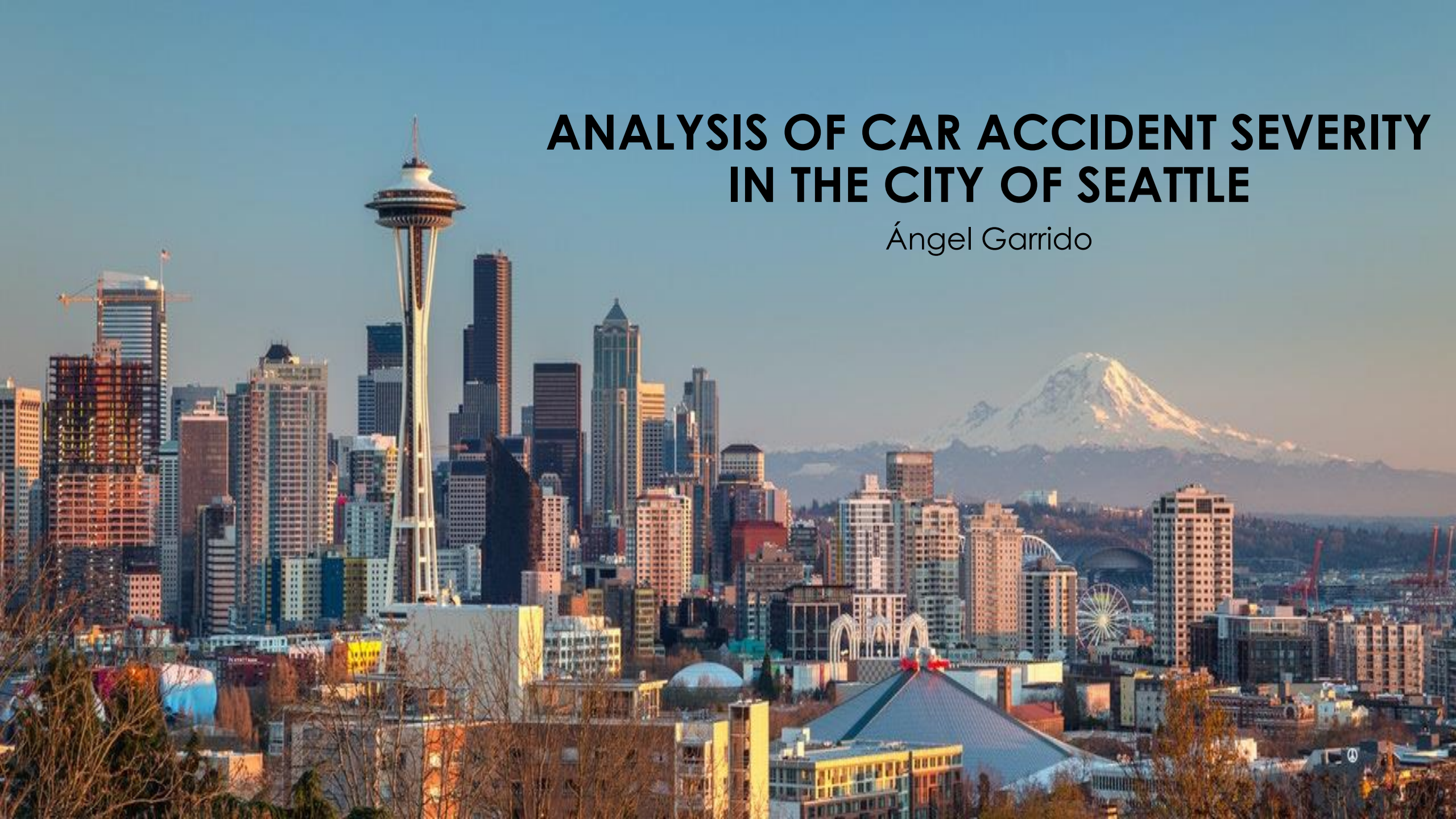


ANALYSIS OF CAR ACCIDENT SEVERITY IN THE CITY OF SEATTLE

Ángel Garrido

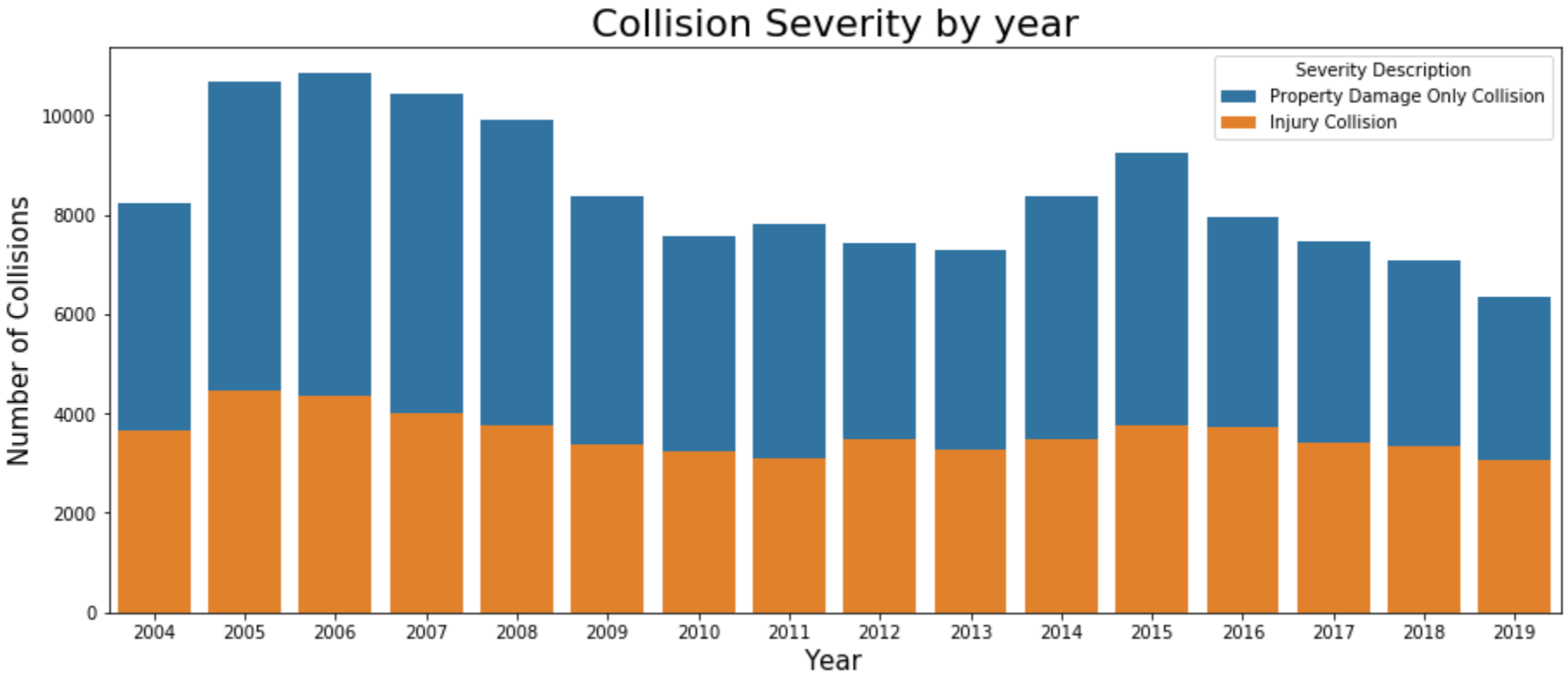
PREDICTING TRAFFIC ACCIDENTS IS VALUABLE FOR THE CITY OF SEATTLE

- There are multiple benefits from traffic accident prediction. Reducing traffic accidents is both an important public safety challenge and critical to prevent traffic congestions.
- Prediction of traffic accident will help us to improve public transportation, to design cost-effectively transportation infrastructure and to enable safer routes.
- In this project the data of Seattle traffic accidents will be explored in order to discover the variables with the strongest influence to traffic collisions.

