SFPD Crime Investigation

Christian Hansen

Questions:

- Where does crime occur?
- How often does crime occur?
- Given new data can I predict the probability that a type of crime will occur?

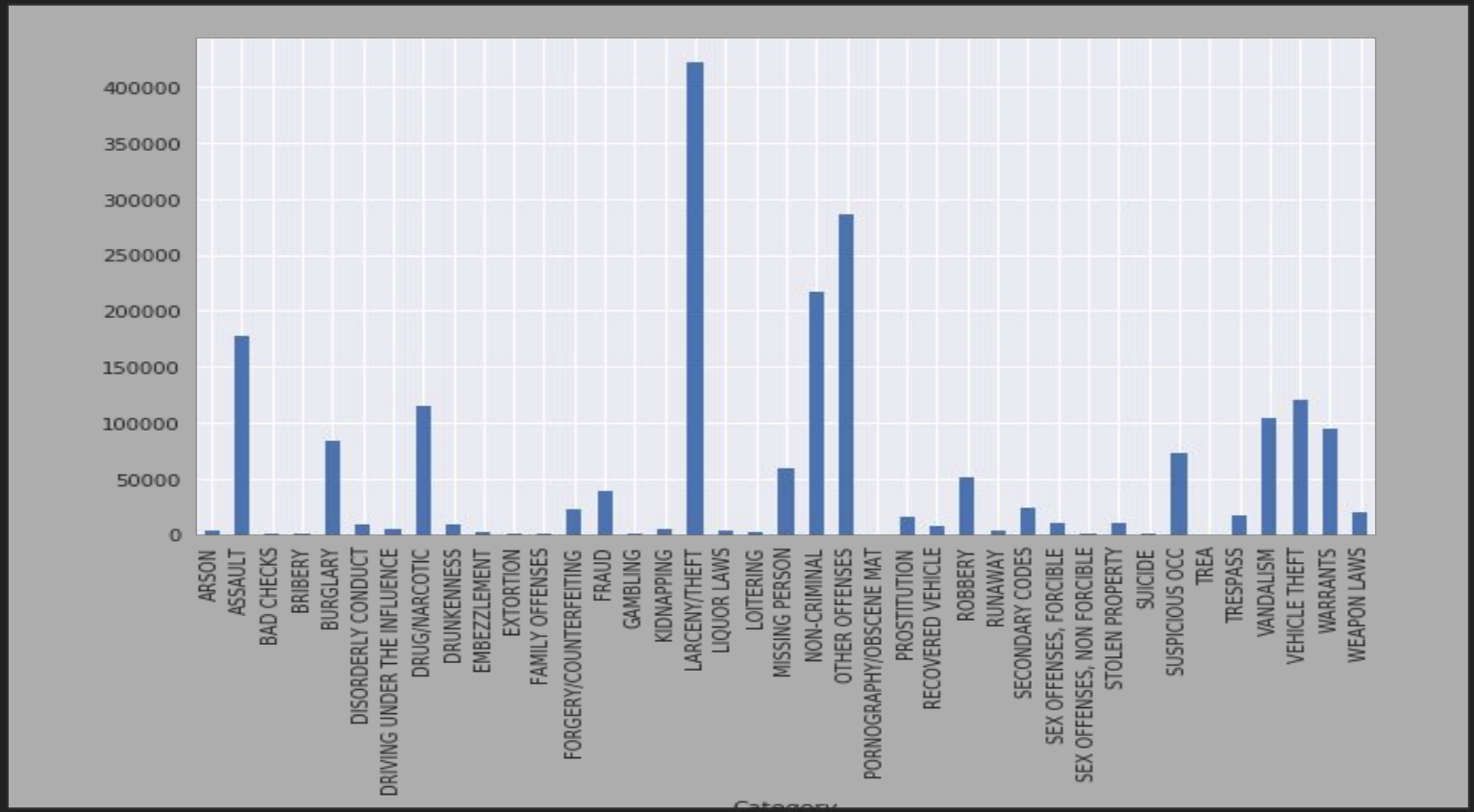
Goals

- Explore Incidents
- Analyze categories
- Predict crimes by categories
- Focusing on violent crimes, non-violent and property related crimes.

