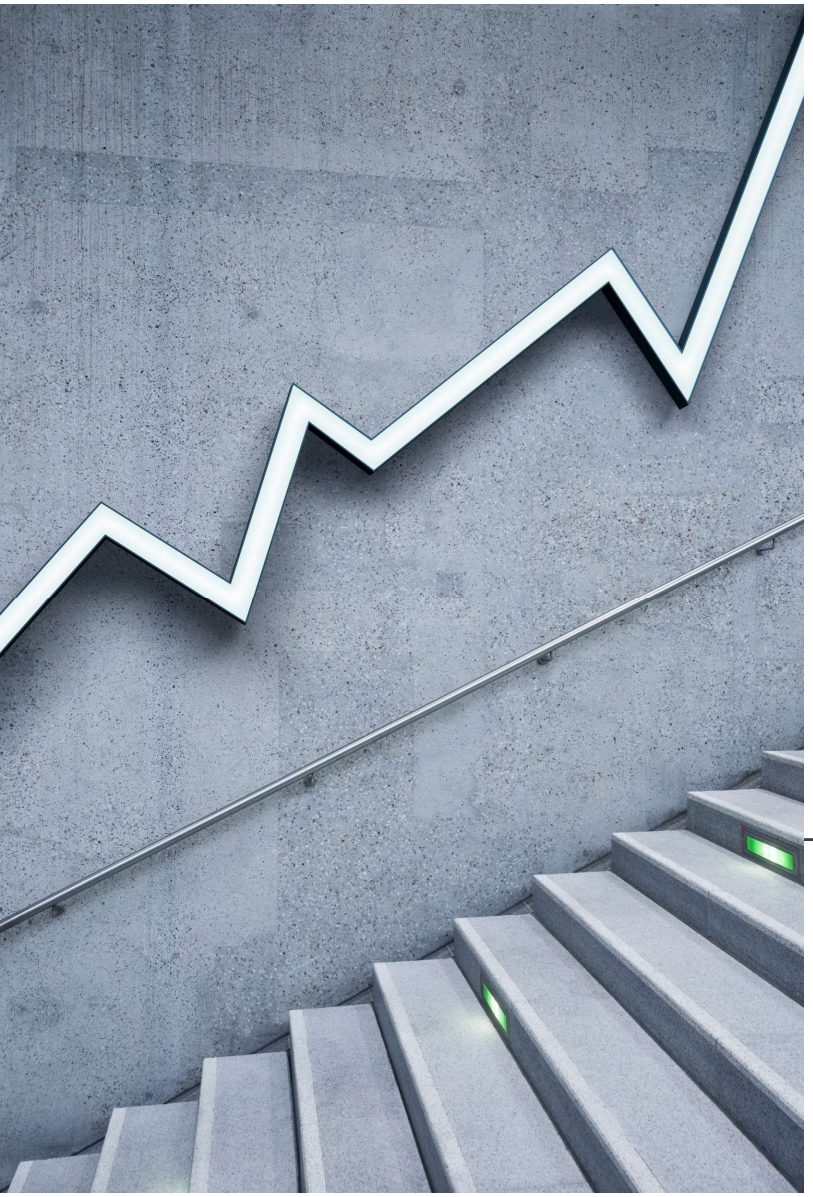


Predicting Car Accident Severity

COURSERA PROJECT

Predicting Car Accident Severity Could be Valuable

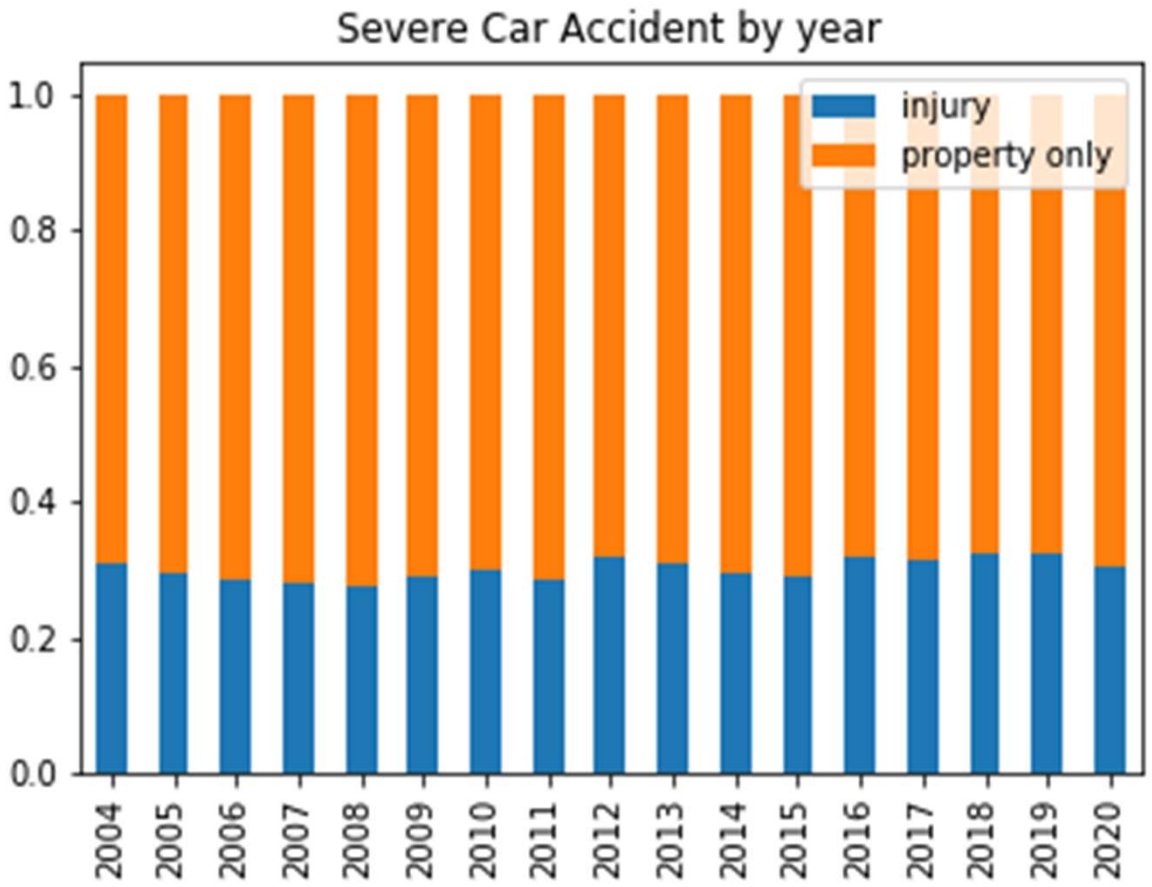
- Allow more efficient deployment of police resource
- Help educate drivers to manage driving habits to reduce risks
- Provide useful information for insurance company for adjudicating claims
- Increase traffic throughput by smart routing

