**Statistical/Hypothetical Question**

1. Do severe accidents tend to happen when visibility is low?
2. Do traffic signals which are intended to control the speed limit in the area causing more accidents than helping reduce?
3. Does wind speed cause more accidents? Should driver be reducing car speed if wind speed is over a certain limit?

**Outcome of your EDA**

Hypothesis testing for visibility over severity came to 3.7%, which means null hypothesis is not correct and this correlation is statistically significant. As per the accidents record of 2019 accidents across USA, hypothesis testing suggests that low visibility may cause severe accidents. Looking at the traffic signal data and categorial hypothesis test (like one in coin test for head and tails) it looks like traffic signal presence does not affect the probability of accidents. Another study should be done to include more parameters (for example, traffic signal and a speed bump presence decrease the probability of accidents), where multiple exploratory variables are used to do the regression.

**What do you feel was missed during the analysis?**

This data set is for entire United states. Regions and weather in all these areas are significantly different. This analysis treats data from all the states in similar fashion. For example, Nebraska weather is much harsher than changes more through out the year as compared to California. A better result and analysis could be established if this study was done by limiting to states of similar weather, or multiple analysis for different regions.

**Were there any variables you felt could have helped in the analysis?**

This analysis completely ignores the regions data. Data set has rich information about the accident location including city state zip and the GPS coordinates. I feel utilizing those variables could have given clearer results. Also, weather condition is a categorical variable with too many categories. A data preparation step could have helped categorize those in more defined weather conditions to help analyze the correlation between weather condition and accident severity.

**What challenges did you face, what did you not fully understand?**

One big challenge was the size of the data set. It has more than 2.5 million records. Most of the times the hypothesis testing was hanging or taking a very long time to provide any results.