UM–SJTU Joint Institute

Applied Regression Analysis (STAT4130)

Term Project Report

Linear Regression on Lisbon House Prices

Instructed by

Dr. Ailin Zhang

July 24, 2023

Name	**ID**
Li Ang	520370910024
Cao Yanzhuo	520370910021

Contents

1 Background and Description of the Problem

This project aims to analyze housing prices in Lisbon, Portugal, using applied regression analysis. The objective is to build a predictive model for estimating housing prices based on property attributes like condition, type, bedrooms, bathrooms, area, parking, and location.

Lisbon's real estate market is dynamic and competitive, attracting diverse buyers and investors. Understanding the factors influencing housing prices is crucial for making informed decisions in this market.

The analysis utilizes a comprehensive dataset[2] of residential properties in Lisbon, containing essential variables like property condition, type, bedrooms, bathrooms, area, parking availability, geographical coordinates, and $price/m^2$.

By employing applied regression techniques, particularly linear regression, we will construct a predictive model that provides valuable insights into the factors driving housing prices in Lisbon, benefiting stakeholders, including buyers, sellers, investors, and policymakers.

2 Design and Analysis

- Process the **categorical and quantitative** variables. Then find the variables which have **collinearity**, and delete one of them.
- Do **linear regression** on all variables. Then Analyze the results and **delete** the variables which are **not significant.**
- Do linear regression on the **remaining variables** and **continue** to see if more variables need to be deleted. Then Use **F-Test** to decide if we will accept the reduced model.
- Analyze the **Residuals vs Fitted values** plot and see if the residuals satisfy our requests. Then List all the points with **high leverage** to **ignore** them and do the regression.
- Analyze the **Normal Q-Q plot**[1] and see the points with **high Standardized Residuals**. Then Analyze the **Leverage plot** and see the points with **high Cook's Distance**.
- **Exclude** points with high Standardized Residuals and Cook's Distance and **do the regression again**. Then **Repeat** the above three steps until there are **no outliers**. And plot the heat map of variables.
- In model 2, very **similar** to the method of model 1, but more focus on the **polynomial** of variables X like X^2, X^3, log(X) ... and the **multiplication** of variables like X_1*X_2, X_1*X_2*X_3 ...

3 Results

Our results will be divided into two parts: Model 1 and Model 2, and will be shown as follows.

3.1 Model 1

3.1.1 Regression on all variables

Firstly, we do the regression on all the variables and get the following result(See Figure 1). Since there are too many variables, I just show some part of the results.

3.1.2 Find the collinearity

Then we find that the variable AreaGross is not included in the regression model since it has collinearity with other variables. Therefore, we do the scatter plot of AreaGross vs each variables to check the collinearity. Then we can find that the variable AreaNet has strong linear relationship with AreaGross(See Figure 2). Since there are too many plots, I just choose the one which can show the collinearity.

```
## Call:
## lm(formula = Price ~ Bedrooms + Bathrooms + AreaNet + AreaGross +
##     Parking + Latitude + Longitude + Price_Sq._M. + Condition +
##     PropertyType + PropertySubType + Parish, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -370299  -91115     170   69735 1648112
##
## Coefficients: (3 not defined because of singularities)
##                                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)                    2.375e+08  1.245e+08   1.907  0.05786 .
## Bedrooms                      -4.878e+04  2.105e+04  -2.318  0.02146 *
## Bathrooms                      6.818e+04  2.387e+04   2.856  0.00473 **
## AreaNet                        5.668e+03  4.953e+02  11.443  < 2e-16 ***
## AreaGross                             NA         NA      NA       NA
## Parking                        5.303e+04  5.256e+04   1.009  0.31412
## Latitude                      -6.672e+06  3.188e+06  -2.093  0.03755 *
## Longitude                     -2.004e+06  2.708e+06  -0.740  0.46010
## Price_Sq._M.                   7.538e+02  8.344e+02   0.903  0.36738
## ConditionNew                  -6.629e+04  4.646e+04  -1.427  0.15514
## Condition'As_New'              7.448e+04  4.684e+04   1.590  0.11333
## Residual standard error: 189800 on 206 degrees of freedom
## Multiple R-squared:   0.83,  Adjusted R-squared:  0.7978
## F-statistic: 25.78 on 39 and 206 DF,  p-value: < 2.2e-16
```

