



### Equipo 3

#### **Investigación:**

Delitos y víctimas de la CDMX entre 2019 y junio de 2021.

#### **Base de Datos:**

Delitos denunciados en la CDMX de enero 2019 a junio 2021

12 de agosto del 2021

Nombre	E-mail
Hugo Martínez Ibarra	<a href="mailto:hugomtzib@gmail.com">hugomtzib@gmail.com</a>
Juan de Jesús Hernández Angulo	<a href="mailto:jhernandezangulo@gmail.com">jhernandezangulo@gmail.com</a>
Jaime Israel Vazquez Bernal	<a href="mailto:israfullshot@gmail.com">israfullshot@gmail.com</a>
Gilberto Silva Tijerina	<a href="mailto:gilberto.silvat2812@gmail.com">gilberto.silvat2812@gmail.com</a>
Fernando Arriaga Palma	<a href="mailto:arriaga141@gmail.com">arriaga141@gmail.com</a>
Bryan Daniel Moreno Abrego	<a href="mailto:abre.go@outlook.com">abre.go@outlook.com</a>

Data Science BEDU

Catedrático: Rafael Arias González.

## Índice

### I Resumen

### II Objetivos

1. Identificación del problema (PW 1).
2. Planteamiento de preguntas (PW 2).
3. Base de datos (PW 3).
4. Carga de datos (PW 4).
5. Limpieza de datos (PW 5).
6. Uso de API (PW 6).
7. Transformación y filtración de datos (PW 7).
8. Notebook y planes a futuro (PW 8).
9. Problemas durante el desarrollo.

## Resumen

En la Ciudad de México suceden hechos delictivos diariamente. Dichos suceden en diversas localidades en diferentes horarios. Y a través de los años se ha buscado recopilar la información de cada uno de ellos de la manera más detalla posible. Y una dependencia encargada de dicha tarea es el C5 de la Ciudad de México. *“El Centro de Comando, Control, Cómputo, Comunicaciones y Contacto Ciudadano de la CDMX (C5), es la dependencia del Gobierno de la Ciudad de México encargada de captar información integral para la toma de decisiones en materia de seguridad pública, urgencias médicas, medio ambiente, protección civil, movilidad y servicios a la comunidad en la capital del país a través del video monitoreo, de la captación de llamadas telefónicas y de aplicaciones informáticas de inteligencia, enfocadas a mejorar la calidad de vida de las y los capitalinos.” [1]*

Luego de recopilar los datos se necesita de personal capacitado para darles un adecuado tratamiento, con la finalidad de que la información resultante sea de ayuda para prevenir dichas situaciones. En el presente proyecto se desarrolló justamente esa parte. La que realiza un científico de datos.

El objetivo del proyecto es presentar el desarrollo de la limpieza de datos de un conjunto de datos emitido por el C5 de la ciudad de México. El cual empieza desde la detección de un problema, hasta la obtención del dataset limpio y adecuado para ser usado en el siguiente módulo de Data Science. Se pretende usar Python como herramienta principal.

## Palabras clave

Delincuencia

Datos

Limpieza

## Objetivos

1. **Identificación del problema:** identificar y escoger un problema (justificar su elección). Presentar investigación preliminar completa del tema.  
Presentar información previa sobre las soluciones al problema
2. **Planteamiento de preguntas:** Las son correctamente respondidas. La entrega tiene al menos 5 preguntas pertinentes sobre el problema.
3. **Colección de datos:** Obtener una colección de datos completa y no previamente procesada que usará para los siguientes pasos.  
Los datos recopilados pueden responder completamente las preguntas planteadas anteriormente.
4. **Análisis exploratorio de los datos:** Convertir de forma adecuada su colección de Datos (en JSON, CSV, etc...) en un DataFrame de Pandas. Realizar un análisis exploratorio de estos datos a través del uso correcto de Pandas.
5. **Limpieza de datos:** El nuevo Dataset no contiene ningún NaN.  
El Dataset está correctamente indexado (en orden y sin índices incoherentes). Las columnas están correctamente nombradas.  
Se aplicaron agregaciones al Dataset para poder comenzar a responder algunas preguntas planteadas en la sesión 2.
6. **Transformación de datos:** Todos los datos de cada columna tienen un valor correcto dependiendo del tipo de dato (fechas, objetos, ints, floats). Todas las columnas de texto están manipuladas para estar en el formato correcto.  
Hay nuevas columnas con nuevos datos resultantes del procesamiento de una o más columnas originales.  
Se crearon otros datasets/subconjuntos/series que presentan información relevante para una pregunta específica, o simplemente para exploraciones más profundas.
7. **Ordenamiento de código:** Entregar un Jupyter Notebook ordenado y limpio que contiene todos los pasos llevados a cabo para el preprocesamiento de datos. El Notebook hace uso de las diversas técnicas aprendidas a través de las clases para la exploración y procesamiento de datos. El Notebook contiene el código (nombrado correctamente) de todas las sesiones pasadas.
8. **Opcional uso de APIs**

## 1. Identificación del Problema (Postwork 1)

*“Hay que señalar que la violencia ha incrementado su presencia en nuestra vida cotidiana difícilmente parece novedoso, pero no deja de ser necesario. La violencia homicida ha arrebatado la vida de 200 mil personas durante los últimos 11 años, principalmente, de nuestros jóvenes. 2017 ha sido el año más violento del que se tiene registro, con más de 25 mil casos de homicidio y una tasa de 20.5 homicidios por cada 100,000 habitantes. [Según datos del Secretariado Ejecutivo del Sistema Nacional de Seguridad Pública (cifras al 20 de mayo de 2018, con corte al 31 de diciembre de 2017).] [2]*

Es evidente que la inseguridad y la delincuencia se sigue elevando año tras año a nivel nacional. La sociedad se acostumbra a que hechos delictivos sean tan cotidianos, que se ha convertido en algo “común y normal” ver como suceden hasta el punto de permitir que sucedan. Cabe resaltar que esto es más reiterado en ciertos lugares del país como lo es la Ciudad de México (CDMX).

### 1.1 Problemática

El problema identificado es el siguiente:

**Alto grado de delincuencia en la CDMX entre 2019 y junio del 2021.**

#### 1.1.1 Justificación

Se detectó un fuerte problema nacional: el aumento de la delincuencia. Como se menciono anteriormente, el proyecto se centra únicamente a las localidades de la CDMX. Esto porque en esta ciudad existe un alto grado de concentración de delincuencia. Es una cuestión que debe ser resuelta con las herramientas que estén a disposición.

Además, al equipo le interesó conocer este problema desde la fuente original de información: los conjuntos de datos generados por las autoridades. De modo que pudiéramos percibir los factores o agentes involucrados en la denuncia de delitos, y la información esencial de la denuncia de un delito para su correspondiente apertura de una carpeta de investigación.

El panorama de la Ciencia de Datos nos dice que este problema (como muchos otros) puede ser abordado y tratado. ¿Cómo? A través de la inmensa cantidad de datos que surgen día con día al suceder los hechos delictivos. Actualmente, en la CDMX, se cuenta con grandes y avanzados sistemas de monitoreo y reportes.

Estos generan los datos que se van a necesitar para realizar modelos que puedan predecir o prevenir incidentes delictivos.

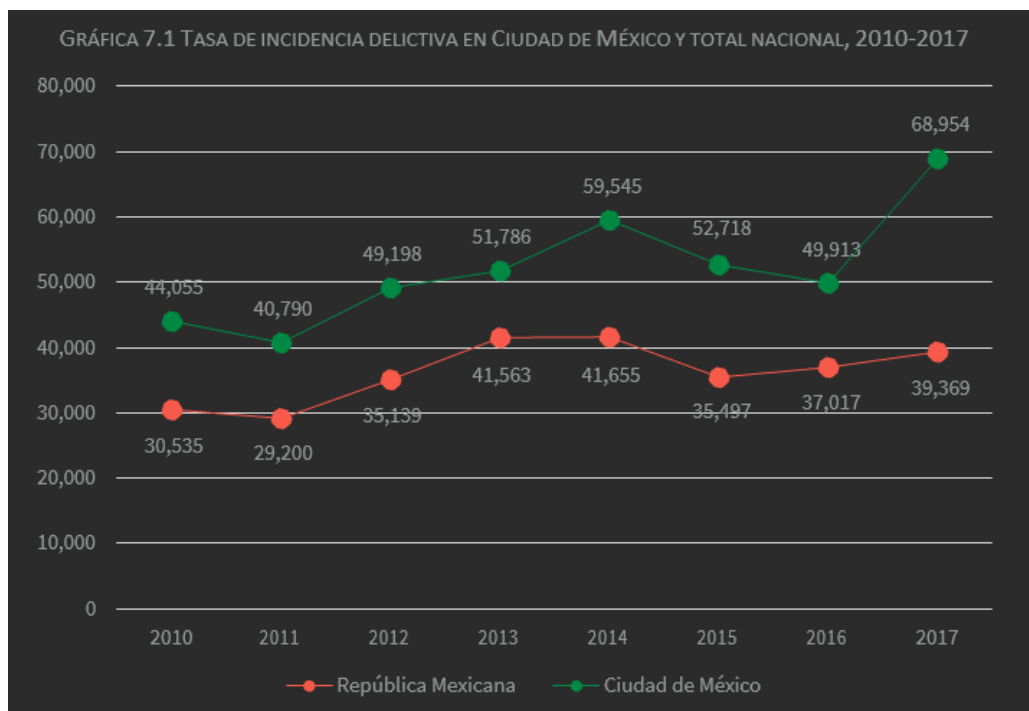
Es por estas razones que se decidió seleccionar este tema y esta problemática. Se cuenta con el tiempo y con capacidad necesaria para desarrollar el proyecto, es decir, que es un proyecto viable en todos sus ámbitos.

## 1.2 Investigación del problema

En varias investigaciones y documentos relacionado con el aumento de la delitos, violencia y denuncias coinciden en que dichos índices han ido creciendo con el paso del tiempo. Cada uno de ellos ha llegado a la misma conclusión. Fragmentos de ellos serán mostrados a continuación.

