

Predicting the Car Accident Severity in Seattle

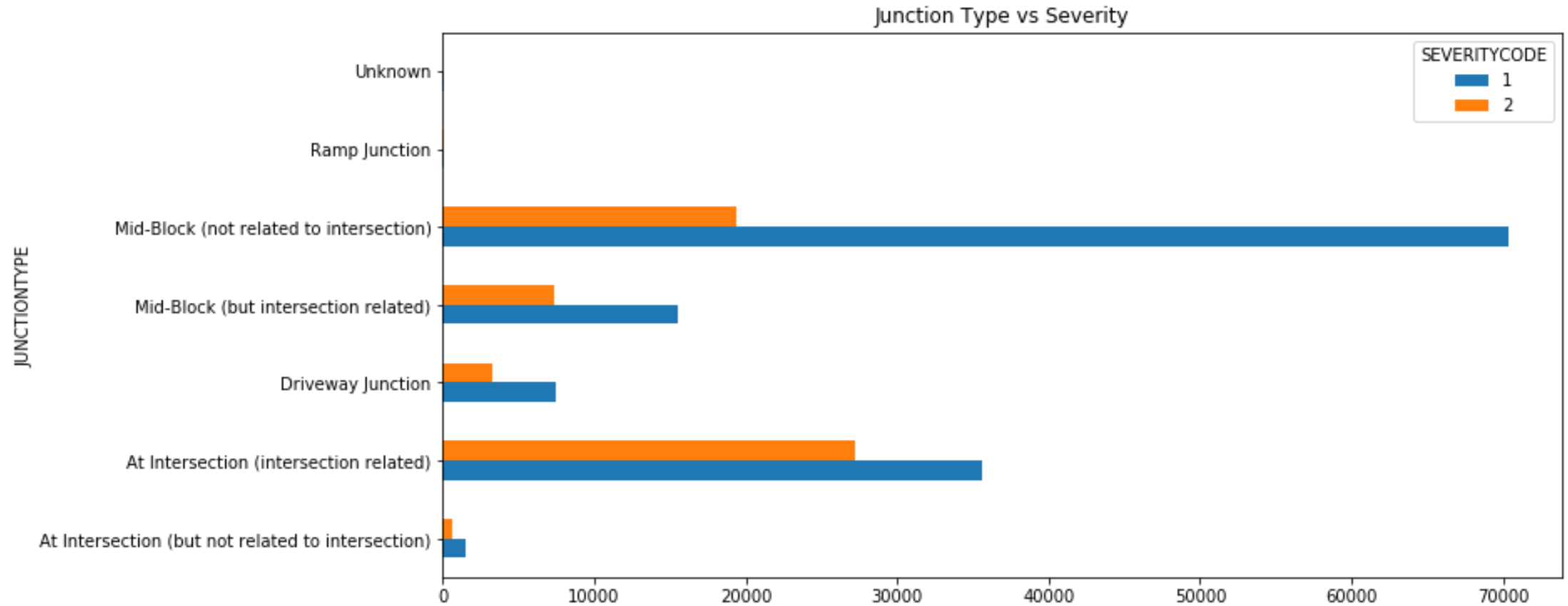
Introduction

- Car accidents is one of most severe problems in the modern world.
- Millions are dying every year due to car accidents.
- Many governments are taking measures to make the roads safer:
 - optimizing public transportation
 - enabling safer routes
 - cost-effectively improving the transportation infrastructure.
- The objective of this project: to build a machine learning algorithm in order to predict the car accident severity in Seattle, USA.

