



Welcome

Project Report on

***USE CAR PREDICTION
FROM CARS24.COM***

BY

Members of Team F

Internship Program

Under the Guidance of:

EVOASTRA VENTURES



Participants Members of Team F

Dhananjay D. Nagare (TL)

Shibasis Karmakar (Co-TL)

Arati Y. Dange

Radha Gohil

Raghav Bhardwaj

Indrajeet

Mobashshir ALi

Nayash Sayyed

Suryansh Gupta

Yugasri

Dange Arati

Tejas Pawar

Nikhil Aswal



Project Overview

Problem Statement

As used car's sale is increasing globally, there is a need of it's price prediction system to determine business effectively.

Solution

A model can be developed to predict the prices.

To determine the significance of used cars by studying their features such as KM Driven, Purchase Year ,Price etc.

Data Collection

- ☐ *For precise and real-time analysis, the data is scrapped from Cars24 website.*
- ☐ *We used BeautifulSoup/Selenium for scrapping to extract the features and stored it in .csv file.*
- ☐ *<https://www.cars24.com>*
- ☐ *Data Set consist of four major City & 7 Brands/ Company of the cars.*

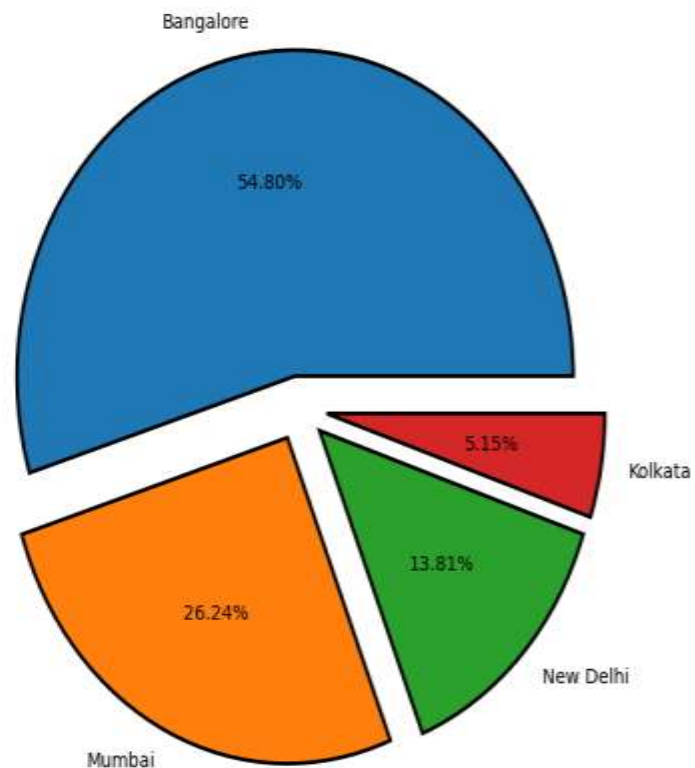
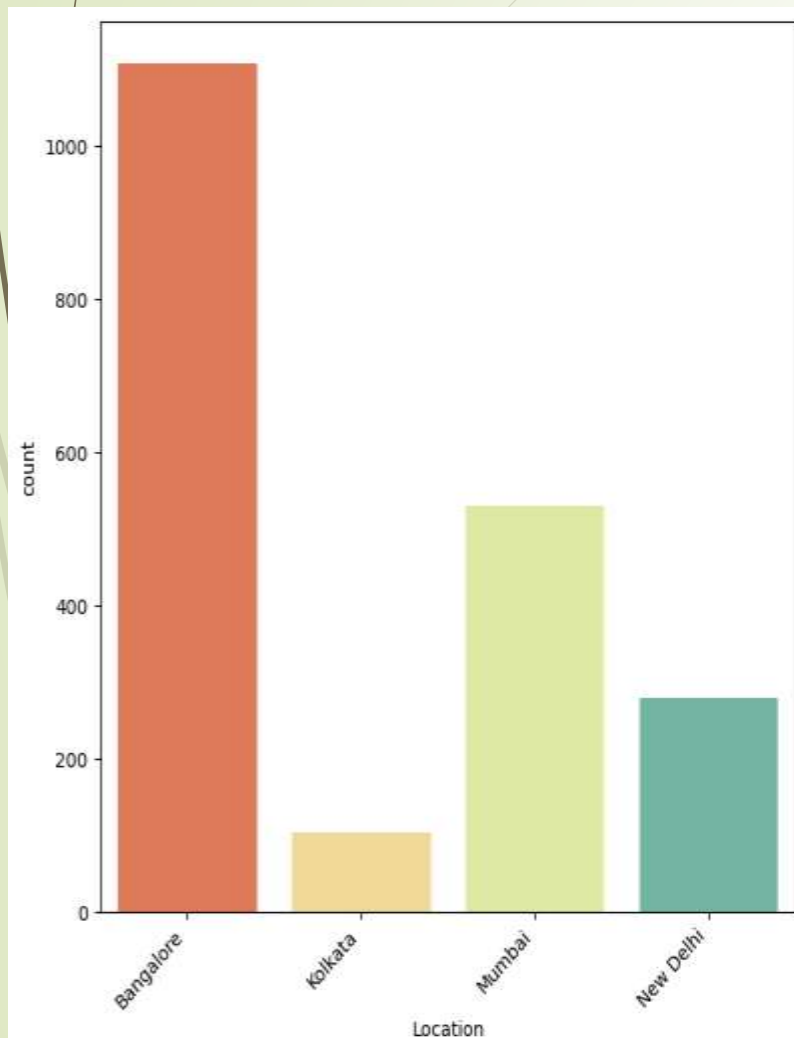
Dataset Description

