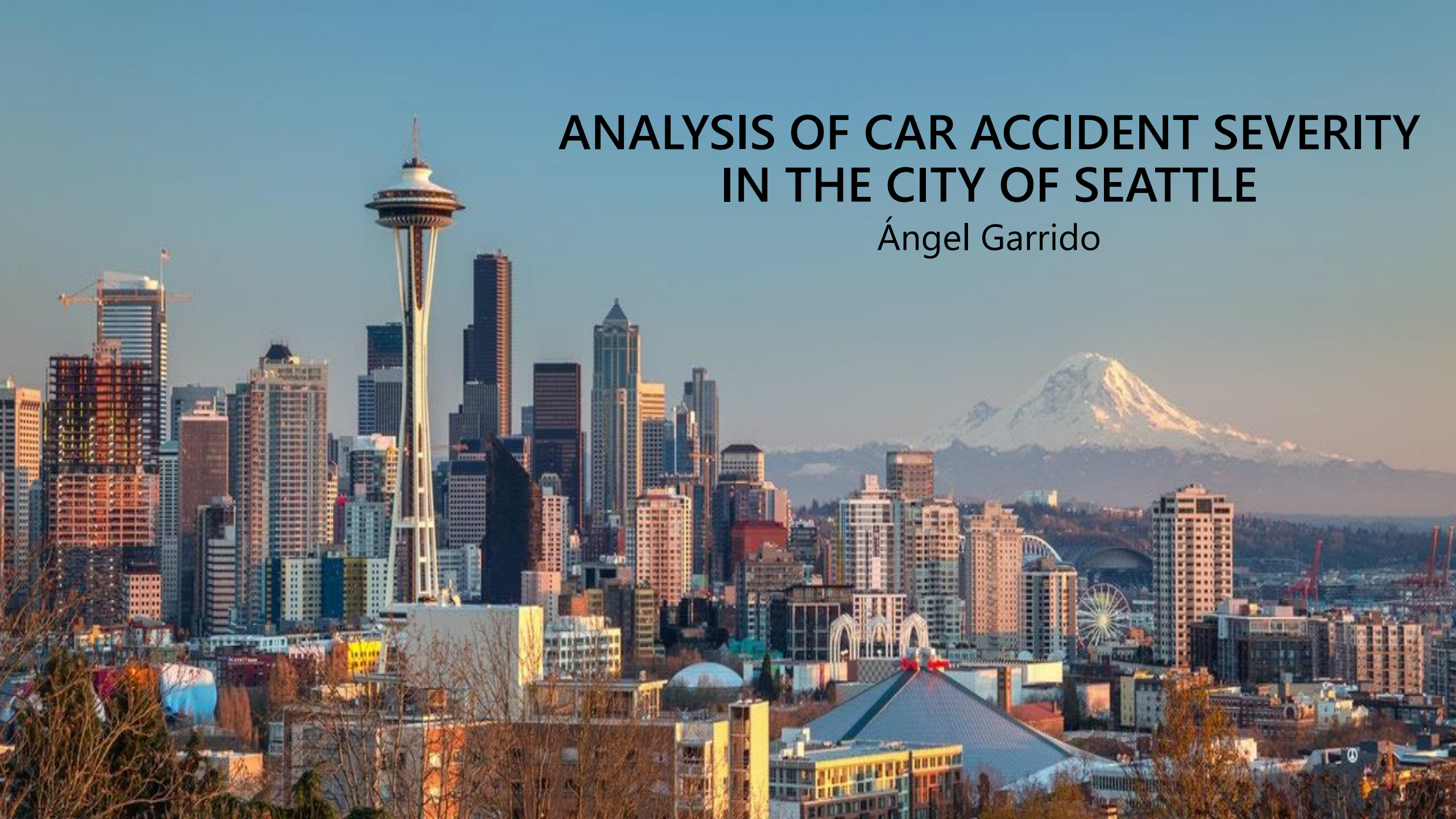


# ANALYSIS OF CAR ACCIDENT SEVERITY IN THE CITY OF SEATTLE

Ángel Garrido

# Predicting traffic accidents is valuable for the city of Seattle

Prediction of traffic accidents is key to reduce both accidents and congestions within urban areas.

- Reducing collisions is critical from the **public health** point of view; personal injuries are prevented and travel times are improved.
- Traffic congestions cost a great **environmental damage**; they increase the levels of greenhouse gases in cities.
- **Economically,** congestions are a loss of money as time and fuel are wasted.

In this project the Seattle traffic accidents data 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 obtained from the IBM server.

In total, the raw dataset has 194,673 rows and 37 features.

