Research on Influencing Factors of Car Rental Price Based on a Linear Regression Model

Yuze Tang1,*

1Department of Industrial Engineering, Purdue University, Indiana, United States

*Corresponding author: tang428@purdue.edu

Abstract. The car rental becomes more and more prevalent as a way of road trips or daily use in the United States. Making the strategy of arranging the type and distribution of vehicles is a great challenge for car rental companies to attract customers. Linear regression is used in this study to analyze the impact of potential factors of vehicles on ratings from customers. These linear regression models are generated based on 5286 sets of data. The result shows that the type of electric vehicles, body style of SUVs, and new vehicles stimulate the increasement of the satisfaction of customers. In contrast, the body style taken has a negative relationship with customers’ ratings. Furthermore, renter trips taken has little impact on customer reflection compared to the age of vehicles. The purpose of this study is to provide a reference for companies to build proper strategies based on customer’s reviews to improve their business.

Keywords: Linear regression; Car rental; Potential factors; Prediction.

1. Introduction

1.1 Background

In 1917, the Saunders family bought 120 Model T’s and a few assorted trucks and started the first car rental business in America. The market for car rental continued since then and grow strong rapidly in the recent 30 years coming with various car rental companies established. According to a study, car rental market expects to expand at a compounded annual rate of 4.8% from 2020 to 2028. Competition among the companies becomes more and more stiff and it is severe for companies to survive in the competitive market. Car rental companies seek a method to make changes or innovations in various areas to improve their competitive strength. Indeed, companies place a high premium on customers and effort to improve the quality of rental experience for the customers. There are lots of potential factors that would affect the satisfaction of customers with their car renting experience. Data analysis on customers is important to summarize their preferences in order to make a reference for companies to build proper strategies fitting the tendency of the market.

1.2 Related research

Back in 2000, the concept of car-sharing was imported to the US from Europe. People in the US started to realize and experience how the convenience of car sharing brings benefits to humans in daily life. Since 2003, the market of car-sharing started to grow with born of several large car rental companies, including Zipcar [1] whose sales and the number of cars doubled each year. A large amount of data analysis research and optimization research has been conducted to improve the scale of car rental for companies and predict the tendency of car rental market. Jun Bi, Zun Yuan, and Qiuyue Sai et al. built a regression model of rental users’ behavior to predict the possibility of silent users in the next month [2]. J.D. Power summarizes a dataset from 9 car rental companies to prove that customers’ satisfaction is challenged by price rises and vehicle supply [3]. Andreas Fink and Torsten Reiners built a minimum-cost network flow model for car rental companies in order to improve efficiency significantly [4]. Beatriz Brito Oliveira, Maria Antónia Caravilla, and José Fernando Oliveira built a new mathematical model that tackles the problems of rental capacity and pricing decisions for car rental companies [5]. A case study [6] was conducted by Maximiliano E. Korstanje to indicate rental car industry in Latin America should explore more in the field of tourism and is concerned of global rises in fuel price may affect the car rental industry.
In 2014, scholars realized most research in that period undermined the effect of customers’ review [7], but the most important and effective form of research on car rental is still one simple thing: information from the customers. Indeed, recent research tends to be based on customers’ feedback. A customer study [8] asked 100 interviewers to share their experience of renting electric cars versus gas cars and compare the advantages between them. Based on 323 completed questionnaires [9], Adamczak Michał, Tobola Adrianna and Fijałkowska Jadwiga et al. get the results of the conducted research may be used by car rental companies to develop incentive systems for their customers, systems encouraging eco-driving which will also reduce the emission of harmful gases into the atmosphere and increase road safety. Nadine Rauh, Thomas Franke, Josef F. Krems designed a comparative study on the experience of electric vehicles to get real feedback on driving an electric vehicle on purpose for further comparison with that of gas vehicles [10].

1.3 Objection

Linear regression is used in this study to describe a relationship between customer’s ratings on rental vehicles and potential factors of customer’s rating. The potential factors of customer’s rating are fuel type, body style, the year vehicle produced, and renter trips taken. Equation based on the result of linear regression predicts the significance of factors to customer’s rating.

