CAPSTONE PROJECT

IBM DATA SCIENCE PROFESSIONAL CERTIFICATE

Imperio Antonio

10/05/2020

PREDICT ACCIDENTS SEVERITY

- A reliable and fast model that can prevent the accident gravity could be of huge help:
    - Prevent more accidents in the future
    - Warn the interested entities in time to have fast and accurate response
    - Learn where to invest more energies and resources to optimize the service
    - Save resources and money, increasing the quality of the service
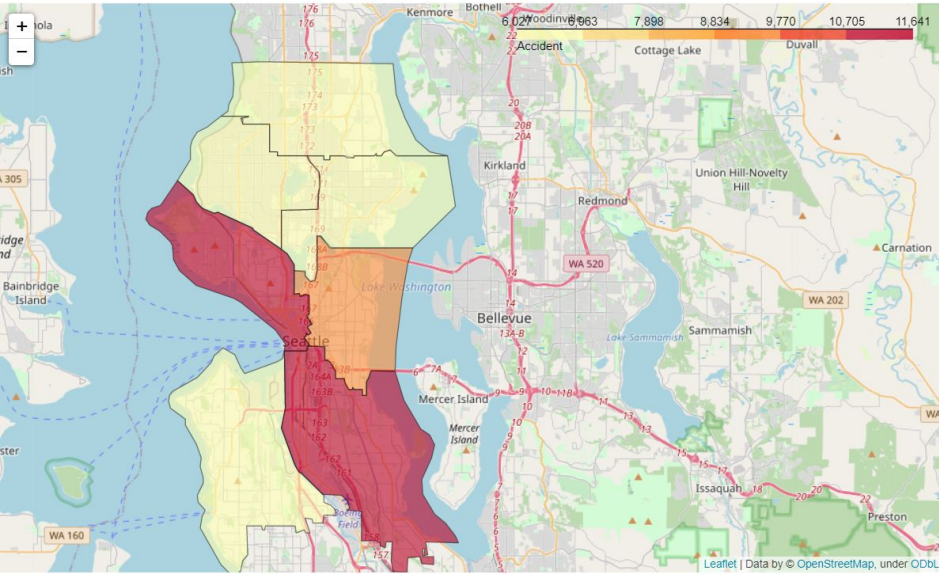
DATA ACQUISITION AND CLEANING

Data collected from online repository describing conditions at the moment of the incident.

Useless Data are removed and missing values taken care of.

The resting data are analyzed and elaborated to obtain until a satisfying dataset is obtained that contains useful features to the model development.
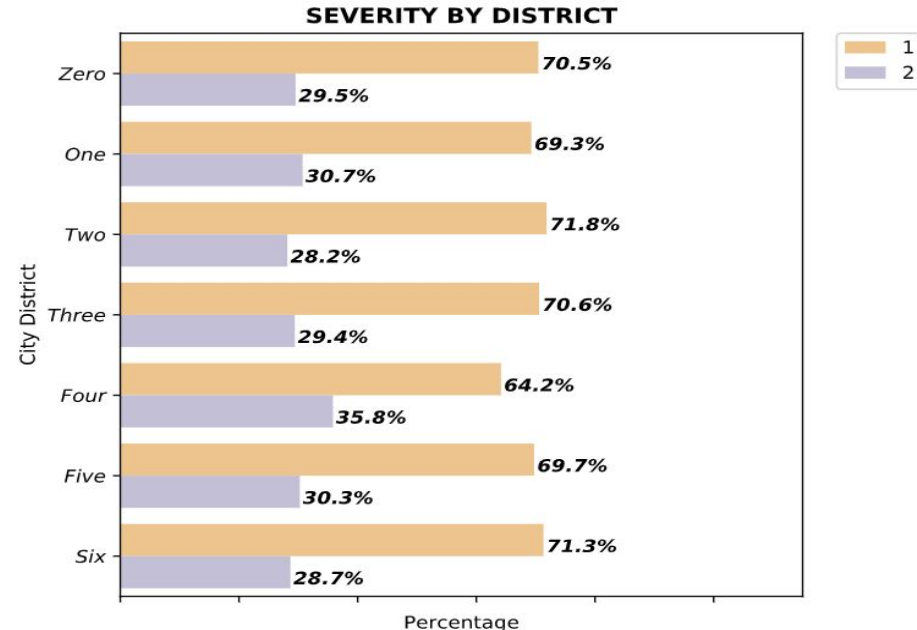
- Weather/Road/Light Conditions
- Geographical position of the incident (District/Block/Intersection/Alley)
- Number of people/pedestrians/cyclists/vehicles involved in the accident
- Incident Date

DISTRICTS SEVERE ACCIDENTS DISTRIBUTION

Graph shows as critics accident are spread throughout the city districts.

