IBM Capstone Project
Seattle Car Accident Severity

Jason Rutherford
September 13th, 2020

Introduction/Business Problem

Did you know that road crashes are the leading cause of death in the United States for people aged between 1 and 54? The United States is one of the busiest countries in the world in terms of road traffic with nearly 280 million vehicles in operations.

They are numerous factors that determine the severity of accidents such as irresponsible driving (speeding, distracted and under the influence of alcohol/dugs), time of day/day of week and environmental conditions (weather, season, road surface and lighting conditions).
We can gain insight and solutions from the numerous factors that affect accident severities by using data from past accidents.

We will be using machine learning to build multiple models that can predict the severity of a future accident base on the similarity of their initial conditions to those of other accidents from historical data.

Data

A comprehensive dataset of 194,673 accidents occurring from 2004 to May 2020 in the Seattle city area was obtained from the Seattle Police Department and recorded by Traffic Records and include Collisions at intersection or mid-block of a segment.

The data also has 37 columns describing the details of each accident including the weather conditions, collision type, date/time of accident and location (latitude and longitude).

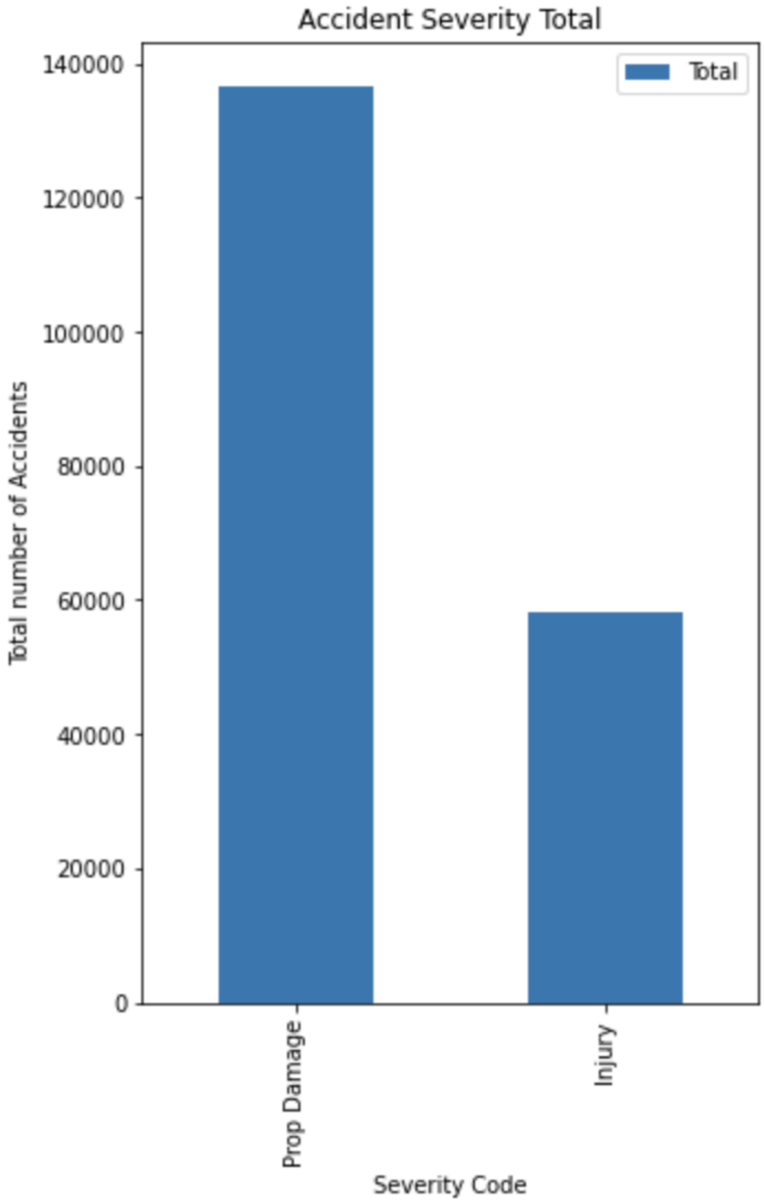
An additional document was provided with the description of each column given; the data is labeled but unbalanced.

