Crime Prediction Using Classification Model

February 11, 2020

Rising Crime Rates in New York City

Crimes doesn’t seem to stop in New York City…

Washington Free Beacon

Crime Is on the Rise in New York City

Major crimes jumped in New York City in January, new data from the NYPD show, bucking the trend of decades of declining crime. Compared ...

4 days ago

CBS New York

Mayor Bill De Blasio Defends Policing Policies As NYPD Statistics Show Rise In Shootings, Other Major Crimes

NEW YORK (CBSNewYork) — Mayor Bill de Blasio is defending his policies in the face of skyrocketing crime in New York City. It's a dramatic ...

6 days ago

Newsday

Data: NYC serious crime in January worst in five years

New York City experienced its worst January for serious crime in five years, according to NYPD data set for release Tuesday and expected to ...

3 days ago

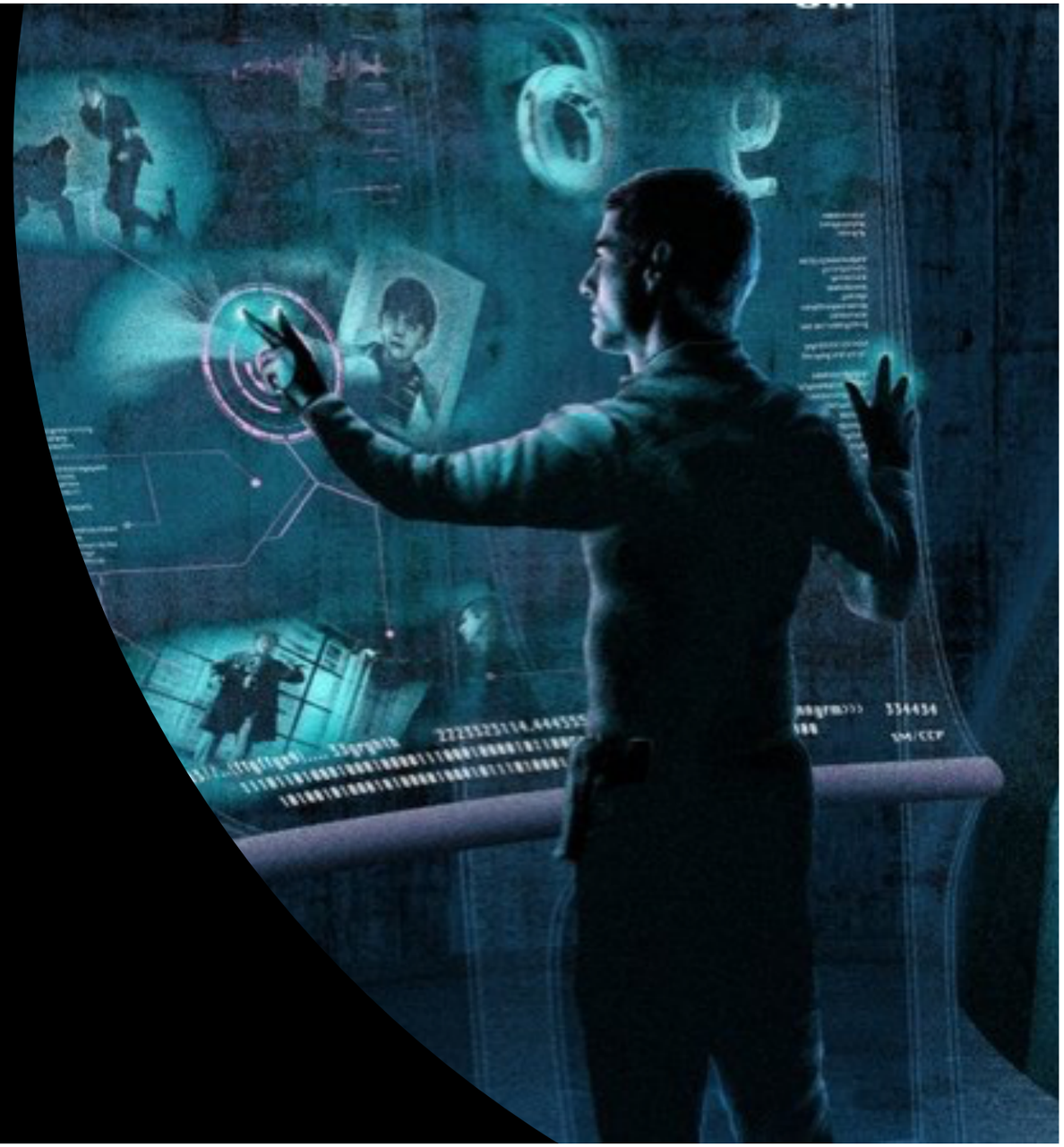
What if machine can predict crimes..?

