DATA

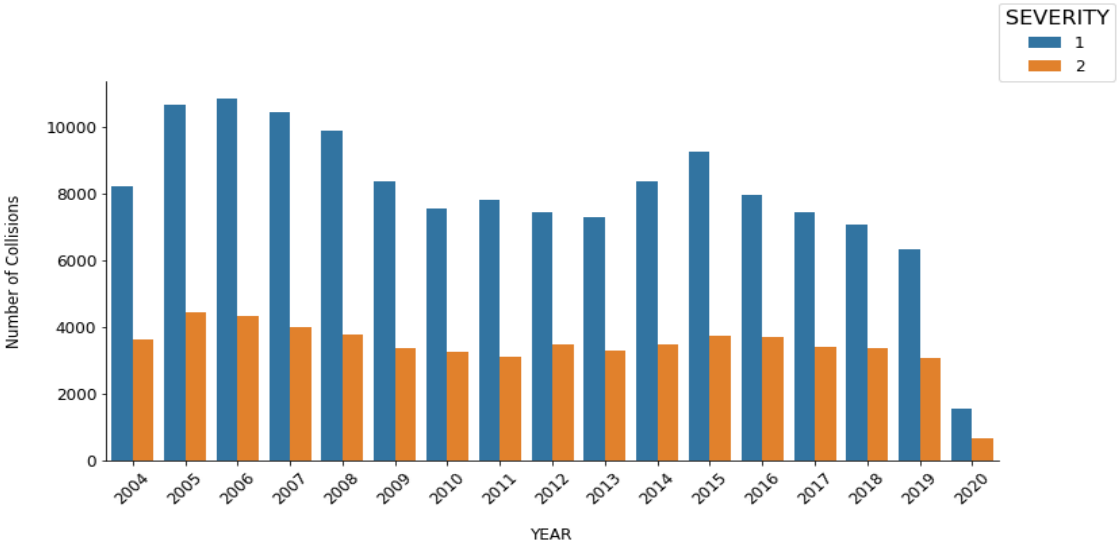
The data for this project was provided by IBM's Applied Data Science Capstone Project. This dataset includes many features to build a robust predictive machine learning algorithm. This data contains all types of collisions which occurred between the years of 2004 and 2020 in the city of Seattle. It has a total of 194673 records and 38 different attributes.

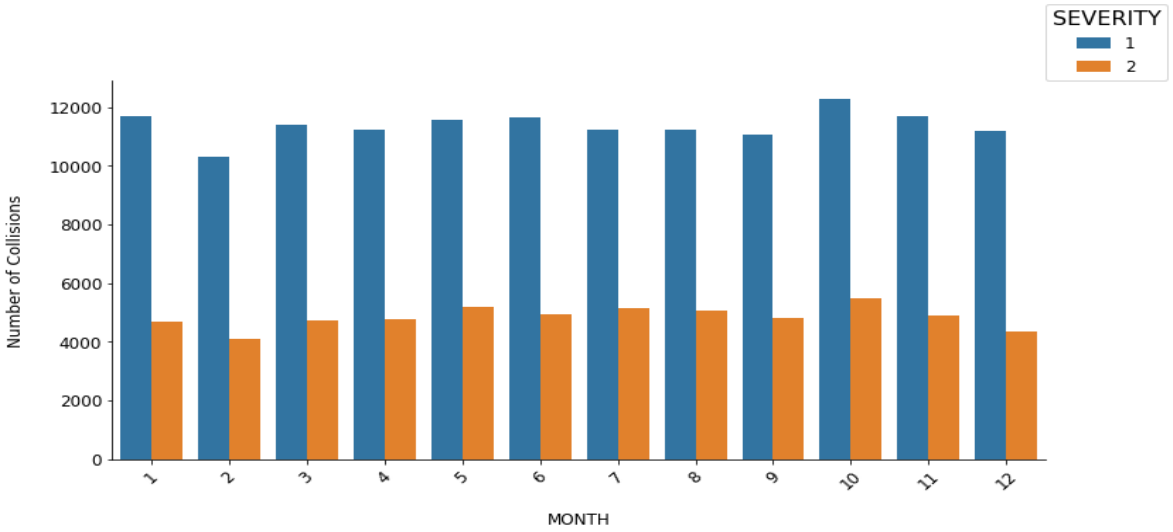
Based on this data and developed models, it is hope that insights from our results will help authorities and stakeholders develop a concrete plan to prevent severity of accident collisions to some extent and help them improve their responses from different perspectives.

