Seattle traffic incidents analysis

Applied Data Science Capstone by IBM/Coursera

Overview

Seattle is a beautiful city, and famous of its America northern west coast climate -- long, greyish and raining fall/winter. Driving in well-developed city like Seattle should be safe, and the data shows it is only becoming safer.

Most of the incident occurred without too significant cause. And it seems more reasonable to keep doing what the city has been doing right: Improve the road (wider/flatter roads with better tractions, and lighting), the interception design (visibility, signals, flow controls), reinforce traffic bylaws (Speeding, Impaired driving), specific road user segments will help reduce the collision to especially cyclists and pedestrians, and so on.

We may have drill down further to find out if there are predictability potentials, by developing more models and try out some other data combinations. HOWEVER, just like in the discussion session … with the yet happened incident(s), we may ONLY take projection from data from elsewhere like Speeding, Impaired Driving, infrastructure construction, population, weather station statistics.

Discussion on the Dataset / Results

- **Result**

  Exploratory Analysis have brought certain insights from the statistics. We are able to answer some of the questions about Seattle's traffic incidents.

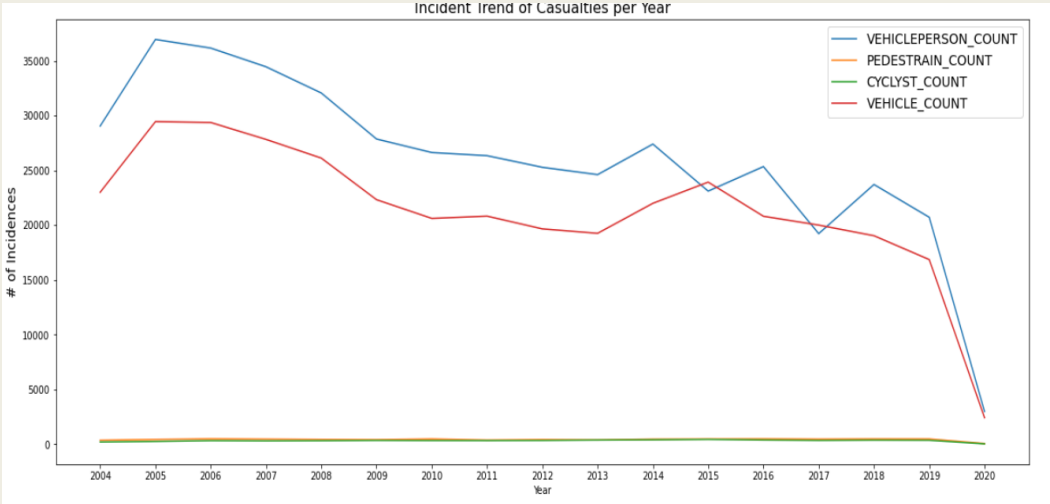
  Modeling analysis have encountered some difficulties in poor model performance, which may indicate underfitting of the model but with the R2 to be relatively close to 100% which may also indicate a good prediction. None of those are good models per the MSE but the "Ridge Regression w/h @14300" with "Polynomial order = 3" may stand out among the models.

