ANALYSIS OF TERRY STOPS IN SEATTLE

- By: Melody Peterson
- Flat Iron School Data Science Project

January 4, 2021

“

A Terry stop in the United States allows the police to briefly detain a person based on reasonable suspicion of involvement in criminal activity. Reasonable suspicion is a lower standard than probable cause which is needed for arrest. When police stop and search a pedestrian, this is commonly known as a stop and frisk. When police stop an automobile, this is known as a traffic stop. If the police stop a motor vehicle on minor infringements in order to investigate other suspected criminal activity, this is known as a pretextual stop. - Wikipedia

WHAT CAN/CAN'T THIS DATA TELL US

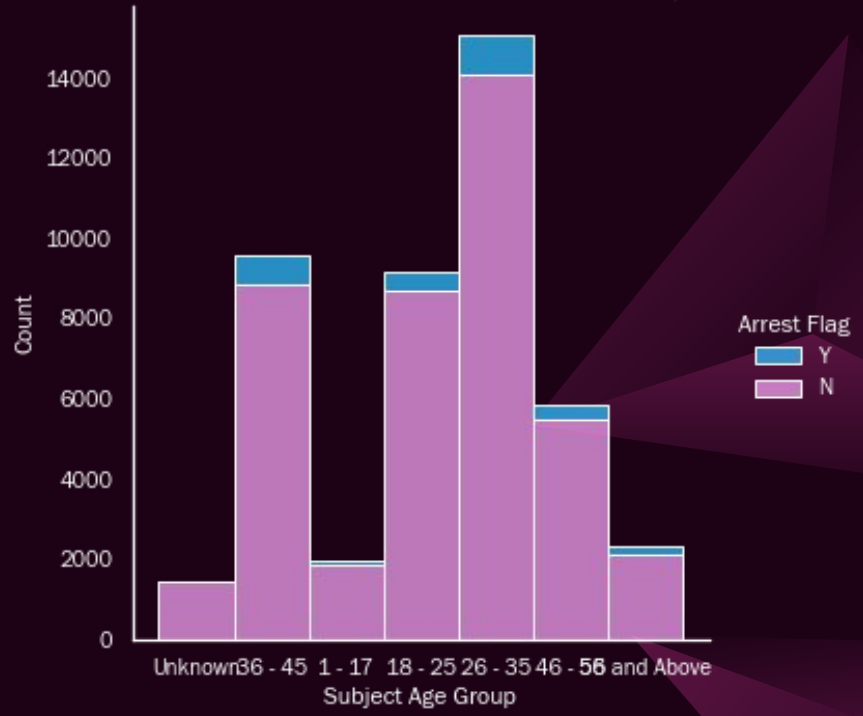
- Modeling done for inference only
- Predictions would perpetuate any bias
- Accuracy measurement rather than predictive ability on unseen data

DATASET INFORMATION

- This data represents records of police reported stops under Terry v. Ohio, 392 U.S. 1 (1968).
- The dataset was created on 04/12/2017 and first published on 05/22/2018 and is provided by the city of Seattle, WA.
- There were 45,317 rows and 23 variables
- Classification target is ‘Arrest Flag’
- Initial ‘Arrest Flag’ distribution ‘N’ - 42585, ‘Y’ - 2732

