ANALYSIS OF USED CARS FOR PRICE PREDICTION

GROUP 2

BY
AKANKSHA VASUDEVA HANUMAPALLY
A SRINIDHI
A YASASWINI
HARIHARAN REDDY
B.SC. HDS

Bhavan's Vivekananda College
Sainikpuri
Autonomous College

CONTENTS

- INTRODUCTION
- DATA CLEANING
- EXPLORATORY DATA ANALYSIS
- MODEL BUILDING
- MODEL EVALUATION
- CONCLUSION
- RECOMMENDATIONS

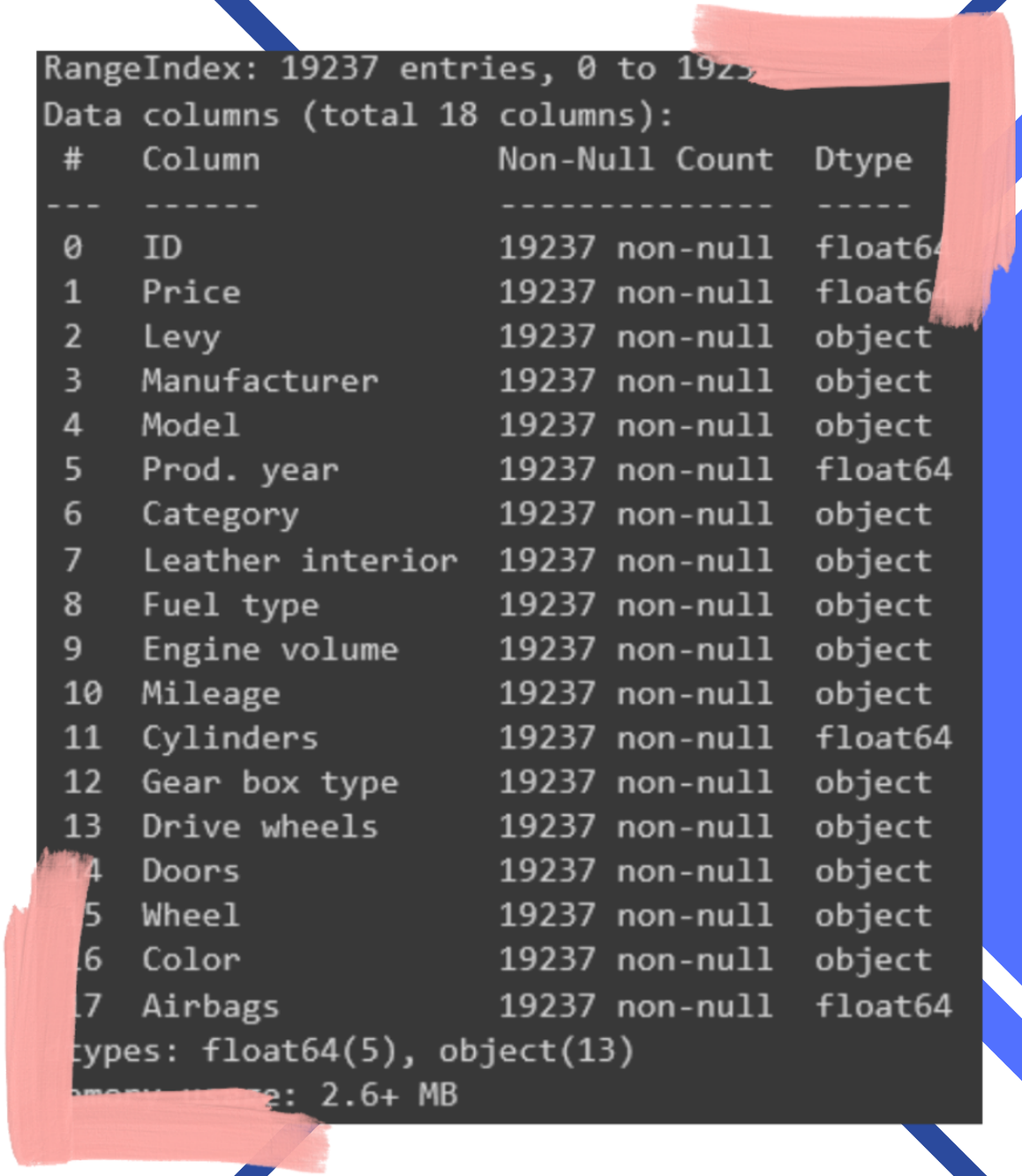
A BRIEF GIST

AIM:
TO DETERMINE THE BEST FIT MODEL FOR USED CAR'S PRICE PREDICTION

PATH TAKEN: THE DATA WAS FIRST CLEANED AND THEN VARIOUS REGRESSION AND ENSEMBLE MODELS WERE APPLIED TO IT

DATA INFORMATION

- NO OF ROWS: 19237
- NO OF COLUMNS: 18
- THERE WERE NO NULL VALUES IN THE DATASET
- NO OF CATEGORICAL FEATURES:13
- NO OF NUMERIC FEATURES:5



```
RangeIndex: 19237 entries, 0 to 192
Data columns (total 18 columns):
 #   Column             Non-Null Count  Dtype
---  ------             --------------  -----
 0   ID                 19237 non-null  float64
 1   Price              19237 non-null  float6
 2   Levy               19237 non-null  object
 3   Manufacturer       19237 non-null  object
 4   Model              19237 non-null  object
 5   Prod. year         19237 non-null  float64
 6   Category           19237 non-null  object
 7   Leather interior   19237 non-null  object
 8   Fuel type          19237 non-null  object
 9   Engine volume      19237 non-null  object
 10  Mileage            19237 non-null  object
 11  Cylinders          19237 non-null  float64
 12  Gear box type      19237 non-null  object
 13  Drive wheels       19237 non-null  object
 14  Doors              19237 non-null  object
 15  Wheel              19237 non-null  object
 16  Color              19237 non-null  object
 17  Airbags            19237 non-null  float64
types: float64(5), object(13)
e: 2.6+ MB
```