- ❑ *18180 DATAPPOINTS WITH 2020 ROWS AND 9 COLUMNS WERE OBTAINED VIA SCRAPPING AS SHOWN IN THE TABLE BELOW.*

	Name	Model	Company	Year	Fuel Type	KM Driven	Transmission Type	Price(in Lakh)	Location
0	Hyundai Creta SX PLUS AT 1.6 PETROL	Creta SX PLUS AT 1.6 PETROL	Hyundai	2017	PETROL	98493	Automatic	973000	Bangalore
1	Renault Kwid 1.0 MARVEL IRON MAN EDITION AMT	Kwid 1.0 MARVEL IRON MAN EDITION AMT	Renault	2018	PETROL	19178	Automatic	407000	Bangalore
2	Hyundai Eon ERA PLUS (O)	Eon ERA PLUS (O)	Hyundai	2017	PETROL	33963	Manual	381000	Bangalore
3	Maruti Swift VXi	Swift VXi	Maruti	2012	PETROL	64557	Manual	463000	Bangalore
4	Hyundai Creta SX 1.6 DIESEL	Creta SX 1.6 DIESEL	Hyundai	2019	DIESEL	43987	Manual	1150000	Bangalore

Exploratory Data Analysis

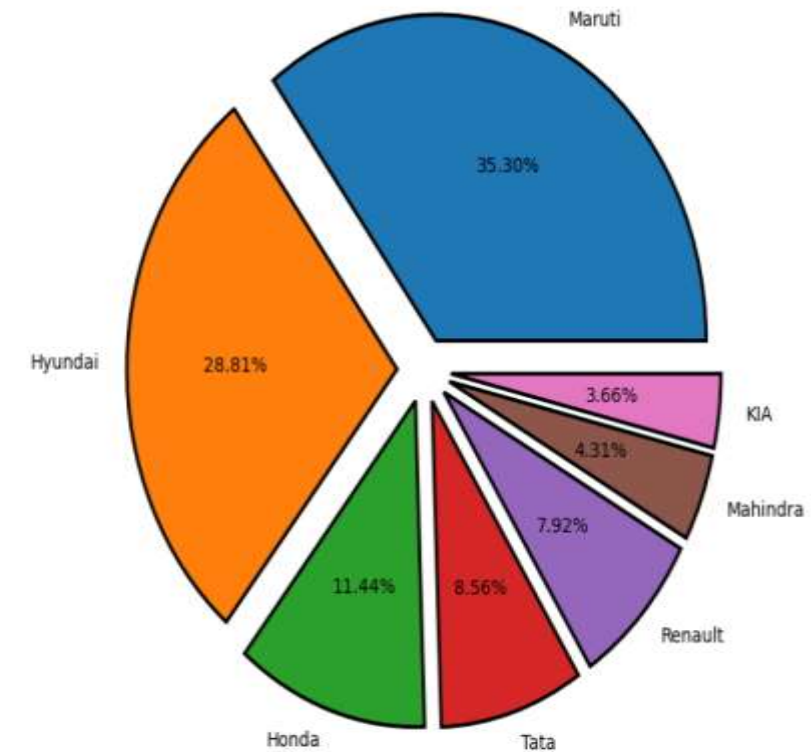
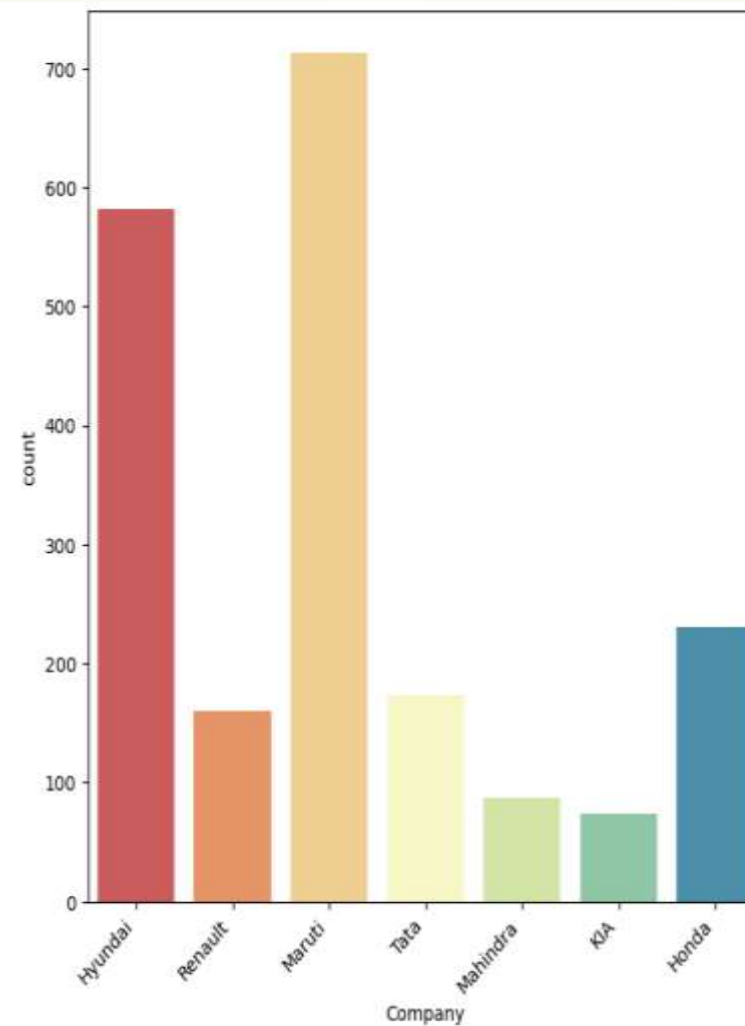
Location Wise Visualization



