Term Project Discussion

DSC530

Veronica Warren

November 20, 2020

My data looked at what variables played a role in Boston housing prices. Using the Boston Housing prices data that I pulled from Kaggle, I started out with the below variables.

CRIM per capita crime rate by town

ZN proportion of residential land zoned for lots over 25,000 sq.ft.

INDUS proportion of non-retail business acres per town

CHAS Charles River dummy variable (= 1 if tract bounds river; 0 otherwise)

NOX nitric oxides concentration (parts per 10 million)

RM average number of rooms per dwelling

AGE proportion of owner-occupied units built prior to 1940

DIS weighted distances to five Boston employment centres

RAD index of accessibility to radial highways

TAX full-value property-tax rate per $10,000

PTRATIO pupil-teacher ratio by town - B 1000(Bk - 0.63)^2 where Bk is the proportion of blacks by town

LSTAT % lower status of the population

MEDV Median value of owner-occupied homes in $1000's

I condensed the information down to just MEDV, CRIM, RM, AGE, TAX, & RAD. I found that MEDV and RM had the most significant correlations on Boston housing prices. The data was great, but I would have preferred there to be other variables like window pane thickness(because of the cold), parking, new appliances, remodeling, front yard, back yard, garage, fireplace.

The challenges I faced were basically getting the code to work. I am still new to Python so it was a great feeling seeing all of the histograms and scatterplots populate like they did.. I really liked the colorful PMF graph as well.

I definitely learned something valuable from this class and project and am thankful I got to study it!