PREDICTING SEATTLE TRAFFIC COLLISIONS

Collisions from 2004 to Present

Source: SPD, Traffic Records

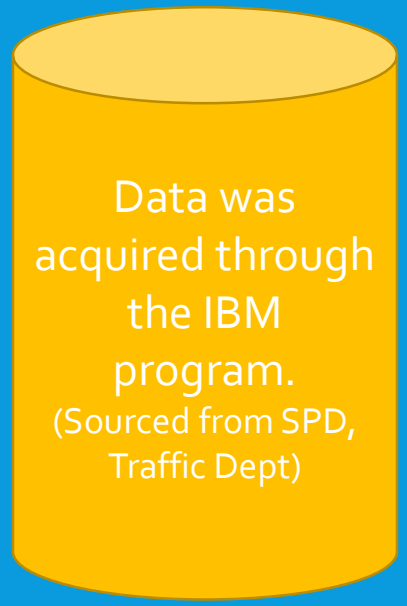
PREDICTING TRAFFIC COLLISIONS

The Seattle administration has been reviewing accident cases to deploy preventive measures to avoid damage to individuals and property.

An ability to convert historic accident reports and use that to help state police and authorities to create targeted awareness measures to prevent accidents will be helpful.
Further, insurance companies would also benefit by preventive and pro-active counter-measures that will reduce third-party property damage.

Using existing fields from records such as environmental factors, the state can use this model to preempt accidents and plan its response accordingly. The target audience for this model are the state police department, emergency services, town planners and insurance companies that can use this model to plan better and reduce accidents.

DATA ACQUISITION AND CLEANING



37 features were provided

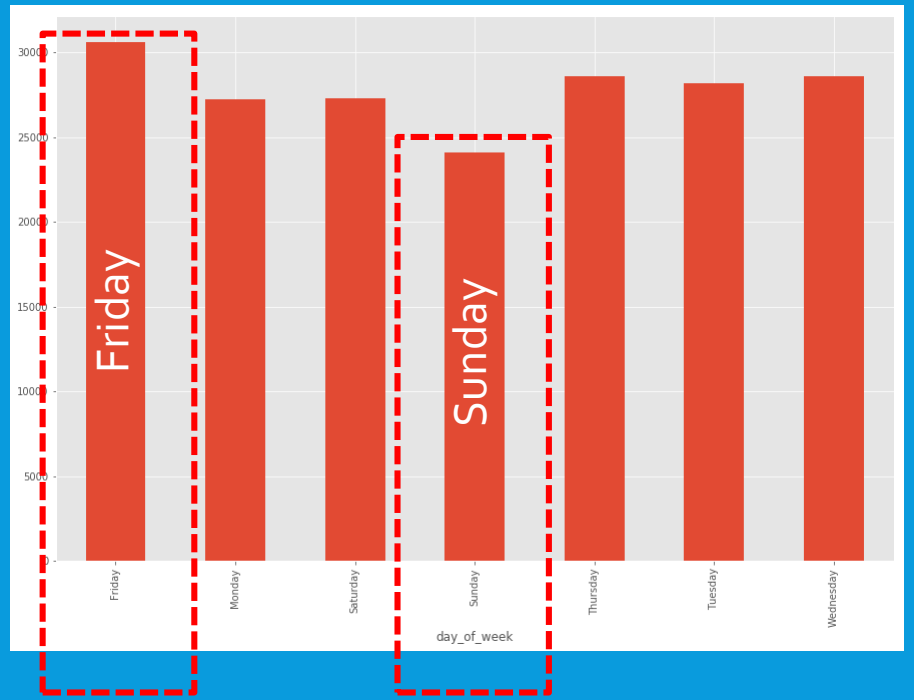
2 severity indicators were provided:
1 – Property Damage
2 - Injury

From an assessment of the data, the following features are considered to evaluate the model for severity prediction.

