



USED CAR PRICE PREDICTION PROJECT

Submitted by:

Roshan Kumar Verma

ACKNOWLEDGMENT

This Project would not have been possible without the kind support and help of many individuals and organizations. I would like to extend my sincere thanks to all of them.

I would like to thank Flip-Robo Technologies Bangalore for their guidance and Constant supervision as well as for providing necessary information regarding The project & also for their support in completing the project.

I would like to express my gratitude towards my parents & my SME of Flip-Robo Mohd. Kashif for their kind co-operation and encouragement which help Me in completion of this project.

My thanks and appreciations also go to my colleague in developing the project And people who have willingly helped me out with their abilities.

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 growth 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, changed 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 growth, 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.

Analytical Problem Framing

Mathematical/ Analytical Modelling of the Problem

On importing the data that is collected I have done some initial play around to understand the data and to cleanse the data.

Data Cleansing.

On data cleansing I have detecting duplicate records in the data collected and the null values in the data.

```
In [33]: 1 df.duplicated().sum()
```

```
Out[33]: 3387
```

As like we can see I had 3387 duplicate records out of 6239, I decided to drop all the duplicate records because it will not help us in creating a perfect model.

Removing the Duplicate Records.

```
In [12]: df = df.drop_duplicates() #Because we have too many duplicate records we will detete them to have a good model
```

After dropping the duplicate records, I checked for null values in the data.

```
In [6]: 1 df.isnull().sum()
```

```
Out[6]: MNF_YEAR      0  
        BRAND        0  
        MODEL        0  
        VARIANT      252  
        DRIVEN_KM     0  
        FUELTYPE      0  
        NOOF_OWNERS   0  
        PRICE         0  
        dtype: int64
```

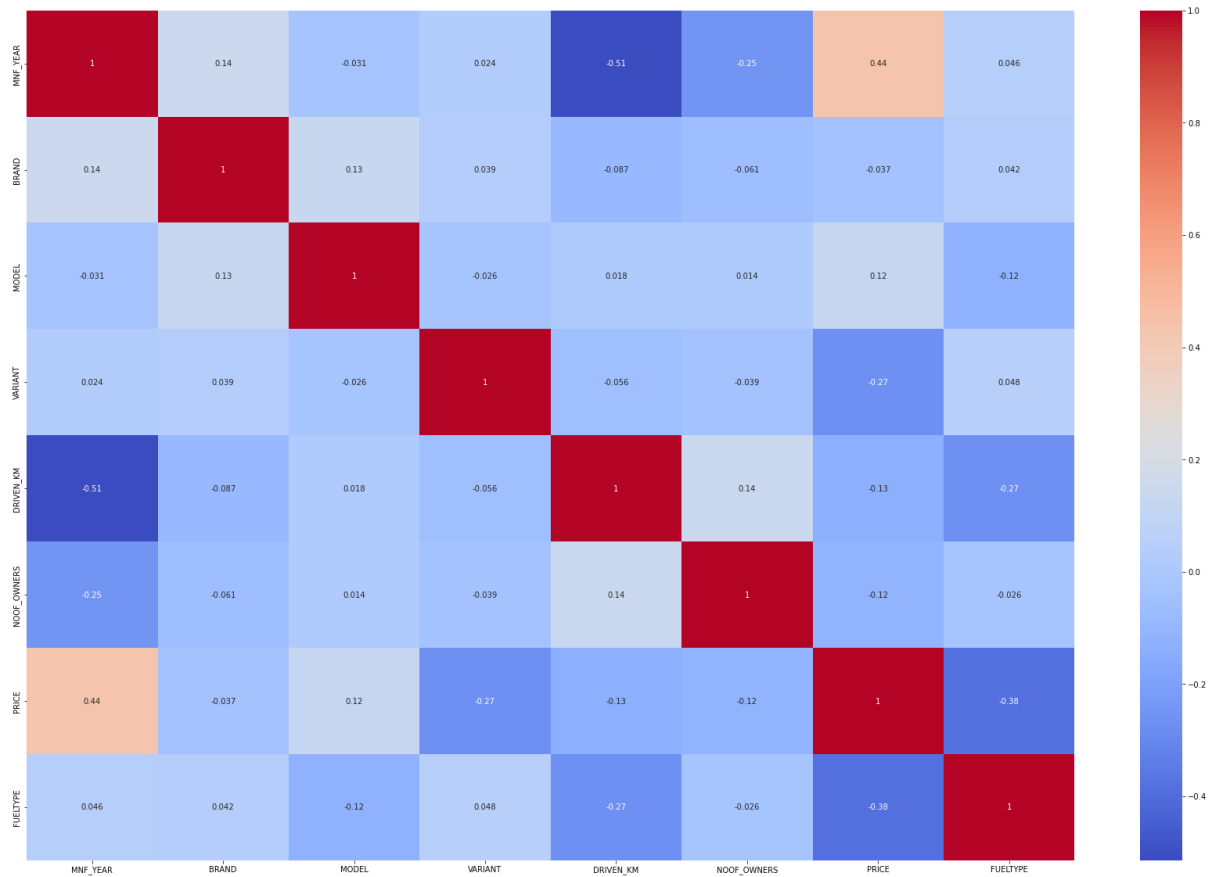
We can see that there are many null values in "Variant" Column

We have 252 null values in Variant column, I have used those null values in the model building but I have changed null values as not mentioned, this will also help the client to predict the values on the used cars without the Variant values.

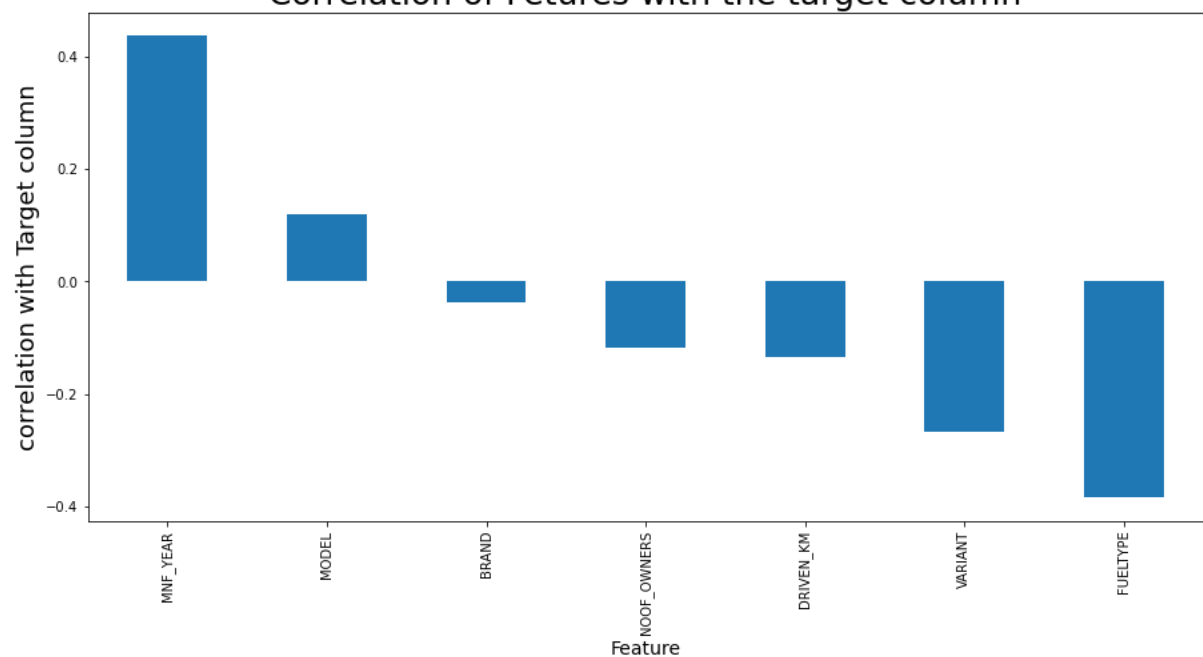
Data correlations

To better understanding the mathematical concept of the problem we have

to see the correlation of the data.



