



ROAD ACCIDENT'S SEVERITY

Capstone Project

Benitez Samuel

August, 2020

BACKGROUND



1.3 million people die every year due to traffic accidents.



Between 20 and 50 million suffer non-fatal injuries.



Pedestrians, cyclist and riders account for half of all these deaths.



1: PROPERTY DAMAGE

BUSINESS PROBLEM



2: INJURY

- Today, having a vehicle is practically a basic necessity for people anywhere in the world. However, road accidents are part of the day to day in cities around the world. For this reason, it is important for any driver to be able to avoid an accident, especially those of high severity, which can result in a catastrophic end.
- Being able to predict the severity or probability of an accident is highly useful for the well-being of any driver or pedestrian.
- This would undoubtedly help all drivers to have a greater perspective on possible accidents and thus make the best decision about whether to drive, not drive or do it safely.



DATA

SEATTLE DEPARTMENT OF TRANSPORTATION (SDOT) DATASET

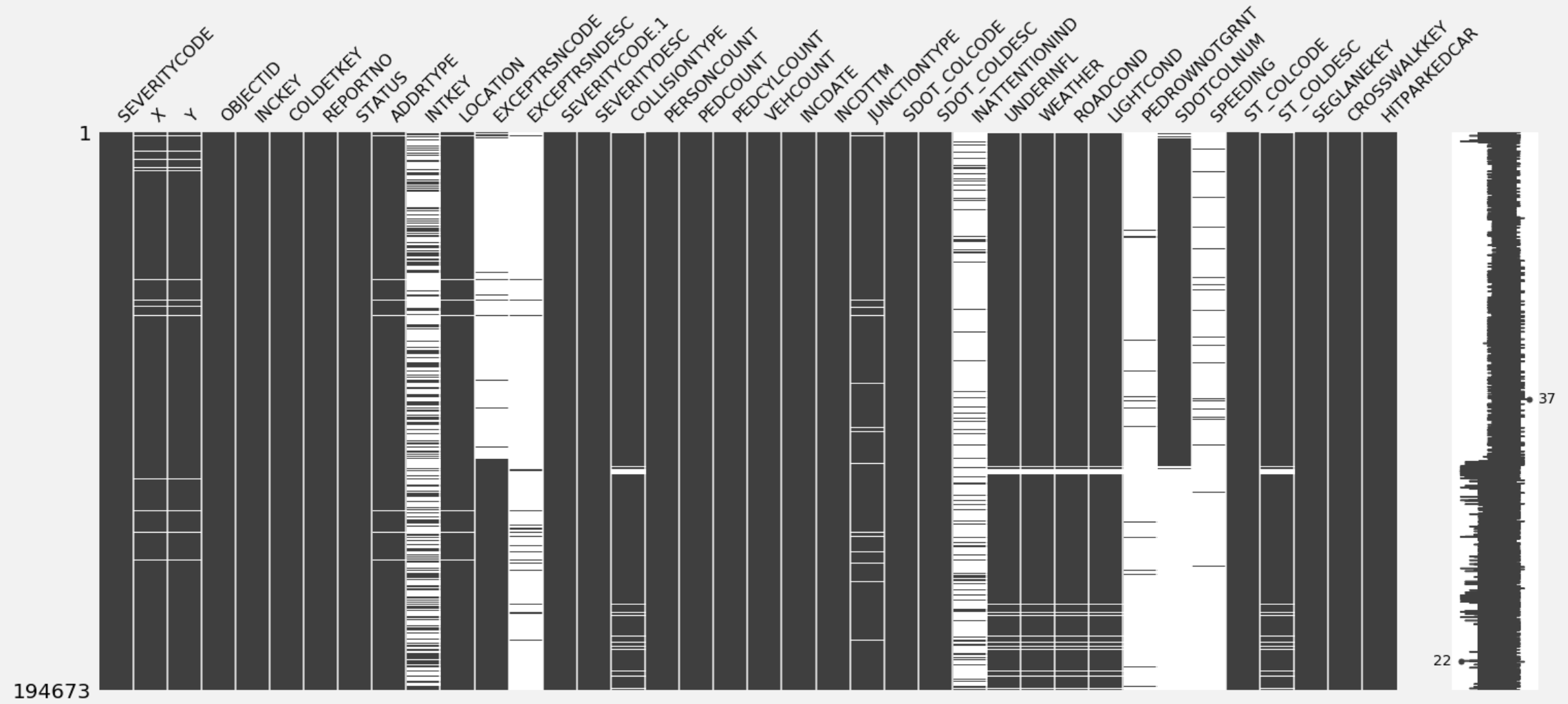
Number of attributes: 38

Number of rows: 194673

IMPORTANT ATTRIBUTES

- Accident details
- Location and time
- People affected
- Driver's fault
- Road conditions
- Weather conditions

