CAR PRICING MODEL

FLIP ROBO

Problem Statement

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

Theoretical Background

- There is no doubt that the Covid-19 pandemic has had a positive influence on used car pricing.
- There is huge pricing being paid for Trade-in values from both Dealers and Auction houses, if you have a late model 2,3- or 4-year-old car – why not make enquiries on what it is worth today, you could be pleasantly surprised.
- The 'fear' of getting back on public transport, working from home, in fact any number of reasons have resulted in the demand for late model secondhand cars to increase significantly.
- If you are thinking that now would be a good time to trade-up and take advantage of the Instant Asset Write-off allowances on offer, you may very well get a great price on your current vehicle, drive away in a new car, and claim the full price of the new car as an instant write-off this financial year. Talk to your accountant and see if this makes perfect business sense for your company.

EDA-Step-wise

EDA

```
In [80]: ds.shape
Out[80]: (5044, 7)
```

```
In [81]: ds.info()
```

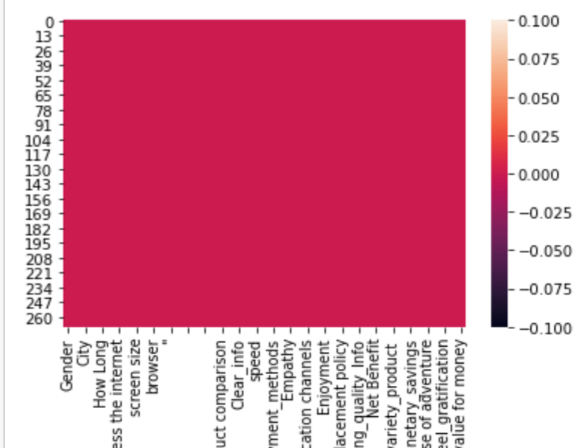
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5044 entries, 0 to 5043
Data columns (total 7 columns):
 #   Column      Non-Null Count  Dtype
---  ------      --------------  -----
 0   Unnamed: 0  5044 non-null   int64
 1   Brand       5044 non-null   object
 2   Price       5044 non-null   object
 3   Model       5044 non-null   object
 4   KMS_driven  5044 non-null   object
 5   Fuel        5044 non-null   object
 6   Variant     5044 non-null   object
dtypes: int64(1), object(6)
memory usage: 276.0+ KB
```

There are 5044 rows and 6 columns.

The dataset contains 5044 rows and 7 columns in the dataset.

The heat map shows there is no missing values in the data set.

```
In [6]: sns.heatmap(ds.isnull())
Out[6]: <AxesSubplot:>
```



Preparing the data for Analysis and Model Building

```
# Extract price in from price variable in numeric form
ds['Price'] = ds['Price'].str.split().str[0]
ds
```

	Unnamed: 0	Brand	Price	Model	KMS_driven	Fuel	Variant
0	0	2017 Maruti Swift	5.79	VDI BSIV	84,730 kms	Diesel	Manual
1	1	2017 Maruti Ignis	4.93	1.2 Zeta BSIV	36,985 kms	Petrol	Manual
2	2	2012 Hyundai i10	3.41	Sportz	73,717 kms	Petrol	Manual
3	3	2017 Maruti Celerio	4.14	ZXI	25,149 kms	Petrol	Manual
4	4	2018 Maruti Ignis	4.7	1.2 Delta BSIV	7,714 kms	Petrol	Manual
...	...	...	...	...	...	...	...
5039	5039	2018 Honda BR-V	8.35	i-VTEC S MT	8,972 kms	Petrol	Manual
5040	5040	2015 Mercedes-Benz S-Class	53.75	S 350 CDI	53,500 kms	Diesel	Automatic
5041	5041	2018 Mahindra Scorpio	9.9	S5 BSIV	65,700 kms	Diesel	Manual
5042	5042	2013 Mercedes-Benz M-Class	20	ML 350 4Matic	53,000 kms	Diesel	Automatic
5043	5043	2015 Hyundai Creta	9.1	1.6 VTVT SX Plus	28,000 kms	Petrol	Manual

