

Delitos Municipales y su Impacto en la Gestión de Seguros

con Enfoque en Análisis de Datos

Estefanía Zavala Mendoza



Contenido

❖ Introducción: Contexto del Proyecto

- Se explica el propósito del análisis y su relevancia en la estrategia de la compañía de seguros.
- Se menciona cómo la incidencia de delitos influye en la determinación del riesgo en las pólizas de seguro.

❖ Descarga y Creación de la Base de Datos

- Se describe el proceso de descarga de los datos desde la página oficial de datos abiertos.
- Se incluye una captura de pantalla o diagrama de la creación de la base de datos en SQL.

❖ Conexión a la Base de Datos desde Python

- Se explica cómo se realizó la conexión a la base de datos utilizando Python.
- Se muestran fragmentos de código relevantes y se describe su función.

❖ Limpieza y Transformación de los Datos

- Muestra el proceso de limpieza de los datos, incluyendo la gestión de valores nulos y transformación de formatos.
- Presenta las visualizaciones que realizaste en el primer tablero de Looker Studio.

Contenido

❖ Análisis de Series de Tiempo

- Objetivo: Identificar patrones y tendencias en los delitos de 2015 a 2021 y predecir los delitos en 2022.
- Selección del Municipio: Explica tu elección del municipio.
- Visualización de Datos: Incluye los gráficos del tablero de looker studio enfocados en esta parte.
- Predicciones: Explica el modelo de series de tiempo elegido, muestra las predicciones para 2022 e incluye gráficos comparativos.

❖ Clasificación de Estados por Peligrosidad (Clustering)

- Objetivo: Clasificar los estados según su peligrosidad en 2021.
- Preparación de Datos: Describe cómo se normalizaron los datos y cómo se preparó el dataframe final.
- Aplicación de Algoritmos: Explica el uso de K-means para clustering, cómo determinaste el número de clusters, y muestra gráficos de dispersión.
- Interpretación de Resultados: Presenta las características de cada cluster, identificando los estados más y menos peligrosos, añade la imagen del mapa con esta clasificación, así como también incluye los gráficos del tablero de looker studio enfocados en esta parte.

Contenido

❖ Enlace y Capturas de Pantalla del Tablero de Google Data Studio

- Incluye capturas de las vistas de tus tableros y explica porqué es relevante tener una herramienta de visualización como esta en la empresa.
- Proporciona el enlace a los tableros en línea.

❖ Conclusiones y Recomendaciones

- Resume los hallazgos más importantes.
- Discute pasos a seguir para futuras mejoras en el análisis.

Introducción

En este análisis, se explora una base de datos de delitos municipales proporcionada por el gobierno de México con el objetivo de comprender patrones delictivos y su impacto en la evaluación de riesgos para compañías de seguros.

La incidencia delictiva es un factor clave en la determinación del riesgo de las pólizas de seguro, ya que un mayor índice de delitos en una región puede influir en las primas y en la viabilidad de ciertos productos aseguradores. A través de un análisis estructurado, este proyecto busca proporcionar información valiosa que permita a las aseguradoras tomar decisiones fundamentadas sobre la asignación de riesgos y estrategias de cobertura.

El estudio abarca desde la carga y limpieza de los datos hasta su análisis mediante modelos de series de tiempo y técnicas de clustering. Finalmente, los resultados se presentan en un tablero interactivo de Google Data Studio, facilitando la visualización de hallazgos y la toma de decisiones basada en datos.



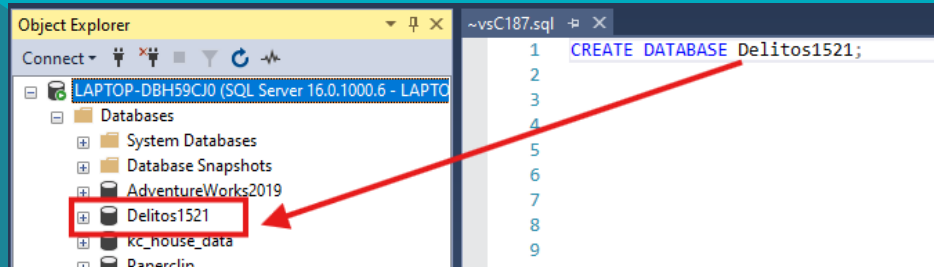
Descarga y Creación de la Base de Datos

Dentro del archivo Zip descargable en la página del gobierno de México se encuentra la información de los delitos cometidos en el país desde 2015 hasta 2021. Podemos encontrar un archivo de xml por cada año y un archivo csv donde se hace la recopilación de los 6 años.

En este caso se hará una lectura del archivo csv para hacer un análisis más completo de la información disponible.

Se crea en SQL una base de datos donde se insertará la tabla con la información del proyecto con:

- CREATE DATABASE Delitos1521;



Conexión a la Base de Datos desde Python

```
import pyodbc
from sqlalchemy import create_engine
from sqlalchemy import text

# Parámetros de conexión
server = 'LAPTOP-DBH59CJ0'
database = 'Delitos1521'
driver = 'ODBC Driver 17 for SQL Server'

# Se crea conexión con SQLAlchemy
engine = create_engine(f"mssql+pyodbc://{server}/{database}?trusted_connection=yes&driver={driver}&database={database}")
# Se guarda el DataFrame en SQL Server
df.to_sql('Delitos1521', con=engine, schema='dbo', if_exists='replace', index=False)
print("Datos insertados correctamente en SQL Server.")
```

[4] ✓ 4m 59.2s

.. Datos insertados correctamente en SQL Server.

