



## **TIME SERIES MODELING & ANALYSIS**

**Instructor Name:** Reza Jafari

**Lab#:** 5

**Submitted by** Dinesh Kumar Padmanabhan

**Date:** 16-Oct-2020

# ABSTRACT

In LAB #5, we learned concepts of Regression. Using dataset auto.csv we implemented the multi regression model. For the very model using Least squares estimation, the sum square of errors was minimized. using datasets, we plotted various time series plots and made a comparison with respect to training, testing and different regression methods and using Coefficient of determination or R-squared measured we assessed how well a model explains and predicts future outcome.

## INTRODUCTION

The auto.csv dataset will be used for this LAB. The dependent variable is 'price' and the independent variables are 'normalized-losses', 'wheel-base', 'length', 'width', 'height', 'curb-weight', 'engine-size', 'bore', 'stroke', 'compression-ratio', 'horsepower', 'peak-rpm', 'city-mpg' and 'highway-mpg'.

In this LAB you want to eliminate features and find the best multiple linear regression model.

The multiple linear regression model is:

$$y_t = \beta_0 + \beta_1 x_{1,t} + \beta_2 x_{2,t} + \dots + \beta_k x_{k,t} + \varepsilon_t$$

$\beta_0, \beta_1, \dots, \beta_k$  are unknown values which needs to be estimated using LSE using the following equation:

$$\hat{\beta} = (X^T X)^{-1} X^T Y$$

where X and Y are given as :

$$X = \begin{pmatrix} 1 & x_{1,1} & x_{2,1} & \dots & x_{k,1} \\ 1 & x_{1,2} & x_{2,2} & \dots & x_{k,2} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{1,T} & x_{2,T} & \dots & x_{k,T} \end{pmatrix} \quad Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{pmatrix}$$

Matrix X has T rows and (k+1) columns where T is the number of samples and k is number of independent variable.

# METHOD, THEORY & PROCEDURES

## Method:

1. Programming Language: Python

*Libraries used:* Some basic libraries used for analysis & model building are mentioned below

library(Numpy) - large collection of high-level mathematical functions to operate on these arrays.

library (Pandas) – For Data manipulation and analysis

library(Matplotlib) – is a system for declaratively creating graphics

library(Math) –To Compute mathematical calculations

library (statsmodels) – Import statistical models

## Theory:

We want to eliminate features and find the best multiple linear regression model. To Plot the forecast accuracy of above-mentioned methods for the given data set and determine which method performs better.

## Procedure:

I shall be looking at the results of various regression accuracy methods and time series plots and infer about it in my analysis. And through my exploration I shall try to identify which methods perform better and draw inferences.

The Dataset will be explored in following stages:

1. **Data Exploration (EDA)** – looking at the models and making inferences about the data.
2. **Data Visualization** – Plotting different time series plots for the regression method and forecast accuracy.
3. **Testing** – Running Autocorrelation, Pearson correlation test to identify the correlation between errors.

## ANSWERS TO QUESTIONS

File - unknown

```
1 C:\ProgramData\Anaconda3\python.exe "C:\Program Files\
  JetBrains\PyCharm 2019.3.1\plugins\python\helpers\pydev\
  pydevconsole.py" --mode=client --port=53461
2
3 import sys; print('Python %s on %s' % (sys.version, sys.
  platform))
4 sys.path.extend(['C:\\Users\\nsree_000\\Desktop\\Python-
  Quiz', 'C:/Users/nsree_000/Desktop/Python-Quiz'])
5
6 Python 3.7.4 (default, Aug 9 2019, 18:34:13) [MSC v.1915
  64 bit (AMD64)]
7 Type 'copyright', 'credits' or 'license' for more
  information
8 IPython 7.8.0 -- An enhanced Interactive Python. Type '?'
  for help.
9 PyDev console: using IPython 7.8.0
10
11 Python 3.7.4 (default, Aug 9 2019, 18:34:13) [MSC v.1915
  64 bit (AMD64)] on win32
12 In[2]: runfile('C:/Users/nsree_000/Desktop/Python-Quiz/TIME
  SERIES/LAB5.py', wdir='C:/Users/nsree_000/Desktop/Python-
  Quiz/TIME SERIES')
13
14
15 Unknown Coefficients: [-58079.71740468      1.65121682
  164.25166205 -59.57672822
16 393.49354813 161.54990942 1.00056352 117.
  72013471
17 -152.30810365 -3019.31506977 319.07356781 48.
  53290032
18 3.07146384 -281.5442421 224.5818695 ]
19
20
21 OLS Regression Results
22 =====
23 Dep. Variable: y R-squared
24 Model: OLS Adj. R-squared
25 Method: Least Squares F-statistic
26 Date: Fri, 16 Oct 2020 Prob (F-statistic)
27 Time: 13:49:07 Log-Likelihood
```

File - unknown

```

27 : -1514.4
28 No. Observations: 160 AIC
   : 3059.
29 Df Residuals: 145 BIC
   : 3105.
30 Df Model: 14

31 Covariance Type: nonrobust

32 =====
33      coef      std err      t      P>|t
34 |      [0.025      0.975]
35 -----
35 const -5.808e+04  1.85e+04  -3.138  0.002  -9
   .47e+04 -2.15e+04
36 x1 1.6512  10.580  0.156  0.876
   -19.259 22.561
37 x2 164.2517  115.175  1.426  0.156
   -63.388 391.891
38 x3 -59.5767  62.746  -0.949  0.344  -
   183.592 64.438
39 x4 393.4935  285.660  1.377  0.170  -
   171.101 958.088
40 x5 161.5499  162.238  0.996  0.321  -
   159.107 482.207
41 x6 1.0006  1.997  0.501  0.617
   -2.947 4.948
42 x7 117.7201  17.209  6.841  0.000
   83.707 151.733
43 x8 -152.3081  1521.939  -0.100  0.920  -
   3160.359 2855.742
44 x9 -3019.3151  849.671  -3.554  0.001  -
   4698.655 -1339.975
45 x10 319.0736  103.795  3.074  0.003
   113.927 524.220
46 x11 48.5329  22.888  2.120  0.036
   3.296 93.770
47 x12 3.0715  0.936  3.283  0.001
   1.222 4.921
48 x13 -281.5442  214.788  -1.311  0.192  -
   706.064 142.976
49 x14 224.5819  196.248  1.144  0.254  -
   163.293 612.457
50 =====

```

Page 2 of 15

File - unknown

```

50 =====
51 Omnibus:                18.174  Durbin-Watson
52   :                0.993
53 Prob(Omnibus):          0.000  Jarque-Bera (JB
54   ):                73.988
55 Skew:                -0.070  Prob(JB
56   ):                8.58e-17
57 Kurtosis:              6.328  Cond. No
58   :                4.10e+05
59 =====
60
61 Warnings:
62 [1] Standard Errors assume that the covariance matrix of
63 the errors is correctly specified.
64 [2] The condition number is large, 4.1e+05. This might
65 indicate that there are
66 strong multicollinearity or other numerical problems.
67
68 variance of prediction error: 10752385.280026225
69 variance of forecast error: 16120350.653554574
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
2036
2037
2038
2039
2040
2041
2042
2043
2044
2045
2046
2047
2048
2049
2050
2051
2052
2053
2054
2055
2056
2057
2058
2059
2060
2061
2062
2063
2064
2065
2066
2067
2068
2069
2070
2071
2072
2073
2074
2075
2076
2077
2078
2079
2080
2081
2082
2083
2084
2085
2086
2087
2088
2089
2090
2091
2092
2093
2094
2095
2096
2097
2098
2099
2100
2101
2102
2103
2104
2105
2106
2107
2108
2109
2110
2111
2112
2113
2114
2115
2116
2117
2118
2119
2120
2121
2122
2123
2124
2125
2126
2127
2128
2129
2130
2131
2132
2133
2134
2135
2136
2137
2138
2139
2140
2141
2142
2143
2144
2145
2146
2147
2148
2149
2150
2151
2152
2153
2154
2155
2156
2157
2158
2159
2160
2161
2162
2163
2164
2165
2166
2167
2168
2169
2170
2171
2172
2173
2174
2175
2176
2177
2178
2179
2180
2181
2182
2183
2184
2185
2186
2187
2188
2189
2190
2191
2192
2193
2194
2195
2196
2197
2198
2199
2200
2201
2202
2203
2204
2205
2206
2207
2208
2209
2210
2211
2212
2213
2214
2215
2216
2217
2218
2219
2220
2221
2222
2223
2224
2225
2226
2227
2228
2229
2230
2231
2232
2233
2234
2235
2236
2237
2238
2239
2240
2241
2242
2243
2244
2245
2246
2247
2248
2249
2250
2251
2252
2253
2254
2255
2256
2257
2258
2259
2260
2261
2262
2263
2264
2265
2266
2267
2268
2269
2270
2271
2272
2273
2274
2275
2276
2277
2278
2279
2280
2281
2282
2283
2284
2285
2286
2287
2288
2289
2290
2291
2292
2293
2294
2295
2296
2297
2298
2299
2300
2301
2302
2303
2304
2305
2306
2307
2308
2309
2310
2311
2312
2313
2314
2315
2316
2317
2318
2319
2320
2321
2322
2323
2324
2325
2326
2327
2328
2329
2330
2331
2332
2333
2334
2335
2336
2337
2338
2339
2340
2341
2342
2343
2344
2345
2346
2347
2348
2349
2350
2351
2352
2353
2354
2355
2356
2357
2358
2359
2360
2361
2362
2363
2364
2365
2366
2367
2368
2369
2370
2371
2372
2373
2374
2375
2376
2377
2378
2379
2380
2381
2382
2383
2384
2385
2386
2387
2388
2389
2390
2391
2392
2393
2394
2395
2396
2397
2398
2399
2400
2401
2402
2403
2404
2405
2406
2407
2408
2409
2410
2411
2412
2413
2414
2415
2416
2417
2418
2419
2420
2421
2422
2423
2424
2425
2426
2427
2428
2429
2430
2431
2432
2433
2434
2435
2436
2437
2438
2439
2440
2441
2442
2443
2444
2445
2446
2447
2448
2449
2450
2451
2452
2453
2454
2455
2456
2457
2458
2459
2460
2461
2462
2463
2464
2465
2466
2467
2468
2469
2470
2471
2472
2473
2474
2475
2476
2477
2478
2479
2480
2481
2482
2483
2484
2485
2486
2487
2488
2489
2490
2491
2492
2493
2494
2495
2496
2497
2498
2499
2500
2501
2502
2503
2504
2505
2506
2507
2508
2509
2510
2511
2512
2513
2514
2515
2516
2517
2518
2519
2520
2521
2522
2523
2524
2525
2526
2527
2528
2529
2530
2531
2532
2533
2534
2535
2536
2537
2538
2539
2540
2541
2542
2543
2544
2545
2546
2547
2548
2549
2550
2551
2552
2553
2554
2555
2556
2557
2558
2559
2560
2561
2562
2563
2564
2565
2566
2567
2568
2569
2570
2571
2572
2573
2574
2575
2576
2577
2578
2579
2580
2581
2582
2583
2584
2585
2586
2587
2588
2589
2590
2591
2592
2593
2594
2595
2596
2597
2598
2599
2600
2601
2602
2603
2604
2605
2606
2607
2608
2609
2610
2611
2612
2613
2614
2615
2616
2617
2618
2619
2620
2621
2622
2623
2624
2625
2626
2627
2628
2629
2630
2631
2632
2633
2634
2635
2636
2637
2638
2639
2640
2641
2642
2643
2644
2645
2646
2647
2648
2649
2650
2651
2652
2653
2654
2655
2656
2657
2658
2659
2660
2661
2662
2663
2664
2665
2666
2667
2668
2669
2670
```