*“Las tasas a lo largo del tiempo presentan una tendencia ascendente tanto para la Ciudad de México como para el conjunto del país, pero notoriamente más elevada en el primer caso, como corresponde a sus tasas medias de crecimiento: entre 2010 y 2017 la tasa de incidencia delictiva en la capital del país ha aumentado a un ritmo interanual de 6.6%, mientras que para el total nacional el crecimiento medio de un año a otro ha sido de 3.7%. El segundo resultado que debe ser destacado es el aumento de la incidencia delictiva en la Ciudad de México entre 2016 y 2017, pues entre ambos años la tasa cre- ció 38.2%, la mayor variación relativa bianual registrada desde el año 2010.” [3]*

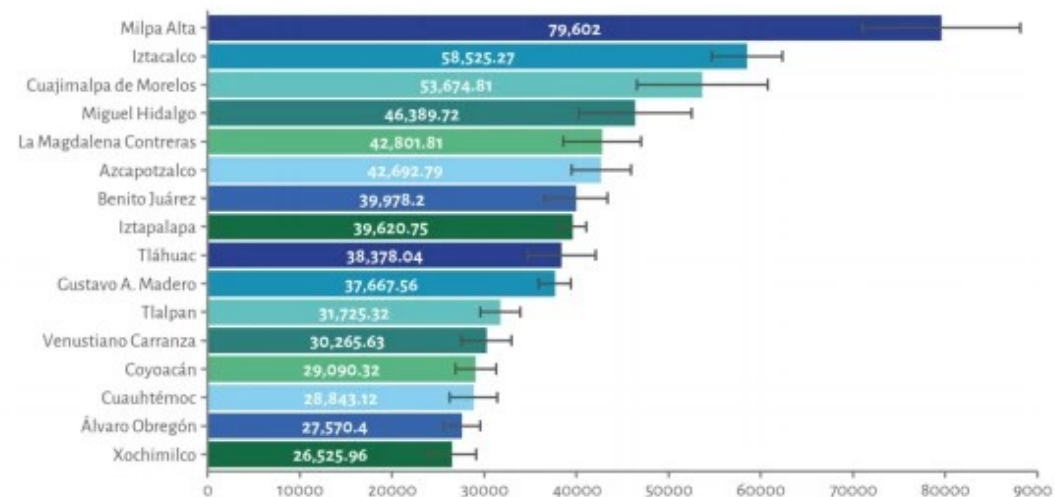
## Tasa de incidencia delictiva en Ciudad de México y Total Nacional, 2010-2017



Fuente: Evalúa Ciudad de México a partir de ENVIPE (2011 a 2018).

Diversas fuentes de divulgación periodística han registrado de igual manera un notable crecimiento en el aumento de la delincuencia. También a su vez han realizado investigaciones para determinar en cuales alcaldías se han registrado un mayor número de incidentes delictivos. Los resultados fueron los siguientes:

“Las alcaldías con mayor cantidad de delitos por cada 100,000 habitantes fueron Milpa Alta, Iztacalco y Cuajimalpa, en ese orden; mientras, los que tuvieron una incidencia menor fueron Xochimilco, Cuauhtémoc y Álvaro Obregón.”



Del mismo sustrato se hace evidente que en los hogares es reportado en una gran parte de la población por lo menos algún incidente delictivo en contra de algún miembro de dicho hogar.

*“Durante 2020, 53.8% de los hogares reportó que al menos uno de sus miembros sufrió un delito, lo que resulta en alrededor de 2 millones 695,682 delitos en la capital de enero a septiembre pasado.”*

*No obstante, 81.1% de las personas no denunció, mientras 11.5% de los denunciante sí acudió a reportar ante las autoridades, pero se le negó la apertura de una carpeta de investigación.”*

El sentimiento inseguridad es una realidad que se comparte en la población nacional. Pero que, en ciertos lugares, como lo es la Ciudad de México se vive en escalas indignantes. Desde fraudes, que es el delito con mayor cifra negra en la ciudad, pues 97.2% de los casos no son denunciados, hasta robos de vehículos



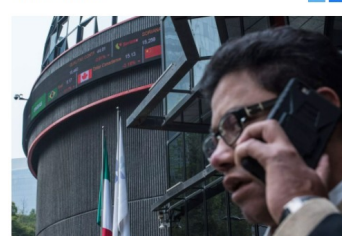
con un porcentaje de denuncia de 53.5% han sido las cifras que se han registrado en esta investigación.

Imágenes que representan el problema:

## Extorsión y fraude, los delitos que crecieron en la CDMX durante la pandemia

La delincuencia se ha adaptado ante el confinamiento causado por el COVID-19, alerta el Observatorio Nacional Ciudadano. El 97.2% de los casos de extorsión no se denuncia.

mié 24 marzo 2021 05:43 PM



Casi 9 de cada 10 imprevistos de extorsión se realizan vía telefónica. (Miguel Valiente / Cuartoscuro)



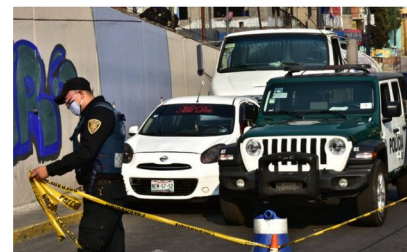
## El Centro, la zona más peligrosa de la CDMX

Sólo en la zona, que coincide con la operación de La Unión Tepito, hay registro de 245 carpetas de investigación por homicidio

CDMX / LUNES 25 DE MAYO DE 2020

## Delincuencia también entra en cuarentena en la ciudad

En el Portal de Datos Abiertos se observan bajas de casi 50 por ciento en delitos como robo a peatón



El hurto de vehículos mostró una baja al registrar solo 310 indagatorias / Foto: David Deleante

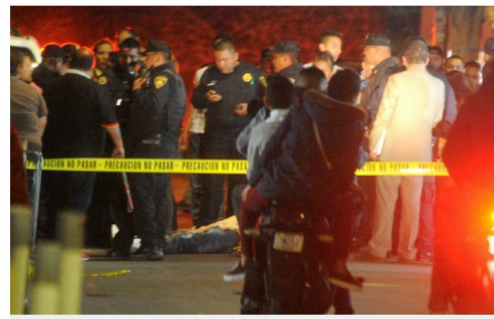


Foto Cuartoscuro

CDMX registró un aumento de delitos del 5.9% en los últimos 10 meses



Cuartoscuro

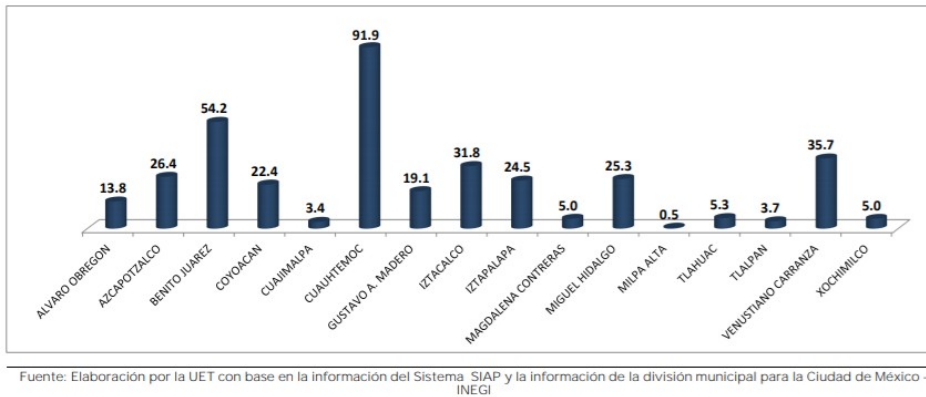
Por Redacción Animal Político | 5 de octubre, 2019

Imágenes de El Sol De México ([www.elsoldemexico.com.mx](http://www.elsoldemexico.com.mx)), EXPANSIÓN política ([politica.expansion.mx](http://politica.expansion.mx)) y Animal Politico ([www.animalpolitico.com](http://www.animalpolitico.com)).

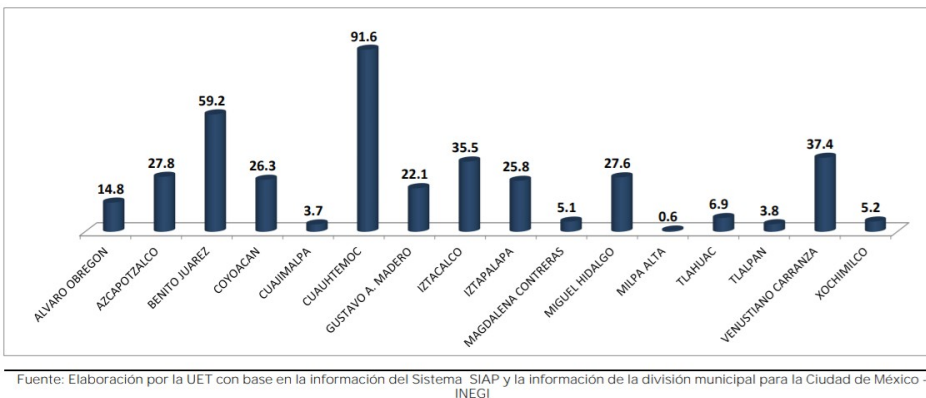
La Fiscalía General de Justicia de la Ciudad de México tiene en su sitio en internet estadísticas delictivas de cada mes de cada año. En ellas se pudo sustraer la siguiente información.

En cada gráfica se muestran los delitos cometidos por kilómetro cuadrado en cada una de las Alcaldías.

