MTH 765P Mini–Project: Car Details from Car Dekho

Dhruv Vudayagiri

220375472

Contents

1 Introduction

The data is related to the Indian car sales as per the Car Dekho portal. Car Dekho is an Indian auto portal that assists users with car research, financing, insurance, used cars, and any other aspect of buying and selling a vehicle. To facilitate vehicle purchases, the company has partnerships with numerous auto manufacturers, car dealers, and financial institutions.

The source of the data found is from the Kaggle website and downloaded the data set was as .CSV document in the local folder and uploaded into the Jupyter notebook for further usage and analysis.

1.1 Description

The dataset includes the following car information:

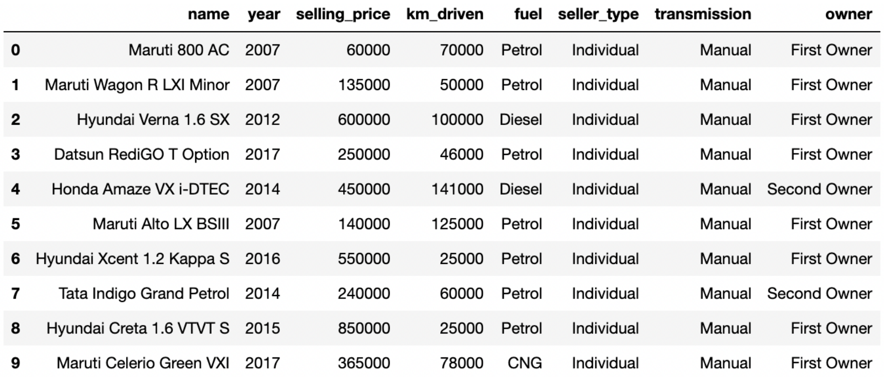
1. Car name: describes the car brand as well as the model's name.
2. Year: describes the year the car was purchased.
3. Selling Price: the price of the car at the time of sale.
4. Kms driven: denotes the number of kilometres already driven with the vehicle at the time of sale.
5. Fuel: refers to the type of fuel used to power the vehicle's engine.
6. Seller type: specifies whether the seller is a Dealer, an individual, or a Trustmark dealer.
7. Transmission: specify whether the engine's transmission is automatic or manual.
8. Owner: Indicates how many previous owners the vehicle has had.

Out of the eight variables, two of them are continuous variables and the other six of them are discrete variables.

The data is 355 KB (Kilobytes) in size as a CSV file. The dataset contains 4340 entries. (4340 rows) and 8 specifics for each vehicle (8 columns)

1.2 Libraries

- Matplotlib.pyplot (to deal with graphic charts)
- pandas (to deal with data like data processing, CSV file I/O)
- seaborn (to deal with graphic charts)



	name	year	selling_price	km_driven	fuel	seller_type	transmission	owner
0	Maruti 800 AC	2007	60000	70000	Petrol	Individual	Manual	First Owner
1	Maruti Wagon R LXI Minor	2007	135000	50000	Petrol	Individual	Manual	First Owner
2	Hyundai Verna 1.6 SX	2012	600000	100000	Diesel	Individual	Manual	First Owner
3	Datsun RediGO T Option	2017	250000	46000	Petrol	Individual	Manual	First Owner
4	Honda Amaze VX i-DTEC	2014	450000	141000	Diesel	Individual	Manual	Second Owner
5	Maruti Alto LX BSIII	2007	140000	125000	Petrol	Individual	Manual	First Owner
6	Hyundai Xcent 1.2 Kappa S	2016	550000	25000	Petrol	Individual	Manual	First Owner
7	Tata Indigo Grand Petrol	2014	240000	60000	Petrol	Individual	Manual	Second Owner
8	Hyundai Creta 1.6 VTVT S	2015	850000	25000	Petrol	Individual	Manual	First Owner
9	Maruti Celerio Green VXI	2017	365000	78000	CNG	Individual	Manual	First Owner

Figure 1: Dataframe preview

2 Preprocessing

The following describes the process that was done to the extracted data, before proceeding to the visualisation part of our analysis.

2.1 Checking duplicates

Firstly, we checked whether any duplicates in the data exist and found there were 763 duplicate rows. We removed the duplicates from the data using the “drop_duplicates() ”function in pandas.

2.2 Creating vehicle age column

We have created a new column called ’vehicle_age’. The column is important when visualising vehicle data, as each year the value of the vehicles usually depreciate, which would result a decrease in the price. From our visualisations, we will be seeing if that is true or not.