Se verifican los parámetros de conexión y utilizando la librería SQLAlchemy se genera la conexión con SQL para después guardar la información descargada en la base de datos previamente creada.

Conexión a la Base de Datos desde Python

The image shows a Python IDE with three code blocks and their outputs, alongside SQL Server Enterprise Manager and a query window.

Code Block 1:

```
# Comprobar que la conexión se realizó con la base de datos correcta
with engine.connect() as conn:
    result = conn.execute(text("SELECT DB_NAME()"))
    print("Conectado a la base de datos:", result.fetchone()[0])
```

Output: Conectado a la base de datos: Delitos1521

Code Block 2:

```
# Comprobar información de la tabla y esquema donde se creo
with engine.connect() as conn:
    result = conn.execute(text("SELECT TABLE_SCHEMA, TABLE_NAME FROM INFORMATION_SCHEMA.TABLES WHERE TABLE_NAME = 'Delitos1521'"))
    tables = result.fetchall()
    print("Tablas encontradas:", tables)
```

Output: Tablas encontradas: [('dbo', 'Delitos1521')]

Code Block 3:

```
# Comprobar información contenida en la tabla
with engine.connect() as conn:
    result = conn.execute(text("SELECT COUNT(*) FROM dbo.Delitos1521"))
    print("Número de filas en la tabla:", result.fetchone()[0])
```

Output: Número de filas en la tabla: 1589462

SQL Server Object Explorer: Shows the database structure for 'LAPTOP-DBH59CJ0 (SQL Server 16.0.1000.6 - LAF)'. The 'Delitos1521' database is selected, showing its tables, including 'dbo.Delitos1521'.

SQL Query Window: Shows the SQL code used to create the database and table, and the result of a query.

```
1 CREATE DATABASE Delitos1521;
2
3 SELECT COUNT (*) FROM Delitos1521;
```

Results: The query result shows a single row with the value 1589462.

Se comprueba que la base tabla se haya creado correctamente y contenga toda la información

Limpieza y Transformación de los Datos

```
# Se verifican los tipos de datos que contiene la tabla, no se presentan nulos
df.info()

[0] ✓ 1.0s

... <class 'pandas.core.frame.DataFrame'>
RangeIndex: 1589462 entries, 0 to 1589461
Data columns (total 21 columns):
#   Column                Non-Null Count  Dtype
---  -
0   Año                    1589462 non-null  int64
1   Clave_Ent               1589462 non-null  int64
2   Entidad                 1589462 non-null  object
3   Cve. Municipio         1589462 non-null  int64
4   Municipio              1589462 non-null  object
5   Bien jurídico afectado  1589462 non-null  object
6   Tipo de delito         1589462 non-null  object
7   Subtipo de delito      1589462 non-null  object
8   Modalidad              1589462 non-null  object
9   Enero                  1589462 non-null  int64
10  Febrero                1589462 non-null  int64
11  Marzo                  1589462 non-null  int64
12  Abril                  1589462 non-null  int64
13  Mayo                   1589462 non-null  int64
14  Junio                  1589462 non-null  int64
15  Julio                  1589462 non-null  int64
16  Agosto                 1589462 non-null  int64
17  Septiembre             1589462 non-null  int64
18  Octubre                1589462 non-null  int64
19  Noviembre              1589462 non-null  int64
20  Diciembre              1589462 non-null  int64
dtypes: int64(15), object(6)
memory usage: 254.7+ MB
```

```
# Datos únicos por columna
df.nunique()

✓ 1.7s

Año                    7
Clave_Ent              32
Entidad                32
Cve. Municipio        2478
Municipio              2331
Bien jurídico afectado 7
Tipo de delito         40
Subtipo de delito      55
Modalidad              59
Enero                  409
Febrero                418
Marzo                  446
Abril                  437
Mayo                   458
Junio                  446
Julio                  446
Agosto                456
Septiembre             445
Octubre                459
Noviembre              432
Diciembre              424
dtype: int64
```

Se verifica que la tabla no contenga datos nulos o columnas tipo ID.

Limpieza y Transformación de los Datos

Debido a que el proyecto se realiza para una empresa de seguros de automóviles, los datos de la tabla se filtran para que esta contenga únicamente información relacionada con este rubro.

Se considera también información de motocicletas y otros vehículos con motor.

Encontrar y filtrar delitos que estén relacionados con automoviles

```
#Bien jurídico afectado (No se identifican categorías directamente relacionadas con automóviles)
print('Bien jurídico afectado')
print(df['Bien jurídico afectado'].unique().__len__())
print(df['Bien jurídico afectado'].sort_values().unique())
```

```
#Tipo de delito (No se identifican categorías directamente relacionadas con automóviles)
print('Tipo de delito')
print(df['Tipo de delito'].unique().__len__())
print(df['Tipo de delito'].sort_values().unique())
```

```
#Subtipo de delito (Se identifican delitos relacionados a automóviles)
print('Subtipo de delito')
print(df['Subtipo de delito'].unique().__len__())
print(df['Subtipo de delito'].sort_values().unique())
# 'Robo a transportista', 'Robo de autopartes', 'Robo de maquinaria', 'Robo de vehículo automotor',
# 'Robo en transporte individual', 'Robo en transporte público colectivo', 'Robo en transporte público individual'
# Resumen de palabras: transportista, autopartes, maquinaria, vehículo, transporte
```

```
#Modalidad (Se identifican delitos relacionados a automóviles)
print('Modalidad')
print(df['Modalidad'].unique().__len__())
print(df['Modalidad'].sort_values().unique().tolist())

# 'En accidente de tránsito', 'Robo de coche de 4 ruedas Con violencia', 'Robo de coche de 4 ruedas Sin violencia',
# 'Robo de motocicleta Con violencia', 'Robo de motocicleta Sin violencia', 'Robo de tractores Con violencia', 'Robo de tractores Sin violencia'
# Resumen de palabras: tránsito, coche, motocicleta, tractores
```