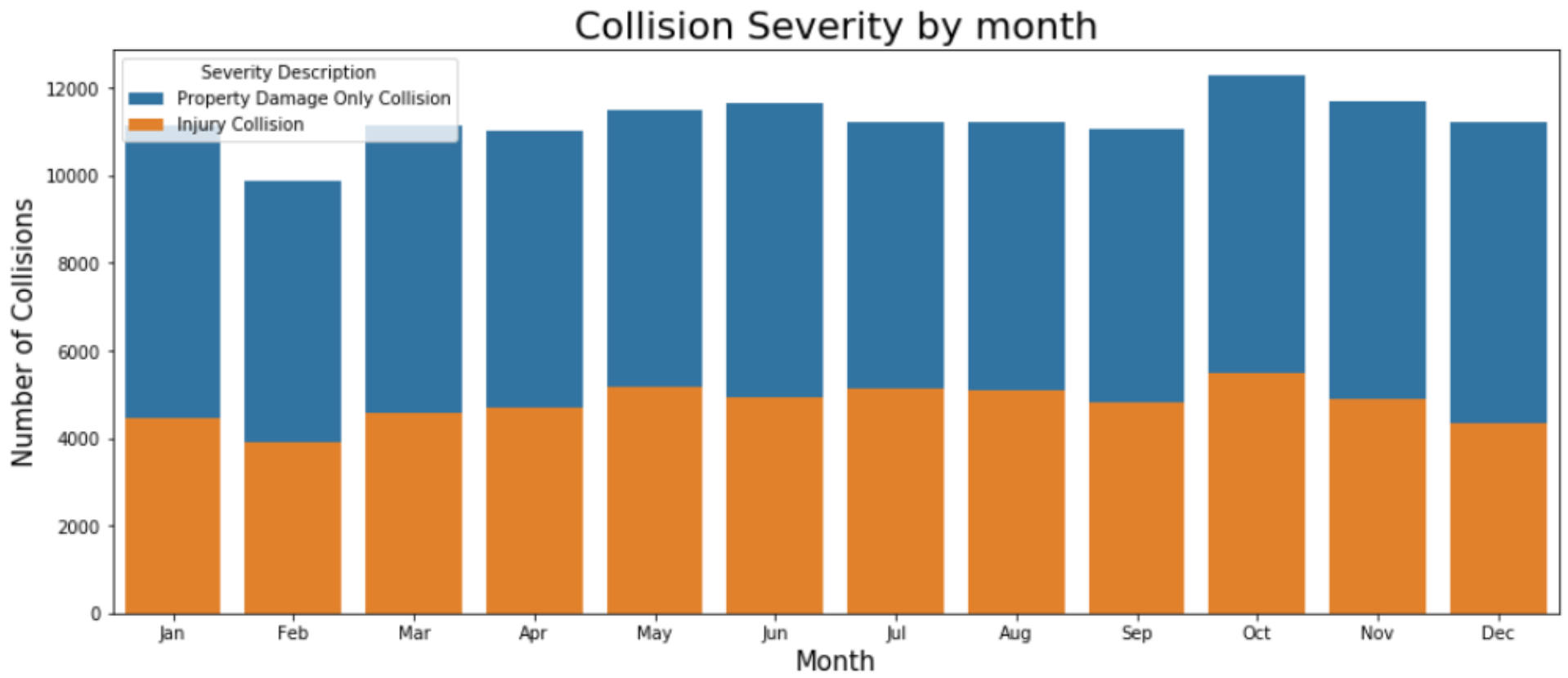
DATA ACQUISITION AND CLEANING

- Traffic accident (2004-2020) from Seattle has been downloaded from the IBM server.
- In total, 194,673 rows and 37 features in the raw dataset.
- The target variable, “severitycode” has two outputs in our data set:
  - 1 - Property Damage Only Collision
  - 2 – Injury Collision.
- Cleaned data has 166,803 rows and 10 features.