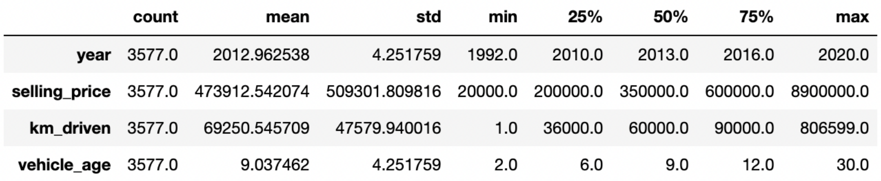
2.3 Creating brand column

Initially, there is a single ‘name’ column that has data as a combined form with both brand and model names. So, we updated the data type of the ‘name’ column to string. Next, using ‘insert()’ in the data frame with the lambda function, generated a new column as ‘brand’ in the data frame.

2.4 Creating model column

We have created a new column called ’model’, which was created by using the column ’name’ from the data frame with the ’insert()’ function. The ’model’ column shows the model of each of the cars that are found in our data set, by using ’name’ to get our data.

3 Visualization



	count	mean	std	min	25%	50%	75%	max
year	3577.0	2012.962538	4.251759	1992.0	2010.0	2013.0	2016.0	2020.0
selling_price	3577.0	473912.542074	509301.809816	20000.0	200000.0	350000.0	600000.0	8900000.0
km_driven	3577.0	69250.545709	47579.940016	1.0	36000.0	60000.0	90000.0	806599.0
vehicle_age	3577.0	9.037462	4.251759	2.0	6.0	9.0	12.0	30.0

Figure 2: Description of the Dataframe

This table describes the data set with columns such as year, selling_price, km_driven, and vehicle_age. It includes the mean, standard deviation, minimum, and maximum values along with IQR values for each column.

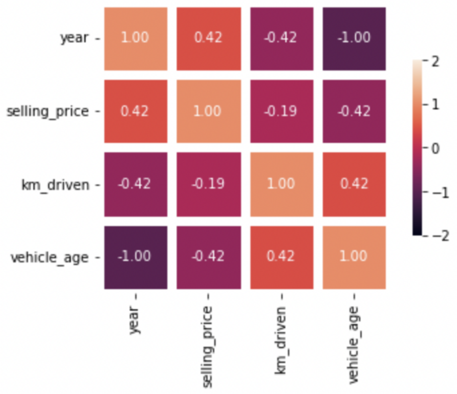


Figure 3: Correlation of the Data

Figure 3 is a correlation plot, showing the correlation between variables. As the points on the graph form a diagonal line, we can see that as the values of one variable are increasing, the values of the other variable also increase. This also suggests that, as the value of one variable decreases, the value of the other variable also tends to decrease, and vice versa. For example, we can to see that 'vehicle_age' shows if the year decreases, then 'vehicle_age' increases since they have a negative correlation value.

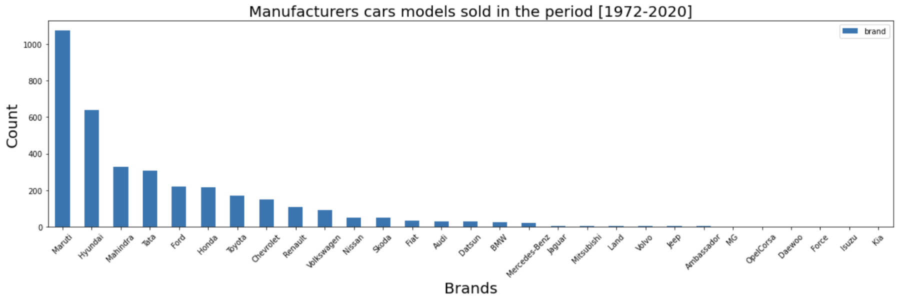


Figure 4: Count of brands being sold in the period [1972-2020]

The above model is a bar plot that shows the total number of cars sold by each car brand, from a range of years [1972-2020]. The x-axis represents the car brands and the y-axis represents the number of cars sold. The plot shows that the brand ”Maruti” sold the most cars, followed by ”Hyundai” and ”Mahindra”. The other car brands ”MG”, ”OpelCorsa”, ”Daewoo”, ”Force”, ”Isuzu” and ”Kia” sold fewer cars.

