CAR PRICE PREDICTION PROJECT

PROBLEM STATEMENT:

With the covid19 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 covid19 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.

UNDERSTANDING:

This project aims to predict the price of an used Car by taking it's Company name, it's variant, price details, how it work whether it work manual or by automatic and many other parameters it can be predicted. So, here in this it required to the making of model of the price of cars with the available independent variables. It will be used by the management to understand how exactly the prices vary with the independent variables to meet certain price levels. Further, the model will be a good way for management to understand the pricing dynamics of a new market.

A primary objective of this project is to estimate used car prices by using some attributes that are highly correlated with a Price. By doing some research on this project, are able to trained the models and predicting things make the previous background to work efficiently.

EDA STEPS AND VISUALIZATIONS:

1. Data Collection

```
1 df=pd.read_csv(r'car prediction.csv')
```

```
1 df.head()
```

	Unnamed: 0	BRAND	VARIANT	PRICE	DISTANCE	FUEL	How it WORK	CUSTOMER CARE
0	0	2018 Maruti Alto 800	LXI	₹3.15 Lakh	44,711 kms	Petrol	Manual	Trustmark
1	1	2016 Mahindra XUV500	W10 1.99 mHawk	₹9.10 Lakh₹ 10,68,500Save ₹1,58,500	1,08,939 kms	Diesel	Manual	Trustmark
2	2	2019 Maruti Ciaz	Alpha BSIV	₹9.34 Lakh₹ 9,40,000Save ₹6,000	33,206 kms	Petrol	Manual	Trustmark
3	3	2021 Nissan Kicks	1.5 XV	₹9.52 Lakh₹ 10,59,000Save ₹1,07,000	9,799 kms	Petrol	Manual	Trustmark
4	4	2019 Maruti Ciaz	Delta BSIV	₹8.17 Lakh₹ 8,26,000Save ₹9,000	34,474 kms	Petrol	Manual	Trustmark

2. Data Cleaning

DATA CLEANING

```
total=oversampled.isnull().sum().sort_values(ascending=False)
percent = (oversampled.isnull().sum()/oversampled.isnull().count()).sort_values(ascending=False)
missing = pd.concat([total, percent], axis=1, keys=['Total' , 'Percent'])
missing.head()
```

	Total	Percent
Unnamed: 0	0	0.0
BRAND	0	0.0
VARIANT	0	0.0
PRICE	0	0.0
DISTANCE	0	0.0

3. Univariate Analysis

```
1 from sklearn.preprocessing import StandardScaler
2 # Data Scaling Formula Z=(x-mean)/std
3 scaler = StandardScaler()
4 X_scaled=scaler.fit_transform(X)
5 X_scaled
```

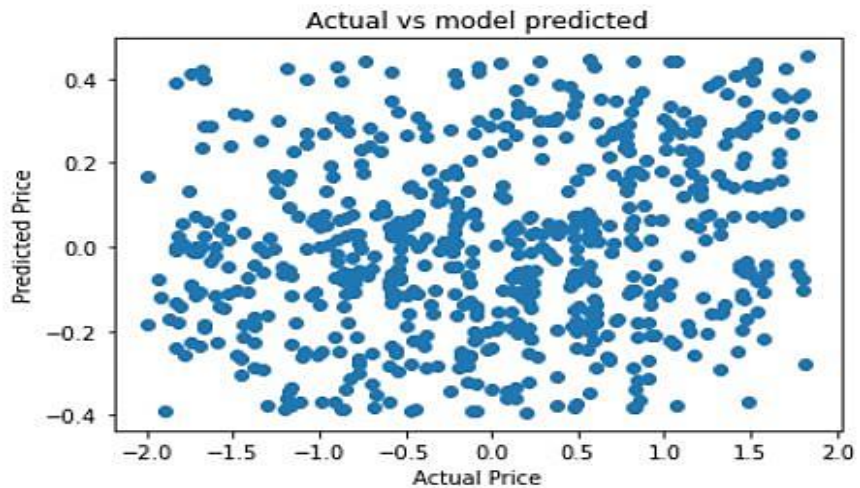
```
array([[-1.73154056,  0.55559882,  0.28924445, ...,  0.92497169,
         0.76791637,  2.36193515],
       [-1.7305199 , -0.13818739,  1.2339643 , ..., -1.019302  ,
         0.76791637,  2.36193515],
       [-1.72949925,  0.93441461, -0.37514309, ...,  0.92497169,
         0.76791637,  2.36193515],
       ...,
       [ 1.72949925,  0.25765383, -0.27142014, ..., -1.66739323,
         0.76791637, -0.29162589],
       [ 1.7305199 ,  0.66200775, -1.76839462, ...,  0.92497169,
         0.76791637, -0.29162589],
       [ 1.73154056,  0.48324075, -1.49366897, ..., -1.019302  ,
         0.76791637, -0.29162589]])
```