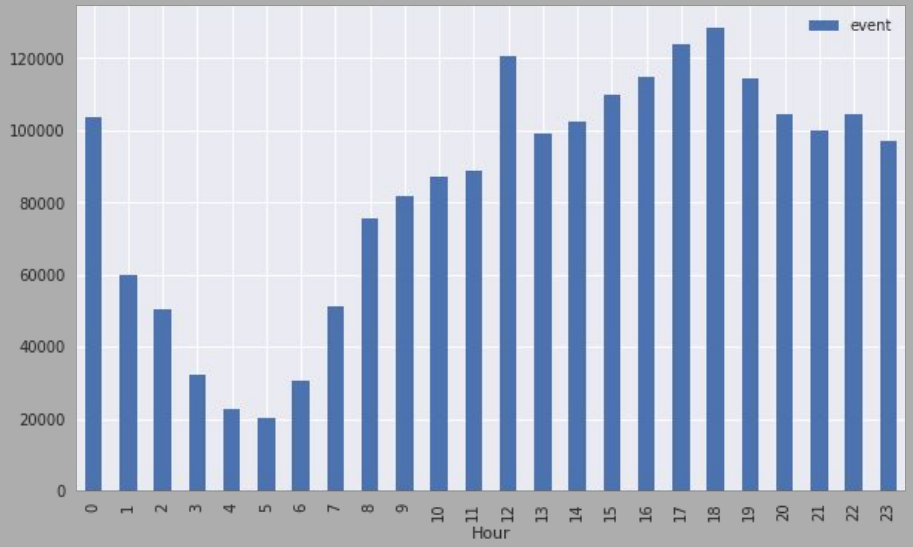
Data/Tools

- Data source: https://data.sfgov.org/
- SFPD crime incidents from January 1st, 2003 until now.
- Amazon AWS
- Python/Pandas
- Basemap
- Xgboost classifier

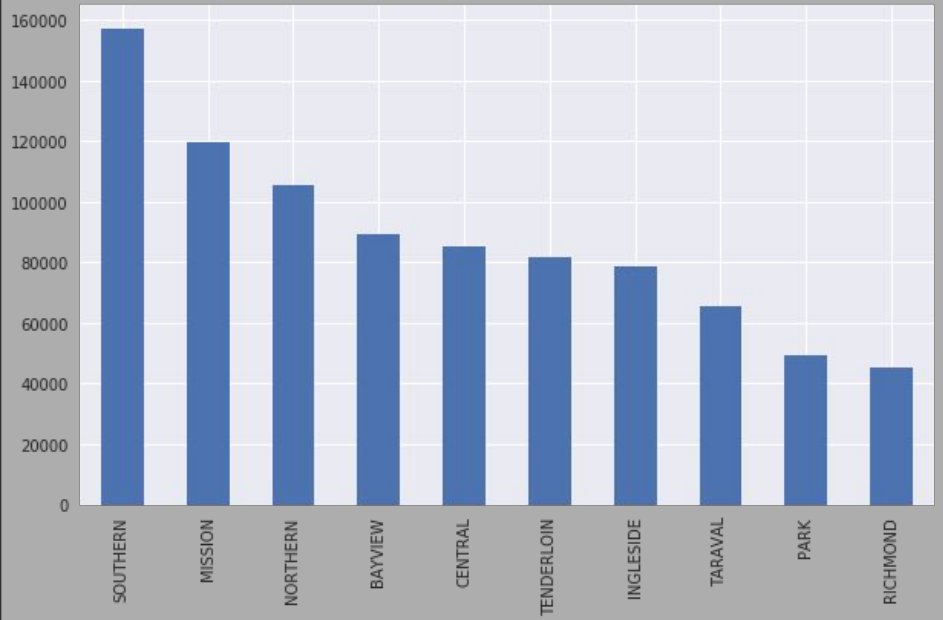
Incidents per category



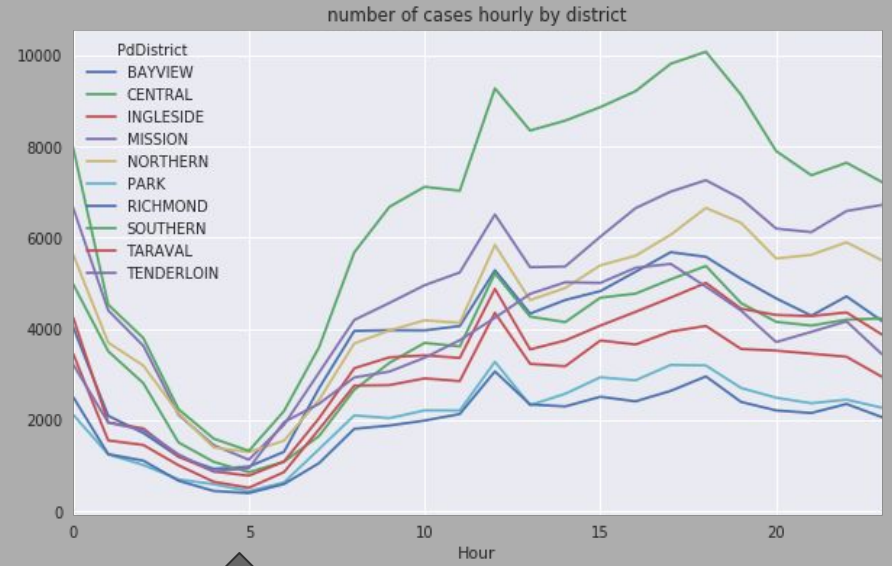
The 2017 line--No Data yet.



