An Analysis of Crime in NYC

Ali Abbas Causer
Sprint 1 - October 2023

Agenda

The Goal

Overview of datasets

Preliminary EDA

Next Steps

The Goal

To identify root causes and predict crime in New York (5 boroughs)

- Can you predict a categorical type and severity of crime based on day of week, weather and location? (Log Regression, KNN, Decision Trees)

  - Why is this useful: Can help provide information to residents about safety

- What underlying demographic features (excluding race/ethnicity) are most heavily correlated crime (type and frequency)? (Log regression)

  - Why is this useful: Can help provide local public organizations areas to improve (e.g. education, shelter, healthcare, jobs)

- Relationships of type/severity of crime with age groups and why? (Linear regression)

  - Why is this useful: "Customer" segmentation to develop a targeted approach for action

Overview of Datasets

Dataframe	Description	Rows / Columns	Primary Key	Cleaning Required	Source
Arrests (Year-to-Date)	Record of each arrest by type and severity, including basic demographic info	112,571 / 20 (500+ / day)	Precinct	Feature reduction / null	NYC Open Data
Demographic Snapshot	School enrollment by grade, disabilities, english language learners, % poverty, Economic index	9251 / 44	School Name	Feature reduction / duplicates	NYC Open Data
Financial Information	Median income, poverty rate, jobs, housing capacity, credit score, etc.	385 / 52	Zip Code	Feature reduction / null values	NYC Open Data
Weather	Daily weather by zip code	TBD	Zip Code	TBD	NOAA

