Capstone Project Plan

Rachalle Ann Olila

[Rachalle.olila@gmail.com](mailto:Rachalle.olila@gmail.com)

General Assembly Data Analytics Bootcamp

DAB 0304

## 1. Introduction

Using the NYPD arrest data to understand the distribution of arrest across different neighborhoods in NYC. I will have to examine frequency and type of arrests in various areas.

The goal is to assess the level of crime activity and it potential impacts on neighborhood safety. Hopefully the findings will provide some insights into how NYPD arrest data can inform decisions to residence, education and employment choices within the city. By the end of this, I should be able to make a recommendation on the “safest” area to live in in NYC.

Will use Python & SQL

Visualization will be made on Tableau

Build an interactive tool on Tableau

## 2. DATA PRIMARY AND SECONDARY SOURCES TO SUPPORT YOUR ANALYSIS

I will be using the NYPD Arrest Data (Historic) available publicly on Data.gov. The dataset contains records from 2006 through the end of the previous calendar year (2023).

The original dataset:

Total number of columns: 19

Total number of rows: 5,725,522

|  |  |
| --- | --- |
| Column name | Column Description |
| ARREST\_KEY | Randomly generated persistent ID for each arrest |
| ARREST\_DATE | Exact date of arrest for the reported event |
| PD\_CD | Three digit internal classification code (more granular than Key Code) |
| PD\_DESC | Description of internal classification corresponding with PD code (more granular than Offense Description) |
| KY\_CD | Three digit internal classification code (more general category than PD code) |
| OFNS\_DESC | Description of internal classification corresponding with KY code (more general category than PD description) |
| LAW\_CODE | Law code charges corresponding to the NYS Penal Law, VTL and other various local laws |
| LAW\_CAT\_CD | Level of offense: felony, misdemeanor, violation |
| ARREST\_BORO | Borough of arrest. B(Bronx), S(Staten Island), K(Brooklyn), M(Manhattan), Q(Queens) |
| ARREST\_PRECINCT | Precinct where the arrest occurred |
| JURISDICTION\_CODE | Jurisdiction responsible for arrest. Jurisdiction codes 0(Patrol), 1(Transit) and 2(Housing) represent NYPD whilst codes 3 and more represent non NYPD jurisdictions |
| AGE\_GROUP | Perpetrator’s age within a category |
| PERP\_SEX | Perpetrator’s sex description |
| PERP\_RACE | Perpetrator’s race description |
| X\_COORD\_CD | Midblock X-coordinate for New York State Plane Coordinate System, Long Island Zone, NAD 83, units feet (FIPS 3104) |
| Y\_COORD\_CD | Midblock Y-coordinate for New York State Plane Coordinate System, Long Island Zone, NAD 83, units feet (FIPS 3104) |
| Latitude | Latitude coordinate for Global Coordinate System, WGS 1984, decimal degrees (EPSG 4326) |
| Longitude | Longitude coordinate for Global Coordinate System, WGS 1984, decimal degrees (EPSG 4326) |
| Lon\_Lat | Column with both longitude and latitude in a single line. |

This dataset comprises arrest records spanning from 2006 to 2023. To streamline the analysis and minimize data volume, I will narrow the focus to the most recent 10 years (2014-2023). The analysis will prioritize crime characteristics rather than individual profiles; therefore, demographic information will be excluded to avoid potential biases in the analysis.

I will most likely have to create some sort of seperations for the years to analyze individually.

## 3. ANALYSIS DATA DRIVEN INSIGHTS

Question to jumpstart my analysis (available on Trello):

1: How has the overall arrest count changed over the years?

2: What is the distribution of arrests by crime type in New York City, and which crime types are the most common offenses in each borough/neighborhood according to NYPD arrest data?

* + Which borough has the highest arrest rate?which crime?
  + Which neighborhood had highest reported arrest?

3: How do arrest rates vary across different boroughs, neighborhoods, or precincts? Are there any high-crime areas or patterns of spatial clustering?

4:How do arrest patterns differ across different time periods, boroughs, or neighborhoods? Are there any notable differences or similarities in arrest trends between various groups or locations?

Create a map to see any high crime patterns. Explore any disparities in arrest rates across NYC. Using "Latitude" & "Longitude" to create a map to see high crime patterns.

Time Analysis: find any seasonality in arrest in activity and explore potential factors contributing.

* Which months had the highest record of arrest? Any holidays near it?
* Do events have a correlation to arrest rates?

Classify crimes based on their severity (e.g., misdemeanor vs. felony) and analyze the distribution of arrests by crime severity level. Determine which types of crimes are more likely to result in arrests for specific locations.