Predicting collision severity

Predicting the severity of a collision is helpful for commuters

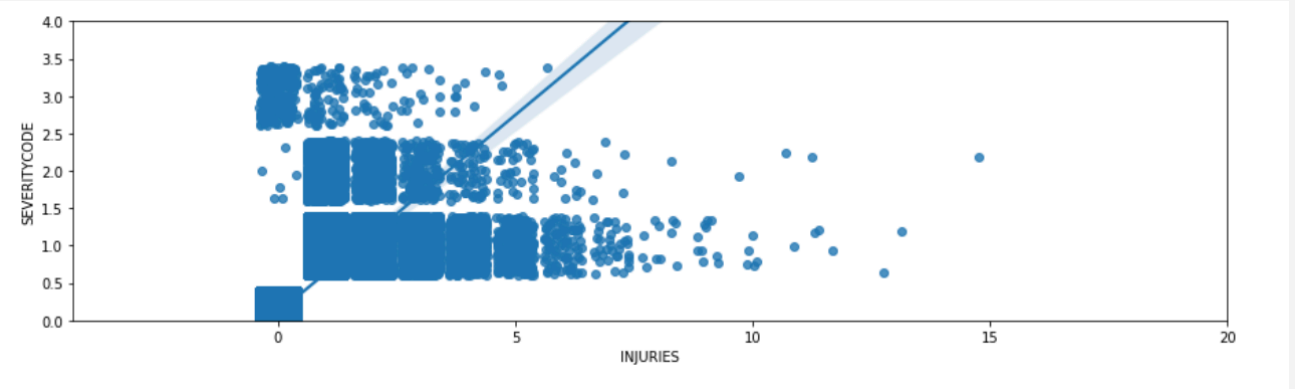
- In 2018, the United States had roughly 276 million vehicles in operation. Out of those vehicles, 12 million were involved in crashes.
- The United States is also among the countries with the highest rate of traffic-related fatalities per one million population.
- Reliable severity prediction of possible car accidents can help commuters to drive with more foresight or change their travel plans accordingly.

Data source and cleaning

- Collision dataset obtained from the Seattle Department of Transportation website
- The data was collected from 2004 to present for Seattle and each observation represents a collision.
- In total, about 220,000 rows, 39 attributes and one label.
- Duplicate, correlated or attributes where majority of data was missing were dropped.
- Cleaned data consists of 24 features.

The more people injured the higher the possibility of a larger severity code

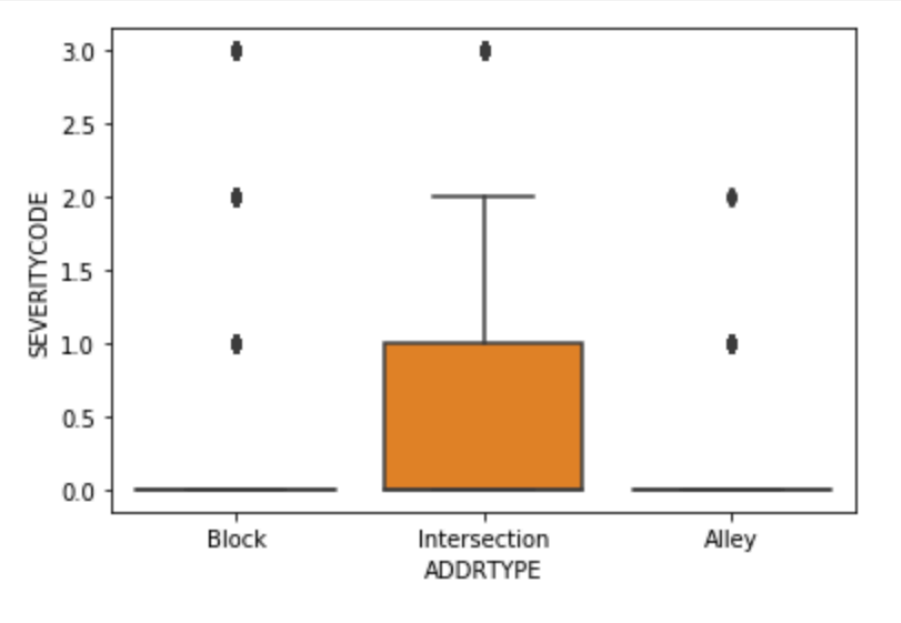
Strong correlation between number of injuries and Severity Code of the collision. ($p<0.001$, Kendall's rank coefficient 0.949)