2. Method and Data

2.1 Data

The dataset used in this study is obtained on Kaggle which is collected by Cornell University. In this dataset, there are 5286 rows of data after sorting, with 1.1% for diesel vehicles, 10.8% for electric vehicles, 83.3% for gasoline vehicles, and 4.8% for a hybrid vehicle. As the total number of this dataset is large, even a diesel vehicle has 61 rows of data and 253 rows of data for a hybrid vehicle. The dependent variable is the rating score, while the independent variable is four different fuel type vehicles.

First, the general distribution of four types of vehicles is shown in Fig. 1. It indicates that both diesel vehicles and electric vehicles get the highest average rating of 4.95, while the hybrid gets 4.93 and gasoline gets the lowest grade 4.92. All of the maximum ratings for four vehicles are perfect 5.00 score, and the minimum rating for a vehicle is only 1.00 which is extremely low compared to the mean of other types of vehicles.

2.2 Regression Model

Linear regression is a general model for predicting the tendency of the dependent variable in the future based on historical statistics. In addition, linear regression can provide a brief relationship between dependent and independent variables. Linear regression is a simple method of predicting the
future independent variable based on the current independent and dependent data. Linear regression on Excel generates a set of data, including intercept on the y-axis and slope of the line. These two data can be used to make up a linear equation between dependent and independent variables. This form of analysis estimates the coefficients of the linear equation, involving one or more independent variables that best predict the value of the dependent variable. [13]

3. Result

In this study, a dummy variable is used to convert the text information into a number format in each cell. The dummy variable of each fuel type of vehicle is the independent variable in the regression model while the dependent variable is rating scores from customers.

3.1 Vehicle type

The result of the regression model for four factors separately is presented in table 1. The p-value for diesel (0.22) and of hybrid (0.63) is a bit larger than 0.05. Therefore, the data of diesel vehicles or hybrid vehicles versus rating is not significant enough and there does not exist a strong relation between rating and diesel or hybrid vehicles.

<table>
<thead>
<tr>
<th></th>
<th>Coefficient</th>
<th>Standard error</th>
<th>T Stat</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diesel</td>
<td>Intercept</td>
<td>4.9201</td>
<td>0.0025</td>
<td>1950.4979</td>
</tr>
<tr>
<td>vs. Rating</td>
<td>Diesel</td>
<td>0.0282</td>
<td>0.0235</td>
<td>1.2014</td>
</tr>
<tr>
<td>Gasoline</td>
<td>Intercept</td>
<td>4.9450</td>
<td>0.0061</td>
<td>807.2252</td>
</tr>
<tr>
<td>vs. Rating</td>
<td>Gasoline</td>
<td>-0.0294</td>
<td>0.0067</td>
<td>-4.3796</td>
</tr>
<tr>
<td>Electric</td>
<td>Intercept</td>
<td>4.9165</td>
<td>0.0027</td>
<td>1855.2277</td>
</tr>
<tr>
<td>vs. Rating</td>
<td>Gasoline</td>
<td>0.0366</td>
<td>0.0081</td>
<td>4.5254</td>
</tr>
<tr>
<td>Hybrid</td>
<td>Intercept</td>
<td>4.9202</td>
<td>0.0026</td>
<td>1914.1275</td>
</tr>
<tr>
<td>vs. Rating</td>
<td>Rating</td>
<td>0.0056</td>
<td>0.1175</td>
<td>0.4805</td>
</tr>
</tbody>
</table>

As the p-value is extremely low for gasoline and electric vehicle, both two data are significant to rating. For gasoline vehicle, based on the result of a regression model, the equation between gasoline vehicle and rating is presented as follow Equations (1), where RGasoline donates the rating of a gasoline vehicle, and NGasoline donates the number of gasoline vehicle is counted in the rating score.

\[ R_{\text{Gasoline}} = -0.029N_{\text{Gasoline}} + 4.945 \]  

(1)

Based on the resulting number from the regression model, the equation of relation between rating and electric vehicle is shown in Equations (2). Where RElectric donates the rating of electric vehicles and NGasoline donates the number of electric vehicles.

\[ R_{\text{Electric}} = 0.037N_{\text{Electric}} + 4.917 \]  

(2)

