

An Analysis of Crime in NYC

Ali Abbas Causer
Sprint 2 - November 2023

The Goal

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

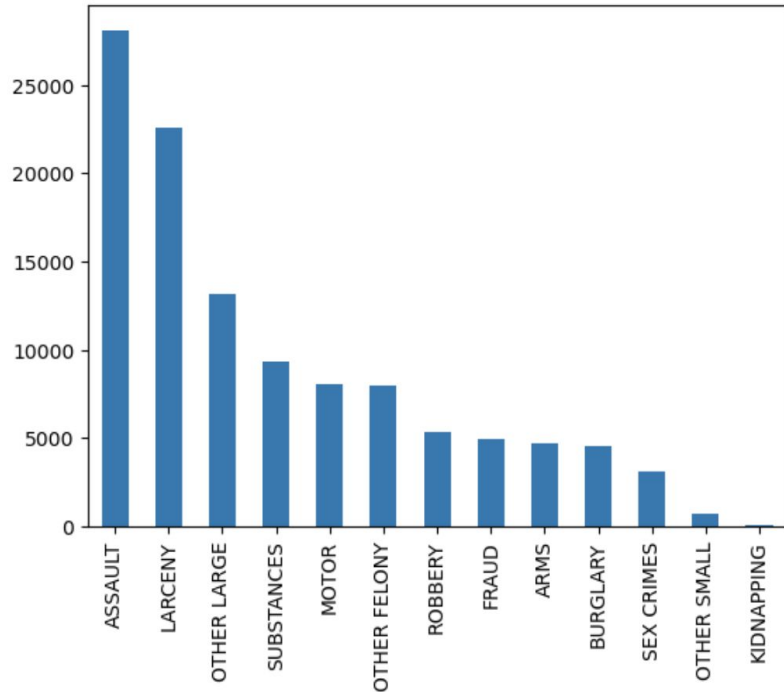
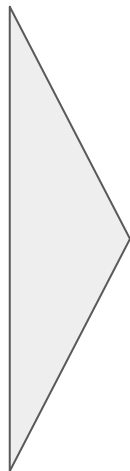
- Can you predict a categorical type and severity of crime based on day of week, weather, location (demographics)? (Log Regression, KNN, Decision Trees)
 - Why is this useful: Can help provide information to residents about safety
- What underlying demographic features (excluding race/ethnicity) are most heavily correlated with 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

Since last time...

- Feature engineering
- Run a model
- Identified a better data source for relevant demographic features to incorporate in the model
- Join arrests data with demographic and economics data using geopandas

Feature engineering

```
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
Name: OFNS_DESC, Length: 63, dtype: int64
```

