**DATA EXPLORATION Project Part3**

**Submitted by**

**Habeeb Omotunde, 200584861**

1. **Introduction**

In this project, hypothesis testing will be executed on using the FINER questions asked earlier in Data Exploration 2 (DE2). As a brief reminder, in **DE2**, univariate descriptive statistics of 10 attributes with a mix of 7 dependent and 3 independent variables was executed. The variable types spanned across 6 qualitative & 4 quantitative variables.

Of the 10 selected variables, only 8 (4 of each quantitative and qualitative) of these variables, as highlighted in red, were treated. Note that all necessary coding and categorization have been executed in the spread sheet titled, “**myCarRentalDatav3.xlsx**”.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **#** | **Variable** | **Description** | **Type** | **Dependency** |
| 1 | **Fuel Type** | This variable represents the type of fuel that powers the vehicle. This ranges between gasoline, electric and hybrid | Categorical, Nominal. | Dependent Variable |
| 2 | **Location state** | This variable tracks the state in which the car is located within the United States. The city variable depends on the state as state can have one or more cities. | Categorical, Nominal | Independent Variable |
| 3 | **Vehicle make** | This variable tracks the car brand | Categorical, Nominal | Independent Variable |
| 4 | **Vehicle type** | This variable tracks the type of vehicle rented out to customers. It could be SUV, Sedan, Truck, Van, Minivan etc. | Categorical, Nominal | Dependent Variable |
| 5 | **Rating** | This is a quantitative measure of user satisfaction within a scale of 1.00 to 5.00. Currently, 490 observations are not rated | Numerical, Continuous | Dependent Variable |
| 6 | **Renter Trips Taken** | Number of trips the rented car has embarked upon. | Numerical, Discrete | Dependent Variable |
| 7 | **Rate daily** | This is the daily charge for renting each car, It depends on several variables like the year, make, model, fuel type etc. | Numerical, Discrete | Dependent Variable |
| 8 | **Vehicle year** | The year the vehicle was made. The current range of inventory is 1955 -2020 | Numerical, Discrete | Independent Variable |
| 9 | **Location city** | This variable tracks the city in which the car is located within the United States | Categorical, Nominal. | Dependent Variable |
| 10 | **Vehicle model** | This is variable tracks the model of a car brand. | Categorical, Nominal | Dependent Variable |

1. **Dataset Properties**

* **Number of Selected Variables**: 10
* **Source**: Kaggle: https://www.kaggle.com/datasets/kushleshkumar/cornell-car-rental-dataset
* **Assumptions about dataset:** This dataset is believed to be authoritative as it includes current information concerning trends and insight on car rentals by customers across the United States. Startups can investigate this to make decisions regarding which states are most suitable & profitable to run a pilot test of the rental service. However, a trivial number of ratings and few data points necessary to make these decisions are missing and will be removed during data transformation stage. Therefore, more data processing and cleaning must be executed to prepare the data for usage.

By no means is this dataset exhaustive but it covers states and cities interesting enough to derive insights and information about how car rental & sharing services can help solve the problem of mobility while providing a decent profit for investors at the same time.

1. **Univariate Descriptive statistics**

Out of 10 variables, 8 attributes were selected for the analysis with detailed charts in myCarRentalData.xslx file.

The selected variable are ***Fuel Type, Location state, Vehicle make, Vehicle type, Rating, Renter Trips Taken, Daily Rate & Vehicle year.*** On the spread sheet, the dataset used for Data Exploration 2 located in the sheet named “DE2\_dataset.

* 1. **Fuel Type**

A screenshot of a computer

Description automatically generatedThe dataset and the bar chart shows clearly that most vehicles used by customers were powered by gasoline. This insight helps rental outlets to decide on which vehicles to invest more based on fuel type.

* 1. **Location State**

This chart shows that California, Florida and Texas respectively are the most patronized states in the business of car rentals. Many other factors might influence this such as tourism, business, Technological advancement, proliferation of startups etc. This is a very important variable in the decision-making process

* 1. A table with numbers and text

     Description automatically generated**Vehicle Make/ Brand**

An interesting insight from the variable shows that not only do customers love gasoline cars, but they also have huge interest in using electric cars as well. This obviously has influenced the automobile market and manufacturers to invest in more research and development to hasten the rate of innovation in this area to satisfy customer demand.

By brand irrespective, showing the top 20 most patronized brands in the table here, Tesla electric vehicles take the lead followed by Toyotas and BMW vehicles. An aggregation of all fuel powered cars will definitely overshadow the electric counterparts.

This is why it's important for these variables to be considered both induvial and collectively to derive concise insights from the dataset.

* 1. A screenshot of a computer

     Description automatically generated**Vehicle Type**

This variable shows that 63% of customers prefer normal cars while a larger portion, approximately 29% opted for SUVs which may be based on needs such as family, comfort, weather etc. As we narrow down, it is safe to say that most customers engaged in borrowing cars (92%) use vehicles such as cars and SUVs.

