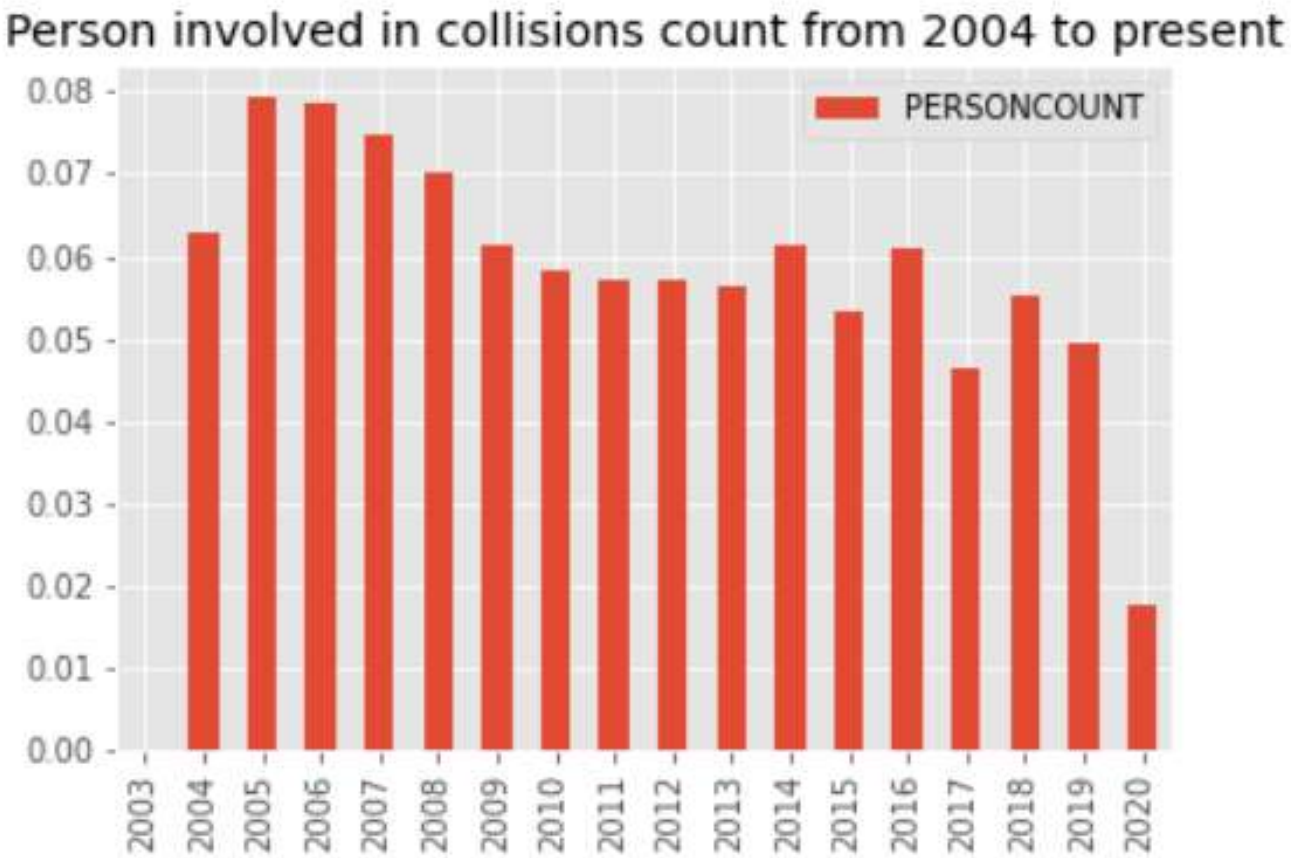
# Seattle Collisions severity recogition using Machine Learning