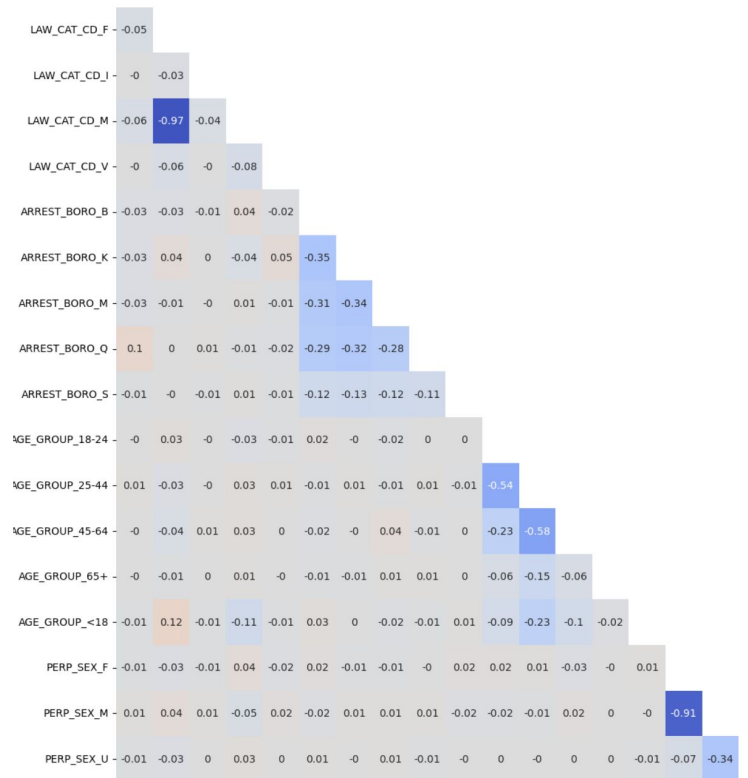
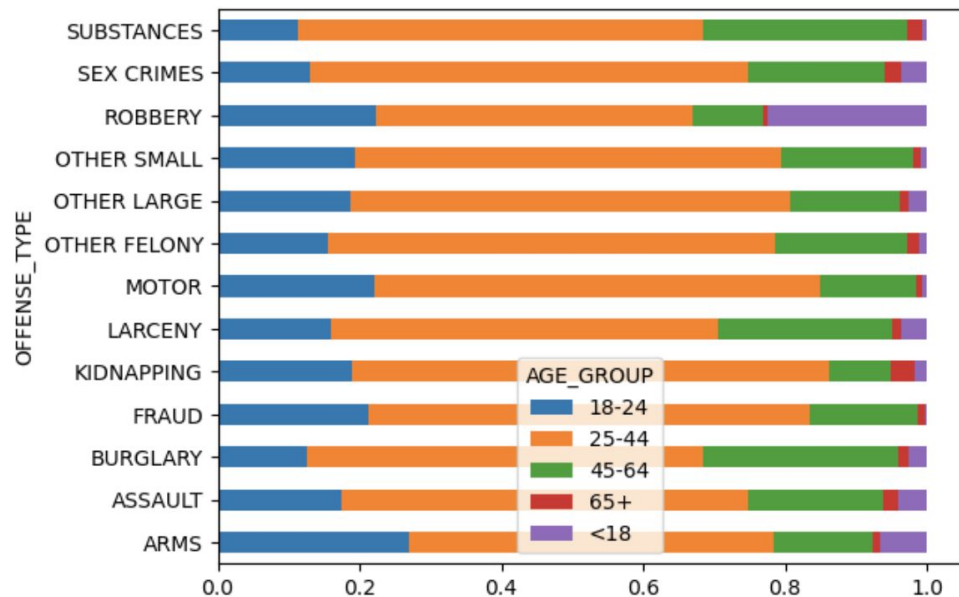


Imputations

```
larceny = arrests[arrests['OFFENSE_TYPE'].str.contains('|'.join(['larceny', 'thef', 'stolen', 'steal']), case=False, na=False) == True]
burglary = arrests[arrests['OFFENSE_TYPE'].str.contains('|'.join(['burglar', 'trespass']), case=False, na=False) == True]
substance = arrests[arrests['OFFENSE_TYPE'].str.contains('|'.join(['drug', 'cannabis', 'alcohol']), case=False, na=False) == True]
```

