfoque en la búsqueda de oportunidades para generar información de valor en el campo.

3. Hipótesis y Objetivos

3.1. Hipótesis

Hipótesis Nula (H0): No existe una relación entre la actividad comercial física y los DMCS.

Hipótesis Alternativa (H1): Existe una relación entre la actividad comercial física y los DMCS.

3.2. Objetivos

Objetivo General:

Analizar la relación entre la tasa de delitos (DMCS) y la actividad comercial, medida a través de transacciones en comercios físicos en diferentes comunas.

Objetivos Específicos:

OE1: Identificar y cuantificar la actividad comercial a nivel país utilizando las variables cantidad de comercios, cantidad de transacciones y cantidad de tarjetas.

OE2: Analizar la variabilidad en la actividad comercial en relación con la tasa de delitos (DMCS) a lo largo del tiempo y en comunas diferentes.

OE3: Evaluar la robustez de la hipótesis H0 y H1 mediante técnicas estadísticas y de modelado apropiadas.

4. Datos y Metodología

4.1. Datos

4.1.1. Fuentes de Datos

Datos de Transacciones de Transbank: Se utilizó como fuente principal de la investigación el conjunto de datos Transbank, que contiene información agregada sobre transacciones o compras realizadas en comercios físicos con tarjeta de crédito/débito entre 06-2021 y 05-2022 en todas las comunas de Chile mediante dispositivos POS. Estos datos incluyen información sobre la cantidad de comercios (cnt_comercios), tarjetas utilizadas (cnt_tarjetas) y transacciones realizadas (cnt_trx) acumuladas por día.

Datos del Censo 2017 (BCN, 2021-2022): Se incorporaron datos demográficos del Censo realizado el año 2017, proporcionando información sobre cantidad de personas, género, escolaridad y nacionalidad a nivel de comuna.

Datos de Delitos de Mayor Connotación Social (DMCS): Los DMCS son la tipificación más común de delitos en el país, caracterizados por la violencia y agresión de los delincuentes; entre ellos se encuentra el robo con violencia e intimidación, las lesiones y el robo en lugar habitado o del vehículo. Afectan al mismo tiempo la integridad y propiedad de las personas, añaden a la pérdida de bienes el riesgo de sufrir daños, especialmente cuando los delincuentes actúan en grupo. Esta tipología de delitos es fuente de interés inagotable y es la causa de iniciativas de ley, proyectos de intervención barrial, programas de televisión y manifestaciones sociales (Padilla, Rodríguez, & Espinosa, 2022). Se utilizaron, entonces, los datos oficiales 2021-2022 sobre delitos de mayor connotación social (tasa por cada 100.000 Habitantes) obtenidos desde la Biblioteca del Congreso Nacional.

4.1.2. Preparación, Limpieza y Enriquecimiento de Datos

Los datos de transacciones se limpiaron para eliminar entradas no estructuradas y posibles valores nulos. Se realizaron transformaciones a nivel de texto para normalizar y estandarizar los nombres de las columnas y se convirtieron los tipos de datos según correspondiera. Una apreciación importante respecto del dataset radica en que su nivel de agregación desde el punto de vista geográfico es la manzana, esto básicamente nos fuerza a enriquecer los datos con alguna otra fuente que nos proporcione la capacidad de ubicar las transacciones a niveles agregados más conocidos (comunas o regiones)

	manzent	fch_transaccion	cnt_comercios	cnt_tarjetas	cnt_trx
0	13119101001039	2021-11-13	16	177	201,0
1	740101	2021-11-04	72	1355	1446,0
2	0	2021-09-07	36339	1374885	1739572,0
3	8301201006025	2021-09-09	1	5	5,0
4	8102111006025	2021-10-13	3	89	89,0
5	1310403	2021-10-24	9	107	112,0
6	560501	2021-06-10	10	235	252,0

Tabla 1: Dataset Transbank

Así, se aplicaron técnicas de preprocesamiento para integrar adecuadamente los datos de transacciones con la información demográfica del censo. Esto incluyó la conversión de formatos, la corrección de inconsistencias, la eliminación de acentos y la estandarización de nombres de comunas. Todo para garantizar una integración precisa y consistente de los conjuntos de datos mediante el mapeo de la unidad de manzana.

fch_transaccion	cnt_comercios	cnt_tarjetas	cnt_trx	ID_MANZENT	REGION	PROVINCIA
2021-11-13	16	177	201.0	13119101001039.0	METROPOLITANA	131.0
2021-11-04	72	1355	1446.0	nan	NaN	NaN
2021-09-07	36339	1374885	1739572.0	nan	NaN	NaN
2021-09-09	1	5	5.0	8301201006025.0	BIOBIO	83.0
2021-10-13	3	89	89.0	8102111006025.0	BIOBIO	81.0

Tabla 2: Dataset Transbank enriquecidos (Censo 2017)

Como se aprecia en la tabla 2, pese a los esfuerzos de preprocesamiento, no todos los registros contienen datos de Censo 2017. Tras eliminarlos, a nivel cuantitativo durante esta etapa se perdió el 7% de los registros del dataset Transbank.

En una dinámica similar, al incorporar los datos de DMCS (anuales por comuna) se aplicaron técnicas de preprocesamiento de texto para, esta vez, mapear a nivel de comuna y quedar con nuestro dataset final enriquecido y listo para el análisis. Durante esta etapa se mantuvo la integridad de registros quedando con una data set de 7.8MM de registros.

	fch_transaccion	comuna	cnt_comercios	cnt_tarjetas	cnt_trx	region	area	personas	hombres	inmigrantes	escolaridad	dmcs_2021
0	2021-06-01	algarrobo	104	3013	3329.0	VALPARAISO	URBANO	2822.0	1364.5	97.0	13.981183	2610.779995
1	2021-06-01	alhue	9	116	123.0	METROPOLITANA	URBANO	279.0	148.0	4.0	16.378802	716.560510
2	2021-06-01	altohospicio	168	4629	4900.0	TARAPACA	URBANO	16567.0	8222.0	1620.0	12.228614	1916.694634
3	2021-06-01	ancud	153	4169	4481.0	LOS LAGOS	URBANO	6759.0	3266.0	79.0	12.184812	760.107309
4	2021-06-01	andacollo	22	385	418.0	COQUIMBO	URBANO	561.0	289.5	0.0	9.123415	1269.250296
5	2021-06-01	angol	273	8175	9089.0	ARAUCANIA	URBANO	9576.0	4433.5	53.0	10.907608	1403.471371
6	2021-06-01	antofagasta	1801	66889	74026.0	ANTOFAGASTA	URBANO	112587.0	55586.5	10702.0	13.325609	1448.887741
7	2021-06-01	antuco	13	449	505.0	BIOBIO	URBANO	495.0	233.0	0.0	11.199853	974.025974
8	2021-06-01	arauco	126	4093	4422.0	BIOBIO	URBANO	4540.0	2247.0	0.0	11.867722	1454.770564
9	2021-06-01	arica	647	19687	21861.0	ARICA Y PARINACOTA	URBANO	29467.0	14325.0	2199.0	12.914882	1533.124664
10	2021-06-01	aysen	126	3670	3991.0	AYSEN	URBANO	4100.0	2054.0	24.0	9.910973	1056.408212
11	2021-06-01	buin	450	18857	20991.0	METROPOLITANA	URBANO	29809.0	14586.5	555.0	11.835593	1231.975986
12	2021-06-01	cabildo	66	1659	1788.0	VALPARAISO	URBANO	1522.0	716.5	24.0	11.368991	666.183925
13	2021-06-01	cabodehornos	10	337	406.0	MAGALLANES	URBANO	349.0	202.0	0.0	10.875909	605.143722

Tabla 3: Dataset enriquecido

Finalmente, se agrupan las variables del set de datos por comuna resultando un conjunto de datos de 107114 filas que se describen como sigue:

- **fch_transaccion:** Fecha asociada a los datos de actividad comercial. En el dataset final quedan referentes a la unidad territorial de comuna.
- **comuna:** Comuna de Chile asociada a todas las variables agrupadas por fila.
- **cnt_comercios:** Cantidad de comercios en los que se realizaron compras físicas para la fecha y comuna indicadas en cada fila.
- **cnt_tarjetas:** Cantidad de tarjetas que realizaron compras físicas en distintos comercios para la fecha y comuna indicadas en cada fila.
- **cnt_trx:** Cantidad de transacciones que se realizaron mediante compras físicas en distintos comercios para la fecha y comuna indicadas en cada fila.
- **región:** Unidad territorial en las que se agrupan las comunas.
- **personas:** Cantidad de personas registradas por el CENSO 2017 para la comuna indicada en la fila.
- **hombres:** Cantidad de hombres registrados por el CENSO 2017 para la comuna indicada en la fila.
- **inmigrantes:** Cantidad de inmigrantes registrados por el CENSO 2017 para la comuna indicada en la fila.
- **escolaridad:** Nivel de escolaridad promedio registrado por el CENSO 2017 para la comuna indicada en la fila.

