# Project #3 Final Report

## Introduction:

The essential role of transportation systems in economic development and daily operations requires thorough investigations into driver behavior and transport trends for developing optimal infrastructure and policy decisions. These important aspects become accessible through the combination of National Household Travel Survey (NHTS) and Next Generation Simulation (NGSIM) datasets. The research from both NHTS and NGSIM offers data about household vehicle possession alongside demographic information and travel data and displays extensive information about vehicles through their location positions and velocities and acceleration rates.

This document explains the variables selected for analysis and plot construction methods along with gained insights and their relevance to the transportation field.

**Selected Variables:** To best convey the transportation systems in the US, we decided to focus on the variables listed below:

NGSIM Data

| Time | time interval in seconds | Rationale |
|---|---|---|
| leader_position(m) | position of the leading vehicle at every time step | Tracks vehicle trajectories and spacing to improve traffic models and adaptive cruise control. |
| follower_position(m) | position of the follower vehicle at every time step | Identifies lane-changing patterns and collision risks, enhancing safety measures. |
| leader_speed(m/s) | speed of the leading vehicle at every time step | Detects congestion patterns and speed variations to support better traffic management. |
| follower_speed(m/s) | speed of the following vehicle at every time step | Highlights differences in speed between leader and follower to analyze reaction times. |
| leader_acc(m/s^2) | acceleration of the leading vehicle at every time step | Provides insight into braking patterns and acceleration for refining autonomous vehicle systems. |

| | | |
|---|---|---|
| follower_acc(m/s^2) | acceleration of the following vehicle at every time step | Identifies sudden braking and acceleration behavior, aiding in collision risk analysis. |
| trajectory_number | an id representing a leader-follower pair dataset | Enables detailed tracking of driver interactions and enhances prediction models. |

NHTS Data

| | | |
|---|---|---|
| census_division | Census division classification for home address | Helps identify regional variations in transportation patterns and optimize resource allocation for public transit. |
| household_income | Household income | Influences vehicle ownership and mode choice, enabling policymakers to address equity in transportation access.. |
| vehicle_age | Age of vehicle | Indicates environmental impact and safety risks, informing emissions control programs and vehicle replacement incentives. |
| vehicles_per_household | Number of vehicles per household | Provides insights into car dependency, predicting congestion and identifying areas with potential transit needs.. |

**Stacked Bar Chart:**

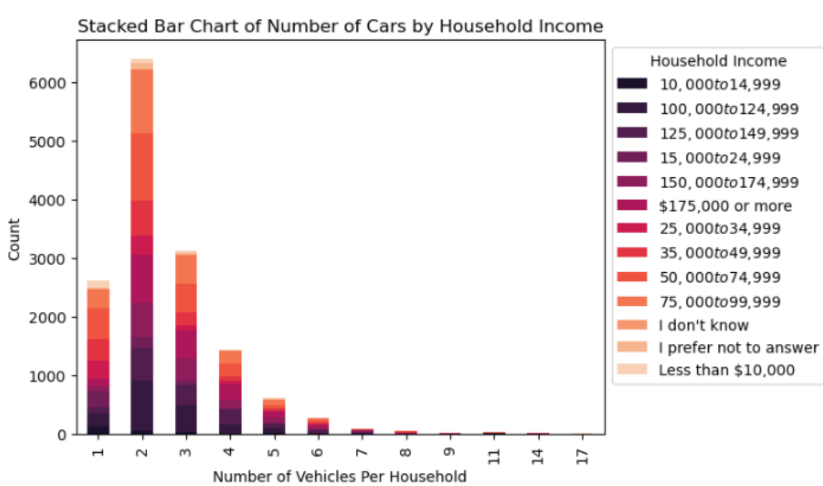


*Figure 1: Bar Chart*

Our stacked bar chart analyzes the number of cars in each household by the household income. To construct this chart, we first used the matplotlib.pyplot package to create a stacked bar chart that would depict the amount of household vehicle ownership levels with respect to income. We did this through using data points collected from the NHTS dataset between household_income and vehicles_per_household variables. Within the clustered income groups, the plot displayed household vehicle counts.