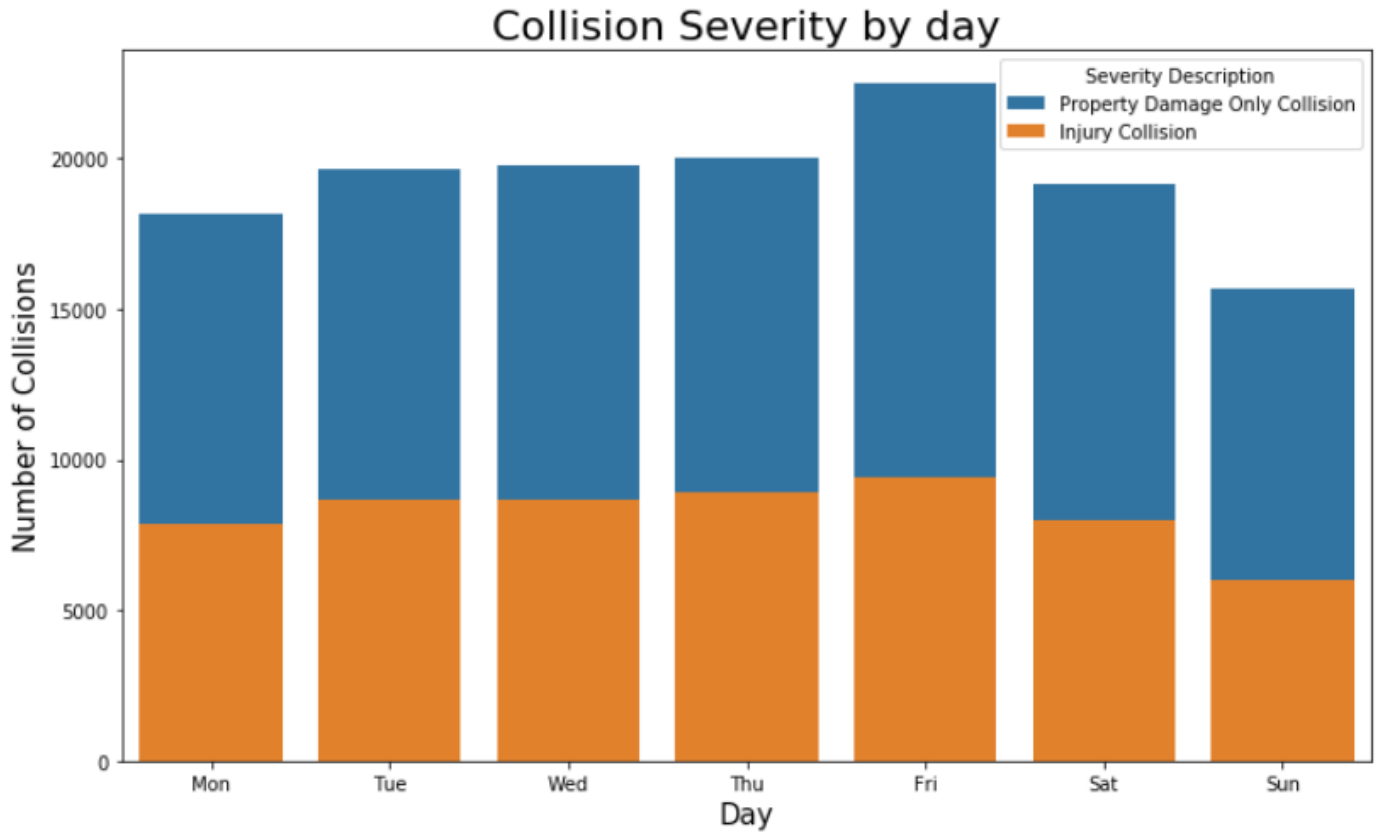
COLLISIONS DECREASE OVER YEARS



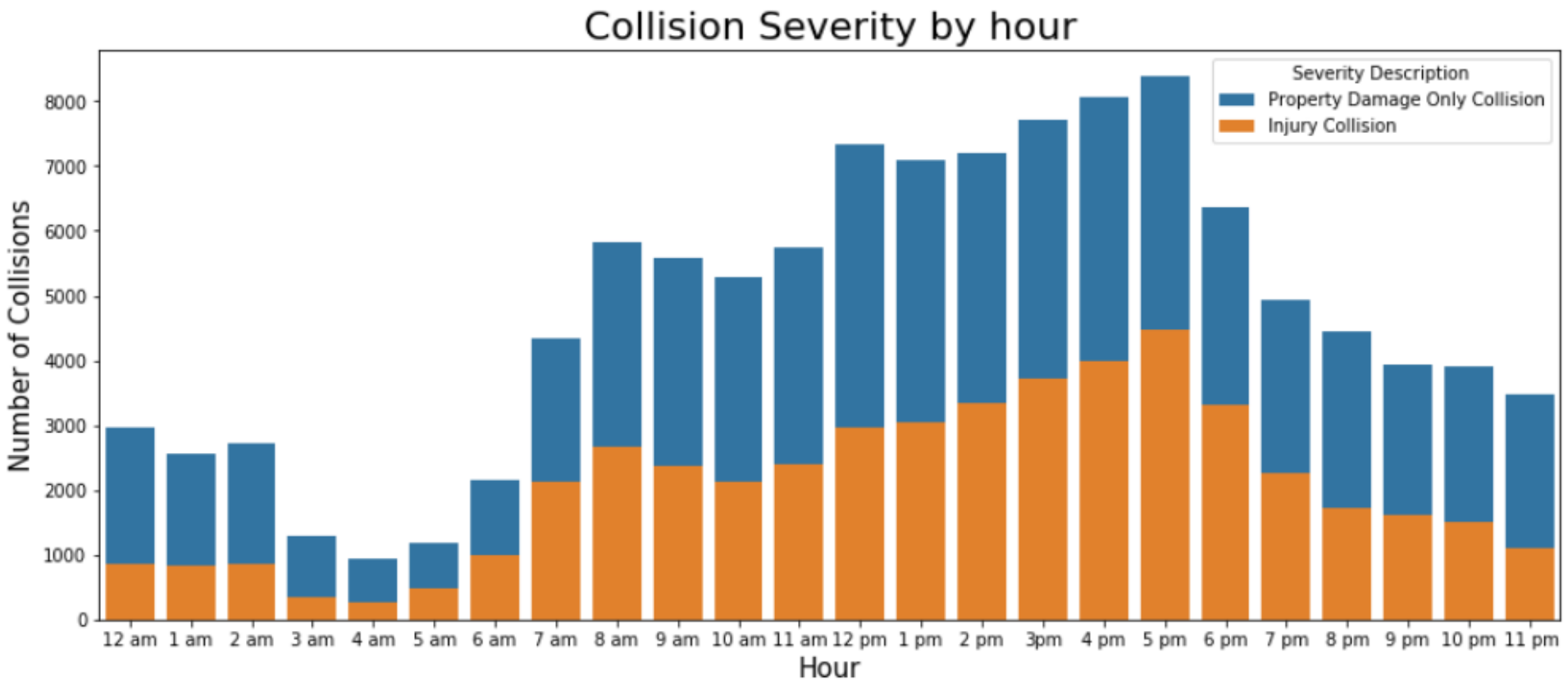
ACCIDENTS DISTRIBUTE MONTHLY EVENLY



SUNDAYS HAVE THE LOWEST RATES OF ACCIDENTS