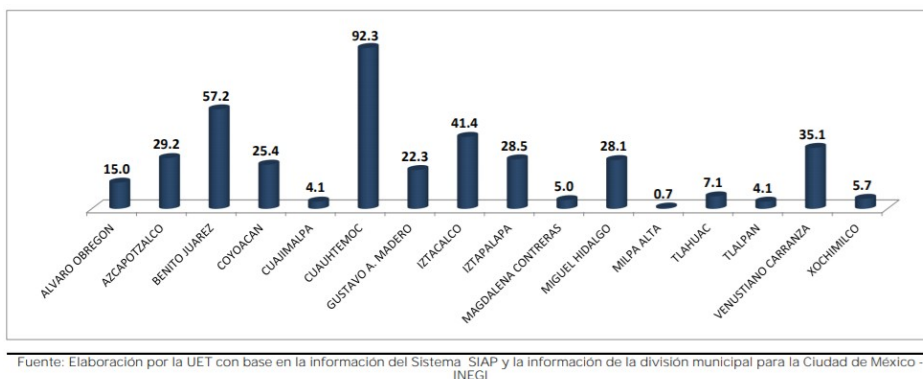
Enero 2020:



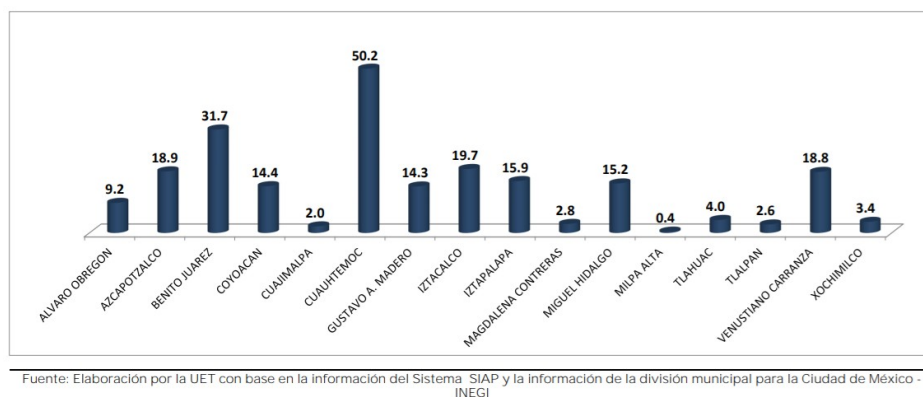
Febrero 2020:



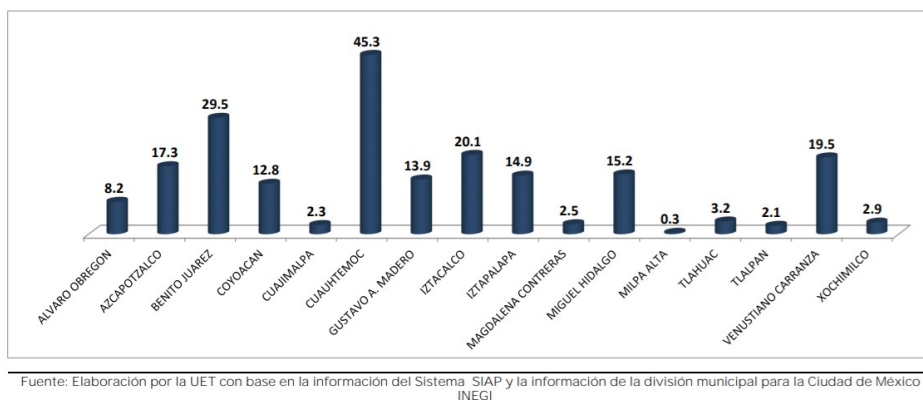
Marzo 2020:



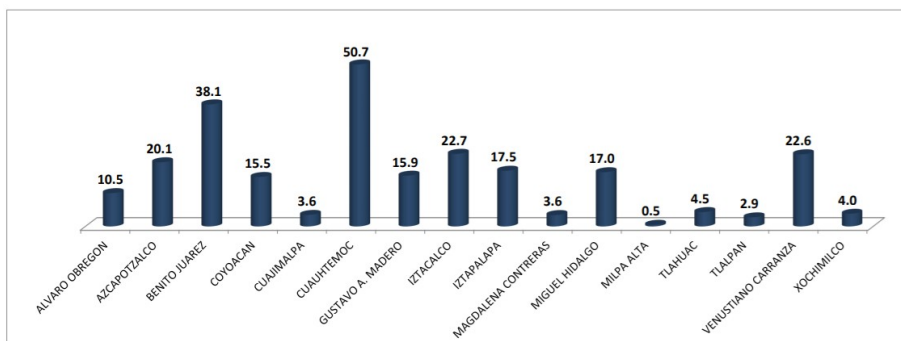
Abril 2020:



Mayo 2020:

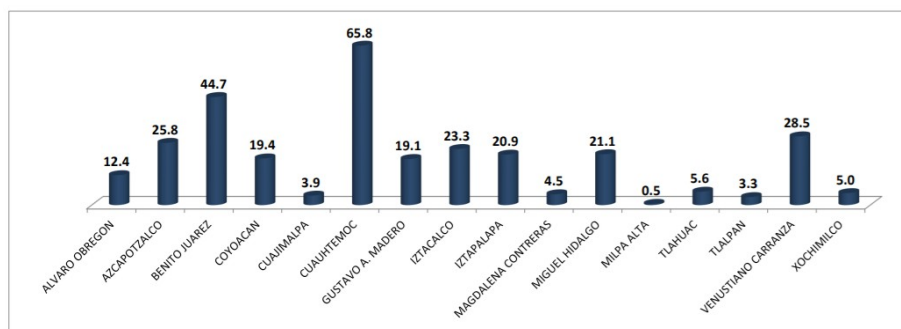


Junio 2020:



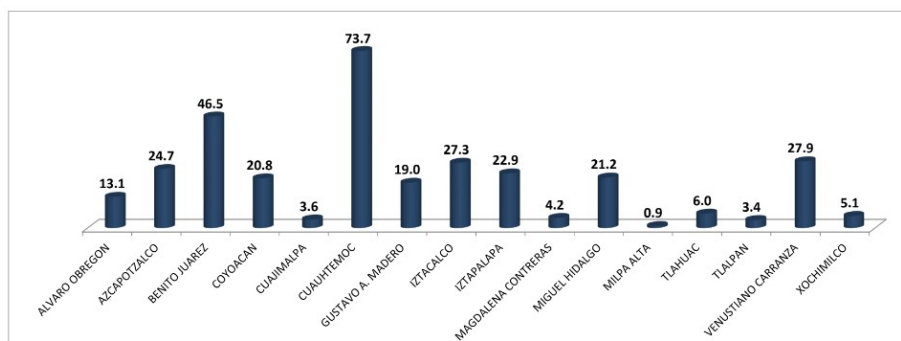
Fuente: Elaboración por la UET con base en la información del Sistema SIAP y la información de la división municipal para la Ciudad de México - INEGI

Julio 2020:



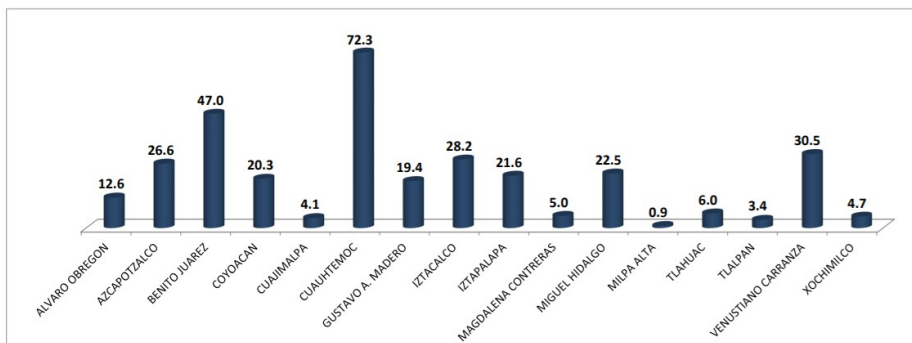
Fuente: Elaboración por la UET con base en la información del Sistema SIAP y la información de la división municipal para la Ciudad de México - INEGI

Agosto 2020:



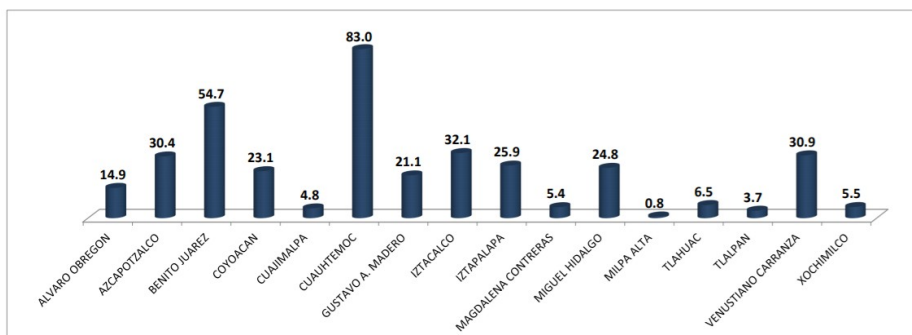
Fuente: Elaboración por la UET con base en la información del Sistema SIAP y la información de la división municipal para la Ciudad de México - INEGI

## Septiembre 2020:



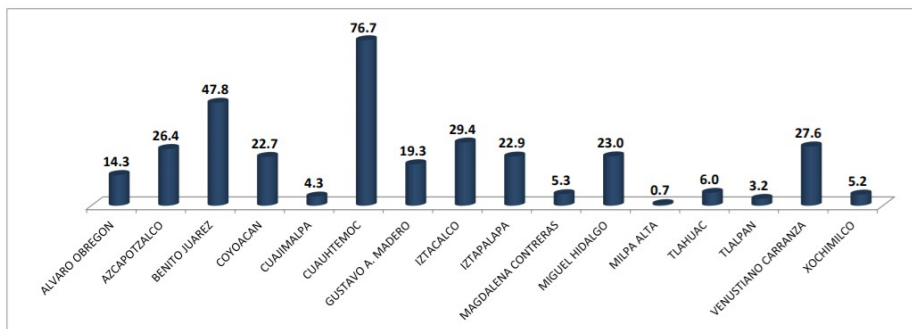
Fuente: Elaboración por la UET con base en la información del Sistema SIAP y la información de la división municipal para la Ciudad de México - INEGI

## Octubre 2020:



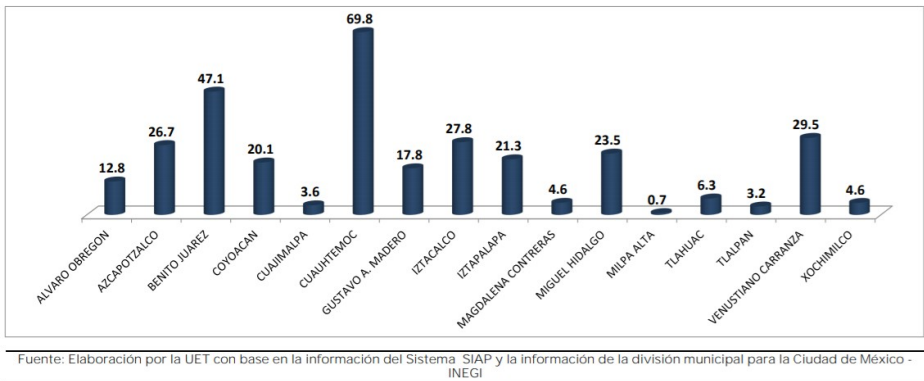
Fuente: Elaboración por la UET con base en la información del Sistema SIAP y la información de la división municipal para la Ciudad de México - INEGI

## Noviembre 2020:



Fuente: Elaboración por la UET con base en la información del Sistema SIAP y la información de la división municipal para la Ciudad de México - INEGI

Diciembre 2020:



### 1.3 Búsqueda de soluciones anteriores

La delincuencia no es un problema nuevo en la CDMX, y durante el paso de los años se han dado propuestas que intentan disminuir los índices de incidentes delictivos.

