EM 624 – Informatics for Engineering Management

Topic – Analyzing Used Cars

Siddharth Gajare
Graduate Student
Master of Engineering in Engineering Management
Stevens Institute of Technology
CWID – 20008696
Email – sgajare@stevens.edu

Instructor: Dr. Carlo Lipizzi

PROJECT GOAL

The Indian used-car industry, which has been valued at $18.3 billion in revenue, is expected to double at a rate over the forecast period of 19.5% until FY'2026-2027. More than 4.4 million used cars were sold in India in a year. Cardekho.com is an Indian website that sells used cars. The website has more than 39 million monthly visits and has more than 5000 dealers in India.

The reason to buy a used car is that it saves money through depreciation, reduces paperwork stress, lowers insurance costs, lowers registration fees, saves on sales tax, avoids extra costs on extra features, and so on.

Given the size of the market and the opportunities for business, it is critical to conduct research on the various factors that influence the market.

Many factors can influence the used car market, such as selling price, brand, model features, the year the car is manufactured, kilometers that are driven, fuel type, transmission type, and owner type of car.

This project will analyze factors that influence the used car market. Also, the correlation between factors. Which would help Cardekho to understand its market.

References: (for numbers produced)

https://www.cardekho.com/

https://www.mahindra.com/news-room/press-release/indian-used-car-market-valued-at-dollar-23-billion-in-fy-2021-2022

https://techcrunch.com/2019/01/03/cardekho-raises-110/#:~:text=CarDekho%20claims%2039%20million%20monthly,million%20downloads%20of%20its%20apps

RESEARCH QUESTIONS

This project would answer following questions.

1. Which brand sold out the most cars? This would answer which brand is producing good cars and which brand people are preferring to purchase.
2. Which models were sold out; this would answer what models are doing well in the market.
3. The percentage sales of the brands? This would answer the market sales with its distribution.
4. What is the average selling price of the brands? This would help us analyze what is the average price of the product the company is producing.
5. Which year models are being purchased? This will assist us in analyzing what role the manufacturing year plays.
6. Some types of factors on which car units are sold play roles like the owner type, fuel type, and transmission type.
7. The relationship between Price to factors like model year and km driven.
8. Correlation Matrix between factors.
9. Wordcloud which shows important words in the project.

UNDERSTANDING DATASET & PREPROCESSING

The data consists of Cardekho.com which is a used car selling website in India.

This is the head (5) of the data:

```
                      name  year  ...  transmission        owner
0            Maruti 800 AC  2007  ...        Manual  First Owner
1   Maruti Wagon R LXI Minor  2007  ...        Manual  First Owner
2       Hyundai Verna 1.6 SX  2012  ...        Manual  First Owner
3    Datsun RediGO T Option  2017  ...        Manual  First Owner
4     Honda Amaze VX i-DTEC  2014  ...        Manual  Second Owner
```

This are the Columns in the data:

(name, year, selling_price, km_driven, fuel, seller_type, transmission, owner)

Size of the data: (4340 rows, 8 columns)

Data Type:

```
Data columns (total 8 columns):
 #   Column         Non-Null Count  Dtype
---  ------         --------------  -----
 0   name           4340 non-null   object
 1   year           4340 non-null   int64
 2   selling_price  4340 non-null   int64
 3   km_driven      4340 non-null   int64
 4   fuel           4340 non-null   object
 5   seller_type    4340 non-null   object
 6   transmission   4340 non-null   object
 7   owner          4340 non-null   object
dtypes: int64(3), object(5)
```

Checking if there are Null Values in the data:

```
name            0
year            0
selling_price   0
km_driven       0
fuel            0
seller_type     0
transmission    0
owner           0
```

Unique strings in the data columns:

```
['Maruti 800 AC' 'Maruti Wagon R LXI Minor' 'Hyundai Verna 1.6 SX' ...
 'Mahindra Verito 1.5 D6 BSIII'
 'Toyota Innova 2.5 VX (Diesel) 8 Seater BS IV'
 'Hyundai i20 Magna 1.4 CRDi']
[1992, 1995, 1996, 1997, 1998, 1999, 2000, 2001, 2002, 2003, 2004, 2005, 2006, 2007, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019, 2020]
['Petrol' 'Diesel' 'CNG' 'LPG' 'Electric']
['Individual' 'Dealer' 'Trustmark Dealer']
['Manual' 'Automatic']
['First Owner' 'Second Owner' 'Fourth & Above Owner' 'Third Owner'
 'Test Drive Car']
```

I created a new column for brand:

df.name.apply(lambda x : ' '.join(x.split(' ')[:1]))

I created this because the brand name was with model, and I wanted just brand name for further analysis. This preprocessing of splitting the string and a new column for it. And named it 'brand'.