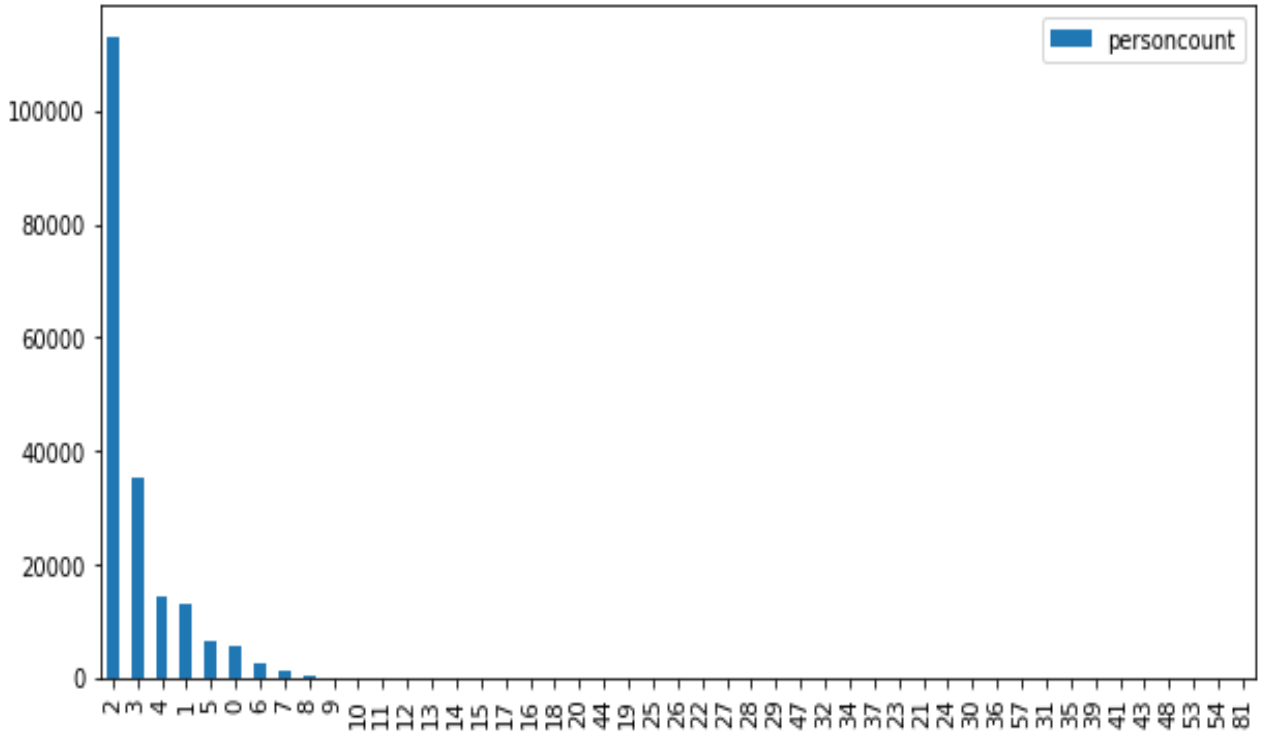


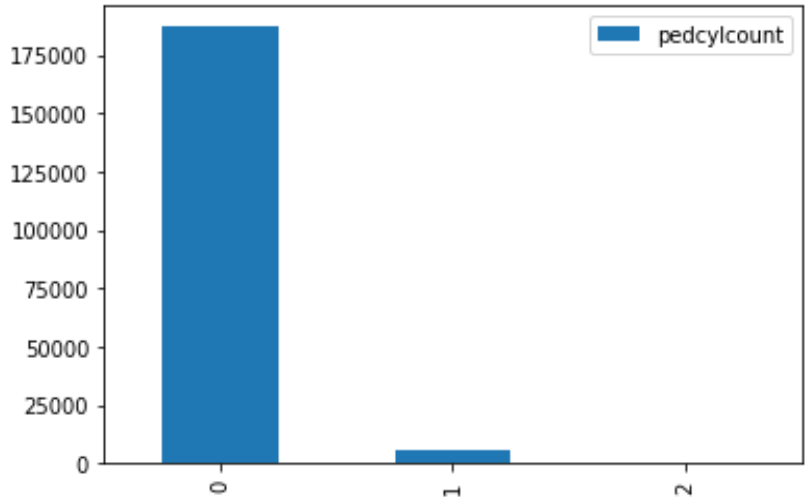
THERE ARE MORE COLLISIONS DURING DAYTIME
AND THEY ARE MORE SERVERE

