Problem 3

We are going to study the price of houses using the designed linear model.

We fit the model on our training set and obtain the following parameters considering interaction with the categorical variable bathrooms for all the covariates:

(Intercept)	2950.82579	485.38799	6.079	1.71e-08	***
two.bathroomsTRUE	76.45448	749.35912	0.102	0.91892	
two.bathroomsFALSE:footage	23.86056	1.58780	15.027	< 2e-16	***
two.bathroomsTRUE:footage	25.50310	1.50318	16.966	< 2e-16	***
two.bathroomsFALSE:age	-10.04729	4.88797	-2.056	0.04215	*
two.bathroomsTRUE:age	5.34187	4.54048	1.176	0.24189	
two.bathroomsFALSE:renovation	-7.99892	18.16029	-0.440	0.66045	
two.bathroomsTRUE:renovation	2.70081	17.10305	0.158	0.87481	
two.bathroomsFALSE:transport	0.70526	1.18230	0.597	0.55203	
two.bathroomsTRUE:transport	-0.17971	1.22098	-0.147	0.88325	
two.bathroomsFALSE:center	-0.11016	0.02268	-4.856	3.91e-06	***
two.bathroomsTRUE:center	-0.07404	0.02637	-2.808	0.00588	**
two.bathroomsFALSE:supermarket	-0.25634	1.01536	-0.252	0.80114	
two.bathroomsTRUE:supermarket	0.27011	1.02672	0.263	0.79297	
two.bathroomsFALSE:park	-0.17230	0.12889	-1.337	0.18398	
two.bathroomsTRUE:park	-0.35496	0.14485	-2.450	0.01581	*

And sigma is 462.8797
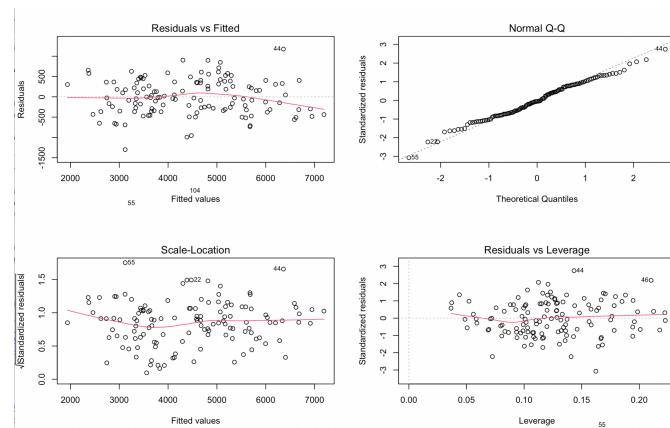
We can trust the model since by looking at the residual we see no patterns so the homoscedastic assumption is met; they seem to be gaussian and there are no leverage points. The Shapiro test on the residuals gives a p-values of 0.6531 so we confirm the gaussianity intuition on the residuals.

The VIF values are big in some cases so we suppose exists collinearity between variables.

By Performing a Lasso regression with parameter lambda = 45 we get keep following covariates.

(Intercept)	2954.28242488
# two.bathroomsFALSE:footage	20.59126191
# two.bathroomsTRUE:footage	24.33863521
# two.bathroomsFALSE:age	-4.25396034
# two.bathroomsTRUE:age	0.84029839
# two.bathroomsFALSE:center	-0.09434332
# two.bathroomsTRUE:center	-0.02832479
# two.bathroomsTRUE:park	-0.13420861

Via cross validations we find that the best lambda is 11 and the corresponding coefficients are:

(Intercept)   3076.30865895
two.bathroomsFALSE:footage   23.01042330
two.bathroomsTRUE:footage   25.29765737
two.bathroomsFALSE:age   -9.09855389
two.bathroomsTRUE:age   4.54539917
two.bathroomsFALSE:renovation   -4.11485707
two.bathroomsFALSE:center   -0.10960538
two.bathroomsTRUE:center   -0.06355008
two.bathroomsFALSE:park   -0.10227143
two.bathroomsTRUE:park   -0.29203071

The price prediction for the chosen apartment is 3580.72€