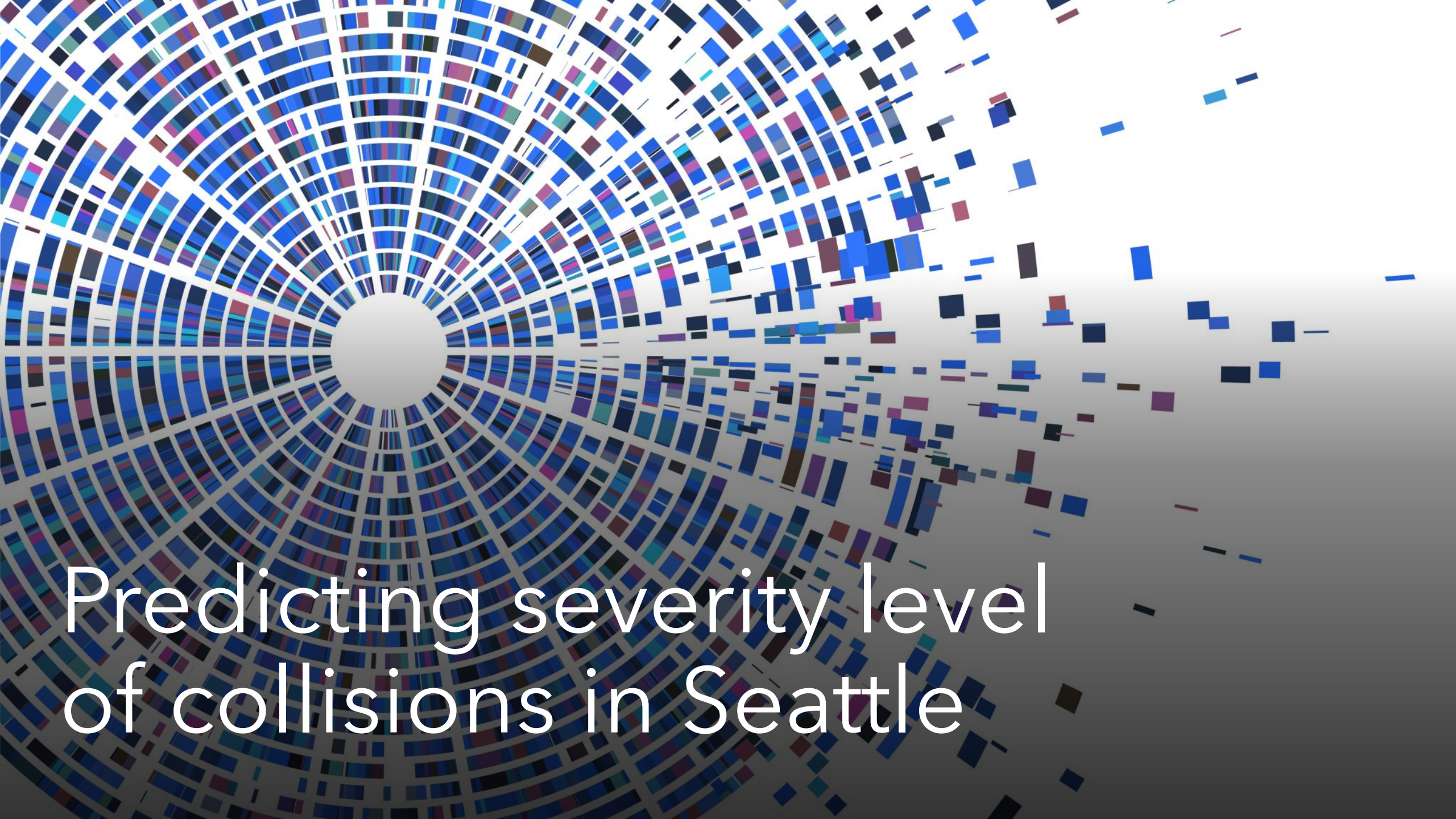


Predicting severity level of collisions in Seattle

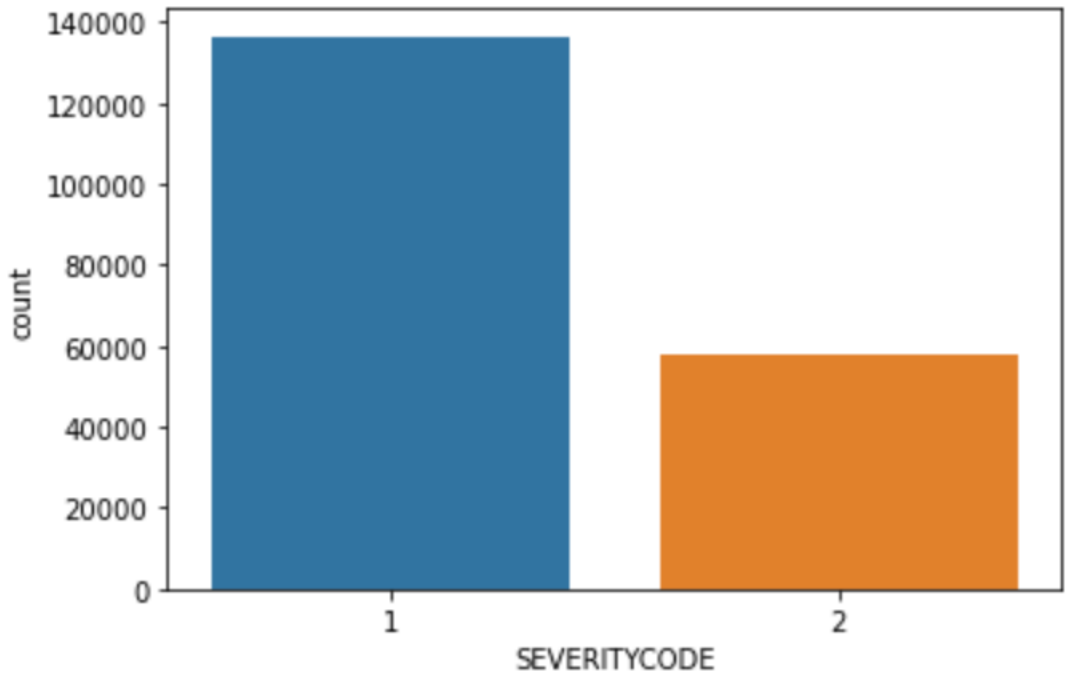
Predicting severity level of collision in Seattle is valuable for citizens and local government