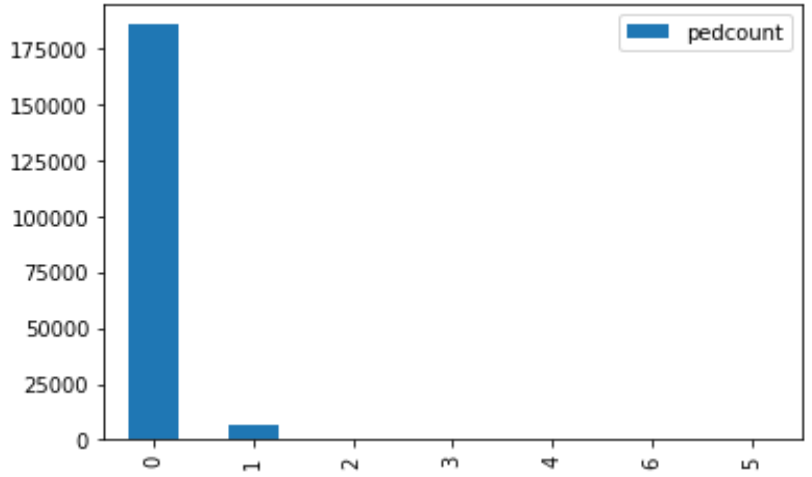


MOST TRAFFIC ACCIDENTS INVOLVE 2-3 PEOPLE.

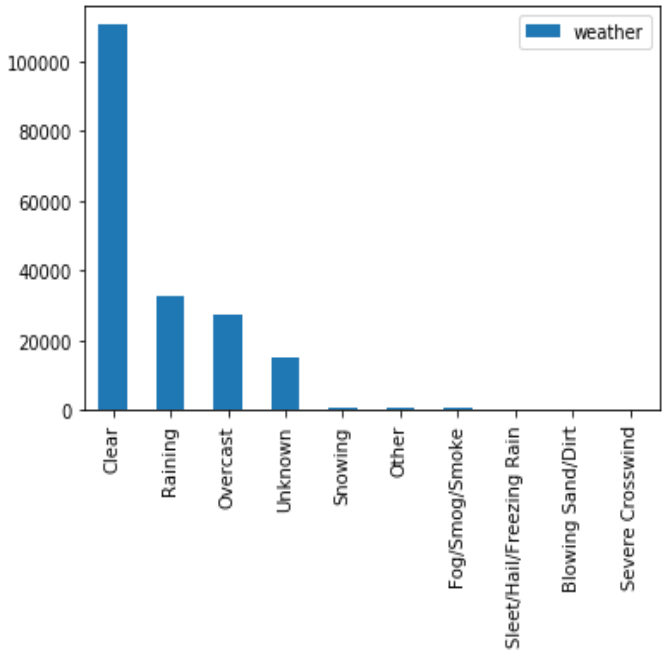
NORMALLY, THERE IS NO BICYCLE OR PEDESTRIAN INVOLVED

