

## Sección A – Datos de crimen en la CDMX

Esta sección plantea el análisis de datos sobre el crimen en la CDMX, dando información sobre un periodo de tiempo del 2016 al 2023 y varias clasificaciones como es la fecha, hora, delito, fiscalía y demás, con la ayuda de Python y librerías de apoyo como pandas se procede a contestar las siguientes preguntas:

### **1. ¿Qué pruebas identificarías para asegurar la calidad de estos datos? No es necesario hacerlas, sólo describe la prueba y lo que te dice cada una.**

Como primer paso y una aproximación al análisis de estos datos, identificar si los datos son verídicos, si están completos y si estos no presentan alteraciones que puedan resultar en una alteración a los resultados, siempre es importante por lo tanto se pueden seguir las siguientes pruebas.

- **Verificación de integridad y completitud:** Se asegura de que todos los campos obligatorios estén llenos y no haya valores faltantes. Esto garantiza que cada registro tenga la información necesaria para su comprensión y uso.
- **Validación de formato:** Comprobar si los datos están en el formato correcto, por ejemplo, fechas en el formato adecuado.
- **Detección de duplicados:** Identifica registros duplicados en función de ciertos campos clave, lo que evita redundancias y errores.
- **Consistencia:** Verifica que los valores en los campos relacionados sean coherentes entre sí. Por ejemplo, si hay un campo que indica el tipo de delito, se debe verificar que coincida con la descripción del delito en otro campo.
- **Control de valores atípicos:** Detecta valores que son significativamente diferentes del resto, lo que podría indicar errores o datos incorrectos.
- **Pruebas de rango y límites:** Verifica que los valores numéricos estén dentro de rangos válidos y límites establecidos. Por ejemplo, edades negativas o muy altas podrían indicar errores.
- **Normalización y estandarización:** Asegura que los datos estén en el mismo formato y sigan las mismas convenciones. Esto podría aplicarse a campos como direcciones, nombres de calles, etc.

## 2. Identifica los delitos que van al alza y a la baja en la CDMX (ten cuidado con los delitos con pocas ocurrencias).

Para esta pregunta se escribe un código en Python el cual se anexa en el correo.  
Donde como primera observación se tiene que en la columna de "ao\_hechos" existen valores atípicos como por ejemplo 222.0 o años anteriores al 2016, por lo que son eliminados  
Y los delitos que van a alza o a la baja son representados en las variables

- delitos\_al\_alza
- delitos\_a\_la\_baja

Delitos al alza:

|      | ao_hechos | delito  | ocurrencias | cambio por año |
|------|-----------|---|-------------|----------------|
| 272  | 2017.0    | ABANDONO DE PERSONA                               | 671         | 8.051530       |
| 274  | 2017.0    | ABUSO DE AUTORIDAD                                | 1291        | 3.362690       |
| 275  | 2017.0    | ABUSO DE AUTORIDAD Y USO ILEGAL DE LA FUERZA P... | 15          | 275.000000     |
| 276  | 2017.0    | ABUSO DE CONFIANZA                                | 3864        | 18.056829      |
| 278  | 2017.0    | ACOSO SEXUAL                                      | 246         | 18.840580      |
| ...  | ...       | ...   | ...         | ...            |
| 1964 | 2022.0    | VARIACION DE NOMBRE O DOMICILIO                   | 20          | 53.846154      |
| 1967 | 2022.0    | VIOLACION EQUIPARADA                              | 1339        | 80.945946      |
| 1970 | 2022.0    | VIOLACION TUMULTUARIA                             | 40          | 25.000000      |
| 1971 | 2022.0    | VIOLACION TUMULTUARIA EQUIPARADA                  | 18          | 1700.000000    |
| 1973 | 2022.0    | VIOLENCIA FAMILIAR                                | 36977       | 5.191739       |

