

An abstract graphic on the left side of the slide, consisting of a dark grey vertical band. Overlaid on this band is a complex network of thin, light blue lines that resemble a circuit board or a neural network. These lines branch out and connect to small, light blue circles, creating a sense of digital connectivity and data flow.

# PREDICTING CAR ACCIDENT SEVERITY

# INTRODUCTION

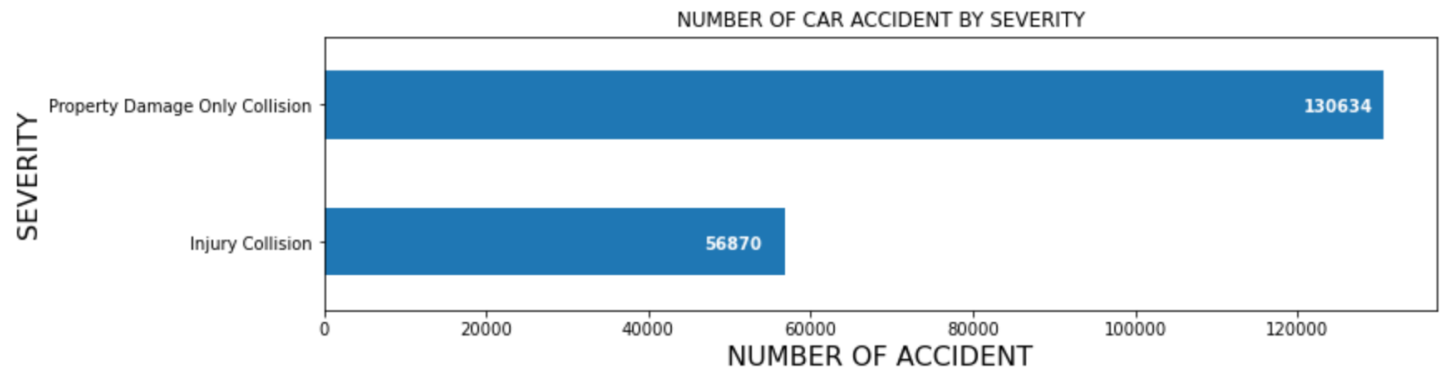
- Predicting accident severity is valuable for Seattle department of transportation.
- Knowing the severity of accident helps to control factors that affect accident.

# DATA ACQUISITION AND CLEANING

- Dataset comes from Transportation SeattleCityGIS.
- Dataset has data of accidents in Seattle from 2004 until 2020.
- In total, 194673 rows and 38 features.
- After cleaned data, I got 187504 rows and 19 features.
- Then, I reduced features to 7 features.

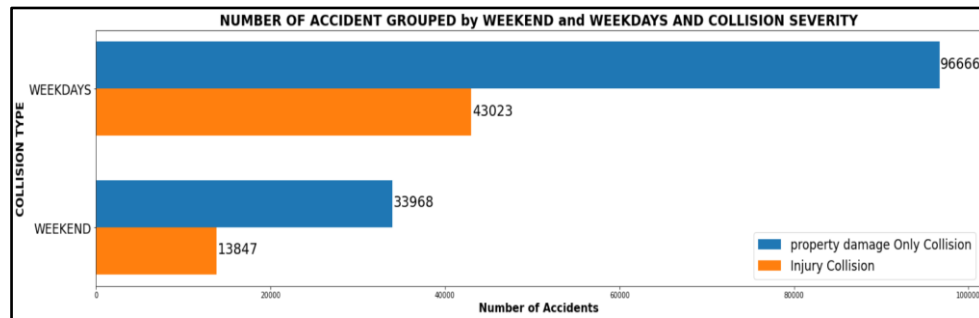
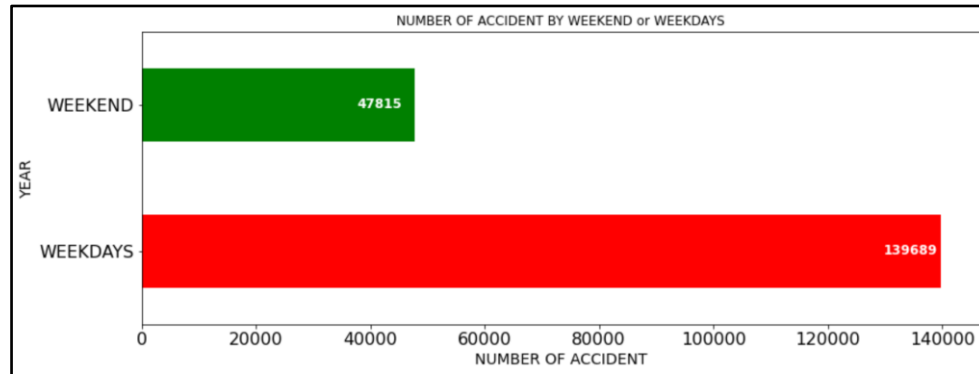
# ANALYSIS