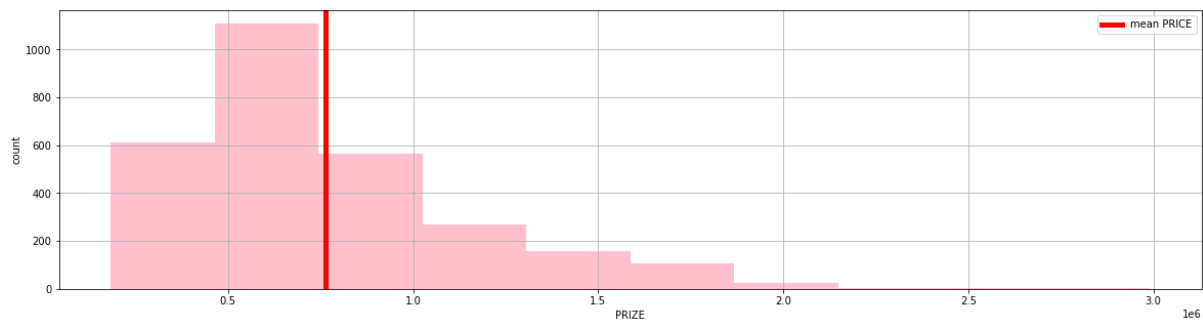
Correlation of Fetures with the target column



Key Observations:

- From above we can clearly see that MNF_YEAR is positively correlated to PRICE and FUEL_TYPE and VARIENT is negative correlated to PRICE.

Univariate Analysis

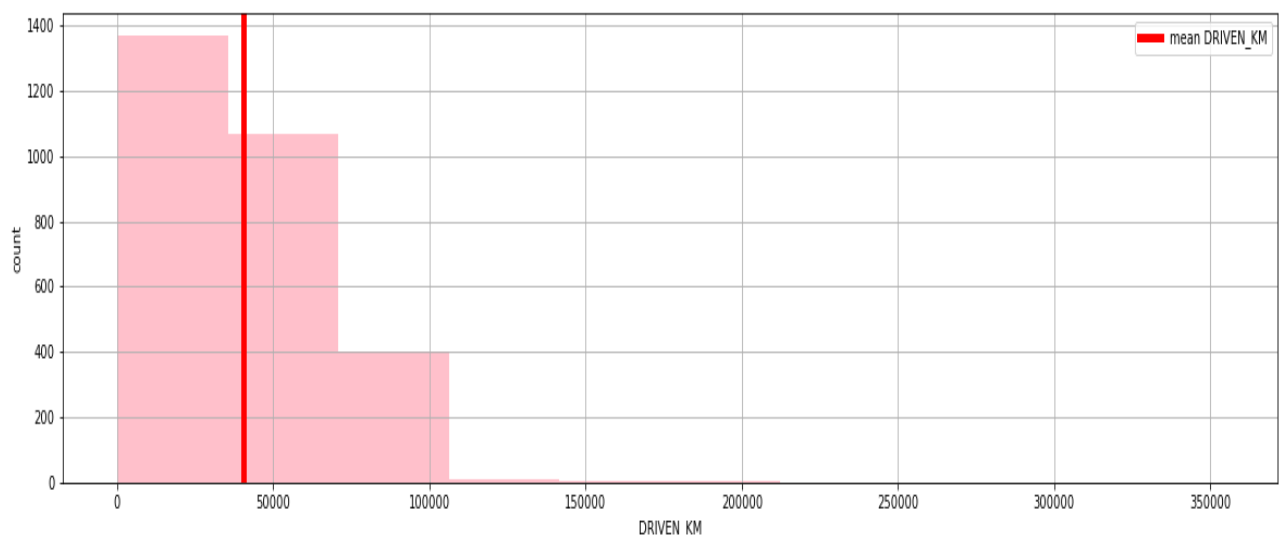


MATHEMATICAL SUMMARY OF PRICE:

```
count      2.852000e+03
mean       7.628008e+05
std        3.830175e+05
min        1.836990e+05
25%        4.907490e+05
50%        6.611490e+05
75%        9.404740e+05
max        2.989899e+06
Name: PRICE, dtype: float64
```

Key observations:

- Mean of the prize is Rs: 5,33,047, the prize is distributed between Rs: 99,162 to Rs: 41,00,000.
- Above we can understand that most of the Car price is lesser than the Mean i.e., Rs: 5,33,047




```
MATHEMATICAL SUMMARY OF PRIZE :
count      2852.00000
mean       40604.47791
std        26853.60534
min        188.00000
25%        19100.75000
50%        37430.00000
75%        57734.75000
max        353693.00000
Name: DRIVEN_KM, dtype: float64
```

Key observations:

- Mean of the DRIVEN_KM is 64489.27888kms and the maximum KMS driven is 312882kms.
- Above we can understand that most of the Car comes to selling around low kilometres driven.

Data Sources and their formats

The Data is scrubbed on multiple ecommerce websites that sells used cars in India, Websites like cars 24, OLX and car Dekho. These data are scrubbed and stored in a CSV format. Data contains following columns.

1. 'MNF_YEAR' – At what year the car is manufactured
2. 'BRAND' – Brand is manufacturer or which company made
3. 'MODEL' – It is basically the model of the car.
4. 'VARIANT' – Gear shift variant is (Automatic, Manual, Semi-Automatic)
5. 'DRIVEN_KM' – no of Kms driven before selling
6. 'FUELTYPE' – Petrol, diesel, CNG, LPG, Electric
7. 'NOOF_OWNERS' – 1end, 2end or 3end car
8. 'PRICE' – our target variable that tells what is the price of the used car.

Data Pre-processing Done

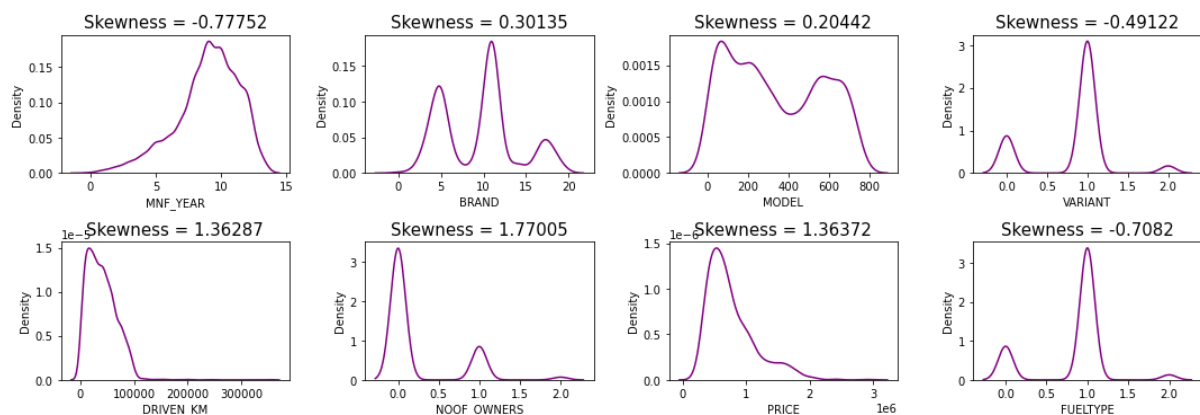
On pre-processing the data, I have tried in finding out the skewness of the data and the outliers, have changed the data into numbers with label encoders.

```
In [40]: 1 categorical_data = [ 'MNF_YEAR', 'BRAND', 'MODEL', 'VARIANT', 'FUELTYPE' , 'NOOF_OWNERS' ]
          2
          3 for i in categorical_data:
          4     DF[i]=DF[i].astype('str')
          5     DF[i]=LE.fit_transform(DF[i])
```

```
In [41]: 1 DF['PRICE'] = DF['PRICE'].str.replace(r'\D', '').astype(int)
          2 DF['DRIVEN_KM'] = DF['DRIVEN_KM'].str.replace(r'\D', '').astype(int)
```

After changing all the data with label encoders, I have tried in identifying skewness and outliers as follows.

