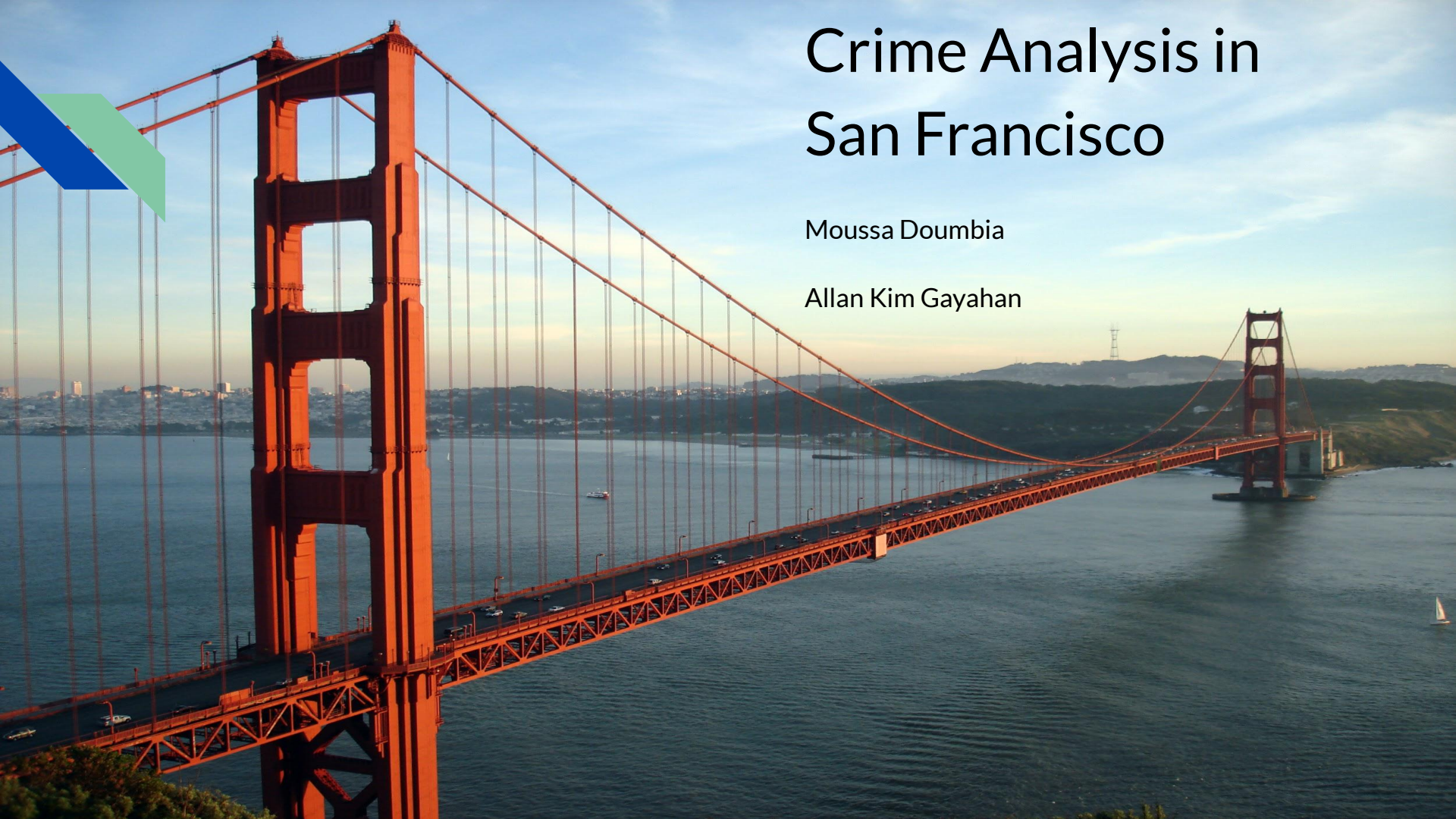
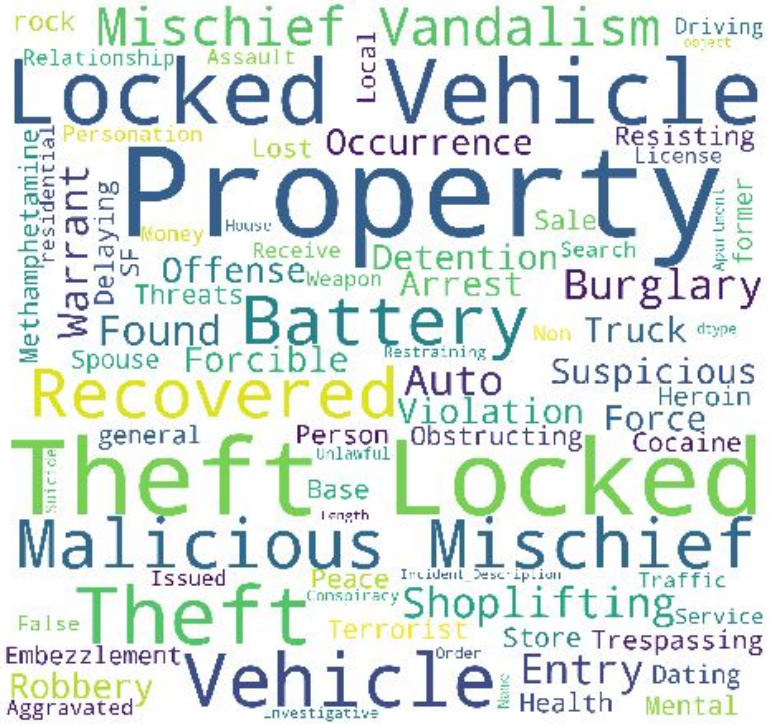


