

TERRY STOP ARREST CLASSIFICATION

BY RACHEL SPIRO

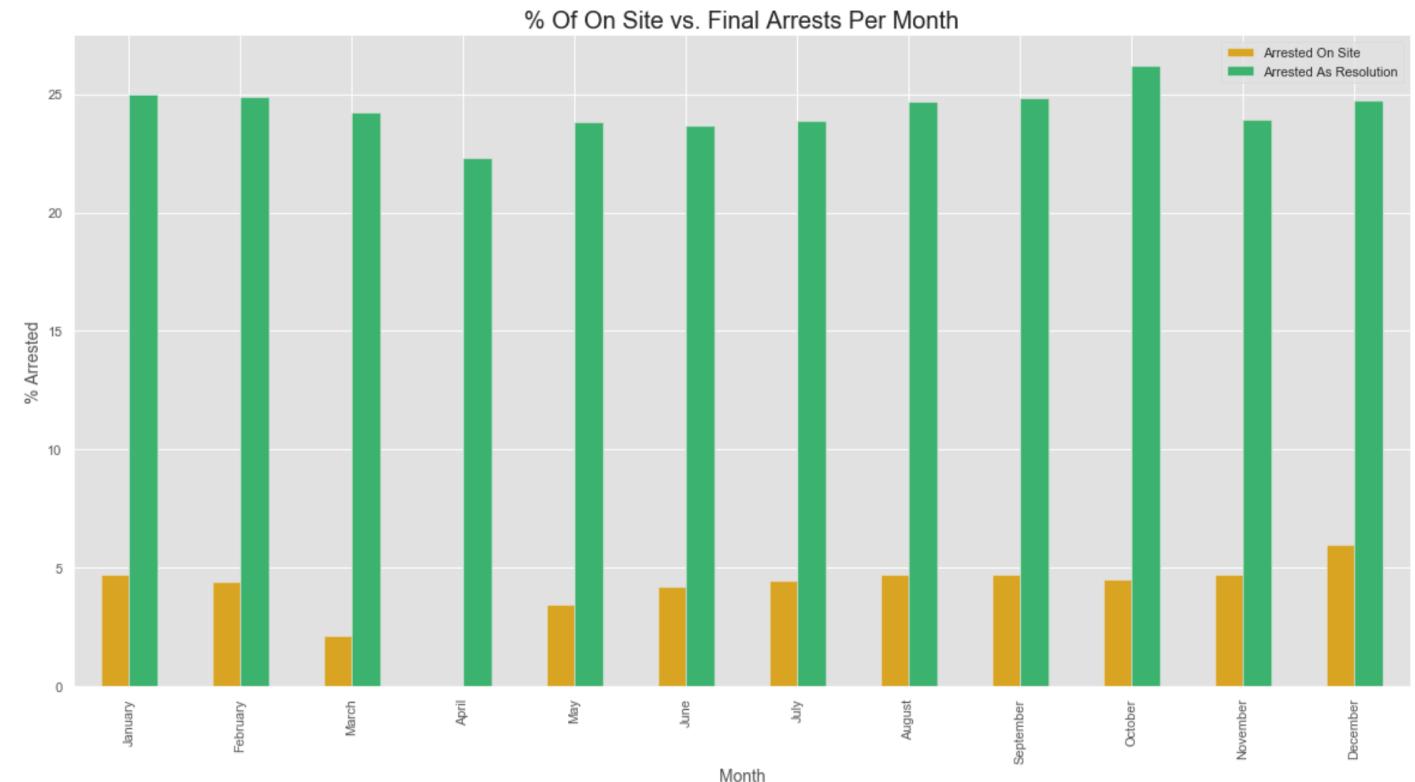
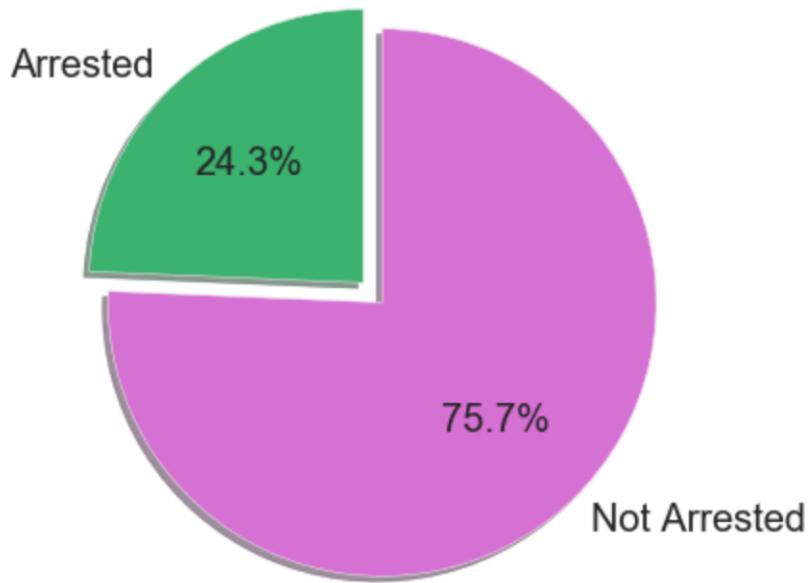