We actually can see there are more skewness in the data let also see about the outliers in the data.

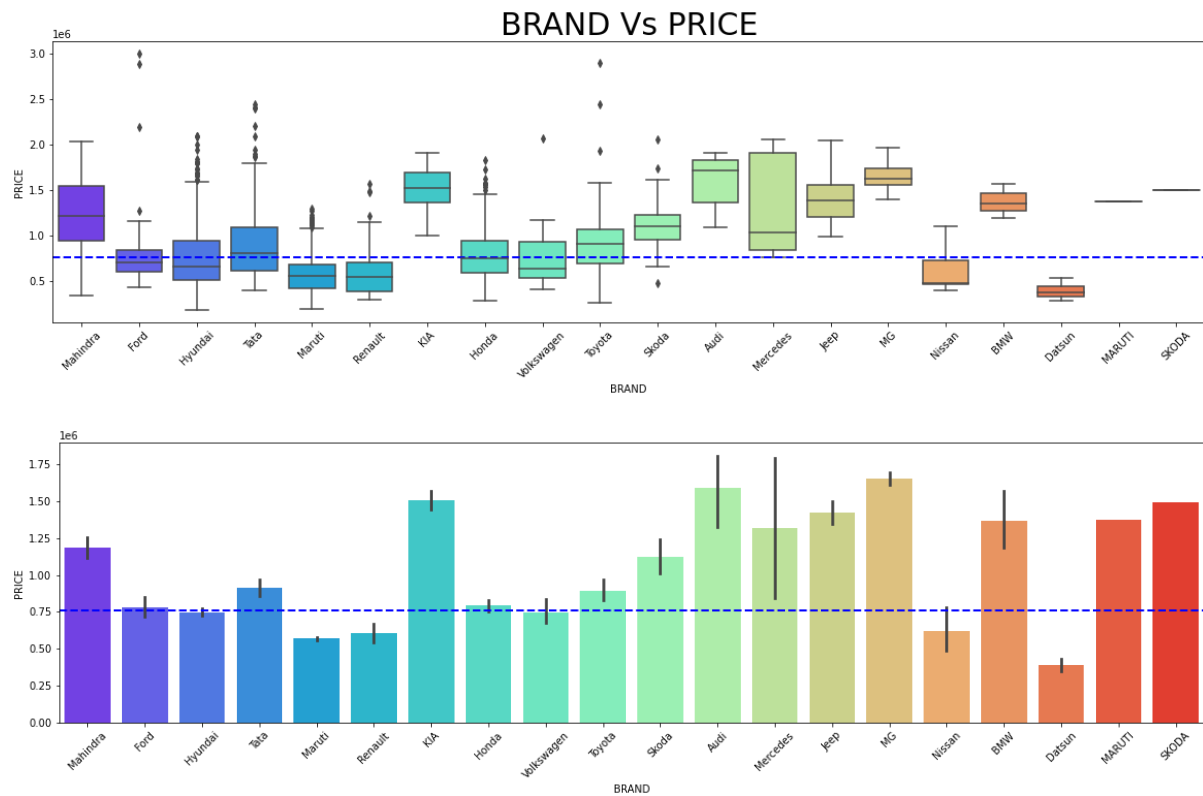


We have many outliers and we also have skewness in the data. Because its more, correcting them will have loss in the data or data will be deformed. I am deciding to build the model with the skewness and outliers present in the data and this will also help the client to get an accurate prediction on the car prices with the data he gets.

```

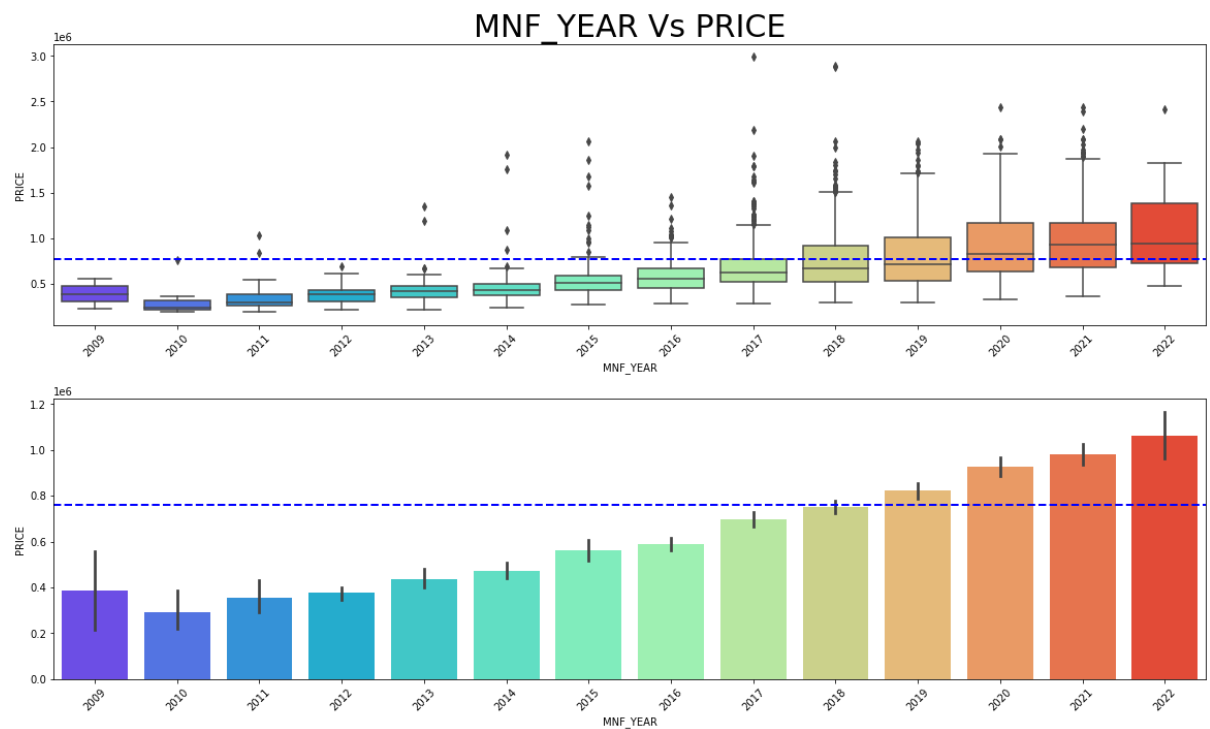
In [63]: 1 from scipy.stats import zscore
          2
          3 z = np.abs(zscore(DF))
          4 threshold = 3
          5 df_new = DF[(z < 3).all(axis=1)]

In [64]: 1 print(f"Original Data {DF.shape}\nAfter Removing outliers {df_new.shape}\nThe percentage of
          2
          3
          4
          5
          6
          7
          8
          9
         10
         11
         12
         13
         14
         15
         16
         17
         18
         19
         20
         21
         22
         23
         24
         25
         26
         27
         28
         29
         30
         31
         32
         33
         34
         35
         36
         37
         38
         39
         40
         41
         42
         43
         44
         45
         46
         47
         48
         49
         50
         51
         52
         53
         54
         55
         56
         57
         58
         59
         60
         61
         62
         63
         64
         65
         66
         67
         68
         69
         70
         71
         72
         73
         74
         75
         76
         77
         78
         79
         80
         81
         82
         83
         84
         85
         86
         87
         88
         89
         90
         91
         92
         93
         94
         95
         96
         97
         98
         99
        100
        101
        102
        103
        104
        105
        106
        107
        108
        109
        110
        111
        112
        113
        114
        115
        116
        117
        118
        119
        120
        121
        122
        123
        124
        125
        126
        127
        128
        129
        130
        131
        132
        133
        134
        135
        136
        137
        138
        139
        140
        141
        142
        143
        144
        145
        146
        147
        148
        149
        150
        151
        152
        153
        154
        155
        156
        157
        158
        159
        160
        161
        162
        163
        164
        165
        166
        167
        168
        169
        170
        171
        172
        173
        174
        175
        176
        177
        178
        179
        180
        181
        182
        183
        184
        185
        186
        187
        188
        189
        190
        191
        192
        193
        194
        195
        196
        197
        198
        199
        200
        201
        202
        203
        204
        205
        206
        207
        208
        209
        210
        211
        212
        213
        214
        215
        216
        217
        218
        219
        220
        221
        222
        223
        224
        225
        226
        227
        228
        229
        230
        231
        232
        233
        234
        235
        236
        237
        238
        239
        240
        241
        242
        243
        244
        245
        246
        247
        248
        249
        250
        251
        252
        253
        254
        255
        256
        257
        258
        259
        260
        261
        262
        263
        264
        265
        266
        267
        268
        269
        270
        271
        272
        273
        274
        275
        276
        277
        278
        279
        280
        281
        282
        283
        284
        285
        286
        287
        288
        289
        290
        291
        292
        293
        294
        295
        296
        297
        298
        299
        300
        301
        302
        303
        304
        305
        306
        307
        308
        309
        310
        311
        312
        313
        314
        315
        316
        317
        318
        319
        320
        321
        322
        323
        324
        325
        326
        327
        328
        329
        330
        331
        332
        333
        334
        335
        336
        337
        338
        339
        340
        341
        342
        343
        344
        345
        346
        347
        348
        349
        350
        351
        352
        353
        354
        355
        356
        357
        358
        359
        360
        361
        362
        363
        364
        365
        366
        367
        368
        369
        370
        371
        372
        373
        374
        375
        376
        377
        378
        379
        380
        381
        382
        383
        384
        385
        386
        387
        388
        389
        390
        391
        392
        393
        394
        395
        396
        397
        398
        399
        400
        401
        402
        403
        404
        405
        406
        407
        408
        409
        410
        411
        412
        413
        414
        415
        416
        417
        418
        419
        420
        421
        422
        423
        424
        425
        426
        427
        428
        429
        430
        431
        432
        433
        434
        435
        436
        437
        438
        439
        440
        441
        442
        443
        444
        445
        446
        447
        448
        449
        450
        451
        452
        453
        454
        455
        456
        457
        458
        459
        460
        461
        462
        463
        464
        465
        466
        467
        468
        469
        470
        471
        472
        473
        474
        475
        476
        477
        478
        479
        480
        481
        482
        483
        484
        485
        486
        487
        488
        489
        490
        491
        492
        493
        494
        495
        496
        497
        498
        499
        500
        501
        502
        503
        504
        505
        506
        507
        508
        509
        510
        511
        512
        513
        514
        515
        516
        517
        518
        519
        520
        521
        522
        523
        524
        525
        526
        527
        528
        529
        530
        531
        532
        533
        534
        535
        536
        537
        538
        539
        540
        541
        542
        543
        544
        545
        546
        547
        548
        549
        550
        551
        552
        553
        554
        555
        556
        557
        558
        559
        560
        561
        562
        563
        564
        565
        566
        567
        568
        569
        570
        571
        572
        573
        574
        575
        576
        577
        578
        579
        580
        581
        582
        583
        584
        585
        586
        587
        588
        589
        590
        591
        592
        593
        594
        595
        596
        597
        598
        599
        600
        601
        602
        603
        604
        605
        606
        607
        608
        609
        610
        611
        612
        613
        614
        615
        616
        617
        618
        619
        620
        621
        622
        623
        624
        625
        626
        627
        628
        629
        630
        631
        632
        633
        634
        635
        636
        637
        638
        639
        640
        641
        642
        643
        644
        645
        646
        647
        648
        649
        650
        651
        652
        653
        654
        655
        656
        657
        658
        659
        660
        661
        662
        663
        664
        665
        666
        667
        668
        669
        670
        671
        672
        673
        674
        675
        676
        677
        678
        679
        680
        681
        682
        683
        684
        685
        686
        687
        688
        689
        690
        691
        692
        693
        694
        695
        696
        697
        698
        699
        700
        701
        702
        703
        704
        705
        706
        707
        708
        709
        710
        711
        712
        713
        714
        715
        716
        717
        718
        719
        720
        721
        722
        723
        724
        725
        726
        727
        728
        729
        730
        731
        732
        733
        734
        735
        736
        737
        738
        739
        740
        741
        742
        743
        744
        745
        746
        747
        748
        749
        750
        751
        752
        753
        754
        755
        756
        757
        758
        759
        760
        761
        762
        763
        764
        765
        766
        767
        768
        769
        770
        771
        772
        773
        774
        775
        776
        777
        778
        779
        780
        781
        782
        783
        784
        785
        786
        787
        788
        789