DATA BEFORE PROCESSING

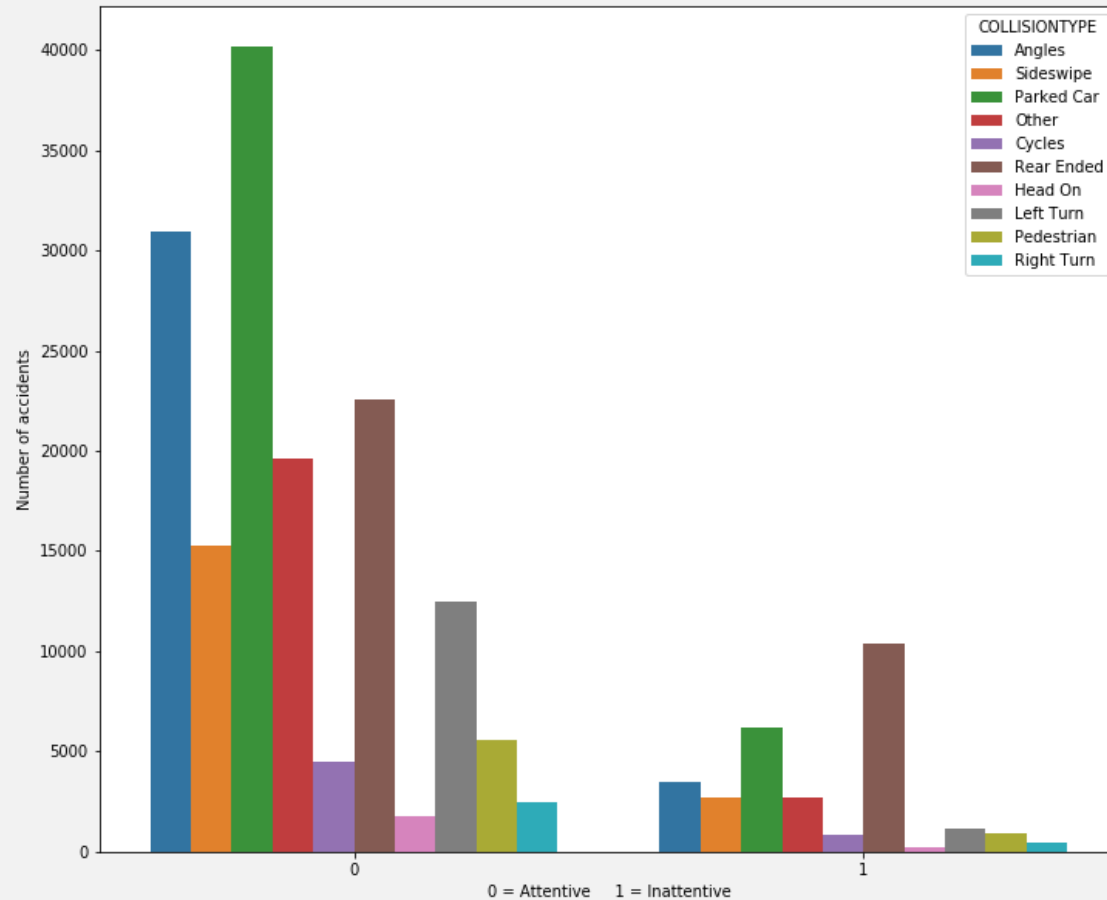


DATA AFTER CLEANING

