**Phase-1 Submission Template**

**Student Name:** Mukesh A P

**Register Number:** 422423104039

**Institution:** University College of Engineering, Tindivanam **Department:** B.E (Computer Science and Engineering)

**Date of Submission:**

# Problem Statement

Crime patterns in urban areas exhibit spatial and temporal correlations that can inform public safety strategies. This project analyzes San Francisco crime data to identify:

* Geographic hotspots requiring increased police presence
* Time-based patterns (daily/weekly/seasonal variations)
* Relationships between crime types and locations  
  Understanding these trends will help law enforcement optimize resource allocation and enable residents to make safer decisions.

# Objectives of the Project

* Identify top 5 crime categories by frequency and location
* Detect spatial clusters of criminal activity using geospatial analysis
* Build a time-series model to forecast crime trends
* Develop an interactive dashboard showing crime patterns
* Correlate crime rates with demographic factors (where data exists)

# Scope of the Project

****Inclusions**:**

* Analysis of violent and property crimes (2018-2023)
* Temporal patterns at hourly/daily/seasonal levels
* Heatmap visualization of crime density
* Predictive modeling for theft and assault trends

****Limitations**:**

* Excludes unreported crimes
* Limited to SFPD jurisdiction boundaries
* Weather/demographic data requires separate sourcing

# Data Sources

****Primary Dataset**:**

* ****Name**:** SF Crime Dataset (2018-Present)
* ****Source**:**[Kaggle](https://www.kaggle.com/datasets/san-francisco/sf-crime" \t "/Users/sureshcs/Documents\\x/_blank)
* ****Type**:** Public, static CSV (updated monthly)
* ****Size**:** ~600,000 records
* ****Key Fields**:** Incident DT, Category, Latitude/Longitude, Police District

****Supplemental Data**:**

* SF Neighborhoods GeoJSON (for mapping boundaries)
* Weather data from NOAA (optional correlation analysis)

# High-Level Methodology

### Data Collection:

Download CSV from Kaggle

Acquire neighborhood boundaries from SF OpenData

Optional: Pull weather data via NOAA API

### Data Cleaning:

Handle missing coordinates (drop or impute using district centroids)

Standardize crime categories (merge similar types)

Filter outliers (verify coordinates fall within SF boundaries)

Convert datetime formats

### EDA:

* **Temporal Analysis:**

Hourly crime heatmaps (seaborn)

Weekly patterns by crime type (plotly)

* **Spatial Analysis:**

Folium choropleth by police district

DBSCAN clustering for hotspots

* **Statistical Analysis:**

Crime type correlations

Arrest rate by offense

### Feature Engineering:

* **Create temporal features:**

is\_weekend, hour\_of\_day, season

* **Generate spatial features:**

distance\_to\_police\_station

neighborhood\_crime\_density

* **Encode categoricals:**

One-hot encoding for crime categories

### Model Building:

* **Time-Series:**

Prophet for crime trend forecasting

* **Classification:**

Random Forest to predict crime category

* **Clustering:**

1. means for hotspot identification

### Model Evaluation:

Time-Series: MAE, RMSE

Classification: Precision-Recall, F1-score

Clustering: Silhouette Score

### Visualization:

* **Interactive dashboard with:**
* Crime heatmap (Folium)

Temporal trend graphs (Plotly)

Model performance metrics

### Deployment:

Streamlit web app hosted on HuggingFace

Jupyter notebook with full analysis

1. Tools and Technologies

* **Programming Language –**

State the main language you will use **“Python”**

* **Notebook/IDE –**

Mention the platform or environment you’ll work in **“VS Code”** or **“Jupyter Lab”**

* **Libraries -**
* Data Processing: pandas, numpy, geopandas
* Visualization: matplotlib, seaborn, plotly, folium
* ML: scikit-learn, fbprophet, xgboost
* Spatial: shapely, pyproj

### Deployment-

* Streamlit (frontend)
* Docker (containerization)

# Team Members and Roles

| **Member** | **Role** | **Tasks** |
| --- | --- | --- |
| [Name 1] | Data Engineer | Data collection, cleaning, pipelines |
| [Name 2] | ML Specialist | Model development, evaluation |
| [Name 3] | Visualization Expert | Dashboard development, EDA visuals |
| [Name 4] | Project Coordinator | Documentation, presentation |

* ****Key Advantages of This Approach**:**
* **Comprehensive Analysis**: Combines spatial, temporal, and predictive elements
* ****Actionable Outputs**:** Heatmaps and forecasts directly usable by law enforcement
* ****Scalable**:** Methodology adaptable to other cities' crime data.