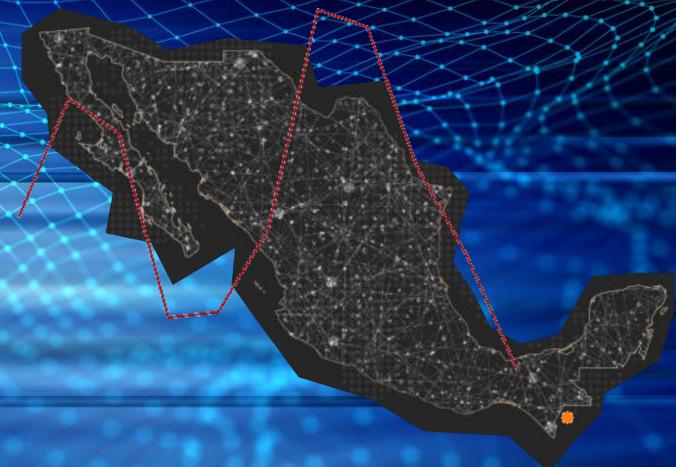


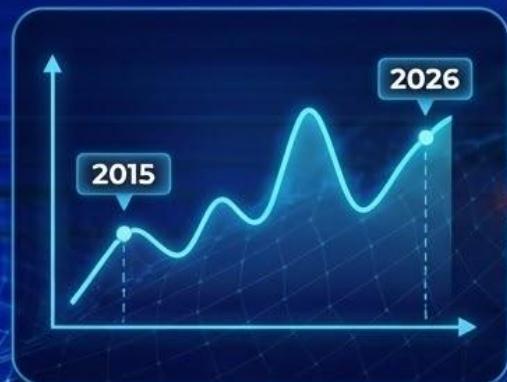
Análisis de Delitos en México (2015-2025)

Series de tiempo y clustering de datos



Proyecto Final - Análisis de Datos
Ruben Tamez

Objetivo del Proyecto



Analizar la evolución
de los delitos entre
2015 y 2026.



Predecir el
comportamiento de
los delitos en 2026.



Clasificar los estados de
México según su nivel
de peligrosidad en 2026.

Descarga y Creación de la Base de Datos



Incidencia delictiva (metodología vigente)

- 2015 -2025 (Fuero Común - Delitos), Incidencia delictiva estatal
- 2015 - 2025 (Fuero Común - Víctimas), Incidencia delictiva estatal
- 2015 - 2025 (Fuero Común - Delitoss), Incidencia delictiva estatal
- 2015 - 2025 (Fuero Común - Delitos), Incidencia delictiva municipal
- Diccionarios de datos incidencia delictiva metodología vigente



Incidencia delictiva (metodología anterior)

- 1997 - 2017 (Fuero común) - Delitos). Incidencia delictiva estatal
- 2014 - 2017 (Fuero común - Víctimas). Incidencia delictiva estatal
- 2011 - 2017 (Fuero común- Delitos). Incidencia delictiva municipal
- Diccionarios de datos incidencia delictiva metodología anterior

Proceso de Carga de Datos

```
1 # Lectura de datos y importación de librerías
2 import pandas as pd
3 import os
4 os.chdir('C:/Users/Rubén/Documents/DATA
ANALIST/git/TareasEBACmx/Proyecto Final/')
5
6 df = pd.read_csv('Estatal-Delitos-2015-2025
_oct2025.csv', encoding='latin-1')
```

Limpieza y Transformación de los Datos



Filtrado de Datos en Nuevo León

```
# Filtrar por entidad "Nuevo León"  
df_nuevo_leon = df[df['Entidad'] == 'Nuevo León']
```



```
# Filtrado por robo de automóviles en Nuevo León  
df_nl_robo_autos = df_nuevo_leon[df_nuevo_leon['Modalidad'].isin([  
    'Robo de coche de 4 ruedas Con violencia',  
    'Robo de coche de 4 ruedas Sin violencia',  
    'Robo de motocicleta Con violencia',  
    'Robo de motocicleta Sin violencia'])]
```



Tratamiento de Valores Nulos

```
# Tratamiento de nulos (No tenemos nulos en este caso)  
df_nl_robo_autos_rango.isnull().sum()
```

Resultado: 0 nulos encontrados

Análisis de Series de Tiempo

Contexto y Metodología



Objetivo:

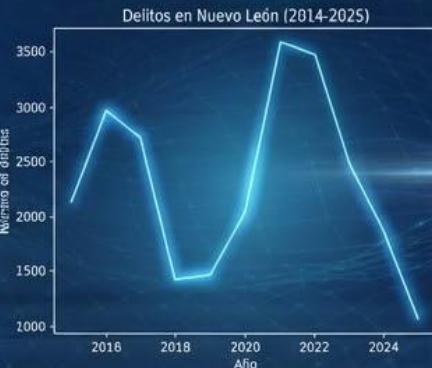
Identificar patrones y tendencias en los delitos de 2015 a 2025 y predecir los delitos en 2026.



