

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 desarrollo de la práctica debe de estar ordenada y comentada dentro de un notebook.

**Puntos a desarrollar :**

- Etiquetado de variables, prefijos :
  - Variables numericas : c\_
  - Variables categórica : v\_
  - Variables texto : t\_
  - Variables de fecha : d\_
- Limpieza de las variables categóricas y variables de texto
- Se debe de revisar cada atributo de la calidad de datos para cada variable dentro de la tabla:
  - Duplicados
  - Completitud
  - Consistencia
  - Orden
  - Precisión
- Solo consideré los delitos que ocurrieron después de 2013.
- Normalice la variable categoría\_delito , disminuyendo el número de posibilidades, por ejemplo: secuestro,violacion, homicidio doloso podría ser una categoría llamada "delito de alto impacto", la forma de normalización es abierta.
- Normalice la variable delegación de la siguiente forma :

Figure 4: ZONAS



- Normalice la variable delito 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.