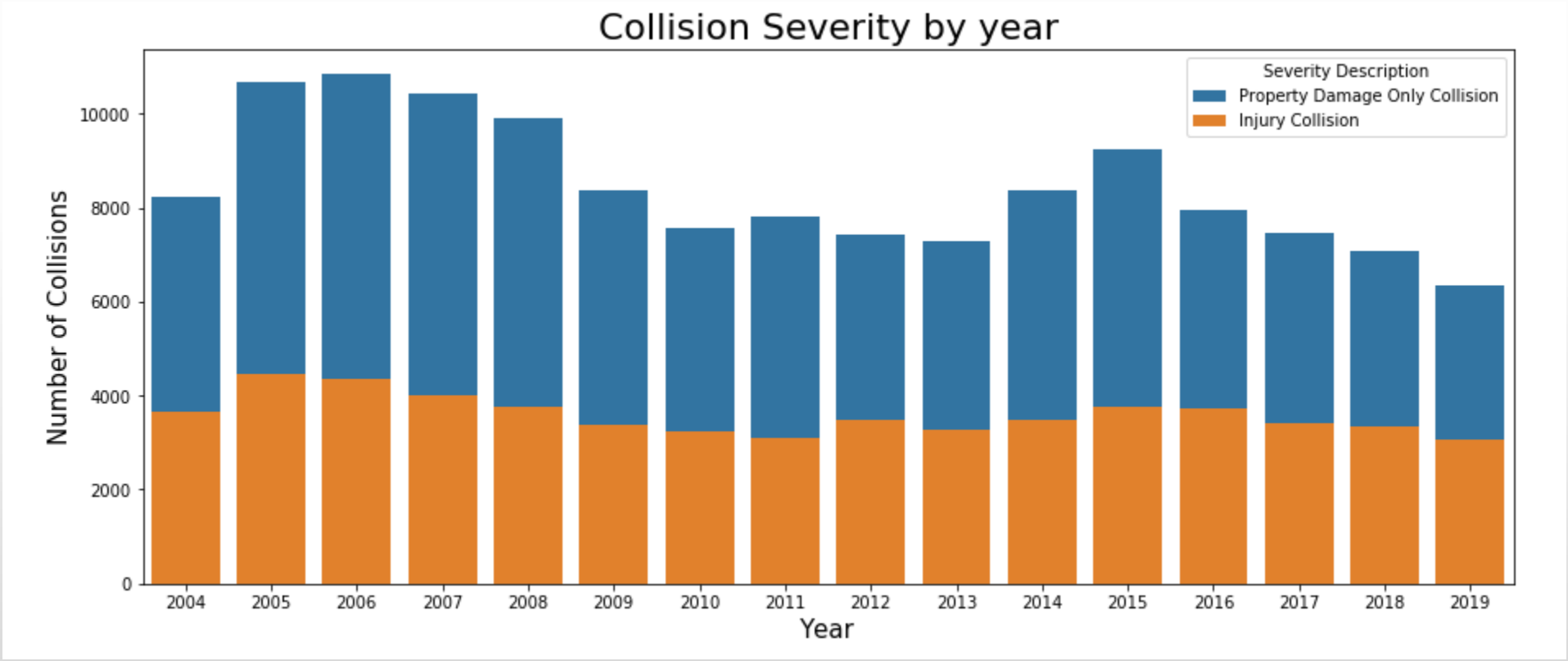
The target variable, “severitycode” has two possible outputs:

