CRIME PATTERN ANALYSIS

PROJECT ABSTRACT

Team members:

1. 21PD05 – Dhisha C S
2. 21PD28 - Rohith Krishna R
3. 21PD37 – V K Deeksha

Overview:

This project focuses on analyzing crime data in **Los Angeles from 2020 to the present**. It explores crime trends, patterns, and hotspots using data mining techniques such as classification, clustering, and time series analysis. The goal is to provide meaningful insights for better understanding urban crime behavior.

Data Collection and Preprocessing:

The dataset was collected from the **Los Angeles Open Data Portal** (Data.gov) and also supports local file uploads. The data was cleaned by handling missing values, removing duplicates, and engineering features like victim age, area, and weapon used. Categorical variables were encoded to prepare the data for modeling.

Visualization and Modeling:

- **EDA (Exploratory Data Analysis)** was done using plots and summaries to understand key crime statistics. Time Series Analysis was performed to study crime trends over time.

- **MODELS:** Classification was implemented using a **Decision Tree Classifier** to predict crime categories. Clustering was performed using **K-Means**(cluster based on weapons) and **AGNES Hierarchical Clustering** to group similar crime incidents and analyze crime hotspots.

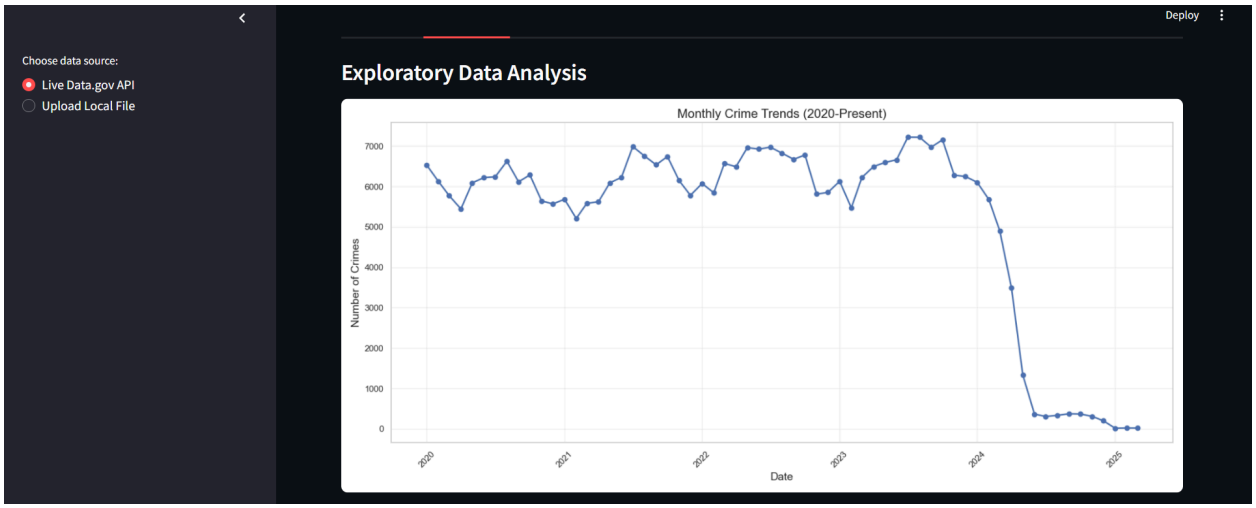
Tools and Technologies:

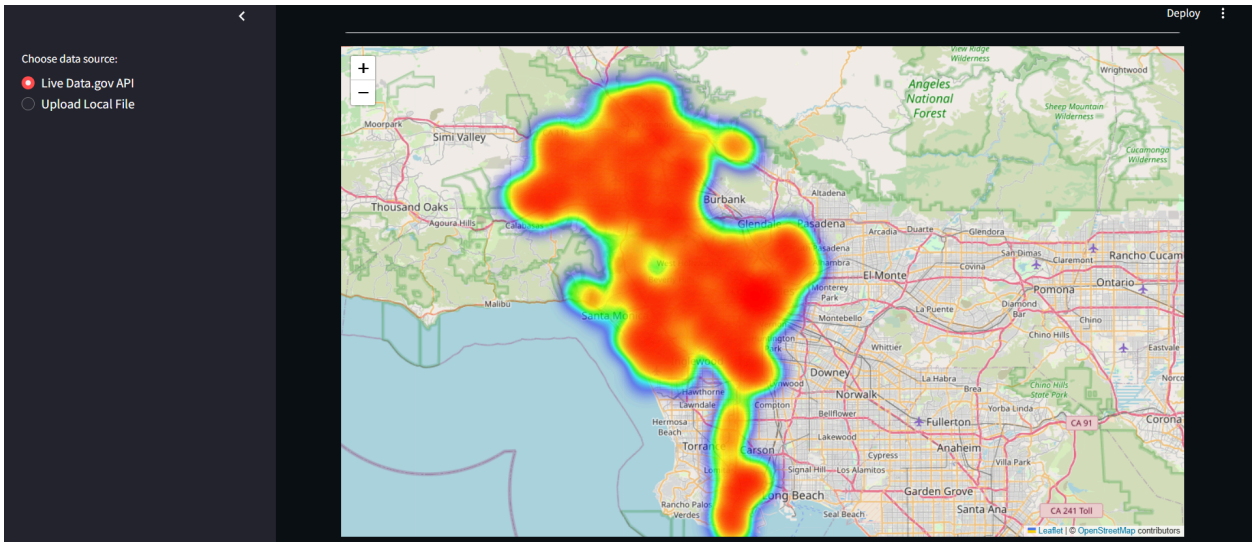
- Python,Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn

- **Streamlit** for building an interactive web application

OUTPUTS:


Deploy
Choose data source:
Live Data.gov API
Upload Local File
Los Angeles Crime Data Analysis (2020-Present)
Data Overview
EDA Visualizations
Model Results
Dataset Preview
DR_NO
Date Rptd
DATE OCC
TIME OCC
AREA
AREA NAME
Rpt Dist No
Part 1-2
Crm Cd
Crm Cd Desc
0 190326475 03/01/2020 12:00:00 AM 03/01/2020 12:00:00 AM 2130 7 Wilshire 784 1 510 VEHICLE - STOLEN
1 200106753 02/09/2020 12:00:00 AM 02/08/2020 12:00:00 AM 1800 1 Central 182 1 330 BURGLARY FROM VEHICLE
2 200320258 11/11/2020 12:00:00 AM 11/04/2020 12:00:00 AM 1700 3 Southwest 356 1 480 BIKE - STOLEN
3 200907217 05/10/2023 12:00:00 AM 03/10/2020 12:00:00 AM 2037 9 Van Nuys 964 1 343 SHOPLIFTING-GRAND THEFT
4 200412582 09/09/2020 12:00:00 AM 09/09/2020 12:00:00 AM 630 4 Hollenbeck 413 1 510 VEHICLE - STOLEN
Dataset Columns
[...]

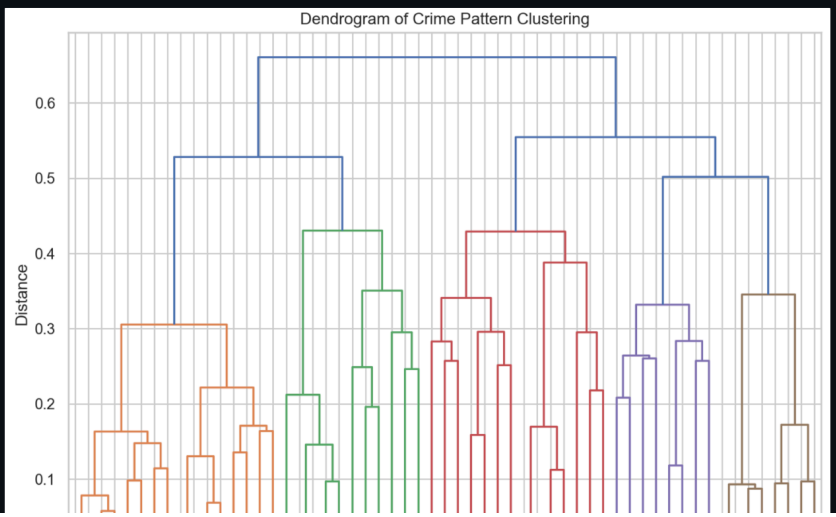

Deploy
Choose data source:
Live Data.gov API
Upload Local File
Exploratory Data Analysis
Monthly Crime Trends (2020-Present)
Number of Crimes
Date
0
1000
2000
3000
4000
5000
6000
7000
2020
2021
2022
2023
2024
2025


Deploy
Choose data source:
Live Data.gov API
Upload Local File
Angeles National Forest
Simi Valley
Thousand Oaks
Burbank
Pasadena
El Monte
Ontario
Rancho Cucamonga
Malibu
Downey
Norwalk
Fullerton
Anaheim
Garden Grove
Santa Ana
Long Beach
CA 91
CA 241 Toll
Leaflet | © OpenStreetMap contributors

Deploy

Choose data source:

- Live Data.gov API
- Upload Local File

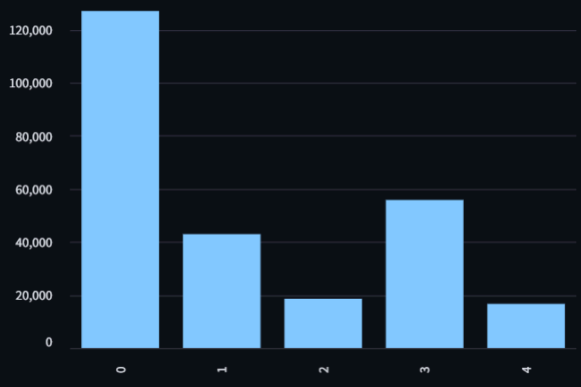


K-Means Clustering:

Number of samples in each cluster: [127064 42947 18587 55829 16662]

Cluster Analysis

Crime Cluster Distribution:



Show Processed Data

Decision Tree Classification

	Value
accuracy	0.4016
precision	0.3905
recall	0.4016
f1_score	0.395