File - unknown

|     |                  |            |           |                 |       |
|-----|------------------|------------|-----------|-----------------|-------|
| 79  |                  |            |           |                 |       |
| 80  | Covariance Type: |            | nonrobust |                 |       |
| 81  | =====            |            |           |                 |       |
| 82  |                  | coef       | std err   | t               | P> t  |
|     |                  | [0.025     | 0.975]    |                 |       |
| 83  | -----            |            |           |                 |       |
| 84  | const            | -5.88e+04  | 1.7e+04   | -3.457          | 0.001 |
|     |                  | 9.24e+04   | -2.52e+04 |                 |       |
| 85  | x1               | 1.7897     | 10.453    | 0.171           | 0.864 |
|     |                  | -18.869    | 22.449    |                 |       |
| 86  | x2               | 163.3571   | 114.438   | 1.427           | 0.156 |
|     |                  | -62.812    | 389.526   |                 |       |
| 87  | x3               | -60.2266   | 62.197    | -0.968          | 0.334 |
|     |                  | -183.150   | 62.697    |                 |       |
| 88  | x4               | 394.0880   | 284.628   | 1.385           | 0.168 |
|     |                  | -168.435   | 956.611   |                 |       |
| 89  | x5               | 163.2132   | 160.836   | 1.015           | 0.312 |
|     |                  | -154.655   | 481.082   |                 |       |
| 90  | x6               | 1.0186     | 1.982     | 0.514           | 0.608 |
|     |                  | -2.899     | 4.937     |                 |       |
| 91  | x7               | 117.8979   | 17.059    | 6.911           | 0.000 |
|     |                  | 84.183     | 151.612   |                 |       |
| 92  | x8               | -3005.5437 | 835.606   | -3.597          | 0.000 |
|     |                  | 4656.990   | -1354.098 |                 |       |
| 93  | x9               | 317.1114   | 101.580   | 3.122           | 0.002 |
|     |                  | 116.354    | 517.869   |                 |       |
| 94  | x10              | 47.9373    | 22.026    | 2.176           | 0.031 |
|     |                  | 4.407      | 91.467    |                 |       |
| 95  | x11              | 3.1064     | 0.865     | 3.590           | 0.000 |
|     |                  | 1.396      | 4.817     |                 |       |
| 96  | x12              | -279.5013  | 213.090   | -1.312          | 0.192 |
|     |                  | -700.640   | 141.638   |                 |       |
| 97  | x13              | 224.7105   | 195.577   | 1.149           | 0.252 |
|     |                  | -161.817   | 611.238   |                 |       |
| 98  | =====            |            |           |                 |       |
| 99  | Omnibus:         |            | 17.965    | Durbin-Watson   |       |
|     | :                | 0.990      |           |                 |       |
| 100 | Prob(Omnibus):   |            | 0.000     | Jarque-Bera (JB |       |
|     | ):               | 72.315     |           |                 |       |
| 101 | Skew:            |            | -0.064    | Prob(JB         |       |
|     | ):               | 1.98e-16   |           |                 |       |
| 102 | Kurtosis:        |            | 6.291     | Cond. No        |       |

Page 4 of 15



File - unknown

```

102 . 3.78e+05
103 =====
104
105 Warnings:
106 [1] Standard Errors assume that the covariance matrix of
107 the errors is correctly specified.
108 [2] The condition number is large, 3.78e+05. This might
109 indicate that there are
110 strong multicollinearity or other numerical problems.
111 OLS Regression Results
112 =====
113 Dep. Variable: y R-squared
114 : 0.869 OLS Adj. R-squared
115 : 0.858
116 Method: Least Squares F-statistic
117 : 81.15
118 Date: Fri, 16 Oct 2020 Prob (F-statistic)
119 : 1.57e-58
120 Time: 13:49:15 Log-Likelihood
121 : -1514.4
122 No. Observations: 160 AIC
123 : 3055.
124 Df Residuals: 147 BIC
125 : 3095.
126 Df Model: 12
127 Covariance Type: nonrobust
128 =====
129
130 coef std err t P>|t
131 | [0.025 0.975]
132 -----
133 const -5.871e+04 1.69e+04 -3.465 0.001 -
134 9.22e+04 -2.52e+04
135 x1 166.0377 112.987 1.470 0.144
136 -57.251 389.327
137 x2 -60.6161 61.950 -0.978 0.329
138 -183.044 61.812
139 x3 398.8909 282.306 1.413 0.160
140 -159.011 956.792

```

Page 5 of 15

File - unknown

```

151 :                               88.93
152 Date:                               Fri, 16 Oct 2020   Prob (F-statistic
   ):                               1.87e-59
153 Time:                               13:49:15   Log-Likelihood
   :                               -1514.6
154 No. Observations:                   160   AIC
   :                               3053.
155 Df Residuals:                       148   BIC
   :                               3090.
156 Df Model:                           11
157 Covariance Type:                   nonrobust

158 =====
159      coef      std err          t      P>|t
160      [0.025      0.975]
-----+-----
161 const      -6.006e+04   1.67e+04   -3.594   0.000   -
   9.31e+04   -2.7e+04
162 x1          178.8338    110.135    1.624   0.107
   -38.807     396.475
163 x2          -53.8332    60.480   -0.890   0.375
   -173.348     65.682
164 x3          425.3626    277.222    1.534   0.127
   -122.461     973.186
165 x4          158.2753    150.296    1.053   0.294
   -138.727     455.278
166 x5          119.7368    16.301    7.345   0.000
   87.523     151.950
167 x6          -2973.4431   828.249   -3.590   0.000   -
   4610.164   -1336.722
168 x7          337.4087    93.440    3.611   0.000
   152.760     522.057
169 x8          51.3240     21.016    2.442   0.016
   9.795      92.853
170 x9          2.9923      0.827    3.620   0.000
   1.359      4.626
171 x10         -285.6199   207.760   -1.375   0.171
   -696.179    124.939
172 x11         208.6321    184.739    1.129   0.261
   -156.436    573.700
173 =====
174 Omnibus:                               17.907   Durbin-Watson

```

File - unknown

```

174 : 1.021
175 Prob(Omnibus): 0.000 Jarque-Bera (JB
): 72.325
176 Skew: -0.043 Prob(JB
): 1.97e-16
177 Kurtosis: 6.293 Cond. No
: 3.35e+05
178 =====
179
180 Warnings:
181 [1] Standard Errors assume that the covariance matrix of
the errors is correctly specified.
182 [2] The condition number is large, 3.35e+05. This might
indicate that there are
183 strong multicollinearity or other numerical problems.
184 OLS Regression Results
185 =====
186 Dep. Variable: y R-squared
: 0.868
187 Model: OLS Adj. R-squared
: 0.859
188 Method: Least Squares F-statistic
: 97.88
189 Date: Fri, 16 Oct 2020 Prob (F-statistic
): 2.72e-60
190 Time: 13:49:16 Log-Likelihood
: -1515.0
191 No. Observations: 160 AIC
: 3052.
192 Df Residuals: 149 BIC
: 3086.
193 Df Model: 10
194 Covariance Type: nonrobust
195 =====
196 coef std err t P>|t
| [0.025 0.975]
197 -----
198 const -5.875e+04 1.66e+04 -3.531 0.001 -
9.16e+04 -2.59e+04

```

Page 8 of 15

File - unknown

```

199 x1      133.0737    97.332    1.367    0.174
      -59.256    325.404
200 x2      348.9836    263.423    1.325    0.187
      -171.544    869.511
201 x3      120.7793    144.171    0.838    0.404
      -164.104    405.662
202 x4      118.9571    16.266    7.313    0.000
      86.814    151.100
203 x5      -2943.7248    826.999    -3.560    0.000 -
      4577.885    -1309.564
204 x6      328.8318    92.877    3.541    0.001
      145.306    512.358
205 x7      50.6306    20.986    2.413    0.017
      9.161    92.100
206 x8      3.0961    0.818    3.786    0.000
      1.480    4.712
207 x9      -218.9137    193.637    -1.131    0.260
      -601.544    163.716
208 x10     171.0471    179.724    0.952    0.343
      -184.089    526.183
209 =====
      =====
210 Omnibus:      17.374    Durbin-Watson
      :      1.006
211 Prob(Omnibus):      0.000    Jarque-Bera (JB
      ):      66.791
212 Skew:      -0.078    Prob(JB
      ):      3.14e-15
213 Kurtosis:      6.161    Cond. No
      :      3.34e+05
214 =====
      =====
215
216 Warnings:
217 [1] Standard Errors assume that the covariance matrix of
      the errors is correctly specified.
218 [2] The condition number is large, 3.34e+05. This might
      indicate that there are
219 strong multicollinearity or other numerical problems.
220 OLS Regression Results
221 =====
      =====
222 Dep. Variable:      y    R-squared
      :      0.867
223 Model:      OLS    Adj. R-squared

```

Page 9 of 15

File - unknown

```

223 : 0.859
224 Method: Least Squares F-statistic
: 108.9
225 Date: Fri, 16 Oct 2020 Prob (F-statistic)
: 3.58e-61
226 Time: 13:49:16 Log-Likelihood
: -1515.4
227 No. Observations: 160 AIC
: 3051.
228 Df Residuals: 150 BIC
: 3082.
229 Df Model: 9

230 Covariance Type: nonrobust

231 =====
232 =====
233 | [0.025 0.975]
234 -----
235 const -5.297e+04 1.51e+04 -3.502 0.001 -
8.29e+04 -2.31e+04
236 x1 178.8177 80.493 2.222 0.028
19.771 337.865
237 x2 305.6295 258.033 1.184 0.238
-204.219 815.478
238 x3 117.4667 16.153 7.272 0.000
85.550 149.383
239 x4 -3048.7616 816.626 -3.733 0.000 -
4662.338 -1435.185
240 x5 331.0819 92.746 3.570 0.000
147.825 514.339
241 x6 50.5417 20.965 2.411 0.017
9.116 91.967
242 x7 3.0132 0.811 3.715 0.000
1.411 4.616
243 x8 -215.8108 193.409 -1.116 0.266
-597.970 166.348
244 x9 168.0934 179.510 0.936 0.351
-186.602 522.789
245 =====
246 Omnibus: 17.709 Durbin-Watson
: 1.012
247 Prob(Omnibus): 0.000 Jarque-Bera (JB