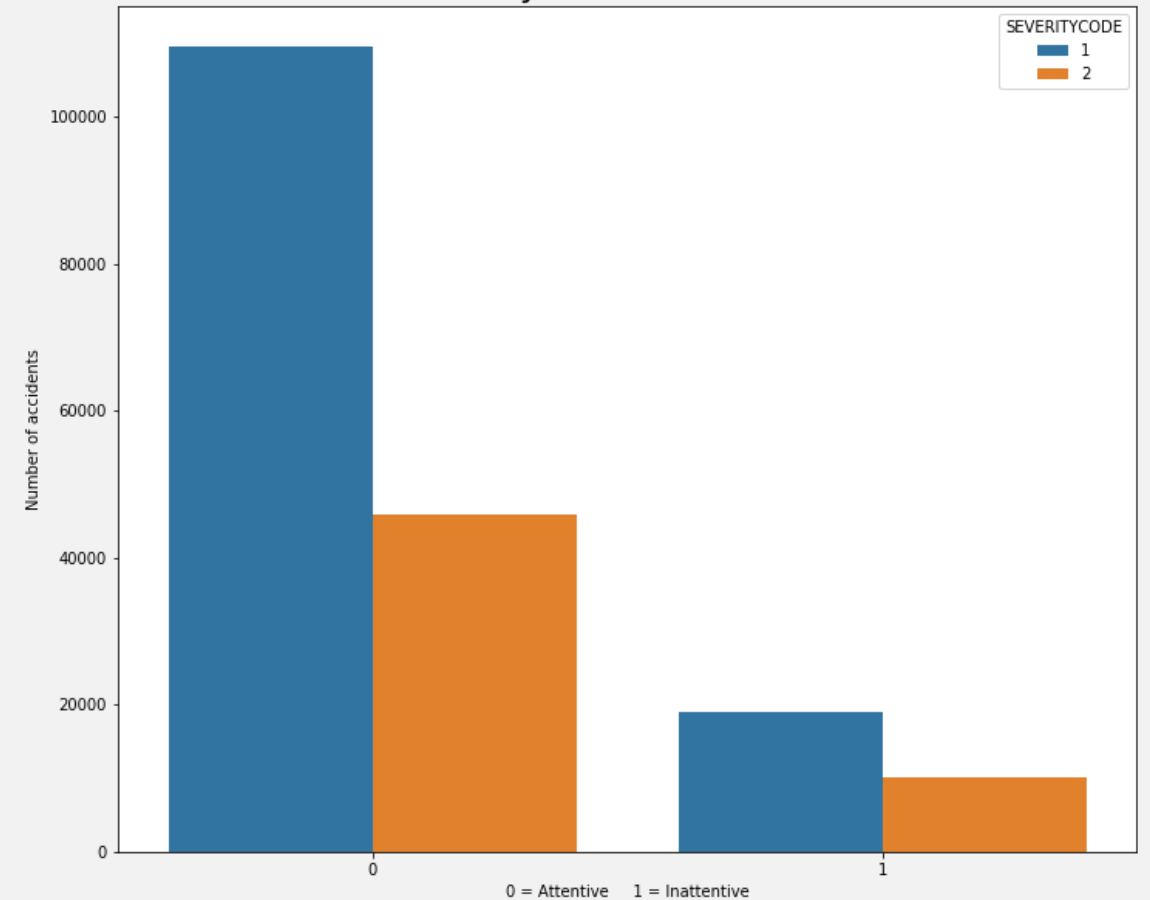
[illegible]

EXPLORATORY DATA ANALYSIS (EDA)

