Data Understanding:

The data selected to analyze is provided by the state of Indiana. It contains all traffic accident in 2019.

Data can be found at the following link:

<https://hub.mph.in.gov/dataset/aries-crash-data-2007-2017/resource/7081439e-33c9-45f5-9c0b-86d8a8c49252?view_id=f56f9538-0d0b-4162-bd5a-09082e412f5f>

The data contains 625,702 rows and 107 columns. The data contains a large variety of columns surrounding each accident. Examples of columns included are: Gender, Age Group, Injury Status, Seatbelt use, Nature of Injury, Local of accident, Day of the week, Light conditions, Weather conditions, conditions, Surface conditions, Type of vehicle, Type of Road, etc. Not all columns have complete data. For each column that will be used in the prediction, the data will be cleaned to avoid skewing the prediction. The Injury Status column indicates the severity of the accident and has a range of 1-7. A status of 1 indicates a fatal accident, with the injury being less severe as it increases. Blanks indicate no injury, and those will be populated with an 8 as part of the data preparation.

The following columns will be used to predict severity:

1. Light Conditions
2. Weather conditions
3. Surface conditions
4. Road Type (2-lane/multi-lane/etc)