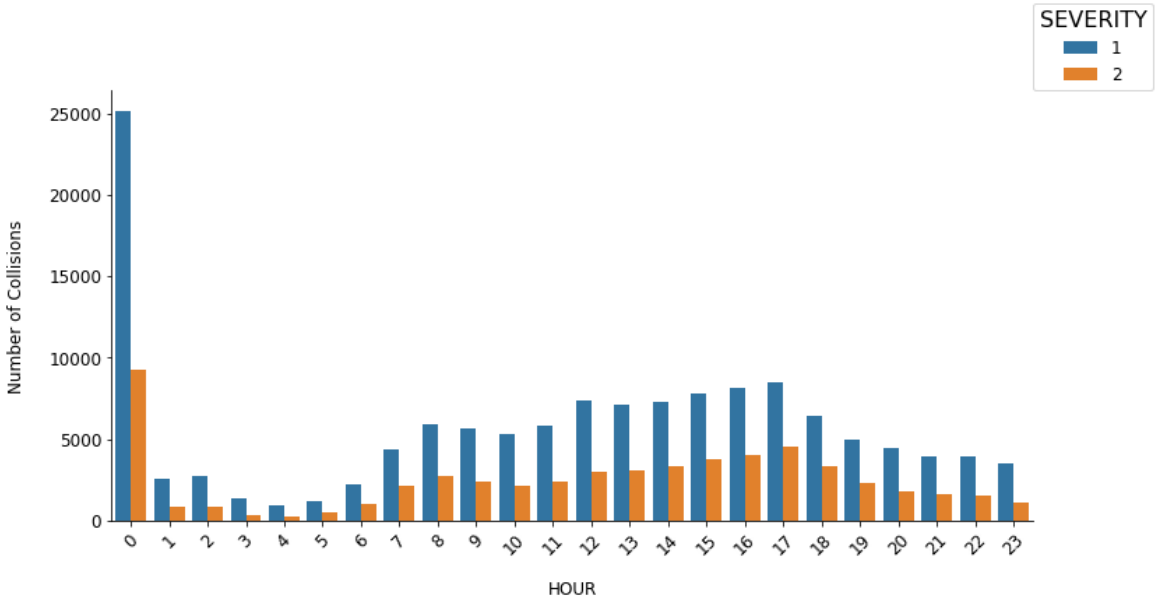
EXPLORATORY DATA ANLAYSIS

1. Severity Levels by Year & Month & Hour

Following tables illustrates the severity level of accidents by the time of year, month, and hour.

