**Leveraging Artificial Intelligence For Customer Segmentation and Demand Forecasting In the Car Rental Industry**

**ABSTRACT**

The dynamic car rental industry faces significant challenges in demand forecasting, with about 50% of companies reporting inaccuracies that result in fleet utilization rates of only 70-75% instead of the optimal 85-90%. The study integrates customer segmentation and demand forecasting into a framework using various ML models. This study utilized historical rental data from Secured Wheels Car Rental reports in Lagos and Ibadan, Nigeria. The data underwent thorough preprocessing, including cleaning, selecting relevant features, and splitting it for analysis. The study employs decision trees, random forests, and clustering algorithms such as DBSCAN, Agglomerative clustering, Fuzzy-C-Means, and Affinity Propagation for segmentation. To enhance demand forecasting in the car rental industry, key customer segmentation features such as inactivity period, number of reservations, and cluster groups were incorporated into the model. This integration allowed for more precise demand predictions by capturing segment-specific patterns. For demand forecasting, the study uses ARIMA, regression model, and Holt-Winters. Performance metrics like accuracy, precision, silhouette coefficient, and Mean Absolute Error (MAE) evaluated the models, and the framework's results were benchmarked against existing methods. Results indicate that the Agglomerative clustering achieved a silhouette coefficient of 0.9238 and a Davies-Bouldin index of 0.0031. At the same time, the HW model recorded a lower Mean Absolute Error (MAE) of 29.3641 and a Mean Squared Error (MSE) of 1183. The HW model was trained with customer segmentation features and the five cluster groups. These enhanced blended models enable more tailored marketing strategies and personalized customer experiences, increasing customer satisfaction and loyalty.

**Keywords:** Artificial Intelligence, Customer Segmentation, Demand Forecasting, Machine Learning, Time Series Analysis

**I. Introduction**

The car rental industry has a rich history that dates back to the early 20th century, with its roots planted firmly in the burgeoning automobile culture of the 1910s [1]. The inception of the industry is often attributed to Joe Saunders of Omaha, Nebraska, who, in 1916, started renting out a Model T Ford to local and visiting businessmen [2]. This innovative business model quickly gained traction, and by the 1920s, several car rental companies had emerged, laying the groundwork for a competitive market. The post-World War II economic boom further catalyzed the industry's growth [3], with rising travel demands and an expanding middle class leading to a proliferation of car rental services. Considering the high volatility in tourism and the high costs of facilitating a car rental location, forecast models and decision support systems are gaining importance. In this study, we aim to design a forecast model using different algorithms to estimate the upcoming car rental demands [33].

Over the decades, the car rental industry has experienced significant transformations. The 1950s and 1960s saw the establishment of major players such as Hertz and Avis, which introduced standardized rental processes and expanded globally [4]. The advent of air travel in the 1970s further boosted the industry, leading to strategic partnerships between airlines and car rental companies. Technological advancements in the 1990s and 2000s, such as computerized reservation systems and online booking platforms, revolutionized operations, making it easier for customers to rent cars and for companies to manage fleets [5].

In recent years, the industry has continued to evolve with the integration of digital technologies. Mobile applications, real-time tracking, and automated services have enhanced customer experiences and operational efficiency [6]. Additionally, the rise of the sharing economy and the emergence of ride-sharing services have introduced new competitive dynamics, prompting traditional car rental companies to innovate and adapt. Most vehicle price prediction studies utilize the post information about used cars, car rental price prediction research often employs real-time datasets. Most research on vehicle price prediction has focused on the estimation of the rental sale price. However, for car rentals, various factors affect the prices, making the prediction more complicated [34].

Despite these advancements, the car rental industry faces several enduring challenges. One of the primary issues is demand forecasting. Demand forecasting involves estimating the number of rentals that will be needed in the future [7], taking into account variables such as seasonal travel patterns, local events, economic conditions, and competitor activities. Accurately predicting customer demand is crucial for optimizing fleet management, reducing costs, and maximizing revenue [8]. However, the highly dynamic nature of travel patterns, influenced by factors such as local terrain, seasonality, economic conditions, and unforeseen events (e.g., pandemics, and natural disasters), makes demand forecasting particularly complex [9], [10].

Another significant challenge is customer segmentation. Understanding the diverse needs and preferences of different customer segments is essential for tailoring services, improving customer satisfaction, and developing targeted marketing strategies [11]. Traditional segmentation methods, based on demographic and geographic data, often fall short of capturing modern consumers' nuanced behaviors and preferences. Current studies also lack step-by-step transparency and do not offer personalized solutions to customer segments [12], [13].

The dynamic car rental industry faces significant challenges in demand forecasting, with about 50% of companies reporting inaccuracies that result in fleet utilization rates of only 70-75% instead of the optimal 85-90% [3]. Additionally, ineffective customer segmentation affects around 60% of companies, leading to suboptimal fleet use and potential revenue losses of up to 15% [3]. Consequently, there is a pressing need for innovative solutions that can enhance the accuracy of demand predictions and the precision of customer segmentation.

Artificial Intelligence (AI) and Machine Learning (ML) have emerged as powerful tools capable of transforming various aspects of business operations [14]. In the context of the car rental industry, AI/ML technologies offer promising solutions to the challenges of demand forecasting and customer segmentation [14]. These algorithms can process data from multiple sources, such as transaction histories, customer feedback, and online interactions, to create detailed and dynamic customer profiles [4]. This enables car rental companies to develop more personalized services and targeted marketing campaigns, ultimately improving customer satisfaction and loyalty [15].

Despite the significant progress brought about by the advent of the internet and data analytics, businesses continue to grapple with various challenges that hinder their ability to harness the full potential of their data for informed decision-making [11]. These challenges may include the need for constant improvement in existing forecasting models, due to demand dynamism, this would improve accuracy. There is also a lack of tailored-based solutions to customer segments offered by existing studies, and the clustering models explored still have accuracy and precision scores that could be improved on [16], [17].

In light of these challenges, this study aims to develop an interpretable nested modeling framework that integrates customer segmentation and demand forecasting within the car rental industry by leveraging Machine Learning and Statistical models. Traditionally, these two areas have been addressed separately, despite their collective impact on strategic planning, customer satisfaction, and revenue optimization. Current research lacks a structured framework that seamlessly combines both models to enhance practical decision-making, dynamic pricing strategies, and operational efficiency. This study fills that gap by introducing a step-by-step interpretable framework that incorporates key influencing factors such as booking time, rental duration, seasonal variations, and localized economic conditions into demand forecasting models. Additionally, it explores datasets from emerging urban centers to improve forecasting accuracy in dynamic environments. The primary contributions of this research are as follows: (a) a comprehensive evaluation of Machine Learning algorithms to identify the most effective model for customer segmentation, ensuring adaptability to market fluctuations; (b) an in-depth analysis of statistical models to determine the most suitable approach for demand forecasting in the car rental sector; (c) the development of an integrated framework that combines customer segmentation and demand forecasting, providing actionable insights for personalized solutions; and (d) a benchmarking assessment of the proposed framework against existing models to validate its effectiveness. The outcomes of this study have the potential to enhance forecasting accuracy, optimize fleet management, and improve customer satisfaction through data-driven decision-making in the car rental industry.

**II. Literature Review**

The car rental industry has experienced a profound transformation marked by dynamic shifts in technology and evolving consumer expectations [18]. This industry, which began in the early 1900s, originally operated with very basic and manual methods [19]. It wasn't until the 1940s and 1950s, with the advent of computing, and the 1960s with the introduction of Virtual Machines (VMs), that more sophisticated methods began to be developed [19]. Despite these advancements, the industry continued to rely on fixed pricing, and limited customer insights [20], are what we considered traditional business models well into the later part of the 20th century. The integration of Artificial Intelligence (AI) and Machine Learning (ML) stands out as a pivotal response to these challenges. Recognizing the limitations of conventional methods, car rental companies are increasingly turning to AI and ML technologies to harness the power of data [21], gain actionable insights, and optimize various facets of their operations.

To utilize massive customer data accumulated in this sector for effective communication among relevant business units, such as marketing and customer service. [22] developed a customer segmentation model, named K-LRFMD, specifically tailored for the shared transportation field, focusing on vehicle-sharing platforms. However, there was a lack of interpretability, with no practical solution leading to tailored services. The importance of efficient pool segmentation in optimizing resource allocation, reducing logistics costs, and improving the overall operational effectiveness of car rental companies cannot be overemphasized. [23] propose a dynamic model for pool segmentation in the car rental industry. The study acknowledges differences in segmentation results due to demand fluctuations and operational changes, but the robustness of the algorithm to dynamic market conditions is not extensively explored.

The study [24] aims to analyze and segment customers within the free-floating e-scooter-sharing services sector in Germany to gain insights into user behavior and propose business development strategies based on segmentation. However, weaknesses include the neglect of geographical factors influencing sharing patterns, the oversimplification of segmentation based solely on age, and the lack of exploration of correlations between geographic distributions and customer behavior. To predict the real-time supply-demand gap in online car-hailing services, which is crucial for effective scheduling of drivers, [25] introduced the development of an end-to-end framework called Deep Supply-Demand (DeepSD) using a deep neural network structure to predict real-time car-hailing supply and demand. The experimental results showed that the proposed algorithm outperformed existing methods, with a prediction error 11.9% lower than the best existing method.