Location of Crime
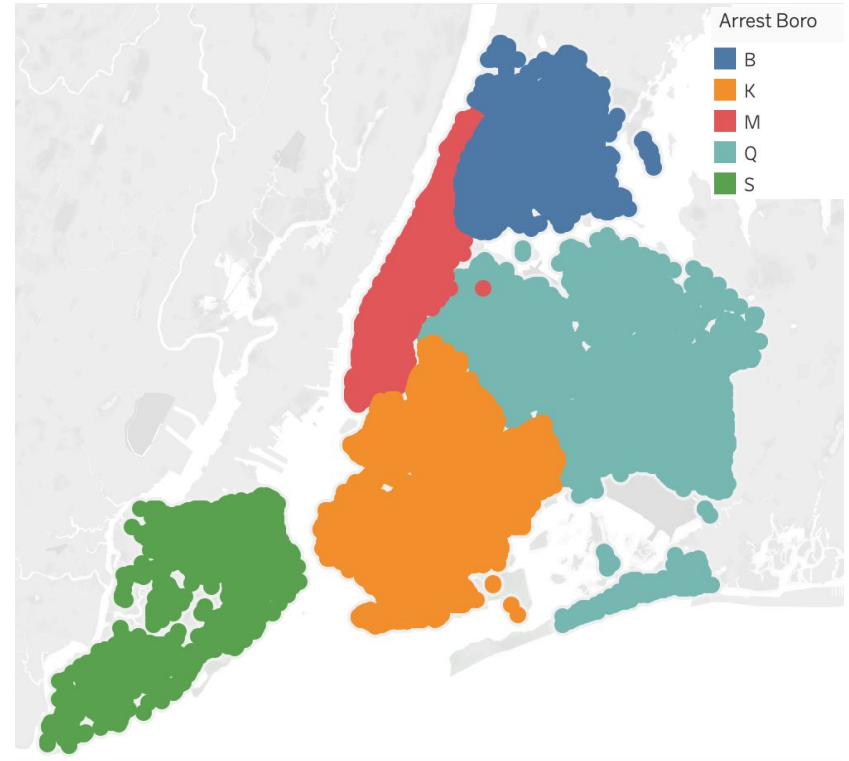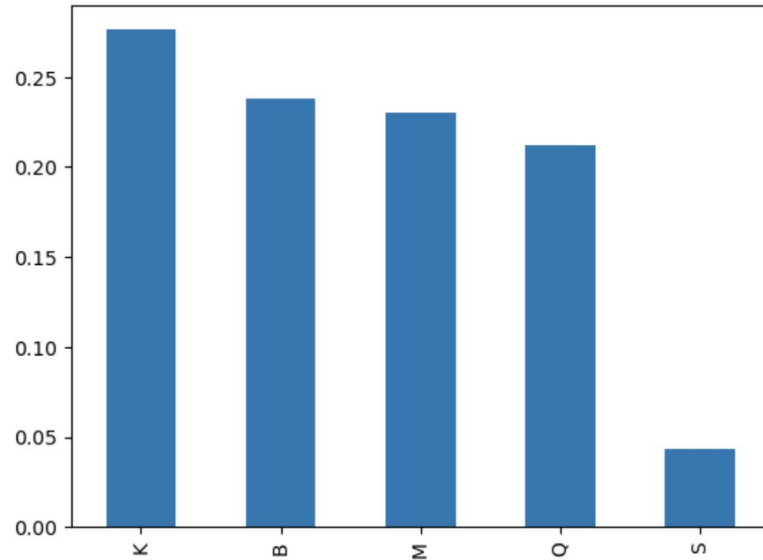
Age / Sex
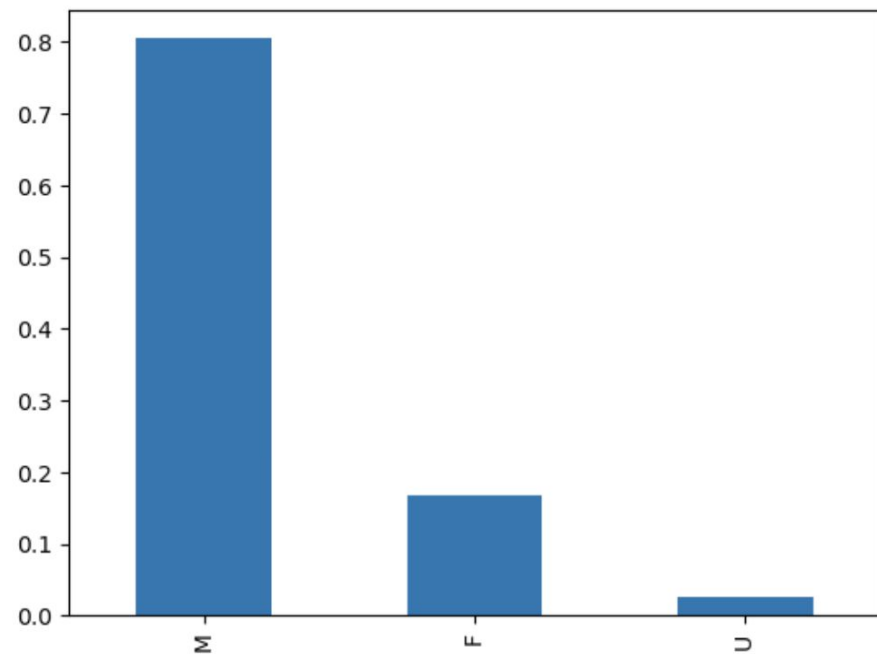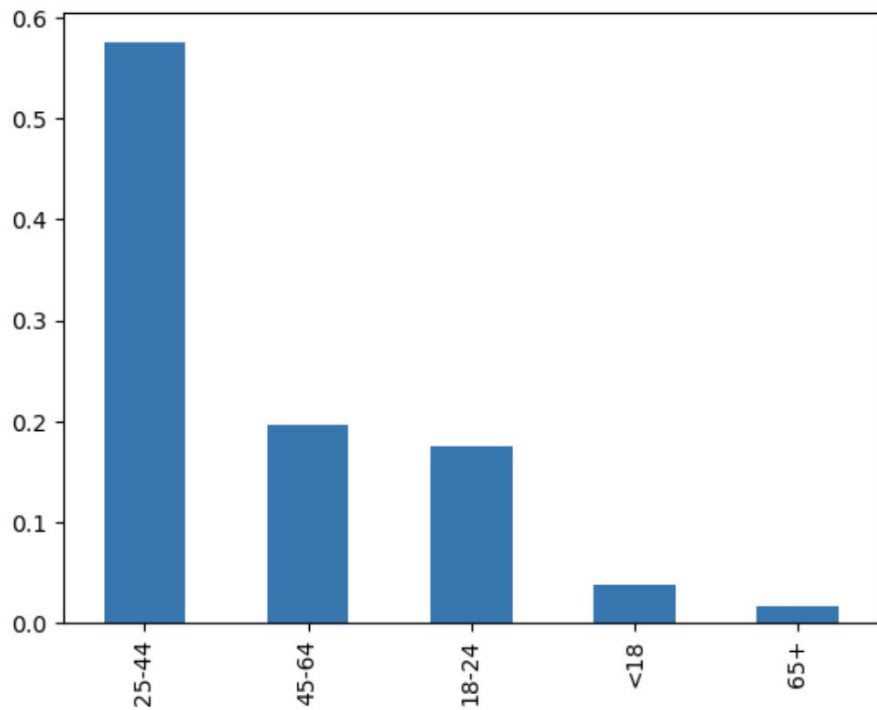
Type / Severity

```
ASSAULT 3 & RELATED OFFENSES        16619
PETIT LARCENY                       12070
FELONY ASSAULT                      10474
