PREDICTING THE CAR ACCIDENT SEVERITY IN SEATTLE

THE PROBLEM

What is the problem?	• What factors contribute accident severity?
Why should this problem be solved?	• To determine the severity of future car accidents.

BACKGROUND INFORMATION

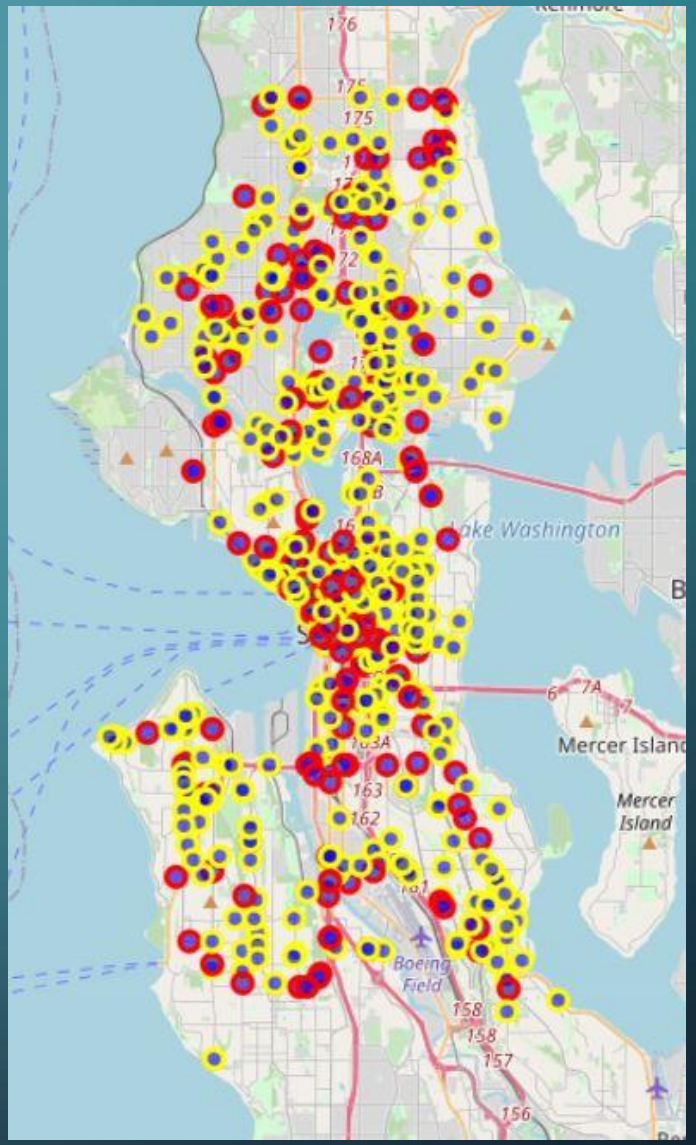
- Car accidents are a common yet avoidable problem in the United States.
  - In Washington, car accidents occur every four minutes.
  - Caused by speeding, driving under influence, weather conditions, and road conditions.
  - If these factors were correlated car accident severity, a predictive can be made.

DATA ACQUISITION AND CLEANING

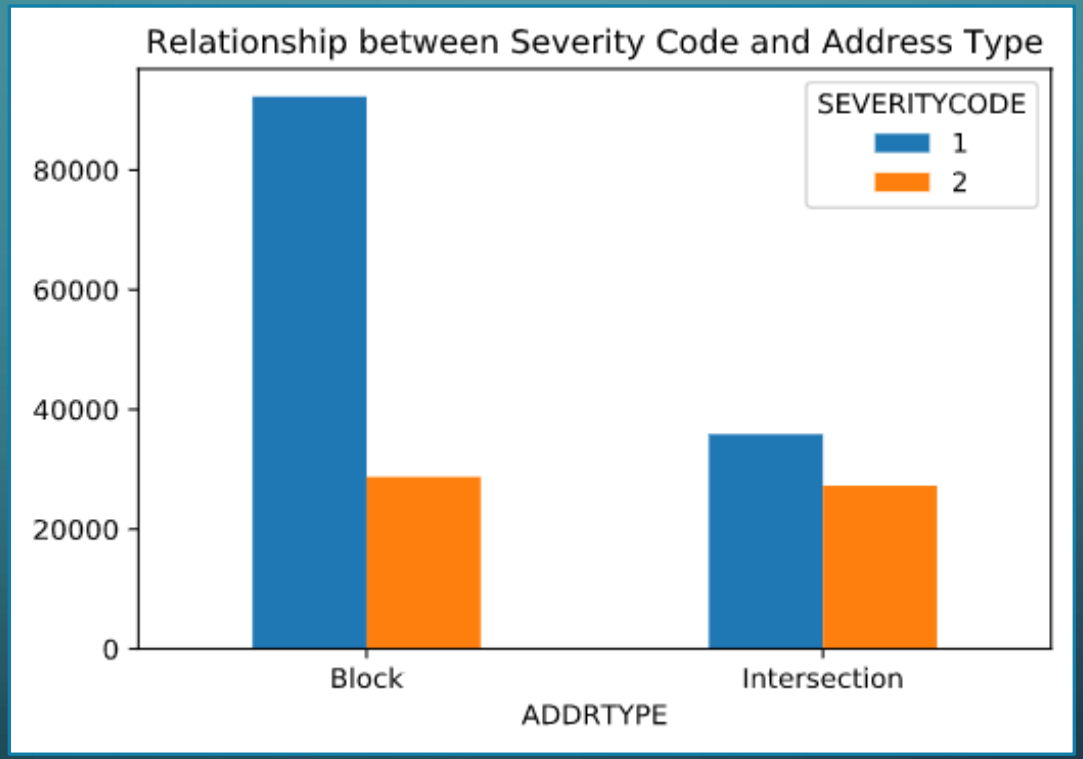
- Data provided by the Seattle Police Department via IBM
  - Car Accident Severity
  - Code 1: Property Damage
  - Code 2: Physical Injury
- Originally contained 194,673 with 38 features.
- Removed samples with missing values.
- Removed features that lacked any data.
- Cleaned dataset contains 184,146 samples with 9 features.

