Decoding Usage and Adoption Behavior of the Low-Carbon Transportation Market: An AI-driven Exploration

Vuban Chowdhury

University of Arkansas-Fayetteville

Follow this and additional works at: https://scholarworks.uark.edu/etd

Part of the Artificial Intelligence and Robotics Commons, Civil Engineering Commons, Transportation Commons, and the Transportation Engineering Commons

Citation

This Thesis is brought to you for free and open access by ScholarWorks@UARK. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of ScholarWorks@UARK. For more information, please contact scholar@uark.edu.
Decoding Usage and Adoption Behavior of the Low-Carbon Transportation Market: An AI-driven Exploration

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in Civil Engineering

by

Vuban Chowdhury
Islamic University of Technology
Bachelor of Science in Civil Engineering, 2021

December 2023
University of Arkansas

This thesis is approved for recommendation to the Graduate Council.

Suman Mitra, Ph.D.
Thesis Director

Andrew Braham, Ph.D.
Committee Member

Lekshmi Sasidharan, Ph.D.
Committee Member
Abstract

The transportation sector stands as a significant contributor to greenhouse gas emissions in the United States, with its environmental impact steadily escalating over the past few decades. This has prompted government agencies to facilitate the adoption and usage of low-carbon transportation (LCT) options as alternatives to fossil-fuel-powered transportation. LCTs include modes of transportation that minimize the overall carbon footprint of the transportation sector by relying on energy sources that are environmentally sustainable. These sustainable transportation options have also garnered significant interest in the transportation research community. For government agencies and researchers alike, a comprehensive understanding of the adoption and usage of LCT is necessary to fully realize their carbon mitigation potential.

The transportation sector is made up of several sub-sectors (e.g., light duty vehicle sector, heavy-duty vehicle sector, off-road equipment sector), which differ with regard to their compatibility and acceptance of LCT. Likewise, the research gaps for these sub-sectors also vary. This heterogeneous nature of the transportation sector necessitates a separate consideration of the factors associated with LCT adoption and usage in the different sub-sectors. Consequently, this thesis explores these factors separately for the the light-duty vehicle (LDV) sector, the heavy-duty vehicle (HDV) sector and off-road equipment (ORE) sector. In the LDV sector, electric vehicles (EVs) have been some of the most widely accepted form of LCT. However, the literature has primarily explored their usage with respect to vehicle miles travelled. This thesis features a unique exploration of EV usage through the lens of vehicle choice in the LDV sector. To this end, the thesis employs a two-step machine learning framework (clustering and decision trees) on the National Household Travel Survey 2017 datasets. For HDVs and OREs, on the other hand, the influence of behavioral factors (awareness and impression) on LCT adoption has
not been widely studied. To address this research gap, this thesis conducts a series of semi-structured interviews and analyzes them using a qualitative content analysis. The results of the analysis were refined by generative artificial intelligence (AI). Moreover, the analysis also explores the varying importance of factors (behavioral and other) of LCT adoption within different types (e.g., long-haul vs short-haul, large fleet vs small fleet) of organizations.

The findings from the exploration of the LDV sector can be valuable to policymakers attempting to boost EV usage within the sector. The exploration of usage patterns can also provide auto manufacturers a unique perspective on the shortcomings of EVs that need to be addressed. For the HDV and ORE sectors, the findings can guide government agencies to develop tailored campaigns that leverage the influence of behavioral factors on LCT adoption. Moreover, the insights uncovered from the utilization of generative AI can serve as a baseline for researchers seeking to conduct AI-assisted qualitative studies in transportation.
Acknowledgement

I extend my heartfelt appreciation and acknowledgement to all those who played a role in the successful completion of my master's thesis. I am profoundly thankful to my academic advisor, Dr. Suman Mitra, whose unwavering guidance and encouraging words were indispensable in making this work possible. I am also grateful to my committee members, Dr. Andrew Braham, and Dr. Lekshmi Sasidharan for their cooperation with my thesis defense. I highly appreciate the assistance from California Air Resources Board (CARB) as their support (financial and non-financial) played an important role in the realization of this thesis.

I want to express my sincere gratitude to my colleague and fellow researcher Farzana Mehzabin Tuli for her assistance with this research study. I would also like to extend this appreciation to my colleagues at the Zero Lab, as their positive attitude and support has made me love the research that I do.

My heartfelt gratitude goes to my parents and family members for their continuous support, which has been an unwavering source of strength and encouragement. And lastly, I want to express my appreciation to the Almighty for navigating me through difficulties. His blessings were a guiding force, leading me to the completion of my thesis and my degree.
Table of Contents

Introduction ............................................................................................................................................... 1

References ............................................................................................................................................... 4

   Abstract ................................................................................................................................................. 7
   1.1. Introduction ........................................................................................................................................ 8
   1.2. Data .................................................................................................................................................. 12
       1.2.1. NHTS 2017 ................................................................................................................................. 12
       1.2.2. Data Cleaning ............................................................................................................................ 12
       1.2.3. Variable Selection .................................................................................................................... 14
   1.3. Methodology ..................................................................................................................................... 14
       1.3.1. Clustering Model ....................................................................................................................... 16
       1.3.2. Classification Model ............................................................................................................... 17
   1.4. Results and Discussion .................................................................................................................... 20
       1.4.1. Clustering Model ....................................................................................................................... 21
       1.4.2. Classification Model ............................................................................................................... 23
   1.5. Conclusion ...................................................................................................................................... 31
   References .............................................................................................................................................. 33

2. Assessing Behavioral Factors in Low-Carbon Transportation Adoption Among Heavy-Duty and Off-Road Transportation Sectors in California: A Generative-AI-Assisted Content Analysis .............................................................................................................................. 39
   Abstract ................................................................................................................................................. 39
2.1. **Introduction** ................................................................. 41

2.2. **Data** .................................................................................. 44

   2.2.1. Indexing and Recruiting............................................... 44

   2.2.2. Semi-structured Interviews ......................................... 45

2.3. **Methodology** ................................................................. 46

   2.3.1. Content Analysis............................................................ 46

   2.3.2. Manual Coding, Segmentation, and Analysis............... 46

   2.3.3. Generative AI assistance ............................................. 48

2.4. **Results and Discussion** .................................................. 50

   2.4.1. Behavioral Factors of LCT Adoption ........................... 50

   2.4.2. Other Factors influencing LCT adoption ...................... 55

   2.4.3. Existing Incentives and Expected Government Support .... 60

   2.4.4. Generative AI in Content Analysis ............................. 62

2.5. **Conclusion** ...................................................................... 64

**References** ............................................................................. 66

**Conclusion** ............................................................................. 72

**Funding Sources** ................................................................... 74
List of Figures

Figure 1-1: The percentages of trips made by different vehicle types and fuel types (top) and the number of trips from 50 states in the US (bottom) ................................................................. 13

Figure 1-2: Two-step modeling framework used to predict vehicle choice ................................. 17

Figure 1-3: Comparison of k-modes clustering models based on the elbow plot (left) and Silhouette score (right) ........................................................................................................ 21

Figure 1-4: Comparison of accuracies of the classification models ............................................. 24

Figure 1-5: The variables in the decision tree model arranged in ascending order of variable importance .......................................................................................................................... 25

Figure 1-6: The accumulated local effect (ALE) plots for the variables in the decision tree model. The black dots (joined by the broken lines or continuous lines) indicate the effect of the variables at a specific value. The light blue bars indicate number of observations for a specific value of a variable. ........................................................................................................ 26

Figure 2-1: Adoption behavior, vocation, and fleet size of the organizations interviewed .......... 45

Figure 2-2: Mixed approach (concept-driven and data-driven) to build the coding framework .. 48

Figure 2-3: Awareness and impression of LCT and incentives .................................................... 51

Figure 2-4: Association between awareness of LCT and potential to adopt LCT ....................... 52

Figure 2-5: Association between awareness of incentives and potential to adopt LCT .......... 54

Figure 2-6: Association between environmental awareness and potential to adopt LCT ........... 55
List of Tables

Table 1-1: The descriptive statistics of the variables used in this study ......................... 15

Table 1-2: Proportion of different trip attributes for the 5 trip clusters .......................... 22

Table 2-1: Awareness, impression, and the three most important factors influencing the LCT adoption behavior of different categories of organizations ................................................................. 58

Table 2-2: Awareness, impression, and the three most important reasons negatively influencing the impression of incentives and expected government support for different categories of organizations ................................................................. 61
Introduction

The United States, home to approximately 5% of the global population, is responsible for 28% of global greenhouse gas (GHG) emissions (Zulinski, 2018). Among all the sectors of the US economy, the transportation sector is the largest contributor to greenhouse gas emissions, accounting for 29% of total emissions (EPA, 2023). In the last 33 years, greenhouse gas emissions from the transportation sector has been growing rapidly, with an average growth of 18.6%. Additionally, the transportation sector is also the largest source of premature deaths caused by air emissions in the U.S., with 58,000 annual fatalities (Caiazzo et al., 2013). These alarming numbers have produced a pressing need among local and federal government agencies to facilitate the reduction of GHG emissions from the sector.

The transportation sector is made up of light-duty vehicle (LDV) sector (<8,500 lbs.), medium-duty vehicle (MDV) sector (10,000 lbs. to 26,000 lbs.), heavy-duty vehicle (HDV) sector (>26,000 lbs.) and the off-road equipment (ORE) sector. Although the vehicles/equipment from these sub-sectors predominantly run on diesel or gasoline, many of them have low-carbon alternatives. These low-carbon alternatives are collectively referred to as low-carbon transportation (LCT). LCTs include any vehicles or equipment that are solely powered by alternative energy sources such as battery electric vehicles (BEVs) (U.S. Department of Energy, 2013) or partially powered by these energy sources such as hybrid electric vehicles (HEVs) and plug-in hybrid electric vehicles (PHEVs) (Chitradeep. Sen, 2010). As a wide number of studies suggest, facilitating the reduction of GHG emissions within the transportation sector is closely associated with promoting these LCT alternatives (Ercan et al., 2016; B. Sen et al., 2017).

Promoting the adoption of LCT is only the first step in the path to decarbonization of the transportation sector. The carbon reduction potential of LCT is ultimately determined by how
much these vehicles are driven after purchase (Davis, 2019). In other words, achieving decarbonization within the transportation sector can be delineated into two distinct phases: first, enabling the widespread adoption of low-carbon technology (LCT), and second, fostering its extensive usage.

The different sub-sectors of transportation (e.g., LDV sector, HDV sector) are at different stages with regard to their acceptance of LCT. The LDV sector has seen a rapid growth in adoption over the last few years (0.2% market share in 2017 to 7% market share in 2023) (Peter Johnson, 2023). Given the market penetration of LCTs in the LDV sector, it is warranted to put emphasis on the usage of the adopted vehicles alongside their adoption. However, the HDV and ORE sectors still haven’t embraced LCT in significant numbers (Leard & Mcconnell, 2020). For instance, 98% of the class 8 trucks (>33,000 lbs.) are powered by internal combustion engines (Fleming et al., 2021). Hence, the adoption of LCTs in the HDV sector still needs further attention. The lower penetration of LCT in the HDV and ORE sectors might also be attributable to the greater emphasis placed on passenger LDVs with regard to emissions reduction strategies (Fleming et al., 2021), which constitute majority of the vehicles on the road. But it is also important to note that the HDVs and MDVs (constituting only 10% of the on-road vehicles) contribute disproportionately (29%) to transportation GHG emissions (O’ Dea, 2019).

