

# Seattle Terry Stops Classification



Naweed Ahmed

 **FLATIRON SCHOOL**

# Understanding the problem

- The data set shows details of Terry Stops in Seattle between 2015 to 2020.
- Terry Stop is when a police officer in the US detains an individual based on reasonable suspicion.
- We used this data to predict whether an arrest was made after a Terry Stop.
- This is a Binary Classification problem, and we will need to use classification learners in order to predict this.

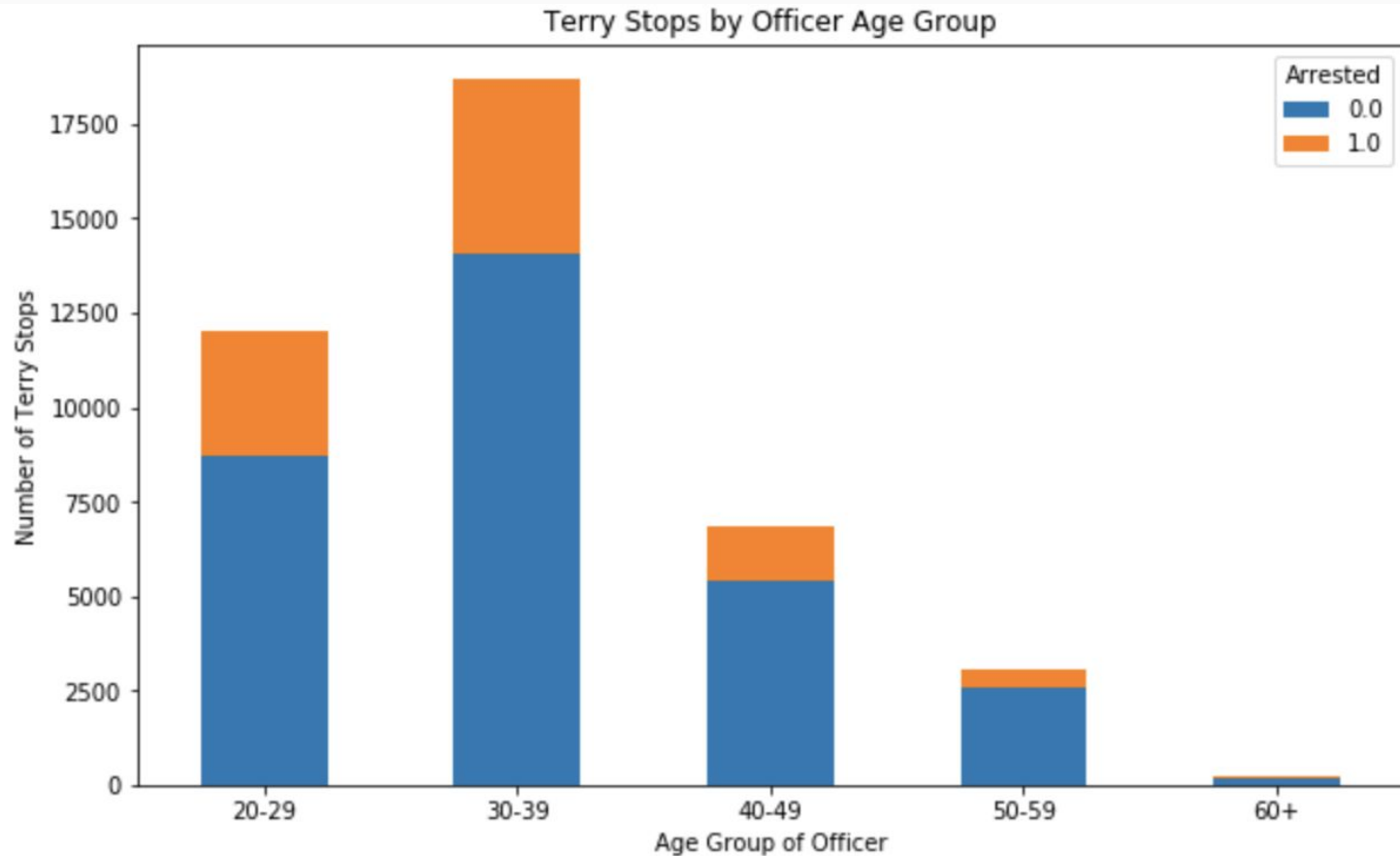
# Model Selection

- We used the following learners and tuned the hyperparameters using Grid Search CV for optimisation.
  - Logistic Regression
  - K-Nearest Neighbours
  - Decision Trees
  - Random Forest Models
- The best model was a Logistic Regression model giving us an Accuracy Score of 76% and a ROC\_AUC score of 62%.