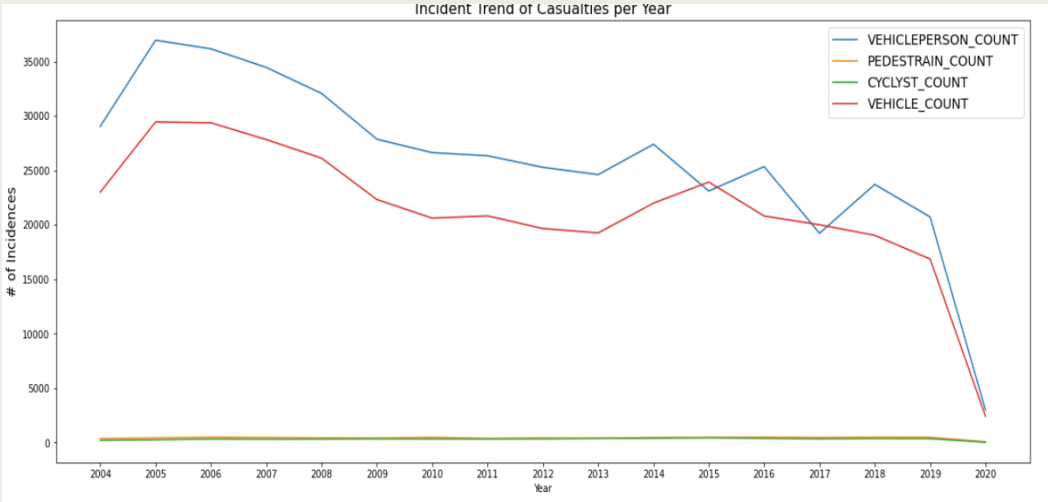
- Discussion

  Besides exploratory analysis, Is the dataset suitable for modeling analysis? It is obvious the "features" are all dependents to the incidents itself, while modeling and prediction will require external factors / data which should be independent to the incidents (i.e. weather statistics, traffic reinforcement records, road construction schedules, and locations of some major amendities)

Seattle is getting safer on the road

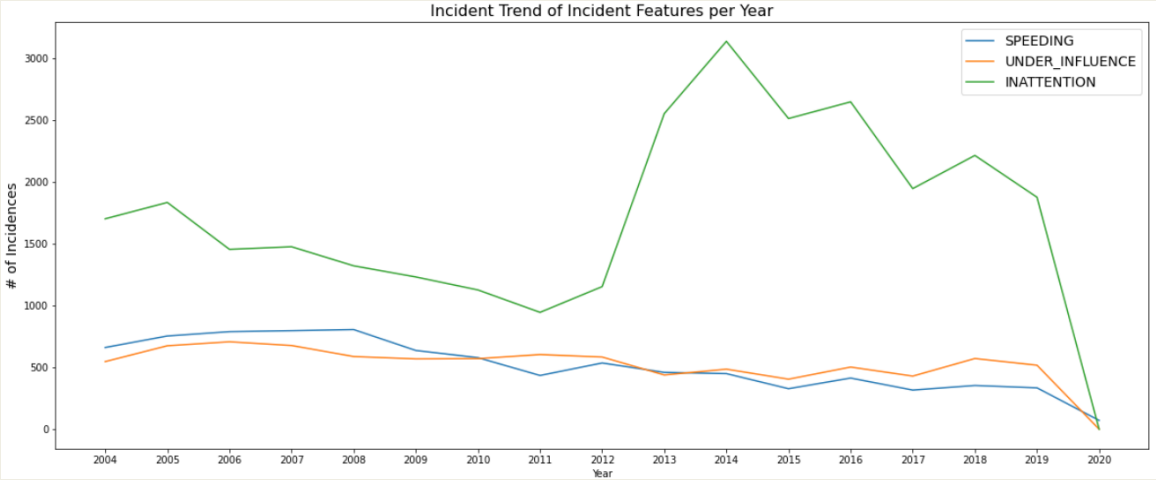
The Incident Trend is trending downwards while the Casualty Trend has the similar trend. Apparently, the pandemic has caused significant drop of records in 2020 which we may ignore in any case coming across the same year.