The number of studies on adoption and usage of LCT in the LDV, HDV and ORE sectors aligns with the respective market penetration levels in each of these sectors. The adoption of LCTs in the LDV sector has received more attention (than HDV and ORE sectors) in the transportation research community, which is representative of their market share. However, a parallel effort in studying LCT usage (after they are adopted) in the LDV sector appears to be missing; the number of studies that explore the determinants of low-carbon LDV usage are few
in number (Chakraborty et al., 2022; Mandev et al., 2022; Srinivasa Raghavan & Tal, 2021). The gap in the literature primarily lies in the exploration of low-carbon LDV usage through the lens of vehicle choice in multi-vehicle households. With regard to the HDV and ORE sectors, the literature is scant on both adoption and usage. Although studies on the adoption of HDVs within organizations exist (Bae et al., 2022; Sierzchula, 2014), there are some research questions that are yet to be answered. Particularly, the influence of behavioral factors on LCT adoption in HDV and ORE sector requires further consideration. The introduction sections of the two chapters of the thesis dissects the literature and presents an elaborate discussion of these research gaps.

Given the gap in the current LCT literature, this thesis intends to fulfill two key objectives, one pertaining to the LDV sector and the other pertaining to the HDV and ORE sectors. The key objectives are as follows:

1) What factors (trip characteristics, socio-demographic traits, built-environment variables) influence the preference of light-duty EVs for individual trips in US multi-vehicle households?

2) What is the association between behavioral factors and LCT adoption in HDV and ORE sectors?

The study uses both quantitative and qualitative AI (artificial intelligence) driven approaches to fulfill its objectives. With regard to the first objective, the study employs a two-step machine learning model on the 2017 national household travel survey dataset and applies explainable AI techniques (variable importance and accumulated local effects) on the model outcomes. By doing so, it seeks to uncover the effect of trip characteristics, socio-demographic traits, and built environment variables on EV preference. And to fulfill the second objective, the study employs a qualitative content analysis, refined by using generative AI assistance. The data for this segment
of the study was collected by conducting semi-structured interviews of 8 HDV and 4 ORE organizations from California.

The thesis is organized as follows. Chapter 1 focuses on first objective as it dissects the determinants of EV preference within US multi-vehicle households. The chapter contains a discussion on the research questions answered in previous studies, the data and two-step machine learning framework utilized in the thesis, the results of the modeling framework, and discussion on their policy implications. Chapter 2 focuses on the second objective by exploring the influence of behavioral factors on LCT adoption in the HDV and ORE sectors. It presents a discussion on the studies that have previously explored LCT adoption in these sectors, the steps involved in data collection and qualitative content analysis, the results from the analysis and their implications.

References


Sen, Chitradeep. (2010). *Performance analysis of batteries used in electric and hybrid electric vehicles*. University of Windsor.


Abstract

Electric vehicles (EVs) play a significant role in reducing carbon emissions. In the US, EVs are mostly owned by multi-vehicle households, and studied in the context of vehicle miles traveled. This study takes a unique approach by analyzing EV usage through the lens of vehicle choice (between EVs and conventional vehicles) within multi-vehicle households. An ensemble machine-learning framework (clustering and decision trees) is utilized. The framework determines the preferred trip category for EV use and captures the effects of household attributes, driver attributes, built-environment factors, and gas price on EV preference. Results indicate that discretionary trips such as weekend shopping and dining trips (accumulated local effect = 0.037) are mostly preferred for EV use. EV preference is more pronounced among households with fewer workers (<2) and lower income levels (<$150,000). These findings are valuable for policymakers and auto manufacturers in targeting specific market segments and promoting EV adoption.
1.1. Introduction

The United States, home to approximately 5% of the global population, is responsible for 28% of global carbon emissions, with the transportation sector contributing 27% (Zulinski, 2018). Conventional vehicles (CV) powered by gasoline or diesel are significant contributors to carbon emissions from the light-duty transportation sector. Hence, augmenting the adoption and usage of alternative fuel vehicles lies at the heart of the endeavor to reduce carbon emissions. Electric vehicles (EVs) are a category of alternative fuel vehicles that have garnered significant interest in the transportation research community. EVs include vehicles that are solely powered by electricity such as battery electric vehicles (BEVs) (U.S. Department of Energy, 2013) or partially powered by electricity such as hybrid electric vehicles (HEVs) and plug-in hybrid electric vehicles (PHEVs) (Chitradeep. Sen, 2010).

The vast majority of studies on EVs are centered around the research question “What factors influence the decision to purchase EVs?”. These studies answer the question by exploring the determinants of EV adoption. These determinants can be broadly classified into four groups namely, demographic, contextual, situational, and psychological (V. Singh et al., 2020). Demographic determinants such as age, education, income, household size, number of vehicles in the household, marital status, and political affiliation were deemed significant in some studies (Shin et al., 2019). Contextual determinants of EV market share pertain to charging station availability (Tanaka et al., 2014), and government policies (HOV lane access, presence of purchase incentives, tax exemptions) (Hackbarth & Madlener, 2013; Vergis & Chen, 2014). With regard to the situational factors, several studies showed that purchase cost (Carlucci et al., 2018), and more importantly, fuel cost (Carlucci et al., 2018; Hackbarth & Madlener, 2013; Soltani-
Sobh et al., 2017; Tanaka et al., 2014) are among the most important considerations among EV consumers.

Apart from exploring the determinants of EV adoption, researchers have also widely sought to answer the question “Which households or consumer groups are most likely to adopt EVs?”. Before EVs became widely available in the market, several studies assessed the potential acceptability of EVs in both single and multi-vehicle households. The studies used a combination of metrics (e.g., vehicle miles traveled, range) and performed their assessments for different economic scenarios. Assessments from the studies led to the conclusion that EVs are technically and economically better suited for multi-vehicle households (Jakobsson, Gnann, et al., 2016; Musti & Kockelman, 2011; Sherman, 1980; Tamor & Milačić, 2015). Moreover, replacing one of the vehicles in a multi-vehicle instead of replacing the only vehicle in a single-vehicle household was expected to electrify roughly twice as many miles (Tamor & Milačić, 2015). With the introduction of EVs into the market, the projections about acceptance of EVs in multi-vehicle households proved to be accurate. The most recent National Household Travel Survey data (NHTS 2017) shows that the majority of EV-owning households are multi-vehicle households (78.7% for HEVs, 83.4% for PHEVs, and 92.7% for BEVs) (Li et al., 2019).

Given the compatibility of EVs with multi-vehicle households, some recent studies aimed to address the question “How much are EVs driven in multi-vehicle households and what factors influence their annual vehicle miles traveled?”. The studies collectively suggest that there exists a large heterogeneity in the EV driving patterns; some household drive their new EVs more than the replaced car and some drive it less. Interestingly, they often found that the owners adapt their driving behavior (e.g., take alternative routes) to suit the new EV in their household (Jakobsson et al., 2022; Jakobsson, Karlsson, et al., 2016). Moreover, PHEVs were found to have higher
total vehicle miles traveled (VMT) and a higher share of household VMT compared to BEVs (Chakraborty et al., 2022; Mandev et al., 2022; Srinivasa Raghavan & Tal, 2021). The studies underscore the importance of charging infrastructure (especially level 2 charging at home) as a determinant of eVMT (Electrified Vehicle Miles Travelled) in multi-vehicle households apart from traditional factors (e.g., population density, attitudes towards technology, and lifestyle preferences) (Chakraborty et al., 2022; Mandev et al., 2022; Srinivasa Raghavan & Tal, 2021).

Unlike the studies on eVMT, there exists a significant shortage of studies exploring EV usage (in multi-vehicle households) through the lens of vehicle choice/preference. To the best of our knowledge, there have been no studies on EV preference in US multi-vehicle households (with at least one CV and one EV). The only studies exploring this topic were conducted in Europe, where the driving needs are very different from the US. A Danish study found that the number of trip legs, the drivetime, requirement to charge the vehicle all had negative effects on the probability of choosing EV. On the other hand, precipitation, urban areas had positive effects on the choice (Jensen & Mabit, 2015). A study in Switzerland modeled household vehicle choice using trip attributes, socio-demographic variables, and spatiotemporal variables (Bucher et al., 2020). After a comparison of different vehicle choice models, they suggest that trip duration, trip distance, weekend (indicator) are some of the most important determinants, negatively influencing EV preference. However, they conclude that the choice cannot be predicted easily by the variables considered in the study.

This study aims to expand research on vehicle choice in US multi-vehicle households, specifically households with at least one EV and one CV. The choice holds significance because it has cascading effects on the gasoline VMT (Srinivasa Raghavan & Tal, 2021) and carbon emissions from CVs within multi-vehicle households. Furthermore, gaining a deeper
understanding of this choice can assist policymakers and manufacturers in understanding user preferences and designing strategies to boost EV usage.

This study also addresses three limitations of the previous studies (Bucher et al., 2020; Jensen & Mabit, 2015) on vehicle choice in multi-vehicle households. Firstly, this study incorporates a clustering model to capture the heterogeneity in trips (overlooked in previous studies). To highlight the significance of capturing heterogeneity in trips, consider the following example: Two trips covering the same distance may be more appropriately regarded as different types of trips because they differ on many other attributes (Ozhegov & Ozhegova, 2019). Conversely, two trips with different distances may be more appropriately regarded as similar due to many other shared attributes. A clustering model would be able to capture this heterogeneity by grouping similar trips based on a number of attributes (instead of only one attribute), something that a simple discrete-choice model fails to do. Secondly, this study utilizes the NHTS 2017 datasets, which provide information on the natural adopters of EVs. This is a representative sample of EV adopters compared to the previous studies, which provided EVs to households solely for research purposes. Thirdly, this study employs interpretable machine learning techniques to capture the non-linear relationships between explanatory variables and EV preference. Although one previous study hinted at the existence of such non-linear effects (Bucher et al., 2020), none of them reported or discussed these effects. By capturing non-linearity, this study aims to unravel the intricacies of these relationships. With the aforementioned research gaps in mind, this study seeks to answer the following questions:

1) What type of trips are most likely to be made by EVs in US multi-vehicle households?

2) How do different socio-demographic and built-environment variables influence the preference for EVs for individual trips in US multi-vehicle households?
The rest of the chapter is organized as follows. The “Data” section describes the dataset used in this study and the data preprocessing steps. The “Methodology” section contains an overview of the models used in this study as well as the metrics used to evaluate and interpret them. The last two sections of the chapter are “Results and Discussion” and “Conclusion”, which present this study's findings, implications, and concluding statements. The rest of this chapter uses the terms “EV choice” and “EV Preference” interchangeably.

1.2. Data

1.2.1. NHTS 2017

This study uses the 2017 National Household Travel Survey (NHTS) datasets, which contain travel information for US residents in all 50 states and the District of Columbia (U.S. Department of Transportation, 2017). These surveys employ professional processing procedures to capture travel behavior and its seasonal variation over 12 months (Liu et al., 2019; U.S. Department of Transportation, 2017). The 2017 NHTS consists of 4 datasets namely, household, vehicle, person, and trip datasets.

1.2.2. Data Cleaning

The NHTS trip dataset is an inventory of all trips taken within a specified 24-hour period by household members older than 5. It contains 923,572 trips made by 117,222 households. Among them, 81,913 (69.88% of all households) households owned multiple vehicles. They made a total of 589,750 trips (63.86% of all trips). This study deals with households owning at least one CV and at least one EV. Consequently, only the trips made by these households were considered. Trips with non-household vehicles and missing/unknown values for important variables were discarded. Additionally, trips with unrealistically high average speeds (e.g., 1495 mph) were removed by looking up the top speed of the fastest model from each vehicle manufacturer (J.
Lastly, return trips (trips ending at home) were excluded to focus solely on outgoing trips. This step is justified as household members typically use the same vehicle when returning home; there is no choice involved with return trips.

Figure 1-1: The percentages of trips made by different vehicle types and fuel types (top) and the number of trips from 50 states in the US (bottom)

The cleaned dataset contained a total of 19,825 trips made by 3917 households. Figure 1-1 (top) shows the proportion of EV trips and CV trips in the dataset. It can be observed that the cleaned
dataset is balanced with regard to the proportion of EV trips (49.5%) and CV trips (50.5%), which is conducive to the performance of machine learning models (Jia, 2019). Among the EV trips, most of the trips were made by conventional hybrid vehicles. Figure 1-1 (bottom) contains a symbol map showing the spatial distribution of trips across the different states. Major shares of the trips are from the states highlighted with blue and orange symbols (California, Texas, New York, Wisconsin, and North Carolina).

