Capstone Project

Predict Car Accident Severity

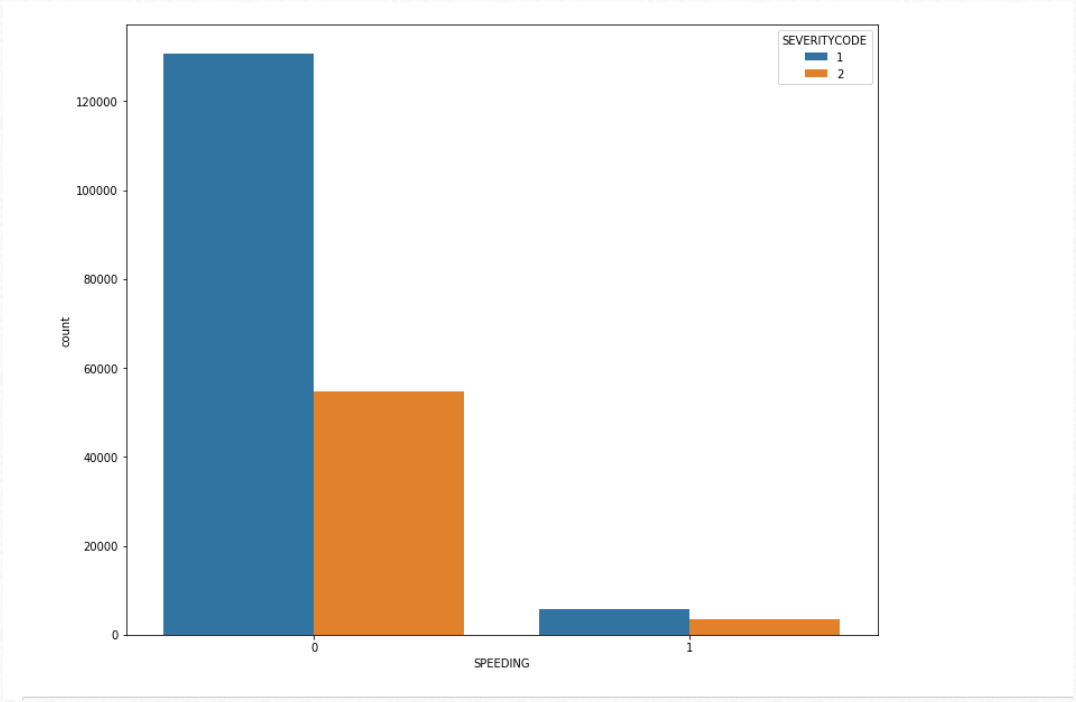
Problem Statement

- 6 million car accidents every year in U.S.
- 3 million injured by car accidents every year
- 2 million experience permeant injuries every year
- Goal:
  - Predict the severity so the doctos and nurses will have enough resources or personnel before the ambulances are back from the accident sites

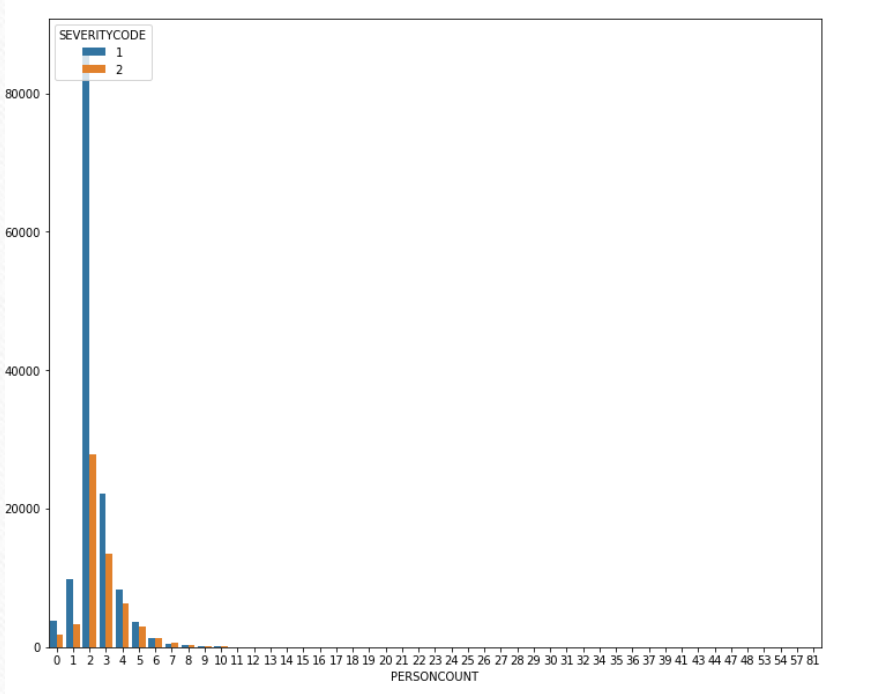
Data Analysis

- 194673 x 38 Entries
- Independent Variable (X): PERSONCOUNT, PEDCOUNT, PEDCYLCOUNT, INATTENTIONIND, SPEEDING
- Dependent Variable (Y): SEVERITYCODE
- Fill NAN values for columns INATTENTIONIND and SPEEDING
- Change YES and NO to numeric value
  - Yes as 1
  - No as 0

