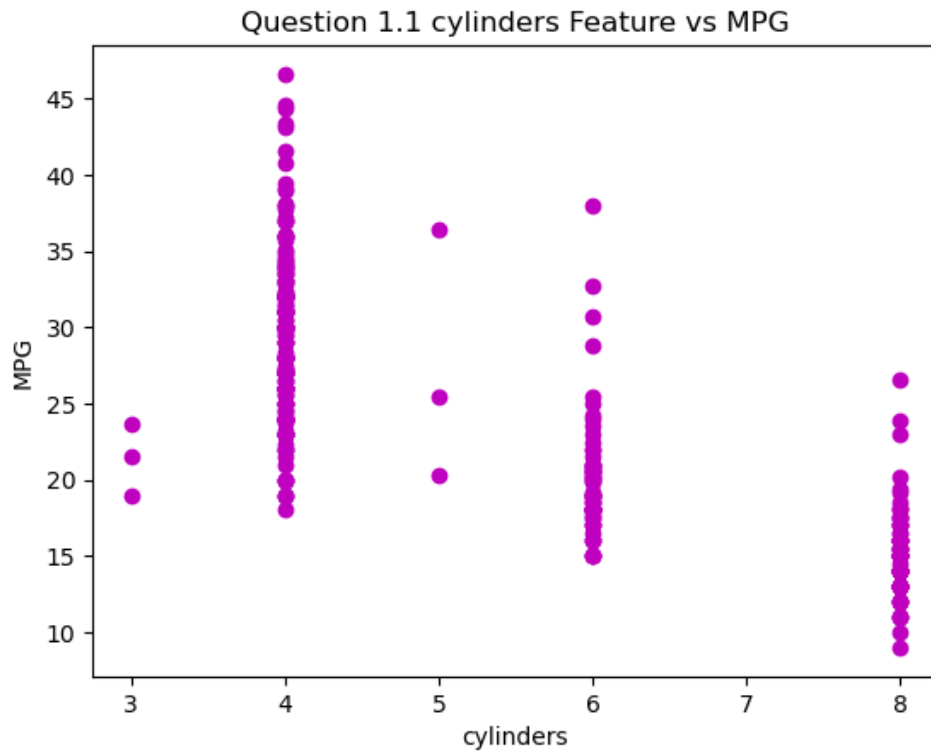


Thanya Nguyen

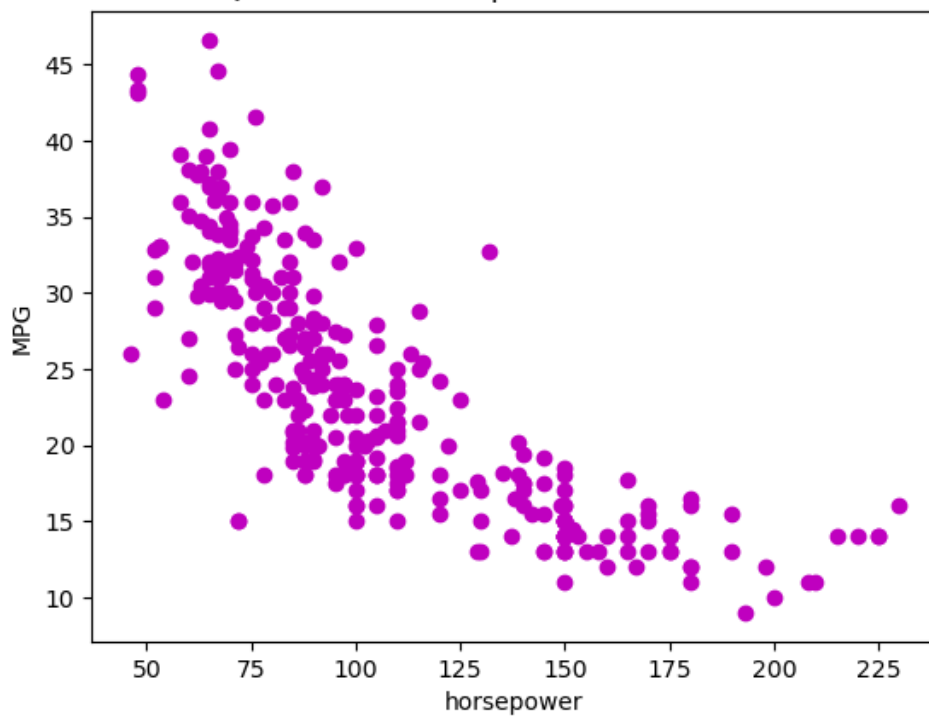
CS 425 – Lab 4

1.1 - In the report, note which of the features you believe is best for predicting MPG and why you selected that feature.