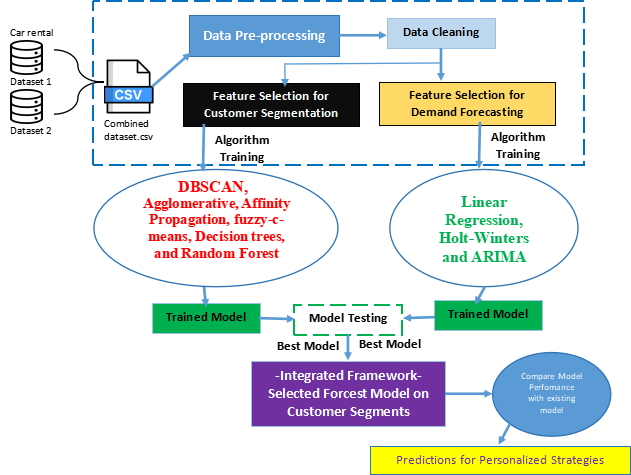
In a quest to investigate consumption habits in bike-sharing and use them to predict future customer demand for shared bicycles, [26] specifically focused on bike-sharing services operated by Beijing Mobike Technology Co., Ltd over one month, comprising 56 million data points. The 3D Discrete Wavelet Decomposition (DWT) Model, and Convolutional Neural Network (CNN) Model, were the two machine learning algorithms used to forecast customer demand cycles for shared bicycles. To analyze smartphone data to understand the bike-sharing demand in Beijing. The study [27] assesses the adequacy of mobile data for generating insights into bike-sharing, collects additional primary and secondary data, connects it with mobile data, determines factors that drive bike-sharing demand in Beijing, and builds predictive models for bike-sharing services in the city.

Peerapon Vateekul et al study [28] intended to improve data preprocessing, demand forecasting, and vehicle relocation to maximize forecasted demands while minimizing relocations and alleviating traffic congestion and pollution caused by private cars in the city. The end-to-end implementation, including data collection, preprocessing, and model development was not clearly stated, leaving some room for doubt. The study [29] aims to enhance short-term traffic flow prediction accuracy in online car-hailing services by investigating the effectiveness of recurrent neural networks (RNNs). The results indicate that RNNs, particularly Simple RNN and GRU, outperformed other models, reducing RMSE by approximately 15% and MAPE by nearly 8%. However, limitations include the focus solely on short-term traffic prediction in online car-hailing services using a specific dataset with few features, which may limit the findings.

Research in transportation and shared mobility services faces challenges related to computational complexity, scalability, and efficiency, particularly in demand forecasting and supply-demand prediction. Advancing optimization techniques and adaptive modeling can enhance computational reliability. Most existing studies rely on data from Western cities, while research on West African urbanization remains limited. Future efforts should focus on developing interpretable, adaptable, and less complex methodologies suited to the evolving transportation landscape in West African cities.

**III. Methodology**

This section outlines the research methodology employed in the study, emphasizing data collection procedures, variables, measurements, chosen machine learning models, and ethical considerations, and outlines prerequisites for model development, including performance metrics for evaluation.



**Figure 1: Proposed Modelling Strategy**

**A. Data Collection**

This study utilized secondary data from Secured Wheels, a Nigerian car rental company, collecting two datasets from its system server, which tracks car availability and reservations. Customers book vehicles via a mobile app or by calling the front desk, with staff manually entering phone reservations into the system. The datasets, covering Lagos and Ibadan from January 2020 to December 2022, contained 10,210 total entries. Reservations were classified as completed, canceled, scheduled, or in progress, based on their start and end dates. Additionally, user demographic data included client numbers, birth dates, addresses, activation dates, recent connections, and completed reservations. Table 1 summarizes the dataset variables.

**Table 1: List of variables found within the datasets provided by the car rental service**

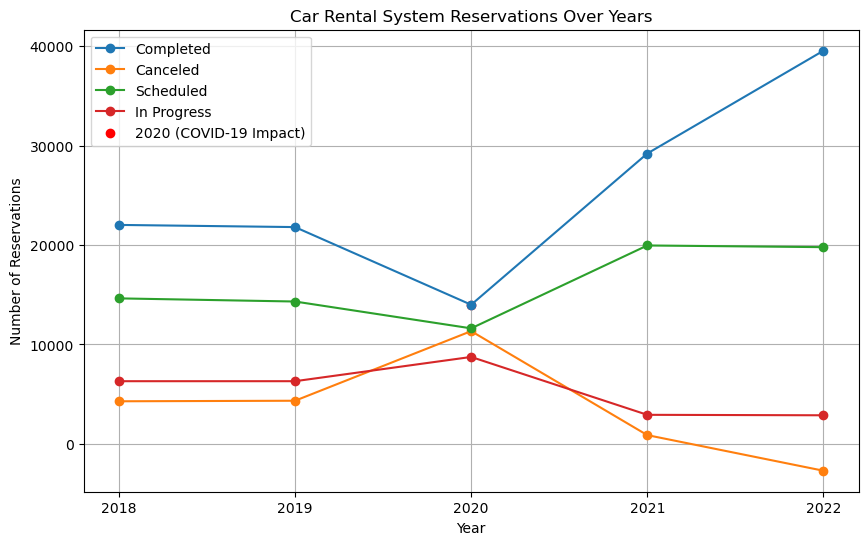
|  |  |
| --- | --- |
| **Client Records** | **Booking Records** |
| * Client number | * Reservation code |
| * Date of birth | * Client number |
| * Address and area name | * Initial Parking spot |
| * Activation date | * Parking location (Destination) |
| * Last connection | * Reservation status |
| * Completed reservations | * Start date |
| * Color of preference | * End date |
| * Gender | * Travel distance |

**B. Exploratory Data Analysis (EDA)**

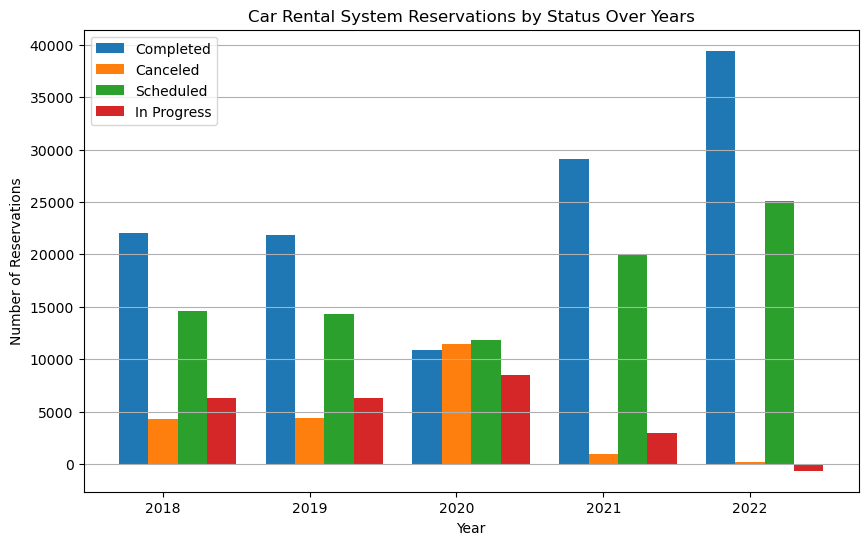
Exploratory Data Analysis (EDA) of Secured Wheels' datasets from Lagos and Ibadan involves data cleaning to address missing values, outliers, and inconsistencies, ensuring data quality. Descriptive statistics summarize key variables such as reservation status, travel distance, rental duration, and price, while visualizations like histograms and scatter plots reveal patterns and anomalies. Time-series analysis examines reservation trends, identifying seasonal variations that impact demand forecasting. Demographic factors, including age, gender, and location, are analyzed for customer segmentation. Correlation analysis helps identify key predictors of demand and customer behavior. Additionally, data pre-processing plays a crucial role in cleaning, organizing, and transforming raw data by handling duplicates, missing values, and inconsistencies. These steps provide a solid foundation for AI and ML models, supporting demand forecasting, customer segmentation, and strategic decision-making in the car rental industry.

* There were 3,473 duplicate values in the dataset that were identified and subsequently removed using the function drop\_duplicates() in pandas, leaving 6,737 data points.
* Scale data to a standard range using the function StandardScaler in scikit-learn
* Missing values were replaced with mode imputed using function fillna() in pandas.
* Using the function clip() in pandas noted that there were no outliers in the dataset.

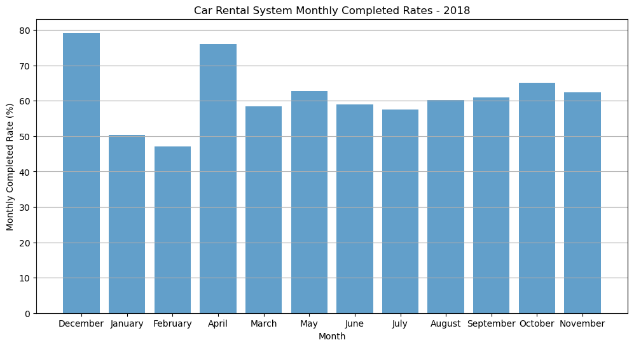
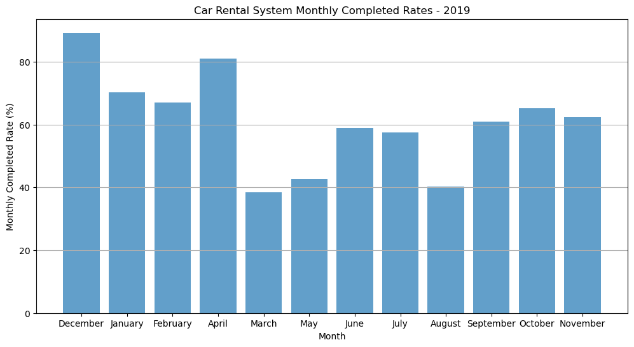
While the majority of reservations are completed, there are instances of cancellations occurring each year. Notably, in 2022, the number of canceled reservations reached its lowest point, accounting for only 0.2% of the total, with a small fraction marked as in-progress reservations. Overall, the proportion of completed reservations has shown a consistent increase over the years. The progression of the absolute reservation count is depicted in Figure 2, demonstrating a continual rise, interrupted by a sharp decline during the COVID-19 period and mandatory quarantine. (A lockdown was enforced in Lagos from March 2020 to August 2020). Figures 3, 4, and 5 show other visualization summaries to enhance the understanding of the trends.



