



Instituto Politécnico Nacional

Escuela Superior de Cómputo



Práctica 3

**Definición del proyecto de datos semestral:
Carga y exploración de datos**

Grupo: 3CV18

Integrantes.

Cazares Martínez Maximiliano

Ramos Nieves Adrián

Profesor:

Roberto Eswart Zagal Flores

Introducción

En la presente práctica determinaremos qué dataset se tomará como referencia para trabajar a lo largo de la materia como nuestro objeto de estudio. Se considerarán como referencia varios lineamientos propuestos en las características del dataset para poder obtener la información más viable para la realización del proyecto, así como para el manejo de filtros y detección de errores en la base de datos.

Como investigación de interés decidimos tomar el apartado de Justicia y Seguridad como filtro en el dataset de la CDMX. Finalmente se escogió el tema de Víctimas en carpetas de investigación FGJ.

Esta base de datos contiene la información actualizada de las víctimas de los delitos en las carpetas de investigación de la Fiscalía General de Justicia (FGJ) de la Ciudad de México a partir de enero de 2019. Los datos de esta base fueron actualizados por la FGJ el 29 de julio de 2020.

Para una correcta interpretación de la información, se realizan las siguientes aclaraciones.

- El campo "FechaHecho" representa la fecha en la que se cometió.
- El campo "FechaInicio" corresponde a la fecha de la apertura de la carpeta de investigación.
- En esta base se señala el sexo de la víctima, así como la fecha en que ocurrieron los hechos denunciados.
- Es importante destacar que, aunque en algunas ocasiones la víctima es la denunciante, en otras, denunciante y víctima son diferentes (por ejemplo, casos en los que menores de edad son víctimas).
- Es posible que en una misma denuncia se incluya a una o más víctimas.

Carpeta de Investigación PGJ CDMX

Dimensiones

- Delito
- Categoria_delito
- Fiscalía
- Agencia
- Unidad_investigación
- Colonia_hechos
- Alcaldía
- Alcaldia_hechos
- Calle_hechos
- Calle_hechos2
- Longitud
- Latitud

Granularidad Temporal

- Año_hechos
- Mes_hechos
- Fecha_hechos
- Fecha_inicio
- Año_inicio
- Mes_inicio

Caso de estudio adicional

Conocer la relación entre delitos y la cantidad de población en cada delegación de la Ciudad de México, ya que con esto podemos saber cómo responder a la tendencia en cada delegación.

Motivos de la elección del dataset

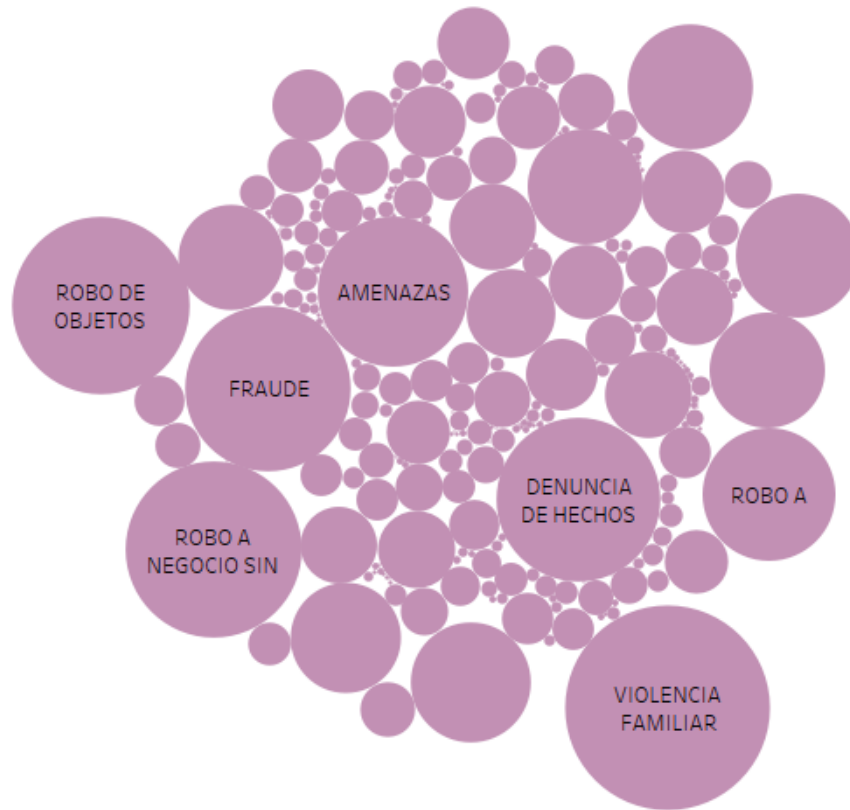
El dataset fue elegido con base en los lineamientos establecidos en la práctica y la importancia de los datos que dicho dataset maneja ya que con esto podríamos saber la tendencia de delitos en cada delegación .

