

# Examining the NYPD Arrests Dataset

Varun Somani (vs472), Mengjia Xia (mx233), Binxin Liu (bl642)

ORIE 4741 Learning with Big Messy Data  
Fall 2019

The dataset we will be using is the NYPD Arrests Data [1] owned by NYC OpenData. The data is manually extracted every quarter by Police Department (NYPD). Each record represents an arrest effected in NYC by the NYPD and includes information about the type of crime, the location and time of enforcement. The dataset covers the period 2006 - 2018, and has 4.8M records in total.

We will use the dataset to try to answer the following question:

- How has the crime type changed in a certain region over the 12 years the dataset presents?

This question is important because the dataset ranges from time the world was reeling from the massive financial collapse which was centered in NYC. We can examine how an increase in prosperity from that time to now has affected the type of crime committed.

In addition, information related to suspect demographics is also included in this dataset. We can also examine the race and gender. Thus, it will allow us a broad view of how crime has changed in the various regions, and whether certain racial or gender groups have fared better over the time than others.

## References

- [1] NYPD, "NYPD Arrests Data (Historic): NYC Open Data," <https://data.cityofnewyork.us/Public-Safety/NYPD-Arrests-Data-Historic-/8h9b-rp9u>, 2019, [Online; accessed 23-September-2019].