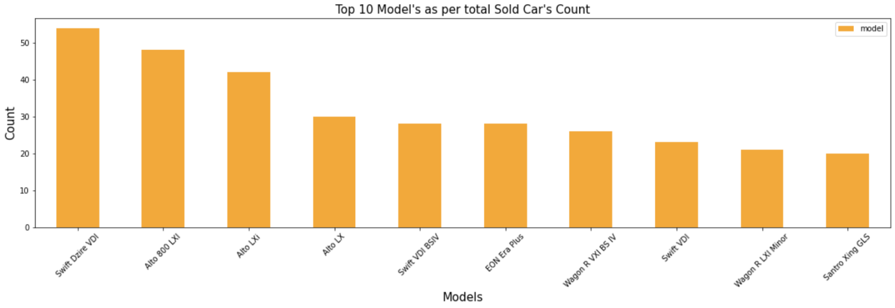


Figure 5: Top 10 Models as per total Sold Car Count

Figure 5 is a bar plot that describes the sale count of the top 10 models. The plot suggests that the car model with the most sales is “Swift Dzire VDI” followed by ”Alto 800 LXI”, “ Alto LXI” and so on.

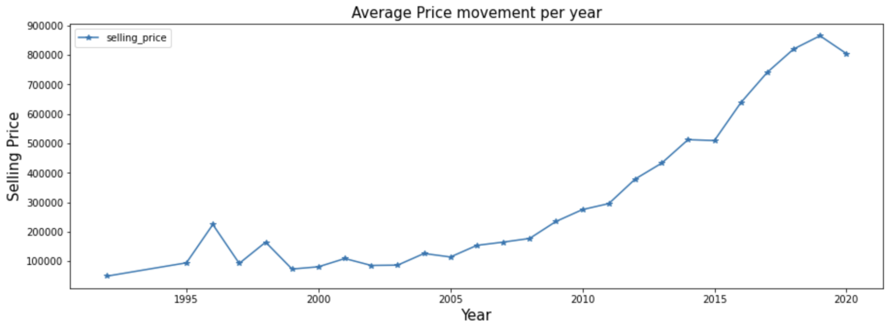


Figure 6: Average Price movement per year

The above illustration is a line graph showing the average price movement per year for the vehicles. The vertical axis of the graph shows the price, while the horizontal axis shows the years. The graph shows that the price generally increased over time, with some fluctuations. The largest price increase occurred between 2010 and 2012, while the largest price decrease occurred between 2015 and 2016. The most recent data point on the graph is from 2020, and the price at that point is higher than it has been at any other point shown on the graph.

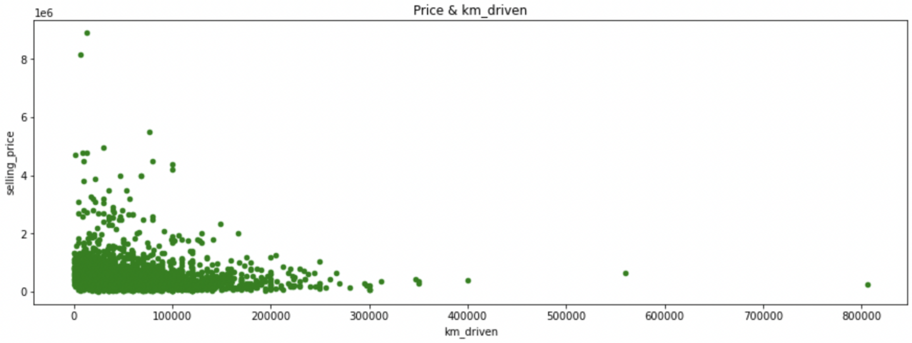


Figure 7: Scatter plot between Price and km driven

Figure 7 is a scatter plot showing the relationship between the price of a car and the number of kilometers driven on the car. Each point on the plot represents a single car, with the x-axis representing the selling price and the y-axis representing kilometers driven. The plot shows that, in general, as the number of kilometers driven by a car increases, the price of the car decreases. The plot also shows a few outliers, where cars with a high number of kilometers driven are still priced at a relatively high value.

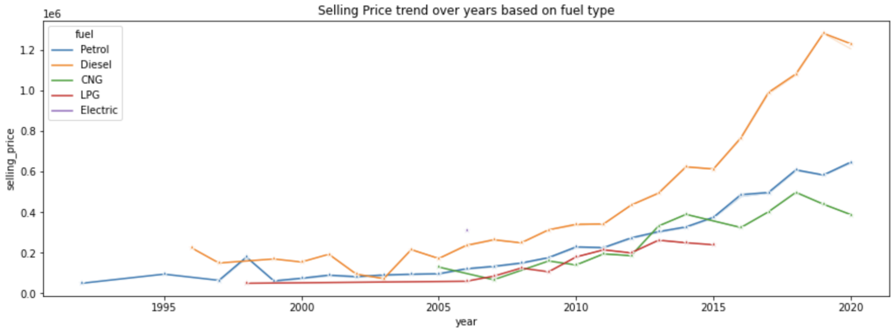


Figure 8: Selling Price trends over years based on fuel type

Figure 8 shows the trend of selling prices of vehicles, based on their fuel type over time. The x-axis represents the years and the y-axis represents the selling price. The graph shows the trend of selling prices of vehicles with diesel, petrol, electric, LPG (liquid petroleum gas), and CNG (compressed natural gas) fuel types over time. It can be observed that the selling price of vehicles with CNG fuel types, is generally lower than those with diesel and petrol fuel types. Additionally, the selling price of vehicles with diesel fuel type is generally higher than those with petrol fuel type. Overall, the trend of selling prices of vehicles with all fuel types appears to be increasing over time.