Frequency of crimes for neighborhoods



Frequency of crimes over 24 hours by district

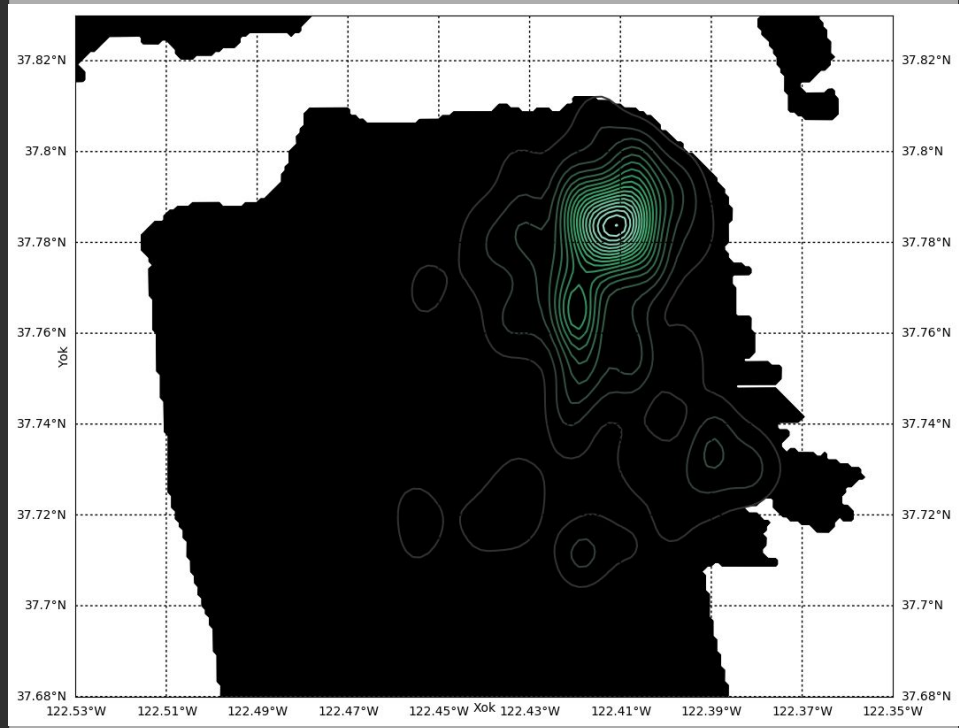


Lowest crime count for all crimes over 24 hours for all data

Frequency Count

Violent Crimes Defined as: Assault, Robbery, Sex Offense, Kidnapping.

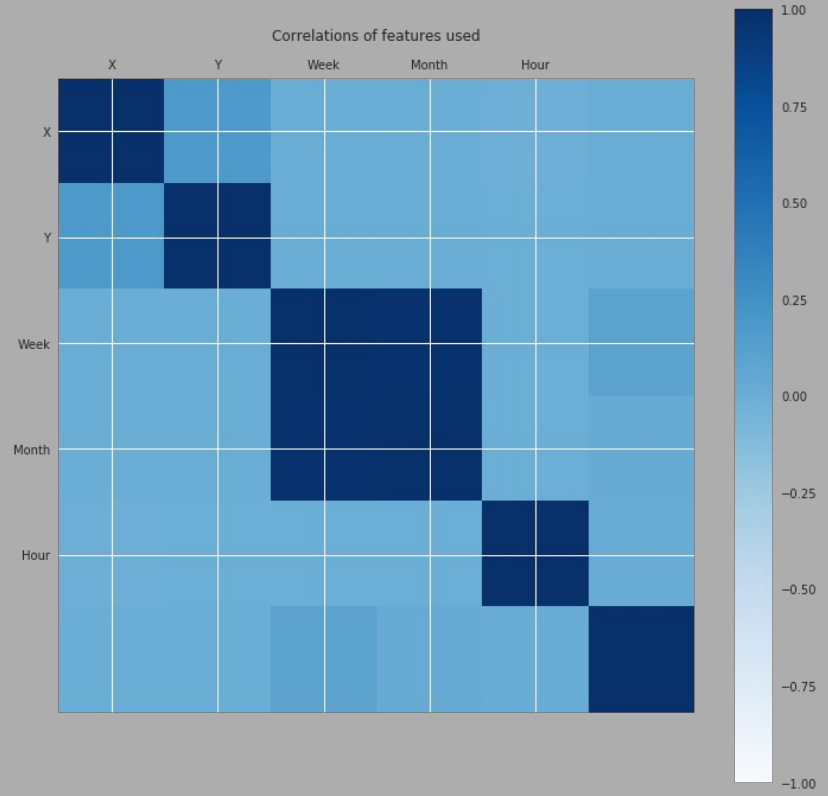
Violent Crimes



Features

Looking at individual events of crime instances:

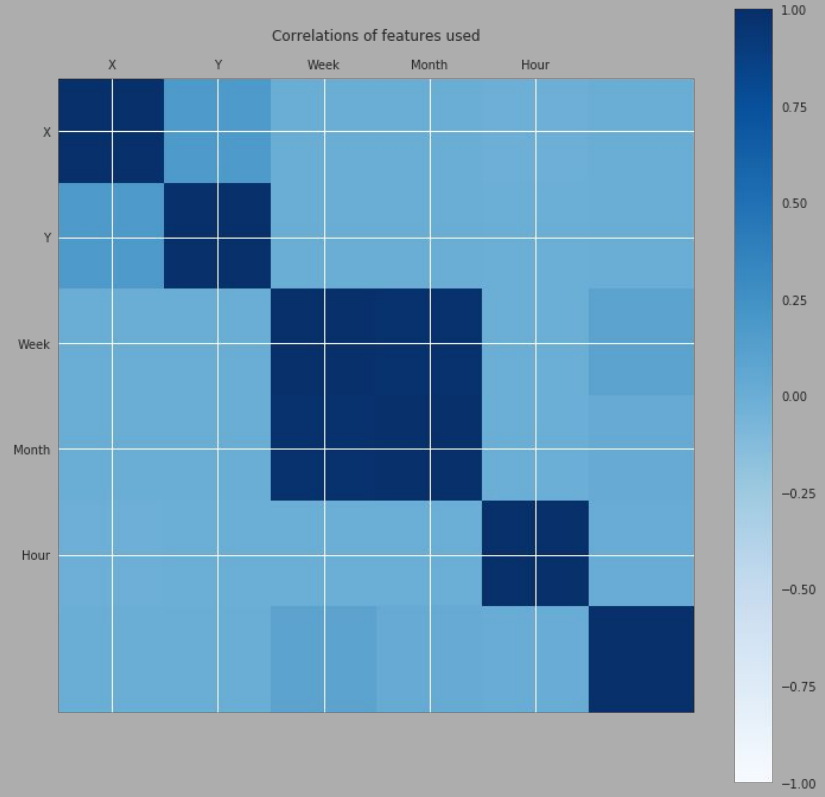
- Longitude
- Latitude
- Hour
- Week
- Month
- Year