Address type influences severity of an accident

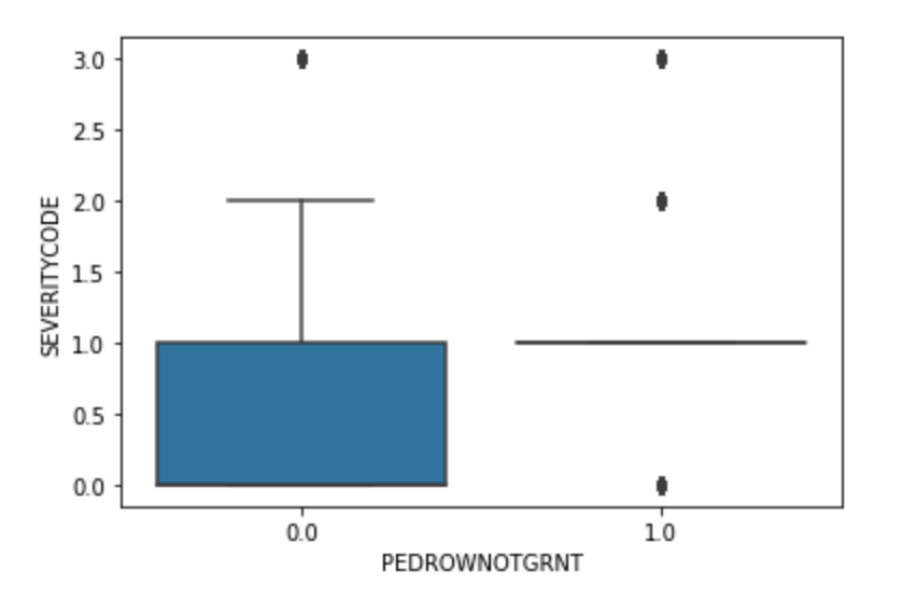
The severity of an accident is influenced by the location of the accident: Alley, Block or Intersection.

This makes sense because the traffic in an alley is significantly less, for example.



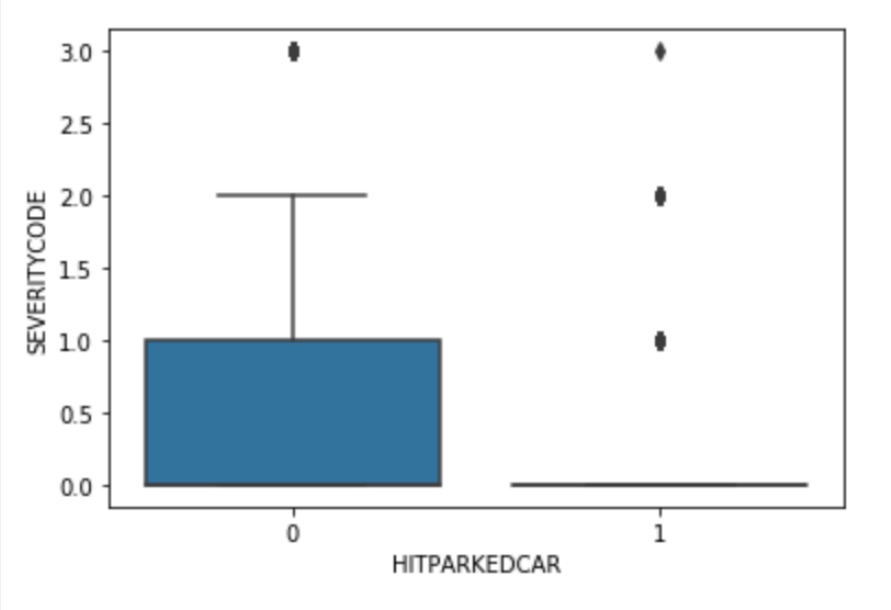
Keep an eye out for pedestrians.

If the pedestrian right of way is not granted, the possibility of a more severe collision increases.



The severity of an accident increases if you hit a parked car.

This one is obvious.



Classification models

- Logistic loss between 0.072 and 0.730
- Accuracy between 0.707 and 0.983
- In total, about 220,000 rows, 39 attributes and one label.
- Decision trees performed best, but differences were very small.

	Naive Bayes	Logistic Regression	K-nearest Neighbors	Decision Trees
Accuracy	0.707	**0.983**	**0.983**	**0.983**
Logarithmic loss	0.730	0.074	0.367	**0.072**

Discussion and conclusion

- Built classification models to predict the severity of an accident , which can be useful for commuters to drive with more foresight or change their travel plans accordingly when a collision occurs on their way to work.
- Most features in the dataset did not have a good correlation with Severity, might be worth combining the set with other collision data
- Would be good to not only consider injuries and fatalities when looking at severity, but also look at the impact the collision has on the traffic, e.g. traffic delays for other commuters.
- Model can be improved by not only predicting the severity of an accident, but to also predict whether or not a person might be involved in an accident when driving a certain route at a certain time and what severity that accident might have.