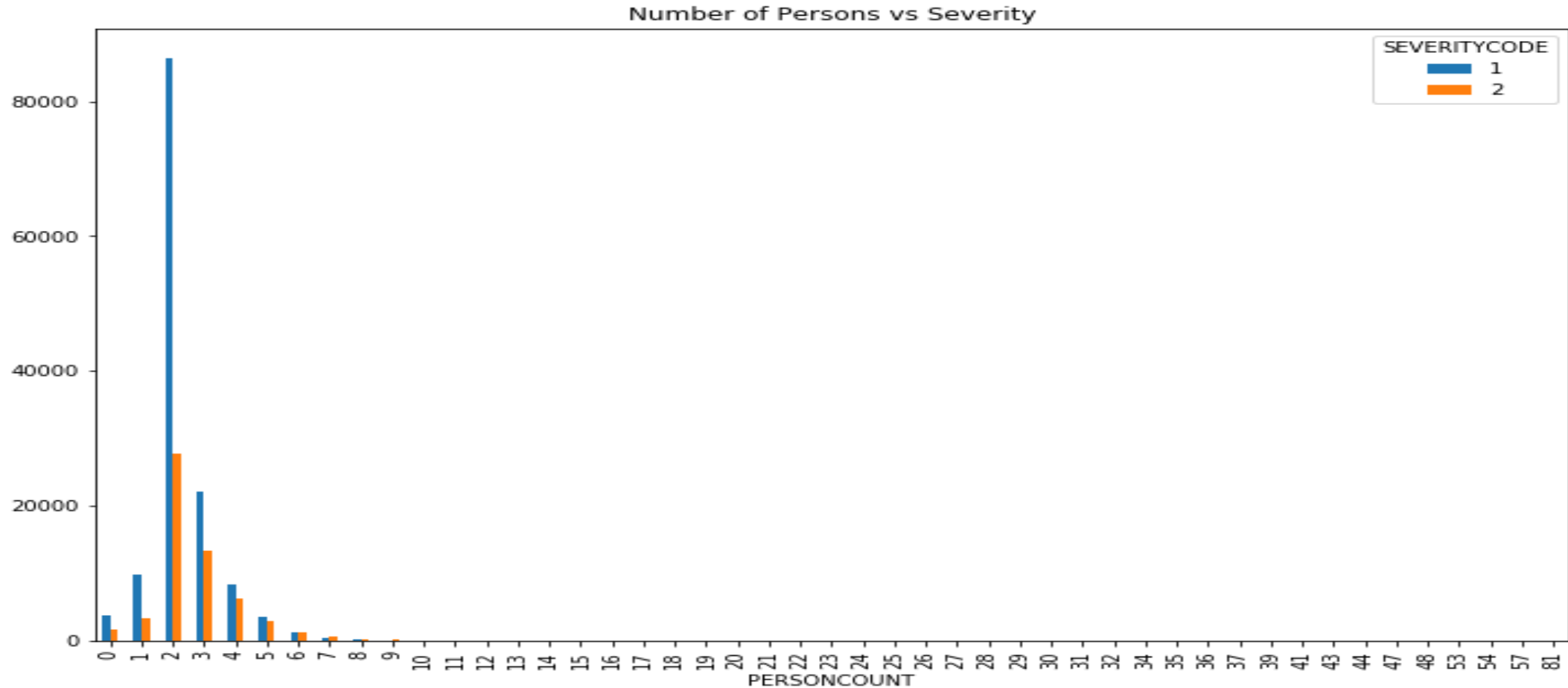
Data Collection, Preprocessing and Exploratory Data Analysis

- Data is collected from the web page of Coursera Course Applied Data Science Capstone.
- Six features were selected for building the machine learning models: SEVERITYCODE, JUNCTIONTYPE, PERSONCOUNT, WEATHER, ROADCOND, LIGHTCOND.
- the categorical features were converted to numerical values.
- The imbalanced target feature SEVERITYCODE was balanced by down sampling the majority class
- The relationships among the features were explored by using bar plot.

