IBM Data Science

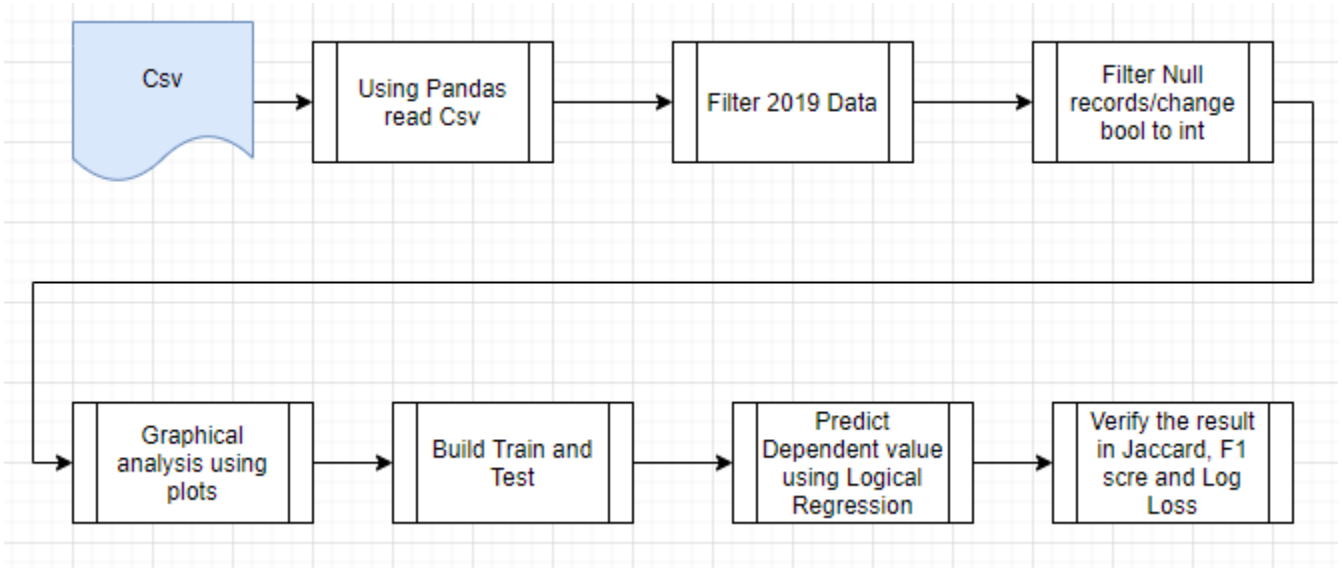
Final Capstone Project

Prepared By Vijayaragavan

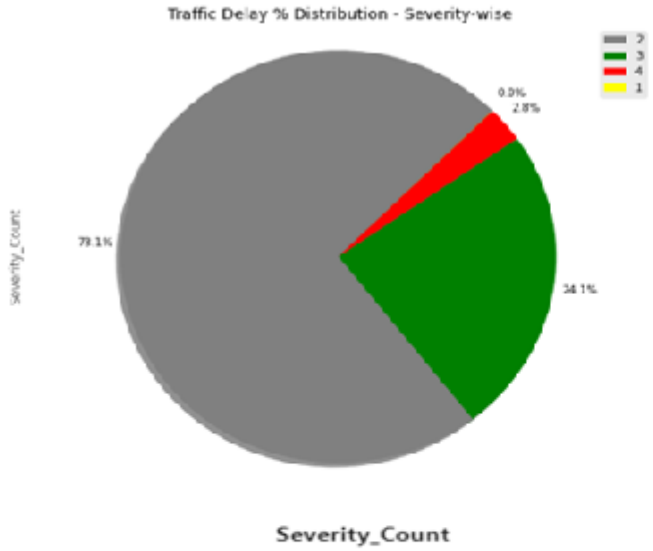
Introduction

- The aim of this project is predicting the traffic delay duration with given weather condition in United states. This model will be useful for who are new comers on the state for safe journey. A model to predict accident severity is built using these predictors as input vectors. Once the model is validated using a Machine Learning algorithm approach, we can be confident in predicting accident severity. In this proactive approach, the results of the analysis would be useful to various Entities like the Police and Insurance Companies. The goal is to reduce the fatalities and economic losses from accidents.
- Approximately 1.35 million people die each year as a result of road traffic crashes. The 2030 Agenda for Sustainable Development has set an ambitious target of halving the global number of deaths and injuries from road traffic crashes by 2020. Road traffic crashes cost most countries 3% of their gross domestic product. More than half of all road traffic deaths are among vulnerable road users: pedestrians, cyclists, and motorcyclists. 93% of the world's fatalities on the roads occur in low- and middle-income countries, even though these countries have approximately 60% of the world's vehicles. Road traffic injuries are the leading cause of death for children and young adults aged 5-29 years.

