



Instituto Politécnico Nacional

Escuela Superior de Cómputo



Práctica 6

**Diseño estructural del datawarehouse, ETL, Lattice y
Visualización de Datos del proyecto semestral.**

Grupo:3CV18

Integrantes:

Cazares Martínez Maximiliano

Ramos Nieves Adrián

Profesor:

Roberto Eswart Zagal Flores

Análisis de las carpetas de investigación de la Fiscalía General de Justicia.

Objetivo

Dados los alcances de la unidad de aprendizaje, se busca aplicar la mayor cantidad de conocimientos adquiridos a lo largo del curso, centrado en datos abiertos de los registros de víctimas de delitos en las carpetas de investigación de la Fiscalía General de Justicia, donde al final del ejercicio se busca conocer las zonas con mayor delitos en la CDMX.

Entendimiento del negocio

Se busca conocer las delegaciones con mayor índice de víctimas por delitos en la CDMX, cuales son los delitos más comunes por alcaldía y colonia en la CDMX, así como el mes y el año donde se han registrado la tasa más grande de delito, todo esto para determinar el índice de peligro en la CDMX.

Para este análisis se tomarán los datos únicamente de la base de datos de las víctimas de los delitos en las carpetas de investigación de la Fiscalía General de Justicia (FGJ) de la Ciudad de México.

Entendimiento de los datos

La fuente de los datos se obtiene de [aquí](#) donde las dimensiones temáticas que incluye son: id, Delito, Categoria_delito, Fiscalía, Agencia, Unidad_investigación, Colonia_hechos, Alcaldía, Alcaldia_hechos, Calle_hechos, Calle_hechos2, Longitud, Latitud. Además, tenemos una granularidad temporal amplia que incluye: Año_hechos, Mes_hechos, Fecha_hechos, Fecha_inicio, Año_inicio, Mes_inicio.

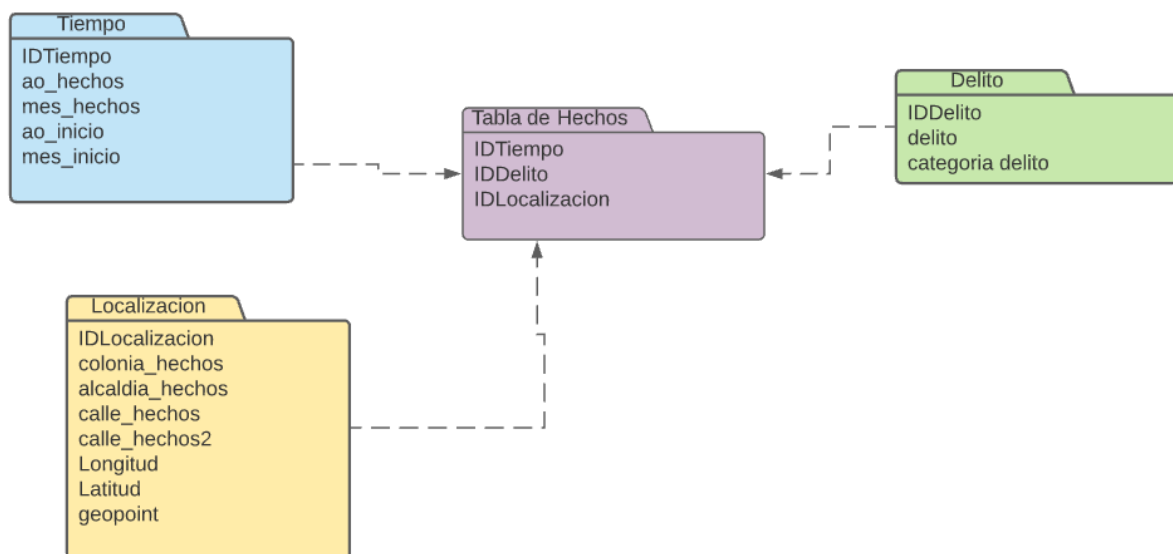
Campo	Tipo	Descripción
id	int	Identificador del delito
delito	int	Tipo de delito cometido
categoria_delito	string	Categoría de delito según carpetas de la PGJ
fiscalía	string	Nombre de la fiscalía donde se levantó el reporte
agencia	string	Agencia del lugar donde se levantó el reporte
unidad_investigacion	string	Número y nombre de la unidad de Investigación
colonia_hechos	string	Nombre de la colonia del lugar del delito
alcaldia_hehco	string	Nombre de la alcaldía del lugar del delito
calle_hechos	string	Nombre de la calle del lugar del delito
calle_hechos2	string	Nombre de la segunda calle del lugar del delito
longitud	float	Número de Longitud del lugar del reporte
latitud	float	Número de Latitud del lugar del reporte
geopoint	string	Coordenadas del lugar del reporte
ao_hechos	int	año del delito
mes_hechos	string	mes del delito
fecha_hechos	string	fecha completa del delito
ao_inicio	int	año del reporte
mes_inicio	string	mes del reporte

