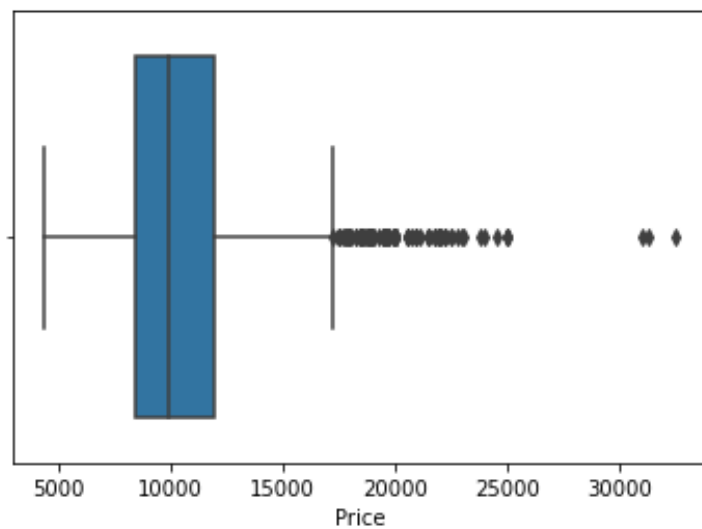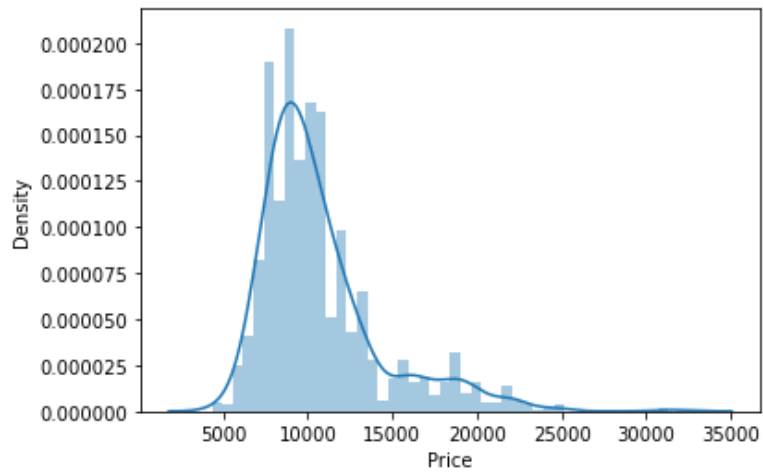1. Variables:

- Numerical variable: Price, Age_08_04, KM, HP, Quarterly Tax, Weight, CC

- Categorical variable: doors, gears

2. Distribution of Price (output)
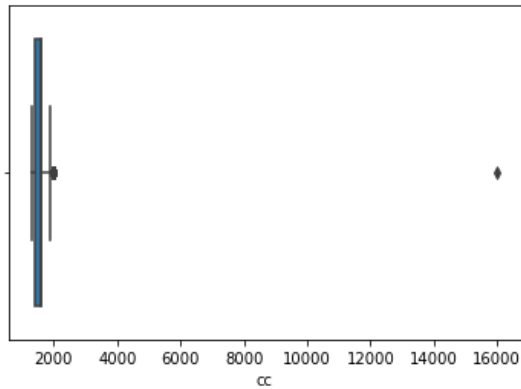

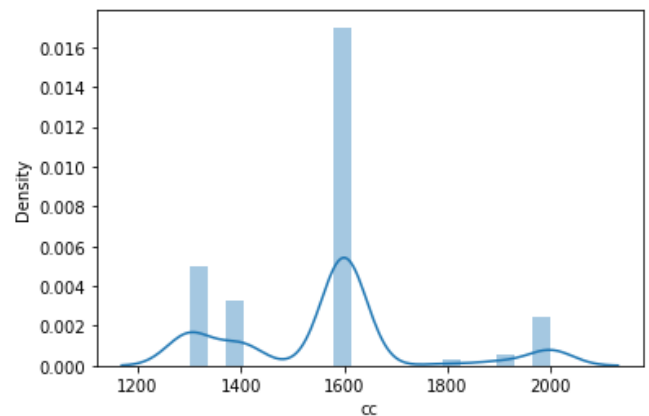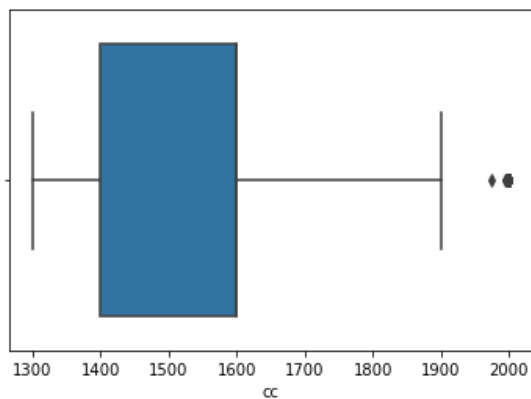Distribution slightly skewed to the right, with outliers

Price > RM 30000 for sports cars

3. Distribution of numerical variables

i) CC

For variable 'CC', there is extreme value of 16000CC, which is mistyped into data frame, so the value will be replaced with 1600CC, which is also the median value of 'CC'.

ii) Age – Highest frequency around 60-80
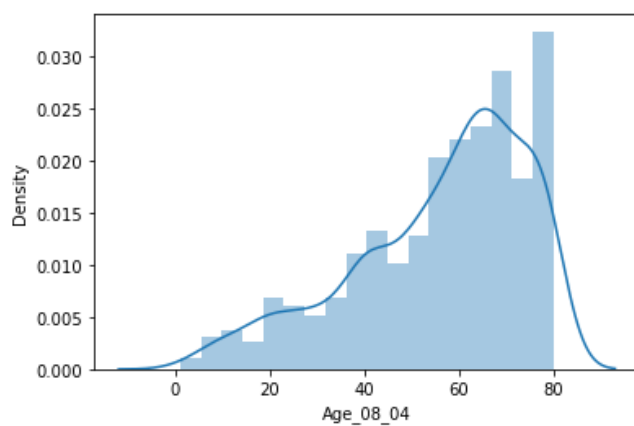
iii) KM – Skewed to the right
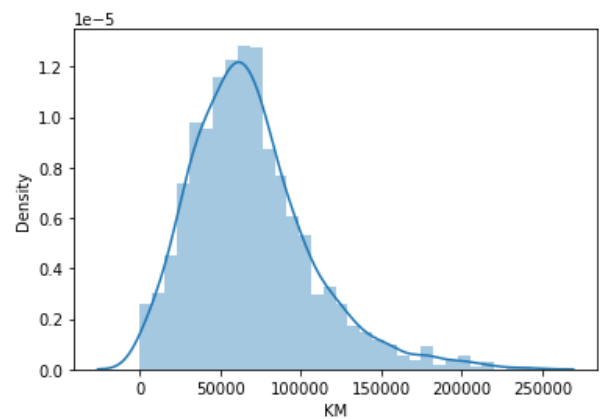
iv) HP – outlier value (around 190 due to sports car)
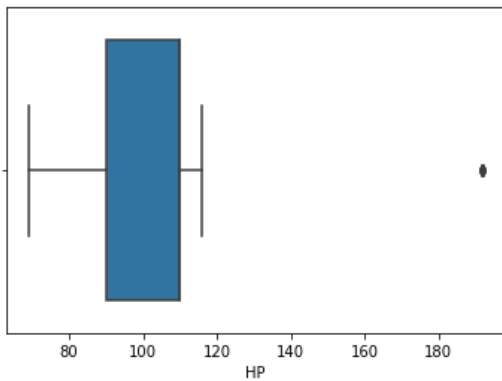
v) Quarterly Tax
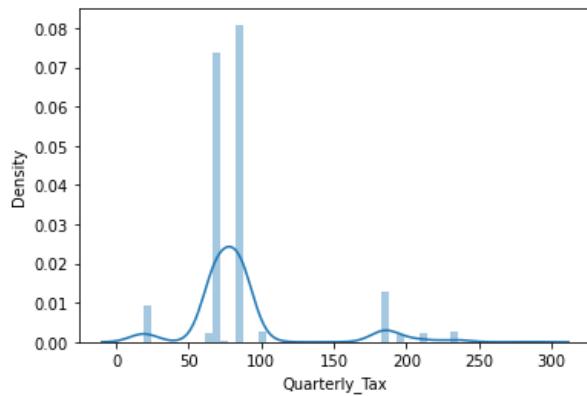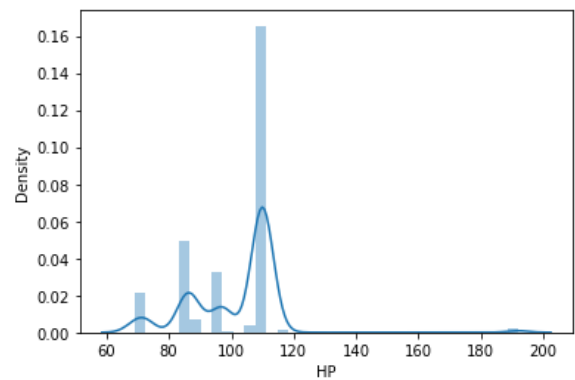
Mode at 85, highest tax at 280