Objectives

- ▶ **The Dataset** – Database containing information on every terry stop made by Seattle police officers (i.e. weapon type, if an arrest was made, reported date, officer precinct, etc.)
- ▶ **Terry Stop** - When a police officer stops a person based on 'reasonable suspicion' that the person may have been involved in criminal activity.
- ▶ **Task** – Utilize the data to predict if an arrest is made as a final resolution to the Terry Stop
- ▶ **Solution** – [Classification](#) – A technique used to identify which set an observation belongs to. We will look at 4 different classification models to find the best performing option.
- ▶ **Importance** – This research can be used to help officers better prepare for what to expect during a terry stop based on certain variables that are present.

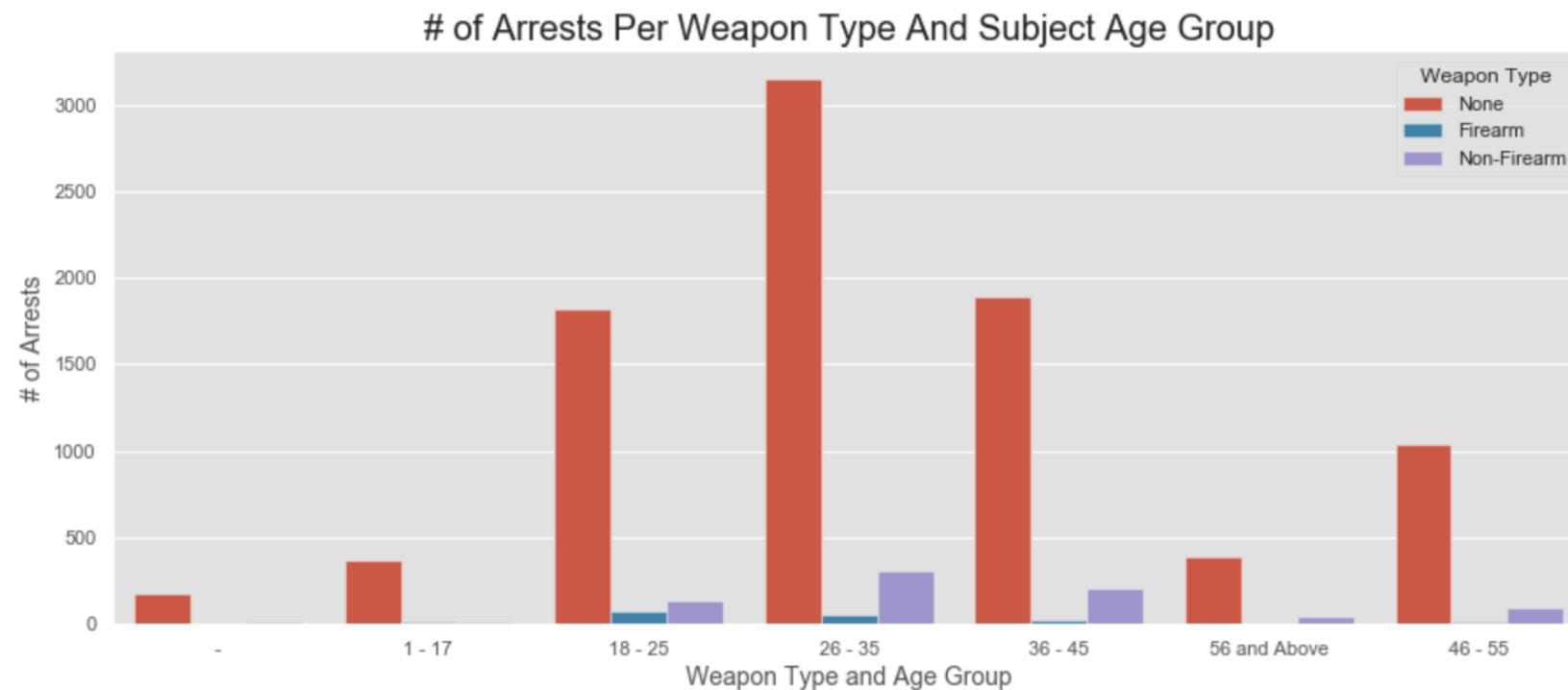
Background - Arrests

- ~25% of terry stops end in an arrest. However, only ~5% of arrests happen during the terry stop, while the rest are resolutions decided later on.