        790
        791
        792
        793
        794
        795
        796
        797
        798
        799
        800
        801
        802
        803
        804
        805
        806
        807
        808
        809
        810
        811
        812
        813
        814
        815
        816
        817
        818
        819
        820
        821
        822
        823
        824
        825
        826
        827
        828
        829
        830
        831
        832
        833
        834
        835
        836
        837
        838
        839
        840
        841
        842
        843
        844
        845
        846
        847
        848
        849
        850
        851
        852
        853
        854
        855
        856
        857
        858
        859
        860
        861
        862
        863
        864
        865
        866
        867
        868
        869
        870
        871
        872
        873
        874
        875
        876
        877
        878
        879
        880
        881
        882
        883
        884
        885
        886
        887
        888
        889
        890
        891
        892
        893
        894
        895
        896
        897
        898
        899
        900
        901
        902
        903
        904
        905
        906
        907
        908
        909
        910
        911
        912
        913
        914
        915
        916
        917
        918
        919
        920
        921
        922
        923
        924
        925
        926
        927
        928
        929
        930
        931
        932
        933
        934
        935
        936
        937
        938
        939
        940
        941
        942
        943
        944
        945
        946
        947
        948
        949
        950
        951
        952
        953
        954
        955
        956
        957
        958
        959
        960
        961
        962
        963
        964
        965
        966
        967
        968
        969
        970
        971
        972
        973
        974
        975
        976
        977
        978
        979
        980
        981
        982
        983
        984
        985
        986
        987
        988
        989
        990
        991
        992
        993
        994
        995
        996
        997
        998
        999
       1000
       1001
       1002
       1003
       1004
       1005
       1006
       1007
       1008
       1009
       1010
       1011
       1012
       1013
       1014
       1015
       1016
       1017
       1018
       1019
       1020
       1021
       1022
       1023
       1024
       1025
       1026
       1027
       1028
       1029
       1030
       1031
       1032
       1033
       1034
       1035
       1036
       1037
       1038
       1039
       1040
       1041
       1042
       1043
       1044
       1045
       1046
       1047
       1048
       1049
       1050
       1051
       1052
       1053
       1054
       1055
       1056
       1057
       1058
       1059
       1060
       1061
       1062
       1063
       1064
       1065
       1066
       1067
       1068
       1069
       1070
       1071
       1072
       1073
       1074
       1075
       1076
       1077
       1078
       1079
       1080
       1081
       1082
       1083
       1084
       1085
       1086
       1087
       1088
       1089
       1090
       1091
       1092
       1093
       1094
       1095
       1096
       1097
       1098
       1099
      1100
      1101
      1102
      1103
      1104
      1105
      1106
      1107
      1108
      1109
      1110
      1111
      1112
      1113
      1114
      1115
      1116
      1117
      1118
      1119
      1120
      1121
      1122
      1123
      1124
      1125
      1126
      1127
      1128
      1129
      1130
      1131
      1132
      1133
      1134
      1135
      1136
      1137
      1138
      1139
      1140
      1141
      1142
      1143
      1144
      1145
      1146
      1147
      1148
      1149
      1150
      1151
      1152
      1153
      1154
      1155
      1156
      1157
      1158
      1159
      1160
      1161
      1162
      1163
      1164
      1165
      1166
      1167
      1168
      1169
      1170
      1171
      1172
      1173
      1174
      1175
      1176
      1177
      1178
      1179
      1180
      1181
      1182
      1183
      1184
      1185
      1186
      1187
      1188
      1189
      1190
      1191
      1192
      1193
      1194
      1195
      1196
      1197
      1198
      1199
     1200
     1201
     1202
     1203
     1204
     1205
     1206
     1207
     1208
     1209
     1210
     1211
     1212
     1213
     1214
     1215
     1216
     1217
     1218
     1219
     1220
     1221
     1222
     1223
     1224
     1225
     1226
     1227
     1228
     1229
     1230
     1231
     1232
     1233
     1234
     1235
     1236
     1237
     1238
     1239
     1240
     1241
     1242
     1243
     1244
     1245
     1246
     1247
     1248
     1249
     1250
     1251
     1252
     1253
     1254
     1255
     1256
     1257
     1258
     1259
     1260
     1261
     1262
     1263
     1264
     1265
     1266
     1267
     1268
     1269
     1270
     1271
     1272
     1273
     1274
     1275
     1276
     1277
     1278
     1279
     1280
     1281
     1282
     1283
     1284
     1285
     1286
     1287
     1288
     1289
     1290
     1291
     1292
     1293
     1294
     1295
     1296
     1297
     1298
     1299
    1300
    1301
    1302
    1303
    1304
    1305
    1306
    1307
    1308
    1309
    1310
    1311
    1312
    1313
    1314
    1315
    1316
    1317
    1318
    1319
    1320
    1321
    1322
    1323
    1324
    1325
    1326
    1327
    1328
    1329
    1330
    1331
    1332
    1333
    1334
    1335
    1336
    1337
    1338
    1339
    1340
    1341
    1342
    1343
    1344
    1345
    1346
    1347
    1348
    1349
    1350
    1351
    1352
    1353
    1354
    1355
    1356
    1357
    1358
    1359
    1360
    1361
    1362
    1363
    1364
    1365
    1366
    1367
    1368
    1369
    1370
    1371
    1372
    1373
    1374
    1375
    1376
    1377
    1378
    1379
    1380
    1381
    1382
    1383
    1384
    1385
    1386
    1387
    1388
    1389
    1390
    1391
    1392
    1393
    1394
    1395
    1396
    1397
    1398
    1399
   1400
   1401
   1402
   1403
   1404
   1405
   1406
   1407
   1408
   1409
   1410
   1411
   1412
   1413
   1414
   1415
   1416
   1417
   1418
   1419
   1420
   1421
   1422
   1423
   1424
   1425
   1426
   1427
   1428
   1429
   1430
   1431
   1432
   1433
   1434
   1435
   1436
   1437
   1438
   1439
   1440
   1441
   1442
   1443
   1444
   1445
   1446
   1447
   1448
   1449
   1450
   1451
   1452
   1453
   1454
   1455
   1456
   1457
   1458
   1459
   1460
   1461
   1462
   1463
   1464
   1465
   1466
   1467
   1468
   1469
   1470
   1471
   1472
   1473
   1474
   1475
   1476
   1477
   1478
   1479
   1480
   1481
   1482
   1483
   1484
   1485
   1486
   1487
   1488
   1489
   1490
   1491
   1492
   1493
   1494
   1495
   1496
   1497
   1498
   1499
  1500
  1501
  1502
  1503
  1504
  1505
  1506
  1507
  1508
  1509
  1510
  1511
  1512
  1513
  1514
  1515
  1516
  1517
  1518
  1519
  1520
  1521
  1522
  1523
  1524
  1525
  1526
  1527
  1528
  1529
  1530
  1531
  1532
  1533
  1534
  1535
  1536
  1537
  1538
  1539
  1540
  1541
  1542
  1543
  1544
  1545
  1546
  1547
  1548
  1549
  1550
  1551
  1552
  1553
  1554
  1555
  1556
  1557
  1558
  1559
  1560
  1561
  1562
  1563
  1564
  1565
  1566
  1567
  1568
  1569
  1570
  1571
  1572
  1573
  1574
  1575
  1576
  1577
  1578
  1579
  1580
  1581
  1582
  1583
  1584
  1585
  1586
  1587
  1588
  1589
  1590
  1591
  1592
  1593
  1594
  1595
  1596
  1597
  1598
  1599
 1600
1601
1602
1603
1604
1605
1606
1607
1608
1609
1610
1611
1612
1613
1614
1615
1616
1617
1618
1619
1620
1621
1622
1623
1624
1625
1626
1627
1628
1629
1630
1631
1632
1633
1634
1635
1636
1637
1638
1639
1640
1641
1642
1643
1644
1645
1646
1647
1648
1649
1650
1651
1652
1653
1654
1655
1656
1657
1658
1659
1660
1661
1662
1663
1664
1665
1666
1667
1668
1669
1670
1671
1672
1673
1674
1675
1676
1677
1678
1679
1680
1681
1682
1683
1684
1685
1686
1687
1688
1689
1690
1691
1692
1693
1694
1695
1696
1697
1698
1699
1700
1701
1702
1703
1704
1705
1706
1707
1708
1709
1710
1711
1712
1713
1714
1715
1716
1717
1718
1719
1720
1721
1722
1723
1724
1725
1726
1727
1728
1729
1730
1731
1732
1733
1734
1735
1736
1737
1738
1739
1740
1741
1742
1743
1744
1745
1746
1747
1748
1749
1750
1751
1752
1753
1754
1755
1756
1757
1758
1759
1760
1761
1762
1763
1764
1765
1766
1767
1768
1769
1770
1771
1772
1773
1774
1775
1776
1777
1778
1779
1780
1781
1782
1783
1784
1785
1786
1787
1788
1789
1790
1791
1792
1793
1794
1795
1796
1797
1798
1799
1800
1801
1802
1803
1804
1805
1806
1807
1808
1809
1810
1811
1812
1813
1814
1815
1816
1817
1818
1819
1820
1821
1822
1823
1824
1825
1826
1827
1828
1829
1830
1831
1832
1833
1834
1835
1836
1837
1838
1839
1840
1841
1842
1843
1844
1845
1846
1847
1848
1849
1850
1851
1852
1853
1854
1855
1856
1857
1858
1859
1860
1861
1862
1863
1864
1865
1866
1867
1868
1869
1870
1871
1872
1873
1874
1875
1876
1877
1878
1879
1880
1881
1882
1883
1884
1885
1886
1887
1888
1889
1890
1891
1892
1893
1894
1895
1896
1897
1898
1899
1900
1901
1902
1903
1904
1905
1906
1907
1908
1909
1910
1911
1912
1913
1914
1915
1916
1917
1918
1919
1920
1921
1922
1923
1924
1925
1926
1927
1928
1929
1930
1931
1932
1933
1934
1935
1936
1937
1938
1939
1940
1941
1942
1943
1944
1945
1946
1947
1948
1949
1950
1951
1952
1953
1954
1955
1956
1957
1958
1959
1960
1961
1962
1963
1964
1965
1966
1967
1968
1969
1970
1971
1972
1973
1974
1975
1976
1977
1978
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024
2025
2026
2027
2028
2029
2030
2031
2032
2033
2034
2035
203
```