Según el sitio de Política Expansión se especifican estrategias que usa el gobierno durante la actual presidencia de Andrés Manuel López Obrador:

*“Nuevo Modelo Policial:*

*Si como objetivo se plantea la creación de la Guardia Nacional como un cuerpo permanente que pueda servir a futuros gobiernos, como una estrategia se prevé articular los esfuerzos y las aportaciones de los tres órdenes de gobierno y tomar en cuenta condiciones, contextos y necesidades locales para impulsar un Nuevo Modelo Policial. En él, se busca que los elementos deban tener una visión de derechos humanos, proximidad y participación ciudadana. El eslabón más importante es la policía municipal, pues es la que está más cerca de los ciudadanos.*

*Prevención del delito*

*Esto incluye a su vez varias acciones:*

- *Desarrollo alternativo. Plantea dar alternativas económicas para los hogares y las comunidades que dependen del ingreso provisto por actividades ilícitas.*
- *Prevención especial de la violencia y el delito. Son intervenciones restaurativas, orientadas a protección (en el caso de los menores que son forzados a cometer delitos), resocialización y reparación del daño.*
- *Disuasión focalizada. Plantea el diseño y la implementación de estrategias que permitan identificar dónde y quiénes se dedican a actividades ilícitas, comunicarles las consecuencias, movilizar a la comunidad para generar conciencia y ofrecer apoyo integral a quienes deseen abandonar la actividad ilícita.*
- *Cumplimiento regulatorio. Se promueve que dependencias o instancias que puedan estar involucradas con actos ilícitos se apeguen a estándares de operación y cumplimiento legal.*
- *Justicia procesal, proximidad y de atención a víctimas. Velar por que las autoridades de seguridad y justicia cuenten con herramientas y protocolos para consolidar su legitimidad frente a la sociedad.*



### *Participación ciudadana*

*En esta estrategia se plantea recuperar espacios públicos, combatir los delitos que más afectan a la sociedad y conformar redes vecinales.*

### *Cambio en el reparto recursos para seguridad*

*De acuerdo con esta estrategia, se hará un cambio en los criterios, las fórmulas y las variables para la asignación, la administración y el ejercicio del Fondo de Aportaciones para la Seguridad Pública (FASP), y se les dará prioridad a los siguientes lineamientos:*

- Profesionalización, certificación y capacitación de policías estatales, municipales, custodios y ministeriales con énfasis en sistema de justicia penal acusatorio, derechos humanos y perspectiva de género.*
- Equipamiento policial y de instituciones, con unidades especializadas en combate al secuestro, cibernética, búsqueda de personas y sistema penal.*
- Fortalecimiento y/o creación de las unidades de inteligencia financiera en los estados.*
- Construcción, mejora y equipamiento de los servicios médicos forenses.*
- Fortalecimiento del sistema penitenciario nacional y de ejecución de medidas para adolescentes.*

### *Combate al lavado de dinero*

*Las operaciones con recursos de procedencia ilícita, defraudación fiscal y finanzas de la delincuencia organizada son áreas a las que hasta el momento no se les había dado atención, según el gobierno, por lo que en la estrategia se prevé que la Unidad de Inteligencia Financiera (UIF) de la Secretaría de Hacienda (SHCP) tenga un papel más importante.*

### *Combate al robo en carreteras*

*Se realizarán operaciones de revisión y supervisión del transporte en rutas con alta incidencia delictiva, así como filtros de revisión y arcos en centrales de autobuses, revisiones de afectaciones en tramos carreteros y blindaje de accesos a autopistas; también, se implementará el uso de arcos gamma para detectar armas y drogas y el acceso a cámaras de vigilancia.”*



## **EL C5 de la Ciudad de México**

*“El Centro de Atención a Emergencias y Protección Ciudadana de la Ciudad de México (CAEPCCM), fue creado el 22 de Junio de 2009 y a partir del año siguiente comenzó a operar el programa “Ciudad Segura” a través de las cámaras de video vigilancia, que permite mejorar la reacción de las autoridades ante emergencias, situaciones de crisis y comisión de ilícitos. Ciudad Segura es el programa en materia de seguridad urbana más ambicioso a nivel mundial.*

*Para fortalecer y crecer los servicios que proporciona, así como su nivel de calidad en la atención ciudadana, el 23 de diciembre de 2015 el Gobierno de la Ciudad de México decretó la anexión de LOCATEL al CAEPCCM para conformar el Centro de Comando, Control, Cómputo, Comunicaciones y Contacto Ciudadano de la Ciudad de México “C5”, el cual actualmente ofrece los servicios de:*

- *Video Monitoreo*
- *Servicio de Atención de llamadas de emergencia 9-1-1 CDMX*
- *Denuncia Anónima 089*
- *LOCATEL 5658 1111*

*Todos los servicios del C5 operan las 24 horas, los 365 días del año.” [10]*

## **Cámaras con IA de la CDMX ya podrán evitar delitos**

*“Jean Marie Letort, vicepresidente del consejo de ciberseguridad de Thales, indicó a Expansión que gracias a una licitación que ganó con el gobierno de la Ciudad de México ahora será posible “no sólo vigilar y seguir la huella de algún criminal, sino que podremos prevenir delitos a través de sistemas de videovigilancia que usan inteligencia artificial con el fin de prevenir y evitar que algún delincuente culmine un delito”.*

*Con aproximadamente 5,000 unidades, la empresa de origen francés será la indicada para soportar y apoyar al C5 en temas de prevención de delito, gracias a una licitación que ganó en la actual gestión de Claudia Sheinbaum. El proyecto en el que participa Thales forma parte de la modernización de la red de cámaras de videovigilancia del gobierno actual y que es denominado Ciudad Segura.*

*El C5 opera y monitorea las más de 15,000 cámaras de vigilancia que hay en el Sistema Tecnológico de Videovigilancia (STVs), así como las 6,000 cámaras instaladas en el Sistema de Transporte Colectivo Metro, “con la finalidad de*

*prevenir y alertar inmediatamente a las autoridades de seguridad y de emergencias capitalinas sobre cualquier situación de riesgo” y por ello es que ha buscado proveedores que le ayuden a cumplir con este objetivo, entre ellos Thales, empresa de seguridad y logística de origen francés.*

*En el momento de su anuncio, Sheinbaum indicó que el plan de esta modernización es integrar al C5 como un organismo que opere de forma cercana en las tareas policiales y de protección civil que requiere la capital mexicana y destacó que existía obsolescencia tecnológica en 6,588 cámaras de videovigilancia, por lo que se harían mejoras en términos de monitoreo, procesamiento y almacenamiento de los datos que recopilan estas unidades.*

*De acuerdo con el gobierno de la Ciudad de México, el contrato de licitación está planteado en conjunto con Thales, Telmex y Cisco, quienes modernizarán esta red de cámaras y puertos de Wifi gracias a una inversión gubernamental de 500 millones de dólares.*

*La empresa de origen francés ha hecho inversiones interesantes en términos de infraestructura de seguridad y operación logística, entre ellas crecer el talento en México y en América Latina y en la compra de algunas empresas como Gemalto, la cual adquirió en abril del presente año por 4,800 millones de euros (5,372 millones de dólares) y con quien buscó aliarse para brindar mayor énfasis en temas de ciberseguridad y salvaguarda de datos.”*

[11]

## 2. Planteamiento de preguntas

A continuación se plantean una serie de preguntas para comenzar a conocer el problema con base en los datos que se tuvieron disponibles.

Estas preguntas permitirán identificar algunos de los factores clave en el problema de las víctimas y las denuncias hechas en la CDMX.

- ¿Cuáles son los delitos más comunes?
- ¿Hay una tendencia de mayor número de delitos o denuncias en algún mes?
- ¿Cuáles son los lugares con mayor/menor incidencia de delitos?
- ¿Cuáles son algunas estadísticas de las víctimas (edad, sexo)?
- Diferencia de tiempo entre el momento de los hechos y el momento de la denuncia.
- Delitos denunciados en la CDMX, pero no ocurridos en la ciudad.

Adicionalmente, durante el procesamiento de la información surgieron otras preguntas que más adelante se comentarán.

### 3. Colección de datos

Para darle solución al problema planteando a través de las herramientas que puede proporcionar un Científico de Datos, se ha elegido una colección de datos del gobierno de la CDMX, dentro del portal de datos abiertos en su sitio web.

El link para entrar a la siguiente colección de datos es el siguiente:

<https://datos.cdmx.gob.mx/dataset/victimas-en-carpetas-de-investigacion-fgj>

En él se encuentra un archivo de tipo de CSV, lleno de datos relacionados justamente con el tema seleccionado. También se encuentran archivos que explican los datos que encuentran en la colección. Dentro del sitio se menciona lo siguiente:

#### ***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 hasta junio de 2021.*

*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.*
- *Estas y otras aclaraciones las puedes encontrar en el Diccionario que se adjunta a la presente base.*

La base de datos puede ser consultada y descargada desde el siguiente enlace directo:

[https://archivo.datos.cdmx.gob.mx/fiscalia-general-de-justicia/victimas-en-carpetas-de-investigacion-fgj/victimas\\_completa\\_junio\\_2021.csv](https://archivo.datos.cdmx.gob.mx/fiscalia-general-de-justicia/victimas-en-carpetas-de-investigacion-fgj/victimas_completa_junio_2021.csv)

Un primer acercamiento de los datos, usando el software de Excel, es el siguiente:

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
idCarpeta	Año_inicio	Mes_inicio	FechaInicio	Delito	Categoría	Sexo	Edad	TipoPersona	CalidadJurídica	competencia	Año_hecho	Mes_hecho	Fechahecho	HoraInicio	Horahecho	Alcaldiahecho	Coloniahecho	Calle_hecho	Calle_hecho	latitud	longitud	
8324429	2019	Enero	04/01/2019	FRAUDE	DELITO DE BAJA IMPACTO	Masculino	62.0	FISICA	OFENDIDO	FUERO COMÚN	2018	Agosto	29/08/2018	12:00:00	12:19:00	ALVARO OBR	GUADALUPE INSURGENTES NA			19.36125	-99.18314	
8324430	2019	Enero	04/01/2019	PRODUCCIÓN, IMPRESIÓN, ENAJENACIÓN, DISTRIBUCIÓN...	DELITO DE BAJA IMPACTO	Femenino	38.0	FISICA	VICTIMA Y DENUNCIANTE	FUERO COMÚN	2018	Diciembre	15/12/2018	15:00:00	12:20:00	AZCAPOTZALCO	VICTORIA DE AV. CUATLA NA			19.47181	-99.10458	
8324431	2019	Enero	04/01/2019	ROBO A TRANSUNTE SALIENDO DEL BANCO CON VIOLENCIA	ROBO A TRANSUNTE SALIENDO DEL BANCO CON VIOLENCIA	Masculino	42.0	FISICA	VICTIMA Y DENUNCIANTE	FUERO COMÚN	2018	Diciembre	22/12/2018	15:30:00	12:23:00	COYOACAN	COPOLCO UN COPOLCO NA			19.33297	-99.18611	
8324435	2019	Enero	04/01/2019	ROBO DE VEHICULO DE SERVICIO PARTICULAR SIN VIOLENCIA	ROBO DE VEHICULO DE SERVICIO PARTICULAR SIN VIOLENCIA	Masculino	35.0	FISICA	VICTIMA Y DENUNCIANTE	FUERO COMÚN	2019	Enero	04/01/2019	06:00:00	12:27:00	ITZACALCO	AGRAICOLA FCALLE 6 ENTRE PRIVA			19.40327	-99.09883	
8324438	2019	Enero	04/01/2019	ROBO DE VEHICULO DE SERVICIO PARTICULAR SIN VIOLENCIA	ROBO DE VEHICULO DE SERVICIO PARTICULAR SIN VIOLENCIA	Masculino	35.0	FISICA	VICTIMA Y DENUNCIANTE	FUERO COMÚN	2019	Enero	03/01/2019	20:00:00	12:35:00	ITZAPALAPA	PROGRESIST UNIVERSIDA NA			19.3548	-99.06324	
8324442	2019	Enero	04/01/2019	PRODUCCIÓN, IMPRESIÓN, ENAJENACIÓN, DISTRIBUCIÓN...	DELITO DE BAJA IMPACTO	Femenino	42.0	FISICA	OFENDIDO	FUERO COMÚN	2018	Octubre	12/10/2018	18:00:00	12:38:00	COYOACAN	PUEBLO DE L PEDRO ENRI NA			19.33537	-99.16016	
8324444	2019	Enero	04/01/2019	ROBO A TRANSUNTE SALIENDO DEL BANCO CON VIOLENCIA	ROBO A TRANSUNTE SALIENDO DEL BANCO CON VIOLENCIA	Masculino	35.0	FISICA	VICTIMA Y DENUNCIANTE	FUERO COMÚN	2019	Enero	04/01/2019	08:30:00	12:40:00	ALVARO OBR	TOLTECA CALLE 10 NA			19.39	-99.19472	
8324451	2019	Enero	04/01/2019	VIOLACIÓN	VIOLACIÓN	Masculino	13.0	FISICA	VICTIMA	FUERO COMÚN	2010	Diciembre	19/12/2010	03:00:00	12:48:00	GUSTAVO A ILA CALILLA	CORONA DE INDETERMIN			19.55832	-99.12933	
8324454	2019	Enero	04/01/2019	PRODUCCIÓN, IMPRESIÓN, ENAJENACIÓN, DISTRIBUCIÓN...	DELITO DE BAJA IMPACTO	Masculino	83.0	FISICA	VICTIMA Y DENUNCIANTE	FUERO COMÚN	2018	Diciembre	27/12/2018	13:00:00	12:51:00	ITZAPALAPA	BANIDAL SUR 69 -A CALZ. ERMIT.			19.35996	-99.13896	
8324455	2019	Enero	04/01/2019	OMISION DE DEBITO DE BAJA IMPACTO	DELITO DE BAJA IMPACTO	Masculino	15.0	FISICA	VICTIMA	FUERO COMÚN	2018	Enero	04/01/2018	13:00:00	12:51:00	VENUSTIANK NA	ORIENTE 156 NOTIFICACION NA					
8324457	2019	Enero	04/01/2019	DESPOJO	DELITO DE BAJA IMPACTO	Masculino	45.0	FISICA	VICTIMA Y DENUNCIANTE	FUERO COMÚN	2018	Mayo	11/05/2018	10:00:00	12:52:00	ITZAPALAPA	SANTIAGO A REFORMA (N NA			19.35325	-99.00041	
8324459	2019	Enero	04/01/2019	ROBO A TRANSUNTE SALIENDO DEL BANCO CON VIOLENCIA	ROBO A TRANSUNTE SALIENDO DEL BANCO CON VIOLENCIA	Masculino	27.0	FISICA	VICTIMA Y DENUNCIANTE	FUERO COMÚN	2018	Enero	04/01/2018	08:30:00	12:53:00	VENUSTIANK COCHILLA	PARRO CHURUB CALLE 1			19.42102	-99.06775	
8324465	2019	Enero	04/01/2019	PRODUCCIÓN, IMPRESIÓN, ENAJENACIÓN, DISTRIBUCIÓN...	DELITO DE BAJA IMPACTO	Femenino	56.0	FISICA	VICTIMA Y DENUNCIANTE	FUERO COMÚN	2018	Diciembre	03/12/2018	14:30:00	12:59:00	GUSTAVO A LUNDAD HCAV	OTON (MIGUEL BERR			19.50059	-99.13985	
8324469	2019	Enero	04/01/2019	PRODUCCIÓN, IMPRESIÓN, ENAJENACIÓN, DISTRIBUCIÓN...	DELITO DE BAJA IMPACTO	Femenino	31.0	FISICA	VICTIMA Y DENUNCIANTE	FUERO COMÚN	2018	Diciembre	19/12/2018	10:00:00	13:05:00	XOCHIMILCO	L NORIA GUADALUPE NA			19.26772	-99.12733	
8324477	2019	Enero	04/01/2019	ROBO DE VEHICULO DE SERVICIO PARTICULAR SIN VIOLENCIA	ROBO DE VEHICULO DE SERVICIO PARTICULAR SIN VIOLENCIA	Masculino	61.0	FISICA	VICTIMA Y DENUNCIANTE	FUERO COMÚN	2019	Enero	03/01/2019	19:00:00	13:14:00	MIGUEL HIDALGO	TLAXAPANA TLALOC NA			19.441	-99.16927	
8324479	2019	Enero	04/01/2019	ROBO A TRANSUNTE SALIENDO DEL BANCO CON VIOLENCIA	ROBO A TRANSUNTE SALIENDO DEL BANCO CON VIOLENCIA	Masculino	39.0	FISICA	VICTIMA Y DENUNCIANTE	FUERO COMÚN	2018	Diciembre	25/12/2018	04:00:00	13:16:00	CUAUHTEMOC	DOCTORES DOCTOR LUCI DOCTOR LUC			19.42444	-99.15237	
8324482	2019	Enero	04/01/2019	ROBO DE VEHICULO DE SERVICIO PARTICULAR SIN VIOLENCIA	ROBO DE VEHICULO DE SERVICIO PARTICULAR SIN VIOLENCIA	Masculino	37.0	FISICA	OFENDIDO	FUERO COMÚN	2018	Diciembre	31/12/2018	11:00:00	13:19:00	CUAUHTEMOC	MORELOS CONSTANCIA NA			19.44962	-99.12782	
8324485	2019	Enero	04/01/2019	ROBO DE VEHICULO DE SERVICIO PARTICULAR SIN VIOLENCIA	ROBO DE VEHICULO DE SERVICIO PARTICULAR SIN VIOLENCIA	Masculino	37.0	FISICA	OFENDIDO	FUERO COMÚN	2018	Diciembre	31/12/2018	11:00:00	13:23:00	CUAUHTEMOC	ROMA NORT CARROBA NA			19.41359	-99.15729	
8324486	2019	Enero	04/01/2019	ABUSO DE CREDITO DE BAJA IMPACTO	DELITO DE BAJA IMPACTO	Femenino	72.0	FISICA	OFENDIDO	FUERO COMÚN	2018	Noviembre	22/11/2018	14:00:00	13:23:00	GUSTAVO A ISAN JUAN	AVENIDA 15 NA			19.46801	-99.06586	
8324487	2019	Enero	04/01/2019	ROBO A TRANSUNTE SALIENDO DEL BANCO CON VIOLENCIA	ROBO A TRANSUNTE SALIENDO DEL BANCO CON VIOLENCIA	Masculino	44.0	FISICA	VICTIMA Y DENUNCIANTE	FUERO COMÚN	2019	Enero	04/01/2019	11:40:00	13:24:00	ITZAPALAPA	ITZLAHUAC VALLE NACIC NA			19.33477	-99.00053	
8324491	2019	Enero	04/01/2019	VIOLACIÓN	VIOLACIÓN	Femenino	76.0	FISICA	VICTIMA Y DENUNCIANTE	FUERO COMÚN	2019	Enero	04/01/2019	12:35:00	13:29:00	VENUSTIANK ARENAL	LA S CUAUHTEMOC			19.42655	-99.06463	
8324495	2019	Enero	04/01/2019	PRODUCCIÓN, IMPRESIÓN, ENAJENACIÓN, DISTRIBUCIÓN...	DELITO DE BAJA IMPACTO	Femenino	56.0	FISICA	VICTIMA Y DENUNCIANTE	FUERO COMÚN	2018	Diciembre	22/12/2018	11:35:00	13:34:00	BENITO JUAREZ	AV. PATRIOT CALLE HOLBE			19.38005	-99.18487	
8324496	2019	Enero	04/01/2019	ROBO A TRANSUNTE SALIENDO DEL BANCO CON VIOLENCIA	ROBO A TRANSUNTE SALIENDO DEL BANCO CON VIOLENCIA	Masculino	36.0	FISICA	VICTIMA Y DENUNCIANTE	FUERO COMÚN	2019	Enero	03/01/2019	15:00:00	13:35:00	CUAUHTEMOC	JUAREZ ESTOCOLMO NA			19.42676	-99.1664	
8324497	2019	Enero	04/01/2019	ROBO A TRANSUNTE SALIENDO DEL BANCO CON VIOLENCIA	ROBO A TRANSUNTE SALIENDO DEL BANCO CON VIOLENCIA	Masculino	30.0	FISICA	VICTIMA Y DENUNCIANTE	FUERO COMÚN	2019	Enero	03/01/2019	20:00:00	13:37:00	MIGUEL HIDALGO	POPOTLA AV. CUTLASH NA			19.45796	-99.17955	
8324503	2019	Enero	04/01/2019	VIOLACIÓN	VIOLACIÓN	Femenino	9.0	FISICA	VICTIMA	FUERO COMÚN	2018	Diciembre	14/12/2018	03:30:00	13:41:00	ITZAPALAPA	LA COLMENA AVENIDA TE NA			19.372	-98.99966	
8324504	2019	Enero	04/01/2019	ROBO A TRANSUNTE SALIENDO DEL BANCO CON VIOLENCIA	ROBO A TRANSUNTE SALIENDO DEL BANCO CON VIOLENCIA	Masculino	40.0	FISICA	VICTIMA Y DENUNCIANTE	FUERO COMÚN	2018	Diciembre	06/12/2018	14:30:00	13:42:00	ITZAPALAPA	TEPAPATES PERIFERICO (CASI ESQUIN			19.39706	-99.05837	
8324506	2019	Enero	04/01/2019	DELITOS DE BAJA IMPACTO	DELITO DE BAJA IMPACTO	Femenino	40.0	FISICA	VICTIMA Y DENUNCIANTE	FUERO COMÚN	2018	Julio	23/07/2018	17:00:00	13:43:00	ITZAPALAPA	UNIDAD VICI NA CALLE DE NA			19.36056	-99.056	
8324507	2019	Enero	04/01/2019	ROBO A TRANSUNTE SALIENDO DEL BANCO CON VIOLENCIA	ROBO A TRANSUNTE SALIENDO DEL BANCO CON VIOLENCIA	Masculino	25.0	FISICA	VICTIMA Y DENUNCIANTE	FUERO COMÚN	2018	Diciembre	11/12/2018	19:00:00	13:44:00	BENITO JUAREZ	DEL VALLE CEFLEX CUEVA INSURGENTE			19.37353	-99.1786	
8324511	2019	Enero	04/01/2019	AMENAZAS	DELITO DE BAJA IMPACTO	Femenino	57.0	FISICA	VICTIMA Y DENUNCIANTE	FUERO COMÚN	2019	Enero	04/01/2019	12:20:00	13:47:00	ITZAPALAPA	TENORIOS QUETZALCO NA			19.32881	-99.02166	
8324512	2019	Enero	04/01/2019	LESIONES IN LESIONES DE BAJA IMPACTO	LESIONES IN LESIONES DE BAJA IMPACTO	Masculino	NA	FISICA	VICTIMA	FUERO COMÚN	2019	Enero	04/01/2019	12:30:00	13:47:00	TLAHUAC	LAS PUERTAS AMAROLA (N NA			19.29405	-99.02049	
8324515	2019	Enero	04/01/2019	ROBO A TRANSUNTE SALIENDO DEL BANCO CON VIOLENCIA	ROBO A TRANSUNTE SALIENDO DEL BANCO CON VIOLENCIA	Masculino	NA	FISICA	VICTIMA Y DENUNCIANTE	FUERO COMÚN	2019	Enero	04/01/2019	08:45:00	13:48:00	VENUSTIANK MOCTEZUMA EE 1	(CALLE ORIENTE 158			19.43023	-99.09511	
8324520	2019	Enero	04/01/2019	ROBO A TRANSUNTE SALIENDO DEL BANCO CON VIOLENCIA	ROBO A TRANSUNTE SALIENDO DEL BANCO CON VIOLENCIA	Masculino	60.0	FISICA	VICTIMA Y DENUNCIANTE	FUERO COMÚN	2018	Diciembre	20/12/2018	12:00:00	13:53:00	MIGUEL HIDALGO	ESCANZANA AV. PROGRES NA			19.40546	-99.18163	
8324522	2019	Enero	04/01/2019	ROBO A TRANSUNTE SALIENDO DEL BANCO CON VIOLENCIA	ROBO A TRANSUNTE SALIENDO DEL BANCO CON VIOLENCIA	Masculino	NA	FISICA	VICTIMA Y DENUNCIANTE	FUERO COMÚN	2018	Diciembre	04/12/2018	19:40:00	13:53:00	GUSTAVO A ESTRELLA	JOYAS NA			19.47576	-99.11819	
8324522	2019	Enero	04/01/2019	ROBO A TRANSUNTE SALIENDO DEL BANCO CON VIOLENCIA	ROBO A TRANSUNTE SALIENDO DEL BANCO CON VIOLENCIA	Masculino	NA	MORAL	OFENDIDO	FUERO COMÚN	2018	Diciembre	04/12/2018	19:40:00	13:53:00	GUSTAVO A ESTRELLA	JOYAS NA			19.47576	-99.11819	
8324527	2019	Enero	04/01/2019	ROBO A TRANSUNTE SALIENDO DEL BANCO CON VIOLENCIA	ROBO A TRANSUNTE SALIENDO DEL BANCO CON VIOLENCIA	Masculino	NA	FISICA	VICTIMA Y DENUNCIANTE	FUERO COMÚN	2018	Diciembre	22/12/2018	18:34:00	13:59:00	CUAUHTEMOC	GUERRERO AV. HIDALGO NA			19.43769	-99.14753	
8324528	2019	Enero	04/01/2019	ROBO A TRANSUNTE SALIENDO DEL BANCO CON VIOLENCIA	ROBO A TRANSUNTE SALIENDO DEL BANCO CON VIOLENCIA	Masculino	77.0	FISICA	VICTIMA Y DENUNCIANTE	FUERO COMÚN	2018	Enero	04/01/2018	11:00:00	14:00:00	ITZAPALAPA	EJERCITO DE FUERTE DE L ESQUINA CO			19.36248	-99.03152	