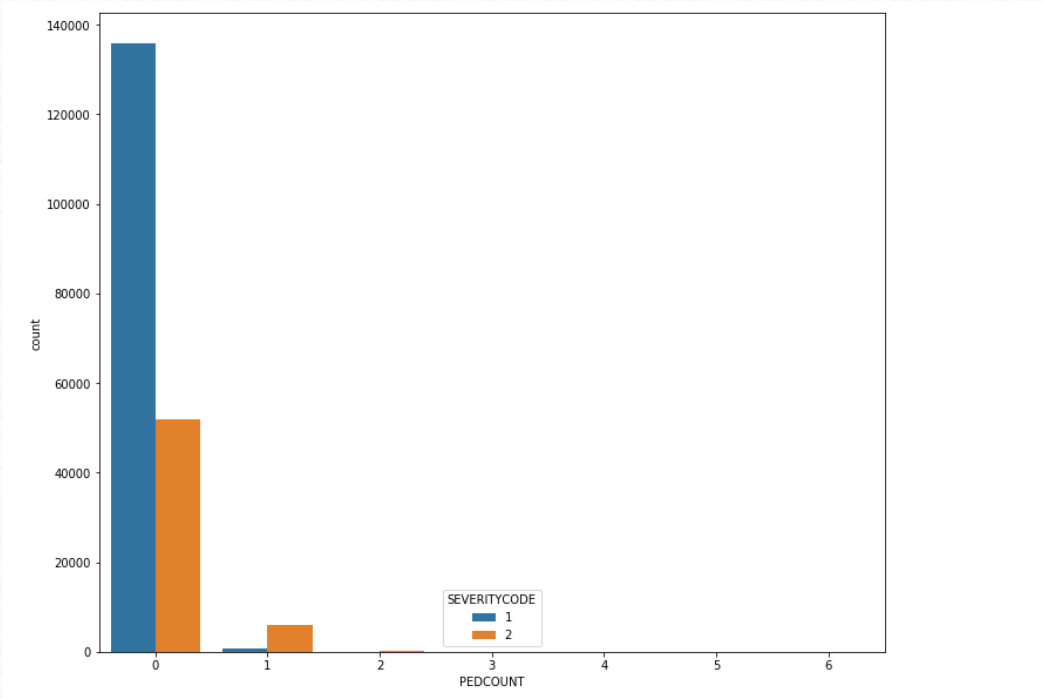
Data Visualization (Speeding)



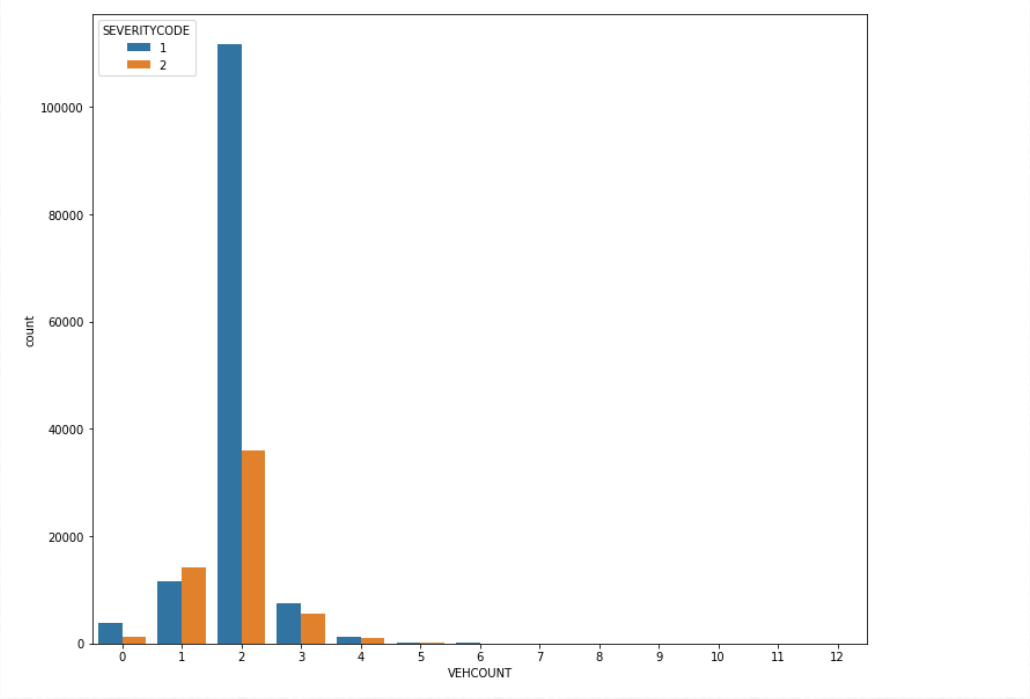
Data Visualization (Person Count)



Data Visualization (Pedestrian Count)



Data Visualization (Vehicle Count)



Machine Learning Model

- Ratio of train and test data was 80% and 20%

```
Out[20]: DecisionTreeClassifier(class_weight=None, criterion='entropy', max_depth=None,
                                max_features=None, max_leaf_nodes=None,
                                min_impurity_decrease=0.0, min_impurity_split=None,
                                min_samples_leaf=1, min_samples_split=2,
                                min_weight_fraction_leaf=0.0, presort=False, random_state=None,
                                splitter='best')
```

Conclusion

- F1 score
  - Train dataset: 0.849 = 85%
  - Test dataset: 0.847 = 85%
- Model show high accuracy