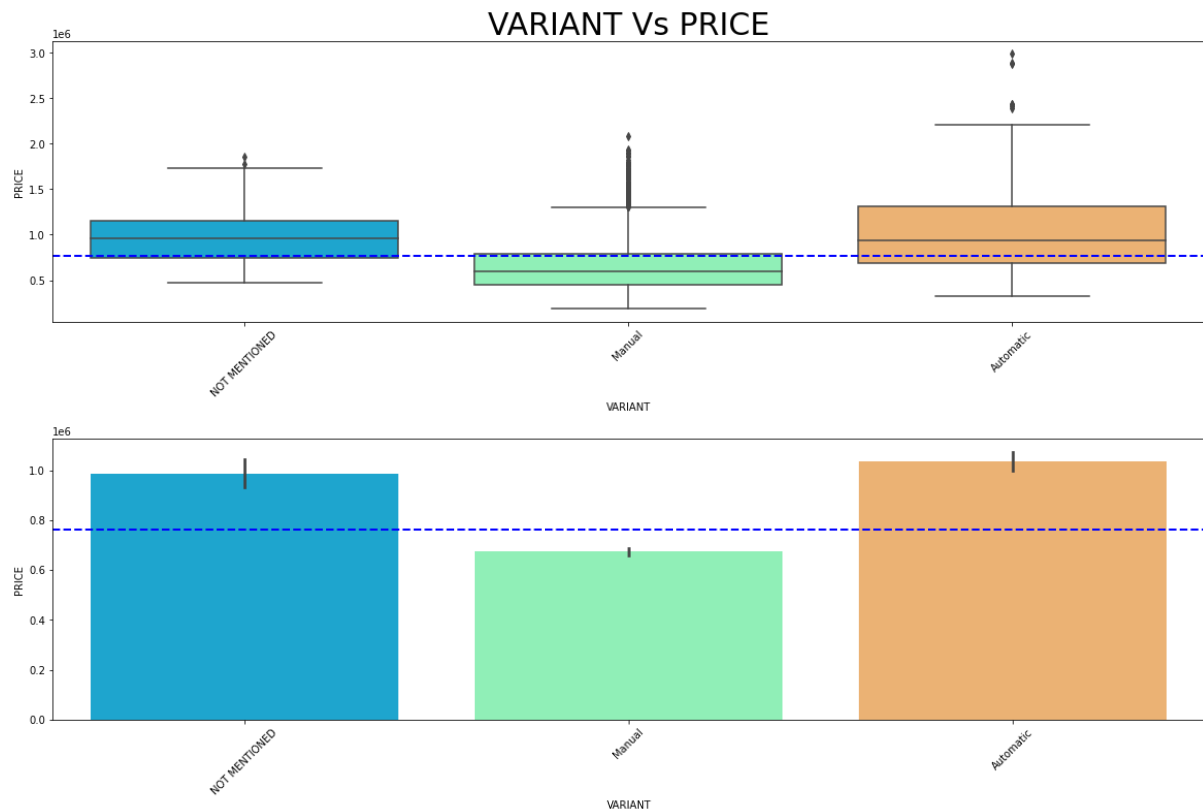
Key observations:

When comes to the Brand MG and Audi are being the costliest in country followed by the likes of BMW , KIA , Jeep etc .



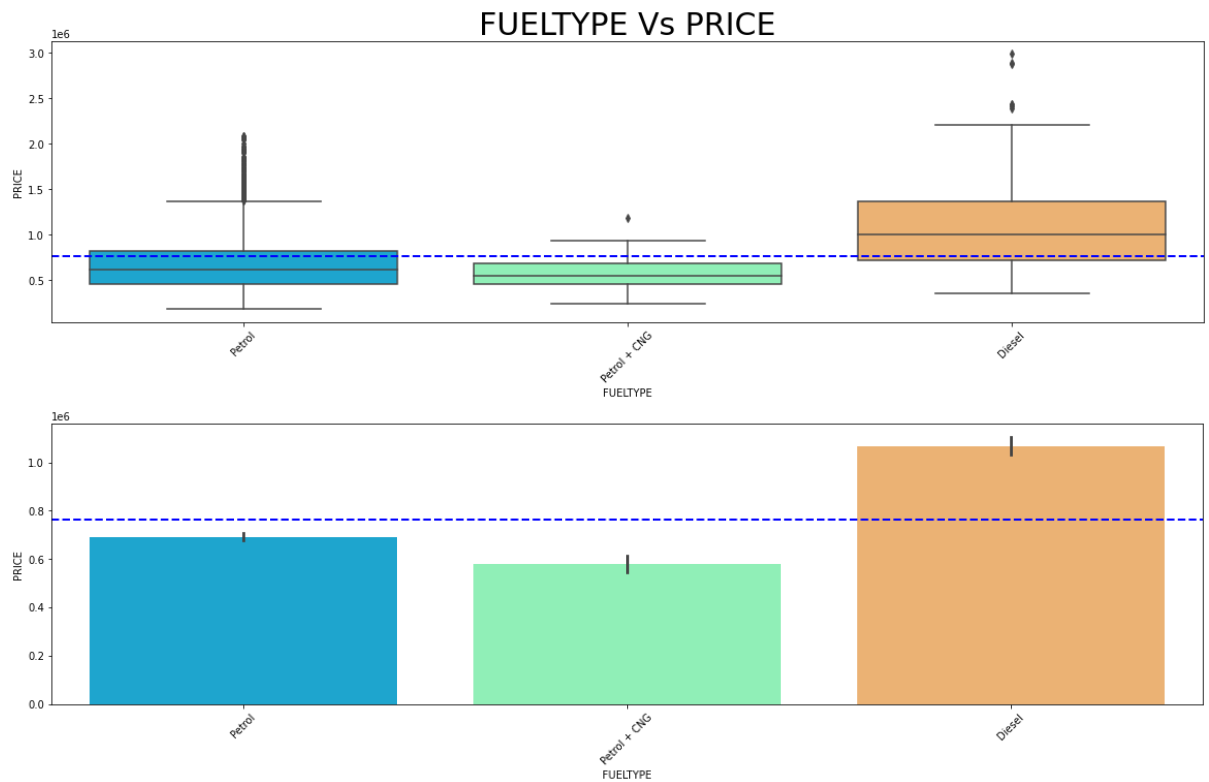
Key observations:

1. 2019, 2020, 2021 , 2022 model are being sold higher in PRICE and also above average PRICE.
2. Above we can understand that cars sold in lesser kms driven and also in lesser years used are sold in high price.