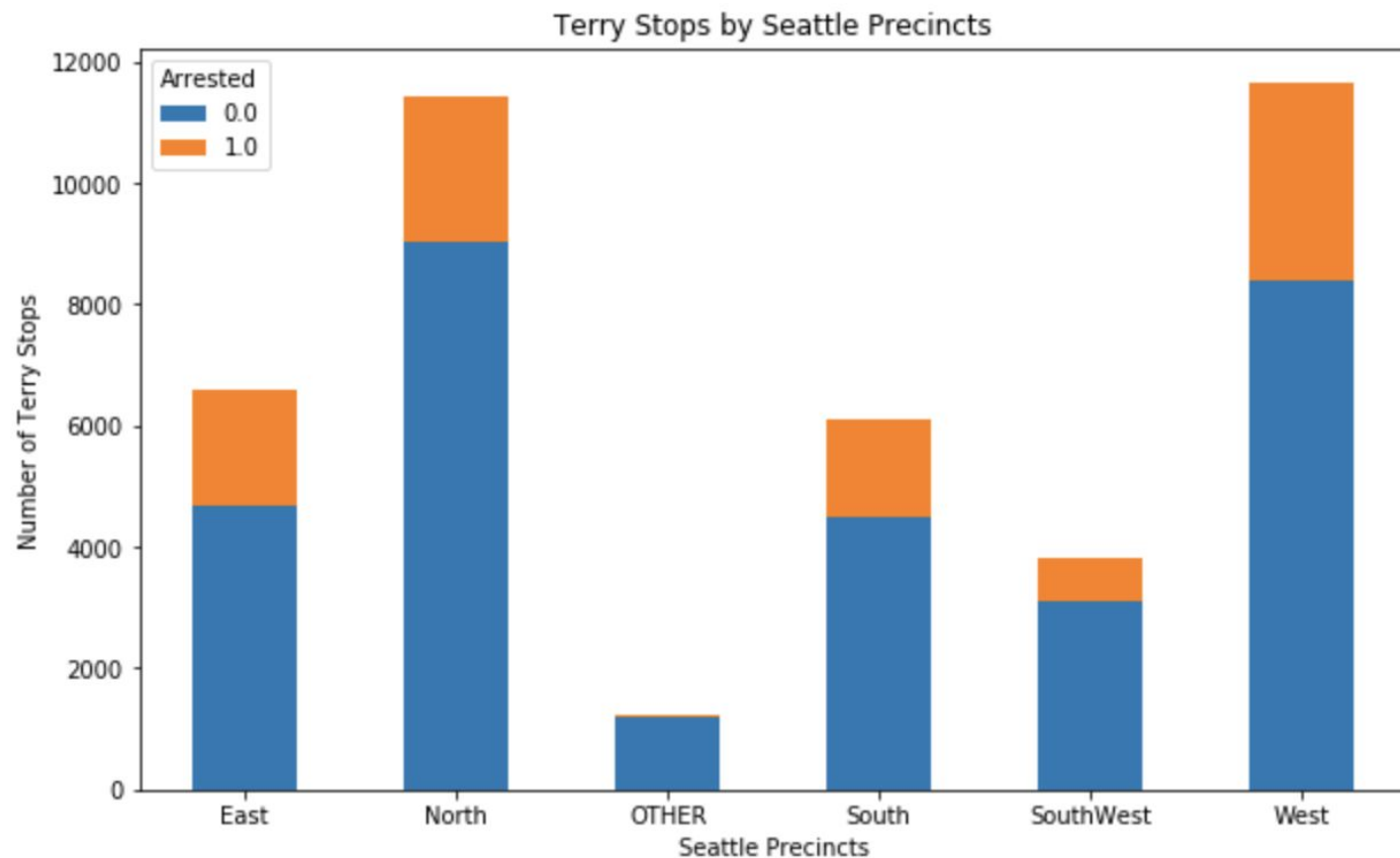
# Interpreting the model.