fecha_inicio	string	fecha completa del delito
--------------	--------	---------------------------

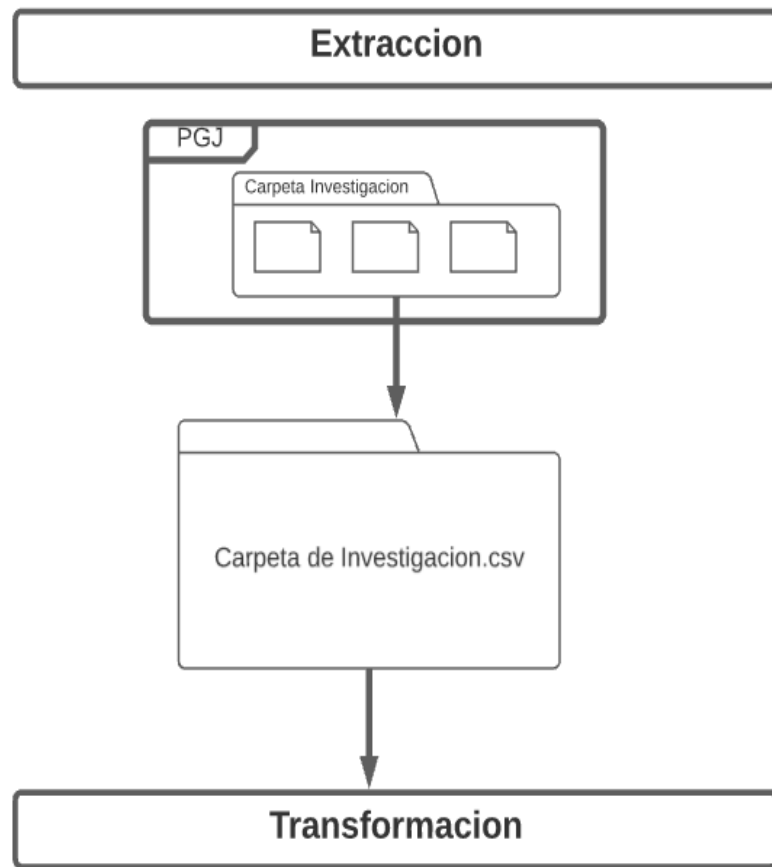
Estructura de tabla de hechos y el modelado conceptual del data warehouse.

El siguiente diagrama muestra un modelo teórico sencillo basado en el modelo multidimensional en estrella. Para este modelo descartamos únicamente dos columnas de la base de datos original, puesto que estas dos columnas son la fecha en donde se realizaron los delitos, ya que tenemos tanto el mes y el año donde esto sucedió, generando redundancia.

Dividimos, los datos en 4 tablas, Tiempo, Delito, Localización y Tabla de hechos, donde esta contiene los ID's de las otras 3, creando una relación entre sí y generando el modelo multidimensional en estrella.



Fases del ETL



- Transformamos el documento csv en un archivos.xlsx.
- Cambiamos el formato de las fechas para mejorar la su apreciación a la hora de realizar cualquier consulta de datos.
- Mediante consultas SQL creamos las tablas antes mencionadas para el modelo en estrella.

Carga.

- Las tablas anteriores se cargaron a SQL server para su posterior minería.

