Will Manley

Dr. Cates

COMSC 415

Project 1 Report

**Part 1:**

There are 506 datapoints within the dataset. Each point has 13 features (not including MEDV).

**Part 2:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **MEDV** | **RM** | **LSTAT** | **PTRATIO** |
| **MIN** | 5.0 | 3.561 | 1.73 | 12.6 |
| **MAX** | 50.0 | 8.78 | 37.97 | 22.0 |
| **MEAN** | 22.53 | 6.28 | 12.65 | 18.45 |
| **STANDARD DEVIATION** | 9.19 | 0.7 | 7.14 | 2.16 |

**Part 3:**

RM: If the average number of rooms among homes in the neighborhood increased, then I believe the median value of owner-occupied homes would increase. I believe this because the value of many houses would go up, and as a result the median value should go up (assuming enough houses increase in price).

LSTAT: If the percentage of all Boston homeowners who have a greater net worth than the homeowners in the neighborhood went up, then I feel as if the median value of the homes would probably go down. If the LSTAT percentage is rising, then I feel as if that given neighborhood may be deteriorating over time. As a result, many of the house values have probably gone down resulting in a lower median value of owner-occupied homes.

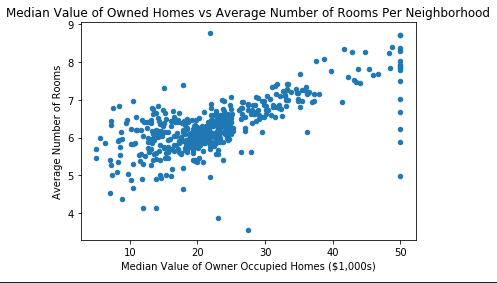
PTRATIO: If the ratio of students to teachers in primary and secondary schools in the neighborhood increased, then I believe that the median value of owner-occupied homes would potentially decrease. This one seems harder to predict; however, it seems more logical to me that the bigger the student to teacher ratio is, the worse a student’s education will be. As a result, I believe that the median house price would decrease slightly.

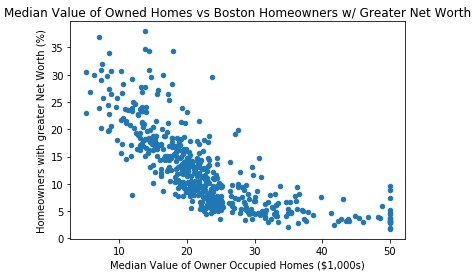
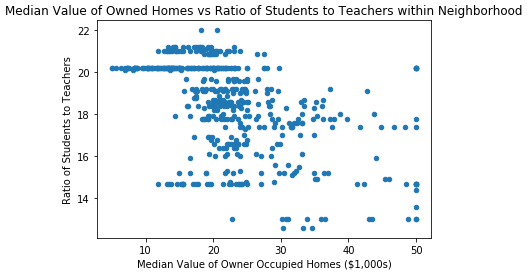
Of the three, I would expect LSTAT to have the strongest correlation to MEDV. It seems to make sense that the lower median value of houses goes, the higher the percentage of people with more net worth than the homeowners within the neighborhood would increase. The average amount of rooms should also have a decent correlation; however, I do not think it will be as strong.

**Part 4:**

It appears that my predictions were pretty accurate for all three features. The increase in RM made MEDV increase, the increase in LSTAT made MEDV decrease, and the increase in PTRATIO made MEDV decrease. I can come to this conclusion due to the correlation between MEDV and RM being positive and the correlation between MEDV and LSTAT/PTRATIO being negative.

**Part 5:**



**Part 6:**

There are quite a few interesting things going on with the above graphs. The median values of owner-occupied homes within these neighborhoods maxes out at $50,000. This seems very low considering how expensive houses are nowadays. After going back to the assignment instructions, I saw that the data was from 1978 which explains why the median house cost is only $50,000 at most. The median cost of housing has gone up a lot since then.

I am not that surprised that the smaller the ratio of students to teachers within a neighborhood, the higher the median housing price is. In most areas, the cost of an average house is typically higher if there is a good public school system. The lower that ratio is, the more a teacher can focus on every individual, resulting in better grades.

I also found it interesting how the median value of a neighborhood's homes can shoot up due to there being an extra two or three rooms on average per house. It is crazy to see that having 8 rooms as opposed to two would increase the median price of the homes by $20,000. It is even crazier to think about how much that price increase would be today.