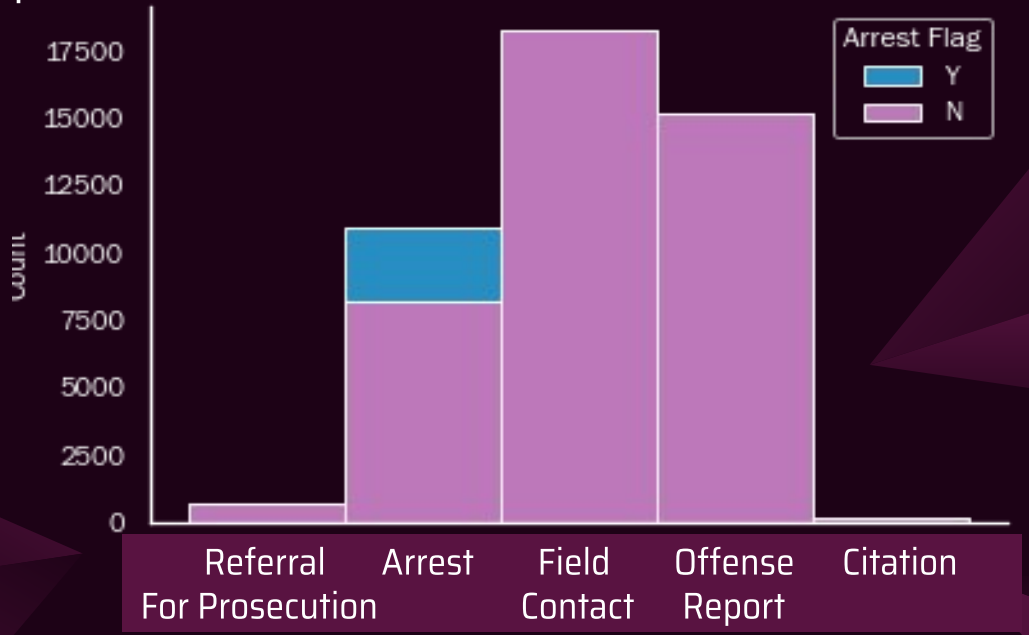
EXPLORATORY DATA ANALYSIS

Subject Age Group



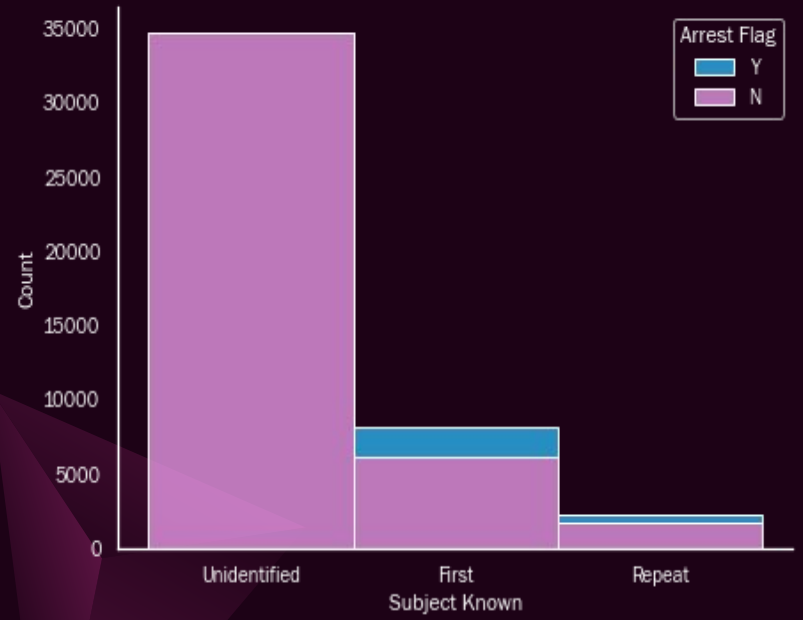
EXPLORATORY DATA ANALYSIS

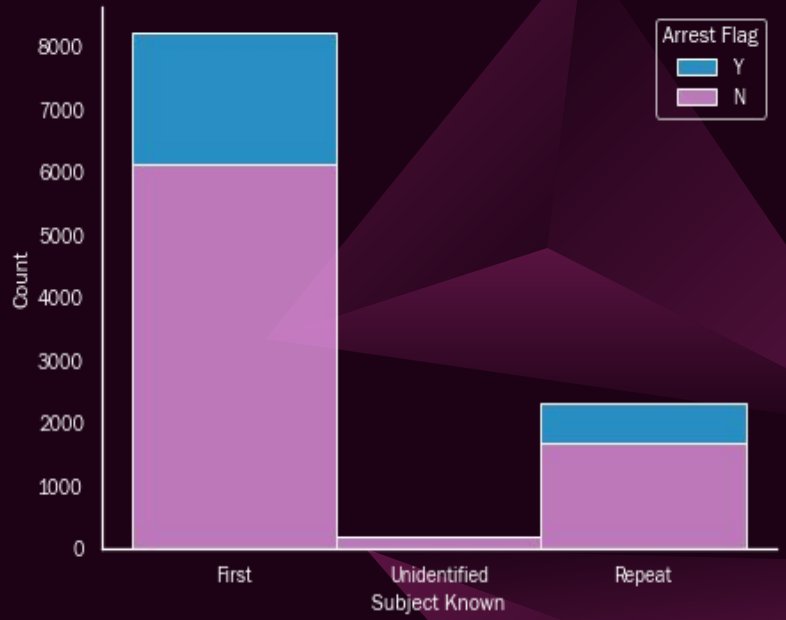
Stop Resolution



EXPLORATORY DATA ANALYSIS

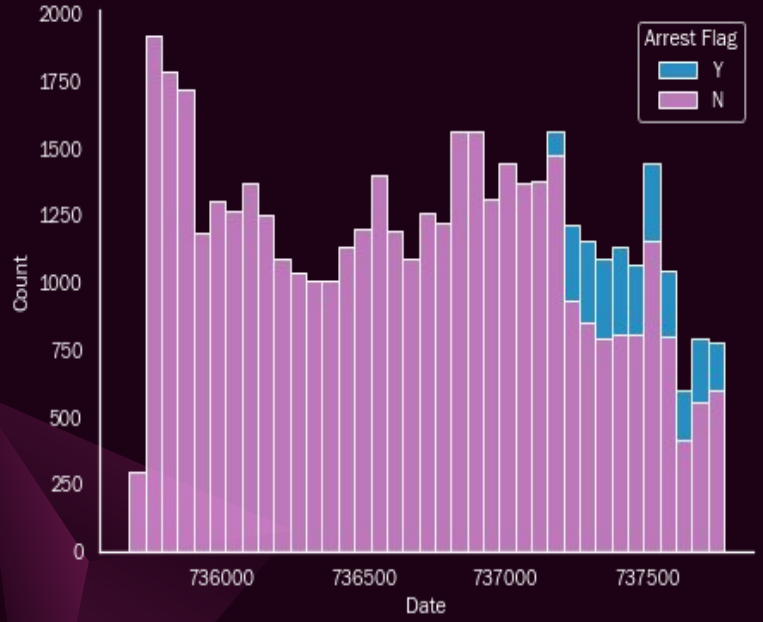
Subject ID

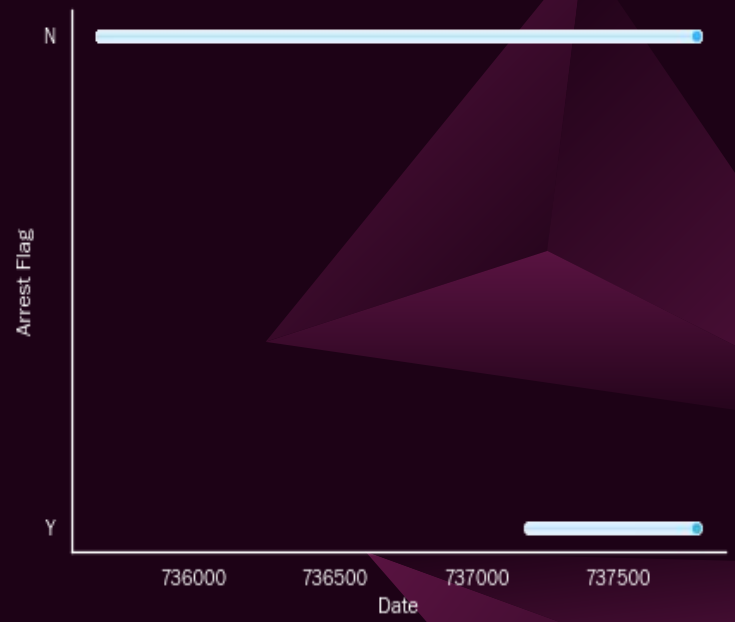