1 - **Property Damage Only Collision**

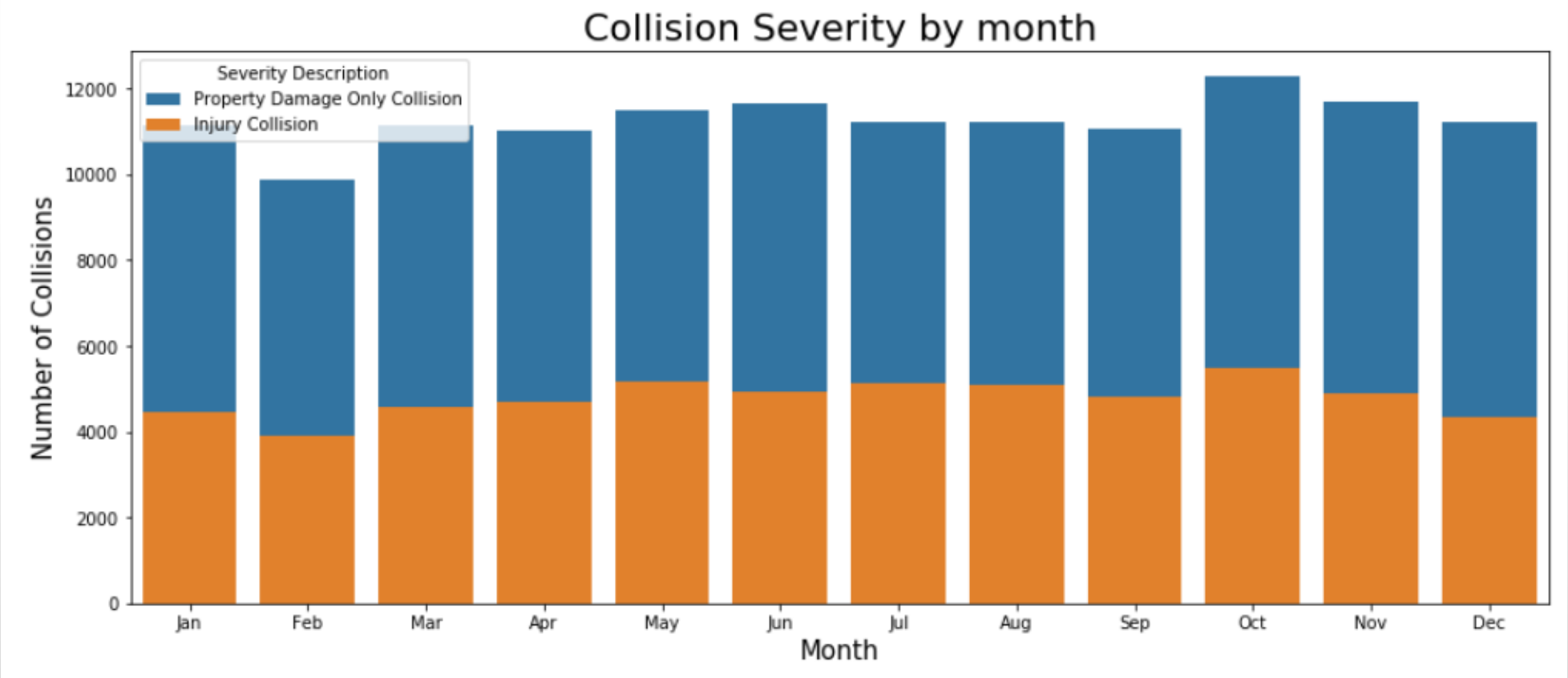
2 – **Injury Collision**

Cleaned data has **166,803 rows** and **10 features**.