## 4. Carga de datos

La carga de datos se llevó a cabo de la manera usual, mediante el uso de funciones del paquete Pandas de Python. Aunque al principio nos encontramos con un pequeño problema que comentaremos más adelante.

Al cargar la base de datos en un notebook de Google Colab, plataforma donde llevamos a cabo nuestro proyecto, se puede observar de la siguiente forma:

```
[2] data = pd.read_csv("https://archivo.datos.cdmx.gob.mx/fiscalia-general-de-justicia/victimas-en-carpeta-de-investigacion-fg/victimas_completa_junio_2021.csv")

# Convertimos la base de datos a un data frame.
df = pd.DataFrame(data)
df.head()
```

	idCarpeta	Año_inicio	Mes_inicio	FechaInicio	Delito	Categoría	Sexo	Edad	TipoPersona	CalidadJurídica	competencia	Año_hecho	Mes_hecho	Fechahecho	HoraInicio	Horahecho	Alcaldiahecho	Coloniahecho	Calle_hecho	latitud	longitud	
0	8324429	2019	Enero	04/01/2019	FRAUDE	DELITO DE BAJA IMPACTO	Masculino	62.0	FISICA	OFENDIDO	FUERO COMÚN	2018.0	Agosto	29/08/2018	12:00:00	12:19:00	ALVARO OBR	GUADALUPE INSURGENTES NA		19.36125	-99.18314	G
1	8324430	2019	Enero	04/01/2019	PRODUCCIÓN, IMPRESIÓN, ENAJENACIÓN, DISTRIBUCIÓN...	DELITO DE BAJA IMPACTO	Femenino	38.0	FISICA	VICTIMA Y DENUNCIANTE	FUERO COMÚN	2018.0	Diciembre	15/12/2018	15:00:00	12:20:00	AZCAPOTZALCO	VICTORIA DE AV. CUATLA NA		19.47181	-99.10458	VI DEB
2	8324431	2019	Enero	04/01/2019	ROBO A TRANSUNTE SALIENDO DEL BANCO CON VIOLENCIA	ROBO A TRANSUNTE SALIENDO DEL BANCO CON VIOLENCIA	Masculino	42.0	FISICA	VICTIMA Y DENUNCIANTE	FUERO COMÚN	2018.0	Diciembre	22/12/2018	15:30:00	12:23:00	COYOACAN	COPOLCO UN COPOLCO NA		19.33297	-99.18611	UN
3	8324435	2019	Enero	04/01/2019	ROBO DE VEHICULO DE SERVICIO PARTICULAR SIN VIOLENCIA	ROBO DE VEHICULO DE SERVICIO PARTICULAR SIN VIOLENCIA	Masculino	35.0	FISICA	VICTIMA Y DENUNCIANTE	FUERO COMÚN	2019.0	Enero	04/01/2019	06:00:00	12:27:00	ITZACALCO	AGRAICOLA FCALLE 6 ENTRE PRIVA		19.40327	-99.09883	
4	8324438	2019	Enero	04/01/2019	ROBO DE VEHICULO DE SERVICIO PARTICULAR SIN VIOLENCIA	ROBO DE VEHICULO DE SERVICIO PARTICULAR SIN VIOLENCIA	Masculino	NaN	FISICA	VICTIMA	FUERO COMÚN	2019.0	Enero	03/01/2019	20:00:00	12:35:00	ITZAPALAPA	PROGRESIST UNIVERSIDA NA		19.3548	-99.06324	PRI

Esta base de datos consiste de 598,402 registros ( renglones) y 22 campos (o columnas)

También se detectó que existían columnas cuyo tipo de datos no era el más adecuado para la información que expresaba o una manipulación posterior, o

incluso diferencia de tipo de datos para columnas que representaban la misma información, como las columnas que indicaban año, una estaba con tipo int64 y otra con float64.

```
[4] # Para ver el tipo de datos de cada columna.
df.dtypes

idCarpeta          int64
Año_inicio         int64
Mes_inicio         object
FechaInicio        object
Delito             object
Categoria          object
Sexo              object
Edad              float64
TipoPersona        object
CalidadJuridica    object
competencia        object
Año_hecho          float64
Mes_hecho          object
FechaHecho         object
HoraHecho          object
HoraInicio         object
AlcaldiaHechos     object
ColoniaHechos      object
Calle_hechos       object
Calle_hechos2      object
latitud            float64
longitud           float64
dtype: object
```

