There are certain criteria we are using to build the model: (1) We are giving equal weightage to all the different amenities (Rooms, Bathrooms, Elevator, Attic, Terrasse, Parking, Kitchen, Type and Yard) and summing them up, and we think that if we have more amenities it will certainly lead to increase in prices when having less amenities so we are using the sum of the amenities to build on our models.

We built 6 models, and our intuition goes like this for each of the model:

Model 1: The first model has dependent variable price, and the rest of the independent variables are area, rooms, bathrooms and amenities (as mentioned above) and even though the p-values that we got were good we thought we could do better by building a model that has better variance between the fitted values and the residuals. The R^2 values also were extremely low for this model.

```
Call:
lm(formula = Price ~ m2 + Rooms + Bathrooms + Amenities, data = bardos)

Residuals:
    Min      1Q  Median      3Q     Max
-206073  -33723   -1121   26965  355437

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -41591.2    11210.2  -3.710 0.000236 ***
m2            3242.6      158.9  20.412  < 2e-16 ***
Rooms       -13476.6     4463.8  -3.019 0.002694 **
Bathrooms    25806.7     7475.3   3.452 0.000614 ***
Amenities    12472.8     3432.3   3.634 0.000315 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 64890 on 408 degrees of freedom
Multiple R-squared:  0.7328,    Adjusted R-squared:  0.7302
F-statistic: 279.7 on 4 and 408 DF,  p-value: < 2.2e-16
```

Model 2: After running the log (Price) versus the same independent variables as the Model 1, the p-values improved further for all the variables. Further we saw that model was not responding very well with the number of rooms present in an apartment as the p-values were very large and this was our intuition behind removing the number of rooms from the mode.

```
Call:
lm(formula = log(Price) ~ m2 + Rooms + Bathrooms + Amenities,
    data = bardos)

Residuals:
     Min       1Q   Median       3Q      Max
-0.65075 -0.13631 -0.00275  0.11540  0.74802

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 1.134e+01  3.985e-02 284.498  < 2e-16 ***
m2          8.849e-03  5.647e-04  15.671  < 2e-16 ***
Rooms       8.799e-03  1.587e-02   0.555 0.579488
Bathrooms   9.455e-02  2.657e-02   3.558 0.000417 ***
Amenities   6.330e-02  1.220e-02   5.188 3.35e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.2306 on 408 degrees of freedom
Multiple R-squared:  0.6903,    Adjusted R-squared:  0.6873
F-statistic: 227.4 on 4 and 408 DF,  p-value: < 2.2e-16
```

Model 3: The model without the rooms has better p-values and the R^2 values also were better and now we had to see what other things that we need to incorporate to improve the variance between the independent variables and the residuals as they seemed to cluster around the same area.

```
Call:
lm(formula = log(Price) ~ m2 + Bathrooms + Amenities, data = bardos)

Residuals:
     Min       1Q   Median       3Q      Max
-0.65024 -0.13731 -0.00529  0.11746  0.74944

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 1.135e+01  3.473e-02 326.694  < 2e-16 ***
m2          8.996e-03  4.979e-04  18.067  < 2e-16 ***
Bathrooms   9.699e-02  2.618e-02   3.704 0.000241 ***
Amenities   6.326e-02  1.219e-02   5.189 3.33e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.2305 on 409 degrees of freedom
Multiple R-squared:  0.6901,    Adjusted R-squared:  0.6878
F-statistic: 303.6 on 3 and 409 DF,  p-value: < 2.2e-16
```

Model 4: We realized that we can incorporate the information about individual city zones too and see how the model behaved. We incorporated the information and saw the p-values being the same but not getting worse, the fitted values and the residuals showed lower variance, but we still saw space for improvement.

```
Call:
lm(formula = log(bartres.Price) ~ ., data = barmodel4)

Residuals:
     Min       1Q   Median       3Q      Max
-0.61357 -0.10993  0.00392  0.11414  0.72168

Coefficients:
                               Estimate Std. Error t value Pr(>|t|)
(Intercept)                  11.8108436  0.0739528 159.708  < 2e-16 ***
bartres.m2                    0.0074590  0.0004863  15.339  < 2e-16 ***
bartres.Bathrooms             0.0766905  0.0232001   3.306  0.00103 **
bartres.Amenities             0.0721610  0.0108132   6.673 8.35e-11 ***
X.City.Zone.Ciutat.Vella     -0.3105847  0.0647176  -4.799 2.25e-06 ***
X.City.Zone.Eixample         -0.1162935  0.0564515  -2.060  0.04004 *
X.City.Zone.Gràcia           -0.3476301  0.0570834  -6.090 2.65e-09 ***
X.City.Zone.Horta...Guinardó -0.2849428  0.0600034  -4.749 2.86e-06 ***
X.City.Zone.Les.Corts        -0.1617665  0.0731701  -2.211  0.02761 *
X.City.Zone.Nou.Barris       -0.4802362  0.0602012  -7.977 1.59e-14 ***
X.City.Zone.Sant.Andreu      -0.3839391  0.0573456  -6.695 7.31e-11 ***
X.City.Zone.Sant.Marti       -0.3765282  0.0603713  -6.237 1.14e-09 ***
X.City.Zone.Sants...Montjuïc -0.3518133  0.0682868  -5.152 4.06e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.201 on 400 degrees of freedom
Multiple R-squared:  0.7694,    Adjusted R-squared:  0.7625
F-statistic: 111.2 on 12 and 400 DF,  p-value: < 2.2e-16
```

Model 5: The final thing that we added here was the interaction effect between the amenities and the area of the house which was kind of a game changer for the complete model as we saw our p-values change even though R^2 value actually didn’t improve. All the fitted values and independent variables versus the residuals showed a better plot. And this was our intuition to finally choose this model.

```
Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)       12.4867048  0.0725925 172.011  < 2e-16 ***
`M2 x Amenities`   0.0010513  0.0001231   8.540 2.81e-16 ***
Bathrooms          0.2271302  0.0252409   8.999  < 2e-16 ***
`Ciutat Vella`    -0.6472549  0.0734162  -8.816  < 2e-16 ***
Eixample          -0.3353144  0.0658408  -5.093 5.44e-07 ***
Gràcia            -0.5623939  0.0667729  -8.422 6.64e-16 ***
`Horta Guinardó`  -0.5800467  0.0682627  -8.497 3.85e-16 ***
`Les Corts`       -0.4342110  0.0859757  -5.050 6.71e-07 ***
`Nou Barris`      -0.7872294  0.0676926 -11.629  < 2e-16 ***
`Sant Andreu`     -0.6403111  0.0656968  -9.746  < 2e-16 ***
`Sant Marti`      -0.6545270  0.0692669  -9.449  < 2e-16 ***
`Sants Montjuïc` -0.6050749  0.0801044  -7.554 2.89e-13 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’

Residual standard error: 0.2466 on 401 degrees of freedom
Multiple R-squared:  0.6521,    Adjusted R-squared:  0.6426
F-statistic: 68.33 on 11 and 401 DF,  p-value: < 2.2e-16
```