Predictive policing using classification models

To predict most likely high crime areas in New York City on any given day

Dataset

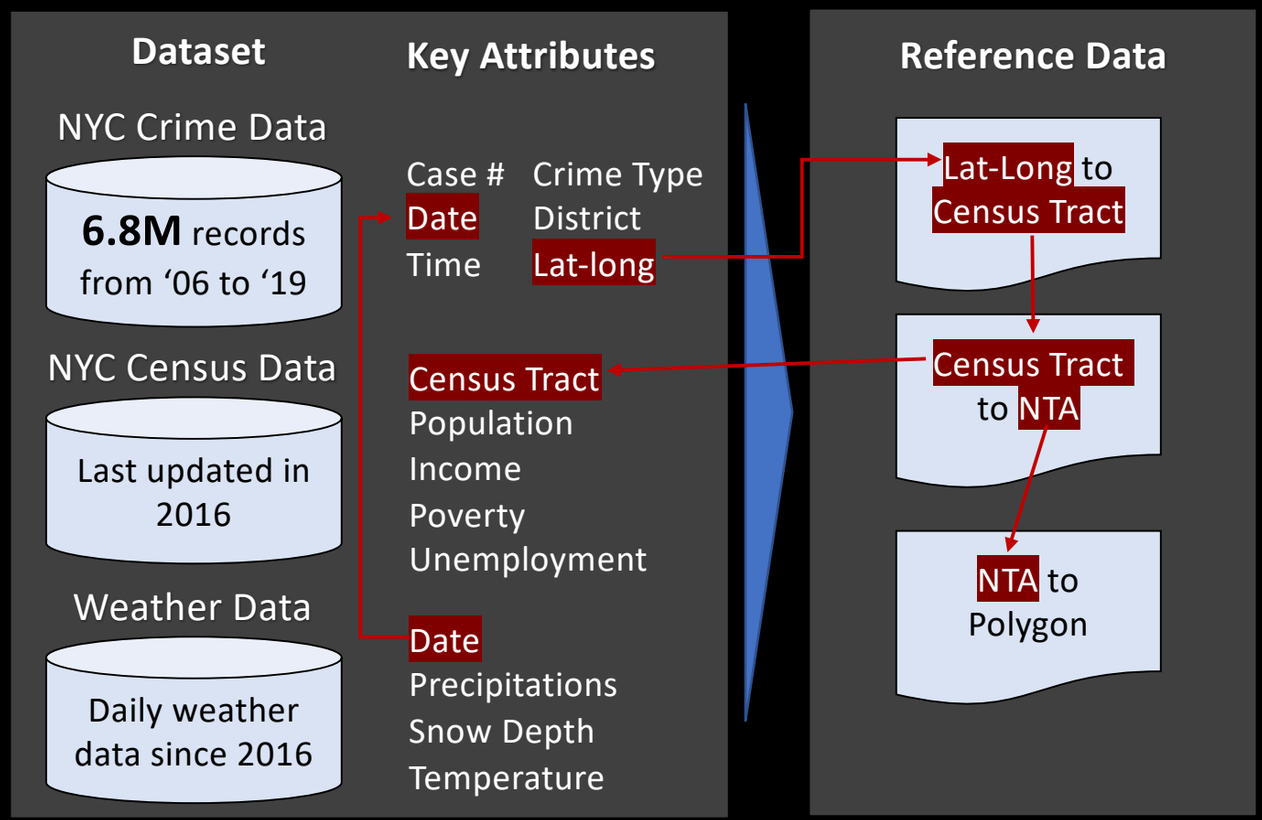
- Crime Data (from NYPD webpage)
- Census Data (from census.gov)
- Weather Data (from weather.com)



