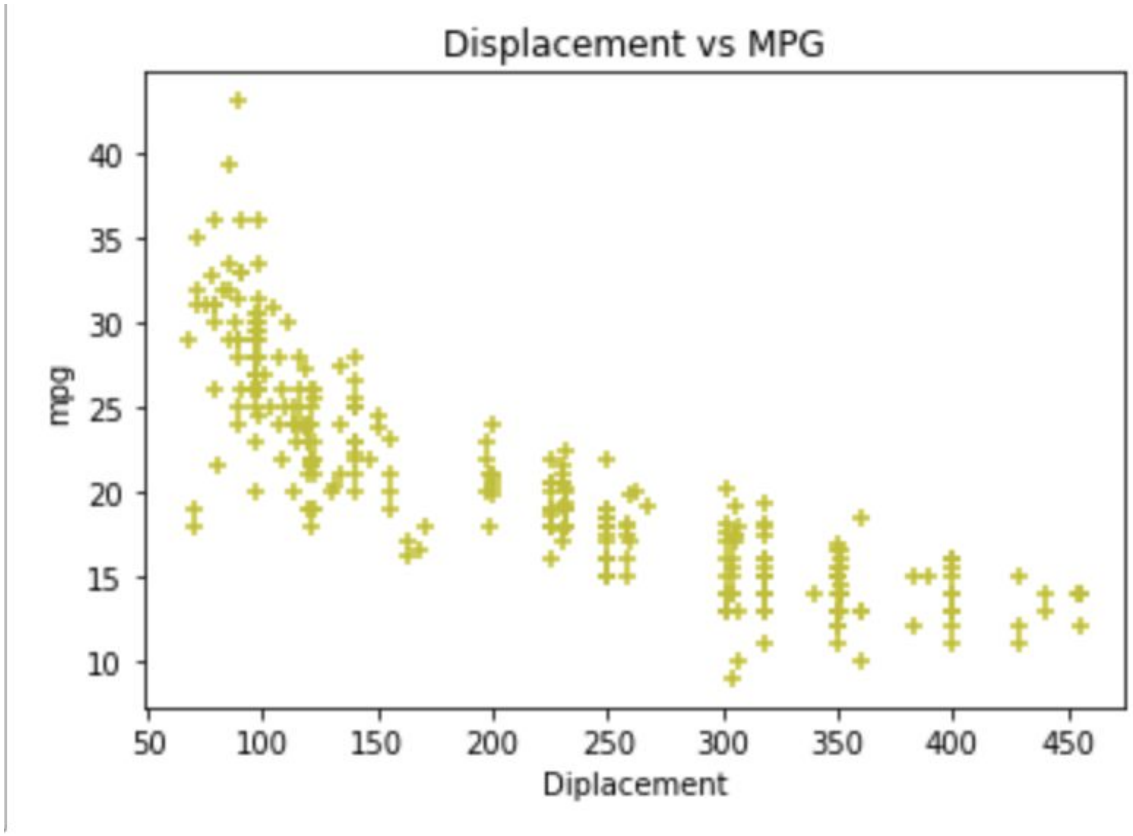