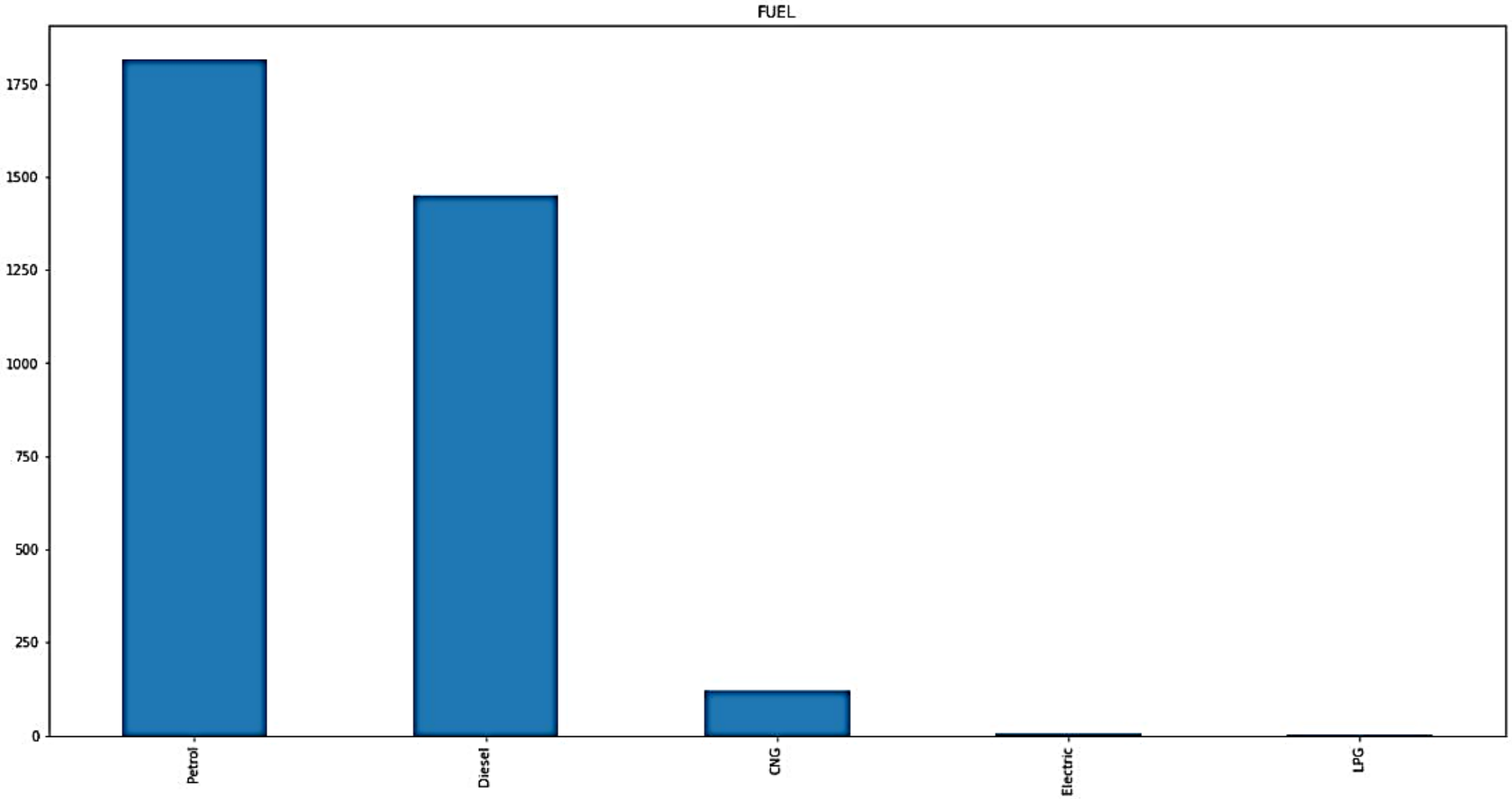
4. Bivariate Analysis

BIVARIATE ANALYSIS

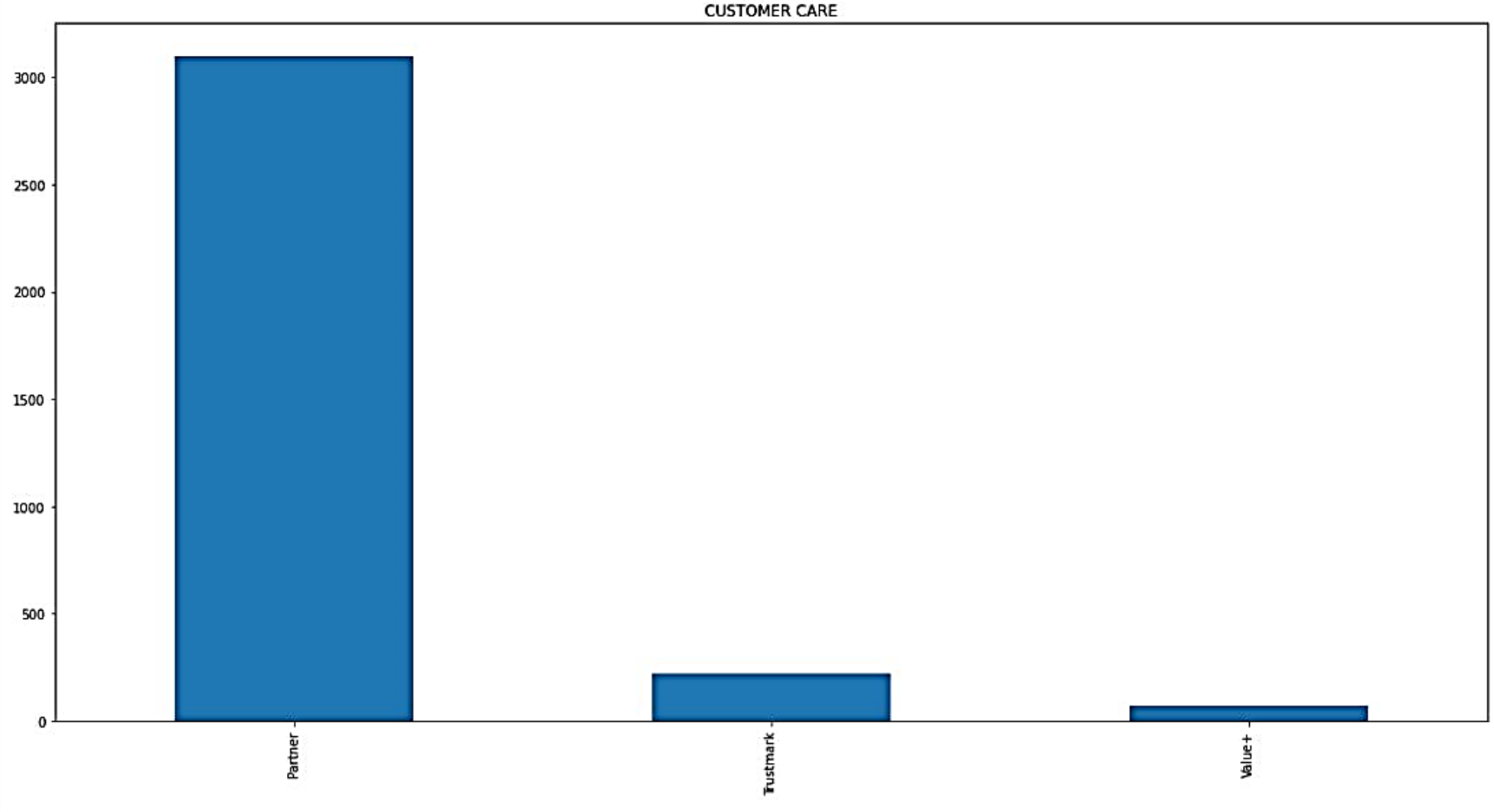
```
1  plt.scatter(y_test,y_pred)
2  plt.xlabel('Actual Price')
3  plt.ylabel('Predicted Price')
4  plt.title('Actual vs model predicted')
5  plt.show
```

```
<function matplotlib.pyplot.show(close=None, block=None)>
```