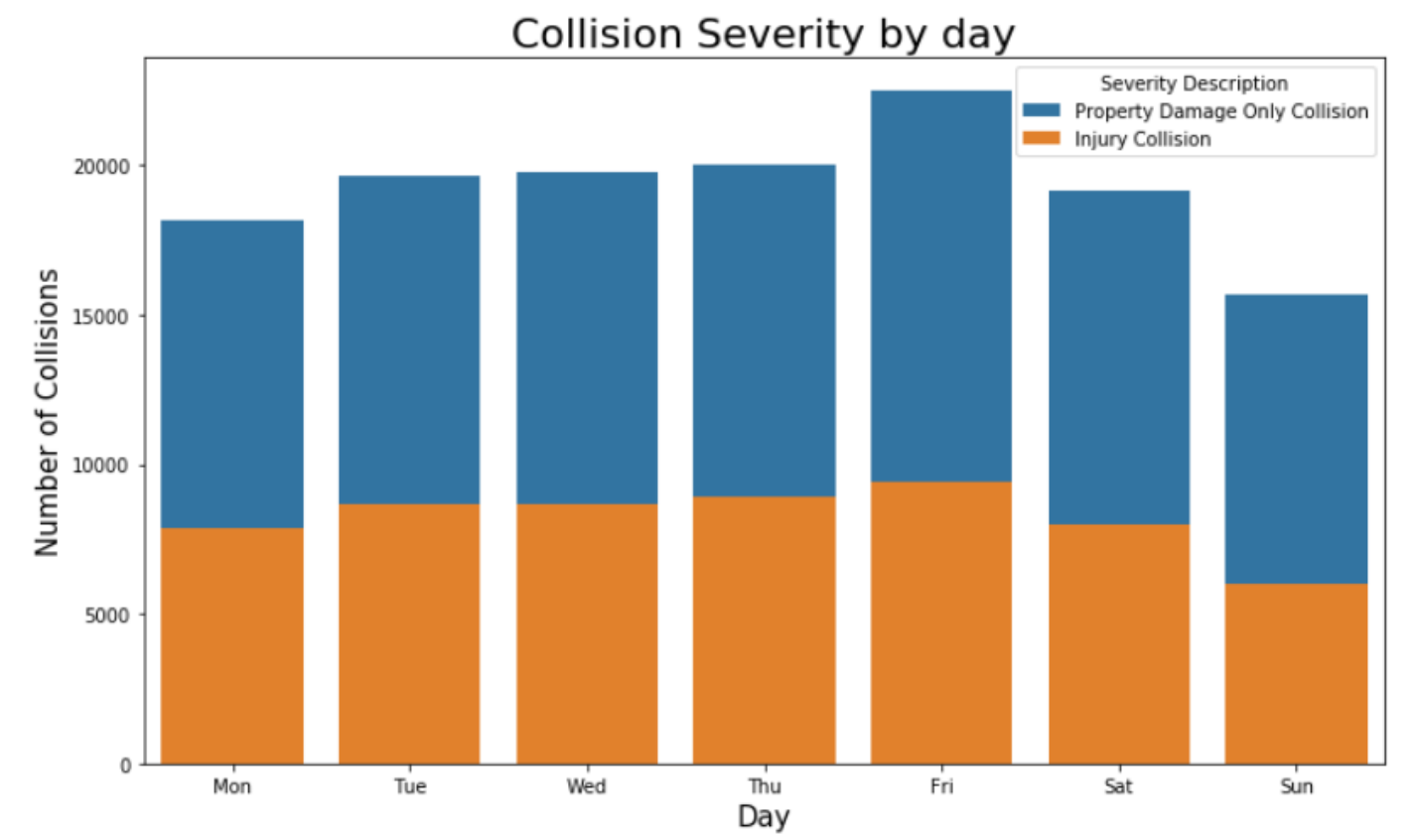
# Collisions decrease over years



# Accidents distribute monthly evenly

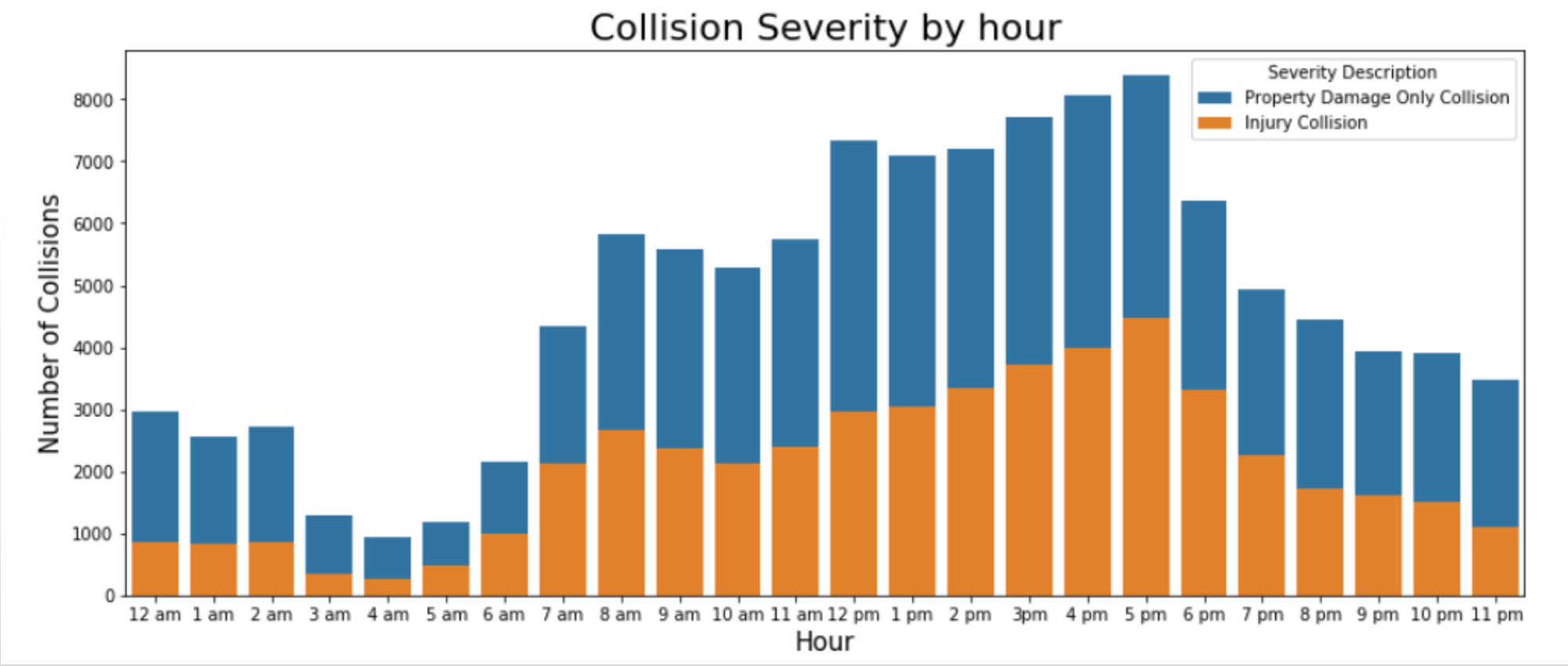


