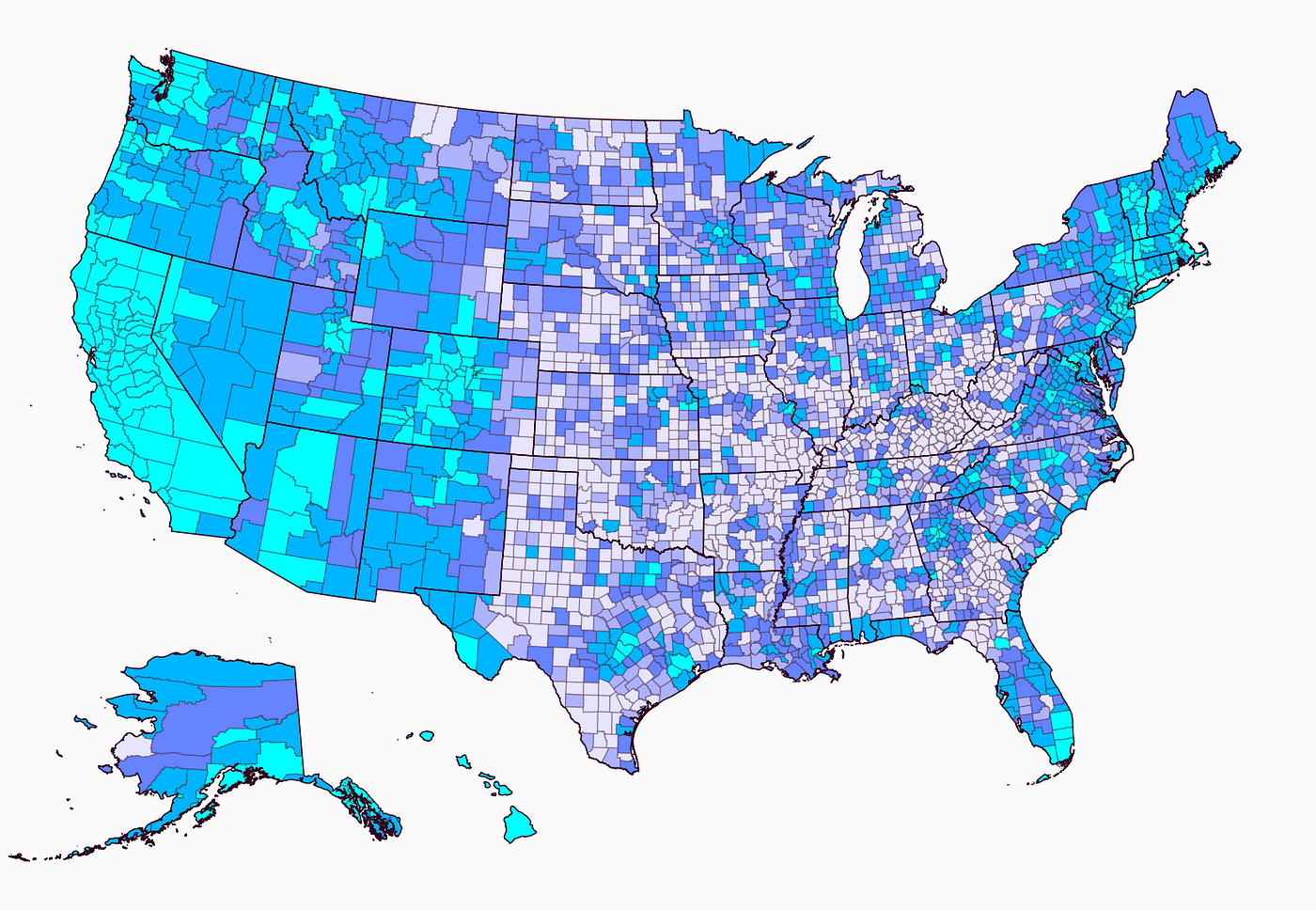
