Project Proposal

Group 1

11/15/2022

Introduction

In general, when people want to buy a house, they look for a house that is affordable and has all the desired features. Home price predictions will help them to decide whether the house they want to buy is worth that price or not.

It's the same with people who want to sell a house. By utilizing a house price prediction system, the seller will be able to decide whether all the features inherent in the property and the surrounding environment can add value to the sale so that the house can be sold at the best price.

Data

The real estate markets, like those in Sydney and Melbourne, present an interesting opportunity for data analysts to analyze and predict where property prices are moving towards. Prediction of property prices is becoming increasingly important and beneficial. Property prices are a good indicator of both the overall market condition and the economic health of a country. Considering the data provided, we are wrangling a large set of property sales records stored in an unknown format and with unknown data quality issues

Sampling method: Each of the observation is taken independently. We can conclude that the method used is **Random sampling** method.

Variables and it's attributes:

1:Date: Date at which the price calculated.

2:Price: The price of the house.(Numeric value).

3:bedrooms: The number of bedrooms in the house(Numeric)

4:bathrooms: The number of bathrooms in the house.(Numeric)

5:sqft_living: The area of living in the house measured in the square feet(Numeric).

6:sqft_lot: The lot area of the house measured in square feet.(Numeric)

7:floors: Number of floors in the house.(Numeric)

8:waterfront: It determines whether it has waterfront or not.(Factorial)

9:view: Rating the view on a scale of 0-4(Factorial)

10:condition: Rating the condition of house on a scale of 0-4(Factorial)

11:sqft_above: The amount of area in the top floors measured in square feet.(Numeric)

12:sqft_basement: The amount of area of the basement measured in square feet.(Numeric)

13:yr_built: The year in which the house is built.(Numeric)

14:yr_renovated: The year in which the house is renovated.(Numeric)

15:street: The street at which the house is located.

16:city: The name of city at which the house is located.

17:statezip: The zipcode at which the house is located.

18:country: The name of country at which the house is located.

```
data <- read.csv("/cloud/project/data.csv")
```

Data Analysis Plan

Response Variable: Price

Price Predictor of interest: sqft_living

Confounding Variables:sqft_lot

Null Hypothesis(H_0): Null Hypothesis: sqft_living and price are dependent.

Alternate Hypothesis(H_A): sqft_living and price are not dependent.

Conditons:

1. Independence - It is randomly taken Satisfied
2. Expected counts are all greater than 5

```
data2<-subset(data,price!="0")
```