METHOD & ANALYSIS

Firstly, I imported all the required libraries like pandas, numpy, ax, matplotlib.pyplot, seaborn and wordcloud from the python interpreter.

Then using pandas, I read the csv file.

Using commands like head, describe, shape, info and isnull and unique I made sure that I understood my data for analysis.

Following that, I was prepared to plot the graphs using the dataset. For the analysis, various graphs such as a bar graph, a pie chart, a matrix, and a wordcloud were used.
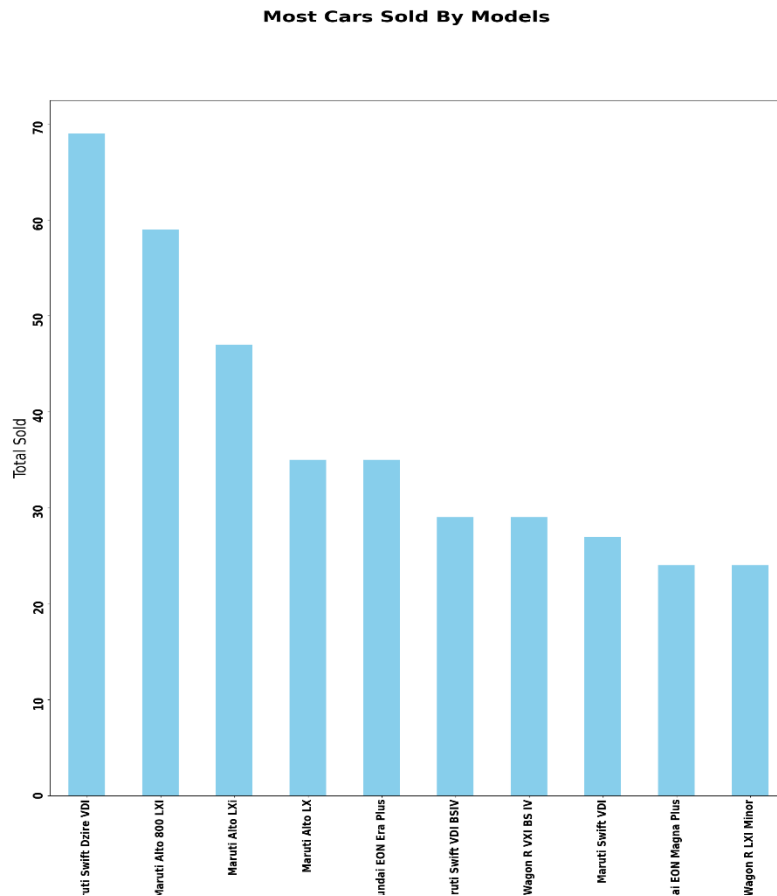
Top 5 brands of cars most sold out.

Brand	Cars Sold Out
Maruti	1280
Hyundai	821
Mahindra	365
Tata	361
Honda	252

Here we see that Maruti is the most popular choice for Indian people when buying a used car.

Followed by Hyundai and Mahindra. We see that the total sale for Maruti was 1280 units securing the top spot.