Implementación del Proceso de ETL

```

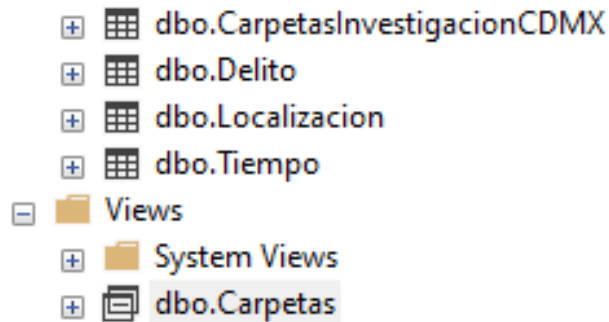
1 import os
2 import openpyxl
3 from conexionbd import *
4
5 connection = connectionBD()
6 cursor = connection.cursor()
7
8 file = f"{os.path.dirname(__file__)}/{'CarpetasInvestigacionCDMX.xlsx'}"
9 book = openpyxl.load_workbook(file)
10 sheet = book.active
11 rows = sheet.max_row
12 data_raw = sheet['A2': f'S{rows}'] # 808872
13
14 query = """Insert into CarpetasInvestigacionCDMX(id, ao_hechos,
15     mes_hechos, fecha_hechos, delito, categoria_delito, fiscalia,
16     agencia, unidad_investigacion, colonia_hechos, alcaldia_hechos,
17     fecha_inicio, mes_inicio, ao_inicio, calle_hechos, calle_hechos2,
18     longitud, latitud, geopoint)
19     values (?, ?, ?, ?, ?, ?, ?, ?, ?, ?, ?, ?, ?, ?, ?, ?, ?)"""
20
21 data = []
22 for r in data_raw:
23     a = [c.value for c in r]
24     data.append(a)
25
26 for i in range(len(data)):
27     ID = data[i][0]
28     AnioHechos = data[i][1]
29     MesHechos = data[i][2]
30     FechaHechos = data[i][3]
31     Delito = data[i][4]
32     CateDelito = data[i][5]
33     Fiscalia = data[i][6]
34     Agencia = data[i][7]
35     UnidadInvest = data[i][8]
36     ColHechos = data[i][9]
37     AlcaHechos = data[i][10]
38     FechaInicio = data[i][11]
39     MesInicio = data[i][12]
40     AnioInicio = data[i][13]
41     CalleHechos = data[i][14]
42     CalleHechos2 = data[i][15]
43     Longitud = data[i][16]
44     Latitud = data[i][17]
45     Geopoint = data[i][18]
46
47     FH = FechaHechos.strftime('%d/%m/%Y')
48     FI = FechaInicio.strftime('%d/%m/%Y')
49
50     values = (ID, AnioHechos, MesHechos, FH, Delito, CateDelito, Fiscalia,
51     Agencia, UnidadInvest, ColHechos, AlcaHechos, FI, MesInicio, AnioInicio,
52     CalleHechos, CalleHechos2, Longitud, Latitud, Geopoint)
53     cursor.execute(query, values)
54
55 cursor.commit()
56 cursor.close()
57 connection.close()

```

Tecnologías usadas en el proceso de ETL.

- [openpyxl](#): Biblioteca de python que permite la manipulación de archivos .xlsx.
- [pyodbc](#): Biblioteca de python que permite la conexión a SQL Server.

Tablas resultantes de la transformación



A continuación se muestran las consultas para crear el modelo en estrella.

```
1  -- Modelo de estrella
2
3  SELECT id as IDTiempo, ao_hechos, mes_hechos, ao_inicio, mes_inicio into Tiempo
4  FROM CarpetasInvestigacionCDMX;
5  SELECT * FROM Tiempo order by IDTiempo asc;
6
7  SELECT id as IDDelito, delito, categoria_delito into Delito
8  from CarpetasInvestigacionCDMX;
9  SELECT * FROM Delito order by IDDelito asc;
10
11 SELECT id as IDLocalizacion, colonia_hechos, alcaldia_hechos,
12 calle_hechos, calle_hechos2, longitud, latitud, geopoint
13 into Localizacion FROM CarpetasInvestigacionCDMX;
14 SELECT * FROM Localizacion order by IDLocalizacion asc;
15
16 create view Carpetas as select Tiempo.IDTiempo, Delito.IDDelito, Localizacion.IDLocalizacion
17 from Delito, Tiempo, Localizacion where IDTiempo = IDLocalizacion and IDDelito = IDTiempo;
```

Creación de tablas mediante código.