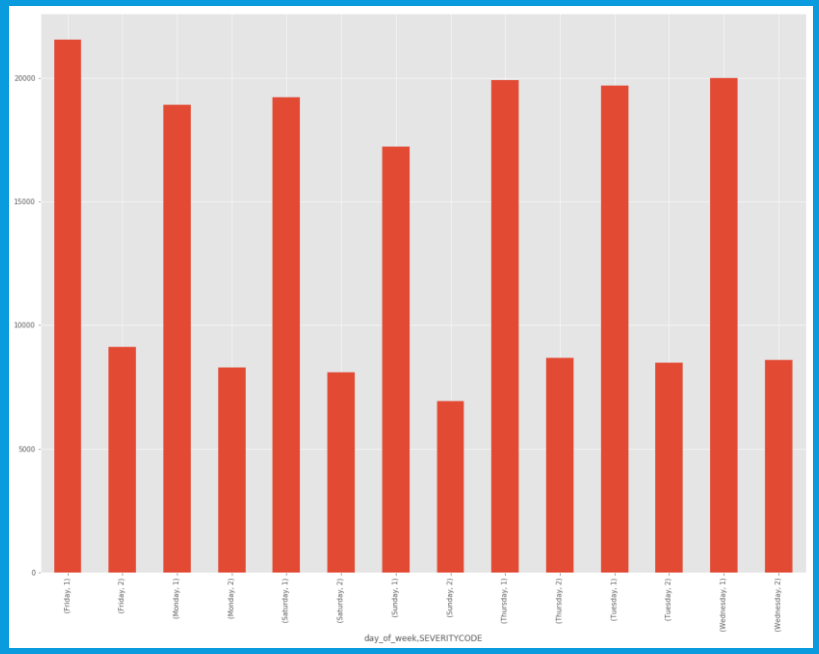
- **Junction Type** - Midblock, intersections have higher incidents.
- **Weather** - Rain/Overcast do indicate a higher incident rate compared to other non-Clear conditions.
- **Road Cond** - Wet conditions indicate higher rates. This is related to the weather. No other road conditions (other than dry) seem to drive the incident rates.
- **Light Conditions** - Dark(with Streetlights) has a higher rate of incidents.
- **Day of the week:** Derived column based on incident date - shows that incidents are higher on a Friday and lowest on Sundays. Under influence and Speeding were not considered as completeness and accuracy of data was not established.

USING DAYS OF THE WEEK

Using the feature containing incident time, the day of the week was derived to determine weekly patterns