## 5. Limpieza de datos

Después de haber cargado exitosamente la base de datos y explorar el tipo de datos que se tenían, se procedió a dar el mejor formato adecuado a cada columna, así como limpiar la base de datos de registros incompletos (con valor NaN) de información como se puede apreciar:

```
[5] # El número de valores NaN por cada columna.
df.isna().sum(axis=0)

idCarpeta          0
Año_inicio         0
Mes_inicio         0
FechaInicio        0
Delito             0
Categoria          0
Sexo              120629
Edad              221401
TipoPersona        5149
CalidadJuridica    1
competencia        0
Año_hecho          288
Mes_hecho          288
FechaHecho         288
HoraHecho          286
HoraInicio         1
AlcaldiaHechos     415
ColoniaHechos      24270
Calle_hechos       1119
Calle_hechos2      380760
latitud            24045
longitud           24047
dtype: int64
```



Entre las primeras modificaciones y transformaciones estuvieron las requeridas para poner el formato correcto de fecha en Python, y aprovechando el hecho de que las fechas en Python pueden estar acompañadas de la hora, se unieron estos datos donde fue posible, para poder manipular estos datos más adelante, como obtener la diferencia entre dos fechas.

```
# Para comenzar, se concatenan los datos de fecha y hora.
df_base["FechaInicio"] = df_base["FechaInicio"] + " " + df_base["HoraInicio"]
df_base["FechaHecho"] = df_base["FechaHecho"] + " " + df_base["HoraHecho"]
df_base.head()
```

	idCarpeta	Año_inicio	Mes_inicio	FechaInicio	Delito	Categoria	Sexo	Edad	TipoPersona	CalidadJuridica	competencia	Año_hecho	Mes_hecho	FechaHecho	HoraHecho	HoraInicio	AlcaldiaHechos	Colo
0	8324429	2019	Enero	04/01/2019 12:19:00	FRAUDE	DELITO DE BAJO IMPACTO	Masculino	62	FISICA	OFENDIDO	FUERO COMUN	2018.0	Agosto	29/08/2018 12:00:00	12:00:00	12:19:00	ALVARO OBREGON	G
1	8324430	2019	Enero	04/01/2019 12:20:00	PRODUCCION, IMPRESION, ENAJENACION, DISTRIBUCI...	DELITO DE BAJO IMPACTO	Femenino	38	FISICA	VICTIMA Y DENUNCIANTE	FUERO COMUN	2018.0	Diciembre	15/12/2018 15:00:00	15:00:00	12:20:00	AZCAPOTZALCO	VI DEH
2	8324431	2019	Enero	04/01/2019 12:23:00	ROBO A TRANSEUNTE SALIENDO DEL BANCO CON VIOLE...	ROBO A CUENTAHABIENTE SALIENDO DEL CAJERO CON ...	Masculino	42	FISICA	VICTIMA Y DENUNCIANTE	FUERO COMUN	2018.0	Diciembre	22/12/2018 15:30:00	15:30:00	12:23:00	COYOACAN	UN
3	8324435	2019	Enero	04/01/2019 12:27:00	ROBO DE VEHICULO DE SERVICIO PARTICULAR SIN VI...	ROBO DE VEHICULO CON Y SIN VIOLENCIA	Masculino	35	FISICA	VICTIMA Y DENUNCIANTE	FUERO COMUN	2019.0	Enero	04/01/2019 06:00:00	06:00:00	12:27:00	IZTACALCO	
4	8324442	2019	Enero	04/01/2019 12:38:00	PRODUCCION, IMPRESION, ENAJENACION, DISTRIBUCI...	DELITO DE BAJO IMPACTO	Femenino	42	FISICA	OFENDIDO	FUERO COMUN	2018.0	Octubre	12/10/2018 18:00:00	18:00:00	12:38:00	COYOACAN	I

Otra cuestión importante fue también reconocer los criterios para eliminar datos. En primer lugar consideramos eliminar registros de columnas donde claramente es una imprecisión en la recolección de datos, como falta de dato sobre la alcaldía donde sucedió el delito, o tipo de persona implicada en la denuncia, o algún dato de la fecha de la denuncia o delito denunciado. Sin embargo, hubo un tipo de caso especial para el cual tomamos un criterio particular, las denuncias por personas morales donde no había dato de sexo y edad (ambos ausentes) pues muchos de estos delitos eran del tipo agresión contra locales, y la denuncia no era hecha por una persona física en particular, sino quizá por el propio negocio afectado. Por eso se decidió dar a estos casos, en los campos de Sexo y Edad la descripción "No aplica".

Finalmente, se decidió eliminar también la columna que indicaba el dato de la segunda calle de referencia donde ocurrió el delito, pues era un dato ausente en muchos casos y eliminar registros debido a la falta de esta información equivalía a remover aproximadamente la mitad de nuestros datos.

Después de todo esto, la base de datos quedó limpia como se muestra a continuación:

```
[10] # Se elimina la columna Calle_hechos2
df_cleaned.drop(columns="Calle_hechos2", inplace=True)
df_cleaned.reset_index(drop=True, inplace=True)
# Ahora la base de datos ya está completamente limpia
df_cleaned.isna().sum(axis=0)

idCarpeta      0
Año_inicio     0
Mes_inicio     0
FechaInicio    0
Delito         0
Categoria      0
Sexo           0
Edad           0
TipoPersona    0
CalidadJuridica 0
competencia    0
Año_hecho      0
Mes_hecho      0
FechaHecho     0
HoraHecho      0
HoraInicio     0
AlcaldiaHechos 0
ColoniaHechos  0
Calle_hechos   0
latitud        0
longitud       0
dtype: int64
```

```
idCarpeta      int64
AñoInicio      int64
MesInicio      object
FechaInicio    datetime64[ns]
Delito         object
Categoria      object
Sexo           object
Edad           object
TipoPersona    object
CalidadJuridica object
Competencia    object
AñoHecho       int64
MesHecho       object
FechaHecho     datetime64[ns]
HoraHecho      object
HoraInicio     object
AlcaldiaHechos object
ColoniaHechos  object
CalleHechos    object
Latitud        float64
Longitud       float64
dtype: object
```

## 6. Uso de API

En nuestro caso no fue necesario el empleo de una API para la extracción de información, sin embargo, no quisimos perdernos la experiencia y llevamos a cabo un pequeño ejercicio (no relacionado con este tema) sobre el uso de API.

Las peticiones que se llevaron a cabo fueron a la API de la USGS (*U.S. Geological Survey*) para obtener información sobre terremotos. La documentación se encuentra en: <https://earthquake.usgs.gov/fdsnws/event/1/>

Vale la pena mencionar que esto nos sirvió para practicar lo visto en la sesión correspondiente de uso de API y reconocer la disponibilidad y beneficios de estas herramientas.



## Postwork 6: Automatización y APIs

Para este postwork utilizamos información que no está relacionada con el dataset que originalmente se está trabajando

Para realizar peticiones HTTP, se utiliza el paquete requests de Python.

```
In [1]: # !pip install requests
import requests
```

Las peticiones se realizarán a la API de la USGS (U.S. Geological Survey) para obtener información sobre terremotos. La documentación se encuentra en: <https://earthquake.usgs.gov/fdsnws/event/1/>

El end-point a utilizar es

```
In [2]: endpoint = 'https://earthquake.usgs.gov/fdsnws/event/1/query'
```

Los parámetros que se utilizarán en esta petición son

- format: Formato de los datos de salida.
- starttime: Fecha de inicio de incidentes
- endtime: Fecha de fin de incidentes
- minlatitude: Latitud mínima en una geometría (región delimitada) rectangular
- minlongitude: Longitud mínima en una geometría (región delimitada) rectangular
- maxlatitude: Latitud máxima en una geometría (región delimitada) rectangular
- maxlongitude: Longitud máxima en una geometría (región delimitada) rectangular

Deseamos obtener los datos de terremotos ocurridos entre el 01/jul/2021 y el 01/ago/2021, en México (aproximadamente entre latitudes 13.607153 a 32.534628 y longitudes -117.124463 a -85.839584) y en formato GeoJSON, el cual es un caso especial de JSON que almacena información geológica.

Por tanto, el objeto de parámetros queda como

```
In [3]: payload = {'format': 'geojson',
                  'starttime': '2021-07-01',
                  'endtime': '2021-08-01',
                  'minlatitude': 13.607153,
                  'minlongitude': -117.124463,
                  'maxlatitude': 32.534628,
                  'maxlongitude': -85.839584}
```

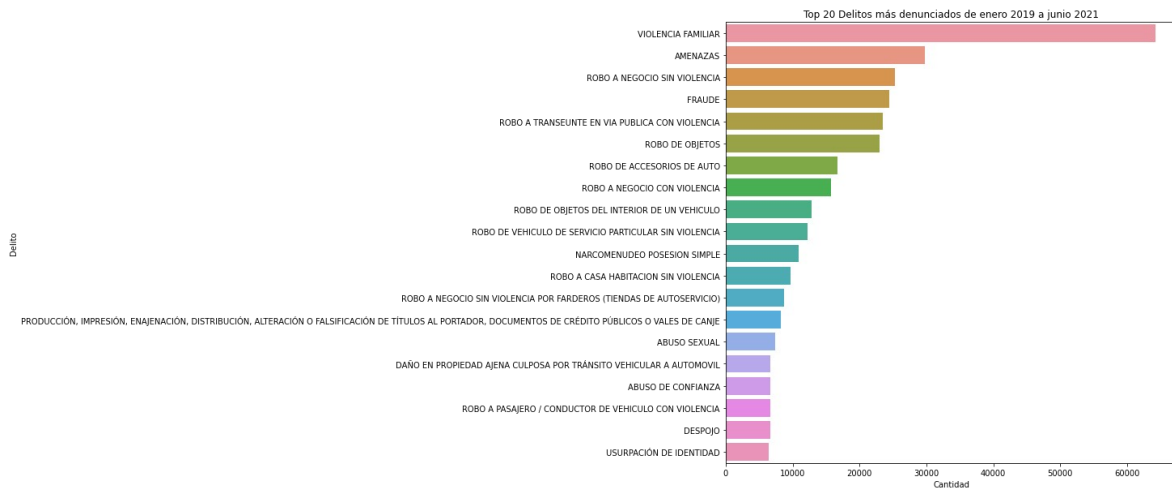
## 7. Transformación y filtración de datos

Para esta parte nos enfocamos en observar el tipo de datos e información disponible y junto con nuestra creatividad imaginar las combinaciones posibles de generar información o de hacernos preguntas durante el proceso sobre las tendencias o hallazgos de información que no era evidente al principio por estar oculta entre la inmensidad de datos.