- Based on our model, the following factors appear to have the biggest impact on whether or not an individual is arrested.
  - Officer Age - Age group 20-29 has 11% increased probability
  - Location in Seattle - South-West Precinct has 2.3% increased probability
  - Officer Race - Mixed Race officer has 3.6% increased probability.
- We assessed this by calculating the log odds after increasing the value of a coefficient, and then working out the odds and the change in probability.

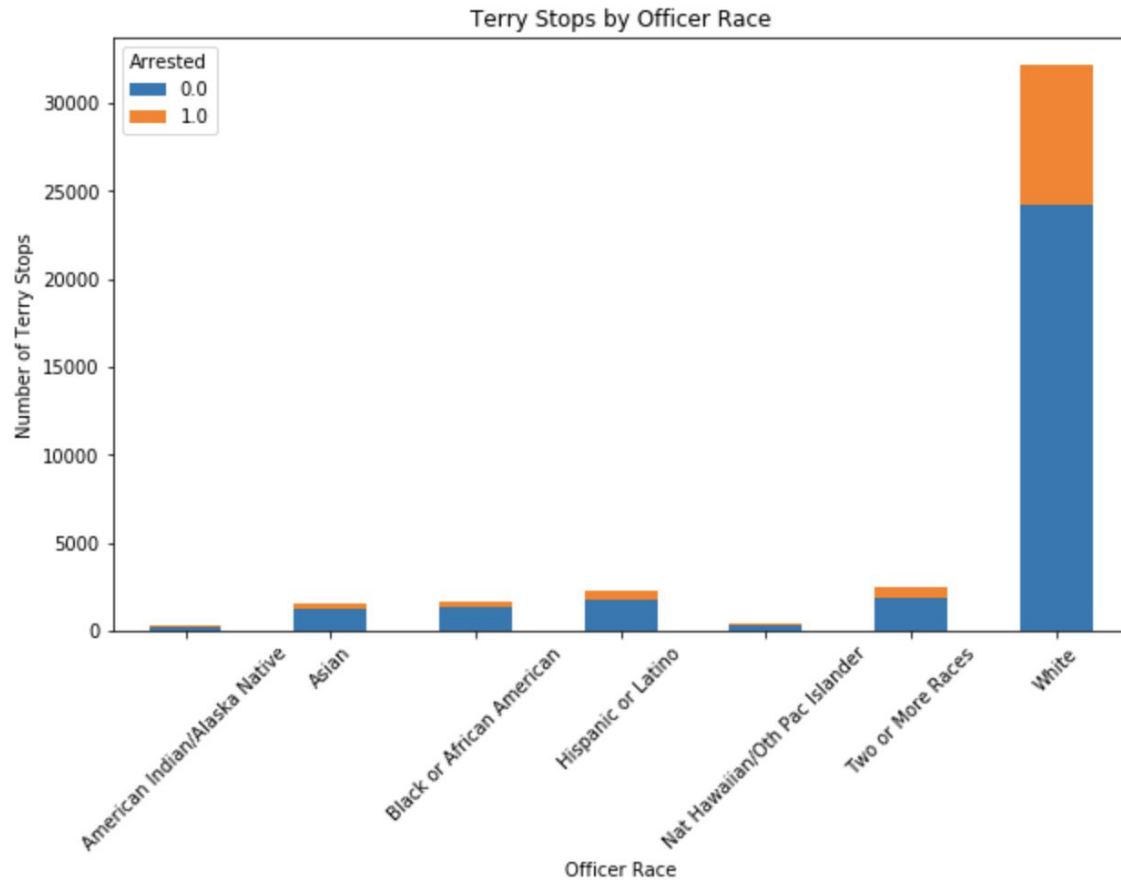
# Officer Age Groups



## Arrests by Location in Seattle



# Officer Race



# Threshold Selection

- Based on data between 2008 to 2017, we found that average arrests per annum is around 3,665 per 100,000 of population.
- This gives a prevalence of arrests in Seattle of 3.6%.
- Based on 2007 data, economic losses due to crime was \$15 bio, and \$179 bio was spent on police protection, judicial and legal activities, and corrections.
- US population in 2007 was 301.2 mio.
- We assigned the following costs to the confusion matrix metrics.

