Project Proposal: Philadelphia Crime Analysis

Group 14 | Team Members:
1. Abhidha Arya
2. Neha Chaudhary

Introduction:
Crime rate in the US has varied over time and there is a lot of data pertaining to *criminogenic factors* such as time, place, socio-demographics. Crime Analysis is the examination of relationships between such factors. With such analysis, it will be possible to identify threats and eventually reduce the frequency and severity of such threats.

Objectives:
Our project tries to predict the probability of crime rate against specific ethnic group/ gender with respect to the type of crime and location in Philadelphia. For example, for a specific zip code in Philadelphia 100 crimes are reported. Out of these, 40 were reported against Hispanic Latinos, of which 25 were against females. The focus of our project are these 40 victims.

Methodology/Scope:
Data Mining Techniques that will be used: -
Data Pre-Processing and Exploration
- Data Cleaning
  - Handling Missing Values
  - Regression Analysis
  - Identifying and removing Outliers
  - Resolving inconsistencies
- Data Transformation
- Data Reduction
- Predictive Analysis
  - Classification: Decision Trees
  - Regression

Data:
We will be using Philadelphia Crime data set which is a second-hand data obtained from online source kaggle.com. The data lists down the attributes such as zip code, demographic data (ethnicity, gender), dispatch time, police districts and type of crime. The description of variables is attached in the excel sheet below:

Variables.xlsx