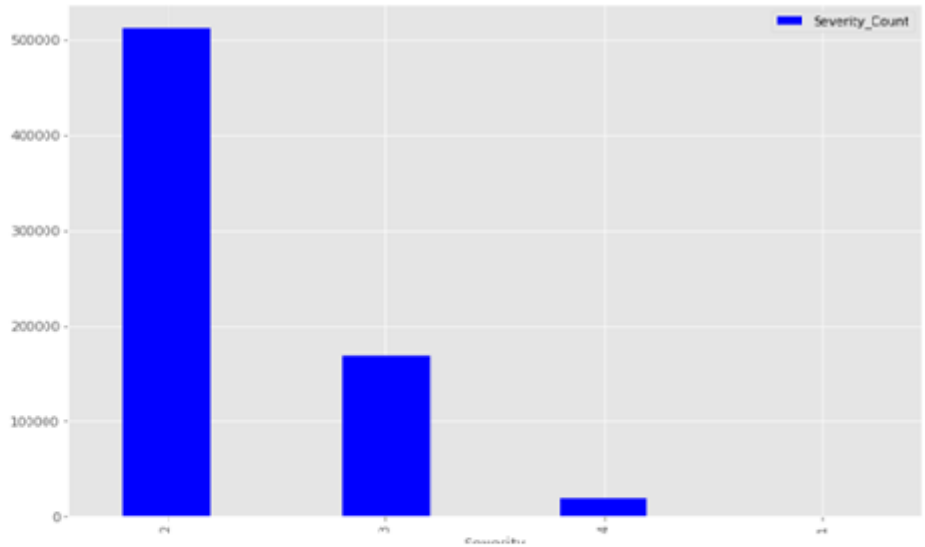
Process Flow



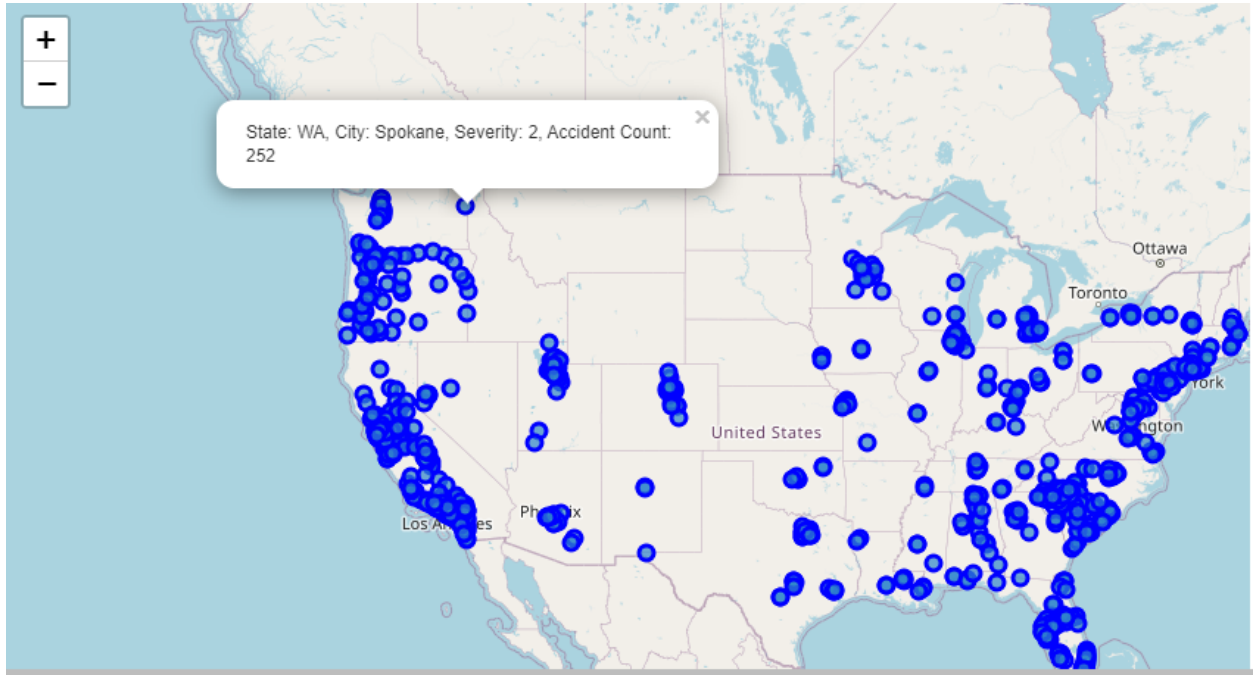
Severity Count Analysis – USA 2019 data



Severity	Severity_Count
2	511665
3	168942
4	19499
1	151



Accident counts with State/City Analysis



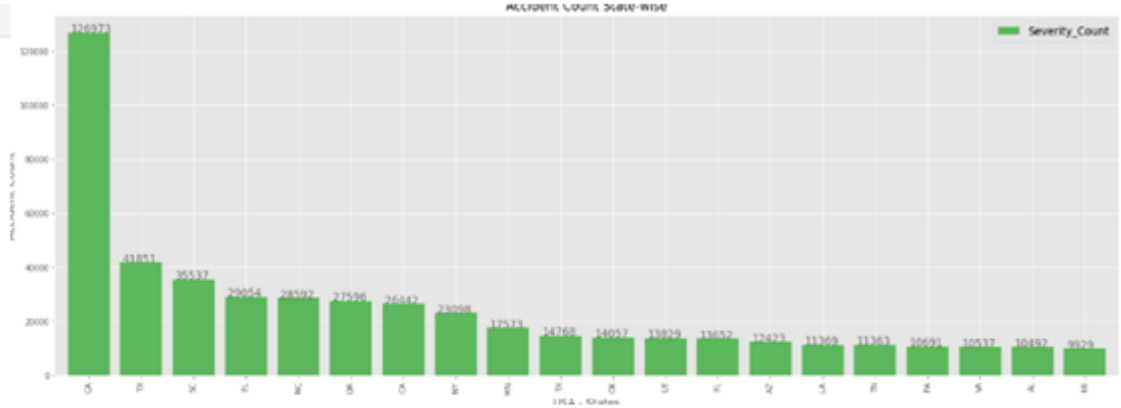
accdf_city

	Severity	State	City	SeverityCount	Lat	Lan
5257	2	NC	Charlotte	15278	35.391026	-80.570168
9062	2	TX	Houston	12746	30.029045	-95.133965
1136	2	CA	Los Angeles	11891	34.156310	-118.123697
8943	2	TX	Austin	11716	30.513420	-97.554733
7304	2	OK	Oklahoma City	8226	35.623718	-97.264999
...	...	...	...	...	...	...