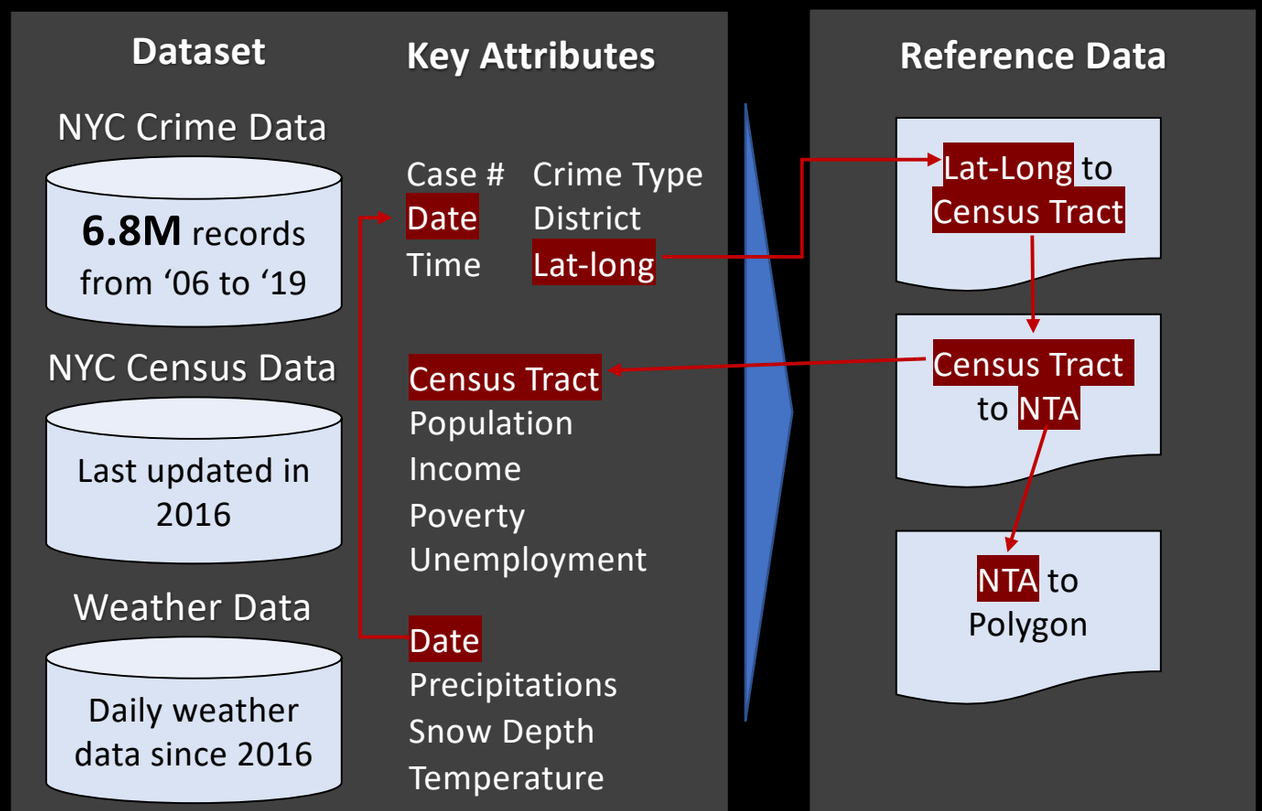
Data Collection & Processing

Dataset	Key Attributes
NYC Crime Data **6.8M** records from ‘06 to ‘19	Case # Crime Type Date District Time Lat-long
NYC Census Data Last updated in 2016	Census Tract Population Income Poverty Unemployment
Weather Data Daily weather data since 2016	Date Precipitations Snow Depth Temperature

Data Collection & Processing



Data Collection & Processing



NTA	Crime Data	Census Data	Crime Level
Neighbor-hood name	**Timing** and **daily crime counts**	Population, Income, Poverty/Un-employment	High (1) Low (0) <u>Threshold</u> **Daily Counts > 8**

- Grouped by NTA on a daily basis
- 2016 ~ 2019
- Manhattan, Queens, Brooklyn
- Approx. **230K** records