MAP OF SEATTLE

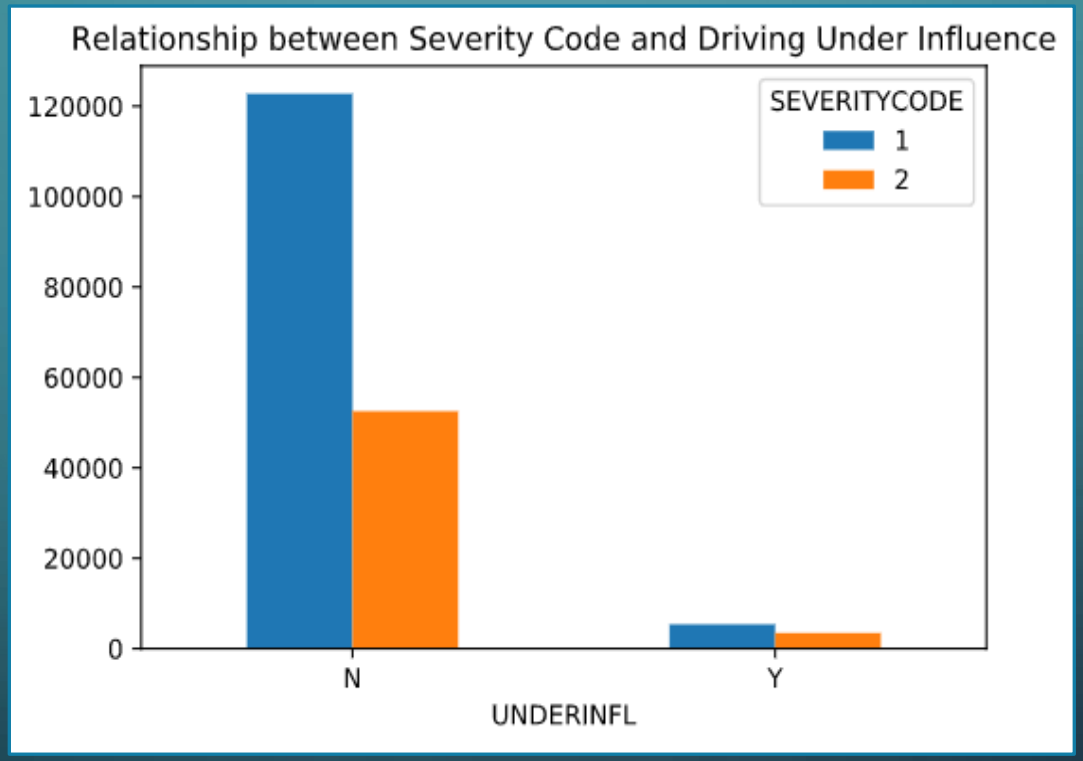
- Concentrated in the International District.
- Red is Code 2 Severity
- Yellow is Code 1 Severity