# Sundays have the lowest rates of accidents



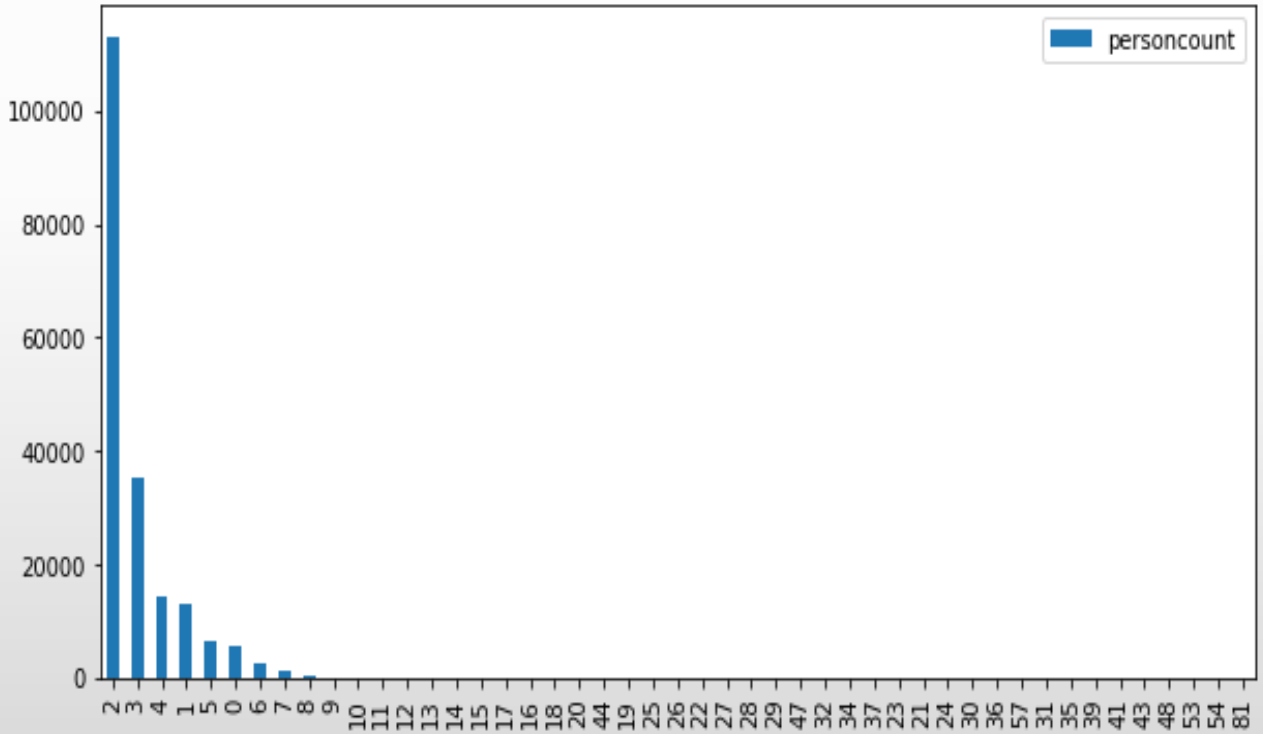
# There are more collisions during daytime