Figure 1: Original regression of Model1

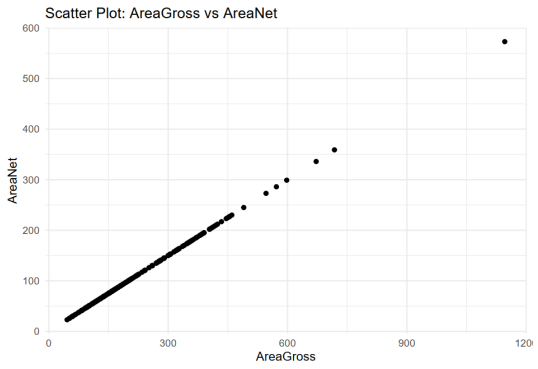

Figure 2: Scatter plot: AreaGross vs AreaNet

3.1.3 Reduce the original model

Then we further do another regression deleting some NA variables and $Pr|t| > 0.1$ variables shown in the first regression (See Figure 3).

We can see that the r squared dropped a little, so we further use F-Test to decide if we will accept the reduced model. The result in Figure 4 shows that we can accept the reduced model.

```
## Residual standard error: 193000 on 212 degrees of freedom
## Multiple R-squared:  0.8191, Adjusted R-squared:  0.7909
## F-statistic: 29.09 on 33 and 212 DF,  p-value: < 2.2e-16
```

Figure 3: Regression result of reduced model

```
## Analysis of Variance Table
##
## Model 1: Price ~ Bedrooms + Bathrooms + AreaNet + PropertySubType + Parish
## Model 2: Price ~ Bedrooms + Bathrooms + AreaNet + AreaGross + Parking +
##     Latitude + Longitude + Price_Sq._M. + Condition + PropertyType +
##     PropertySubType + Parish
##   Res.Df        RSS Df  Sum of Sq      F  Pr(>F)
## 1    212 7.8960e+12
## 2    206 7.4221e+12  6 4.7383e+11 2.1919 0.04513 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 4: Result of F-test

According to the Residual Plot (See Figure 5) below, we can notice that the variance of residuals increases as Fitted value getting large, with some outliers.

3.1.4 Delete high leverage points

To exclude the points with high leverage, we need to first find them by running the following codes (See Figure 6).

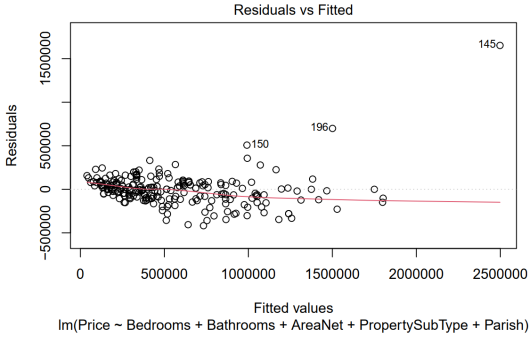

Figure 5: Residual plot of model 1

```r
leverage <- hatvalues(lm3)
high_leverage_points <- which(leverage > mean(leverage) + 2 * sd(leverage))
print(high_leverage_points)
```

```
##  15  20  50  51  52 127 128 141 155 167 188 203 245
##  15  20  50  51  52 127 128 141 155 167 188 203 245
```

Figure 6: High leverage points of model 1

3.1.5 Delete high outliers based on Q-Q plot and leverage plot

Further more, we do the Q-Q Plot (See Figure 7) to check if the standardized residual obey normal distribution. Then we find 3 outliers: point 150, point 196, point 145. The other residuals basically obey the normal distribution since they are almost on a straight line and ranges from (-2,2).

To determine whether these points will have great influence on the regression, we do the leverage plot (See Figure 8). We can see from the plot that point 145, point 196 and point 246 will have great influence on the regression result.

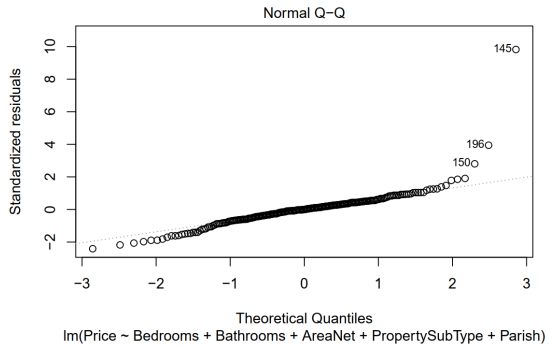

Figure 7: Q-Q Plot of model 1 after deleting high leverage points

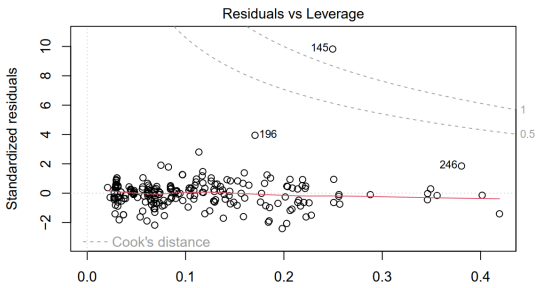

Figure 8: Leverage plot of model 1 after deleting high leverage points

Combine the results of Q-Q plot and leverage plot, we decided to delete the outliers: point 145 and point 196. Then we do a new regression to see the results (See Figure 9).