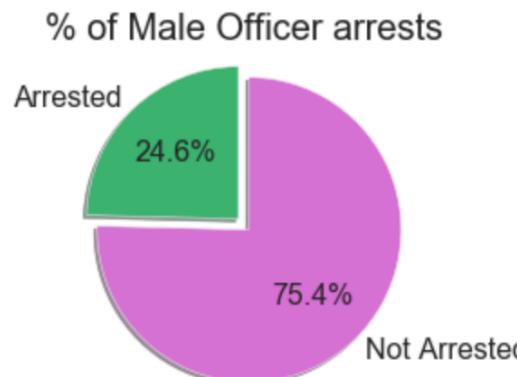
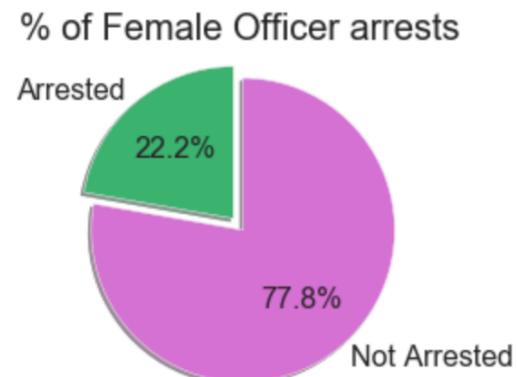
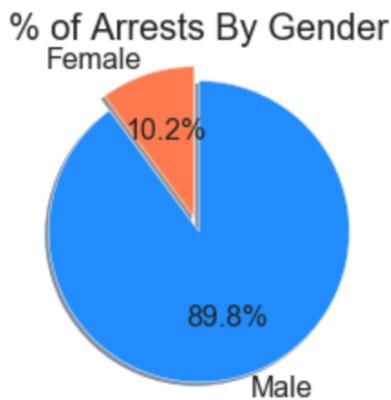
Background - Subjects

- Most subjects are in the 26-35 age range, followed by 18-25 and 36-45. While the majority of Terry Stops do not find a weapon present, it is more likely for a weapon to be found if the subject is in these age ranges.



Background - Officers

- ▶ While there are many more male officers, the percentage of arrests made is similar for each gender.
- ▶ The West, East, and South precinct appear to have the greatest percentages of arrests made from a terry stop.

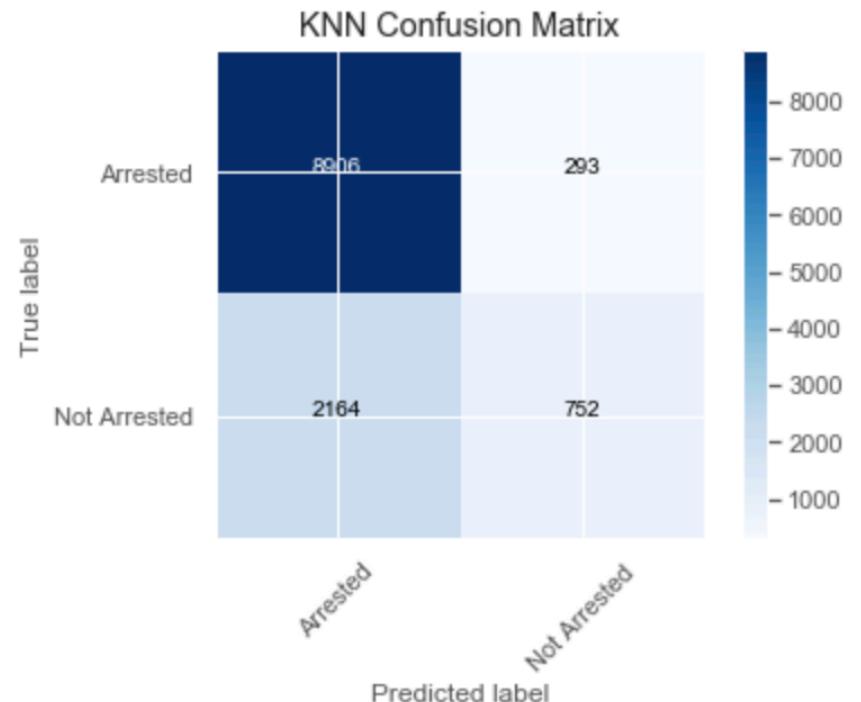


% Male Officers – 89%
% Female Officers – 11%

	# of Terry Stops	% Arrested	% Not Arrested
-	9432	1.887193	98.112807
East	5417	34.539413	65.460587
FK ERROR	12	8.333333	91.666667
North	9057	25.902617	74.097383
OOJ	16	6.250000	93.750000
South	4828	32.145816	67.854184
SouthWest	495	27.474747	72.525253
Southwest	2320	23.836207	76.163793
Unknown	200	25.500000	74.500000
West	8605	36.513655	63.486345