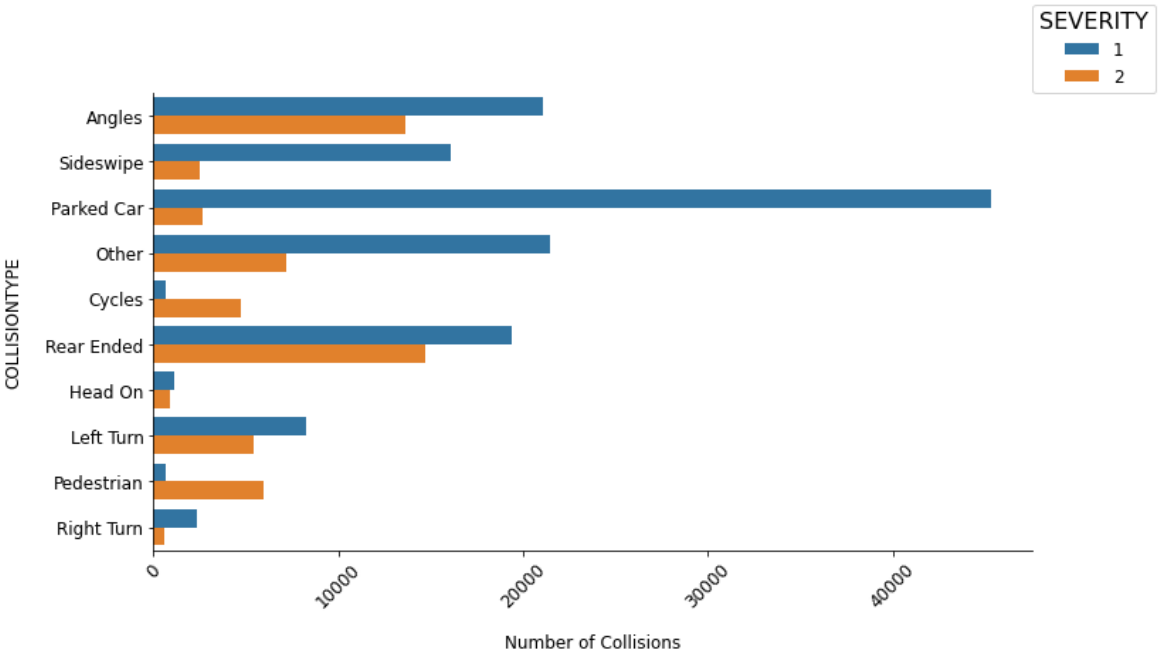




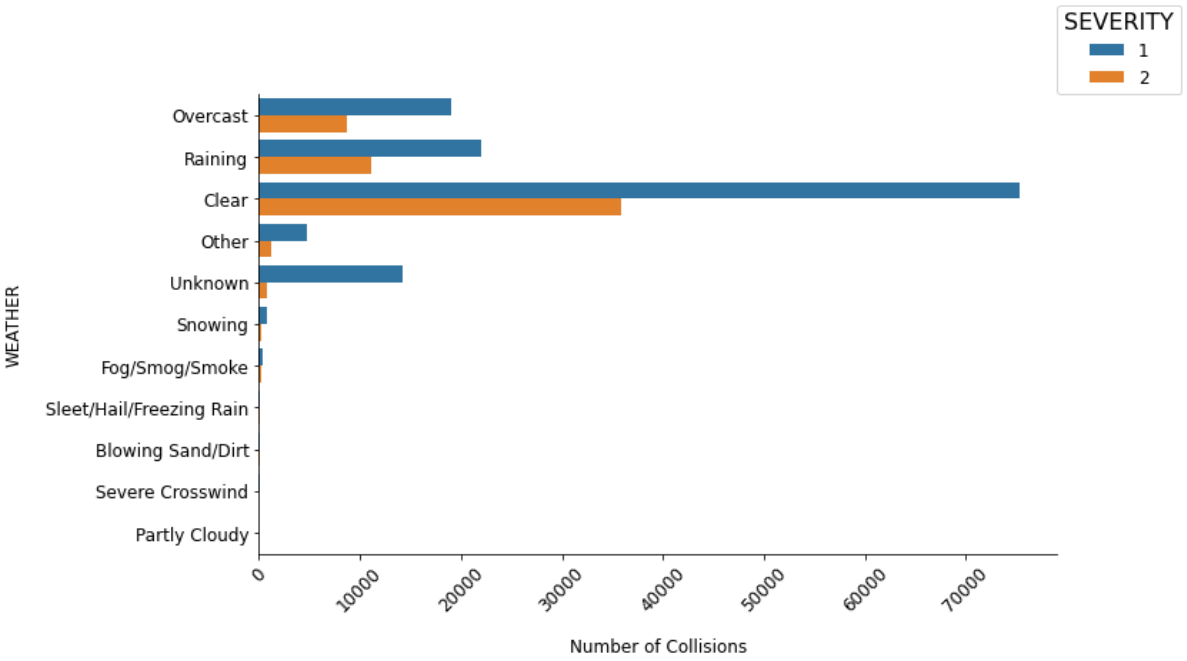


The number of accidents for both severity classes have been decreasing from the year of 2005 to 2013. Then it seems increasing to the year of 2015, and then decrease again. The month of October has slightly higher than rest of the months in terms of severity level of accidents for both types. The highest number of accidents happen at midnight.

