Analysis of Road Accidents in United States

Name: Aswini Pusuluri

Student ID: 015310269

GitHub ID: https://github.com/AswiniPusuluri/Data-Visualization

Abstract

Road Safety has become a crucial factor in determining the day-to-day lives of every individual. Improvement in living standards along with social and economic statuses has introduced new trends in the motor vehicle sector, thus resulting in an increased number of vehicles on the road each year. Road accidents have become a major unnatural global problem claiming staggering numbers of fatalities every year. Despite the remote work options and reduced traffic trends in major cities in the last couple of years, the traffic mortality rate is still on the rise. The death estimates from motor vehicle crashes in the first half of 2021 have risen by 18.4% over the preceding year with an estimate of 20160 deaths. This is a shocking revelation to the concerned authorities as road accidents claim a fair share of the economy in the form of shattered lives, disabled persons, cost for the treatment of injuries and repair costs to the roads and vehicles involved. There are also intangible effects that prevail from road accidents that might not reflect in any of the statistics, such as the effect on the mental and physical wellbeing of the individuals involved in accidents, long term, or permanent disabilities, etc. As per WHO statistics 1.35 million succumb to road accidents every year globally of which 38,000 deaths occur in the United States involving a total of 280 million vehicles and 227.5 million drivers in the US. The impact on the US economy is estimated as high as $800 billion each year. Although there are numerous causes for road accidents, most of them would broadly fall under the categories which are humanly controllable and can be avoided by exercising caution and necessary safety measures in place. The most common causes were the behavior of the drivers, weather, age, and time.

The main aim of this project is to perform an analysis on road accidents in the United States caused by major factors such as climatic conditions, location, traffic flow interruptions, time period (day or night), and determine which one has a significant effect on the economy and lives of people involved. This analysis would further help in identifying the key factors that are

responsible for the cause of accidents and suggest preventive and precautionary measures to curb road accidents. This analysis will be performed using a public dataset from Kaggle which has road accidents details from 49 states of the United States. This data was collected from multiple APIs that provide continuous streaming data from 2016 to 2020. These APIs transmit traffic events collected by a range of institutions inside the road networks, including US transportation authorities, traffic cameras and sensors, and law enforcement agencies. This dataset reportedly contains around 1.5 million accident records.

The following are some of the visualizations proposed as part of this analysis:

- Top ten states in the United States with the most road accident cases

- Cities having a high rate of car accidents

- Environmental conditions that cause the majority of accidents in the United States

- The average number of road accidents per day

- The course of road accidents from 2016 to 2020

- Percentage of the severity of road accidents

- The proportion of traffic accidents for various time periods

- Which periods of the day did the most accidents occur?

- The overall number of accidents when there are any traffic flow interruptions

The dataset used for this project has 47 attributes with the description as listed below:

ID: This is used for the unique identification of accident records.

Severity: Shows the accident's severity as a number between 1 and 4, with 1 as low severity and 4 as high severity

Start_Time: Start time of the accident

End_Time: End time of the accident with all the traffic cleared

Start_Lat & Start_Lng: Accident starting point GPS coordinates

End_Lat & End_Lng: Accident ending point GPS coordinates

Distance(mi): The accident's impact on the entire path

Description: Description of the accident

Number: Street number where the accident took place

Street: Street name where the accident took place

Side: Gives information regarding the side of the road the accident took place

City, County, State, Zipcode & Country: This attribute shows the entire address of the incident

Timezone: Timezone depending on the accident's location.

Airport_Code: The nearest weather station to the accident site is an airport-based weather station.

Weather_Timestamp: Timestamp when the weather details are noted

Temperature(F): Temperature at the time of the accident

Wind_Chill(F): Windchill at the time of the accident

Humidity (%): Humidity during the time of the accident

Pressure(in): Air pressure during the time of the accident

Visibility(mi): Visibility during the time of the accident

Wind_Direction & Wind_Speed(mph): Direction and speed of wind at the time of the accident

Precipitation(in): Precipitation measured in inches

Weather_Condition: Condition of weather for example snow, rain, etc

Amenity: Shows the presence of a local amenity using POI annotation

Bump: Indicates the existence of a speed bump or hump in the vicinity using POI annotation

Crossing: Shows the existence of a crossing in the immediate vicinity.

Give_Way: Shows the existence of a Give_Way in the immediate vicinity.

Junction: Shows the existence of a Junction in the immediate vicinity.

No_Exit: Shows the existence of a No_Exit in the immediate vicinity.

Railway: Shows the existence of a Railway in the immediate vicinity.

Roundabout: Shows the existence of a Roundabout in the immediate vicinity.

Station: Shows the existence of a Station in the immediate vicinity.

Stop: Shows the existence of a Stop in the immediate vicinity.

Traffic_Calming: Shows the existence of a Traffic_Calming in the immediate vicinity.

Traffic_Signal: Shows the existence of a Traffic_Signal in the immediate vicinity.

Turning_Loop: Shows the existence of a Turning_Loop in the immediate vicinity.

Sunrise_Sunset: The time of day or night is displayed based on the Sunrise_Sunset.

Civil_Twilight: The time of day or night is displayed based on theCivil_Twilight.

Nautical_Twilight: The time of day or night is displayed based on the Nautical_Twilight.

Astronomical_Twilight: The time of day or night is displayed based on Astronomical_Twilight.

Milestone and different phases of the project with tentative end dates:

Phases	Tentative Start Date	Tentative End Date
Abstract	02/01/2022	02/20/2022
Data collection	02/01/2022	02/20/2022
Data Exploration	02/21/2022	02/28/2022
Data Cleaning & Preparation	03/01/2022	03/15/2022
Data Visualizations	03/16/2022	04/19/2022
Documentation	04/20/2022	05/01/2022