


An aerial photograph of a winding asphalt road that curves through a dense, green forest. The road is light gray and contrasts with the dark green trees. The forest covers a hillside, and the road appears to be a two-lane road with a white line marking. The overall scene is captured from a high angle, looking down at the road as it snakes through the landscape.

DATA SCIENCE CAPSTONE PROJECT

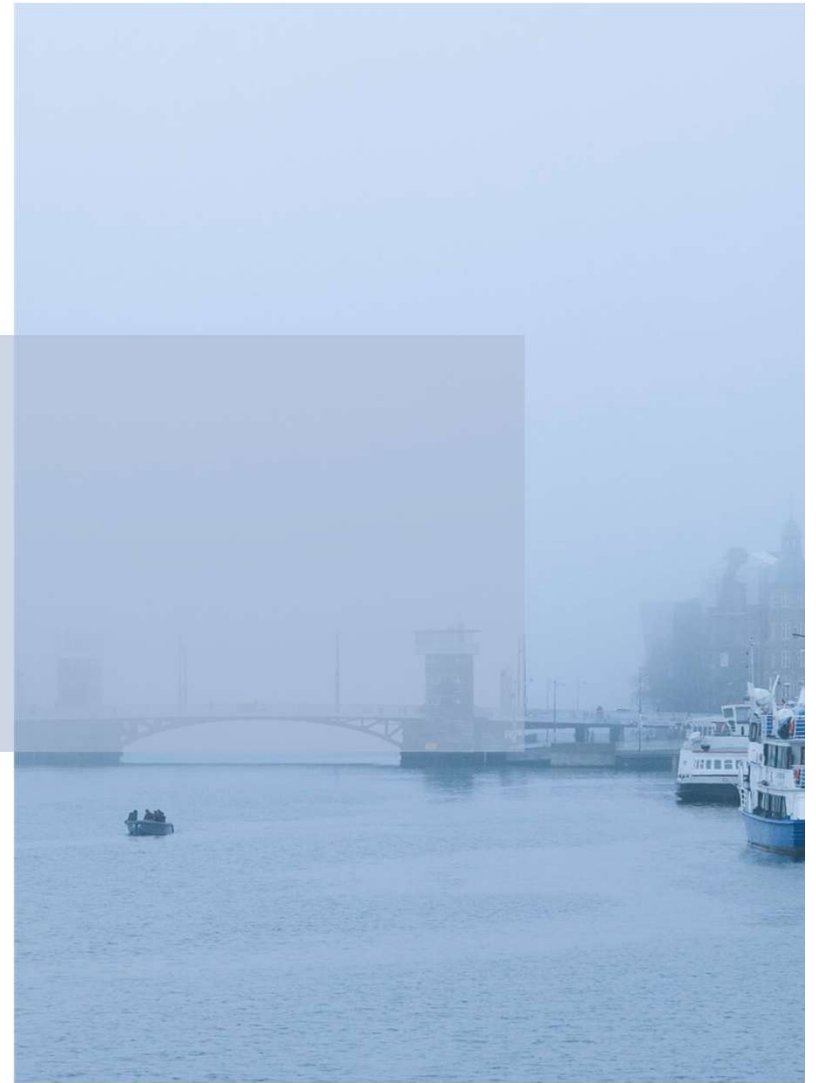
CAR ACCIDENT SEVERITY
SEATTLE

TABLE OF CONTENTS



Introduction
Data
Methodology
Predictive Modeling
Discussion
Conclusion

INTRODUCTION





TRAFFIC

Traffic injuries happen all the time and it is increasingly becoming the main cause of death globally as the world has more and more vehicles. Many researches have conducted to analyze a range of factors, including weather conditions, special events, roadworks, traffic jams among others, in order to perform an accurate prediction of the severity of the accidents.