Below Bar chart indicates Accident count analysis for statewise

```
df_state_vis.head(10)
```

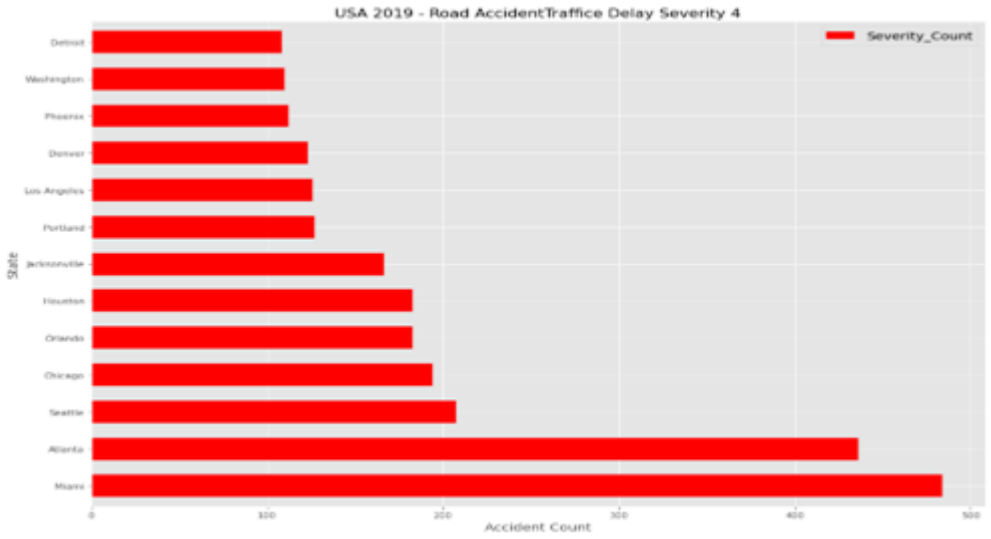
State	Severity_Count
CA	126973
TX	41851
SC	35537
FL	29054
NC	28592
OR	27596
CA	26442
NY	23098
MN	17573
TX	14768



Below Bar chart indicates Severity4 count analysis for Citywise

```
df_Sev4.head(10)
```

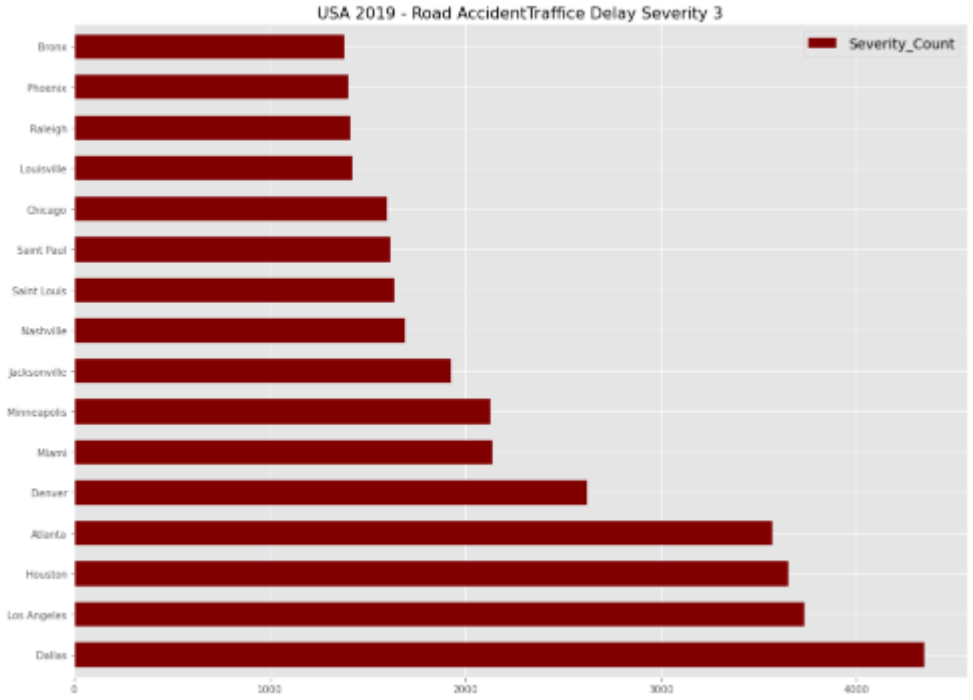
City	Severity_Count
Miami	484
Atlanta	436
Seattle	207
Chicago	194
Orlando	183
Houston	183
Jacksonville	166
Portland	127
Los Angeles	126
Denver	123