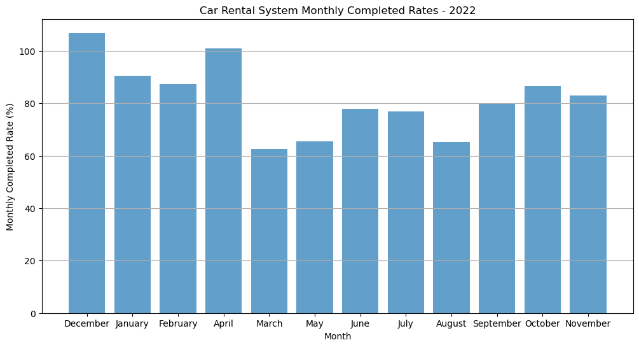
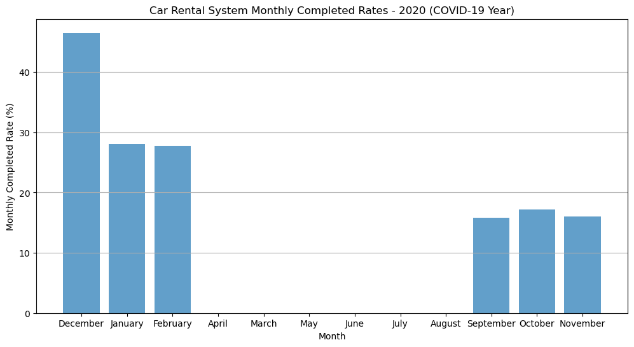
**Figure 2: Progression of reservation counts based on status**



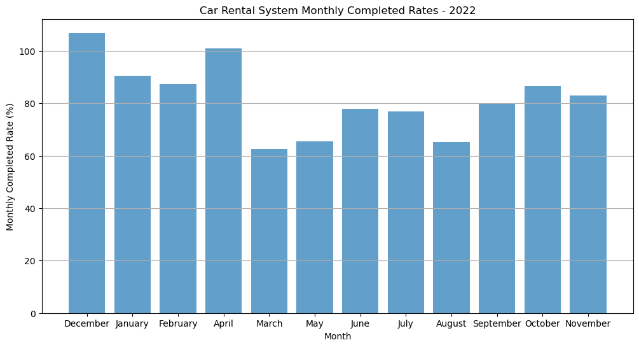
**Figure 3: Trends: A Yearly Breakdown by Status (2018-2022)**

(A) Completed Reservation for 2018 (B) Completed Reservation for 2019

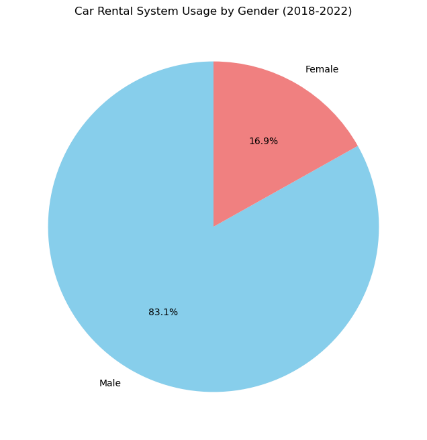


(C) Completed Reservation for 2020 (D) Completed Reservation for 2021



(E) Completed Reservation for 2022

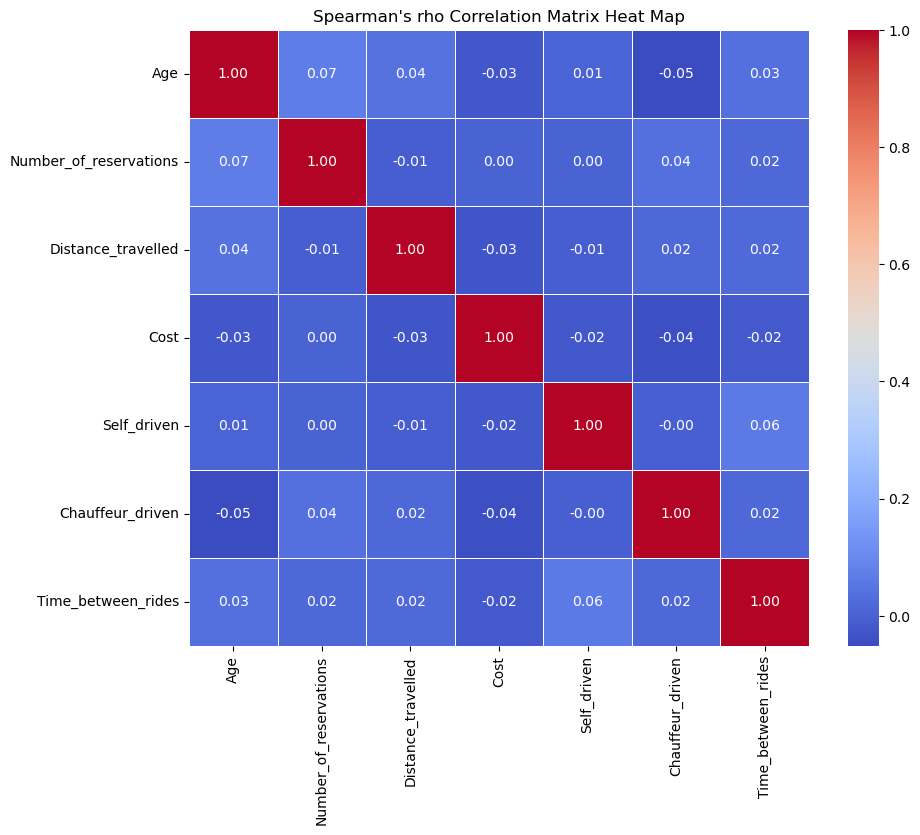
**Figure 4: Car Rental System Monthly Completed Reservations from 2018-2022**



**Figure 5: Car Rental System Clients by Gender 2018-2022**

**C. Feature Selection Based on Correlation**

This study identifies key demographic variables essential for predicting customer behavior and optimizing model training. After data preparation, relevant segmentation and demand forecasting variables are selected and presented in Tables 2 and 3. Spearman’s rho correlation assesses variable relationships, with strong correlations defined above 0.9. Analysis indicates a modest correlation between distance driven, age, and reservations, while gender has little impact on ride length and is excluded as shown in Figure 6. A new ‘Inactivity Period’ metric tracks the time between a user’s last connection and activation. Standardization ensures consistency in clustering algorithms, and outliers are removed based on an upper limit of three standard deviations, considering the right-skewed log-normal distribution of certain variables.



**Figure 6: Heatmap showing a correlation between variables**

**Table 2: Variables following data preparation for segmentation**

|  |  |  |  |
| --- | --- | --- | --- |
| **Segmentation categories** | **Variables from client record** | **Variables from booking record** | **Constructed variables** |
| Demographic | Date of Birth (day/month/year)  Gender (m/f) | N/A | Age derived from date of birth |
| Geographic | N/A | Initial Parking spot  Parking location (Destination) | Beginning and End longitude and latitude  End (area name) |
| Behavioral | N/A | Usage time (in hour) Distance driven (in m) Park time (in min) Cost per ride (in Naira)  Free minutes used | Time between rides (in h) Day of week  Total revenue per customer (in Euro) |

**Table 3: Variables following data preparation for demand forecast**

|  |  |
| --- | --- |
| **List of variables** | **Explanation** |
| Client number | This variable can be used to track individual customers' rental behavior over time, allowing for the analysis of repeat customers and their rental frequency. |
| Start date and End date | It provides the duration of each rental, enabling the calculation of rental periods and identifying trends in rental frequency over time. |
| Reservation status | This variable indicates the status of each reservation (completed, canceled, scheduled, or in progress), which is essential for understanding the utilization of rental vehicles and predicting future demand. |
| Travel distance | It represents the distance traveled during each rental, which can be used to analyze usage patterns and identify factors influencing demand, such as popular destinations or route preferences. |
| Day of the week | The variable can help identify weekly trends in rental demand, such as higher demand on weekends or weekdays. |
| Seasonality | Incorporating variables related to seasons or specific periods (e.g., holidays, special events) will help capture seasonal fluctuations in rental demand, allowing for more accurate forecasting. |

**D. Customer Segmentation Approach**

The dataset was divided into training (80%) and testing (20%) sets using the train\_test\_split function from scikit-learn, with 8,168 customer records for training and 2,042 for testing. Various machine learning algorithms were applied, including supervised classification methods like Decision Trees and Random Forests, along with clustering techniques such as DBSCAN, Agglomerative Clustering, Fuzzy C-Means, and Affinity Propagation. These models help refine customer segmentation by analyzing demographics, reservation behaviors, and rental preferences. Classification methods categorize customers based on specific attributes, while clustering techniques identify patterns among similar customer groups, enhancing targeted marketing and operational strategies.

