

**USED CAR PRICE PREDICTION**.

Submitted by:

Kanishk chahande

# ACKNOWLEDGMENT

With the covid 19 impact, we have seen lot of changes in the Automobile market. Now some cars are in demand hence making them costlier and some are not in demand hence it is cheaper. One of our clients who works with small traders, sells used cars. With the change in market due to covid 19 impact, our client is facing problems with their previous car price evaluation machine learning models. So, they are looking for new machine learning models from new data.

Based on the business requirements of the Client, I have scraped the Data from the well-known e-commerce websites such as cars 24, OLX and cardekho.com. Based on the Data collected, we will be predicting the prices of used cars. We will be building various Machine Learning models. In the end, we will see how all the machine learning models performs. And based on which we will sort the best machine learning model and hyperparameter tune the same to get the improved performance.

# INTRODUCTION

## Business Problem Framing.

### Impact of COVID-19 on Indian automotive sector

The Indian automotive sector was already struggling in FY20. before the Covid-19 crisis. It saw an overall degrowth of nearly 18 per cent. This situation was worsened by the onset of the Covid-19 pandemic and the ongoing lockdowns across India and the rest of the world. These two years (FY20 and FY21) are challenging times for the Indian automotive sector on account of slow economic growth, negative consumer sentiment, BS-VI transition, changes to the axle load norms, liquidity crunch, low-capacity utilisation and potential bankruptcies.

The return of daily life and manufacturing activity to near normalcy in China and South Korea, along with extended lockdown in India, gives hope for a U-shaped economic recovery. Our analysis indicates that the Indian automotive sector will start to see recovery in the third quarter of FY21. We expect the industry demand to be down 15-25 per cent in FY21. With such degrowth, OEMs, dealers and suppliers with strong cash reserves and better access to capital will be better positioned to sail through.

Auto sector has been under pressure due to a mix of demand and supply factors. However, there are also some positive outcomes, which we shall look at.

* With India’s GDP growth rate for FY21 being downgraded from 5% to 0% and later to (-5%), the auto sector will take a hit. Auto demand is highly sensitive to job creation and income levels and both have been impacted. CII has estimated the revenue impact at $2 billion on a monthly basis across the auto industry in India.
* Supply chain could be the worst affected. Even as China recovers, supply chain disruptions are likely to last for some more time. The problems on the Indo-China border at Ladakh are not helping matters. Domestic suppliers are chipping in but they will face an inventory surplus as demand remains tepid.
* The Unlock 1.0 will coincide with the implementation of the BS-VI norms and that would mean heavier discounts to dealers and also to customers. Even as auto companies are managing costs, the impact of discounts on profitability is going to be fairly steep.
* The real pain could be on the dealer end with most of them struggling with excess inventory and lack of funding options in the post COVID-19 scenario. The BS-VI price increases are also likely to hit auto demand.

There are two positive developments emanating from COVID-19. The China supply chain shock is forcing major investments in the “Make in India” initiative. The COVID-19 crisis has exposed chinks in the automobile business model and it could catalyse a big move towards electric vehicles (EVs). That could be the big positive for auto sector.

## Conceptual Background of the Domain Problem

Understanding the above business problem, there are certain factors that will influence the automotive industries in the future. Some of them include digital technologies, changing customer preferences, electrical vehicles, intelligent ability, and technical advancements. Technologies such as artificial intelligence, machine learning, cloud computing, and internet of things will also play an important role in developing new business models. Apart from that, they enable customers to ensure a better mobility experience. In other words, technologies may impact automotive industry units significantly that will change the markets. The introduction of electrical cars and hybrid vehicles may transform the automobile industries in coming years.

## Review of Literature.

As per the requirement of our client, I have scrubbed data from different used cars selling merchants websites, and so based on the data collected I have tried analysing based on what factors the used car price is decided? What is the relationship between cost of the used cars and other factors like Fuel type, Brand and Model, year the car is purchased and No. Of owners before selling? And so based on all the above consideration I have developed a model that will predict the price of the used cars.

## Motivation for the Problem Undertaken

I have taken this problem based on the requirement of the client and also, with a curiosity to know how the used cars markets are at the time of pandemic.