Feature Engineering

Baseline Model Using Logistic Regression

Baseline Features

Total Population

Income level

Men/Women

Poverty/Unemployment Rate

Accuracy: 69%

Recall: 17%

Baseline Model Using Logistic Regression

Baseline Features

Total Population

Income level

Men/Women

Poverty/Unemployment Rate

Accuracy: 69%

Recall: 17%

12%

in accuracy

44%

in recall

Additional Features

Total Population

Income level

Men/Women

Poverty/Unemployment Rate

Location

Seasonality

Day time vs. Night time

Accuracy: 81%

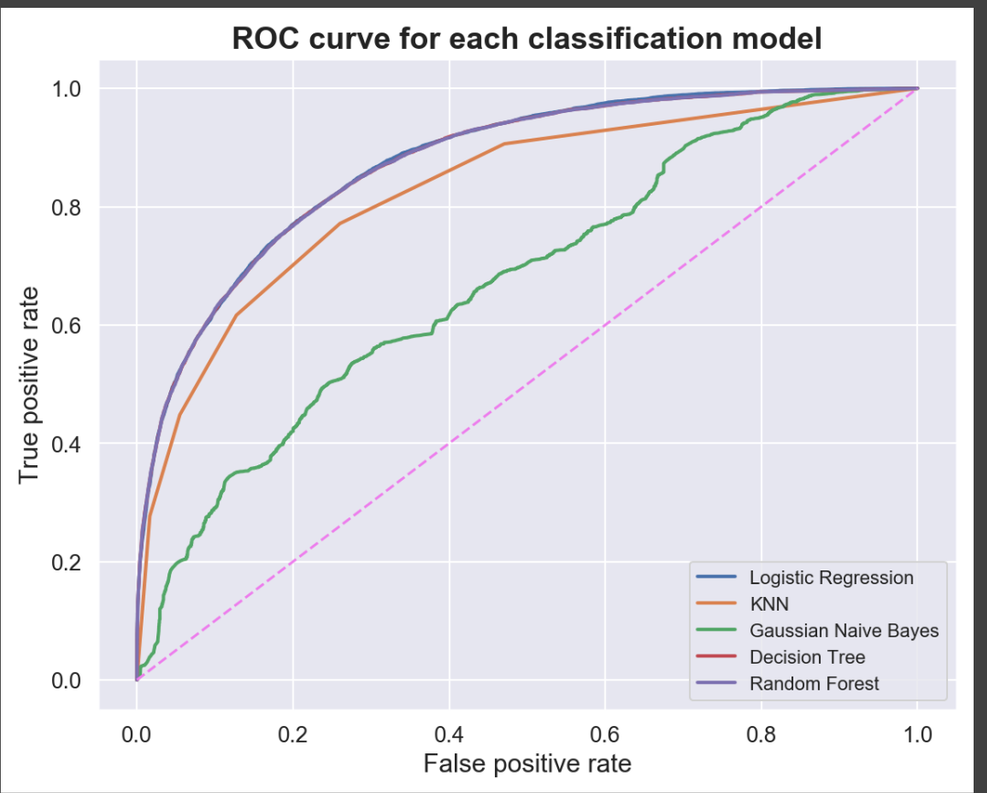
Recall: 62%

Model Selection

Logistic Regression, Decision Tree, Random Forest are all very good predictors!

Final model: <u>Logistic Regression</u>

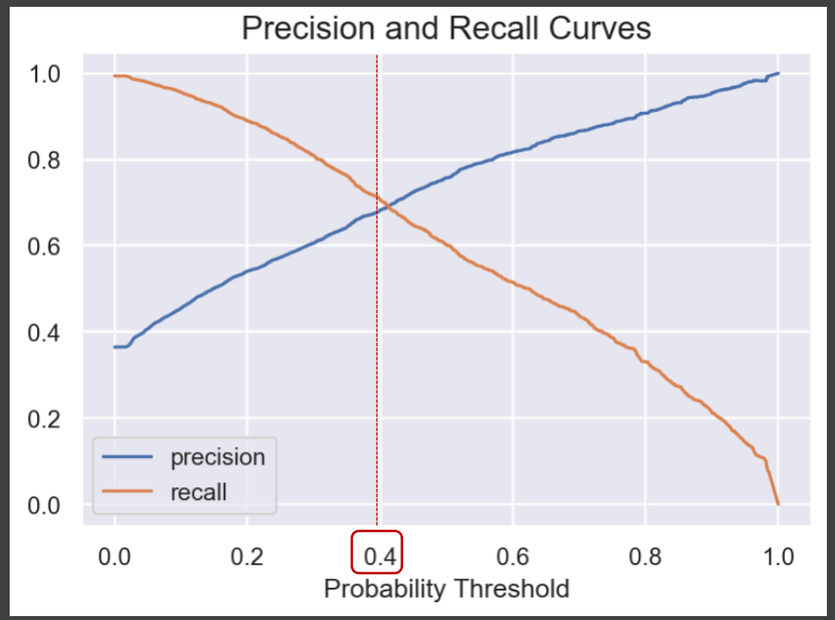
- Slightly higher recall rate and AUC
- Simpler Model
- Interpretability