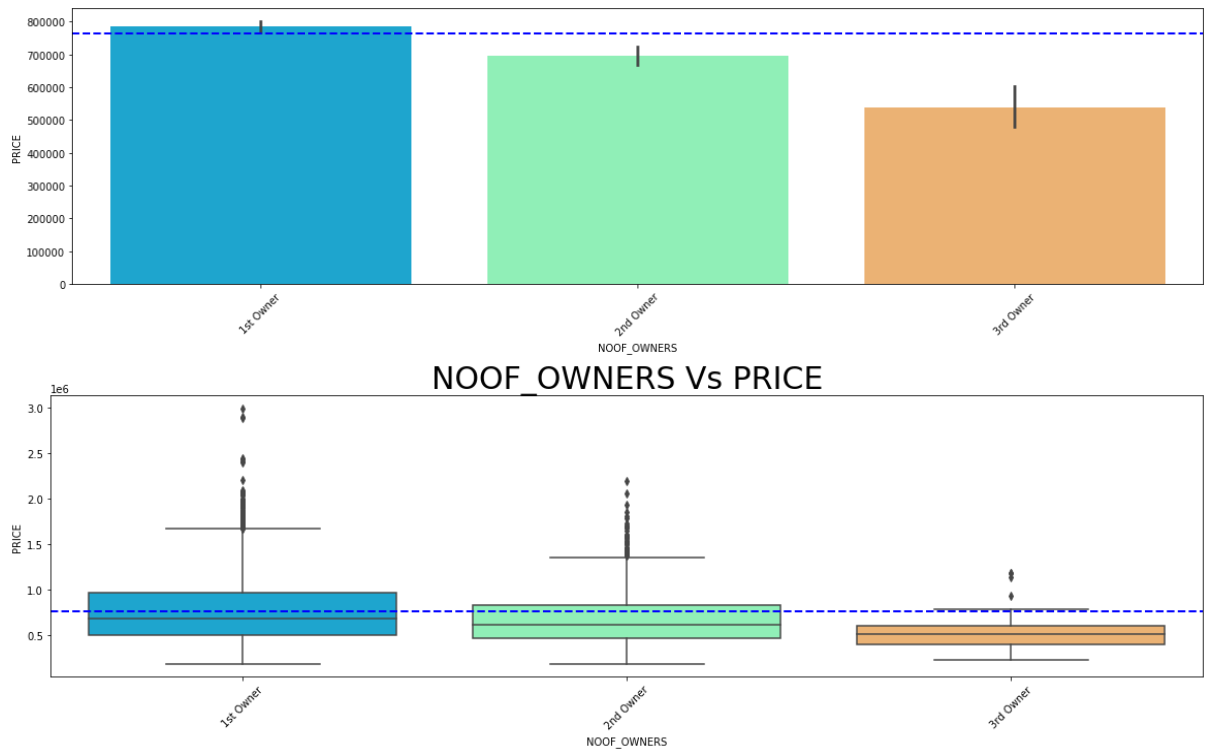
Key observations:

1. We can see that the automatic engines are costliest in the market. And also, most costlier cars come in Automatic shift.



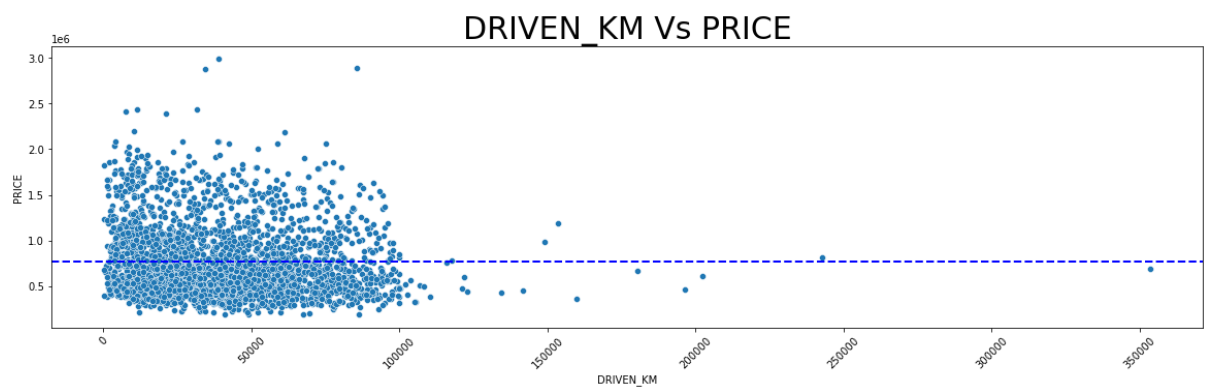
Key observations:

1. Diesel being the cheapest and most efficient fuel type, and so the Diesel engines are being the costliest fuel types.
2. Petrol bend second preferred followed by CNG and LPG fuel types



Key observation:

1st owner cars are costliest followed by second and third.



Key Observations:

The lesser kms driven are evidently sold costlier.

Hardware and Software Requirements and Tools Used

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

Model/s Development and Evaluation

Identification of possible problem-solving approaches (methods).

Considering the business requirement of the client, I have collected the precise data to predict the used car price but there were multiple data of the car that are available. Data like colour of the car, sun roof attached, music system brand, electronics in the car, tyre brands, seat colours and much more. But after analysing all these data I have selected the data that have more correlation with the price of the car. Data like manufacturing year, number of owners used before, mode, fuel variant, gear shift variant, Brand of the car. I experimented and visualized how these variables contributed more towards the deciding factor of the car price. Based on such visualization I have built the model.

Testing of Identified Approaches (Algorithms)

After the pre-processing of the data that is collected, I have split the data as `x_1` and `y_1` and I have imported the required libraries to train my model.

```
In [36]: x_1=DF.drop(["PRICE"], axis = 1)
         y_1=DF.PRICE
```

```
In [70]: 1 from sklearn.linear_model import LinearRegression, Lasso, Ridge, ElasticNet
         2 from sklearn.svm import SVR
         3 from sklearn.neighbors import KNeighborsRegressor
         4 from sklearn.tree import DecisionTreeRegressor
         5 from sklearn.ensemble import RandomForestRegressor, AdaBoostRegressor
```

on selecting the best random state parameter, I have used nine regression algorithms to train my model. Based on the best model score and best CV score I have selected Random Forest Regressor as the final model.

Run and evaluate selected models

I have used nine different regression algorithms to shortlist the best model.

Shortlisting the best model

```
In [42]: models = [LinearRegression(), Lasso(), Ridge(alpha=1, random_state=42), ElasticNet(), SVR(), KNeighborsRegressor(), DecisionTreeRegressor(),
                  RandomForestRegressor(), AdaBoostRegressor()]
         model_names = ["LinearRegression", "Lasso", "Ridge", "ElasticNet", "SVR", "KNeighborsRegressor", "DecisionTreeRegressor", "AdaBoostRegressor"]
```

Cross validation mean score and the model score.

As live saw above Random forest Regression model stands at the top with the model score of 98.76 with the CV score of 88.54 further I am going to hyperparameter tune the model to reduce over fitting and to increase the performance of the model.

```
In [74]: 1 Regression_result = pd.DataFrame({"MODEL": model_names,
2                                     "SCORE": score,
3                                     "CV_mean_score": mean_score,
4                                     "CV_STD": STD,
5                                     "MBE": mean_abs_e,
6                                     "MSE": mean_sqr_e,
7                                     "RMSE": root_mean_e,
8                                     "R2": r2
9                                     })
10 Regression_result.sort_values(by="CV_mean_score", ascending=False)
```