```
# Extracting year from Brand
ds['Manuf_Year'] = ds['Brand'].str.split().str[0]
ds['Brand'] = ds['Brand'].str.split().str[1]
ds
```

	Unnamed: 0	Brand	Price	Model	KMS_driven	Fuel	Variant	Manuf_Year
0	0	Maruti	5.79	VDI BSIV	84,730	Diesel	Manual	2017
1	1	Maruti	4.93	1.2 Zeta BSIV	36,985	Petrol	Manual	2017
2	2	Hyundai	3.41	Sportz	73,717	Petrol	Manual	2012
3	3	Maruti	4.14	ZXI	25,149	Petrol	Manual	2017
4	4	Maruti	4.7	1.2 Delta BSIV	7,714	Petrol	Manual	2018
...	...	...	...	...	...	...	...	...
5039	5039	Honda	8.35	i-VTEC S MT	8,972	Petrol	Manual	2018
5040	5040	Mercedes-Benz	53.75	S 350 CDI	53,500	Diesel	Automatic	2015
5041	5041	Mahindra	9.9	S5 BSIV	65,700	Diesel	Manual	2018
5042	5042	Mercedes-Benz	20	ML 350 4Matic	53,000	Diesel	Automatic	2013
5043	5043	Hyundai	9.1	1.6 VTVT SX Plus	28,000	Petrol	Manual	2015

5044 rows × 8 columns

```
# Extract price in from price variable in numeric form
ds['KMS_driven'] = ds['KMS_driven'].str.split().str[0]
ds
```

	Unnamed: 0	Brand	Price	Model	KMS_driven	Fuel	Variant
0	0	2017 Maruti Swift	5.79	VDI BSIV	84,730	Diesel	Manual
1	1	2017 Maruti Ignis	4.93	1.2 Zeta BSIV	36,985	Petrol	Manual
2	2	2012 Hyundai i10	3.41	Sportz	73,717	Petrol	Manual
3	3	2017 Maruti Celerio	4.14	ZXI	25,149	Petrol	Manual
4	4	2018 Maruti Ignis	4.7	1.2 Delta BSIV	7,714	Petrol	Manual
...	...	...	...	...	...	...	...
5039	5039	2018 Honda BR-V	8.35	i-VTEC S MT	8,972	Petrol	Manual
5040	5040	2015 Mercedes-Benz S-Class	53.75	S 350 CDI	53,500	Diesel	Automatic
5041	5041	2018 Mahindra Scorpio	9.9	S5 BSIV	65,700	Diesel	Manual
5042	5042	2013 Mercedes-Benz M-Class	20	ML 350 4Matic	53,000	Diesel	Automatic
5043	5043	2015 Hyundai Creta	9.1	1.6 VTVT SX Plus	28,000	Petrol	Manual

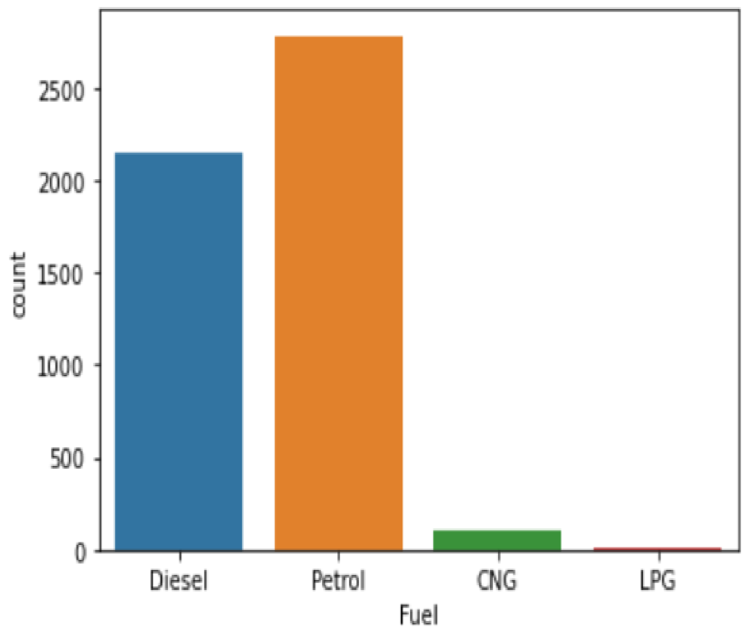
5044 rows × 7 columns

Firstly we need to extract manufacturing year from Brand column, secondly extract price in numeric form from price column, thirdly kilometer driven in numeric form. Lastly drop the unnamed column from the data set.

Data Analysis

