

Predicting Severe Accidents in Seattle

By Heather Spero



Solving the Accident Problem

Seattle had 187 serious accidents in 2017

Many factors impact including weather, vehicle, light etc.

Goal: Predict the probability of a severe accident

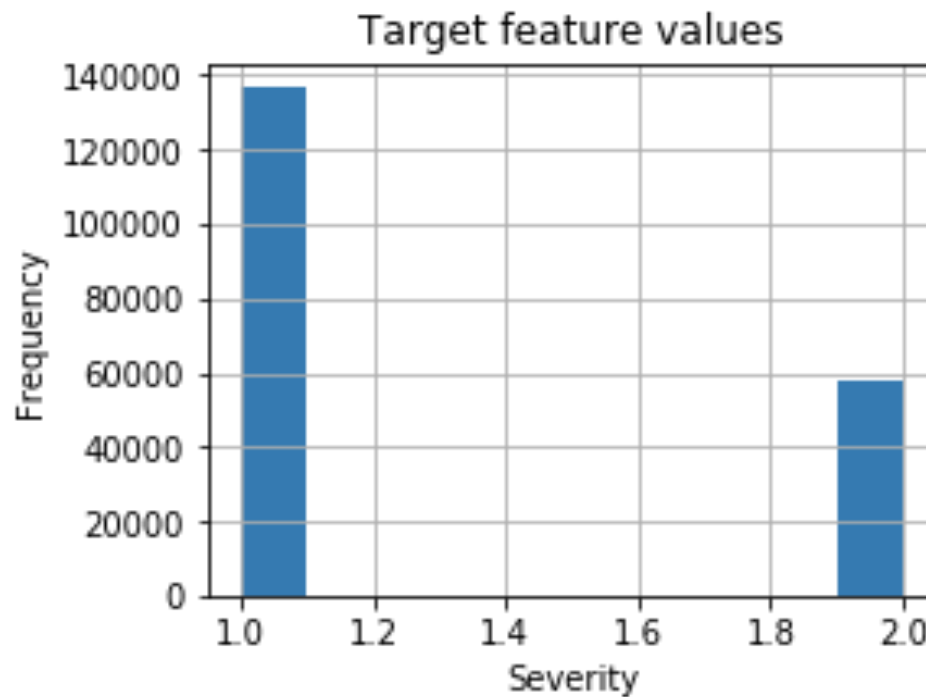
People could determine the risk of an accident before they go out, impact their decision on timing and whether to go

Police, Hospitals could use the data to forecast employment

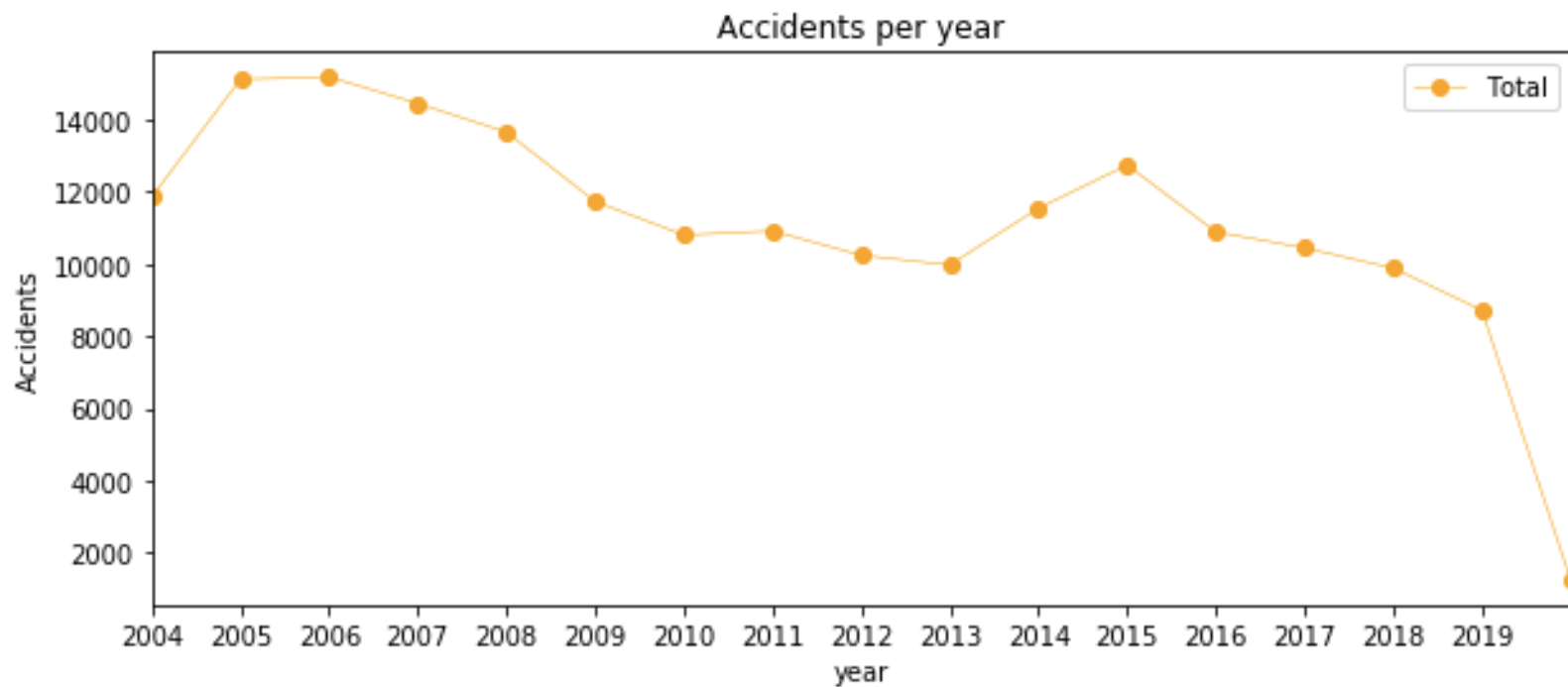
Data Acquisition/Cleaning

- ◆ Data was provided by SDP recorded by Traffic Records
- ◆ Target Value is SEVERITYCODE: (1 – not severe, 2 – severe)
- ◆ Primary variables included in analysis: Road Conditions, Light Conditions, year, month