We will be using a predictive analytic approach to determine the severity of an accident base on its attributes such as location, weather condition, speeding, light conditions and road condition. The attribute **SEVERITYCODE** (dependent variable) will be used to determine the severity of an accident.

According to the additional document, the possible values for **SEVERITYCODE** are:

- 0 - Unknown
- 1 - Prop Damage
- 2 - Injury
- 2b - Serious Injury
- 3 – Fatality

Unfortunately, the data set only provided two values for the **SEVERITYCODE** (1 & 2).



A bar graph showing the amount of accidents base on the severity

As mentioned above, the dataset has almost 40 attributes, but we want to focus only a set of attributes that has useful information being able to predict the severity of a future accident. Within the data, the following columns was selected:

SEVERITYCODE – This variable corresponds to the severity of the collision

COLLISIONTYPE – This variable determines the collision type.

UNDERINFL – This variable determines whether or not a driver involved was under the influence of drugs or alcohol.

INATTENTIONIND – This variable determines whether or not collision was due to inattention.

WEATHER – This variable determines the description of the weather conditions during the time of the collision.

ROADCOND – This variable determines the condition of the road during the collision.

LIGHTCOND – This variable determines the light conditions during the collision.

SPEEDING – This variable determines whether or not speeding was a factor in the collision.

During data analysis, it was discovered the data set had a lot of missing values for certain columns:

SPEEDING

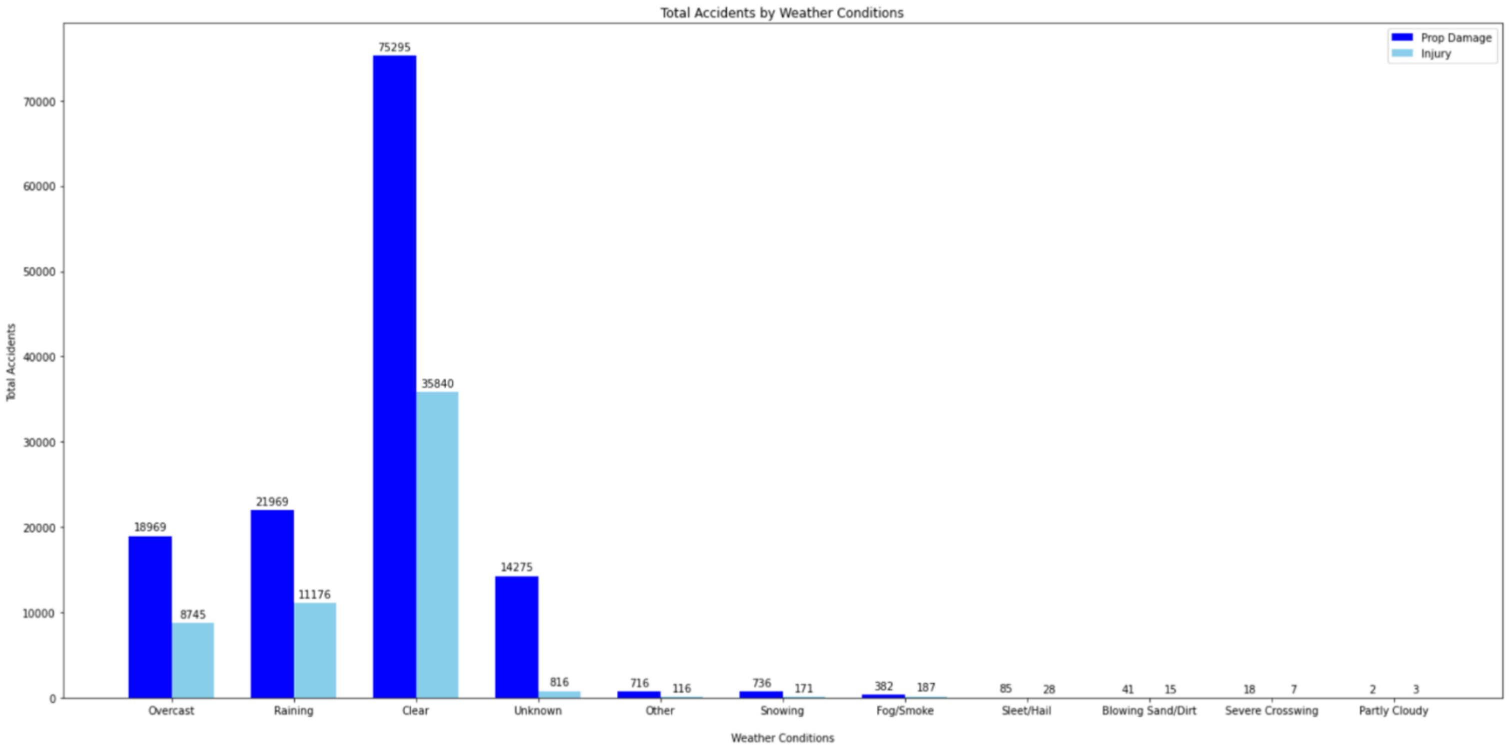
- In the SPEEDING column, 95.21% of the values were unknown.

