Project Proposal

Hunter Garrison, Kevin Smith, Reese Madsen, Giulio Martini

Due: April 2, 2024

1. Team Members: Hunter Garrison, Kevin Smith, Reese Madsen, Giulio Martini

2. Title: Housing Rent Regression

3. Narrative

The research question we are exploring is which combination of predictor variables will be good for representing the variance in the price of houses. We will perform tests on our full linear model to see which features are the best for regression and which points or features may be causing problems. We will use metrics (such as R squared, RMSE, and RMSE LOOCV) to test the performance of our models on how well it describes the variance of the price of the houses. Some of the predictor variables include area of the house, number of bedrooms, number of bathrooms, stories, and presence of basement. Our hypothesis is that the area of the house, preferred area, and bedrooms will be the variables that will best represent the variance in the price of the houses, and that the other predictors will not have much affect on the linear model. Also, we predict that there will be sever collinearity between the number of stories of a house and the area, number rooms, and status of a guest room in a house.

4. Data Description

Dataset source: https://www.kaggle.com/datasets/yasserh/housing-prices-dataset

Acknowledgement

- Harrison, D. and Rubinfeld, D.L. (1978) Hedonic prices and the demand for clean air. J. Environ. Economics and Management 5, 81–102.
- Belsley D.A., Kuh, E. and Welsch, R.E. (1980) Regression Diagnostics. Identifying Influential Data and Sources of Collinearity. New York: Wiley.

Description

The Housing dataset consists of thirteen variables with 545 samples collected in the Delhi region of India. There is no information on how this data was collected. Of the features, six of them are numerical: price of the house in Rupees, area of the house in square feet, as well as the number of bedrooms, bathrooms, stories, and parking spots for the house. There are also six boolean type predictors: whether the house is connected to a main road, whether it has a guest room, a basement, a hot-water heater, air conditioning, and whether it is in a preferred area. There is one categorical variable which is the furnishing status of the house, consisting of "semi-furnished", "unfurnished", and "furnished". In our case, we are considering the price of the house to be the response variable and the other twelve variables to be the predictors.

5. Data Dictionary

Variables

price - price of the houses in rupees. **This is the response variable** (integer)

area - area of a house in square feet (integer)

bedrooms - number of bedrooms in house (integer)

bathrooms - number of bathrooms in house (integer)

stories - number of stories in house (integer)

mainroad - whether or not house is connected to main road (boolean)

guestroom - whether or not house has a guest room (boolean)

basement - whether or not house has a basement (boolean)

hotwaterheating - whether or not house has a hot-water heater (boolean)

airconditioning - whether or not house has air conditioning (boolean)

parking - number of parking spots at house (integer)

prefarea - whether or not the house is in a preferred area (boolean)

furnishingstatus - furnishing status of the house (string)

6. Evidence of Data

```
##       price  area bedrooms bathrooms stories mainroad guestroom basement
## 1  13300000  7420        4         2       3      yes        no       no
## 2  12250000  8960        4         4       4      yes        no       no
## 3  12250000  9960        3         2       2      yes        no      yes
## 4  12215000  7500        4         2       2      yes        no      yes
## 5  11410000  7420        4         1       2      yes       yes      yes
## 6  10850000  7500        3         3       1      yes        no      yes
## 7  10150000  8580        4         3       4      yes        no       no
## 8  10150000 16200        5         3       2      yes        no       no
## 9   9870000  8100        4         1       2      yes       yes      yes
## 10  9800000  5750        3         2       4      yes       yes       no
##    hotwaterheating airconditioning parking prefarea furnishingstatus
## 1               no             yes       2      yes        furnished
## 2               no             yes       3       no        furnished
## 3               no              no       2      yes   semi-furnished
## 4               no             yes       3      yes        furnished
## 5               no             yes       2       no        furnished
## 6               no             yes       2      yes   semi-furnished
## 7               no             yes       2      yes   semi-furnished
## 8               no              no       0       no      unfurnished
## 9               no             yes       2      yes        furnished
## 10              no             yes       1      yes      unfurnished
```