Crime Analysis in San Francisco

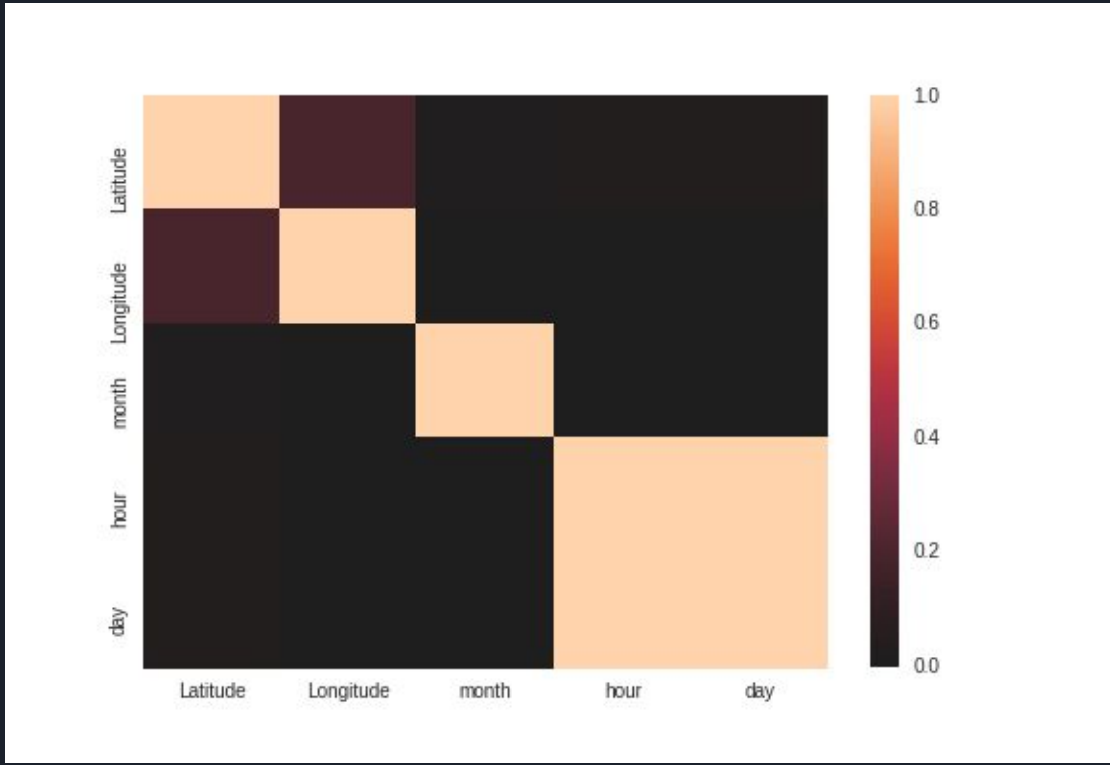
Moussa Doumbia

Allan Kim Gayahan

Top Crimes


rock
Mischief
Vandalism
Driving
Relationship
Assault
Local
Locked
Vehicle
Personation
Lost
Occurrence
Resisting
License
Methamphetamine
residential
Warrant
Delaying
Property
Money
House
Sale
former
SF
Receive
Search
Offense
Detention
Weapon
Arrest
Burglary
Threats
Found
Battery
Non
Truck
dtype
Spouse
Forcible
Restraining
Recovered
Auto
Suspicious
Heroin
Violation
Force
general
Person
Obstructing
Cocaine
Theft
Locked
Unlawful
Suicide
Base
Length
Malicious
Mischief
Issued
Peace
Incident Description
Traffic
Theft
Conspiracy
Shoplifting
Service
False
Terrorist
Store
Trespassing
Embezzlement
Order
Robbery
Vehicle
Entry
Dating
Aggravated
Investigative
Health
Mental

Is clean-ready data good to use?