```
##
## Call:
## lm(formula = Price ~ Bedrooms + Bathrooms + AreaNet + PropertySubType +
##     Parish, data = new_data3)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -264140  -69474     597   73880  569097
##
## Coefficients:

##                                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)                       -7649.7    55907.5  -0.137 0.891301
## Bedrooms                         -43897.0    12040.3  -3.646 0.000338 ***
## Bathrooms                         93442.7    14589.8   6.405 1.02e-09 ***
## AreaNet                            4381.6      325.7  13.454  < 2e-16 ***
## PropertySubTypeDuplex            121574.8    46826.5   2.596 0.010109 *
## PropertySubTypeTownhouse_Dwelling 174498.9    75419.9   2.314 0.021680 *
## ParishAlcantara                   77155.9    90062.1   0.857 0.392618
## ParishAlvalade                   -28023.4    72469.5  -0.387 0.699387
## ParishArroios                    -26013.0    64336.1  -0.404 0.686394
## ParishAvenidas_Novas             137868.0    77276.6   1.784 0.075896 .
## ParishBelem                       49452.6    68629.2   0.721 0.471995
## ParishSao_Vicente                 16156.3    64430.2   0.251 0.802254
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 120800 on 204 degrees of freedom
## Multiple R-squared:  0.8747, Adjusted R-squared:   0.86
## F-statistic: 59.35 on 24 and 204 DF,  p-value: < 2.2e-16
```

Figure 9: Regression on model 1 after first deleting of outliers

3.1.6 Analyze the new model and further improve it

Using the similar methods mentioned above (See Figure 10 and Figure 11), we delete new outliers: points ”23”.

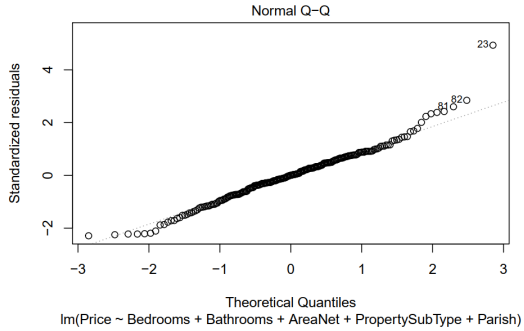

Figure 10: Q-Q Plot of model 1 after deleting of outliers

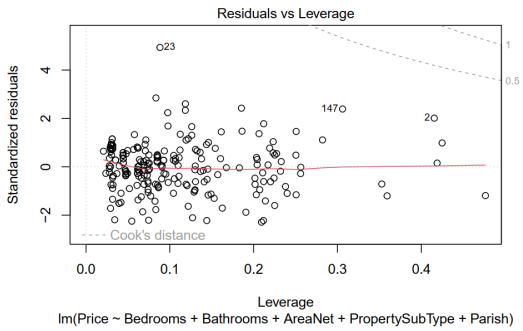

Figure 11: Leverage plot of model 1 after first deleting of outliers

Then we do a new regression to see the results (See Figure 12).

