

Article

Electric Vehicle Usage Patterns in Multi-Vehicle Households in the US: A Machine Learning Study

Vuban Chowdhury , Suman Kumar Mitra *  and Sarah Hernandez

Department of Civil Engineering, University of Arkansas, Fayetteville, AR 72701, USA; vubanc@uark.edu (V.C.); sarahvh@uark.edu (S.H.)

* Correspondence: skmitra@uark.edu

Abstract: Electric vehicles (EVs) play a significant role in reducing carbon emissions. In the US, EVs are mostly owned by multi-vehicle households, and their usage is primarily 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 internal combustion engine vehicles) within multi-vehicle households. A two-step machine-learning framework (clustering and decision trees) is proposed. The framework determines the preferred trip category for EV use and captures the effects of household attributes, driver attributes, built-environment factors, and gas prices on EV use in multi-vehicle households. Results indicate that discretionary 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. These findings are valuable for policymakers and auto manufacturers in targeting specific market segments and promoting EV adoption.

Keywords: electric vehicles; multi-vehicle household; machine learning; clustering; decision tree; NHTS



Citation: Chowdhury, V.; Mitra, S.K.; Hernandez, S. Electric Vehicle Usage Patterns in Multi-Vehicle Households in the US: A Machine Learning Study. *Sustainability* **2024**, *16*, 5200. <https://doi.org/10.3390/su16125200>

Academic Editor: Bin Ji

Received: 10 May 2024

Revised: 13 June 2024

Accepted: 17 June 2024

Published: 19 June 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

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% [1]. Since internal combustion engine vehicles (ICEVs) powered by gasoline or diesel are significant contributors to carbon emissions in the transportation sector, it is imperative to augment the usage of vehicles that rely on alternative fuels, such as electric vehicles (EVs). EVs include vehicles that are solely powered by electricity, such as battery electric vehicles (BEVs) [2], or partially powered by electricity, such as hybrid electric vehicles (HEVs) and plug-in hybrid electric vehicles (PHEVs) [3]. By 2050, 50% of the global light-duty vehicle sector is likely to be made up of EVs, resulting in a 50% reduction in carbon emission levels in the sector [4]. Understanding the usage pattern of EVs, therefore, lies at the heart of the endeavor to reduce carbon emissions.

EVs have garnered significant interest within the transportation research community. Well before their introduction to the market, researchers predicted the acceptance of EVs as a mode of transport. Several studies analyzing vehicle miles traveled (VMT) data from US households projected that EVs would be more suitable for multi-vehicle households, which were expected to be early adopters of this new technology [5–7]. With the introduction of EVs into the market, these projections proved to be accurate; the most recent National Household Travel Survey data [8] show that the majority of households owning EVs are multi-vehicle households (78.7% for HEVs, 83.4% for PHEVs, and 92.7% for BEVs) [9]. Since multi-vehicle households represent the major portion of the EV market, it is essential to investigate the usage patterns of these households in relation to their EVs.

Although studies on the usage of EVs in multi-vehicle households with at least one ICEV and one EV are limited in number, they do exist. Some recent studies have utilized GPS data to identify factors influencing EV usage and assess their potential for reducing

greenhouse gas emissions [10,11]. Other studies have evaluated how EVs are used when replacing an ICEV in multi-vehicle households [12]. To evaluate usage, these studies have collectively used metrics like the VMT, VKT (vehicle kilometers traveled), utility factor (fraction of VMT electrified), and daily driving distance. Even fewer in number are studies that specifically explore vehicle choice in multi-vehicle households (with at least one EV and one ICEV). This choice holds significance because the use of EVs in multi-vehicle households has cascading effects on the VMT of gasoline-powered vehicles and carbon emissions from ICEVs [11]. Furthermore, gaining a deeper understanding of this choice can assist policymakers in formulating policies based on the motivators for and barriers to EV usage. Two European studies have modeled this choice by creating artificial multi-vehicle households [13,14]. However, the EV markets in those countries differ significantly from the US EV market. To the best of our knowledge, there have been no studies in the US that model vehicle choices for daily trips in such multi-vehicle households.

This study aims to expand research on vehicle choice in US multi-vehicle households, specifically those with at least one EV and one ICEV. It also addresses some limitations of previous studies on the same topic. Firstly, unlike previous studies, this research incorporates a clustering analysis to capture the heterogeneity in trips. To highlight the significance of capturing heterogeneity in trips, consider the following example: Two trips covering the same distance might be regarded as different because they serve completely different purposes [15]. Conversely, two trips with different distances might be considered similar due to other shared attributes. A clustering analysis would be able to capture this heterogeneity by grouping similar trips together 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 multi-vehicle households in the US that adopted EVs. Consequently, these households offer a more representative sample of the overall population. Thirdly, this study employs interpretable machine learning techniques to capture the non-linear relationships between explanatory variables and household vehicle choice. By doing so, it aims to address the complexity and intricacies of these relationships. With the aforementioned research gaps in mind, this study seeks to answer the following questions:

- (1) What types 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 choice of using EVs for individual trips in US multi-vehicle households?

The rest of the article is organized as follows. Section 2 (Literature Review) summarizes the articles and the findings that are relevant to the research questions of this study. Section 3 (Data) provides a description of the dataset used in this study and the data preprocessing steps. Section 4 (Methodology) 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 article are Section 5 (Results and Discussion) and Section 6 (Conclusions), which present this study's findings, implications, and concluding statements.

2. Literature Review

In the context of EVs, previous studies have explored a wide range of topics like vehicle market share (e.g., [16]), user perceptions (e.g., [17]), incentives (e.g., [18]). However, the literature review for this study mainly consists of studies that explore the adoption and usage of electric vehicles in the light-duty vehicle sector (Table 1).

Table 1. Reviewed articles on electric vehicle usage (from 2015 to 2022).

Serial	Authors	Data	Variables Considered	Method(s)	Key Findings
1	Jensen and Mabit [14]	Data from 667 Danish households	Journey time, driving time, number of trip legs, journey distance, at least one charge, windspeed, precipitation, Citroën dummy, number of driving licenses, city dummy, first-week dummy.	Logit model	The number of trip legs, the drive time, and the requirement to charge the vehicle all had negative effects on the choice of EV. Meanwhile, precipitation and urban area had positive effects on the choice of EV.
2	Tamor and Milačić [7]	Data from 446 vehicles in the Puget Sound region in Washington	Range, DRA (days requiring adaptation)/threshold for inconvenience	Trip counting, analytic estimations	Electric vehicles of the same range if deployed as a second car in two-vehicle households would electrify roughly twice as many miles as those deployed in one-car households (replacing the household's only vehicle).
3	Jakobsson and Karlsson [19]	German household survey data (from 6339 vehicles) and Swedish GPS data (from 700 vehicles)	VKT (vehicle kilometers traveled), DRA (days requiring adaptation), range, capital expenditure, operating expenditure	Extrapolation and economic analysis	From the economic analysis, it was found that BEVs are best suited for multi-car households. Secondary household cars in these households are better suited to be replaced by a BEV.
4	Jakobsson, Karlssona, and Sprei [20]	GPS data from 10 Swedish households	DRA (days requiring adaptation), annual VKT, daily driving distance	Extrapolation	For most households, the EV is driven more than the car it replaced. There exists a large heterogeneity in the usage and adaptation among the households. The EVs mainly replaced the 40–70 km trips of the previous cars.
5	Karlsson [21]	GPS logging data for both cars in 64 commuting Swedish two-car households	VKT (vehicle kilometers traveled), SOC (state of charge), TCO (total cost of ownership)	Mixed integer quadratically constrained programming (MIQCP)	Two-car households in Sweden could gain USD 6700 due to the flexibility of owning an EV. This is because they can drive more on electricity, which is cheaper, and rely on their internal combustion engine vehicles for longer trips.
6	Bucher et al. [13]	A dataset of 129 Swiss drivers over a period of 1 year	Weekday/weekend, temperature, precipitation, sex, age, number of cars in household, work status, household size, long-distance trip leg, duration of activity, trip duration, household income, hour of day, month of year	Random forest and logit model	The variables of duration, distance, and weekday/weekend have a larger effect than household size, but they do not possess a high predictive power. This indicated that the range of the vehicles is not a deciding factor in this choice.
7	Karlsson [12]	GPS data from 20 Swedish two-car households	Range, VKT (vehicle kilometers traveled), flexibility utilization index	Ex-post analysis	The electric vehicles made a significant portion of short-range trips on weekends.
8	Mandey, Sprei, and Tal [22]	A dataset of 650 vehicles from 287 Californian households	Range, charging frequency, frequency of long-distance travel, frequency of overlaps, household VMT, ICEV mileage	Statistical analysis and regression	A short-range PHEV can electrify up to 70% of the eVMT of long-range BEVs (Bolt and Model S). Hence, PHEVs with a 35-mile all-electric range can be used as tools to decarbonize the transport sector.
9	Jakobsson, Karlsson, and Sprei [23]	GPS data from 25 Swedish two-car households	DRA (days requiring adaptation), daily driving distance	Quantitative, qualitative, and mixed methods	There exists a large heterogeneity in driving adaptation and behavior. Some households use their electric vehicle more than their previous car; some use it less. Some households change their driving style when they use their electric vehicle.
10	Chakraborty, Hardman, and Tal [10]	Survey data of 4125 Californian Households with BEVs or PHEVs	PEV characteristics, other household vehicle characteristics, built-environment variables, household characteristics, respondent characteristics, other factors influencing PEV use	OLS regression, SUR model, hypothesis tests	eVMT is correlated with traditional factors such as population density, attitudes towards technology, and lifestyle preferences. PEVs are driven as much as ICEVs. The availability of level 2 charging at home greatly influences the eVMT.