#1 – K-Nearest-Neighbor

- ▶ The K-Nearest Neighbors (KNN) technique is able to generate a prediction for a data point by finding the k-nearest data points and then predicting the majority class of these k points. This model assumes that the smaller the distance between 2 points, the more similar they are.
- ▶ Confusion Matrix – A table used to visually describe the performance of a model (i.e. how accurate our model is).
- ▶ KNN Accuracy = 79.72%



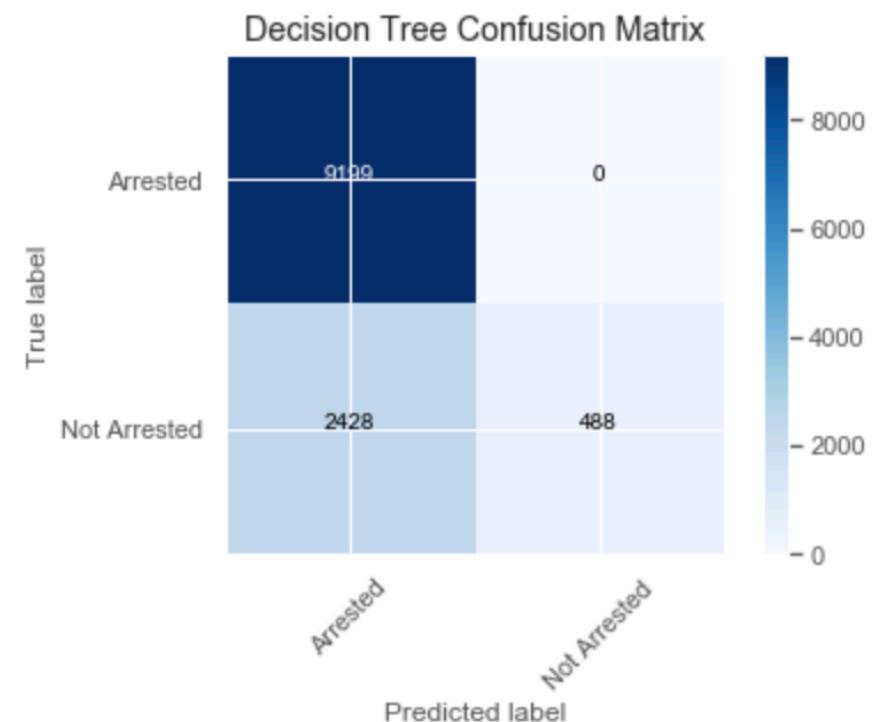
#2 – Logistic Regression

- ▶ Regression is a technique used to predict a response variable from one or more predictor variables.
- ▶ Logistic regression is used when the response variable is binary, and we get a sigmoid function (S-shaped).
- ▶ Logistic Regression Accuracy = 79.92%