Limpieza y Transformación de los Datos

Se crea un nuevo df que únicamente contenga delitos relacionados con automóviles

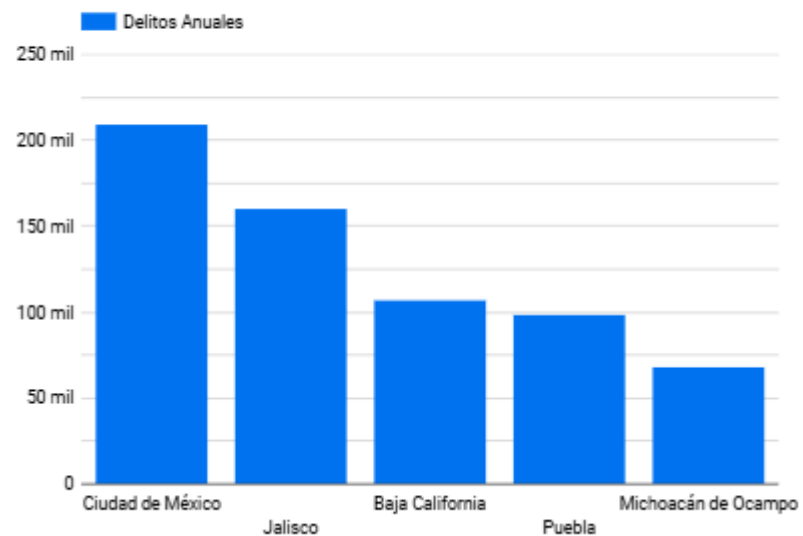
```
#Crear copia de base de datos original para crear una nueva con información solo relacionada a automóviles
delitos_autos = df.copy()
#Crear columna adicional con toda la información del delito
delitos_autos["Delito"] = delitos_autos[['Tipo de delito', 'Subtipo de delito', 'Modalidad']].agg('-'.join, axis=1)
delitos_autos.sample(5)

#Filtrar para que solo contenga delitos relacionados con autos con las palabras identificadas
#palabras relacionadas: transportista, autopartes, maquinaria, vehículo, transporte, tránsito, coche, motocicleta, tractores
delitos_autos = delitos_autos[delitos_autos['Delito'].str.contains('transportista|autopartes|vehículo|transporte|tránsito|coche|motocicleta|tractores|transito')]
delitos_autos
```

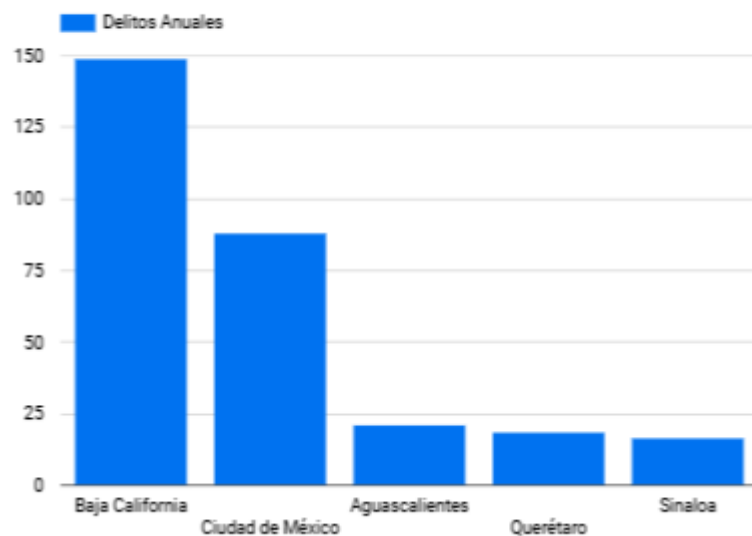
Se crea un nuevo Data Frame que sólo contenga información útil para la compañía, de esta forma reducimos la cantidad de datos a procesar de 1,589,462 a 324,380 .