Most sold out car models.
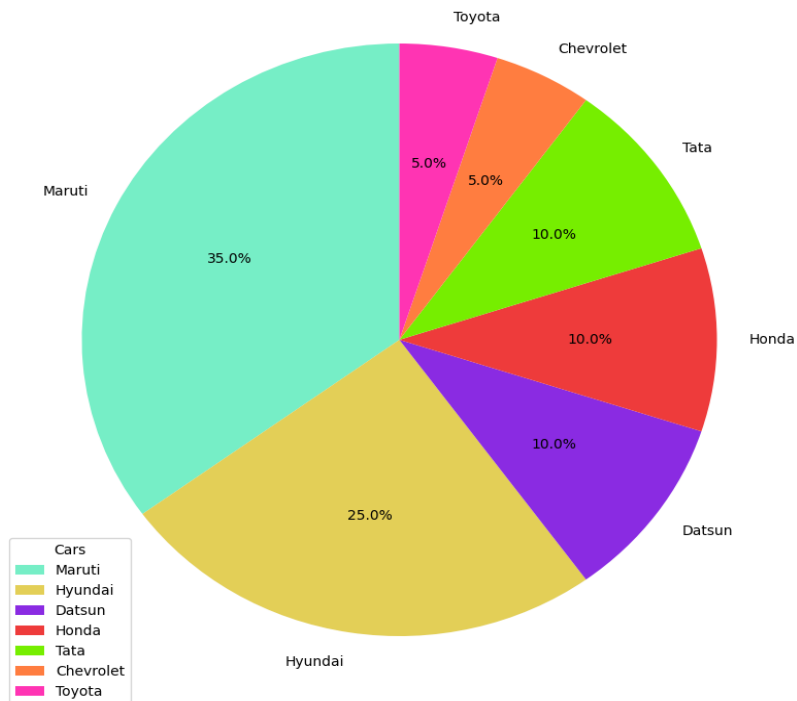
Most Cars Sold By Models

This graph gives out the Models people are preferring to buy.

The Maruti Swift Dzire VDI is the best-selling model here. Maruti Alto 800 LXI is the runner-up, and Maruti Alto LXi is third.

The Maruti Dezire and Alto are the cars that people want the most as those are topping the chart.

# Percentage Distribution of Brands

Percent Sales by Brands

This is a Pie chart which represents the percentage of sale by the brands.

Here we see Maruti accounts for 35% while Hyundai 25% and rest brands like Datsun, Honda, Tata, Chevrolet and Toyota fall at 10% or below.

Maruti and Hyundai collectively make up 60% of the market ie. More than half of the market.

Average Selling Price of Brands
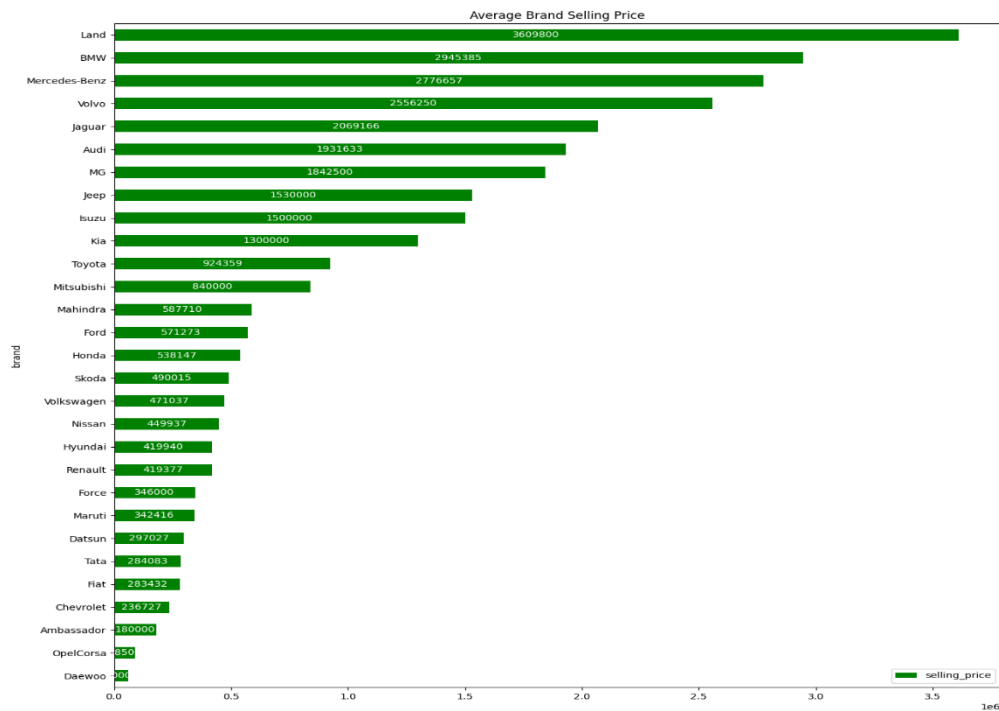
This is horizontal Bar Graph. This graph shows the average price at which the brands sell their vehicles.