- **DMCS:** Tasa de denuncias de delitos de mayor connotación social por cada 100000 habitantes para cada comuna con granularidad temporal anual.

	fch_transaccion	comuna	cnt_comercios	cnt_tarjetas	cnt_trx	region	personas	hombres	inmigrantes	escolaridad	dmcs
0	2021-06-01	algarrobo	104	3013	3329.0	VALPARAISO	2822.0	1364.5	97.0	13.981183	2610.779995
1	2021-06-01	alhue	9	116	123.0	METROPOLITANA	279.0	148.0	4.0	16.378802	716.560510
2	2021-06-01	altohospicio	168	4629	4900.0	TARAPACA	16567.0	8222.0	1620.0	12.228614	1916.694634
3	2021-06-01	ancud	153	4169	4481.0	LOS LAGOS	6759.0	3266.0	79.0	12.184812	760.107309
4	2021-06-01	andacollo	22	385	418.0	COQUIMBO	561.0	289.5	0.0	9.123415	1269.250296
...
107109	2022-05-31	vinadelmar	1588	34710	38022.0	VALPARAISO	67904.0	31679.5	3278.0	13.747708	2367.697510
107110	2022-05-31	vitacura	984	21564	24163.0	METROPOLITANA	37680.0	16864.5	4130.0	14.850534	2496.646220
107111	2022-05-31	yerbasbuenas	14	166	174.0	MAULE	921.0	432.0	0.0	9.979868	1040.225468
107112	2022-05-31	yumbel	28	466	495.0	BIOBIO	1285.0	696.0	11.0	10.742240	2110.579959
107113	2022-05-31	zapallar	33	450	478.0	VALPARAISO	1267.0	636.0	43.0	11.990871	2367.586039

Tabla 4: Dataset Final

4.2. Metodología

4.2.1. Análisis exploratorio (EDA)

Cuando se comienza un análisis exploratorio es imprescindible poder entender el contexto de los datos para desarrollar una exploración eficaz y que vaya enfocada en describir los patrones y aspectos más importantes hacia los objetivos del estudio, como lo es determinar relaciones entre el consumo y los DMCS.

En primer lugar, se realiza una inmersión en el comportamiento del dataset para evidenciar posibles tendencias y correlaciones, analizando su estructura y principales estadísticas descriptivas, como también visualizando las relaciones bidimensionales entre múltiples pares de variables, a través de una matriz de dispersión. Esto permitirá encontrar

anomalías, comprender la distribución de los datos, y tener percepciones iniciales para guiar análisis estadísticos más profundos.

Adicionalmente se realizó un análisis temporal de las transacciones, agrupando los datos por fecha y sumando los montos de las transacciones para visualizar la evolución en el tiempo. Se utilizó un gráfico de líneas para representar las tendencias en el monto total de las transacciones y sus distintas relaciones con las demás variables.

Durante este análisis exploratorio se realiza también una matriz de correlación para observar cómo interactuaban las variables del dataset entre ellas y comenzar a observar indicios de relaciones importantes.

Posteriormente se realiza un análisis agregado para generar rankings de las comunas con más actividad comercial y poder detectar tendencias y congruencias con respecto a la percepción de la realidad.

Por último, se realiza una visualización de la densidad comercial por cantidad de comercios a través de un mapa de coropletas de la región Metropolitana.

4.2.2. Validación Estadística de las relaciones entre variables

Se emplearon métodos estadísticos para validar la relación entre DMCS y la actividad comercial:

Correlación de Pearson: Se calculó el coeficiente de correlación y el valor p para determinar la significancia estadística de las relaciones. Esto como una primera aproximación en caso de que hubiera una relación directamente lineal muy fuerte.

Coefficiente de Kendall: Se utilizó para medir la correlación y proporcionar otra perspectiva de la relación entre las variables. Sobre todo, porque este coeficiente no requiere que los datos se distribuyan normalmente, es más robusto a outliers que Pearson y es capaz de cuantificar relaciones que no necesariamente pudieran ser lineales. Todo esto de mucha utilidad para relaciones complejas como la que tratamos de observar.

4.2.3. Visualización Temporal y Correlación Móvil

Para investigar la relación entre las variables de actividad comercial (cnt_comercios, cnt_tarjetas, cnt_trx) y DMCS, se utilizaron técnicas de visualización temporal. Estas visualizaciones revelaron que, a partir de 2022, hay una tendencia decreciente en las variables de actividad comercial, mientras que DMCS muestra una tendencia ascendente.

Se implementó un análisis de correlación móvil para explorar cómo la relación entre DMCS y las variables de actividad comercial cambia con el tiempo. Y de esta forma entender mejor los resultados de correlación y tendencia de cada una de las variables.

4.2.4. Análisis de sensibilidad con XGBoost

Para otorgar más evidencia al principal propósito de esta investigación se implementó un análisis de sensibilidad basado en modelos XGBoost para predecir DMCS utilizando las variables de actividad comercial. Más allá de la calidad predictiva del modelo, lo que nos interesa conocer es si este es capaz de reconocer y capturar la relación compleja que existe entre estas variables, que pudiera no quedar muy clara con métodos de relación lineal, mediante su impacto en la performance de estos.

Este modelo proporcionó una medida de la efectividad de estas variables para predecir DMCS. Se evaluó el rendimiento del modelo utilizando el Mean Squared Error (MSE) y el coeficiente R-squared, visualizando los resultados comparando las predicciones con los valores reales.

5. Resultados

5.1. Exploración y Análisis General de Correlaciones

En primera instancia se realizó un análisis descriptivo de los datos, que ya fueron procesados, limpiados y fueron consolidados en un dataset añadiéndole los datos externos. Hay que considerar que, para efectos de la interpretación de la magnitud de las variables cuantitativas, se está realizando el proceso a nivel de comuna.

Es cuanto, al análisis descriptivo, para la cantidad de comercios sugiere una disparidad considerable en la densidad comercial entre las comunas. Para la variable cantidad de tarjetas el valor máximo es extremadamente alto, lo podría indicar la presencia de outliers o comunas con una actividad financiera excepcionalmente alta. Para el caso de cantidad de transacciones hay una coherencia con lo que se presenta en la cantidad de tarjetas. La variable “personas” indica que algunas comunas son muy densamente pobladas. Para la variable “hombres” su desviación considerable implica una distribución desigual de la población masculina, a su vez, para los inmigrantes, se observa presencia de comunas con concentraciones específicas de poblaciones inmigrantes. Por otro lado, para la escolaridad se reporta que la variabilidad en los años de educación entre las comunas no es tan pronunciada. Por último, hay una variabilidad significativa en la incidencia de delitos y que existen comunas cuyo índice de criminalidad es considerablemente más alto que en otros.

	count	mean	std	min	25%	50%	75%	max
Unnamed: 0	107114.0	53556.500000	30921.292704	0.000000	26778.250000	53556.500000	80334.750000	107113.000000
cnt_comercios	107114.0	226.794033	546.512352	1.000000	17.000000	47.000000	194.000000	10292.000000
cnt_tarjetas	107114.0	10274.279739	29755.875166	1.000000	457.000000	1655.000000	7939.000000	656309.000000
cnt_trx	107114.0	11657.952088	34467.618813	1.000000	502.250000	1805.000000	8680.000000	819931.000000
personas	107114.0	13223.448578	30362.028448	13.000000	649.000000	2059.000000	11241.000000	690885.000000
hombres	107114.0	6416.417620	14903.425630	5.000000	327.000000	1011.250000	5450.000000	354379.000000
inmigrantes	107114.0	1051.593601	6012.508852	0.000000	0.000000	17.000000	248.000000	193308.000000
escolaridad	107114.0	11.520251	1.407023	7.913649	10.544951	11.448564	12.415444	16.378802
dmcs	107114.0	1408.343606	714.046388	293.685756	917.251189	1287.063795	1721.529929	6235.145573

Tabla 5: Describe del dataset

La presencia de valores extremos en distintas variables sugiere que el conjunto de datos podría beneficiarse de una transformación o normalización antes de realizar análisis más complejos. Además, la comparación entre el promedio y la mediana (50%) en algunas de las variables indica la presencia de sesgo en la distribución de los datos, posiblemente hacia la derecha (sesgo positivo), lo que es particularmente notable en variables como cantidad de tarjetas y cantidad de transacciones.

Estos hallazgos preliminares son cruciales para comprender las dinámicas de los distritos y formular hipótesis para análisis predictivos o causales más detallados. Estos resultados invitan a considerar para el análisis posterior, un análisis focalizado en las principales comunas.