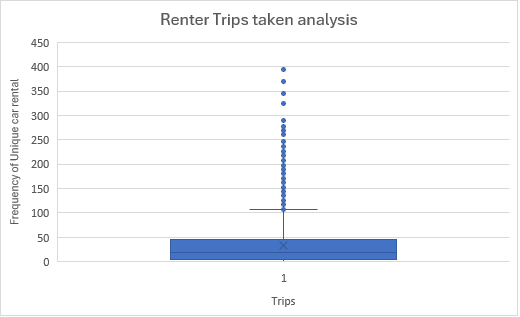
Such insight will guide investors on what vehicle types to purchase. This is obvious in the United States and Canada where certain car types are becoming very scarce and expensive.

* 1. **Rating**

A screenshot of a computer

Description automatically generatedThe rental services have a huge percentage of excellent rating according to the observations in the dataset noting that the worse cases are majorly cases where rating were absent. Of the observed 504 worse cases, only 14 were reported by customers while the rest were not specified at all. Overall, over 90% of rental services were rated excellent.

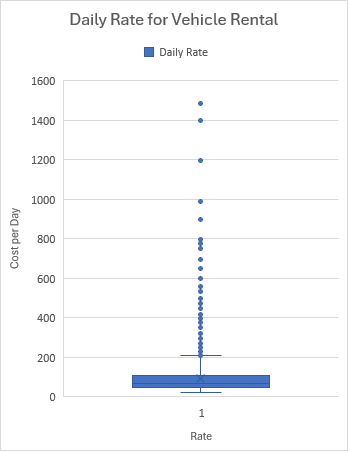
* 1. **Renters Trip Taken**



This numerical variable tends to ascertain the number of unique trips each car has taken when borrowed by customers. Looking closely at the box and whiskers chart, we'd notice many vehicles tend to be overused by rental outlets for many reasons hence giving these vehicles higher trips than others that barely get rented out.

A screenshot of a computer screen

Description automatically generatedThis might be better understood when this variable is compared with the daily rates. A low daily rate will attract more customers hence the huge value, Now, can we consider these values outside the upper as an outlier? That depends on further analysis that helps us to understand the dataset better without losing important observations that could help improve the quality of rental services.

* 1. **Daily Rate**

Quite a few of the rates fall in the region of outliers however careful observation of the dataset and other circumstances suggest otherwise. While the cheapest rate is $20 and most expensive is $1500, we must consider the car type. These could be exotic cars or collection vehicles that are very rare hence the cost. They are barely borrowed compared to cheap cars.

A screenshot of a computer screen

Description automatically generated

Nonetheless, the average rate is very close to the daily cost of rentals in North America in the range of $80 to $100. Furthermore, the median rate is $69 while the mode is $35. This shows that most customers borrow cheaper cars, and such insight could influence investment in vehicles of such grade.

* 1. **Vehicle Year**

A graph with numbers and text

Description automatically generatedA table with numbers and numbers

Description automatically generatedIt is not surprising that customers prefer very recent vehicles as they give peace of mind and comfort. Any issues with recent cars can be easily remedied with the insurance and warranty that come along with it due to the year of the car. Quite obviously, the most rented vehicles were from 2016 to 2020 with 2018 taking the highest number of rents.

1. **Finer Questions**
   1. Does increasing the distribution of rental cars across the state promote access to cost-effective car rental services?
   2. What is the total costs of ownership associated with buying and owning a car, compared to the expenses incurred through frequent car rentals over a specified period?
   3. To what degree do these costs vary across different locations and regions considering frequently visits? if any?
   4. How often will the rental services be required for an extended period versus short-term or occasional trips?
   5. What is the costs of limitations or challenges associated with renting a car in certain regions, such as availability during peak seasons or restricted rental policies?.
2. **Hypothesis testing**

|  |  |  |
| --- | --- | --- |
| **RQ1:** | Does increasing the distribution of rental cars across the state promote access to cost-effective car rental services? |  |
| **H0**  **(Null Hypothesis):** | Increasingthe distribution of rental cars across the state promotes access to cost-effective car rental services. | Odd Ratio, Risk Ratio |
| **RQ2** | What is the total costs of ownership associated with buying and owning a car, compared to the expenses incurred through frequent car rentals over a specified period? |  |
| **H0**  **(Null Hypothesis):** | The total costs of ownership associated with buying and owning a car, compared to the expenses incurred through frequent car rentals over a specified period is $35000 | **Chi Square Test** |
| **RQ3** | To what degree do these costs vary across different locations and regions considering frequently visits? |  |
| **H0**  **(Null Hypothesis):** | **The degree of variation is very minimal if the frequency is high** | **T-Test** |
| **RQ4** | **How** often will the rental services be required for an extended period versus short-term or occasional trips? |  |
| **H0**  **(Null Hypothesis):** | **Rental services will be required often in every condition by different clients** | Odd Ratio, Risk Ratio |
| **RQ5** | **What** is the costs of limitations or challenges associated with renting a car in certain regions, such as availability during peak seasons or restricted rental policies?. |  |
| **H0**  **(Null Hypothesis):** | There is no significant difference in the costs associated with limitations or challenges of renting a car in different regions, such as availability during peak seasons or restricted rental policies. | **Chi Square Test** |
|  |  |  |