Limpieza y Transformación de los Datos

Delitos Cometidos por Entidad (2015-2021)



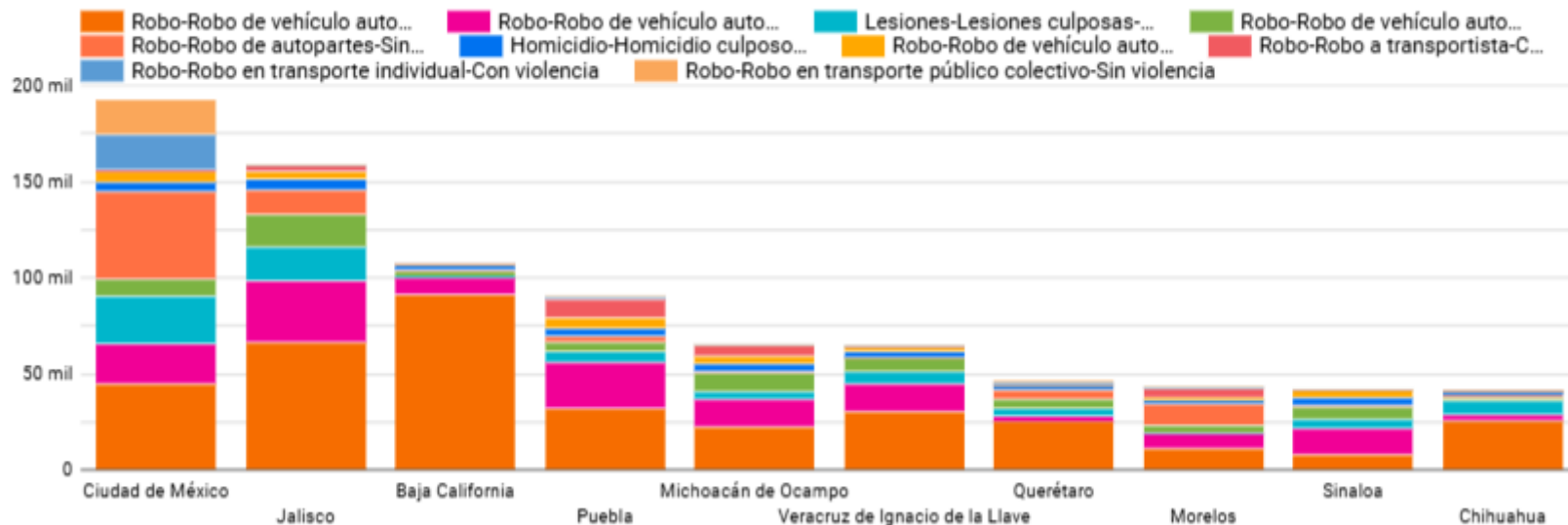
Promedio de Delitos Cometidos por Entidad (2015-2021)



Primeras visualizaciones de Looker Studio después de la limpieza de datos.

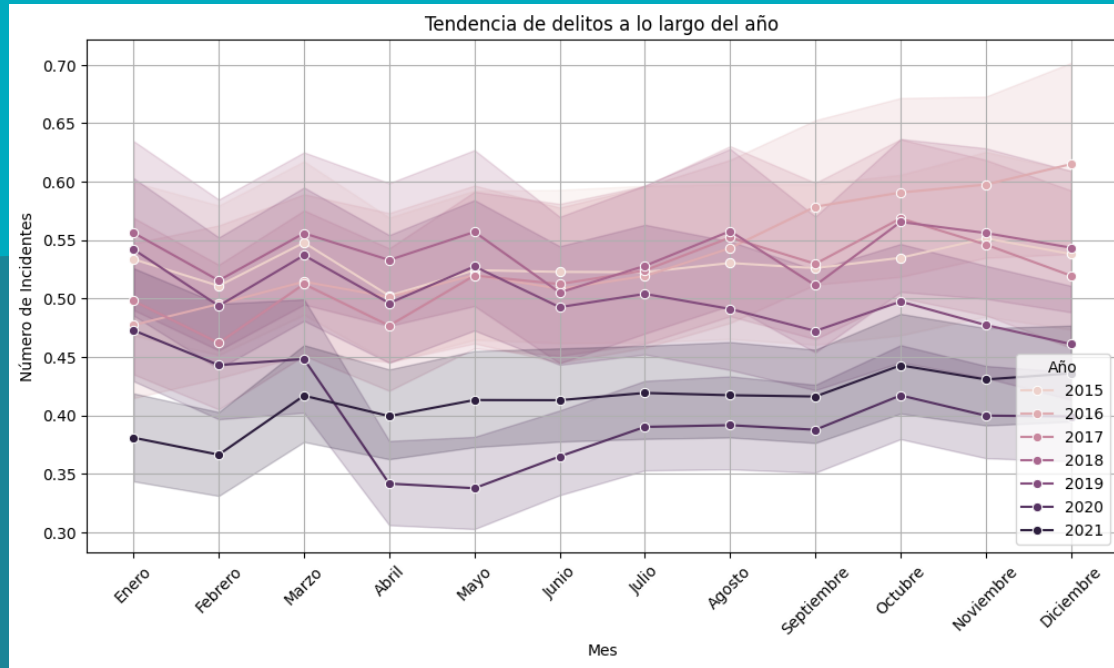
Limpieza y Transformación de los Datos

Delitos más cometidos por estado (2015-2021)



Primeras visualizaciones de Looker Studio después de la limpieza de datos.