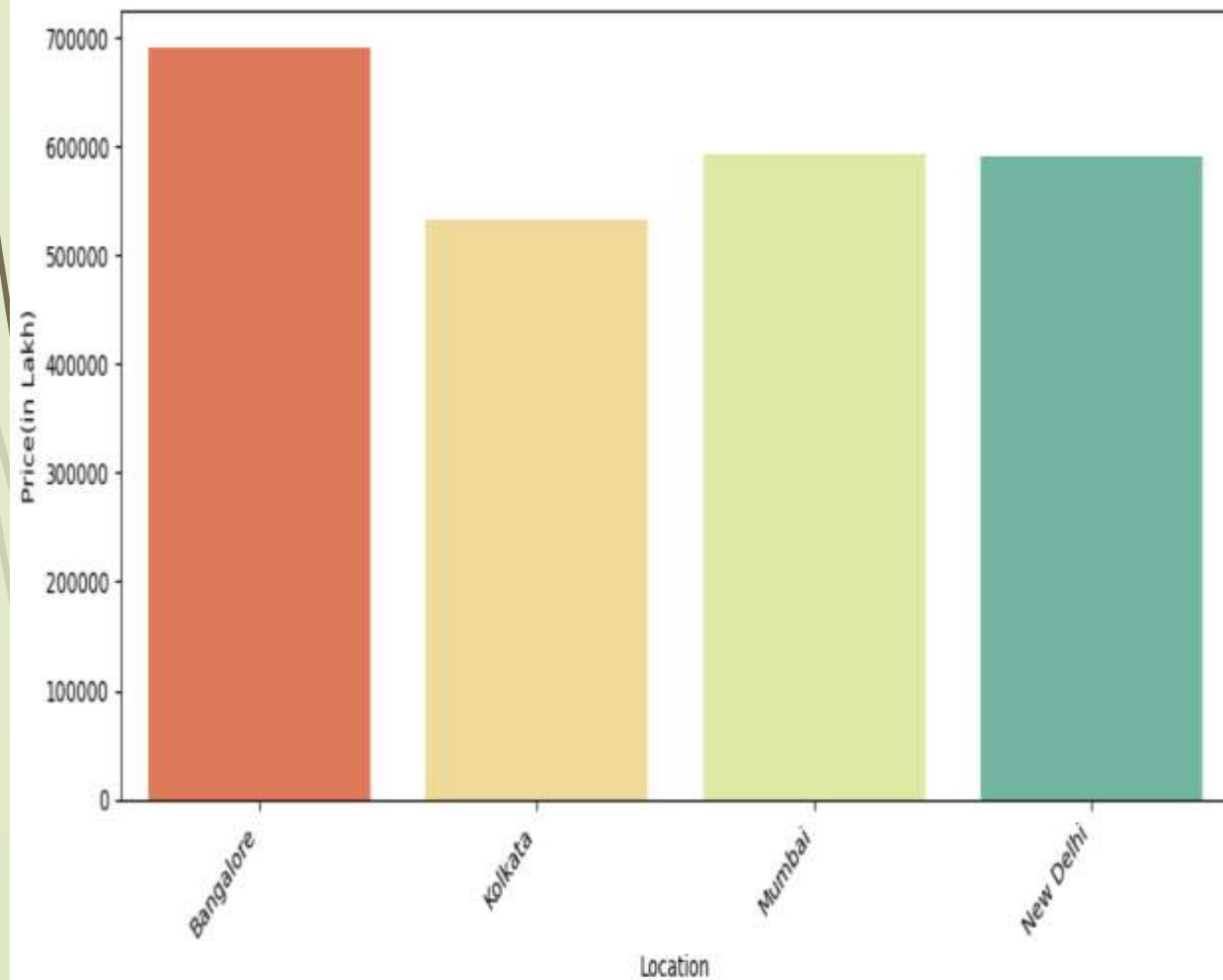
According to the plot we can determined that location wise Bangalore has the highest percentage in selling the used car then Mumbai, New Delhi & Kolkata.

Company Wise Visualization

As per the company details of used car we can see that Maruti has the highest selling used car then Hyundai & Honda respectively.