```
##
## Call:
## lm(formula = Price ~ Bedrooms + Bathrooms + AreaNet + PropertySubType +
##     Parish, data = new_data3)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -251171  -65800     368   72684  316944
##
## Coefficients:
##                                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)                      -1969.8    53298.8  -0.037  0.97055
## Bedrooms                        -46037.4    11485.5  -4.008 8.57e-05 ***
## Bathrooms                       102191.9    13841.1   7.383 3.86e-12 ***
## AreaNet                           4140.0      309.1  13.395  < 2e-16 ***
## PropertySubTypeDuplex           123401.9    44632.8   2.765  0.00622 **
## PropertySubTypeTownhouse_Dwelling 270027.1    62278.1   4.336 2.28e-05 ***
## ParishAlcantara                  84946.8    85844.6   0.990  0.32357
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 115100 on 204 degrees of freedom
## Multiple R-squared:  0.8834, Adjusted R-squared:  0.8696
## F-statistic: 64.37 on 24 and 204 DF,  p-value: < 2.2e-16
```

Figure 12: Regression on model 1 after second deleting of outliers

3.1.7 Analyze the second model and plot the heat map

Using the similar methods mentioned above (See Figure 13 and Figure 14), we cannot find more outliers.

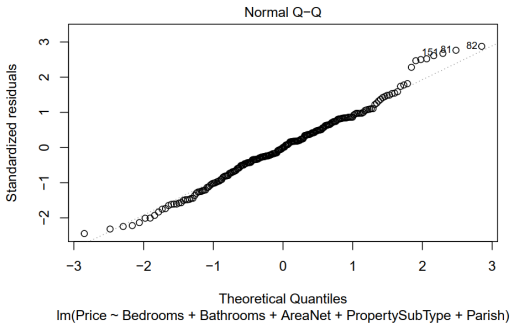

Figure 13: Q-Q Plot of model 1 after second deleting of outliers

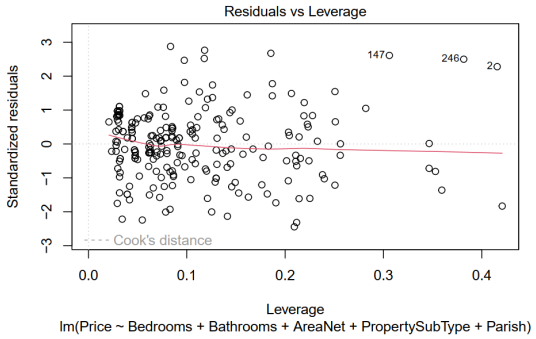

Figure 14: Leverage plot of model 1 after second deleting of outliers

Finally, we use the heat map to visualize the degree of variable influence —- with deeper red colors indicating greater influence.

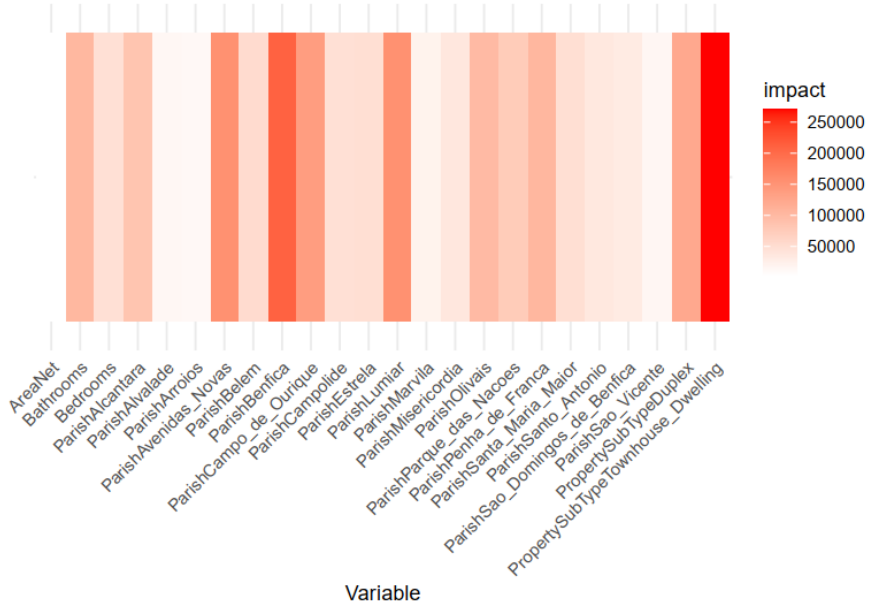

Figure 15: Heat map of model 1

Now we can see from the regression result that the R squared value is 0.88, from residuals plot that the residuals seems randomly assigned, from Q-Q Plot that standardized residual ranges from (-3,3), from leverage plot that no points have cook's distance larger than 0.5, which suggests that our first model is acceptable.

3.2 Model 2

3.2.1 Using boxplot to have a glance at the data

The boxplot result is shown below.

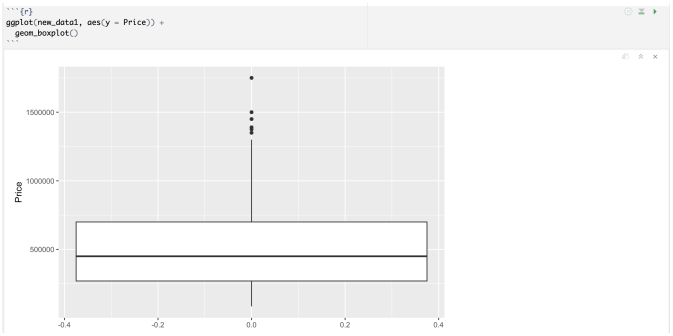

Figure 16: Boxplot for model 2

There are some points much higher than medium. They are some houses that are very expensive. From the shape of the box, we can conclude that in the data, there are more houses with low price. There are less number of houses with high price but they are much more expensive.

3.2.2 Basic ideas

This model is developed in a way similar to model 1. We intend to create a more advanced model while avoiding overfitting. The approach we take is taking higher order polynomial of the variables, logarithm of the variables and interaction terms into account. We also tried to devide between certain dummy varaible and others, like assigning those likely to have a relationthip with the dependent variable 1 and others not likely to have a relationship 0, but this doesn't work.

We only include the variables that appeared to be of high correlation in model 1. We used the data got in model 1 expect that the outliers are included, since they may not be outliers in the advanced model.

The first regression result is shown in Figure 17 and 18.