Data Acquisition and Preparation

- Data collected from Seattle area from 2004 to 2020
- Data dictionary provided with detailed field coding
- Missing data were removed from the data set prior to modeling
- Categorical data fields were converted to on-hot coding
- Certain features were created from the original data field

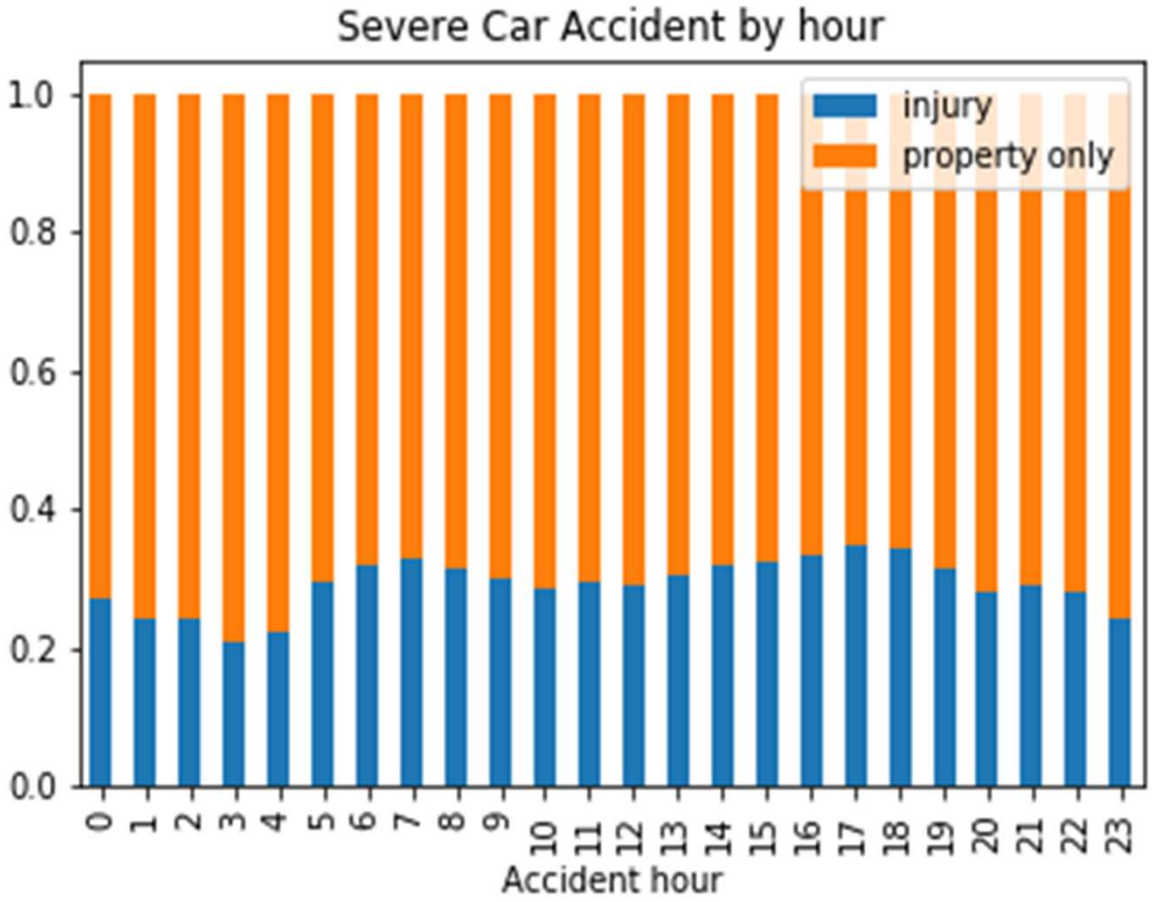
Data Exploratory Analysis – Long-term trend