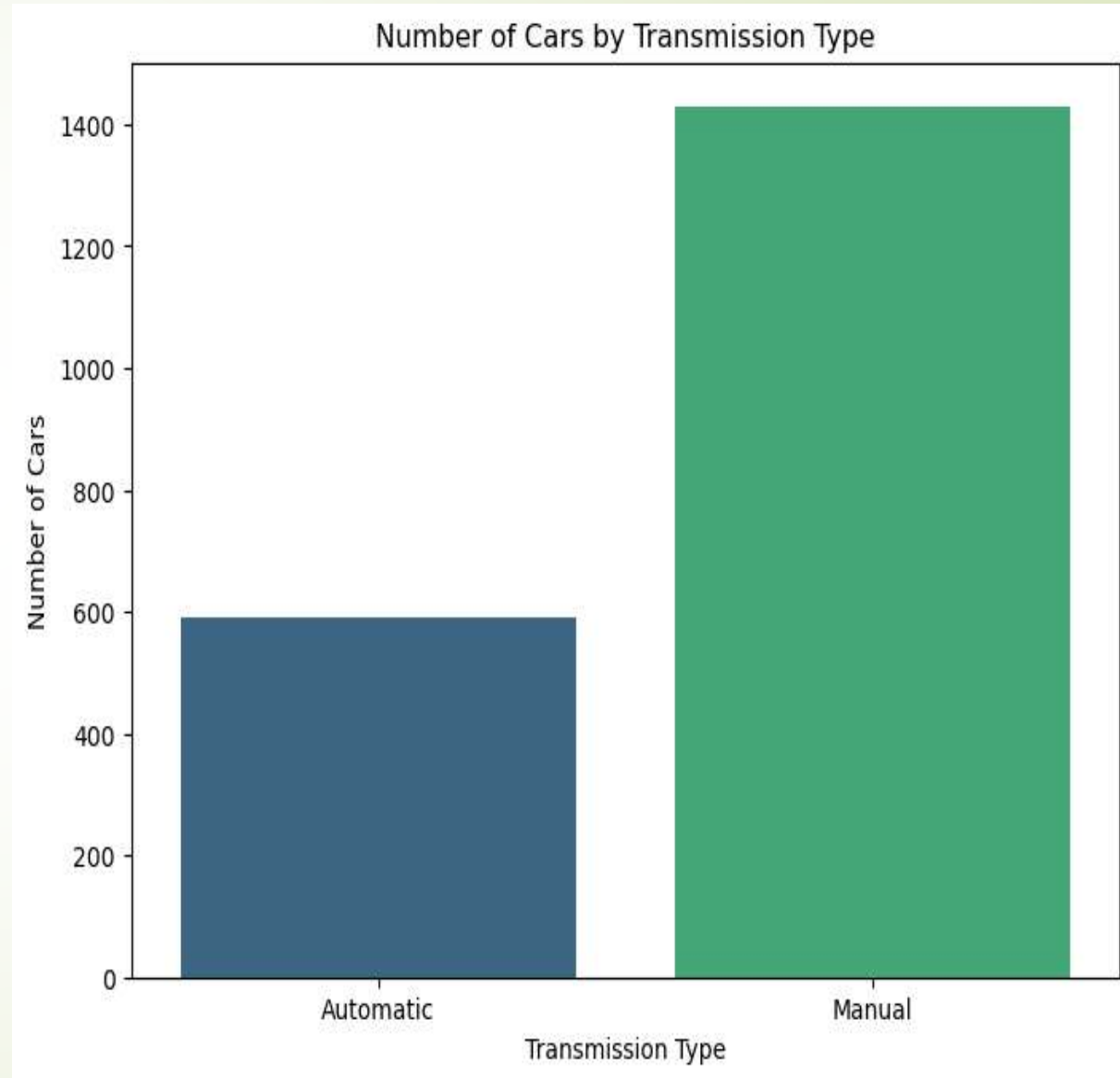
Price of the used car Location Wise



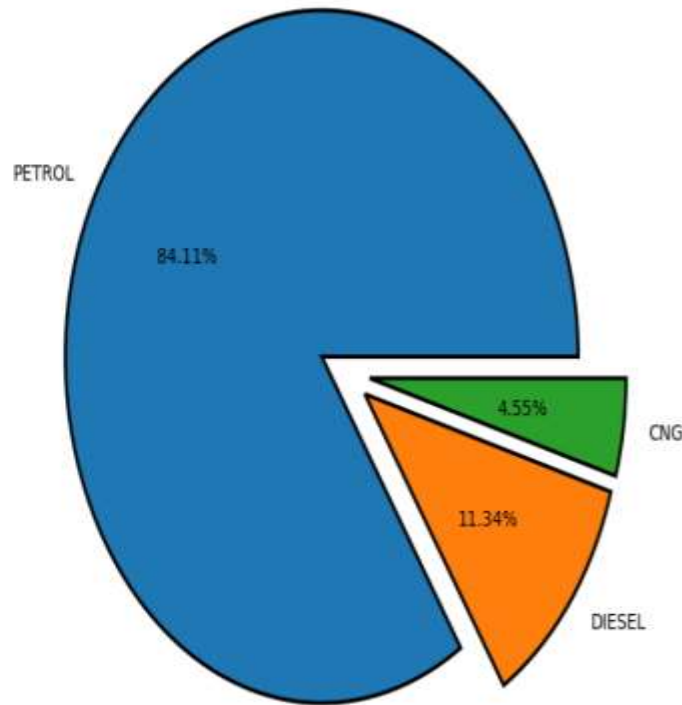
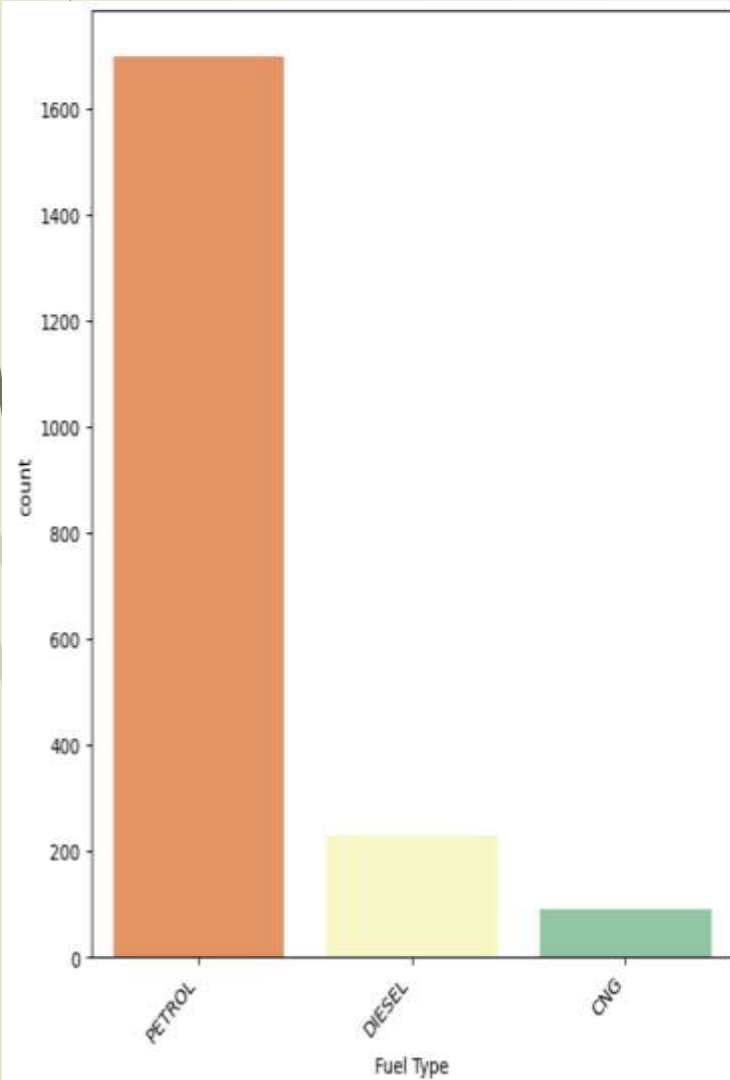
As per the data of selling price based on location Bangalore has the highest selling price of used car rather than New Delhi & Mumbai.