Alcance del proyecto

Conocer las zonas de la Ciudad de México más peligrosas y con delitos más graves para llevar a cabo un plan de acción en cada Alcaldía para disminuir los índices de inseguridad y mejorar la calidad de vida de las personas en dichas zonas.

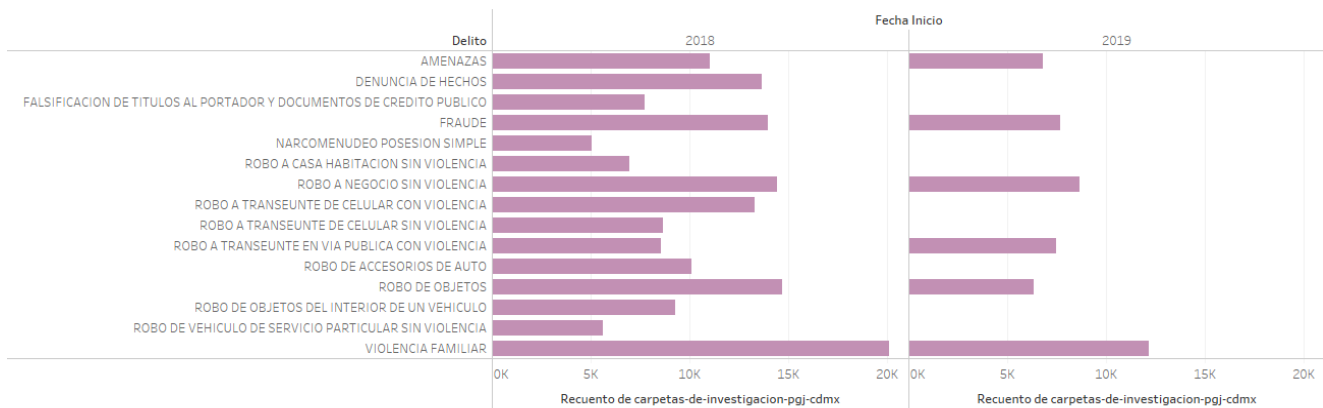
Análisis exploratorio básico con Tableau

- Distribución de la dimensión categórica en el dataset son las carpetas de investigación de delitos cometidos en la Ciudad de México.