3.2 Renter trips taken analysis

Renter trips taken is defined as the number of trips taken for this car before the customer rent and rate it. A histogram is a graph that can show the distribution of data by using the frequency of the numeric bin in a triangle. A histogram of the distribution of renter trips taken is shown in Figure 2. Majority of renter trips taken is between 1 trip to 21 trips. Only a few cars have taken larger than 200 trips of which 1% of it.
Fig. 2. Histogram for renter trips taken  (Photo credit: Original)

Linear regression model is used to predict the relation between number of trips taken and rating of vehicles. In this regression model, independent variable is the number of trips taken by each vehicle and dependent variable is rating score and the result of regression model is shown in the following Table 2.

<p>| Table 2. Regression Model for Renter Trips Taken |</p>
<table>
<thead>
<tr>
<th>COEFFICIENTS</th>
<th>STANDARD ERROR</th>
<th>T STAT</th>
<th>P VALUE</th>
</tr>
</thead>
<tbody>
<tr>
<td>INTERCEPT</td>
<td>4.9286</td>
<td>0.0034</td>
<td>1438.1336</td>
</tr>
<tr>
<td>RENTER TRIPS TAKEN</td>
<td>-0.0002384</td>
<td>6.0784E-05</td>
<td>-3.920</td>
</tr>
</tbody>
</table>

Equation (3) of the relation between renter trips taken and rating is obtained. where \( R_{\text{RenterTripsTaken}} \) donates the rating of the vehicles and \( N_{\text{RenterTripsTaken}} \) Donates the number of renter trips taken by each vehicle. The slope of number of renter trips taken is negative which indicates that there is a negative relation between renter trips taken and rating score. When the number of trips increases, the rating score will decline. However, as the slope of renter trips is tiny, the declining impact of renter trips taken to rating can be ignored.

\[ R_{\text{RenterTripsTaken}} = -0.00024N_{\text{RenterTripsTaken}} + 4.93 \]  

(3)

3.3 Car Production Year

In order to obtain the relation between years and rating score, linear regression is also used in this section. Vehicle year is defined as the year of the vehicle produced in this model. In this regression model, the independent variable is the year of the vehicle, and the dependent variable is still rating. The result of this regression model is shown in Table 3 followed by the equation of the relationship between the year of the vehicle and its rating score. where \( R_{\text{Vehicle_Year}} \) donates the rating of the vehicle, and \( Y_{\text{Vehicle}} \)donates the year the vehicle is produced. The slope is 0.0052 which is negative, indicating that there is a positive of the year of vehicle produced, Equation (4). The latest vehicle can get a higher rating score in general.
Table 3. Regression Between Year of Vehicle and Rating

<table>
<thead>
<tr>
<th>COEFFICIENTS</th>
<th>STANDARD</th>
<th>T STAT</th>
<th>P VALUE</th>
</tr>
</thead>
<tbody>
<tr>
<td>INTERCEPT</td>
<td>-5.5056</td>
<td>1.3492</td>
<td>-4.0807</td>
</tr>
<tr>
<td>VEHICLE YEAR</td>
<td>0.00517</td>
<td>6.6945E-04</td>
<td>7.7272</td>
</tr>
</tbody>
</table>

\[ R_{\text{Vehicle\_Year}} = 0.0052Y_{\text{Vehicle}} - 5.5056 \]  

(4)

3.4 Body style

Linear regression is used to predict the relation between rating and body style of SUVs and car separately. In this regression model, SUV and car are independent variables for each regression model, while rating is the dependent variable. The result of the regression model is shown in Table 4 followed by the equation of each relation between SUVs or cars and rating based on the numbers in the regression model.