- Number of Property only collision is 130634.
- Number of Injury Collision is 56870.



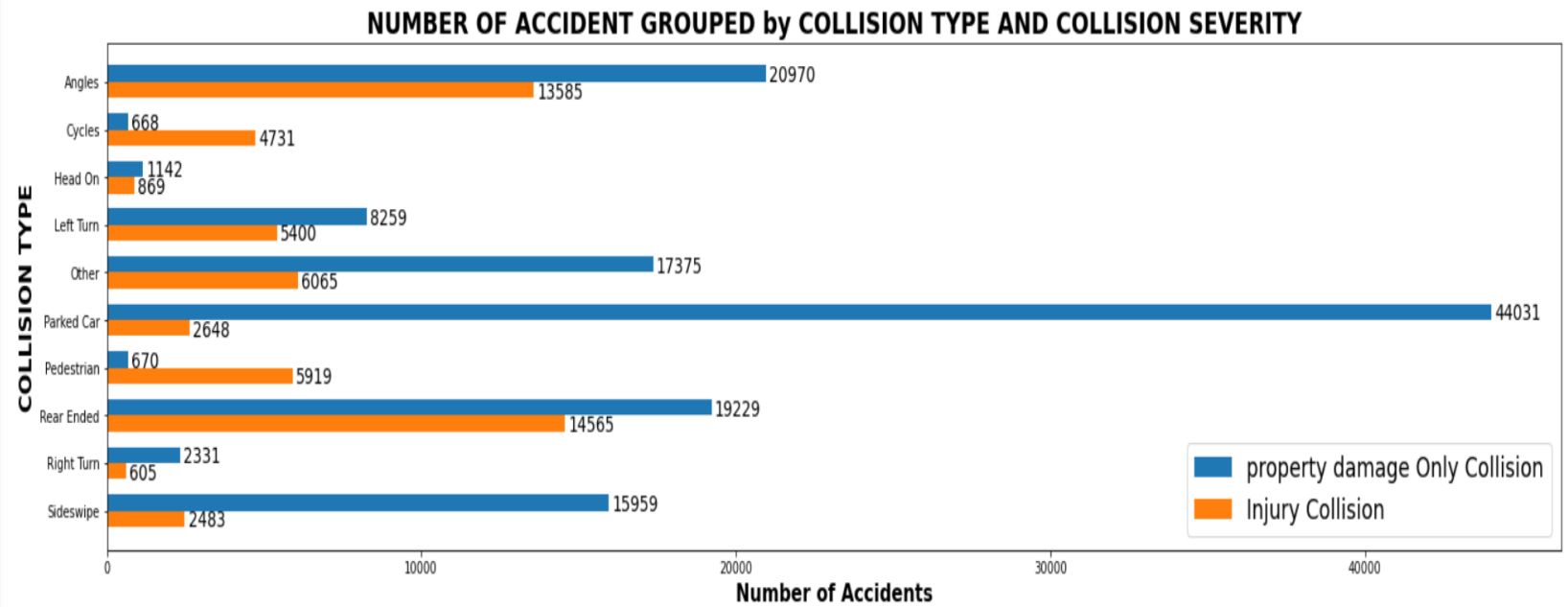
# ANALYSIS

- Number of Weekend and Weekday.
- Most of accidents at weekdays.