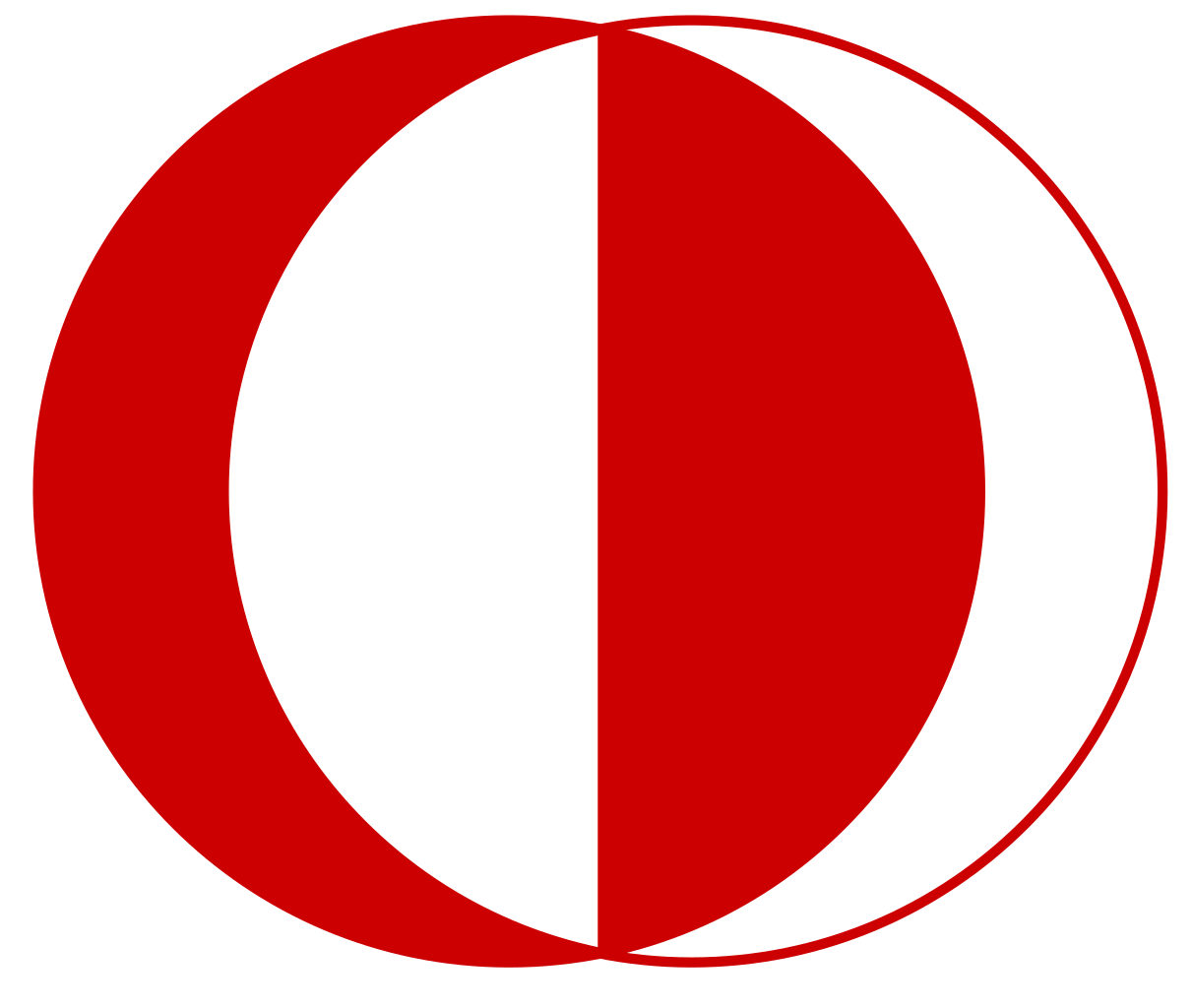
BEHIND THE WHEEL