2.1. Adoption

The vast majority of studies on EVs are related to “adoption” or the decision to purchase these vehicles. The studies falling under this category used discrete choice models to explore EV adoption behavior and its determinants in the market. These determinants can be broadly classified into four groups, namely, demographic, contextual, situational, and psychological [24].

Before EVs became widely available in the market, studies relied upon stated preference surveys to explore the adoption behavior of potential EV owners. These studies found fuel cost to be one of the most important determinants of EV adoption. A study found that Americans were willing to pay USD 7600 more to own an EV [25] and American EV adoption was more sensitive to fuel cost reductions and charging availability than Japanese EV adoption [26]. Among the US states, fuel cost reduction had the greatest impact on consumers’ willingness to pay for EVs in California. In Germany, a stated preference study found that consumers’ willingness to pay was determined by factors such as fuel cost, emissions reductions, and tax exemptions [27]. The results from a study in Ireland, which explored similar factors, showed that respondents place a higher utility on fuel cost reduction compared to tax exemptions and emissions reductions [28]. Similar results were drawn from a systemic literature review study, which concluded that consumers rank the cost components (fuel cost and purchase cost) as the most important determinant of adoption [29].

The more recent studies have used revealed preference data to explore the determinants of EV adoption. Although these studies reiterated some of the findings from the stated preference surveys, they found a wider range of determinants. These determinants and their impact on EV adoption were found to vary widely across the different states in the US [9]. Nevertheless, the results from the revealed preference studies suggest the use of a combination of social, economic, infrastructural, and policy tools to increase EV adoption. Socio-economic 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 [30]. Infrastructural and policy factors, such as publicly available charging stations, gasoline and electricity prices, HOV lane access, and the presence of purchase incentives, were also found to be important determinants of EV market share [16]. A study exploring EV market share in different US states found that electricity prices affects the electric vehicle adoption rate the most [31].

2.2. Usage

Although it is important to understand the adoption of EVs, their impact on GHG emissions reductions is determined by their usage patterns within the households that own them. With regard to EV usage, the existing studies (on single- and multi-vehicle households) mostly explore determinants of EV usage by modeling vehicle miles travelled (VMT) or vehicle kilometers travelled (VKT). But there exists a shortage in the number of studies on vehicle choice in multi-vehicle households. The studies on EV usage that were reviewed for this study are presented in Table 1.

Before the availability of EVs in the market, several studies assessed the potential acceptability of EVs in both single- and multi-vehicle households. These studies used a combination of metrics (e.g., VMT, range, and the number of days one’s daily driving distance exceeds the EV range) and performed assessments for different scenarios (e.g., in which gas costs USD 5 per gallon or gas costs USD 7 per gallon). Assessments from these studies led to the conclusion that EVs are technically and economically better suited to multi-vehicle households [5,7,19]. As compared to single-vehicle household driving requirements, multi-vehicle household driving requirements supported the adoption of EVs with a lower range [20]. Replacing one vehicle in a multi-vehicle household instead of replacing the only vehicle in a single-vehicle household was expected to electrify roughly twice as many miles [7]. More specifically, EVs were found to be better suited as the second

car (the car with the lower annual VMT) in multi-vehicle households because of the greater frequency of driving and shorter distances covered by these second cars [19].

Given the potential acceptability of EVs in multi-vehicle households, researchers became interested in monitoring EV usage in such households. To study EV usage during the early adoption phase, some researchers provided an EV as a replacement for an ICEV in multi-vehicle households. They often found that owners adapt their driving behavior (e.g., take alternative routes) to suit their new EV in their household [20,23]. These studies collectively suggest that there exists a large heterogeneity in EV driving patterns; some households drive their EVs more than their previous car and some drive it less. Some recent studies, however, did not provide an EV as a replacement but collected data from households owning EVs instead [10,11,22]. They find that a range of factors can cause heterogeneity in EV driving patterns, such as population density, attitudes towards technology, and lifestyle preferences. These studies underscore the importance of charging infrastructure (especially the availability of level 2 charging at home) as a determinant of eVMT (electrified vehicle miles travelled) in multi-vehicle households. Moreover, PHEVs were found to have a higher total VMT and a higher share of household VMT compared to BEVs.

To the best of our knowledge, only two European studies have explored vehicle choice in multi-vehicle households (with at least one electric and one vehicle with an internal combustion engine). A Danish study found that the number of trip legs, the drive time, and the requirement to charge the vehicle all had negative effects on the probability of choosing an EV. On the other hand, precipitation and urban areas had positive effects on the probability of choosing an EV [14]. A study in Switzerland modeled household vehicle choice as a function of trip attributes, socio-demographic variables, and spatio-temporal variables [13]. After comparing different vehicle choice models, they suggest that trip duration, trip distance, and weekend indicators are among the most important determinants. While the first two variables positively influence these predictions, the weekend variable has a negative influence. However, they conclude that the choice cannot be predicted easily by the features considered in the study.

2.3. Summary

From the literature review, it was observed that studies on vehicle choice within multi-vehicle households with an EV and an ICEV are rare for the US. The studies that did explore this topic were conducted in two European markets (Denmark and Switzerland), where people's driving needs are very different from those in the US. The studies were limited in certain aspects. Firstly, the households considered in the existing studies may not be representative of the broader population of multi-vehicle households adopting EVs. This is due to the fact that the household samples in these studies did not previously own EVs and were provided with one solely for research purposes. Secondly, although Bucher et al. [13] hinted at the existence of non-linear effects of explanatory variables on vehicle choice, none of the studies reported or discussed these effects. Thirdly, the studies on this topic incorporated trip attributes in their models, but they considered these attributes in isolation. This approach introduces bias in the results since they fail to capture the heterogeneity in trips, as highlighted by Ozhegov and Ozhegova [15]. It is important to recognize that a trip's similarity or dissimilarity to another trip is based on multiple attributes rather than just a single attribute. This study fulfills these research gaps by modeling the vehicle choice of a representative sample of multi-vehicle households and employing machine learning techniques that capture the heterogeneity in trips and the non-linear effects of variables on vehicle choice.

3. Data

3.1. 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 [8]. These surveys employ professional processing procedures (e.g., weighted response rates to

account for disproportionate sampling across a region) to capture travel behavior and its seasonal variation over 12 months [8,9]. The 2017 NHTS consists of four datasets, namely, household, vehicle, person, and trip datasets.

3.2. Data Cleaning

The NHTS trip dataset is an inventory of all trips taken within a specified 24 h period by household members older than five. 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). The trip data for the multi-vehicle households were merged with the vehicle and person datasets to retrieve the vehicle, household, and driver attributes. This study deals with households owning at least one ICEV and at least one EV. Hence, trips made by households owning only EVs or only ICEVs were taken out. As this study is only interested in the choice of household vehicles for trips, any trips involving non-household vehicles were discarded. Observations with missing/unknown values for important variables were also taken out. To drop observations with excessive average trip speeds (e.g., 1495 mph) (derived from the trip distance and duration), the top speed of the fastest model for each vehicle manufacturer was checked [32]. Lastly, observations with return trips (trips in which one's destination is their home) were excluded. This data-cleaning step is justified because household members are bound to use the vehicle for return trips that they chose when leaving home. Since there is not a vehicle choice involved for these trips, they were not considered in this study. The cleaned dataset contained a total of 19,825 trips made by 3917 households. Figure 1 (top) shows the proportion of EV trips and ICEV 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 ICEV trips (50.5%), which is conducive to the performance of machine learning models [33]. Among the EV trips, most of the trips were made by HEVs. Figure 1 (bottom) also contains a symbol map showing the spatial distribution of trips across the different states. The cleaned dataset contains trips from 49 states (except Mississippi) and the district of Columbia. The states highlighted with green and light-green symbols (California, Texas, New York, Wisconsin, North Carolina, and Georgia) contribute greatly to the number of trips.

3.3. Variable Selection

The variable selection for this study was informed by existing studies on EVs and a multicollinearity assessment. Since this study intends to identify the types of trips that are most likely to be made by EVs in multi-vehicle households, individual trip attributes (e.g., trip purpose, trip distance, and number of passengers) were considered in the modeling framework. Trip attributes have been previously used in studies that modeled EV usage [13,14]. Based on these studies, a number of trip attributes were considered to be included in the model. In addition to trip attributes, the models also included household attributes, which have been commonly linked to EV usage and adoption [9,10,30,33]. Existing EV studies also identified the effects of personal attributes [9,13,25,28], built-environment variables [5,16], and gas prices [34]. Based on previous studies, an initial list of variables was produced. Some of these variables had multi-collinearity and were excluded from the list. The final subset of variables selected for modeling had VIFs below four. Table 2 shows the variables which were included in the final modeling framework and their descriptive statistics.