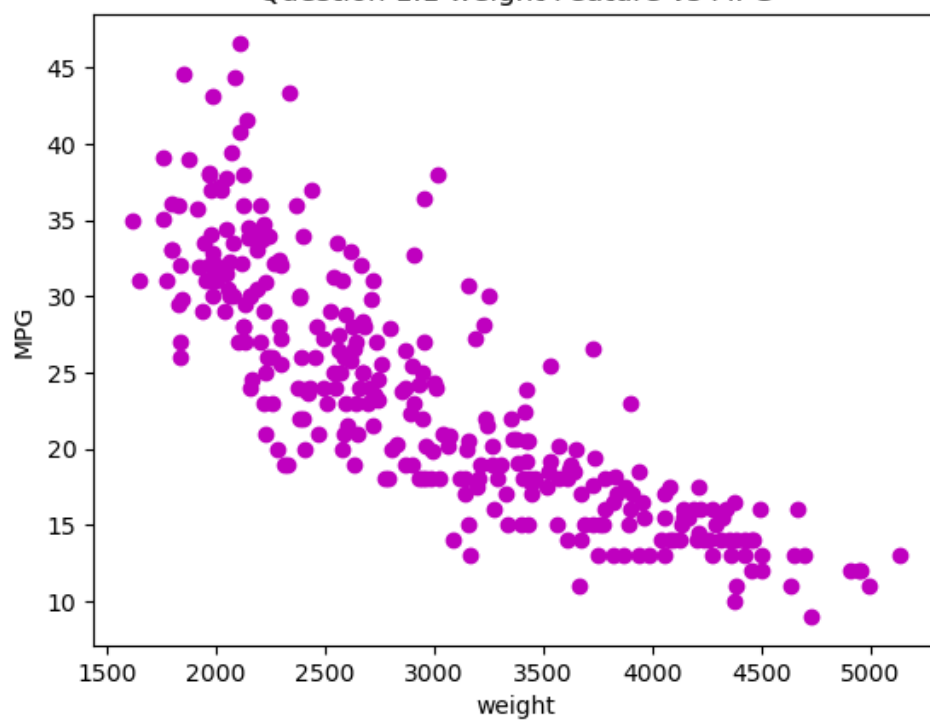
I think horsepower has the best predicting MPG because it is the closest to the linear regression line.