Matriz de Dispersión

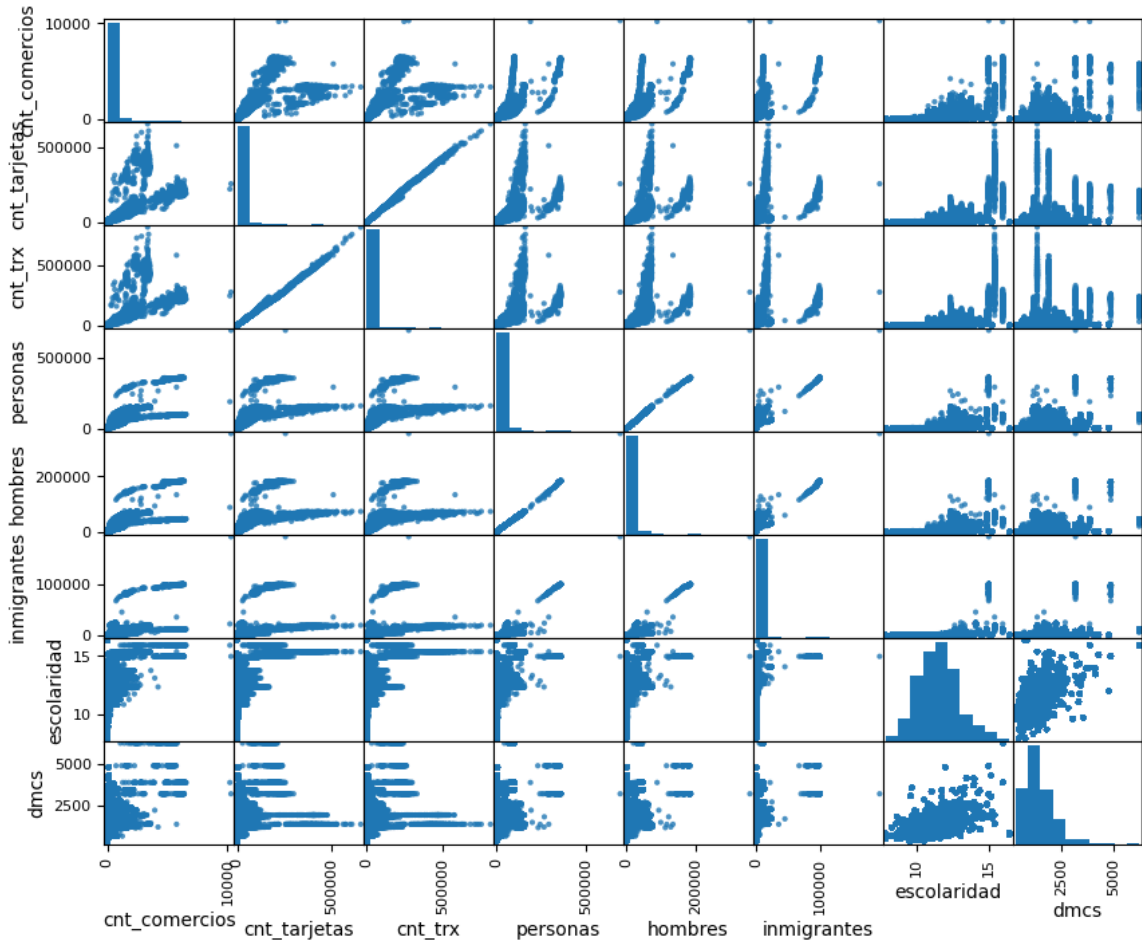


Gráfico 1: Matriz de dispersión

La matriz de dispersión indica relaciones proporcionales entre la cantidad de comercios, tarjetas y transacciones, así como entre la población total y la población masculina, sugiriendo una correlación entre la actividad económica y la demografía de los distritos. Se observa una notable dispersión y presencia de valores atípicos en distintas variables, lo cual destaca la variabilidad entre los distritos. Las distribuciones de las variables están mayormente sesgadas hacia la derecha, particularmente en lo que respecta a la economía y la demografía, mientras que la escolaridad muestra una distribución más uniforme y no

parece relacionarse directamente con las otras variables. La incidencia de delitos de alta connotación social no muestra una correlación lineal clara con la población, lo que sugiere que otros factores podrían influir en su distribución.

Por otra parte, se realiza una exploración del total de transacciones en el tiempo, donde se observa varios patrones interesantes, como valores más elevados en fechas relativas a festividades, valores bottom para fechas posteriores a periodos conocidos de gasto como marzo y leves tendencias a mayor cantidad de transacciones durante los periodos de altas temperaturas. Todo esto nos ayuda en gran parte a otorgarle confiabilidad al dataset ya que expresa tendencias normales y conocidas dando menos margen a la existencia de periodos anómalos dentro del conjunto de datos.

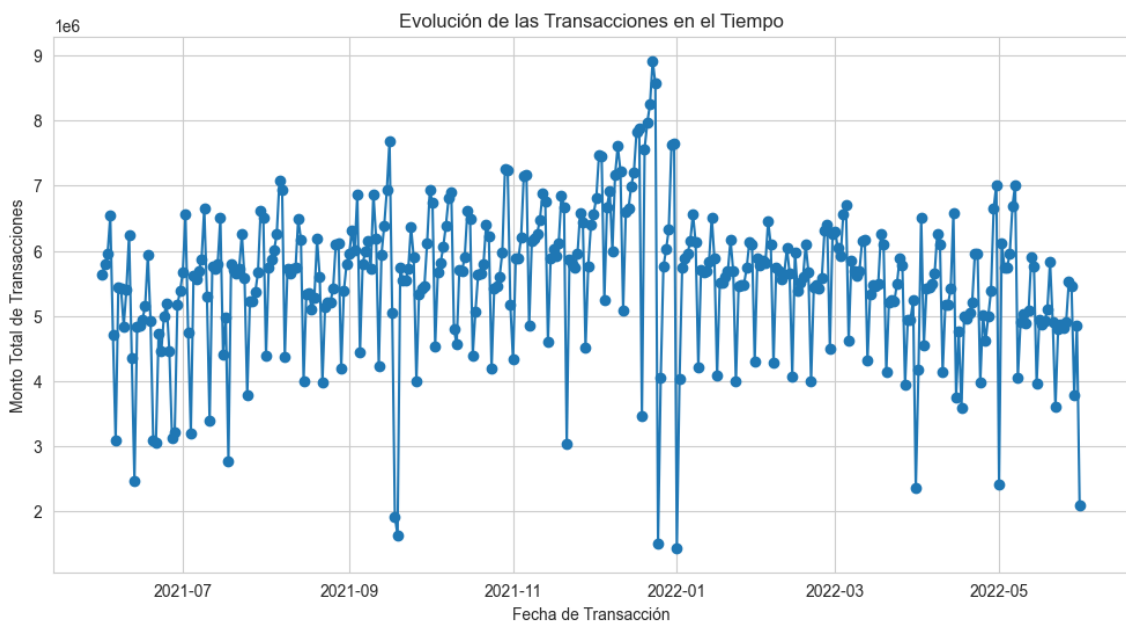


Gráfico 2: Tendencia del total de transacciones en el tiempo

En el gráfico 2, se realiza una exploración del total de transacciones en el tiempo, donde podemos observar varios patrones interesantes, como valores más elevados en fechas relativas a festividades, valores bottom para fechas posteriores a periodos conocidos de gasto como marzo y leves tendencias a mayor cantidad de transacciones durante los periodos de altas temperaturas. Todo esto nos ayuda en gran parte a otorgarle confiabilidad al dataset ya que expresa tendencias normales y conocidas dando menos margen a la existencia de periodos anómalos dentro del conjunto de datos.

Posterior, se realizaron agrupaciones por mes para ver como interactuaban en el tiempo las variables de actividad comercial con DMCS.



Gráfico 3: Transacciones, tarjetas y comercios vs DMCS

Como podemos apreciar, de manera visual y a nivel agregado nacional, se observa en las 3 variables por separado una relación compleja con DMCS. Si bien es cierto entre 2021 y 2022 DMCS tiene un alza y esta parece coincidir con una baja en nuestras 3 variables de actividad comercial, este quiebre no necesariamente representa el comportamiento de los datos estadísticamente como veremos más adelante. Otra consideración importante al respecto radica en que los datos de DMCS se presentan de manera anual, mientras que las transacciones tienen variaciones a nivel diario, esto impacta fuertemente en las interpretaciones visuales que se pueden realizar durante estas comparaciones.

De esta forma, nos adentramos en el uso de técnicas estadísticas para analizar los datos y acercarnos a conclusiones respecto a las distintas relaciones. En primera instancia se realiza una matriz de correlación con el método Pearson para todas las variables.

	cnt_comercios	cnt_tarjetas	cnt_trx	personas	hombres	inmigrantes	escolaridad	dmcs
cnt_comercios	1.000000	0.851372	0.838324	0.857531	0.852933	0.668113	0.454158	0.421173
cnt_tarjetas	0.851372	1.000000	0.999105	0.779860	0.766150	0.559629	0.419824	0.306480
cnt_trx	0.838324	0.999105	1.000000	0.766391	0.752056	0.541806	0.414970	0.297629
personas	0.857531	0.779860	0.766391	1.000000	0.999049	0.800242	0.457912	0.377329
hombres	0.852933	0.766150	0.752056	0.999049	1.000000	0.816914	0.450486	0.377086
inmigrantes	0.668113	0.559629	0.541806	0.800242	0.816914	1.000000	0.301470	0.309612
escolaridad	0.454158	0.419824	0.414970	0.457912	0.450486	0.301470	1.000000	0.514354
dmcs	0.421173	0.306480	0.297629	0.377329	0.377086	0.309612	0.514354	1.000000

Tabla 6: Matriz de correlación (Pearson) dataset final

- **Correlaciones Altas en Variables Económicas:** Hay una correlación muy alta entre cnt_comercios, cnt_tarjetas y cnt_trx, todas superando el valor de 0.83. Esto sugiere una relación fuerte y positiva entre la presencia de comercios, el uso de

tarjetas y el número de transacciones, lo cual es consistente con la idea de que más comercios pueden conducir a un mayor uso de tarjetas y transacciones en un área.

