1. **Introduction**

Car accidents in the United States are happening in high volumes annually and many drivers don’t have many insights to know how unfavorable some road conditions are. The result of these accidents could be catastrophic in nature, due to: weather, speed, and car safety features. While being a vigilant driver and purchasing cars with solid safety features can help, actionable driving insights could help save a life. For instance, do drivers know how the number of car accidents in their state compares to the rest of the country? Do they know how driving at certain times of the day or in inclement weather could impact their safety? I’m not convinced this data is well known by the average driver, so I aim to use my data science project to inform them.

1. **Data**

The dataset I will use is titled “A Countrywide Traffic Accident Dataset”, created by: Sobhan Moosavi, Mohammad Hossein Samavatian, Srinivasan Parthasarathy, and Rajiv Ramnath. The dataset is a compilation of car accident data in the contiguous United States, spanning from February 2016 to December 2019. This data was scraped from Foursquare and Mapquest and holds over two million car accidents. I will clean the data and examine relationships between different columns of the dataset such as: daypart, visibility, state, accident severity, and total accidents. Through examining these relationships through proper data analysis and visualization, I will solve my problem by delivering some key insights for drivers. They will know if driving after sun down is more dangerous, which states are the most dangerous to drive in, and the role visibility plays in accident frequency and severity.

1. **Methodology**

In order to pass on insights to potential drivers. I have to develop them from proper data cleaning and analysis. To begin, I used Pandas to read the csv file into the data frame and use the isnull().sum() function to isolate values that were missing or uninterpretable in the data set. Then, I dropped the columns with missing values so I could utilize rows that have values in every column.

After cleaning the data, I analyzed and modeled car accident data in the contiguous United States using a choropleth map. To provide more insight on car accident data I created a bar chart to display the 10 states with the highest accident totals. Those visualizations demonstrated the prevalence of car accidents and are intended to build a sense of urgency in readers to learn information to avoid this outcome. Then, I use the columns: “Visibility(mi)”, “Sunrise\_Sunset”, and “Severity” to show correlations between these columns and the “Accidents” column the scatter plots. Afterwards, I evaluated the information from the data I modeled.

1. **Results**

The choropleth map showed California lead the United States in car accidents by a wide margin with 640,762 collisons . After California: Texas, Florida, South Carolina, North Carolina, New York, Pennsylvania, Michigan, Illinois, and Georgia completed the list of the highest ten accident totals in the country. After those ten states, there was a significant drop off among accident totals across states (~81,000 or lower)..

When I analyzed the columns,I started with comparing day and night driving (Sunrise\_Sunset column) and tried to see which time of day held the highest accident totals. Three times as many accidents occurred during the day (2,127,689) compared to at night (753,803). This can be assumed due to the fact that there are usually more drivers on the road during this time compared to night. I compared Severity and Visibility(mi) and noticed that most severe accidents occur when Visibility is at 80 miles or lower for drivers. I modeled the visibility data for day and night drivers and night driving visibility topped out at 80 miles, therefore visibility is indeed reduced at night. The data also showed that accidents due to reduced visibility was very prevalent in the day time as well which could imply exposure to inclement weather, which could create a very significant threat to drivers given how many more drivers are on the road. Given our analysis, drivers do have some takeaways that can protect them while on the road: Daytime driving during inclement weather is a scenario that holds the highest risk of accidents for drivers.

* For more severe accidents (rating of 3 or 4), there is strong correlation between reduced visibility and accident severity.
* Daytime driving during inclement weather is a scenario that holds the highest risk of accidents for drivers.
* Severe accidents aren't occurring more frequently at night than during the day, reduced visibility does suggest this, but it can’t be confirmed without knowing the proportion of day drivers and night drivers and the total accidents for both dayparts.
* Aside from California, states with large populations don't necessarily have a large number of accidents. Part of this could be due to the accessibility of public transportation in some states.

1. **Discussion**

Some limitations of the data for this study was a lack of information of total cars on the road, and an organized weather condition column in place of one for temperature. For future research, it would be helpful to know the total cars on the road during the day as well as at night. This task would be arduous in nature, but could produce a rate or car accidents to give more insight as to the proportion of outcomes for day and night driving. Adding a column for driving conditions and limiting it to: clear, rain, fog, snow, or severe (natural disaster). Another helpful column would be one titled “weather conditions” could confirm danger of car accidents during inclement weather for both the night and day accidents. Aside from those topics, this was a fairly interesting data set.

1. **Conclusion**

I completed this study to better inform people who drive everyday about situations that could be more dangerous to them. By mapping accident instances in the contiguous United States and listing the top 10 states with the highest accident totals, I offered a visual for drivers to see what states put them at the highest risk of an accident. Travelers also know that driving during the day when facing inclement weather is a high risk scenario because: the inclement weather negates the added visibility drivers gain by not driving at night, and there will be many more drivers on the road because more people drive during the day. These insights don’t mean a perfectly clean driving record for one’s insurance company, but they can help avoid high risk situations and encourage driving when less people are on the road. It was a pleasure to work on this project and to give people more information on safe driving.