Analysis of US Accidents

Team Members

- Matthew Scott
- Paul Schneider
- Daniel Simonson
- Kevin Qian

Description

- We intend to explore and answer questions about how weather, time of day, precipitation, geolocation, and other external factors lead to an increase of accidents in the US. Our particular interest is finding a correlation between why an accident might be more likely to occur near a certain location vs another, during extreme weather instances vs not, or during daytime rush hour vs nighttime low visibility

Prior Work

- We have found the dataset we plan to use for the project and investigated the data.
- Discussed options for analysis
- Discussed potential questions to ask and explore with regards to our data

Dataset

- List of datasets to use: US Accidents (2016 - 2023)
- Where found: https://www.kaggle.com/datasets/sobhanmoosavi/us-accidents/data
  - Data Sourced from local traffic APIs
- Uploading primary dataset to shared Google Drive.

Proposed Work

- Data cleaning:
  - Review and scrub data for values that are null, negative, or otherwise unnecessary data.
- Questions:
  - Create and focus on 2-4 questions to guide our analysis.
- Data integration:
  - Avoid redundant data, columns, etc. through correlation analysis.
- Data Reduction:
  - If specific data is not integral to answering our questions then remove

List of Tools to Use

- Python
  - Pandas
  - Numpy
  - MatplotLib
- Excel
- Github
- Zoom
- Text Message
- Google Drive
- Tableau or PowerBI

Evaluation

- Present a collection of visualizations that show our correlation analysis.
- Draw conclusions from our results to see if they align with our previous predictions.