Transmission type of cars

***We can see from the figure that
Transmission of used car has more Manual
impact than Automatic***



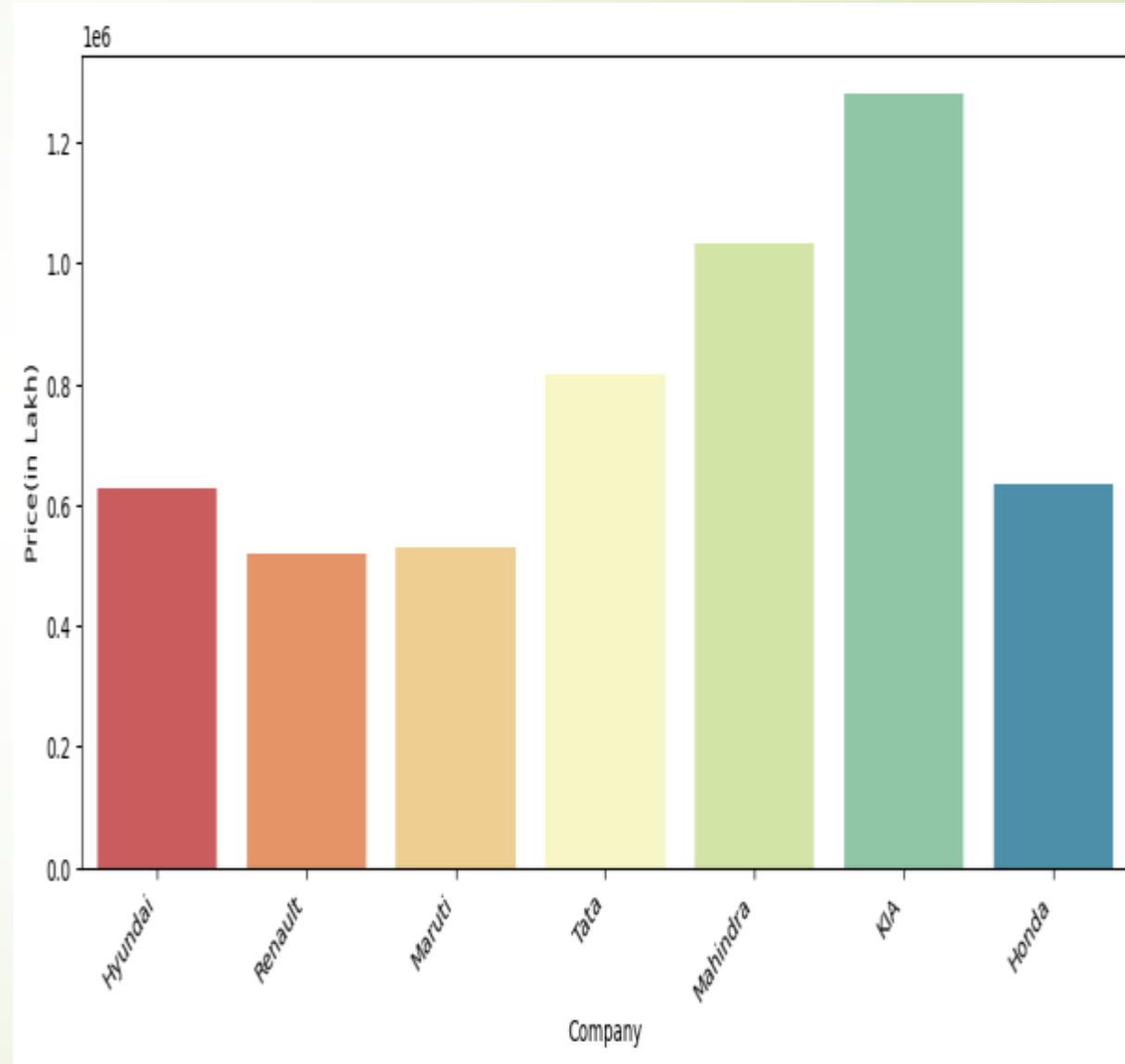
Cars based on Fuel



As per the fuel category of used car it shows clearly that Petrol has highest uses than Deisel & CNG

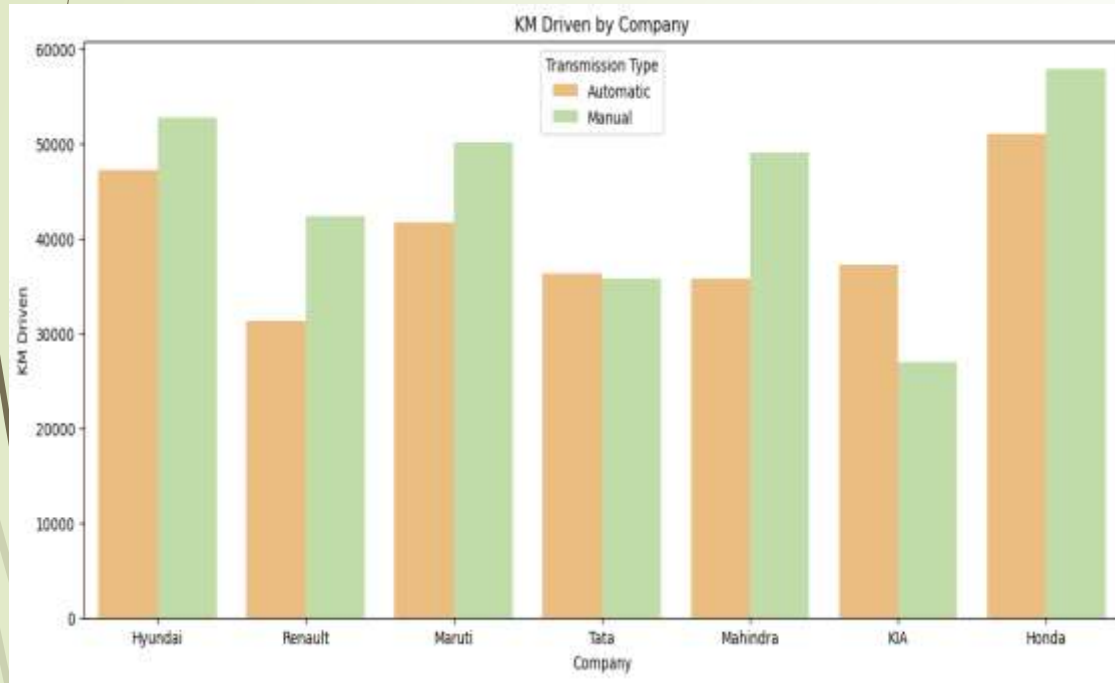
Price Distribution for various companies

Company like KIA has the most expensive used car sold out of the top 7 company.