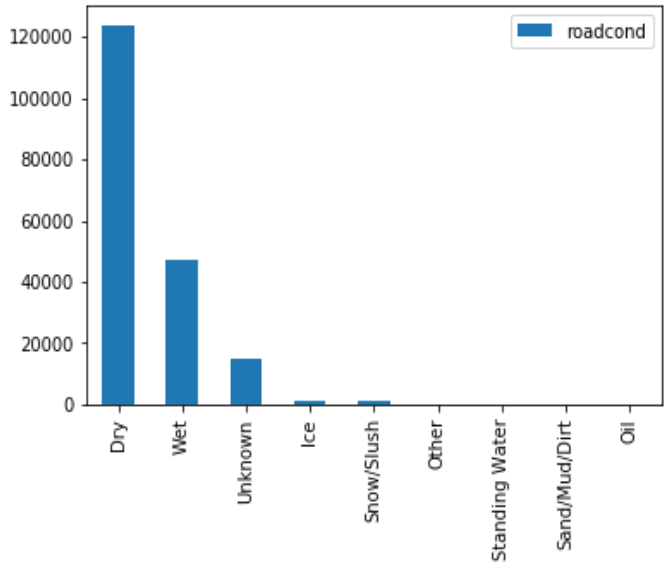


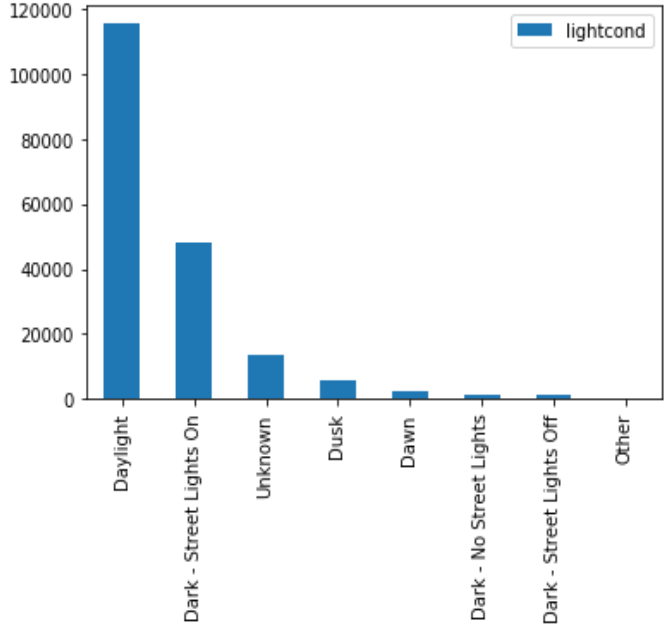




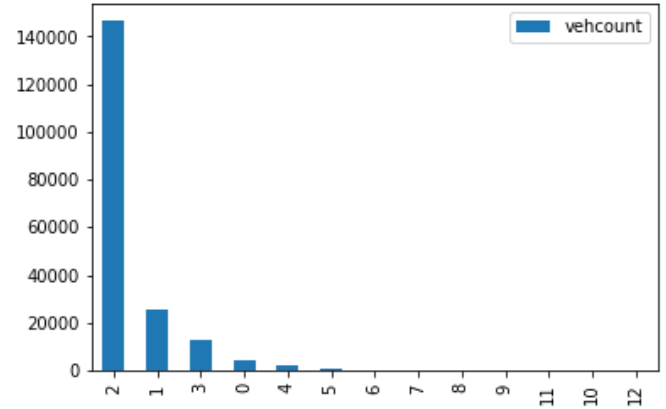
GOOD EXTERNAL CONDITIONS FOR DRIVING

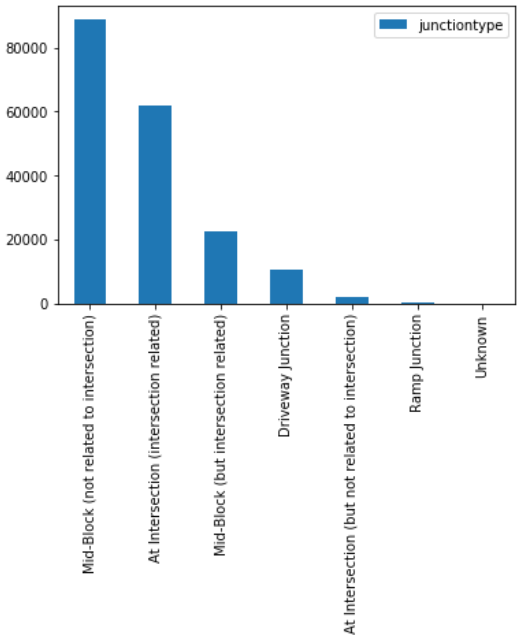