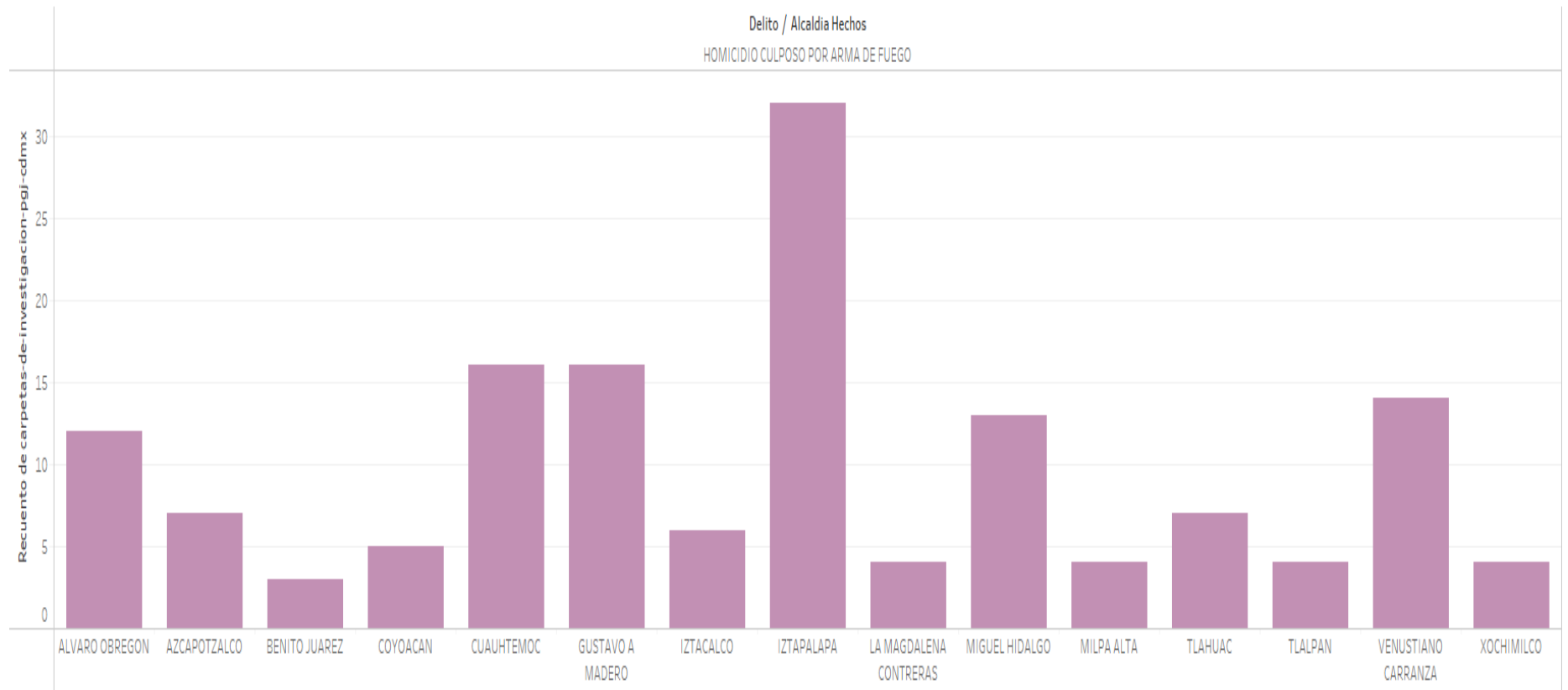
- Distribución del fenómeno en el tiempo.