```
Call:
lm(formula = Price ~ poly(Bedrooms, 3, raw = TRUE) + poly(Bathrooms,
    3, raw = TRUE) + poly(AreaNet, 3, raw = TRUE) + Bedrooms:Bathrooms +
    Bedrooms:AreaNet + Bathrooms:AreaNet + log(AreaNet) + Latitude +
    PropertySubType + Parish, data = new_data0)

Residuals:
    Min      1Q  Median      3Q     Max
-713598  -76713    8269   71533 1227432

Coefficients:
                                Estimate Std. Error t value Pr(>|t|)
(Intercept)                    2.588e+08  1.165e+08   2.221  0.02744 *
poly(Bedrooms, 3, raw = TRUE)1 -2.734e+05  1.000e+05  -2.733  0.00684 **
poly(Bedrooms, 3, raw = TRUE)2  7.649e+04  3.501e+04   2.184  0.03008 *
poly(Bedrooms, 3, raw = TRUE)3 -6.586e+03  3.025e+03  -2.177  0.03061 *
poly(Bathrooms, 3, raw = TRUE)1 -6.342e+04  1.767e+05  -0.359  0.72012
poly(Bathrooms, 3, raw = TRUE)2  2.037e+04  7.350e+04   0.277  0.78200
poly(Bathrooms, 3, raw = TRUE)3  5.971e+03  9.097e+03   0.656  0.51233
poly(AreaNet, 3, raw = TRUE)1    2.224e+04  9.899e+03   2.246  0.02578 *
poly(AreaNet, 3, raw = TRUE)2   -8.750e+01  3.908e+01  -2.239  0.02625 *
poly(AreaNet, 3, raw = TRUE)3    1.458e-01  5.607e-02   2.601  0.00998 **
log(AreaNet)                    -3.280e+05  3.584e+05  -0.915  0.36127
Latitude                        -6.662e+06  3.011e+06  -2.212  0.02806 *
PropertySubTypeApartment        -2.390e+04  2.116e+05  -0.113  0.91017
PropertySubTypeDuplex            2.484e+05  2.209e+05   1.125  0.26211
PropertySubTypeDwelling         -1.060e+05  2.533e+05  -0.419  0.67599
PropertySubTypeIsolated Villa   -6.739e+06  2.837e+06  -2.376  0.01846 *
PropertySubTypePenthouse         2.375e+05  3.010e+05   0.789  0.43099
PropertySubTypeStudio           -3.000e+05  2.402e+05  -1.249  0.21316
PropertySubTypeTownhouse Dwelling 2.371e+05  2.320e+05   1.022  0.30815
```

Figure 17: Original Regression of Model 2 part 1