Features

Latitude

Longitude

Day of the Month

Day of the Week

Hour of the Day

Incident Description

Resolution

Target

Open or Active

Cite or Arrest Adult

Cite or Arrest Juvenile

Unfounded

Exceptional Adult

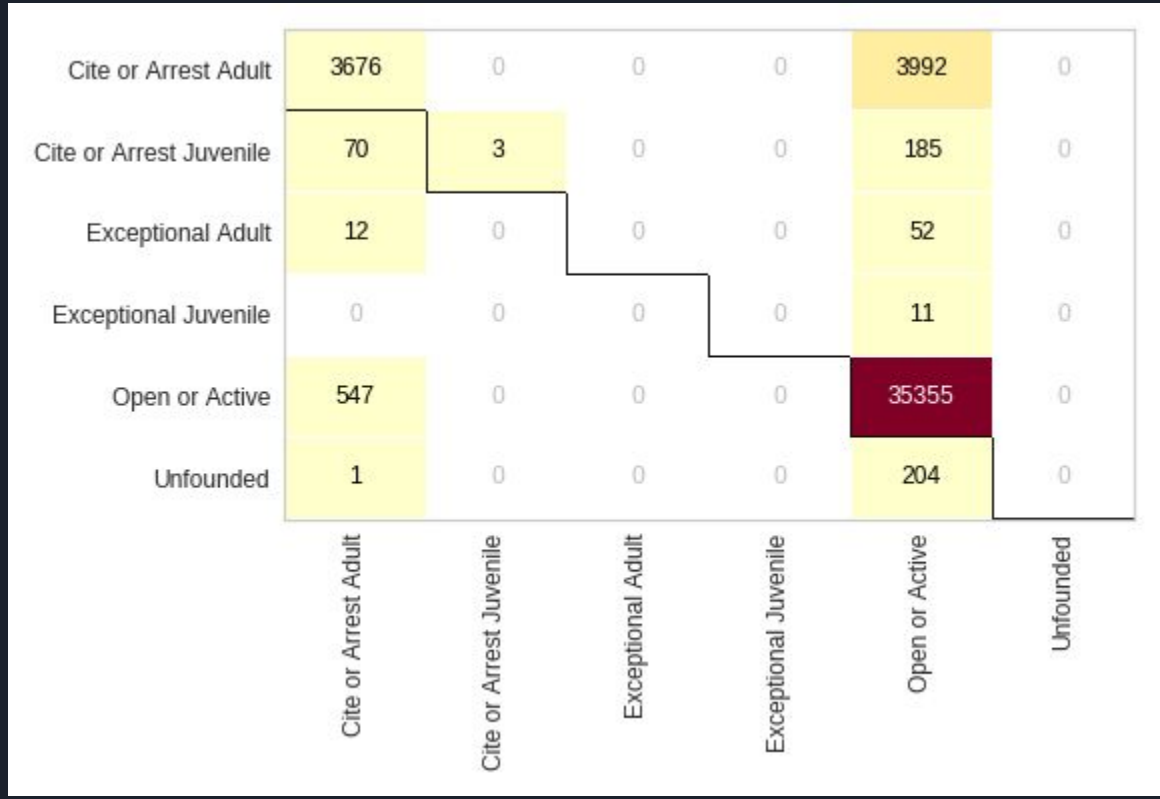
Exceptional Juvenile

Random Forest

Max_Depth = 100

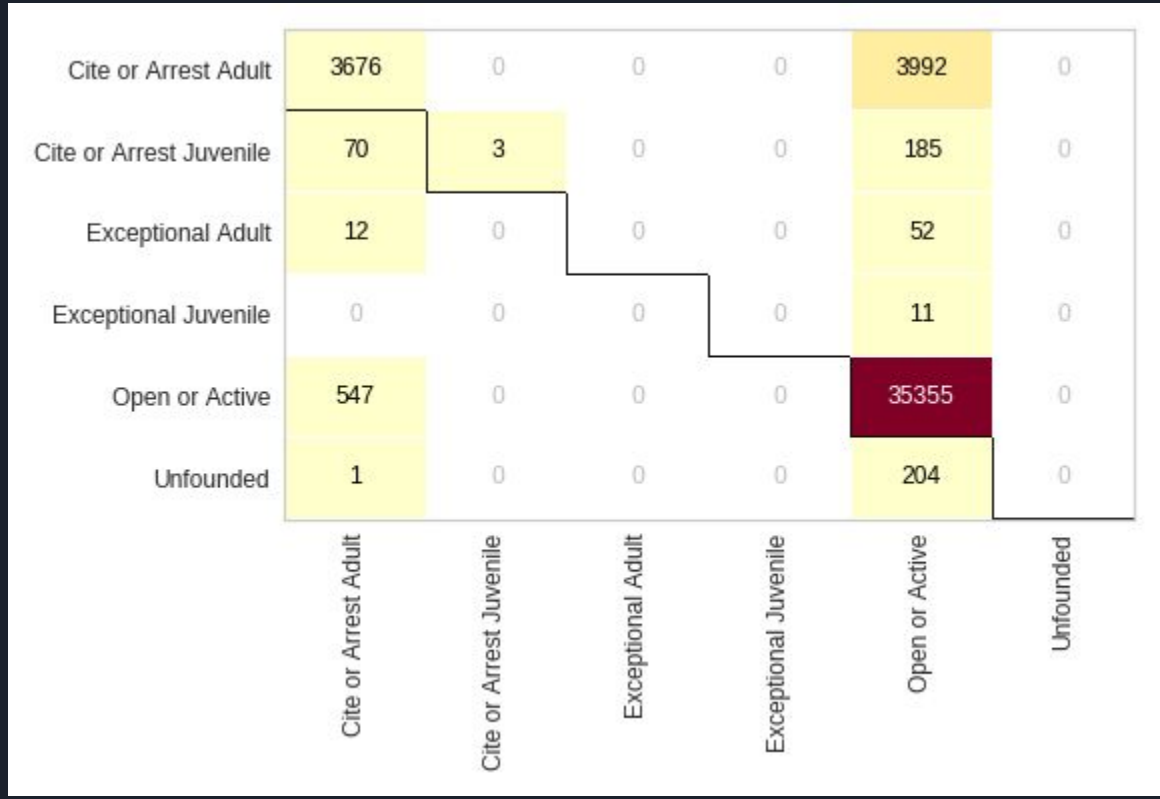
Estimators = 100

Accuracy: 88.49%



Random Forest

Open or Active	0.813957
Cite or Arrest Adult	0.173846
Cite or Arrest Juvenile	0.005849
Unfounded	0.004648
Exceptional Adult	0.001451
Exceptional Juvenile	0.000249



Weighted Class RF

Open or Active	0.813957
Cite or Arrest Adult	0.173846
Cite or Arrest Juvenile	0.005849
Unfounded	0.004648
Exceptional Adult	0.001451
Exceptional Juvenile	0.000249