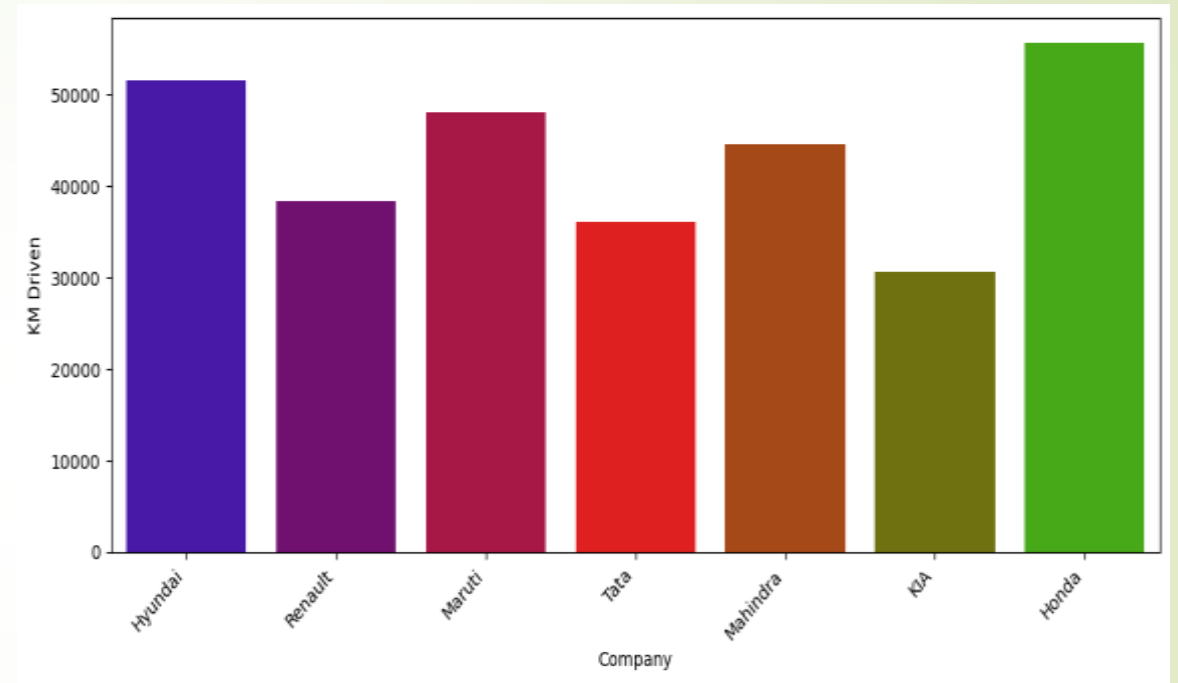


Cars based on KM Driven

According to Transmission type and Company



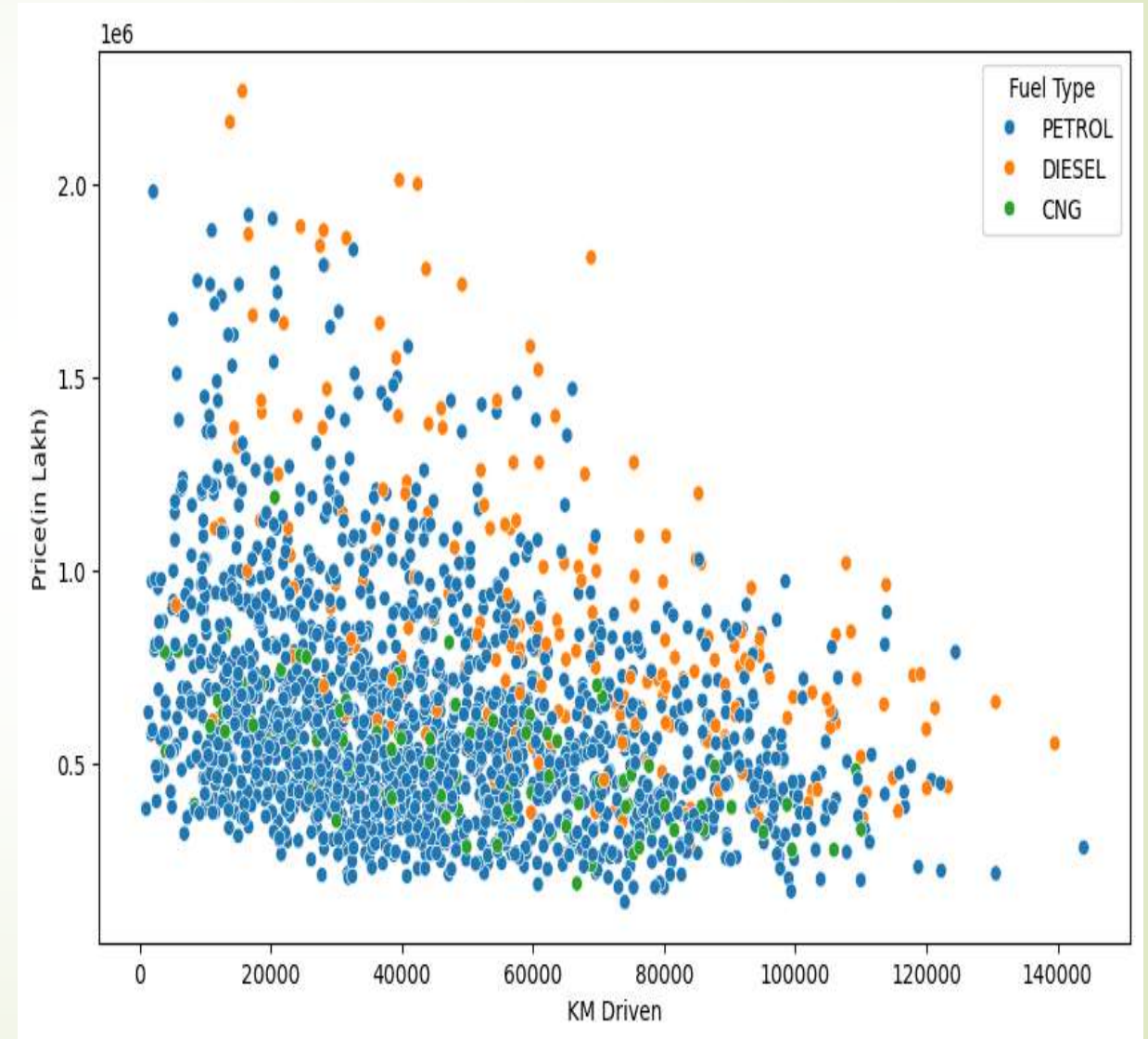
According to Company



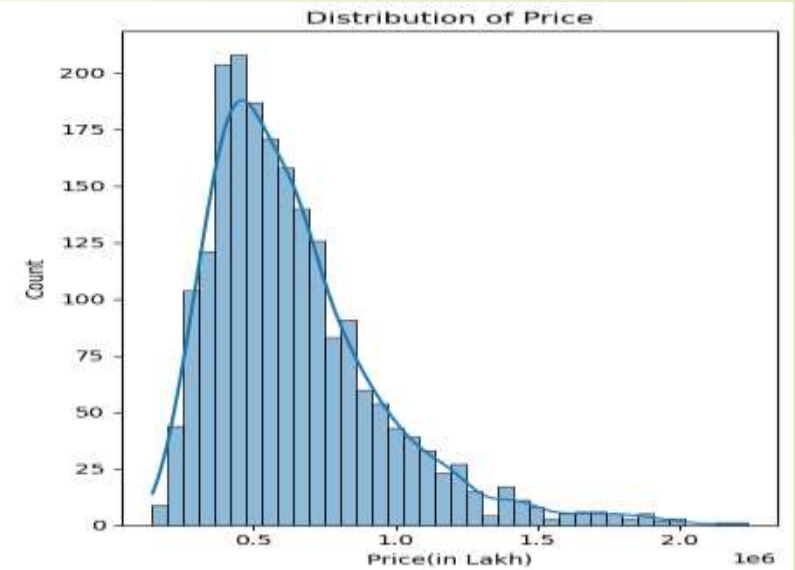
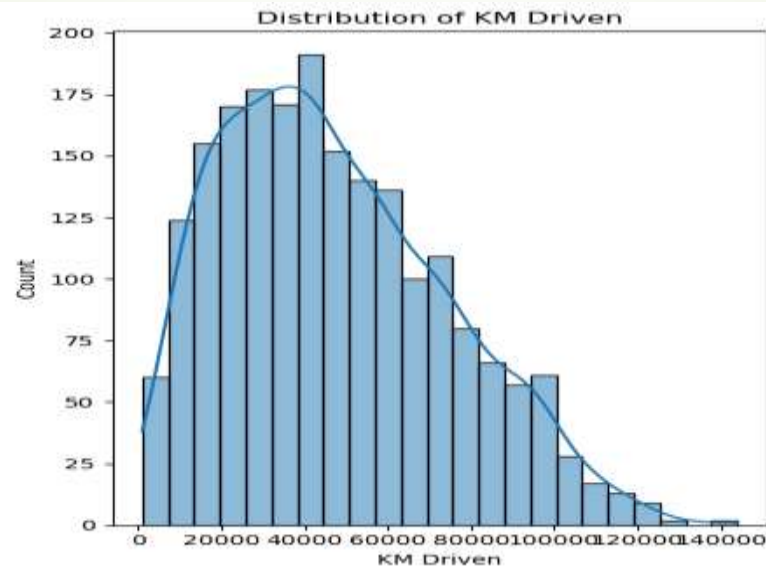
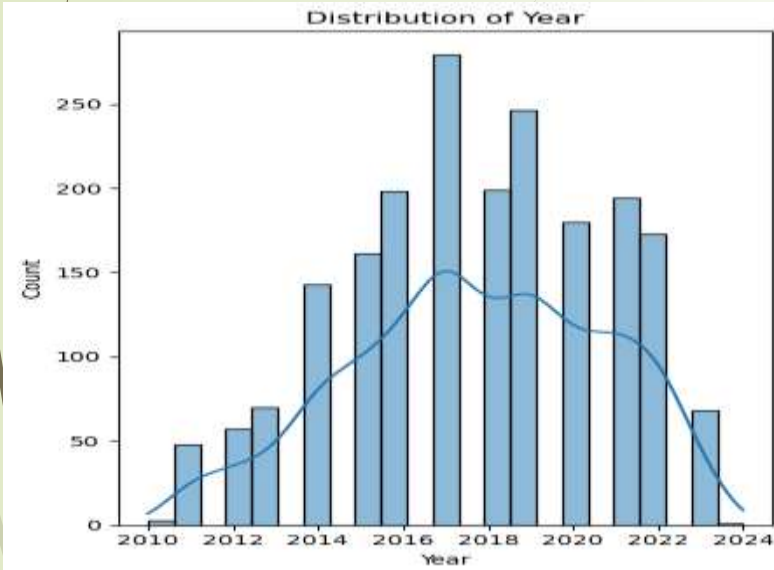
***One of the most Important Features in Used Car is KM driven.
Honda has the most km driven car company & also the most highest manual
transmission car as well.***

Car price according to KM Driven & Fuel type

Another most Important Features in Used Car is KM driven & Price . Here in the plot we can observe that less KM driven and Diesel features has the highest impact in used car sold.



