Predicting Severity of Traffic Accidents in Seattle

Predicting severity (injury or not) of traffic accidents may be useful for local government and city council as it is needed to

- Prevent severe traffic accidents by understanding the factors involved such as road conditions
- Optimize resources to be deployed on scene based on the predicted severity of the accident
- To gauge the time needed for traffic to return to normal, in order to plan for road diversions to ease traffic, based on predicted severity of the accident

Data Acquisition and Cleaning

Data Source:

Collisions dataset provided by the Traffic Records Group, Seattle Department of Transportation (SDOT). Only records from years 2017, 2018 and 2019 were used for this study.

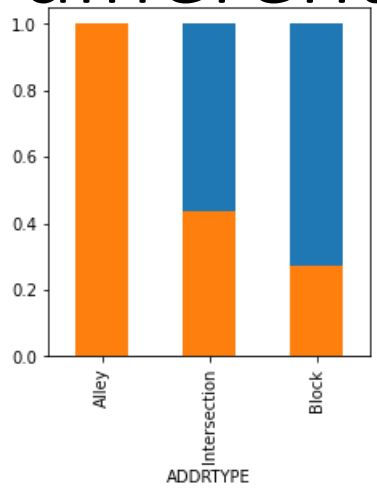
Dropped variables:

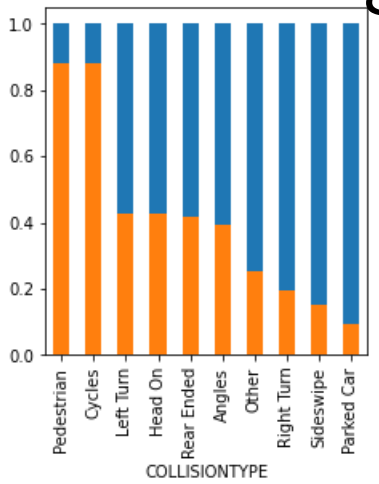
Granular location (specific intersection), coordinates, date and time, descriptions of codes (duplicate since same as code)

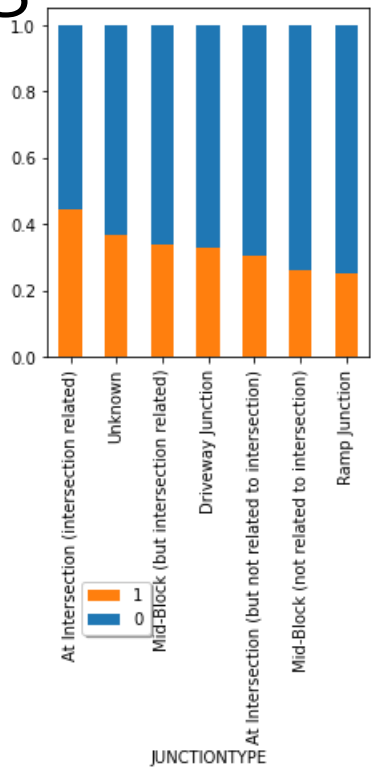
Final Dataset contains 28,218 records and 17 features

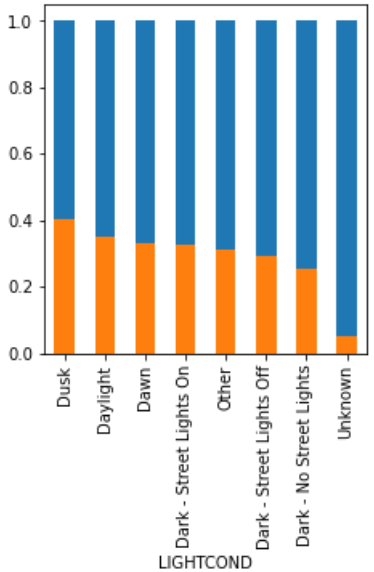
Feature	Description
ADDRTYPE	Collision address type: Alley, Block, Intersection
COLLISIONTYPE	Collision type
PERSONCOUNT	The total number of people involved in the collision
PEDCOUNT	The number of pedestrians involved in the collision
PEDCYLCOUNT	The number of bicycles involved in the collision
VEHCOUNT	The number of vehicles involved in the collision
JUNCTIONTYPE	Category of junction at which collision took place
SDOT_COLCODE	A code given to the collision by SDOT
INATTENTIONIND	Whether or not collision was due to inattention
UNDERINFL	Whether or not a driver involved was under the influence of drugs or alcohol
WEATHER	A description of the weather conditions
ROADCOND	The condition of the road during the collision
LIGHTCOND	The light conditions during the collision
PEDROWNOTGRNT	Whether or not the pedestrian right of way was not granted
SPEEDING	Whether or not speeding was a factor in the collision
ST_COLCODE	A code provided by the state that describes the collision
HITPARKEDCAR	Whether or not the collision involved hitting a parked car

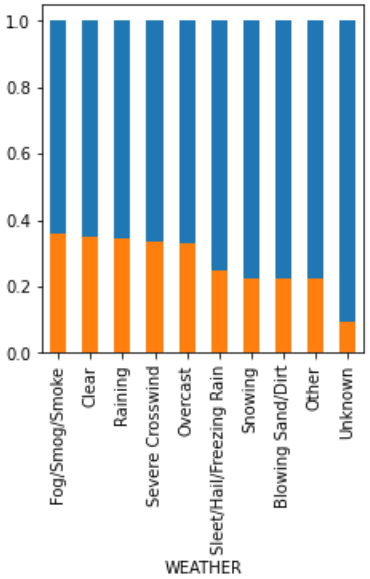
Various factors such as light condition, road condition, weather, and junction type have different percentages of severe accidents