Key Findings:

The provided chart reveals an undisputed relationship that exists between household earnings and the number of vehicles they possess. The greatest numbers of vehicles belong to affluent households whereas less fortunate groups normally possess fewer vehicles. Half of middle-income families own two vehicles but a significant segment also owns one or two vehicles.

Relevance to the Transportation Industry:

Strategic transportation policy development requires knowledge of household vehicle possession rates between different financial groups. The information enables organizations to improve their transit services for places that do not have many vehicle owners.

**Histogram:**

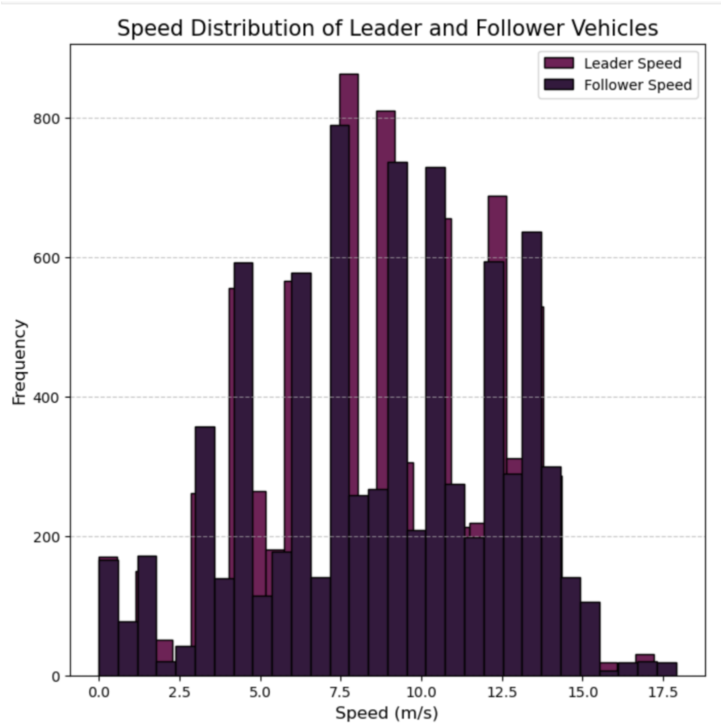


*Figure 2: Histogram*

We built a histogram using the matplotlib.pyplot package that showed comparisons between leader_speed and follower_speed, using data points in the NGSIM dataset. The variable speed range was distributed across 30 bins in order to summarize the data in an effective manner. The vehicle speed was a defined as the x variable and count was defined as the y variable.

Key Findings:

Both leader and follower vehicle speeds follow nearly identical distributions that peak at speeds ranging between 10-25 m/s. Followers adapt their speeds to leader vehicle modifications when traffic becomes congested, which would cause both vehicles to follow the same pattern.

Relevance to the Transportation Industry:

Speed distribution analysis helps improve simulation model predictions and enables better speed limit policies. The definition of speed ranges, which encompass typical vehicle speeds, enables the development of adaptive cruise control systems alongside collision avoidance technology.