```

File - unknown

```

246 ):              70.213
247 Skew:              -0.059   Prob(JB
):              5.67e-16
248 Kurtosis:          6.243   Cond. No
:              3.04e+05
249 =====
250
251 Warnings:
252 [1] Standard Errors assume that the covariance matrix of
the errors is correctly specified.
253 [2] The condition number is large, 3.04e+05. This might
indicate that there are
254 strong multicollinearity or other numerical problems.
255 OLS Regression Results

256 =====
257 Dep. Variable:          y   R-squared
:              0.866
258 Model:              OLS   Adj. R-squared
:              0.859
259 Method:          Least Squares   F-statistic
:              122.5
260 Date:              Fri, 16 Oct 2020   Prob (F-statistic)
):              4.83e-62
261 Time:              13:49:16   Log-Likelihood
:              -1515.9
262 No. Observations:          160   AIC
:              3050.
263 Df Residuals:              151   BIC
:              3077.
264 Df Model:              8
265 Covariance Type:          nonrobust

266 =====
267
268 |              coef      std err          t      P>|t
269 |              [0.025      0.975]
270 -----+-----
269 const          -4.952e+04   1.47e+04   -3.377   0.001   -
7.85e+04   -2.06e+04
270 x1              168.3419    79.679     2.113   0.036
10.911    325.772

```

Page 11 of 15

File - unknown

```

271 x2      281.3419    256.621    1.096    0.275
      -225.690    788.374
272 x3      115.4832    16.007    7.215    0.000
      83.857    147.109
273 x4     -2979.3949    812.928   -3.665    0.000  -
      4585.577   -1373.213
274 x5      328.8373    92.677    3.548    0.001
      145.726    511.948
275 x6      52.5411    20.848    2.520    0.013
      11.350    93.732
276 x7      3.0272     0.811    3.735    0.000
      1.426     4.629
277 x8     -52.2924    83.114   -0.629    0.530
      -216.508    111.923
278 =====
      =====
279 Omnibus:      17.933    Durbin-Watson
      :      0.987
280 Prob(Omnibus):    0.000    Jarque-Bera (JB
      ):      72.250
281 Skew:      -0.056    Prob(JB
      ):      2.05e-16
282 Kurtosis:      6.290    Cond. No
      :      2.94e+05
283 =====
      =====
284
285 Warnings:
286 [1] Standard Errors assume that the covariance matrix of
      the errors is correctly specified.
287 [2] The condition number is large, 2.94e+05. This might
      indicate that there are
288 strong multicollinearity or other numerical problems.
289 OLS Regression Results
290 =====
      =====
291 Dep. Variable:      y    R-squared
      :      0.866
292 Model:      OLS    Adj. R-squared
      :      0.860
293 Method:      Least Squares    F-statistic
      :      140.5
294 Date:      Fri, 16 Oct 2020    Prob (F-statistic)
      ):      4.77e-63
295 Time:      13:49:16    Log-Likelihood

```

Page 12 of 15

File - unknown

```

295 : -1516.1
296 No. Observations: 160 AIC
: 3048.
297 Df Residuals: 152 BIC
: 3073.
298 Df Model: 7
299 Covariance Type: nonrobust

300 =====
301 coef std err t P>|t
302 | [0.025 0.975]
303 -----
304 const -5.448e+04 1.23e+04 -4.414 0.000 -
7.89e+04 -3.01e+04
305 x1 183.9395 75.574 2.434 0.016
34.628 333.251
306 x2 314.8122 250.547 1.256 0.211
-180.193 809.817
307 x3 112.2363 15.122 7.422 0.000
82.360 142.113
308 x4 -2984.4499 811.271 -3.679 0.000 -
4587.273 -1381.627
309 x5 305.5303 84.783 3.604 0.000
138.026 473.035
310 x6 60.6620 16.339 3.713 0.000
28.381 92.943
311 x7 2.9749 0.805 3.697 0.000
1.385 4.565
312 =====
313 Omnibus: 17.915 Durbin-Watson
: 0.991
314 Prob(Omnibus): 0.000 Jarque-Bera (JB
): 71.425
315 Skew: -0.078 Prob(JB
): 3.09e-16
316 Kurtosis: 6.269 Cond. No
. 2.48e+05
317 =====
318 Warnings:
319 [1] Standard Errors assume that the covariance matrix of

```

Page 13 of 15



```

319 the errors is correctly specified.
320 [2] The condition number is large, 2.48e+05. This might
    indicate that there are
321 strong multicollinearity or other numerical problems.
322                                OLS Regression Results
323 =====
324 Dep. Variable:                    y    R-squared
325 :                               0.865
326 Model:                            OLS    Adj. R-squared
327 :                               0.859
328 Method:                          Least Squares    F-statistic
329 :                               163.0
330 Date:                            Fri, 16 Oct 2020    Prob (F-statistic)
331 :                               7.81e-64
332 Time:                            13:49:16    Log-Likelihood
333 :                               -1516.9
334 No. Observations:                160    AIC
335 :                               3048.
336 Df Residuals:                    153    BIC
337 :                               3069.
338 Df Model:                        6
339 Covariance Type:                 nonrobust
340 =====
341 |               coef      std err          t      P>|t
342 |-----+-----+-----+-----+-----+-----+
343 | [0.025      0.975]
344 -----+-----+-----+-----+-----+-----+
345 const      -4.164e+04    6932.165     -6.006     0.000    -
346 5.53e+04    -2.79e+04
347 x1          250.5779     53.941      4.645     0.000
348 144.012     357.144
349 x2          113.9081     15.092      7.548     0.000
350 84.093     143.723
351 x3         -2945.3206     812.205     -3.626     0.000    -
352 4549.905    -1340.736
353 x4          321.4083     83.994      3.827     0.000
354 155.471     487.346
355 x5           66.9018     15.595      4.290     0.000
356 36.091      97.712
357 x6           3.0111      0.806      3.737     0.000
358 1.419       4.603

```

```

343 =====
344 Omnibus:                                18.109   Durbin-Watson
      :                                1.032
345 Prob(Omnibus):                          0.000   Jarque-Bera (JB
      ):                                70.826
346 Skew:                                  -0.128   Prob(JB
      ):                                4.17e-16
347 Kurtosis:                             6.249   Cond. No
      :                                1.39e+05
348 =====
349
350 Warnings:
351 [1] Standard Errors assume that the covariance matrix of
      the errors is correctly specified.
352 [2] The condition number is large, 1.39e+05. This might
      indicate that there are
353 strong multicollinearity or other numerical problems.
354

```

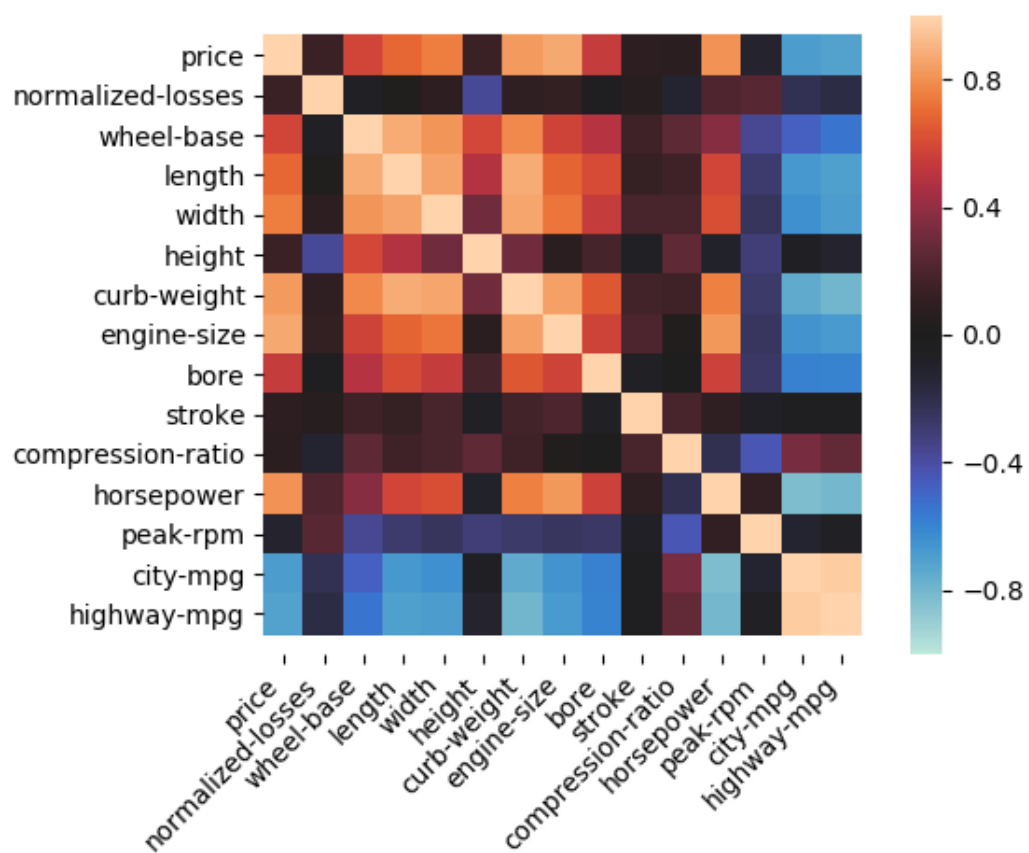
```

# %%=====
#1: Using Pandas library load the time series data from the BB.
# Split the dataset into training set and test set.
# Use 80% for training and 20% for testing. Display the training set and testing
set array as follow:
# Hint: This can be done using the following library in python.
# "from sklearn.model_selection import train_test_split" . Make sure the
"Shuffle=False"
# %%-----

# See Appendix section for code.

# %%=====
#2: Plot the correlation matrix using the seaborn package and heatmap function.
# %%=====