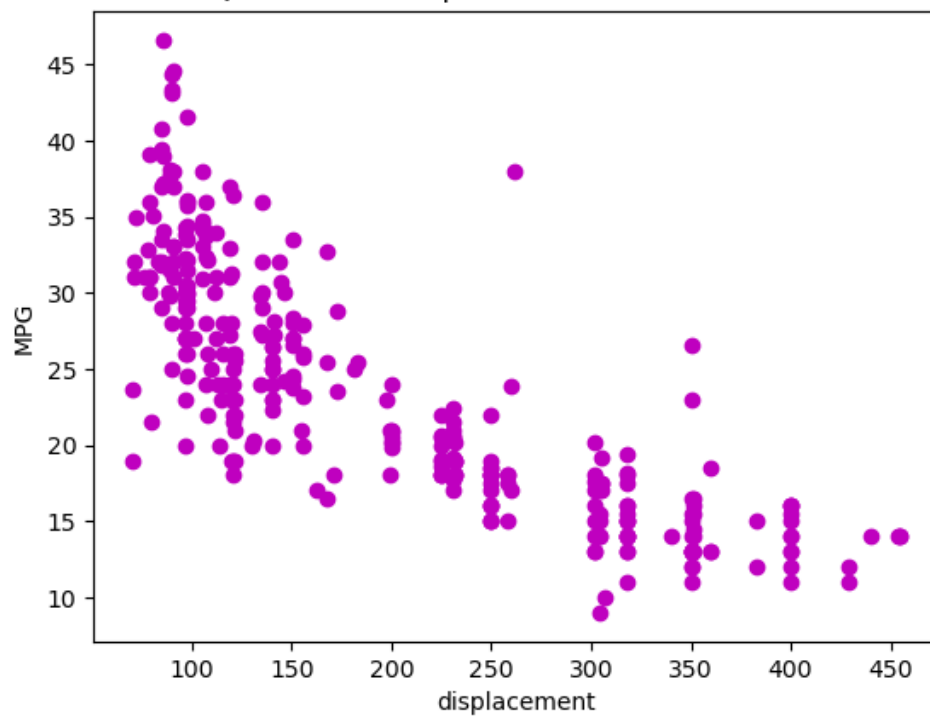
Question 1.1 horsepower Feature vs MPG



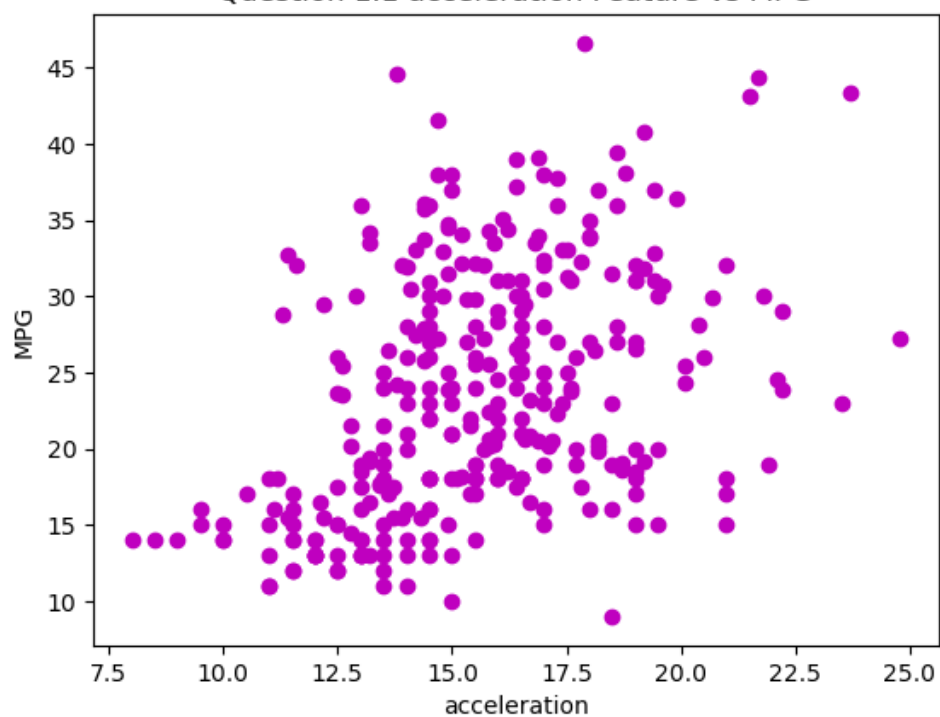
Question 1.1 weight Feature vs MPG



Question 1.1 displacement Feature vs MPG



Question 1.1 acceleration Feature vs MPG



1.2 - Does using all of the data improve performance over using each of the features individually?

