**NYPD Arrest Data**

**Wei Chen** wc2755, **Yini Fang** yf2517, **Andrew Li** hl3534,

**Nan Liao** nl2783, **Lingfei Zhang** lz2803

**Introduction**

NYPD Arrest Data from NYC OpenData contains the information of every arrest in NYC made by the NYPD in the year of 2021. The data has 155507 rows, each represents an arrest effected in NYC by the NYPD and includes information about the type of crime, the location and time of enforcement. It has 19 columns and contains three major groups of data.

The first group of columns contains the detailed classifications of the reason for arrestments. It has five columns that details the PD code and the KY code for the classification, and the level of offenses. The second group contains the location of the arrestment, which has the borough, the precinct, the longitude, and the latitude. The third group contains the demographics of the arrestees, which included the age group, the sex, and the race.

**Data Preprocessing**

The data is very clean. After dropping columns that contain NAs, the resulting data frame is almost free of human error. The arrest data is converted into a datetime object for easier analysis.

**Exploratory Data Analysis**

We first performed a simple breakdown of the dataset by age. As shown, the majority of Chart, pie chart

Description automatically generatedperpetrators is still in the range of 25-44.

We want to investigate how the number of arrests differ across different boroughs in New York. In the graphs, B stands for Bronx, S stands for Staten Island, K stands for Brooklyn, M stands for Manhattan, and Q stands for Queens.

The number of incidents, which is the first figure, does not signify anything because the boroughs have different sizes and different populations. Thus, we researched the population of each borough online and divided the number of incidents for each borough by the population of that borough to get the second figure. From the second figure, we can see that Bronx and Manhattan has relatively higher numbers of incidents by borough per capita. We also researched the area of each borough and get the number of incidents by area. And Manhattan has the largest number of incidents per kilometers.

Chart, bar chart

Description automatically generated

We also collected national holiday in 2021, which in total has 15 days, and tried to figure whether the number of incidents would differ when there was a holiday. And according to the figure, there seems to be a decrease in number of incidents when it is a holiday.

Chart, bar chart

Description automatically generated

**Data Preprocessing**