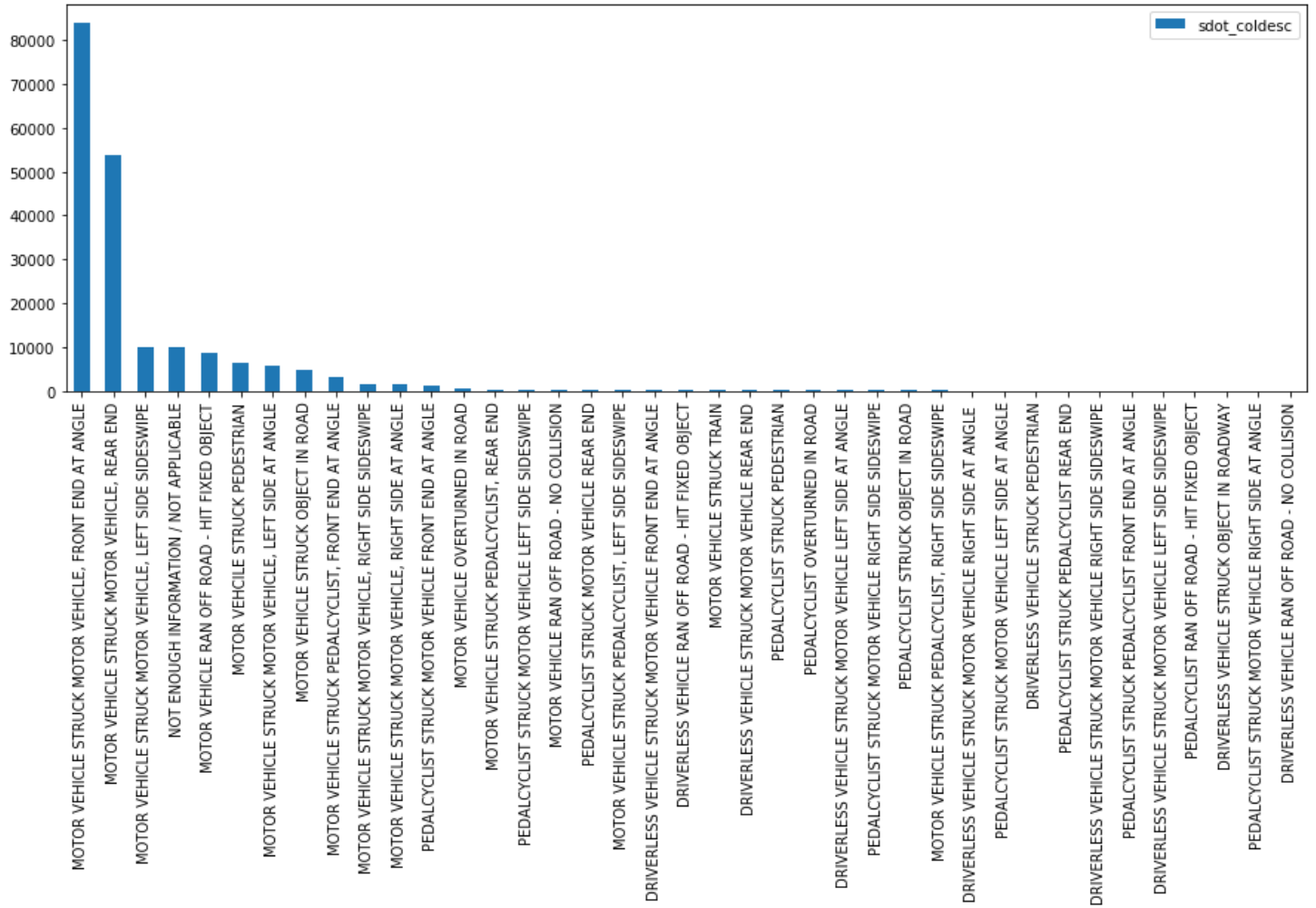




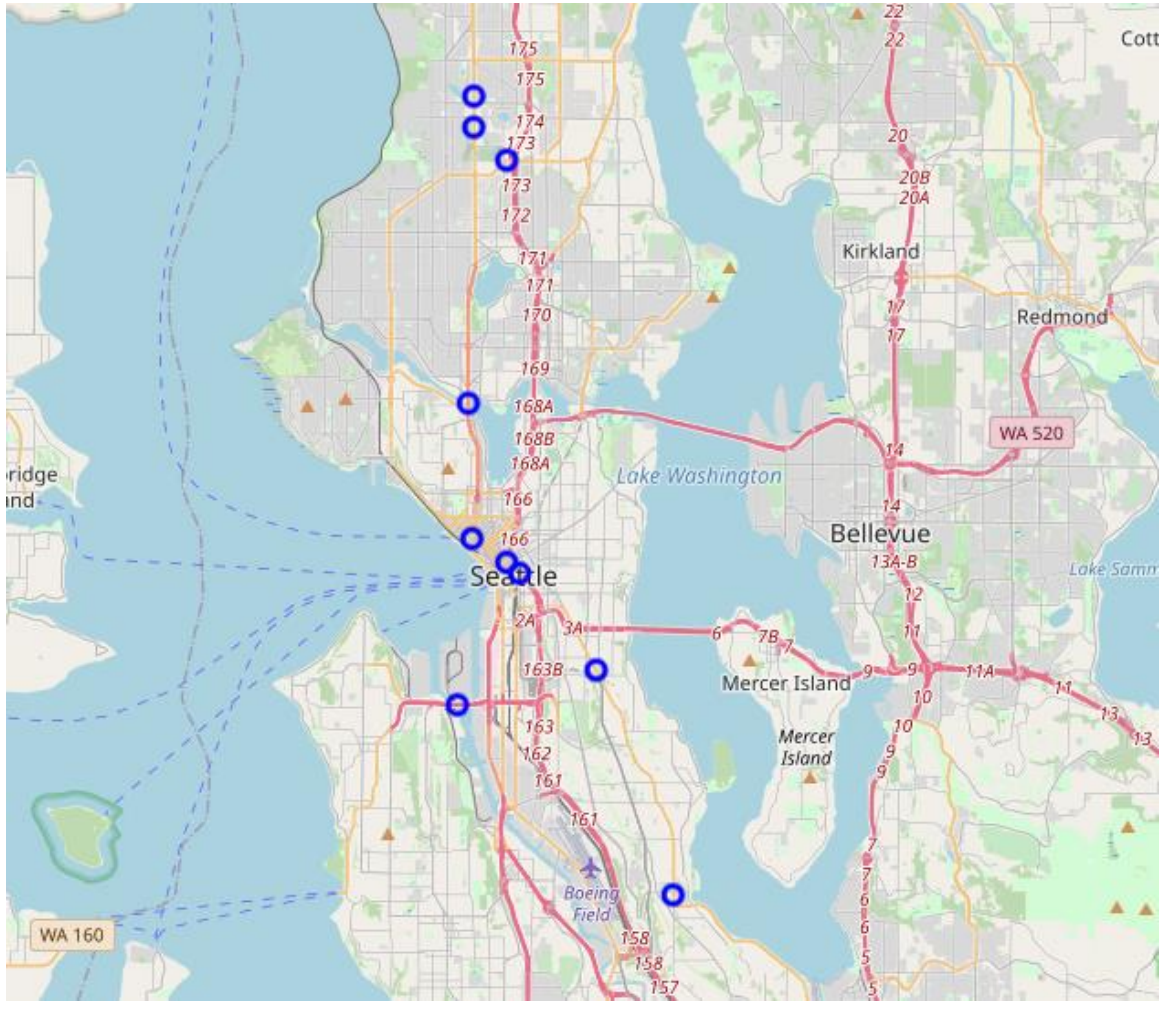
MOST ACCIDENTS ARE CAUSED BY TWO VEHICLES, IMPACTING AT THE REAR OR FRONT END