Delitos a la baja:

|      | ao_hechos | delito  | ocurrencias | cambio por año |
|------|-----------|---|-------------|----------------|
| 273  | 2017.0    | ABORTO  | 82          | -28.695652     |
| 277  | 2017.0    | ABUSO SEXUAL                                      | 1811        | -7.318321      |
| 281  | 2017.0    | AMENAZAS  | 9532        | -4.479407      |
| 282  | 2017.0    | ATAQUE A LAS VIAS DE COMUNICACION (DAÑO A VIAS... | 22          | -77.777778     |
| 290  | 2017.0    | COHECHO   | 238         | -4.032258      |
| ...  | ...       | ...   | ...         | ...            |
| 2231 | 2023.0    | VIOLACION   | 200         | -73.684211     |
| 2233 | 2023.0    | VIOLACION EQUIPARADA                              | 408         | -69.529500     |
| 2235 | 2023.0    | VIOLACION TUMULTUARIA                             | 12          | -70.000000     |
| 2236 | 2023.0    | VIOLACION TUMULTUARIA EQUIPARADA                  | 13          | -27.777778     |
| 2237 | 2023.0    | VIOLENCIA FAMILIAR                                | 11712       | -68.326257     |

**3. ¿Cuál es la alcaldía que más delitos tiene y cuál es la que menos? ¿Por qué crees que sea esto?**

Para resolver esta pregunta, se agrupan los valores por alcaldía y se cuenta quien contiene mas delitos, sin embargo, nuevamente se encuentra que estos datos presentan errores al considerar un valor de columna de "alcaldía\_hechos" la cadena "CDMX" que podemos alucir que no se sabe exactamente en que alcaldía se habrá cometido el delito, eliminando este valor podemos encontrar entonces que la alcaldía con mas delitos es CUAUHEMOC y la que menos delitos tiene es MILPA ALTA

|    | alcaldia_hechos        | total_delitos |
|----|------------------------|---------------|
| 6  | CUAUHEMOC              | 251561        |
| 10 | IZTAPALAPA             | 246274        |
| 8  | GUSTAVO A. MADERO      | 167440        |
| 2  | BENITO JUAREZ          | 131717        |
| 4  | COYOACAN               | 112389        |
| 0  | ALVARO OBREGON         | 112352        |
| 12 | MIGUEL HIDALGO         | 104584        |
| 15 | TLALPAN                | 98136         |
| 16 | VENUSTIANO CARRANZA    | 94983         |
| 1  | AZCAPOTZALCO           | 80713         |
| 9  | IZTACALCO              | 71013         |
| 17 | XOCHIMILCO             | 52900         |
| 14 | TLAHUAC                | 44245         |
| 11 | LA MAGDALENA CONTRERAS | 28441         |
| 5  | CUAJIMALPA DE MORELOS  | 25454         |
| 7  | FUERA DE CDMX          | 16969         |
| 13 | MILPA ALTA             | 13601         |
| 3  | CDMX                   | 3683          |