Las siguiente imágenes muestran los datos ingresados en las tablas del modelo en estrella.

<pre>-- Tiempo SELECT TOP (1000) [IDTiempo] ,[ao_hechos] ,[mes_hechos] ,[ao_inicio] ,[mes_inicio] FROM [CarpetasInvestigacion].[dbo].[Tiempo]</pre>					
100 %					
Results Messages					
	IDTiempo	ao_hechos	mes_hechos	ao_inicio	mes_inicio
1	0	2016	Enero	2016	Enero
2	1	2017	Noviembre	2017	Noviembre
3	2	2017	Noviembre	2017	Noviembre
4	3	2015	Diciembre	2016	Enero
5	4	2018	Mayo	2018	Mayo
6	5	2018	Mayo	2018	Mayo
7	6	2018	Mayo	2018	Mayo
8	7	2016	Enero	2016	Enero
9	8	2017	Noviembre	2017	Noviembre
10	9	2016	Septiembre	2018	Mayo
11	10	2017	Noviembre	2017	Noviembre
12	11	2016	Enero	2016	Enero
13	12	2017	Noviembre	2017	Noviembre
14	13	2018	Mayo	2018	Mayo
15	14	2017	Octubre	2017	Noviembre
16	15	2017	Octubre	2017	Noviembre
17	16	2017	Octubre	2017	Noviembre
18	17	2017	Octubre	2017	Noviembre

Datos Insertados en Tabla Tiempo

<pre>-- Localizacion SELECT TOP (1000) [IDLocalizacion] ,[colonia_hechos] ,[alcaldia_hechos] ,[calle_hechos] ,[calle_hechos2] ,[longitud] ,[latitud] ,[geopoint] FROM [CarpetasInvestigacion].[dbo].[Localizacion]</pre>				
100 %				
Results Messages				
	IDLocalizacion	colonia_hechos	alcaldia_hechos	calle_hechos
1	0	GRANJAS MODERNAS	GUSTAVO A MADERO	SAN JUAN DE ARAGON
2	1	NULL	TLALNEPANTLA DE BAZ	BENITO JUAREZ
3	2	GABRIEL RAMOS MILLÁN SECCIÓN TLACOTAL	IZTACALCO	SUR 117
4	3	CONSEJO AGRARISTA MEXICANO	IZTAPALAPA	CUARTA CERRADA DE FRESNO
5	4	DESARROLLO URBANO QUETZALCOATL	IZTAPALAPA	TARASCO
6	5	INFONAVIT IZTACALCO	IZTACALCO	FRANCISCO DEL PASO Y TRONCOSO
7	6	TLACOQUEMECATL	BENITO JUAREZ	AV. DE LOS INSURGENTES SUR
8	7	LOS REYES	IZTACALCO	AV PLUTARCO ELIAS CALLES
9	8	CENTRO	CUAUHTEMOC	PLAZA DE LA CONSTITUCION
10	9	LA PRADERA	GUSTAVO A MADERO	FRANCISCO MORAZAN
11	10	ROMA SUR	CUAUHTEMOC	AV. INSURGENTES SUR
12	11	PARQUE NACIONAL CERRO DE LA ESTRELLA	IZTAPALAPA	JAIME M. PULLMAN (JORGE PULLMAN)
13	12	LA MAGDALENA	LA MAGDALENA CONT...	HEROES
14	13	DEL RECREO	AZCAPOTZALCO	CALLE NORTE 87 B
15	14	POLANCO	MIGUEL HIDALGO	AVENIDA EJERCITO NACIONAL