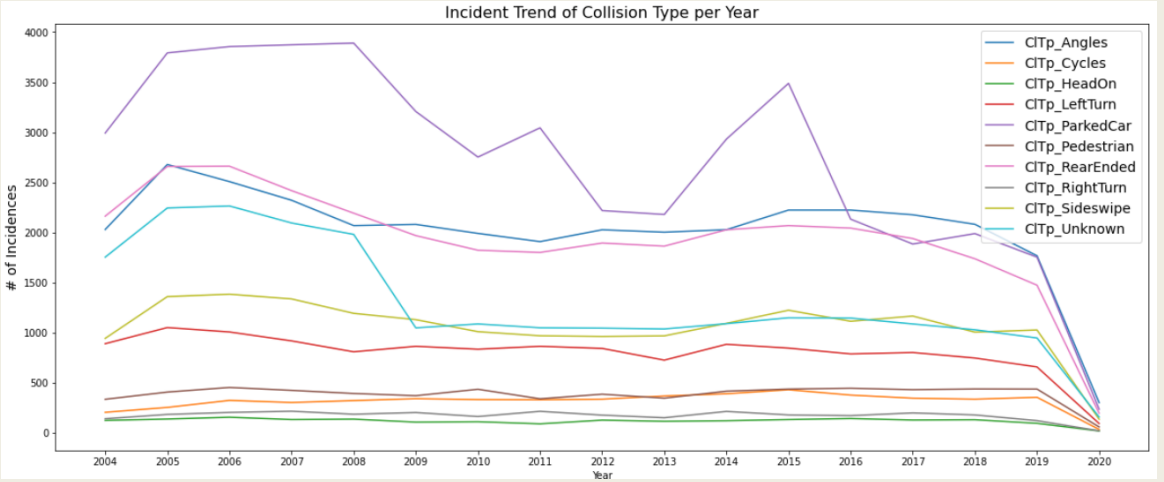


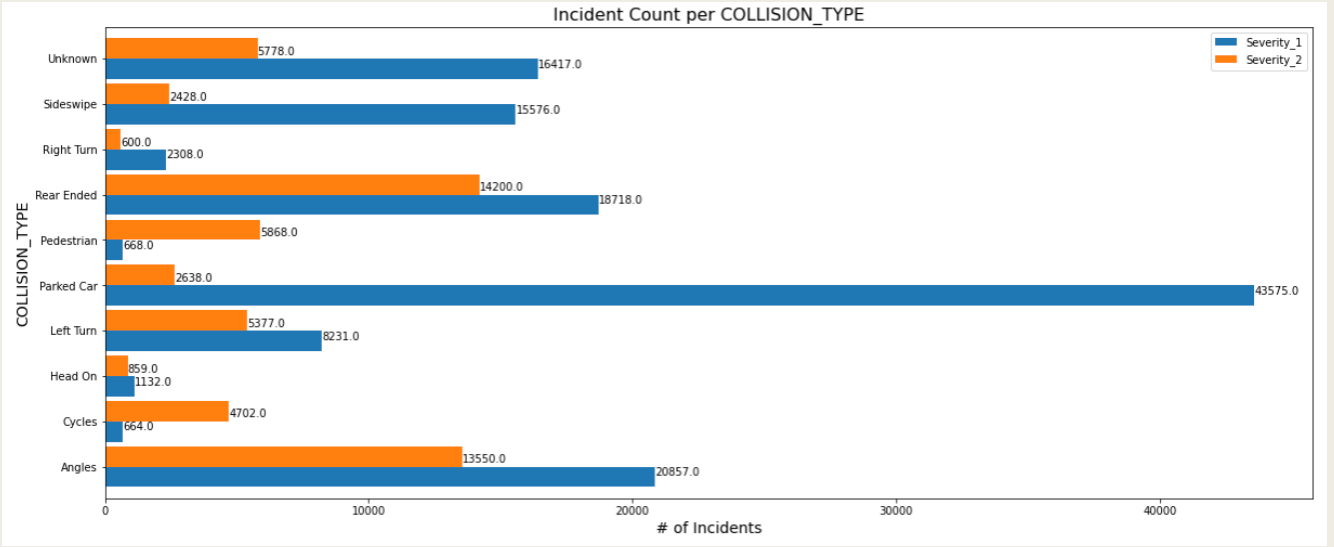


Major causes of incidents

Inattention driver who may also Hitting a Parked car are the major causes of incidents