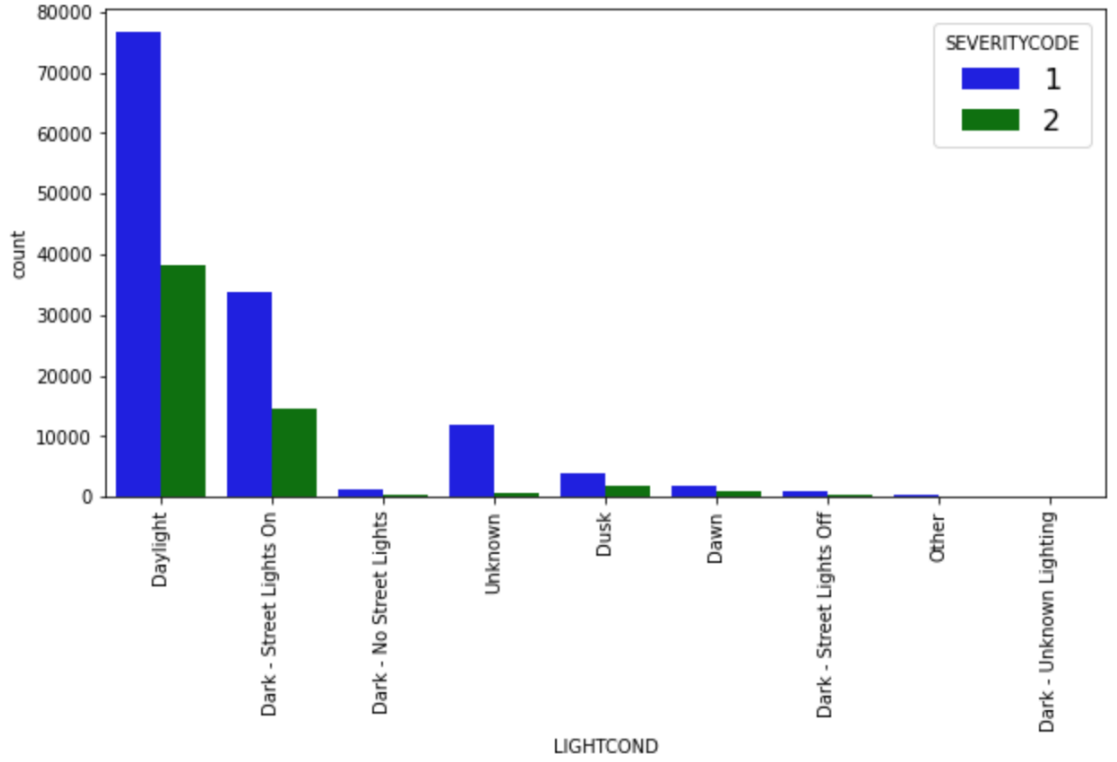
- Predictive model which will provide reasonable prediction if under certain conditions is higher probability of collision with a given severity
- Citizens of Seattle should be interested in this problem because knowing the relationships between conditions and likelihood of collision can save their money and life
- Goal is to provide them information about current situation on a roads and possible dangers – so it will be possible to avoid them