Yes, they the dots are more concise and closer to the linear regression line

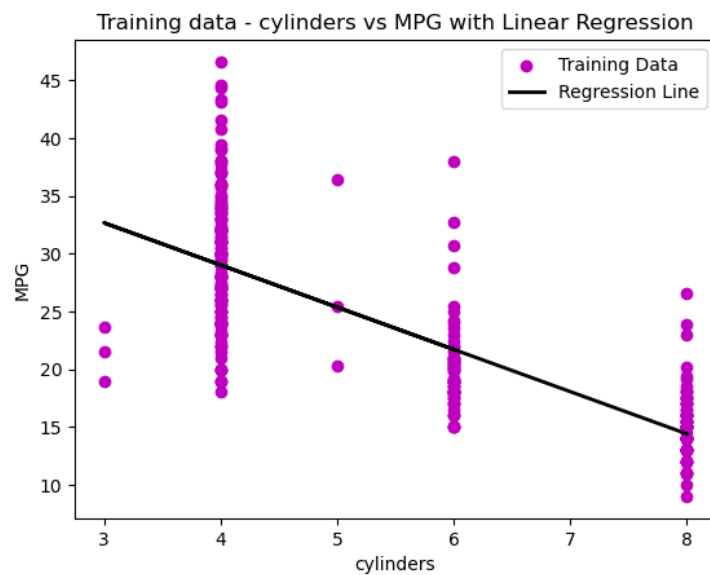
Feature: cylinders

Coefficient: -3.652154589303549

Intercept: 43.62203602314661

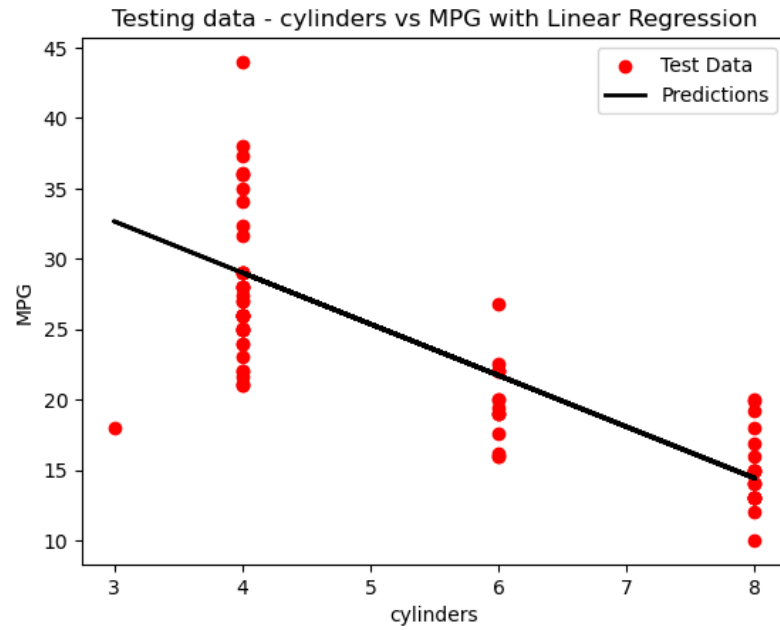
Mean squared error on train: 24.655233534565706