Datos insertados en la tabla Localización

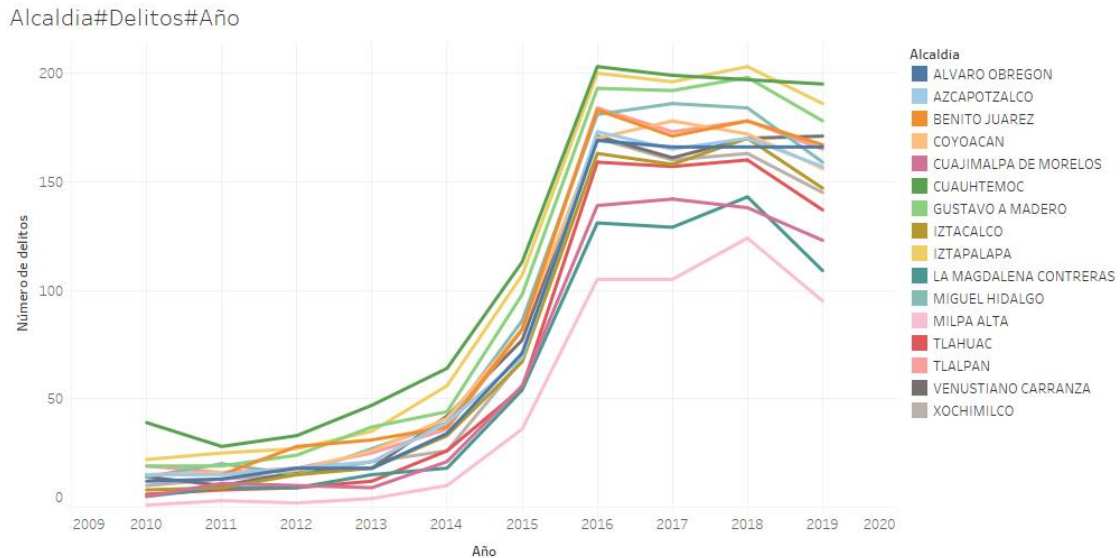
<pre>-- Delitos SELECT TOP (1000) [IDDelito] , [delito] , [categoria_delito] FROM [CarpetasInvestigacion].[dbo].[Delito] order by IDDelito</pre>			
100 %			
Results Messages			
	IDDelito	delito	categoria_delito
1	0	LESIONES CULPOSAS	DELITO DE BAJO IMPACTO
2	1	ROBO DE VEHICULO DE SERVICIO PARTICULAR SIN VIOLEN...	ROBO DE VEHÍCULO CON Y !
3	2	ROBO A TRANSEUNTE EN VIA PUBLICA CON VIOLENCIA	ROBO A TRANSEUNTE EN V
4	3	FRAUDE	DELITO DE BAJO IMPACTO
5	4	ROBO A REPARTIDOR CON VIOLENCIA	ROBO A REPARTIDOR CON Y
6	5	ROBO A TRANSEUNTE DE CELULAR CON VIOLENCIA	DELITO DE BAJO IMPACTO
7	6	ROBO A TRANSEUNTE DE CELULAR CON VIOLENCIA	DELITO DE BAJO IMPACTO
8	7	ROBO DE DINERO	DELITO DE BAJO IMPACTO
9	8	LESIONES INTENCIONALES	DELITO DE BAJO IMPACTO
10	9	FRAUDE	DELITO DE BAJO IMPACTO
11	10	ABUSO DE AUTORIDAD	DELITO DE BAJO IMPACTO
12	11	DAÑO EN PROPIEDAD AJENA CULPOSA POR TRÁNSITO VEH...	DELITO DE BAJO IMPACTO
13	12	ROBO DE VEHICULO DE SERVICIO PÚBLICO SIN VIOLENCIA	ROBO DE VEHÍCULO CON Y !
14	13	DELITOS AMBIENTALES	DELITO DE BAJO IMPACTO
15	14	ROBO DE OBJETOS	DELITO DE BAJO IMPACTO
16	15	ROBO DE OBJETOS	DELITO DE BAJO IMPACTO
17	16	DAÑO EN PROPIEDAD AJENA INTENCIONAL A AUTOMOVIL	DELITO DE BAJO IMPACTO
18	17	ROBO DE OBJETOS	DELITO DE BAJO IMPACTO
19	18	VIOLENCIA FAMILIAR	DELITO DE BAJO IMPACTO
20	19	NARCOMENUDEO POSESION SIMPLE	DELITO DE BAJO IMPACTO

