**Used Car Sales Analysis & Report**

Matt Riveira

MDA 620

**Table of Contents**

**Overview/Background 3**

**Scenario/Objective 3**

**Problem Scenario 4**

**Data Exploration 4**

**Data Visualization 7**

**Model Building 8**

**Conclusion 9**

**Works Cited 10**

**Overview/Background**

In India, the used car market has seen significant growth in recent years, with an increasing number of consumers opting to purchase pre-owned vehicles rather than new ones. This trend can be attributed to a variety of factors, including the availability of a wide range of used cars at competitive prices, the ease of financing options, and the increasing reliability of used cars. As a result, used car sales have become an important aspect of the automotive industry in India, with a large number of dealerships and private sellers offering a variety of pre-owned vehicles for sale. In order have a better understanding of the Indian used car market, we will analyze our dataset which contains 7,906 rows and 19 columns. I also created an extra column in the dataset which is labeled “Sales Price USD”. This was created from the original column “Selling Price” which was accomplished by multiplying the selling price column by the US exchange rate. There were no Na’s in this dataset.

**Scenario/Objective**

This dataset will analyze the top car brands sold, their average sale price, number of seats, fuel used, mileage, and total kilometers driven. Each column will show different statistics and descriptions of each car. We will be analyzing all these statistics and descriptions and seeing which leads to more sales and what the popular demand is in the Indian used car market.

**Problem Scenario**

The Indian used car market’s biggest issue is overpriced vehicles. Despite the availability of a wide range of used cars at competitive prices, some sellers may attempt to charge excessively high prices for their vehicles, either due to a lack of awareness of current market values or because of fraudulent or deceptive practices. This can make it difficult for buyers to find good deals on used cars, leading to frustration and dissatisfaction with the market. In addition, overpriced vehicles can also discourage potential buyers from entering the market, as they may feel that they are unable to afford a used car at the prices being offered. To address this problem, there is a need for greater transparency and accuracy in pricing, as well as more robust systems and processes for ensuring fair and reasonable prices for used cars. This could involve the use of industry data and market analysis to inform pricing decisions, as well as stricter penalties and enforcement measures to deter fraudulent or deceptive pricing practices.

**Data Exploration**

This first pie chart shows the top ten car brands sold by total sales in quantity. As we can see Maruti is the leader in sales with 33.1% of total used car sales in all of India. The second pie chart shows the top selling car brands based on total sale price. It is interesting to see how not a single car in the top 10 of total sale price is in the top 10 sold in quantity.

**Chart, pie chart

Description automatically generated**Chart, pie chart

Description automatically generated

Chart, treemap chart

Description automatically generated

The heat map to the right shows how different car statistics correlate to each other. As you can see, this heat map uses: Sales ID, Year manufactured, Selling price, KM driven, mileage, engine, max power, seats, and sales price USD. Based on the heat map we can conclude that max power has the highest correlation to sales price.

The next three pie charts will show; how the car was sold, the number of previous owners each car had, and the type of transmission the car had. As we can see below, 83% of all used cars sold were sold by the individual themselves. 65.96% of all cars sold had only one previous owner, and 86.83% of all cars had a manual transmission.

Chart, pie chart

Description automatically generatedChart, pie chart

Description automatically generatedChart, pie chart

Description automatically generated

The next two bar charts show the type of transmissions each car has and the fuel they use. As we can see below there are more than 75% of all cars with a manual transmission. The fuel bar graph shows that diesel and petrol account for more than 90% of all the fuel in cars.

Chart, bar chart

Description automatically generatedChart

Description automatically generated with low confidence

Below we can see scatterplots of each combination of every column. We will go into more depth of two of the scatterplots later.

A picture containing shoji, square

Description automatically generated

**Data Visualizations**

Based on the heat map above we saw a lot of different columns have positive correlation with each other. The scatter plots below will visualize their correlation. As we can see below, the first plot shows how the higher the max power is, the more the car will for. We also see the newer the car is, the more it will sell for. Showing that both max power and year manufactured have a positive correlation with sale price.

Chart, scatter chart

Description automatically generatedChart, scatter chart

Description automatically generated

**Model Building**

Chart, scatter chart

Description automatically generatedThe first model made was linear regression. This model was picked because I wanted to compare the year the car was manufactured vs how much it sold for. As we can see the newer the car is the more it sold for.

Graphical user interface, text, application, email

Description automatically generated

The second model built was a logistic regression model. This model compared the year, mileage, and km driven to if the car was sold or not. After splitting up the data between test data and train data we can see that the model has a accuracy of 0.75.

**Conclusion**

The Indian used car market is far different from the US car market. Here in the US we would expected most the fuel, seller type, and transmission statistics to be the exact opposite of what is shown in this report. We can also see max power and year manufactured have a direct positive correlation with sale price.

**Works Cited**

<https://www.kaggle.com/datasets/shubham1kumar/usedcar-data>