Accidents due to inattention

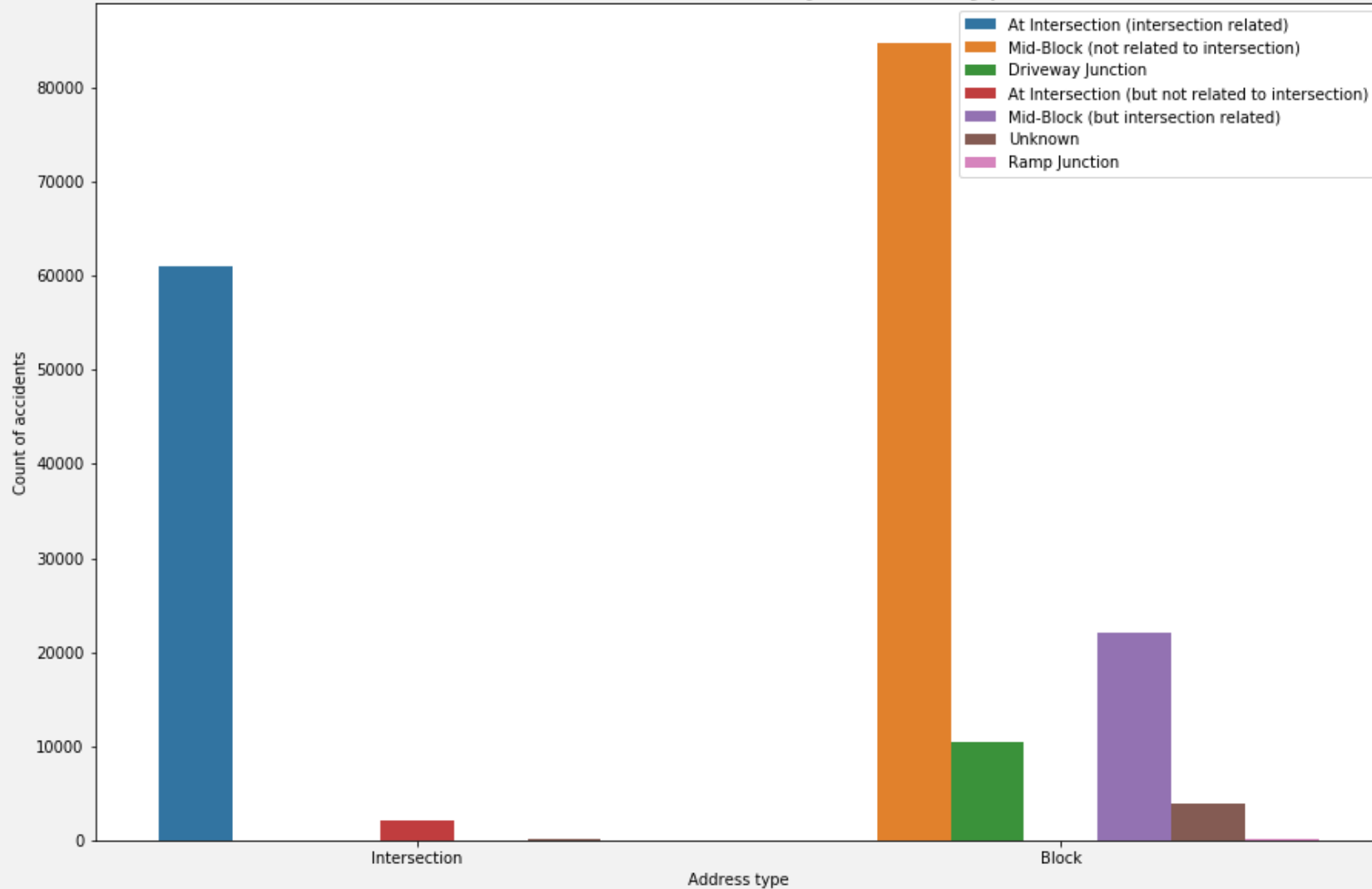


Severity due to inattention



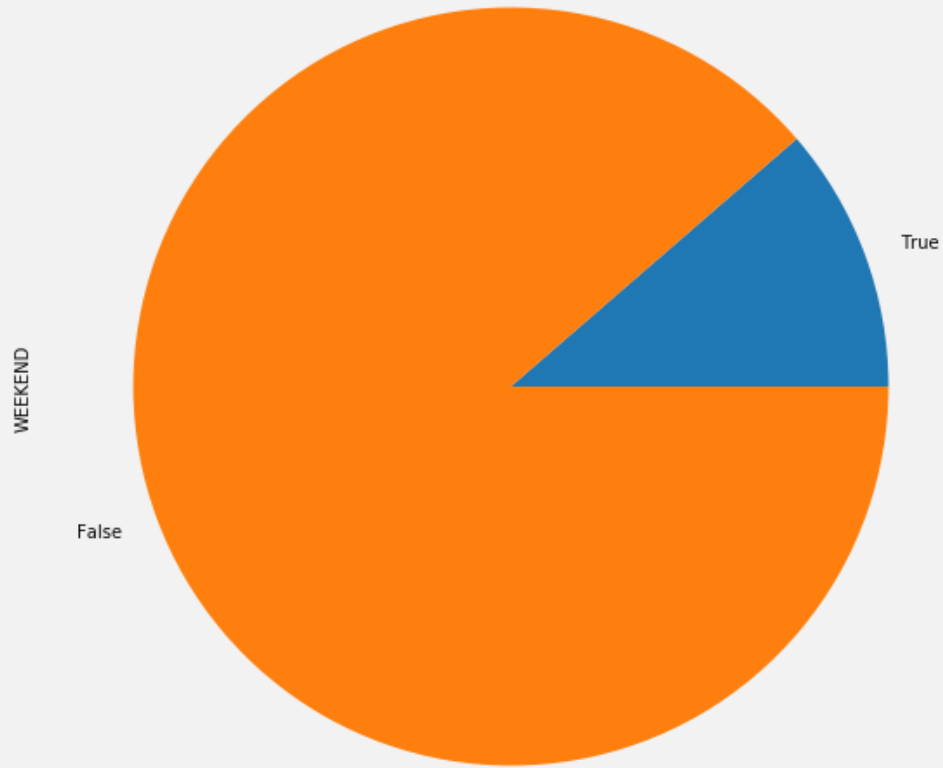
Inattention is the most common cause of accidents, speeding and intoxication don't cause many accidents in Seattle.