```
ParishAlcantara                 4.782e+04  1.342e+05   0.356  0.72194
ParishAlvalade                  2.072e+05  1.684e+05   1.230  0.22018
ParishAreeiro                   8.633e+04  2.245e+05   0.385  0.70096
ParishArroios                   7.834e+04  1.176e+05   0.666  0.50602
ParishAvenidas Novas            1.482e+05  1.590e+05   0.932  0.35226
ParishBeato                     7.963e+04  2.147e+05   0.371  0.71105
ParishBelem                    -7.729e+04  1.045e+05  -0.740  0.46045
ParishBenfica                   8.918e+04  1.746e+05   0.511  0.61008
ParishCampo de Ourique          2.515e+05  1.027e+05   2.448  0.01522 *
ParishCampolide                 9.417e+04  1.333e+05   0.706  0.48073
ParishCarnide                   2.669e+05  2.380e+05   1.121  0.26344
ParishEstrela                  -2.620e+04  9.451e+04  -0.277  0.78188
ParishLumiar                    5.538e+04  2.101e+05   0.264  0.79232
ParishMarvila                   2.028e+05  1.503e+05   1.349  0.17886
ParishMisericordia              5.842e+04  1.244e+05   0.470  0.63914
ParishOlivais                   3.350e+05  2.227e+05   1.504  0.13406
ParishParque das Nacoes         3.067e+05  2.291e+05   1.338  0.18226
ParishPenha de Franca           2.078e+04  1.224e+05   0.170  0.86529
ParishSanta Clara               4.087e+05  2.750e+05   1.486  0.13886
ParishSanta Maria Maior         2.694e+05  9.739e+04   2.766  0.00620 **
ParishSanto Antonio             1.628e+05  1.165e+05   1.398  0.16378
ParishSao Domingos de Benfica   1.613e+05  1.883e+05   0.856  0.39282
ParishSao Vicente               7.297e+04  1.027e+05   0.710  0.47836
Bedrooms:Bathrooms             -2.161e+04  2.072e+04  -1.043  0.29822
Bedrooms:AreaNet               -3.578e+01  7.054e+02  -0.051  0.95960
Bathrooms:AreaNet               2.376e+02  6.418e+02   0.370  0.71162
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 174900 on 201 degrees of freedom
Multiple R-squared:  0.8591,    Adjusted R-squared:  0.8283
F-statistic: 27.85 on 44 and 201 DF,  p-value: < 2.2e-16
```

Figure 18: Original Regression of Model 2 part 2

3.2.3 Reducing the original model

Variables with too high p-value is deleted one after another. In order to determine whether we should keep the high degree terms, we used t-test. In order to make the report shorter, we only show some example and the result.

Figure 19: Anova 1

Figure 20: Anova 2

The anova tables show that the quadratic term for bathroom should be kept while the cubic term should be deleted, since the quadratic term changes the model just a little bit while the cubic term changes the model a lot.

3.2.4 Getting rid of outliers

We applied Q-Q Plot and Leverage plot. However, after we compared all measures, we found the most effective way is to getting rid of points that lies

further than 2 * standard deviation.

The result shows that we managed to increase the R-square by 0.02.

```
Call:
lm(formula = Price ~ poly(Bedrooms, 2, raw = TRUE) + poly(Bathrooms,
    3, raw = TRUE) + poly(AreaNet, 4, raw = TRUE) + Latitude +
    PropertySubType + Parish, data = new_data1)

Residuals:
    Min      1Q  Median      3Q     Max
-256947  -53100   13024   62649  281353