1.2.3. Variable Selection

The variable selection for this study was informed by existing studies on EVs and multicollinearity assessment. Since the study intends to identify the type of trips that are most likely to be made by EVs in multi-vehicle households, individual trip attributes (e.g., trip purpose, trip distance, number of passengers) were considered in the modeling framework. Trip attributes have been previously used in studies that modeled EV usage (Bucher et al., 2020; Jensen & Mabit, 2015). In addition to trip attributes, the framework also included household attributes (Chakraborty et al., 2022; Jia, 2019; Liu et al., 2019; Shin et al., 2019), person attributes (Bucher et al., 2020; Li et al., 2019), built-environment variables (Musti & Kockelman, 2011; Vergis & Chen, 2014) and gas price (De Borger et al., 2016), which have been commonly linked to EV usage and adoption. Based on previous studies, an initial list of variables was produced. Variables displaying high multicollinearity (indicated by variance inflation factors over 4) were excluded from the modeling framework. Table 1-1 and Table 1-2 present the final subset of variables that were included in the modeling framework.

1.3. Methodology

This study used a combination of two machine-learning models to predict household vehicle choice (Figure 1-2). The first model (clustering model) captured the heterogeneity in trips by
clustering them based on trip attributes. The second model (classification model) predicted vehicle choice. It was hypothesized that individuals in multi-vehicle households choose between their electric and conventional vehicles based on the type of trip they are making. Hence, the classification model (which predicted vehicle choice) accepted the trip cluster from the clustering model as an input variable, along with household attributes, driver attributes, gas price, and built-environment variables. Apart from vehicle choice prediction, the classification model was also used to interpret the effects of the input variables. The following subsections describe the two models in detail.

Table 1-1: The descriptive statistics of the variables used in this study

<table>
<thead>
<tr>
<th>Variable</th>
<th>Categories</th>
<th>Distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Household Attributes</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Home Ownership (Discrete)</td>
<td>Rent's a House</td>
<td>91.26%</td>
</tr>
<tr>
<td></td>
<td>Owns a House</td>
<td>8.74%</td>
</tr>
<tr>
<td>Household Income (Discrete)</td>
<td>Low (&lt;$50,000)</td>
<td>9.73%</td>
</tr>
<tr>
<td></td>
<td>Medium ($50,000 - $150,000)</td>
<td>57.24%</td>
</tr>
<tr>
<td></td>
<td>High (&gt; $150,000)</td>
<td>33.03%</td>
</tr>
<tr>
<td>Number of Household Workers</td>
<td>0</td>
<td>17.53%</td>
</tr>
<tr>
<td>(Discrete)</td>
<td>1</td>
<td>25.61%</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>45.92%</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>8.49%</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>2.12%</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.33%</td>
</tr>
<tr>
<td>Children (Discrete)</td>
<td>No children</td>
<td>90.42%</td>
</tr>
<tr>
<td></td>
<td>1 or more children</td>
<td>9.58%</td>
</tr>
<tr>
<td><strong>Driver Attributes</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Driver's Age (continuous)</td>
<td></td>
<td>52.49 ± 15.65*</td>
</tr>
<tr>
<td>Driver's Education Level (Discrete)</td>
<td>No high school degree</td>
<td>8.64%</td>
</tr>
<tr>
<td></td>
<td>High school or associate degree</td>
<td>19.31%</td>
</tr>
<tr>
<td></td>
<td>Bachelor's degree or higher</td>
<td>72.06%</td>
</tr>
<tr>
<td>Driver's Sex (Discrete)</td>
<td>Female</td>
<td>43.54%</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>56.44%</td>
</tr>
<tr>
<td><strong>Built Environment Variables</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Employment Density in workers</td>
<td>-</td>
<td>3.74 ± 4.52*</td>
</tr>
<tr>
<td>per 0.01 square miles (Continuous)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Population Density in persons</td>
<td>-</td>
<td>1.54 ± 1.52*</td>
</tr>
<tr>
<td>per 0.01 square miles (Continuous)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>**Trip Day Gas Price in cents</td>
<td>-</td>
<td>246.86 ±</td>
</tr>
<tr>
<td>per gallon (Continuous)</td>
<td></td>
<td>25.64*</td>
</tr>
</tbody>
</table>

*Note: The distributions of continuous variables are presented as mean ± standard deviation

**Note: The NHTS 2017 contains gas price data on the Petroleum Administration for Defense District (PADD) level
1.3.1. Clustering Model

Since trips can vary based on a number of attributes, a k-modes clustering model (Huang, 1998) was used to cluster the trips. The clustering model maximized the heterogeneity between different trip clusters. The trips were clustered based on seven attributes namely, weekend/weekday trip, starting time, number of passengers, dwelling time at destination, distance, trip purpose, home-based/non-home-based trip. Although k-prototype clustering may make more sense for mixed data (with categorical and continuous variables), it was found to provide an unclear delineation among the clusters for this specific dataset. Hence, the k-modes clustering was used by converting all variables to categorical.

The k-modes clustering analysis in this study consisted of the following steps:

1) A set of observations from the dataset is initialized as the cluster centroids using a density-based initialization algorithm (F. Cao et al., 2009). Then every trip in the dataset is assigned to the nearest trip cluster based on the hamming distance (Pandit & Gupta, 2011) from the cluster centroids.

2) After every trip is assigned to a cluster, the mode of the individual clusters acts as the updated cluster centroids. And then, the observations are reassigned to the closest clusters based on the updated centroids. This step is repeated until no observation in the dataset changes clusters.

Models with clusters ranging from 2 to 10 were evaluated using the elbow plot and Silhouette Score (Rousseeuw, 1987). The final clustering model was chosen as the one with the best evaluation scores and clearest separation among the clusters based on the trip attributes.
1.3.2. Classification Model

The model classifying household vehicle choice (CV or EV) considered variables including trip cluster (from the clustering model), household attributes, driver attributes, built-environment attributes, and trip-day gas price. Four modeling techniques were compared: decision tree, random forest, extreme gradient boosting, and binary logit model. Tree-based machine learning models were picked as candidate models because of their proven success in predicting mode choice (Kim, 2021; Zhao et al., 2020). The dataset was randomly shuffled to produce training (85%) and testing (15%) subsets. Apart from training accuracy, the models were evaluated based on 10-fold-cross-validation accuracy (Zhao et al., 2020) and testing accuracy to ensure maximum generalizability. The most accurate model was selected for interpretation. The following sections provide a general overview of the four classification models evaluated for this study along with the methods used to interpret the most accurate classification model.
1.3.2.1. **Decision Tree**

A decision tree performs classification by producing recursive partitions (or decisions) of the dataset based on the variables (Myles et al., 2004). The decision tree model takes a series of consecutive decisions creating a tree structure and leading to the final predictions. The mathematical formulation can be described as follows:

Given training vectors $x_i \in \mathbb{R}$, $i = 1,2,3,\ldots l$ and a label vector $y \in \mathbb{R}^l$, a decision tree performs partitioning such that observations with the same labels fall in the same group. The data at node $m$ is represented by $Q_m$ where the number of observations is $n_m$. For each candidate split, $\Theta = (j, t_m)$ consisting of a variable $j$ and a threshold value $t_m$, the data is split into $Q_m^{left}(\Theta)$ and $Q_m^{right}(\Theta)$ subsets such that,

$$Q_m^{left}(\theta) = \{(x,y)|x_j < t_m\}$$

$$Q_m^{right}(\theta) = Q_m \setminus Q_m^{left}(\theta)$$

After a split is performed, the quality of the split is assessed based on the gini impurity function, and the candidate split which provides the minimum impurity is selected as the final split for node $m$. In this way, subsets $Q_m^{left}(\Theta)$ and $Q_m^{right}(\Theta)$ are recursively produced until $n_m = 1$. Since the set of variables considered in this study consists of categorical variables, they were converted into dummy variables as suggested by (Breiman, 2001).

1.3.2.2. **Random Forest**

A random forest is an ensemble technique that is applied to decision trees to improve generalization (improve the prediction accuracy for unseen data) (Myles et al., 2004). The technique is referred to as random forest because it combines many randomized decision trees.
Biau & Scornet, 2016. The random forest algorithm for this study and the hyperparameters involved in each step are discussed below:

1) At first, a random subset of the observations is sampled with replacement.

2) Next, a decision tree is grown using the random sample and a set of randomly selected variables. As recommended by previous studies (Liaw & Wiener, 2002), the number of randomly selected variables was set as the square root of the total number of variables.

3) The number of decision trees to be grown is set to 1000; the first two steps were repeated 1000 times.

4) The vehicle choice for a trip is the most common prediction from the decision trees in the forest.

1.3.2.3. **Extreme Gradient Boosting**

Extreme Gradient Boosting model (XG Boost) is another ensemble technique, applied to decision trees, which is based on the gradient boosting model (Friedman, 2001). It grows a sequence of decision trees with low depth and each tree is trained by putting more weight on the incorrect predictions of the preceding trees (Chen & Guestrin, 2016). The technique minimizes a loss function using gradient descent. It works by iteratively adding decision trees to the model, with each new tree attempting to correct the errors made by the previous trees. At each iteration, the algorithm calculates the gradient and the hessian of the loss function with respect to the current model and uses this information to create a new decision tree that minimizes the loss function.

1.3.2.4. **Binary Logit**

This study also employed a binary logit model, which served as a baseline for accuracy comparison. The binary logit has been applied in a wide number of studies to model mode choice
(Bhat, 1997; Yang et al., 2018). The dependent variable of the model $Y_i$ could take either of the values 1 (for EV) or 0 (for CV). The probability that the dependent variable equals 1 for an observation $i$ is given by:

$$
\Pr(Y_i = 1|X_i) = \frac{\exp (X_i \beta)}{1 + \exp (X_i \beta)}
$$

(3)

$X_i$ is a matrix of variables and $\beta$ is a vector of unknown coefficients, estimated via maximum likelihood estimation (MLE). Robust standard errors were used in the process to account for possible heteroskedasticity (Cameron & Trivedi, 2010).

1.3.2.5. **Classification Model Interpretation**

Although machine learning models have been criticized for their lack of interpretability, some interpretation techniques successfully illustrate the importance and effects of the variables in machine learning models (Kim, 2021; Wang & Ross, 2018). Two such methods were used to interpret the best-performing classification model. Firstly, the Gini-impurity-based variable importance (Zhao et al., 2020) was estimated as a measure of the predictive powers of the variables in the model. In addition, the accumulated local effects (ALE) (Apley & Zhu, 2020) were estimated to decipher the marginal effects of the independent variables on EV choice. ALE plots can illustrate any type of relationship (e.g., linear, multi-linear, non-linear) between a variable and the predicted outcome.

1.4. **Results and Discussion**

This section discusses the different models used for clustering trips and predicting vehicle choice. For each purpose, different model specifications are compared, and interpretations are drawn from the best model.
1.4.1. Clustering Model

1.4.1.1. Model Comparison

As mentioned earlier, k-modes clustering models with clusters ranging from 2 to 10 were tested and compared based on the elbow plot, silhouette score, and cluster separation. Figure 1-3 shows the elbow plot and the silhouette scores. An “elbow” (the point with the most significant cost reduction) in the elbow plot and a higher value for the silhouette score indicates the optimal number of clusters. The silhouette score indicates that there are two candidate models: the 2-cluster model and the 5-cluster model. Even though the 2-cluster model ranks higher based on the silhouette score, its cost is also higher in the elbow plot. And it was found that there wasn’t a clear delineation between the 2 clusters based on the trip attributes used for clustering. Hence, based on the elbow plot and upon closer inspection of the cluster separation, the 5-cluster model was selected as the final model.