**Boxplot:**

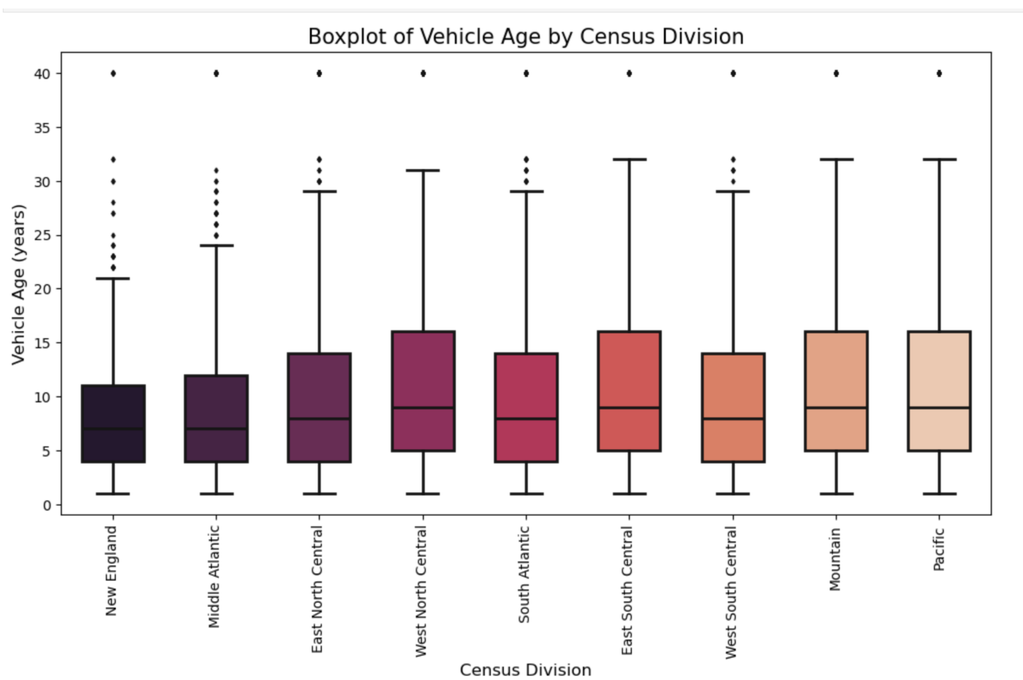


*Figure 3: Boxplot*

Our boxplot analyzes the vehicle age by census division using the matplotlib.pyplot package. The NHTS data was analyzed and used to create a boxplot for vehicle_age data alongside census_division information. The plot demonstrates how different census division their vehicles in terms of age. Census_division was assigned as our x variable, while vehicle_age was assigned to be the y value.

Key Findings:

The data reveals significant differences in vehicle age distribution between different geographical areas. The lower age of vehicles in urban locations corresponds to their greater availability of modern transport options. The vehicles in rural locations tend to be older than those in urban areas resulting in increased vehicle maintenance expenses and environmental emissions.

Relevance to the Transportation Industry:

Vehicle age distribution information helps authorities develop environmental policies together with vehicle replacement programs. Older vehicles within the population need specific interventions for emission control and road safety improvement.

**Time-chart 1 (Leader and Follower Acceleration vs Time):**

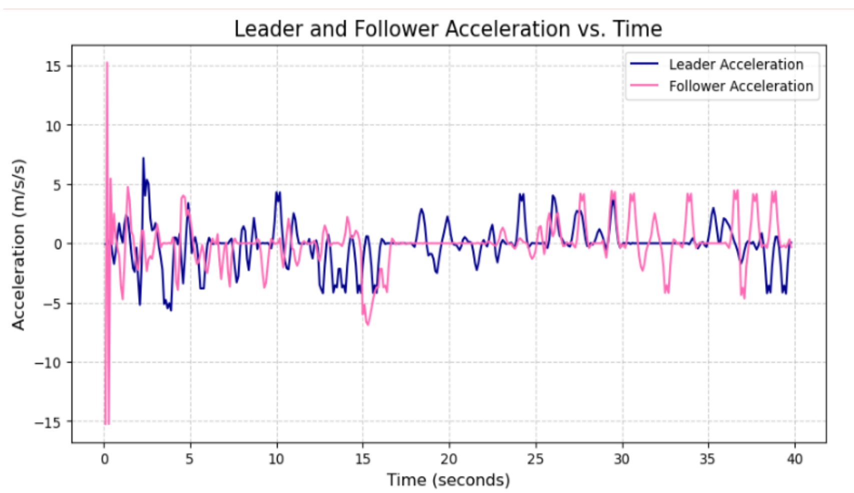


*Figure 4 Time Series (Acceleration vs Time)*

Our first time-series plot looks at the leader and follower acceleration vs time. To create this, the seaborn package was used to create a basic lineplot, before customizations were done. The analytical chart examined both leader and follower vehicle acceleration dynamics through time, with time being assigned the x variable, and acceleration being assigned the y variable. The NGSIM dataset was analyzed when plotting this graph.