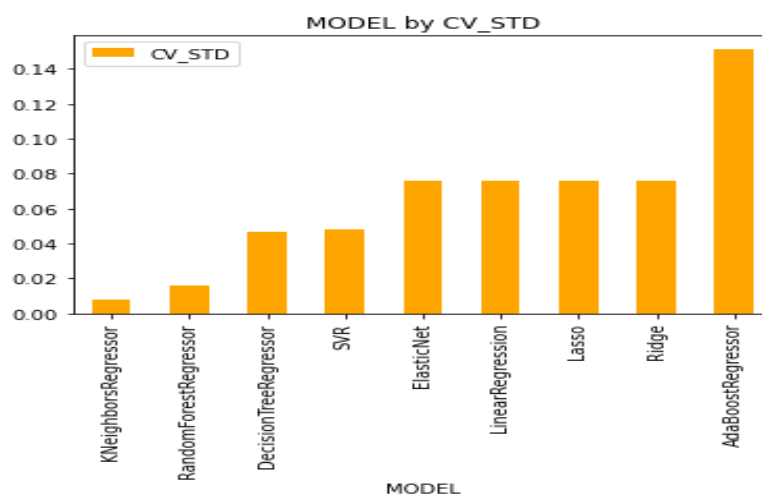
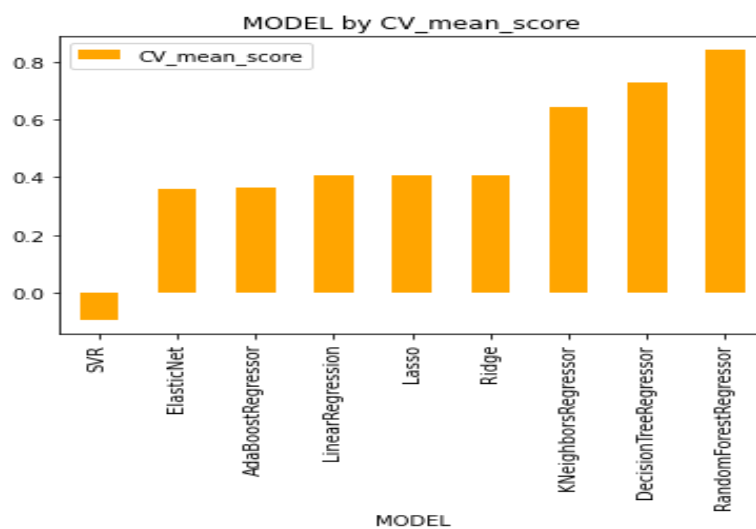
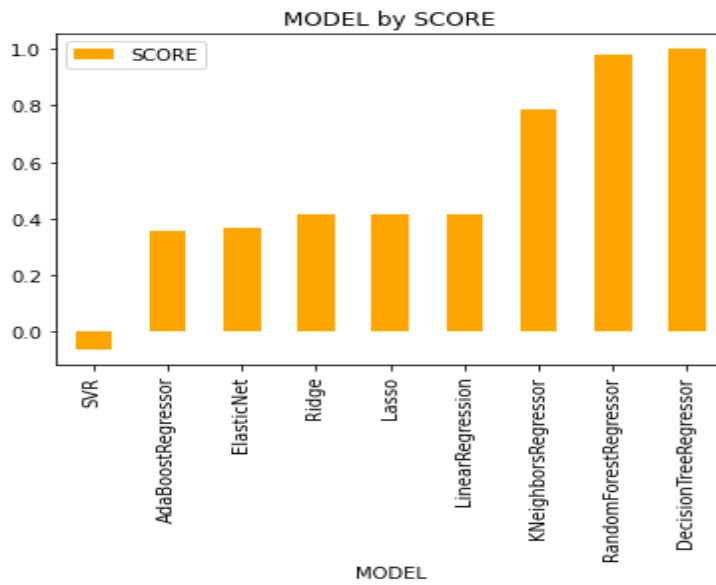
```
Out[74]:
```

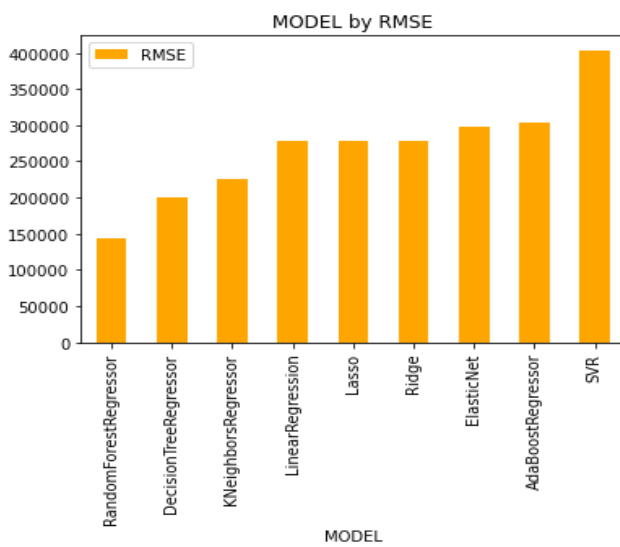
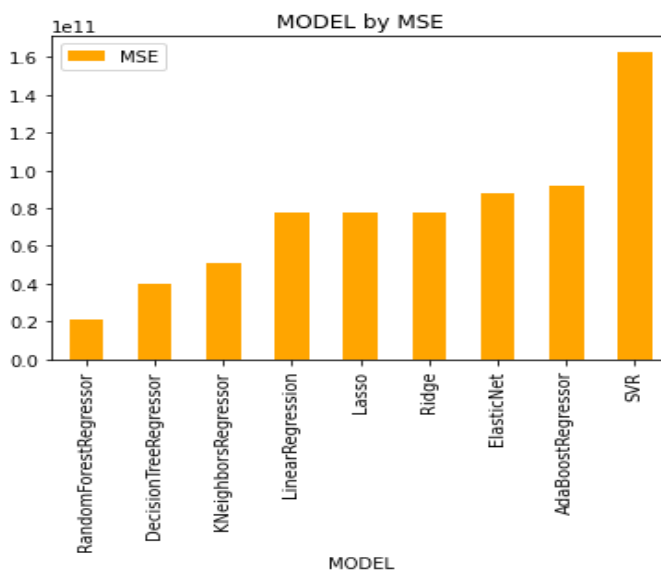
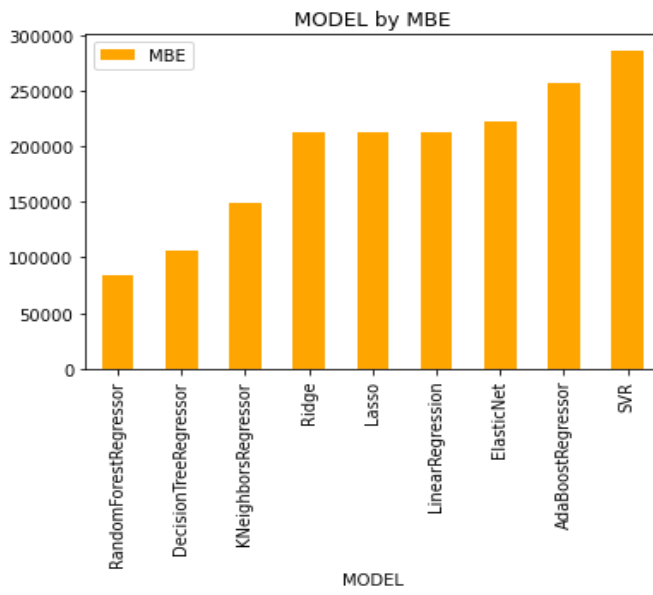
	MODEL	SCORE	CV_mean_score	CV_STD	MBE	MSE	RMSE	R2
8	RandomForestRegressor	0.979314	0.841111	0.015778	84026.630210	2.064815e+10	143694.632216	0.866157
6	DecisionTreeRegressor	1.000000	0.730536	0.046780	106765.739130	4.028384e+10	200708.340982	0.738877
5	KNeighborsRegressor	0.783535	0.645154	0.007712	149230.757363	5.076580e+10	225312.676994	0.670932
2	Ridge	0.416783	0.404763	0.076225	212431.861675	7.737795e+10	278168.928456	0.498430
1	Lasso	0.416784	0.404751	0.076217	212441.131081	7.737442e+10	278162.584271	0.498453
0	LinearRegression	0.416784	0.404751	0.076216	212441.300911	7.737434e+10	278162.437604	0.498453
7	AdaBoostRegressor	0.357893	0.364234	0.151647	256851.710333	9.162189e+10	302691.083921	0.406100
3	ElasticNet	0.367546	0.360062	0.075804	221929.688231	8.828648e+10	297130.416271	0.427720
4	SVR	-0.062405	-0.093185	0.048456	286877.591140	1.630589e+11	403805.551205	-0.056961

Model

selecting

Visualization:





From above observation, we can come to a conclusion that Random Forest is the best model with Score of 98.78 let's try in Hyper tuning the same for improved performance and also to reduce the over fitting the Data.

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. 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

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 cars sell 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 help him to understand based on what factors the Car Price is decided.

Limitations of this work and Scope for Future Work

The limitation of the study is that in the volatile changing market we have taken the data, to be more precise we have taken the data at the time of pandemic, so when the pandemic ends the market correction might happen slowly. So based on that again the deciding factors of the used car price might change and we have shortlisted and taken these data from the important cities across India, if the seller is from the different city our model might fail to predict the accurate price of that used car.