ATTRIBUTE INFORMATION

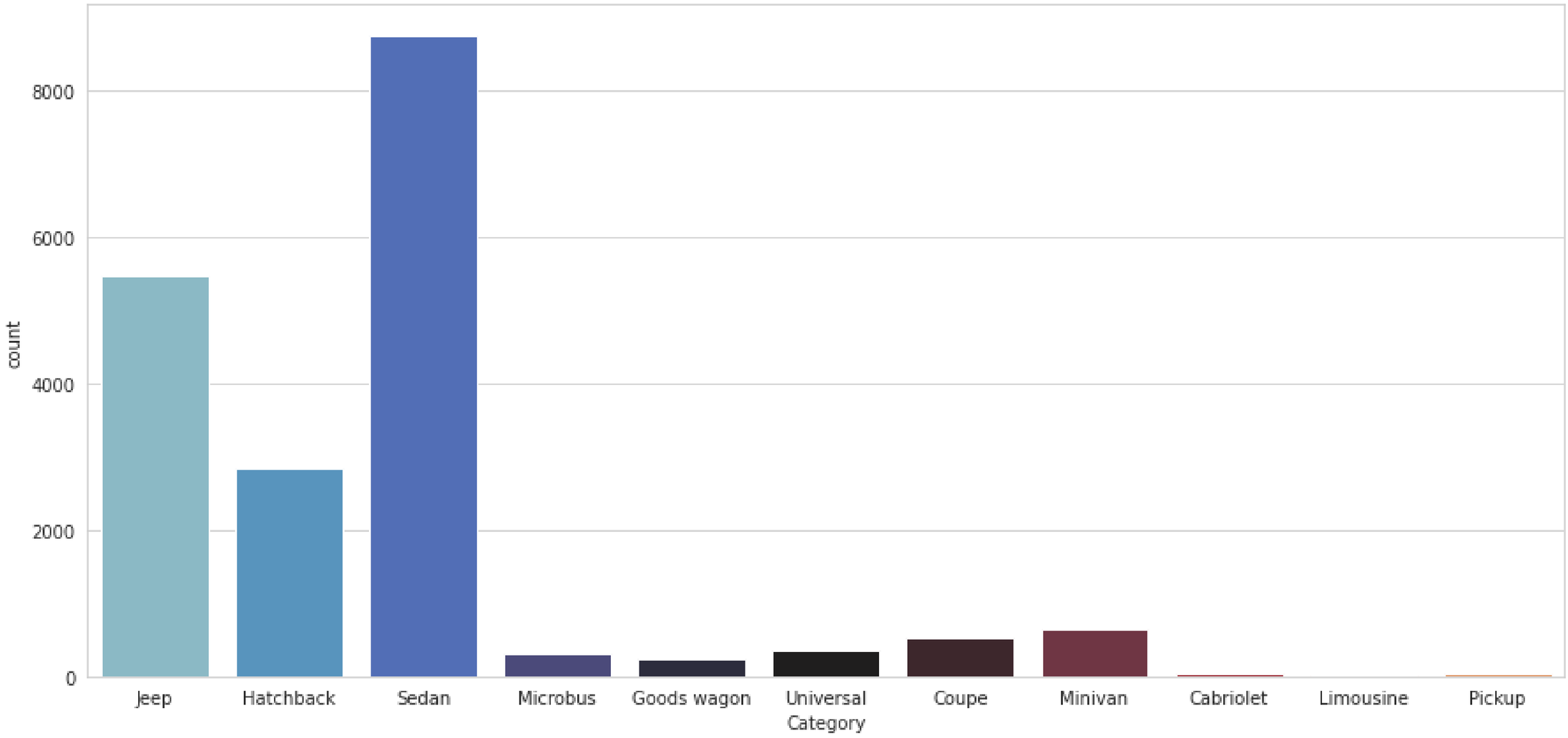
NUMERIC FEATURES	CATEGORICAL FEATURES	
PRICE [TARGET VARIABLE] ID PRODUCT YEAR CYLINDERS AIRBAGS	MANUFACTURER MODEL LEVY CATEGORY LEATHER INTERIOR FUEL TYPE ENGINE VOLUME	MILEAGE GEARBOX TYPE DRIVE WHEELS DOORS WHEEL COLOR