CENG790 Big Data Analytics  
2023-2024 Spring  
Term Project Final Report



**Berk GÜLER - 2310092  
Mustafa Barış EMEKTAR - 2304533**



Middle East Technical University

Department of Computer Engineering

# INTRODUCTION

According to information from the National Highway Traffic Safety Administration (NHTSA), there are 30,000 to 40,000 deaths in traffic accidents in the United States in a year. The analysis of traffic accident data in the United States presents an opportunity to gain valuable insights into the factors contributing to road safety and accident occurrence. In this project, we delve into a comprehensive dataset encompassing various variables related to traffic accidents across the nation.

We now have the tools to identify important patterns and trends from this enormous information thanks to the development of big data analytics. Through the utilization of sophisticated analytical methods, our objective is to detect underlying trends, identify risk variables, and create models for prediction that will help in improving traffic management systems and devising focused safety measures.

We utilized methodologies such as exploratory data analysis and machine learning algorithms. Our aim is to simplify the understanding of road safety, facilitating informed decision-making and ultimately contributing to a safer transportation environment for all.

As we navigate through the dataset, our objective is to highlight the potential of big data analytics in reshaping the future of road safety initiatives. By unraveling hidden insights within the data, we endeavor to pave the way towards a transportation ecosystem that is safer and more resilient.

# DATASET

For this project, we used the [US Accidents dataset from Kaggle](https://www.kaggle.com/datasets/sobhanmoosavi/us-accidents/data). It covers accidents that occurred in the United States from 2016 to 2023. Here is why this dataset is appropriate to use:

* **Time Scope and Size:** The dataset covers 7 years of accidents that occurred in the 49 states of the USA. It consists of 7.7 million accident records with 3.06GB of data.
* **Rich Features:** The dataset provides a rich source of data for our analysis. It contains detailed information about each accident, including features such as location, severity, weather conditions, time of day, and road characteristics. It contains 46 columns (different features). These (possibly) relevant features can help us understand accident patterns and severity.
* **Severity Labels:** The dataset includes labels for accident severity (e.g., fatal, minor), which is crucial for our analysis. Without these labels, we would treat all accidents equally, missing critical insights into the impact of accidents.