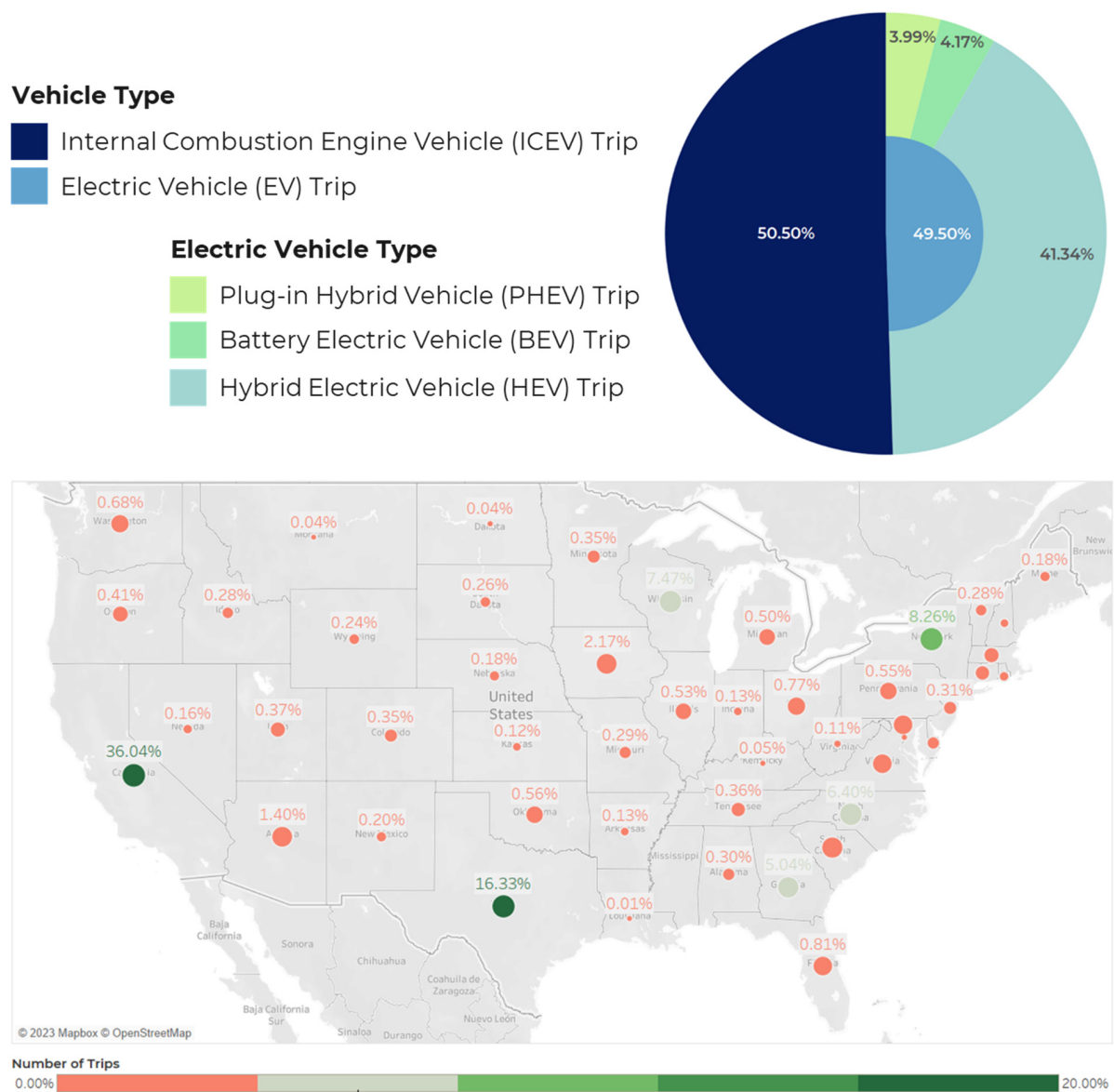


Figure 1. The percentages of trips made by different vehicle types and fuel types (**top**) and number of trips in the 50 states of the US (**bottom**).

Table 2. The descriptive statistics of the variables used in this study.

Variable		Categories	Distribution
Household Attributes	Home Ownership (Binary)	Rent's a House	91%
		Own's a House	9%
	Household Income (Discrete)	Low (<USD 50,000)	10%
		Medium (USD 50,000–150,000)	57%
		High (>USD 150,000)	33%
	Number of Household Workers (Discrete)	0	18%
		1	26%
		2	46%
		3	8%
		4	2%
5		<1%	
Children (Binary)	No children	90%	
	1 or more children	10%	

Table 2. Cont.

	Variable	Categories	Distribution
Driver Attributes	Driver's Age (Continuous)	-	52.4 ± 15.65 *
	Driver's Education Level (Discrete)	No high school degree	9%
		High school or associate degree	19%
		Bachelor's degree or higher	72%
	Driver's Sex (Binary)	Female	44%
		Male	56%
Built-Environment Variables	Employment Density in Workers Per 0.01 Square Miles (Continuous)	-	3.74 ± 4.52 *
	Population Density in Persons Per 0.01 Square Miles (Continuous)	-	1.54 ± 1.52 *
Gas price on the day of the trip in Cents (USD) per Gallon (Continuous) **		-	246.86 ± 25.64 *
Trip Attributes	Weekday /Weekend (Binary)	Weekday	77%
		Weekend	23%
	Starting Time (Discrete)	12 AM–6 AM	3%
		6 AM–10AM	30%
		10 AM–3 PM	38%
		3 PM–7 PM	24%
		7 PM–12 AM	5%
	Home/Non-home Based (Binary)	Home-based trip	49%
		Non-home-based trip	51%
	Trip Purpose (Discrete)	Errands	16%
		Others	10%
		Shopping or Dining	38%
		Social or recreational	16%
	Dwelling time (Discrete)	Work	20%
		1–15 min	29%
		15–50 min	24%
		50–150 min	25%
	Trip Distance (Discrete)	More than 150 min	22%
		0–2 miles	28%
		2–5 miles	28%
		5–15 miles	28%
	Number of Passengers (Discrete)	More than 15 miles	16%
		1 passenger	57%
		2–4 passengers	41%
		5–10 passengers	2%

* 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.

4. Methodology

This study used a combination of two machine learning techniques to model household vehicle choice (Figure 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 make a choice between their electric and internal combustion engine vehicles based on the type of trip they are making. Hence, the classification model accepted the trip cluster from the clustering model as an input variable. In addition, the classification model also captured the effects of household attributes, driver attributes, gas prices, and built-environment variables by accepting them as inputs. The following subsections describe the two models in detail. All the models were estimated using Python 3.11.7 on a Dell (Round Rock, TX, USA) Precision Tower 3630 computer, equipped with an Intel (Hillsboro, OR, USA) Core i7-9700K processor (8 cores, 12 MB cache, 3.6 GHz base frequency, and 4.9 GHz turbo frequency) and 32 GB of memory.

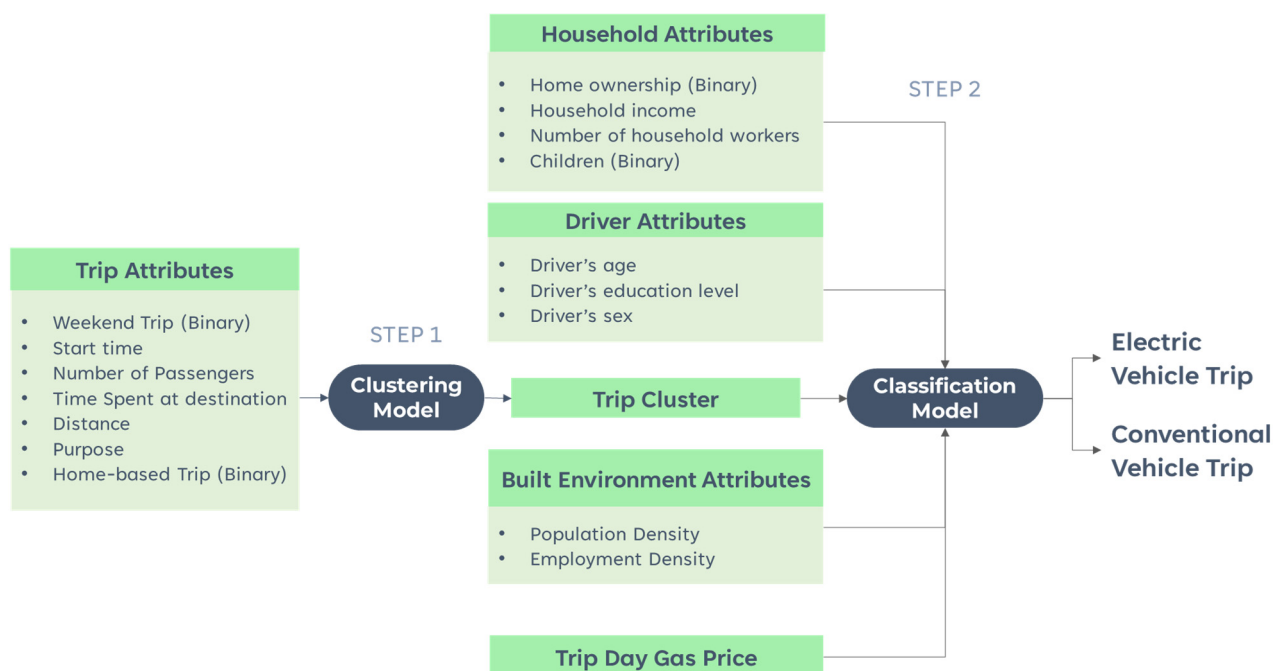


Figure 2. Two-step modeling framework used to predict vehicle choice.