DATA PREPROCESSING AND CLEANING

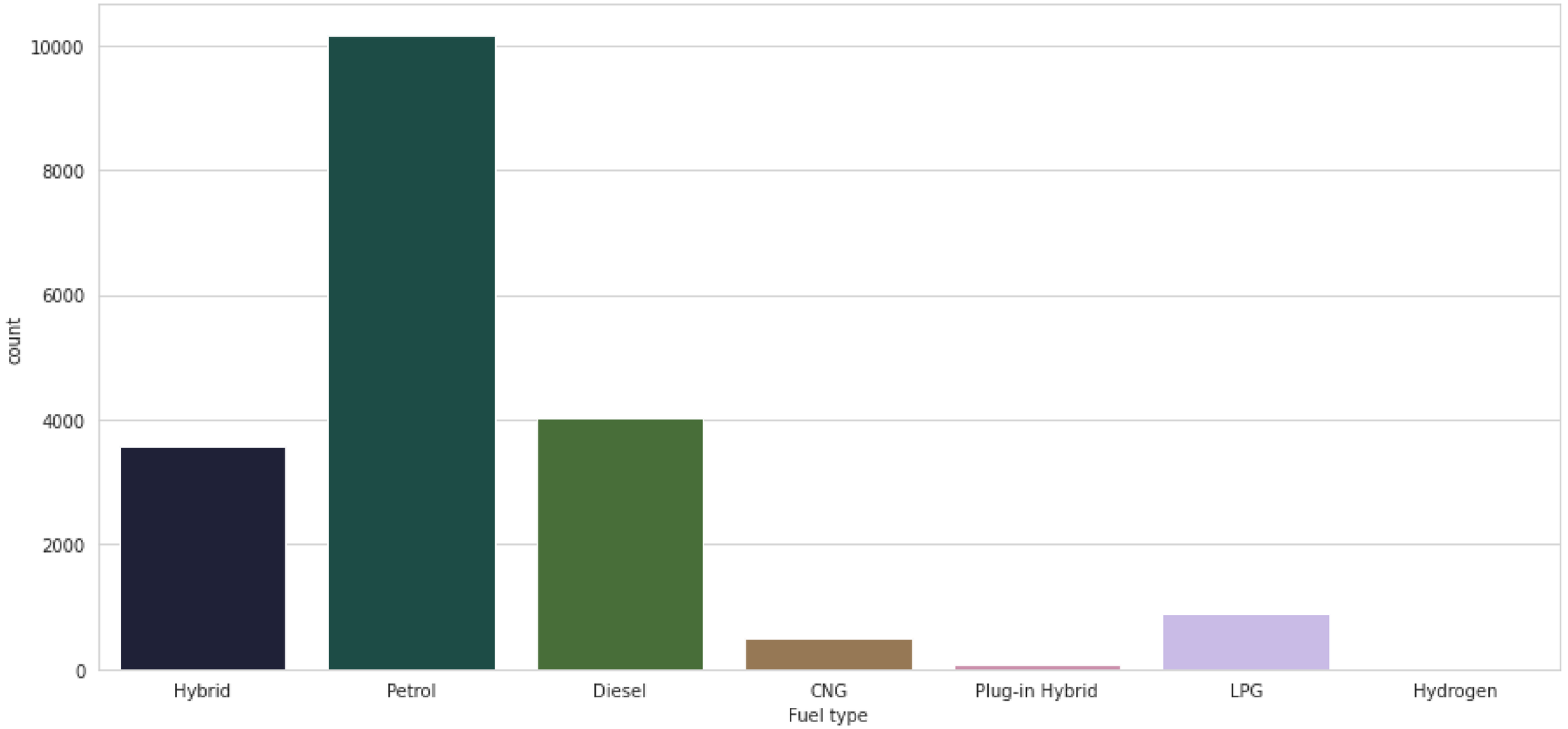
- OUTLIERS WERE REMOVED IN THE COLUMNS : **PRICE, MILEAGE, ENGINE VOLUME**
- THE COLUMNS LEATHER INTERIOR, WHEEL, DRIVE WHEELS, GEARBOX TYPE, FUEL TYPE, CATEGORY, COLOR, DOORS WERE DUMMIFIED