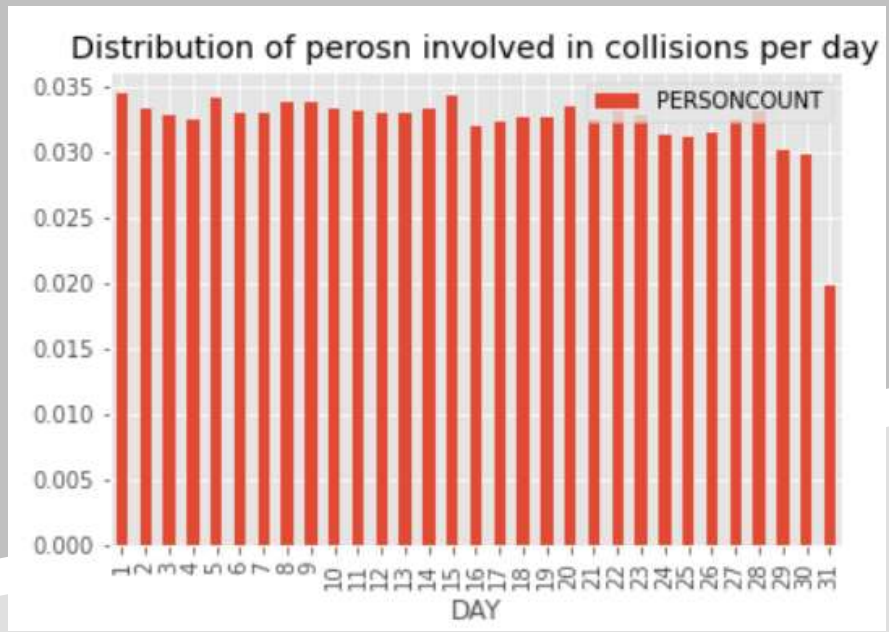
# Data exploratory analysis

- Incidents have decreased over the years

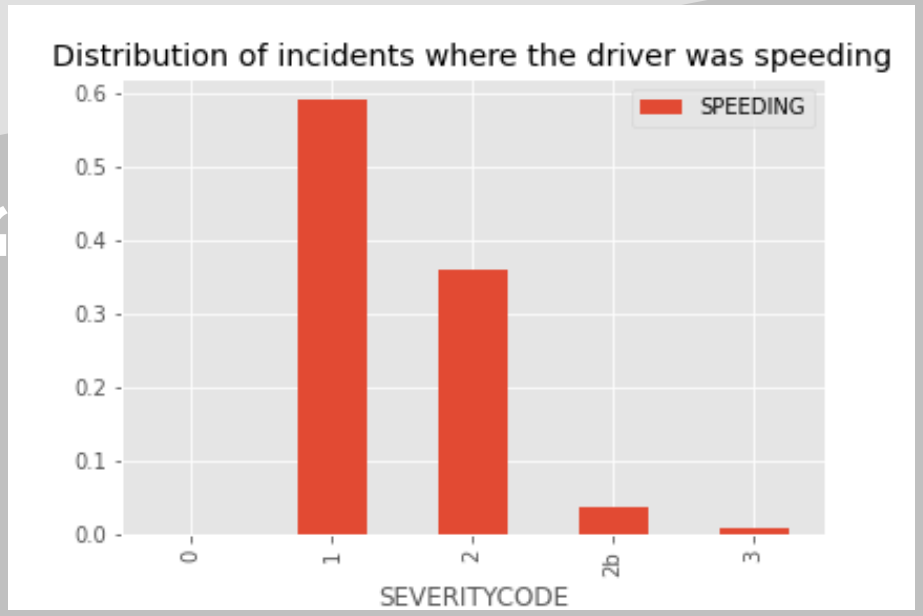


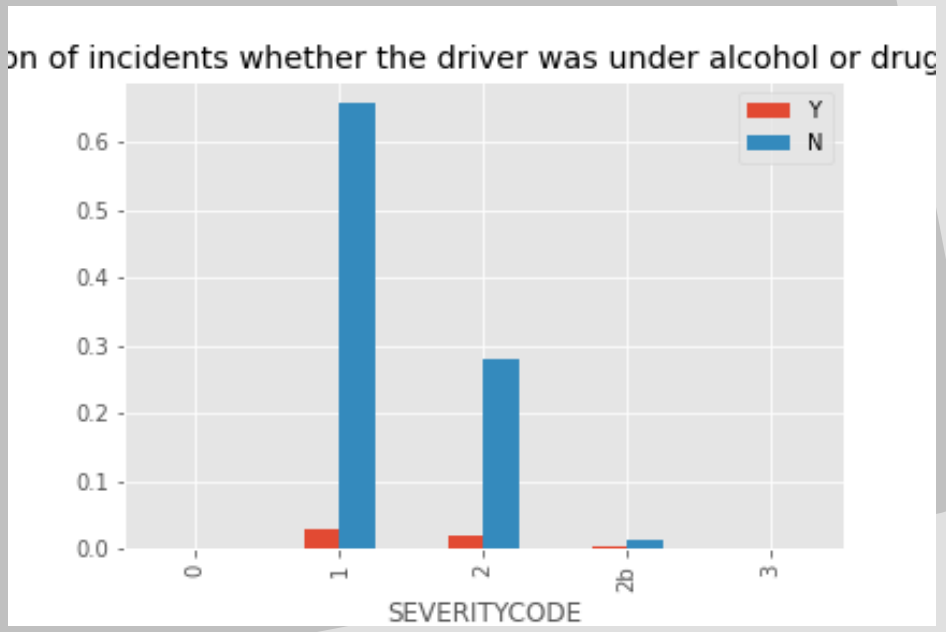
# Data exploratory analysis

- There is no prefered month of day
- February and September have the least number of incidents