Accidents at different junction types



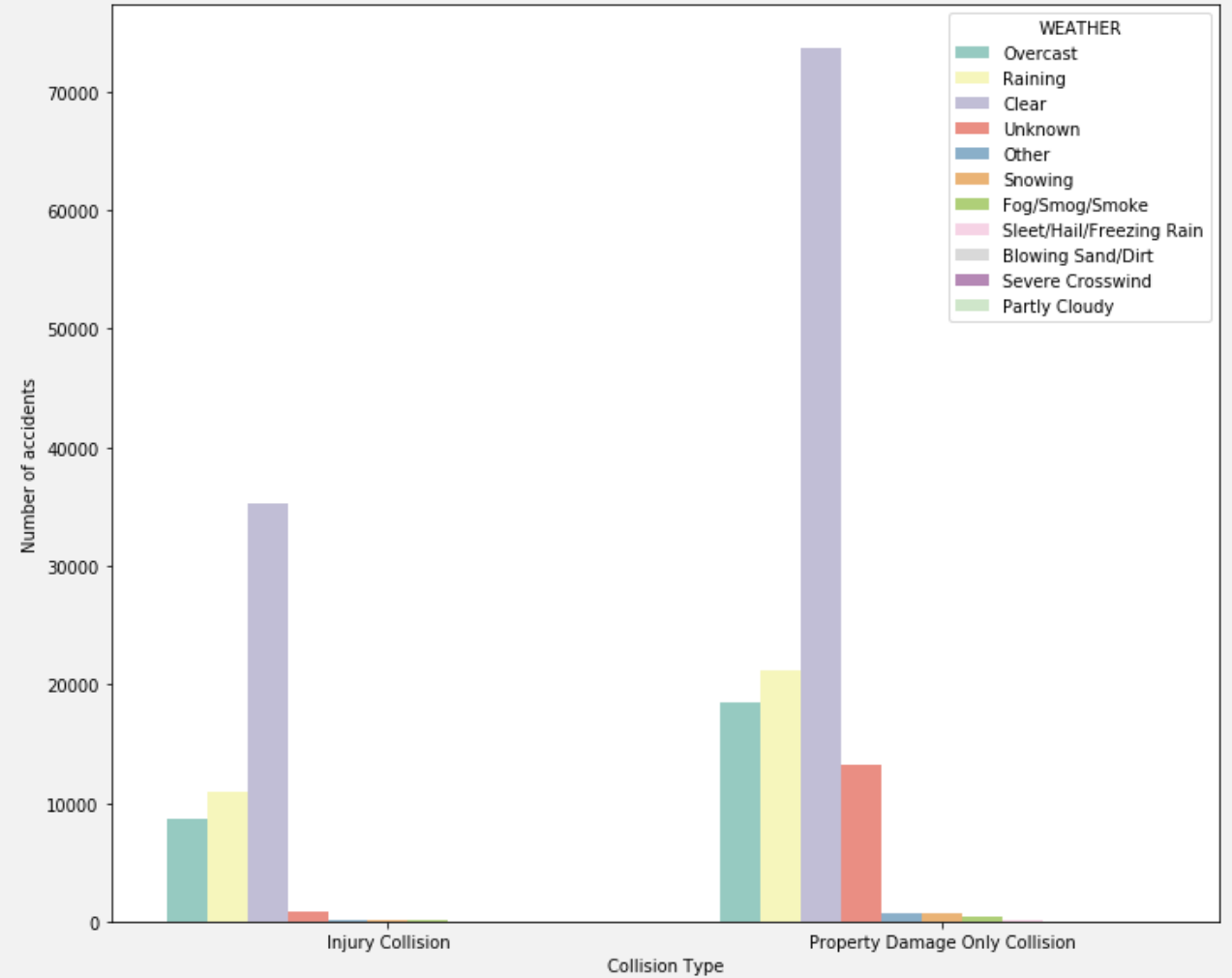
- After cleaning the data, 751 Alley collisions were dropped.
- Most accidents occur at mid-block and intersections.