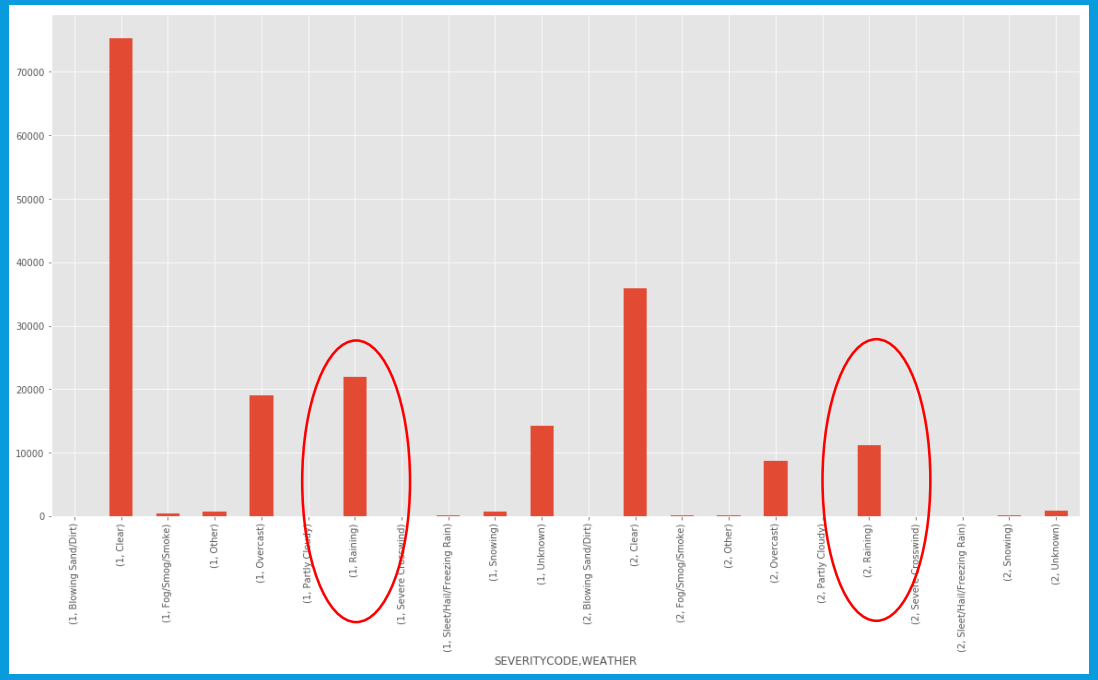


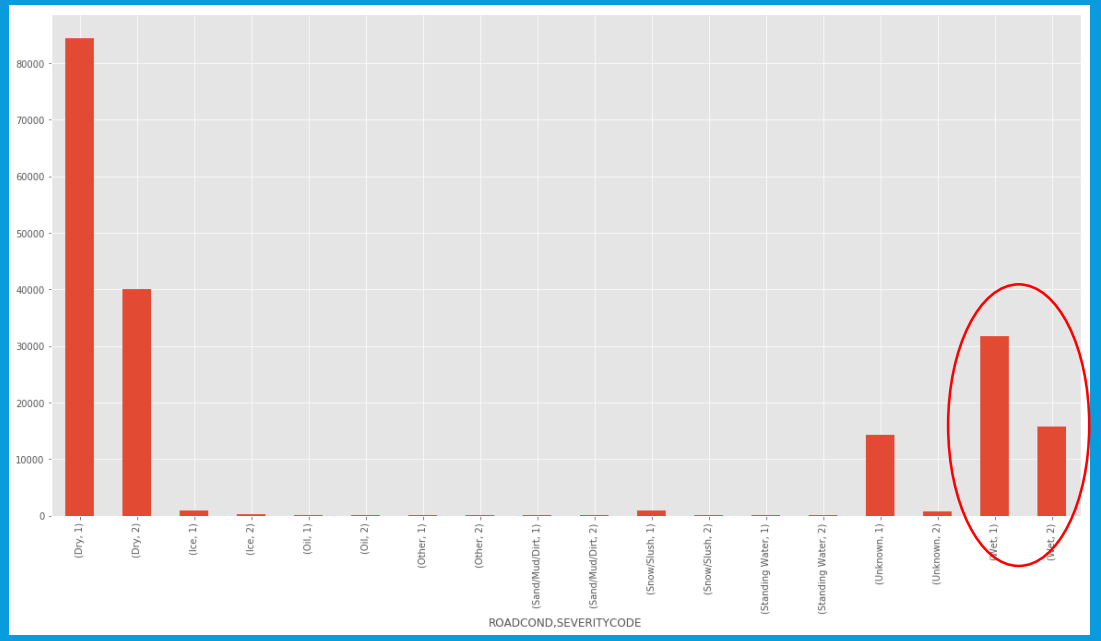
Severity by day



Incidents were higher on Friday and lowest on Sundays. Accordingly, authorities can plan awareness programs to reduce property damage and injuries.