Análisis de Series de Tiempo

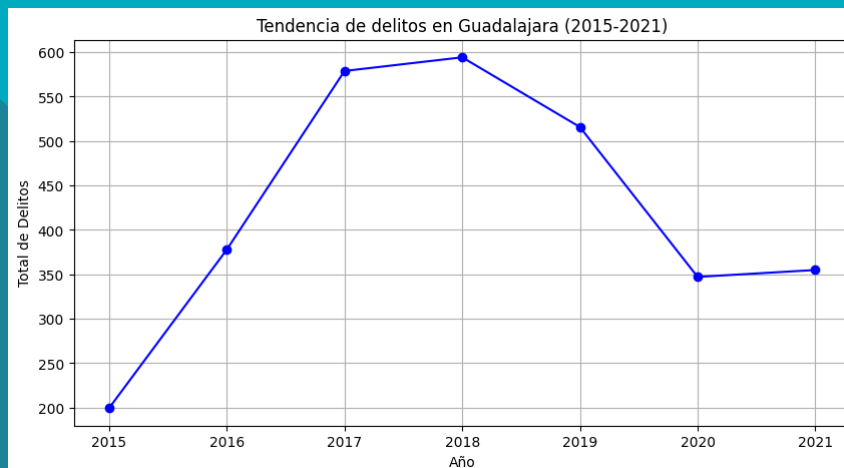


Se identifica cierto aumento en la incidencia de delitos conforme avanza el año .

Análisis de Series de Tiempo

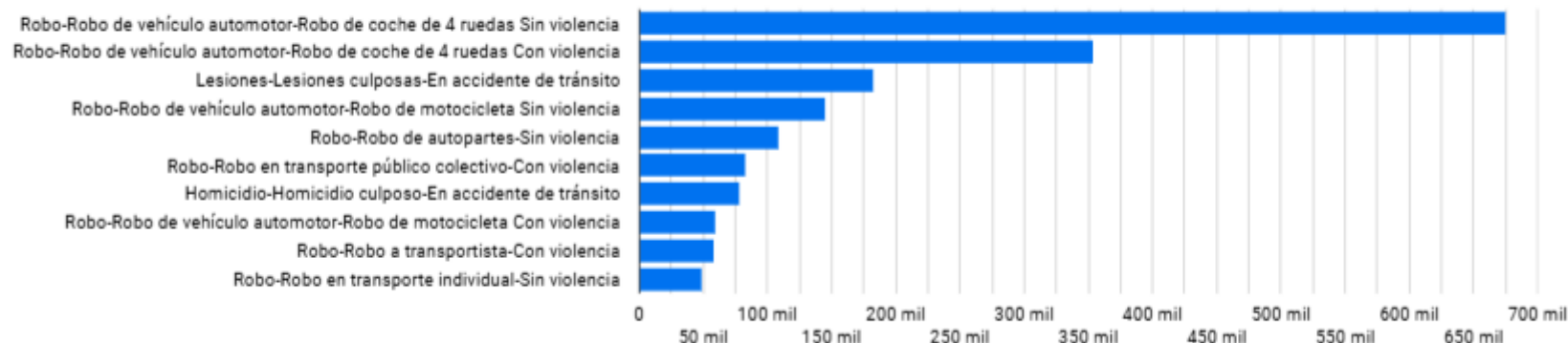
	Enero	Febrero	Marzo	Abril	Mayo	Junio	Julio	Agosto	Septiembre	Octubre	Noviembre	Diciembre	Total_Delitos
Municipio													
Ecatepec de Morelos	52.171429	50.971429	55.578571	51.235714	57.492857	53.514286	53.978571	55.792857	53.521429	57.164286	57.371429	53.721429	652.514286
Guadalajara	33.935714	33.585714	33.350000	29.592857	32.685714	36.814286	38.950000	40.507143	35.721429	38.685714	36.350000	33.864286	424.042857
Tijuana	37.592857	33.371429	35.271429	32.885714	33.521429	31.028571	32.064286	33.135714	33.114286	33.721429	34.128571	38.150000	407.985714
Tlalneantla de Baz	25.292857	24.292857	27.200000	25.228571	27.114286	25.407143	26.450000	26.857143	26.957143	27.957143	27.550000	24.878571	315.185714
Toluca	24.335714	23.214286	24.407143	23.914286	23.928571	21.992857	23.928571	25.335714	25.171429	25.907143	25.471429	25.292857	292.900000

Debido a que es uno de los municipios con mayor incidencia de delitos relacionados con automóviles, se selecciona a Guadalajara para hacer los siguientes análisis