The top five brands are Land Rover, BMW, Mercedes-Benz, Volvo, and Jaguar.

We can see that none of these brands are among the best-selling brands or vehicles. We see that brands that sell more cars sell cars at a lower price.

The model year people are purchasing

Model Year Most Purchased
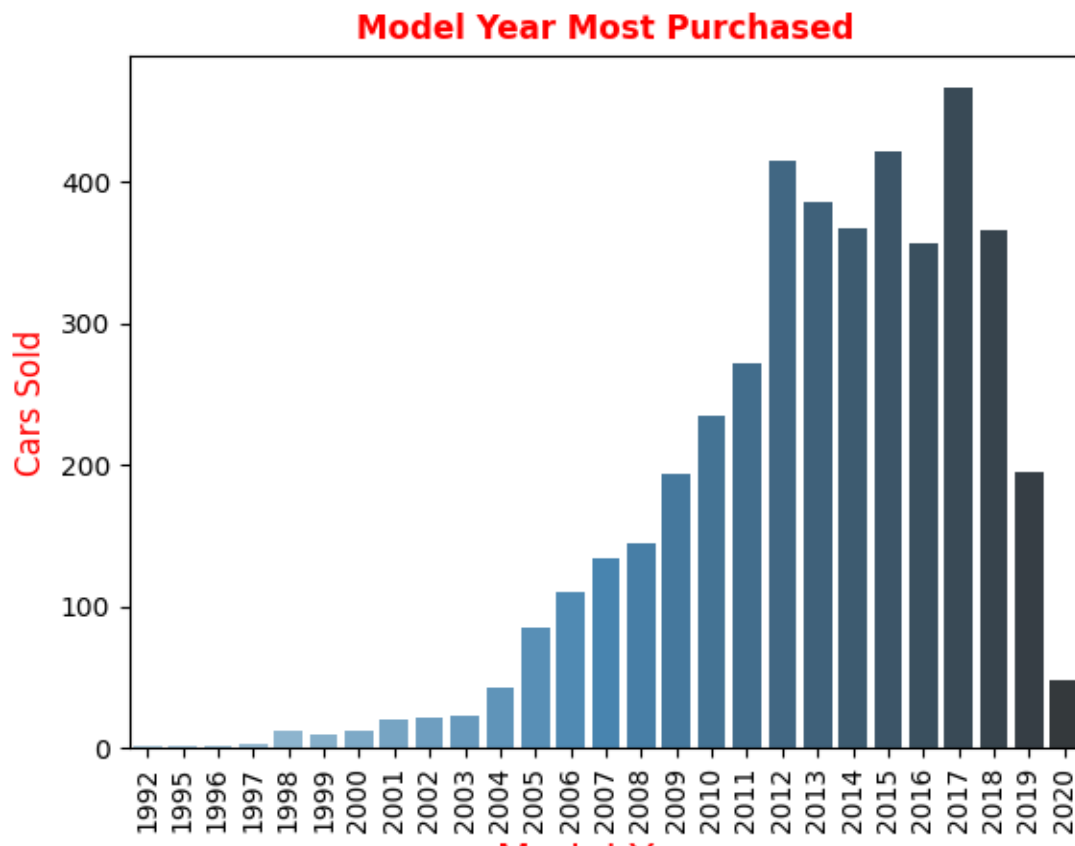
This graph shows the year the car was manufactured and the units that were sold out to the customer.

Here year 2017 tops the chart. We see a Upward trends as the model gets newer and downward trends just couple of years before the car is sold.

The factors that would be leading to such a graph could be the depreciation value of the car.

Also, the fact that car depreciates in the initial years faster and gets slower in depreciation as a couple of years pass by.

It is seen that people's best choice to purchase the car is right between 3-8 years old models.

