

## **Crime Rate NYPD Predictions**

Milestone: - Final Project Proposal

Group 14

Mayur Mahanta

Aniket Sakharkar

857-437-9190 (Mayur Mahanta)

857-230-5126 (Aniket Sakharkar)

[mahanta.m@northeastern.edu](mailto:mahanta.m@northeastern.edu)

[sakharkar.a@northeastern.edu](mailto:sakharkar.a@northeastern.edu)

Percentage of Effort Contributed by Mayur: 50%

Percentage of Effort Contributed by Aniket: 50%

Signature of Mayur: *Mayur*

Signature of Aniket: *Aniket*

Submission Date: 20<sup>th</sup> May 2022

**Problem Setting:** This is a breakdown of every arrest made in New York City by the NYPD this year. Every quarter, this data is manually extracted and reviewed by the Office of Management Analysis and Planning. Each record represents an arrest made by the NYPD in NYC and contains information about the type of crime, the location, and the time of enforcement. Additionally, information about suspect demographics is included.

**Problem Definition:** Crime is a worldwide issue that can harm a country in both social and economic terms. Data mining provides us with several practical and convenient methods for evaluating large and diverse sets of information. It assists organizations and users in uncovering hidden information from a large database of criminal records to investigate, control, and prevent crime.

As the above setting suggests, we are in possession of data in regards of the crime rates reported by the NYPD during the year 2021.

We can use the same to display insights on:

- The crime type (Rape, Assault, etc.) that is frequently occurring.
- The age groups most likely to commit the crimes.
- The gender and race categories are mostly involved in the crimes throughout the city.

**Data Sources:** Website Link: [NYPD Arrest Data \(Year to Date\) | NYC Open Data \(cityofnewyork.us\)](https://nypd-arrest-data.cityofnewyork.us/)

**Agency:** Police Department NYPD

**Dataset Owner:** NYC Open Data

**Data Description:**

**Rows:** - 156K

**Columns:** - 19

Data Mining techniques or algorithms will be applied to understand which are the areas that are prone to crime and profiling of the individuals involved in committing the crime will be done. Our target variable is a column, "*CRIME\_AREA\_PERPETRATOR\_PROFILE*". A few of the prediction variables/columns of interest are, "*AGE GROUP*" which represents the Perpetrator's age within a category, and "*PERP\_SEX*" which gives the Perpetrator's sex description, etc.