It's clear as there are zones that require more attention than others

The second graph shows as the distribution of cases is constant for all the districts.

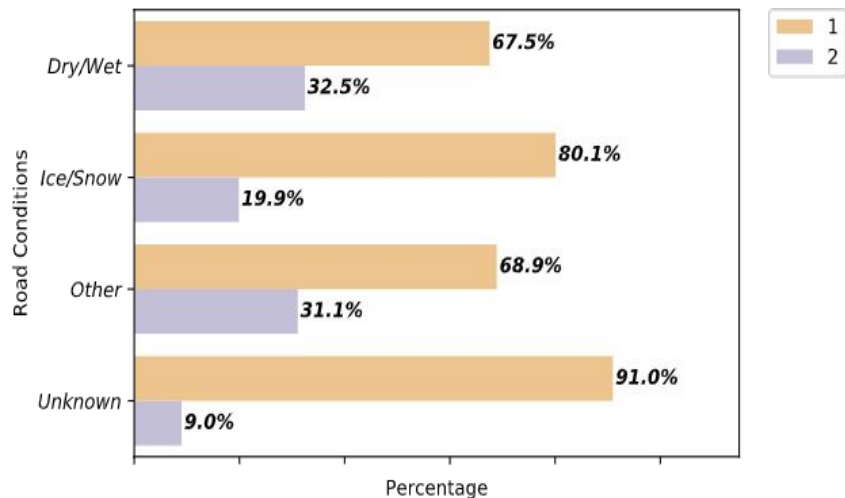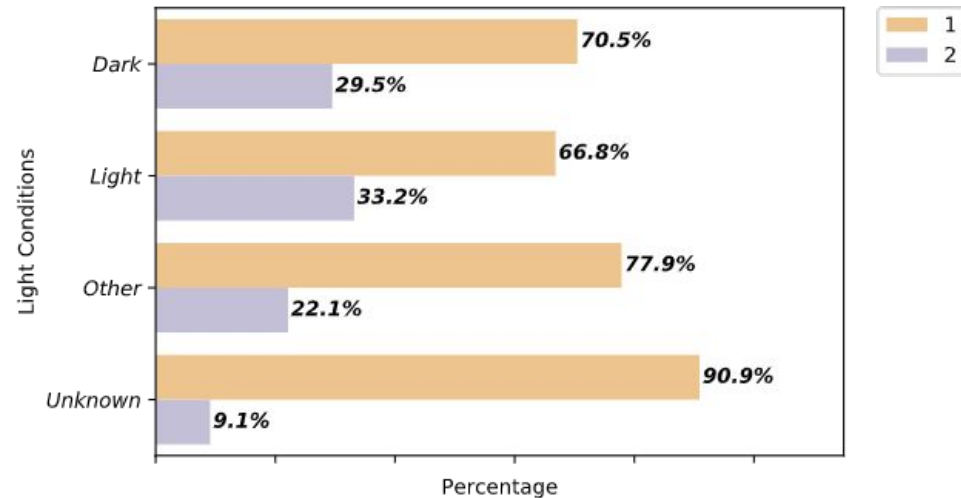This means that cases happen with same frequency just in greater quantity for certain districts

CONDITIONS

The statistics distributions show that no particular conditions influence in a strong way the critic/non critic ratio.