4.1. Clustering Model

Since these trips can vary based on a number of attributes (e.g., trip distance, starting time, and trip purpose), a k-modes clustering analysis [35] was performed to cluster the trips. The clustering analysis maximized the homogeneity within the same trip cluster and minimized the homogeneity 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, and home-based/non-home-based trip. The k-modes clustering approach used in this study is less vulnerable to the local optima compared to statistical approaches such as the latent class analysis [36]. K-modes clustering was chosen over k-means clustering because some of the variables of interest were unordered categorical variables (e.g., trip purpose and home-based/non-home-based trip). The continuous variables of interest (e.g., dwelling time at destination and trip distance) were converted into categorical variables before applying the k-modes clustering analysis. Although k-prototype clustering may make more sense for mixed data (with categorical and continuous variables), the k-modes clustering approach was found to provide clearer delineation among clusters for this specific dataset. Apart from the delineation among clusters, the model results are easier to interpret when the variable types are the same. Hence, k-modes clustering was chosen over the other k-prototype clustering approach.

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

- (1) A set of observations from the dataset $Q = [Q_1, Q_2, Q_3, \dots, Q_k]$ is initialized as the cluster centroids using a density-based initialization algorithm [37]. This initialization method helps avoid the necessity of running the algorithm multiple times to search for an effective solution. The observation with the maximum density is initialized as the first centroid. The remaining centroids are initialized based on density as well as the distance from other centroids.
- (2) Every trip/observation (denoted by X) outside Q is assigned to a cluster from Q whose centroid has the smallest hamming distance [38] from X . Hamming distance can be defined as follows:

$$d(X, Q) = \sum_{j=1}^I \delta(x_j, q_j) \text{ where, } \delta(x_j, q_j) = \begin{cases} 0 & (x_j = q_j) \\ 1 & (x_j \neq q_j) \end{cases} \quad (1)$$

- (3) After every trip is assigned to a cluster, the cluster centroids in Q are updated based on the newly assigned trips. The hamming distances for each trip are recalculated, and trips are assigned to new clusters based on the newly calculated hamming distances.
- (4) Step 3 is repeated until no trip in the dataset changes clusters. And the cost or the sum of the hamming distances for all the observations were recorded for the model with k clusters.

Models with clusters ranging from 2 to 10 were assessed. To assist the selection of the final clustering model, the elbow plot, the silhouette score [39], and cluster separation were evaluated. Moreover, the trip attributes for different clusters were observed. The model that provided the best scores for the evaluation metrics and a clear delineation among the clusters was selected as the final model.

4.2. Classification Model

Household vehicle choice (ICEV or EV) was modeled as a function of trip cluster (from the clustering model), household attributes, driver attributes, built-environment attributes, and gas price on the day of the trip. To accomplish this, 4 different modeling techniques were tested and compared. The first three models were tree-based machine learning models (decision tree, random forest, and extreme gradient boosting). These models were chosen as candidate models because previous studies have found tree-based machine learning models to be highly accurate in predicting mode choice [40,41]. The last one among the four models was a logit model, which has been the most popular method applied to mode choice modeling. In order to generate training and testing subsets for the models, the cleaned dataset underwent random shuffling, followed by a split of 85% for training and 15% for testing purposes. The four modeling approaches were then compared based on cross-validation accuracy (10-fold cross-validation) [41] and testing accuracy. Assessment of both cross-validation and testing accuracies ensures the maximum generalizability of the models. The most accurate model was then chosen for us to interpret. The following sections provide a general overview of the four modeling techniques and the specific configurations used for this study.

4.2.1. Decision Tree

A decision tree performs classification by recursively partitioning the dataset based on its features [42]. These partitions are also called “decisions”. The decision tree model makes a series of consecutive decisions to form a tree structure, which leads to its final predictions. For this study, a decision tree was implemented using the “DecisionTreeClassifier” class in Python’s scikit-learn library. This class implements the CART version of the decision tree introduced by [43]. The mathematical formulation can be described as follows:

Given training vectors $x_i \in R$, $i = 1, 2, 3, \dots, l$ and a label vector $y \in R^l$, a decision tree performs partitioning such that observations with the same labels fall into the same group. The datum 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 feature j and a threshold value t , the datum is split into $Q_m^{left}(\theta)$ and $Q_m^{right}(\theta)$ subsets such that the following is seen:

$$Q_m^{left}(\theta) = \{(x, y) | x_j < t_m\} \quad (2)$$

$$Q_m^{right}(\theta) = Q_m \setminus Q_m^{left}(\theta) \quad (3)$$

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 features considered in this study consist of categorical features, they were converted into dummy variables as suggested by Breiman [44].

4.2.2. 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) [42]. The technique is referred to as random forest because it combines a number of randomized decision trees [45]. The “RandomForestClassifier” class in Python’s scikit-learn library was used to implement the random forest model in this study. The class implements the version of random forests introduced by Breiman [44]. The random forest algorithm for this study and the hyperparameters involved in each step are discussed below:

- (1) At first, a subset of the data is formed by bootstrapping [43]. In this step, a random sample of the data is drawn with replacement. This implies that some observations may be duplicated, and others may be left out of the sample.
- (2) Next, a decision tree is constructed using the bootstrapped sample and a set of randomly selected features. As recommended by previous studies [46], the number of randomly selected features was set as the square root of the total number of features rounded to the nearest integer. The decision tree is grown until the number of observations at a node reaches 1.
- (3) The number of decision trees to be grown was set to 1000; the first two steps were repeated 1000 times.
- (4) To make a prediction for a new observation, each decision tree in the forest is traversed, and the predictions from each tree are recorded. Finally, the majority vote of predictions is taken as the final prediction.

4.2.3. 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 [47]. 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 [48]. 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. The gradient and hessian are used to split the data into regions, with each region corresponding to a specific leaf node in the decision tree. The algorithm assigns weights to each region based on the objective function and the current model and uses these weights to make predictions. The XGBoost package in Python was used to implement the model.

4.2.4. Binary Logit

This study also employed a binary logit model, which served as a baseline for comparison. The binary logit has been applied in a great number of studies to model mode choice [49,50]. The dependent variable of model Y_i could take either a value of 1 (for EV) or 0 (for ICEV). The probability that the dependent variable equals 1 for an observation i is given as follows:

$$\Pr(Y_i = 1|X_i) = \exp(X_i \beta) / (1 + \exp(X_i \beta)) \quad (4)$$

X_i is a matrix of features; β is a vector of unknown coefficients estimated via maximum likelihood estimation (MLE) on Stata 16. Robust standard errors were used in the process to account for possible heteroskedasticity [51]. The base case in this model [52] was the ICEV.

4.2.5. Classification Model Interpretation

Although machine learning models have been widely successful in terms of their predictive accuracy, they have been criticized for their lack of interpretability compared to

traditional discrete choice models, such as the binary logit. However, some recent studies have utilized a number of interpretation methods to make machine learning models interpretable [40,53]. Two such methods were used to interpret the best-performing classification model. Firstly, the impurity-based variable importance of the predictors was estimated using the Gini index [41]. This metric provided a measure of the predictive powers of the variables in the model. In addition, the accumulated local effects (ALEs) [54] were estimated to decipher the marginal effects of the independent variables on EV choice. ALE plots are able to illustrate any type of relationship (e.g., linear, multi-linear, or non-linear) between a variable and the predicted outcome.

5. Results and Discussion

This section discusses the results from each of the two steps in the modeling framework. For each step, different model specifications are compared, and we choose the best model to interpret.

5.1. Clustering Model

5.1.1. Model Comparison

As mentioned earlier, k-modes clustering models with clusters ranging from two to ten were tested and compared based on the elbow plot, silhouette score [39], and cluster separation. Figure 3 shows the elbow plot and the silhouette scores. An “elbow” (the point with the most significant reduction in the value of the cost function) in the elbow plot and a higher value for the silhouette score indicate the optimal number of clusters. The silhouette score indicates that there are two candidate models: the two-cluster model and the five-cluster model. Even though the two-cluster model ranks higher based on the silhouette score, its cost is also higher in the elbow plot. And it was found that there was not a clear separation between the two clusters based on the trip attributes used for clustering. Hence, based on the elbow plot and upon a closer inspection of the cluster separation, the five-cluster model was selected as the final model.

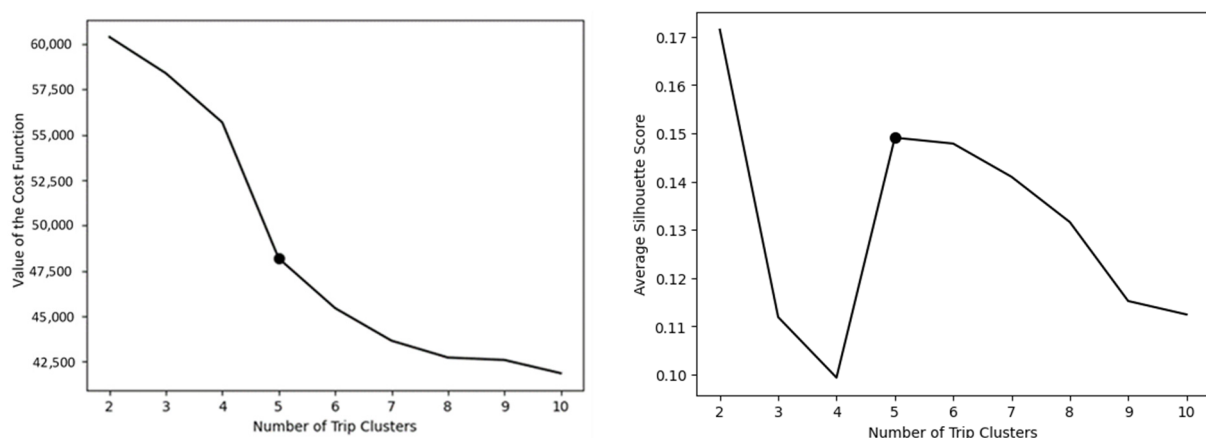


Figure 3. Comparison of k-modes clustering models based on the elbow plot (left) and silhouette score (right).