Below Bar chart indicates Severity3 count analysis for Citywise

```
df_Sev3.head(10)
```

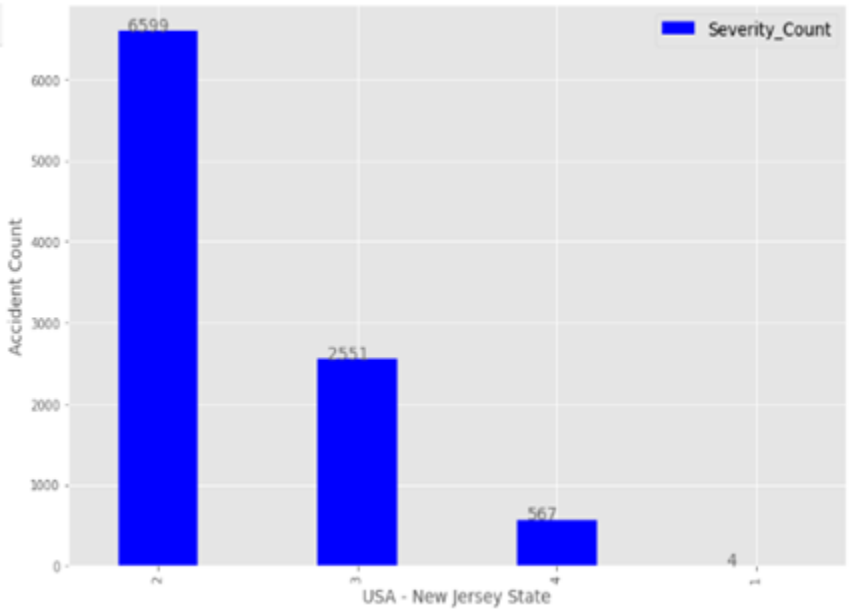
City	Severity_Count
Dallas	4348
Los Angeles	3734
Houston	3652
Atlanta	3569
Denver	2621
Miami	2138
Minneapolis	2130
Jacksonville	1927
Nashville	1694
Saint Louis	1636

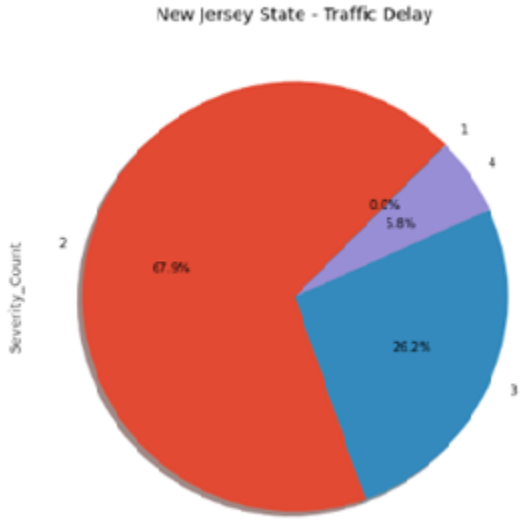


Below Bar/Pie chart indicates Severity count analysis for New Jersey

df_nj_cnt

Severity	Severity_Count
2	6599
3	2551
4	567
1	4





Conclusion

3. **Conclusion**:

- Built useful models to predict traffic delays and plan for travel in.
- Below is the final score in Logical Regression model.

```
knn_Jaccard = jaccard_similarity_score(y_test, yhat)
print (knn_Jaccard)
knn_f1_score = f1_score(y_test, yhat, average='weighted')
print (knn_f1_score)
ll_log_loss = log_loss(y_test, yhat_prob)
print (ll_log_loss)
```

```
0.9720675177791106
0.9592660214774792
0.05748875901860401
```

```
yhat = LR.predict(X_test)
yhat
```

```
array([2, 2, 3, ..., 2, 2, 2])
```