	Logistic Reg.	KNN	Gaussian NB	Decision Tree	Random Forest
Accuracy	0.81	0.79	0.71	0.81	0.81
Precision	0.75	0.69	0.55	0.76	0.76
Recall	0.62	0.62	0.35	0.60	0.60
F1	0.68	0.65	0.43	0.67	0.67

Improving Model Performance

Moving Probability Threshold



Probability Threshold: 40%

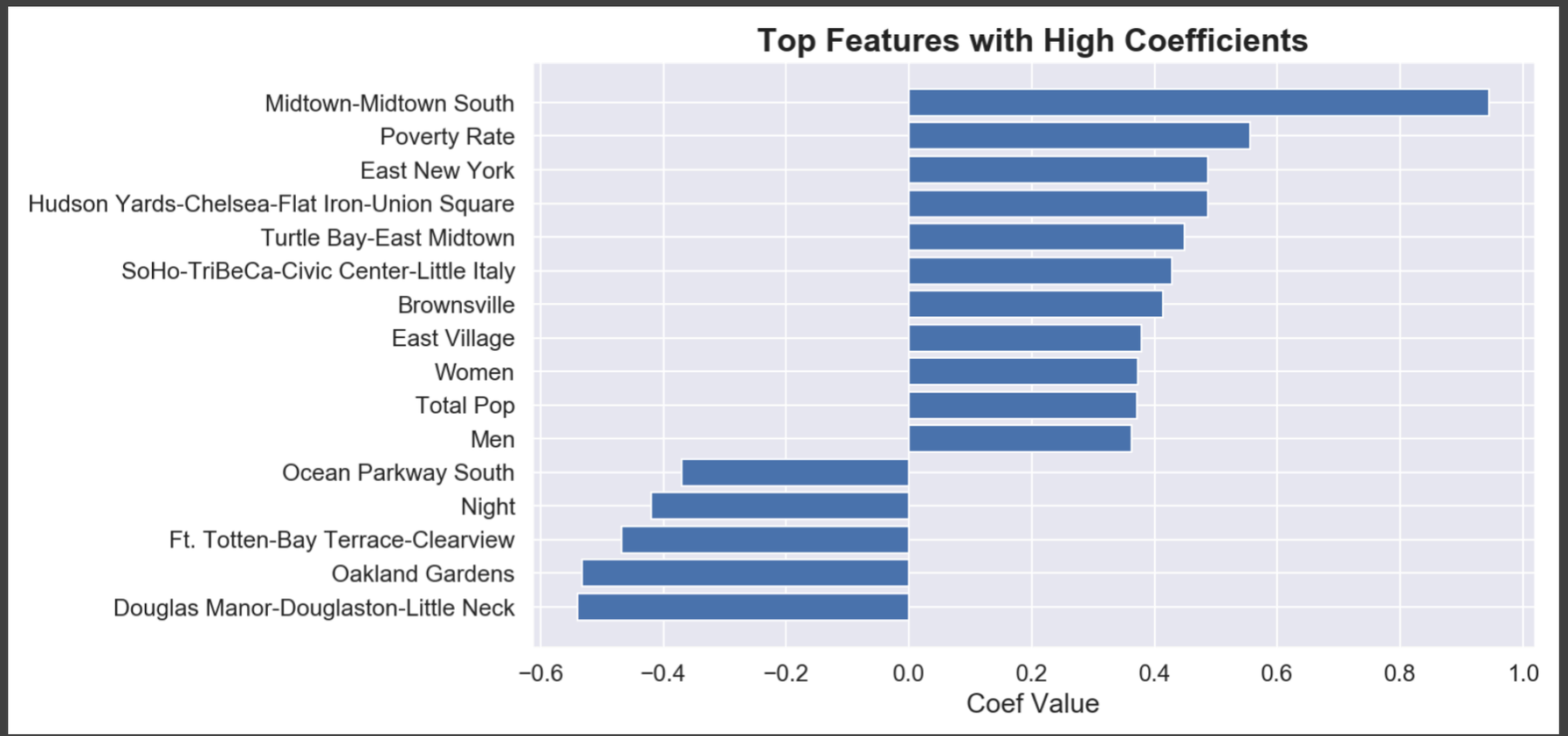
Accuracy: **80%** **1%** **frm default**