Cars Sold to Types (Owner, Fuel, Transmission)
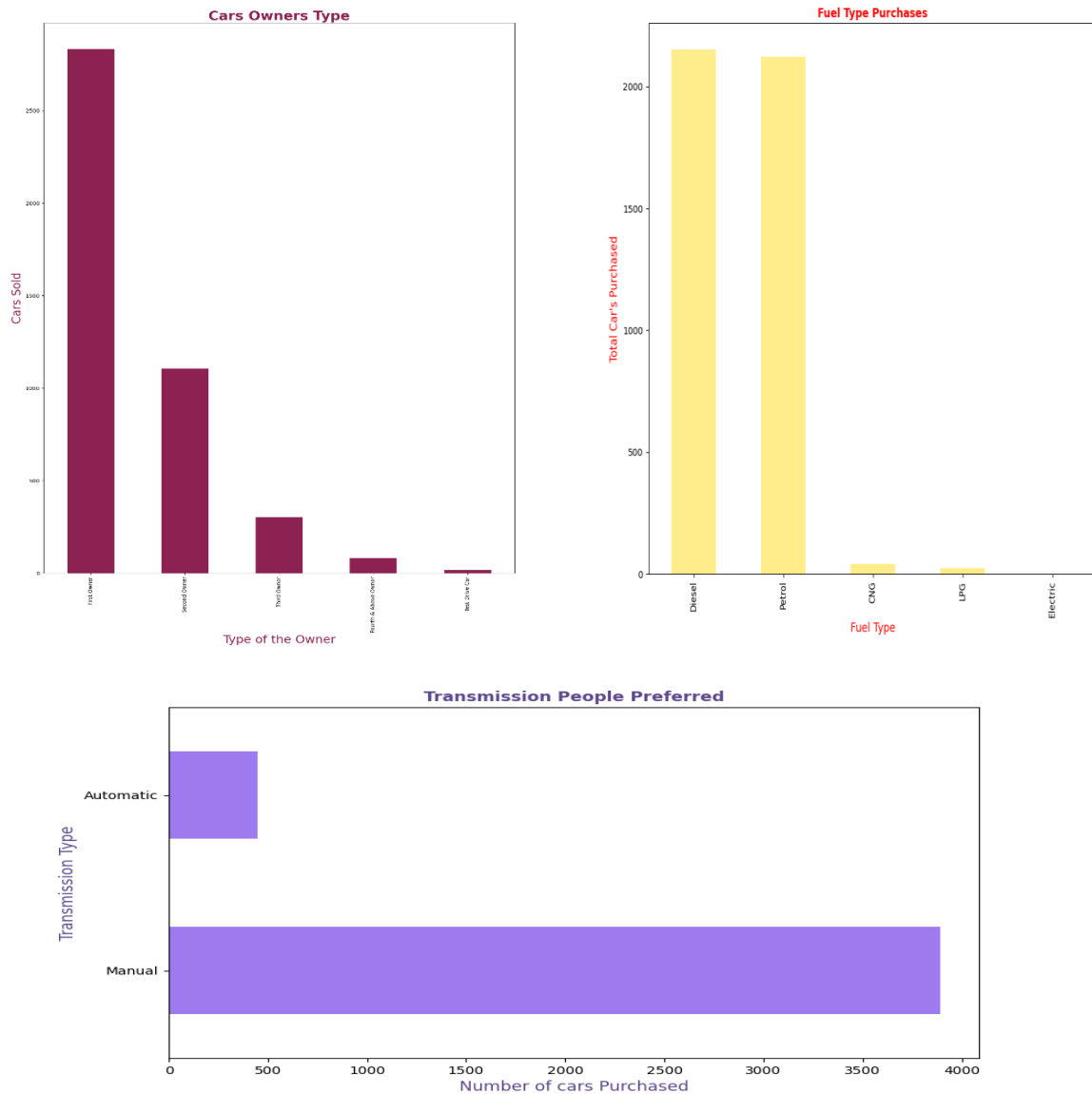
These graphs show Engine Type, Transmission Type and Owner Type.

We see that for

Owner Type: First Owner is the most preferred, followed up by second owner, third owner, fourth & above and test drive car. Here we observe that people feel less secure about their cars when the number of owners increases.