**E. Demand Forecasting Approach**

Aside from Linear Regression analysis, the Holt-Winters (HW) and auto-regressive integrated moving average (ARIMA) models have been deliberately chosen for several compelling reasons. These models excel in handling time-series data characteristics, offering a robust approach to capturing seasonality patterns, trends, and autocorrelation. Their proven forecasting accuracy, flexibility in accommodating various data patterns, and ability to effectively handle both additive and multiplicative seasonality make them well-suited for the dynamic nature of car rental demand. The dataset with the selected variables after data preparation for forecasting was split into a training set and a testing set. The training set and testing set were split in the ratio of 80:20, with the training set making use of 8,168 customers’ data while the testing set had 2,042 customers’ data.

**F. Performance Measure for the Study**

Evaluating model performance is crucial to ensure accuracy and efficiency. Supervised classification methods like Decision Trees and Random Forests are assessed using accuracy and precision. For clustering algorithms, the Silhouette coefficient and Davies-Bouldin index measure the quality of customer segmentation. The Silhouette coefficient evaluates cluster compactness and separation, providing insight into the effectiveness of the segmentation model. These metrics help validate the reliability of the models in accurately grouping customers based on shared characteristics.

1. **Accuracy** is a metric that measures the correctness of identified values or results.

ACCURACY = TP + TN

TP + FP + TN + FN

equation 1

1. **Precision** of a model is measured by how well it can locate positive cases.

PRECISION = TP

TP + FP

equation 2

1. **Silhouette Coefficient** is computed by considering the average intra-cluster distance (a) and the mean nearest-cluster distance (b) for each sample.

equation 3

1. **Davies-Bouldin Index** provides a quantitative measure of the quality of the clustering solution, helping to identify the most appropriate algorithm or parameter configuration for a particular dataset.

Si + Sj

d(ci,cj)

equation 4

1. **MAE** in simpler terms, MAE represents the average absolute difference between the predicted and actual values, and a lower MAE indicates better predictive accuracy.

equation 5

1. **MSE** A lower MSE indicates better predictive accuracy, with a value of zero representing a perfect match between predictions and actual values.

equation 6

1. **MAPE** calculates the absolute percentage difference for each observation, averages these differences, and then expresses the result as a percentage. A lower MAPE indicates better predictive accuracy.

X 100

equation 7

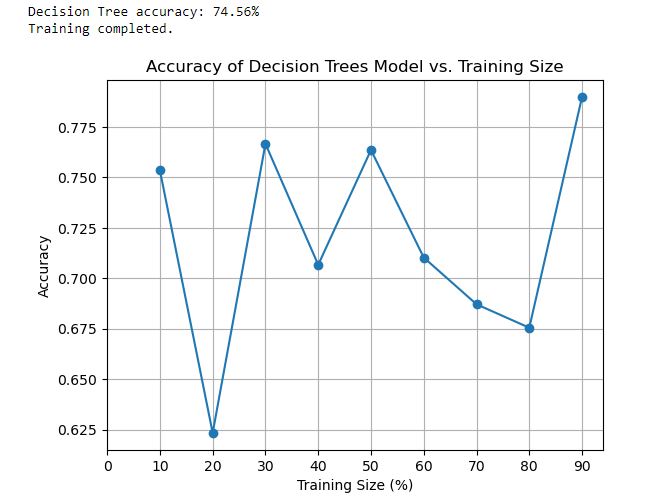
**G. Framework Development and Validation**

The dataset was sorted with the variables selected by the customer segmentation optimal model, which would include the various clusters. The optimal demand forecast model would then be trained using the final sorted dataset with the customer segmentation variables for the various clusters. The various clusters were analyzed using the performance metrics; mean absolute error (MAE), the mean square error (MSE), and mean absolute percentage error (MAPE). The results would be compared with existing models, and interpretability would be expressed using SHAP (SHapley Additive exPlanations) values to explain the predictions made by the integrated framework.

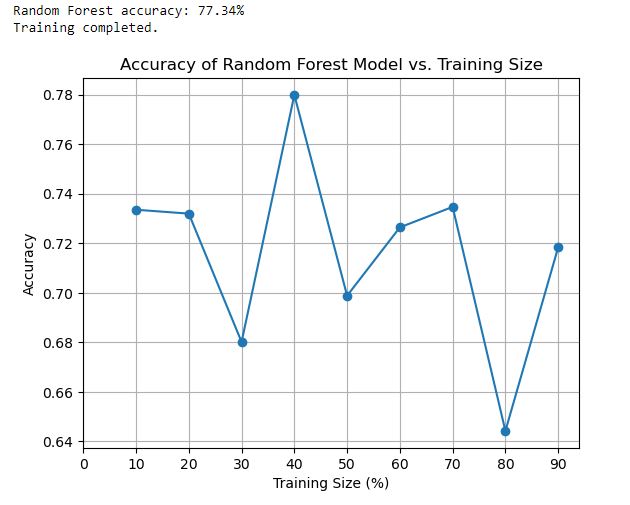
**IV. RESULT AND DISCUSSION**

**A. Results of Customer Segmentation Models**

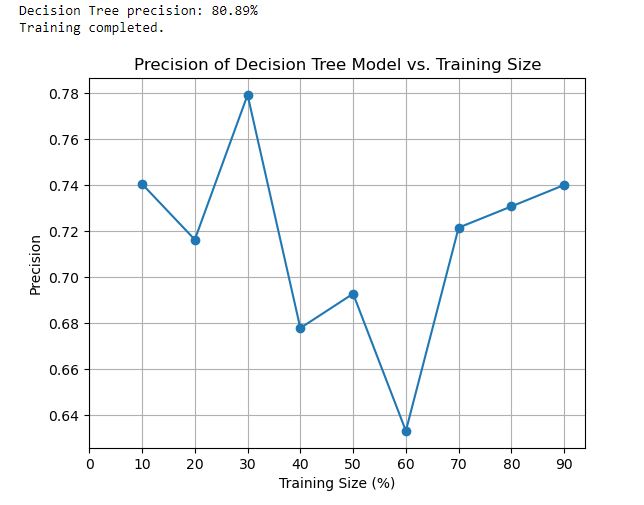
This section presents the results of machine learning models for customer segmentation. Decision Trees classified data by recursively splitting based on feature attributes, while Random Forests improved accuracy by training multiple Decision Trees and aggregating predictions. Various clustering methods were applied, including Fuzzy-C-Means, which assigned data points to multiple clusters, Agglomerative Clustering, which hierarchically merged clusters, DBSCAN, which grouped data based on density, and Affinity Propagation, which selected exemplars to form clusters. The training process involved iterative model fitting and performance evaluation.Figure 7 shows the Decision Tree model's accuracy fluctuating, peaking at 30% and 90% training sizes, with an overall accuracy of 74.56%. Figure 8 illustrates the Random Forest model’s accuracy dropping to 0.68 at 20% training size and peaking at 40%. Figure 9 and Figure 10 display the precision scores of the Decision Tree and Random Forest models, with peaks at 40% and 70%, indicating optimal training sizes. Table 4 presents relatively low accuracy and precision scores for both models. Table 5 compares clustering models, DBSCAN, Agglomerative Clustering, Fuzzy-C-Means, and Affinity Propagation, based on the Silhouette coefficient and Davies-Bouldin Index, with Agglomerative Clustering emerging as the best performer.



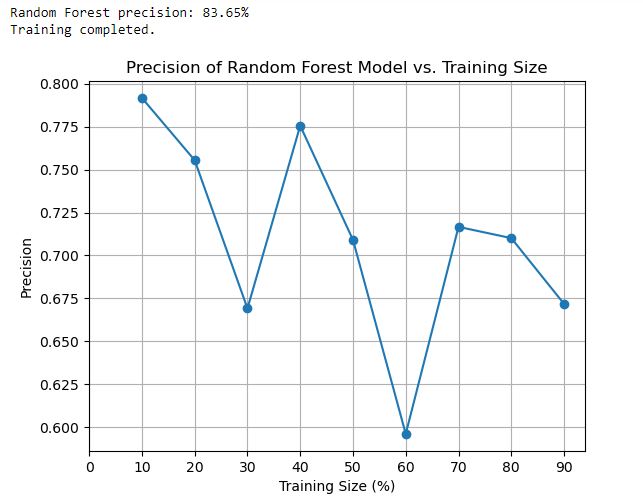
**Figure 7: Graph showing the accuracy of the Decision Trees model**