Coefficients:
                                 Estimate Std. Error t value Pr(>|t|)
(Intercept)                     1.121e+08  7.098e+07   1.579 0.115914
poly(Bedrooms, 2, raw = TRUE)1 -1.004e+05  2.982e+04  -3.368 0.000912 ***
poly(Bedrooms, 2, raw = TRUE)2  8.935e+03  4.333e+03   2.062 0.040518 *
poly(Bathrooms, 3, raw = TRUE)1 -1.018e+05  1.186e+05  -0.858 0.392008
poly(Bathrooms, 3, raw = TRUE)2  4.858e+04  5.205e+04   0.933 0.351787
poly(Bathrooms, 3, raw = TRUE)3 -4.918e+02  6.879e+03  -0.071 0.943077
poly(AreaNet, 4, raw = TRUE)1   9.314e+03  4.270e+03   2.181 0.030342 *
poly(AreaNet, 4, raw = TRUE)2  -3.647e+01  5.275e+01  -0.691 0.490171
poly(AreaNet, 4, raw = TRUE)3   1.142e-01  2.633e-01   0.434 0.664851
poly(AreaNet, 4, raw = TRUE)4  -1.669e-04  4.482e-04  -0.372 0.710068
Latitude                       -2.893e+06  1.834e+06  -1.577 0.116377
PropertySubTypeApartment       -7.449e+04  1.235e+05  -0.603 0.547199
PropertySubTypeDuplex           3.845e+04  1.294e+05   0.297 0.766678
PropertySubTypeDwelling        -1.893e+05  1.480e+05  -1.279 0.202491
PropertySubTypeIsolated Villa   3.796e+06  1.396e+07   0.272 0.785942
PropertySubTypePenthouse        2.031e+05  1.773e+05   1.146 0.253266
PropertySubTypeStudio          -2.190e+05  1.442e+05  -1.519 0.130309
PropertySubTypeTownhouse Dwelling  1.759e+05  1.352e+05   1.301 0.194787
ParishAlcantara                 7.321e+04  7.957e+04   0.920 0.358658
ParishAlvalade                  8.883e+04  1.014e+05   0.876 0.382099
ParishAreeiro                  -3.704e+04  1.352e+05  -0.274 0.784395
```

Figure 21: Result part 1