Accuracy: 88.70

Weighted Class RF w/ GS

random_state = 24

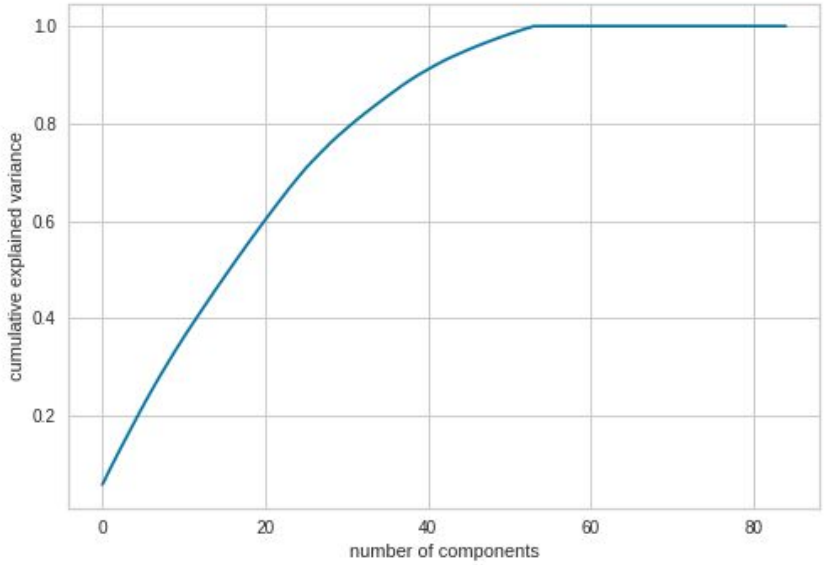
criterion= 'gini'

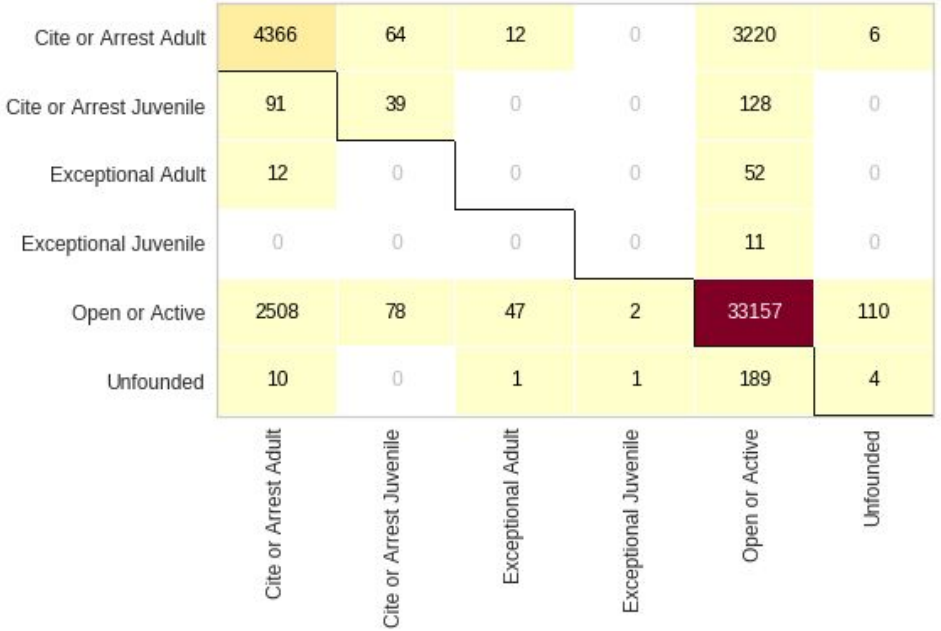
max_depth= 20

n_estimators= 1000

Accuracy: 87.88

Weighted Class RF w/ PCA





Accuracy: 85.17

Conclusion

A lot faster and more realistics with PCA

Grid Search was not a viable option

Dilemma between High Accuracy and Realistic

Future Exploration/Improvement

Over/Undersampling of target classes

Clustering

Geographical Classification/Time Series

Data Source

Kaggle

https://www.kaggle.com/psmavi104/san-francisco-crime-data