**Exploratory Data Analysis**

1. Checking the Missing Values

2.categorical data visualisation

3.numerical data and label relationship

4.checking outlier

5.checking skewness and distribution of numerical data

* we will find missing values
* LOCATION 0
* MNF\_YEAR 0
* BRAND 0
* MODEL 0
* VARIANT 287
* DRIVEN\_KM 0
* FUELTYPE 0
* NOOF\_OWNERS 0
* PRICE 0
* dtype: int64
* We have seen there are 287 null values in variant so we have to fill it.after knowing
* Manual 6461
* Automatic 1110
* Name: VARIANT, dtype: int64
* Manual is mode so we will fill missing value with manual

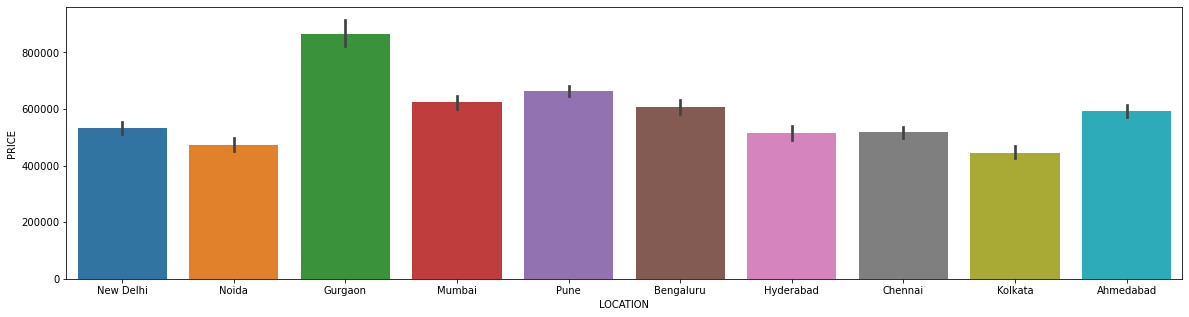
we have to convert data type of columns which has wrong data type

df['PRICE'] = df['PRICE'].str.replace(r'\D', '').astype(int)

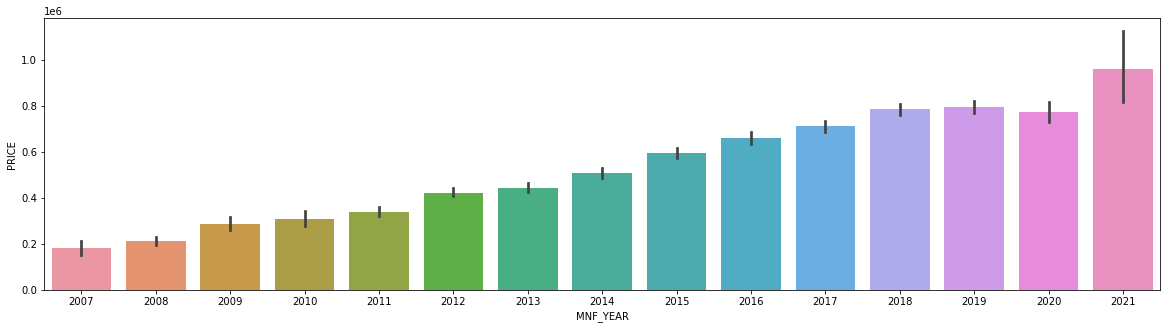
df['DRIVEN\_KM'] = df['DRIVEN\_KM'].str.replace(r'\D', '').astype(int)