EXPLORATORY DATA ANALYSIS

Date

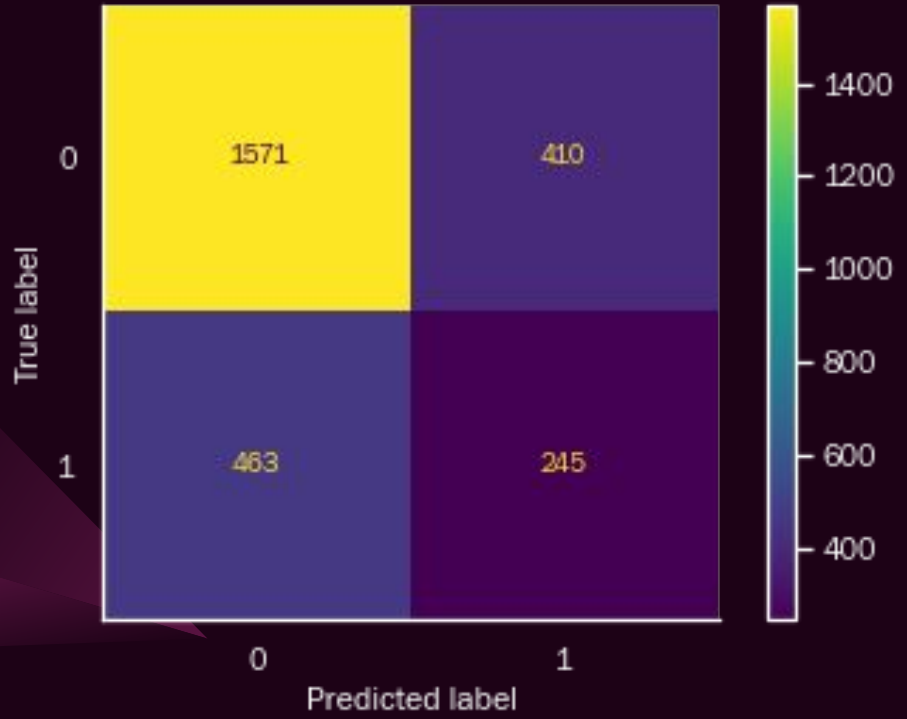




MODELS: XGBOOST AND RANDOM FOREST

	Baseline Model	Random Forest	XGBoost
Train Accuracy	0.497707	0.825958	0.992810
Test Accuracy	0.507996	0.666791	0.675344
Train F1 Score	0.334210	0.664916	0.985686
Test F1 Score	0.351153	0.382069	0.359501