Figure 1-3: Comparison of k-modes clustering models based on the elbow plot (left) and Silhouette score (right)

1.4.1.2. Model Interpretation

Table 1-2 shows the names of the 5 trip clusters resulting from the final model and the proportions of different trip attributes for the clusters. Every trip cluster had a dominant characteristic (represented by bold typeface in Table 1-2) for each trip attributes.
### Table 1-2: Proportion of different trip attributes for the 5 trip clusters

<table>
<thead>
<tr>
<th>Trip Attribute</th>
<th>Trip Clusters</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cluster 1</td>
</tr>
<tr>
<td></td>
<td>(8060 trips)</td>
</tr>
<tr>
<td>Weekday/Weekend</td>
<td></td>
</tr>
<tr>
<td>Weekday trip</td>
<td>86%</td>
</tr>
<tr>
<td>Weekend trip</td>
<td>14%</td>
</tr>
<tr>
<td>Starting time</td>
<td></td>
</tr>
<tr>
<td>10 AM - 3 PM</td>
<td>60%</td>
</tr>
<tr>
<td>12 AM - 6 AM</td>
<td>1%</td>
</tr>
<tr>
<td>3 PM - 7 PM</td>
<td>18%</td>
</tr>
<tr>
<td>6 AM - 10 AM</td>
<td>15%</td>
</tr>
<tr>
<td>7 PM - 12 AM</td>
<td>6%</td>
</tr>
<tr>
<td>Home-based/non-home based</td>
<td></td>
</tr>
<tr>
<td>Home-based trip</td>
<td>22%</td>
</tr>
<tr>
<td>non-home-based trip</td>
<td>78%</td>
</tr>
<tr>
<td>Trip Purpose</td>
<td></td>
</tr>
<tr>
<td>Errands</td>
<td>19%</td>
</tr>
<tr>
<td>Others</td>
<td>7%</td>
</tr>
<tr>
<td>Shopping or Dining</td>
<td>58%</td>
</tr>
<tr>
<td>Social or recreational</td>
<td>6%</td>
</tr>
<tr>
<td>Work</td>
<td>9%</td>
</tr>
<tr>
<td>Dwelling time</td>
<td></td>
</tr>
<tr>
<td>1-15 minutes</td>
<td>51%</td>
</tr>
<tr>
<td>15-50 minutes</td>
<td>29%</td>
</tr>
<tr>
<td>50-150 minutes</td>
<td>14%</td>
</tr>
<tr>
<td>More than 150 minutes</td>
<td>5%</td>
</tr>
<tr>
<td>Trip Distance</td>
<td></td>
</tr>
<tr>
<td>0-2 miles</td>
<td>33%</td>
</tr>
<tr>
<td>2-5 miles</td>
<td>43%</td>
</tr>
<tr>
<td>5-15 miles</td>
<td>15%</td>
</tr>
<tr>
<td>More than 15 miles</td>
<td>9%</td>
</tr>
<tr>
<td>Number of Passengers</td>
<td></td>
</tr>
<tr>
<td>1 passenger</td>
<td>67%</td>
</tr>
<tr>
<td>2-4 passengers</td>
<td>32%</td>
</tr>
<tr>
<td>5 - 10 passengers</td>
<td>2%</td>
</tr>
</tbody>
</table>

**Note:** The bold typeface indicates the dominant characteristic of the corresponding cluster

For instance, in trip cluster 1, 58% of trips are made for shopping or dining purposes, occurring mainly on weekdays (86%). These trips are typically short in duration (51% lasting 1-15 minutes) and cover short distances (43% between 2-5 miles).

Trip cluster 2 is primarily made up of work trips (63%), typically starting between 6 AM and 10 AM (75%). As expected from work trips, these trips often involve longer durations at the destination (over 150 minutes in 66% of cases).

Trip cluster 3 contains trips that are mostly made to run some errands 47% of the time. They are usually home-based trips (76%) that take place during the weekdays (88%).

---

22
Trip cluster 4 is predominantly made up of social or recreational trips (51%). As expected for social and recreational trips, these trips usually have 2-4 passengers (86%) and involve spending 50-150 minutes at the destination in 60% of the cases. The majority of the trips in this cluster are greater than 15 miles (38%).

Trip cluster 5, similar to Cluster 1, also has shopping or dining as the dominant trip purpose. However, there are some key distinctions. Unlike Cluster 1, the trips in Cluster 5 mostly have 2-4 passengers (86%) and they are primarily made during the weekends (80%). Since people have more time to spare during the weekends, these trips involve spending a longer time (50-150 minutes) at the destination in 52% of the cases.

1.4.2. Classification Model

1.4.2.1. Model Comparison

The classification models were evaluated based on cross-validation accuracy, training accuracy, and testing accuracy (Figure 1-4). Based on these metrics, the decision tree demonstrated superior performance. While the machine learning models had similar accuracies, the binary logit model showed a noticeable drop in accuracy. This supports the preference for machine learning over traditional discrete choice models and possibly indicates the presence of non-linear relationships. For this specific study, ensemble techniques (random forest and XG boost) did not improve prediction accuracy compared to the decision tree. However, in studies with more complex relationships or higher-dimensional datasets, ensemble techniques may outperform the decision tree. The following section focuses on interpreting the best-performing machine learning model (decision tree), confirming non-linear relationships, and providing insights into their nature.
1.4.2.2. Model Interpretation

1.4.2.2.1. Variable Importance

The variable importance generally represents the predictive power of a variable in the model. The importance of the variables used in the decision tree is shown in Figure 1-5. Upon initial inspection, it can be noticed that the continuous variables (trip day gas price, driver’s age, employment density, and population density) contribute more to the predictive power of the model as compared to the discrete variables. The trip day gas price (0.331) is the most important continuous predictor in the model. Among the discrete variables, the trip cluster (0.073) was found to be the most important, which underscores the importance of clustering the trips in the modeling framework.
Figure 1-5: The variables in the decision tree model arranged in ascending order of variable importance

Among the driver attributes, the age (0.252) of the driver was significantly more important than the sex and education level of the driver in predicting the vehicle choice of a driver in a multi-vehicle household. Among the household attributes, the number of household workers (0.054) was found to be the most important. While variable importance offers insights into the contribution of each variable to the model's accuracy, it is important to note that it does not provide information about the magnitude and direction of their effects on vehicle choice.

1.4.2.2. Accumulated Local Effects

Accumulated local effects (ALE) allow for the estimation of the marginal effects of variables on vehicle choice. ALE is the main effect of a variable at a specific value, relative to the average prediction value of the data. Through this method, complex non-linear relationships between variables can be captured. Figure 1-6 and Figure 1-7 show the ALE plots of each of the variables in the model.
Figure 1-6: The accumulated local effect (ALE) plots of the trip cluster and household attributes for the decision tree model. The black dots (joined by the broken lines or continuous lines) indicate the effect of the variables at a specific value. The light blue bars indicate number of observations for a specific value of a variable.

From the ALE plot of trip clusters, we can see that the first three trip clusters (Weekday shopping and dining trips, work trips, and errand trips) have a negative effect on EV choice probability, implying that multi-vehicle households prefer their CVs for making a trip from these trip clusters. On the other hand, trip Clusters 4 (social and recreational trips) and 5 (weekend shopping and dining trips) have positive effects (ALE of 0.009 and 0.037 respectively) on the
Figure 1-7: The accumulated local effect (ALE) plots of driver attributes, gas price and built environment variables for the decision tree model. The black dots (joined by the broken lines or continuous lines) indicate the effect of the variables at a specific value. The light blue bars indicate number of observations for a specific value of a variable. Choice probability of EV. This trip-making behavior has some important implications. Firstly, it implies that multi-vehicle households prefer to reserve their EVs for trips that are less frequent, less time-sensitive and discretionary in nature. For instance, both trip clusters 4 and 5 (which have positive effects on EV choice probability) include carpooling (2-4 passengers) trips. These trips are less common in the US compared to the single occupancy trips observed in cluster 1,2, and 3; the average car occupancy in 2017 was 1.5 passengers per vehicle (U.S. Department of
Energy (DOE) Oak Ridge National Laboratory, 2022). Additionally, trip cluster 5 (which has the largest positive effect on EV choice probability) primarily serve the discretionary purposes of shopping and dining (Table 1-2). Although trip cluster 1 also predominantly serve shopping and dining purposes, it has a negative effect on EV choice probability. This may be because trips in cluster 1 are more time-sensitive in nature (mostly involve spending 1-15 minutes at the destination compared to 50-150 minutes for trip cluster 5). Moreover, trip cluster 5 (which has the largest positive effect on EV choice probability) are made during the weekends 80% of the time (Table 1-2). This finding aligns well with a Swedish study on two-car households (Karlsson, 2020), which found that the weekend driving distance of EVs is 80% greater than CVs during the weekends. The discretionary nature of EV trips can be partly explained by the disparity between the charging time for EVs and CVs (at least 30 minutes for HEVs compared to 5 minutes for a CV) (K. V. Singh et al., 2019). Moreover, EVs are usually associated with higher insurance payments (Parker et al., 2021) and greater sensitivity to external environments. Given the disparity in charging time, sensitivity and higher insurance payments, multi-vehicle households might prefer to reserve their EVs for trips that are less frequent, less time-sensitive and discretionary in nature. Secondly, the trip-making behavior also has some implications on recharging/refueling station locations. If government agencies and other stakeholders want to keep station locations consistent with trip-making behavior, then there should be active efforts to place more charging stations near discretionary trip attractors (e.g., shopping malls, restaurants). On the other hand, if government agencies intends to change people’s behavior and encourage more frequent trips from EVs, placing stations near frequent trips attractors could be a possible course of action.
As the number of household workers goes up, the choice probability of EV goes down steadily. A household with a higher number of workers is going to make more trips from trip cluster 2 (predominantly work trips). Since trip cluster 2 had a negative effect (ALE = -0.005) on EV choice probability, it makes sense for the number of household workers to have a negative effect as well.

From the ALE plot of household income, it is evident that households falling under the low-income (ALE = 0.0057) and medium-income categories (ALE = 0.0058) are more likely to choose their EV for a trip than households in the high-income category (ALE = 0.0155). Two key observations deserve attention here. Firstly, high-income households are likely to have a higher number of workers. Since a higher number of workers is negatively associated with EV choice probability, it is logical that higher income will also have a negative effect. Secondly, lower-income households might use EVs more to save money. Previous studies have shown that EVs can be expected to achieve cost-effectiveness in multi-vehicle households within six years (Jakobsson, Gnann, et al., 2016) as they make up for their high purchase price through their fuel efficiency. Hence, low, and medium-income households may find their EVs to be a cost-effective alternative to CVs. Given the effects of the number of household workers and income, policies, and incentives aiming to increase EV usage should focus on households with a lower income and a lower number of workers. Currently, households with a higher income are more likely to own EVs (Liu et al., 2019). If incentives reduce the purchase price of EVs for lower-income households, we can expect more EV usage.

The ALE plot of children indicates that households with one or more children prefer CVs and households with no children prefer EVs. This could be true because households with children are likely to have more members, thereby needing higher occupancy vehicles (e.g.,
Most of these higher occupancy vehicles in the NHTS dataset are CVs, which explains their preference for CVs over EVs. Hence households with children may represent a segment of the market that EV manufacturers still haven’t tapped into.

The ALE plot indicates a multi-linear relationship between driver age and EV choice. The probability of choosing an EV increases linearly with age until 30. However, the relationship becomes unclear after 30. Despite age being the second most important predictor, its marginal effect above 30 is indiscernible. Geographical variation of the effect of age on EV adoption was previously found within the US (Liu et al., 2019). A similar variation may explain why the national sample couldn't capture a clear relationship between EV choice and age beyond 30.