- **Demografía y Economía:** Las variables personas y hombres tienen correlaciones fuertes con las variables económicas mencionadas anteriormente, lo que indica que las áreas con mayores poblaciones tienden a tener una mayor actividad económica en términos de comercios y transacciones.
- **Inmigrantes:** La presencia de inmigrantes (inmigrantes) tiene una correlación moderada con las variables económicas y demográficas, sugiriendo que los distritos con mayores números de inmigrantes podrían tener una actividad económica más dinámica.
- **Escolaridad:** La escolaridad (escolaridad) tiene correlaciones más bajas con todas las variables, especialmente con las económicas, lo que podría indicar que los años de educación no están directamente asociados con la actividad económica o demográfica en los distritos, pero si tiene una correlación especialmente más alta con los delitos, lo que sugiere un punto de atención.
- **Delitos de Alta Connotación Social (DMCS):** Los delitos tienen correlaciones bajas a moderadas con las demás variables, siendo más alta con escolaridad (0.514354) y más baja con cnt_tarjetas (0.306480). Esto podría sugerir que los factores económicos y demográficos no tienen una relación tan directa o fuerte con la incidencia de delitos de alta connotación social como podrían tener con otros factores no observados en esta matriz.

La lógica detrás de esto es inherente a la naturaleza de los datos, ya que encontraremos una mayor cantidad de comercios en zonas altamente pobladas produciendo una mayor cantidad de transacciones. Y como apunta, dentro del contexto nacional, J. Núñez et al. (Núñez, Rivera, Villavicencio, & Molina, 2003) la densidad poblacional está positivamente asociada con el delito de robo y otros.

Adicionalmente, tenemos un indicio de como la escolaridad estaría relacionada con el DMCS, lo cual es plausible en estudios que demuestran que, por ejemplo, un aumento de 0,1 años en la escolaridad promedio de la población entre 18 y 60 tiene asociado entre 9,2 y 11,6 menos delitos por cada 100000 habitantes (Param Fuentealba & Pérez Chávez, 2011).

	cnt_comercios	cnt_tarjetas	cnt_trx	personas	hombres	inmigrantes	escolaridad	dmcs
cnt_comercios	1.000000	0.878242	0.876411	0.789176	0.788709	0.695515	0.433782	0.342543
cnt_tarjetas	0.878242	1.000000	0.991365	0.771054	0.769794	0.680873	0.439416	0.347878
cnt_trx	0.876411	0.991365	1.000000	0.768246	0.767031	0.680322	0.439875	0.347460
personas	0.789176	0.771054	0.768246	1.000000	0.984343	0.751562	0.434461	0.328982
hombres	0.788709	0.769794	0.767031	0.984343	1.000000	0.756902	0.436147	0.330794
inmigrantes	0.695515	0.680873	0.680322	0.751562	0.756902	1.000000	0.516442	0.360984
escolaridad	0.433782	0.439416	0.439875	0.434461	0.436147	0.516442	1.000000	0.347528
dmcs	0.342543	0.347878	0.347460	0.328982	0.330794	0.360984	0.347528	1.000000

Tabla 7: Matriz de Correlación (Kendall) data set final

Para el caso de la correlación de Kendall (tabla 7) podemos ver que existe una fuerte relación entre las variables económicas, así como también en las correlaciones entre variables demográficas. La variable inmigrante está moderadamente correlacionada con las variables económicas y demográficas, lo que podría reflejar que los distritos con mayor número de inmigrantes poseen una dinámica económica particular, posiblemente debido

a diferentes patrones de consumo o de emprendimiento. La variable escolaridad muestra correlaciones relativamente bajas con todas las demás variables, sugiriendo que el nivel educativo promedio no tiene una relación directa o fuerte con la actividad económica o la demografía de los distritos. El DMCS presenta correlaciones bajas a moderadas con las demás variables. La correlación más notable es con la variable escolaridad, lo cual podría indicar una relación entre el nivel de educación y la incidencia de delitos, aunque esta relación no es muy fuerte.

En conjunto, estas correlaciones indican relaciones significativas entre la actividad económica y la demografía de los distritos, mientras que las asociaciones con los delitos de alta connotación social y la escolaridad son menos marcadas, lo que sugiere la influencia de otros factores no capturados en este conjunto de datos.

Continuando con el análisis exploratorio de los datos, se realizó una agregación por comuna y región de las variables comerciales, obteniendo los siguientes resultados:

- Existe una fuerte predominancia de la región Metropolitana en la cantidad de transacciones totales registradas en el dataset.

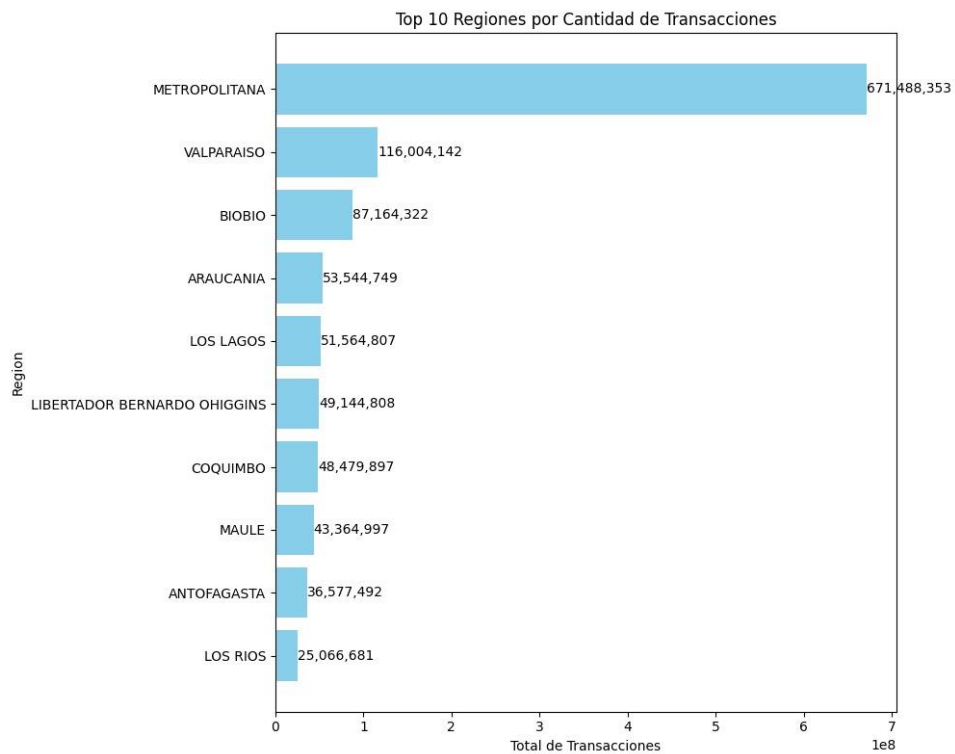


Gráfico 4: Cantidad de transacciones por región

A su vez, en concordancia con lo anterior, se observa que las 3 comunas con más transacciones pertenecen a la región Metropolitana.

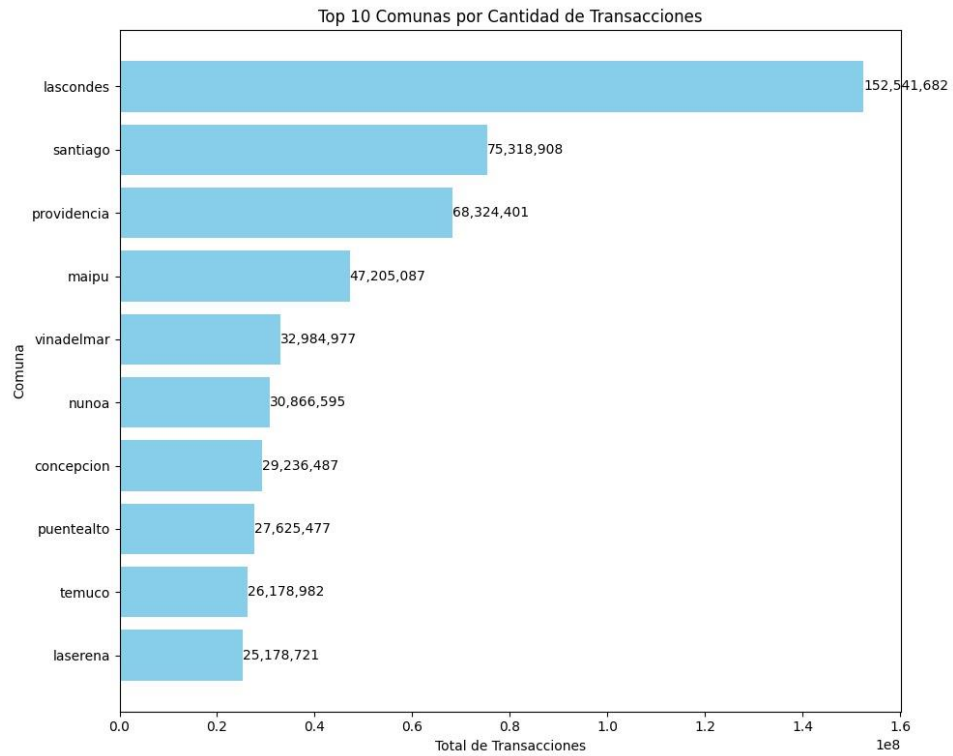


Gráfico 5: Cantidad de transacciones por comuna

Continuando con la exploración se construyeron tres rankings para las comunas con mayor cantidad de comercios, cantidad de tarjetas y cantidad de transacciones. Estos resultados fueron calculados utilizando el promedio y no la totalidad de registros como lo anterior.