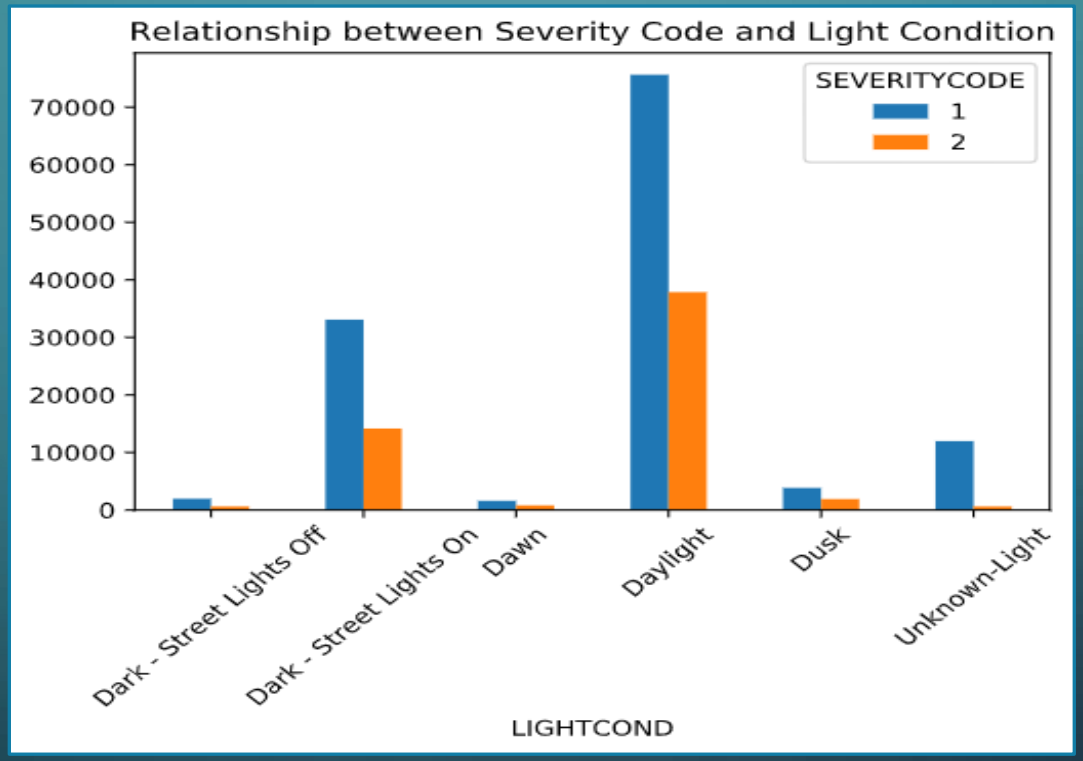
USING ADDRESS TYPE



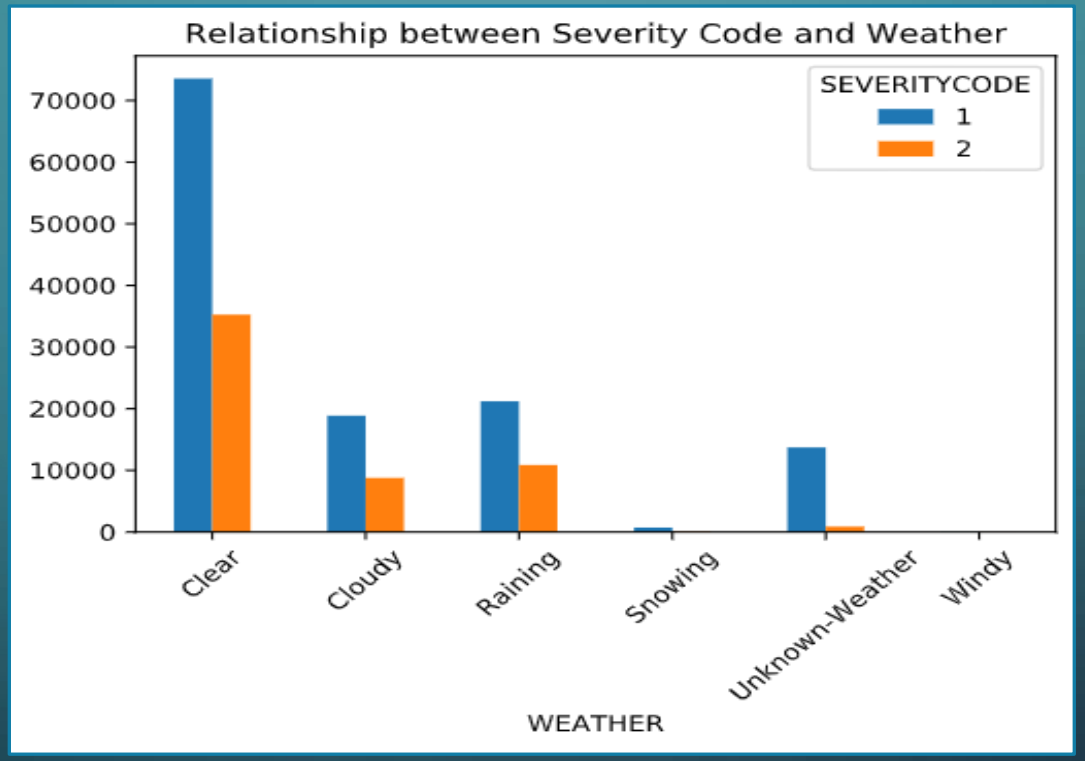
USING DRIVING UNDER INFLUENCE



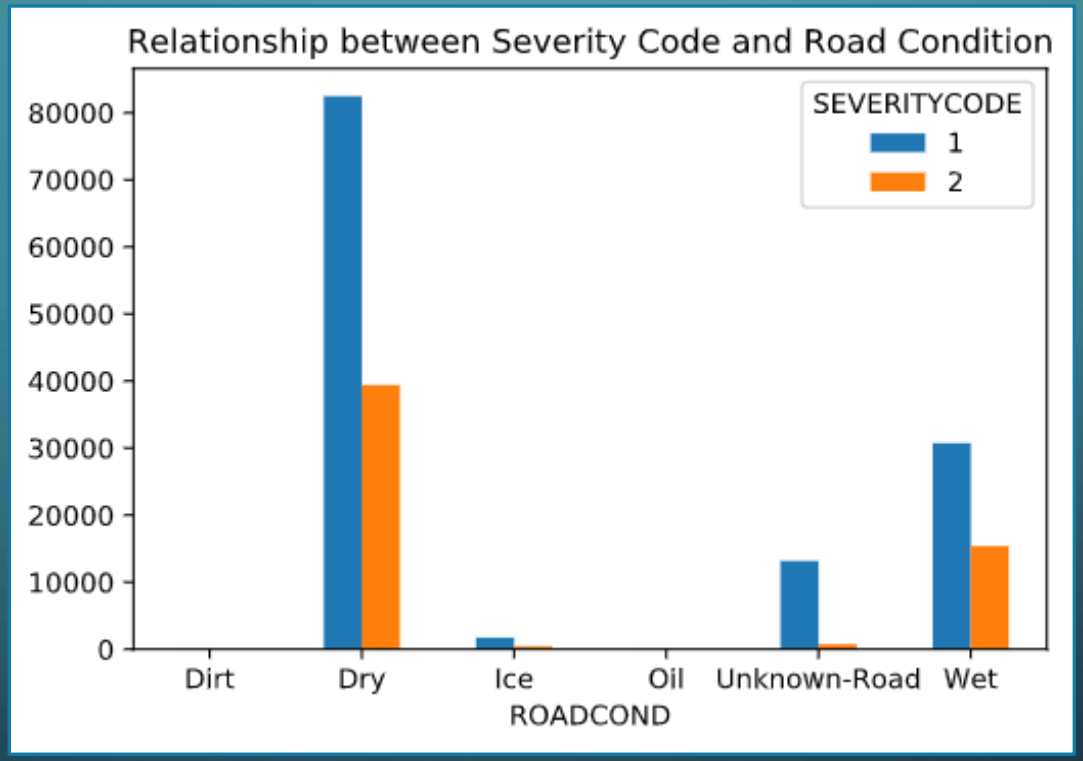