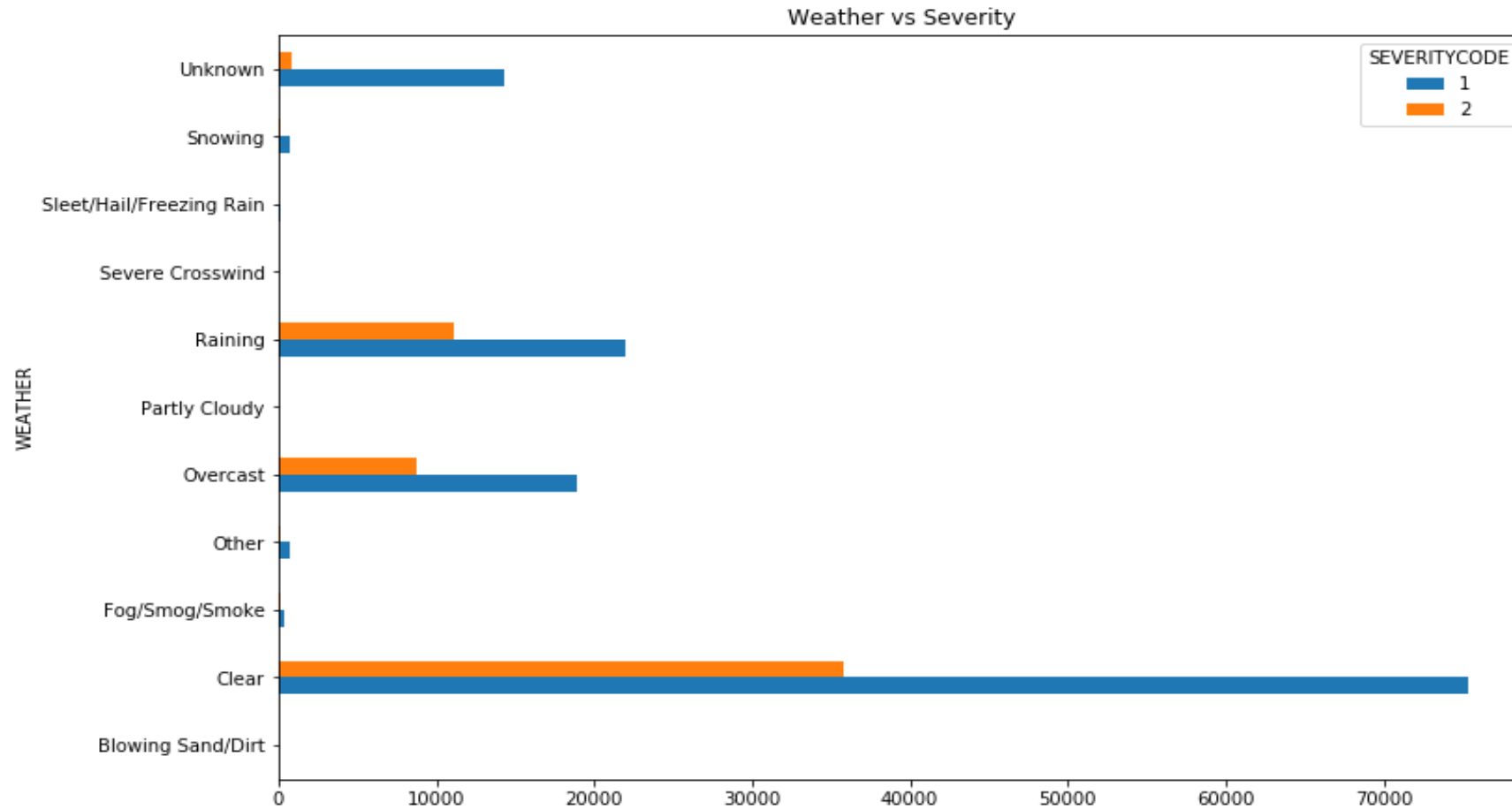
Relationship between Junction type and Severity



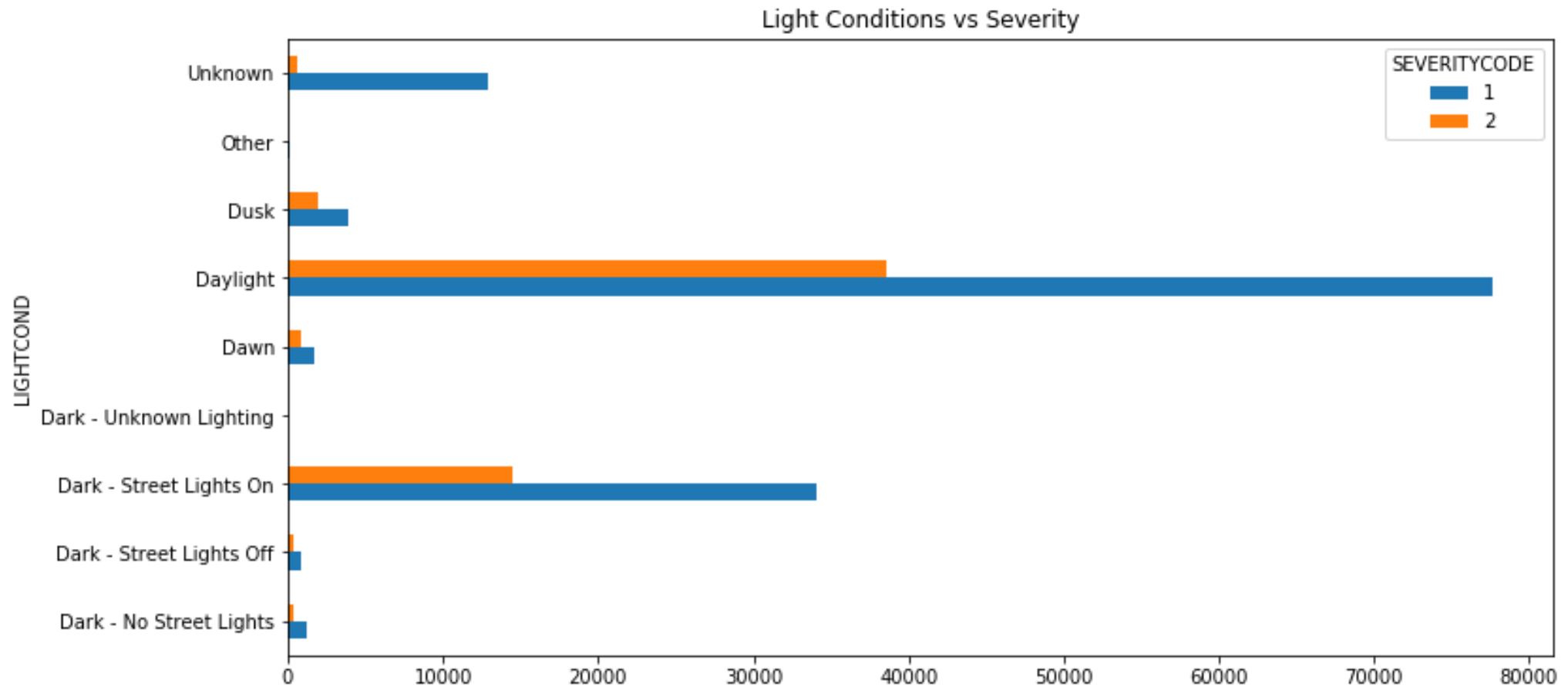
Relationship between Number of Persons and Severity