```
PropertySubTypePenthouse        2.031e+05  1.773e+05   1.146 0.253266
PropertySubTypeStudio          -2.190e+05  1.442e+05  -1.519 0.130309
PropertySubTypeTownhouse Dwelling  1.759e+05  1.352e+05   1.301 0.194787
ParishAlcantara                 7.321e+04  7.957e+04   0.920 0.358658
ParishAlvalade                  8.883e+04  1.014e+05   0.876 0.382099
ParishAreeiro                  -3.704e+04  1.352e+05  -0.274 0.784395
ParishArroios                  -1.344e+04  7.150e+04  -0.188 0.851046
ParishAvenidas Novas            1.065e+05  9.572e+04   1.112 0.267387
ParishBeato                    -5.466e+04  1.287e+05  -0.425 0.671440
ParishBelem                    -1.069e+04  6.207e+04  -0.172 0.863457
ParishBenfica                  -9.412e+04  1.045e+05  -0.901 0.368863
ParishCampo de Ourique          1.816e+05  6.057e+04   2.998 0.003074 **
ParishCampolide                 6.214e+03  7.948e+04   0.078 0.937754
ParishCarnide                   4.294e+04  1.427e+05   0.301 0.763761
ParishEstrela                  -4.557e+04  5.571e+04  -0.818 0.414401
ParishLumiar                    1.336e+04  1.269e+05   0.105 0.916230
ParishMarvila                   1.025e+05  8.995e+04   1.140 0.255744
ParishMisericordia              3.077e+04  7.384e+04   0.417 0.677371
ParishOlivais                   8.712e+04  1.346e+05   0.647 0.518271
ParishParque das Nacoes         9.438e+04  1.383e+05   0.682 0.495773
ParishPenha de Franca          -5.690e+04  7.289e+04  -0.781 0.435919
ParishSanta Clara               1.069e+05  1.661e+05   0.643 0.520831
ParishSanta Maria Maior         1.062e+05  5.988e+04   1.774 0.077640 .
ParishSanto Antonio             9.069e+04  6.810e+04   1.332 0.184473
ParishSao Domingos de Benfica   3.446e+04  1.123e+05   0.307 0.759302
ParishSao Vicente               3.026e+04  6.082e+04   0.498 0.619385
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 105200 on 196 degrees of freedom
Multiple R-squared:  0.907,     Adjusted R-squared:  0.888
F-statistic: 47.79 on 40 and 196 DF,  p-value: < 2.2e-16
```

Figure 22: Result part 2

4 Discussion

4.1 Model 1

In the result section ”Reduce the original model”. Notice that although some subvaribales like ParishAvenidas_Novas also has significant influence on the model, we cannot just deleting other Parish subvariables and keep some significant variables, since the dataset we use will change (we will only focus on data whose Parish column is Avenidas_Novas or Santa_Maria_Maior and etc.). Therefore, the following data is meaningless in a way.

4.2 Model 2

The R-square and adjusted R-square of model 2 is lower of lower 1. Since data is different, as the outliers deleted are different, we can’t conclude that the more complexed model is not as good as the simple one, and a f-test is not possible. Still, to avoid overfitting, it’s possible that using model 1 to predict is better.

5 Conclusions

The model can be useful in predicting the price of houses in Lisbon and it shows that the price is correlated with number of bedrooms, number of bathrooms, net area, latitude and the district where the house is located in.

Model 1 shows that in order for the price to be higher, the house has to have a large area, less bedrooms, more bathrooms and be at a good place. Since number of bedrooms and bathrooms are mainly small, the result is the same for model 2.

In model 1, the result is that: price = -60310.7 - 74587.9 * Bedrooms + 99780.1 * Bathrooms + 5785.5 * Areanet + 325135.3 * PropertySubTypeDuplex + 234872.4 * PropertySubTypeTownhouseDwelling - 119.1 * ParishAlcantara - 77896.5 * ParishAlvalade - 30854.6 * ParishArroios + 139182.3 * ParishAvenidasNovas - 48450.8 * ParishBelem - 234985.4 * ParishBenfica + 126794.2 * ParishCampodeOurique - 63518.2 * ParishCampolide - 53855.9 * ParishEstrela - 267615.3 * ParishLumiar - 88231.2 * ParishMarvila + 35976.2 * ParishMisericordia - 82982.8 * ParishOlivais - 158743.8 * ParishParquedasNacoes - 124265.3 * ParishPenhade-Franca + 238416.6 * ParishSantaMariaMaior + 35358.1 * ParishSantoAntonio - 127298.9 * ParishSaoDomingosdeBenfica + 6808.5 * ParishSaoVicente.

In model 2, the result is that: price = $1.121*10^8$ - $1.004*10^5$Bedrooms + $8.935*10^3 Bedrooms^2$ - $1.018*10^5$Bathrooms + $4.858*10^4 Bathrooms^2$-$4.918*10^2$ $Bathrooms^3$ + $9.314*10^3$AreaNet - $3.647*10 AreaNet^2$ + $1.142*10^{-1} AreaNet^3$ - $1.669*10^{-4} AreaNet^4$ - $2.893*10^6$Latitude - $7.449*10^4$PropertySubTypeApartment + $3.845*10^4$PropertySubTypeDuplex - $1.893*10^5$PropertySubTypeDwelling + $3.796*10^6$PropertySubTypeIsolated Villa + $2.031*10^5$PropertySubTypePenthouse - $2.190*10^5$PropertySubTypeStudio + $1.759*10^5$PropertySubTypeTownhouse Dwelling + $7.321*10^4$ParishAlcantara + $8.883*10^4$ParishAlvalade - $3.704*10^4$ParishAreeiro - $1.344*10^4$ParishArroios + $1.065*10^5$ParishAvenidas Novas - $5.466*10^4$ParishBeato - $1.069*10^4$ParishBelem - $9.412*10^4$ParishBenfica + $1.816*10^5$ParishCampodeOurique + $6.214*10^3$ParishCampolide + $4.294*10^4$ParishCarnide - $4.557*10^4$ParishEstrela + $1.336*10^4$ParishLumiar + 1.025e+05ParishMarvila + $3.077*10^4$ParishMisericordia + 8.712e+04ParishOlivais + $9.438*10^4$ParishParque das Nacoes - $5.690*10^4$ParishPenha de Franca + $1.069*10^4$ParishSanta Clara + $1.062*10^4$ParishSanta Maria Maior + $9.069*10^4$ParishSanto Antonio + $3.446*10^4$ParishSao Domingos de Benfica + $3.026*10^4$ParishSao Vicente

References

[1] Ailin Zhang. "STAT4130_all_lecture_slides.pdf"(2023). UMJI-SJTU, Shanghai. [Online; accessed 24-7-2023].

[2] OpenML. "Lisbon-House-Prices dataset"(2023). https://www.openml.org [Online; accessed 24-7-2023].