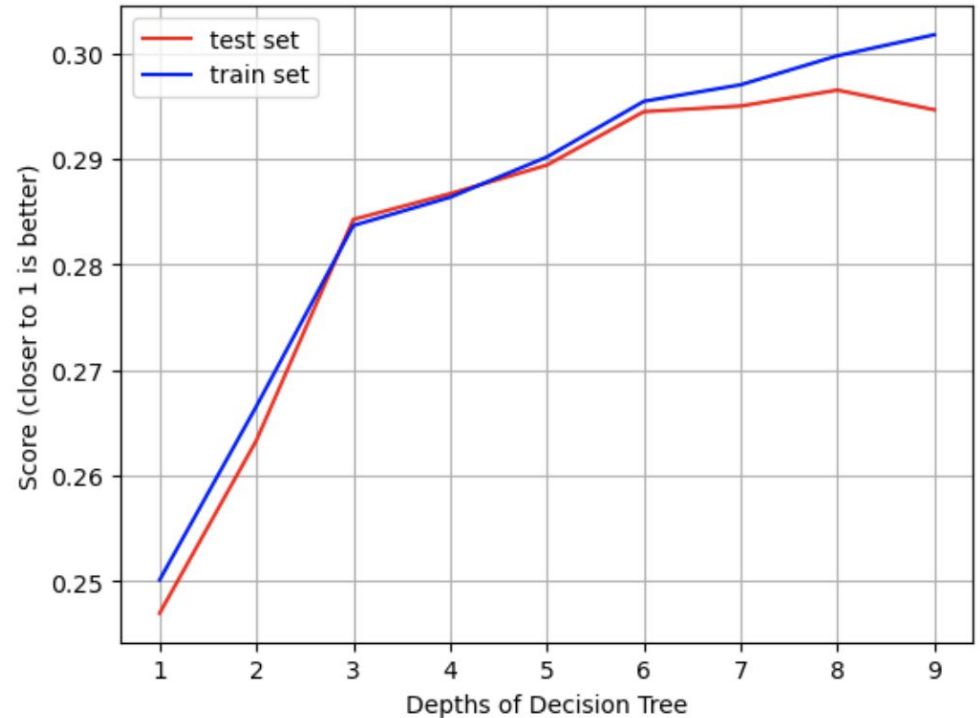
Felony > 80% of category = Other Felony, Category > 100 instances = Other Large

Pre-model assessment



Logistic Regression / Decision Tree

31% Train
30% Test



Additional location based features

Bronx Community District 2

Indicators Built Environment Floodplain Community Board Projects Resources

Estimate Margin of error

Age Under 18¹

27.5%

of residents are under the age of 18

Bronx: 25.1% NYC: 20.9%

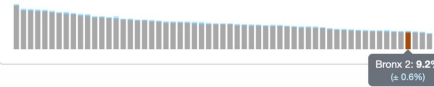


Age 65 & Over¹

9.2%

of residents are age 65 and over

Bronx: 12.1% NYC: 14.1%

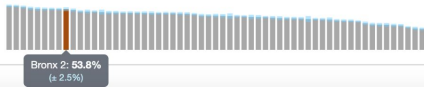


Rent Burden¹

53.8%

of households spend 35% or more of their income on rent

Bronx: 50.9% NYC: 44.2%

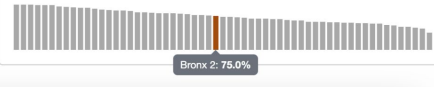


Access to Parks¹

75.0%

of residents live within walking distance of a park or open space

Bronx 2: 75.0%



Mean Commute to Work¹

44.9
minutes

Bronx: 44.8 NYC: 41.2



Street Cleanliness¹

81.3%

of streets rated "acceptable" on the Street Cleanliness Scorecard in FY 2017

Bronx: 88.3% NYC: 95.0%



Crime¹

864

major felonies were reported in 2019

Bronx: 18,859 NYC: 92,480

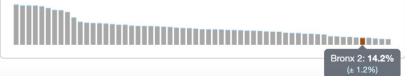


Educational Attainment¹

14.2%

of residents age 25 years and over have earned a bachelor's degree or higher

Bronx: 19.8% NYC: 37.4%

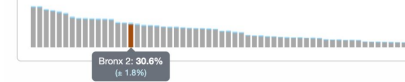


Limited English Proficiency¹

30.6%

of residents self-identify as having limited English proficiency

Bronx: 26.0% NYC: 22.8%

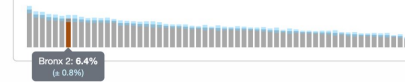


Unemployment¹

6.4%

of the civilian labor force is unemployed

Bronx: 6.3% NYC: 4.4%

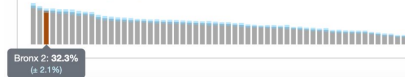


NYCgov Poverty Measure¹

32.3%

of residents have incomes below the NYCgov poverty threshold. Compare to federal poverty rate.

Bronx: 27.9% NYC: 19.7%



Next steps for model improvement...

- Further tighten target categories
- Re-run analysis using expanded feature set with demographic data
- Get daily weather data
- Break down the analysis into small studies
- GridSearch to assess performance across multiple models and params