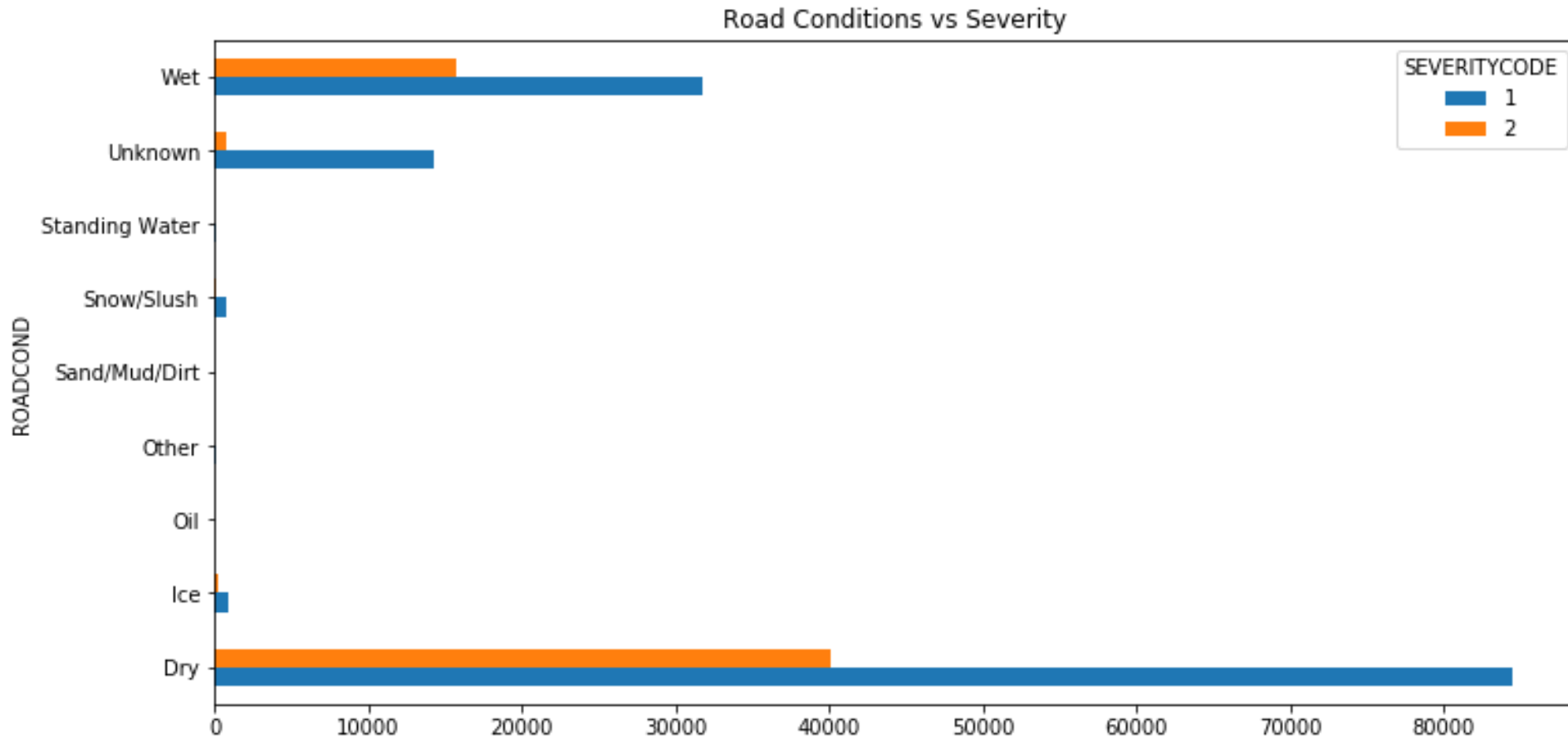
Relationship between Weather and Severity