Fuel Type: Diesel tops with an edge against petrol when people choose to buy a car. Followed up by CNG and LPG. Here we see that there is a vast difference in sales between diesel-petrol and CNG-LPG vehicles. Electric is still not a viable option in India for people to purchase.

Transmission Type: Manuel is preferred to Automatic in India. In this Market, people tend to love driving Manuel vehicles. Automatic is less preferred.

Finding Relationship of Price against model year and then Km driven.
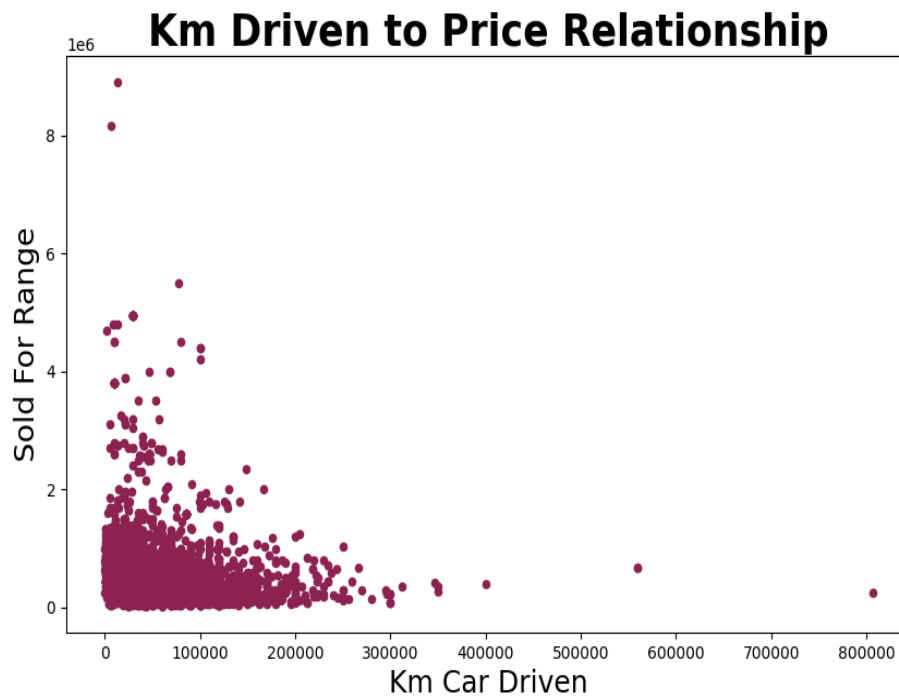
1. Relationship of price against year manufactured.
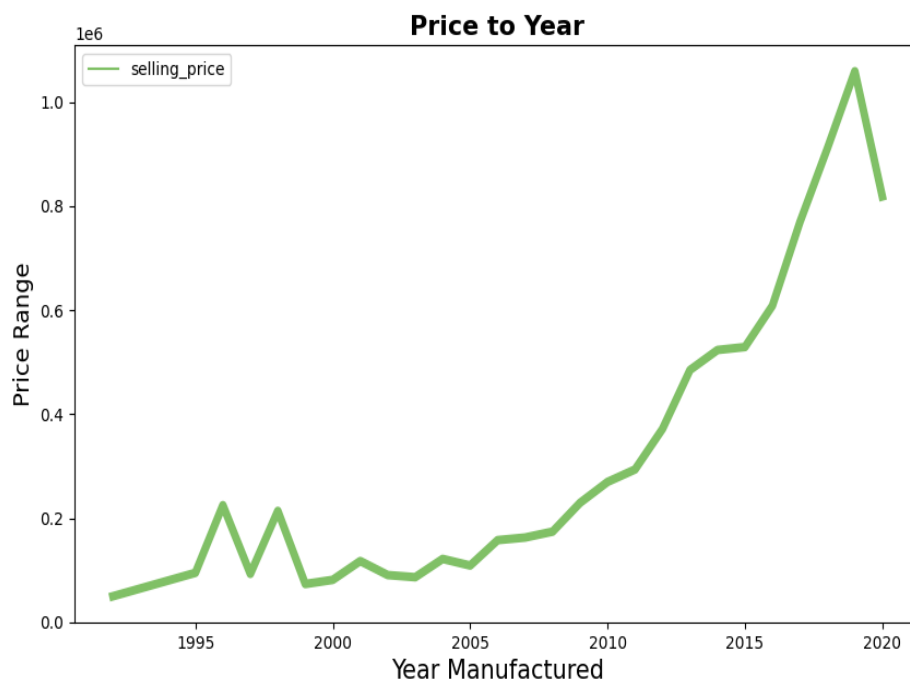
The graph is plotted with Price against year manufactured.
In this graph, we can see that as the car gets older, the price tends to fall, owing to the fact that the value of the car depreciates over time.