Datos Insertados en la tabla Delito

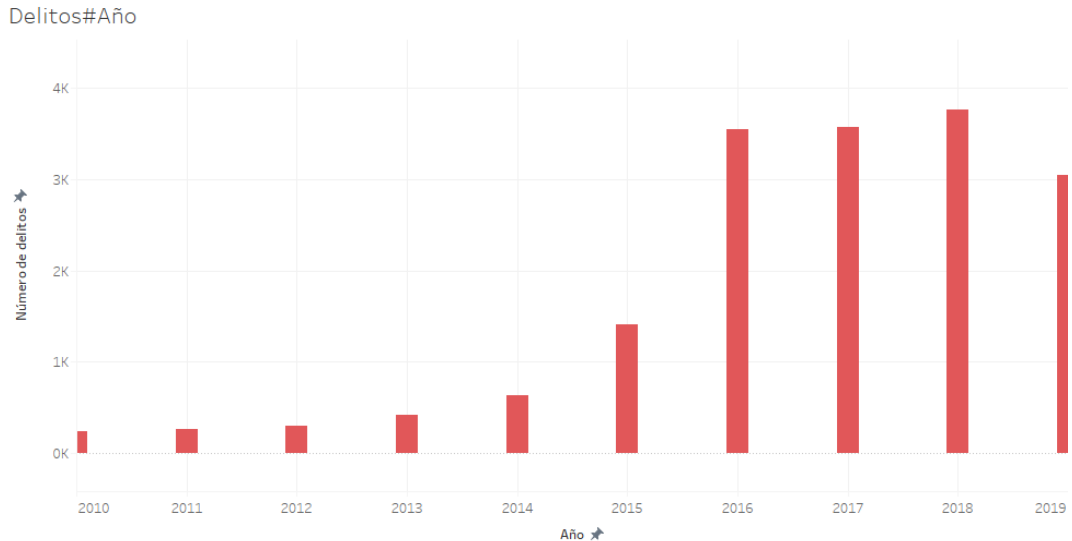
<pre>-- Carpetas SELECT TOP (1000) [IDTiempo] , [IDDelito] , [IDLocalizacion] FROM [CarpetasInvestigacion].[dbo].[Carpetas]</pre>			
100 %			
Results Messages			
	IDTiempo	IDDelito	IDLocalizacion
1	0	0	0
2	1	1	1
3	2	2	2
4	3	3	3
5	4	4	4
6	5	5	5
7	6	6	6
8	7	7	7
9	8	8	8
10	9	9	9
11	10	10	10
12	11	11	11
13	12	12	12
14	13	13	13
15	14	14	14
16	15	15	15
17	16	16	16
18	17	17	17
19	18	18	18