Data acquisition and cleaning

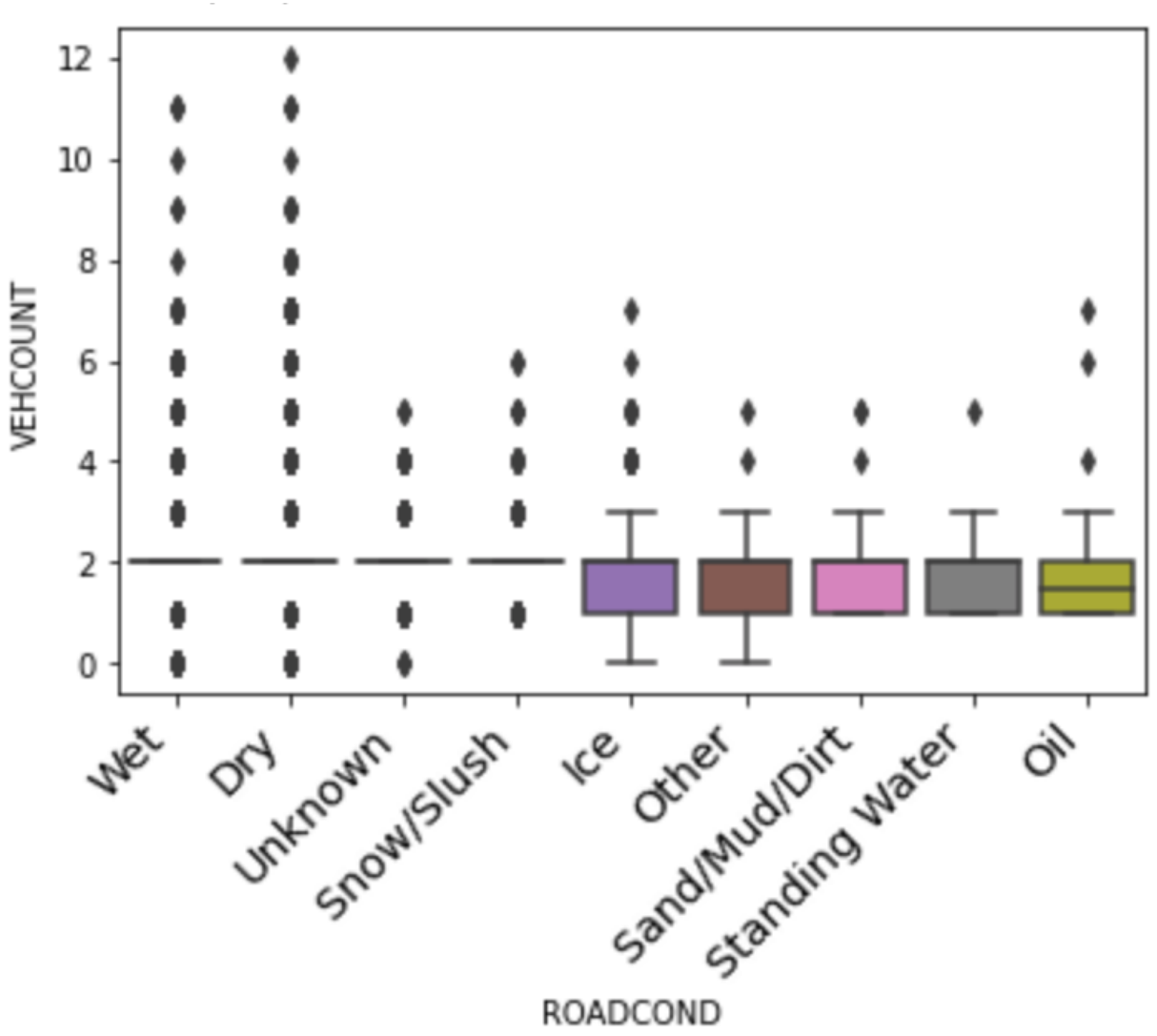
- Dataset Collisions—All Years provided by SPD and recorded by Traffic Records is available here and metadata about the dataset can be found here
- Dataset includes all types of collisions. Collisions will display at the intersection or mid-block of a segment. Timeframe: 2004 to Present.
- In the raw dataset were 194 673 rows and 38 columns
- Duplicate, highly similar or highly correlated features were dropped.