INATTENTIONIND

- In the INATTENTIONIND column, 84.69% of the values were unknown.

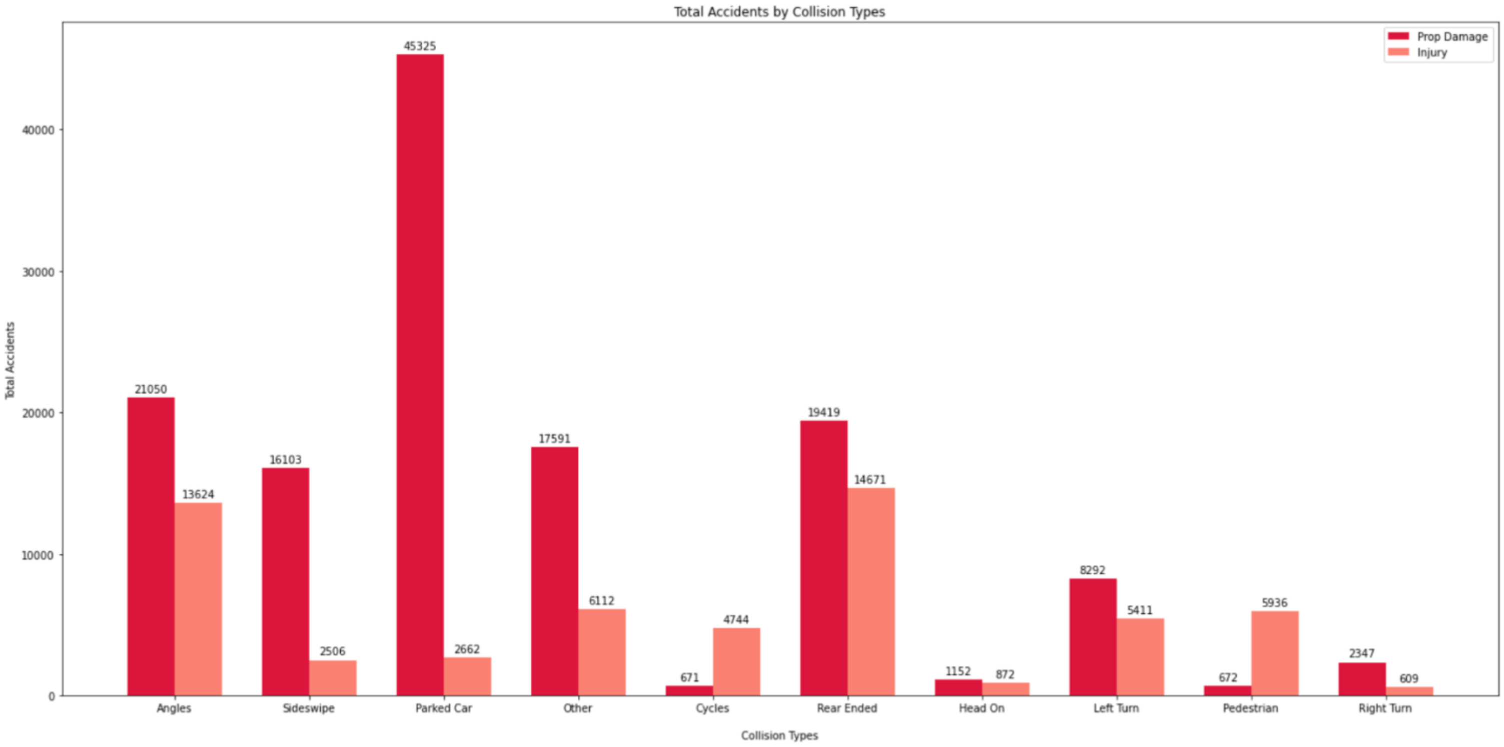
So, both columns had to be removed to further increase the accuracy of the model and the analysis of the data.