As indicated by the ALE plot of the driver’s education level, the probability of a driver choosing the EV is higher when he/she has a bachelor’s degree or higher education (ALE=0.03). This is consistent with the findings from studies on EV adoption which suggest that education level has a positive effect on adoption (Liu et al., 2019). A higher level of education is generally associated with a higher concern for the environment, which might cause people to choose their EVs.

Trip-day gas price, similar to age, has a multi-linear relationship with EV choice probability. The ALE plot indicates that the gas price does not have a clear effect on EV choice as long as the gas price is below 2.80 USD per gallon. However, when the gas price goes above 2.80 USD per gallon, we start to notice a clear positive effect on EV choice probability. Previous studies have suggested making conventional fuels more expensive as a strategy to promote EV use (Parker et al., 2021). The threshold of $2.80 for gas prices can be used to implement such strategies. However, this threshold has to be adjusted for inflation since the data for this study is
from 2017. Nevertheless, this valuable information could not have been extracted from a discrete-choice model that assumes a linear relationship.

The ALE plots of employment density and population density show contrasting effects on vehicle choice. The marginal effects of these built environment variables do not show a steady increase or decrease. Generally, higher employment density values show a positive effect on EV choice probability, while higher population density values show a negative impact. Population density was also found to have a negative effect on the VMT of PEVs in multi-vehicle households (Chakraborty et al., 2022). These negative effects of population density might be explained by the higher number of publicly available charging stations in suburban areas compared to urban areas (Brown et al., 2022).

1.5. Conclusion

This study investigates the factors that influence vehicle choice for trips in multi-vehicle households in the US, specifically those with at least one EV and one CV. A two-step machine learning modeling framework was utilized, starting with k-modes clustering to identify 5 distinct trip clusters and capture the heterogeneity in trips. Subsequently, a decision tree model was employed to predict vehicle choice (EV or CV). A comparison of four different modeling approaches was performed before the decision tree was chosen as the final model in the modeling framework. The comparison of the models revealed that the decision tree (cross-validation accuracy of 88%) outperformed the binary logit (cross-validation accuracy of 57.7%) by a large margin. Notably, trip day gas price and the driver's age were found to be major contributors to the decision tree's predictive power. Both of these variables had non-linear effects on EV preference, which the binary logit model would be unable to capture. This further
underscores the significance of employing machine learning approaches, such as decision trees, within the context of this study.

ALE plots were produced to analyze the effects of different variables on vehicle choice, as captured by the decision tree model. The analysis revealed that weekend trips primarily intended for shopping and dining were most likely to be made using EVs, indicating the discretionary use of EVs within multi-vehicle households. This underscores the significance of having charging stations located near popular shopping and dining destinations. Other factors such as the number of household workers, income, and trip day gas price also exhibited discernible effects on vehicle choice. Overall, multi-vehicle households with lower incomes and fewer workers were more inclined to choose EVs for their daily trips. However, these households face challenges in adopting EVs initially due to the higher purchase prices. To promote higher EV usage, targeted incentives should be implemented to make EVs more affordable for these households. Additionally, gas prices exceeding $2.80 per gallon were found to discourage the use of CVs within multi-vehicle households, suggesting that gas prices can serve as a tool to increase EV usage.

It is important to acknowledge a few limitations of this study. Due to the limited number of trips made by PHEVs and BEVs, EVs were treated as a single category, potentially overlooking differences between vehicle types. Furthermore, the NHTS 2017 dataset predominantly includes older EV models, which are gradually being replaced by newer models with extended ranges. Future studies on EV usage can leverage datasets that encompass trip data from newer EV models, enabling the modeling of vehicle choices for different EV categories separately. Furthermore, future studies can explore the lagged/delayed effect of gas price in addition to the effect of trip day gas price explored in this study.
References


2. Assessing Behavioral Factors in Low-Carbon Transportation Adoption Among Heavy-Duty and Off-Road Transportation Sectors in California: A Generative-AI-Assisted Content Analysis

Abstract

The transportation sector has been identified as the largest contributor to greenhouse gas emissions in California. This produces a pressing need among government agencies in California to facilitate the adoption of low-carbon transportation (LCT) among the heavy-duty vehicles (HDV) and off-road equipment (ORE) sectors. By performing content analysis, this study explores the influence of behavioral factors (awareness and impression) on LCT adoption within these sectors. Moreover, the varying importance of the factors (behavioral and other) of LCT adoption within different types of organizations were assessed and discrepancies between available and expected government support for LCT adoption were identified. The results of the analysis were refined using a generative artificial intelligence (AI) tool. Among the behavioral factors, positive associations were found between the potential to adopt LCT and awareness of LCT as well as aspects linked to LCT (government incentives, environment). Regarding the importance of other factors, subtle differences were noticed among different types of organizations (with respect to adoption behavior, vocation, and fleet size). For instance, HDV organizations were mostly driven by mandates, and ORE organizations were mostly driven by environmental concern and green public image. Compared to large-fleet organizations, financial barriers, and incentives held greater importance among small-fleet organizations. The findings can serve as a guideline for government agencies to develop tailored mandates and incentives for different types of organizations in HDV and ORE sectors. Moreover, the use of generative AI
revealed some of its possible contributions and shortcomings in qualitative research that merit consideration.
2.1. **Introduction**

The greatest hazard to public health in the twenty-first century is greenhouse gas emissions which lead to global warming and climate change (A. Singh & Purohit, 2014). California is the second largest greenhouse gas emitter in the U.S. (P. Perez et al., 2020), with its transportation sector accounting for the highest share of emissions. To comply with federal air quality regulations and climate change targets, the State of California has taken substantial steps, in the form of mandates, to convert medium-duty (MD) and heavy-duty (HD) trucks to low-carbon transportation alternatives. (Gordon et al., 2022). Low-carbon transportation modes are defined by their minimal energy consumption, which supports sustainable urban development and greenhouse gas reduction (primarily CO2) (Y. Cao et al., 2023). The California Air Resources Board (CARB) developed several models to cut emissions by mandating low-carbon transportation (LCT) adoption in specific sectors. To facilitate compliance with such mandates in the on-road sector, the Hybrid and Zero-Emission Truck and Bus Voucher Incentive Project (HVIP) was established by CARB to provide vouchers that lower the upfront cost of clean trucks (California Air Resources Board, 2023). Similarly, to encourage the off-road equipment (ORE) sector to adopt zero-emission vehicles, several incentive and regulatory programs were introduced, such as the Clean Off-Road Equipment (CORE) Voucher Incentive Project (Gordon et al., 2022), Agricultural Replacement Measures for Emission Reductions (FARMER) Program (Mccullough et al., 2021), In-Use Off-Road Diesel-Fueled Fleets Regulation (Huang & Fan, 2022), etc.

Given the emerging importance of LCT in recent years, several studies in Europe sought to analyze the determinants of alternative fuel vehicle (AFV) adoption in HDV sectors. Performing a choice experiment in Switzerland and Germany, (Walter et al., 2012) discovered
that two financial determinants namely, the vehicle purchase price, and operating costs, had the greatest impact on the decision to purchase. Another German study (Seitz et al., 2015), found that corporate social responsibility with environmental attitudes can also have a profound influence on the choice of CO$_2$-reducing powertrain technologies. By employing a Delphi study, (Anderhofstadt & Spinler, 2019) found that the availability of fueling or charging infrastructure, the ability to enter low-emission zones, current and projected fuel costs, are crucial considerations while adopting alternative fuel-powered HDTs.

Some studies in the US also sought to identify the determinants of AFV adoption in HDV sectors. A study from 2014 found three motivating rationales namely, first-mover advantage, specialized functional capabilities, and a desire to build an appealing business brand (Sierzchula, 2014). By conducting a series of in-depth qualitative interviews in California, USA, a more recent study (Bae et al., 2022) detected 38 motivators or barriers (including functional suitability, monetary costs, fuel infrastructures, and reliability/safety of the vehicles and engines) of AFV adoption in the HDV sectors.

Since past research has consistently identified cost components as important determinants of LCT adoption, CARB introduced several monetary incentives for electric vehicle fleets, many in the form of vouchers for MD and HD sectors (Burke & Miller, 2020; Jin et al., 2014). These vouchers level the playing field of LCT with conventional technologies in terms of cost. However, due to a lack of infrastructural or financial backing, many fleet operators may be ignorant or hold incorrect impressions about LCT and their incentive programs. A noticeable knowledge gap exists within the literature concerning these behavioral factors (awareness and impression of LCT and other aspects linked to LCT), which may determine the adoption of new technologies such as LCT (Frambach & Schillewaert, 2002).
Moreover, the HDV sector is complex and very heterogeneous (wide variety of applications, fleet sizes, engine configurations, and duty cycles). It has many stakeholders making decisions having different rational choices (Winebrake et al., 2012). Therefore, it is challenging to create a standard or set of mandates and incentives for the sector. Further research in this area is needed to confirm and explore the variation of factors influencing LCT adoption for different fleet sizes (small vs large) and hauling types (short haul vs long haul) in the HDV sector.

Although the existing body of literature on the barriers and opportunities of electrifying on-road vehicles has witnessed noticeable growth, a parallel endeavor in the ORE sector appears to be lacking. In fact, it’s a greater challenge to control emissions in the ORE sector, attributable to a variety of reasons (Hall et al., 2018). Due to the cross-boundary nature of aviation, maritime, and rail as well as the widespread use of off-road construction and agricultural equipment, calculating the precise emission consequences is more challenging. Moreover, (Hall et al., 2018) indicated that government regulation of off-road land vehicles is uneven because of the wide variety of vehicles in the sector, slow vehicle turnover, and operating models that frequently include leases and rentals. Nevertheless, California has been at the forefront of agricultural regulations, which is unique in providing millions of dollars in incentives to encourage growers to prepare for possible mandatory air quality implementation plans (Mccullough et al., 2021).

The major contribution of this study is twofold. The first contribution is contextual as the study fills a key knowledge gap on the association between behavioral factors and LCT adoption in HDV and ORE sectors. The research collects data from semi-structured interviews and analyzes the data using qualitative content analysis. The second contribution is methodological as the study incorporates the assistance of generative artificial intelligence (AI) to refine the
analysis and uncover potential contributions/shortcomings of generative AI technologies in qualitative research.

The study makes additional contributions to the literature by exploring the variation of factors influencing LCT adoption among organizations (with respect to their adoption behavior, vocation, and fleet size) and identifying possible discrepancies between available and expected government support for LCT adoption.

The rest of the chapter is organized as follows. The data collection steps are outlined in the “Data” section. The “Methodology” section outlines the methods employed and the results are discussed in the “Results and Discussion” section. The “Conclusion” section presents a summary of the findings, recommendations, and future work.

2.2. Data

2.2.1. Indexing and Recruiting

An index of HDV and ORE organizations in California was prepared from the website of Dun & Bradstreet (Dun & Bradstreet, 2022) and cross-referenced with FMCSA Company Snapshot (U.S. Department of Transportation, 2022). Using these resources, 74 HDV (general freight trucking, public transit, and waste collection) and 56 ORE (construction & demolition, farming, and landscaping) organizations were added to the index. To recruit participants in the study, indexed organizations were contacted through five rounds of emailing and phone calls. Representatives from 12 organizations (8 HDV and 4 ORE organizations) consented to participate in the study.

Given the qualitative nature of this study, collecting a statistically representative sample was not the intention. However, a sample size (12 participants) large enough to produce data saturation was collected (Boddy, 2016). Apart from the size of the sample, the participants
exhibited as much variability as possible in terms of adoption behavior, vocation, and fleet size (Figure 2-1). Heavy-duty fleets covering less than 250 miles were considered short-haul fleets and those covering more than 250 miles were considered long-haul fleets (Indeed, 2023). Small and large fleet organizations were distinguished by the presence of 10 or more vehicles/equipment in the fleet (Collins, 2021).