Datos Insertados en la tabla Principal de Hechos Carpetas

Los cubos de datos realizados fueron los siguientes

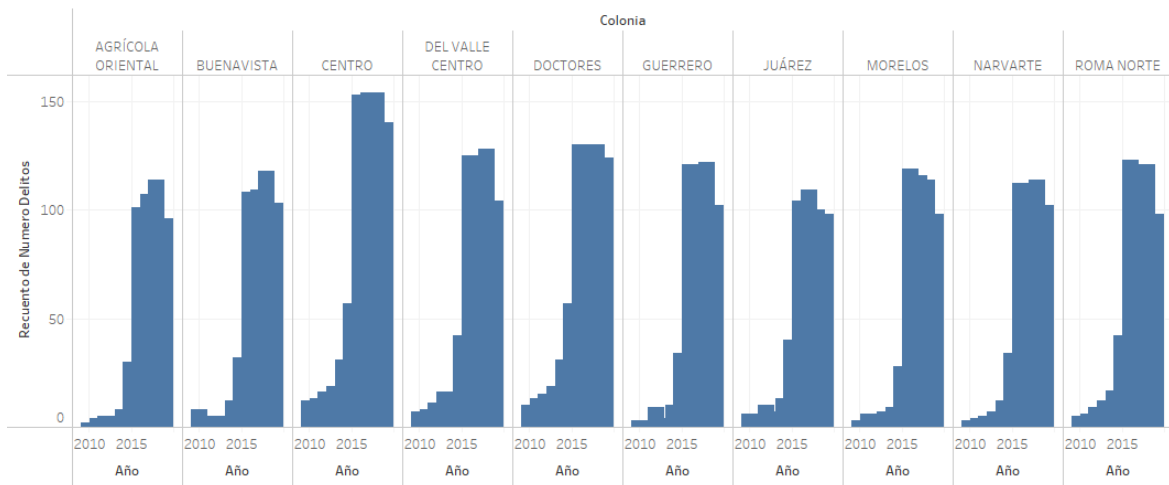


Como primer cubo realizamos el análisis del número de delitos cometidos por años tomando en cuenta desde el 2010 al 2019 , en las diferentes alcaldías de la CDMX, en el cual podemos destacar que Cuauhtemoc es una de las delegaciones con más reportes de delitos a lo largo de los años, manteniéndose al margen con Iztapalapa y Gustavo A Madero, podemos pensar que son los lugares en los cuales la policía no hace mucho ruido. Por lo cual ha sido más fácil para todos los delincuentes hacer de su zona de trabajo estas alcaldías.



En la presente grafica podemos denotar que en los años del 2010 al 2015 se tenía un número considerable de reportes de delitos en la CDMX, pero a partir del 2016 se presentó un aumento descomunal acerca de los casos, podemos suponer que un cambio ya sea tanto político, económico o social surgió a partir de ese año para que las tasas de delincuencia subieron drásticamente.