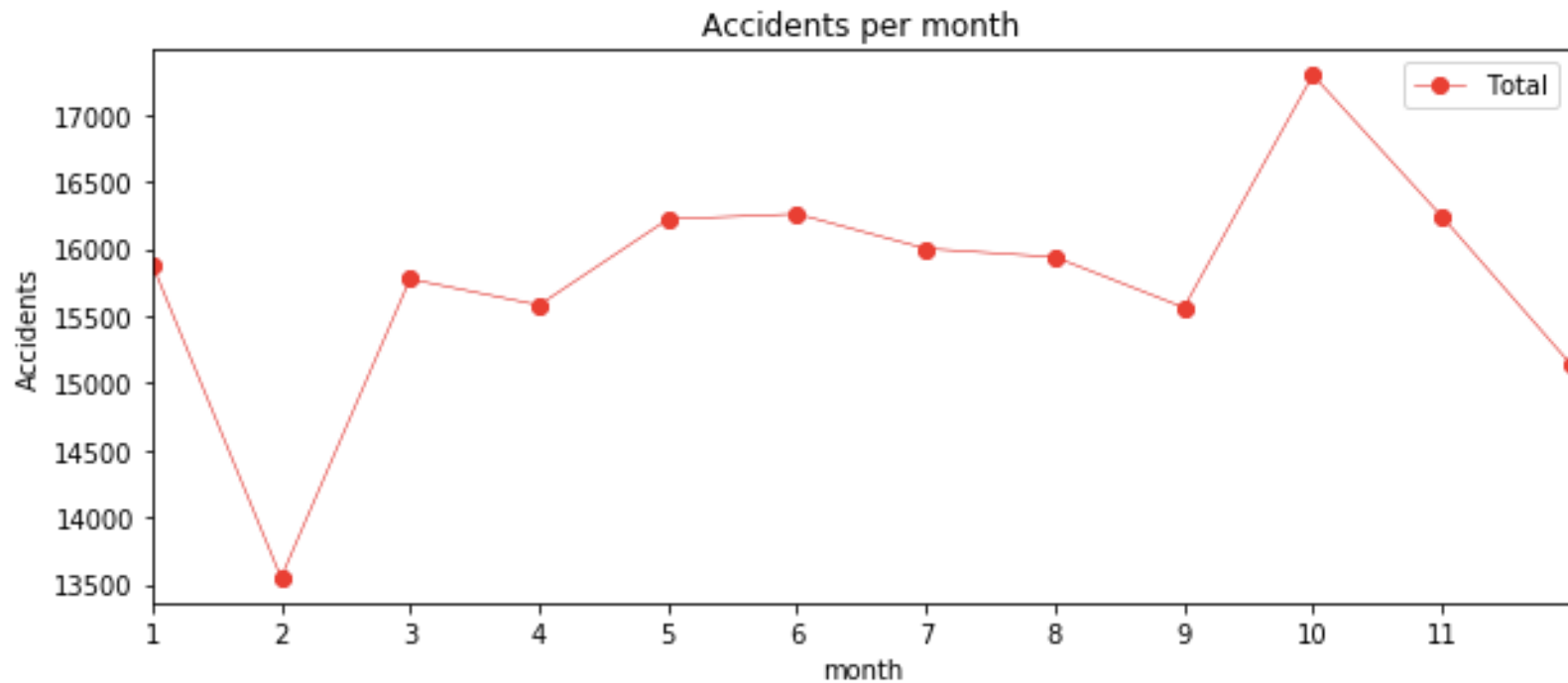
Target Distribution



Accidents Per Year



Accidents Per Month



Correlation of Each Variable

	RD_CAT	LC_CAT	year	month
SEVERITYCODE	-.035	-.104	.023	.004

Review of Correlation

- ◆ There appears to be no significant correlation between the individual variables and the Target variable
- ◆ Most significant variable appears to be day light conditions however the value is still not significant
- ◆ The data will be analyzed through clustering and Logistic Regression to try and improve the predictive method

Predictive Methods

- ◆ KNN
- ◆ Decision Tree
- ◆ Logistic Regression

Summary of Results for Severe Accidents

Algorithm	Jaccard	F1 Score	Precision	Recall
KNN	0.68	0.35	0.05	0.09
Decision Tree	0.7	0.0	0.0	0.0
Logistic Regression	0.7	0.0	0.0	0.0

Discussion

- ◆ The predictive model for severe accidents is not strong/
existent in most of these models
 - ◆ The number of Severe accidents in the data was low and
therefore maybe more data points are needed
- ◆ The Jaccard score and the predictive model for non-severe
accident statistics are high
- ◆ A model to predict non severe accidents could be possible
however does not address all the concerns