EXPLORATORY DATA ANALYSIS

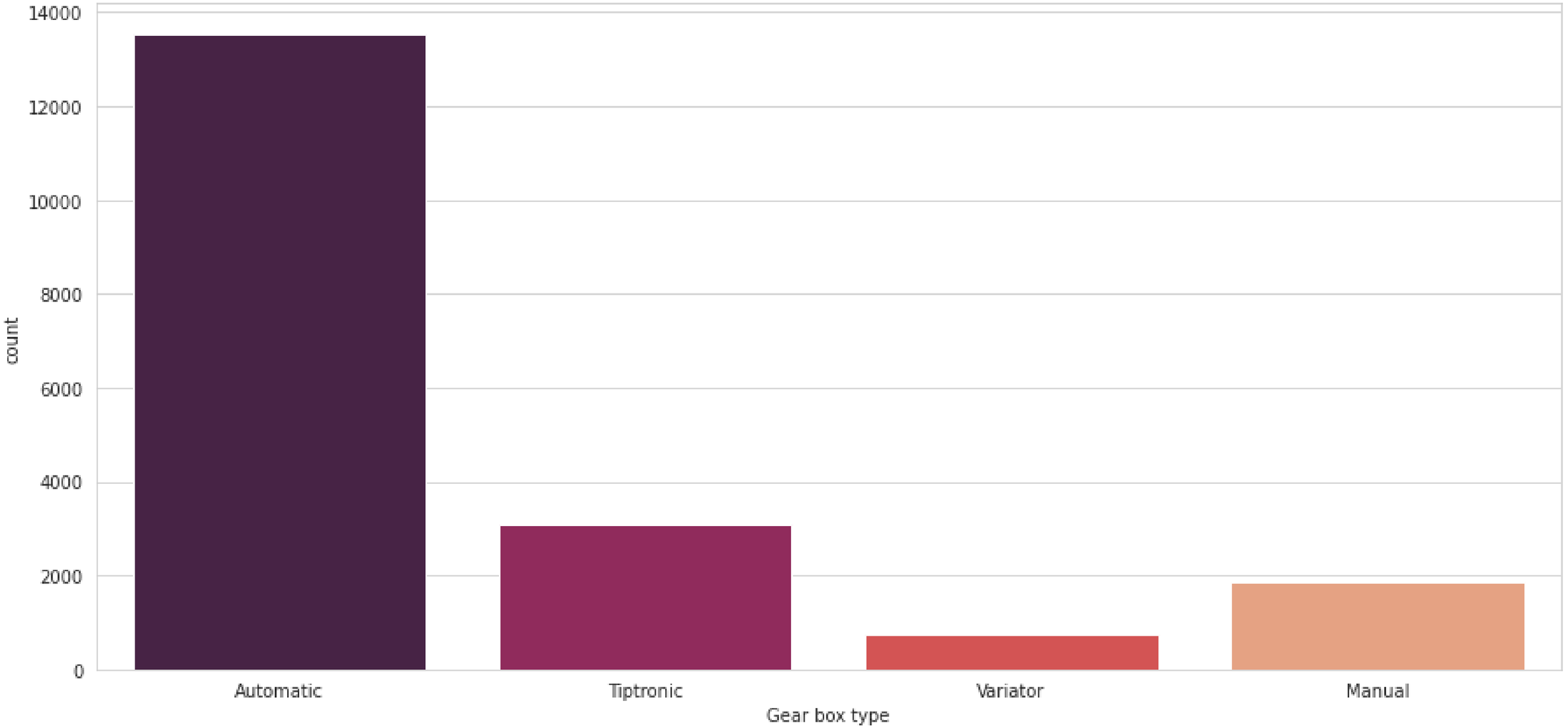
CATEGORY OF CARS



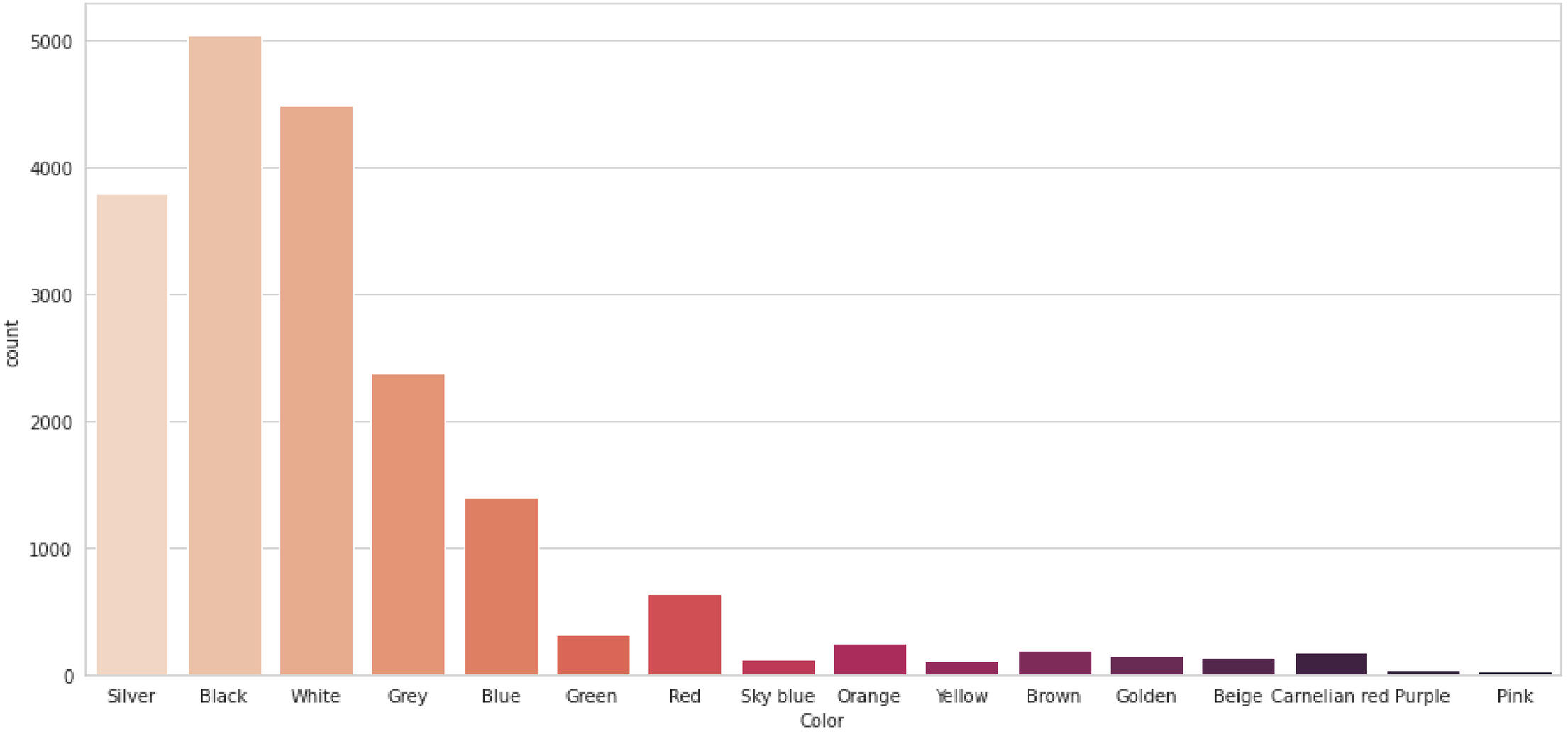
FUEL TYPE


10000
8000
6000
4000
2000
0
count
Hybrid
Petrol
Diesel
CNG
Plug-in Hybrid
LPG
Hydrogen
Fuel type

