

My exploratory data analysis on housing data showed me that there is not just one indicator of the price of a house. The question I was hoping to answer is, "What variable is the best predictor for price?" Through my analysis, I found that the area was the best available variable to predict it but I found that area and price did not have a strong correlation with the Pearson Correlation. The other ones didn't amount to anything so I think there is probably another variable out there that would be a better fit. Maybe the location of each house or the distance from the house to a beach or some other prime location. Maybe some other data that would be hard to find like the property tax for each house would be a better predictor.

I think one thing that I missed is being able to split up each variable and get more granular with my analysis. I am not very well versed in python so my skills are limited. I would have also liked to do more in depth hypothesis testing. I don't feel like the data I chose was very well suited for testing a difference in means or testing proportions.

One thing I didn't fully understand is the regression analysis. It made sense to me that the regression model said that the area increased the price by \$428 but then the RMSE was super high so that means the model wasn't a good fit. Using your intuition is good but doesn't always land you in the proper spot.

An assumption I made that was incorrect is thinking that the area of a house would be the best predictor. I also thought that this project would be a breeze but it took a lot of work!