Compare counts of SEVERITYCODE 1 and 2 in the dataset

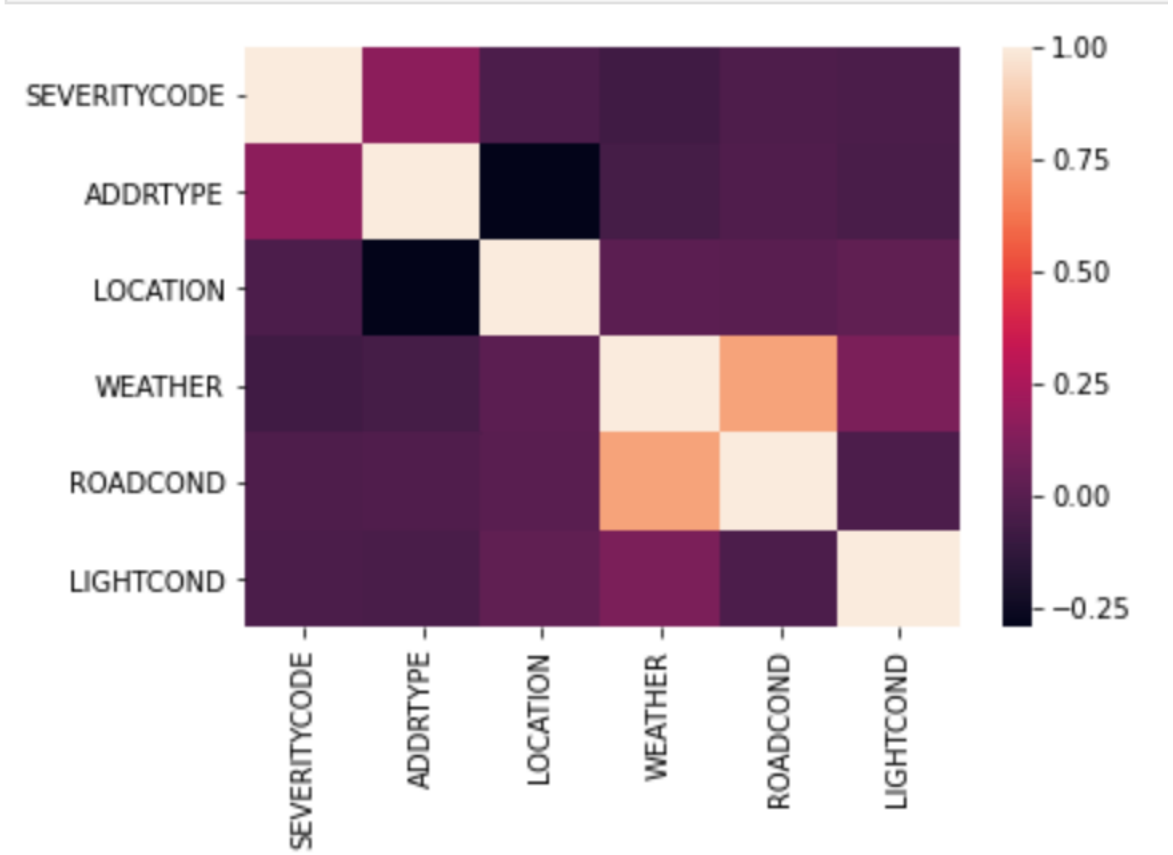




Collisions which happen on a wet or dry road has more outliers and collisions on ice, oil or standing water have similarly average 2 vehicle involved



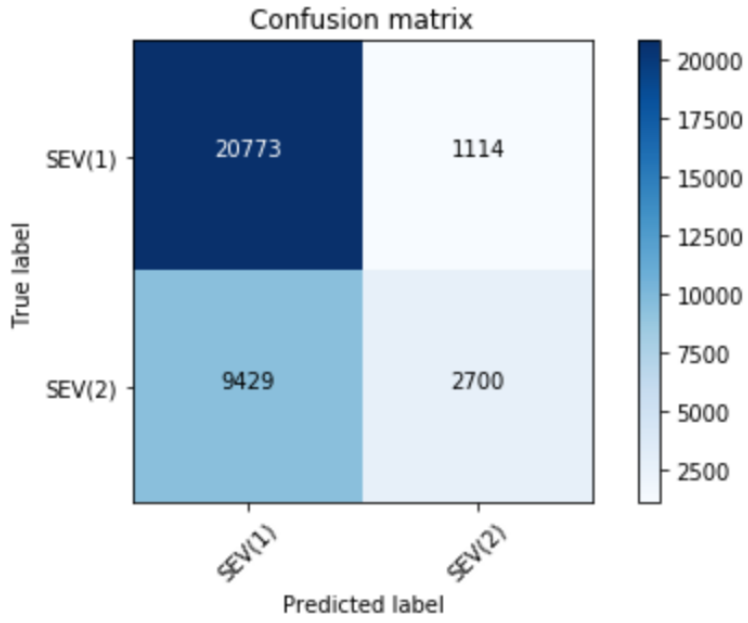
ADDRTYPE, LOCATION and ROADCONDITION have the biggest impact on SEVERITYCODE according to the matrix



Decision tree classifier model performance

	precision	recall	f1-score	support
1	0.69	0.95	0.80	21887
2	0.71	0.22	0.34	12129
accuracy			0.69	34016
macro avg	0.70	0.59	0.57	34016
weighted avg	0.69	0.69	0.63	34016

```
Confusion matrix, without normalization
[[20773  1114]
 [ 9429  2700]]
```



Logistics regression model performance

```
LR Jaccard index: 0.64
LR F1-score: 0.60
LR LogLoss: 0.65
```

Conclusion and future directions

- Built useful models to predict severity code according to a given conditions
- Accuracy of the models has room for improvement
- Capture more data about binary conditions, traffic (number of vehicles) and hour of an accident