Figure 2-1: Adoption behavior, vocation, and fleet size of the organizations interviewed

2.2.2. Semi-structured Interviews

The recruited participants took part in semi-structured phone interviews. Two sets of questions were prepared, one for HDV and another for ORE organizations. Confidentiality of the information, disclosed during interviews, was maintained according to the Institutional Review Board’s guidelines. Participation was voluntary, and participants could skip questions if desired.
The interviewees were free to express their opinions. Therefore, some interviews were up to 1 hour long while some lasted 25 minutes.

2.3. **Methodology**

Since this study intends to understand the influence of behavioral factors on LCT adoption, a qualitative analysis approach has been employed in the study. A qualitative assessment of the semi-structured interviews would be able to provide a more nuanced understanding of the experiences of the interviewees, compared to quantitative assessments (Carduff et al., 2015). Additionally, since there is a gap in literature with regard to behavioral factors of LCT adoption, a qualitative study would provide preliminary knowledge to inform future quantitative studies and allow them to select relevant features/variables.

2.3.1. **Content Analysis**

In this study, the interviews were summarized using qualitative content analysis. In content analysis, data is analyzed within a context, considering the attributed meaning (Krippendorff, 1989). The analysis was manually conducted by two researchers and then further refined using the assistance of a generative AI tool, as discussed in the following sub-sections.

2.3.2. **Manual Coding, Segmentation, and Analysis**

The first step in content analysis involves building a coding framework to structure the material for analysis (Schreier, 2012). Coding frameworks are made up of main categories and subcategories. The main categories are aspects that the researchers are interested in, and the subcategories capture what is said about the aspects of interest (Schreier, 2012). The initial coding framework for this study was built manually using a mix of concept-driven and data-driven strategies (Figure 2-2) (Schreier, 2012). The main categories were directly translated from the research questions i.e., they were concept driven. For example, if one of the research
questions was “What are the barriers to LCT adoption in the heavy-duty vehicle sector?”, then a main category named “Barriers to LCT Adoption” was created. The data-driven part of the coding framework comprised of subcategories, generated from the transcribed interviews using the “subsumption” strategy (Schreier, 2012).

Before using the coding framework to structure the material, the interviews were segmented into units of analysis, context units, and coding units (Schreier, 2012) using ATLAS.ti. The interviews themselves were selected as the units of analysis in the study. Since the interviews in the study were semi-structured, the interviewees sometimes made a point using only a word, and sometimes, using several sentences. Hence, changes in topics signaled the end of one coding unit and the beginning of another (Schreier, 2012; Stemler, 2000). The larger body of text around the coding units was selected as context units, which helped in their interpretation (Prasad, 2008).

After segmentation, the manual analysis proceeded to the coding phase, where the coding units were manually assigned to one or more lowest-level sub-categories using ATLAS.ti. A pilot coding phase was conducted to confirm that the framework fulfilled four prerequisite conditions namely, one-dimensionality, mutual exclusiveness, saturation, and exhaustiveness (Schreier, 2012).
In the final steps, statements linked to subcategories were evaluated, considering the importance or sentiments expressed by the interviewees on various topics. Two raters independently assessed the statements, and a data abstraction sheet was used to record importance/sentiment ratings for each subcategory (for every interviewee). Inter-rater agreeability was calculated using Cohen’s Kappa (Cohen, 1960), yielding a score of 0.48 on a scale of 0 to 1. The value indicates that the agreement between the two raters was moderate (Stemler, 2000). Since the interviews were semi-structured, many of the statements didn’t have standardized meanings, the moderate level of agreement was deemed reasonable (Neuendorf et al., 2017). However, all the disagreements were settled through a follow-up discussion and both coders agreed on the final interpretation (Schreier, 2012).

2.3.3. Generative AI assistance

AI has been utilized in qualitative content analysis for studies spanning various disciplines such as psychology (McNamara et al., 2019), marketing (Lee et al., 2020), and healthcare (Lennon et al., 2021). These studies collectively emphasize the efficiency gains enabled by AI in content analysis and demonstrate its ability to support humans in analyzing vast
quantities of text featuring natural language. Hence, to further refine the initial coding framework and the results from the manual analysis, the AI coding feature on ATLAS.Ti was used. This feature capitalizes on the latest breakthroughs in generative AI, offered by the introduction of generative pre-trained transformer (GPT) language models (Gillioz et al., 2020). The feature is designed to generate codes and assign the codes to relevant segments of the material being studied. For this study, it was used as an assistive tool instead of a standalone approach. The aim of using the tool in this study was to uncover new insights, potentially overlooked by human coders. Additionally, by using the tool, the study intended to uncover potential contributions and shortcomings of generative AI in qualitative research.

To get the best results out of AI coding, timestamps, and incomplete sentences were removed, grammatical errors were corrected. Unlike manual coding, where the change of topics marked the boundary of coding units, the AI coding feature on ATLAS.Ti could only assign codes to individual paragraphs. Hence, consecutive statements discussing the same topic were kept in one paragraph.

After analyzing the interviews, AI identified 399 codes and assigned them to the paragraphs from the interviewees. Since the AI served as an assistive tool, only the codes capturing novel aspects (overlooked by the initial coding framework) pertaining to the research questions were deemed relevant and worthy of further analysis. Among the 399 codes, 18 were relevant and the remaining 381 were irrelevant. Some of the relevant codes with related meanings were aggregated, reducing the number of relevant codes to 11. All of the relevant codes were added to the initial coding framework as a main category (1 code out of 11) or a sub-category (10 codes out of 11). Some of the names of the sub-categories in the original coding framework were changed to the more representative names provided by AI.
The final coding framework consisted of 13 main categories and one or more levels of subcategories under each main category. The importance/sentiment ratings were assigned for the AI-generated categories, and Cohen’s Kappa was recalculated as discussed in the previous section. The value of Cohen’s Kappa increased slightly (from 0.48 to 0.5) for the final coding framework.

2.4. Results and Discussion

In the final phase of the content analysis, the ratings for awareness levels were denoted on a scale of 0 to 3 (0=No awareness, 1=low awareness, 2=moderate awareness, 3=high awareness), and the ratings for impression were denoted on a scale of -2 to +2 (-2=highly negative, -1=somewhat negative, 0=neutral, +1=somewhat positive, +2=highly positive) (Murdoch et al., 2019). The importance/emphasis placed on all other subcategories had a scale of 1 to 3 (0=not stated, 1=implied, 2=explicitly stated, 3=emphasized) (Carley, 1993). The results from the content analysis were aggregated and compared based on LCT adoption behavior (3 adopting organizations vs 9 non-adopting organizations), vocation (8 HDV organizations vs 4 ORE organizations), hauling type (3 long-haul organizations vs 3 short-haul organizations vs 2 mixed-haul organizations), and fleet size (8 small fleet organizations vs 4 large fleet organizations) (Table 2-1 and Table 2-2). The following sub-sections present the results of the content analysis and their implications.

2.4.1. Behavioral Factors of LCT Adoption

As previously discussed, behavioral factors (awareness and impression) may play an important role in the adoption of new technologies such as LCT. The following sub-sections discuss the association of these behavioral factors with LCT adoption and how the factors vary with respect to adoption behavior, vocation, and fleet size.
2.4.1.1. **Awareness and Impression of LCT**

On a scale of 0 to 3, the average awareness of LCT among the interviewed organizations was 2.5 (Figure 2-3). However, subtle differences in awareness levels were observed among different groups of organizations with respect to adoption behavior, vocation, and fleet size (Table 2-1). Compared to large ORE companies and large mixed-haul HDV firms, smaller long-haul and short-haul trucking companies in the HDV sector showed a lower level of LCT awareness.

![Diagram](image)

**Figure 2-3: Awareness and impression of LCT and incentives**

On average the adopters demonstrated a higher awareness of LCT compared to the non-adopters (also included organizations having diesel fleets equipped with emission reduction technologies). In the technology adoption process for organizations, awareness is recognized as the initial stage (Frambach & Schillewaert, 2002). Consequently, a higher level of awareness of LCT among current adopters is expected. Apart from assessing the role of awareness in current LCT adoption, its role in future LCT adoption was also assessed. It was found that interviewees...
who demonstrated a higher awareness of LCT were more likely to represent organizations planning to adopt LCT or expand their LCT fleet (Figure 2-4). These findings highlight a positive association between awareness of LCT and LCT adoption.

![Awareness of LCT](image)

**Figure 2-4: Association between awareness of LCT and potential to adopt LCT**

When the behavioral factors were compared among different vocations and fleet sizes of organizations, the positive association between awareness of LCT and its adoption held true; subgroups of organizations with higher awareness had a larger share of adopters. The ORE organizations had a higher awareness of LCT compared to the HDV organizations. At the same time, 3 out of 4 ORE organizations were adopters (compared to 0 adopters among the HDV organizations) (Figure 2-1). Similarly, small-fleet organizations, who had a lower share of adopters (1 adopter out of 8 small-fleet organizations compared to 2 adopters out of 4 among large-fleet organizations), were found to possess a lower awareness level compared to large-fleet organizations. These smaller organizations are usually aware of LCTs but lack comprehensive knowledge on factors like financing and technology (Wong, 2022).

Although none of the HDV organizations adopted LCT, the mixed-haul organizations had a higher awareness of LCT compared to the other HDV organizations (short-haul and long-haul).
The diverse vocations and vehicle range (Hughes, 1973) prevalent within mixed-haul organizations might make them more aware of available technologies such as LCT.

Although a positive association between awareness of LCT and its adoption was noticed, the same cannot be said about the impression of LCT. Adopters and non-adopters alike had negative impressions of LCT. The dissatisfaction with LCT among adopters might be explained by the technical limitations of LCT (discussed in the next sub-section). The negative impression may also be attributed to the intrinsic beliefs, business values, and strategic motives of the organizations (Bae et al., 2022). The overall impression of LCT among the organizations was somewhat negative to highly negative, with a rating of -1.17 (Figure 2-3). Among all the groups of organizations, the mixed-haul HDV organizations had the most unfavorable impression of LCT (Table 2-1).

2.4.1.2. Awareness and Impression of Incentives

On a scale of 0 to 3, the overall awareness of the organizations was 2.17 (Figure 2-3). This was lower than the awareness of LCT. On a scale of -2 (highly negative) to +2 (highly positive), the overall impression of incentives that they had on incentives was 0.08, in contrast to their somewhat negative (-1.17) impression of LCT (Figure 2-3).

Similar to the findings on awareness and impression of LCT, the awareness of incentives was higher among adopters compared to non-adopters (Table 2-2). The adopters also had the better impression of incentives. In addition, Figure 2-5 shows that interviewees who demonstrated a higher awareness of incentive programs were more likely to represent organizations planning to adopt LCT or expand their LCT fleet. These findings suggest that the awareness of incentives is also positively associated with LCT adoption. The positive association
necessitates information dissemination of available incentive programs (Wong, 2022) to foster LCT adoption.

The assessment of awareness of incentives for different adoption behaviors, vocations, and fleet sizes (Table 2-2) further underscores the positive association between awareness of incentives and LCT adoption. The overall awareness of incentives among ORE organizations, who were mostly adopters (3 out of 4), was higher than that of the HDV organizations, who were all non-adopters. Similarly, the large-fleet organizations, who had a higher share of adopters, were more aware of incentives than the small-fleet organizations. While it is important to raise awareness levels among these small-fleet organizations, it is also important to acknowledge that government agencies find it more difficult to reach out to them (Brito, 2022). Although none of the HDV organizations adopted LCT, the mixed-haul HDV organizations demonstrated the highest awareness of incentives compared to the other HDV organizations (short-haul and long-haul).

![Figure 2-5: Association between awareness of incentives and potential to adopt LCT](image)

The analysis of impression revealed subtle differences with respect to adoption behavior, vocation, and fleet size of organizations (Table 2-2). More importantly, it was found that lower
awareness of incentives doesn’t necessarily lead to a more negative impression. For instance, the small-fleet organizations had a lower level of awareness of incentives (compared to the large-fleet organizations), but their impression was better, suggesting that they perceived incentives as more valuable.