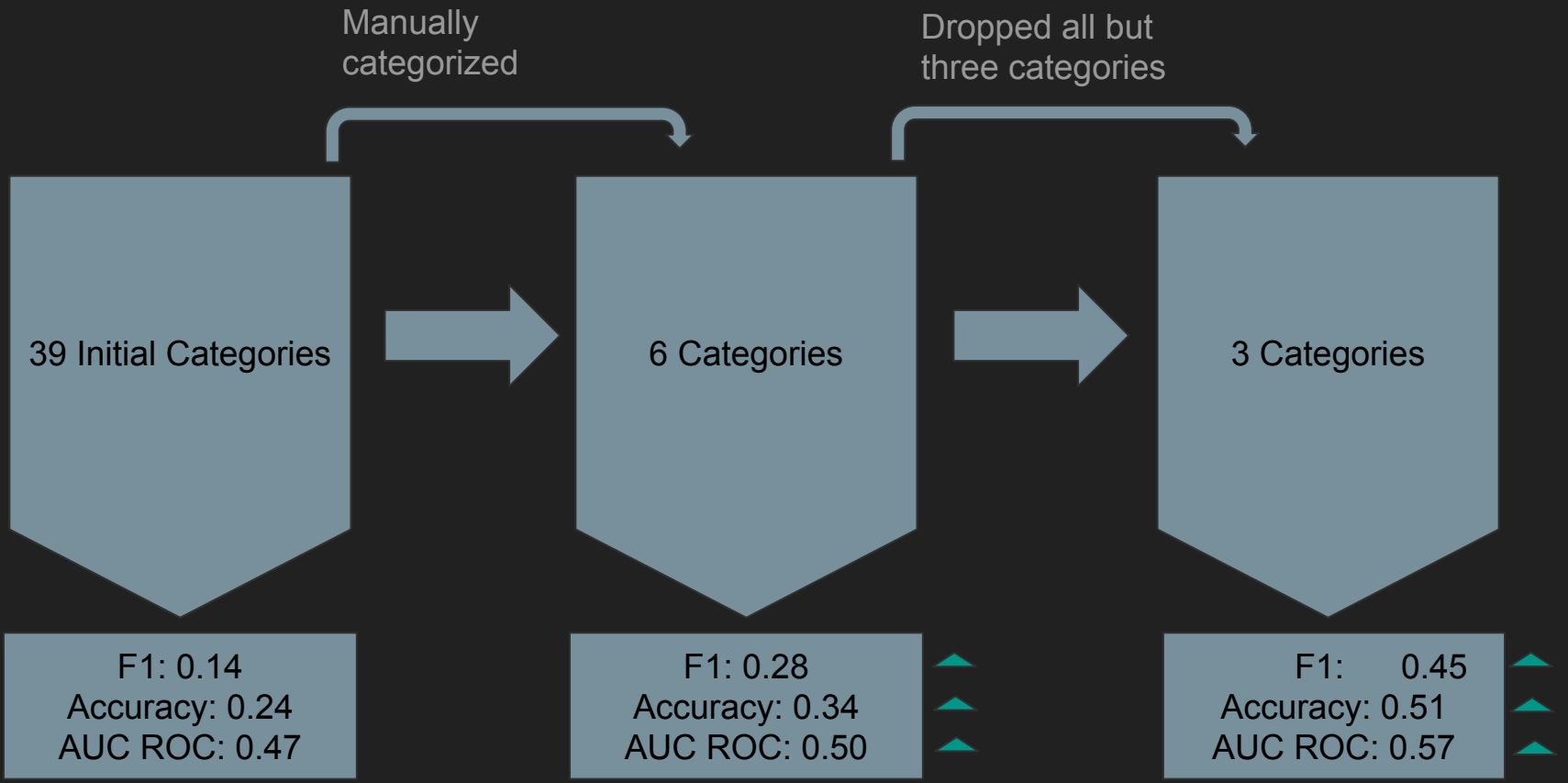
Features

Looking at individual events of crime instances:

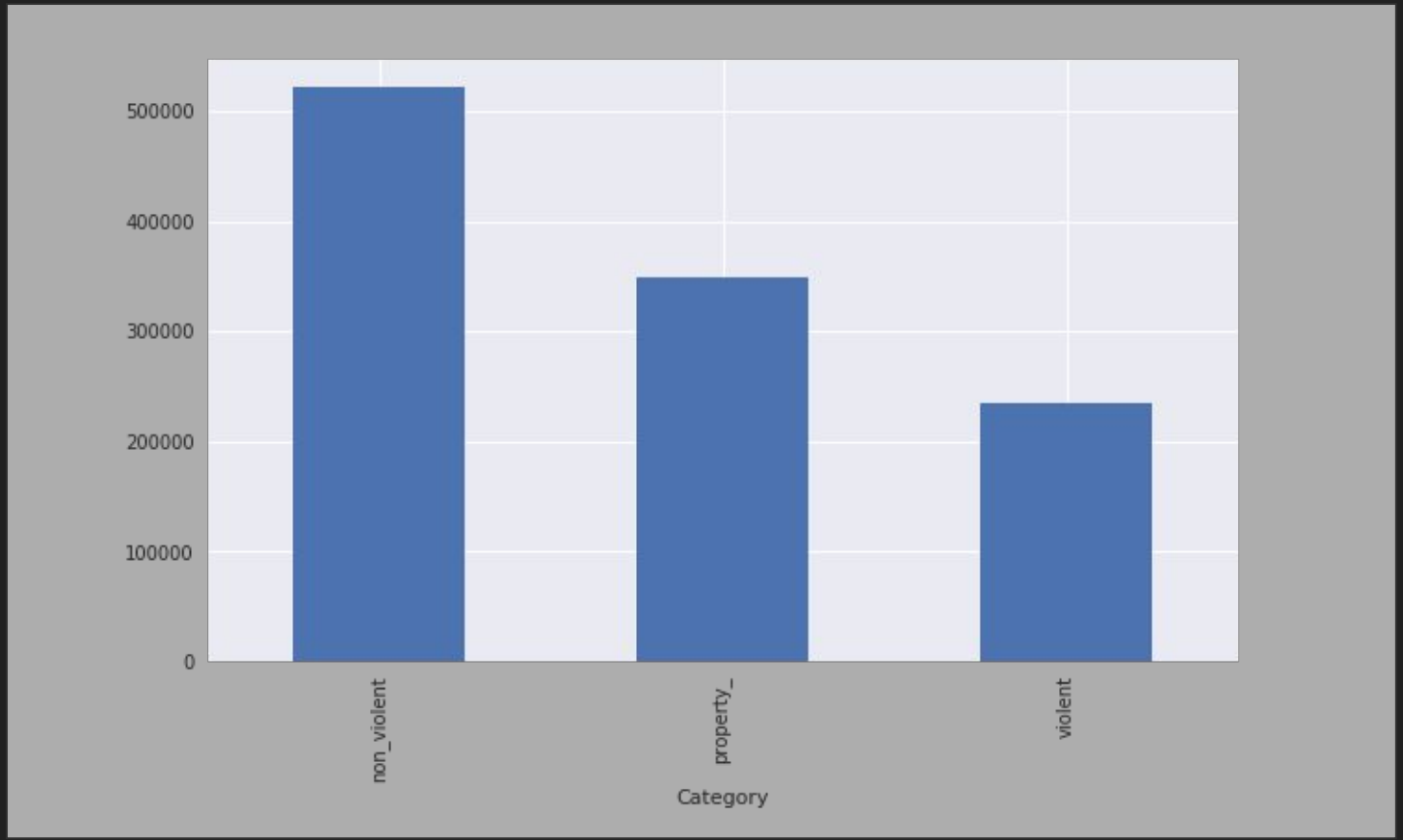
- Longitude ←
- Latitude ←
- Hour ←
- Week ←Not Important...
- Month--Kind of Important...
- Year--Not Important…
- + day!