2. Severity Levels by Collision Types

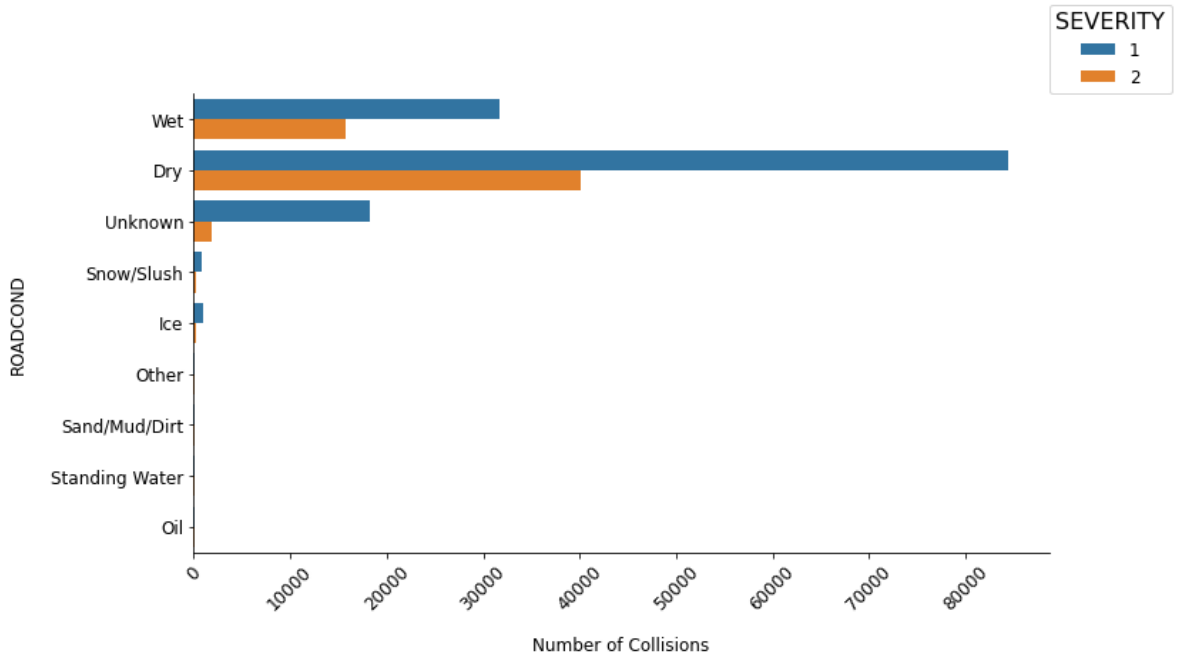


3. Severity Levels by Weather Conditions



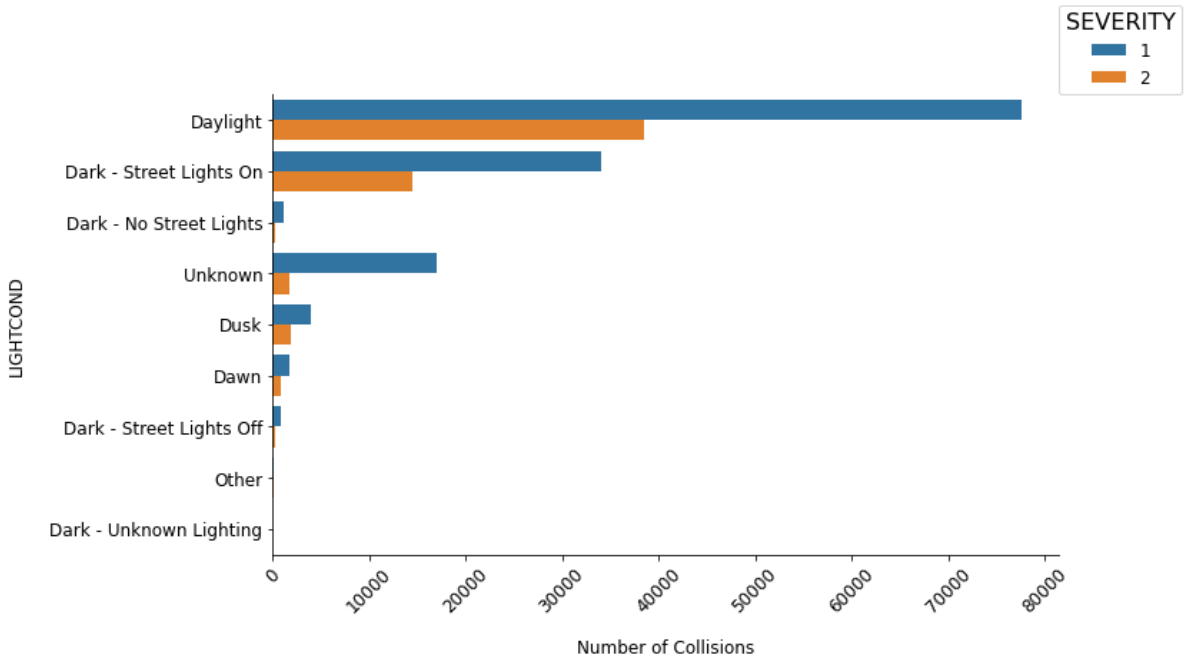
Majority of accidents occur at clear weather conditions even though it is more for severity type 1, which is injury collision. Rainy and overcast are other important weather conditions for severity levels of accidents.

4. Severity Levels by Road Conditions



It seems that accidents occur in the presence of dry and wet conditions. Snowy and Icy road conditions accounts for small fraction of accidents.

5. Severity Levels by Light Conditions



It is interesting that most accidents occur in daylight, and then dark but streetlights are on.

METHODOLOGY

After careful exploratory data analysis, some variables were selected that researcher think, drive the severity level of accidents. For this undertaking, the following variables were selected to build a robust machine learning model.

	SEVERITYCODE	WEATHER	ROADCOND	LIGHTCOND	COLLISIONTYPE	INATTENTIONIND	UNDERINFL	SPEEDING
0	2	Overcast	Wet	Daylight	Angles	NaN	N	NaN
1	1	Raining	Wet	Dark - Street Lights On	Sideswipe	NaN	0	NaN
2	1	Overcast	Dry	Daylight	Parked Car	NaN	0	NaN
3	1	Clear	Dry	Daylight	Other	NaN	N	NaN
4	2	Raining	Wet	Daylight	Angles	NaN	0	NaN

The target variable of the study is the SEVERITY CODE, which is used to measure the severity level of an accident.