**IFT 598 Data Visualization & Reporting for IT**

Hima sree Chalasani

Nikhil Jain Sanghami Padmaprabha

Venkata Bala Nagendra Srirangam

Project - Phase I Planning

Asmaa Elbadrawy

April 09, 2023

**Section 1: Dataset Description**

The US Accidents dataset consists of data on over 2.8 million traffic accidents that occurred in the United States between February 2016 and December 2021. The data was collected from various sources, including traffic cameras, traffic sensors, and other public safety agencies. The dataset includes information on a range of variables, including the severity of the accident, the time and location of the accident, the weather conditions at the time of the accident, and the number of vehicles involved. The dataset covers accidents from all 49 states in the US.

The dataset is provided in CSV format and consists of 47 columns, including information on the accident's severity, location, time, weather, road conditions, and more. However, it's important to note that the dataset is not a complete representation of all traffic accidents that occurred during the time period, as it only includes accidents that were reported through the sources used to collect the data.

Attributes in Dataset

* ID: This is a unique identifier for each accident record.
* Source: This is the source of the accident report (e.g., mapquest, Bing, etc.).
* TMC: The Traffic Message Channel (TMC) code is a unique identifier assigned to road segments for traffic information exchange.
* Severity: This is a measure of the severity of the accident, ranging from 1 (least severe) to 4 (most severe).
* Start\_Time: The date and time at which the accident started.
* End\_Time: The date and time at which the accident ended.
* Start\_Lat: The latitude of the starting point of the accident.
* Start\_Lng: The longitude of the starting point of the accident.
* End\_Lat: The latitude of the ending point of the accident.
* End\_Lng: The longitude of the ending point of the accident.
* Distance(mi): The length of the road extent affected by the accident in miles.
* Description: A textual description of the accident.
* Number: The street number of the address where the accident occurred.
* Street: The street name of the address where the accident occurred.
* Side: The relative side of the street (Right/Left) in the direction of the address where the accident occurred.
* City: The city where the accident occurred.
* County: The county where the accident occurred.
* State: The state where the accident occurred.
* Zipcode: The postal code where the accident occurred.
* Country: The country where the accident occurred.
* Timezone: The timezone in which the accident occurred.
* Airport\_Code: The airport code nearest to the location where the accident occurred.
* Weather\_Timestamp: The time at which the weather conditions were observed.
* Temperature(F): The temperature in Fahrenheit.
* Wind\_Chill(F): The wind chill temperature in Fahrenheit.
* Humidity(%): The relative humidity.
* Pressure(in): The air pressure in inches.
* Visibility(mi): The visibility in miles.
* Wind\_Direction: The direction of the wind.
* Wind\_Speed(mph): The wind speed in miles per hour.
* Precipitation(in): The amount of precipitation in inches.
* Weather\_Condition: The current weather condition.
* Amenity: A boolean indicating whether there was an amenity (e.g., a restaurant, gas station) near the accident location.
* Bump: A boolean indicating whether there was a speed bump near the accident location.
* Crossing: A boolean indicating whether there was a crossing (e.g., a pedestrian crossing, a railroad crossing) near the accident location.
* Give\_Way: A boolean indicating whether there was a give way sign near the accident location.
* Junction: A boolean indicating whether there was a junction (e.g., an intersection, a roundabout) near the accident location.
* No\_Exit: A boolean indicating whether there was a no exit sign near the accident location.
* Railway: A boolean indicating whether there was a railway near the accident location.
* Roundabout: A boolean indicating whether there was a roundabout near the accident location.
* Station: A boolean indicating whether there was a station (e.g., a bus or train station) near the accident location.
* Stop: A boolean indicating whether there was a stop sign near the accident location.
* Traffic\_Calming: A boolean indicating whether there was a traffic calming measure (e.g., a speed bump, a roundabout) near the accident location.
* Traffic\_Signal: A boolean indicating whether there was a traffic signal (e.g., a traffic light) near the accident location.
* Turning\_Loop: A boolean indicating whether there was a turning loop near the accident location.
* Sunrise\_Sunset: Indicates whether the accident occurred before or after the sunrise or sunset.
* Civil\_Twilight: Indicates whether the accident occurred during civil twilight.
* Nautical\_Twilight: Indicates whether the accident occurred during nautical twilight. Nautical twilight is the period after civil twilight.
* Astronomical\_Twilight: Indicates whether the accident occurred during astronomical twilight.

**Section 2: Prospective Dashboard Users**

The potential users of the dashboard could vary depending on the specific focus and purpose of the dashboard. However, some potential users of the US Accidents dashboard could be:

* **Transportation companies:** They can use the dashboard to analyze the frequency and severity of accidents on different routes, and make data-driven decisions to improve the safety of their vehicles and drivers.
* **Government agencies:** They can use the dashboard to identify high-risk areas and take necessary measures to improve the road infrastructure, traffic management, and emergency response systems.
* **Insurance companies:** They can use the dashboard to evaluate the risk of insuring vehicles and drivers in different locations, and set premiums accordingly.
* **Researchers and analysts:** They can use the dashboard to study the patterns and trends of accidents over time and across different regions, and generate insights to inform public policy and transportation planning.
* **General public:** They can use the dashboard to stay informed about the latest traffic and weather conditions, and plan their routes accordingly to avoid potential accidents and delays.

**Section 3: List of User Requirements & Potential Questions**

1. What is the distribution of accidents across different states in the US? Which states have the highest and lowest numbers of accidents?
2. Can we identify any patterns or trends in the types of accidents that occur in different regions of the US? For example, are certain types of accidents more common in certain regions?
3. What is the distribution of accidents across different times of day or days of the week? Are there any patterns in the times or days when accidents are most likely to occur?
4. How do weather conditions and other factors (such as road surface conditions or visibility) impact the likelihood of accidents occurring in different areas or at different times of day?
5. Are there any patterns or trends in the severity of accidents that occur in different regions or at different times of day?
6. How do traffic volumes and congestion impact the likelihood of accidents occurring in different areas or at different times of day?
7. Can we use clustering algorithms to identify different groups or clusters of accidents based on their geographic features, such as proximity to major highways or urban areas?
8. How does the frequency and severity of accidents vary across different types of roads, such as highways vs. local streets?
9. Are there any patterns or trends in the demographics of drivers involved in accidents, such as age, gender, or experience level?
10. Can we use heat maps or other visualization techniques to identify areas of high accident frequency or severity, and how does this vary across different times of day or days of the week?
11. Are there any relationships between the severity of accidents and the distance from major urban areas or highways?
12. How do the characteristics of the vehicle involved (e.g. make, model, age, etc.) impact the likelihood of accidents occurring and their severity?

**References**

**Mural -** <https://app.mural.co/t/greeshmasworkspace0156/m/greeshmasworkspace0156/1680555917958/0960ed43596c4be3fb8a716bdd2332f73c5e5a3d?sender=ufd0807b59c2de6a557262362>

**Dataset source –** [https://www.kaggle.com/datasets/sobhanmoosavi/us-accidents](Dataset%20Link)

Moosavi, Sobhan, Mohammad Hossein Samavatian, Srinivasan Parthasarathy, and Rajiv Ramnath. *“A Countrywide Traffic Accident Dataset.”*, arXiv preprint arXiv:1906.05409 (2019).

Moosavi, Sobhan, Mohammad Hossein Samavatian, Srinivasan Parthasarathy, Radu Teodorescu, and Rajiv Ramnath. *“Accident Risk Prediction based on Heterogeneous Sparse Data: New Dataset and Insights.”* In proceedings of the 27th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, ACM, 2019