DANGEROUS DRUGS                      7530
MISCELLANEOUS PENAL LAW              7327
                                      ...
PARKING OFFENSES                        3
FELONY SEX CRIMES                       2
DISRUPTION OF A RELIGIOUS SERV          2
UNLAWFUL POSS. WEAP. ON SCHOOL          1
ADMINISTRATIVE CODES                    1
```
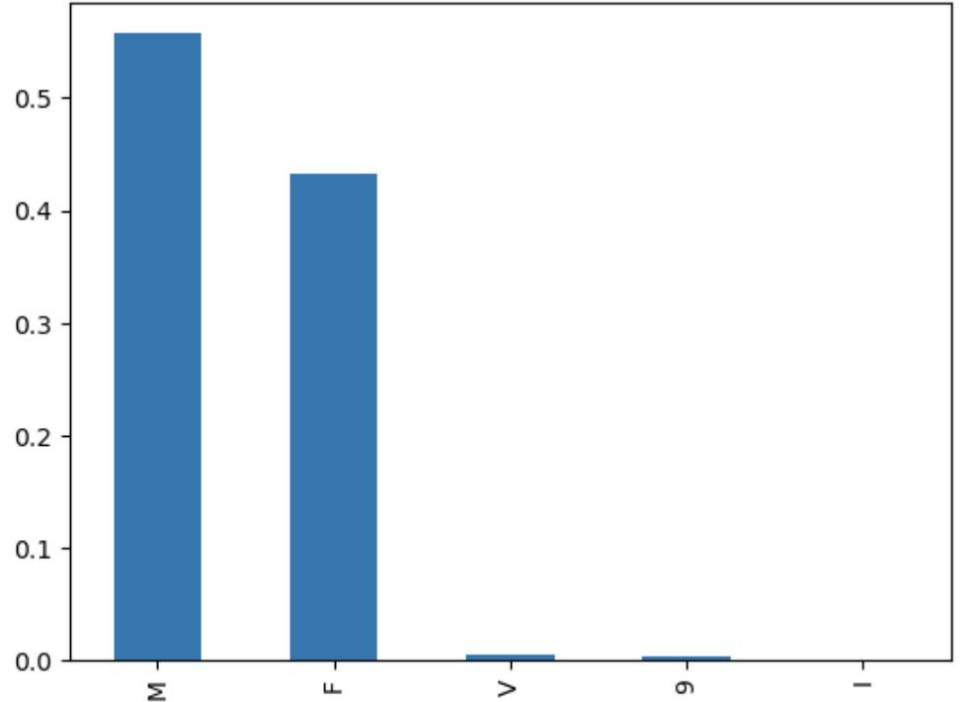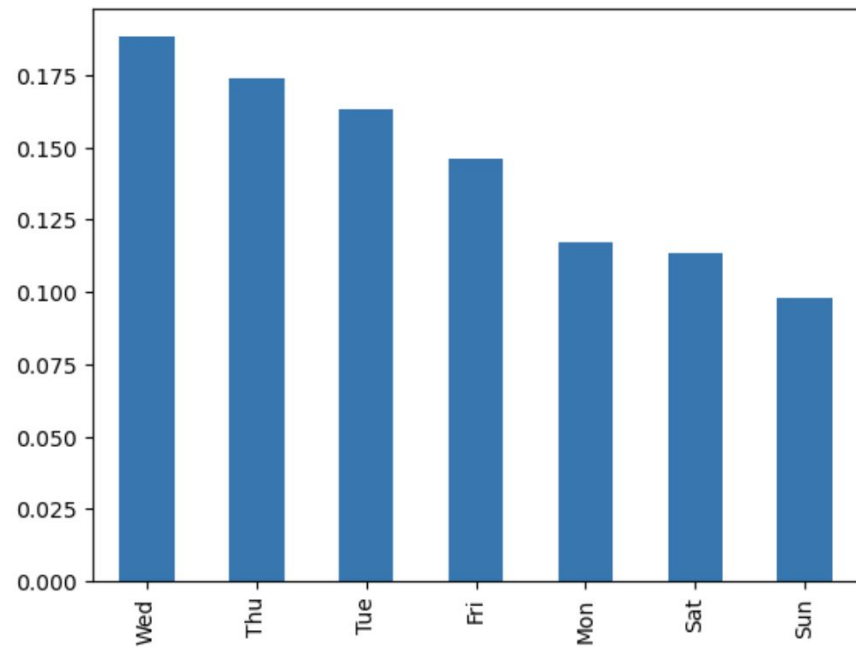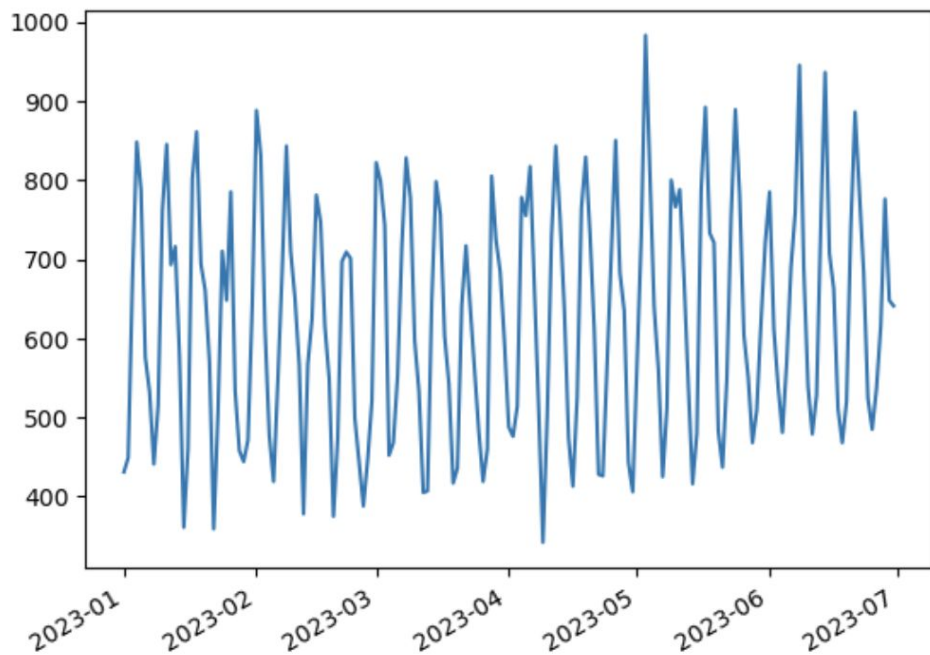
Time Series

Next Steps

- Join tables with a common location feature

- Get daily weather data

- Run contingency table tests with various variables

- Look for multicollinearity for feature reduction

- Create continuous buckets for type of crime and age brackets