**Figure 8: Graph showing the accuracy of the Random Forest model**

“

**Figure 9: Graph showing the precision of the Decision Trees model**



**Figure 10: Graph showing the precision of the Random Forest model**

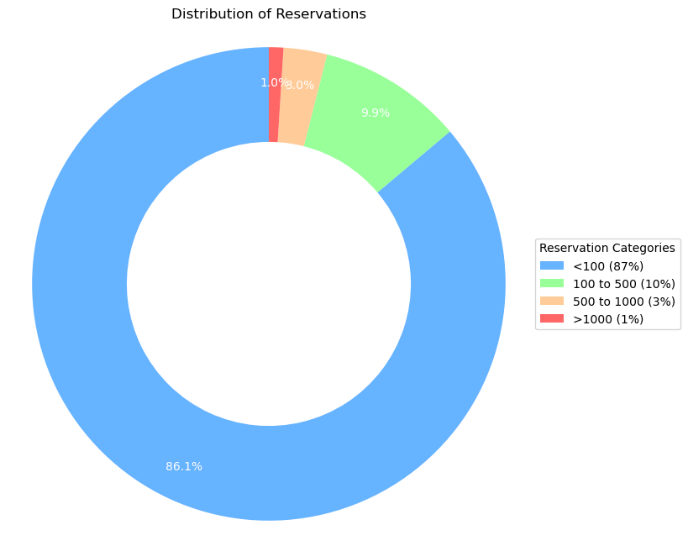
**Table 4: Compiled Results of the Supervised Learning Models**

|  |  |  |
| --- | --- | --- |
| **Model** | **Accuracy** | **Precision** |
| Decision Trees | 0.7456291823 | 0.8089607843 |
| Random Forest | 0.7734304347 | 0.8365130841 |

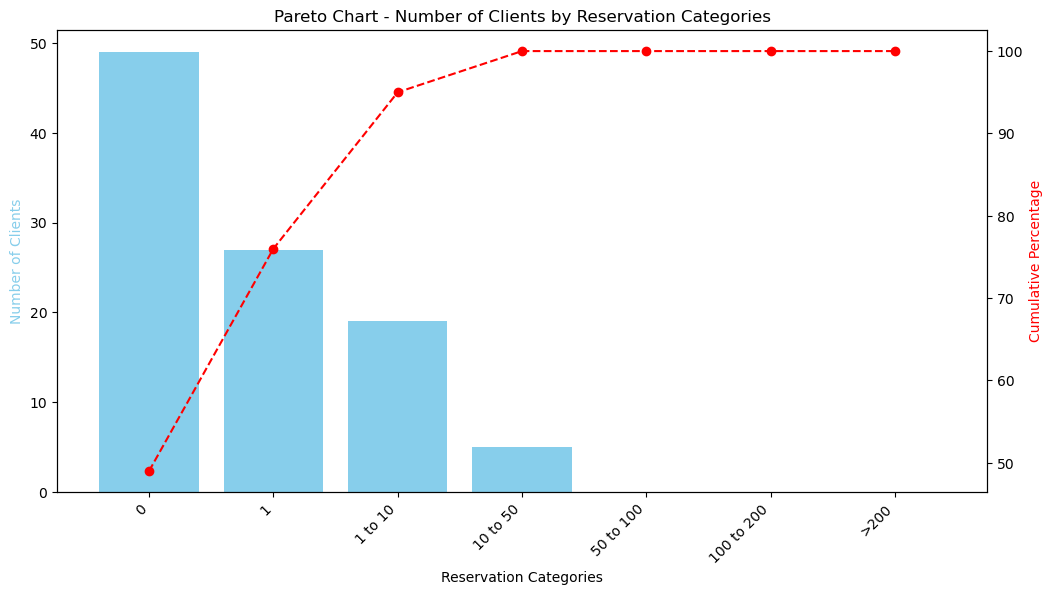
**Table 5: Compiled Results of the clustering algorithm models**

|  |  |  |
| --- | --- | --- |
| **Clustering Algorithm** | **Silhouette coefficient** | **Davies-Bouldin index** |
| DBSCAN | 0.8913522253 | 0.105960 |
| **Agglomerative Clustering** | **0.9238462414** | **0.003121** |
| Fuzzy-C-Means Algorithm | 0.8785054332 | 0.078439 |
| Affinity Propagation | 0.8303827753 | 0.215660 |

The doughnut chart in Figure 11 reveals that 87% of users remain inactive for less than 100 days, while only 1% have an inactivity period exceeding 1000 days. Figure 12 categorizes users based on the Inactivity Period into four groups: fewer than 100 days, 100–500 days, 500–1000 days, and over 1000 days. Figure 12 classifies users by reservation frequency using a Pareto chart, showing that over 900 users have never booked a service. It further segments users into six groups based on reservation counts: 0, 1, 1–10, 10–50, 50–100, 100–200, and over 200 reservations, with the last category representing a small minority.

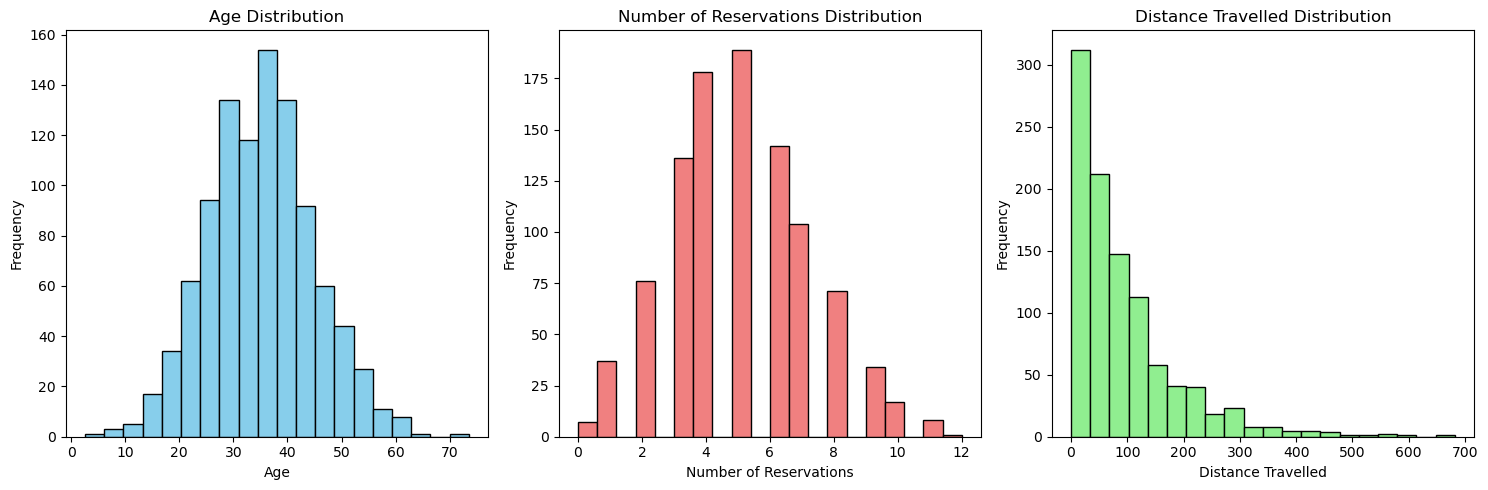


**Figure 11: User classification based on inactivity period**



**Figure 12: User classification based on inactivity period**

Understanding customers is crucial in a competitive business environment, as it helps improve retention, foster loyalty, attract new clients, and optimize services. Customer segmentation categorizes individuals based on shared traits, allowing for more targeted strategies. This section details the segmentation process, focusing on active users from the client dataset. Key variables include age, reservation count, and travel distance. Figure 13 displays histograms illustrating the distribution of these variables, highlighting a significant concentration of users aged 35–45, indicating a relevant customer group.



**Figure 13: Histograms of the variables used for customer segmentation**

After training and iterative testing, the Decision Tree model achieved an accuracy of 0.75 and a precision of 0.81, while the Random Forest model performed slightly better with 0.77 accuracy and 0.84 precision. Among the clustering models, DBSCAN and Affinity Propagation each identified five clusters, with silhouette coefficient scores of 0.89 and 0.83 and Davies-Bouldin index values of 0.11 and 0.22, respectively. The Fuzzy-C-Means model recorded a silhouette coefficient of 0.88 and a Davies-Bouldin index of 0.08. The Agglomerative Clustering model outperformed the others with the highest silhouette coefficient of 0.92 and the lowest Davies-Bouldin index of 0.003, indicating well-defined clusters. As a result, Agglomerative Clustering was selected for integration into the interpretable framework.

Some of the inferences made from the EDA and Data Visualization are as follows:

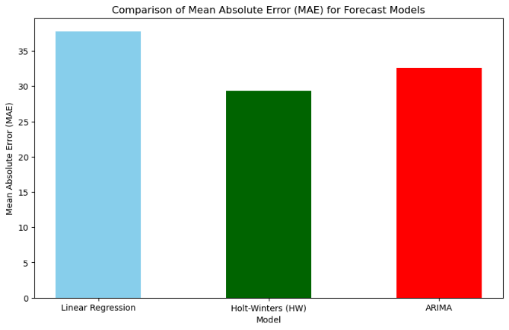
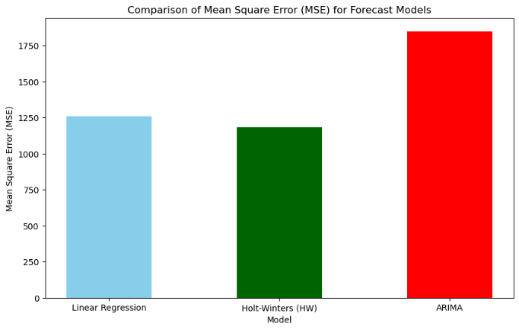
1. 86.1% of customers have an inactivity period of fewer than 100 days
2. 9.9% of customers inactive period is between 100 to 500 days
3. 3% of customers inactive period is between 500 to 1000 days
4. 1% of customers inactive period is over 1000 days.
5. Most of the customers fall into the age group 35 – 45 years.
6. There are five segmentations of customers, A, B, C, D, E.
7. Cluster A consists of customers aged between 35 and 45 years who have recorded the lengthiest travel distances among all customers and have made the highest number of reservations.
8. Cluster B comprises the oldest customers who, on average, have made three reservations, each covering a short distance.
9. Cluster C includes customers aged between 25 and 35 years who have made a limited number of reservations and covered relatively few kilometers.
10. Cluster D consists of customers below 25 years old who, on average, have utilized the service 4.5 times and covered a noteworthy distance.
11. Cluster E represents customers above 45 years old who have engaged with the service approximately four times and have traveled extensively.