Percentage of collisions in weekends



- Most accidents occur on weekdays, less than 15% on the weekend.

Collisions and Weather condition



- It seems that the weather is not a key factor. However, even though most accidents happen in clear weather, there are also many others during overcast and rainy days.

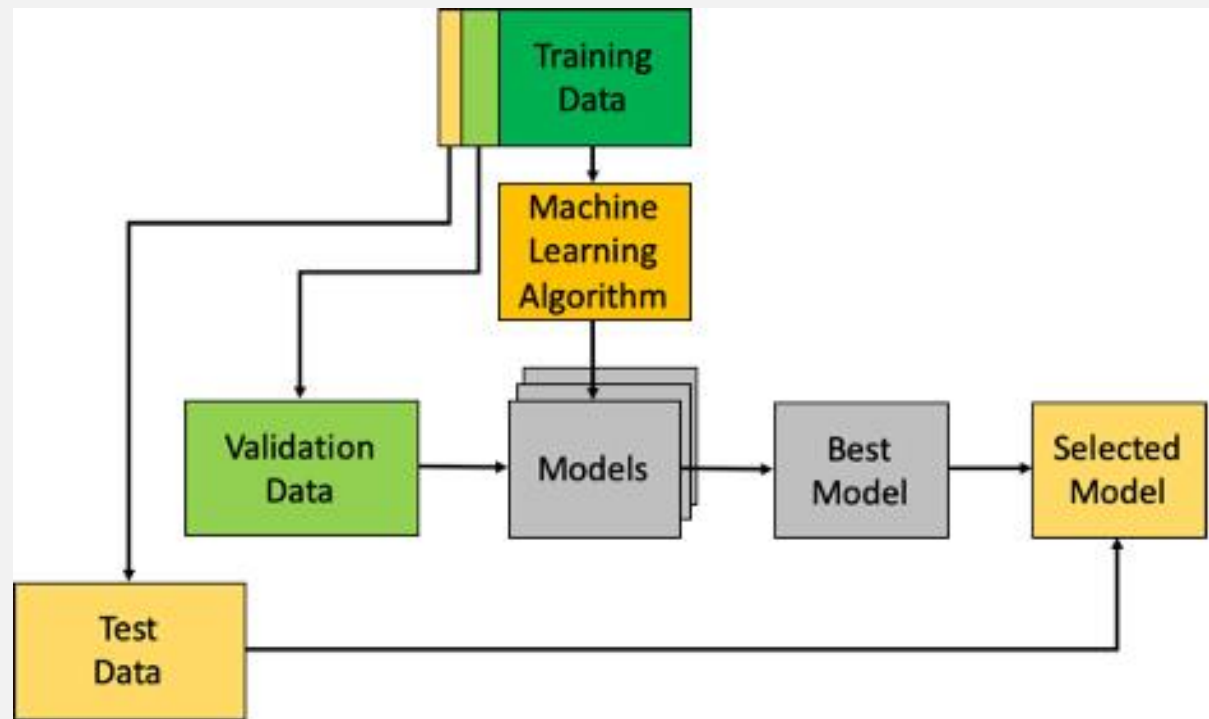
FEATURE SELECTION

```
In [20]: features.corr().style.background_gradient(cmap='Blues').format("{:.3f}")
```

	SEVERITYCODE	PERSONCOUNT	PEDCOUNT	PEDCYLCOUNT	VEHCOUNT	SDOT_COLCODE	INATTENTIONIND	UNDERINFL	PEDROWNOTGRNT
SEVERITYCODE	1.000	0.128	0.249	0.216	-0.085	0.183	0.044	0.043	0.209
PERSONCOUNT	0.128	1.000	-0.027	-0.042	0.395	-0.143	0.072	0.022	-0.031
PEDCOUNT	0.249	-0.027	1.000	-0.018	-0.321	0.270	-0.007	0.015	0.495
PEDCYLCOUNT	0.216	-0.042	-0.018	1.000	-0.311	0.397	0.002	-0.018	0.323
VEHCOUNT	-0.085	0.395	-0.321	-0.311	1.000	-0.469	0.048	-0.008	-0.280
SDOT_COLCODE	0.183	-0.143	0.270	0.397	-0.469	1.000	0.025	0.106	0.247
INATTENTIONIND	0.044	0.072	-0.007	0.002	0.048	0.025	1.000	-0.028	-0.029
UNDERINFL	0.043	0.022	0.015	-0.018	-0.008	0.106	-0.028	1.000	-0.019
PEDROWNOTGRNT	0.209	-0.031	0.495	0.323	-0.280	0.247	-0.029	-0.019	1.000