Selección del Municipio:

A fecha de hoy el dataset de crímenes de México cambió, omitiendo la columna de municipios, por lo que tomare mi entidad de residencia que es "Nuevo Leon"

Análisis y Predicciones



Predicciones:

Se utilizó un modelo ARIMA para analizar la evolución de los delitos entre 2015 y 2025 en Nuevo León. El modelo permitió identificar la tendencia histórica y generar una predicción para 2026, la cual sugiere que el nivel de delitos continuará acorde al comportamiento observado en años anteriores.

Clasificación de Estados por Peligrosidad (Clustering)

Objetivo y Preparación de Datos



Objetivo: Clasificar los estados según su peligrosidad en 2021



Preparación de Datos:

```
# Filtrar año 2021
df_2021 = df[df['Año'] == 2021]

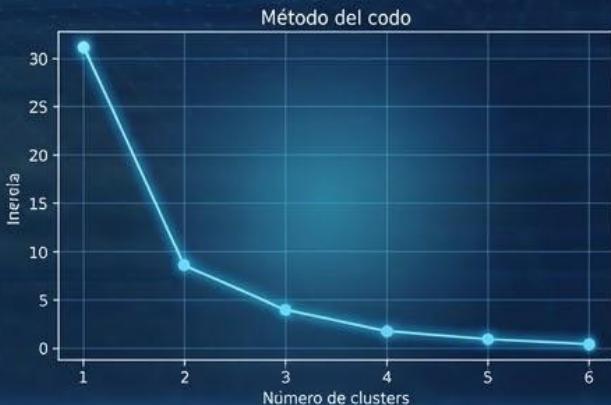
# Calcular total de delitos por estado
meses = ['Enero', 'Febrero', ..., 'Diciembre']
df_2021['Total_delitos'] = df_2021[meses].sum(axis=1)
df_estado = df_2021.groupby('Entidad')['Total_delitos'].sum()

# Normalización de datos
from sklearn.preprocessing import StandardScaler
scaler = StandardScaler()
X_scaled = scaler.fit_transform(df_estado[['Total_delitos']])
```

Método del Codo (K-Means)

```
# Elegir número de clusters (Método del codo)
from sklearn.cluster import KMeans
import matplotlib.pyplot as plt

inercia = []
for k in range(1, 7):
    kmeans = KMeans(n_clusters=k, random_state=42)
    kmeans.fit(X_scaled)
    inercia.append(kmeans.inertia_)
```



Descarga y Creación de la Base de Datos

Clasificación de Estados por Peligrosidad (Clustering)

```
# Interpretación de clusters  
df_estado.sort_values('Total_delitos', ascending=False)
```

✓ 0.0s

	Entidad	Total_delitos	Cluster
16	México	389492.0	1
6	Ciudad de México	223742.0	2
10	Guanajuato	134626.0	2
13	Jalisco	128588.0	2
1	Baja California	98090.0	2
18	Nuevo León	94321.0	2
29	Veracruz de Ignacio de la Llave	88306.0	2
20	Puebla	75141.0	0
5	Chihuahua	73006.0	0

```
# Aplicar K-means  
kmeans = KMeans(n_clusters=3, random_state=42)  
df_estado['Cluster'] = kmeans.fit_predict(X_scaled)
```

Interpretación de los resultados del clustering

El modelo de clustering permitió agrupar a los estados de México en **tres niveles de peligrosidad** con base en la similitud del total de delitos registrados en 2021.

Cluster de alta peligrosidad:

Estados con el mayor volumen de delitos, concentrando una incidencia delictiva significativamente superior al promedio nacional. Representan las zonas con mayor presión en materia de seguridad.

Cluster de peligrosidad media:

Estados con niveles intermedios de delitos, que muestran patrones mixtos y una incidencia moderada en comparación con el resto del país.

Cluster de baja peligrosidad:

Estados con menor número de delitos registrados, caracterizados por una incidencia delictiva relativamente baja.

Esta clasificación facilita la **comparación objetiva entre entidades**, permite identificar patrones regionales y apoya la toma de decisiones en el análisis de seguridad pública.