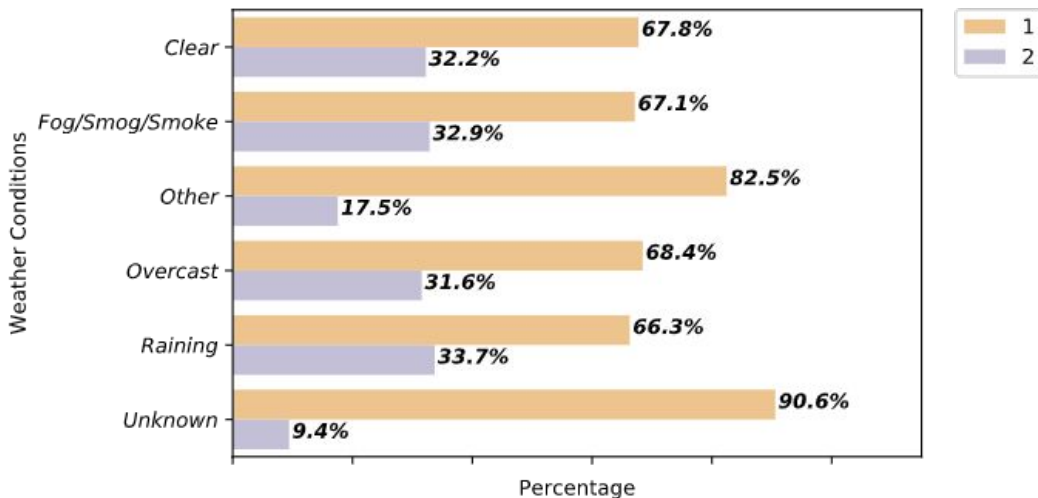### SEVERITY BY LIGHT CONDITIONS

Light Conditions	1	2
Dark	70.5%	29.5%
Light	66.8%	33.2%
Other	77.9%	22.1%
Unknown	90.9%	9.1%

Percentage

SEVERITY BY ROAD CONDITIONS

Road Conditions	1	2
Dry/Wet	67.5%	32.5%
Ice/Snow	80.1%	19.9%
Other	68.9%	31.1%
Unknown	91.0%	9.0%

Percentage

SEVERITY BY WEATHER CONDITIONS

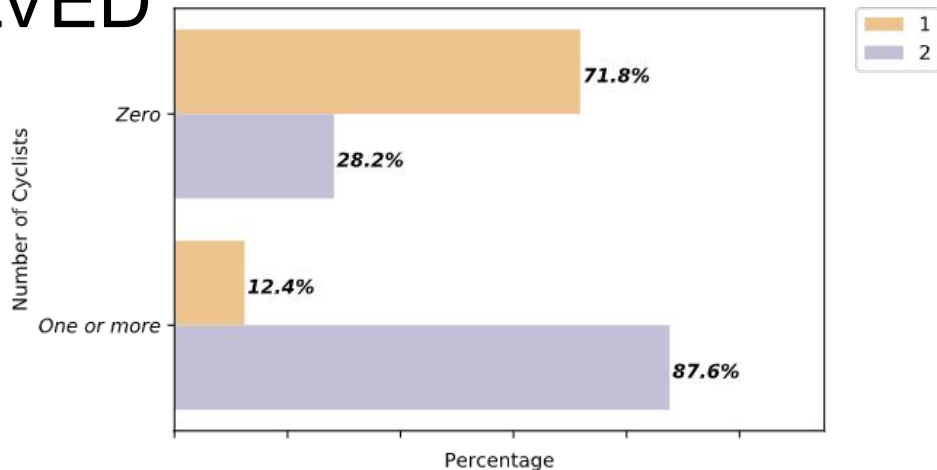| Weather Conditions	1	2
Clear	67.8%	32.2%
Fog/Smog/Smoke	67.1%	32.9%
Other	82.5%	17.5%
Overcast	68.4%	31.6%
Raining	66.3%	33.7%
Unknown	90.6%	9.4%