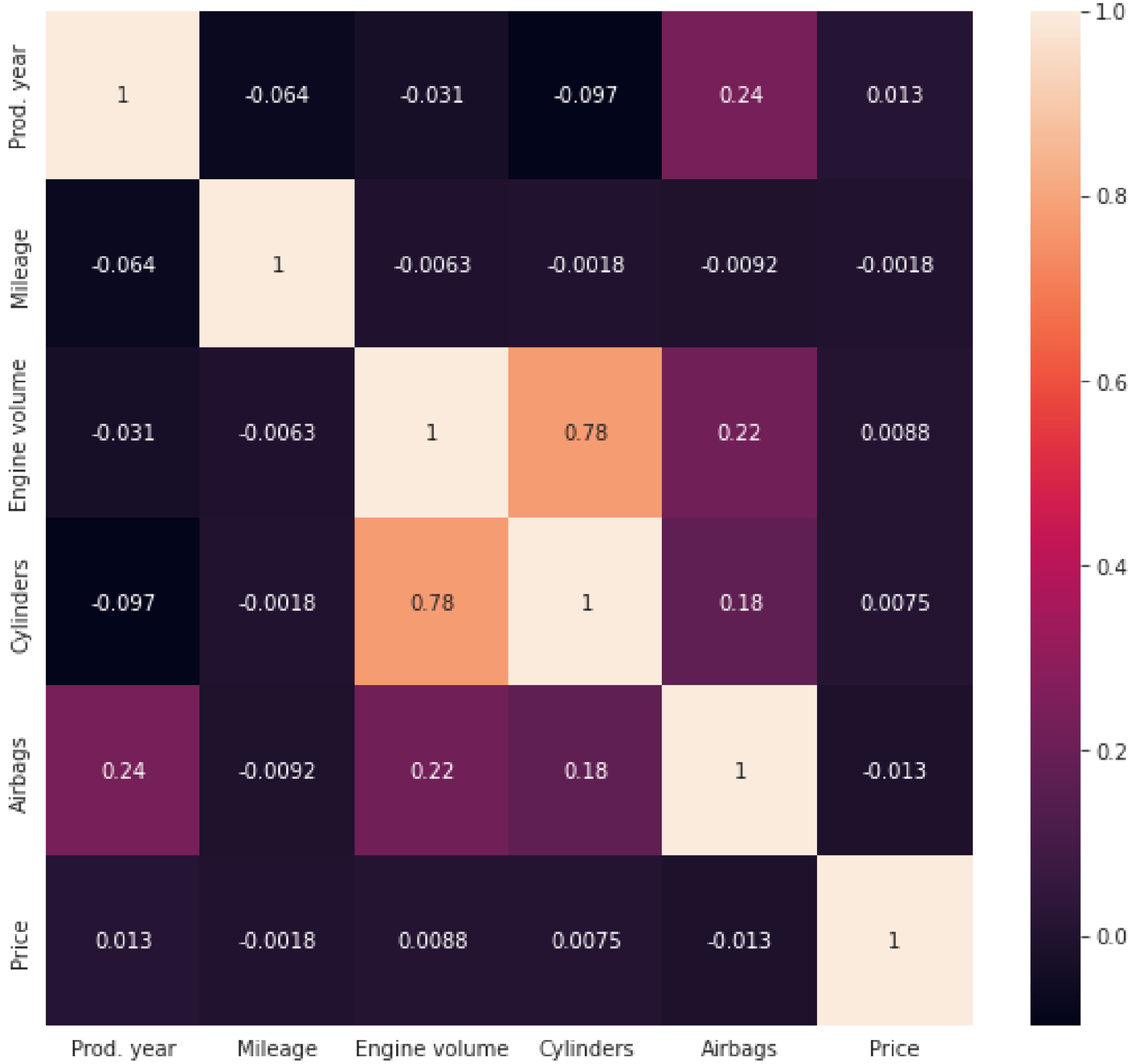
GEAR BOX TYPE



COLOR OF CARS



HEATMAP OF NUMERICAL FEATURES



MODEL BUILDING

- NUMERICAL FEATURES WERE SCALED USING MinMaxScaler()
- Recursive Feature Elemination was used