comuna	region	cnt_comercios	comuna	region	cnt_tarjetas	comuna	region	cnt_trx			
201	providencia	METROPOLITANA	4976.268493	115	lascondes	METROPOLITANA	354740.082192	115	lascondes	METROPOLITANA	417922.416438
264	santiago	METROPOLITANA	4766.813699	264	santiago	METROPOLITANA	188496.583562	264	santiago	METROPOLITANA	206353.172603
115	lascondes	METROPOLITANA	2737.950685	201	providencia	METROPOLITANA	164292.887671	201	providencia	METROPOLITANA	187190.139726
145	maipu	METROPOLITANA	2041.156164	145	maipu	METROPOLITANA	104470.210959	145	maipu	METROPOLITANA	129329.005479
49	concepcion	BIOBIO	1801.213699	291	vinadelmar	VALPARAISO	80830.873973	291	vinadelmar	VALPARAISO	90369.800000
291	vinadelmar	VALPARAISO	1714.008219	166	nunoa	METROPOLITANA	75377.501370	166	nunoa	METROPOLITANA	84566.013699
6	antofagasta	ANTOFAGASTA	1425.934247	49	concepcion	BIOBIO	71938.095890	49	concepcion	BIOBIO	80099.964384
205	puntealto	METROPOLITANA	1332.441096	205	puntealto	METROPOLITANA	65865.920548	205	puntealto	METROPOLITANA	75686.238356
166	nunoa	METROPOLITANA	1317.980822	271	temuco	ARAUCANIA	63823.084932	271	temuco	ARAUCANIA	71723.238356
116	laserena	COQUIMBO	1222.191781	6	antofagasta	ANTOFAGASTA	59802.975342	116	laserena	COQUIMBO	68982.797260

Tabla 8: Ranking por comuna de variables de actividad comercial

En los resultados se pueden apreciar claras diferencias en la actividad comercial para las comunas pertenecientes a la región Metropolitana. En particular existe una mayor cantidad de comercios en Providencia, seguido de Santiago y Las Condes. Para lo que es cantidad de transacciones y cantidad de tarjetas se puede apreciar la dominancia de Las Condes como la comuna con mayor actividad.

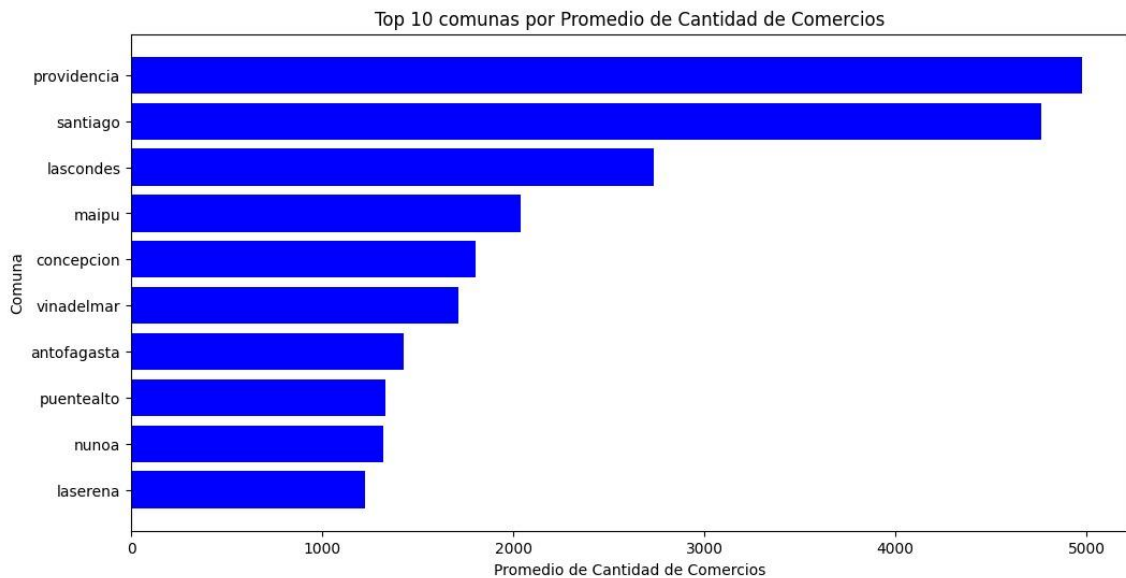


Gráfico 6: Cantidad de comercios promedio por comuna

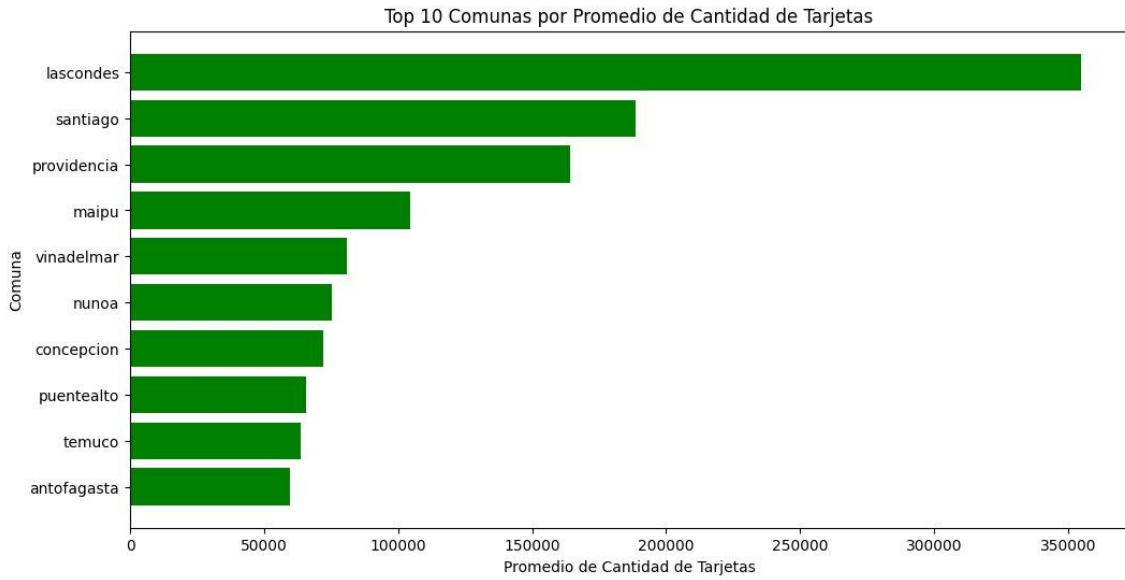


Gráfico 7: Promedio de cantidad de tarjetas por comuna

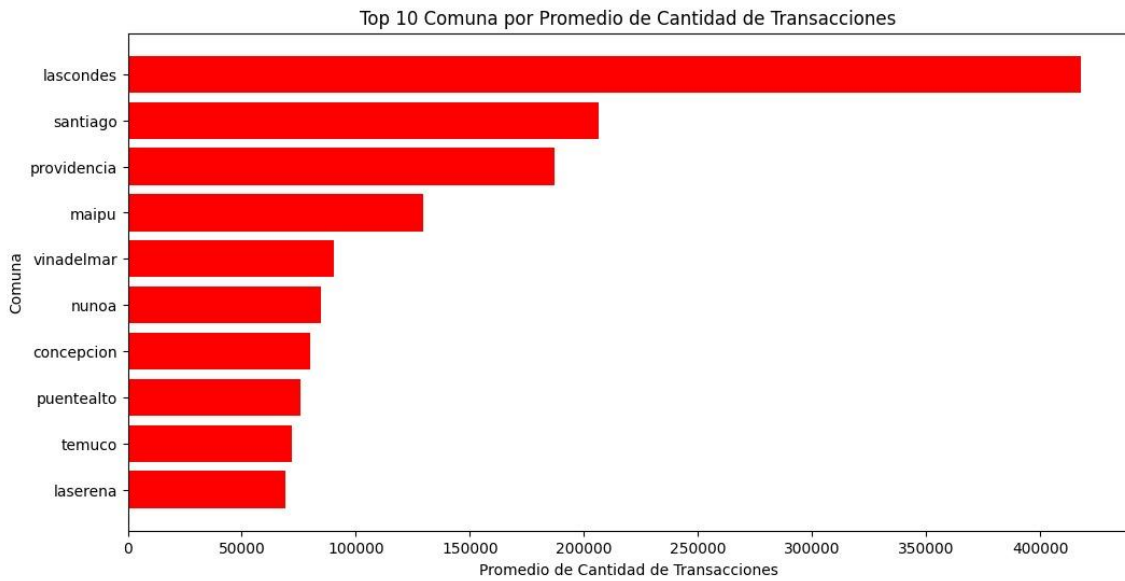


Gráfico 8: Promedio cantidad de transacciones por comuna

Finalmente, y dado que la mayor parte de la actividad comercial se realiza en la región Metropolitana, será interesante visualizar como geográficamente esto se ve reflejado en la cantidad de comercios. Por lo que se construyó un mapa de coropletas:

Total de Comercios en la Región Metropolitana de Santiago por Comuna (Top 10)

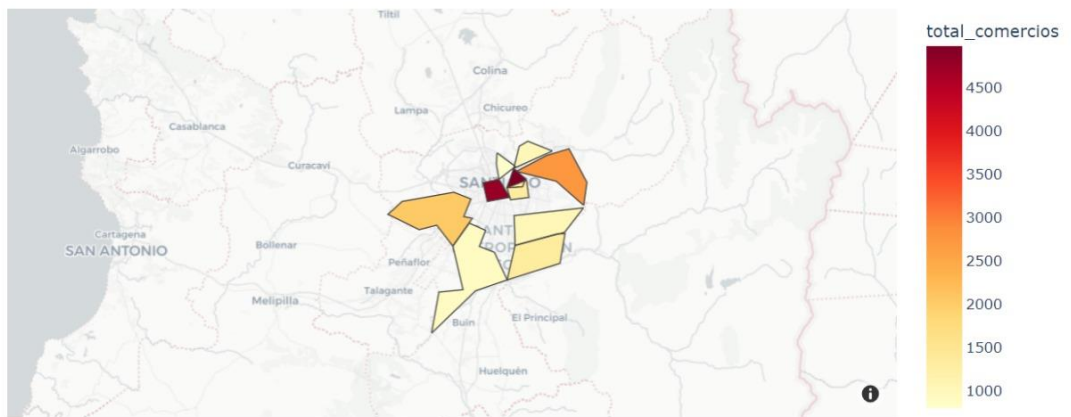


Gráfico 9: Mapa de calor por cantidad de comercios en RM (Top 10)

El mapa indica que tanto en Providencia, Santiago y Las Condes, tienen la mayor actividad comercial.

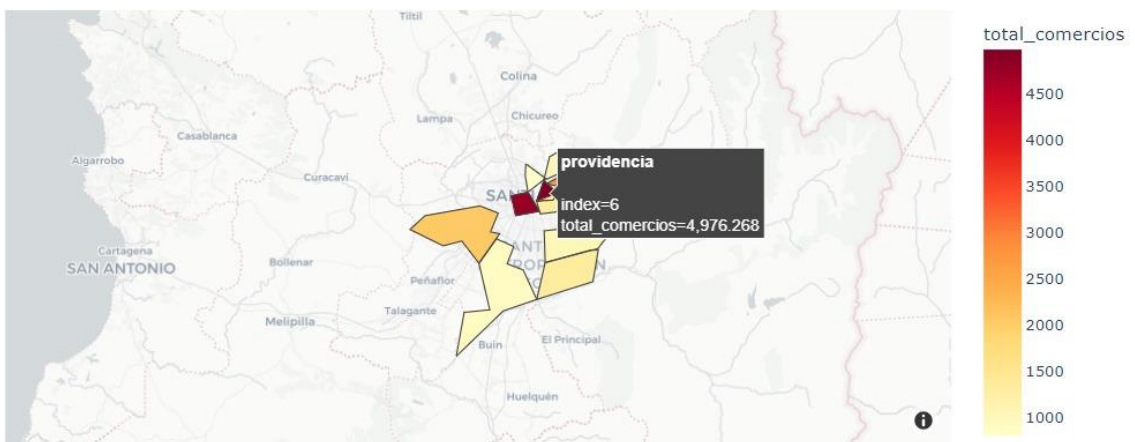


Gráfico 10: Comuna con mayor cantidad de comercios promedio

5.2. Correlación Móvil

El método de correlación móvil reveló patrones subyacentes en las variables comerciales y DMCS, mostrando componentes de tendencia y estacionalidad. Esta se examinó en ventanas de 3 meses, revelando las fluctuaciones en la relación entre ambas, con períodos donde la correlación se volvía negativa, explicando las discrepancias observadas en el análisis visual.

En el gráfico 11, podemos observar como la cantidad de comercios muestra una tendencia al alza en su correlación con DMCS hasta noviembre 2022, donde logra su valor más alto para luego volver a su correlación negativa, aunque con algunos signos de recuperación. Esta componente puede tener una influencia fuerte estacional respecto de por ejemplo comercios flotantes que se establecen por periodos de tiempo determinados (navidad, fiestas patrias, etc.).

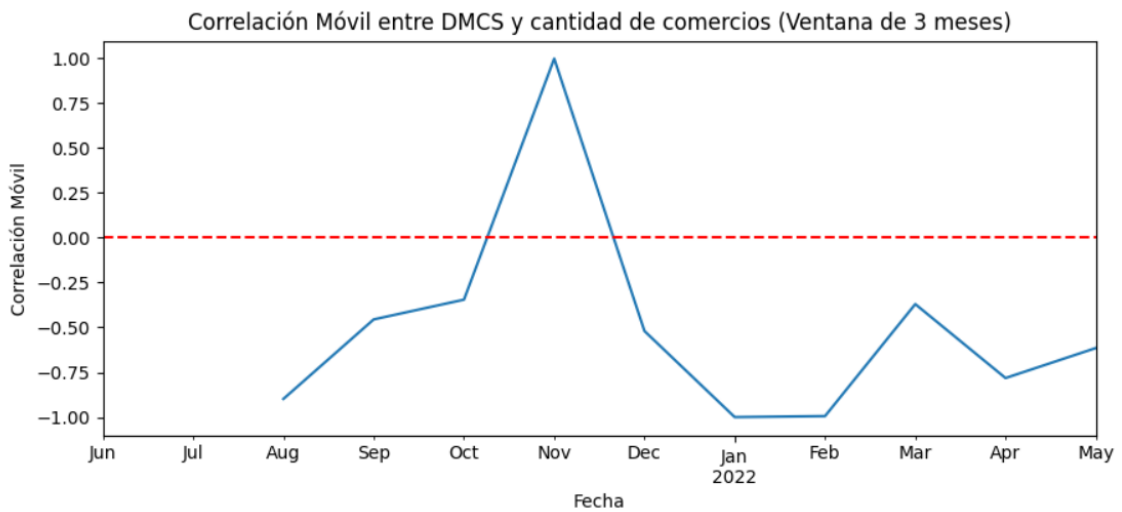


Gráfico 11: Correlación móvil entre cantidad de comercios y DMCS

En lo que respecta a la cantidad de transacciones, podemos ver ciclos mucho más marcados que básicamente se modelan por el comportamiento de la cantidad de transacciones, lo importante es que a partir de 2022 (que sabemos marca un alza en el DMCS) la tendencia cambia y no se visualizan picos de correlaciones positivas como durante el año anterior.

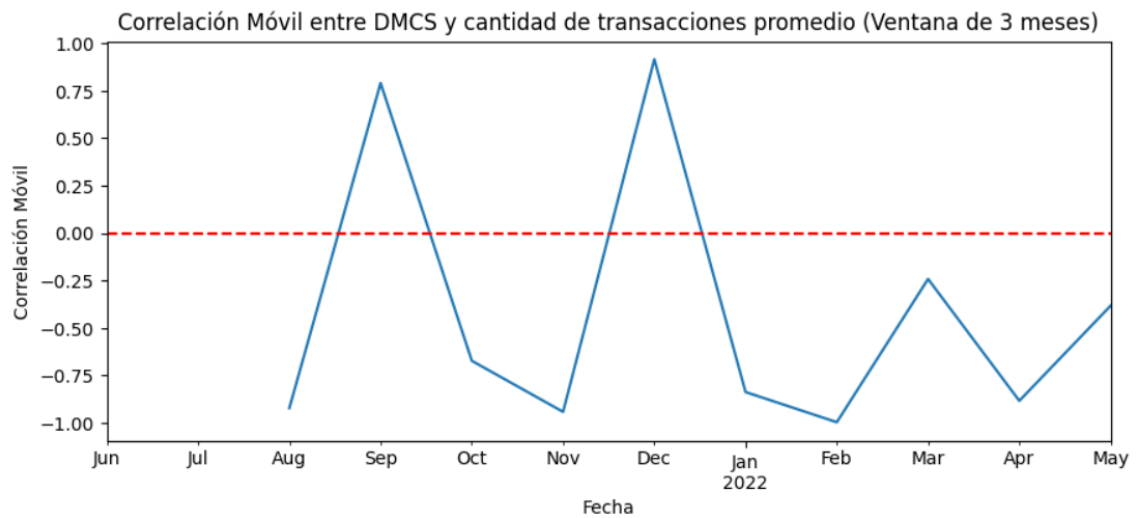


Gráfico 12: Correlación móvil entre cantidad de transacciones y DMCS.

5.3. Análisis de sensibilidad con XGBoost

XGBoost es capaz de capturar patrones complejos y no lineales en los datos. Puede aprender de relaciones no solo lineales, sino también interacciones entre múltiples variables, lo que lo convierte en una herramienta poderosa para modelar situaciones del mundo real donde las relaciones son intrincadas y no se pueden describir fácilmente con modelos lineales.