R2 on training: 0.6092657121665905



Mean squared error on test: 21.81398135353217

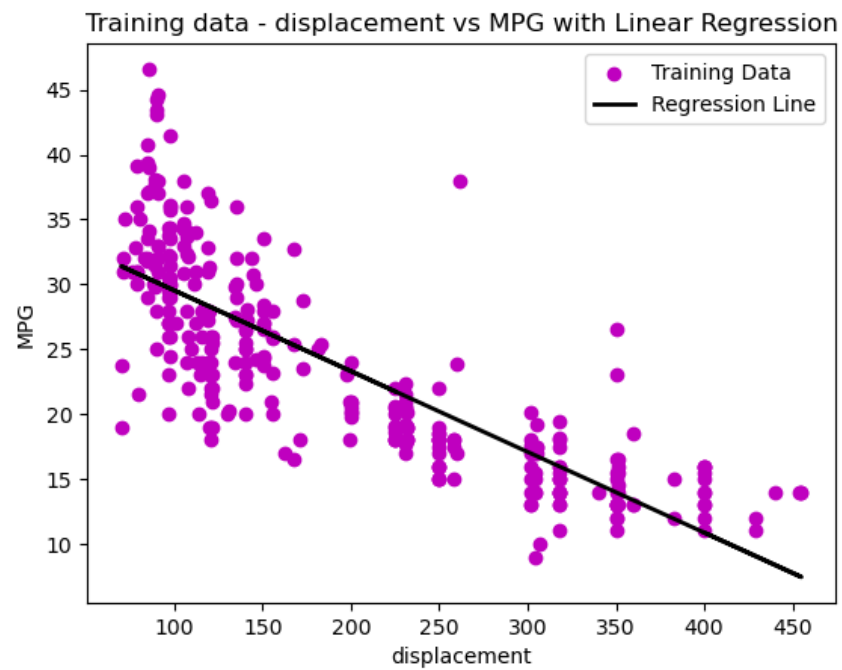
R2 on testing: 0.5726149669760378



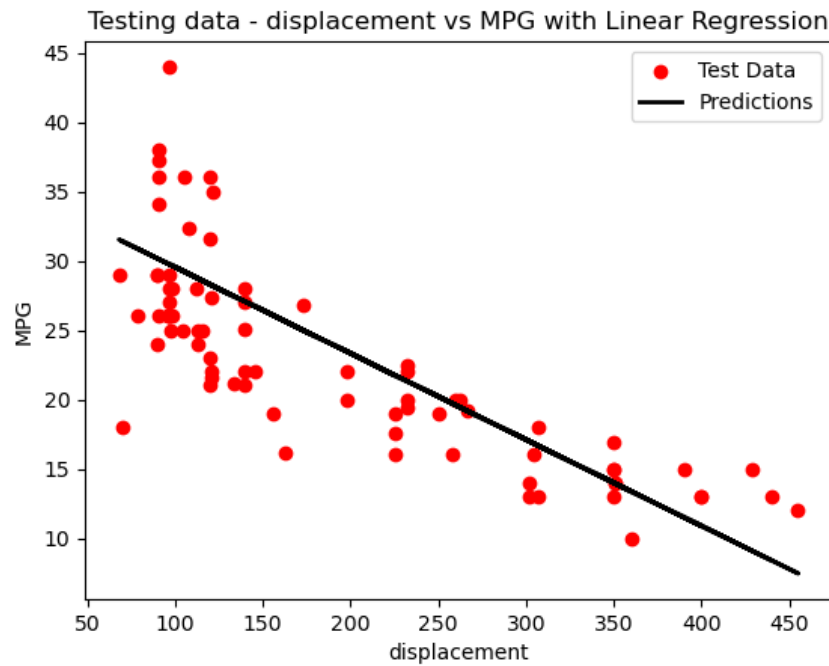
Feature: displacement

Coefficient: -0.06222755649501378

Intercept: 35.765941772604975



Mean squared error on train: 21.538297693471065 R² on training: 0.658662677090289



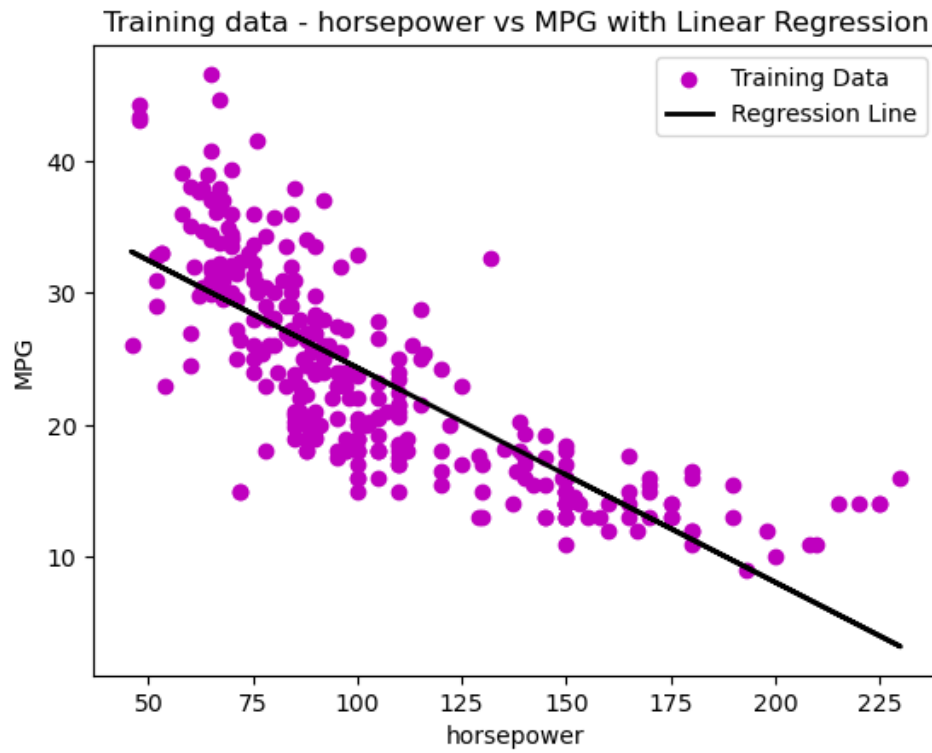
Mean squared error on test: 21.227484289611393