Category Encapsulating



New Crime Category Count



Predicting

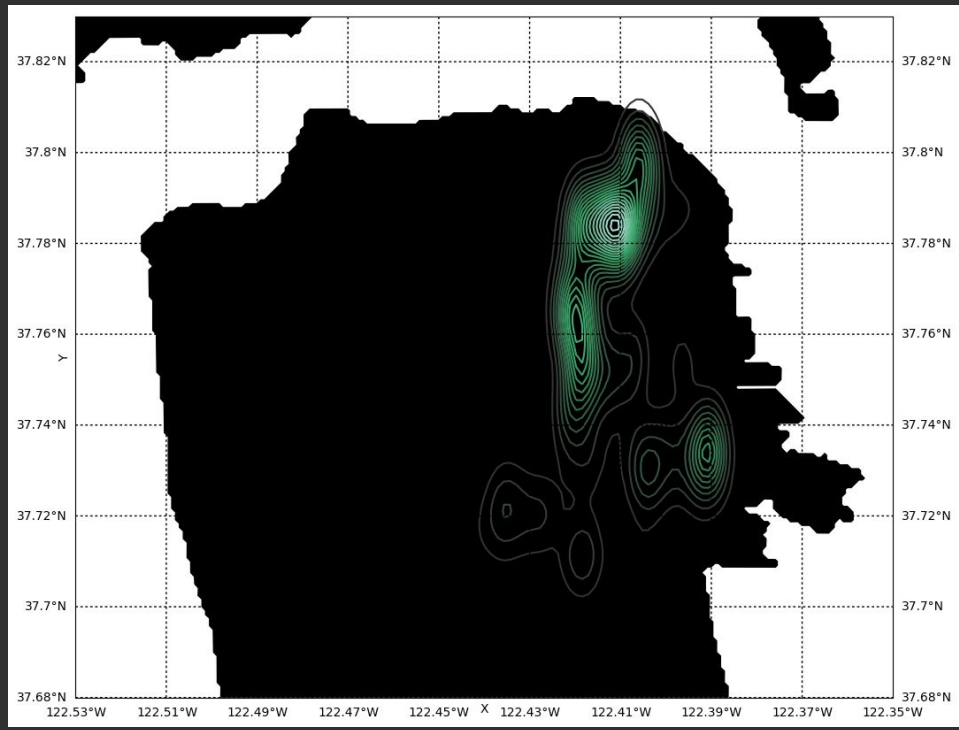
Final model:
xgboost: Extreme Gradient Boosting

Parameters:

N_estimators = 400,

Max_depth = 7

For Violent Crimes*



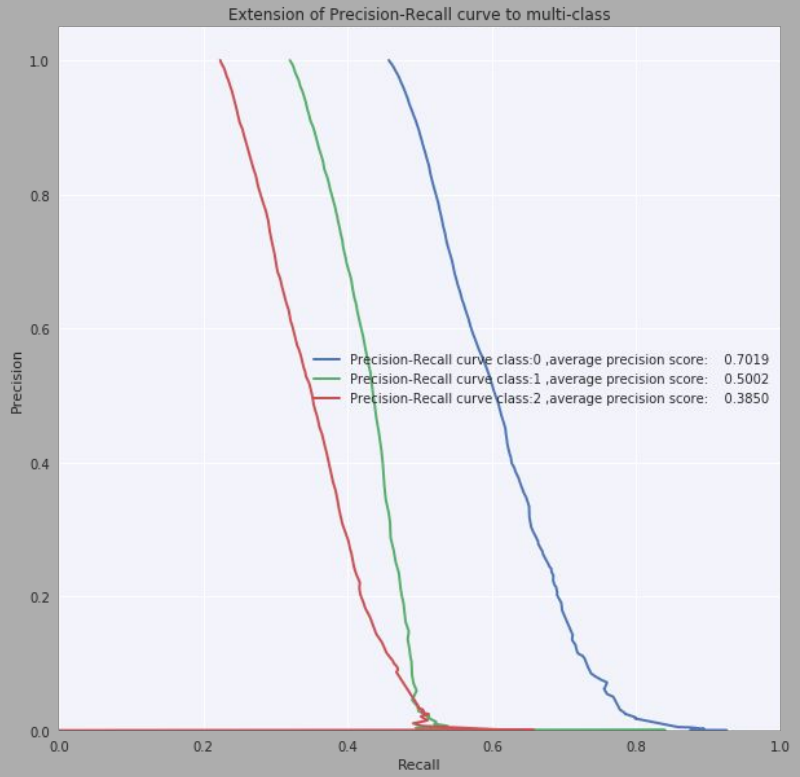
Accuracy Score = 0.497

*Violent crime subcategories at the end

Precision, Recall, and F1

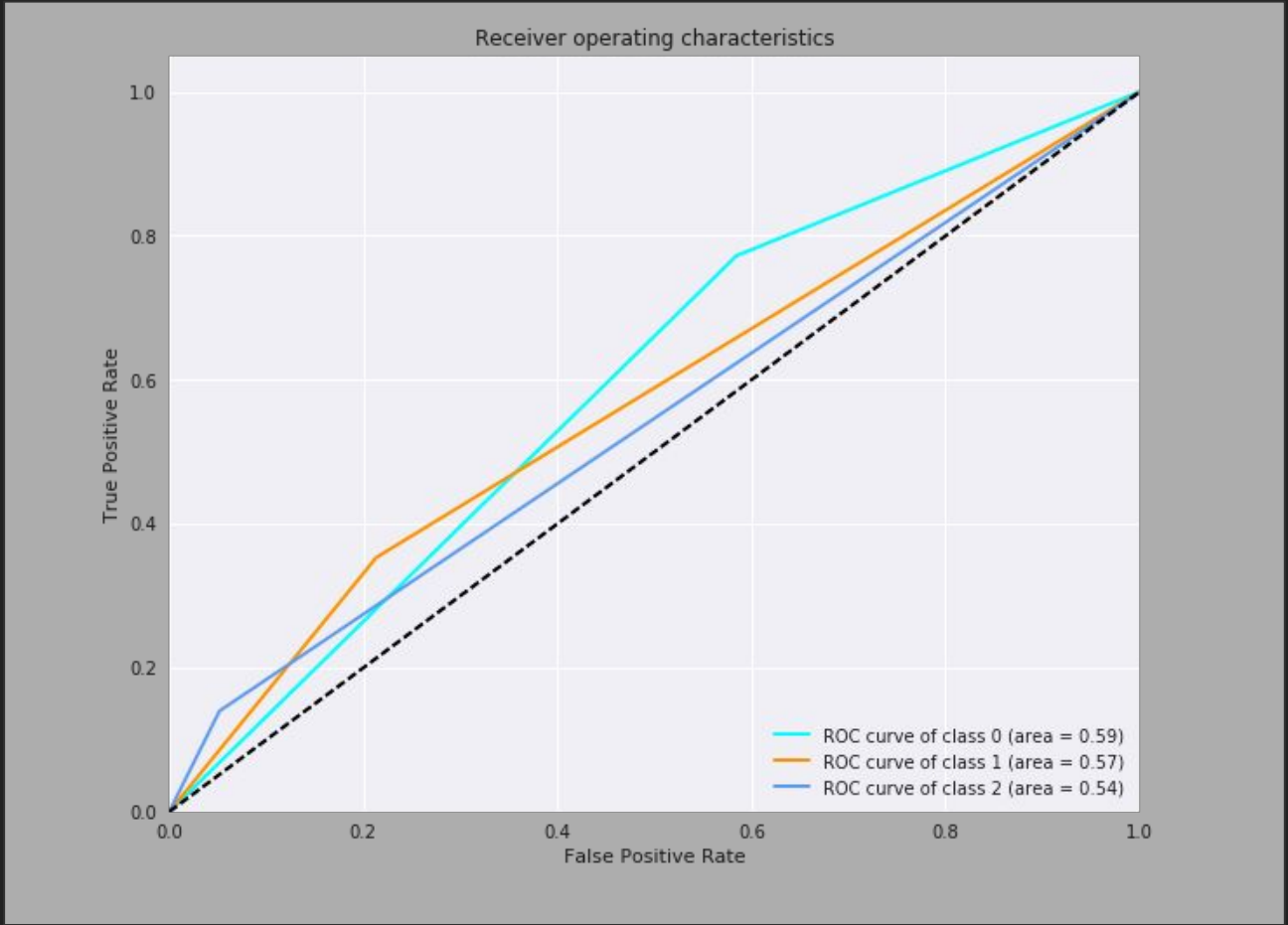
	Precision	Recall	F1-score*
Non-violent:	0.53	0.77	0.63
Property:	0.44	0.35	0.39
Violent:	0.44	0.14	0.21
Avg / Total:	0.48	0.50	0.46