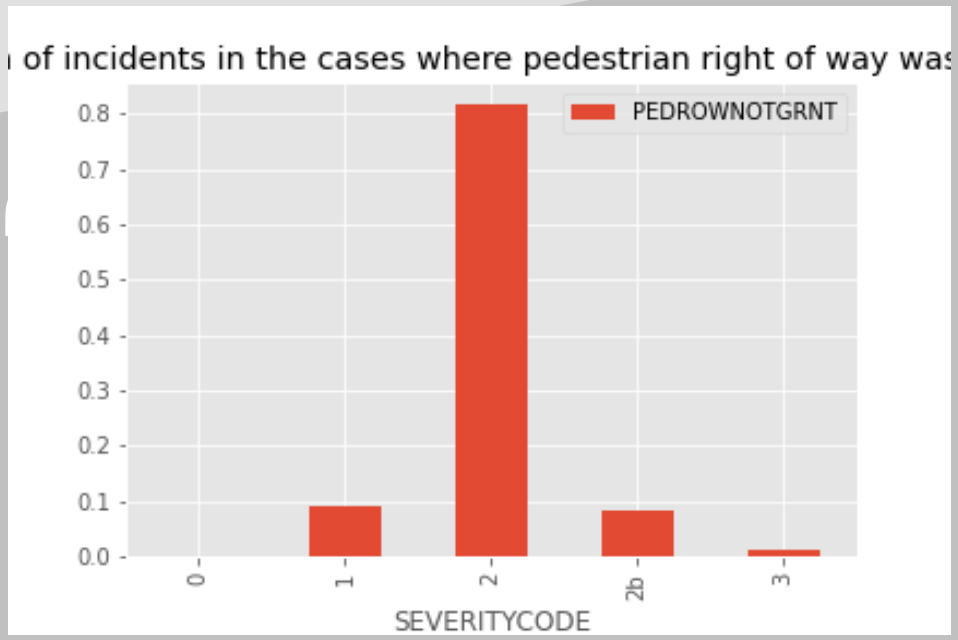



# Data exploratory analysis: Driver-Pedestrian

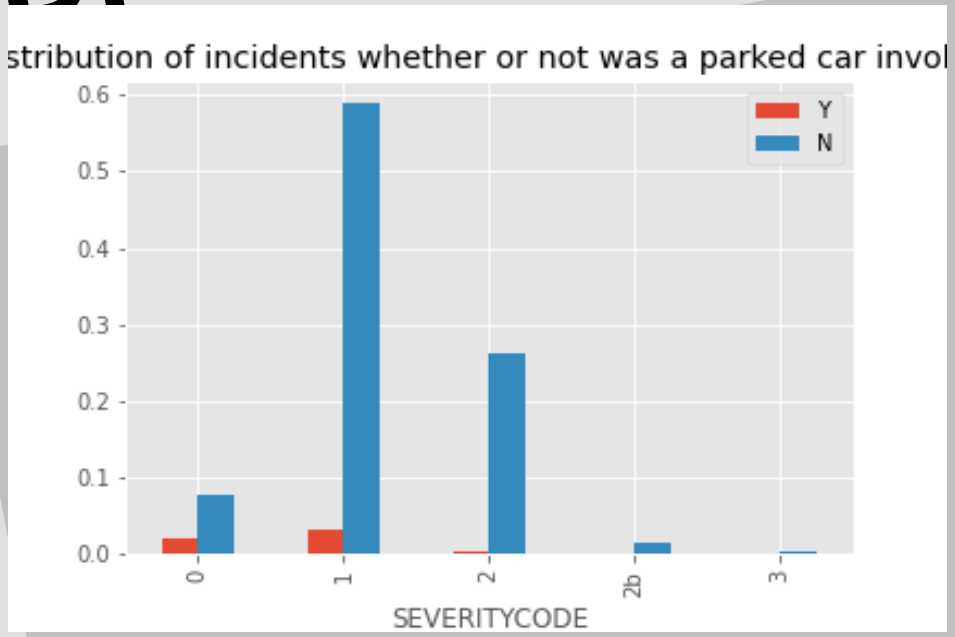


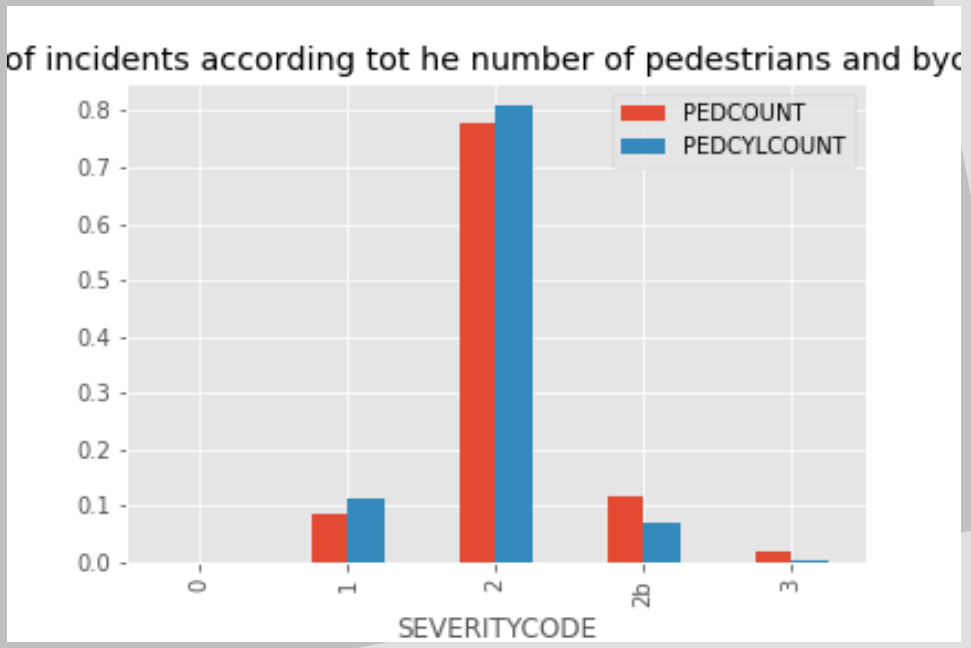