TOP 10 LOCATIONS
WITH MOST TRAFFIC ACCIDENTS



Top 10 locations concentrate in:

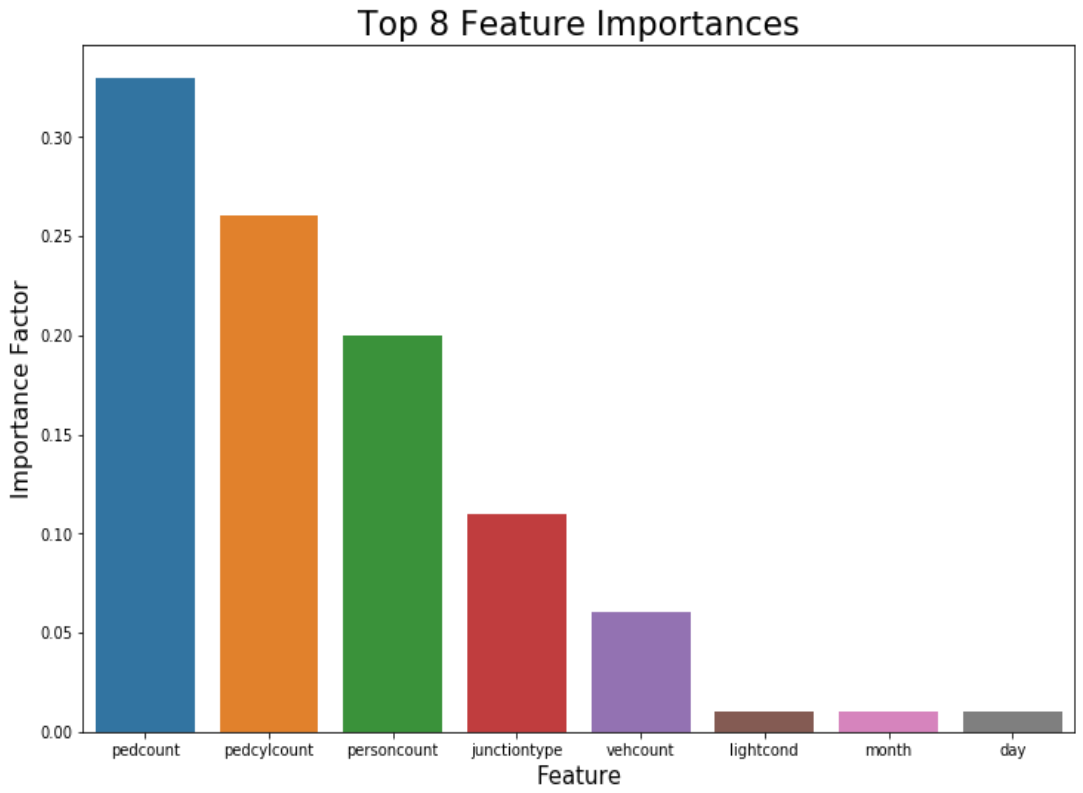
- the city center (3 locations)
- at the north of the city (3 locations)
- 2 bridges
- 2 locations of the same avenue

DECISION TREE IS THE MOST ACCURATE MODEL

	Decision Tree	K-Nearest Neighbors	Logistic Regression	Support Vector Machine
Jaccard Score	0.6472	0.6399	0.6240	0.6444
F1 Score	0.6472	0.6376	0.6111	0.6409
Injury Class F1 Score	0.6418	0.6070	0.5381	0.6041
Injury Class Recall Score	0.6370	0.5605	0.4414	0.5469

Jaccard and F1 Score were used to analyses the results of the ML models. Each model had similar metrics but Decision Tree was the strongest.

INFLUENCE OF FEATURES IN THE MODEL



- According to the graph, Decision Tree Classifier assigns more importance to ‘pedcount’ and ‘pedcylcount’
- This seems logical, as pedestrians and cyclists receive the impact with low protection

CONCLUSION

- In this project, the traffic accidents of Seattle since 2004 have been studied. Several models have been built helped by python and sklearn. In the process, data has been analyzed, plotted and processed to achieve the best model possible.
- It is interesting the influences of the variables studied in the outcome of an accident. From this analysis it can be concluded that a special attention has to be made in protect pedestrians and cyclists as they have contributed the most to the severity of the accident.
- On the other hand, specific actions should be taken to areas with higher rates of accidents, and possible of traffic jams. Studying in more detail these areas and lowering their rates would help the city and people to be less stressed and to reduce the number of injured.