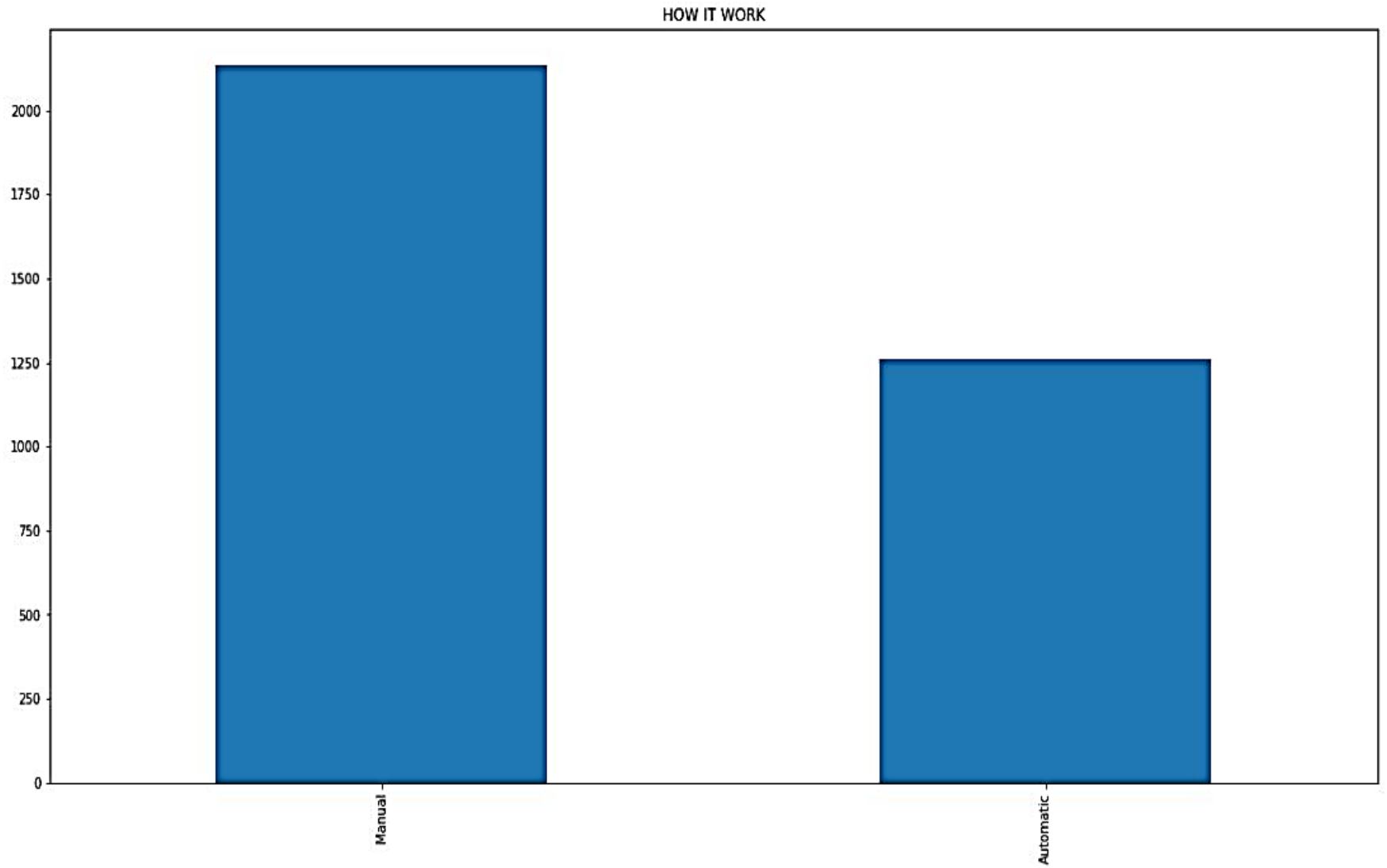




Petrol has highest accuracy among them of used cars



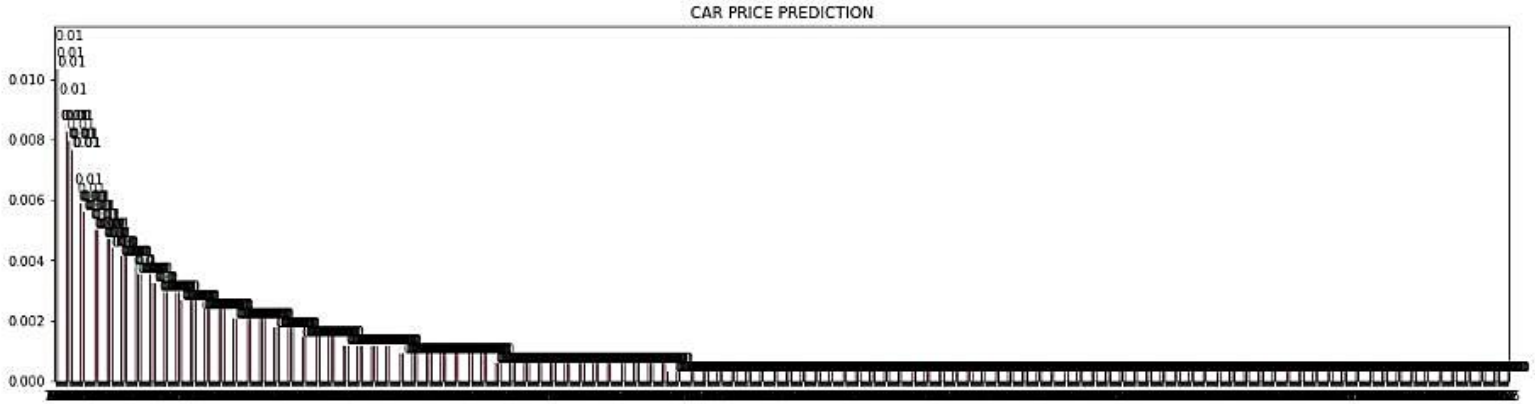
Most of the Used car has the customer care and among them Partner has the highest accuracy.