# Data exploratory analysis: Driver-Pedestrian

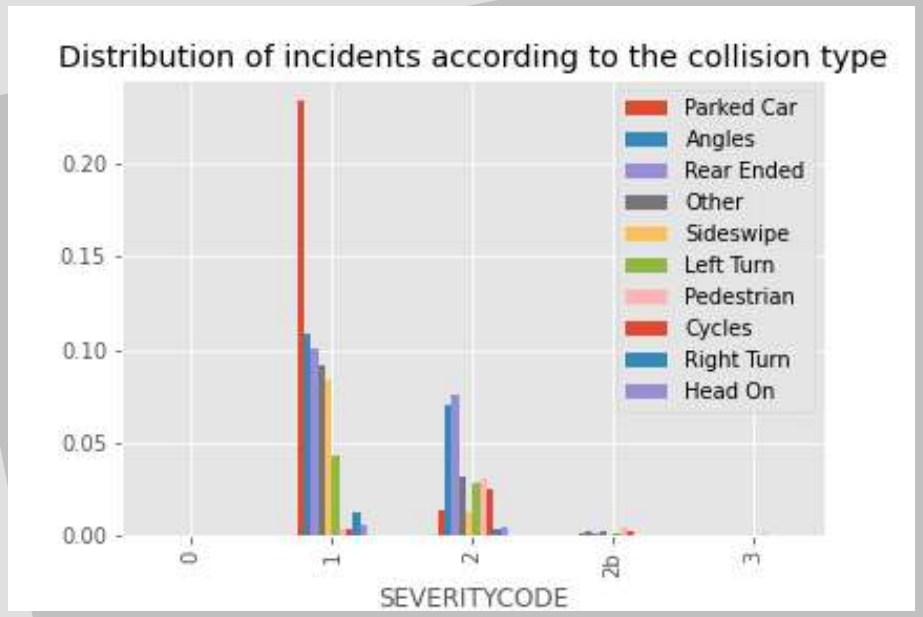


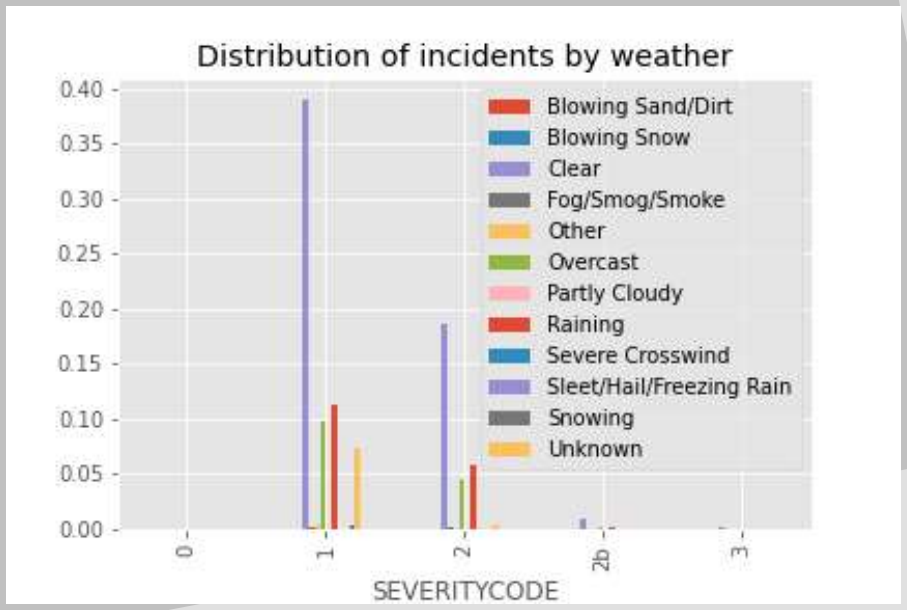

# Data exploratory analysis: Driver-Pedestrian



# Data exploratory