Percentage

PEOPLE/VEHICLES INVOLVED

The analysis about the vehicles and people involved showed that whenever a pedestrian/cyclist was involved the criticity was high
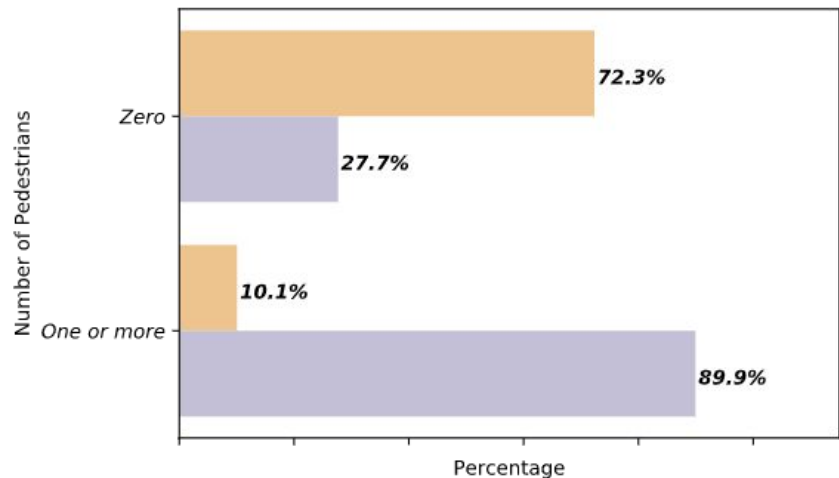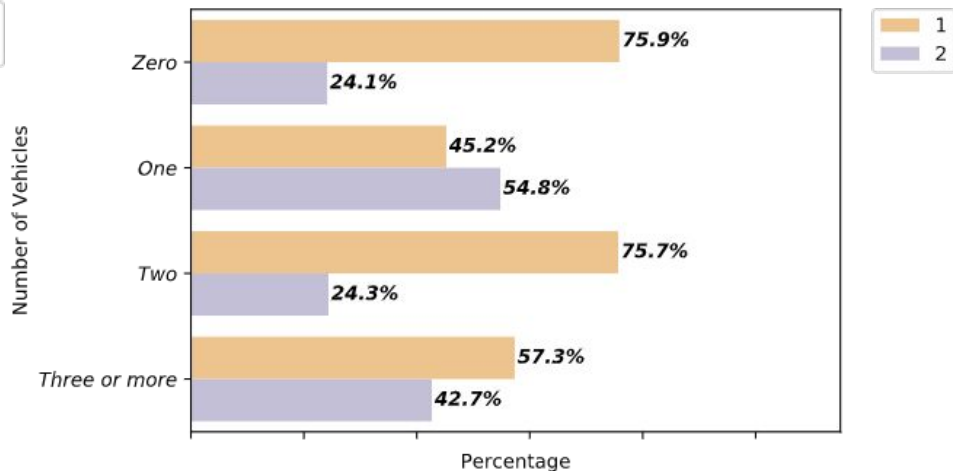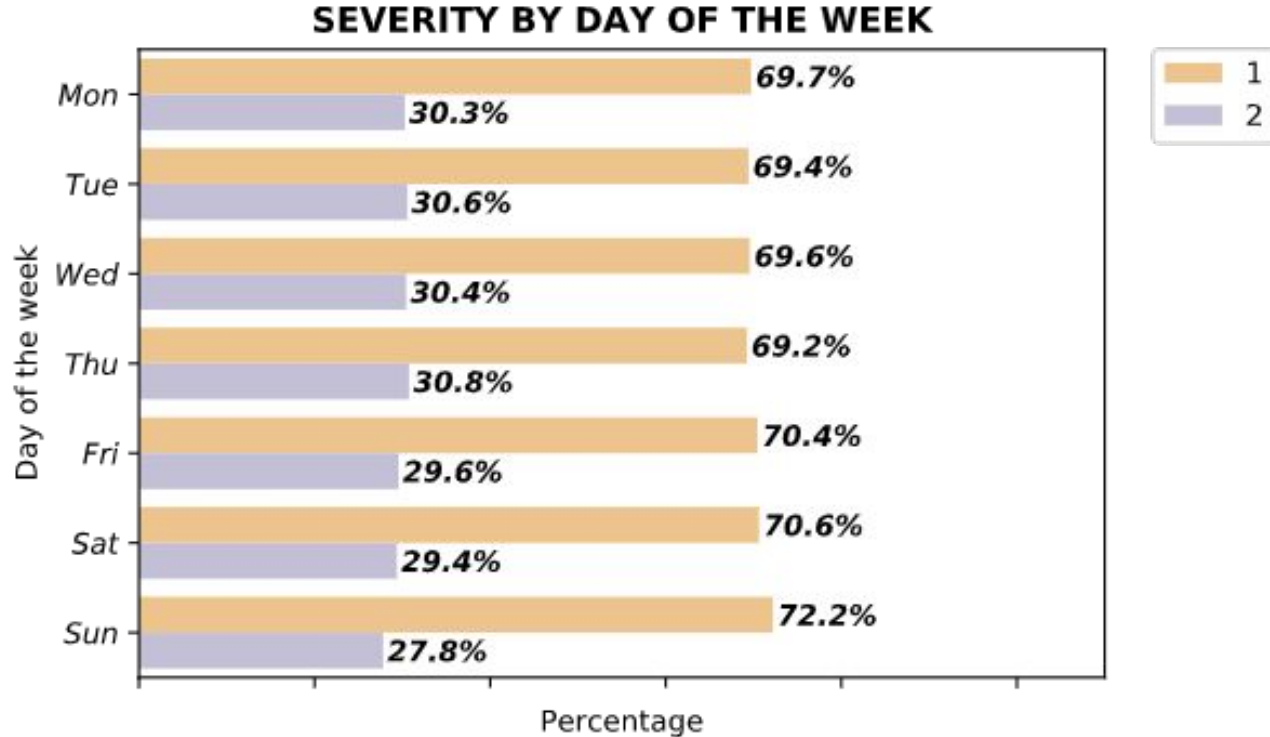
DAY OF THE WEEK