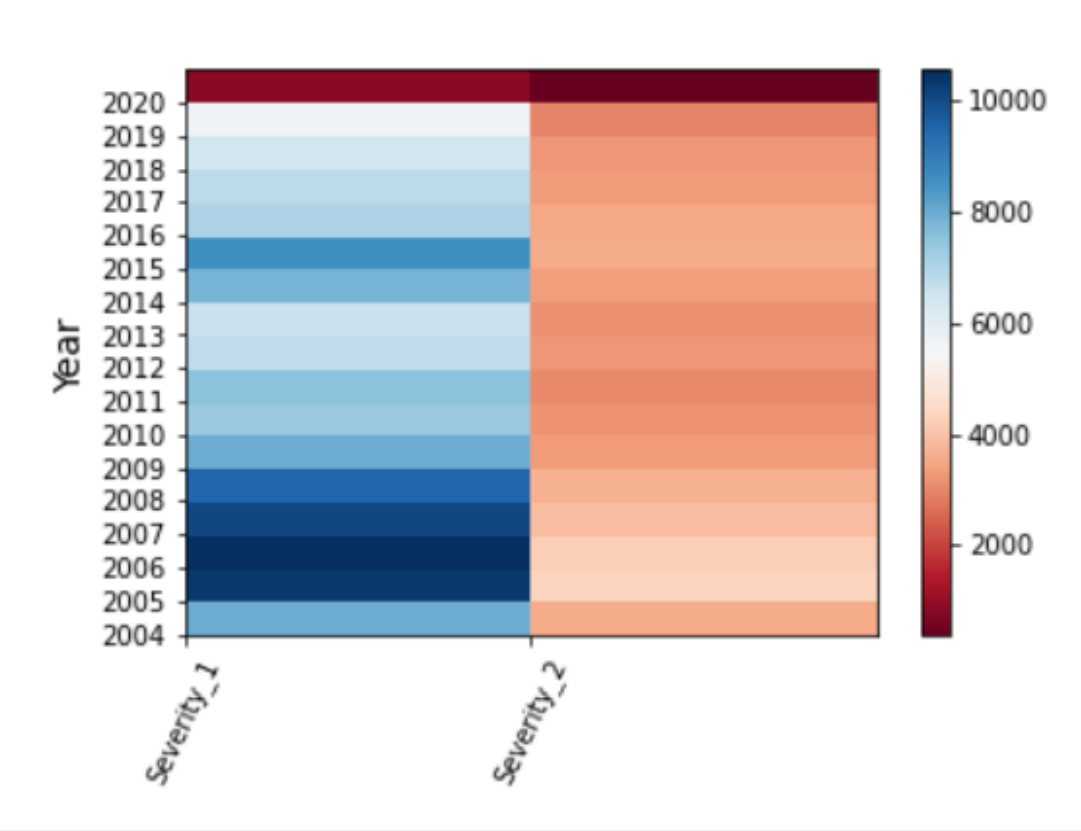






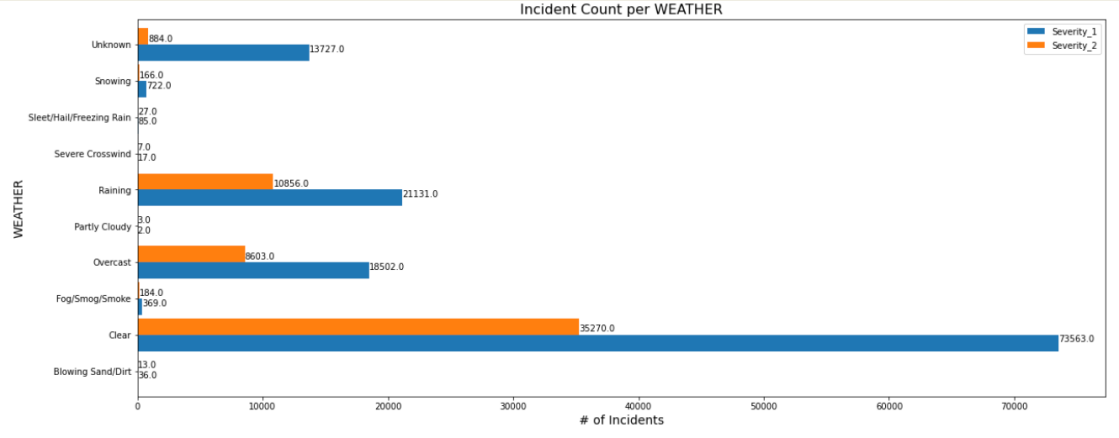
Severity of the incidents

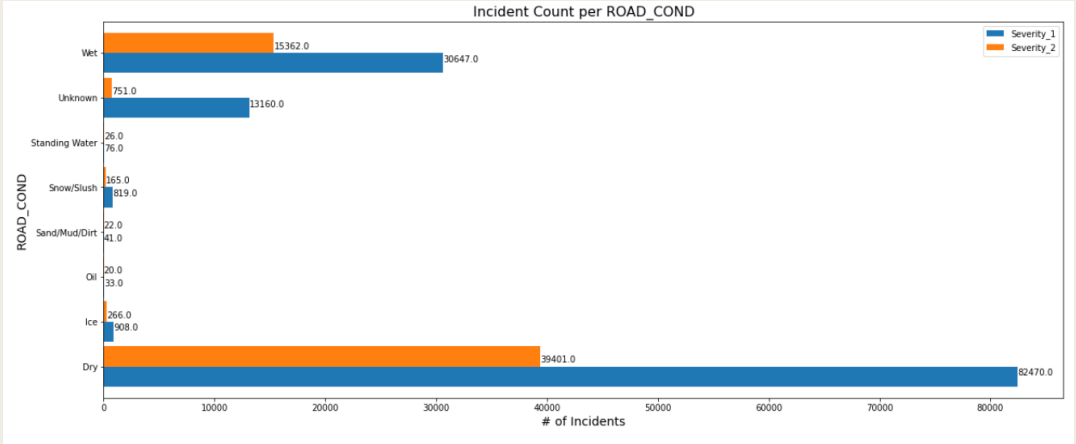
There are more incidents have less severity which involves properties damage, significantly more than those of higher severity which involves injuries