analysis: Environment

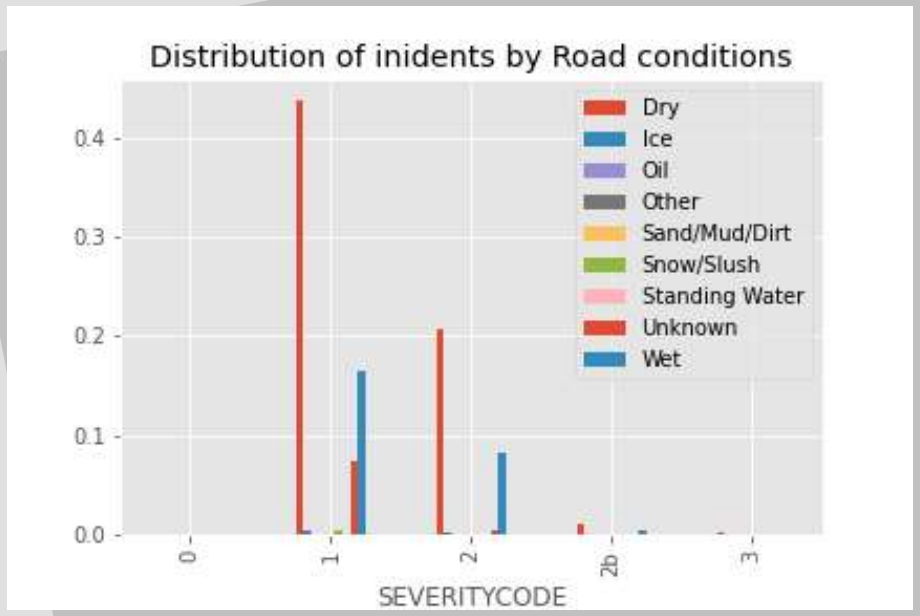


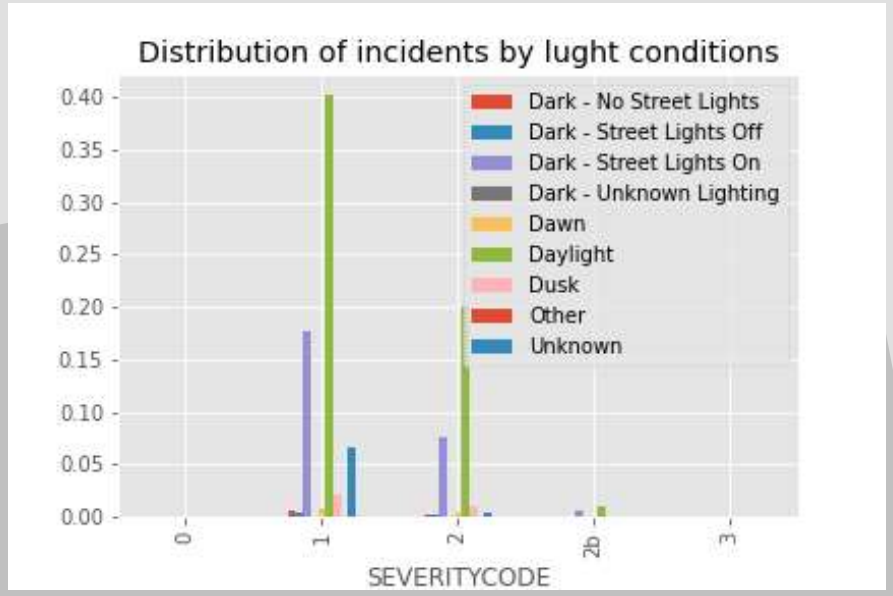



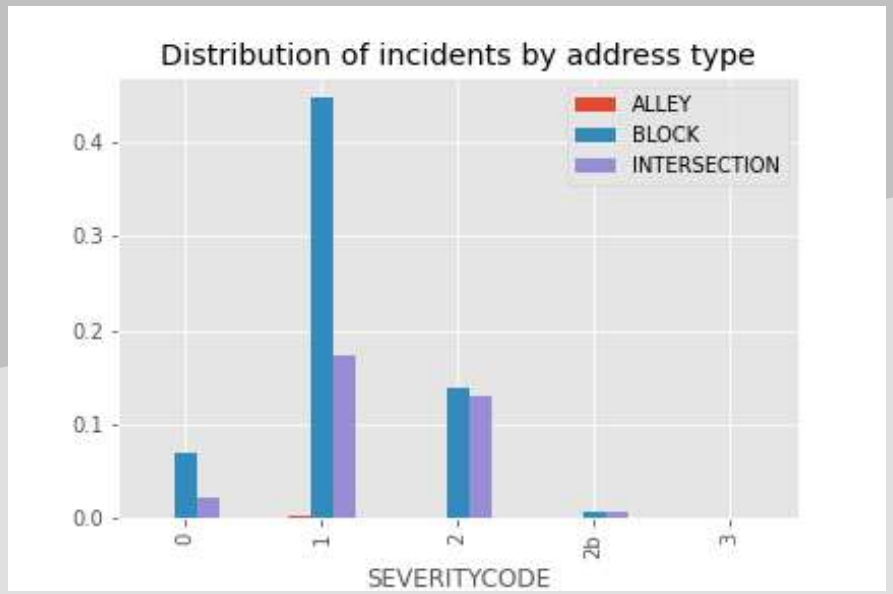
# Data exploratory analysis: Environment

- There is a small fatalitie percentage
- Most accidents implies property damage
- Just a small percetage involves serious injuries althogh the injuries percentage is quite elevated.