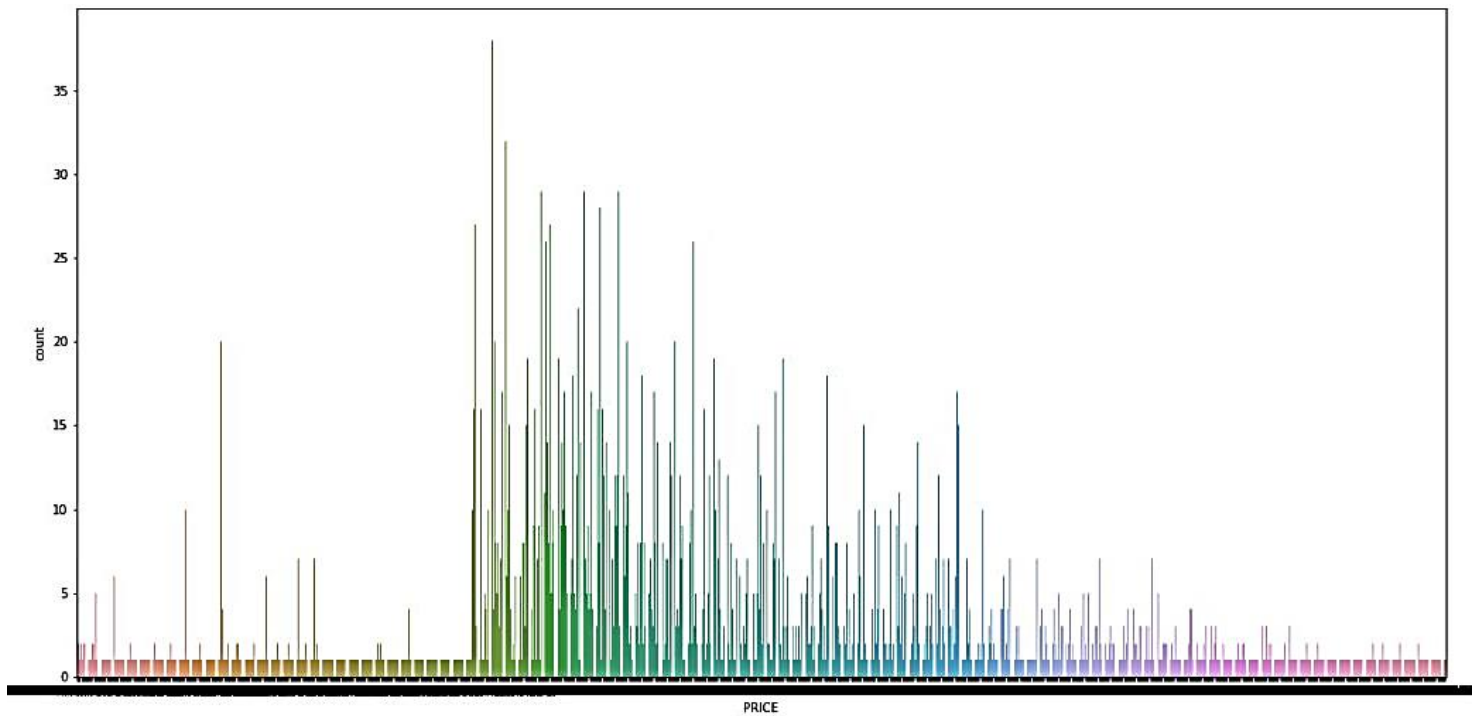


Most of the Used Cars are Manually work than Automatically work

```
plt.figure(figsize=(20,5))
ax=df.PRICE.value_counts(normalize=True).plot(kind='bar', color=['black', 'pink'], alpha=0.9, rot=0)
plt.title('CAR PRICE PREDICTION')
for i in ax.patches:
    ax.annotate(str(round(i.get_height(),2)),(i.get_x() * 1.01, i.get_height() * 1.01))

plt.show()
```





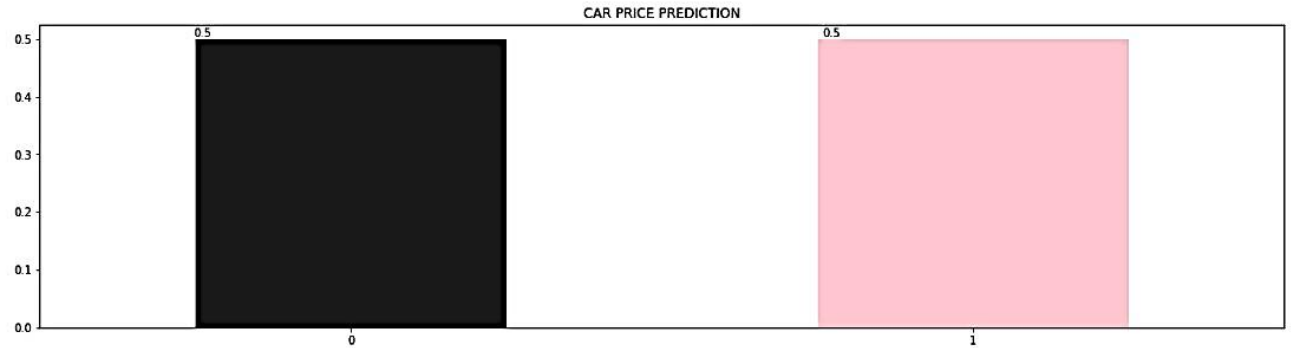
```
from sklearn.utils import resample

Car_price=df[df.PRICE == 0]
Prediction_price=df[df.PRICE == 1]
Prediction_price_oversampled=resample(Prediction_price, replace=True, n_samples=len(Car_price), random_state=42)
oversampled = pd.concat([Car_price, Prediction_price_oversampled])

plt.figure(figsize=(20,5))

ax=oversampled.PRICE.value_counts(normalize=True).plot(kind='bar', color=['black', 'pink'], alpha=0.9, rot=0)
plt.title('CAR PRICE PREDICTION')
for i in ax.patches:
    ax.annotate(str(round(i.get_height(),2)),(i.get_x() * 1.01, i.get_height() * 1.01))

plt.show()
```