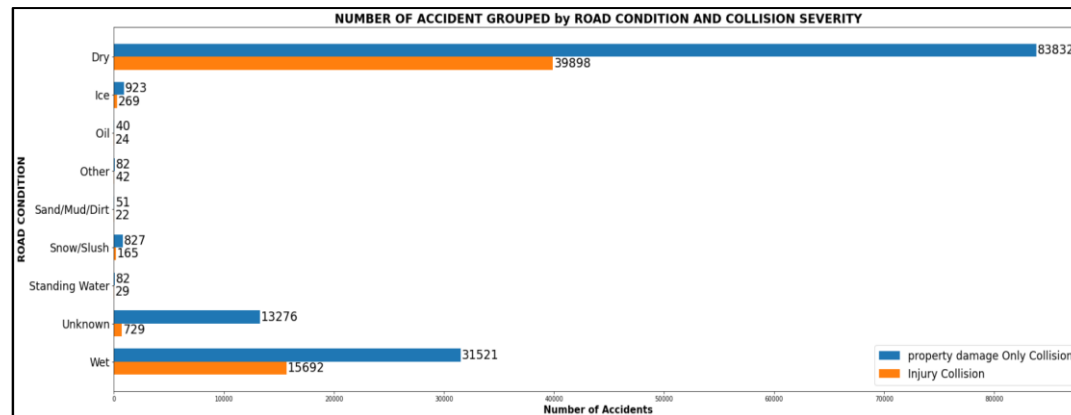
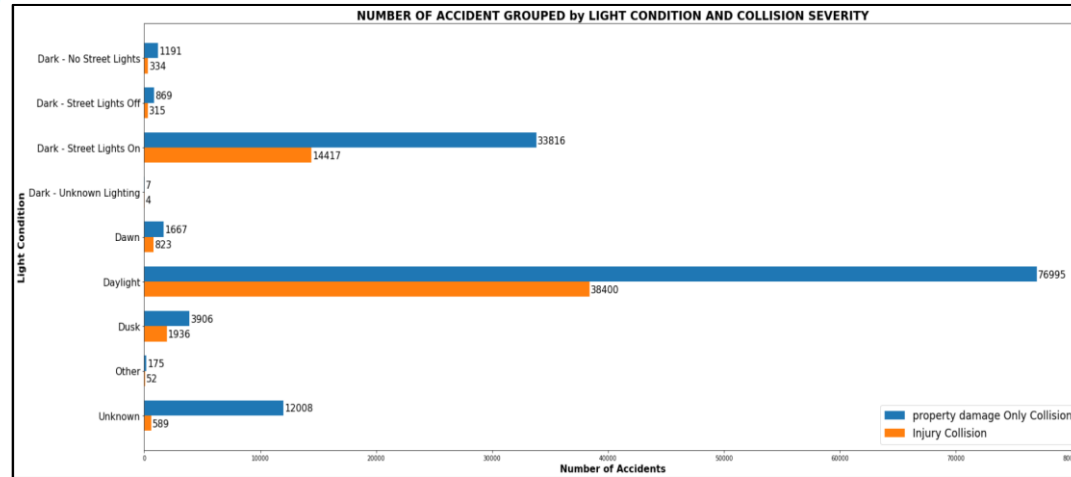
# ANALYSIS

- The most dangerous type of collision is Rear Ended.