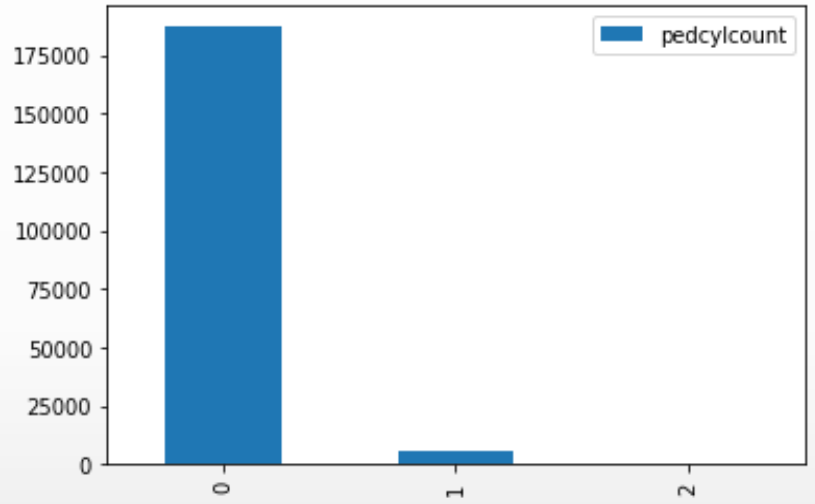
## And they are more severe

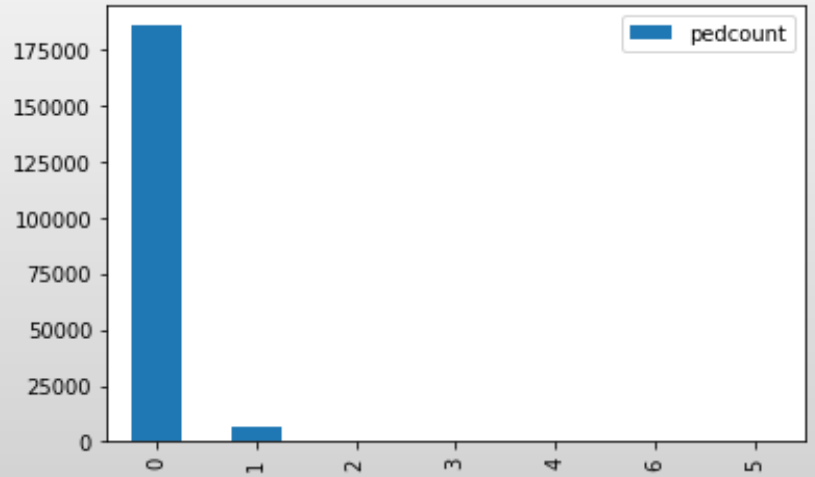


# Most traffic accidents involve 2-3 people.

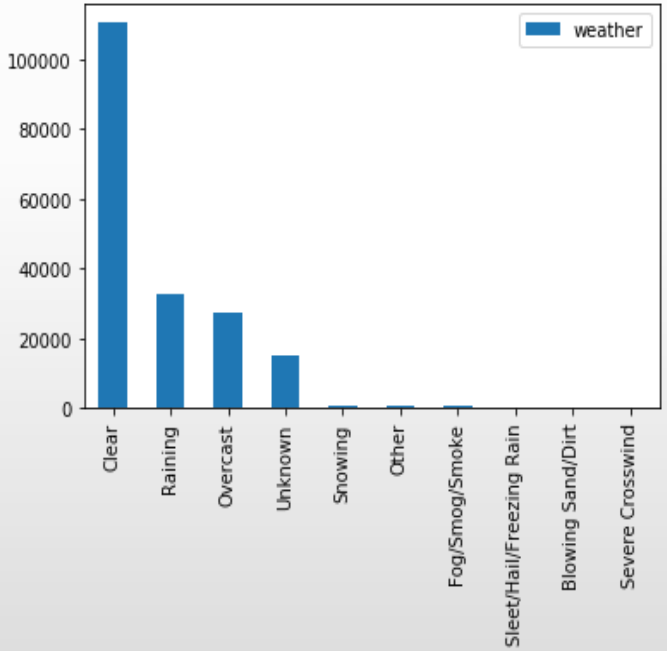
## Usually, there is no bicycle or pedestrian involved