vi) Weight
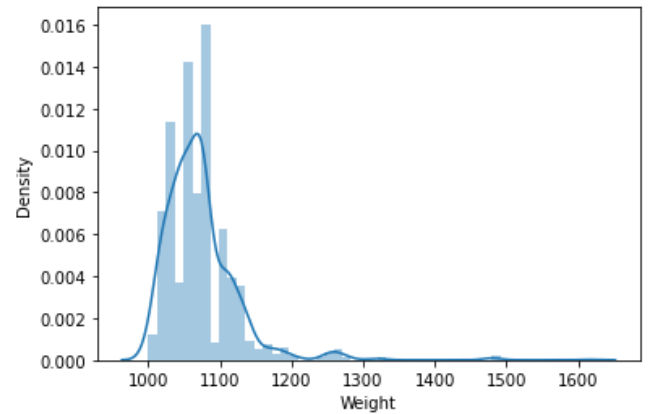
4. Multicollinearity between variables based on heatmap:
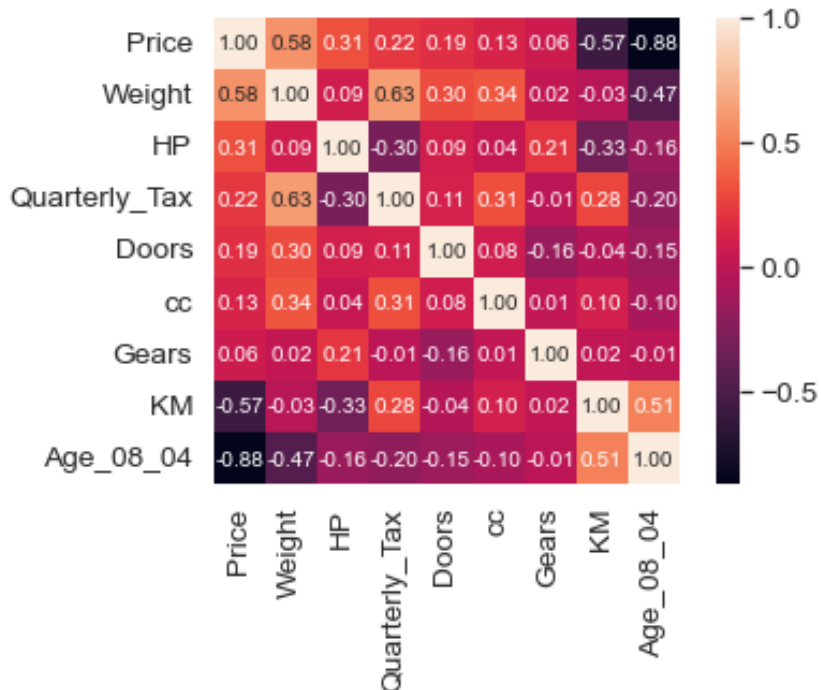
- Price is highly correlated with Age
- Correlation between Weight and Quarterly Tax is 0.63
- Correlation between Price and Weight is 0.58
- Correlation between Price and KM is -0.57
- Correlation between Age and KM is -0.51
- Gears and Price have lowest correlation
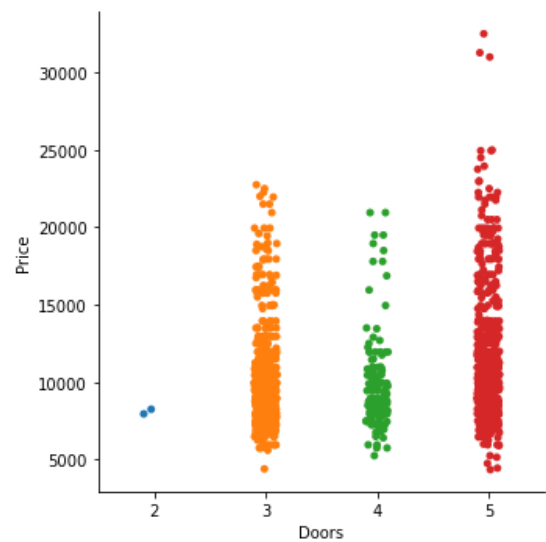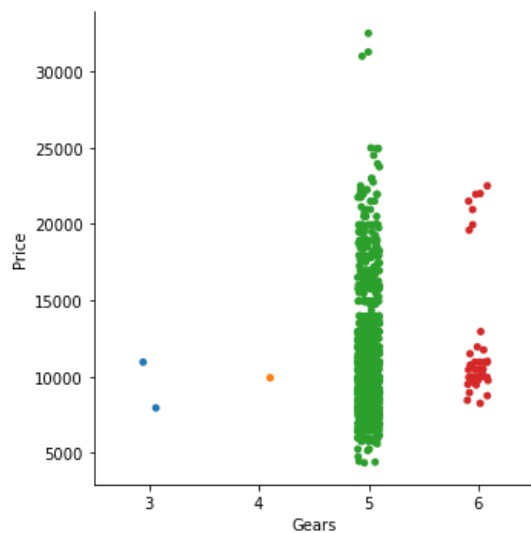
i) Cars with 5 gears are the most cars sold, followed by cars with 6 gears.