PERSONCOUNT, *PEDCOUNT*, *PEDCLYCOUNT* are directly related to the target variable, those are the 3 most correlated, followed by *SDOT_COLCODE*, *PEDROWNOTGRNT*, *ADDRTYPE* and *COLLISIONTYPE*.

MODELING



	Model	Accuracy Score	F1 Score	Precision	Recall
0	Logistic Regression	0.746350	0.841056	0.743927	0.967358
1	K-Nearest Neighbour	0.733158	0.825822	0.754637	0.911835
2	Decision Tree	0.747325	0.841646	0.744526	0.967904
3	Support Vector Machine	0.746323	0.843224	0.738044	0.983367

DISCUSSION

1

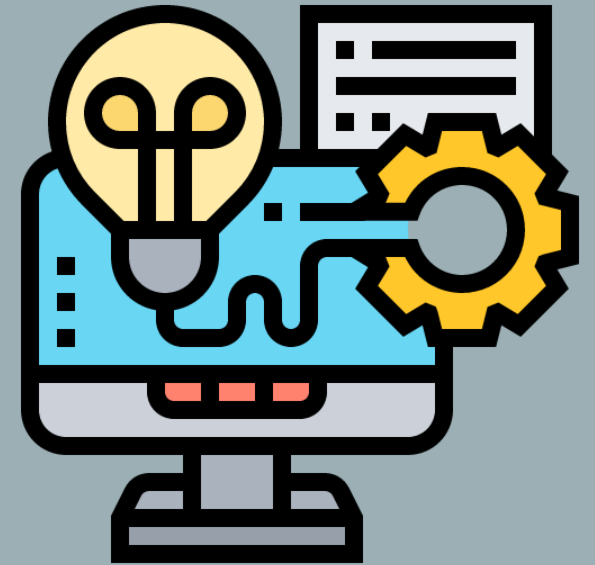
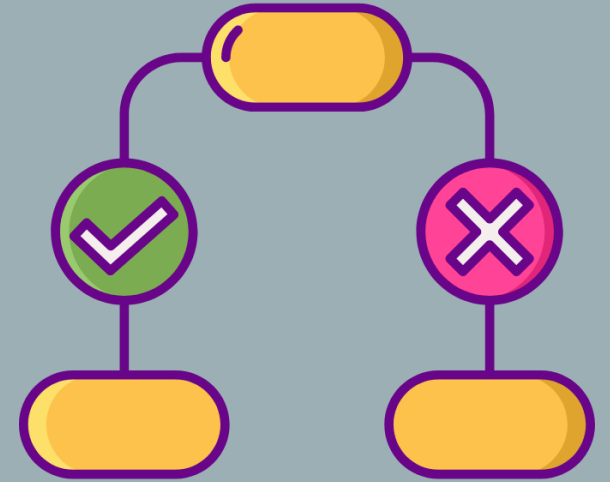
The best model was the Decision Tree, although it is possible that the KNN model may work better finding the optimal K.

2

The Logistic Regression model performed well and, on its advantages, it could tell the probability for both classes of the target value.

3

Obviously, these models can improve if the feature sets are handled better. For now, we can consider them good models for project purposes.



CONCLUSION

- Most collisions occur during **day light** on **weekdays** and commonly against **parked cars**.
- Most accidents happen at **street intersections**.
- Surprisingly, **speeding** and **intoxication** do not cause too many collisions in Seattle.
- On the other hand, **inattention** is the most common cause of an accident.
- **Weather conditions** do not play a significant role.



FUTURE DIRECTIONS

Undoubtedly, a model with these characteristics working with real time data and alert notifications (whether it is appropriate or not to drive under certain conditions) would significantly reduce crashes in many cities around the world.

Additionally, this model serves to make all the population aware of the importance of being careful while driving, not only driving properly but also taking care of others on the road.