Key Findings:

The follower vehicle initiation matches the leader vehicle triggering by a short time interval. The numerous decelerations in congested traffic environments point toward automatic stopping followed by starting which leads to higher fuel usage and less driving efficiency.

Relevance to the Transportation Industry:

In addition to improving autonomous vehicle algorithms, the analysis of acceleration patterns facilitates the development of better adaptive cruise control systems. Such traffic patterns help improve the timing of traffic signals along with their management of intersections. It also shows insight into the environmental impact of high-traffic areas, with lots of stop and go traffic, which contributes to high emissions. Understanding the acceleration patterns can help show areas in need of improvements for smoother traffic flow in order to help reduce emissions.

**Time-chart 2 (Leader and Follower Position vs Time with Velocity Overlay):**


*Figure 5 Time Series (Leader and Follower vs Time)*

Our second time-series plot analyzes NGSIM data to show the leader and follower position over time, with a color bar showing the velocity at any given position. The matplotlib.pyplot package was used to create a lineplot, where the time was assigned x variable, and position was assigned the y variable. The speed was then set equal to variable 'c' in order to create the color bar. The matplotlib.animations, matplotlib.cm, and matplotlib.colors were all used to create the color bar properly. The visualization added a color bar element which displayed vehicle velocities to offer better understanding of driving behaviors.

Key Findings:

Observations within the plot show that vehicles decrease distance between one another during slowing phases while increasing distance when they accelerate. The peaks in velocity data indicate both quick acceleration and possible lane shifting events that affect road traffic operations.

Relevance to the Transportation Industry:

Examining the effects of position and velocity variations has effects on the creation of safety systems that enhance lane control and collision avoidance. Road network efficiency and smart transportation system design gain improvement through the obtained insights. By observing this graph, it is important to analyze how vehicles interact with each other with following distance and braking. This can help the industry predict traffic patterns in order to best understand were improvements are needed.

**Conclusion:**

In conclusion, a significant impact of infrastructure planning and transportation policy decisions were produced by the combined analysis of NHTS and NGSIM data. The strongest connection exists between individual family income levels and the number of cars they possess. Households earning more money tend to have larger vehicle collections than income levels below them that prioritize using public transit systems. In order to create equal mobility opportunities, the analysis highlights the need for specific transportation policies that improve public transit services in low-income areas. Typical vehicle speeds fall between 10-25 m/s including a minimal speed response from following vehicles whose behavior depends on the preceding vehicle. Better traffic simulations, optimized speed limits, and adaptive cruise control technologies for safer travel systems are all supported by the data.

According to a boxplot analysis of vehicle age, rural communities drive older cars, while cities tend to have newer models. Areas possessing older vehicle fleets need specialized incentives for vehicle substitution along with emission reduction programs because these vehicles increase both owner expenses and environmental hazard levels.

Results from the acceleration analysis indicated that trailing vehicles exactly duplicate leader vehicle actions thus demonstrating stop-and-go traffic conditions. Better traffic control system optimization, adaptive cruise control, and autonomous vehicle technology are all made possible by research into these driving patterns.

The time-series analysis of vehicle positions together with velocity data pointed out crucial events of slowing down and lane shifting. The acquired insights serve as fundamental elements for better developing systems designed to stop collisions and manage traffic functions. In order to promote sustainable future mobility, planners can use this data to create safer driving conditions, reduce traffic, and increase system operational efficiency.