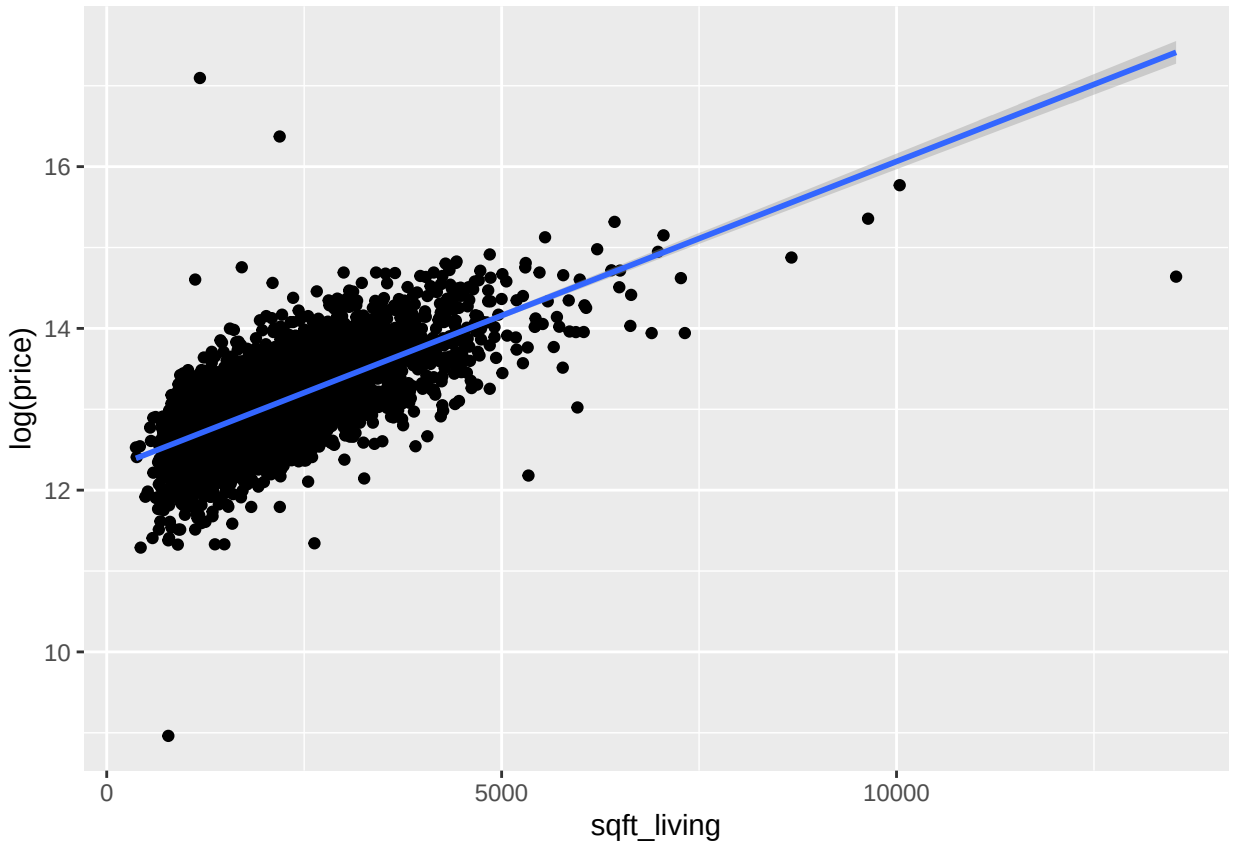
Exploratory Analysis:

variables	sd	mean
sqft_living	963.2069158	2139.3469565
price	5.638347×10^5	5.5196299×10^5

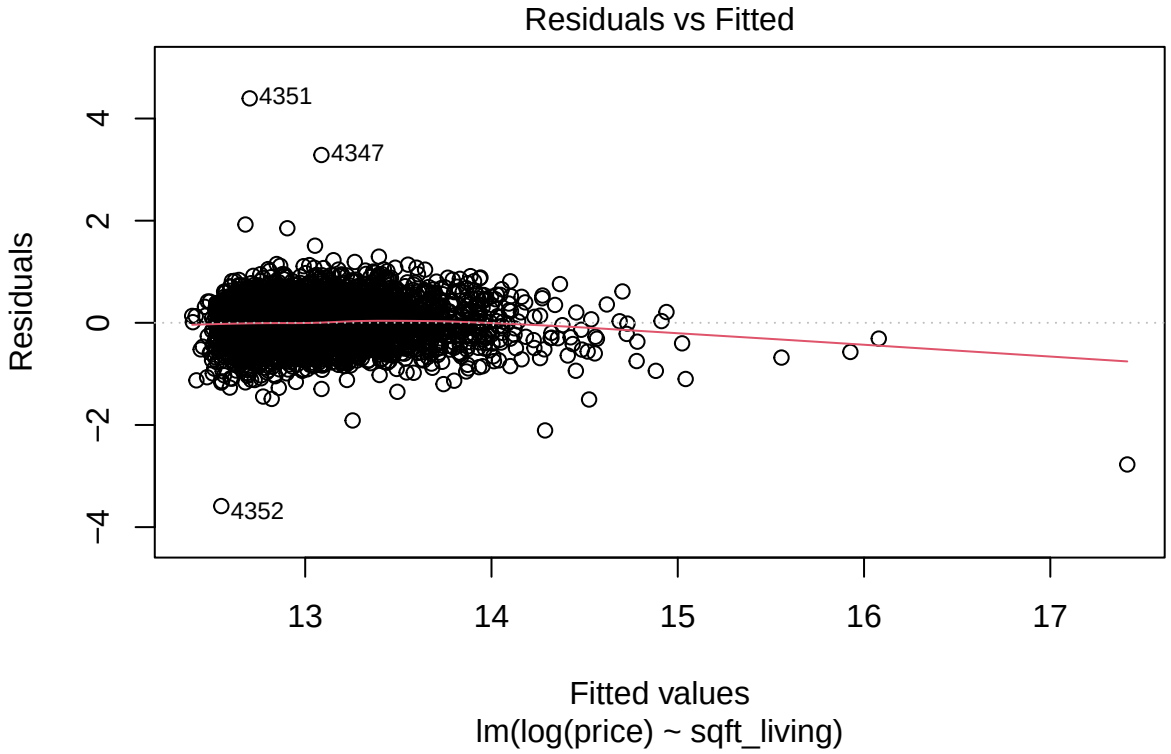
#Linearity between price and sqft_living:

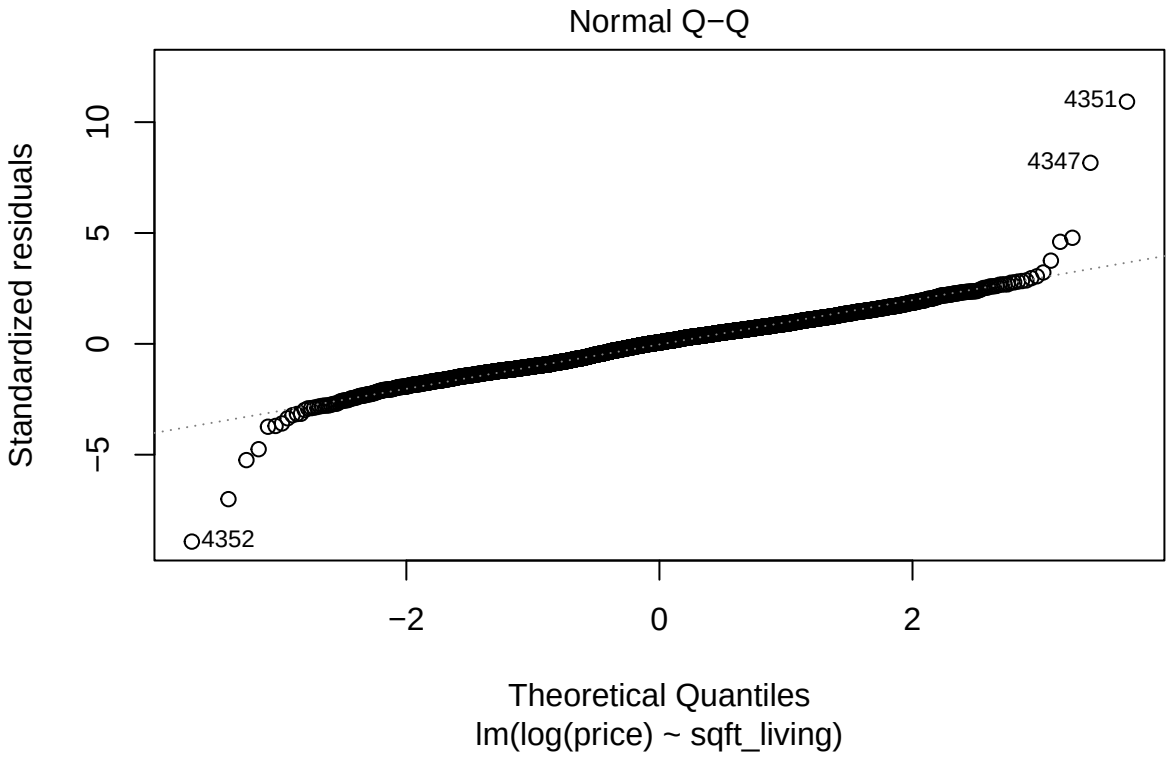
```
library(ggplot2)

ggplot(data2,aes(x=sqft_living,y=log(price)))+geom_point()+geom_smooth(method = lm)
```

```
## `geom_smooth()` using formula 'y ~ x'
```




```
mode <- lm(data = data2,log(price)~sqft_living)
plot(mode,1:2)
```





#Analysis Result

We observe that there is linearity between the price and sqrt_living. With increase in sqft of living, we can see that price also increases.