USING LIGHT CONDITION



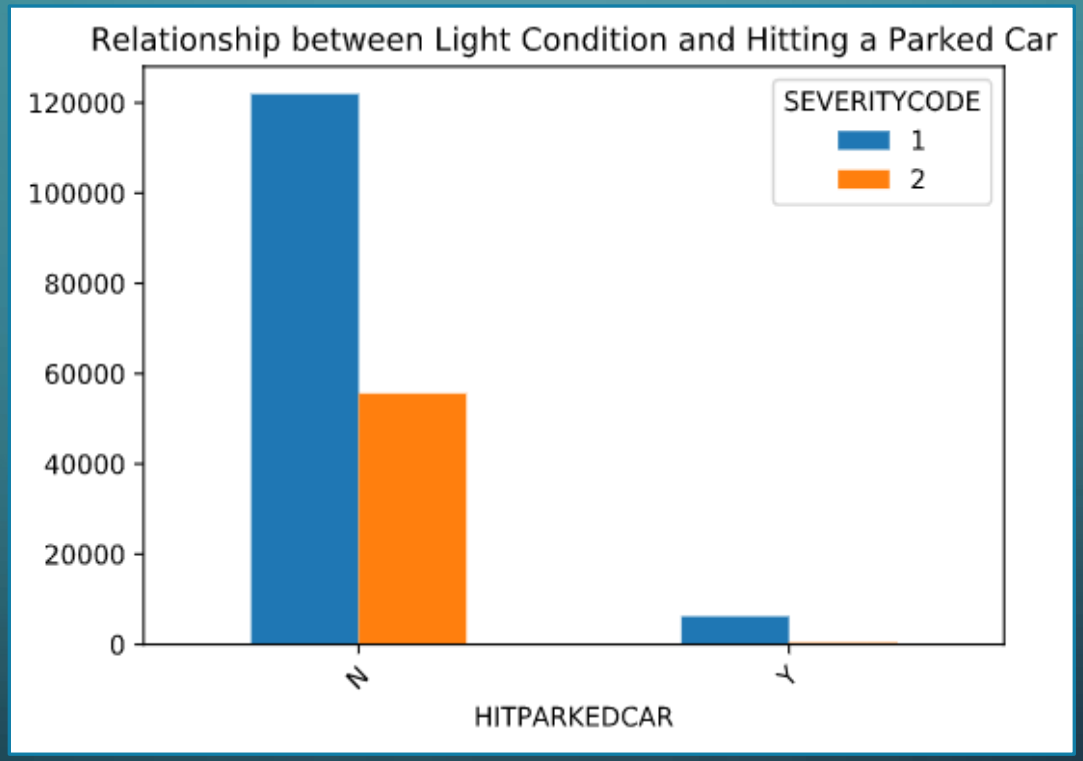
USING WEATHER



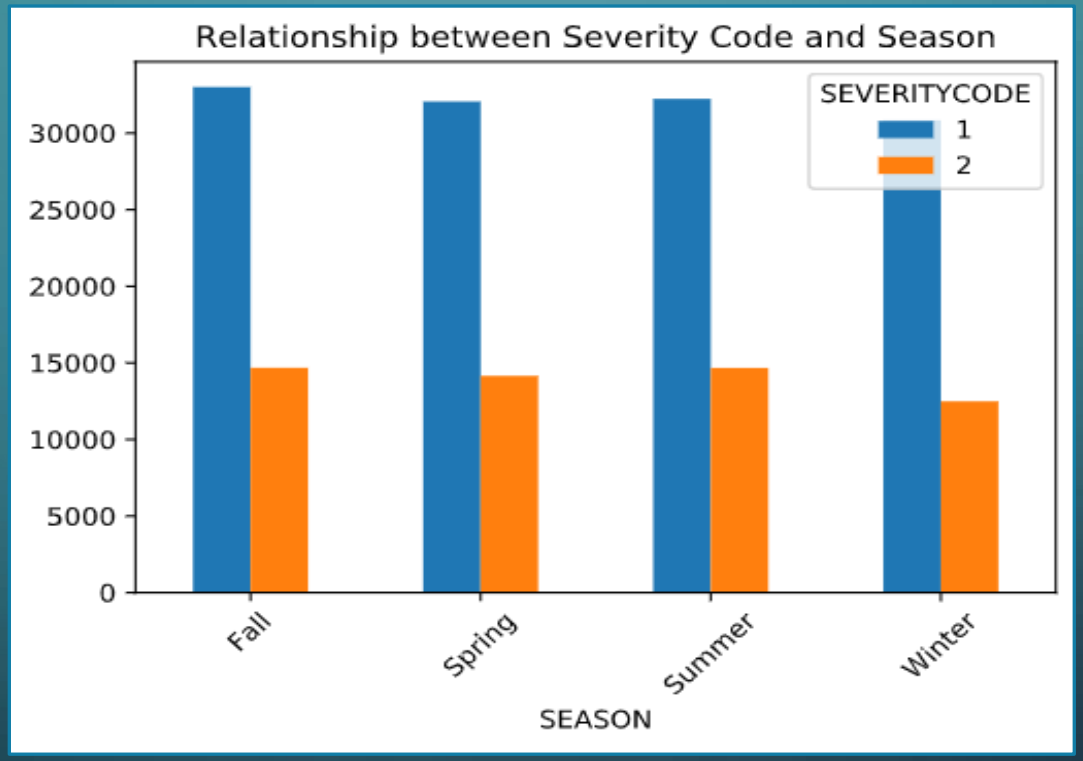
USING ROAD CONDITION



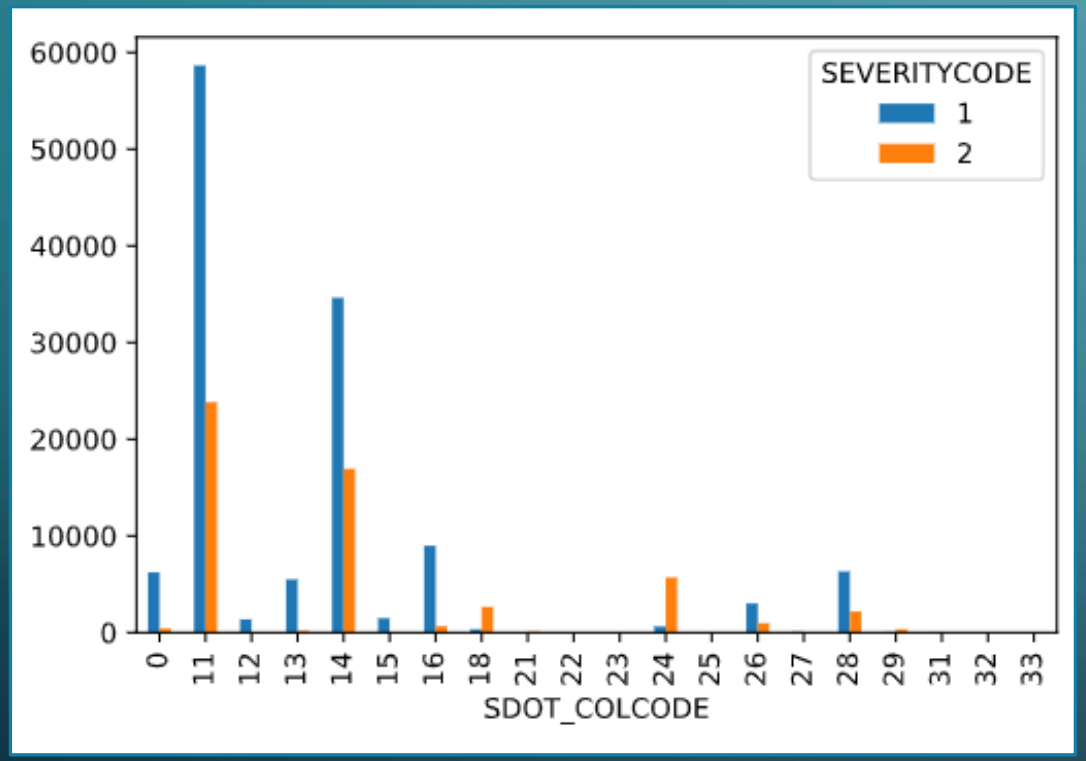