Recall: **71%** **9%** **frm default**

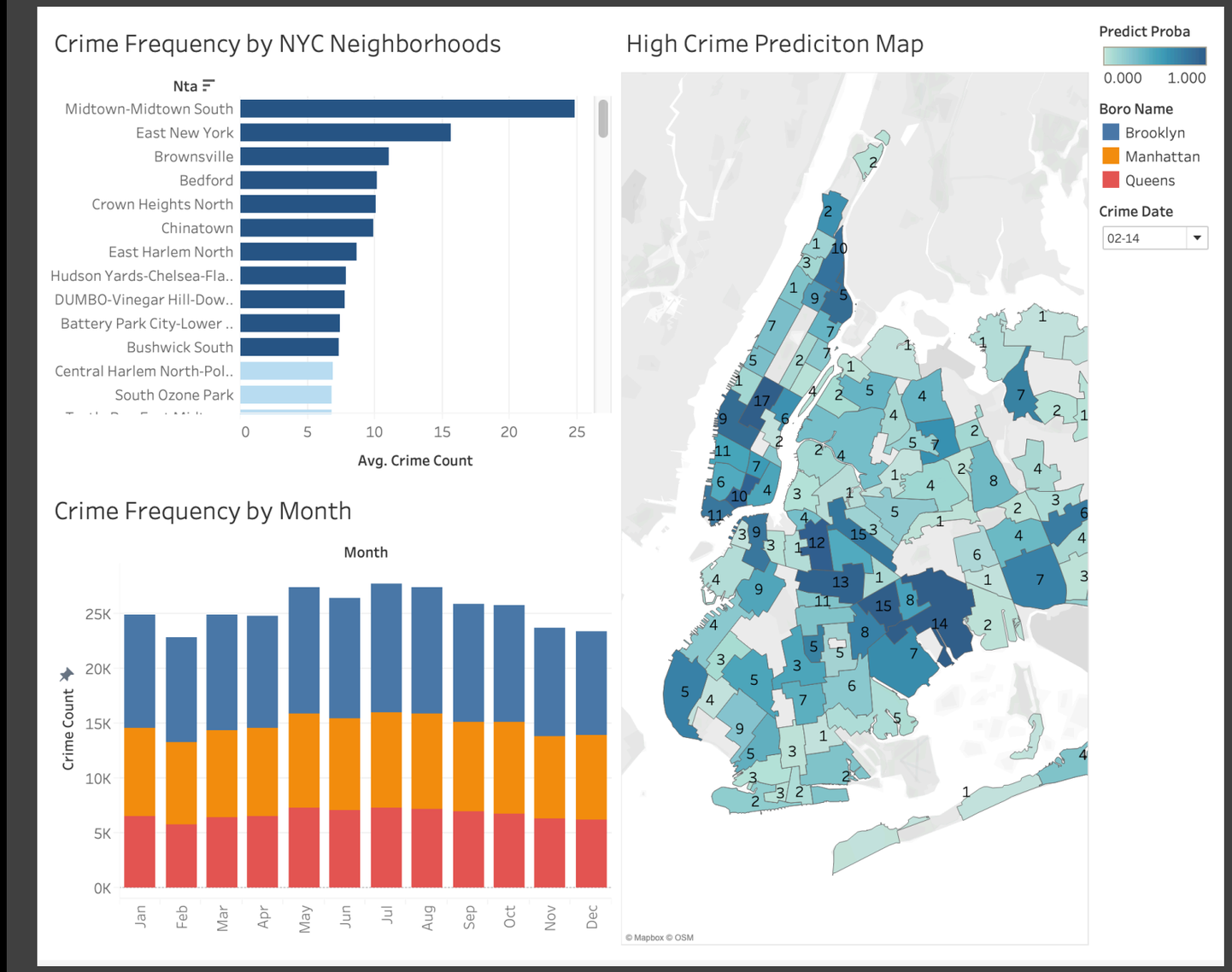
Adding Weather Related Features (e.g., precipitation, snow level, temprature)

No improvement found

Feature Importance



Visualization with Final Model



Future Work

- Obtain more granular weather data
- Predict crime counts and crime types
- Scale it to other cities

Question?