Table 4. Regression Model for Car and SUV

<table>
<thead>
<tr>
<th>COEFFICIENTS</th>
<th>STANDARD</th>
<th>T STAT</th>
<th>P VALUE</th>
</tr>
</thead>
<tbody>
<tr>
<td>DIESEL</td>
<td>Intercept</td>
<td>4.9145</td>
<td>0.0031</td>
</tr>
<tr>
<td>VS. SUV</td>
<td>SUV</td>
<td>0.0180</td>
<td>0.0057</td>
</tr>
<tr>
<td>GASOLINE</td>
<td>Intercept</td>
<td>4.9271</td>
<td>0.0042</td>
</tr>
<tr>
<td>VS. CAR</td>
<td>Car</td>
<td>-0.0117</td>
<td>0.0054</td>
</tr>
</tbody>
</table>

\[ R_{\text{SUV}} = 0.018N_{\text{SUV}} + 4.91 \]  

(5)

\[ R_{\text{Car}} = -0.017N_{\text{Car}} + 4.93 \]  

(6)

where RSUV donates the rating of SUV, NSUV donates the number of SUV, RCar donates the rating of car, and NCar donates the number of cars. As the slope (0.018) of the number of SUV is positive, it has a positive relationship between SUV and rating, while car has a negative relation with the rating because their slope (-0.017) is smaller than 0. If the company owns more SUV vehicles, its entire rating will have a tendency to rise, while car has a negative effect on rating, Equation (5), (6).

4. Discussion

Based on the regression model above, factors of an electric vehicle, new vehicle and SUV body style all have a significant positive effect on raising the rating of vehicles. Compared to gasoline vehicles, electric vehicle has some advantages for customers. The first advantage is electric vehicle is environmentally friendly. Unlike gasoline car, the electric vehicle doesn’t emit greenhouse gasses into the atmosphere despite the process of electricity generation. It can provide the resident with an environment with fresh breath and higher air quality. The second one is electric vehicle is quiet not only for passengers but also for resident or pedestrians around. There isn’t an engine inside an electric vehicle, which can generate loud noise when starting the engine or running on the road. Driver and passenger sitting inside the vehicle won’t hear any annoying sound from the engine, which can provide them a silent atmosphere for conversation. Besides environmentally friendly and silent, electric vehicles is also saved money on energy. As the fuel price is increasing gradually due to COVID-19 and the international situation, driving a gasoline car one day may spend half of the price of the rental fee on filling the vehicle with gasoline. Considering, the fuel price, lots of customers prefer the electric vehicle with its low-cost electricity fill. The fourth advantage of electric vehicles...
has a positive relationship with the year of the vehicle produced. Most electric cars rental in the dataset is the brand Tesla, which is produced after 2018. The newer interior makes customers feel comfortable when sitting inside.

Many families prefer to rent a vehicle for their road trip, and SUV is a better choice for them. Compared to car, SUV has more space to store packages and they can be driven in more complex terrain than car. Except for road trips, SUV provide a better sitting experience in urban areas. Passengers sitting in SUV will feel smooth and won’t follow the uneven ground going up and down.

For the year of the vehicle produced, the newer vehicle can provide the customer with a better driving experience. The newly manufactured car has a new media system and control system, which provides the customer with a more humanized service. There is hardly a musty odor inside new cars. In addition, it is unlikely for a new car to encounter engine breakdown or problems with hardware.

5. Conclusion

Renting a car is one of the popular ways for not only daily use for the long-term in substitution of spending money on purchasing vehicles but also for short-term traveling. The car rental market expands rapidly in the recent 10 years, gradually forming into a perfect competition market with hundreds of companies including some famous companies such as National, Enterprise, Hertz and so on. As the market is more and more competitive, companies look highly at costumers’ experiences to strengthen themselves in order to survive in the market. This study builds linear regression between potential factors and rating of costumers, which includes four fuel types of vehicles (diesel, gasoline, electric and hybrid), two body style (SUV and car), produced year of the vehicles, and renter trips taken. Based on the result of the regression model above, as the technology of electric vehicles is more and more mature in the last five years, it is a better choice to buy more electric car and gasoline SUV and replaces the old car with new ones. However, linear regression makes a simple linear approximation while the actual situation is distinct from it, as there are other variables that affect the customer’s experience, such as weather, location of the renting store, renting price of each vehicle and brand of it. The linear regression model can only provide a brief relation of the upcoming situation but not describe the tendency in detail and precisely.

As the renting car market is growing rapidly in recent several years, this study provides a foundation reference for car rental companies working for further systematic and technical prediction. In addition, the summary about the tendency of this dataset is a reference for the customer when they are frustrated in making a decision of selection among various types of vehicles on car rental websites.

References