LINEAR REGRESSION

Train:Test	MAE
70:30	0.188
80:20	**0.186**

BAGGING

Train:Test	MAE
70:30	0.0865
80:20	**0.0844**

BOOSTING

ADABOOST

Train:Test	MAE
70:30	0.184
80:20	0.183

GRADIENT BOOST

Train:Test	MAE
70:30	0.097
80:20	0.096

XG BOOST

Train:Test	MAE
70:30	0.0979
80:20	0.0967

OTHER ALGORITHMS

KNN

Train:Test	MAE
70:30	0.129
80:20	**0.125**

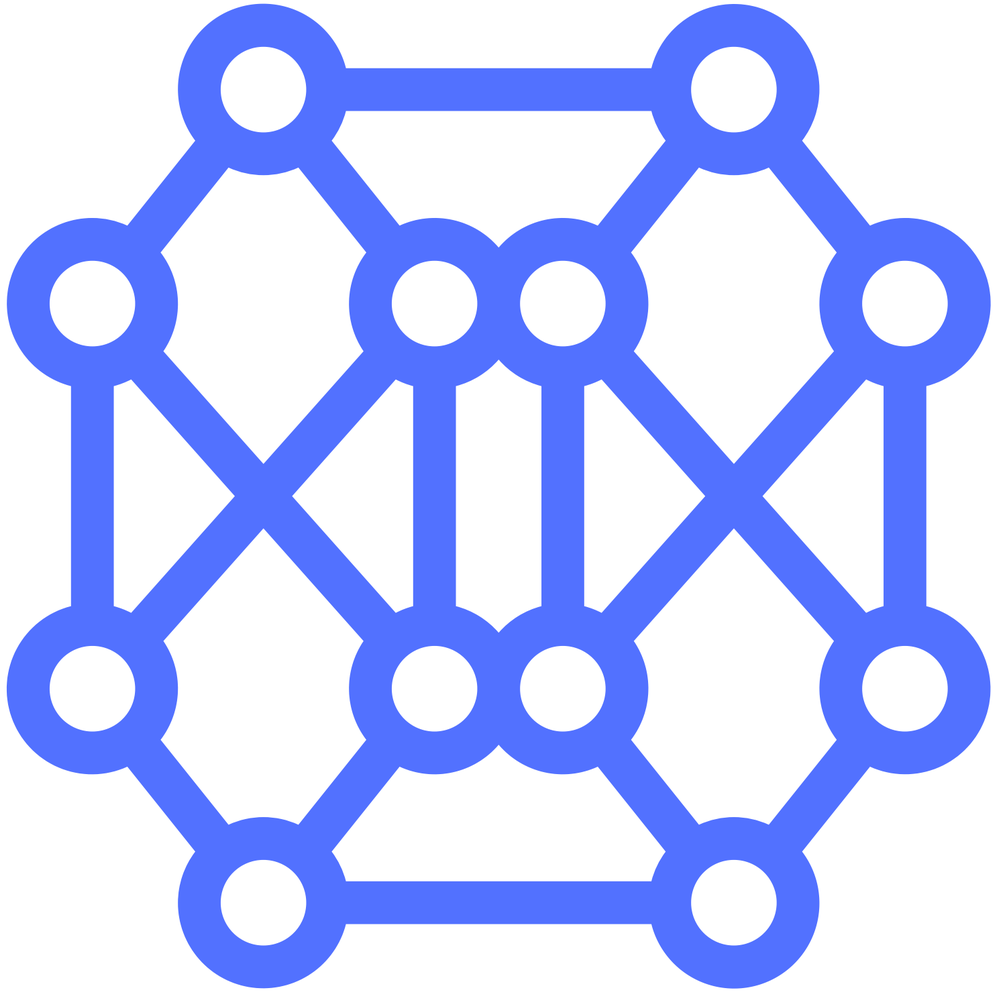
DECISION TREE REGRESSOR

Train:Test	MAE
70:30	0.153
80:20	**0.152**

RANDOM FOREST REGRESSOR

Train:Test	MAE
70:30	0.0905
80:20	**0.0875**

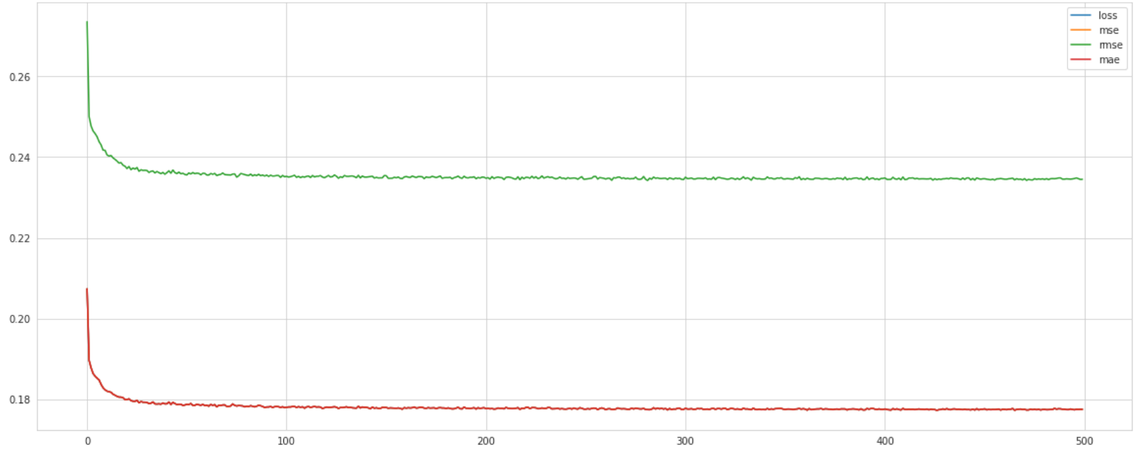
NEURAL NETWORKS



70:30

BEST PERFORMER

ARCHITECTURE	EPOCHS	MAE
24-12-6-3-1	1500	0.1769

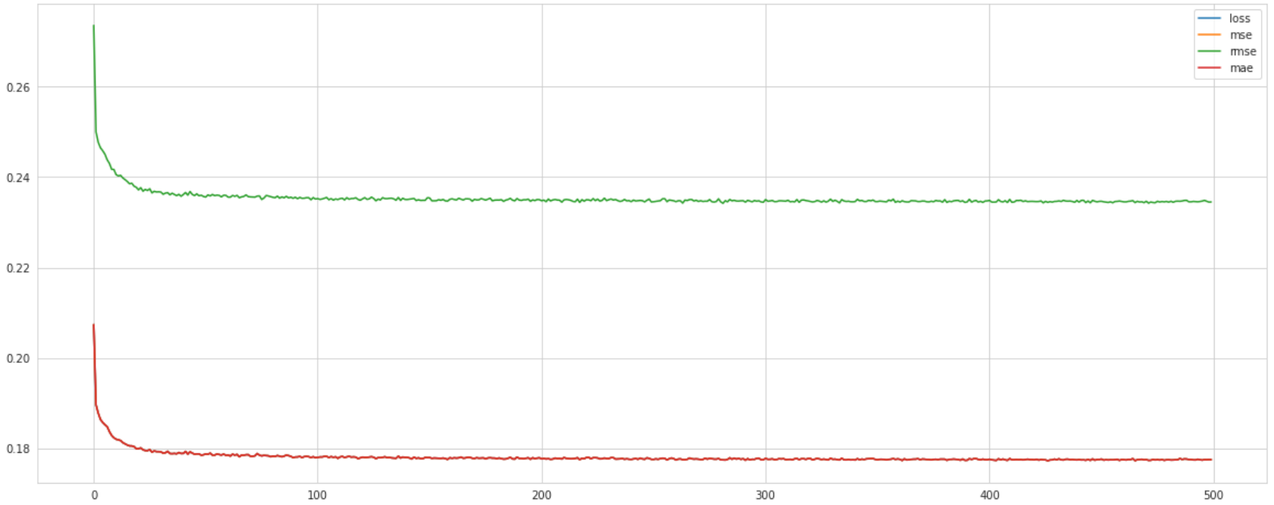


Optimizer	Architectures	Epochs	MAE	RMSE	MSE