ii) Cars with 5 doors sold with higher price range (5000 – 30000), followed by cars with 3 doors (5000 – 20000+).

Based on observation, correlation between independent variables are low. All features do have correlation with dependent variable (Price).

R^2 (after replacing outlier in CC) = 0.869

Scatter plot of Price against Age – Apply polynomial transformation on Age
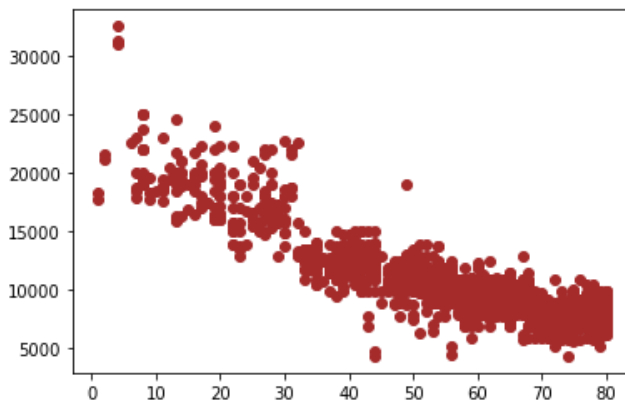
R^2 (after polynomial transformation on Age) = 0.888

Consider only the below columns and prepare a prediction model for predicting Price.