El análisis de sensibilidad realizado con el algoritmo XGBoost consistió en evaluar cómo la inclusión o exclusión de ciertas variables afecta el rendimiento predictivo del modelo respecto a la variable dependiente DMCS. Este análisis será nuestra técnica clave al identificar qué tan relacionadas son las predicciones de los distintos modelos a los cambios en las variables de entrada, especialmente al incorporar o quitar aquellas de actividad comercial.

Así, se compararon diferentes combinaciones, incluyendo siempre las variables demográficas, para observar sus efectos en el Error Cuadrático Medio (MSE) y el coeficiente de determinación (R-cuadrado) de los modelos. Las variaciones en estas métricas nos darán un indicio de la importancia relativa de la actividad comercial y su contribución a la capacidad del modelo para explicar la variabilidad en los datos de DMCS.

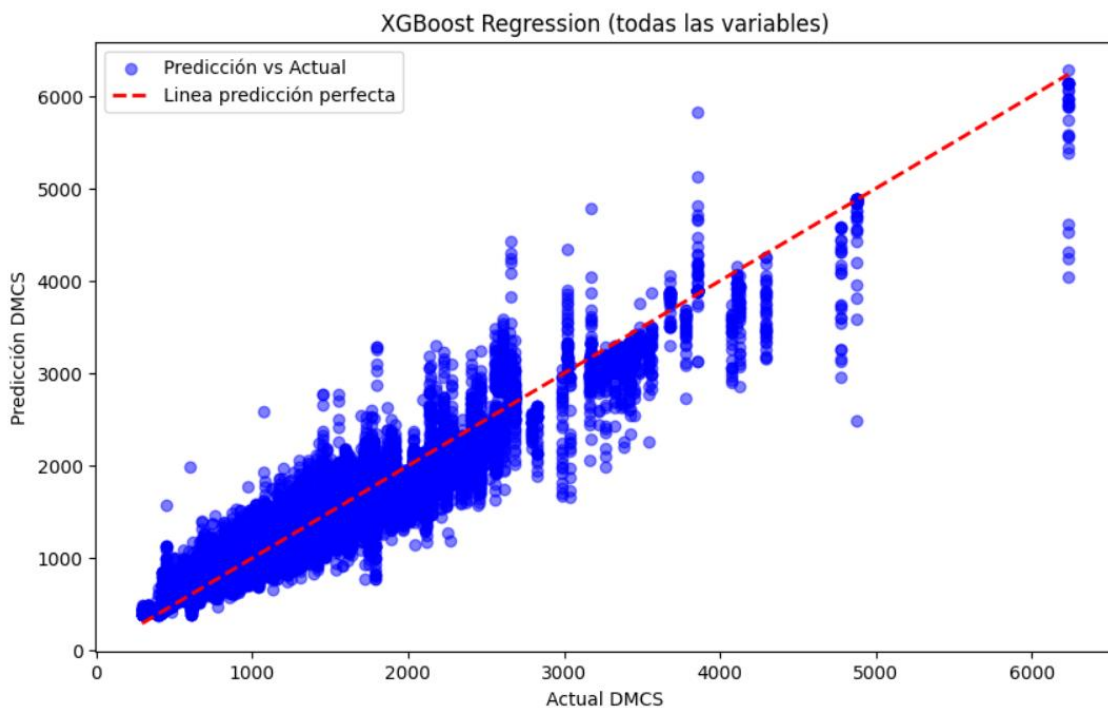


Gráfico 13: Modelo XGBoost DMCS, todas las variables independientes incluidas.

En el gráfico 13, se muestran los resultados de la ejecución del modelo entrenado tomando como variables independientes todo el CDF, los puntos representan las predicciones del modelo; cada punto muestra el valor real de DMCS frente al valor predicho para una observación particular.

La línea roja representa la "línea de predicción perfecta", donde los valores predichos de DMCS serían exactamente iguales a los valores reales. Cuanto más cerca estén los puntos de esta línea, mejor es el rendimiento del modelo.

Análisis de sensibilidad XGBoost		
Variables	MSE	R-squared
'personas', 'hombres','inmigrantes'	134495	0.74
'cnt_comercios','cnt_trx','cnt_tarjetas', 'personas', 'hombres','inmigrantes'	97929	0.81
personas', 'hombres','escolaridad'	75523	0.85
'cnt_comercios','cnt_trx','cnt_tarjetas', 'personas', 'hombres','escolaridad'	60056	0.88
'personas', 'hombres','inmigrantes','escolaridad'	70476	0.86
cnt_comercios', 'cnt_tarjetas', 'cnt_trx', 'personas', 'hombres','inmigrantes','escolaridad'	58484	0.89

Gráfico 14: Modelo predictivo DMCS XGBoost

Al examinar los cambios en el Error Cuadrático Medio (MSE) y el coeficiente de determinación (R-cuadrado), se observa una disminución consistente del MSE y un aumento del R-cuadrado en los modelos que incluyen las variables comerciales en comparación con aquellos que no las incluyen. Específicamente, el modelo que únicamente considera variables demográficas (personas, hombres, inmigrantes) presenta un MSE de 134495 y un R-cuadrado de 0.74. Al añadir las variables comerciales, el MSE disminuye a 97929 y el R-cuadrado aumenta a 0.81, indicando una mejor precisión y capacidad explicativa del modelo.

Esta tendencia se mantiene al incluir la variable escolaridad, donde el modelo con variables demográficas y escolaridad tiene un MSE de 75523 y un R-cuadrado de 0.85,

pero al añadir las variables comerciales, el MSE baja aún más a 60056 y el R-cuadrado sube a 0.88. El modelo más completo, que incluye tanto las variables demográficas, de escolaridad, como las comerciales, proporciona el mejor resultado con un MSE de 58484 y un R-cuadrado de 0.89.

Dados los resultados, podemos establecer que las variables `cnt_comercios`, `cnt_trx` y `cnt_tarjetas` tienen una relación significativa con DMCS, evidenciada por la mejora en los indicadores de desempeño de los modelos cuando estas variables están presentes. Esto sugiere que la actividad comercial tiene un impacto importante en el DMCS. Las mejoras observadas en el R-cuadrado demuestran que las variables comerciales no solo mejoran la precisión de las predicciones, sino que también aumentan la proporción de la varianza explicada por los modelos, reforzando la hipótesis de que capturan efectos sistemáticos y relevantes en la predicción de DMCS.

5.4. Análisis Focalizado en las Principales Comunas

5.4.1. Selección de Comunas con Mayor Actividad Comercial

Se identificaron las cinco comunas con mayor actividad comercial, basándose en el promedio anual de las distintas componentes de actividad comercial. Las comunas seleccionadas fueron Las Condes, Santiago, Providencia, Maipú y Concepción, con los valores promedio más altos.

	comuna	cnt_comercios	cnt_trx	cnt_tarjetas
201	providencia	4976.268493	187190.139726	164292.887671
264	santiago	4766.813699	206353.172603	188496.583562
115	lascondes	2737.950685	417922.416438	354740.082192
145	maipu	2041.156164	129329.005479	104470.210959
49	concepcion	1801.213699	80099.964384	71938.095890

Tabla 9: TOP comunas según actividad comercial

5.4.2. Correlación en Comunas Seleccionadas

Se calculó una matriz de correlación específicamente para estas comunas, revelando un patrón de correlación diferenciado en comparación con el análisis general. Notablemente, se encontró una correlación positiva más fuerte entre la cantidad de comercios y DMCS (0.532158), mientras que las correlaciones entre otras variables de actividad comercial y dmcs presentaron patrones variados.

	cnt_comercios	cnt_tarjetas	cnt_trx	personas	hombres	inmigrantes	escolaridad	dmcs
cnt_comercios	1.000000	0.349488	0.307765	0.469616	0.470199	0.522984	0.619852	0.532158
cnt_tarjetas	0.349488	1.000000	0.997400	0.224322	0.193127	0.200080	0.570889	-0.172154
cnt_trx	0.307765	0.997400	1.000000	0.185680	0.153758	0.154942	0.541583	-0.208813
personas	0.469616	0.224322	0.185680	1.000000	0.999078	0.967339	0.184617	0.150076
hombres	0.470199	0.193127	0.153758	0.999078	1.000000	0.969959	0.170618	0.167061
inmigrantes	0.522984	0.200080	0.154942	0.967339	0.969959	1.000000	0.344915	0.325502
escolaridad	0.619852	0.570889	0.541583	0.184617	0.170618	0.344915	1.000000	0.534076
dmcs	0.532158	-0.172154	-0.208813	0.150076	0.167061	0.325502	0.534076	1.000000

Tabla 10: Análisis de correlación con dataset comunas TOP

Por ejemplo, vemos como la cantidad de tarjetas y transacciones muestra para este grupo una correlación negativa, que pudiera ser más representativa de grandes urbes en comparación con el promedio país. Esto nuevamente nos alerta sobre la complejidad de la relación y como un factor netamente lineal podría no explicar como la actividad comercial afecta la DMCS y viceversa.

5.4.3. Análisis de sensibilidad para comunas TOP actividad comercial

El análisis de sensibilidad aplicado a las comunas con mayor actividad comercial (Comunas TOP) utilizando XGBoost refuerza y complementa las conclusiones previas, mostrando un impacto aún más pronunciado de las variables comerciales (cnt_comercios, cnt_trx, cnt_tarjetas) sobre la variable DMCS.