R² on testing: 0.5841057656051312

Feature: horsepower

Coefficient: -0.16259724322918448

Intercept: 40.606097600118346



Mean squared error on train: 24.475168273683455

R2 on training: 0.6121193728863612

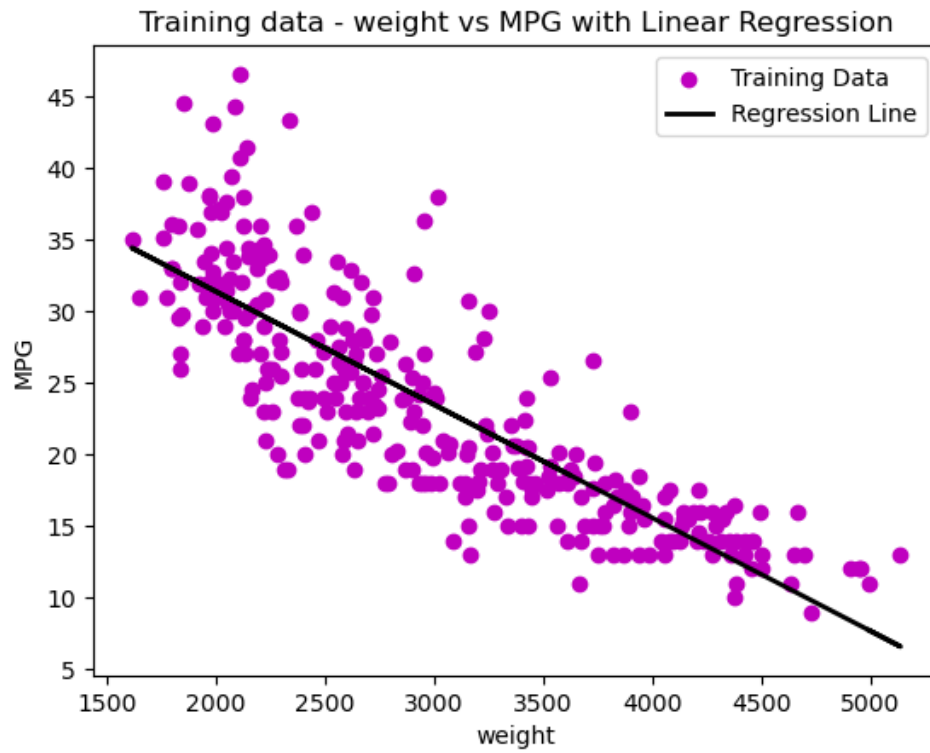
Mean squared error on test: 22.153237123863413