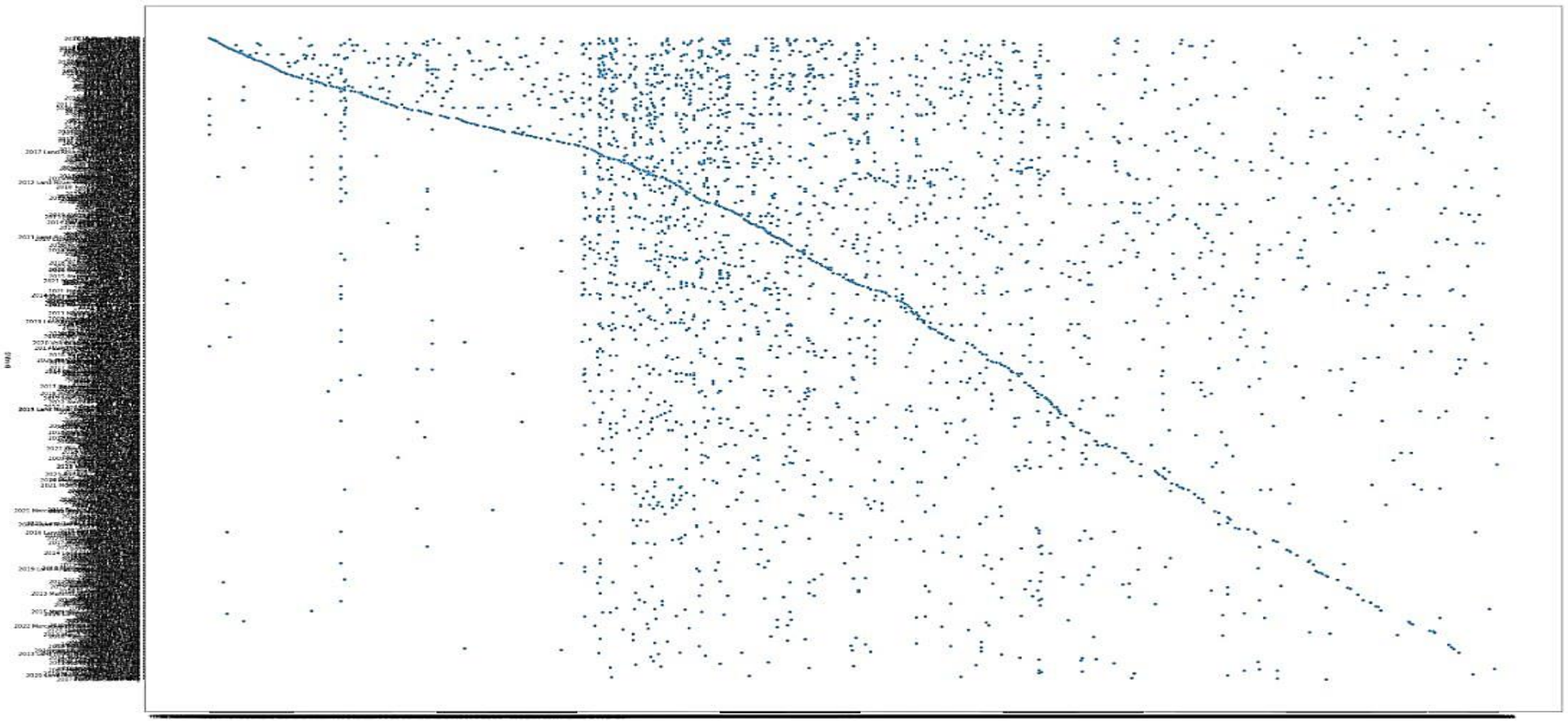



```
plt.figure(figsize=(20,5))
sns.heatmap(df.isnull(),cbar=False,cmap='crest')
```