Corolla<-Corolla[c("Price","Age_08_04","KM","HP","cc","Doors","Gears","Quarterly_Tax","Weight")]

test_rmse

1183.3255007296425

train_rmse

1223.9115702176152

Index	Price
399	9111.945280461656
478	10239.597798584455
1218	8995.842362382136
1365	8052.307490644231
445	10827.146493089658
1167	7200.075859881825
26	15945.636854085635
102	18072.995688475534
610	6565.988408446901
819	9767.250919482987
247	11984.32240909592
222	12450.287167134387
837	8502.658997632278
1066	7226.4489049973035
1041	10235.94261074131
880	9950.111122674425
1207	8074.12524631118
851	9599.01457645599
58	17758.641673675185
226	12577.871390523364
453	10530.44795692099

1077	6604.609098316292
528	10509.113318483483
657	8482.33168453392
1379	8486.239642452747
1168	7006.244856719559
628	8361.94773575765
931	9638.92269263697
385	9824.21197347961
682	9169.712301508589
1047	8211.893399927376
254	12289.117618250106
143	17858.57601629479
1337	8304.563009521295
483	10033.954975875018
1004	10479.472724999934
301	12198.608864104612
637	8496.52383683677
903	9230.031117948394
1302	8314.502093112373
16	19286.26642904436
390	9129.57953530231
940	9476.112131719014
466	10634.176702956414
969	9317.863271453074
156	18234.9662782293
731	9082.447084388892
65	14819.16331159496
252	12347.587239549037
1243	8682.754866596824
924	8992.077736848032
1375	7353.934440481191

442	10755.293308791845
488	10063.808950697916
910	9718.191780346362
830	9785.959757069346
1059	6295.014100801207
474	10758.316663637655
856	9538.190989179728
1268	7638.510322688788
59	15528.05707700845
1275	8130.509178712502
1013	8832.885013674255
285	13070.42930362228
619	7795.437351022572
234	12506.325761494489
751	10383.461937424285
1211	7076.301883316943
127	17071.949416482992
231	13089.119966674536
911	9960.680687787382
1371	8406.900580128346
1271	8627.061412201037
932	10741.965021089534
1082	7175.906041233176
1319	9364.47756191197
502	10282.006671916799
111	28600.71404931112
1317	7943.619341610418
1223	8639.414601605502
868	9372.542243888318
1015	11161.7909287558
1177	7366.168610141462

394	9858.596445915393
316	12572.081910876983
683	9951.275777527386
1112	8021.656637667202
1212	8901.252742558145
2	16716.57875435878
1359	9136.950602098586
757	9197.424316642524
434	10138.55105594049
732	8591.51315694253
651	8330.298519928661
618	9017.711690421831
675	8467.533570099678
116	22594.99316844277
53	19736.49965739942
1278	8592.863676727085
591	11254.389644038503
1049	5771.043595799914
1343	8633.254723230182
1035	10704.518907173524
7	16360.183508949593
645	8457.800208901177
105	17446.5085567649
1130	7907.426503637242
529	11162.910135697639
126	15905.819823924308
922	9473.867919702407
323	11758.852477134333
250	13837.815630438372
124	18056.207311898062
71	14142.87204475232

1057	5995.617018757521
266	12368.45768397244
1053	6350.906535961714
978	9502.963209835169
644	8853.661975820127
726	9476.069437478018
831	9046.791846307622
914	9913.501391966325
668	8060.480930507441
537	10799.320745424482
198	10540.926049897811
1136	6895.926267768142
575	11649.179825144834
38	14840.764612384168
719	8541.969144633485
262	11773.846103862303
180	19312.39996248355
1179	7848.828326243107
1023	9719.17009431784
1341	8917.241709353013
173	20091.529348999808
979	8670.077700376802
364	13599.57471287625
814	9501.916697177632
1203	8122.988301672984
1242	8544.605971385296
17	16473.80839475222
555	11365.973187751495
542	10438.209005116958
1360	8723.560179720862
217	11194.77020909735

