

1. CONJUNTO DE DATOS:

Carpeta de investigación de PGJ de la Ciudad de México

Esta tabla contiene las carpetas de investigación de delitos de la Procuraduría General de Justicia de la Ciudad de México.

Figure 1: TABLA

id	ao_hechos	mes_hechos	fecha_hechos	delito	categoria_delito	fiscalia	agencia	unidad_investigacion	colonia_hechos	a
0	2016.0	Enero	2016-01-05 17:00:00	LESIONES CULPOSAS	DELITO DE BAJO IMPACTO	INVESTIGACIÓN EN GUSTAVO A. MADERO	GAM-6	UI-1CD	GRANJAS MODERNAS	
2	2017.0	Noviembre	2017-11-01 16:40:00	ROBO A TRANSEUNTE EN VIA PUBLICA CON VIOLENCIA	ROBO A TRANSEUNTE EN VÍA PÚBLICA CON Y S...	INVESTIGACIÓN EN IZTACALCO	IZC-2	UI-1SD	GABRIEL RAMOS MILLÁN SECCIÓN TLACOTAL	
3	2015.0	Diciembre	2015-12-30 20:00:00	FRAUDE	DELITO DE BAJO IMPACTO	INVESTIGACIÓN EN IZTAPALAPA	IZP-8	UI-1CD	CONSEJO AGRARISTA MEXICANO	
4	2018.0	Mayo	2018-05-16 16:00:00	ROBO A REPARTIDOR CON VIOLENCIA	ROBO A REPARTIDOR CON Y SIN VIOLENCIA	INVESTIGACIÓN EN IZTAPALAPA	IZP-9	UI-2SD	DESARROLLO URBANO QUETZALCOATL	
5	2018.0	Mayo	2018-05-21 20:40:00	ROBO A TRANSEUNTE DE CELULAR CON VIOLENCIA	DELITO DE BAJO IMPACTO	INVESTIGACIÓN EN CUAUHTEMOC	CUH-2	UI-2SD	INFONAVIT IZTACALCO	

Figure 2: DICCIONARIO

Diccionario de datos	
Columna	Tipo
id	numeric
ao_hechos	numeric
mes_hechos	text
fecha_hechos	timestamp
delito	text
categoria_delito	text
fiscalia	text
agencia	text

Figure 3: DICCIONARIO

unidad_investigacion	text
colonia_hechos	text
alcaldia_hechos	text
fecha_inicio	timestamp
mes_inicio	text
ao_inicio	numeric
calle_hechos	text
calle_hechos2	text
longitud	numeric
latitud	numeric

- id : Identificador
- ao_hechos : Año en el que se reporta el evento
- mes_hechos : Mes en el que sucedió el evento
- fecha_hechos : Momento en el que sucedió el evento
- delito : Descripción del evento
- categoria_delito : Categoría del evento
- fiscalía : Fiscalía en donde se denunció el delito
- agencia : Agencia
- unidad_investigacion : Unidad encargada de la investigación del delito
- colonia_hechos : Colonia donde ocurrió el delito
- alcaldia_hechos : Alcaldía de la CDMX donde ocurrió el delito
- fecha_inicio : Momento en el que se reporta el evento
- mes_inicio : Mes en el que se reporta el evento
- ao_inicio : Año en el que se reporta el evento
- calle_hechos : Nombre de la calle donde sucedió el delito
- calle_hechos2 : Nombre de la calle donde sucedió el delito
- longitud : Longitud donde sucedió el delito

- latitud : Longitud donde sucedió el delito
- geopoint : Desconocida

El examen consiste de tres apartados, cada punto de cada apartado debe documentarse en un PDF a manera de resumen, se deben añadir descripciones detalladas además de gráficos de apoyo.

PARTE I

- Etiquetado de variables
- Calidad de datos: Duplicados, Completitud, Consistencia
- Limpieza de Texto

Los registros eliminados o modificados se deben documentar en el PDF, justificando el por qué de dicha decisión, si se puede realizar un gráfico para detallar casos inconsistentes dichos graficos se deben añadir.

PARTE II - Análisis Exploratorio de Datos En cada punto de debe contestar la siguiente pregunta: ¿Qué puede concluir de esto?