DATA VISUALISATION

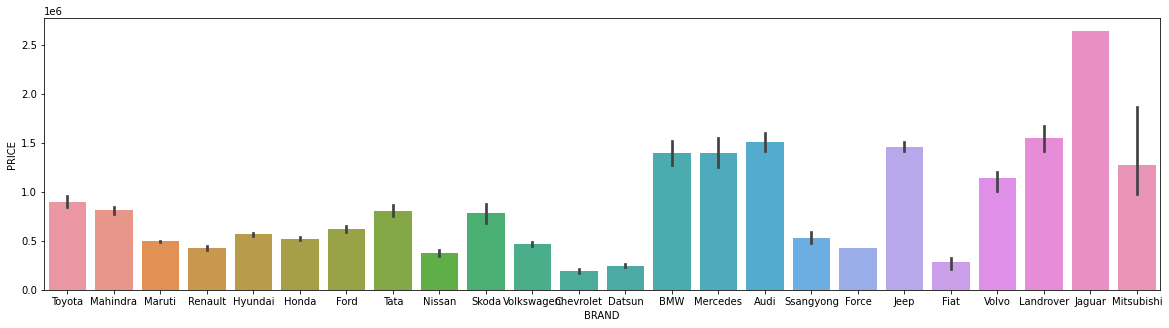
PRICE VS LOCATION



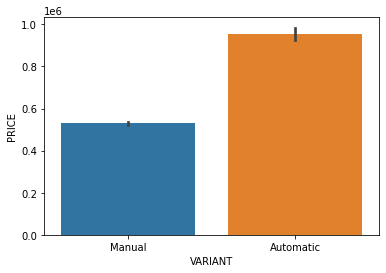
Price vs MNF YEAR



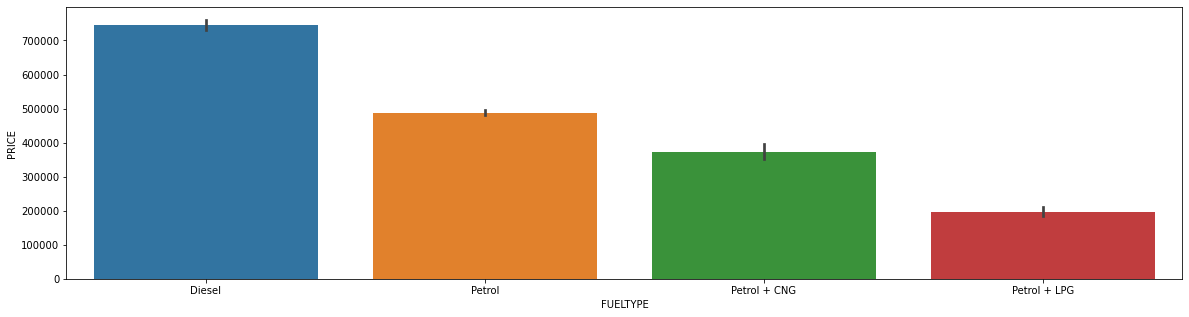
PRICE VS BRAND



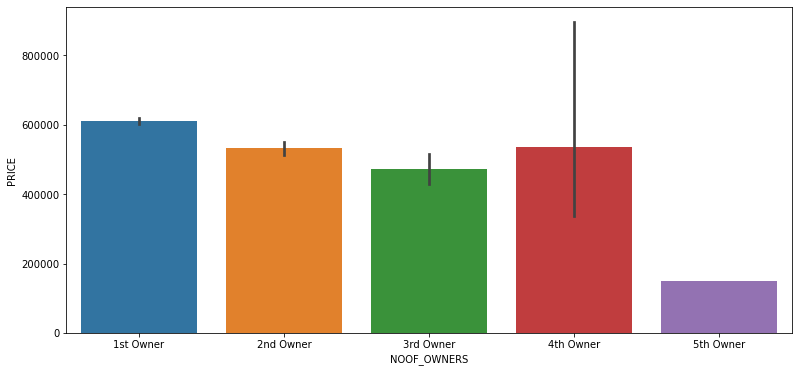
PRICE VS VARIANT



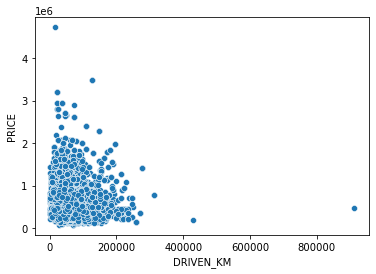
PRIVE VS FUELTYPE



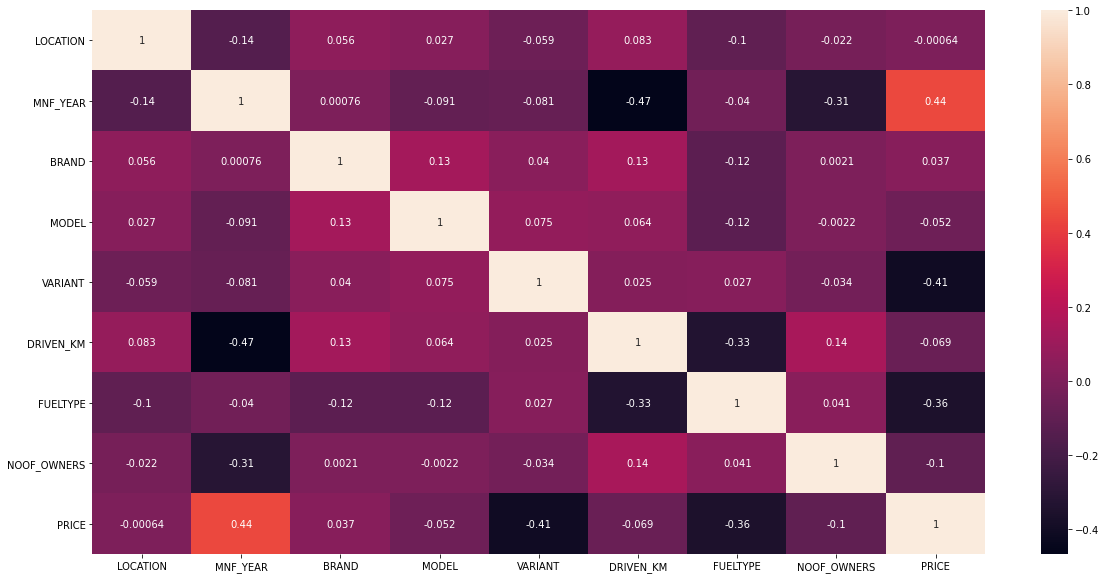
PRICE VS NOOF\_OWNERS



PRICE VS DRIVEN-KM



WE WILL DEFINE NOW CORRELATION



KEY OBSERVATION:

1. New Delhi, Noida, Gurgaon have the costliest cars and Mumbai, Pune, Ahmedabad have most cars being sold.
2. We have Bengaluru, Hyderabad, Chennai, Kolkata has least cars being sold and also comparatively cheaper.
3. 2019, 2020, 2021 model are being sold higher in PRICE and also above average PRICE.
4. When comes to the Brand Land rover are being the costliest in country followed by Jaguar
5. We can see that the automatic engines are costliest in the market. And also, most costlier cars come in Automatic shift.
6. The lesser kms driven are evidently sold costlier.
7. 1st owner cars are costliest followed by second and third.

## Hardware and Software Requirements and Tools Used

1. Python 3.8.
2. NumPy.
3. Pandas.
4. Matplotlib.
5. Seaborn. 6. Data science.
6. SciPy
7. Sklearn.
8. Anaconda Environment, Jupyter Notebook.

MODEL BUILDING

* We will use five model
* 1) linear regression
* 2)randomforest regressor
* 3)knneighbour regressor
* 4)gradientbossting regressor
* 5) AdaBoostRegressor

WE HAVE USED THESE MODEL AND WE GOT HIGHEST ACCURACY IN gradientbossting REGRESSOR

WE WILL THEN HYPERTUNE THAT MODEL

FROM ABOVE AFTER HYPWERTUNING OUR ACCURACY IS 87%

## Interpretation of the Results

From the visualization above we can clearly understand that the used car price factors are decided by the factors such as brand, location, model, year made, number of owners used the car before, fuel type of the car.

From that we can clearly say that the used car price depending on the Brand that is the manufacturer and model it varies.

# CONCLUSION

## Key Findings and Conclusions of the Study

The manufacturer like Land Rover, Benz, BMW cars are costliest used car in the market comparatively to other cars, the low kilometres driven and also if the manufacturing year is lesser on these brands those card sells in much higher rates or closest to the buying new car rates. The Diesel variant and Automatic shift variants are also costliest user car variants in the used car market.

## Learning Outcomes of the Study in respect of Data Science

The above research will help our client to study about the latest used car market and with the help of the model built he can easily predict the price ranges of the cars, and also will helps him to understand based on what factors the Car Price is decided.