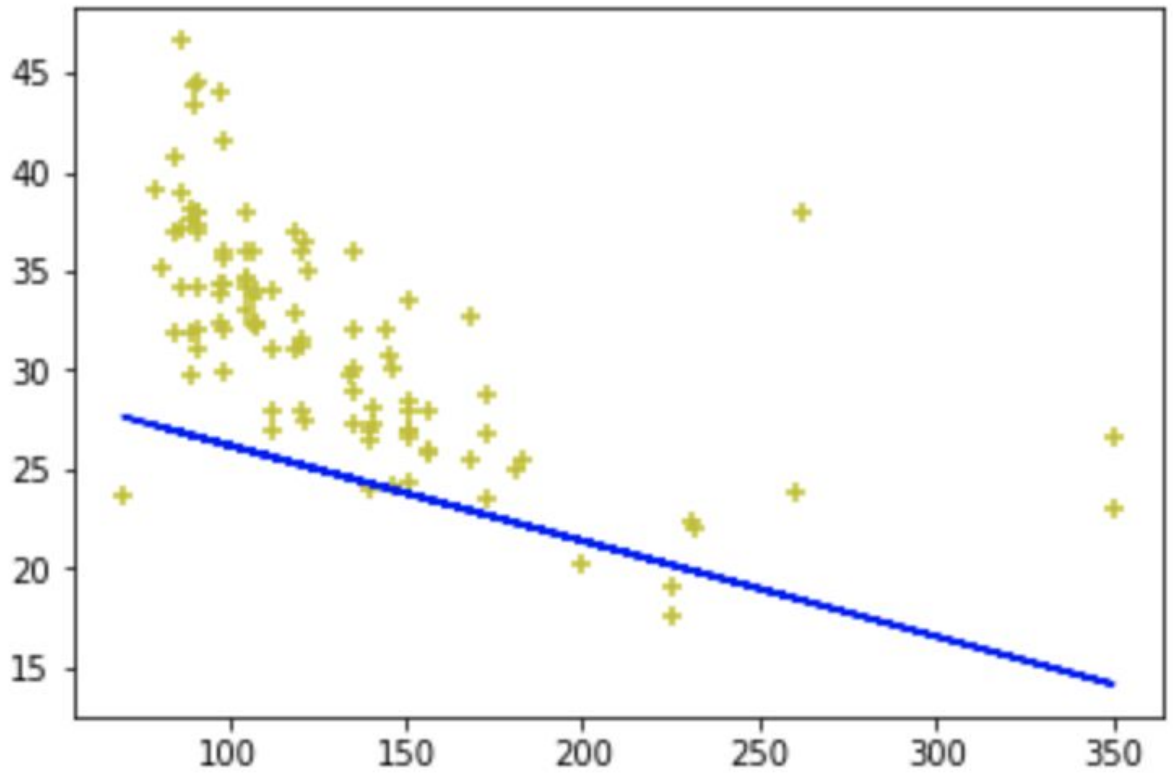
Q1)
Ignoring horsepower, the plot is