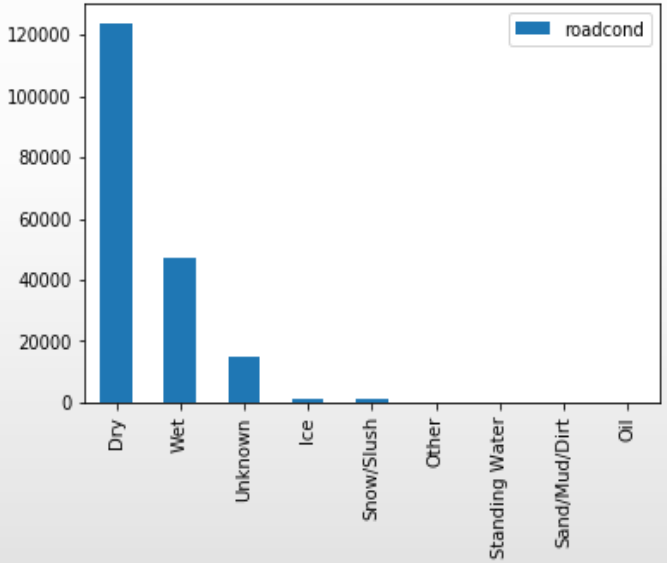


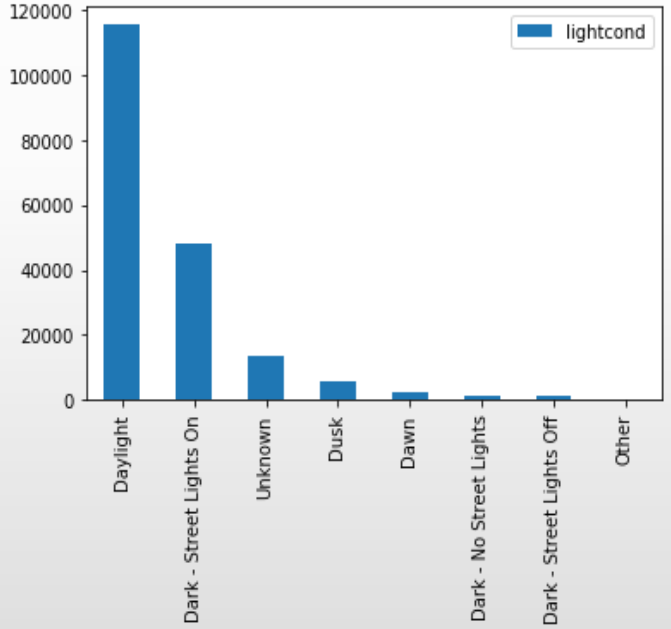



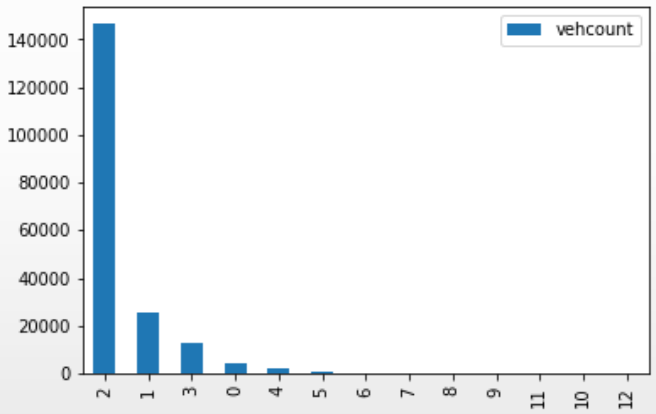
# Fair external conditions for driving

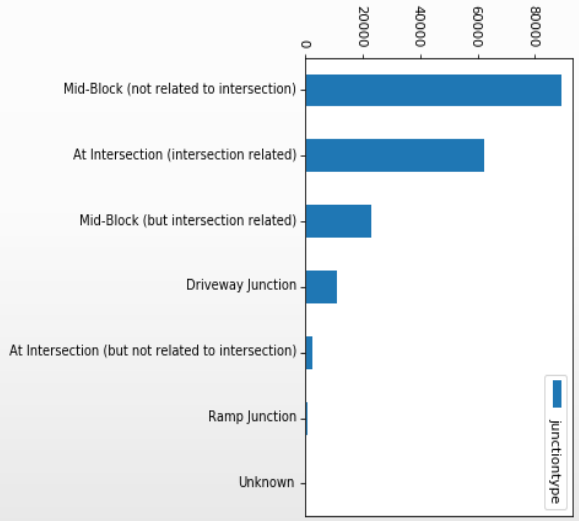


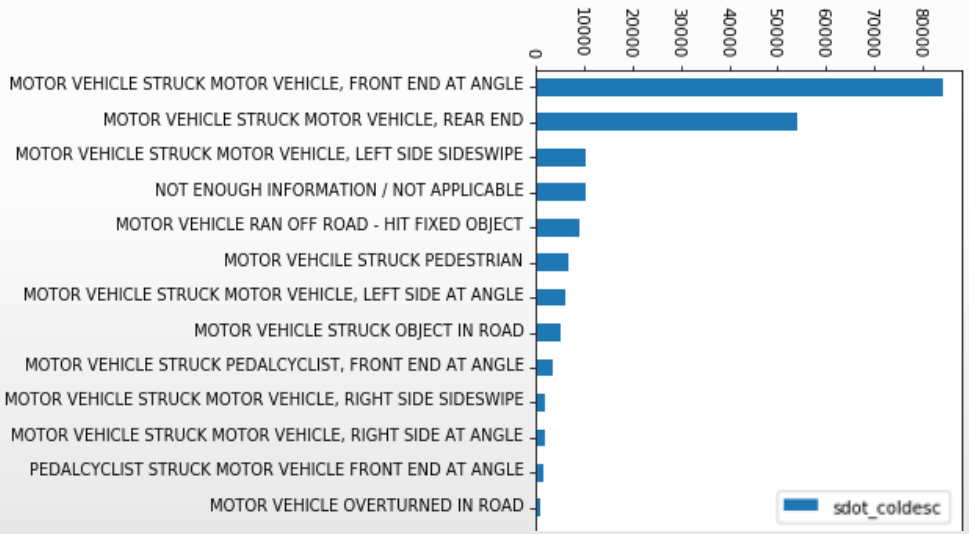



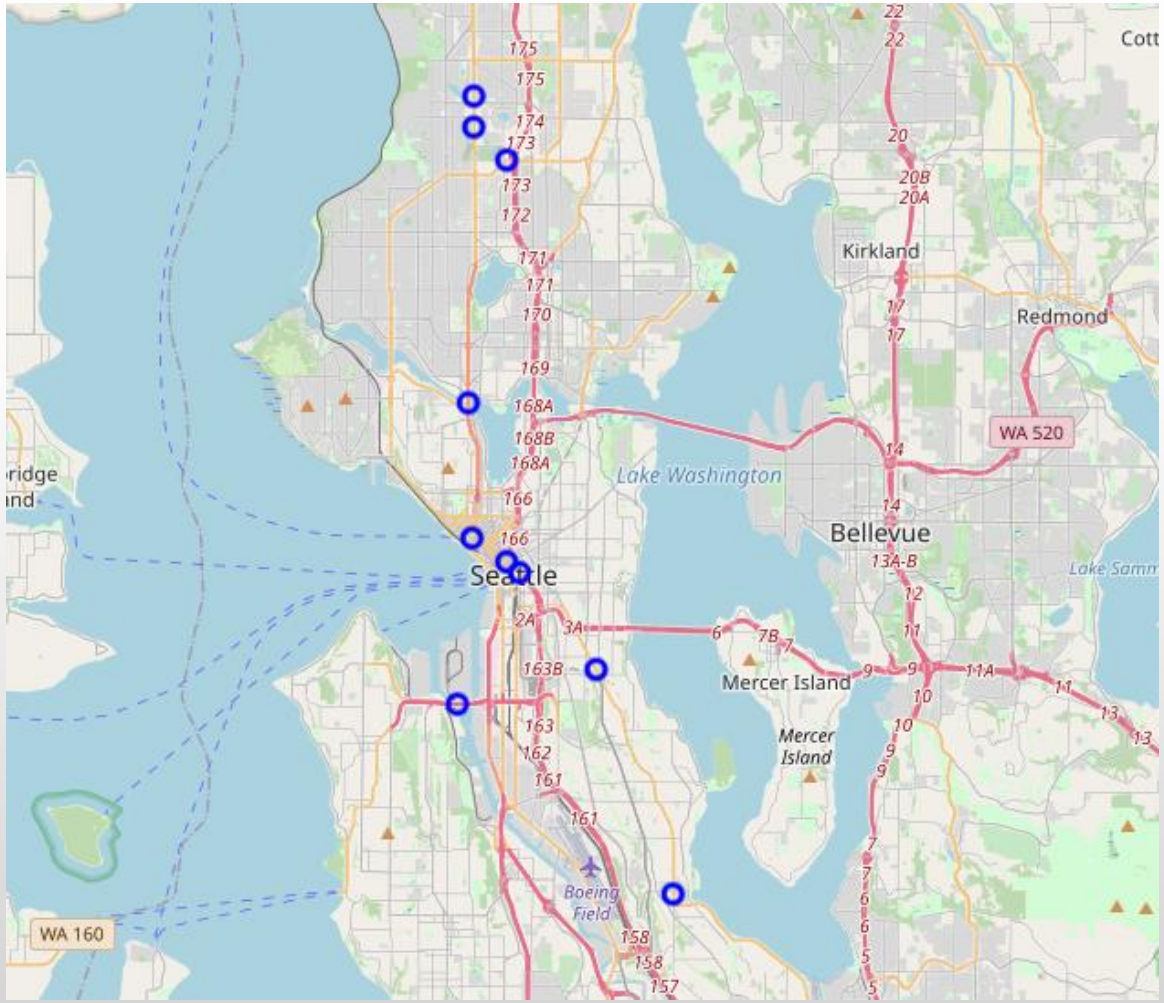
# Most accidents are caused by two vehicles, impacting at the rear or front end