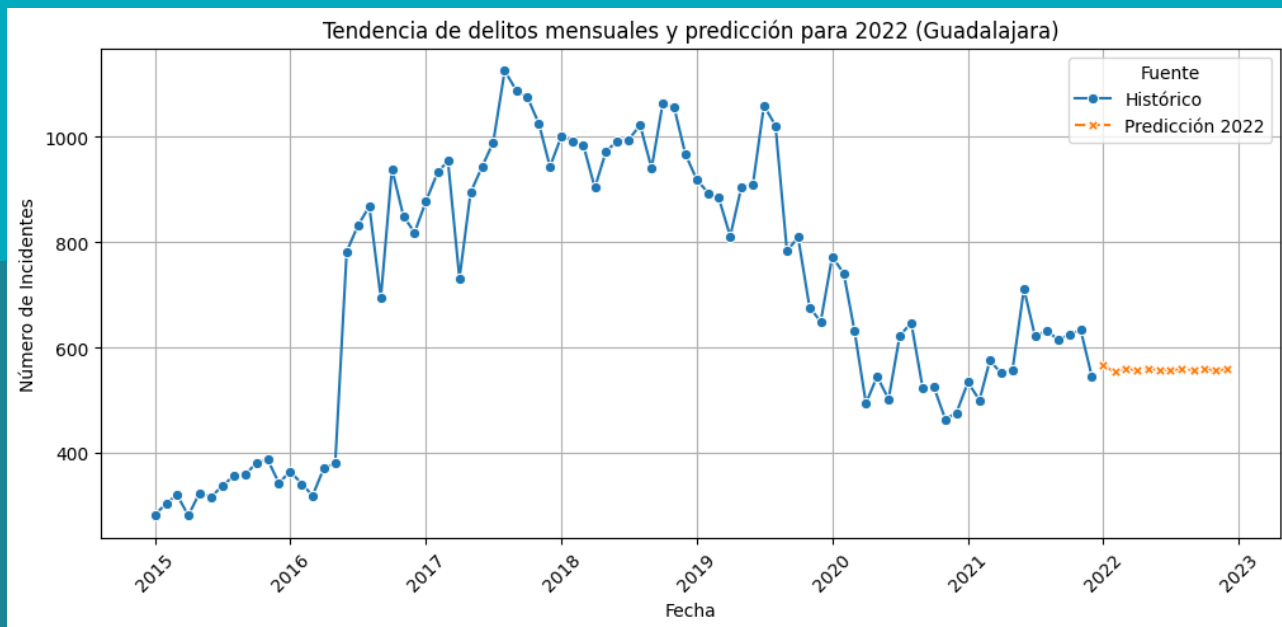
Análisis de Series de Tiempo

Delitos más cometidos en Guadalajara (2015-2021)



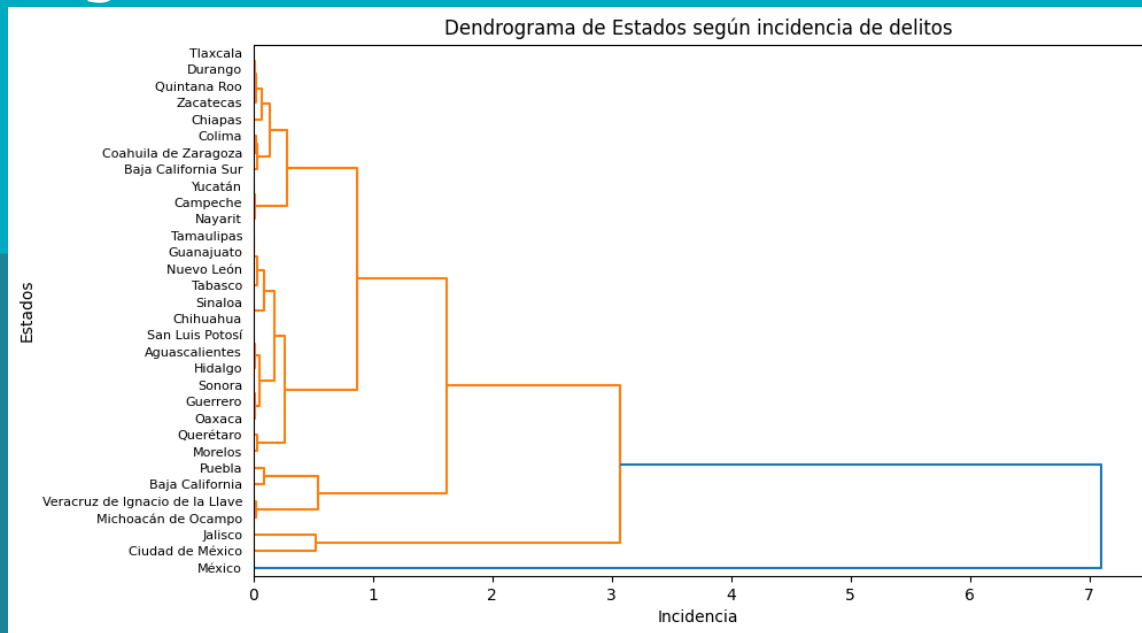
Gráficos del tablero de looker studio enfocados en el municipio seleccionado.

Análisis de Series de Tiempo



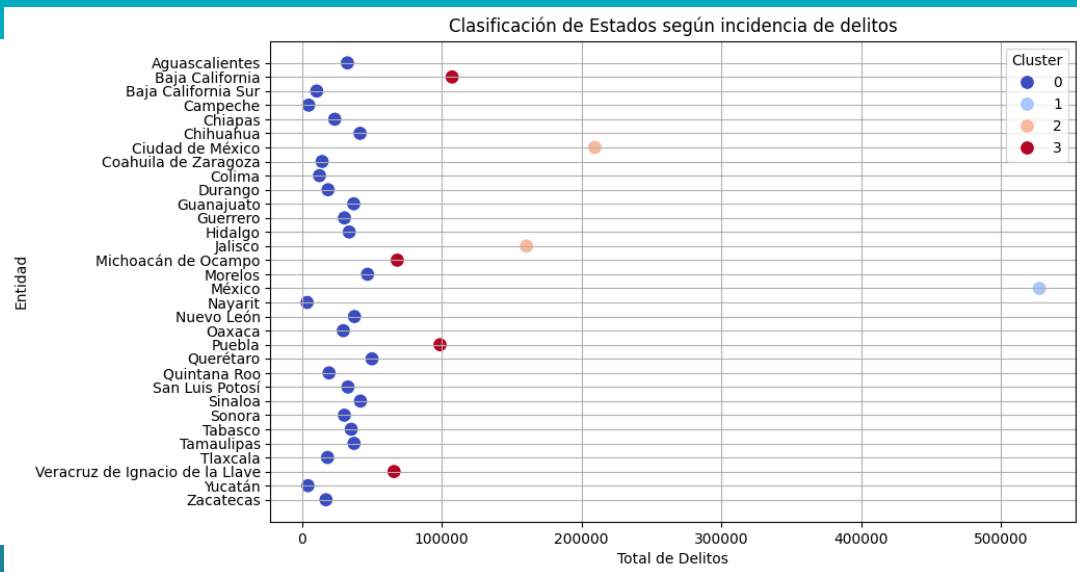
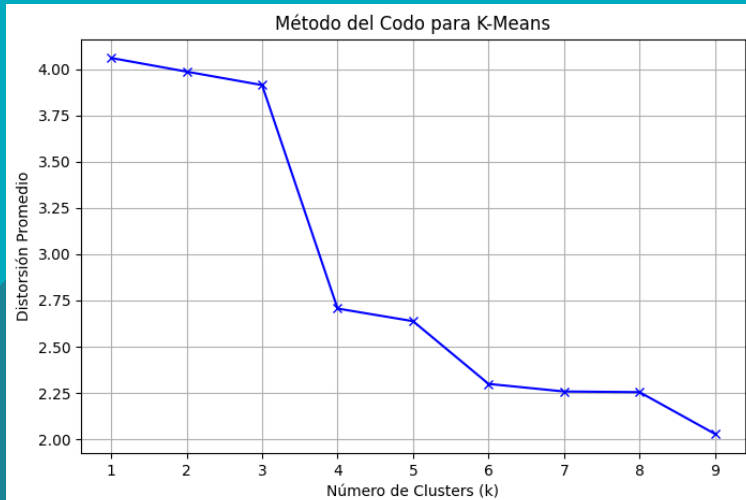
Para realizar las predicciones se utiliza el modelo ARIMA (Autoregressive Integrated Moving Average).