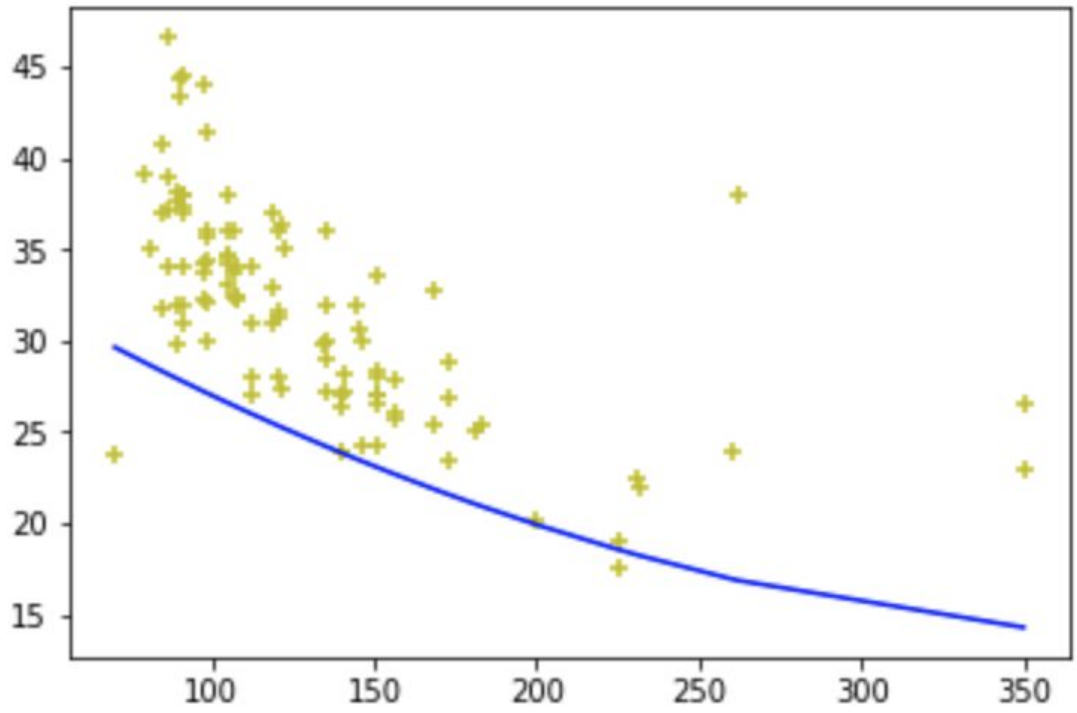


Q2) Training Error: 1557.33
Test Error = 3565.76

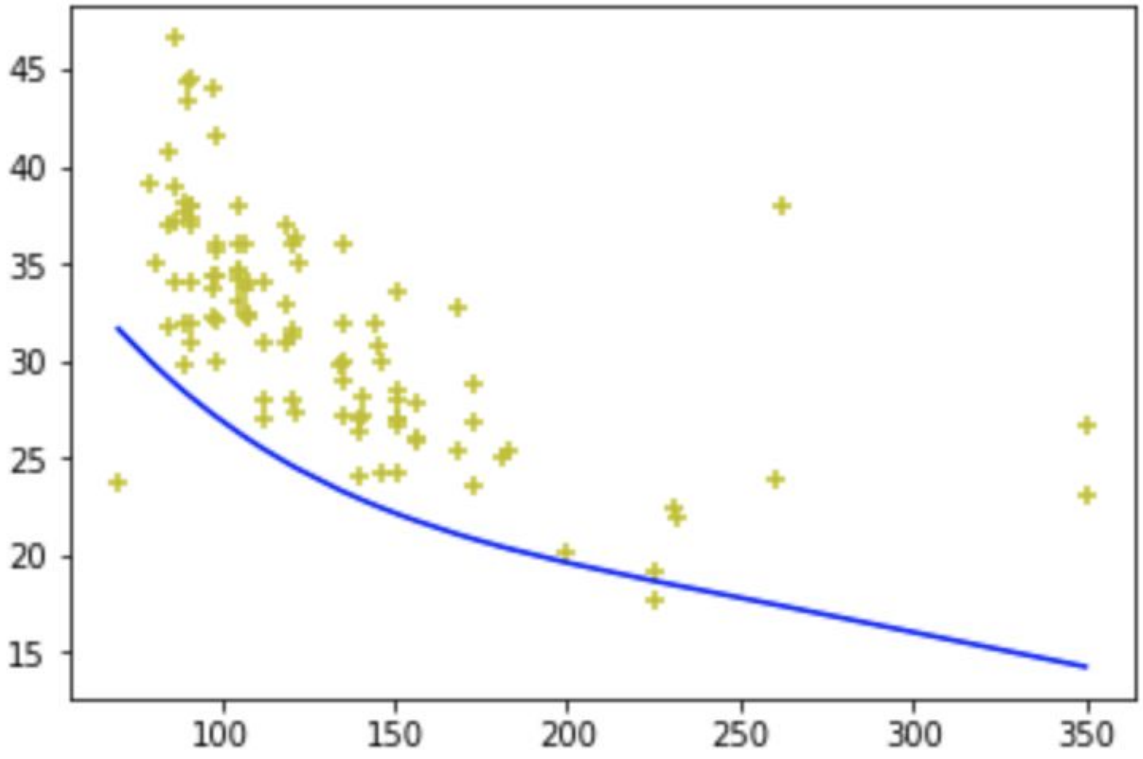


Q3)
Training Error for degree 2 is: 1296.01058799
Testing Error for degree 2 is: 3282.09475116
Training Error for degree 4 is: 1242.23889296
Testing Error for degree 4 is: 3411.64000136
Training Error for degree 6 is: 1219.71902014
Testing Error for degree 6 is: 3445.12492549