# Top 10 locations
## with most traffic accidents




Top 10 locations concentrate in:

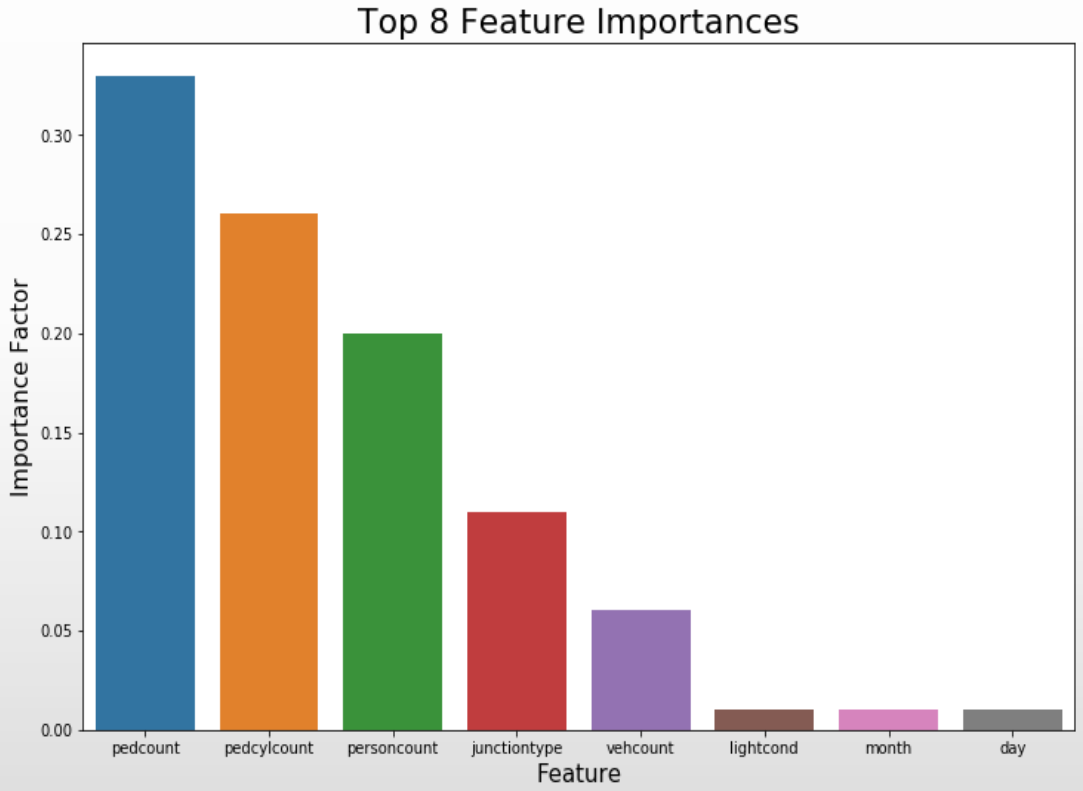
- City center (3 locations)
- North of the city (3 locations)
- 2 bridges
- 2 locations on the same avenue

# 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 |

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 makes sense, as pedestrians and cyclists receive the impact of a collision with low protection.

On the contrary, ‘lightcond’, ‘month’ and ‘day’ have low influence in the algorithm.

# 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 most accurate model.

It is interesting the influence of the studied variables on the outcome of an accident. From this analysis it can be concluded that a special attention has to be made to protect pedestrians and cyclists as they have contributed the most to the severity of an accident.

On the other hand, specific actions should be taken to areas with a high concentration of accidents, and possible traffic jams. Studying in more detail these areas and lowering their collision rates would help to reduce the number of injured and the effects of traffic congestions.