USING PARKED CAR COLLISION

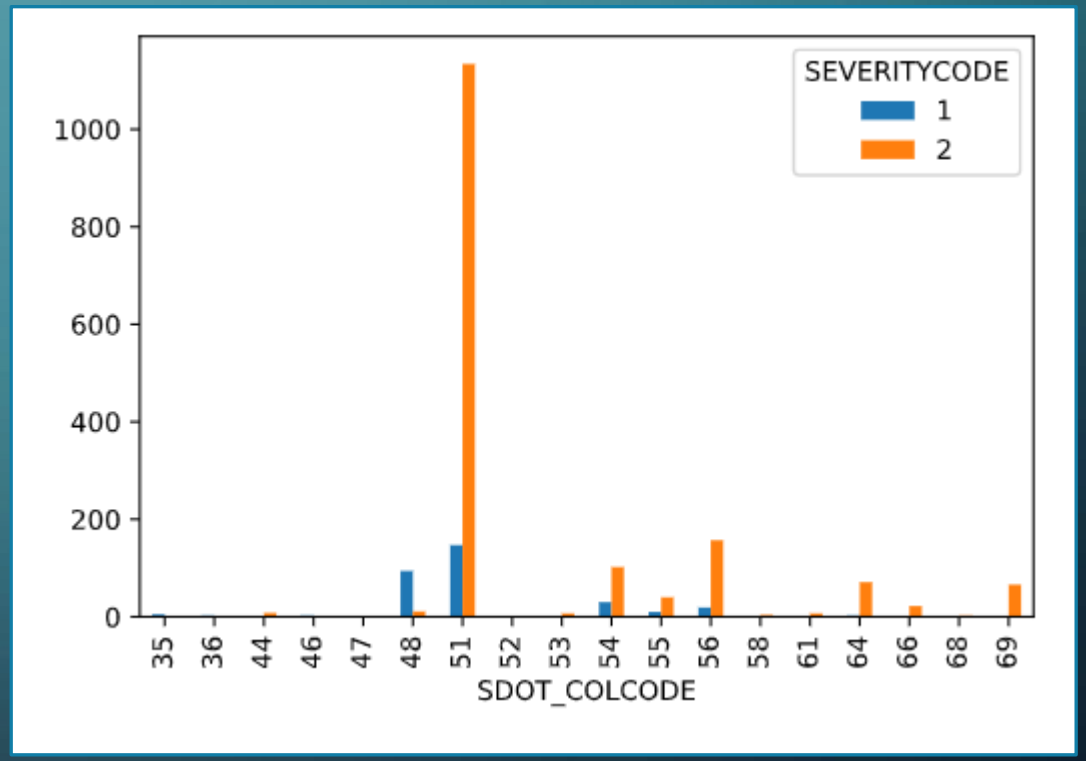


USING SEASON

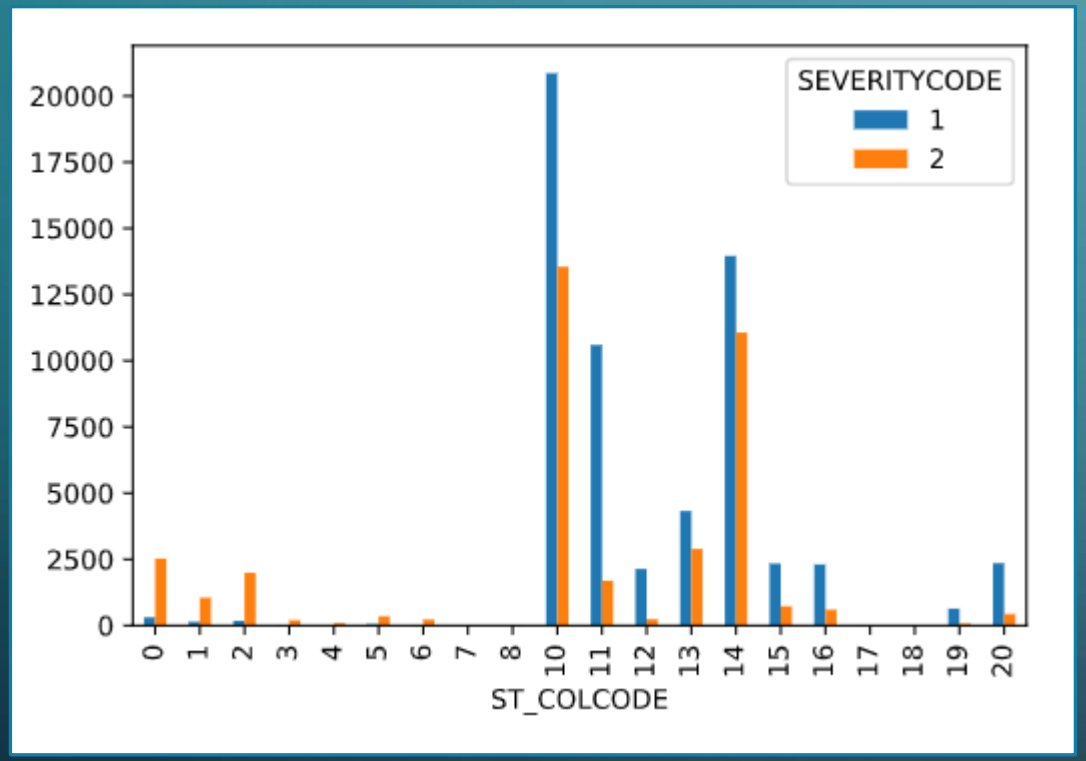


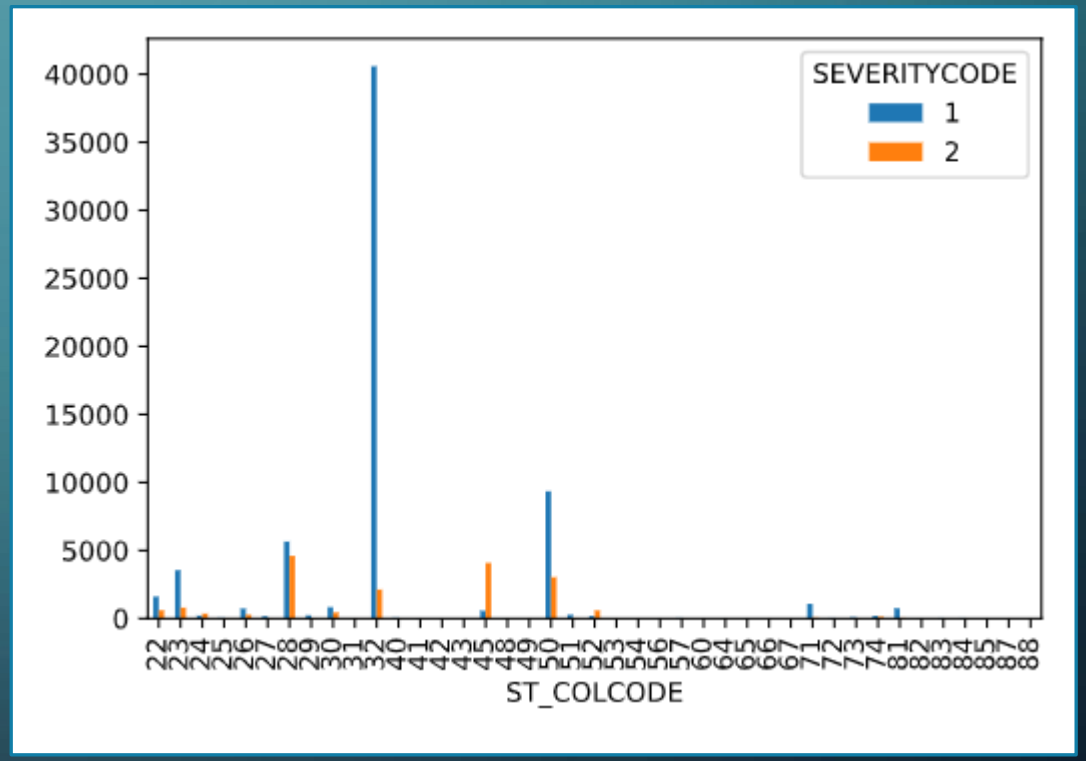