addresstype



junctiontype



weather



**Road
condition**



Light condition

*This project aims to
predict Severity of
accident based on
these metrics.*

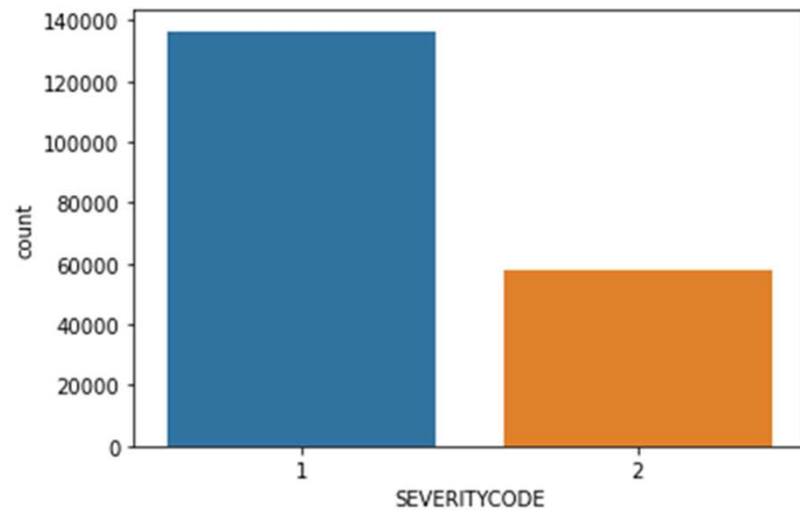


DATA



DATA

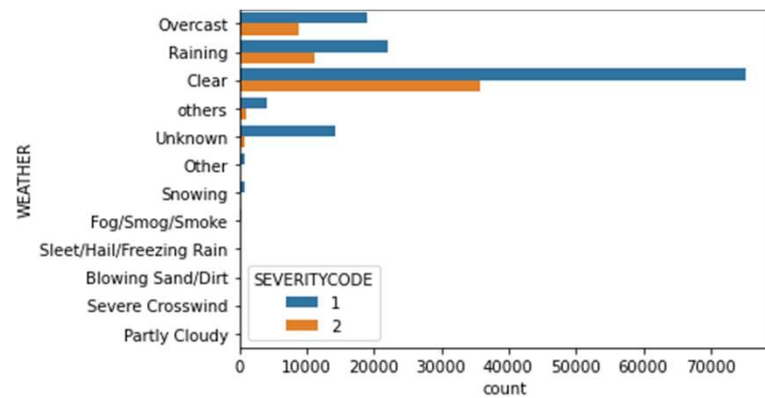
The original link of dataset is provided below. It has around 200000 samples and about 60 attributes, which is provided by Seattle Department of Transportation. It is a well-rounded dataset, and includes how and when the accidents took place with number of people involved and weather condition. Meta-data of the dataset also offered. To simplify the analyzing process, some key attributes are chosen and some features are dropped.



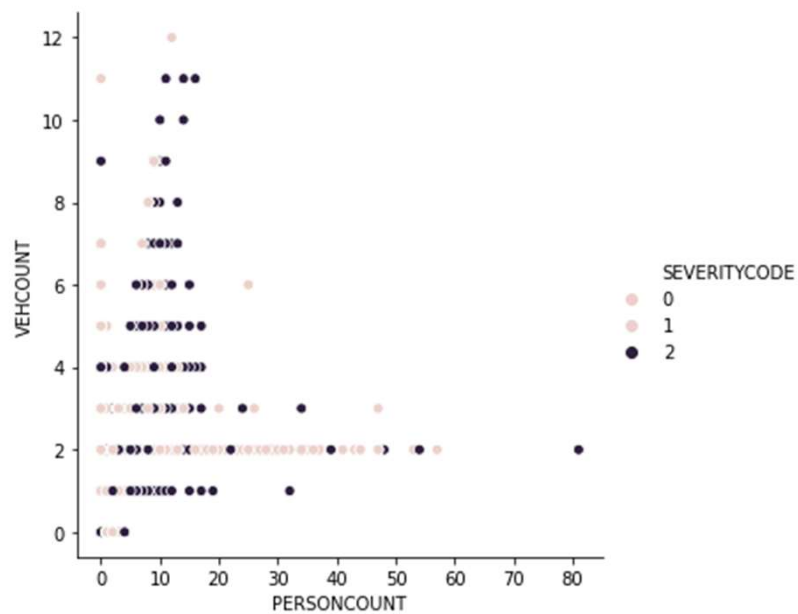
From the diagram above, we can find the severitycode has only two types, one stands for property damage and two stands for injury. Below, we will simply discuss the relation between some factors and severitycode, so called the severity of the accident.