5.1.2. Model Interpretation

Table 3 shows the names of the five 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) for each trip attribute.

Table 3. Proportion of different trip attributes for the five trip clusters.

Trip Attribute		Trip Clusters				
		Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
		(8060 Trips)	(4956 Trips)	(2075 Trips)	(2249 Trips)	(2485 Trips)
Weekday/Weekend	Weekday trip	86%	90%	88%	74%	20%
	Weekend trip	14%	10%	12%	26%	80%
Starting time	12 AM–6 AM	1%	7%	2%	1%	1%
	6 AM–10 AM	15%	75%	14%	13%	17%
	10 AM–3 PM	60%	11%	17%	19%	56%
	3 PM–7 PM	18%	6%	64%	60%	17%
	7 PM–12 AM	6%	2%	3%	6%	8%
Home-based/non-home based	Home-based trip	22%	81%	76%	25%	75%
	Non-home-based trip	78%	19%	24%	75%	25%
Trip Purpose	Errands	19%	6%	47%	10%	6%
	Others	7%	13%	13%	11%	12%
	Shopping or Dining	58%	6%	23%	24%	64%
	Social or recreational	6%	12%	15%	51%	16%
	Work	9%	63%	2%	4%	2%
Dwelling time (time spent at destination)	1–15 min	51%	9%	32%	10%	16%
	15–50 min	29%	8%	49%	16%	25%
	50–150 min	14%	18%	16%	60%	52%
	More than 150 min	5%	66%	4%	14%	8%
Trip Distance	0–2 miles	33%	29%	13%	27%	21%
	2–5 miles	43%	17%	19%	19%	19%
	5–15 miles	15%	28%	62%	16%	49%
	More than 15 miles	9%	26%	6%	38%	11%
Number of Passengers	1 passenger	67%	82%	63%	11%	11%
	2–4 passengers	32%	17%	35%	86%	86%
	5–10 passengers	2%	1%	1%	3%	3%

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 typically involve spending little time at the destination (1–15 min for 51% of the trips) and cover short distances (43% between two and five 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 min 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%).

Trip cluster 4 is predominantly made up of social or recreational trips (51%). As expected for most social and recreational trips, these trips usually have two to four passengers (86%) and involve spending 50–150 min 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 two to four 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 min dwelling time) at the destination in 52% of the cases. On the other hand, the majority (51%) of the trips in cluster 1 involve spending 1–15 min at the destination. These distinctions (between the trips in clusters 1 and 5) have been captured by the clustering model (as highlighted in Table 3), which underscores the importance of capturing trip heterogeneity with a two-step modeling framework.

5.2. Classification Model

5.2.1. Model Comparison

As previously mentioned, the classification models were evaluated based on the cross-validation accuracy and testing accuracy. Table 4 presents the results of these metrics. Among

the models, the decision tree demonstrated a superior performance. While the accuracies of the three machine learning models were quite similar, there was a notable decrease in accuracy for the binary logit model. This discrepancy in accuracy further supports the preference for a machine learning model over traditional discrete choice models in the context of this study. Additionally, the significant difference in accuracy suggests the presence of non-linear relationships that the binary logit model failed to capture. For this specific study, ensemble techniques (random forest and XG boost) did not improve the prediction accuracy of a decision tree for unseen data, as suggested by the cross-validation and testing accuracies. However, if more complex relationships were modeled or a higher dimensional dataset was used, then the ensemble techniques would likely outperform the decision tree. In the following section, the best-performing machine learning model (the decision tree) is examined in terms of its interpretation. This analysis aims to confirm the existence of non-linear relationships and provide insights into their nature.

Table 4. Comparison of accuracies of the classification models.

Model	Cross-Validation Accuracy	Training Accuracy	Testing Accuracy
Decision Tree	88%	99%	87%
Random Forest	84%	99%	83%
XG Boost	87%	97%	86%
Binary Logit	58%	58%	58%

5.2.2. Model Interpretation

Variable Importance

The variable importance generally represents the change in the performance of the model in response to the change in the value of an input variable. The metric represents the predictive power of a variable in the model. The importance of the variables used in the decision tree is shown in Figure 4. Upon an initial inspection, it is noted that the continuous variables (gas price on the day of the trip, driver's age, employment density, and population density) contribute more to the predictive power of the model compared to the discrete variables. Among the discrete variables, the least important predictors are the presence of children, other driver attributes (driver's sex or education level), and household attributes (home ownership, household income, and number of household workers). The most important discrete variable is the trip cluster (0.073), underscoring the importance of clustering the trips in the modeling framework. Among the continuous variables, the least important predictor in the model is population density (0.079), followed by employment density (0.088), driver's age (0.252), and gas price on the day of the trip (0.331).

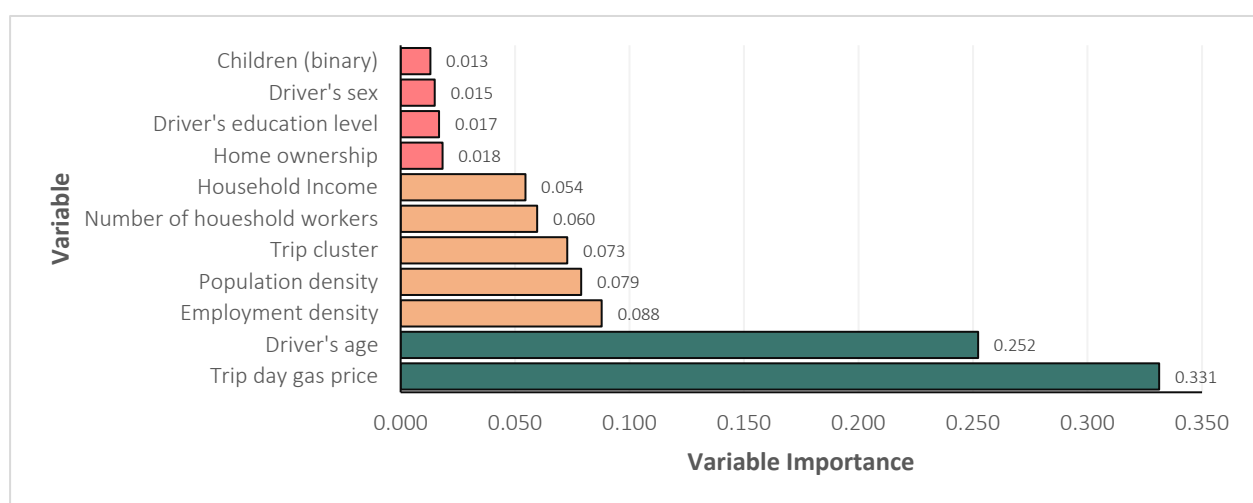


Figure 4. The variables in the decision tree model arranged in ascending order of variable importance.

While the 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.

Accumulated Local Effects

Accumulated local effects (ALEs) allow for the estimation of the marginal effects of variables on vehicle choice. The 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. Figures 5 and 6 show the ALE plots of each of the variables in the model. The plots show the marginal effects on the choice probability of EV for different values of the variables.

