# How Alcohol Consumption Fuels Traffic Accidents Across U.S. States

## TEAM B.I.G

**Vivek Venkateshprasad**
vv002@ucr.edu
*862468868*

**Shubhank Joshi**
sjosh052@ucr.edu
*862467912*

**Prathik Somanath**
psoma005@ucr.edu
*862467832*

**Kyle Russell**
kruss020@ucr.edu
*862302220*

## 1 Background

This project investigates the relationship between alcohol consumption and traffic accidents in the U.S. and the Per Capita Alcohol Consumption. By analyzing the correlation between alcohol consumption trends and the occurrence of traffic accidents across states, we aim to provide insights that contribute to public health policies focused on road safety and alcohol regulations.

Numerous studies have documented the strong link between alcohol consumption and traffic accidents. According to the National Highway Traffic Safety Administration [1], alcohol was involved in 31% of all U.S. traffic fatalities in 2021, resulting in over 13,000 deaths caused by drivers with blood alcohol concentrations (BAC) of 0.08 or higher[2]. Despite these alarming statistics, research indicates that BAC testing is not consistently performed in fatal crashes, as only 38% of drivers involved undergo testing [4] . This highlights the potential underreporting of alcohol-related accidents.

Recent studies, including those from Jiang et al. (2022), have emphasized the need for better data collection and analysis of alcohol-impaired driving incidents. Their research suggests that regional alcohol consumption levels are a significant predictor of traffic accident frequency and severity . Another study by Voas et al. (2020) [3] explores the temporal patterns of alcohol consumption and finds that increases in alcohol consumption during specific times, such as weekends and holidays, lead to spikes in traffic accidents . These studies align with the project's goal of leveraging big data to uncover temporal and geographical trends in alcohol-related accidents.

This project will apply geospatial analysis techniques to show accident hotspots based on alcohol consumption data. By mapping per capita alcohol consumption to traffic accident rates, we aim to provide actionable insights to develop more targeted interventions for reducing alcohol-related road fatalities.

## 2 Motivation

According to the National Highway Traffic Safety Administration [1], approximately 37 individuals in the United States lose their lives daily due to drunk-driving crashes, equating to one fatality every 39 minutes (NHTSA, 2023). This statistic underscores the significant role of drunk driving in traffic-related accidents. However, it raises critical questions: Does an increase in alcohol consumption directly correlate with a rise in drunk driving incidents, and consequently, with a higher rate of accidents? Additionally, do different types of alcohol have varying effects on the likelihood of traffic accidents? This project seeks to investigate these questions. Addressing them could provide valuable insights into the relationship between alcohol consumption and traffic accident rates, offering data that may assist policymakers in developing more targeted strategies to mitigate alcohol-related accidents and fatalities.

## 3 Dataset

US Accidents (2016 - 2023)

- This data set compiles a large data set of US accidents from 2016 to 2023.
- State: This column will be important to organize our data to determine where our accident took place. We can also look at County, City, and Street for more specific information.
- Start_Time: This is essential for determining when this accident occurred.

Per Capita Alcohol Consumption by State

- The data set compiles various aspects of data related to alcohol consumption, age, and population.

- Year and State: Data will be grouped based on these columns, as the year will determine when our data applies and the state will help us determine the geographical position of the data.

- Ethanol_Per_Capita_21_Plus: This data will be important to attain alcohol consumption proportional to the population.

## 4 Outcome

This project will provide an analysis of the relationship between alcohol consumption and accident rates over time. As such, we plan to represent our data using a combination of time series and a geographical map of the United States. We can compare the data between alcohol consumption and accident rates side by side and use the Pearson Correlation (or other such statistical measurement) to determine the strength of the correlation.

## 5 Relevance

Finding out the correlation between alcohol consumption and traffic accidents. This research can identify regions with higher instances of DUI-related accidents, helping law enforcement agencies focus on interventions like sobriety checkpoints or public awareness campaigns. Alcohol-related accidents result in significant economic costs, including medical expenses, lost productivity, legal costs, and property damage. Reducing the rate of such accidents can lead to substantial savings for individuals, insurers, and governments. Due to the time series nature of the data, we can also identify the trends in increase of road accidents during particular days. By providing a clear understanding of the relationship between alcohol consumption and traffic accidents, this project can serve as a foundation for safer roads, informed policymaking, and public health improvements across the U.S.
This is relevant to big data for the following reasons:
1. Volume
The project would involve massive datasets, such as traffic accident reports, severity of accidents, alcohol sales data and demographic information. These datasets often span multiple years and cover a large geographic area (all U.S. states), making the data volume significant.
2. Veracity
Big data often involves dealing with inconsistencies, inaccuracies, and missing data, which is true for this project. Addressing this requires sophisticated data cleaning, normalization, and verification techniques.
3. Data Management
This project requires significant amount of ETL processing which requires passing data through multiple pipelines to get the desired cleaned data.

## 6 Evaluation

The first step is to visualise this information onto the map of the united states. This will provide us a clear understanding of the hot spots for traffic accidents and alcohol consumption. Then, we then use correlation constants to find out the statistical strength of the correlation between alcohol consumption and traffic accidents? Is it statistically significant , and if so, how strong is the relationship. These visualizations will helps us find the correlation between alcohol consumption and traffic accidents.

## 7 Milestones

• 3rd week of October - Project proposal report.
• 4th week of October - Project proposal presentation.
• 5th week of October - Data Collection and Cleaning.
• 1st week of November - Literature survey.
• 2nd week of November - Implement Visualization for alcohol consumption vs traffic accidents across US states.
• 3rd week of November - Comparative analysis of data, integrating different datasets for visualization, performance evaluation.
• 4th week of November - Report outline.
• 5th week of November - Final project presentation.
• 1st week of December - Final project deliverables.

## References

[1] NHTSA. (2023). Traffic Safety Facts Annual Report 2021. https://www.nhtsa.gov/risky-driving/drunk-driving

[2] SafeTREC Traffic Safety Facts: Alcohol-Impaired and Alcohol-Involved Driving (2023)

https://safetrec.berkeley.edu/2023-safetrec-traffic-safety-facts-alcohol-impaired-and-alcohol-involved-driving

[3] Voas, R. B., Torres, P., Romano, E. O., & Lacey, J. H. (2020). *Crashes: Temporal Trends and Regional Variations." Accident Analysis & Prevention, 134, 105316.*

[4] International Transport Forum (ITF). (2024). Alcohol-Related Road Casualties in Official Crash Statistics. https://www.itf-oecd.org/alcohol-related-road-casualties.

[5] The Theory of Parsing, Translation and Compiling, volume 1. Prentice-Hall, Englewood Cliffs, NJ.

[6] Sobhan Moosavi. (2023). US Accidents (2016 - 2023) [Data set]. Kaggle. https://doi.org/10.34740/KAGGLE/DS/199387

[7] Decades of drinking: Per capita alcohol consumption by state: Spreadsheet download. Gigasheet. (n.d.). https://www.gigasheet.com/sample-data/per-capita-us-alcohol-consumption .