```

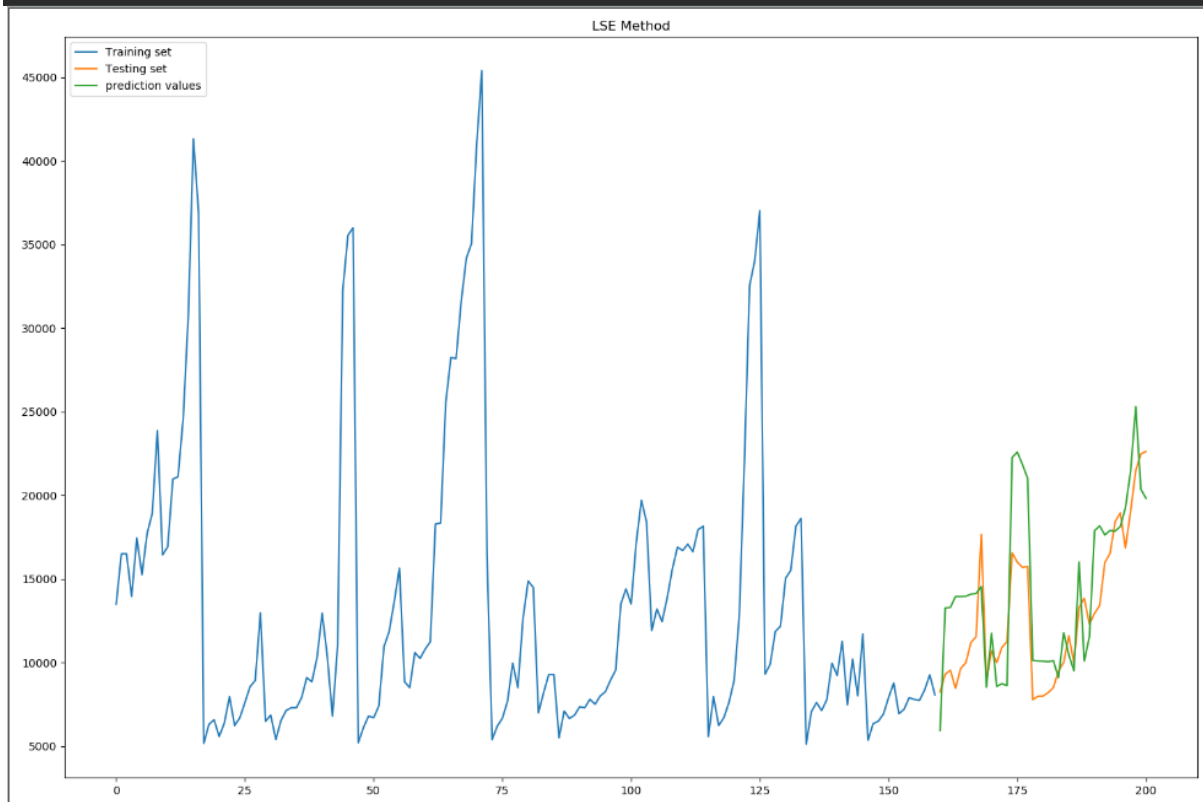


```

# %%=====
#3: Using python, construct matrix X and Y using x-train and y-train dataset and
estimate the regression model unknown coefficients using the Normal equation (LSE
method, above equation).
Display the unknown coefficients on the console.
Note: You are not allowed to use OLS for this part.
# %%-----

Unknown Coefficients: [-58079.71740468      1.65121682    164.25166205      -
59.57672822
      393.49354813    161.54990942      1.00056352    117.72013471
     -152.30810365   -3019.31506977    319.07356781     48.53290032
      3.07146384   -281.5442421    224.5818695 ]

```



```

# %%=====
#4: Using python, statsmodels package and OLS function, find the unknown
# coefficients. Compare the results with the step 3.
# Display the result on the console.
# Are the unknown coefficient calculated using step 3 and step 4 the same?
# %%-----

OLS Regression Results
=====
Dep. Variable:          y      R-squared:          0.869
Model:                OLS      Adj. R-squared:       0.856
Method:             Least Squares  F-statistic:        68.63
Date:                Fri, 16 Oct 2020  Prob (F-statistic):    1.25e-56
Time:                13:31:46   Log-Likelihood:     -1514.4
No. Observations:    160      AIC:                3059.
Df Residuals:        145      BIC:                3105.
Df Model:             14
Covariance Type:      nonrobust
=====

```

|       | coef       | std err  | t      | P> t  | [0.025    | 0.975]    |
|-------|------------|----------|--------|-------|-----------|-----------|
| const | -5.808e+04 | 1.85e+04 | -3.138 | 0.002 | -9.47e+04 | -2.15e+04 |
| x1    | 1.6512     | 10.580   | 0.156  | 0.876 | -19.259   | 22.561    |
| x2    | 164.2517   | 115.175  | 1.426  | 0.156 | -63.388   | 391.891   |
| x3    | -59.5767   | 62.746   | -0.949 | 0.344 | -183.592  | 64.438    |
| x4    | 393.4935   | 285.660  | 1.377  | 0.170 | -171.101  | 958.088   |
| x5    | 161.5499   | 162.238  | 0.996  | 0.321 | -159.107  | 482.207   |
| x6    | 1.0006     | 1.997    | 0.501  | 0.617 | -2.947    | 4.948     |
| x7    | 117.7201   | 17.209   | 6.841  | 0.000 | 83.707    | 151.733   |
| x8    | -152.3081  | 1521.939 | -0.100 | 0.920 | -3160.359 | 2855.742  |
| x9    | -3019.3151 | 849.671  | -3.554 | 0.001 | -4698.655 | -1339.975 |
| x10   | 319.0736   | 103.795  | 3.074  | 0.003 | 113.927   | 524.220   |
| x11   | 48.5329    | 22.888   | 2.120  | 0.036 | 3.296     | 93.770    |
| x12   | 3.0715     | 0.936    | 3.283  | 0.001 | 1.222     | 4.921     |
| x13   | -281.5442  | 214.788  | -1.311 | 0.192 | -706.064  | 142.976   |
| x14   | 224.5819   | 196.248  | 1.144  | 0.254 | -163.293  | 612.457   |

```

=====
Omnibus:                18.174   Durbin-Watson:          0.993
Prob(Omnibus):           0.000   Jarque-Bera (JB):        73.988
Skew:                    -0.070   Prob(JB):                8.58e-17
Kurtosis:                 6.328   Cond. No.                4.10e+05
=====

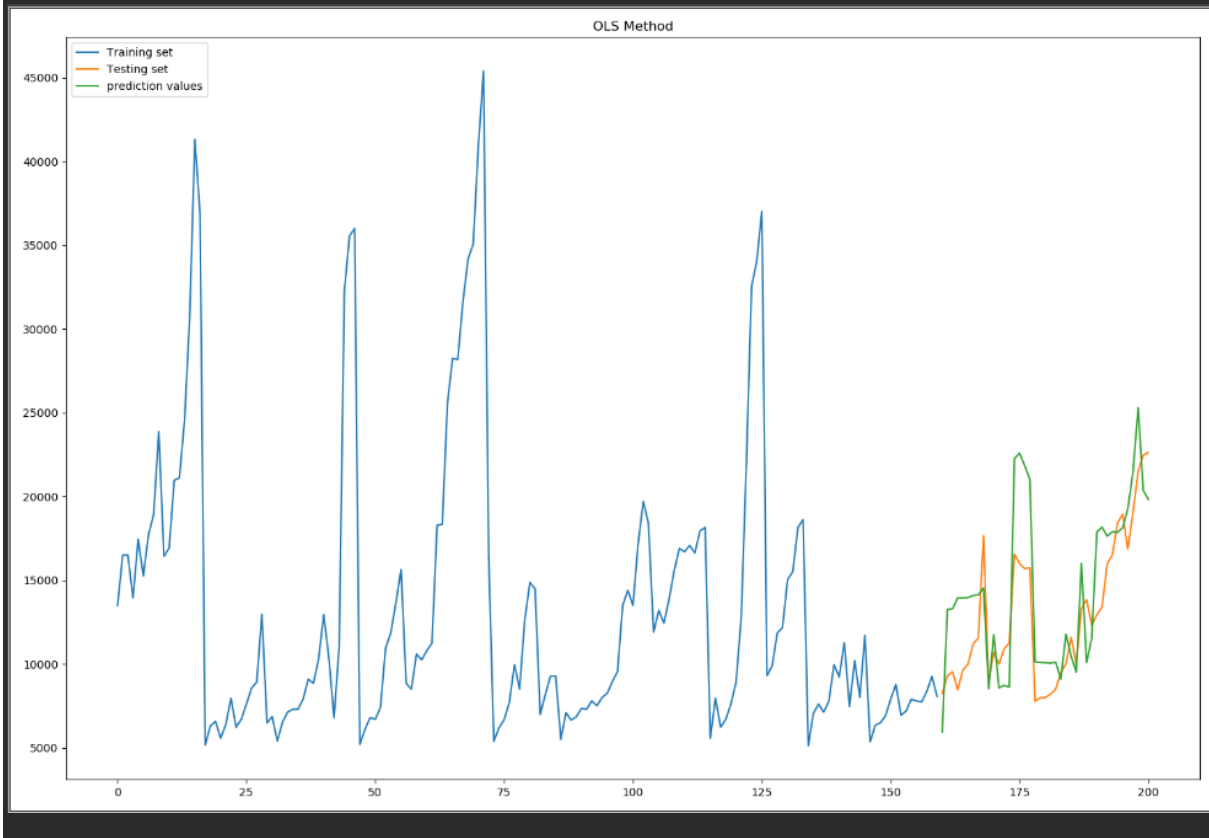
'''
Yes, the unknown coefficient calculated using step 3 and step 4 are same.
'''