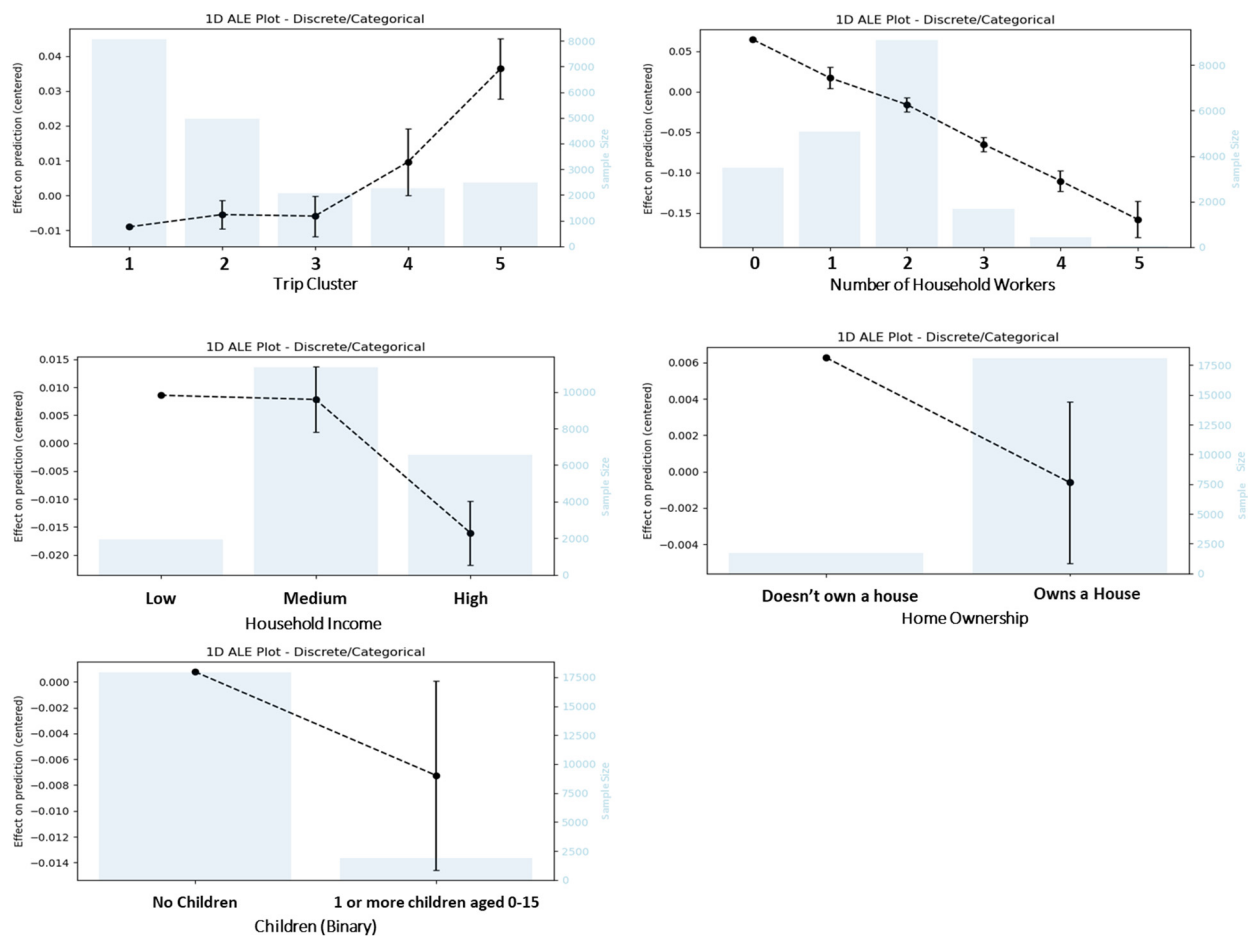


Figure 5. The accumulated local effect (ALE) plots of the trip cluster and household attributes for the decision tree model. The black dots (joined by the broken lines or continuous lines) indicate the effect of the variables at a specific value. The light blue bars indicate number of observations for a specific value of a variable.

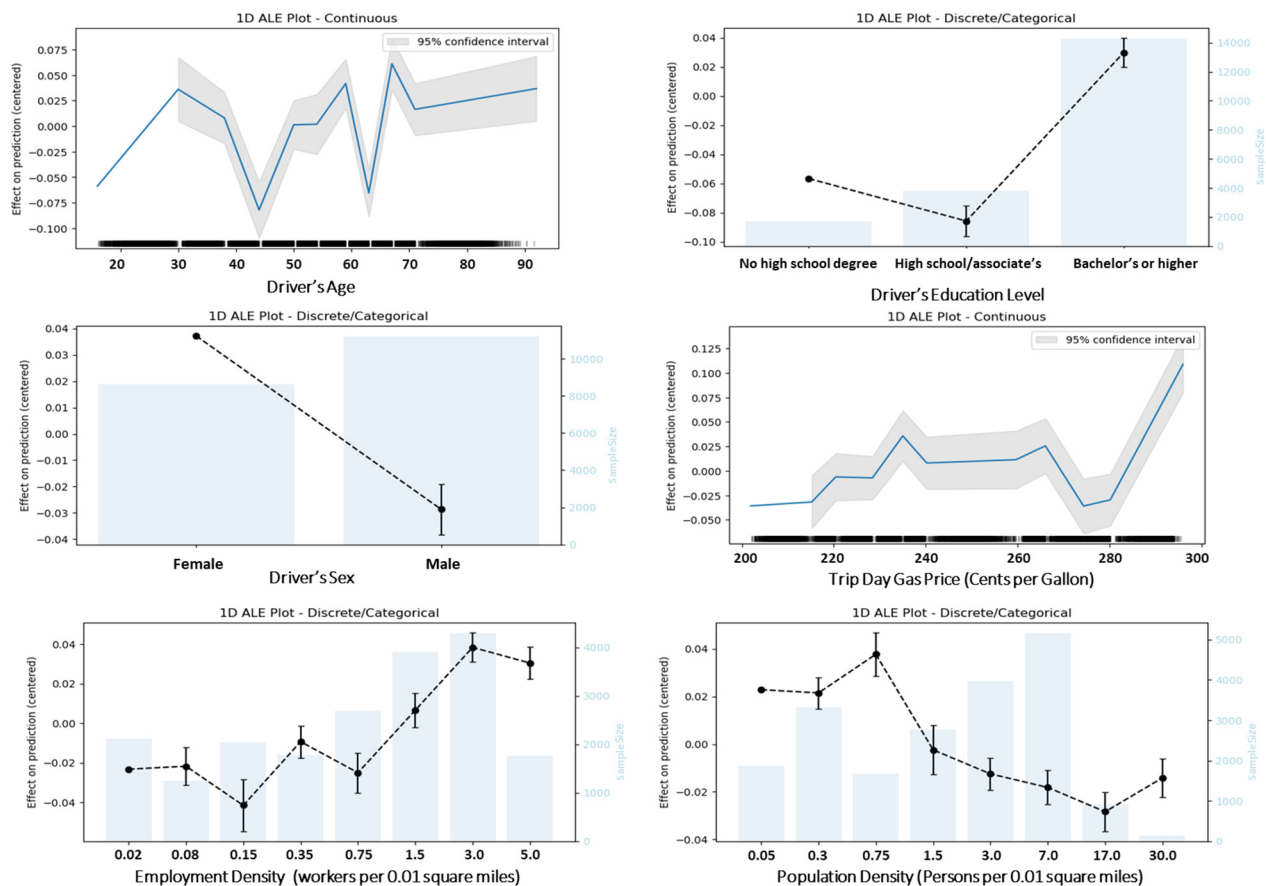


Figure 6. The accumulated local effect (ALE) plots of driver attributes, gas price, and built-environment variables for the decision tree model. The black dots (joined by the broken lines or continuous lines) indicate the effect of the variables at a specific value. The light blue bars indicate number of observations for a specific value of a variable.

From the ALE plot of trip clusters, we can see that the first three trip clusters have a negative effect on the probability of choosing an EV, implying that multi-vehicle households prefer their CVs for making trips from these trip clusters. On the other hand, trip clusters 4 and 5 have positive effects (ALEs of 0.009 and 0.037, respectively) on the choice probability of EV. This 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 the probability of choosing an EV) include carpooling trips (with two to four passengers). These trips are less common in the US compared to the single-occupancy trips observed in clusters 1, 2, and 3; the average car occupancy in 2017 was 1.5 passengers per vehicle [2]. Additionally, trip cluster 5 (which has the largest positive effect on the probability of choosing an EV) primarily serves the discretionary purposes of shopping and dining (Table 2). Although trip cluster 1 also predominantly serves shopping and dining purposes, it has a negative effect on the probability of choosing an EV. This may be because trips in cluster 1 are more time-sensitive in nature (mostly involving spending 1–15 min at the destination compared to 50–150 min for trip cluster 5). Moreover, trips in cluster 5 (which has the largest positive effect on the probability of choosing an EV) are made during the weekends 80% of the time (Table 2).

This finding aligns well with a Swedish study on two-car households [12], which found that the 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 min for HEVs compared to 5 min for CVs) [55].

Moreover, EVs are usually associated with higher insurance payments [56] and a greater sensitivity to external environments. Given the disparity in charging time and sensitivity, as well as the 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 drivers' behavior also has some implications for the locations of recharging/refueling stations and the formulation of charging infrastructure funding policies. These findings can inform the federal government in establishing criteria for the allocation of charging infrastructure funding. For instance, if government agencies and stakeholders aim to align station locations with trip patterns, proactive measures could involve placing more charging stations near discretionary trip attractors, such as shopping malls or restaurants. Conversely, if the goal is to influence behavior and promote more frequent EV trips, strategically siting stations near frequent trip attractors may be a viable strategy.

As the number of household workers goes up, the probability of choosing an EV goes down steadily (Figure 5). This may be attributed to the higher number of ICEVs (compared to EVs) in households with a higher number of workers. For instance, an analysis of the dataset used in this study shows that the average number of ICEVs in households with three or more workers is three. In the same households, the average number of EVs is one. From the ALE plot of household income (Figure 5), 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 ($ALE = 0.0155$). The lower-income households might be inclined to use EVs more because of their fuel efficiency. Previous studies have shown that electric vehicles (EVs) can achieve cost-effectiveness in multi-vehicle households within six years, primarily due to their fuel efficiency offsetting the higher purchase price [19]. Recent federal policies on EV purchase incentives are expected to drive market growth by alleviating the substantial upfront costs of EVs [56]. Given the effects of income and number of household workers, it is crucial for policymakers to prioritize these incentives for lower-income households and those with fewer workers. Currently, higher-income households are disproportionately represented among EV owners [57]. By focusing incentives on reducing the purchase price of EVs for lower-income households, we can anticipate increased adoption rates.