True Positive Cost (TPc) = -1383 + 49

True Negative Cost (TNc) = 0

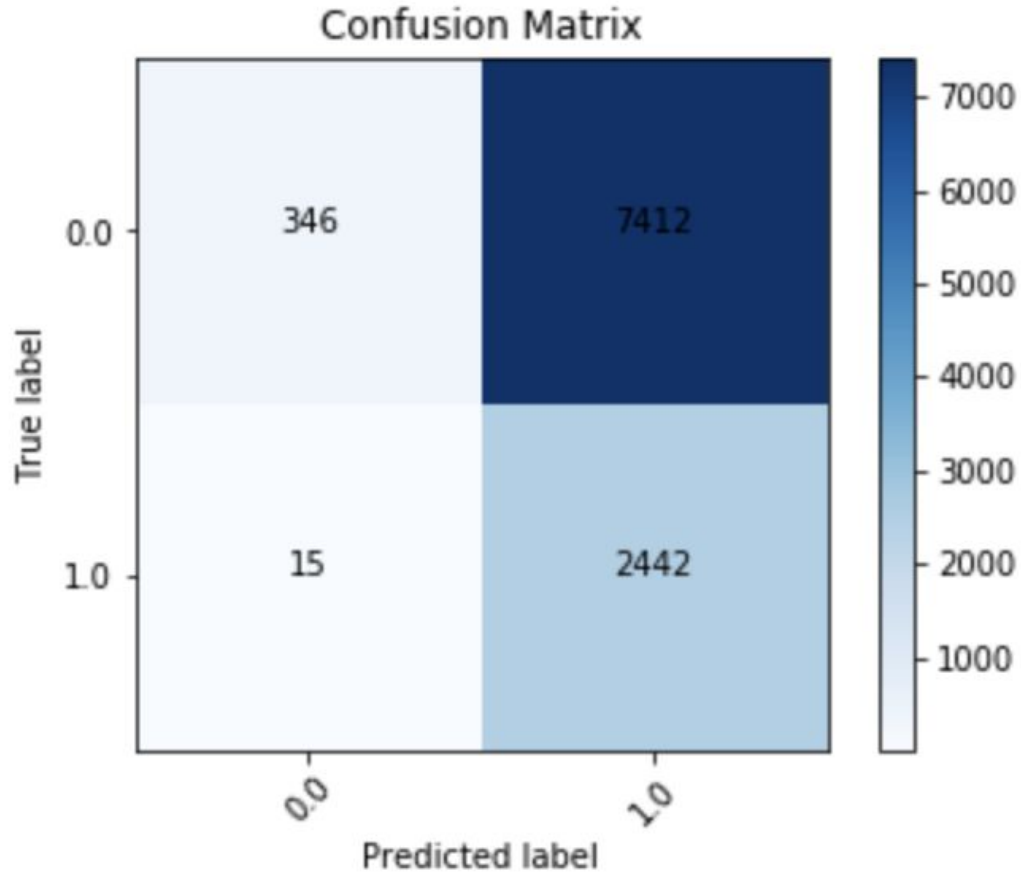
False Positive Cost (FPc) = 49

False Negative Cost (FNc) = 1383

Using this, our fm value came to 0.47 and our threshold level at 0.11



# Confusion Matrix



Alpha - 0.96

Power - 0.99

Accuracy - 0.27

Precision - 0.25

# Interpreting the Confusion Matrix

- Based on the metrics, it is clear that more considerations need to be taken when selecting the threshold level. We cannot make our threshold selection based on the cost of crime and the amount spent on policing alone, as this would lead to increased number of arrests.
- We would need to assess the economic benefit of low crime on society, and not just the loss based on the crime taking place. In addition to that, we would need to work out the social cost due to false arrests.
- These metrics are difficult to quantify, but once we are able to measure these and incorporate it into our model, it would give us a much more balanced threshold and would improve our models performance.

# Future Work

- Ensure that Police Officers do a much better job in filling in the data for Terry Stops.
- Assess Economic impact of low crime and of false arrests to help with our threshold selection.



Thank You!

Questions