2. Relationship of price against km driven

Km Driven to Price Relationship

The scatterplot shows the car which is driven in kilometers to the price sold.
We here observe that as the km on the car increases the scatterplot points drop lower
meaning that as the car gets more driven the price of the car decreases.
We see a direct relationship between price to km driven here in the graph clearly.
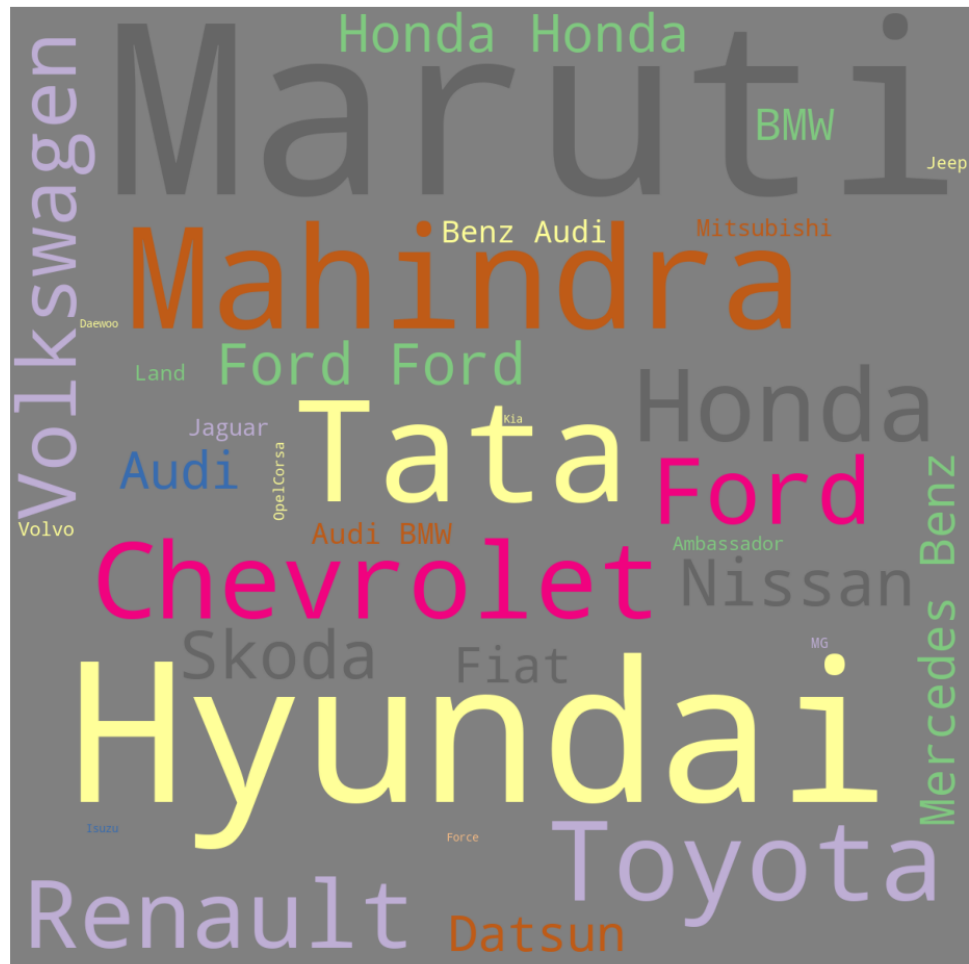
Next, we find out the correlation matrix
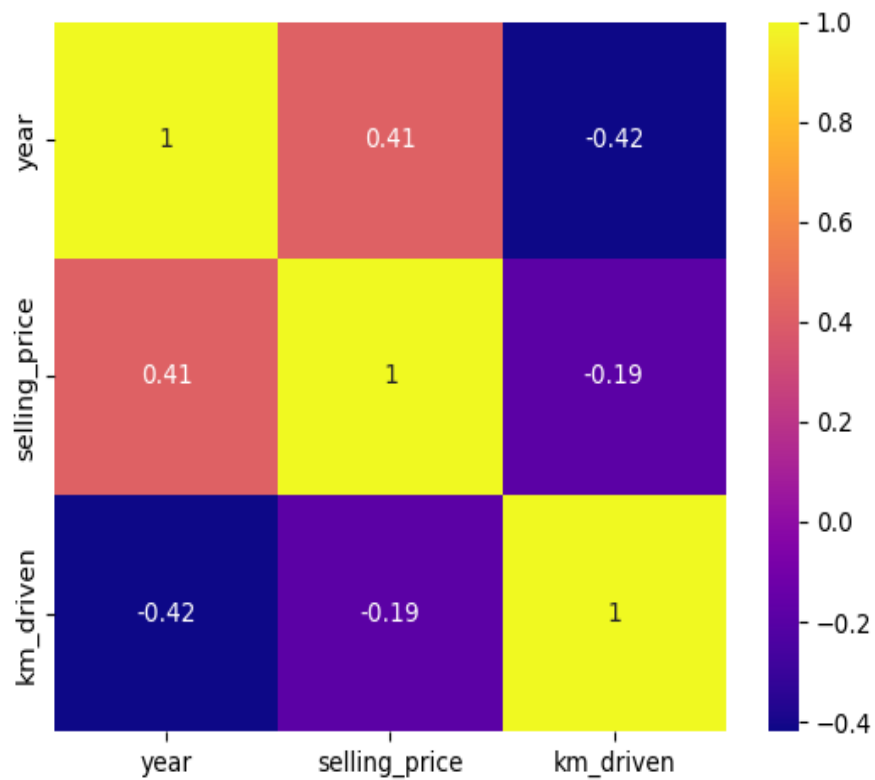
Using this graph, I found out all the numerical values relationship.