```

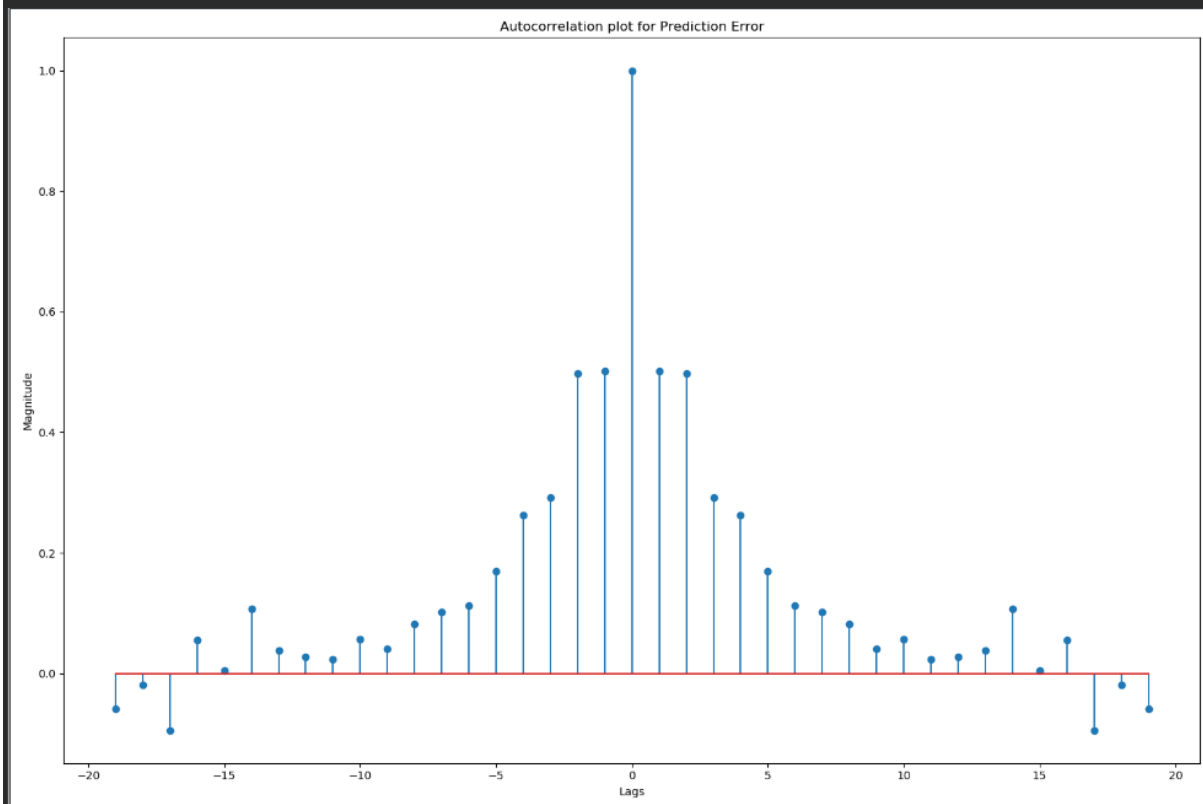
```

%%=====
# 5: Perform a prediction for the length of test-set and plot the train, test and
predicted values in one graph.
#Add appropriate x-label, y-label, title, and legend to your graph.
# %%-----

```



```
##%=====
# 6: Calculate the prediction errors and plot the ACF of prediction errors.
# Write down your observation
# %%-----
```

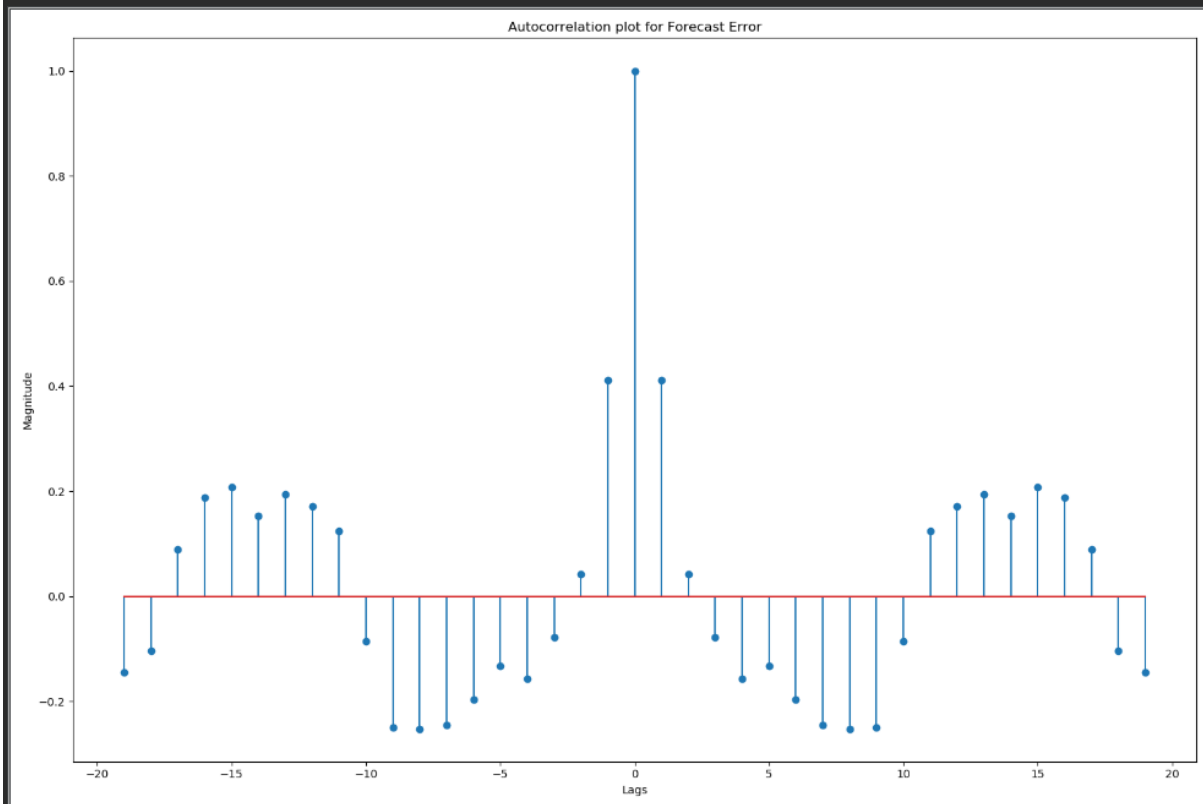


Autocorrelation of have a strong peak at "0". The values are distributed identically with mean of zero and variance of  $\sigma^2$ .

```

#####
# 7: Calculate the forecast errors and plot the ACF of forecast errors.
# Write down your observation.
# %%-----

```



Autocorrelation of have a strong peak at "0". The values are distributed identically with mean of zero and variance of  $\sigma^2$ .



```

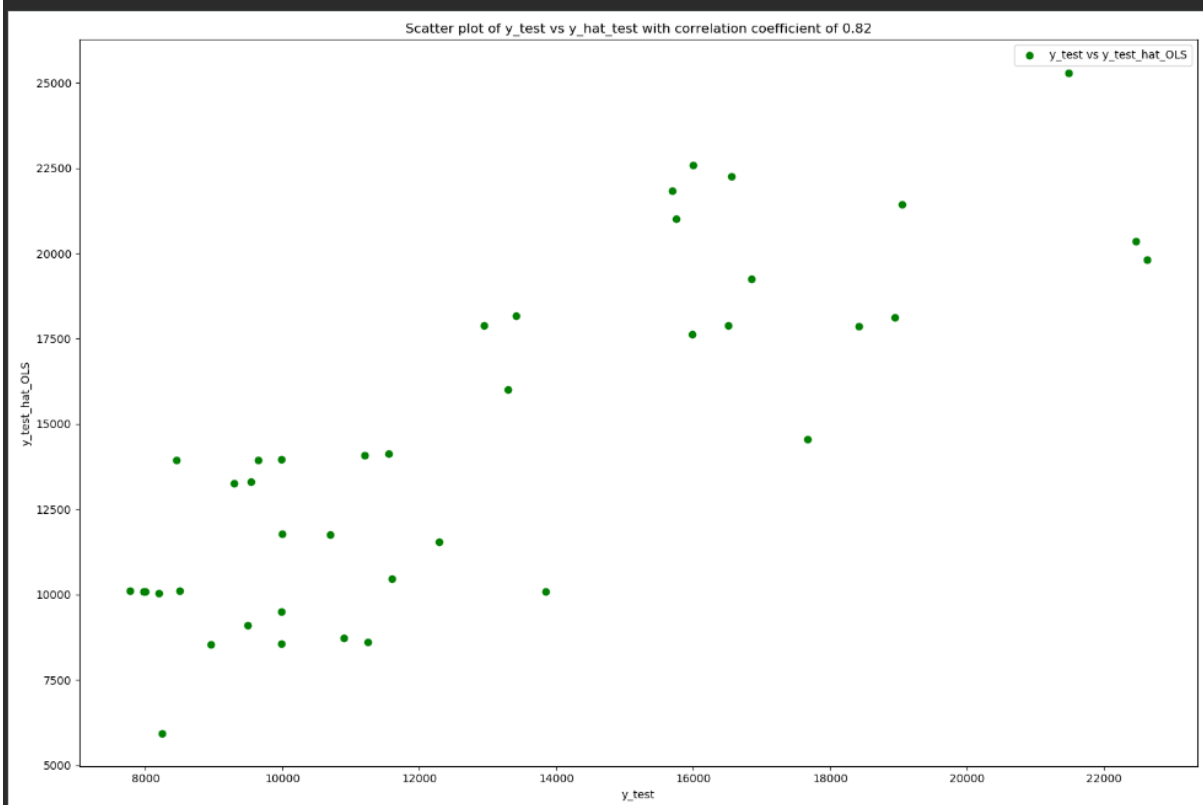
#%=====
#8: Calculate the estimated variance of the prediction errors and the forecast
errors. Compare the results?
# What is your observation?
Hint: Use need to the following equation to estimate the variance:
# %%-----

variance of prediction error: 10752385.280026225
variance of forecast error: 16120350.653554574.

The condition number for forecast error is large. This might indicate that there
strong multicollinearity.

#%=====
#9: Plot the scatter plot between y-test and  $\hat{y}_{t+h}$  and display the correlation
coefficient between them on the title.
# Justify the accuracy of this predictor by observing the correlation coefficient
between y-test and  $\hat{y}_t$ .
# %%-----

```



Accuracy of the predictor kind of matches with correlation coefficient which is positive in nature. Here we can see the extend of correlation between  $y_{test}$  v/s  $y_{test\_hat\_ols}$

```

###=====
#10: Using a stepwise regression, try to reduce the feature space dimension.
# You need to use the AIC, BIC and Adjusted R2 as a predictive accuracy for your
analysis.
# If your analysis recommends an elimination, which feature(s) would you
eliminate?
# You can use the backward or forward or hybrid stepwise regression for feature
selection.
# %%-----

# Removing bore
OLS Regression Results
=====
Dep. Variable:          y      R-squared:          0.869
Model:                OLS      Adj. R-squared:       0.857
Method:             Least Squares      F-statistic:        74.41
Date:                Fri, 16 Oct 2020    Prob (F-statistic):   1.43e-57
Time:                15:42:51      Log-Likelihood:      -1514.4
No. Observations:      160      AIC:                3057.
Df Residuals:          146      BIC:                3100.
Df Model:              13
Covariance Type:      nonrobust
=====