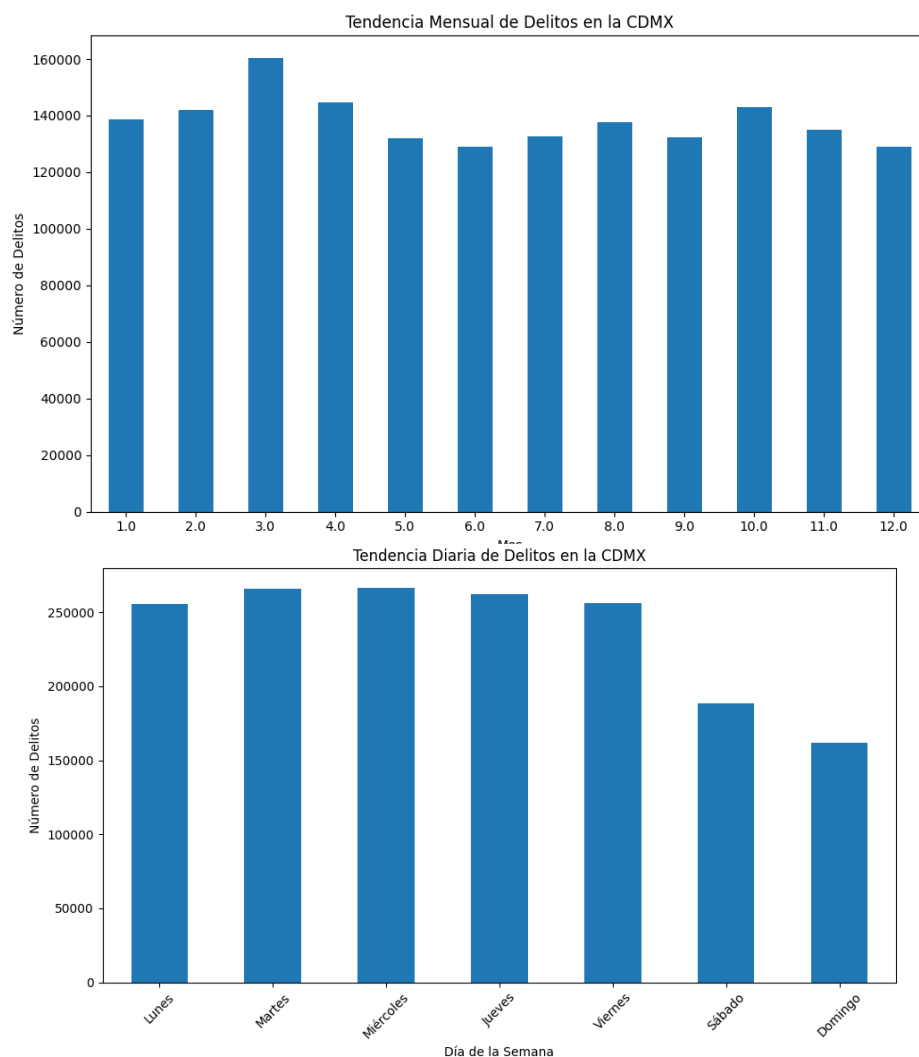
El porque surge estos valores puede estar relacionado a varias características como la densidad poblacional, la actividad económica, la presencia policial, la ubicación geográfica,

la infraestructura, entre otros. Es importante tener en cuenta que el análisis de los datos de los delitos es solo un punto de partida y que se requiere un análisis más profundo y contextualizado para comprender completamente las razones detrás de estas diferencias.

#### 4. ¿Existe alguna tendencia estacional en la ocurrencia de delitos (mes, semana, día de la semana, quincenas) en la CDMX? ¿A qué crees que se deba?

Para esta pregunta podemos analizar en diferentes tiempos la respuesta, por ejemplo, por mes y por día de la semana, donde los resultados del mes presentan cierta tendencia para los primeros y últimos meses del año y una baja en los delitos para los meses de mayo, junio y julio, sin embargo, esta tendencia es poco indicativa ya que la cantidad no decrece tan intensamente.

Para el análisis por día tenemos que, los días entre semana tienen una mayor actividad a los fines de semana, esto puede ser debido a varios factores como la baja de movilidad de las personas en días de descanso.



5. ¿Cuáles son los delitos que más caracterizan a cada alcaldía? Es decir, delitos que suceden con mayor frecuencia en una alcaldía y con menor frecuencia en las demás.

En este caso podemos contar la cantidad de delitos y su tipo por alcaldía, donde con observar la cantidad de delitos podremos saber los que se presentan mas y los que son poco recurrentes.