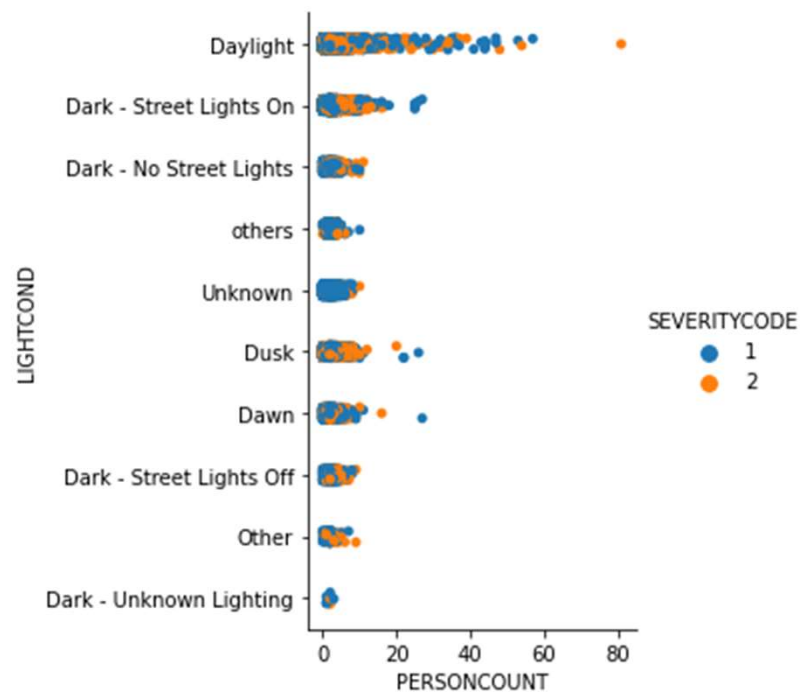


METHODOLOGY



**Relationship between
weather and
seveitycode and
accident**





**Relationship between
LIGHTCOND and
personcount and
SEVERITY of
Accidents**

	SEVERITYCODE	ADDRTYPE	JUNCTIONTYPE	WEATHER	ROADCOND	LIGHTCOND
0	2	2	1	4	8	5
1	1	1	4	6	8	2
2	1	1	4	4	0	5
3	1	1	4	1	0	5
4	2	2	1	6	8	5
...
194668	2	1	4	1	0	5
194669	1	1	4	6	8	5
194670	2	2	1	1	0	5
194671	2	2	1	1	0	6
194672	1	1	4	1	8	5

```
array([[ 1.28668708, -1.19903701,  0.22981187,  1.39847224,  0.25900713],
       [-0.69338623,  0.73412368,  0.87758556,  1.39847224, -1.36653782],
       [-0.69338623,  0.73412368,  0.22981187, -0.73846749,  0.25900713],
       ...,
       [ 1.28668708, -1.19903701, -0.74184867, -0.73846749,  0.25900713],
       [ 1.28668708, -1.19903701, -0.74184867, -0.73846749,  0.80085545],
       [-0.69338623,  0.73412368, -0.74184867,  1.39847224,  0.25900713]])
```

**Convert Categorical
features to numerical
values & Normalize
Data**

DISCUSSION



K-Nearest Neighbor (KNN) KNN will help us predict the severity code of an outcome by finding the most similar to data point within k distance.

KNN Jaccard index: 0.69

KNN F1-score: 0.60



Decision Tree A decision tree model gives us a layout of all possible outcomes so we can fully analyze the consequences of a decision. Its context, the decision tree observes all possible outcomes of different weather conditions.

DT Jaccard index: 0.70

DT F1-score: 0.58



Logistic Regression Because our dataset only provides us with two severity code outcomes, our model will only predict one of those two classes. This makes our data binary, which is perfect to use with logistic regression.

LR Jaccard index: 0.70

LR F1-score: 0.59

LR LogLoss: 0.58



A nighttime photograph of a city street. The street is wet, reflecting the lights from the buildings and street lamps. On the right side, there are several parked cars. In the background, there are multi-story buildings with lit windows. A semi-transparent grey rectangular box is overlaid on the center of the image, containing white text.

It can be suggested that the features like Weather, Road and Light conditions, addresstype, junctiontype have impact on severity in accidents, they can also help to predict the severity of the accident, that may result in injury or property damage.



*ADDRTYPE, JUNCTIONTYPE, WEATHER, ROADCOND, LIGHTCOND
these five factors to analyze whether they are relative to the severity of
accident.*

*By performing three models listed here: K-Nearest Neighbor (KNN)
Decision Tree Logistic Regression. Obviously, the model and its results
provide some key insights for the target audience to make impactful
decisions in reducing the number of accidents and injuries in their
localities.*

CONCLUSION

THANKS