Using the severity code for each accident, we can detect the amount of accidents for each type of value in each column.

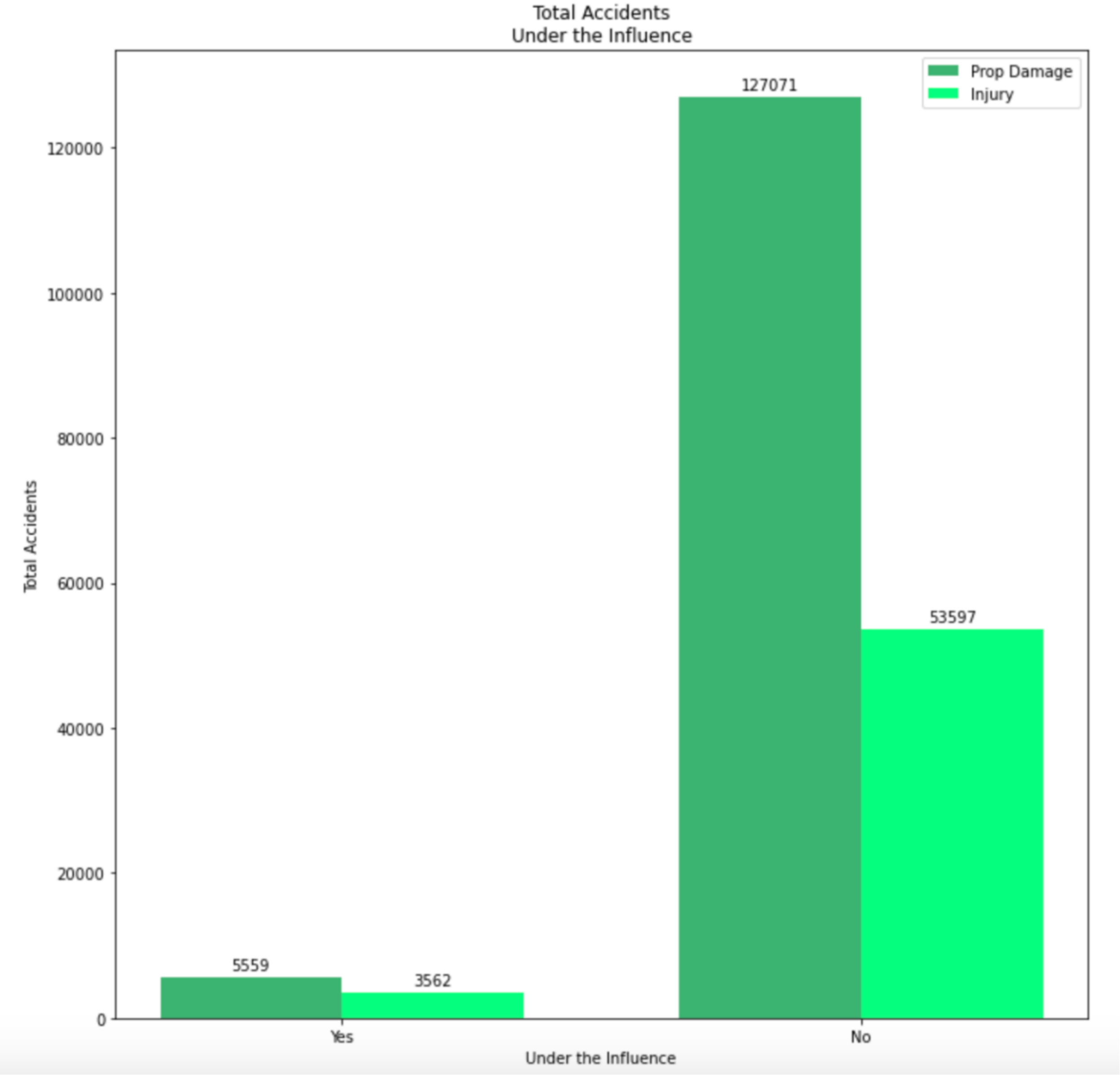


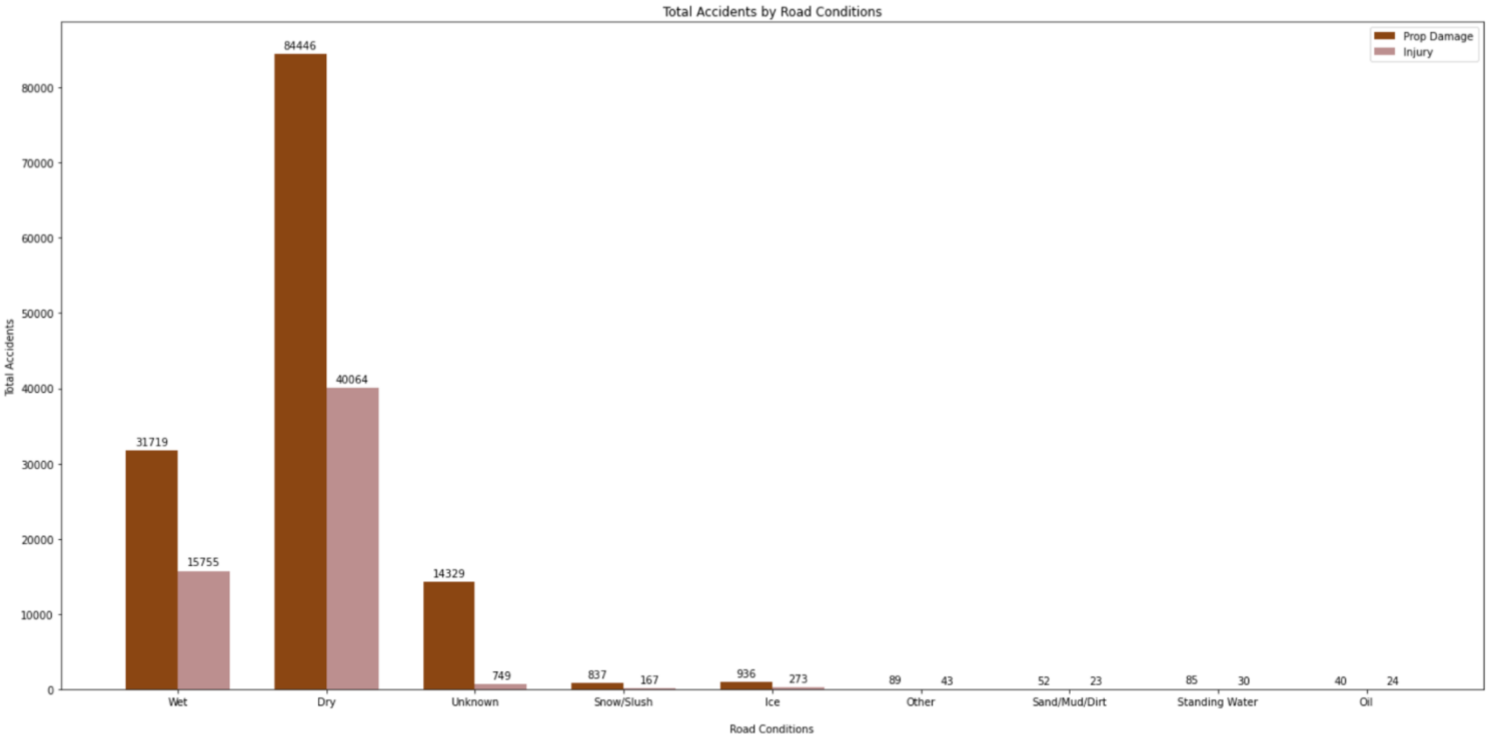
A bar graph that shows the amount of accidents for each type of weather condition base on the severity of the accident

A bar graph that shows the amount of accidents for each type of collision base on the severity of the accident