| alcaldia_hechos        |                        | delito                                   | ocurrencias |
|------------------------|------------------------|--|-------------|
| alcaldia_hechos        |                        |  |             |
| ALVARO OBREGON         | ALVARO OBREGON         | VIOLENCIA FAMILIAR                       | 14554       |
| AZCAPOTZALCO           | AZCAPOTZALCO           | VIOLENCIA FAMILIAR                       | 9380        |
| BENITO JUAREZ          | BENITO JUAREZ          | FRAUDE                                   | 13860       |
| CDMX                   | CDMX                   | FRAUDE                                   | 466         |
| COYOACAN               | COYOACAN               | VIOLENCIA FAMILIAR                       | 13057       |
| CUAJIMALPA DE MORELOS  | CUAJIMALPA DE MORELOS  | VIOLENCIA FAMILIAR                       | 3627        |
| CUAUHTEMOC             | CUAUHTEMOC             | FRAUDE                                   | 25249       |
| FUERA DE CDMX          | FUERA DE CDMX          | LESIONES INTENCIONALES POR ARMA DE FUEGO | 1571        |
| GUSTAVO A. MADERO      | GUSTAVO A. MADERO      | VIOLENCIA FAMILIAR                       | 23617       |
| IZTACALCO              | IZTACALCO              | VIOLENCIA FAMILIAR                       | 11856       |
| IZTAPALAPA             | IZTAPALAPA             | VIOLENCIA FAMILIAR                       | 34700       |
| LA MAGDALENA CONTRERAS | LA MAGDALENA CONTRERAS | VIOLENCIA FAMILIAR                       | 7047        |
| MIGUEL HIDALGO         | MIGUEL HIDALGO         | FRAUDE                                   | 9945        |
| MILPA ALTA             | MILPA ALTA             | VIOLENCIA FAMILIAR                       | 3435        |
| TLAHUAC                | TLAHUAC                | VIOLENCIA FAMILIAR                       | 8952        |
| TLALPAN                | TLALPAN                | VIOLENCIA FAMILIAR                       | 13652       |
| VENUSTIANO CARRANZA    | VENUSTIANO CARRANZA    | VIOLENCIA FAMILIAR                       | 10647       |
| XOCHIMILCO             | XOCHIMILCO             | VIOLENCIA FAMILIAR                       | 9628        |

| alcaldia_hechos        |                        | delito                                     | ocurrencias |
|------------------------|------------------------|--|-------------|
| alcaldia_hechos        |                        |  |             |
| ALVARO OBREGON         | ALVARO OBREGON         | BIGAMIA                                    | 1           |
| AZCAPOTZALCO           | AZCAPOTZALCO           | BIGAMIA                                    | 1           |
| BENITO JUAREZ          | BENITO JUAREZ          | ASOCIACION DELICTUOSA                      | 1           |
| CDMX                   | CDMX                   | CAMBIO DE USO DE SUELO                     | 1           |
| COYOACAN               | COYOACAN               | COALICIÓN DE SERVIDORES PÚBLICOS           | 1           |
| CUAJIMALPA DE MORELOS  | CUAJIMALPA DE MORELOS  | BIGAMIA                                    | 1           |
| CUAUHTEMOC             | CUAUHTEMOC             | CALUMNIAS                                  | 1           |
| FUERA DE CDMX          | FUERA DE CDMX          | CONTRA FUNCIONARIOS PUBLICOS               | 1           |
| GUSTAVO A. MADERO      | GUSTAVO A. MADERO      | ATAQUES A LA PAZ PUBLICA                   | 1           |
| IZTACALCO              | IZTACALCO              | COALICIÓN DE SERVIDORES PÚBLICOS           | 1           |
| IZTAPALAPA             | IZTAPALAPA             | COALICIÓN DE SERVIDORES PÚBLICOS           | 1           |
| LA MAGDALENA CONTRERAS | LA MAGDALENA CONTRERAS | BIGAMIA                                    | 1           |
| MIGUEL HIDALGO         | MIGUEL HIDALGO         | CONCUSION                                  | 1           |
| MILPA ALTA             | MILPA ALTA             | ACOSO SEXUAL AGRAVADO EN CONTRA DE MENORES | 1           |
| TLAHUAC                | TLAHUAC                | CONCUSION                                  | 1           |
| TLALPAN                | TLALPAN                | BIGAMIA                                    | 1           |
| VENUSTIANO CARRANZA    | VENUSTIANO CARRANZA    | ATAQUES A LA PAZ PUBLICA                   | 1           |
| XOCHIMILCO             | XOCHIMILCO             | ATAQUES A LA PAZ PUBLICA                   | 1           |

**6. Diseña un indicador que mida el nivel de “inseguridad”. Génalo al nivel de desagregación que te parezca más adecuado (ej. manzana, calle, AGEB, etc.). Analiza los resultados ¿Encontraste algún patrón interesante? ¿Qué decisiones se podrían tomar con el indicador?**

Este punto puede ser abordado bajo distintos tipos de enfoques, por ejemplo, saber el número de delitos por cantidad de habitantes por alcaldía, la hora donde se comenten mas delitos, los delitos de mas alto riesgo para la población, delitos resueltos favorablemente, y muchas otras opciones, de las cuales necesitaríamos mas datos de los proporcionados por el dataset.