R2 on testing: 0.5659681822256184

Feature: weight

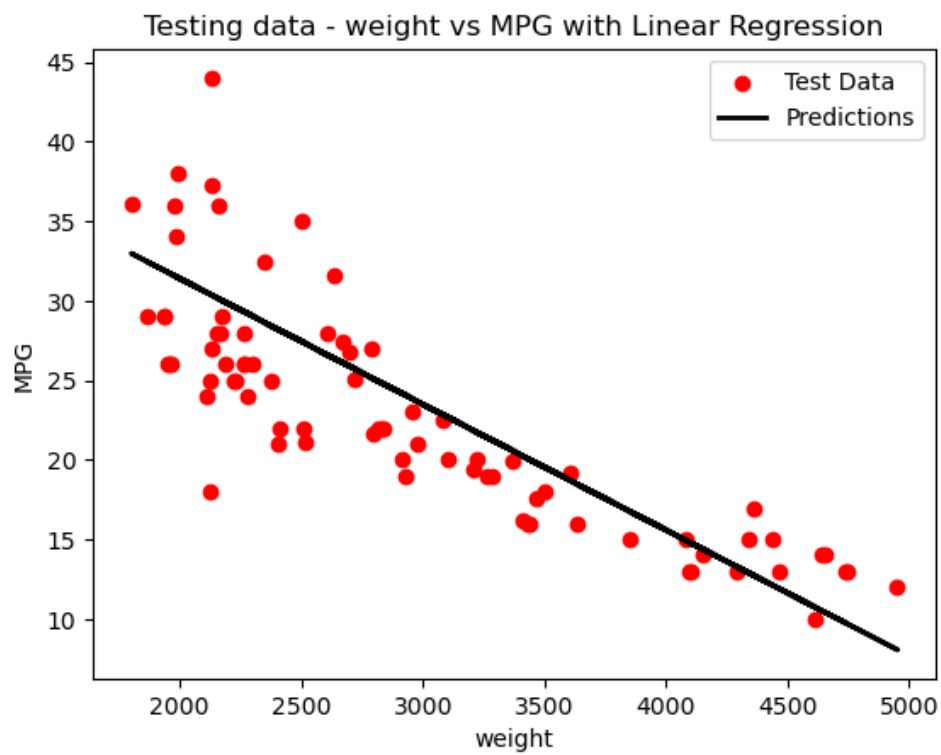
Coefficient: -0.007903610385225605

Intercept: 47.200526427552106



Mean squared error on train: 19.04510619658238

R2 on training: 0.6981745885310532



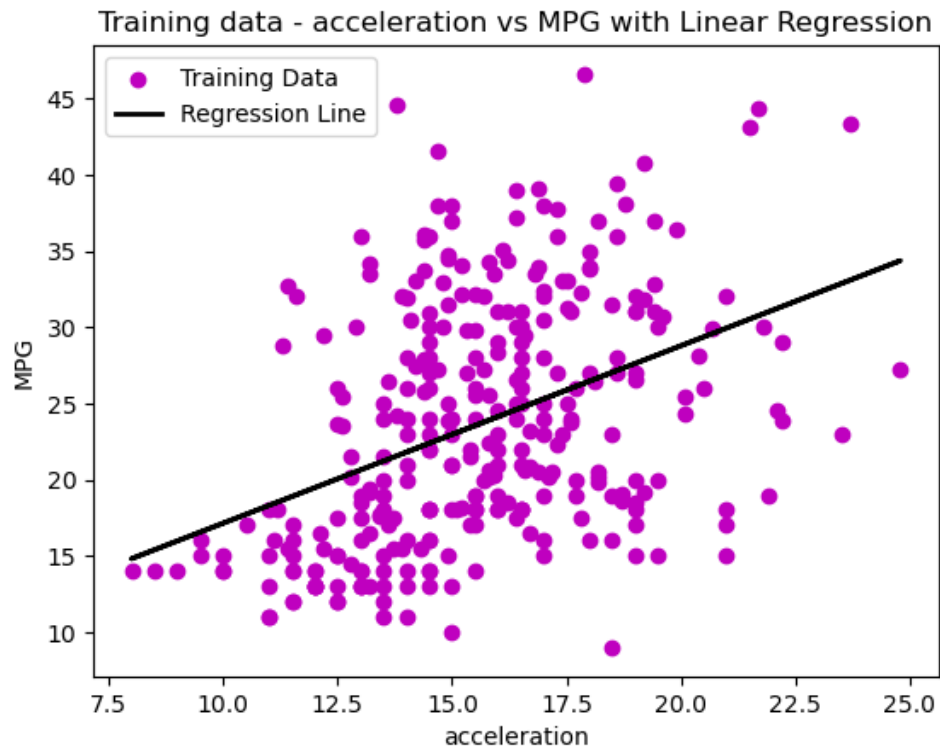
Mean squared error on test: 17.693388269545686