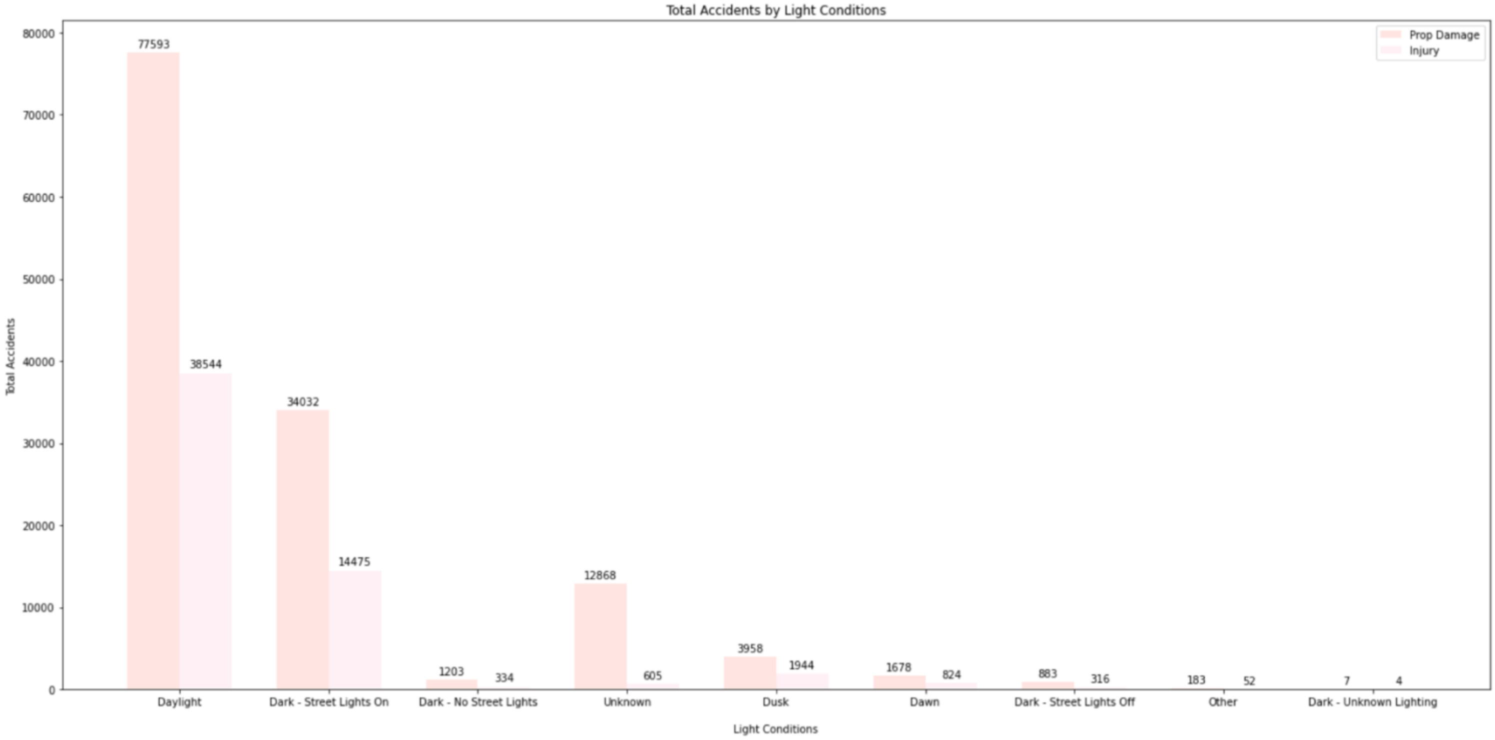
A bar graph that shows the amount of accidents whether the driver was under the influence of alcohol or drugs base on the severity of the accident





A bar graph that shows the amount of accidents for each type of road condition base on the severity of the accident

A bar graph that shows the amount of accidents for each type of light condition base on the severity of the accident


Total Accidents by Light Conditions
Prop Damage
Injury
Total Accidents
Light Conditions
Daylight
Dark - Street Lights On
Dark - No Street Lights
Unknown
Dusk
Dawn
Dark - Street Lights Off
Other
Dark - Unknown Lighting