**John Cason**

**DS160**

**4-7-24**

**Professor Sarkar**

**Housing Prices** **Exploratory Analysis**

John Cason, [jcason@bellarmine.edu](mailto:jcason@bellarmine.edu)

1. **INTRODUCTION**

The Data set I chose to use for this project was a Housing Prices dataset. This data set was found on Kaggle and was made up of housing prices and other housing variables. There are only 545 entries in this dataset, so it is important to remember that this data is limited. I chose to use this because of its large variety of categories, data types, and variables. This data set also seemed very intriguing. Another variable is that the data doesn’t tell us where the housing prices are located so we also don’t know where these prices are relevant too. With this dataset we are looking for how housing prices relate to general area of a house.

1. **Description of Dataset and Independent and dependent variables**

This data set contains 545 samples with 13 columns with 6 int64 and 7 object data types. The dependent variable in my data set is the prices and the independent is area. Due to all the data being complete no data cleaning was necessary.

**0. price - numerical**

**1. area - numerical**

**2. bedrooms - numerical**

**3. bathrooms - numerical**

**4. stories - numerical**

**5. main road - categorical**

**6. guestroom - categorical**

**7. basement - categorical**

**8. hot water heating - categorical**

**9. air conditioning - categorical**

**10. parking - numerical**

**11. prefarea - categorical**

**12. furnishing status – categorical**

**IV Variable Encoding:**

For my specific regression model, no categorical variables needed to be changed to answer my questions. However, in a bigger study you could continue to test how prices were affected by variables by changing some of the categorical variables to numerical using one-hot and ordinal encoding. For example, you could use basement and yes equals 1 and 0 equals no. After that you can see how having a basement affects prices.

**VI Preliminary Analysis:**

**Table 1** shows a statistical analysis of allnumerical values

A table of numbers and a few ones

Description automatically generated with medium confidence

**Table 2** Shows the relationship between price and area. This shows there is a positive relationship to price and area. However, this relationship seems to be on the weaker side which hints at other variables being a factor in housing prices.

A graph of blue dots

Description automatically generated