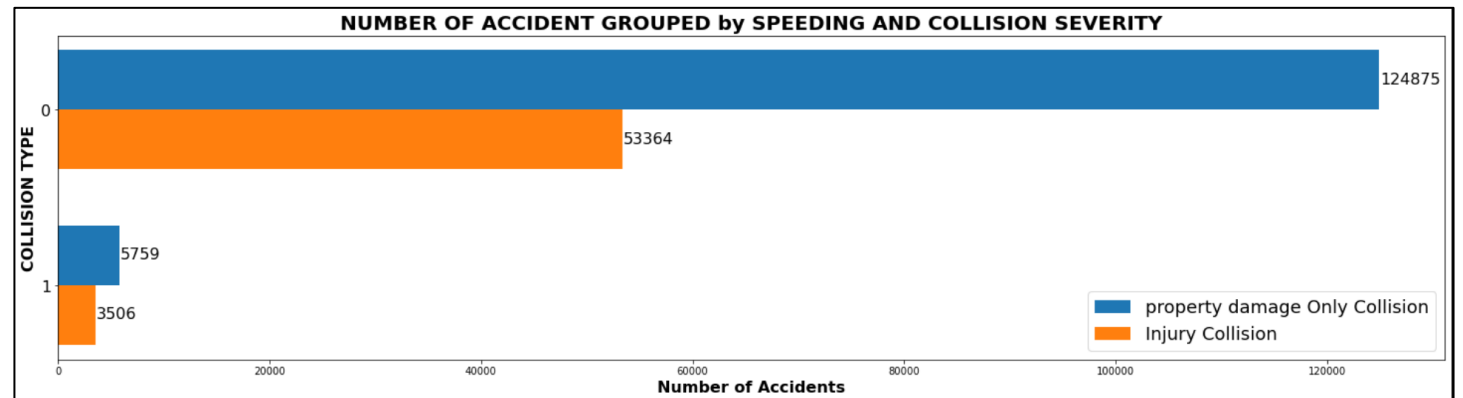
# ANALYSIS

- Most of accidents in daylight and dry road.

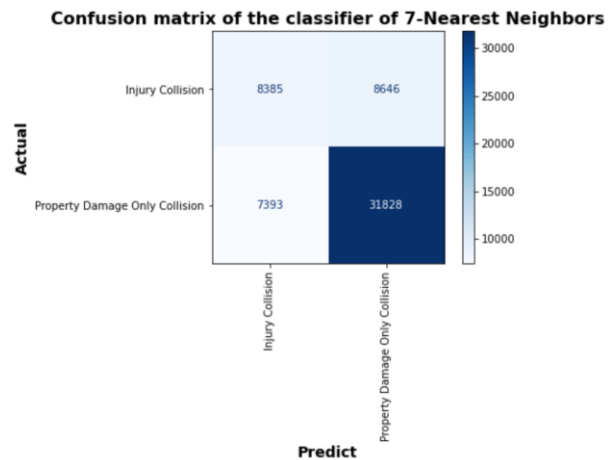
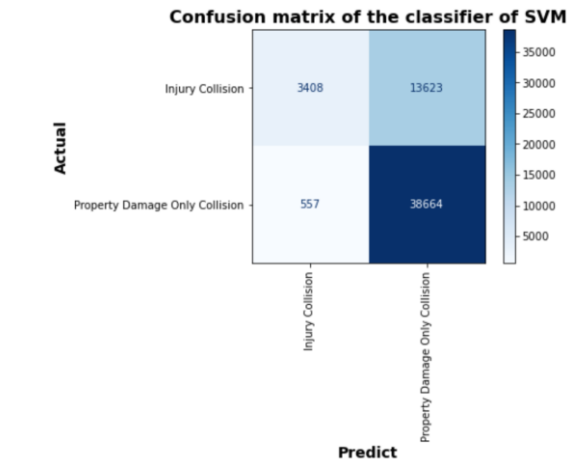


# ANALYSIS

- Speeding is not a high factor that influencing accidents.







# CLASSIFICATION MODELS

- The **accuracy** of 6 models are between 0.6207 - 0.7479.
- The F1-Score of 6 models are between 0.6317 - 0.7116.
- The highest accuracy is for SVM.
- The highest F1-Score is for 7-Nearest Neighbours.



## CONCLUSION AND FUTURE

- Useful models are built to predict severity of accidents.
- The accuracy and F1-score for Models could be improved by add more data to built balanced dataset.