The dataset for this data analysis project is Recalls Data from the U.S. Department of Transportation. The dataset is at this link: Recalls Data | Tyler Data & Insights (transportation.gov). This dataset includes recall information related to specific NHTSA campaigns as far back as 1966. Unfortunately, vehicle recalls are relatively common. Problems range from minor repairs to major repairs. Recently, Tesla recalled 362,758 vehicles due to the beta version of its Full Self-Driving software may cause crashes. Another notable case was the Ford Pinto in the 1970s. According to the U.S. Department of Transportation: "manufacturers who determine that a product or piece of original equipment either contains a safety defect or is not in compliance with Federal safety standards are required to notify NHTSA within five business days."

However, vehicle owners must be made aware of recalls for their vehicles. Manufacturers may require the owners to register their cars on the manufacturer's website for recall notifications. If the owner does not, they will know about the recall if they see it in the news. For example, recently, Tesla recalled 362,758 vehicles due to the possibility of Full Self-Driving Beta software may cause crashes. Then there are cases like Ford and the Ford Pinto where the company refused to acknowledge a problem, and severe injuries and deaths occurred. This dataset is interesting because of cases like the Ford Pinto, Tesla’s recent recall, and Honda in the early 1990s. Recall information is a valuable consideration when deciding which vehicle to purchase. That is why this dataset is meaningful, engaging, and relevant.

Data from this dataset will be collected via the website's API and imported into Python, specifically a Jupyter notebook. An ANOVA test will compare the means of the recalls numbers between three automobile manufacturers – Ford, Volkswagen, and Honda -- to determine which company has the best overall track record regarding recalls (the lowest number of vehicles impacted) and if there is a difference between the three companies mean recall numbers on average. In short, the business question for this project is: Is there a significant difference between the number of Ford, Volkswagen, and Honda recalls? These companies are rivals and constantly compare their vehicles to one another. Ads tout one brand's vehicle better than others. In addition, I chose one company from each of the three major regions – the United States, Germany, and Japan. Hypotheses to test are the following: Null Hypothesis: The average of the manufacturers' recall numbers is not statistically different. Hypothesis 1: There is a statistical difference between the manufacturers' recall numbers. The dataset has twelve columns. The data type is in parentheses: Reported Received Date (date and time); NHTSA ID (string); Recall Link (website URL/string); Manufacturer (string); Subject (string); Component (string); Mfr Campaign Number (string)’ Recall Type (string); Potentially Affected (Number); Recall Description (string); Consequence Summary (string); and Corrective Action (string).

I will filter the dataset into a subset with the manufacturers selected. Appropriate descriptive statistics will explore the dataset – including data visualizations. The "Potentially Affected" column will be used as the metric to find the averages between the manufacturers for the ANOVA and posthoc tests. Prospective vehicle buyers, renters, and rental car companies will find this research valuable. Sites like CarFax do provide recall information – for a fee. This project’s recall information will be free and compares different auto manufacturers for statistical differences. Car Fax and sites like it don't do either of these things. Therefore, vehicle buyers can compare vehicle options/manufacturers during their car shopping research. This research can lead to safer vehicle operation and less injury and death on the roads in the United States. The targeted audience will benefit from the research findings by ensuring vehicles get repaired, increasing negotiating power on car price if a car is recalled, and the knowledge of which companies are better than others (out of the three selected) on a critical safety metric: average number of vehicle recalls. After all, a manufacturer with larger-than-average recall numbers may have a poor track record for safety and vehicle construction.