Análisis de sensibilidad XGBoost - Comunas TOP		
VARIABLES	MSE	R-squared
'personas', 'hombres', 'inmigrantes'	369988	0.83
'cnt_comercios', 'cnt_trx', 'cnt_tarjetas', 'personas', 'hombres', 'inmigrantes'	103740	0.95
'personas', 'hombres', 'escolaridad'	394908	0.81
'cnt_comercios', 'cnt_trx', 'cnt_tarjetas', 'personas', 'hombres', 'escolaridad'	119996	0.94
'personas', 'hombres', 'inmigrantes', 'escolaridad'	322554	0.85
'cnt_comercios', 'cnt_tarjetas', 'cnt_trx', 'personas', 'hombres', 'inmigrantes', 'escolaridad'	105170	0.95

Tabla 11: Análisis de sensibilidad comunas TOP

Análisis de sensibilidad XGBoost - Nacional		
VARIABLES	MSE	R-squared
'personas', 'hombres', 'inmigrantes'	134495	0.74
'cnt_comercios', 'cnt_trx', 'cnt_tarjetas', 'personas', 'hombres', 'inmigrantes'	97929	0.81
'personas', 'hombres', 'escolaridad'	75523	0.85
'cnt_comercios', 'cnt_trx', 'cnt_tarjetas', 'personas', 'hombres', 'escolaridad'	60056	0.88
'personas', 'hombres', 'inmigrantes', 'escolaridad'	70476	0.86
'cnt_comercios', 'cnt_tarjetas', 'cnt_trx', 'personas', 'hombres', 'inmigrantes', 'escolaridad'	58484	0.89

Tabla 12: Análisis sensibilidad, todas las comunas.

Los modelos que incluyen estas variables no solo presentan una disminución dramática en el MSE, sino que también alcanzan un R-cuadrado alto de hasta 0.95. Esto sugiere que, en las comunas con alta actividad comercial, la influencia de las transacciones comerciales, la cantidad de comercios y el uso de tarjetas en el índice de delitos de mayor connotación social es aún más fuerte que en el análisis nacional. Por ejemplo, el modelo solo con variables demográficas (personas, hombres, inmigrantes) tiene un R-cuadrado de 0.83 con un MSE de 369988, mientras que, al incluir las variables de actividad comercial, el R-cuadrado aumenta a 0.95 con una reducción significativa del MSE a 103740. Además, cuando se compara la inclusión de escolaridad con y sin las variables comerciales, se observa que el modelo que las incluye (cnt_comercios, cnt_trx, cnt_tarjetas, personas, hombres, escolaridad) no solo mejora en términos de MSE (119996 comparado con 394908), sino que también alcanza un R-cuadrado muy similar al modelo más completo (0.94 frente a 0.95), lo que indica que las variables comerciales tienen una

capacidad predictiva significativa incluso cuando se toman en cuenta otros predictores importantes como la escolaridad.

La consistencia de estos resultados en un subconjunto de comunas con alta actividad comercial sugiere que las conclusiones son robustas y que las variables comerciales tienen una relación importante con el índice DMCS, especialmente en áreas densas con un dinamismo comercial elevado.

6. Conclusiones

6.1. Conclusiones Principales

Validación de Hipótesis: Los análisis realizados proporcionan evidencia favorable respecto a las hipótesis H1. Determinada principalmente por:

- Correlación positiva moderada de entre 0.30 y 0.40 con p-value de 0.0 para las variables de actividad comercial con la tasa DMCS para el total de los datos según coeficiente de Pearson.
- Correlación positiva moderada de 0.35 con p-value de 0.0 para las variables de actividad económica con la tasa DMCS para el total de los datos según coeficiente de Kendall.
- El análisis de sensibilidad realizado a través de la comparación de modelos XGBoost para la predicción del índice DMCS sugieren que la participación de las variables de actividad comercial impacta significativamente en el desempeño de la predicción del DMCS, disminuyendo entre 11 y 21% el MSE en los distintos modelos que utilizaron las variables de actividad comercial y aumentando hasta en un 9% el R-squared bajo la misma premisa.
- Para el caso del análisis de sensibilidad para comunas top de actividad comercial, se pronuncia el mismo fenómeno, en el que la incorporación de variables de

actividad comercial aumenta la performance de los modelos, incrementando (en el caso de la inclusión de todas las variables independientes) el R-squared de 0.85 a 0.95.

A pesar de los resultados favorables a la hipótesis H1, es importante reconocer la naturaleza compleja de la relación entre ambas variables, que puede estar modelada por una diversidad de componentes sociológicos, temporales, económicos y culturales que nos hablan de una relación no lineal sujeta a una diversidad de factores y que a medida que estos cambien, podría verse afectada.

6.2. Trabajo Futuro

Análisis Extendido a Otras Variables: Se recomienda extender el análisis a otras variables que podrían influir en la relación entre actividad comercial y seguridad, como factores socioeconómicos, tipos de comercios, desagregación de delitos, índices de percepción de seguridad, entre otros.

Estudio Longitudinal Ampliado: Un análisis longitudinal (o temporal) que abarque un período más extenso (en este caso modelado por el periodo de datos Transbank) podría proporcionar una comprensión más profunda de las tendencias a largo plazo y de cómo las dinámicas comerciales y sociales evolucionan con el tiempo, esto cobra suma importancia considerando que desde 2020 a la fecha el escenario nacional ha ido cambiando sustantivamente (COVID-19, estallido social, inmigración, fenómenos políticos, etc) haciendo más difícil extrapolar tendencias en distintos periodos de tiempo.

Validación del Modelo: Es esencial validar el modelo utilizando conjuntos de datos independientes, como a través de la validación cruzada, para asegurar que las mejoras en las métricas no sean resultado de algún sobreajuste.

Avanzar en el modelo predictivo: Explorar mejoras y variaciones a los modelos predictivos implementados que puedan capturar mejor las complejidades y las interacciones no lineales entre las variables y de esta forma consolidar un producto de valor de impacto de cara a la predicción del índice de delitos.

6.3. Contribuciones y Aplicaciones

Los resultados obtenidos apoyan la idea de que la actividad económica local, reflejada en las variables mencionadas, es un determinante significativo en el índice de DMCS y deberían ser consideradas eventualmente en la toma de decisiones y políticas públicas enfocadas en las distintas comunas. Esto puede ser de gran utilidad, por ejemplo, para la planificación de la dotación de personal de seguridad y orden, rutas de vigilancia, planes cuadrante y estrategias de inversión.

Por otra parte, nos permite comprender inicialmente cual es la dinámica de esta relación y cuáles son los elementos determinantes al momento de relacionarlas.

7. Bibliografía

BCN. (2021-2022). *Biblioteca del congreso nacional de Chile - Estadísticas territoriales*. Obtenido de

<https://www.bcn.cl/siit/estadisticasterritoriales/tema?id=63>

Brittany, S. (19 de Sept de 2019). The Impact of Economic Activity on Criminal Behavior: Evidence from the Fracking Boom. *State and Local Policy*, págs. 1-10.

Contreras, H., Candia, C., Troncoso, R., Ferres, L., Bravo, L., Lepri, B., & Rodríguez-Sickert, C. (04 de Diciembre de 2023). Linking physical violence to women's. *EPJ Data Science volume 12, Article number: 54 (2023)*.

Di Clemente, R., Luengo-Oroz, M., Travizano, M., Xu, S., Vaitla, B., & González, M. (Agosto de 2018). Sequences of purchases in credit card data reveal lifestyles in urban populations. *Nature Communications volume 9, Article number: 3330 (2018)*.

Greenbaum. (2004). The impact of violence surges on neighbourhood business activity. *Urban Studies*, 2495-2514.

Hamermesh. (1999). Crime and the timing of work. *Journal Urban Economy*, 311-330.

Núñez, J., Rivera, J., Villavicencio, X., & Molina, O. (Junio de 2003). Determinantes Socioeconómicos y Demográficos del Crimen en Chile. *Estudios de Economía*, 55-85.

- Padilla, G., Rodríguez, C., & Espinosa, D. (2022). Inseguridad ciudadana y delitos de mayor connotación social: formas y extensiones del temor a la delincuencia en Chile. *CUADERNO URBANO. Espacio, cultura, sociedad*, vol. 33, núm. 33.
- Param Fuentealba, C., & Pérez Chávez, F. (2011). *Educación y criminalidad juvenil en Chile*. Santiago: FACULTAD DE ECONOMÍA Y NEGOCIOS.
- Sanhueza, A. (24 de Noviembre de 2023). *EL PAÍS*. Obtenido de <https://elpais.com/chile/2023-11-24/la-percepcion-de-inseguridad-en-chile-llega-al-90-la-mas-alta-en-una-decada.html>
- Vergara, R. (Junio de 2023). El momento económico internacional y nacional. Junio 2023. *PUNTOS DE REFERENCIA N° 662*.
- Wilcox, Land, & Hunt. (2017). *Criminal Circumstance*. New York: Routledge.
- Zuñiga Sergio, R. S. (2015). Crimen, Desempleo y Actividad Económica en Chile. *Revista Brasileira de Políticas Públicas*, 81-99.