| **Column** | **Explanation** | **Column** | **Explanation** |
| --- | --- | --- | --- |
| ID | *This is a unique identifier of the accident record.* | grid\_3x3Visibility(mi) | *Shows visibility (in miles).* |
| text\_formatSource | *Source of raw accident data* | text\_formatWind\_Direction | *Shows wind direction.* |
| grid\_3x3Severity | *Shows the severity of the accident, a number between 1 and 4, where 1 indicates the least impact on traffic (i.e., short delay as a result of the accident) and 4 indicates a significant impact on traffic (i.e., long delay).* | grid\_3x3Wind\_Speed(mph) | *Shows wind speed (in miles per hour).* |
| calendar\_todayStart\_Time | *Shows start time of the accident in local time zone.* | grid\_3x3Precipitation(in) | *Shows precipitation amount in inches, if there is any.* |
| calendar\_todayEnd\_Time | *Shows end time of the accident in local time zone. End time here refers to when the impact of accident on traffic flow was dismissed.* | text\_formatWeather\_Condition | *Shows the weather condition (rain, snow, thunderstorm, fog, etc.)* |
| grid\_3x3Start\_Lat | *Shows latitude in GPS coordinate of the start point.* | checkAmenity | *A POI annotation which indicates presence of amenity in a nearby location.* |
| grid\_3x3Start\_Lng | *Shows longitude in GPS coordinate of the start point.* | checkBump | *A POI annotation which indicates presence of speed bump or hump in a nearby location.* |
| vpn\_keyEnd\_Lat | *Shows latitude in GPS coordinate of the end point.* | checkCrossing | *A POI annotation which indicates presence of crossing in a nearby location.* |
| vpn\_keyEnd\_Lng | *Shows longitude in GPS coordinate of the end point.* | checkGive\_Way | *A POI annotation which indicates presence of give\_way in a nearby location.* |
| grid\_3x3Distance(mi) | *The length of the road extent affected by the accident in miles.* | checkJunction | *A POI annotation which indicates presence of junction in a nearby location.* |
| text\_formatDescription | *Shows a human provided description of the accident.* | checkNo\_Exit | *A POI annotation which indicates presence of no\_exit in a nearby location.* |
| text\_formatStreet | *Shows the street name in address field.* | checkRailway | *A POI annotation which indicates presence of railway in a nearby location.* |
| text\_formatCity | *Shows the city in address field.* | checkRoundabout | *A POI annotation which indicates presence of roundabout in a nearby location.* |
| text\_formatCounty | *Shows the county in address field.* | checkStation | *A POI annotation which indicates presence of station in a nearby location.* |
| text\_formatState | *Shows the state in address field.* | checkStop | *A POI annotation which indicates presence of stop in a nearby location.* |
| text\_formatZipcode | *Shows the zipcode in address field.* | checkTraffic\_Calming | *A POI annotation which indicates presence of traffic\_calming in a nearby location.* |
| flagCountry | *Shows the country in address field.* | checkTraffic\_Signal | *A POI annotation which indicates presence of traffic\_signal in a nearby location.* |
| text\_formatTimezone | *Shows timezone based on the location of the accident (eastern, central, etc.).* | checkTurning\_Loop | *A POI annotation which indicates presence of turning\_loop in a nearby location.* |
| text\_formatAirport\_Code | *Denotes an airport-based weather station which is the closest one to location of the accident.* | text\_formatSunrise\_Sunset | *Shows the period of day (i.e. day or night) based on sunrise/sunset.* |
| calendar\_todayWeather\_Timestamp | *Shows the time-stamp of weather observation record (in local time).* | text\_formatCivil\_Twilight | *Shows the period of day (i.e. day or night) based on civil twilight.* |
| grid\_3x3Temperature(F) | *Shows the temperature (in Fahrenheit).* | text\_formatNautical\_Twilight | *Shows the period of day (i.e. day or night) based on nautical twilight.* |
| grid\_3x3Wind\_Chill(F) | *Shows the wind chill (in Fahrenheit).* | text\_formatAstronomical\_Twilight | *Shows the period of day (i.e. day or night) based on astronomical twilight.* |
| grid\_3x3Humidity(%) | *Shows the humidity (in percentage).* | text\_formatAstronomical\_Twilight: | *Shows the period of day (i.e. day or night) based on*  *astronomical twilight.* |
| grid\_3x3Pressure(in) | *Shows the air pressure (in inches).* |  |  |

# HYPOTHESIS

Duis autem vel eum iriure dolor in hendrerit in vulputate velit esse molestie consequat, vel illum dolore eu feugiat nulla facilisis at vero eros et accumsan.

# MATERIALS

1. Lorem ipsum dolor sit amet
2. Consectetuer adipiscing elit
3. Sed diam nonummy nibh euismod

# PROCEDURE

1. Lorem ipsum dolor sit amet
2. Consectetuer adipiscing elit
3. Sed diam nonummy nibh euismod

# DATA

| LOREM IPSUM | DOLOR SIT |  |
| --- | --- | --- |
| Lorem ipsum |  |  |
| Lorem ipsum |  |  |
| Lorem ipsum |  |  |

# RESULTS

Nam liber tempor cum soluta nobis eleifend option congue nihil imperdiet doming id quod mazim placerat facer possim assum. Typi non habent claritatem insitam; est usus legentis in iis qui facit eorum claritatem. Investigationes demonstraverunt lectores legere me lius quod ii legunt saepius.

1. Lorem ipsum dolor sit amet
2. Consectetuer adipiscing elit
3. Sed diam nonummy nibh euismod

# CONCLUSION

Lorem ipsum dolor sit amet, consectetuer adipiscing elit, sed diam nonummy nibh euismod tincidunt ut laoreet dolore magna aliquam erat volutpat. Ut wisi enim ad minim veniam, quis nostrud exerci tation ullamcorper suscipit lobortis nisl ut aliquip ex ea commodo consequat.

# REFERENCES

1. Lorem ipsum dolor sit amet
2. Consectetuer adipiscing elit
3. Sed diam nonummy nibh euismod