The ALE plot of children indicates that households with one or more children prefer ICEVs and households with no children prefer EVs (Figure 5). This could be true because households with children are likely to have more family members; therefore, they need higher-occupancy vehicles (e.g., minivans). Most of these higher-occupancy vehicles in the NHTS 2017 dataset are ICEVs, which explains their preference for ICEVs over EVs. Hence, households with one or more children may represent a segment of the market that EV manufacturers can further tap into by producing higher-occupancy light-duty vehicles.

Observing the ALE plot for driver age (Figure 6) tells us that age has a multi-linear relationship with EV choice. It is clear from the plot that the probability of choosing an EV linearly goes up as one's age goes up until the age of 30. However, the relationship is no longer clear after 30 years. Even though age was the second most important predictor, the marginal effect of age above 30 is not discernible. It is possible that the effect of age on vehicle choice in these households may also vary geographically within the US, similar to the effect of age on adoption [57]. This might be a reason why the national sample could not capture a clear relationship between EV choice and drivers who are older than 30.

As indicated by the ALE plot of driver's education level (Figure 6), the probability of a driver choosing an 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 [57]. A higher level of education is generally associated with a higher concern for the environment, which might cause people to choose to drive their EVs.

The ALE plots of driver's sex indicate that women in multi-vehicle households are more likely to drive EVs compared to men (Figure 6). This finding may be attributable to vehicle access in these households. A previous study on car-deficient households

(houses with a lower number of cars than drivers) found that men have more access to the household car than women [58]. In multi-vehicle households with EVs and ICEVs, the same phenomenon may apply to ICEVs; female members may have less access to their ICEVs, which may make them more likely to drive their EVs.

Gas price on the day of the trip, similar to age, has a multi-linear relationship with the probability of choosing an EV (Figure 6). Gas price on the day of the trip had the highest predictive power among all the variables. 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 the probability of choosing an EV. Previous studies have suggested making conventional fuels more expensive as a strategy to promote EV use [56]. The threshold of USD 2.80 for the price of gas can be used to implement such strategies. However, this threshold has to be adjusted for inflation since the data for this study are 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 the built-environment variables (employment density and population density) appear to have opposite effects on vehicle choice (Figure 6). The marginal effects for the built-environment variables do not show a steady increase or decrease. In general, the larger values of employment density have positive effects on the probability of choosing an EV, while the smaller values have negative effects. Whereas for the larger values of population density, the marginal effects on the probability of choosing an EV are negative. Population density was also found to have a negative effect on the VMT of PEVs in multi-vehicle households [10]. This negative effect might be explained by the higher number of publicly available charging stations in suburban areas compared to urban areas [59].

6. Conclusions

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 ICEV. A two-step machine learning modeling framework was employed, starting with k-modes clustering to identify five distinct trip clusters that captured the heterogeneity in trips. Subsequently, a decision tree model was employed to predict vehicle choice (EV or ICEV). 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 (with a cross-validation accuracy of 88%) outperformed the binary logit (with a cross-validation accuracy of 57.7%) by a large margin. Notably, gas price on the day of the trip 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. To keep the locations of charging stations consistent with travel behavior, these stations should be placed near popular shopping and dining destinations. Conversely, these stations could be placed near frequent trip attractors (e.g., locations with a high employment density) if policymakers intend to encourage owners to use EVs for non-discretionary/frequent trips. Other factors, such as the number of household workers, income, and gas price on the day of the trip, also exhibited noticeable effects on vehicle choice. Overall, multi-vehicle households with lower income and fewer workers were more inclined to choose EVs for their daily trips. However, these households face challenges in adopting EVs initially due to their 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 USD

per gallon were found to discourage the use of ICEVs 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. Because of the limited number of trips made by PHEVs and BEVs, they were grouped together with HEVs as a single category. This approach might lead to the oversight of potential differences between these vehicle types. This issue can be addressed by acquiring future datasets with a larger representation from each category. The models we selected have shown a strong performance in capturing the relationships between predictor and output variables. However, to address more intricate problems in future studies, integrating advanced techniques such as support vector machines or neural networks will provide valuable insights and further enhance the robustness of the analysis. This study used 2017 NHTS data, which are somewhat outdated. Future studies should use more recent data, such as the 2022 NHTS dataset, which has been recently released. Furthermore, the NHTS 2017 dataset predominantly includes older EV models, which are gradually being replaced by newer models with extended ranges. These limitations of the study should be taken into account by decision makers, and the findings of this study should be complemented with those from future studies before making policy decisions. Such future studies can leverage datasets that encompass trip data from newer EV models, enabling the modeling of vehicle choices for different EV categories separately. While our current study aimed to provide a straightforward analysis as a foundation, we acknowledge the potential for deeper insights. Future studies should use advanced interpretation techniques, such as SHapley Additive exPlanations (SHAP) values, to enhance the clarity and depth of model interpretations. In addition, due to data limitations, this study could not include some important variables, such as charging infrastructure availability. Future studies should consider this variable to develop a more rigorous modeling framework.

Author Contributions: Conceptualization, S.K.M.; Methodology, V.C., S.K.M. and S.H.; Software, V.C.; Validation, V.C. and S.H.; Formal analysis, V.C.; Resources, S.K.M.; Data curation, V.C.; Writing—original draft, V.C.; Writing—review & editing, S.K.M. and S.H.; Project administration, S.K.M.; Funding acquisition, S.K.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research was partly funded by the California Air Resources Board, grant number 2020-1404.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are publicly available.

Acknowledgments: 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.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Zulinski, J. U.S. Leads in Greenhouse Gas Reductions, but Some States Are Falling Behind [WWW Document]. 2018. Available online: <https://www.eesi.org/articles/view/u.s.-leads-in-greenhouse-gas-reductions-but-some-states-are-falling-behind> (accessed on 20 September 2023).
2. U.S. Department of Energy. Alternative Fuels Data Center. 2023. Available online: <https://afdc.energy.gov/vehicles/electric> (accessed on 21 September 2023).
3. Sen, C. *Performance Analysis of Batteries Used in Electric and Hybrid Electric Vehicles*; University of Windsor: Windsor, ON, Canada, 2010.
4. Ghandi, A.; Paltsev, S. Global CO₂ impacts of light-duty electric vehicles. *Transp. Res. D Transp. Environ.* **2020**, *87*, 102524. [CrossRef]
5. Musti, S.; Kockelman, K.M. Evolution of the household vehicle fleet: Anticipating fleet composition, PHEV adoption and GHG emissions in Austin, Texas. *Transp. Res. Part A Policy Pract.* **2011**, *45*, 707–720. [CrossRef]