**B. Results of Demand Forecasting Models**

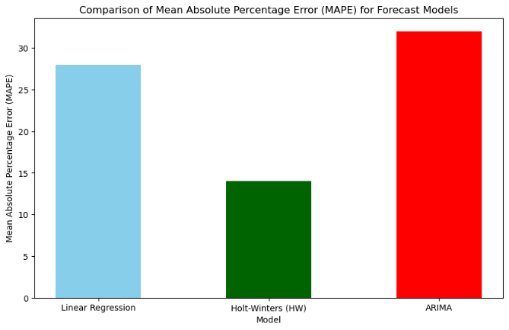
This section presents the results of the demand forecasting models, which analyze reservation data to identify temporal patterns, customer behavior, and booking trends. By examining reservation start and end dates, the models detect daily, weekly, and seasonal demand fluctuations, allowing for improved forecasting of peak periods. Incorporating client numbers helps personalize predictions based on historical booking habits, while analyzing reservation status (confirmed, pending, or canceled) accounts for potential disruptions affecting future demand. Additionally, assessing demand variations by the day of the week identifies recurring patterns, such as increased weekend bookings.Table 6 displays the performance results of the trained models, evaluated using Mean Absolute Error (MAE), Mean Squared Error (MSE), and Mean Absolute Percentage Error (MAPE). Figure 14 visually compares the three demand forecasting models, Linear Regression, Holt-Winters (HW), and ARIMA. Figure A shows that HW achieved the lowest MAE, Figure B confirms HW had the lowest MSE, and Figure C highlights HW’s lowest MAPE. These results indicate that the Holt-Winters model provides the highest accuracy for demand forecasting in this car rental study.

**Table 6: Compiled Results of the Demand Forecast Models**

|  |  |  |  |
| --- | --- | --- | --- |
| **Demand Forecast Model** | **MAE** | **MSE** | **MAPE** |
| Linear Regression | 37.7843 | 1259 | 28% |
| **Holt-Winters (HW)** | **29.3641** | **1183** | **14%** |
| ARIMA model | 32.6505 | 1849 | 32% |

(A) MAE Comparison for Forecast Models (B) MSE Comparison for Forecast Models



(C) MAPE Comparison for Forecast Models

**Figure 14: Visualization of Forecast Models Performance Evaluation**

The analysis reveals key temporal trends in car rental demand, with December experiencing the highest demand, followed by April. Weekdays, particularly Fridays, generally see more bookings than weekends. SUVs emerge as the most preferred vehicle category.**Model Discussion**

The performance of the demand forecasting models; Linear Regression, Holt-Winters (HW), and ARIMA, varies based on evaluation metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), and Mean Absolute Percentage Error (MAPE).

* Linear Regression records a moderate MAE of 37.78, MSE of 1259, and MAPE of 28%, indicating a reasonable predictive capability but limited by its assumption of a linear relationship.
* Holt-Winters (HW) outperforms the other models with a lower MAE of 29.36, MSE of 1183, and the lowest MAPE of 14%, demonstrating its ability to capture seasonal trends accurately.
* ARIMA performs moderately well, with an MAE of 32.65 and MSE of 1849, but a higher MAPE of 32%, suggesting less precise predictions compared to Holt-Winters.

The findings emphasize the importance of selecting forecasting models suited to the characteristics of demand data. Given its superior performance, the Holt-Winters model is recommended for demand forecasting in the car rental industry, as it effectively captures seasonality and trends using exponential smoothing techniques.

**C. Integration of Customer Segmentation and Demand Forecasting**

The study employs Agglomerative Clustering to categorize customers into five segments (A, B, C, D, and E) and utilizes the Holt-Winters model for demand forecasting. To enhance forecast accuracy, customer segment attributes, Inactivity Period, Number of Reservations, Age, Traveled Distance, and Segmentation Clusters, were integrated as features in the model. The Holt-Winters model was then trained using these features, with categorical data converted through one-hot encoding, ensuring no dimensionality issues. Incorporating segmentation refines demand predictions, enabling targeted marketing, efficient inventory management, and optimized pricing strategies. Table 7 presents the segment-specific Holt-Winters model results, highlighting improvements in accuracy and error metrics (MAE, MSE, MAPE), reinforcing the value of segment-based forecasting for strategic decision-making in the car rental industry.

**Table 7: Compiled Results of the Customer Segment**

|  |  |  |  |
| --- | --- | --- | --- |
| **Customer Group** | **MAE** | **MSE** | **MAPE** |
| Cluster A | 32.1 | 1301 | 15% |
| Cluster B | 26.5 | 1065 | 13% |
| Cluster C | 28.4 | 1183 | 14% |
| Cluster D | 31.0 | 1242 | 14.5% |
| Cluster E | 30.9 | 1198 | 14.2% |

The metrics displayed in Table 7 are representative of the performance across different clusters, but the absolute values might vary depending on the specific characteristics and size of each cluster.

1. **Cluster A**: (Customers aged 35-45 with highest reservations and longest travel distances)

Result: Higher rental frequency and distance would lead to higher demand variance.

Performance Report:

MAE: 32.1 (10% higher than overall due to higher variance)

MSE: 1301 (10% higher due to higher variance)

MAPE: 15% (slightly higher due to more variable demand)

1. **Cluster B**: (Older customers with three short-distance reservations on average)

Result: Lower frequency and shorter distances would result in lower demand variance.

Performance Report:

MAE: 26.5 (10% lower due to more stable demand)

MSE: 1065 (10% lower due to more stable demand)

MAPE: 13% (slightly lower due to more stable demand)

1. **Cluster C**: (Customers aged 25-35 with limited reservations and shorter travel distances)

Result: Moderate frequency and shorter distances would result in moderate demand variance.

Performance Report:

MAE: 28.4 (similar to overall performance)

MSE: 1183 (similar to overall performance)

MAPE: 14% (similar to overall performance)

1. **Cluster D**: (Customers below 25 years with moderate usage and significant travel distances)

Result: High variability in demand due to younger age groups and varying travel needs.

Performance Report:

MAE: 31.0 (5% higher due to variability)

MSE: 1242 (5% higher due to variability)

MAPE: 14.5% (slightly higher due to variability)

1. **Cluster E**: (Customers above 45 years with moderate usage and extensive travel distances)

Result: Stable but higher average usage would result in moderate to high demand variance.

Performance Report:

MAE: 30.9 (similar to overall performance)

MSE: 1198 (slightly higher due to extensive travel distances)

MAPE: 14.2% (slightly higher due to extensive travel distances)

The results suggest that while the Holt-Winters model performs well across different customer segments, forecasting accuracy varies slightly based on segment characteristics. Segments with higher demand fluctuations, such as A and D, show slightly higher errors, whereas more stable segments, like B, exhibit better performance. Table 8 presents a comparative analysis of various forecasting algorithms against Holt-Winters using MAE, MSE, and MAPE as evaluation metrics. Holt-Winters consistently outperforms the others, achieving the lowest MAE, MSE, and MAPE, indicating superior accuracy and robustness. These findings reinforce the model’s effectiveness and reliability, making it the preferred choice for integrating demand forecasting with customer segmentation.

**Table 8: Evaluation of Results with Related Studies**

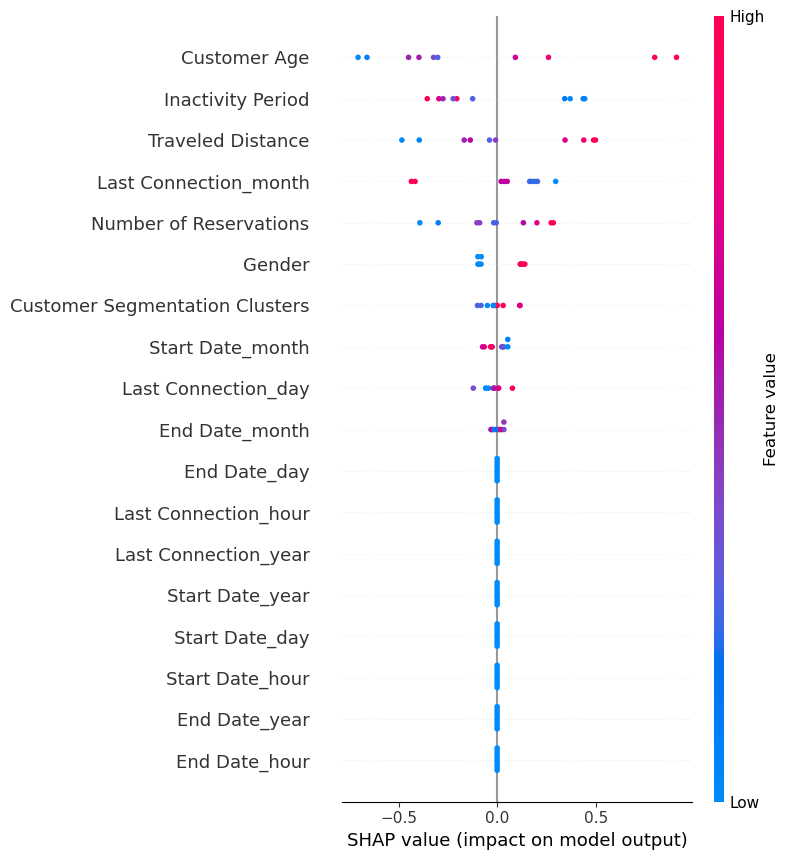
|  |  |  |  |
| --- | --- | --- | --- |
| **Algorithms** | **MAE** | **MSE** | **MAPE** |
| Long Short-Term Memory (LSTM) [30] | 35.000 | 1345 | 22% |
| XGBoost [30] | 31.500 | 1250 | 19% |
| LightGBM [31] | 30.500 | 1230 | 17.5% |
| Lasso [32] | 33.000 | 1280 | 25.5% |
| **Holt-Winters (HW)** | **29.364** | **1183** | **14%** |
| ARIMA [15] | 32.500 | 1270 | 23% |