Relationship between Light Conditions and Severity



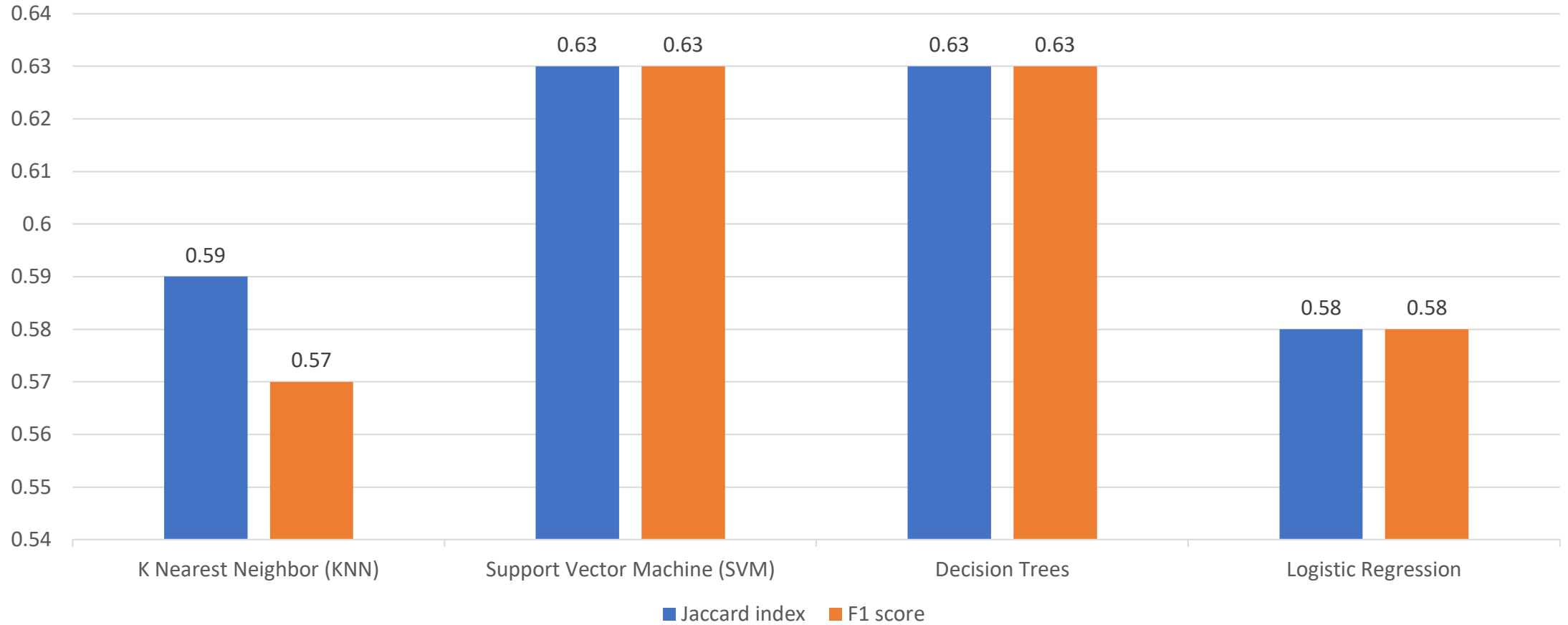
Relationship between Road Conditions and Severity



Building the Machine learning Models

- The selected features were split into train and test sets.
- The following machine learning models were built:
 - K Nearest Neighbor (KNN)
 - Support Vector Machine (SVM)
 - Decision Trees
 - Logistic Regression
- Jaccard index and F1 score were used to evaluate the machine learning algorithms.

Evaluation of the Machine learning Models



Conclusion

- Support Vector Machine (SVM) and Decision Trees are both suitable for predicting the car accident severity in Seattle.
- Further improvement should be done for better result.
- An attribute specifying the reason behind car accidents during ideal driving conditions will increase the effectiveness of the model exponentially.