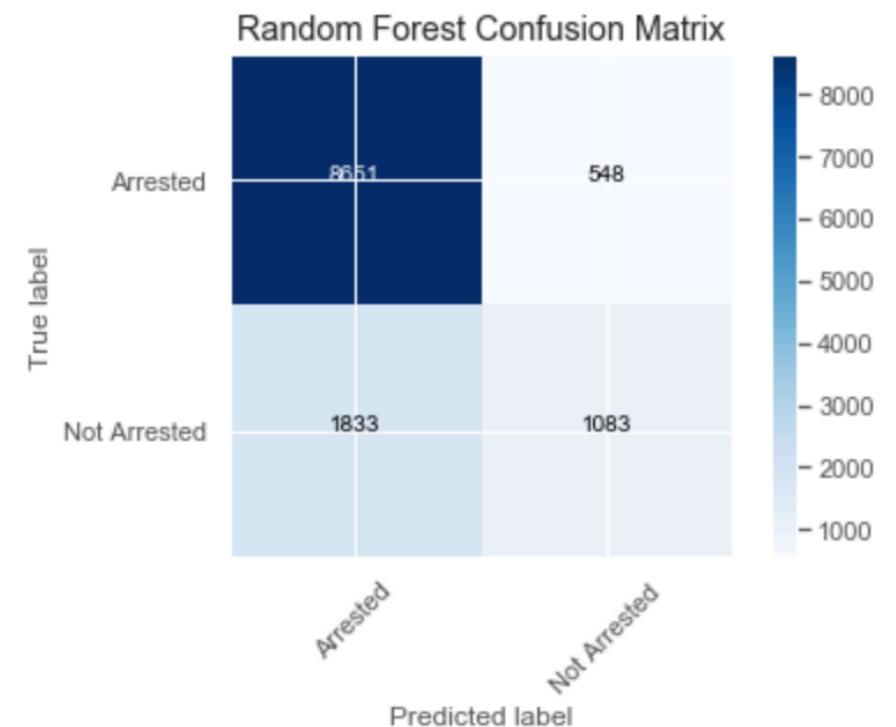
#3 – Decision Tree

- ▶ The decision tree model uses a tree-like structure to classify by efficiently partitioning each sample into sets with similar data points until you get to a homogenous set and can reasonably predict values for new data.
- ▶ Decision Tree Accuracy = 79.96%



#4 – Random Forest

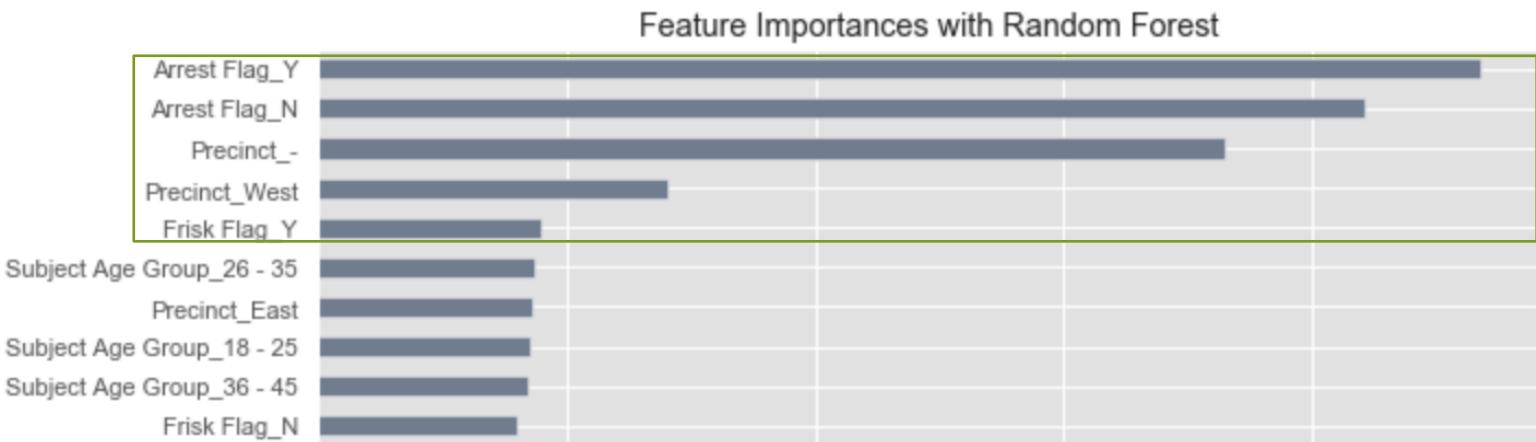
- ▶ Random forests are an ensemble method for decision trees that both trains trees on different samples of data (bagging) and randomly selects a subset of features to use as predictors (subspace sampling method) to create a 'forest' of decision trees.
- ▶ Random Forest Accuracy = 80.35%



Conclusion

► **Conclusion** – Random Forest is our best model, classifying arrests with an accuracy of 80.35%. According to this model, the 5 most important features from the dataset are:

- Arrest flag given
- Arrest flag not given
- Unknown precinct
- West precinct
- Frisk flag given



Recommendations & Next Steps

► Recommendations:

- Train officers on when is an appropriate time to arrest someone during the terry stop vs. when it is appropriate to wait until later as this is another key indicator of arrests.
- Try to capture the officer's precinct for all terry stops to have a better opportunity to predict if an arrest will occur.
- Train officers on when is an appropriate time to 'frisk' someone as this is another key indicator of arrests.
- Next Steps – Improve the model, either through cleaning up the data more or through adjusting model inputs to get an outcome more accurate than 80%. Additionally, looking into other types of classification models, specifically deep learning models, could also help generate a more accurate model.

Thank you!

- ▶ Any questions/concerns/comments?