- Años y meses con más delitos
- Delegaciones con más actividad delictiva
- ¿Cuáles son los delitos más frecuentes?
- ¿Qué tipo de delitos son los más frecuentes en la delegación con mayor incidencia delictiva?
- ¿En qué delegaciones suceden los delitos más graves (categoría_delito)?
- En la fiscalía de "juzgados familiares" , ¿Cuáles son los delitos más frecuentes?
- ¿Cuál es el comportamiento del tiempo que pasa desde que el delito sucede hasta que se reporta?
- Añada al menos tres análisis más que considere relevantes para mostrar

PARTE III - Ingeniería de variables/Normalización

- Reduzca la tabla de tal forma que solo se consideren los delitos que ocurrieron después de 2013.
- Normalice la variable categoría_delito , disminuyendo el número de posibilidades, por ejemplo: secuestro, violación, homicidio doloso podría ser una categoría llamada "delito de alto impacto", la forma de normalización es abierta, solo se debe justificar en el PDF ¿por qué se hizo de esa forma?
- Normalice la variable delegación de la siguiente forma :

Figure 4: ZONAS



- Normalice la colonia de los hechos
- Normalice la variable delito, unidad de investigación y la fiscalía , justifique la forma en que lo hizo
- Genere las siguientes variables para fecha_hechos : Cuatrimestre, día de la semana, si es o no fin de semana y día del evento:

Figure 5: FIN DE SEMANA

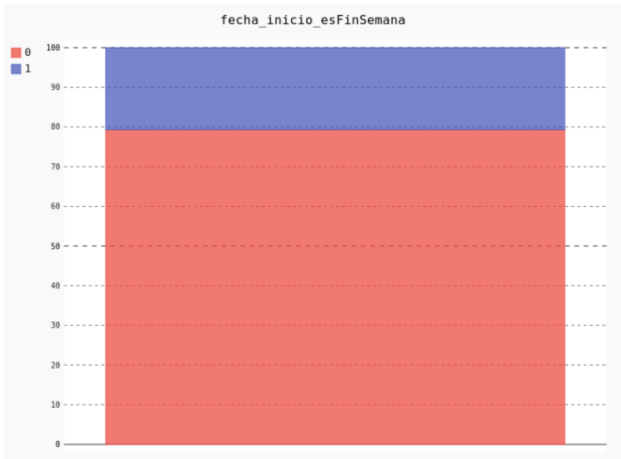


Figure 6: DIA DE LA SEMANA

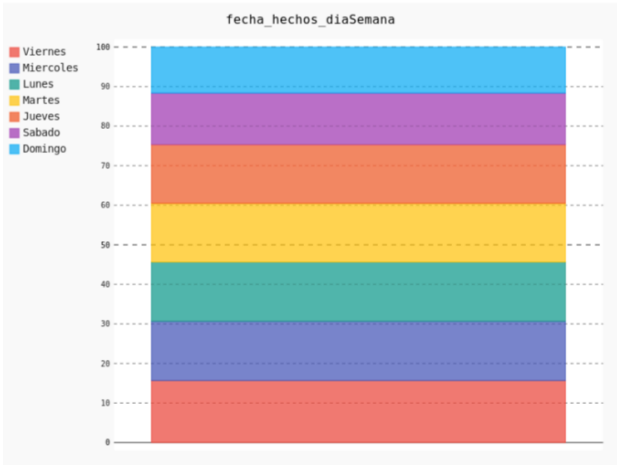


Figure 7: CUATRIMESTRE

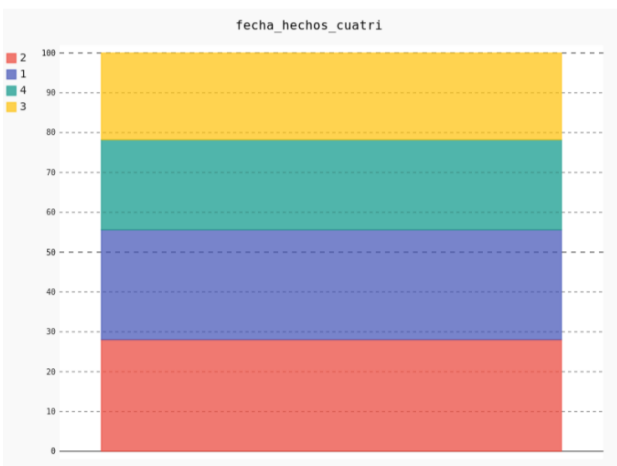
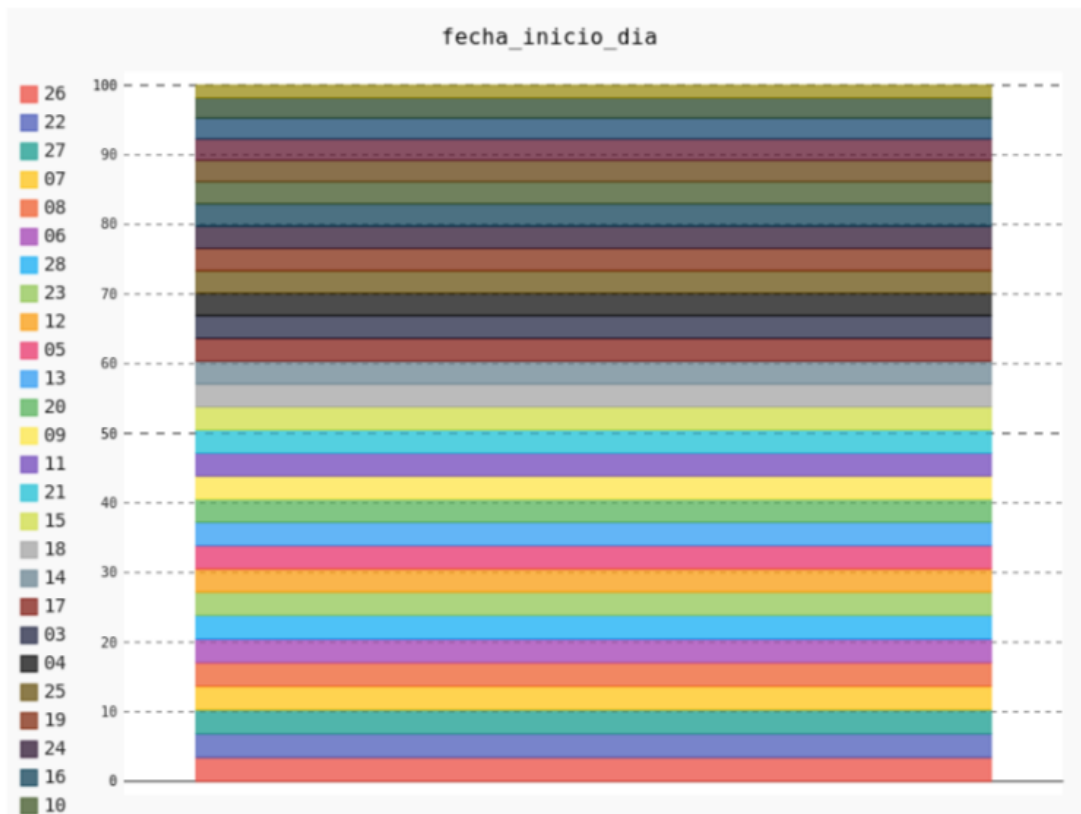
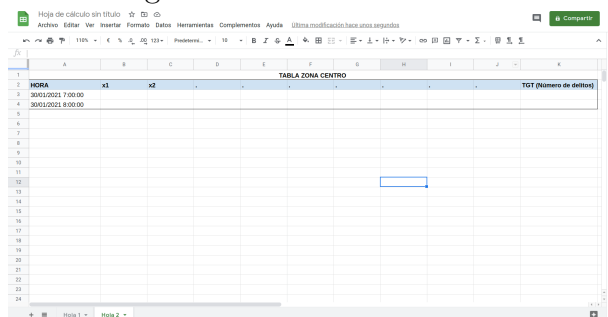


Figure 6. Bill Bush with



- Si el objetivo es predecir el número de delitos que habrá la siguiente hora por zona (Se puede generar una tabla para cada zona), ¿Qué variables propone para que tenga la TAD (x1,x2,x3,...) considerando las variables con las que contamos después de la ingeniería de variables?, no es necesario construirla en python basta con definirla a nivel conceptual, añade un excel donde se ejemplifique qué estructura debería de tener.

Figure 9. ESEM EDX EXCEL



PARTE IV - Adicional

Generé mapa en Kepler u en otra herramienta mostrando los delitos.

PARTE IV - Cuestionario

- ¿Qué es la ciencia de datos?
- ¿Qué habilidades debe dominar un científico de datos?
- ¿Qué es una tabla analítica de datos?
- ¿Qué es la ingeniería de variables?
- Describe la ingeniería de variables posible por cada tipo de variable