MISCLASSIFIED DATA



Test accuracy 67% means
32.46% of data misclassified

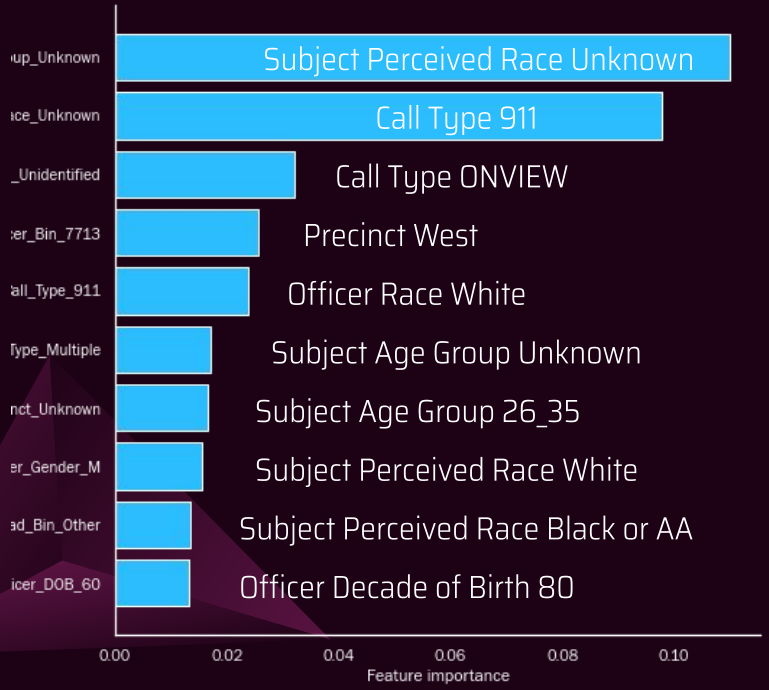
Of 708 arrests, 35% were
classified as arrests

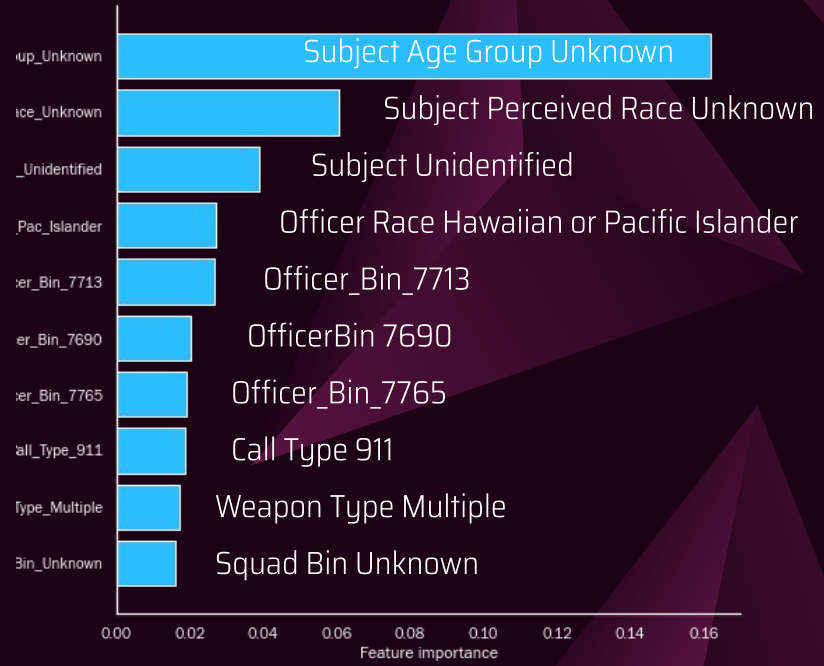
True negatives classified
correctly the majority of the time

410 false positives
463 false negatives

245 true positives
63% of positive misclassified

MODEL DETAILS AND PARAMETERS





SUMMARY AND CONCLUSIONS

- Analysis of misclassified data
- And some text
- But remember not to overload your slides with content

Your audience will listen to you or read the content, but won't do both.

NEXT STEPS

Better Understand Data

Further Analyze unknown or missing values

Update ‘Arrest Flag’ with arrest values from ‘Stop Resolution’

Update model parameters

Try no SMOTE

Tune Support Vector Classification

THANKS!

Any questions?

github.com/melodygr

LinkedIn

melodygr@aol.com