```

|       | coef       | std err | t      | P> t  | [0.025    | 0.975]    |
|-------|------------|---------|--------|-------|-----------|-----------|
| const | -5.88e+04  | 1.7e+04 | -3.457 | 0.001 | -9.24e+04 | -2.52e+04 |
| x1    | 1.7897     | 10.453  | 0.171  | 0.864 | -18.869   | 22.449    |
| x2    | 163.3571   | 114.438 | 1.427  | 0.156 | -62.812   | 389.526   |
| x3    | -60.2266   | 62.197  | -0.968 | 0.334 | -183.150  | 62.697    |
| x4    | 394.0880   | 284.628 | 1.385  | 0.168 | -168.435  | 956.611   |
| x5    | 163.2132   | 160.836 | 1.015  | 0.312 | -154.655  | 481.082   |
| x6    | 1.0186     | 1.982   | 0.514  | 0.608 | -2.899    | 4.937     |
| x7    | 117.8979   | 17.059  | 6.911  | 0.000 | 84.183    | 151.612   |
| x8    | -3005.5437 | 835.606 | -3.597 | 0.000 | -4656.990 | -1354.098 |
| x9    | 317.1114   | 101.580 | 3.122  | 0.002 | 116.354   | 517.869   |
| x10   | 47.9373    | 22.026  | 2.176  | 0.031 | 4.407     | 91.467    |
| x11   | 3.1064     | 0.865   | 3.590  | 0.000 | 1.396     | 4.817     |
| x12   | -279.5013  | 213.090 | -1.312 | 0.192 | -700.640  | 141.638   |
| x13   | 224.7105   | 195.577 | 1.149  | 0.252 | -161.817  | 611.238   |

```

=====
Omnibus:                17.965      Durbin-Watson:          0.990
Prob(Omnibus):           0.000      Jarque-Bera (JB):       72.315
Skew:                    -0.064      Prob(JB):               1.98e-16
Kurtosis:                 6.291      Cond. No.                3.78e+05
=====

```

```
#Removing normalized losses
OLS Regression Results
=====
Dep. Variable:          y      R-squared:          0.869
Model:                  OLS    Adj. R-squared:      0.858
Method:                 Least Squares    F-statistic:        81.15
Date:                  Fri, 16 Oct 2020    Prob (F-statistic):  1.57e-58
Time:                  15:42:51    Log-Likelihood:     -1514.4
No. Observations:      160    AIC:                3055.
Df Residuals:          147    BIC:                3095.
Df Model:              12
Covariance Type:       nonrobust
=====

```

|       | coef       | std err  | t      | P> t  | [0.025    | 0.975]    |
|-------|------------|----------|--------|-------|-----------|-----------|
| const | -5.871e+04 | 1.69e+04 | -3.465 | 0.001 | -9.22e+04 | -2.52e+04 |
| x1    | 166.0377   | 112.987  | 1.470  | 0.144 | -57.251   | 389.327   |
| x2    | -60.6161   | 61.950   | -0.978 | 0.329 | -183.044  | 61.812    |
| x3    | 398.8909   | 282.306  | 1.413  | 0.160 | -159.011  | 956.792   |
| x4    | 153.9118   | 150.883  | 1.020  | 0.309 | -144.269  | 452.092   |
| x5    | 1.0483     | 1.968    | 0.533  | 0.595 | -2.842    | 4.938     |
| x6    | 117.5226   | 16.862   | 6.970  | 0.000 | 84.200    | 150.845   |
| x7    | -3007.6949 | 832.748  | -3.612 | 0.000 | -4653.400 | -1361.990 |
| x8    | 316.9487   | 101.240  | 3.131  | 0.002 | 116.876   | 517.022   |
| x9    | 48.0568    | 21.942   | 2.190  | 0.030 | 4.695     | 91.419    |
| x10   | 3.1160     | 0.861    | 3.621  | 0.000 | 1.415     | 4.817     |
| x11   | -286.6455  | 208.273  | -1.376 | 0.171 | -698.242  | 124.951   |
| x12   | 231.9652   | 190.300  | 1.219  | 0.225 | -144.113  | 608.043   |

```
=====
Omnibus:              17.991    Durbin-Watson:        0.994
Prob(Omnibus):        0.000    Jarque-Bera (JB):     72.485
Skew:                 -0.066    Prob(JB):             1.82e-16
Kurtosis:             6.295    Cond. No.             3.77e+05
=====
```

```
#Removing curb-weight
```

### OLS Regression Results

```
=====
Dep. Variable:          y      R-squared:          0.869
Model:                  OLS    Adj. R-squared:       0.859
Method:                 Least Squares    F-statistic:       88.93
Date:                  Fri, 16 Oct 2020    Prob (F-statistic): 1.87e-59
Time:                  15:42:52    Log-Likelihood:    -1514.6
No. Observations:      160    AIC:              3053.
Df Residuals:          148    BIC:              3090.
Df Model:              11
Covariance Type:       nonrobust
=====
```

|       | coef       | std err  | t      | P> t  | [0.025    | 0.975]    |
|-------|------------|----------|--------|-------|-----------|-----------|
| const | -6.006e+04 | 1.67e+04 | -3.594 | 0.000 | -9.31e+04 | -2.7e+04  |
| x1    | 178.8338   | 110.135  | 1.624  | 0.107 | -38.807   | 396.475   |
| x2    | -53.8332   | 60.480   | -0.890 | 0.375 | -173.348  | 65.682    |
| x3    | 425.3626   | 277.222  | 1.534  | 0.127 | -122.461  | 973.186   |
| x4    | 158.2753   | 150.296  | 1.053  | 0.294 | -138.727  | 455.278   |
| x5    | 119.7368   | 16.301   | 7.345  | 0.000 | 87.523    | 151.950   |
| x6    | -2973.4431 | 828.249  | -3.590 | 0.000 | -4610.164 | -1336.722 |
| x7    | 337.4087   | 93.440   | 3.611  | 0.000 | 152.760   | 522.057   |
| x8    | 51.3240    | 21.016   | 2.442  | 0.016 | 9.795     | 92.853    |
| x9    | 2.9923     | 0.827    | 3.620  | 0.000 | 1.359     | 4.626     |
| x10   | -285.6199  | 207.760  | -1.375 | 0.171 | -696.179  | 124.939   |
| x11   | 208.6321   | 184.739  | 1.129  | 0.261 | -156.436  | 573.700   |

```
=====
Omnibus:                17.907    Durbin-Watson:          1.021
Prob(Omnibus):          0.000    Jarque-Bera (JB):       72.325
Skew:                   -0.043    Prob(JB):               1.97e-16
Kurtosis:               6.293    Cond. No.               3.35e+05
=====
```

```
#Removing length

OLS Regression Results
=====
Dep. Variable:          y      R-squared:                0.868
Model:                  OLS    Adj. R-squared:           0.859
Method:                 Least Squares    F-statistic:             97.88
Date:                   Fri, 16 Oct 2020  Prob (F-statistic):      2.72e-60
Time:                   15:42:52    Log-Likelihood:          -1515.0
No. Observations:       160      AIC:                     3052.
Df Residuals:           149      BIC:                     3086.
Df Model:                10
Covariance Type:        nonrobust
=====

```

|       | coef       | std err  | t      | P> t  | [0.025    | 0.975]    |
|-------|------------|----------|--------|-------|-----------|-----------|
| const | -5.875e+04 | 1.66e+04 | -3.531 | 0.001 | -9.16e+04 | -2.59e+04 |
| x1    | 133.0737   | 97.332   | 1.367  | 0.174 | -59.256   | 325.404   |
| x2    | 348.9836   | 263.423  | 1.325  | 0.187 | -171.544  | 869.511   |
| x3    | 120.7793   | 144.171  | 0.838  | 0.404 | -164.104  | 405.662   |
| x4    | 118.9571   | 16.266   | 7.313  | 0.000 | 86.814    | 151.100   |
| x5    | -2943.7248 | 826.999  | -3.560 | 0.000 | -4577.885 | -1309.564 |
| x6    | 328.8318   | 92.877   | 3.541  | 0.001 | 145.306   | 512.358   |
| x7    | 50.6306    | 20.986   | 2.413  | 0.017 | 9.161     | 92.100    |
| x8    | 3.0961     | 0.818    | 3.786  | 0.000 | 1.480     | 4.712     |
| x9    | -218.9137  | 193.637  | -1.131 | 0.260 | -601.544  | 163.716   |
| x10   | 171.0471   | 179.724  | 0.952  | 0.343 | -184.089  | 526.183   |

```
=====
Omnibus:                17.374    Durbin-Watson:           1.006
Prob(Omnibus):           0.000    Jarque-Bera (JB):        66.791
Skew:                    -0.078    Prob(JB):                3.14e-15
Kurtosis:                 6.161    Cond. No.:               3.34e+05
=====
```

```
#Removing height

OLS Regression Results

=====
Dep. Variable:          y      R-squared:          0.867
Model:                  OLS    Adj. R-squared:       0.859
Method:                 Least Squares    F-statistic:        108.9
Date:                  Fri, 16 Oct 2020    Prob (F-statistic):  3.58e-61
Time:                  15:42:52    Log-Likelihood:     -1515.4
No. Observations:      160    AIC:                3051.
Df Residuals:          150    BIC:                3082.
Df Model:               9
Covariance Type:       nonrobust
=====

```

|       | coef       | std err  | t      | P> t  | [0.025    | 0.975]    |
|-------|------------|----------|--------|-------|-----------|-----------|
| const | -5.297e+04 | 1.51e+04 | -3.502 | 0.001 | -8.29e+04 | -2.31e+04 |
| x1    | 178.8177   | 80.493   | 2.222  | 0.028 | 19.771    | 337.865   |
| x2    | 305.6295   | 258.033  | 1.184  | 0.238 | -204.219  | 815.478   |
| x3    | 117.4667   | 16.153   | 7.272  | 0.000 | 85.550    | 149.383   |
| x4    | -3048.7616 | 816.626  | -3.733 | 0.000 | -4662.338 | -1435.185 |
| x5    | 331.0819   | 92.746   | 3.570  | 0.000 | 147.825   | 514.339   |
| x6    | 50.5417    | 20.965   | 2.411  | 0.017 | 9.116     | 91.967    |
| x7    | 3.0132     | 0.811    | 3.715  | 0.000 | 1.411     | 4.616     |
| x8    | -215.8108  | 193.409  | -1.116 | 0.266 | -597.970  | 166.348   |
| x9    | 168.0934   | 179.510  | 0.936  | 0.351 | -186.602  | 522.789   |

```
=====
Omnibus:                17.709    Durbin-Watson:         1.012
Prob(Omnibus):           0.000    Jarque-Bera (JB):      70.213
Skew:                    -0.059    Prob(JB):              5.67e-16
Kurtosis:                 6.243    Cond. No.              3.04e+05
=====
```

```
#Removing highway-mpg
```