Distribucion_tiempo



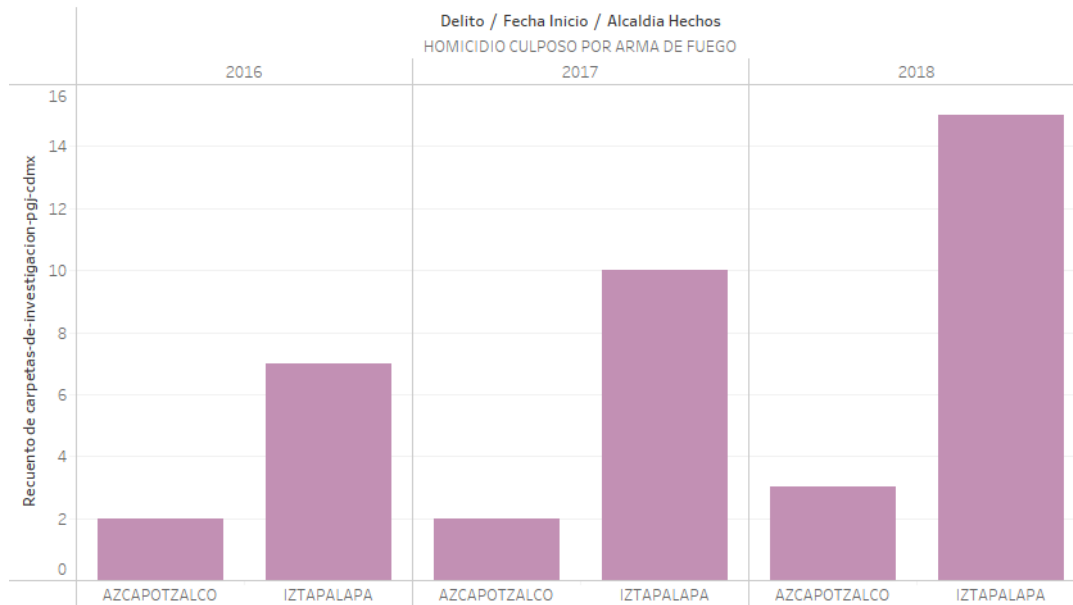
- Distribución en el espacio

Distribucion_espacio



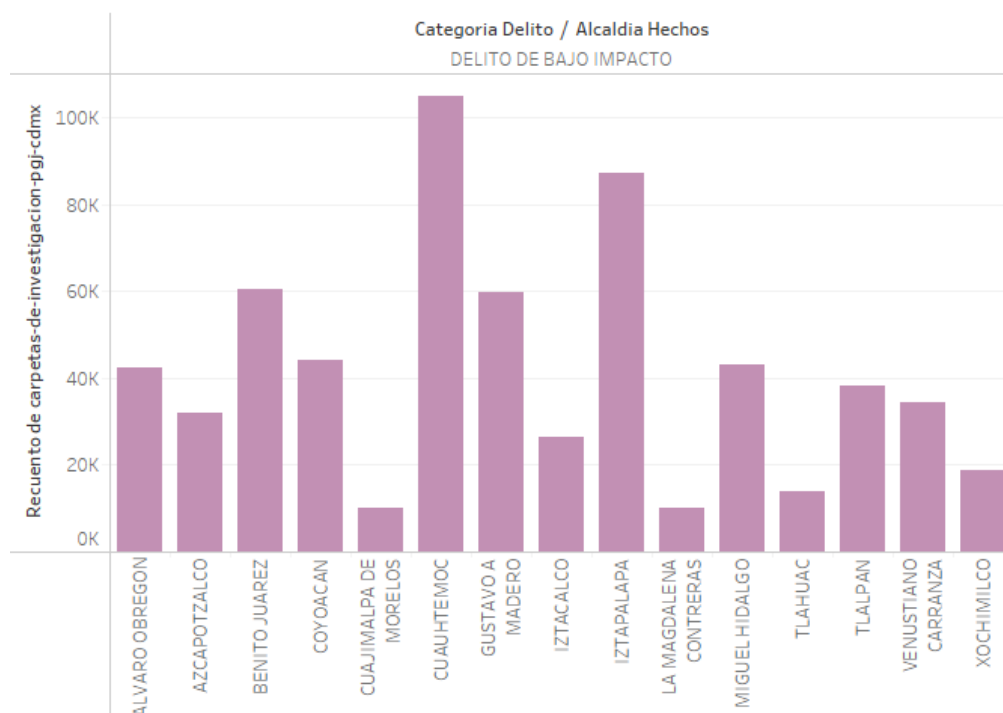
- Distribución del fenómeno en el espacio y tiempo

Distribucion_espacio/tiempo



- Otras dimensiones temáticas

Otras_dimensiones



Valores atípicos e inconsistentes

Se encontró que en el dataset existen varias columnas con elementos nulos, además de tener valores inconsistentes en las alcaldías, ya que solo existen en CDMX 16 de ellas y se tienen registradas 486, tal vez porque se consideraron otras partes de los estados, pero como objeto de estudio solo se tomarán las 16 existentes marcadas como oficiales.

Al analizar el dataset, se encontró que era conveniente utilizar algunas variables de tiempo para acortar los datos, así como filtros para descartar las alcaldías de las cuales no eran de interés para nuestro caso de estudio, ya que consideramos que sería más fácil procesar la información de esta manera.

Conclusión

Con el análisis de la base de datos anterior encontramos que es factible para realizar el proyecto contemplado, ya que, aunque tiene varias inconsistencias en columnas de las dimensiones, también cumplen con los requisitos planteados en las especificaciones. Además, consideramos que es un tema de estudio sumamente importante y bastante interesante el encontrar algunas tendencias de los delitos más comunes en diferentes alcaldías de la CDMX. Se tendría que llevar a cabo una limpieza de datos exhaustiva para que no se muestren inconsistencias al momento de utilizar Tableau