San Francisco Crime Analysis and Classification

Dylan Karman

Flatiron School Data Science, Capstone Project

10/19/2020

Problem Statement

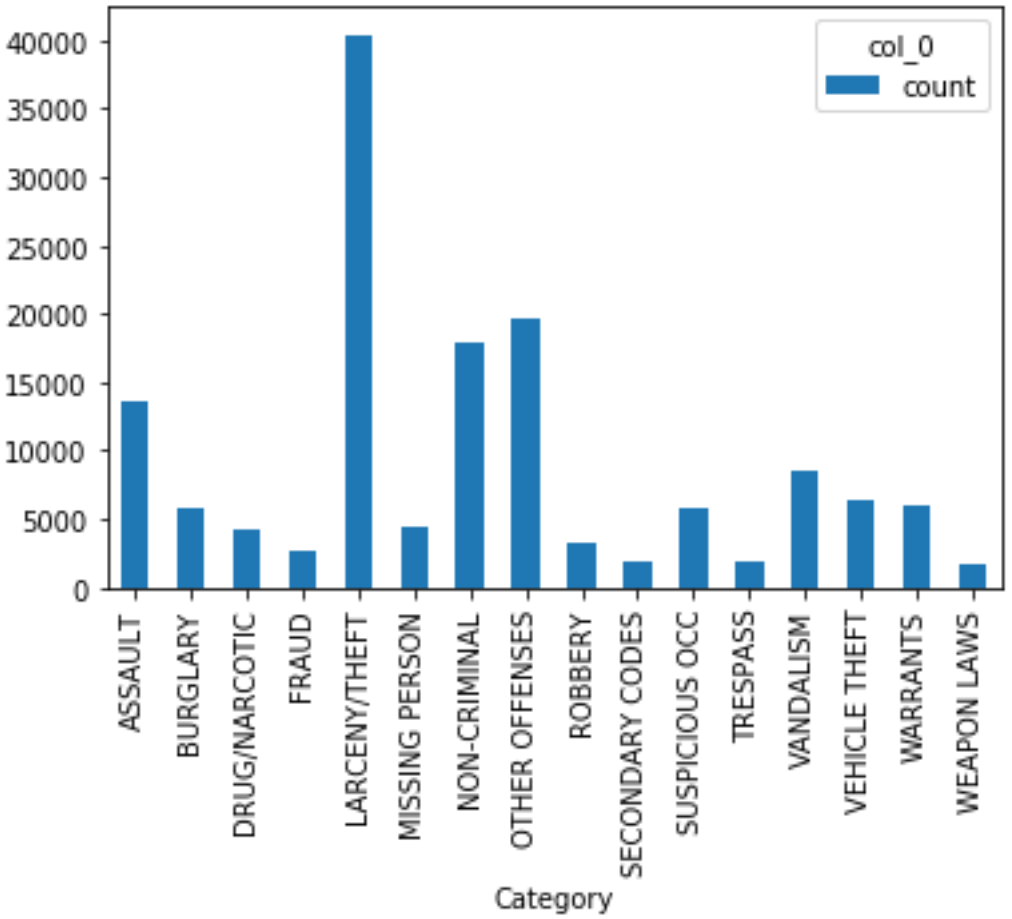
- How can the San Francisco Police Department better allocate their time and resources?
- How can they become more proactive instead of reactive?

Business Value

- Better allocation of resources
- Save time and money
- Prevent Crime
- Make the community safer

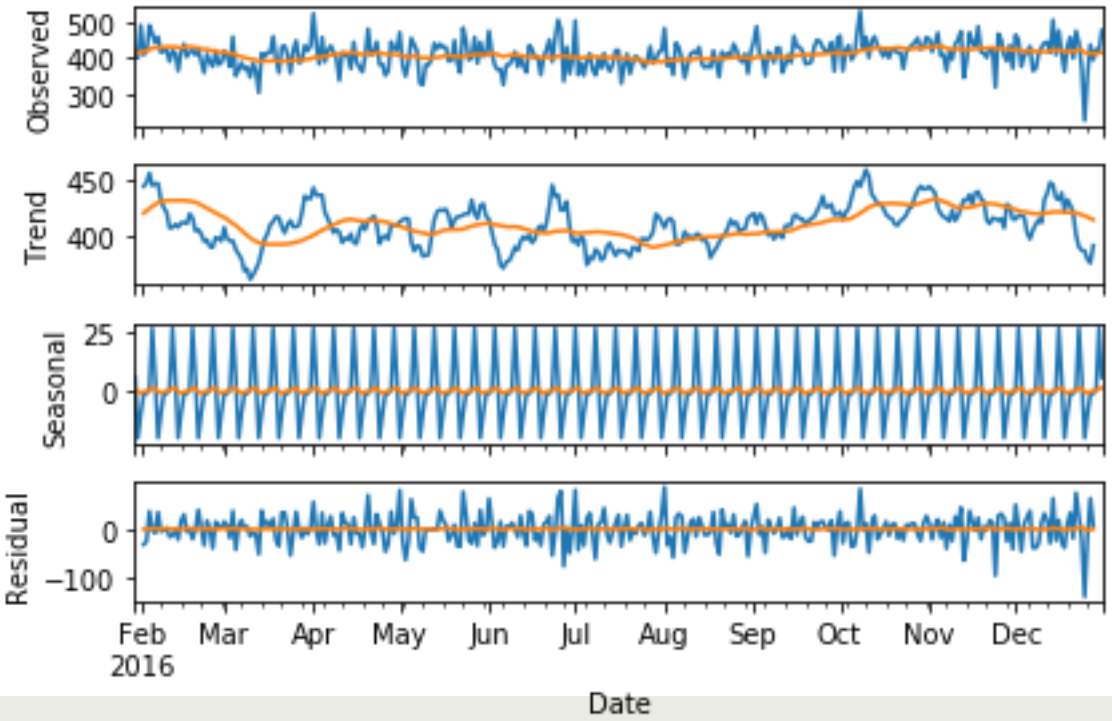
Methodology

- Data Analysis
  - Group and count the different variables
  - Graph results



Methodology

- Seasonal Decomposition
  - Are there any trends in the data?



Methodology

- Classification of Pd District
  - Predicting which district the next crime is going to happen

	precision	recall	f1-score	support
0	0.91	0.93	0.92	4291
1	0.93	0.88	0.90	5300
2	0.89	0.89	0.89	3478
3	0.87	0.89	0.88	5851
4	0.90	0.81	0.85	6030
5	0.78	0.86	0.82	2610
6	0.87	0.89	0.88	2677
7	0.96	0.81	0.88	8533
8	0.92	0.91	0.92	3397
9	0.63	0.98	0.77	2983
micro avg	0.87	0.87	0.87	45150
macro avg	0.87	0.89	0.87	45150
weighted avg	0.89	0.87	0.88	45150

Accuracy Score: 87.3 %

Findings

- **Category**: Larceny/theft
- **Day of Week:** Friday
- **Date:** January 1
- **Time:** 12:00 noon
- **Pd District:** Southern
- **Resolution**: None
- **High Seasonality**

Findings

- Classification
  - 87% accurate in predicting the Pd District

Future Work

- More data from past years
  - More data = more accurate predictions
- Time series analysis

Thank You

- Dylan Karman
- Flatiron School Data Science, Capstone Project
- 10/19/2020