Adam	24,12,6,3,1	1500	0.1769	0.2337	0.1769
Adam	24,12,6,3,1	500	0.1769	0.2338	0.1769
Adam	14,2,2,1	500	0.177	0.2339	0.177
Adam	14,2,2,1	300	0.177	0.2337	0.177
Adam	13,3,3,1	400	0.1771	0.2337	0.1771
Adam	14,2,2,1	400	0.1771	0.2336	0.1771
Adam	24,12,6,3,1	200	0.1774	0.2352	0.1774
Adam	16,2,1	200	0.1774	0.2356	0.1774
Adam	14,2,2,1	200	0.1774	0.2342	0.1774
Adam	10,5,1	500	0.1775	0.235	0.1775
Adam	16,2,1	500	0.1775	0.2348	0.1775
Adam	13,3,3,1	200	0.1775	0.2343	0.1775
Adam	13,3,3,1	500	0.1775	0.2354	0.1775
Adam	10,5,1	100	0.1777	0.2359	0.1777
Adam	14,2,2,1	100	0.1778	0.2348	0.1778
Adam	16,2,1	100	0.1779	0.2366	0.1779
Adam	24,12,6,3,1	100	0.178	0.2351	0.178
Adam	13,3,3,1	100	0.178	0.2349	0.178
Adam	10,5,1	200	0.1774		