*F1-score = 2 (precision x recall) / (precision + recall)



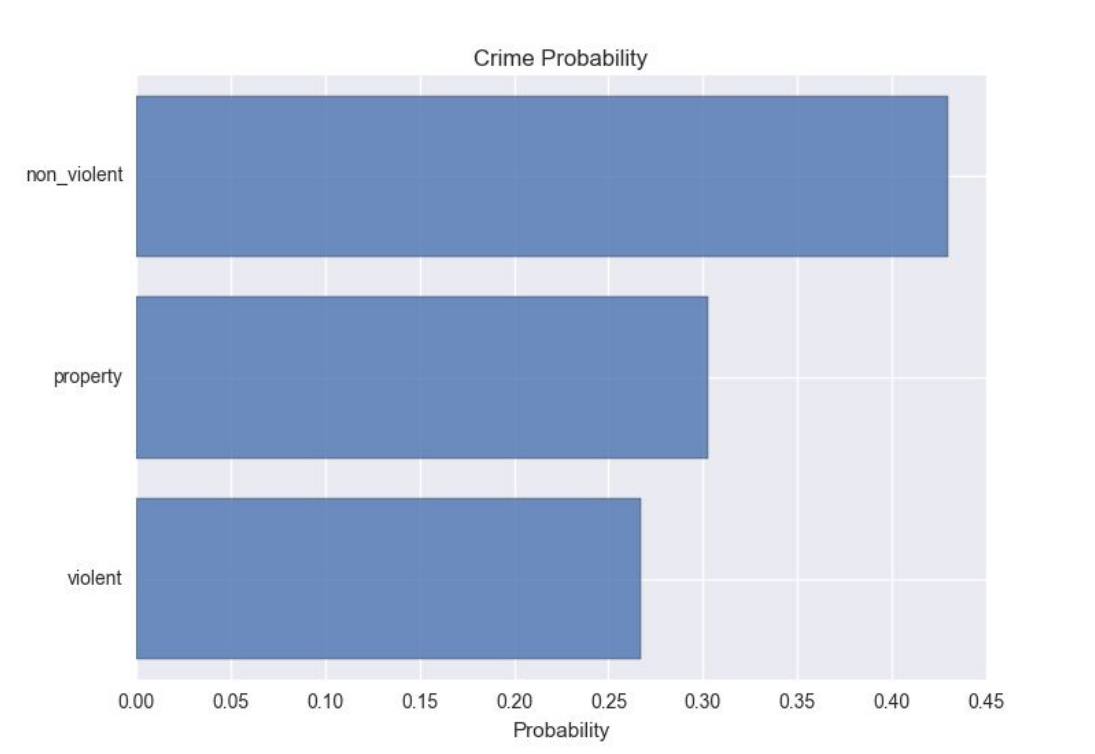
Class 0: non_violent
Class 1: property
Class 3: violent

ROC


Receiver operating characteristics
True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
ROC curve of class 0 (area = 0.59)
ROC curve of class 1 (area = 0.57)
ROC curve of class 2 (area = 0.54)

Web-app

- Build a web app to show location to show crime probabilities for a location and a time.



Baseline Probabilities:

Non-violent: 0.4569
Property: 0.3198
Violent: 0.2232

Example location, Metis:

37.784737, -122.397205

Issues

- Subjective super-categories are subjective and this leads to very different model behaviors
- Dropping sub-categories that were non-essential was also subjective and my model was highly dependent upon this.
- Flat features that didn’t take into account sociological or economic factors.

Conclusions

- Predictions seem sensible, but yield less than desirable metrics.
- Could be enriched with other sources of data
- While my model doesn’t have many results, but correctly labels the categories it does find.

*Example of crime categories

Violent crimes : Assault, Sex offenses, Robbery, Kidnapping

Non-violent crimes: Suicide, Suspicious occupant, Larceny/Theft, Drunkenness, Driving under the influence, disorderly conduct.

Property: Vandalism, Vehicle theft,Trespassing, Stolen Property, recovered vehicle, Arson, Burglary.