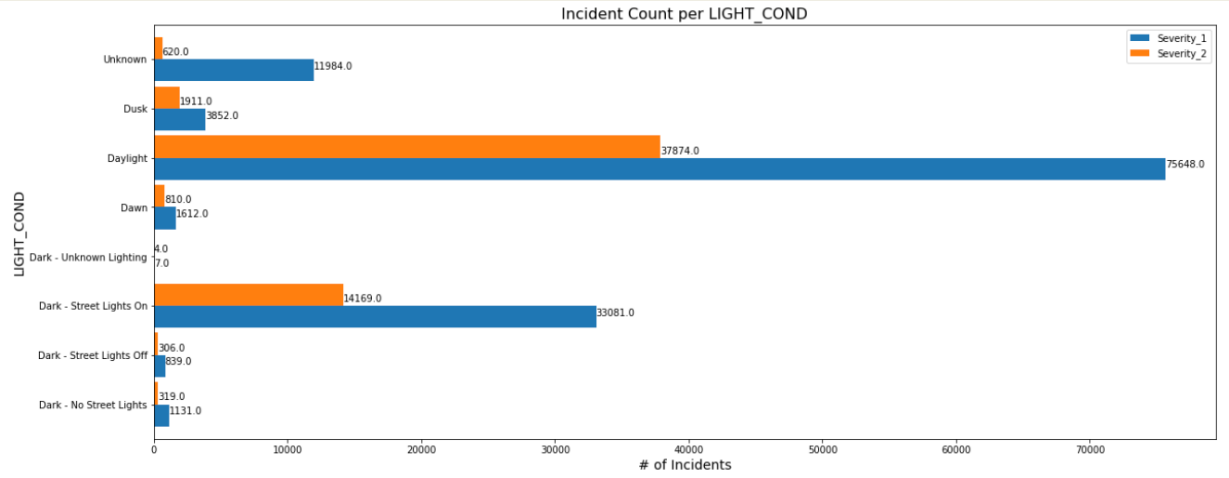


Environment where the incidents occurred

Incidents mostly occurred during board daylights, otherwise in good lighting conditions. Clear weather with decent road conditions have most incidents record.