80:20

BEST PERFORMER

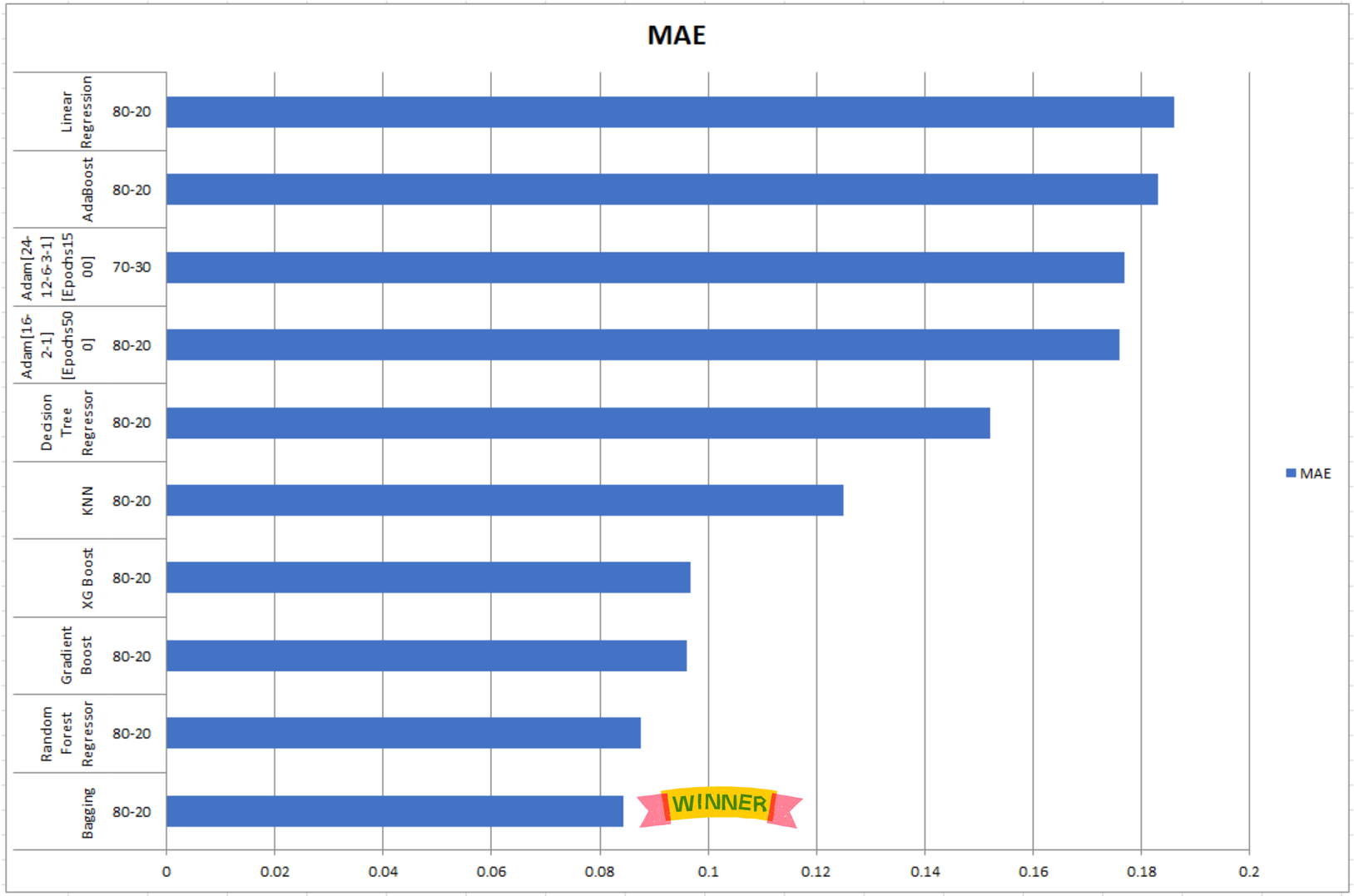
ARCHITECTURE	EPOCHS	MAE
16-2-1	500	0.1759



Optimizer	Architectures	Epochs	MAE	RMSE	MSE
Adam	16,2,1	500	0.1759	0.232	0.1759
Adam	14,2,2,1	500	0.1759	0.2322	0.1759
Adam	10,5,1	500	0.176	0.2316	0.176
Adam	13,3,3,1	500	0.176	0.2317	0.176
Adam	24,12,6,3,1	1500	0.1771	0.2342	0.1771
Adam	24,12,6,3,1	700	0.1773	0.2344	0.1773
Adam	24,12,6,3,1	1000	0.1773	0.2346	0.1773
Adam	16,2,1	300	0.1774	0.2344	0.1774
Adam	16,2,1	400	0.1774	0.2347	0.1774
Adam	10,5,1	300	0.1775	0.2344	0.1775
Adam	10,5,1	400	0.1775	0.2347	0.1775
Adam	24,12,6,3,1	300	0.1775	0.2345	0.1775
Adam	24,12,6,3,1	400	0.1775	0.2348	0.1775
Adam	24,12,6,3,1	500	0.1775	0.2346	0.1775
Adam	24,12,6,3,1	600	0.1775	0.2346	0.1775
Adam	13,3,3,1	300	0.1775	0.2345	0.1775
Adam	13,3,3,1	400	0.1775	0.2348	0.1775
Adam	14,2,2,1	300	0.1776	0.2346	0.1776
Adam	14,2,2,1	400	0.1776	0.235	0.1776
Adam	10,5,1	200	0.1778	0.235	0.1778
Adam	24,12,6,3,1	200	0.1779	0.2351	0.1779
Adam	16,2,1	100	0.1779	0.2351	0.1779
Adam	16,2,1	200	0.1779	0.235	0.1779
Adam	13,3,3,1	200	0.1779	0.235	0.1779
Adam	10,5,1	100	0.178	0.2353	0.178
Adam	13,3,3,1	100	0.178	0.2352	0.178
Adam	24,12,6,3,1	100	0.1781	0.2354	0.1781
Adam	14,2,2,1	100	0.1781	0.2355	0.1781
Adam	14,2,2,1	200	0.1781	0.2352	0.1781

MODEL EVALUATION

Algorithm	Ratio	MAE
Bagging	80-20	0.0844
Random Forest Regressor	80-20	0.0875
Gradient Boost	80-20	0.096
XG Boost	80-20	0.0967
KNN	80-20	0.125
Decision Tree Regressor	80-20	0.152
Adam[16-2-1] [Epochs500]	80-20	0.1759
Adam[24-12-6-3-1] [Epochs1500]	70-30	0.1769
AdaBoost	80-20	0.183
Linear Regression	80-20	0.186



TOP 3 PERFORMERS

BAGGING

MAE=
0.0844

RANDOM FOREST REGRESSOR

MAE=
0.0875

GRADIENT BOOST

MAE=
0.096

CONCLUSION

- *Bagging fits the data the best, followed by Random Forest Regressor and GradientBoost*
- *Train:Test Ratio is 80:20 gave us less MAE values*
- *For Neural Networks, an Epochs value of 500, gave us the best results. There was hardly any change in MAE for values beyond it.*
- *The Best Architecture for Neural Networks was Optimizer Adam, Architecture 16-2-1 and Epochs 500*

THANK YOU!

GitHub

CLICK THE LOGOS TO ACCESS THE CODE