R2 on testing: 0.6533466675646016

Feature: acceleration

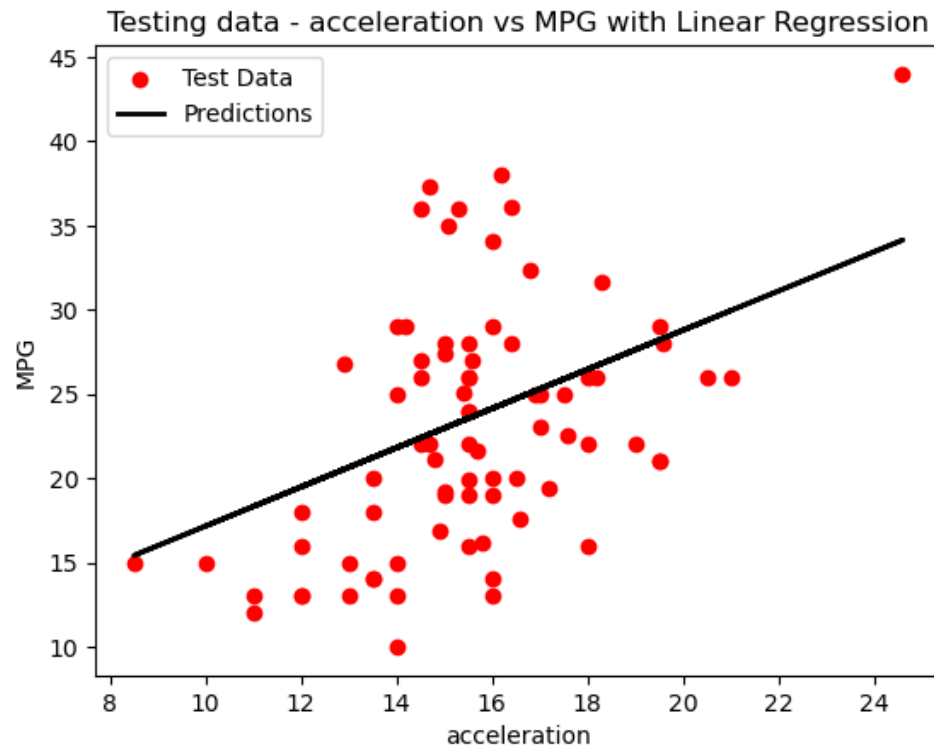
Coefficient: 1.1634207761239546

Intercept: 5.515044402313496



Mean squared error on train: 52.386427206105246

R2 on training: 0.16978383928843055

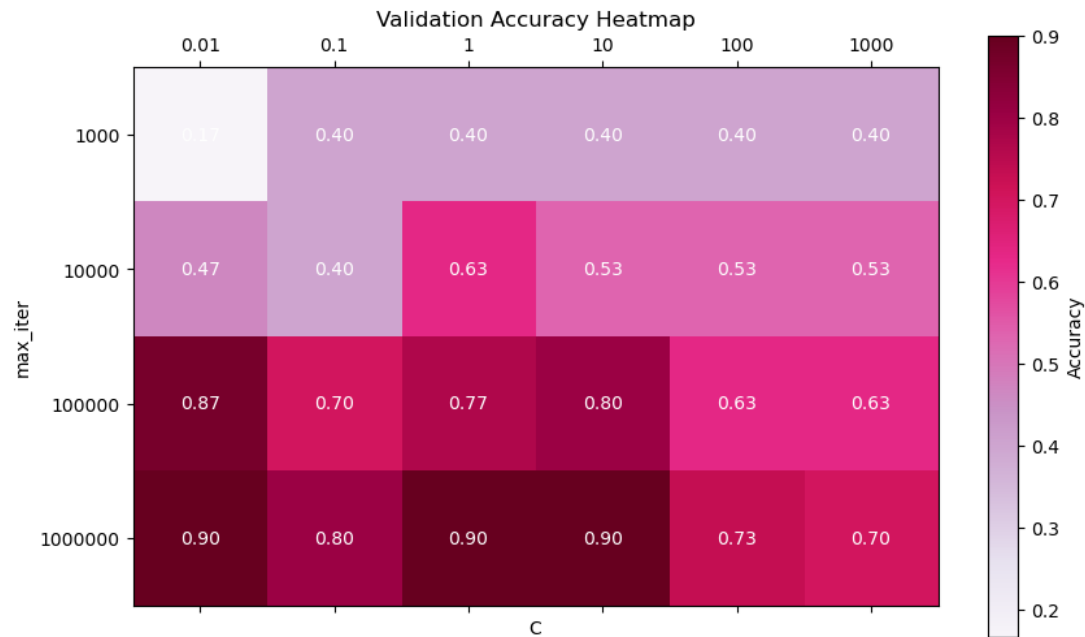


Mean squared error on test: 40.073906749162774

R2 on testing: 0.21486189605562878

2.1 - Note the best parameter combination and discuss why those parameter values might be the best for this dataset (1-3 sentences).

The best parameters are $C=0.01$ and $M_iter = 1000000$, $C=1$ & $m_iter = 1000000$, $C=10$ and $M=1000000$. These 3 are good combinations because they are the highest accuracy score in the heatmap. Their score was .90. Some of the other combinations didn't score as well because they were probably not able to converge all the way (given the convergence warnings).



2.2

training accuracy is 0.9010416666666666

testing accuracy is 0.8541666666666666