```
sns.countplot(x='Fuel',data=data)
```

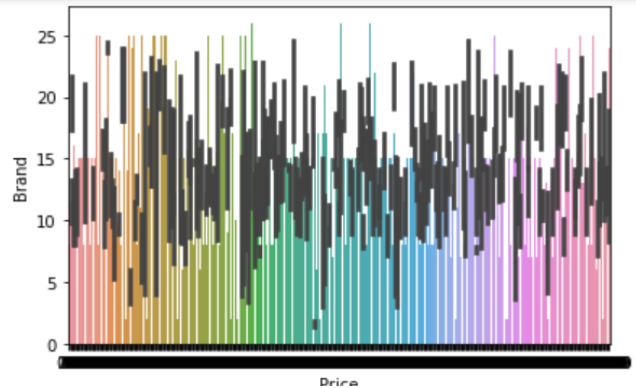
<AxesSubplot:xlabel='Fuel', ylabel='count'>

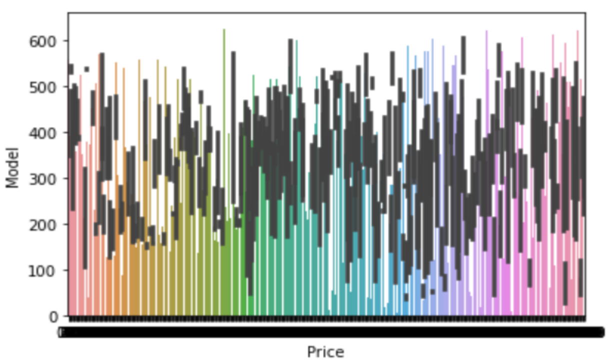


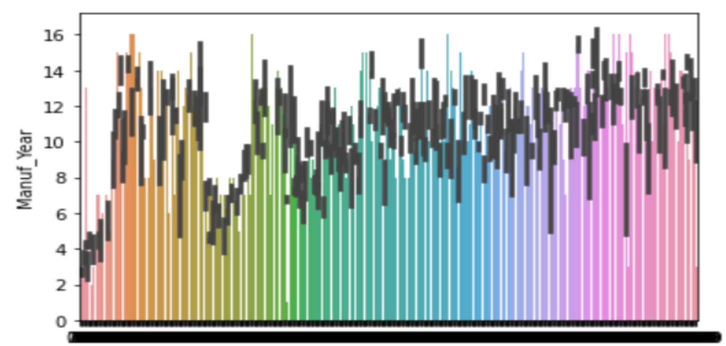
- Key Observation: We can see a trend in data. The above plot Very few CNG and LPG fuel used cars are available on sale.
- Majority of the customer agree and strongly agree that clear & adequate info helps in buying the product.

Visualization

```python
for i in ds.columns[:7]:
    sns.barplot(x = ds['Price'], y = ds[i])
    plt.show()
```







Price of the car is inversely proportional to the price and the Year of Manufacturing.

The price of the car is ihgly influenced with the Brand and Model.

Model Building

We built LinearRegression, Decision Tree Regessor, and Random Forest as machine learning. We built base models of LinearRegression, XGBoost, and RandomForest so there is not much to show about these models but we can see the model summary and how they converge with deep learning models that we built.

- Majority of the customer agree and strongly agree that monetary savings and empathy toward the customers helps in purchasing the product.

Model Evaluation

Based on the results of above models, and capomaring the R2 score and other evalution matrics result of MAE,MSE and RMSE.

We can find the Random Forest Regression model is best model to predict the price of used cars Score R2 Score MAE MSE RMSE 1

LinearRegression 0.11 0.16 110 1800 134

DecisionTreeRegression 1 0.99 .86 113 10 3

KNeighbors Regression 0.76 0.53 70 10229 101

4 Random Forest Regression 0.97 0.97 10 498 22

Since the DecsionTree Regression model has the hishest score(1) and R2 score(0.99) and lowest values of MAE, MSE,RMSE , it is the best model among the above four models

It can be observed that the "DecisionTree Regressor" algorithim has the almost some positive accuracy score after the cross validation. So the best model with highet accuracy score and best on evalution with other matrix is "DecisionTree" is the best model for predicting the Price.

Conclusion

- In this project, we tried predicting the car price using the various parameters that were provided in the data about the car. We build machine learning models to predict car prices and saw that machine learning-based models performed well at this data.

- By performing different ML models, we aim to get a better result or less error with max accuracy. Our purpose was to predict the price of the used cars having 7 predictors and 5044 data entries. Initially, data cleaning is performed to remove the null values and outliers from the dataset then ML models are implemented to predict the price of cars. Next, with the help of data visualization features were explored deeply. The relation between the features is examined.

- From the below table, it can be concluded that Decision Tree is the best model for the prediction for used car prices. The regression model gave the best MSLE and RMSLE values

Thank You

Submitted by
Dinesh Sharma