Distribution of Year, KM Driven, and Price



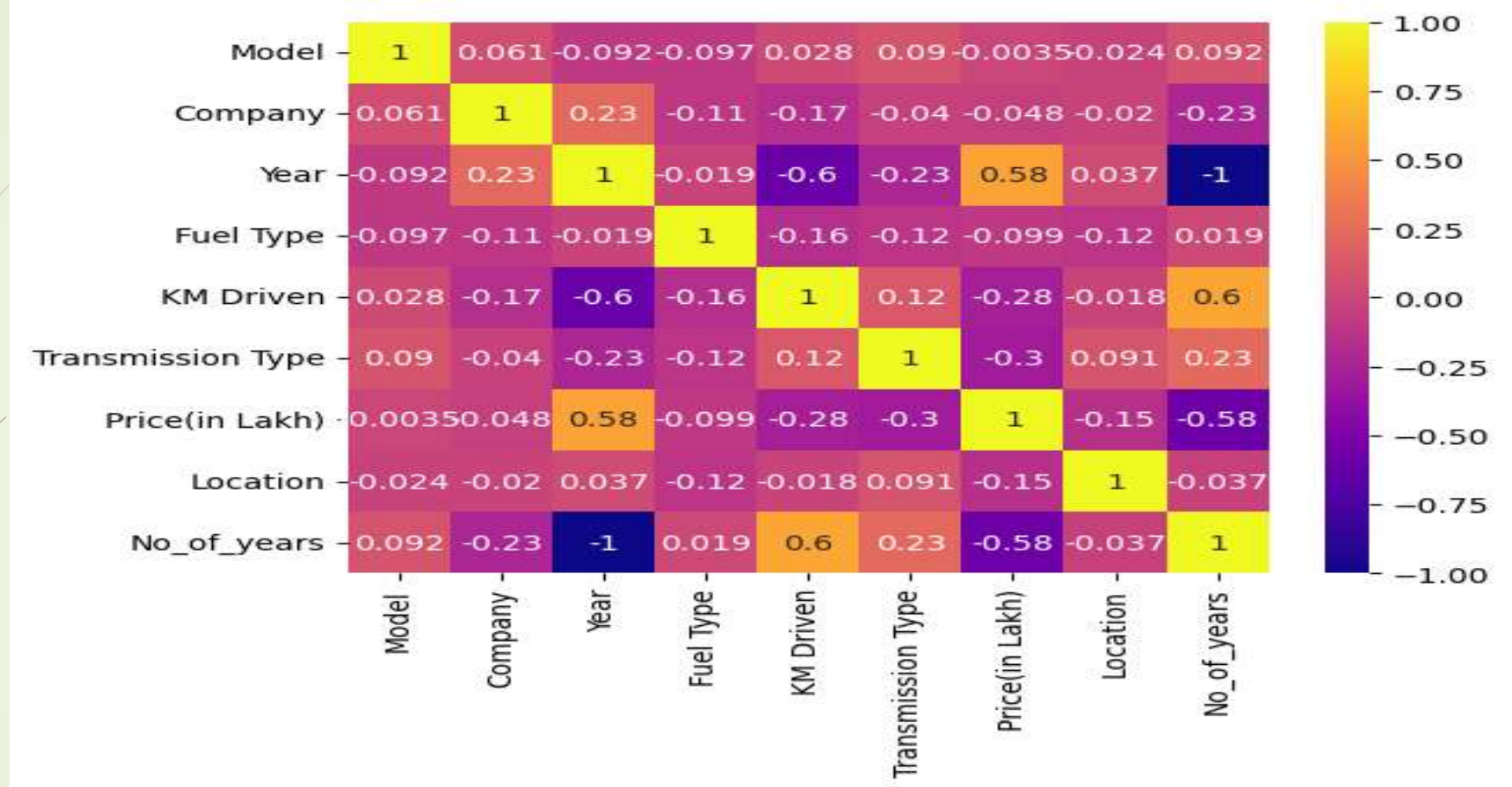
***According to the above plots we can see that
In the year 2017 has the highest selling Car & most of the car has been driven in
between 40-60 thousands KM. And Price of the use car sold in between 4-7 lakhs.***

Descriptive Statistics

	count	mean	std	min	25%	50%	75%	max
Year	2020.0	2017.724752	3.003792	2010.0	2016.00	2018.0	2020.0	2024.0
KM Driven	2020.0	47236.418812	27135.172888	1011.0	25748.75	43399.5	65590.0	143991.0
Price(in Lakh)	2020.0	642743.069307	307385.776635	144000.0	426750.00	576000.0	780000.0	2240000.0
No_of_years	2020.0	6.275248	3.003792	0.0	4.00	6.0	8.0	14.0

From the above stats we can clearly observed that maximum 14 years used car has been sold & the highest price is 22 lakhs 40 thousands only.

Correlation Matrix



WE CAN CONCLUDE THAT, THERE IS,

- 1. POSITIVE CORRELATION BETWEEN KM DRIVEN & NUMBER OF YEARS.*
- 2. POSITIVE CORRELATION BETWEEN YEAR & PRICE.*

Machine Learning Models

❓ *Decision Tree*






☺ *AdaBoost Regressor*

❓ *Random Forest*

🍍 *Linear Regression*

❓ *Gradient Boosting*

Results of the Models

	Models	Train R2_score	Test R2_Score	Mean Square Error on Testing	Mean absolute Error on Testing	Mean absolute percentage error
0	 Decision Tree	100.000000	88.640000	0.002751	0.034689	0.157409
1	 AdaBoost Regressor	55.260000	59.150000	0.009886	0.078815	0.553272
2	 Random Forest	98.590000	92.040000	0.001928	0.028746	0.144613
3	 Linear Regression	45.900000	44.550000	0.013420	0.083037	0.406110
4	 Gradient Boosting	87.790000	84.370000	0.003783	0.038095	0.172848

Models and their Accuracy

Decision Tree Regressor

$$R^2 = 88.64\%$$

$$MAE = 3.5\%$$

Gradient Boosting Regressor

$$R^2 = 84.37\%$$

$$MAE = 3.8\%$$

Random Forest Regressor

$$R^2 = 92.04\%$$

$$MAE = 2.87\%$$

Linear Regression

$$R^2 = 44.55\%$$

$$MAE = 8.3\%$$

Ada Boost Regressor

$$R^2 = 59.15\%$$

$$MAE = 7.88\%$$

Conclusion

CONSIDERING ALL MODELS AND THEIR ACCURACY, MODELS THAT CAN GIVE BEST PREDICTIONS FOR USED CARS PRICES ARE :

- ☐ *Random Forest*
- ☐ *Decision Tree*
- ☐ *Gradient Boosting regressor*



Thank You