426	10341.316006846544
139	17665.948067610872
279	12846.209436545572
614	9030.577475107963
1304	8486.040255570897
181	20487.906874375527
976	9334.631160772362
1273	7035.580966833157
968	9875.775713043053
287	12318.40001162681
1323	8844.576009355766
1111	8562.258830208719
855	8670.990243405553
271	13167.672434056365
701	8991.222216253434
977	9179.62412351386
106	17535.647304534257
131	17160.65885737593
1146	7941.933173121208
482	10913.695087161785
9	14604.748491609564
531	10184.204234079338
1182	8791.329012242752
46	17252.515433638957
205	10925.544669800427
958	8577.139940948127
985	10469.228446319286
1153	7915.118706561185
915	10778.377915724786
1272	9468.464229987185
1241	9045.934135817575

900	8436.241455547544
1024	11011.212153403108
107	17921.167103121235
690	7617.589642502078
1267	9044.32518785437
298	13757.588625517157
1187	8070.441214063521
584	10453.770287062238
85	16398.311836195004
578	10642.031611545368
392	10243.2166228346
1279	7526.547636785675
1294	8678.977661964444
257	11178.454351840875
766	9273.39841141154
96	17179.050359707915
558	12132.805445221604
22	15393.999270436245
1410	7665.414963048157
1229	8429.866386427653
841	7935.876634718876
280	13359.860680836222
305	12566.637242066046
1409	9385.309481589607
1382	9068.803566151117
1122	9273.859701978261
909	8452.677338888076
1062	5831.675862716451
1230	8477.791538086556
1355	8962.396566225532
1036	9615.02710832392

294	12760.507895810224
274	12446.014993205565
313	12887.578369045836
1420	9526.335982004643
379	7964.136999322071
1038	9814.56003063079
1254	8165.324681604343
617	7433.691004271393
200	11476.543330038909
937	9627.860323712111
8	20061.686377043196
345	13305.547877002864
1303	8725.138808690268
27	15471.415849731291
123	17391.192611145438
310	13114.444122051731
663	8175.742200489671
1431	8965.037034965735
807	7982.611633001026
906	9288.677605462553
697	9003.508313523991
626	8884.663248248376
253	12980.7811412835
177	20532.103054589068
1419	8959.359937844998
993	10858.100917130154
1257	7600.097118631069
1095	7427.799144774712
272	12888.018611815154
517	9745.74729540461
411	10081.649453629663

556	12290.862725768662
735	9679.604317453288
1092	7589.966642289518
1110	7627.827074798977
708	9043.261945301054
1165	8831.995404407935
606	7454.343210842139
580	10958.486808771973
755	8328.041212914419
1009	10175.625488269594
1265	7961.713210584076
660	7701.692755687771
365	12086.706192735193
1309	8948.83789080368
278	11435.278684581357
887	9089.41472282258
727	9324.267190835828
472	9956.316237987936
566	10354.024266525474
615	7356.003394764581
56	15417.745982805824
77	16076.953375690293
221	21481.124413289443
860	8589.725985635121
882	8773.618737721574
265	13145.456712072222
1331	7679.892382392263
191	8834.604248842317
1014	10903.610780201476
1434	8795.494379915057
954	9497.97239082087