### OLS Regression Results

```
=====
Dep. Variable:          y      R-squared:          0.866
Model:                  OLS    Adj. R-squared:       0.859
Method:                 Least Squares    F-statistic:      122.5
Date:                  Fri, 16 Oct 2020    Prob (F-statistic): 4.83e-62
Time:                  15:42:52    Log-Likelihood:    -1515.9
No. Observations:      160    AIC:              3050.
Df Residuals:          151    BIC:              3077.
Df Model:               8
Covariance Type:       nonrobust
=====
```

|       | coef       | std err  | t      | P> t  | [0.025    | 0.975]    |
|-------|------------|----------|--------|-------|-----------|-----------|
| const | -4.952e+04 | 1.47e+04 | -3.377 | 0.001 | -7.85e+04 | -2.06e+04 |
| x1    | 168.3419   | 79.679   | 2.113  | 0.036 | 10.911    | 325.772   |
| x2    | 281.3419   | 256.621  | 1.096  | 0.275 | -225.690  | 788.374   |
| x3    | 115.4832   | 16.007   | 7.215  | 0.000 | 83.857    | 147.109   |
| x4    | -2979.3949 | 812.928  | -3.665 | 0.000 | -4585.577 | -1373.213 |
| x5    | 328.8373   | 92.677   | 3.548  | 0.001 | 145.726   | 511.948   |
| x6    | 52.5411    | 20.848   | 2.520  | 0.013 | 11.350    | 93.732    |
| x7    | 3.0272     | 0.811    | 3.735  | 0.000 | 1.426     | 4.629     |
| x8    | -52.2924   | 83.114   | -0.629 | 0.530 | -216.508  | 111.923   |

```
=====
Omnibus:                17.933    Durbin-Watson:          0.987
Prob(Omnibus):           0.000    Jarque-Bera (JB):       72.250
Skew:                   -0.056    Prob(JB):               2.05e-16
Kurtosis:                6.290    Cond. No.:              2.94e+05
=====
```

```
#Removing city-mpg
```

### OLS Regression Results

```
=====
Dep. Variable:          y      R-squared:          0.866
Model:                  OLS    Adj. R-squared:       0.860
Method:                 Least Squares    F-statistic:       140.5
Date:                  Fri, 16 Oct 2020    Prob (F-statistic): 4.77e-63
Time:                  15:42:52    Log-Likelihood:    -1516.1
No. Observations:      160    AIC:              3048.
Df Residuals:          152    BIC:              3073.
Df Model:              7
Covariance Type:       nonrobust
=====
```

|       | coef       | std err  | t      | P> t  | [0.025    | 0.975]    |
|-------|------------|----------|--------|-------|-----------|-----------|
| const | -5.448e+04 | 1.23e+04 | -4.414 | 0.000 | -7.89e+04 | -3.01e+04 |
| x1    | 183.9395   | 75.574   | 2.434  | 0.016 | 34.628    | 333.251   |
| x2    | 314.8122   | 250.547  | 1.256  | 0.211 | -180.193  | 809.817   |
| x3    | 112.2363   | 15.122   | 7.422  | 0.000 | 82.360    | 142.113   |
| x4    | -2984.4499 | 811.271  | -3.679 | 0.000 | -4587.273 | -1381.627 |
| x5    | 305.5303   | 84.783   | 3.604  | 0.000 | 138.026   | 473.035   |
| x6    | 60.6620    | 16.339   | 3.713  | 0.000 | 28.381    | 92.943    |
| x7    | 2.9749     | 0.805    | 3.697  | 0.000 | 1.385     | 4.565     |

```
=====
Omnibus:                17.915    Durbin-Watson:          0.991
Prob(Omnibus):          0.000    Jarque-Bera (JB):       71.425
Skew:                   -0.078    Prob(JB):               3.09e-16
Kurtosis:               6.269    Cond. No.               2.48e+05
=====
```



```
#Removing width
```

### OLS Regression Results

```
=====
Dep. Variable:          y      R-squared:          0.865
Model:                  OLS    Adj. R-squared:       0.859
Method:                 Least Squares    F-statistic:       163.0
Date:                   Fri, 16 Oct 2020    Prob (F-statistic): 7.81e-64
Time:                   15:42:52    Log-Likelihood:    -1516.9
No. Observations:      160    AIC:              3048.
Df Residuals:          153    BIC:              3069.
Df Model:               6
Covariance Type:       nonrobust
=====
```

|       | coef       | std err  | t      | P> t  | [0.025    | 0.975]    |
|-------|------------|----------|--------|-------|-----------|-----------|
| const | -4.164e+04 | 6932.165 | -6.006 | 0.000 | -5.53e+04 | -2.79e+04 |
| x1    | 250.5779   | 53.941   | 4.645  | 0.000 | 144.012   | 357.144   |
| x2    | 113.9081   | 15.092   | 7.548  | 0.000 | 84.093    | 143.723   |
| x3    | -2945.3206 | 812.205  | -3.626 | 0.000 | -4549.905 | -1340.736 |
| x4    | 321.4083   | 83.994   | 3.827  | 0.000 | 155.471   | 487.346   |
| x5    | 66.9018    | 15.595   | 4.290  | 0.000 | 36.091    | 97.712    |
| x6    | 3.0111     | 0.806    | 3.737  | 0.000 | 1.419     | 4.603     |

```
=====
Omnibus:                18.109    Durbin-Watson:          1.032
Prob(Omnibus):          0.000    Jarque-Bera (JB):       70.826
Skew:                   -0.128    Prob(JB):               4.17e-16
Kurtosis:                6.249    Cond. No.                1.39e+05
=====
```

```

#%#=====
#11: Perform a complete t-test and F-test analysis on the final model and write
down your observations
# %%-----

```

| OLS Regression Results |                  |                     |          |  |  |  |
|------------------------|------------------|---------------------|----------|--|--|--|
| Dep. Variable:         | y                | R-squared:          | 0.865    |  |  |  |
| Model:                 | OLS              | Adj. R-squared:     | 0.859    |  |  |  |
| Method:                | Least Squares    | F-statistic:        | 163.0    |  |  |  |
| Date:                  | Fri, 16 Oct 2020 | Prob (F-statistic): | 7.81e-64 |  |  |  |
| Time:                  | 15:42:52         | Log-Likelihood:     | -1516.9  |  |  |  |
| No. Observations:      | 160              | AIC:                | 3048.    |  |  |  |
| Df Residuals:          | 153              | BIC:                | 3069.    |  |  |  |
| Df Model:              | 6                |                     |          |  |  |  |
| Covariance Type:       | nonrobust        |                     |          |  |  |  |

|       | coef       | std err  | t      | P> t  | [0.025    | 0.975]    |
|-------|------------|----------|--------|-------|-----------|-----------|
| const | -4.164e+04 | 6932.165 | -6.006 | 0.000 | -5.53e+04 | -2.79e+04 |
| x1    | 250.5779   | 53.941   | 4.645  | 0.000 | 144.012   | 357.144   |
| x2    | 113.9081   | 15.092   | 7.548  | 0.000 | 84.093    | 143.723   |
| x3    | -2945.3206 | 812.205  | -3.626 | 0.000 | -4549.905 | -1340.736 |
| x4    | 321.4083   | 83.994   | 3.827  | 0.000 | 155.471   | 487.346   |
| x5    | 66.9018    | 15.595   | 4.290  | 0.000 | 36.091    | 97.712    |
| x6    | 3.0111     | 0.806    | 3.737  | 0.000 | 1.419     | 4.603     |

|                |        |                   |          |
|----------------|--------|-------------------|----------|
| Omnibus:       | 18.109 | Durbin-Watson:    | 1.032    |
| Prob(Omnibus): | 0.000  | Jarque-Bera (JB): | 70.826   |
| Skew:          | -0.128 | Prob(JB):         | 4.17e-16 |
| Kurtosis:      | 6.249  | Cond. No.         | 1.39e+05 |

An F-test is a type of statistical test that is very flexible. You can use them in a wide variety of settings. F-tests can evaluate multiple model terms simultaneously, which allows them to compare the fits of different linear models. In contrast, t-tests can evaluate just one term at a time.

In statistical output, we can find the overall ols result in the above table. Comparing the p-value for the F-test to your significance level. Here the p-value is less than the significance level, the sample data provided has sufficient evidence to conclude that your regression model fits the data better than the model with no independent variables.

This finding is good news because it means that the independent variables in your model improve the fit!

## CONCLUSION

For the data set provided we applied multi linear regression. For the very model using Least squares estimation, the sum square of errors was minimized. using datasets, we plotted various time series plots and made a comparison with respect to training, testing and different regression methods. using Coefficient of determination or R-squared measured we assessed how well a model explains and predicts future outcome. Below table summarizes the regression model and their accuracy after stepwise regression.

| OLS Regression Results |                  |                     |          |       |           |           |
|------------------------|------------------|---------------------|----------|-------|-----------|-----------|
| =====                  |                  |                     |          |       |           |           |
| Dep. Variable:         | y                | R-squared:          | 0.865    |       |           |           |
| Model:                 | OLS              | Adj. R-squared:     | 0.859    |       |           |           |
| Method:                | Least Squares    | F-statistic:        | 163.0    |       |           |           |
| Date:                  | Fri, 16 Oct 2020 | Prob (F-statistic): | 7.81e-64 |       |           |           |
| Time:                  | 17:22:49         | Log-Likelihood:     | -1516.9  |       |           |           |
| No. Observations:      | 160              | AIC:                | 3048.    |       |           |           |
| Df Residuals:          | 153              | BIC:                | 3069.    |       |           |           |
| Df Model:              | 6                |                     |          |       |           |           |
| Covariance Type:       | nonrobust        |                     |          |       |           |           |
| =====                  |                  |                     |          |       |           |           |
|                        | coef             | std err             | t        | P> t  | [0.025    | 0.975]    |
| -----                  |                  |                     |          |       |           |           |
| const                  | -4.164e+04       | 6932.165            | -6.006   | 0.000 | -5.53e+04 | -2.79e+04 |
| x1                     | 250.5779         | 53.941              | 4.645    | 0.000 | 144.012   | 357.144   |
| x2                     | 113.9081         | 15.092              | 7.548    | 0.000 | 84.093    | 143.723   |
| x3                     | -2945.3206       | 812.205             | -3.626   | 0.000 | -4549.905 | -1340.736 |
| x4                     | 321.4083         | 83.994              | 3.827    | 0.000 | 155.471   | 487.346   |
| x5                     | 66.9018          | 15.595              | 4.290    | 0.000 | 36.091    | 97.712    |
| x6                     | 3.0111           | 0.806               | 3.737    | 0.000 | 1.419     | 4.603     |
| =====                  |                  |                     |          |       |           |           |
| Omnibus:               | 18.109           | Durbin-Watson:      | 1.032    |       |           |           |
| Prob(Omnibus):         | 0.000            | Jarque-Bera (JB):   | 70.826   |       |           |           |
| Skew:                  | -0.128           | Prob(JB):           | 4.17e-16 |       |           |           |
| Kurtosis:              | 6.249            | Cond. No.           | 1.39e+05 |       |           |           |
| -----                  |                  |                     |          |       |           |           |

## CHALLENGE

Calculations was little tricky to understand in the beginning, after lot of clarifications it provided clarity.