# Data exploratory analysis: Environment



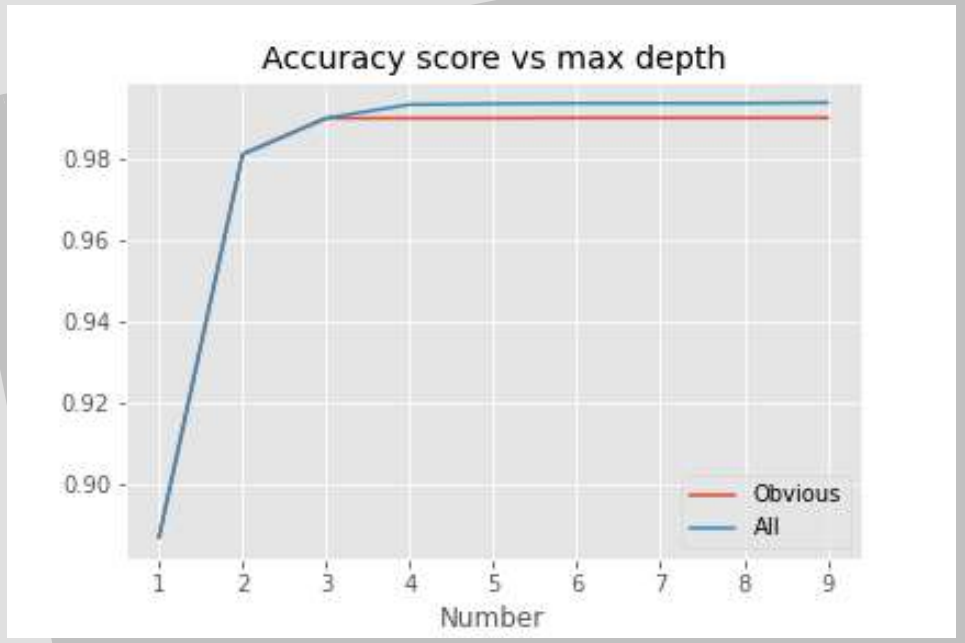
# Data exploratory analysis: Environment

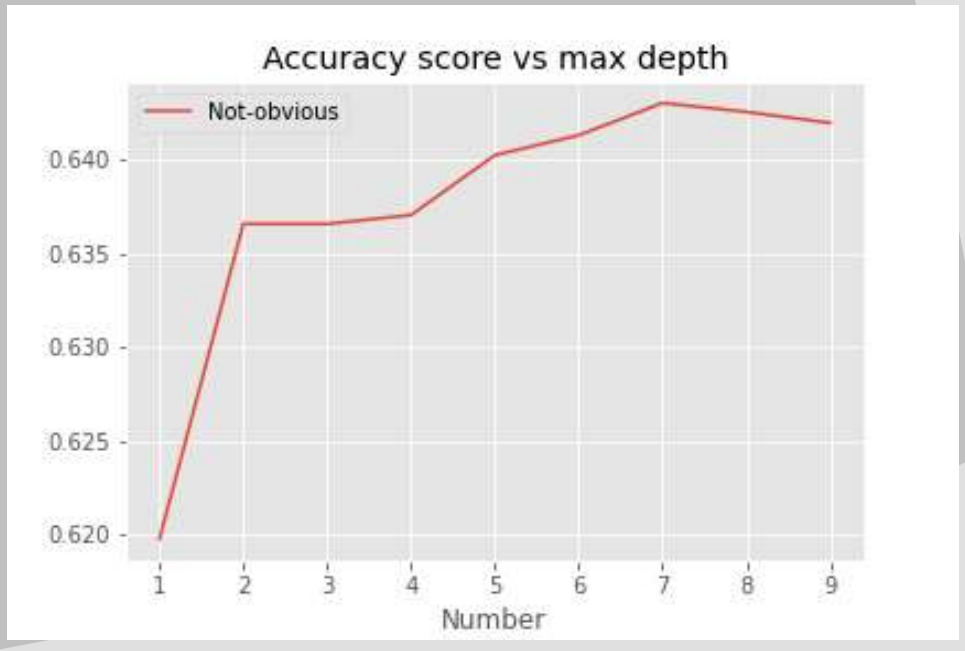
"IT IS ALL ABOUT DRIVER RESPONSIBILITY!!"

