

# Project Report on Crime Data Analysis in Philadelphia

## Introduction

This comprehensive report delves into a detailed analysis of crime trends in Philadelphia, spanning from 2006 to 2016 for temporal insights and focusing specifically on the year 2023 for spatial analysis. My primary aim in this project is to dissect and understand various facets of crime through advanced data analytics, thereby providing substantial insights for informed policy planning and decision-making.

## Concept

The concept of this project was to analyze a comprehensive dataset of crime incidents to uncover trends, patterns, and predictive insights. I tackled a complex dataset covering over a decade, which required careful handling of both temporal and geospatial data. By analyzing crime trends and predicting potential future incidents, my project has practical implications for law enforcement and public policy. The application of advanced clustering and predictive modeling techniques showcased a creative approach to understanding crime data and illuminated pathways for future research endeavors.

## Technical Implementation

**Data Preparation and Cleaning:** The datasets were obtained from an open-source platform and were initially in a relatively clean state. Nonetheless, I undertook a series of meticulous data preparation steps to ensure the utmost accuracy and relevance of the information. This included the integration of the crime incident data with GeoJSON files representing various administrative district levels, a process that required careful data joining. Additionally, I carried out column renaming and data transformation to streamline the dataset for more efficient analysis.

**Exploratory Data Analysis (EDA):** For my exploratory data analysis, I harnessed the capabilities of several powerful Python libraries: seaborn, matplotlib, and geopandas. These tools were instrumental in visualizing the spatial distribution of crime incidents, allowing me to observe how these incidents were dispersed across different district levels and how different types of incidents distributed across the city, such as drunkenness, theft, robbery, etc. I also used line charts and bar charts, which facilitated an exploration of how crime trends fluctuated over the years. I paid particular attention to identifying any significant patterns related to the timing of crimes, such as specific hours of the day, days of the week, or seasons of the year when crimes were more prevalent. This multifaceted approach to EDA provided me with a comprehensive understanding of the temporal and spatial dynamics of crime incidents.

I discovered that in the year 2023, vandalism, fraud, robbery, burglary, and narcotics-related offenses emerged as some of the most prevalent crime incidents. Examining the period from 2006 to 2017, I noted a discernible decline in the overall number of crime incidents, indicating a possible shift in the crime landscape over this timeframe. Moreover, there is a distinct seasonal trend in crime occurrences. Specifically, the summer months consistently showed a higher incidence of crimes, while the winter months registered comparatively fewer incidents. This seasonal fluctuation in crime rates is a notable aspect of the data. In terms of month-to-month variability, the pattern became more pronounced with crime rates reaching their zenith in August, a trend characteristic of the summer period. Conversely, the winter months, particularly those at the beginning of the year, tended to record lower crime figures. There was a marked increase in crime occurrences during the night, particularly between 10 PM and 1 AM. Post-midnight, a gradual decrease in crime incidents was observed, culminating in the lowest frequency around 6 AM. As daylight hours progressed, there was a noticeable uptick in crime incidents, highlighting a daily pattern in crime activities.

**Clustering and Predictive Modeling:** I meticulously aggregated the data by year and police district. My primary objective was to forecast which police districts might experience higher crime rates in the forthcoming year. This predictive analysis was aimed at identifying potential hotspots and directing attention to areas needing heightened vigilance. Also, I employed clustering techniques to group the police districts based on their distinct crime patterns. By doing so, I tried to understand the crime landscape and to reveal the unique characteristics of each district.

**Visualization:** I used the folium package to implement several interactive map, which enhances the depth and clarity of the insights derived from the crime data.