## APPENDIX

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
import statsmodels.api as sm
import warnings
import seaborn as sns

from Autocorrelation import cal_auto_corr
from Pearson_Correlation_Coefficient import correlation_coefficient_cal

warnings.filterwarnings("ignore")
np.set_printoptions(suppress=True)

df = pd.read_csv('autos.clean_corr.csv')
df = df[['price', 'normalized-losses', 'wheel-base', 'length', 'width', 'height',
        'curb-weight', 'engine-size',
        'bore', 'stroke', 'compression-ratio', 'horsepower', 'peak-rpm', 'city-
        mpg', 'highway-mpg']]

features = df.drop(columns='price').to_numpy()
target = df['price'].to_numpy()

features = sm.add_constant(features)
x_train, x_test, y_train, y_test = train_test_split(features, target,
                                                    shuffle=False, test_size=0.2)

corr = df.corr()
ax = sns.heatmap(corr, vmin=-1, vmax=1, center=0, square=True)
bottom, top = ax.get_ylim()
ax.set_ylim(bottom + 0.5, top - 0.5)
ax.set_xticklabels(ax.get_xticklabels(), rotation=45, horizontalalignment='right')
plt.figure(figsize=(15,10))
plt.show()

def plot_fun(train, test, predicted, title):
    plt.figure(figsize=(15,10))
    plt.plot(range(0, len(train)), train, label='Training set')
    plt.plot(range(len(train), len(train)+len(test)), test, label='Testing set')
    plt.plot(range(len(train), len(train)+ len(predicted)), predicted,
    label='prediction values')
    plt.xlabel('')
    plt.title(title)
    plt.legend(loc='upper left')
    plt.show()

x_transpose = np.transpose(x_train)
beta_hat = np.matmul(np.matmul(np.linalg.inv(np.matmul(x_transpose, x_train)),
x_transpose), y_train)
print('\n')
```

```

print('Unknown Coefficients:',beta_hat)

y_hat = np.matmul(x_train, beta_hat)
eps = y_train - y_hat
y_test_hat = np.matmul(x_test, beta_hat)
plot_fun(y_train, y_test, y_test_hat, 'LSE Method')

print('\n')
model = sm.OLS(y_train, x_train).fit()
print(model.summary())

y_hat_OLS = model.predict(x_train)
eps_OLS = y_train - y_hat_OLS
y_test_hat_OLS = model.predict(x_test)
plot_fun(y_train, y_test, y_test_hat_OLS, 'OLS Method')

prediction_error = y_train - y_hat_OLS
forecast_error = y_test - y_test_hat_OLS
lags = 20
prediction_error_acf = cal_auto_corr(prediction_error, lags)
forecast_error_acf = cal_auto_corr(forecast_error, lags)

plt.figure(figsize=(15, 10))
plt.stem(range(-(lags-1),lags), prediction_error_acf, use_line_collection=True)
plt.xlabel('Lags')
plt.ylabel('Magnitude')
plt.title('Autocorrelation plot for Prediction Error')
plt.show()

plt.figure(figsize=(15, 10))
plt.stem(range(-(lags-1),lags), forecast_error_acf, use_line_collection=True)
plt.xlabel('Lags')
plt.ylabel('Magnitude')
plt.title('Autocorrelation plot for Forecast Error')
plt.show()

print('\n')
T = len(x_train)
K = len(x_train[0, 1:])
pred_var = (1/(T-K-1)) * (np.sum((prediction_error)**2))
print("variance of prediction error: ", pred_var)

T = len(x_test)
K = len(x_test[0, 1:])
forecast_var = (1/(T-K-1)) * (np.sum((forecast_error)**2))
print("variance of forecast error: ", forecast_var)
print('\n')

corr_coeff = round(correlation_coefficient_cal(y_test, y_test_hat_OLS),2)
print('correlation coefficient between y-test and  $\hat{y}_t$ :',corr_coeff)
print('\n')

plt.figure(figsize=(15, 10))
plt.scatter(y_test, y_test_hat_OLS, c='green', alpha=1, label='y_test vs y_test_hat_OLS')

```

```

plt.xlabel('y_test')
plt.ylabel('y_test_hat_OLS')
plt.title("Scatter plot of y_test vs y_hat_test with correlation coefficient of
{}".format(corr_coeff))
plt.legend()
plt.show()

# Removing bore
df = pd.read_csv('autos.clean_corr.csv')
df = df[['price', 'normalized-losses', 'wheel-base', 'length', 'width', 'height',
'curb-weight', 'engine-size',
'stroke', 'compression-ratio', 'horsepower', 'peak-rpm', 'city-
mpg', 'highway-mpg']]

features = df.drop(columns='price').to_numpy()
target = df['price'].to_numpy()
features = sm.add_constant(features)
x_train, x_test, y_train, y_test = train_test_split(features, target,
shuffle=False, test_size=0.2)
model = sm.OLS(y_train, x_train).fit()
print(model.summary())

#Removing normalized losses
df = pd.read_csv('autos.clean_corr.csv')
df = df[['price', 'wheel-base', 'length', 'width', 'height', 'curb-weight',
'engine-size', 'stroke',
'compression-ratio', 'horsepower', 'peak-rpm', 'city-mpg', 'highway-mpg']]
features = df.drop(columns='price').to_numpy()
target = df['price'].to_numpy()
features = sm.add_constant(features)
x_train, x_test, y_train, y_test = train_test_split(features, target,
shuffle=False, test_size=0.2)
model = sm.OLS(y_train, x_train).fit()
print(model.summary())

#Removing curb-weight
df = pd.read_csv('autos.clean_corr.csv')
df = df[['price', 'wheel-base', 'length', 'width', 'height', 'engine-size',
'stroke',
'compression-ratio', 'horsepower', 'peak-rpm', 'city-mpg', 'highway-mpg']]
features = df.drop(columns='price').to_numpy()
target = df['price'].to_numpy()
features = sm.add_constant(features)
x_train, x_test, y_train, y_test = train_test_split(features, target,
shuffle=False, test_size=0.2)
model = sm.OLS(y_train, x_train).fit()
print(model.summary())

#Removing length
df = pd.read_csv('autos.clean_corr.csv')
df = df[['price', 'wheel-base', 'width', 'height', 'engine-size',
'stroke', 'compression-ratio',
'horsepower', 'peak-rpm', 'city-mpg', 'highway-mpg']]
features = df.drop(columns='price').to_numpy()
target = df['price'].to_numpy()
features = sm.add_constant(features)
x_train, x_test, y_train, y_test = train_test_split(features, target,
shuffle=False, test_size=0.2)

```

```

model = sm.OLS(y_train, x_train).fit()
print(model.summary())

#Removing height
df = pd.read_csv('autos.clean_corr.csv')
df = df[['price', 'wheel-base', 'width', 'engine-size', 'stroke', 'compression-
ratio', 'horsepower', 'peak-rpm',
        'city-mpg', 'highway-mpg']]
features = df.drop(columns='price').to_numpy()
target = df['price'].to_numpy()
features = sm.add_constant(features)
x_train, x_test, y_train, y_test = train_test_split(features, target,
shuffle=False, test_size=0.2)
model = sm.OLS(y_train, x_train).fit()
print(model.summary())

#Removing highway-mpg
df = pd.read_csv('autos.clean_corr.csv')
df = df[['price', 'wheel-base', 'width', 'engine-size', 'stroke', 'compression-
ratio', 'horsepower', 'peak-rpm',
        'city-mpg']]
features = df.drop(columns='price').to_numpy()
target = df['price'].to_numpy()
features = sm.add_constant(features)
x_train, x_test, y_train, y_test = train_test_split(features, target,
shuffle=False, test_size=0.2)
model = sm.OLS(y_train, x_train).fit()
print(model.summary())

#Removing city-mpg
df = pd.read_csv('autos.clean_corr.csv')
df = df[['price', 'wheel-base', 'width', 'engine-size', 'stroke', 'compression-
ratio', 'horsepower', 'peak-rpm']]
features = df.drop(columns='price').to_numpy()
target = df['price'].to_numpy()
features = sm.add_constant(features)
x_train, x_test, y_train, y_test = train_test_split(features, target,
shuffle=False, test_size=0.2)
model = sm.OLS(y_train, x_train).fit()
print(model.summary())

#Removing width
df = pd.read_csv('autos.clean_corr.csv')
df = df[['price', 'wheel-base', 'engine-size', 'stroke', 'compression-ratio',
'horsepower', 'peak-rpm']]
features = df.drop(columns='price').to_numpy()
target = df['price'].to_numpy()
features = sm.add_constant(features)
x_train, x_test, y_train, y_test = train_test_split(features, target,
shuffle=False, test_size=0.2)
model = sm.OLS(y_train, x_train).fit()
print(model.summary())

```

```

#AUTO CORRELATION
import numpy as np

def correlation_coefficient_cal(x, y):
    result = 0
    cov_res = 0
    var_res1 = 0
    var_res2 = 0
    mean_x = np.mean(x)
    mean_y = np.mean(y)
    if len(x) == len(y):
        for i in range(0, len(x)):
            cov_res += ((x[i]-mean_x)*(y[i]-mean_y))
            var_res1 += (x[i]-mean_x)**2
            var_res2 += (y[i]-mean_y)**2
    result += cov_res/(np.sqrt(var_res1)*np.sqrt(var_res2))
    return result

#PEARSON CORRELATION COEFFICIENT
import numpy as np
def correlation_coefficient_cal(x, y):
    result = 0
    cov_res = 0
    var_res1 = 0
    var_res2 = 0
    mean_x = np.mean(x)
    mean_y = np.mean(y)
    if len(x) == len(y):
        for i in range(0, len(x)):
            cov_res += ((x[i]-mean_x)*(y[i]-mean_y))
            var_res1 += (x[i]-mean_x)**2
            var_res2 += (y[i]-mean_y)**2
    result += cov_res/(np.sqrt(var_res1)*np.sqrt(var_res2))
    return result

```



## REFERENCES

<https://otexts.com/fpp2/#>