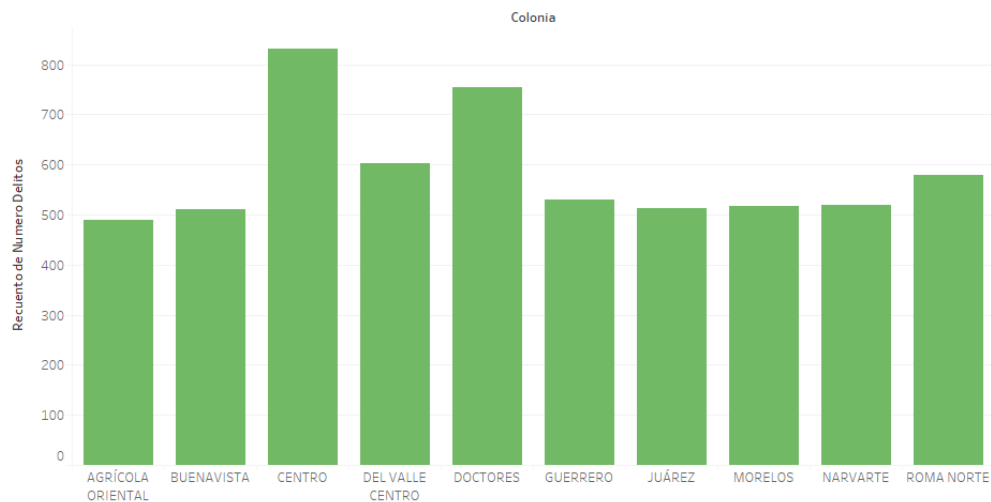
Colonia#Delitos#Año



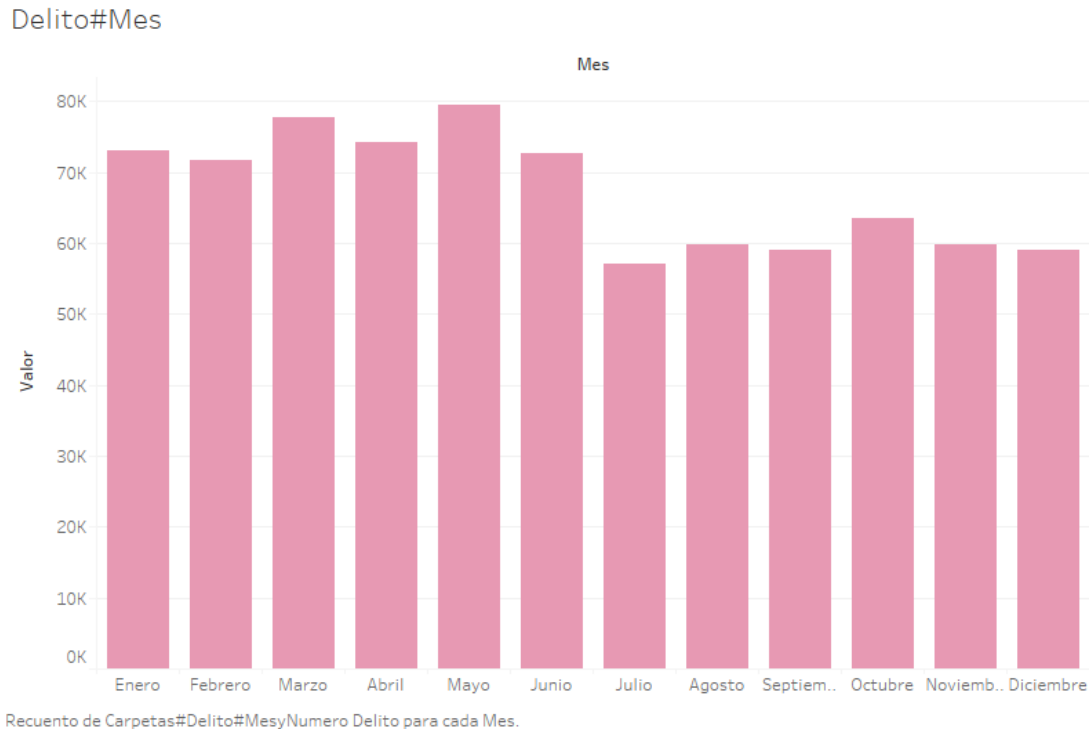
El diagrama de recuento de Numero Delitos para Año desglosado por Colonia. La vista se filtra en AñoColonia. El filtro Año va de 2010 a 2019. El filtro Colonia conserva 10 miembros.

En el siguiente cubo realizamos el número de delitos de las 10 colonias más peligrosas en el periodo del 2010-2019, podemos denotar que en el centro de la CDMX se presentan más reportes en las carpetas de investigación. Podemos concluir que gracias a la gran concurrencia de personas ubicadas en este lugar, es más fácil presentar robos de extorsión, carteristas y muchos otros delitos categorizados.

Colonia#Delitos



Este cubo tiene mucha relación con el anterior, la diferencia que aquí solo nos enfocaremos en los números de delitos cometidos en las colonias más peligrosas entre todas las alcaldías, de las cuales como vimos en la anterior gráfica se encuentra en el primer top el centro de México, seguidas por la colonia doctores, que como es bien sabido el crimen organizado es abundante en este lugar, además que la policía que hace caso omiso a estos actos no ayuda mucho a mejorar este lugar.



La siguiente gráfica presenta el número de delitos cometidos por cada mes entre el periodo de 2010-2019, notamos que los meses con más reportes de delincuencia son Mayo, Marzo, Abril, Enero y Febrero, podemos decir que siempre la primer parte del año tiene más casos de delitos registrados en las carpetas, puede ser por la falta de dinero que presenta la población al haber terminado las festividades de diciembre y empezar con dificultades el año.

Conclusiones.

El explorar la base de datos de víctimas de delitos en la Ciudad de México fue una tarea un tanto ardua, ya que trabajar con datos de hace bastantes años nos obligó a enfocarnos en los últimos 10 años para realizar esta práctica. Así mismo, logramos identificar las alcaldías y colonias donde el índice de reportes por delitos es mayor, por ello son consideradas unas de las más peligrosas en la CDMX. Adicionalmente, logramos observar que de unos años para acá, los reportes por delitos han incrementado, suponemos que esto es debido a los cambios económicos que el país ha sufrido y con la pandemia de COVID-19 esto se podría ver más acentuado.

La creación de los cubos de datos fue la parte más complicada de la práctica debido a que crear cubos con sentido y con la suficiente relevancia es complicado si tenemos en cuenta el volumen de datos y la cantidad de posibilidades a la hora de crear los mismos.