Proposal for Crime Data Analysis Project

Project Title: Analyzing Crime Trends in Philadelphia

Introduction:

This project aims to perform a comprehensive analysis of crime incidents in Philadelphia. The primary focus will be on exploring various aspects of the crime data, including the identification of trends, seasonal variations, daily patterns, and spatial distributions. Additionally, the project will employ clustering and predictive modeling techniques to uncover deeper insights into the nature and spread of crime across different police districts.

Data Sets:

- Crime Incidents Data (2006-2016 & 2023): This dataset includes detailed records of crime incidents, encompassing types of incidents, geographical coordinates, dates, and times. The data spans over ten years (2006-2016) and also covers the year 2023, providing a broad temporal scope for trend analysis.
- Police District: Geospatial data files containing the boundaries of police districts in Philadelphia. This data will facilitate spatial analysis and enable the visualization of crime trends across different districts.

Research Questions:

- How have different types of crime incidents evolved over the years?
- Are there discernible seasonal trends or patterns in crime incidents based on days or months?
- How are crime incidents distributed spatially across Philadelphianow, and are there noticeable hotspots?
- Can clustering algorithms reveal distinct groupings or patterns in the crime data?
- Is it possible to predict the number of crime incidents based on historical data, using machine learning techniques?

Analysis Methods and Techniques:

- Exploratory Data Analysis (EDA): Conduct a thorough EDA to understand the distribution and nature of the data. This will include trend analysis over years, seasonal trend analysis, and day-wise trend analysis.
- Clustering: Employ clustering techniques (such as K-Means or hierarchical clustering) to categorize police districts or crime types based on similarities in crime patterns or frequencies.

- Predictive Modeling: Implement machine learning models (such as Random Forest, Gradient Boosting, or Support Vector Machines) to predict crime incident counts for 2023, using historical data for training.

Alignment with Project Requirements:

The project satisfies the requirements outlined in the final project description in the following ways:

- Comprehensive Data Analysis: It encompasses a wide range of analyses, including temporal, spatial, and clustering analyses, thus offering a multi-faceted understanding of crime trends.
- Application of Advanced Techniques: By utilizing clustering and predictive modeling, the project goes beyond basic analysis and incorporates advanced data science techniques.
- Real-World Relevance: The findings of this project could have practical implications for law enforcement and public policy, particularly in resource allocation and preventive strategies.