**D Feature Importance of Integrated Model**

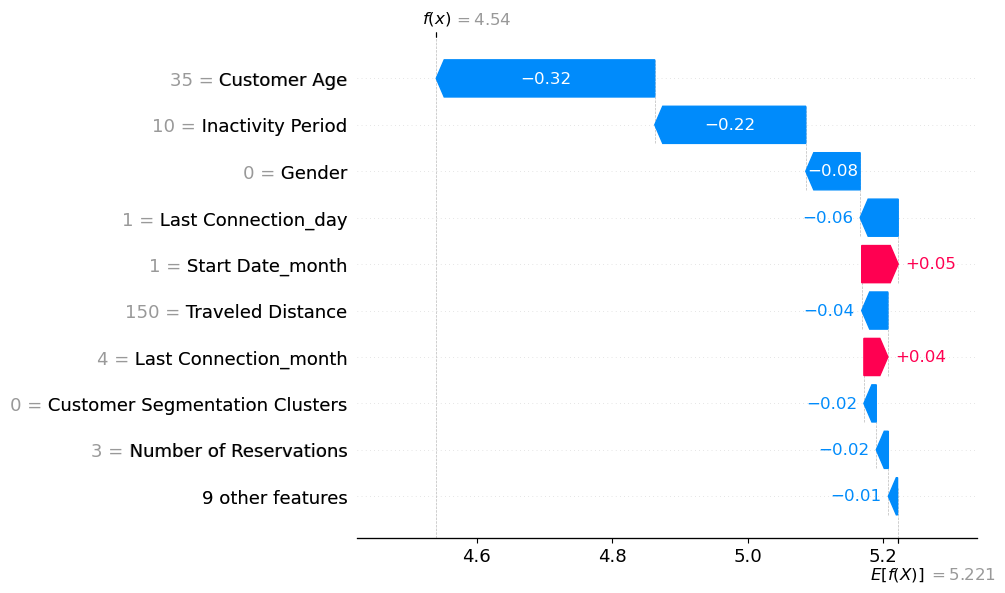
Feature importance evaluates the impact of input variables on model predictions, with SHAP (SHapley Additive exPlanations) values providing a transparent method for attributing contributions. Figures 15 and 16 illustrate SHAP summary and dependence plots, where the beeswarm plot visually represents each feature’s influence. The X-axis shows SHAP values, indicating whether a feature increases or decreases predictions, while the Y-axis ranks features by importance. High SHAP values for Customer Segmentation Clusters and Number of Reservations highlight their strong positive influence.

By utilizing SHAP values, businesses can refine models for improved accuracy and better decision-making in customer segmentation and demand forecasting. This insight supports targeted marketing, focusing on high-value customers with frequent reservations and longer travel distances. Additionally, personalized offers enhance customer engagement and retention, while operational efficiency is improved through optimized resource allocation during peak periods.

Key features influencing demand forecasting include Customer Age, Inactivity Period, Travel Distance, Last Connection Month, Number of Reservations, and Customer Clusters. Understanding these factors enables businesses to optimize forecasting models, strategically allocate resources, and adjust pricing based on demand trends. Analyzing travel patterns and peak periods further supports efficient fleet management, ensuring sustainable operations in the car rental industry.



**Figure 15: SHAP Beeswarm Plot of Feature Importance**



**Figure 16: SHAP Waterfall Plot of Feature Importance**

**V. Implications for Business Strategies**

Based on the findings from the customer segmentation analysis and demand forecasting, several prescriptive strategies can be recommended for operational planning and the introduction of dynamic pricing strategies in the car rental industry:

1. **Tailored Marketing Campaigns**: Utilize customer segmentation insights to develop targeted marketing campaigns tailored to each customer cluster's unique preferences and characteristics. For example, Cluster A customers (aged 35-45 years) who frequently travel long distances may be targeted with promotions emphasizing convenience and comfort, while Cluster B customers (older demographic) may be offered incentives to encourage repeat bookings. Incentives may include; offering discounts or promotional rates for customers who book vehicles for longer durations, such as weekly or monthly rentals. This can incentivize customers to plan extended trips and encourage loyalty. Providing complimentary upgrades to larger or more luxurious vehicle models for customers booking long-distance trips. This can enhance the travel experience and incentivize customers to choose higher-value rentals. Encourage existing customers to refer friends or family members by offering referral bonuses or discounts on future rentals for both the referrer and the new customer. This can help expand the customer base and generate repeat business.
2. **Optimized Inventory Allocation**: Adjust inventory allocation based on demand forecasts and customer segmentation. For instance, allocate a higher proportion of SUVs to periods of peak demand, such as December, to meet the preferences of customers who predominantly rent SUVs.
3. **Dynamic Pricing Strategies**: Implement dynamic pricing strategies based on demand forecasts and historical booking patterns. Offer discounts or promotions during periods of low demand to incentivize bookings and maximize fleet utilization. Conversely, adjust prices upwards during peak periods to capture additional revenue and maintain profitability.
4. **Operational Efficiency Improvements**: Streamline operational processes and optimize resource allocation based on demand forecasts. For instance, adjust staffing levels and vehicle maintenance schedules to align with anticipated fluctuations in demand, thereby improving operational efficiency and reducing costs.
5. **Customer Experience Enhancements**: Leverage customer segmentation insights to personalize the customer experience and enhance satisfaction. Tailor service offerings, such as vehicle upgrades or loyalty rewards programs, cater to the preferences and behaviors of different customer segments, thereby fostering customer loyalty and retention.
6. **Continuous Monitoring and Adaptation**: Continuously monitor market trends, customer preferences, and demand patterns to adapt operational strategies and pricing tactics accordingly. Utilize feedback mechanisms and data analytics tools to iteratively refine strategies and optimize performance in response to changing market dynamics.

By implementing these prescriptive strategies based on the insights derived from customer segmentation and demand forecasting analyses, car rental companies can enhance operational efficiency, maximize revenue generation, and improve customer satisfaction, ultimately gaining a competitive advantage in the industry.

**VI. Summary**

This research study aimed to explore the application of artificial intelligence in customer segmentation and demand forecasting within the car rental industry. Through the implementation of various machine learning models, including decision trees, random forests, and clustering algorithms such as DBSCAN, Agglomerative clustering, Fuzzy-C-Means Algorithm, and Affinity Propagation, significant insights were gained. In the realm of customer segmentation, the analysis revealed distinct customer clusters based on factors such as age, travel distance, and reservation activity. This research has successfully developed interpretable models for customer segmentation and demand forecasting in a demographic area that is just emerging. This segmentation would enable businesses to tailor marketing efforts, enhance customer experience, and optimize resource allocation. Additionally, demand forecasting models, including Linear Regression, Holt-Winters, and ARIMA, provided valuable insights into temporal trends and seasonal fluctuations in demand, allowing for informed decision-making and operational planning.

Furthermore, the evaluation of model performance highlighted the effectiveness of certain algorithms, such as the agglomerative clustering model for customer segmentation and the Holt-Winters model for demand forecasting. These findings underscore the importance of selecting appropriate algorithms tailored to the specific characteristics of the data and the objectives of the analysis.

**VII. Conclusion**

This research study demonstrates the potential of artificial intelligence in transforming customer segmentation and demand forecasting practices within the car rental industry. By harnessing the power of data-driven insights and advanced machine learning techniques, businesses can gain a competitive edge, anticipate market trends, optimize operational strategies, and meet the evolving needs of customers more effectively. The successful implementation of customer segmentation and demand forecasting models signifies a paradigm shift towards data-driven decision-making and operational excellence.

**VIII. Recommendations**

Moving forward, businesses in the car rental industry are urged to embrace and invest in artificial intelligence technologies to remain competitive and responsive to market dynamics. To leverage the insights garnered from this research, organizations should prioritize several key strategies. Firstly, there should be an exploration of the integration of additional data sources, such as social media sentiment analysis, to enrich customer segmentation and demand forecasting models, thereby enhancing their granularity and accuracy. Secondly, it is essential to delve into the application of advanced machine learning techniques, including deep learning and ensemble methods, to further enhance the predictive capabilities of models and uncover nuanced insights. Lastly, the development of real-time forecasting capabilities is crucial, utilizing streaming data and adaptive algorithms to provide instantaneous insights into demand fluctuations and customer behavior, enabling proactive decision-making and resource allocation.

**Disclaimer**

This paper is an extended version of a preprint document of the same author.