It appears that there is no long -term trend in terms of accident severity.



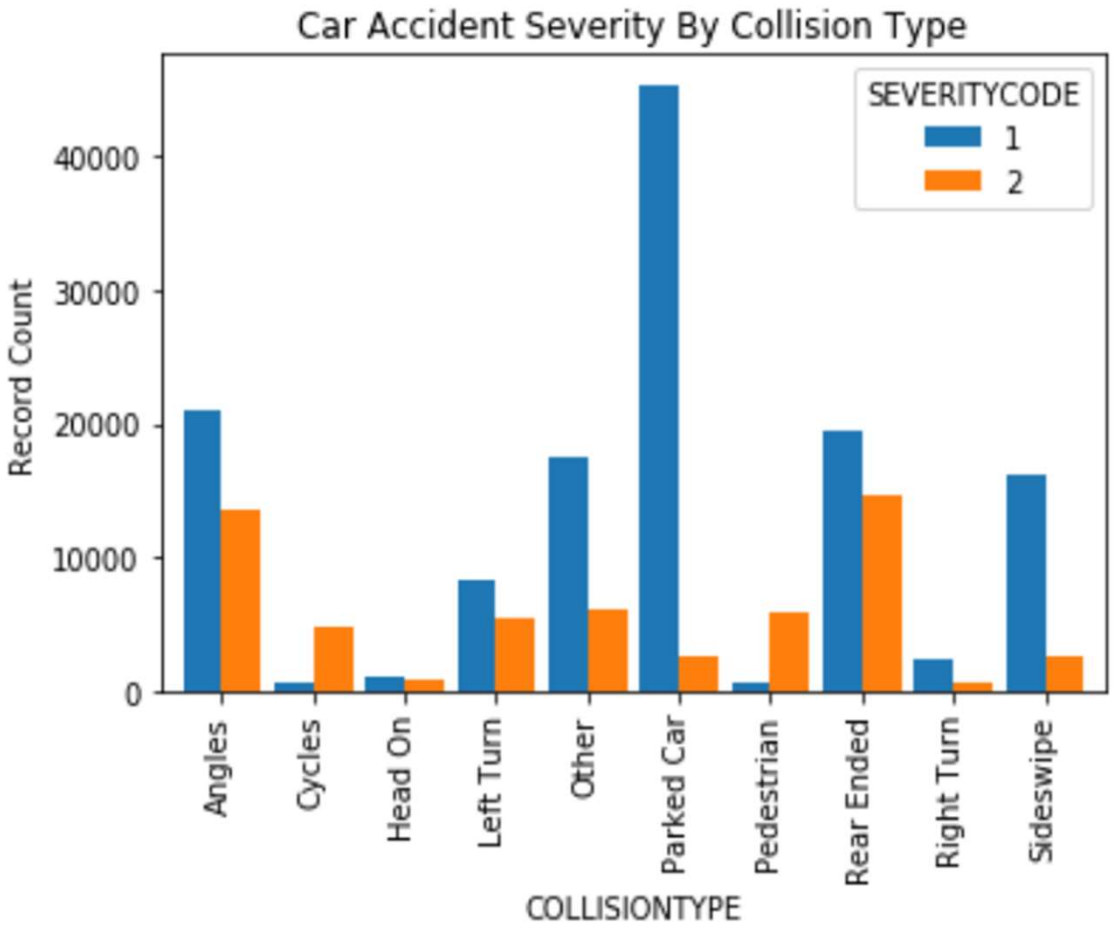
Data Exploratory Analysis – Time of day

It appears that traffic hours tend to observe more severe accidents.