Here 1 indicates perfectly positive correlation. (Yellow Box)
Negative numbers indicate a negative correlation. (Bluish Box)
Positive values indicate positive correlations. (Red Box)

For Final Analysis I made a WordCloud

This is a Wordcloud, and we can see that Maruti, and Hyundai have larger fonts. Medium font is used for words like Mahindra, Chevrolet, Tata, Toyota, and Renault. Small fonts are used for words like Fiat, Skoda, Volvo, Land, and Mitsubishi.

The font size in this graph represents the importance of words. The larger the font, the more important the word.

RESULT

The result of the project is as follows:

1. The most sold-out brand in India is Maruti, which sold over 1280 units.
2. The most sold-out model is Maruti Swift Dzire VDI.
3. The brand with the most market share is Maruti with 35% sales.
4. The highest average car value in the brand is Land Rover with 3609800 Rupees. It does affect as Land Rover doesn't have the maximum number of cars unit sold as well as market share.
5. Cars that are old between 3-8 years are widely purchased as it provides better economy to the customers.
6. Factors that play role in car units sold are:
7. Owner Type: First-hand owners are widely purchased. Also, it is seen that consecutive hand owners have the importance ie. It's sequential.
8. Fuel Type: Diesel tops in terms of cars desired but only with a slight edge margin over petrol. Electric cars suffer in India as of today as it stays bottom.
9. Transmission Type: Surprisingly Indians prefer Manuel over Automatic transmission by a good margin.
10. The relationship between price and model year is that the price of the car decreases as the model gets older. In addition, as the car is driven more kilometers, the price decreases.
11. The Wordcloud shows Maruti in bold as playing the most important role in the market. The least important are Daewoo and Force.

All of these results will assist CarDekho as a company in analyzing the market and the market in which they want to make more profits through revenue generation with strategies they intend to implement in the future.