Por lo tanto, para este ejemplo se genera un indicador de inseguridad basado en la hora y alcaldía donde se presentan más delitos.

Con estos datos podemos realizar ciertas recomendaciones a las autoridades y población en general.

- Reforzar la presencia policial en las horas del día y las alcaldías con los indicadores de inseguridad más altos.
- Implementar patrullajes preventivos y operativos en las áreas y horas identificadas como más inseguras.
- Realizar campañas de concientización y prevención dirigidas a la comunidad en las áreas y horas con mayores indicadores de inseguridad.
- Evaluar la efectividad de las estrategias de seguridad implementadas a lo largo del tiempo mediante un seguimiento del indicador y ajustar las acciones en consecuencia.

|     | alcaldia_hechos | hora_1 | ocurrencias | indicador_inseguridad |
|-----|-----------------|--------|-------------|-----------------------|
| 0   | ALVARO OBREGON  | 0      | 2436        | 6.073904              |
| 1   | ALVARO OBREGON  | 1      | 2244        | 5.595173              |
| 2   | ALVARO OBREGON  | 2      | 2096        | 5.226151              |
| 3   | ALVARO OBREGON  | 3      | 1961        | 4.889543              |
| 4   | ALVARO OBREGON  | 4      | 1531        | 3.817384              |
| ... | ...             | ...    | ...         | ...                   |
| 427 | XOCHIMILCO      | 19     | 2560        | 6.383085              |
| 428 | XOCHIMILCO      | 20     | 2791        | 6.959058              |
| 429 | XOCHIMILCO      | 21     | 2487        | 6.201067              |
| 430 | XOCHIMILCO      | 22     | 2290        | 5.709869              |
| 431 | XOCHIMILCO      | 23     | 1814        | 4.523014              |

En generar un análisis de estos datos nos puede indicar que tan insegura es la CDMX, punto de delitos altos, bajos y horarios mas destacados, información que analizada y presentada a las autoridades correspondientes puede ayudar a tomar medidas preventivas y reducir estos delitos.

## Sección B – La Michoacana

Para esta sección se presenta un problema relacionado a la paletería Michoacana donde se busca la optimización de costos para su venta de forma de suscripción de paletas.

Los puntos a notar para poder resolver el problema son:

- Los puntos de venta son surtidos mediante camiones a lo máximo una vez por día, donde la cantidad de paletas no está limitada
- El surtir un punto de venta tiene un costo de 100 pesos sin importar la cantidad de paletas que se surtan
- Cada punto de venta contiene una cantidad de paletas por espacios únicos y cada uno de los espacios ocupados genera un costo de 1 peso por día
- Se busca que ningún punto de venta se quede sin producto cuando un usuario busque una paleta

De acuerdo al problema para poder dar una solución tenemos la siguiente información

- Cantidad de paletas retiradas en cada punto de venta por día con una historia de años
- Cantidad de paletas disponibles por punto de venta con una historia de 5 años
- Cantidad máxima de paletas por punto de venta

El jefe de operaciones, Luis, debe indicar diariamente si un punto de venta debe ser visitado cada día además de definir la cantidad de paletas a dejar por punto de venta.

Los parámetros óptimos para poder operar son:

- Mantener los costos de operación más bajos posibles
  - Menor cantidad de envíos de paletas
  - Menor cantidad de paletas no retiradas por punto de venta

### Solución:

De acuerdo a los datos presentados anteriormente se puede diseñar un sistema auxiliar para Luis el cual emita una recomendación del número de paletas a enviar a cada punto de venta y si un punto de venta debe ser surtido con una cantidad específica de paletas al iniciar sus operaciones.