6. Sherman, L. Implications of current household vehicle ownership and use patterns on the feasibility of electric cars. *Transportation* **1980**, *9*, 209–227.
7. Tamor, M.A.; Milačić, M. Electric vehicles in multi-vehicle households. *Transp. Res. Part C Emerg. Technol.* **2015**, *56*, 52–60. [CrossRef]
8. NHTS. National Household Travel Survey. 2017. Available online: <https://nhts.ornl.gov/> (accessed on 4 March 2023).
9. Li, X.; Liu, C.; Jia, J. Ownership and Usage Analysis of Alternative Fuel Vehicles in the United States with the 2017 National Household Travel Survey Data. *Sustainability* **2019**, *11*, 2262. [CrossRef]
10. Chakraborty, D.; Hardman, S.; Tal, G. Integrating plug-in electric vehicles (PEVs) into household fleets- factors influencing miles traveled by PEV owners in California. *Travel Behav. Soc.* **2022**, *26*, 67–83. [CrossRef]
11. Srinivasa Raghavan, S.; Tal, G. Behavioral and technology implications of electromobility on household travel emissions. *Transp. Res. D Transp. Environ.* **2021**, *94*, 102792. [CrossRef]
12. Karlsson, S. Utilization of battery-electric vehicles in two-car households: Empirical insights from Gothenburg Sweden. *Transp. Res. Part C Emerg. Technol.* **2020**, *120*, 102818. [CrossRef]
13. Bucher, D.; Martin, H.; Hamper, J.; Jaleh, A.; Becker, H.; Zhao, P.; Raubal, M. Exploring Factors that Influence Individuals' Choice Between Internal Combustion Engine Cars and Electric Vehicles. *AGILE GIScience Ser.* **2020**, *1*, 1–23. [CrossRef]
14. Jensen, A.F.; Mabit, S.L. Modelling real choices between conventional and electric cars for home-based journeys. In *Annual Transport Conference at Aalborg University*; Aalborg University: Aarhus, Denmark, 2015.
15. Ozhegov, E.M.; Ozhegova, A. Heterogeneity in demand and optimal price conditioning for local rail transport. *arXiv* **2019**, arXiv:1905.12859v1.
16. Vergis, S.; Chen, B. Understanding Variations in U.S. Plug-in Electric Vehicle Markets. In Proceedings of the Transportation Research Board 94th Annual Meeting, Washington, DC, USA, 11–15 January 2015; Transportation Research Board: Davis, CA, USA, 2014.
17. Egbue, O.; Long, S. Barriers to widespread adoption of electric vehicles: An analysis of consumer attitudes and perceptions. *Energy Policy* **2012**, *48*, 717–729. [CrossRef]
18. Hardman, S.; Chandan, A.; Tal, G.; Turrentine, T. The effectiveness of financial purchase incentives for battery electric vehicles—A review of the evidence. *Renew. Sustain. Energy Rev.* **2017**, *80*, 1100–1111. [CrossRef]
19. Jakobsson, N.; Gnann, T.; Plötz, P.; Sprei, F.; Karlsson, S. Are multi-car households better suited for battery electric vehicles? Driving patterns and economics in Sweden and Germany. *Transp. Res. Part C Emerg. Technol.* **2016**, *65*, 1–15. [CrossRef]
20. Jakobsson, N.; Karlsson, S.; Sprei, F. How are driving patterns adjusted to the use of a battery electric vehicle in two-car households? In Proceedings of the Electric Vehicle Symposium, Montreal, QU, Canada, 19–22 June 2016.
21. Karlsson, S. What are the value and implications of two-car households for the electric car? *Transp. Res. Part C Emerg. Technol.* **2017**, *81*, 1–17. [CrossRef]
22. Mandev, A.; Sprei, F.; Tal, G. Electrification of Vehicle Miles Traveled and Fuel Consumption within the Household Context: A Case Study from California, U.S.A. *World Electr. Veh. J.* **2022**, *13*, 213. [CrossRef]
23. Jakobsson, N.; Sprei, F.; Karlsson, S. How do users adapt to a short-range battery electric vehicle in a two-car household? Results from a trial in Sweden. *Transp. Res. Interdiscip. Perspect.* **2022**, *15*, 100661. [CrossRef]
24. Singh, V.; Singh, V.; Vaibhav, S. A review and simple meta-analysis of factors influencing adoption of electric vehicles. *Transp. Res. D Transp. Environ.* **2020**, *86*, 102436. [CrossRef]
25. Tompkins, M.; Bunch, D.; Santini, D.; Bradley, M.; Vyas, A.; Poyer, D. Determinants of alternative fuel vehicle in the continental united states choice. *Transp. Res. Rec.* **1998**, *1641*, 130–138. [CrossRef]
26. Tanaka, M.; Ida, T.; Murakami, K.; Friedman, L. Consumers' willingness to pay for alternative fuel vehicles: A comparative discrete choice analysis between the US and Japan. *Transp. Res. Part A Policy Pract.* **2014**, *70*, 194–209. [CrossRef]
27. Hackbarth, A.; Madlener, R. Consumer preferences for alternative fuel vehicles: A discrete choice analysis. *Transp. Res. D Transp. Environ.* **2013**, *25*, 5–17. [CrossRef]
28. Caulfield, B.; Farrell, S.; McMahon, B. Examining individuals preferences for hybrid electric and alternatively fuelled vehicles. *Transp. Policy* **2010**, *17*, 381–387. [CrossRef]
29. Carlucci, F.; Cirà, A.; Lanza, G. Hybrid electric vehicles: Some theoretical considerations on consumption behaviour. *Sustainability* **2018**, *10*, 1302. [CrossRef]
30. Shin, H.-S.; Farkas, Z.A.; Nickkar, A. An Analysis of Attributes of Electric Vehicle Owners' Travel and Purchasing Behavior: The Case of Maryland. In Proceedings of the International Conference on Transportation and Development 2019: Innovation and Sustainability in Smart Mobility and Smart Cities, Alexandria, VA, USA, 9–12 June 2019; American Society of Civil Engineers: Reston, VA, USA, 2019; pp. 77–90.
31. Soltani-Sobh, A.; Heaslip, K.; Stevanovic, A.; Bosworth, R.; Radivojevic, D. Analysis of the Electric Vehicles Adoption over the United States. *Transp. Res. Procedia* **2017**, *22*, 203–212. [CrossRef]
32. Perez, J. The Fastest Cars You Can Buy from Every Automaker. 2020. Available online: <https://www.motor1.com/features/428317/fastest-cars-from-every-automaker/> (accessed on 11 November 2023).
33. Jia, J. Analysis of Alternative Fuel Vehicle (AFV) Adoption Utilizing Different Machine Learning Methods: A Case Study of 2017 NHTS. *IEEE Access* **2019**, *7*, 112726–112735. [CrossRef]

34. De Borger, B.; Mulalic, I.; Rouwendal, J. Substitution between cars within the household. *Transp. Res. Part A Policy Pract.* **2016**, *85*, 135–156. [\[CrossRef\]](#)
35. Huang, Z. Extensions to the k-Means Algorithm for Clustering Large Data Sets with Categorical Values. *Data Min. Knowl. Discov.* **1998**, *12*, 283–304. [\[CrossRef\]](#)
36. Chaturvedi, A.; Green, P.E.; Carroll, J.D. K-modes Clustering. *J. Classif.* **2001**, *18*, 35–55. [\[CrossRef\]](#)
37. Cao, F.; Liang, J.; Bai, L. A new initialization method for categorical data clustering. *Expert Syst. Appl.* **2009**, *36*, 10223–10228. [\[CrossRef\]](#)
38. Pandit, S.; Gupta, S. A comparative study on distance measuring approaches for clustering. *Int. J. Res. Comput. Sci.* **2011**, *2*, 29–31. [\[CrossRef\]](#)
39. Rousseeuw, P.J. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* **1987**, *20*, 53–55. [\[CrossRef\]](#)
40. Kim, E.J. Analysis of Travel Mode Choice in Seoul Using an Interpretable Machine Learning Approach. *J. Adv. Transp.* **2021**, *2021*, 1–13. [\[CrossRef\]](#)
41. Zhao, X.; Yan, X.; Yu, A.; Van Hentenryck, P. Prediction and behavioral analysis of travel mode choice: A comparison of machine learning and logit models. *Travel Behav. Soc.* **2020**, *20*, 22–35. [\[CrossRef\]](#)
42. Myles, A.J.; Feudale, R.N.; Liu, Y.; Woody, N.A.; Brown, S.D. An introduction to decision tree modeling. *J. Chemom.* **2004**, *18*, 275–285. [\[CrossRef\]](#)
43. Breiman, L. *Classification and Regression Trees*, 1st ed.; Chapman and Hall/CRC: Boca Raton, FL, USA, 1984.
44. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [\[CrossRef\]](#)
45. Biau, G.; Scornet, E. A random forest guided tour. *Test* **2016**, *25*, 197–227. [\[CrossRef\]](#)
46. Liaw, A.; Wiener, M. Classification and Regression by random Forest. *R News* **2002**, *2*, 18–22.
47. Friedman, J.H. Greedy Function Approximation: A Gradient Boosting Machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [\[CrossRef\]](#)
48. Chen, T.; Guestrin, C. XGBoost: A scalable tree boosting system. In Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Association for Computing Machinery, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794. [\[CrossRef\]](#)
49. Bhat, C.R. Work Travel Mode Choice and Number of Non-Work Commute Stops. *Transp. Res. Part B Methodol.* **1997**, *31*, 41–54. [\[CrossRef\]](#)
50. Yang, Y.; Wang, C.; Liu, W.; Zhou, P. Understanding the determinants of travel mode choice of residents and its carbon mitigation potential. *Energy Policy* **2018**, *115*, 486–493. [\[CrossRef\]](#)
51. Cameron, A.C.; Trivedi, P.K. *Microeconometrics Using Stata*; Stata Press: College Station, TX, USA, 2010.
52. Kim, S.; Ulfarsson, G.F. Travel mode choice of the elderly-effects of personal, household, neighborhood, and trip Characteristics. *Transp. Res. Rec.* **2004**, *1894*, 117–126. [\[CrossRef\]](#)
53. Wang, F.; Ross, C.L. Machine Learning Travel Mode Choices: Comparing the Performance of an Extreme Gradient Boosting Model with a Multinomial Logit Model. *Transp. Res. Rec.* **2018**, *2672*, 35–45. [\[CrossRef\]](#)
54. Apley, D.W.; Zhu, J. Visualizing the effects of predictor variables in black box supervised learning models. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **2020**, *82*, 1059–1086. [\[CrossRef\]](#)
55. Singh, K.V.; Bansal, H.O.; Singh, D. A comprehensive review on hybrid electric vehicles: Architectures and components. *J. Mod. Transp.* **2019**, *27*, 77–107. [\[CrossRef\]](#)
56. Parker, N.; Breetz, H.L.; Salon, D.; Conway, M.W.; Williams, J.; Patterson, M. Who saves money buying electric vehicles? Heterogeneity in total cost of ownership. *Transp. Res. D Transp. Environ.* **2021**, *96*. [\[CrossRef\]](#)
57. Liu, J.; Khattak, A.J.; Li, X.; Fu, X. A spatial analysis of the ownership of alternative fuel and hybrid vehicles. *Transp. Res. D Transp. Environ.* **2019**, *77*, 106–119. [\[CrossRef\]](#)
58. Tiikkaja, H.; Liimatainen, H. Car access and travel behaviour among men and women in car deficient households with children. *Transp. Res. Interdiscip. Perspect.* **2021**, *10*, 100367. [\[CrossRef\]](#)
59. Brown, A.; Cappellucci, J.; Schayowitz, A.; White, E.; Heinrich, A.; Cost, E. *Electric Vehicle Charging Infrastructure Trends from the Alternative Fueling Station Locator: First Quarter 2022*; NREL: Golden, CO, USA, 2022.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.