

Proyecto Final del Curso: Análisis de Datos con Python

Maestría en Matemáticas Aplicadas

Título del proyecto

Exploración y Predicción de la Incidencia Delictiva en la CDMX (2015-2025)

Objetivo general

Aplicar herramientas de Python para explorar, visualizar y modelar patrones de delitos registrados en la Ciudad de México entre 2015 y 2025, enfocándose en la evolución temporal, distribución geográfica por alcaldía y predicción futura de incidentes delictivos seleccionados.

El conjunto de datos contiene registros mensuales de incidencia delictiva en la CDMX, desagregados por año, municipio (alcaldía), tipo de delito, subtipo y modalidad. Este dataset es apto estudiar la evolución de la violencia urbana y evaluar el impacto de eventos críticos como la pandemia por COVID-19.

Actividades a desarrollar

1. Carga y limpieza del conjunto de datos

- Reorganizar el dataset al formato largo: una fila por mes y año.
- Convertir las columnas de los meses a valores numéricos.

2. Análisis Exploratorio de Datos

- Identificar los delitos más frecuentes por año y por alcaldía.
- Análisis de tendencias anuales y mensuales.
- Comparación entre modalidades del mismo subtipo de delito.

3. Visualización geográfica

- Crear mapas de calor con folium.
- Comparar distribución espacial de diferentes tipos de delitos.

4. Modelado predictivo (opcional)

- Elegir un delito (por ejemplo, homicidio doloso) y predecir su incidencia mensual.

- Aplicar modelos de series de tiempo como regresión lineal, ARIMA o Prophet.
- Evaluar con métricas de error como MAE y RMSE.

5. Análisis del impacto de la pandemia

- Comparar los niveles de incidencia antes (2015–2019), durante (2020–2021) y después (2022–2025).
- Visualizar caídas abruptas o aumentos significativos.

6. Discusión e interpretación

- Destacar patrones temporales o espaciales interesantes.
- Proponer explicaciones para variaciones detectadas.
- Discutir limitaciones del dataset.

Entregables

- **Notebook en Python.**
- **Informe técnico (PDF)** con secciones de introducción, metodología, resultados, visualizaciones y conclusiones.
- **Presentación oral** (20 minutos) donde se expongan los hallazgos principales.

Criterios de evaluación

Criterio	Ponderación
Preprocesamiento y limpieza de datos	15%
Visualizaciones y análisis exploratorio	25%
Modelado y evaluación del modelo	25%
Interpretación de resultados	20%
Calidad del código y reproducibilidad	15%

Recomendaciones Finales

- Las visualizaciones deben ser claras y apoyar la interpretación.
- El informe debe incluir referencias si se utilizan fuentes externas.