Este problema al tener una historia de 5 años para los cambios que sufren los puntos de venta podemos tratarlo mediante un sistema de regresión el cual tendrá múltiples salidas para cada uno de los puntos de venta:

- Predicción de cantidad de paletas retiradas del punto de venta en el día
- Visitar el punto de venta
- Cantidad de paletas a dejar en el punto de venta

Estos dependerán de factores como son:

- Cantidad máxima de paletas en el punto de venta
- Cantidad de paletas consumidas en el día para el punto de venta

- Costo de visitar el punto de venta
- Costo de almacenar una paleta por 24 horas

Para diseñar la solución se tendría que verificar los datos, de los 5 años anteriores como por ejemplo datos anómalos (por ejemplo, contener datos del tipo carácter en datos numéricos), datos vacíos, datos duplicados y demás.

Posteriormente los datos deben ser tratados para facilitar el proceso de modelos inteligentes como es la categorización de los datos, estandarización de datos numéricos, y de manera personal incluir algunos datos que generalmente no se toman en cuenta cuando se recolectan este tipo de historias como es:

- Indicar el día de la semana
- Indicar el día del mes
- Indicar si es día festivo
- Indicar si es fin de semana
- Investigar e indicar el clima de la zona
- Indicar el número de usuarios cercanos en un rango
- Investigar e indicar la recurrencia de los usuarios para el punto de venta

Estos parámetros pueden servir para predecir los comportamientos en el punto de venta futuros, facilitando la tarea de minimizar los costos, además, pueden ayudar a mejorar el proceso y futuras mejoras al sistema.

Después de poder preprocesar los datos antes mencionados, es necesario entrenar y probar algunos modelos que permiten trabajar con este tipo de series como son los modelos de machine learning para regresión, de los cuales podemos elegir como ejemplos:

- Máquinas de vector soporte
- Bosques aleatorios
- Redes neuronales (perceptrones)
- Regresión Lineal

Para los cuales se debe tener en cuenta las limitaciones que presentan, como son el overfitting para las redes neuronales, parámetro C para las máquinas de vector soporte, número máximo de árboles, máxima profundidad para los bosques aleatorios y demás parámetros que pueden afectar en gran parte a los resultados de estos sistemas

Cabe destacar que generalmente cuando se han seleccionado una cantidad de modelos viables es una buena práctica realizar un benchmark sobre los hiperparámetros de los modelos para definir la arquitectura correcta.

Todas estas arquitecturas deben ser analizadas en sus respuestas mediante métricas que nos indiquen si son viables para su uso, en este caso se ha definido que la indisponibilidad de los puntos de venta no debe ser menor a 2%.



En este punto podemos tener un modelo que permita predecir la cantidad de paletas a retirar del punto de venta, y con la ayuda de los datos de cantidad máxima de paletas por punto y cantidad disponible de paletas podremos saber si es necesario un envío de paletas y la cantidad recomendada de paletas.

Un punto a notar, es que, si podemos además verificar a futuro el consumo de paletas para el punto de venta, donde si para los próximos días el consumo de paletas es menos a 100 es recomendable si la capacidad máxima lo permite mandar esas paletas y así ahorrar viajes futuros.

En este proyecto, el despliegue de ese sistema debe ser escalado y basado en pruebas de segmentación, ya que actualmente se tiene a un humano (Luis) que permite realizar la tarea y, por lo tanto, no es necesario su despliegue a gran escala por lo que se recomienda tomar una muestra aleatoria de al menos 500 puntos de venta y poder comparar las recomendaciones del humano y del modelo en un periodo de prueba.

El sistema al ser de pocas variables no es necesario su despliegue en un sistema de cómputo paralelo o la nube, ya que una computadora de servicio medio lo puede realizar al inicio de horario de trabajo.

La forma mas sencilla de evaluar el desempeño del sistema es verificar que las recomendaciones realizadas permitan disminuir el costo de envío de producto o de almacenamiento. El cual es verificable dada la historia de cada punto de venta