<AxesSubplot:>

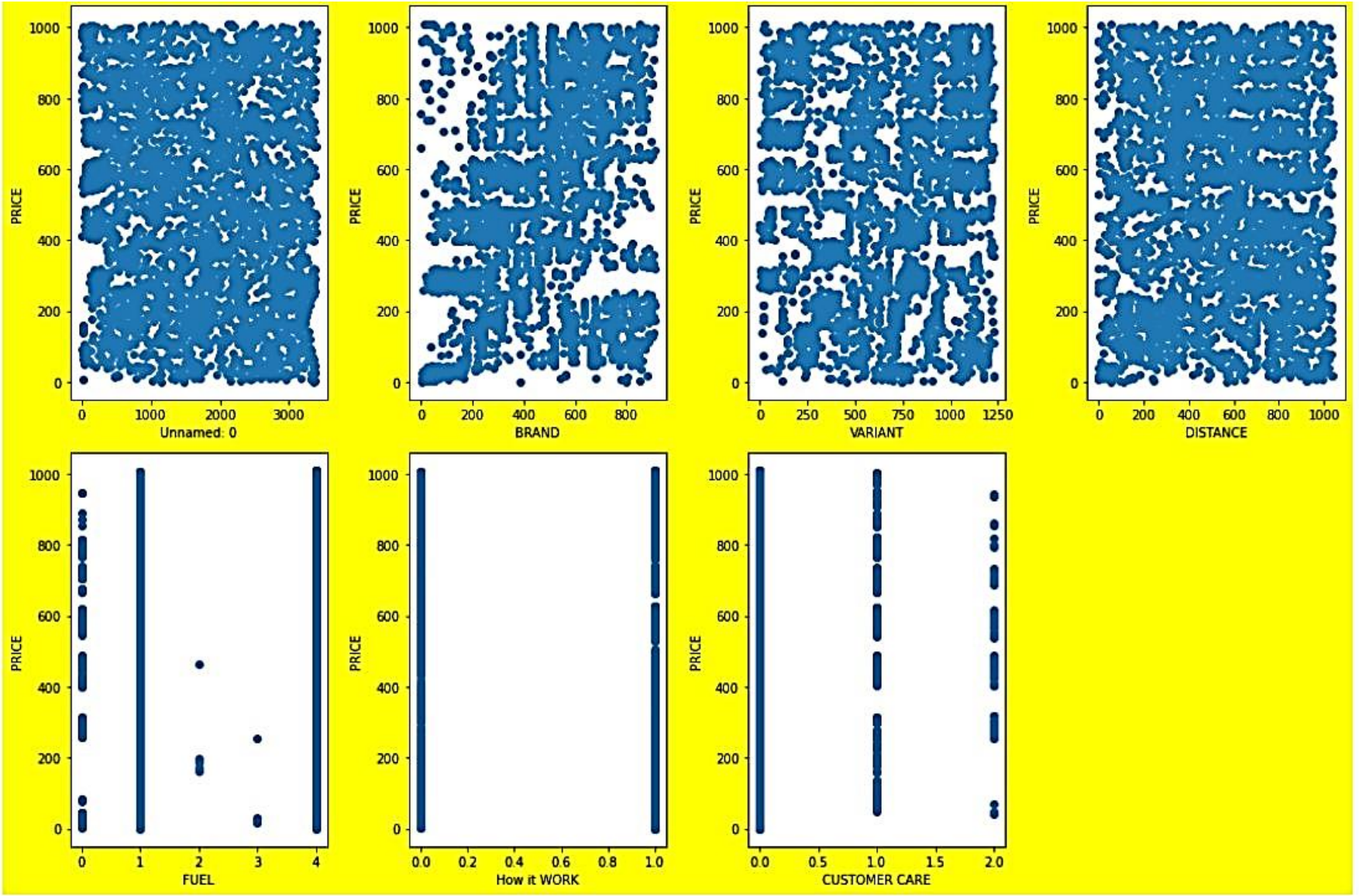
```
:  1  plt.figure(figsize=(40,28))
   2  sns.scatterplot(df['PRICE'],df['BRAND'])
   3  print(df[['PRICE','BRAND']].corr())
   4  plt.show()
```

```
Empty DataFrame
Columns: []
Index: []
```



In this scatterplot, plot between price and brand of the used cars

VISUALIZING THE RELATIONSHIP WITH PRICE



STEPS AND ASSUMPTIONS TO COMPLETE THE PROJECT:

1. The purpose of this case is to understand and evaluate used car prices and to develop a strategy that utilizes data mining techniques to predict used cars prices, to help guide the individuals looking to buy or sell of used cars to give them a better insight into the automotive sector.