Clasificación de Estados por Peligrosidad (Clustering)



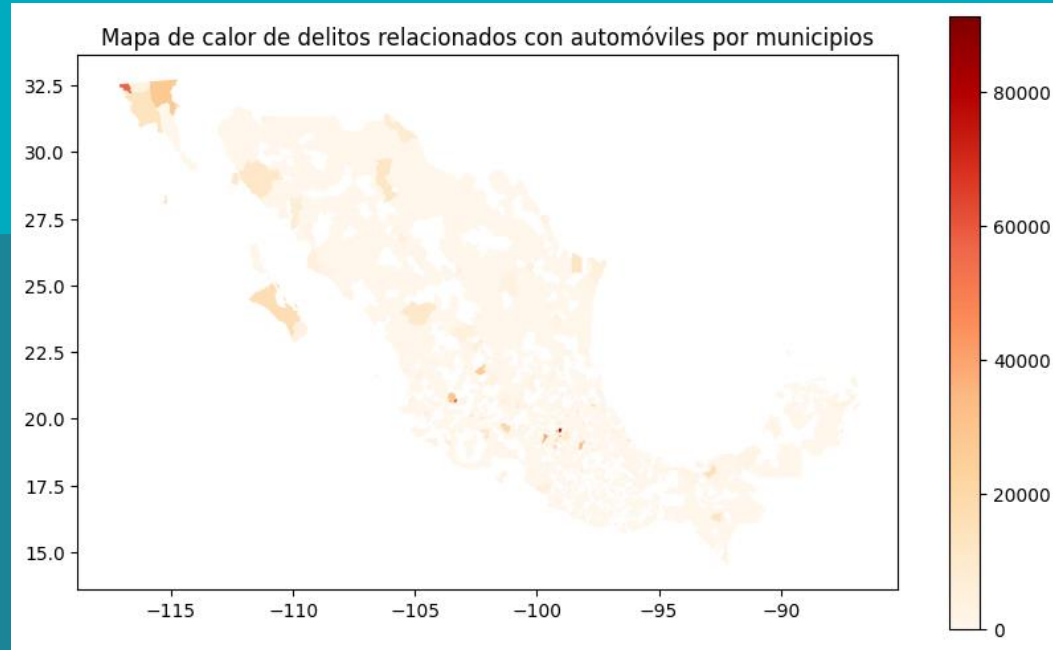
Una vez agrupados los datos por entidad estos se normalizaron usando el método StandardScaler para después crear el Dendrograma que agrupa los estados por peligrosidad.

Clasificación de Estados por Peligrosidad (Clustering)



También podemos agrupar los estados en 4 diferentes grupos, esto como resultado del método del codo que nos indica que el número ideal de clusters es 4.

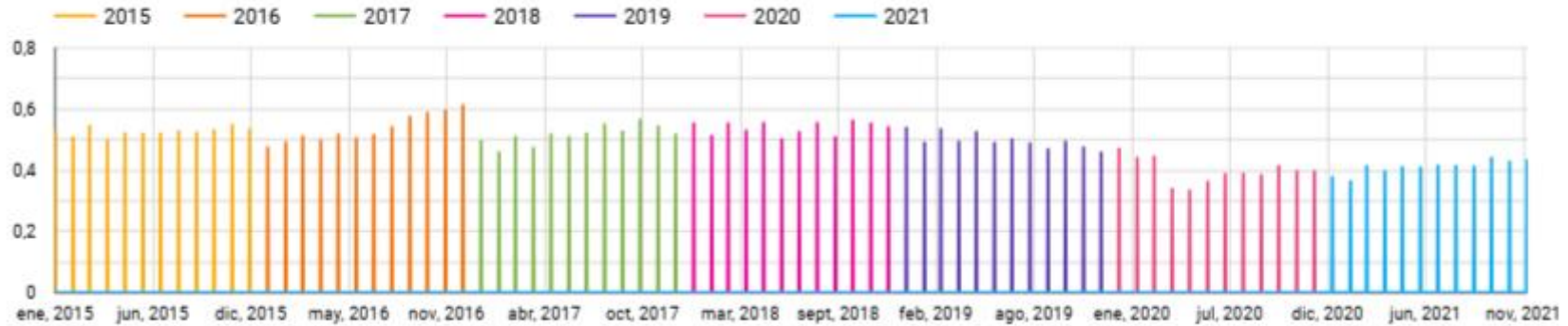
Clasificación de Estados por Peligrosidad (Clustering)