**2.4.1.3. Environmental Awareness**

A positive association between environmental awareness and potential to adopt LCT was also observed from the analysis. The 5 interviewees, who demonstrated some level of environmental concern/awareness during the interviews, represented organizations who are planning to adopt LCT (Figure 2-6). This suggests that there exists a positive association between general environmental awareness and the intention to adopt LCT in the HDV and ORE sectors. Similar associations have been confirmed in the light-duty sector by several studies (Kowalska-Pyzalska, 2018).

**Figure 2-6: Association between environmental awareness and potential to adopt LCT**

**2.4.2. Other Factors influencing LCT adoption**

Although behavioral factors are important determinants of LCT adoption, they aren’t the sole determinants. This section discusses the other factors influencing LCT adoption (facilitators,
barriers, general consideration), and their varying importance for different adoption behaviors, vocations, and fleet sizes.

2.4.2.1. *Facilitators*

The three most important facilitators of LCT adoption were environmental regulations (1.67 on a scale of 0 to 3), environmental friendliness of the vehicles (1.17 on a scale of 0 to 3), and green public relations (0.5 on a scale of 0 to 3). Although environmental regulations were the most important facilitator, they didn't always lead to AFV adoption. Instead of adopting AFVs, many interviewed organizations complied with environmental regulations by adding DEF filters to diesel vehicles.

Table 2-1 presents the most important facilitators for different groups of organizations. The contrast between the facilitators for HDV and ORE organizations offers valuable insight. Environmental regulations stood out as the pivotal facilitator for LCT adoption among HDV organizations, while ORE organizations emphasized environmental friendliness and green public relations as their top facilitators. Past research has verified the tendency of organizations to expand their low-carbon fleet as a strategic move to enhance public image (Sierzchula, 2014). In this regard, one of the ORE interviewees mentioned, “Well, we're located right near Silicon Valley in California. So, a lot of high-tech companies use biodiesel or use low-CARB emission-type equipment. We try to cater to them a little bit.”

Interestingly, infrequent maintenance, a facilitator acknowledged in prior research (Bae et al., 2022), was among the most important ones for small organizations and adopting organizations. An interviewee from a small fleet organization stated, “They're relatively maintenance-free, so I don't have to worry about checking the oil on them every time I start it up, I don't have to worry about air filters clogging up”.

56
2.4.2.2. **Barriers**

When the barriers were individually ranked, lack of refueling/recharging facilities (1.92 on a scale of 0 to 3), high purchase cost (1.5 on a scale of 0 to 3) and low range (1.42 on a scale of 0 to 3) were found to be the three most important ones. Table 2-1 presents the biggest barriers for different groups of organizations.

Significant insights can be drawn from the biggest barriers faced by adopters and non-adopters as well as those faced by small-fleet and large-fleet organizations. The most important barriers for adopting organizations and large-fleet organizations were from the technical category, suggesting that technical barriers may hinder their LCT fleet expansion, even after they have overcome the financial barriers to adopt them. However, these financial barriers were among the biggest barriers for small-fleet organizations and non-adopters. Hence, incentives aimed at mitigating financial barriers are particularly vital for smaller non-adopting organizations. These incentives may foster adoption among small-fleet organizations as these organizations are argued to be innovative, flexible, and more receptive toward new technologies (Frambach & Schillewaert, 2002).

One of the less common but important barriers that came up during the interviews was resistance to change. As one of the interviewees mentioned, “The reason we wouldn't change is because it's a total change in our operation and that would be expensive”. Hence, to make the change in operations smoother for these organizations, the regulators might want to consider a phased transition towards LCT rather than implementing an “all-at-once” policy (Alp et al., 2019). Another less common but important barrier was inability of LCT to fulfill extreme requirements. This might be a legitimate barrier because the already limited range of LCT
(specifically electric vehicles) is further reduced by extreme conditions like heavy loads and high temperatures (Quak et al., 2016).

2.4.2.3.  **General Considerations**

Apart from facilitators and barriers to LCT adoption, the interviews featured some general considerations that organizations make when purchasing any vehicle (Table 2-1). Among the different considerations, operating cost (1.92 on a scale of 0 to 3), purchase cost (1.58 on a scale of 0 to 3), and presence of incentives (1.33 on a scale of 0 to 3) were among the most important considerations.

The general considerations for adopters and non-adopters align with the findings from the barriers section; LCT-adopting organizations were more concerned about technical considerations while the non-adopting organizations were more concerned about the financial considerations. Operating cost was the most important consideration for long-haul and mixed-haul organizations, attributable to their heavier load and higher energy requirements (Mareev et al., 2018). For short-haul organizations, the most important consideration was purchase cost. For small organizations, incentives were among the top three considerations for vehicle purchase, unlike the large organizations.

**Table 2-1: Awareness, impression, and the three most important factors influencing the LCT adoption behavior of different categories of organizations**

<table>
<thead>
<tr>
<th>Category of interviewees (Number of interviewees)</th>
<th>Behavioral factors</th>
<th>Factors influencing LCT adoption</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Awareness rating on a scale of 0 to 3</td>
<td>Impression rating on a scale of -2 to +2</td>
</tr>
<tr>
<td>Adoppers (3)</td>
<td>3</td>
<td>-1</td>
</tr>
<tr>
<td></td>
<td>Less frequent maintenance (1)</td>
<td>Low operational/load-carrying capacity (2)</td>
</tr>
<tr>
<td></td>
<td>Surveillance (1)</td>
<td>Low range (2)</td>
</tr>
<tr>
<td>Non-adopters (9)</td>
<td>2.33</td>
<td>-1.22</td>
</tr>
<tr>
<td>Category</td>
<td>Code</td>
<td>Rating</td>
</tr>
<tr>
<td>-------------------</td>
<td>------</td>
<td>--------</td>
</tr>
<tr>
<td>Environmentally friendly</td>
<td>HDV (8)</td>
<td>2.25</td>
</tr>
<tr>
<td>Low range</td>
<td>Lack of refueling/recharging facilities</td>
<td>Purchase cost</td>
</tr>
<tr>
<td>Green public relations</td>
<td>0.56</td>
<td>1.22</td>
</tr>
<tr>
<td>Environmental regulations</td>
<td>(2.13)</td>
<td>High purchase cost</td>
</tr>
<tr>
<td>Environmentally friendly</td>
<td>(1.25)</td>
<td>Lack of refueling/recharging facilities</td>
</tr>
<tr>
<td>Green public relations</td>
<td>(0.25)</td>
<td>Low range (1.38)</td>
</tr>
<tr>
<td>Environmental regulations</td>
<td>(0.75)</td>
<td>Low operational/load-carrying capacity (1.5)</td>
</tr>
<tr>
<td>HDV (8)</td>
<td>2.25</td>
<td>-1.38</td>
</tr>
<tr>
<td>Environmental regulations</td>
<td>(2.67)</td>
<td>Low range (2.33)</td>
</tr>
<tr>
<td>Environmentally friendly</td>
<td>(2)</td>
<td>High purchase cost (2)</td>
</tr>
<tr>
<td>Green public relations</td>
<td>(0.67)</td>
<td>Lack of refueling/recharging facilities (2)</td>
</tr>
<tr>
<td>Environmental regulations</td>
<td>(0.75)</td>
<td>Low operational/load-carrying capacity (1.5)</td>
</tr>
<tr>
<td>Long-haul (3)</td>
<td>2.33</td>
<td>-1.33</td>
</tr>
<tr>
<td>Environmental regulations</td>
<td>(3)</td>
<td>Low range (2.33)</td>
</tr>
<tr>
<td>Environmentally friendly</td>
<td>(2)</td>
<td>High purchase cost (2)</td>
</tr>
<tr>
<td>Green public relations</td>
<td>(0.67)</td>
<td>Lack of refueling/recharging facilities (2)</td>
</tr>
<tr>
<td>Environmental regulations</td>
<td>(0)</td>
<td>Low range (1)</td>
</tr>
<tr>
<td>Short-haul (3)</td>
<td>1.67</td>
<td>-1.33</td>
</tr>
<tr>
<td>Environmental regulations</td>
<td>(3)</td>
<td>Lack of refueling/recharging facilities (2)</td>
</tr>
<tr>
<td>Environmentally friendly</td>
<td>(0.67)</td>
<td>High purchase cost (1.67)</td>
</tr>
<tr>
<td>Green public relations</td>
<td>(0)</td>
<td>Low range (1)</td>
</tr>
<tr>
<td>Mixed-haul (2)</td>
<td>3</td>
<td>-1.5</td>
</tr>
<tr>
<td>Environmentally friendly</td>
<td>(1)</td>
<td>High purchase cost (2)</td>
</tr>
<tr>
<td>Environmental regulations</td>
<td>(0)</td>
<td>Expensive batteries (1.5)</td>
</tr>
<tr>
<td>Green public relations</td>
<td>(0)</td>
<td>High repair &amp; maintenance cost (1.5)</td>
</tr>
<tr>
<td>Small Fleet (8)</td>
<td>2.25</td>
<td>-1.25</td>
</tr>
<tr>
<td>Environmental regulations</td>
<td>(1.88)</td>
<td>Lack of refueling/recharging facilities (1.88)</td>
</tr>
<tr>
<td>Environmentally friendly</td>
<td>(0.88)</td>
<td>High purchase cost (1.5)</td>
</tr>
<tr>
<td>Less frequent maintenance</td>
<td>(0.38)</td>
<td>High repair &amp; maintenance cost (1)</td>
</tr>
<tr>
<td>Large Fleet (4)</td>
<td>3</td>
<td>-1</td>
</tr>
<tr>
<td>Environmentally friendly</td>
<td>(1.75)</td>
<td>Low range (2.5)</td>
</tr>
<tr>
<td>Green public relations</td>
<td>(1.5)</td>
<td>High refueling/recharging time (2)</td>
</tr>
<tr>
<td>Environmental regulations</td>
<td>(1.25)</td>
<td>Lack of refueling/recharging facilities (2)</td>
</tr>
</tbody>
</table>
2.4.3. Existing Incentives and Expected Government Support

Although monetary incentives were stated as important considerations for vehicle purchase, the interviewees mentioned six different factors that negatively influence the impression of incentives. These factors unveil discrepancies between the expectations of the organizations and the current state of government support for LCT adoption. Hence, these factors need to be addressed to realize the potential of incentives as facilitators of LCT adoption. The factors can be ranked in the following order: conditions/restrictions (0.67 on a scale of 0 to 3), difficulty to acquire (0.58), cost ineffectiveness (0.50), distrust of government (0.50), bureaucracy (0.25), and waiting period (0.25).

Conditions/restrictions frequently came up as an important factor (especially among small and long-haul organizations), determining the impression of incentives. One interviewee stated that applying for incentives would mean that a huge chunk of his operations would be restricted within California only and it would be detrimental to his/her business. To make incentives more lucrative for organizations, it is imperative to remove burdensome restrictions and streamline the application process to the extent possible (Wong, 2022). Another important factor that came up during the interviews was distrust of government. When talking about the lack of trust the industry has in the government, one of the interviewees stated, “Any government involvement, the more it's under the radar, I think the more effective it's going to be, knowing the industry like I do”. Hence, even though regulatory/punitive environmental mandates cause diffusion of LCTs, they may make organizations less receptive to rewarding interventions (e.g., incentives) coming from the government. Therefore, it is recommended that policymakers recognize this tradeoff and use punitive interventions sparingly (Shi et al., 2021).
When asked about expected forms of government support, the organizations mentioned charging infrastructure support (1.75 on a scale of 0 to 3), monetary incentives (0.92 on a scale of 0 to 3), and collaboration with manufacturers (0.5 on a scale of 0 to 3) as the three most important ones. Since the lack of charging/refueling infrastructure came up as the most important barrier, it is expected that the organizations would be seeking charging infrastructure support. However, some interviewees also suggested potential technological improvements such as swappable batteries (Quak et al., 2016) and solar chargers, where the government can allocate resources. The government may also consider supporting manufacturers to come up with range extender technologies (Jahangir Samet et al., 2021) to mitigate the challenges of limited range. A statement from one of the interviewees emphasizes the need for technological improvements (Bae et al., 2022) and suggests that financial incentives are not enough to ensure long-term diffusion of LCT. He/she mentioned, "The incentives don’t make up for the short-range on the electric vehicles. You know the incentives don't overcome the problems they just offer you a little cash to deal with the problems indefinitely".