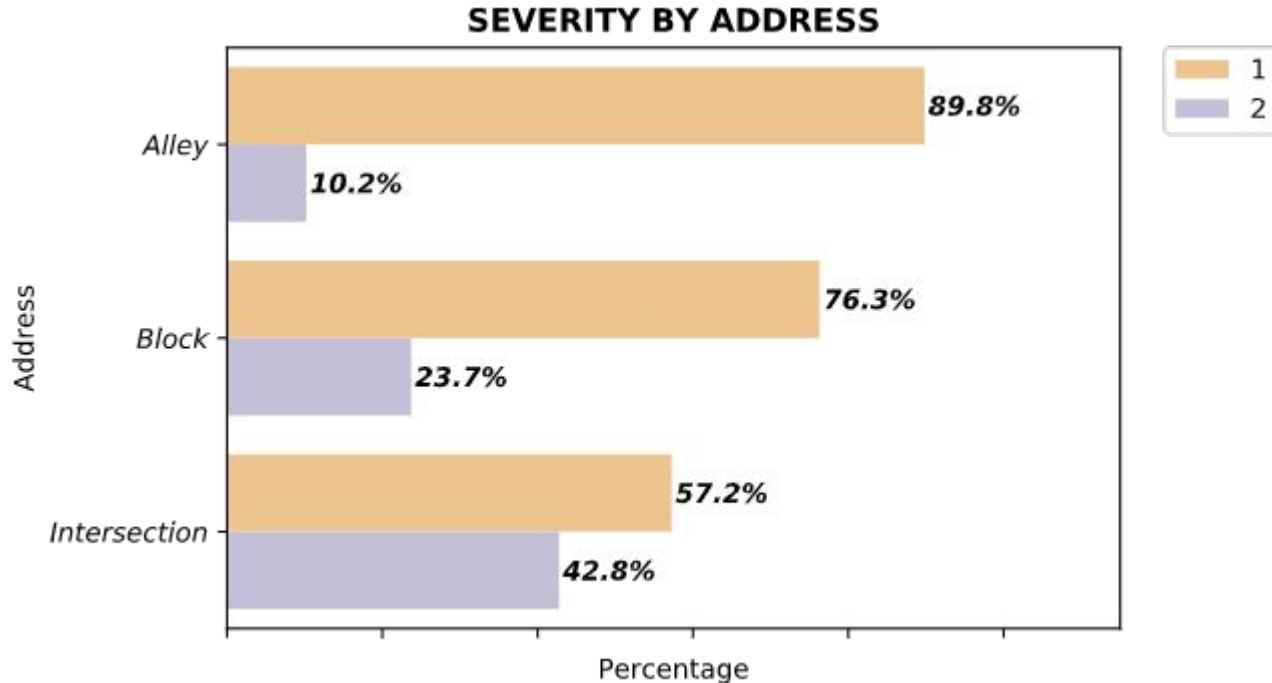Graph analysis shows that no day in particular presents a different critic/non critic incident ratio.

The distribution is the same throughout the whole week

SEVERITY BY DAY OF THE WEEK

Day of the week	1	2
Mon	69.7%	30.3%
Tue	69.4%	30.6%
Wed	69.6%	30.4%
Thu	69.2%	30.8%
Fri	70.4%	29.6%
Sat	70.6%	29.4%
Sun	72.2%	27.8%

Percentage

ADDRESS OF THE ACCIDENT

Graph shows as Intersection accident result more dangerous than the rest

SEVERITY BY ADDRESS

Legend: 1, 2

Alley: 89.8%, 10.2%
Block: 76.3%, 23.7%
Intersection: 57.2%, 42.8%

Y-axis: Address
X-axis: Percentage

CLASSIFICATION MODEL

A Logistic Regression model is implemented:

- Classification distinguishes between critic and non critic cases

- F-Score value of 0.85

CONCLUSION

- The analysis showed that:
  - Lots of conditions don't influence the statistic of the accidents
  - The districts show same behaviour in the incidents distributions, this gives that:
    - Districts with higher number of incidents have to be monitored more and more resources have to be invested in them
    - Districts with low number can reduce number of observations in favor of the high risk ones
  - The critical factors that present higher incidence of critics accident tell us that improvement can be made such as:
    - Pedestrian/Cycles interaction with traffic
    - Control of the intersections
  - This model could help in preventing the incidents in the city making hospitals and police departments to act faster and prepared
  - Much more data could be collected to improve the quality of the model as well as introduction of new features could help into making the model more accurate