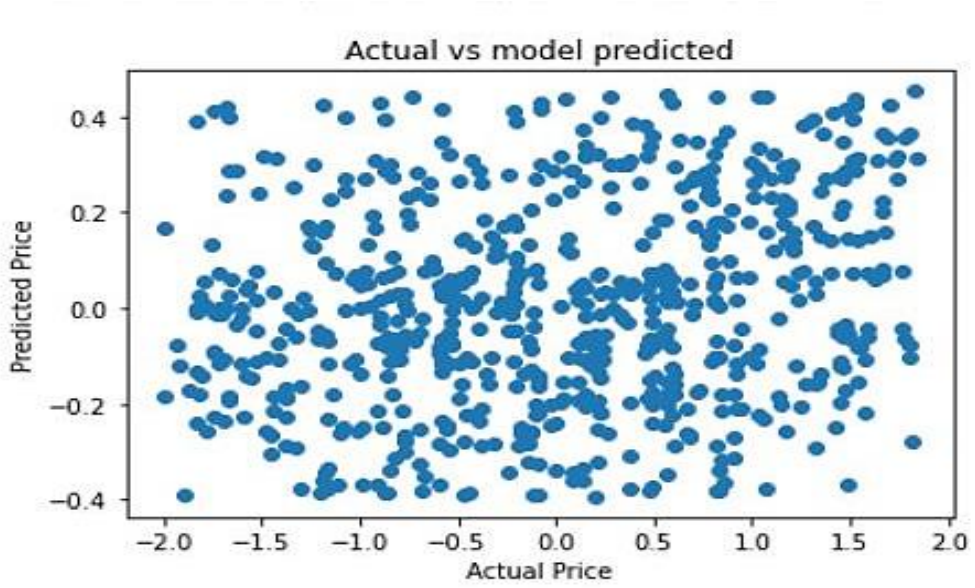
2. The model of the independent variables and dependent variables are exactly vary with the variables.

3. It can accordingly manipulating the strategy of the areas that will yield high returns as it make easier for the clients.

4. By visualizations there are many things to be noted when it will according to work each other it means that from which aspect it is going to work when between the brand of the car and the price of cars get compare and between the price of the car and which fuel is used in cars compare so, it will be predicted.

5. By preprocessing the data it means that from the help of label encoder helps the dataset column to transform to fit another column in to it.

MODEL DASHBOARD



It shows the effect of the model and compares it against the null model.

FINALIZED MODEL:

the accuracy score was 1%

```
====================Train Result=====================
Accuracy Score: 1.5157894736842104
_____________________________________________________
CLASSIFICATION REPORT:
                0    2    3    4    5    6    7    8    9   10  ...  1002  1003  \
precision     0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  ...   0.0   0.0
recall        0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  ...   0.0   0.0
f1-score      0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  ...   0.0   0.0
support       2.0  1.0  1.0  3.0  1.0  1.0  1.0  1.0  1.0  6.0  ...   7.0   1.0

              1004  1007  1008  1009  1010  accuracy     macro avg  weighted avg
precision      0.0   0.0   0.0   0.0   0.0  0.015158     0.000895      0.002710
recall         0.0   0.0   0.0   0.0   0.0  0.015158     0.002287      0.015158
f1-score       0.0   0.0   0.0   0.0   0.0  0.015158     0.000624      0.002263
support        1.0   1.0   2.0   2.0   2.0  0.015158  2375.000000   2375.000000

[4 rows x 828 columns]
_____________________________________________________
Cofusion Matrix:
 [[0 0 0 ... 0 0 0]
 [0 0 0 ... 0 0 0]
 [0 0 0 ... 0 0 0]
 ...
 [0 0 0 ... 0 0 0]
 [0 0 0 ... 0 0 0]
 [0 0 0 ... 0 0 0]]


==================Test Result=====================
Accuracy: 1.0794896957801767
_____________________________________________________
CLASSIFICATION REPORT:
                1    6   10   12   17   21   24   26   27   28  ...  994  995  \
precision     0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  ...  0.0  0.0
recall        0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  ...  0.0  0.0
f1-score      0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  ...  0.0  0.0
support       1.0  2.0  1.0  1.0  2.0  1.0  1.0  1.0  1.0  1.0  ...  1.0  2.0

              997  1002  1005  1006  1008  accuracy     macro avg  weighted avg
precision     0.0   0.0   0.0   0.0   0.0  0.010795     0.000163      0.000628
recall        0.0   0.0   0.0   0.0   0.0  0.010795     0.003306      0.010795
f1-score      0.0   0.0   0.0   0.0   0.0  0.010795     0.000274      0.001055
support       4.0   3.0   1.0   2.0   1.0  0.010795  1019.000000   1019.000000

[4 rows x 492 columns]
_____________________________________________________
Confusion Matrix:
 [[0 0 0 ... 0 0 0]
 [0 0 0 ... 0 0 0]
 [0 0 0 ... 0 0 0]
 ...
 [0 0 0 ... 0 0 0]
 [0 0 0 ... 0 0 0]
 [0 0 0 ... 0 0 0]]
```

CONCLUSION

I would like to conclude here that doing research in the topic found that in the market, some of the cars are in demand and the clients are ready to buy or sell their used cars in the market through some websites. By using some parameters to predict the data according to the statement. There are many things to be noted when it will according to work each other like pre-processing the data, cleaning, visualizing, scaling and label encoder helps that dataset column to transform to fit another column into it. The model of the project are ready to analyse the independent and dependent variable.

THANK YOU