WEATHER AND ROAD CONDITIONS



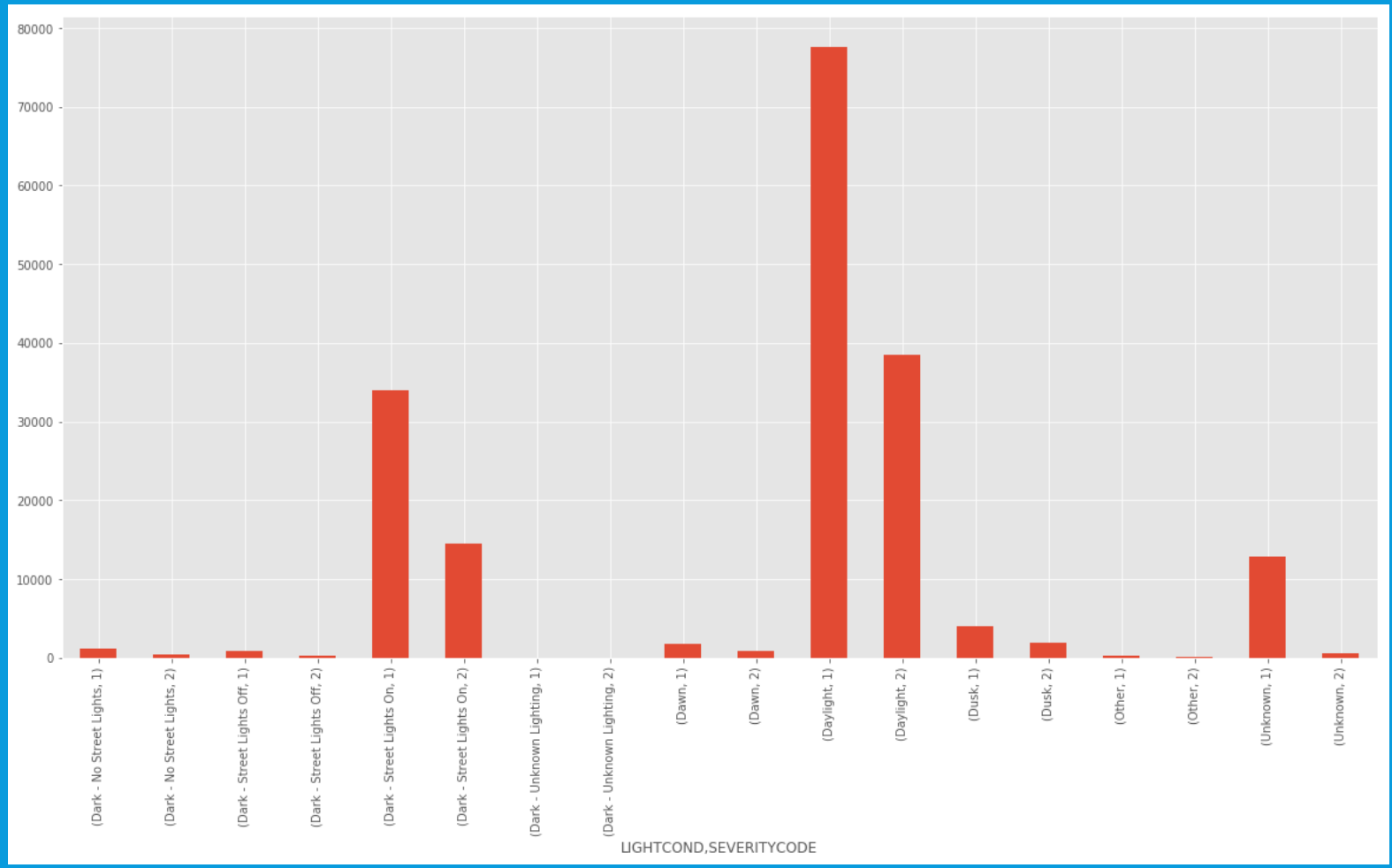


Apart from Dry road conditions, the next significant feature was a wet road – that is largely due to 'Raining' weather conditions. Only weather was considered in the model.

LIGHT CONDITIONS

Dark light conditions caused a significant number of incidents (despite street lights).

There were unknown elements in this feature and all these rows were dropped.



REGRESSION MODELS

Model	F1 score	Jaccard Score
Logistic Regression	59%	59.3%
K-nearest Neighbour	52.3%	55.2%
Decision Tree	61.3%	59.1%
Support Vector Machine	58.5%	58.9%

4 regression models were evaluated with accuracy levels varying from 52% to 61%. Accuracy rates could improve with more data. Currently, to prevent bias, a huge number of Severity 1 instances were deleted from the dataset to maintain a more proportionate Severity 1 and Severity 2 ratio.

CONCLUSION AND WAY FORWARD

From the analysis of the data and the algorithms, predictive capabilities is better than a random guess (i.e 50% accuracy). However, from the data, the following efforts could be made to reduce incidents and create more awareness:

1. Public programs on Friday to reduce higher incident rate.
2. During bad weather conditions where there is rain (wet road conditions), public could be cautioned over the radio to be cautious.
3. During poor lighting conditions and dark roads, better lighting can be made available

WAY FORWARD:

1. Continue to collect more data so the models can be trained better
2. Ensure completeness of data such as speeding, driving under the influence, etc.