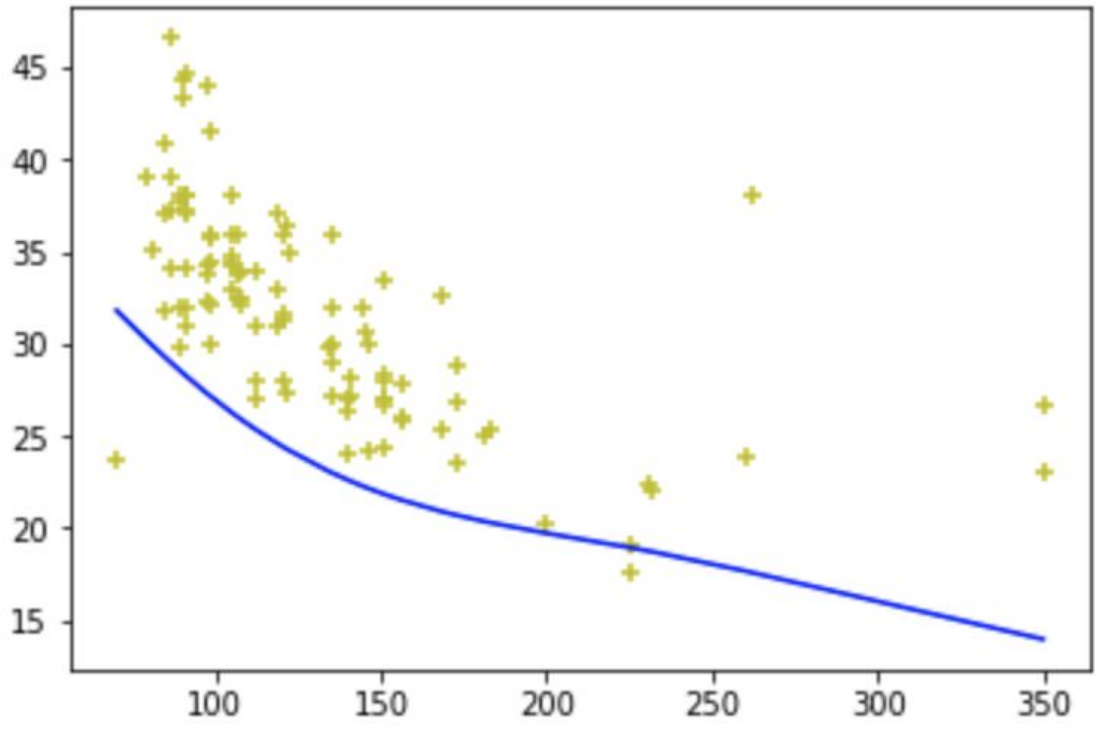
Plots for each are
Degree 2


45
40
35
30
25
20
15
100
150
200
250
300
350

Degree 4


45
40
35
30
25
20
15
100
150
200
250
300
350

Degree 6



Here, we can see evidence of overfitting. As the degree of polynomial increases, the training error keeps decreasing whereas the test error keeps increasing. Thus, as degree of polynomial increases, it performs better and better in training dataset but worse in test data set. Hence it overfits.

Q4)
Using multiple linear regression, we get the testing error to be
Testing Error: 3443.39

Q5)

Test Error for Euclidean Distance with K = 1 is: 2868.005
Test Error for Euclidean Distance with K = 3 is: 2794.73
Test Error for Euclidean Distance with K = 20 is: 2746.1914125

Please find the implementation in the jupyter notebook attached

Q6)

K = 20 performed better in knn algorithm, i.e lowest error on the test dataset.

When the value of k is small, the model is exhibits low bias but high variance.
When k is too large, the model shows high bias but low variance.

K = 20 seems optimal choice for the given dataset.

Q7)
Implemented manhattan distance (code in jupyter notebook)
Took average of the results.

K = 1 test error : 2838.83
K = 3 test error : 2676.45277778
K = 20 test error : 2800.3250375

Note : Performed better than euclidean distance.