Given the heterogeneous landscape of the HDV and ORE sectors, the factors influencing the impression of incentives and expected government support differed for different groups of organizations (Table 2-2). Hence, targeted incentive programs tailored to the varying expectations of organizations may prove to be an effective approach (Wong, 2022).

<table>
<thead>
<tr>
<th>Category of interviewees (Number of interviewees)</th>
<th>Behavioral factors</th>
<th>Reason behind negative impression (Importance rating on a scale of 0 to 3)</th>
<th>Expected government support (Importance rating on a scale of 0 to 3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adopters (3)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Awareness rating on a scale of 0 to 3</td>
<td>Impression rating on a scale of -2 to +2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.67</td>
<td>0.33</td>
<td>Difficult to acquire (1.67)</td>
<td>Less restrictive environmental regulations (1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Conditions/restrictions (1)</td>
<td>Charging infrastructure support (0.67)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Cost ineffective (1)</td>
<td>Collaboration with manufacturers (0.67)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Distrust of Government (0.67)</td>
<td>Charging infrastructure support (2.11)</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
2.4.4. Generative AI in Content Analysis

Although the AI coding tool on ATLAS.Ti is still at its preliminary stage of development, using it as an assistive tool revealed ways in which generative AI can positively contribute to qualitative research.

Firstly, the codes produced by a generative AI tool can provide a summary of the studied material and the human coders could use the codes to validate the original framework or identify new concepts overlooked by the framework. In this study only 18 (4.5%) out of the 399 AI codes were found to capture new concepts, which weren’t already identified by the original coding framework. Hence, it was an indication that the original framework captured most of the concepts relevant to the research questions. However, the AI tool offered a new perspective, and the original framework was refined using the 18 codes that captured new concepts. For instance,
the subcategories “resistance to change” and “inability to fulfill extreme requirements” produced by the AI tool were newly added under the pre-existing main category “barriers to LCT adoption”. The AI tool also produced a main category named “environmental awareness” which was found to have a positive association with the potential to adopt LCT.

Secondly, a generative AI tool can identify new statements that belong to existing categories of the original coding framework. For instance, the sentence “Well, we would like to see some history of fleets that have changed over and what they ran into, to decide on how soon we’d want to change” was assigned the code “unproven technology” by the AI tool. Although “unproven technology” was a pre-existing subcategory (under “barriers to LCT adoption”), the human coders did not assign a category to the aforementioned statement during the manual coding phase. However, upon reevaluation of the statement, the code “unproven technology” was found to be a suitable subcategory for the statement.

Thirdly, generative AI can be used as a suggestive tool to find more suitable names for categories in the original coding framework. For instance, the subcategory “paperwork” (under the main category “reasons behind negative impression of incentives”) was more aptly changed to “bureaucracy” as it provided a more general description of the content being analyzed (Elo & Kyngäs, 2008).

Apart from the positive contributions of generative AI in this study, some potential areas of improvement were also discovered. Firstly, a generative AI tool needs to be fine-tuned to gather numerous statements sharing the same meaning under one code. Unless the tool can accomplish this task, the purpose (saving time) of using AI in a qualitative study would be defeated. This was especially prevalent in this study as 296 (74.2%) out of 399 codes generated had only one statement associated with them. This is an issue because coding is meant to reduce
the data instead of proliferating it (Bernard, 2000). Hence, even though the generation and assignment of codes using an AI tool takes very little time, aggregating codes to summarize the data may consume a considerable amount of time depending on the number of codes generated. Secondly, a generative AI tool needs to detect the nuances in natural language almost in the same way a human being can. Unless the AI tool can detect such nuances, it may assign wrong codes to statements. For instance, the AI tool inappropriately assigned the code “environmental concern” to the statement “We were motivated to acquire low-carbon vehicles because California was imposing some restrictions”. Although the organization did purchase low-carbon vehicles, they did so because of legal restrictions not because of environmental concern.

2.5. Conclusion

This study employed generative-AI-assisted content analysis to understand the association of behavioral factors with LCT adoption and explore the variation in factors influencing LCT adoption among different groups of organizations (with respect to adoption behaviors, vocations, and fleet sizes). Moreover, the study also identified some discrepancies between expected and available government support for LCT adoption.

With regard to the behavioral factors, a positive association was found between potential to adopt LCT and awareness of LCT. A positive association was also noticed between the potential to adopt LCT and awareness of different aspects (incentives and the environment) linked to LCT. This underscores the importance of raising awareness levels within organizations regarding factors linked to LCT. To this end, the government could consider introducing accessible vehicle leasing schemes for non-adopters to experiment with new low-carbon technologies. Moreover, they should invest more in educational/outreach programs to increase the awareness of available incentive programs. The smaller long-haul and short-haul trucking
companies operating in the HDV sector were found to have a lower level of awareness of LCT (compared to the ORE companies and larger mixed-haul HDV companies). Therefore, these organizations may be prioritized when planning initiatives designed to increase organizational awareness of LCT. However, further research is needed to confirm this.

The factors (including behavioral factors) influencing LCT adoption varied with respect to different adoption behaviors and vocations, suggesting a need for tailored incentives to facilitate LCT adoption among different groups of organizations. Environmental mandates were the most important reason behind LCT adoption in the HDV organizations, while the ORE organizations were mostly driven by environmental concern and opportunities to form green public relations. The financial barriers received greater importance among the non-adopters and smaller organizations compared to the larger organizations. On the other hand, for adopters and larger organizations, technical barriers received paramount importance.

The analysis of statements on available incentives identified a range of issues (e.g., conditions/restrictions, difficulty to acquire) with existing incentive programs that may limit their effectiveness as facilitators of LCT adoption. Hence, in order to ensure that government support initiatives facilitate LCT adoption, these issues should be addressed. Moreover, the analysis results suggest that the government should extend further support to charging infrastructure and technological improvements.

By using generative AI as an assistive tool, some additional insights were uncovered from the data, which the human coders overlooked in their analysis. This suggests the potential of this emerging technology to validate and refine the coding framework used in content analysis. However, to achieve better results, the ability of generative AI tools to produce aggregate-level codes and identify the nuances of natural language needs to be ensured.
The organizations, which were interviewed for this study, exhibited as much variability as possible in terms of adoption behavior, vocation, and fleet size. However, there are other organizational differences that were not captured in the sample. Hence, future research can explore heterogeneity in LCT adoption behavior based on other such differences (e.g., operations, energy demands, duty cycles) among HDV and ORE organizations. These future studies can leverage the power of generative AI, finding more creative ideas for its application as the technology evolves.

References


Indeed. (2023, March 3). *Short Haul vs. Long Haul Trucking: Definitions and Differences*.


Schreier, M. (2012). *Qualitative Content Analysis in Practice*. Sage. www.sagepub.co.uk/schreier


Conclusion

This thesis performs an artificial intelligence (AI) driven analysis of the low-carbon transportation (LCT) usage and adoption behavior in the transportation sector. For the light-duty vehicle (LDV) sector, the study employs a quantitative analysis, featuring a two-step machine learning model. And for the heavy-duty vehicle (HDV) and off-road equipment (ORE) sectors, the study employed a qualitative content analysis, refined by generative AI. By performing separate analyses for the LDV, HDV and ORE sectors the thesis intended to mend LCT research gaps that are specific to these sub-sectors of transportation. The results of the study can assist government agencies and manufacturers in their effort to realize the carbon mitigation potential of LCT.

For the LDV sector, the study investigates the factors that influence vehicle choice for trips in multi-vehicle households in the US, specifically those with at least one electric vehicle (EV) and one conventional vehicle (CV). A two-step machine learning modeling framework was utilized, starting with k-modes clustering to identify 5 distinct trip clusters and capture the heterogeneity in trips. Subsequently, a decision tree model was employed to predict vehicle choice (EV or CV). The variable importance assessment showed that trip day gas price and the driver's age were found to be major contributors to the decision tree's predictive power. Both of these variables had non-linear effects on EV preference, which a simple binary logit model would be unable to capture. This further underscores the significance of employing machine learning approaches, such as decision trees, within the context of this study. ALE plots were produced to analyze the effects of different variables on vehicle choice, as captured by the decision tree model. The analysis revealed that weekend trips primarily intended for shopping and dining were most likely to be made using EVs, indicating the discretionary use of EVs within multi-vehicle households.
This underscores the significance of having charging stations located near popular shopping and dining destinations. With regard to socio-economic attributes, multi-vehicle households with lower incomes and fewer workers were more inclined to choose EVs for their daily trips. However, these households face challenges in adopting EVs initially due to the higher purchase prices. To promote higher EV usage, targeted incentives should be implemented to make EVs more affordable for these households. Additionally, gas prices exceeding $2.80 per gallon were found to discourage the use of CVs within multi-vehicle households, suggesting that gas prices can serve as a tool to increase EV usage.

It is important to acknowledge a few limitations of the analysis of the LDV sector performed in this thesis. Due to the limited number of trips made by PHEVs and BEVs, EVs were treated as a single category, potentially overlooking differences between vehicle types. Furthermore, the NHTS 2017 dataset predominantly includes older EV models, which are gradually being replaced by newer models with extended ranges. Future studies on EV usage can leverage datasets that encompass trip data from newer EV models, enabling the modeling of vehicle choices for different EV categories separately.

For the HDV and ORE sectors, this thesis employed a generative-AI-assisted content analysis to understand the association of behavioral factors (awareness and impression) with LCT adoption and explore the variation in factors influencing LCT adoption among different groups of organizations (with respect to adoption behaviors, vocations, and fleet sizes). With regard to the behavioral factors, a positive association was found between potential to adopt LCT and awareness of LCT. A positive association was also noticed between the potential to adopt LCT and awareness of different aspects (incentives and the environment) linked to LCT. This underscores the importance of raising awareness levels within organizations regarding factors
linked to LCT. The smaller long-haul and short-haul trucking companies operating in the HDV sector were found to have a lower level of awareness of LCT (compared to the ORE companies and larger mixed-haul HDV companies). Therefore, these organizations may be prioritized when planning initiatives designed to increase organizational awareness of LCT. However, further research is needed to confirm this. The factors (including behavioral factors) influencing LCT adoption varied with respect to different adoption behaviors and vocations, suggesting a need for tailored incentives to facilitate LCT adoption among different groups of organizations. By using generative AI as an assistive tool, some additional insights were uncovered from the data, which were initially overlooked. This suggests the potential of this emerging technology to validate and refine the coding framework used in content analysis. However, to achieve better results, the ability of generative AI tools to produce aggregate-level codes and identify the nuances of natural language needs to be ensured.

The sample used in the analysis of the HDV and ORE sectors restricted the scope of the analysis; there are other organizational differences that were not captured in the sample. Hence, future research can explore heterogeneity in LCT adoption behavior based on other such differences (e.g., operations, energy demands, duty cycles) among HDV and ORE organizations. These future studies can leverage the power of generative AI, finding more creative ideas for its application as the technology evolves.

**Funding Sources**

The research reported in this paper is partly supported by the California Air Resources Board. The contents of this paper reflect the views of the authors who are responsible for the facts and the accuracy of the data presented herein. This paper does not constitute a standard, specification, or regulation.