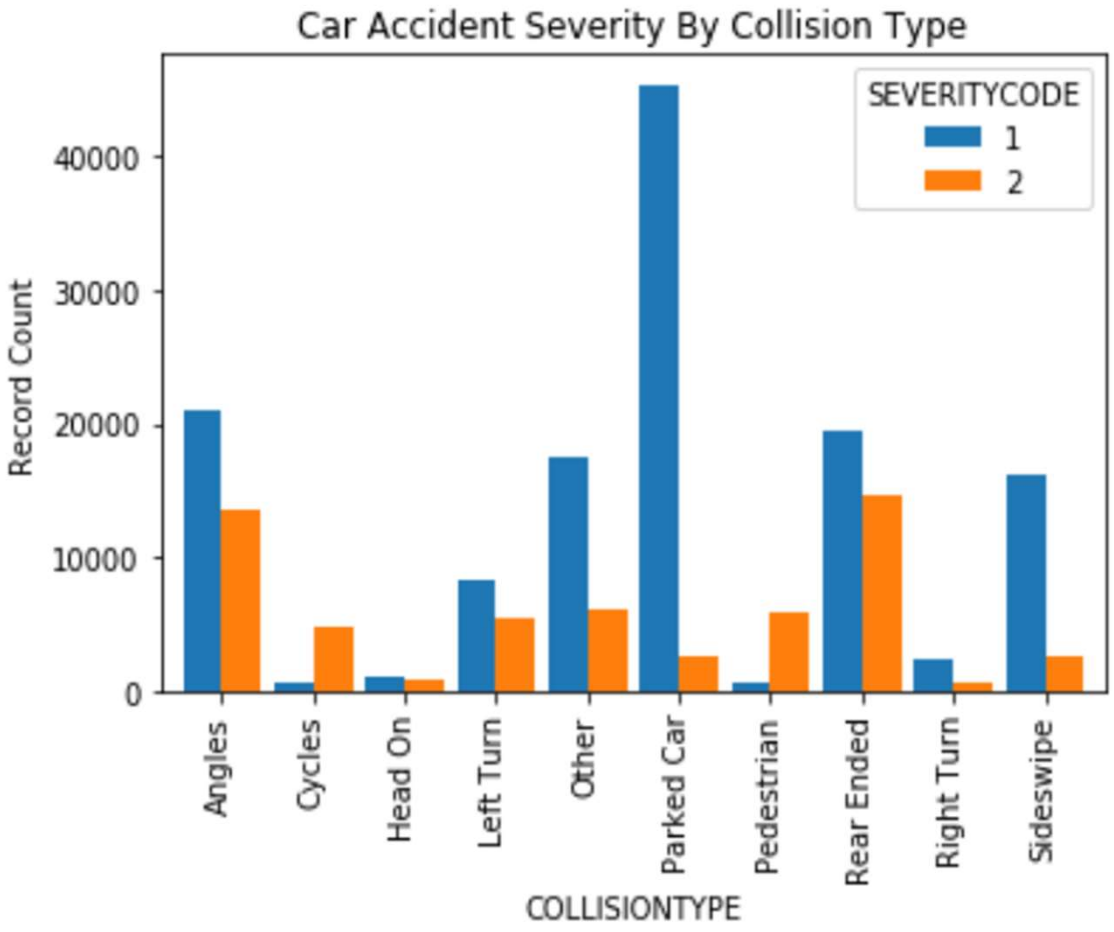
Data Exploratory Analysis – Collision Type

It appears that accidents vary by collision types.



Data Exploratory Analysis – Collision Type

It appears that accidents vary by collision types.



Classification Models

Following models were built to predict accident severity:

- K-nearest-neighbor
- Decision Tree
- Support Vector Machine
- Logistic Regression
- Random Forest

Model Performances Comparison - Training

Table 1 - Training Data Performance

Model	Precision	Recall	F1-score	Accuracy
KNN	0.768	0.915	0.835	0.748
Decision Tree	0.739	0.990	0.846	0.748
SVM	0.749	0.975	0.847	0.755
LogisticRegression	0.753	0.965	0.846	0.755
Random Forest	0.738	0.990	0.846	0.748

Model Performances Comparison - Test

Table 2 – Test Data Performance

Model	Precision	Recall	F1-score	Accuracy
KNN	0.763	0.905	0.828	0.737
Decision Tree	0.741	0.989	0.847	0.751
SVM	0.752	0.974	0.849	0.758
LogisticRegression	0.756	0.964	0.847	0.757
Random Forest	0.741	0.989	0.847	0.751

Conclusion and Future Directions

- All models have similar performance with accuracy around 75%
- The model selection may depend on the intended purpose
- Each model has advantages and disadvantages
- Other factors may need to be included to improve model performance
- More discussion with model users to determine improvement needed.