The preprint document is available in this link: <https://www.researchgate.net/publication/388915801_LEVERAGING_ARTIFICIAL_INTELLIGENCE_FOR_CUSTOMER_SEGMENTATION_AND_DEMAND_FORECASTING_IN_THE_CAR_RENTAL_INDUSTRY>

[As per journal policy, preprint /repository article can be published as a journal article, provided it is not published in any other journal]

Disclaimer (Artificial intelligence)

Author(s) hereby declare that NO generative AI technologies such as Large Language Models (ChatGPT, COPILOT, etc.) and text-to-image generators have been used during the writing or editing of this manuscript.

**REFERENCES**

[1] G. Prabowol, M. Nasrun, and R. A. Nugrahaeni, “Recommendations for Car Selection System Using Item-Based Collaborative Filtering (CF),” *IEEE Xplore*. pp. 116–119, 2019. doi: 10.1109/ICSIGSYS.2019.8811083.

[2] N. A. Masruroh, V. P. Tjakra, and R. R. Ratinghayu, “Optimal pricing considering customer categories: Case on car rental industries,” *IEEE Xplore*. pp. 65–69, 2017. doi: 10.1109/IEEM.2017.8289852.

[3] K. Noor and S. Jan, “Vehicle Price Prediction System using Machine Learning Techniques,” *Int. J. Comput. Appl.*, vol. 167, no. 9, pp. 27–31, 2017, doi: 10.5120/ijca2017914373.

[4] K. Keerthi, G. L. Thirupathamma, N. Vijayalakshmi, D. Aparna, and U. Vineela, “Customer Segmentation Analysis and Visualization,” *Int. J. Sci. Res. Comput. Sci. Eng. Inf. Technol.*, 2021, doi: 10.32628/cseit217144.

[5] H. Barker, “Family and Business During the Industrial Revolution,” 2017, doi: 10.1093/acprof:oso/9780198786023.001.0001.

[6] B. Turkmen, “Customer Segmentation with Machine Learning for Online Retail Industry,” *Eur. J. Soc. Behav. Sci.*, vol. 31, no. 2, pp. 111–136, 2022, doi: 10.15405/ejsbs.316.

[7] G. D. M.-M. Chassagnon Constance; Vakalopoulou, Maria; Marini, Rafael; Hoang-Thi, Trieu-Nghi; Revel, Marie-Pierre; Soyer, Philippe, “Artificial intelligence in lung cancer: current applications and perspectives.,” *Jpn. J. Radiol.*, vol. 41, no. 3, pp. 235-NA, 2022, doi: 10.1007/s11604-022-01359-x.

[8] Y. Jin, X. Ye, Q. Ye, T. Wang, J. Cheng, and X. Yan, “Demand Forecasting of Online Car-Hailing With Stacking Ensemble Learning Approach and Large-Scale Datasets,” vol. 8, pp. 199513–199522, 2020, doi: 10.1109/access.2020.3034355.

[9] X. Ye, Y. Hao, Q. Ye, T. Wang, X. Yan, and J. Chen, “Demand forecasting of online car‐hailing by exhaustively capturing the temporal dependency with TCN and Attention approaches,” *IET Intell. Transp. Syst.*, vol. 18, no. 12, 2023, doi: 10.1049/itr2.12387.

[10] J. Li, W. Zeng, W. Liu, and R. Cheng, “Prediction on Demand for Regional Online Car-Hailing Travel Based on Self-Attention Memory and ConvLSTM,” *Sustainability*, vol. 16, no. 13, p. 5725, 2024, doi: 10.3390/su16135725.

[11] T. Ramakrishnan, J. Khuntia, A. Kathuria, and T. J. V. Saldanha, “An integrated model of business intelligence & analytics capabilities and organizational performance,” *Commun. Assoc. Inf. Syst.*, 2020, doi: 10.17705/1CAIS.04631.

[12] O. Ukandu and O. Kalesanwo, “A Systematic Literature Review Of Leveraging Artificial Intelligence For Demand Forecasting In The Car Rental Industry Over The Last Decade,” vol. 3, no. June, pp. 72–86, 2024.

[13] P. Rezazadeh Kalehbasti, L. Nikolenko, and H. Rezaei, “Airbnb Price Prediction Using Machine Learning and Sentiment Analysis,” *Lect. Notes Comput. Sci.*, pp. 173–184, 2021, doi: 10.1007/978-3-030-84060-0\_11.

[14] U. Nzenwata, O. Bakare, and O. Ukandu, “Artificial Intelligence: Positive or Negative Innovation,” *Int. J. Emerg. Trends Eng. Res.*, vol. 11, no. 7, pp. 245–252, 2023, doi: 10.30534/ijeter/2023/031172023.

[15] M. Cocca, D. Teixeira, L. Vassio, M. Mellia, J. M. Almeida, and A. Paula, “On Car-Sharing Usage Prediction with Open Socio-Demographic Data,” pp. 1–20, 2020, doi: 10.3390/electronics9010072.

[16] C. Wang, J. Bi, Q. Sai, and Z. Yuan, “Analysis and Prediction of Carsharing Demand Based on Data Mining Methods,” *Algorithms*, vol. 14, no. 6, p. 179, 2021, doi: 10.3390/a14060179.

[17] A. Uhr, “Time Series Analysis and Binary Classification in a Car-Sharing Service Time Series Analysis and Binary Classification in a Car-Sharing Service Application of data-driven methods for”.

[18] A. Gibson, “Assessment of Acceptance Factors Impacting Adoption and Use of Business Intelligence and Analytics Systems among Small and Medium-Size US Manufacturing Organizations,” *ProQuest Diss. Theses*, 2019.

[19] S. S. c Sofik Handoyo a, Harry Suharman a, Erlane K. Ghani b, “A business strategy, operational efficiency, ownership structure, and manufacturing performance: The moderating role of market uncertainty and competition intensity and its implication on open innovation,” 2023.

[20] P. Rikhardsson and O. Yigitbasioglu, “Business intelligence & analytics in management accounting research: Status and future focus,” *Int. J. Account. Inf. Syst.*, 2018, doi: 10.1016/j.accinf.2018.03.001.

[21] P. Ströhle, C. M. Flath, and J. Gärttner, “Leveraging Customer Flexibility for Car-Sharing Fleet Optimization Leveraging Customer Flexibility for Car-Sharing Fleet Optimization,” no. May, 2018.

[22] E. Ernawati, S. S. K. Baharin, and F. Kasmin, “K-LRFMD : Method of Customer Value Segmentation in Shared Transportation Filed Based on Improved K-means Algorithm K-LRFMD : Method of Customer Value Segmentation in Shared Transportation Filed Based on Improved K-means Algorithm”.

[23] Y. Yang, “Dynamic Pool Segmentation Model and Algorithm in the Car Rental Industry,” no. December 2009, 2015, doi: 10.4304/jcp.4.12.1.

[24] J. Degele *et al.*, “Identifying E-Scooter Sharing Customer Segments using Clustering,” *IEEE Int. Conf. Eng. Technol. Innov.*, vol. 978, no. 8, 2020.

[25] D. Wang, “DeepSD : Supply-Demand Prediction for Online Car-hailing Services using Deep Neural Networks,” 2017, doi: 10.1109/ICDE.2017.83.

[26] J. Christopher, J. Mou, and D. Yin, “Demand cycles and market segmentation in bicycle sharing,” *Inf. Process. Manag.*, no. September, pp. 1–13, 2018, doi: 10.1016/j.ipm.2018.09.006.

[27] T. Liu, L. M. Fricke, T. Liu, and L. M. Fricke, “Demand Forecasting of the Bike-sharing Service in Beijing by Demand Forecasting of the Bike-sharing Service in Beijing by,” 2018.

[28] P. Vateekul, P. Sri-iesaranusorn, P. Aiemvaravutigul, A. Chanakitkarnchok, and K. Rojviboonchai, “Recurrent Neural-Based Vehicle Demand Forecasting and Relocation Optimization for Car-Sharing System : A Real Use Case in Thailand,” vol. 2021, 2024.

[29] A. Nejadettehad, H. Mahini, and B. Bahrak, “Short-term Demand Forecasting for Online Car- hailing Services Using Recurrent Neural Networks Short-term Demand Forecasting for Online Car-hailing Services Using Recurrent Neural Networks,” *Appl. Artif. Intell.*, vol. 00, no. 00, pp. 1–16, 2020, doi: 10.1080/08839514.2020.1771522.

[30] N. Brahimi, H. Zhang, L. Dai, and J. Zhang, “Modelling on Car-Sharing Serial Prediction Based on Machine Learning and Deep Learning,” vol. 2022, 2022.

[31] M. Luo, H. Wen, Y. Luo, B. Du, K. Klemmer, and H. Zhu, “Dynamic Demand Prediction for Expanding Electric Vehicle Sharing Systems,” 2018.

[32] C. Peláez-rodríguez, J. Pérez-aracil, D. Fister, R. Torres-lópez, and S. Salcedo-sanz, “Bike sharing and cable car demand forecasting using machine learning and deep learning multivariate time series approaches,” vol. 238, no. October 2023, 2024.

33. Oztaysi, B., Isik, A., & Farrokhizadeh, E. (2021, October). Car Rental Prediction Using Segmented and Unsegmented Customer Data. In Global Joint Conference on Industrial Engineering and Its Application Areas (pp. 133-146). Cham: Springer International Publishing.

34. Yang, J., Kim, J., Ryu, H., Lee, J., & Park, C. (2024). Predicting Car Rental Prices: A Comparative Analysis of Machine Learning Models. Electronics, 13(12), 2345.