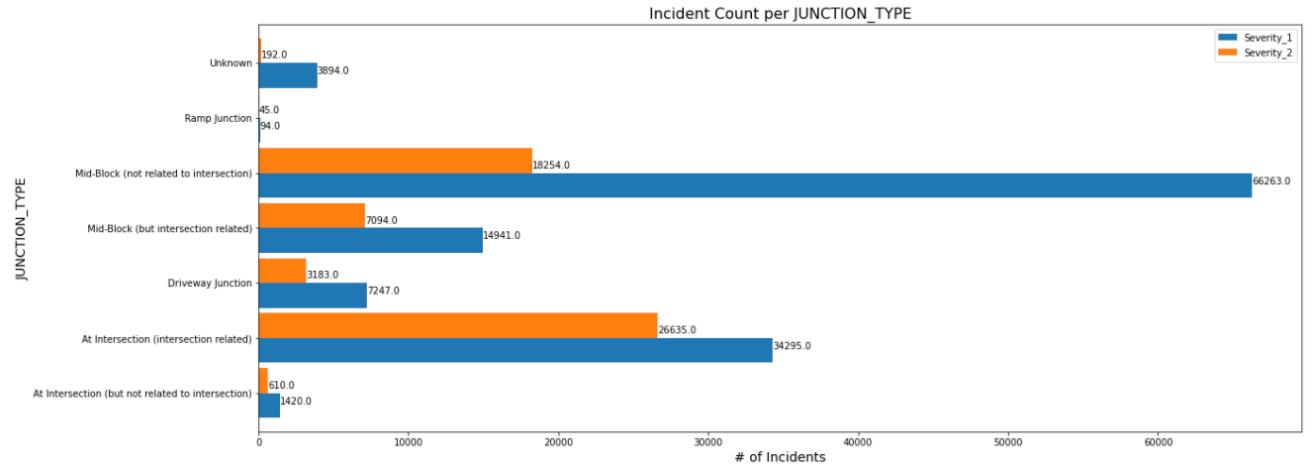


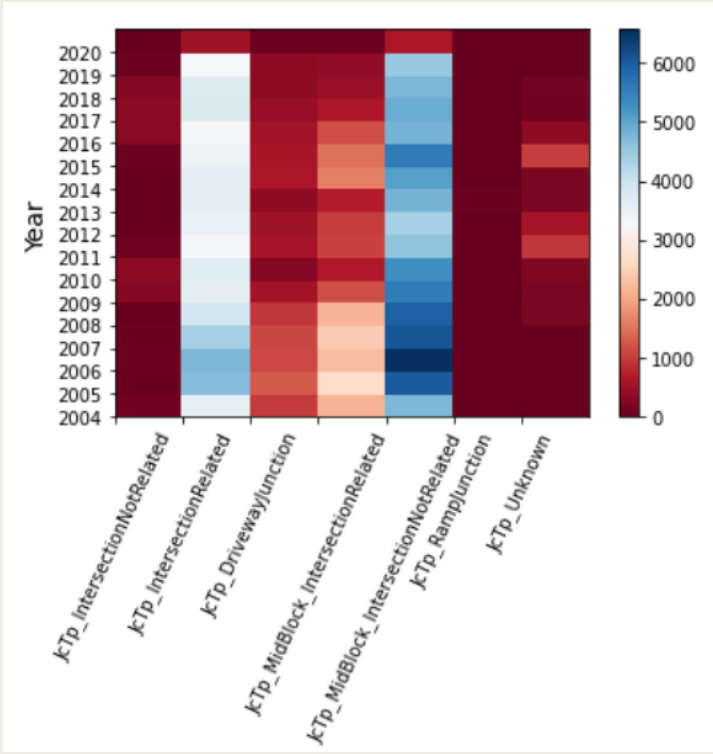




Location of incidents

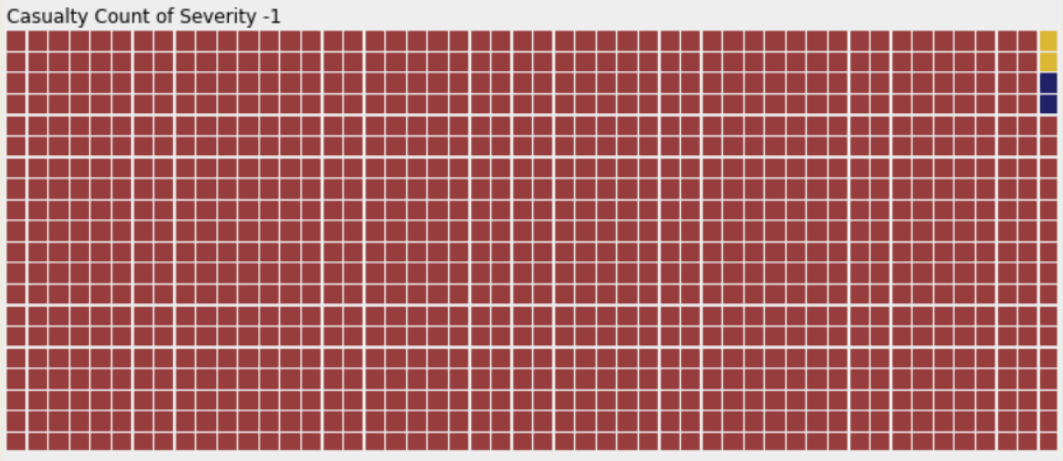
Incidents happened mostly in the Mid-Block, but NOT related to the interception

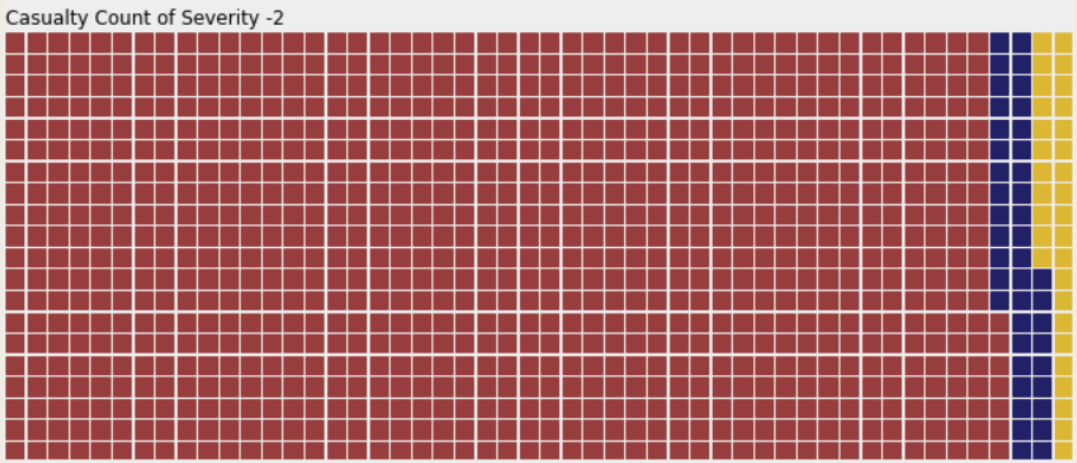




Casualties of the incidents

Persons in the vehicles are mostly involved





VEHICLEPERSON_COUNT PEDESTRAIN_COUNT CYCLYST_COUNT

SEVERITY	1	2
VEHICLEPERSON_COUNT	995.40	926.46
PEDESTRAIN_COUNT	2.37	42.11
CYCLYST_COUNT	2.23	31.43

Performance of Regression Models

	Simple Linear Reg.	Polynomial + LR	Ridge Regression	Grid CV + RR
Configuration	Order = 1	Order = 2	Order = 3 Alpha = 14300	n/a
Cross Val Evaluation	0.57234358 0.57152581 0.74871677	0.57234358 0.57152581 0.74871677	0.5898494 0.57036562 0.7483212	0.5898494 0.57036562 0.7483212
Train R^2	0.98817332	0.98865026	0.98791998	0.98814988
Test R^2	0.43791580	- 0.1621182	0.98728149	0.45117706
MSE	2190969.91	4529866.03	49575.97	2139278.34