# Modeling

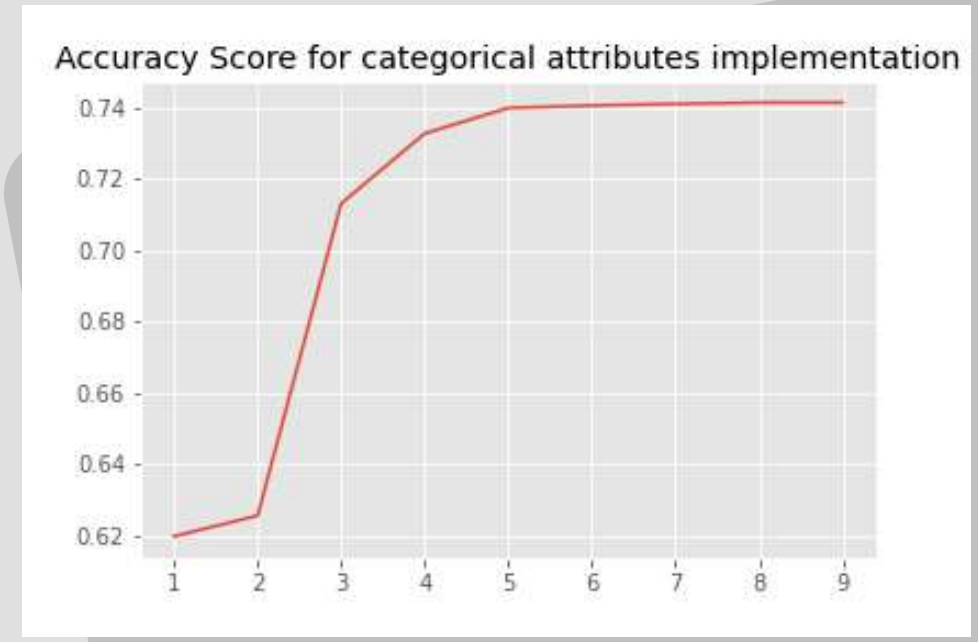
- Decision tree algorithm was accurate
- Other algorithms take high computing times

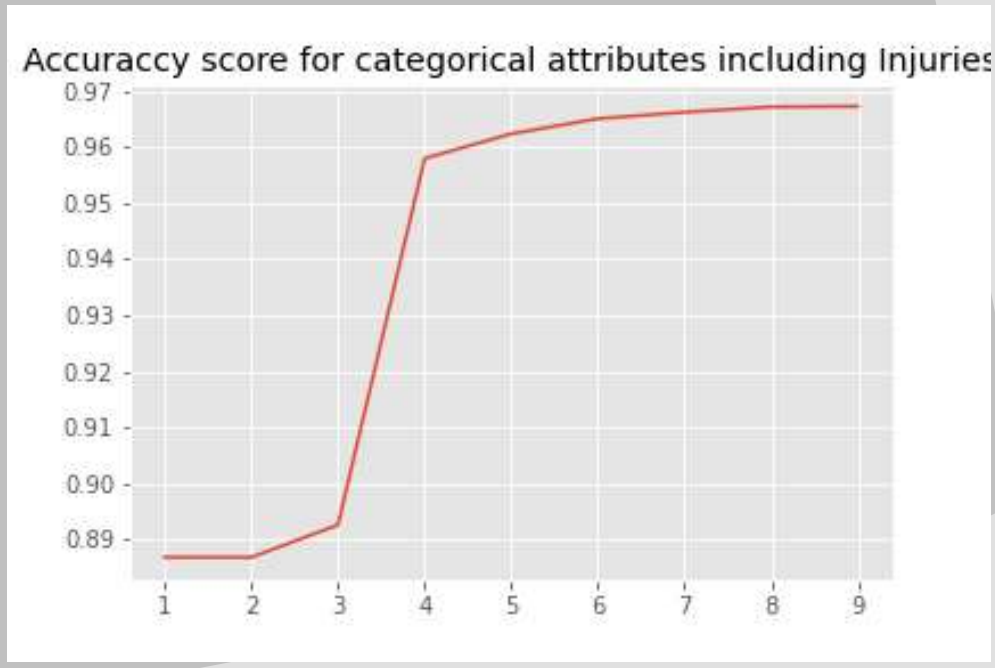
# Modeling: Numerical Attributes



# Modeling: Categorical Attributes




Including injuries highly improve accuracy!

# Conclusions

- Incidents have decreased over the years
- There is no prefered month of day
- February and September have the least number of incidents
- There is a small fatalitie percentage
- Most accidents implies property damage
- Just a small percetage involves serious injuries althogh the injuries percentage is quite elevated.
- Most collisions occurs in good environmental conditions, so they are a result of a lack of responsibility
- Decision tree algorithm was accurate
- One can either model with numerical or categorical attributes with high accuracies.
- Modeling with attributes improves its accuracy by including injuries rather than fatalities
- A tree depth between 5 and 7 is optimal