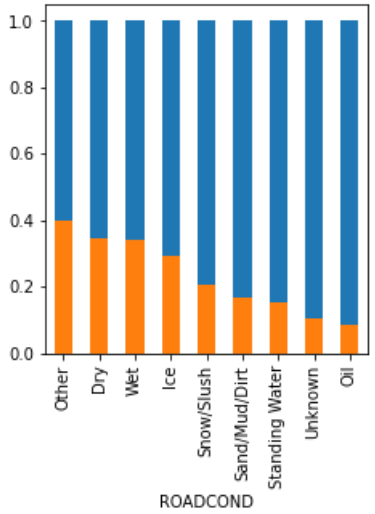




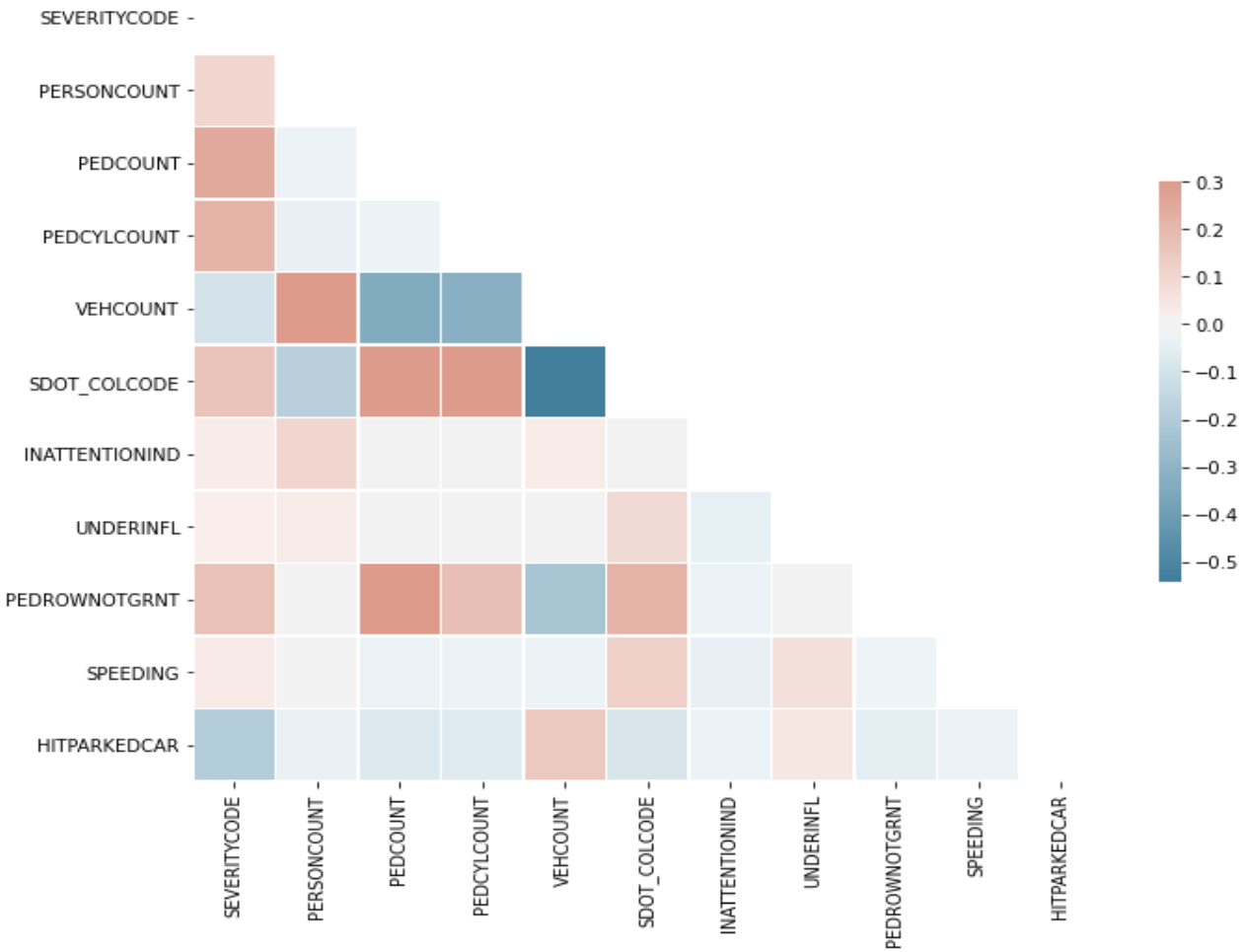








No strong correlation among numerical variables – no multicollinearity



4 Classification Models were built

- K Nearest Neighbors (KNN)
- Decision Tree (DT)
- Support Vector Machine (SVM)
- Logistic Regression (LR)

Model Evaluation on Test Set – Logistic Regression is selected as the best model

	KNN	Tree	SVM	LR
Accuracy	0.72	0.72	0.72	0.73
Precision	0.64	0.85	0.79	0.73
Recall	0.37	0.21	0.24	0.32
F1-Score	0.69	0.66	0.67	0.70
Jaccard Score	0.31	0.20	0.23	0.29
AUC	0.72	0.75	0.73	0.77

Conclusion

- Models were built to predict severity (injury or not) of traffic accidents in Seattle
- Logistic Regression model was selected as the best model
- Modelling is a reiterative process and there is still room for improvement.