Así, nuestra búsqueda de conocimiento comenzó guiada por las preguntas que nos planteamos al inicio. Para esto, creamos por ejemplo una nueva columna llamada IntervaloTiempo\_denuncia que cuantificaba la diferencia de tiempo entre el momento que se hizo la denuncia y el momento en que ocurrió el delito, para conocer después el tiempo promedio que las personas tardaban en hacer una denuncia.

Por otra parte, se hicieron agrupamientos para conocer los delitos más denunciados, las alcaldías con más ocurrencias de delitos, las edades más frecuentes de las víctimas, otras estadísticas básicas de las víctimas como se muestran a continuación:

Delitos más comunes por los que se presentó una denuncia.

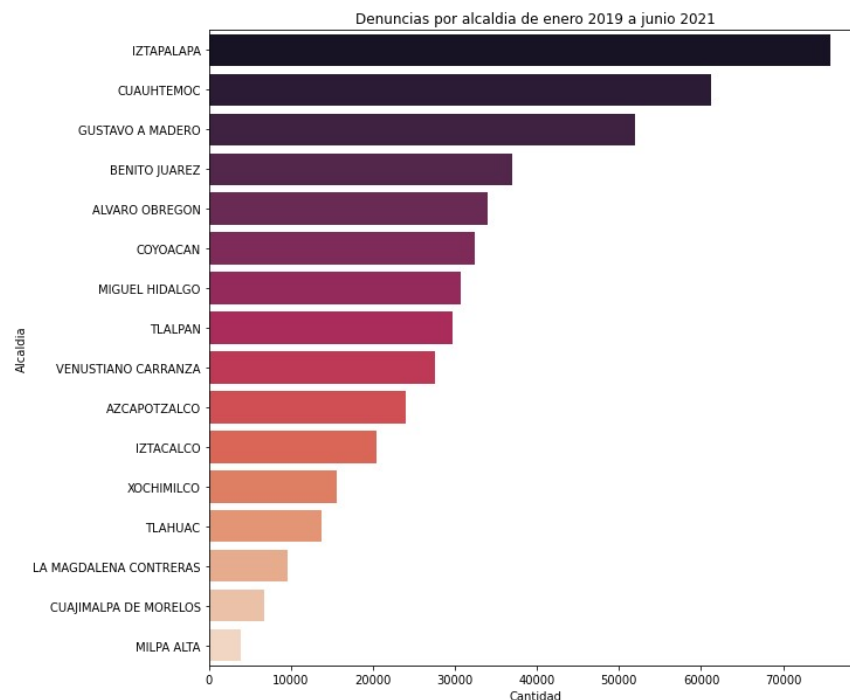


Tiempo transcurrido después del delito para denunciar: aproximadamente 33 días.

```
# Y después se calcula el tiempo promedio de denuncia sin considerar intervalos de tiempo negativos
df_base[df_base["IntervaloTiempo_denuncia"] > pd.Timedelta(0)][["IntervaloTiempo_denuncia"].mean()

Timedelta('33 days 20:38:14.729374029')
```

Número de denuncias hechas por delitos cometidos en cada alcaldía.



Edad más frecuentes de las víctimas y cantidad de víctimas hombres y mujeres.

Cantidad	
Edad	
30.0	9871
28.0	9694
29.0	9587
35.0	9526
27.0	9436
32.0	9301
31.0	9287
33.0	9214
26.0	9136
34.0	8977

Cantidad	
Sexo	
Femenino	170757
Masculino	191047

Adicionalmente, al explorar los datos surgieron cuestiones como las siguientes.

Resulta que la Fiscalía General de la Justicia maneja una clasificación de 16 categorías para los delitos y así abrir una carpeta de investigación. Hallamos que muchas de las denuncias están catalogadas por la FGJ como delitos de bajo impacto, y al explorar que delitos estaban contenidos ahí, sorpresivamente hay delitos como

- Abuso sexual
- Homicidio culposo
- Tortura
- Tráfico de infantes

entre otros por mencionar. Esto abre muchas preguntas sobre la precisión y efectividad de esas clasificaciones de la FGJ.

Cantidad	
Categoria	
DELITO DE BAJO IMPACTO	373820
ROBO A TRANSEUNTE EN VIA PÚBLICA CON Y SIN VIOLENCIA	27467
ROBO DE VEHICULO CON Y SIN VIOLENCIA	23832
ROBO A NEGOCIO CON VIOLENCIA	17702
ROBO A REPARTIDOR CON Y SIN VIOLENCIA	7744
HECHO NO DELICTIVO	5939
ROBO A PASAJERO A BORDO DEL METRO CON Y SIN VIOLENCIA	3823
HOMICIDIO DOLOSO	3634
VIOLACIÓN	3206
ROBO A PASAJERO A BORDO DE MICROBUS CON Y SIN VIOLENCIA	1833
LESIONES DOLOSAS POR DISPARO DE ARMA DE FUEGO	1740
ROBO A CASA HABITACIÓN CON VIOLENCIA	1389
ROBO A CUENTAHABIENTE SALIENDO DEL CAJERO CON VIOLENCIA	1092
ROBO A PASAJERO A BORDO DE TAXI CON VIOLENCIA	626
ROBO A TRANSPORTISTA CON Y SIN VIOLENCIA	264
SECUESTRO	84

Por otra parte, pudimos cuantificar cuantas de las denuncias presentadas en las CDMX pertenecen realmente a delitos cometidos en la CDMX. Esto porque existe un campo (Competencia) que proporciona información sobre las denuncias que realmente fueron consideradas hechos delictivos ocurridos en la CDMX, así calculamos un porcentaje que resulto ser de 98.59% de lo que representan esas denuncias en toda nuestra base de datos.

Finalmente, dejamos un conjunto de datos filtrado del tipo de delitos comentados al inicio que involucra a personas morales, para las cuales no existe dato de edad ni sexo, pues parece un conjunto de datos interesante para relacionar otro tipo de variables no exploradas.

## 8. Notebook y planes a futuro

Para concluir, todo nuestro desarrollo de procesamiento puede ser consultado con detalle en el siguiente enlace donde se encuentra nuestro notebook de Google Colab:

[https://colab.research.google.com/drive/1efUj\\_3yX5szHQMhd8Bnn3js6RfcEL-O1?authuser=1#scrollTo=RICmfYIWSqmJ](https://colab.research.google.com/drive/1efUj_3yX5szHQMhd8Bnn3js6RfcEL-O1?authuser=1#scrollTo=RICmfYIWSqmJ)

Cabe mencionar que aún tenemos planes para sacar todo el provecho posible a este conjunto de datos, pues debido a la gran cantidad de datos y variables involucradas no hemos tenido toda la oportunidad probar todas nuestras combinaciones de cruce de datos. Esperamos estar en los siguientes módulos y no sólo generar conclusiones más interesantes que involucren análisis más profundo sino también ser capaces de usar técnicas de predicción que en este tipo de problemas pueden hacer la diferencia entre la seguridad y una tragedia para los ciudadanos de la CDMX.

En concreto, algunos de nuestros planes para el siguiente módulo son:

- Cuantificar algunos aspectos que permitan la predicción de delitos, como la hora más probable en la que pueden suceder, la zona donde se podría requerir más presencia policíaca, y en qué época del año según el tipo de delito.
- Un análisis más profundo donde se puedan obtener conclusiones después de cruzar información como edad, sexo y calidad jurídica de la víctima.

- Creación de algún mapa de zonas de riesgos por tipo de delito e incluso hora de ocurrencia. Información útil para la población y sistemas de seguridad pública. Los datos de latitud y longitud aún no han sido explotados.
- Generación de algún sistema de clasificación de delitos mejorado y quizá más justo para la FGJ de modo que haya más concordancia entre los delitos agrupados en cada categoría, quizá con un algoritmo que reconozca palabras similares y por tanto delitos más similares en una misma clase.

## Problemas durante el desarrollo

En realidad, los problemas que se tuvieron durante el desarrollo fueron mínimos. Por ejemplo, un problema inicial era que el código fuese reproducible desde cualquier computador. Sobre todo, porque no era posible cargar la base de datos en algún lugar fijo donde se pudiera acceder fácilmente mediante su URL sin cargar manualmente los datos porque la base de datos tiene un tamaño superior a los 100 MB. Finalmente pudimos solucionarlo al indicar directamente el URL de descarga de la base de datos de la plataforma del C5 de la CDMX. Todo el equipo se esforzó en la solución de cada pequeño problema que surgió. Por esa razón se creemos que los problemas fueron mínimos.

Es interesante mencionar que la selección del tema fue algo complicado. Cada miembro del equipo detectó áreas de oportunidad en tantos lugares. Casi todos tenían posibles soluciones con el uso de herramientas que proporciona el Data Science.

Dentro de los problemas menores que se obtuvieron en el desarrollo del proyecto, fueron cosas simples, pero durante las clases en vivo se pudieron solucionar los problemas, puesto que eran parte de los temas que se desarrollaron en cada una de las clases. O incluso regresar a nuestras notas o retos llevados a cabo para modificarlos de acuerdo a nuestras necesidades o consultar la documentación.

Se decidió darle solución a las preguntas que se eligieron y esto fue un verdadero reto de creatividad para poder aprovechar el tipo y cantidad de datos disponibles. Finalmente, se concluyó con el proyecto satisfactoriamente. Los problemas fueron resueltos y se convirtieron en aprendizaje.

## REFERENCIAS

1. <https://www.c5.cdmx.gob.mx/dependencia/acerca-de>
2. <https://www.mexicoevalua.org/homicidioscdmx/documentos/5013HomicidiosCDMX.pdf>
3. <https://www.evalua.cdmx.gob.mx/storage/app/media/DIES20/8-cap-7-inseguridad-y-violencia-publicas.pdf>
4. <https://politica.expansion.mx/cdmx/2021/03/24/extorsion-y-fraude-los-delitos-que-crecieron-en-la-cdmx-durante-la-pandemia>
5. [www.elsoldemexico.com.mx](http://www.elsoldemexico.com.mx)
6. <https://politica.expansion.mx>
7. [www.animalpolitico.com](http://www.animalpolitico.com)
8. <https://www.fgjcdmx.gob.mx/procuraduria/estadisticas-delictivas>
9. <https://politica.expansion.mx/mexico/2019/02/04/una-a-una-estas-son-las-9-estrategias-especificas-de-amlo-contr-la-violencia>
10. <https://www.c5.cdmx.gob.mx/dependencia/acerca-de/el-c5-de-la-cdmx>
11. <https://expansion.mx/tecnologia/2019/06/05/camaras-con-ia-de-la-cdmx-ya-podran-evitar-delitos>