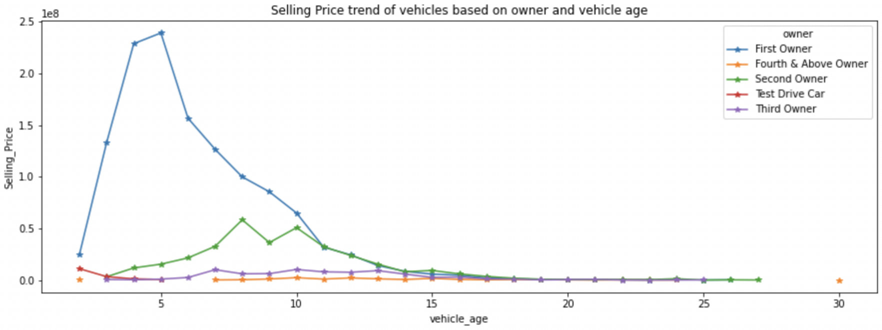


Figure 9: Selling Price trend of vehicles based on owner and vehicle age

The above illustration is a line plot that shows the trend of the selling price of vehicles based on the owners and the age of the vehicle. The x-axis represents the age of the vehicle in years and the y-axis represents the selling price of the vehicle. Each point on the plot represents a single vehicle, with the colour of the point indicating the owner the vehicle had.

From this plot, it can be seen that as the age of the vehicle increases, the selling price tends to decrease. Additionally, the selling price of vehicles that had one owner tends to be higher than those that had multiple owners. This is likely because vehicles that have had only one owner are typically in better condition and have been maintained more regularly, than those that have had multiple owners.

Overall, this visualisation provides insights into the relationship between the age and number of owners of a vehicle and its selling price. It can be used as a reference for car sellers and buyers to know the trend of the selling price of vehicles based on the number of owners and the age of the vehicle.

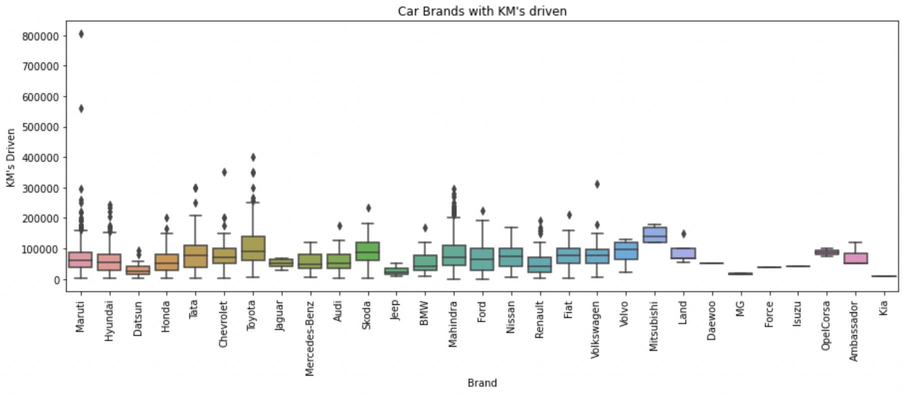


Figure 10: Car Brands with KM’s driven

The above picture is a box plot showing the total kilometers driven by different car brands. The horizontal axis of the chart lists the car brands, and the vertical axis shows the total kilometers driven, in thousands. The box plot shows that the car brand that has driven the most kilometers is ”Maruti”, with a total of around 180 thousand kilometers driven. The car brand that has driven the least kilometers is ”KIA”. The other brands have driven between 120 and 160 thousand kilometers.

4 Conclusion

To conclude, we found that after Hyundai and Mahindra, the Maruti model car has the most units in the Car Dekho dataset. Based on our analysis, diesel fuel engine prices were higher compared to other types of fuels. Apart from that, we saw that the sale price of a vehicle is correlated with its age, type of owner, and kilometers are driven. Moreover, the quality of the vehicle is determined by its number of kilometers driven and the year when the vehicle was purchased. From our visualisations, we can also see that the price is also influenced by the type of owner.

5 References

https://www.kaggle.com/datasets/akshaydattatraykhare/car-details-dataset