En el mapa se pueden apreciar los municipios con mayor incidencia de delitos en colores más intensos. Entre ellos Ecatepec, Guadalajara, Tijuana, Toluca y Tlalnepantla.

Enlace y Capturas de Pantalla del Tablero de Google Data Studio

Incidencia de delitos de 2015 a 2021

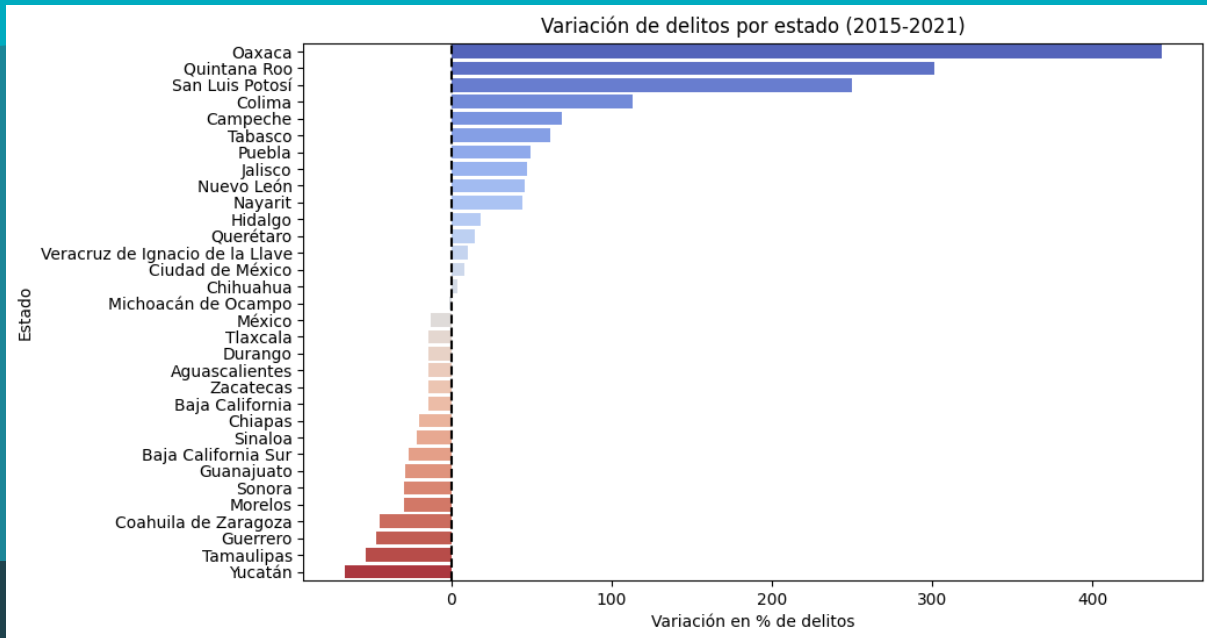


Además de los gráficos ya mostrados es interesante analizar la incidencia de delitos a lo largo del tiempo. Para una empresa de seguros es sumamente valioso hacer uso de herramientas de visualización ya que facilitan la transformación de grandes volúmenes de datos en insights accionables, facilitando la toma de decisiones estratégicas.

<https://lookerstudio.google.com/s/v7rpv2HEqY>

Conclusiones y Recomendaciones

- A pesar de que en algunos estados se observa cierta disminución en los delitos relacionados con autos del año 2015 al año 2021, hay muchos que, al contrario, han ido aumentando y **estos son para los que debería considerarse un aumento en el precio de los seguros.**



Conclusiones y Recomendaciones

- ❑ Se observan ciertos **picos de incidencia de delitos en meses como noviembre, diciembre, marzo**. Valdría la pena hacer un análisis más a profundidad para entender la raíz de esto, saber si esta relacionado con las festividades o si hay algún otro evento que lo propicie.
- ❑ Estados como **Estado de México, Jalisco, Michoacán, Veracruz y Baja California** son los que presentan **mayor incidencia de delitos**, entre ellos Jalisco es el que presenta además una alta tasa de aumento.
- ❑ La categoría de delitos con mayor incidencia es la de **Robo de vehículo** ya sea con o sin violencia, se podría considerar el **aumento de las primas** de dicho rubro.
- ❑ Considerando el aumento de incidencias y las predicciones realizadas se podría hacer un plan para prever a los clientes del aumento de los costos o buscar formas de facilitar la compra de los seguros a pesar del aumento de los precios. A nivel gobierno se podría hacer un plan para reforzar las políticas de seguridad y evitar los aumentos observados.

Fuentes de información

1. <https://datos.gob.mx/busca/dataset/incidencia-delictiva-del-fuero-comun-a-nivel-municipal>