USING SDOT COLLISION NUMBER





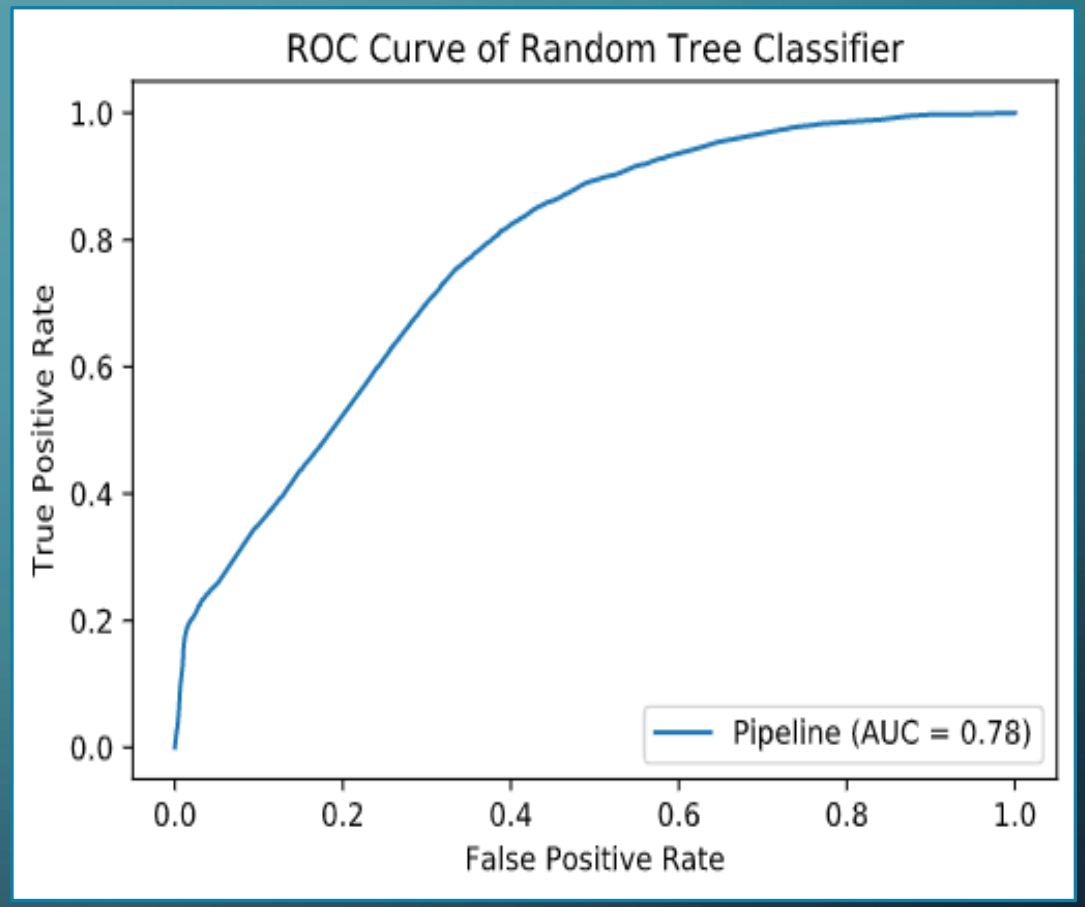
USING STATE COLLISION NUMBER





CLASSIFICATION MODELS

- Accuracy:
  - 0.711 for random tree classifier
  - 0.649 for logistic regression
- Random tree classifier is more accurate, but logistic regression has better code 2 prediction score.



CONCLUSION

- Gather more features that correlates to car accident severity.
- More improvements on the accuracy of the models.
- Better recording for the dataset.
  - Speeding was missing 95% of the samples.
  - Samples contained no information.