1052	6808.325327701425
79	15427.646416241996
45	18166.358512802522
1105	7438.865063441139
429	10945.14640299367
410	9836.08553857342
1139	7901.406103929876
536	10806.268596448383
455	9727.06370619535
193	11324.173063560509
693	7789.905458082558
1134	7983.612834024885
1174	7784.469045642534
499	9966.64569032519
422	10329.820801939402
827	10187.81500175202
858	10016.369429219054
317	12292.14518098576
1055	6774.007876099292
275	12739.053777675264
1108	7703.943696367795
509	11638.72561280433
742	9524.900118355672
787	9366.841790453223
1151	7531.021458441763
1370	9085.078749405475
1005	9919.682501364141
745	8176.832814285255
553	11351.088824518834
325	12966.666892401518
885	10076.776415396964

147	19183.12397317453
1417	9290.570885447387
658	8651.920076517692
175	20505.244773527287
471	9517.2816971026
1351	8943.667325273625
695	8491.884170416131
1256	7436.792315387727
363	13288.713926641914
1021	8410.025436715652
744	8048.600123980792
100	17473.83070381575
368	12882.929956369873
711	8805.214911290339
54	16972.494105233483
144	18746.092941700623
1239	7288.904146519109
1143	8853.694738743521
935	10653.50485894925
743	9456.41063260919
1027	10975.35662017393
508	11433.466044856941
677	8693.523699670093
286	12927.022195121108
72	16056.81480266586
696	8744.479489755364
81	15779.349918560612
1310	8891.189884936037
1407	9811.42859658213
414	10147.307794676504
905	9275.878634925884

146	18436.229675891147
1430	8167.480860281759
1353	8958.243546363272
1196	7423.302059388481
157	19507.429942934483
510	10907.925917781542
473	9949.762747660123
842	10726.217976613974
1219	8280.95070718951
852	9254.183911296815
391	9672.643861991712
360	12668.603331503793
211	14053.303162840493
1001	9192.667633306895
135	17395.70220081323
4	15766.062087496734
284	12774.227250010863
1107	7958.928793948962
853	9455.232399621977
810	8939.05654588352
1018	10087.452487149289
590	11393.502175931784
638	9145.418258518741
351	12633.99418795419
532	10737.722048723728
292	13134.021451275916
1100	7561.998293079272
1295	7941.850269610957
92	20241.809081977226
1390	9259.79469107911
341	12513.770369132173

1067	6589.162386581762
1071	7842.538043953843
1233	7733.702932544689
57	15783.782836516972
938	9193.52930227602
793	9760.755230535458
661	9611.774219175739
642	7788.4034692229525
302	12235.03386711246
462	11450.185233156022
291	13101.579053480322
1220	8705.942558548142
799	9017.081670998556
1169	8794.040969952832
403	10336.672664831465
733	8373.42884791973
809	8299.700144592933
1030	8470.500621579624
803	9618.737587915639
782	10096.609798717618
1352	9350.971928111283
138	21829.6874032644
90	16659.29455164037
133	17964.840297592
862	9780.70225343489
1056	7319.290572973601
1081	7988.416151791696
31	16484.76646799529
1358	8816.3255560685
894	9453.049620210122
32	15642.07935721706

1106	7364.916846249278
1096	7485.330587009168
194	11292.328488178207
1094	7400.532241436191
1320	8745.779233048686
261	12971.677606206964
435	10655.052798517841
1017	9459.360819845138
1026	11410.525130992575
547	9977.742276795429
792	9286.866893410608
1347	8636.45980618214
342	12545.33766255901
704	7854.097471336816
428	10937.286504425221
764	9612.36451156953
140	17283.569939328107
218	10790.36972881473
1046	6022.137057332595
1127	8509.418420413353
511	11415.423716638554
51	14829.441641464065
884	10445.972171302521
1259	9066.12605892603
507	10554.736574059802
734	9533.230567204104
269	12250.869329712936
1398	9387.623124881055
236	12095.940846510079
538	9829.23900481213
634	10552.308766230592

1114	8113.0400929205825
1069	7246.976709501452
605	5670.177323586093
877	10434.361345000754
1102	8031.838358036875
934	8508.941254887824
1128	8122.612854742565