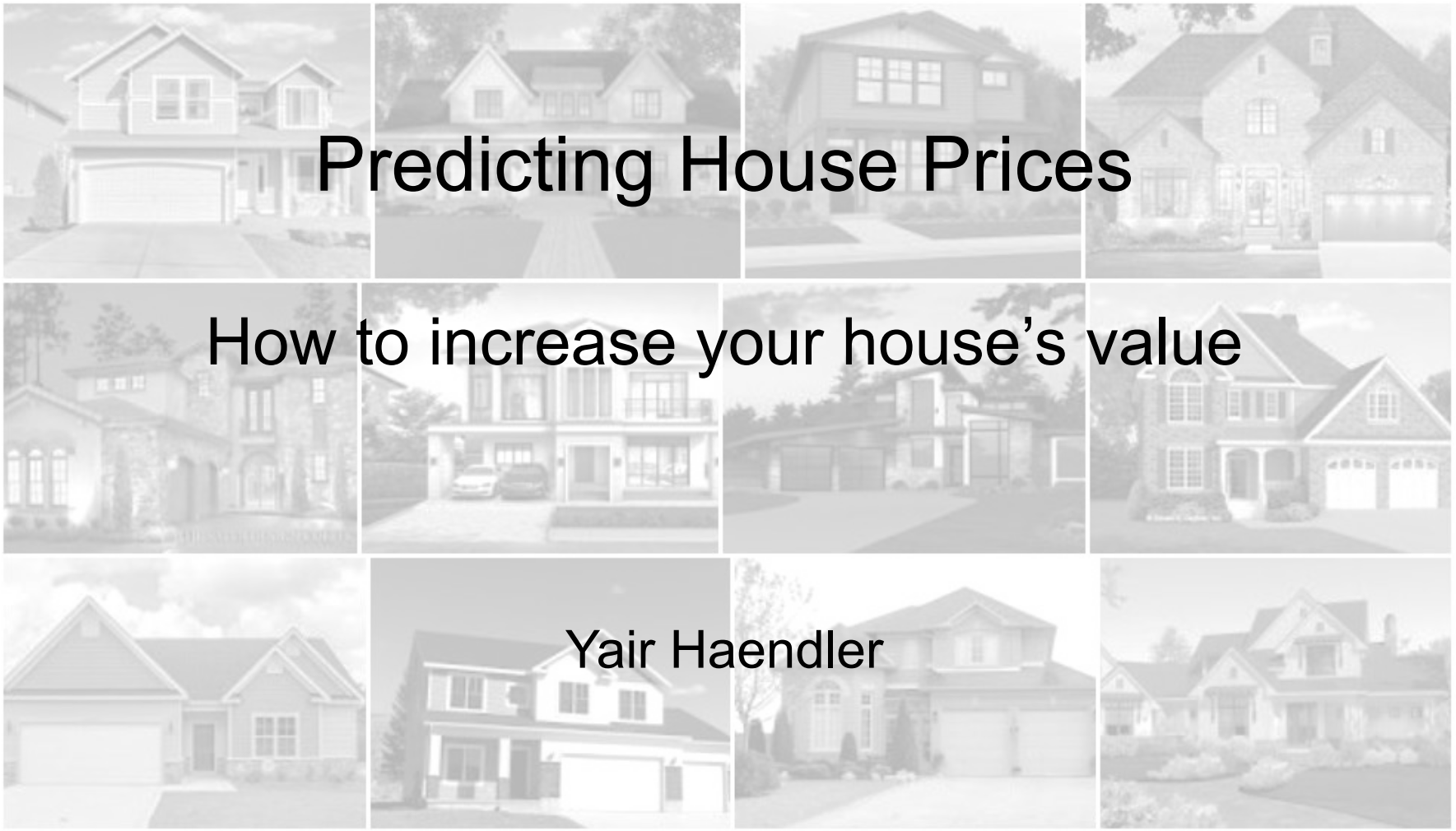


Predicting House Prices

How to increase your house's value

Yair Haendler

Dear residents,

thank you again for contacting us last week concerning your future plans to sell all of your houses due to increasing cases of bear sighting in your area.

As requested, we are presenting you today the results of an analysis of house prices.

Based on this analysis, we will advise you on what characteristics of your homes you should focus on in order to increase their value for sale.

There is an enormous number of house characteristics that might potentially affect its market value.

Our goal in this analysis is to individuate those characteristics related to **things you can modify** (add, remove, etc.) in your home, and see which ones are most important in order to increase your house's sale price.

Step 1 - Selecting features:

Among all the features in our data set, extract those related to things that homeowners can modify in their house.

For example:

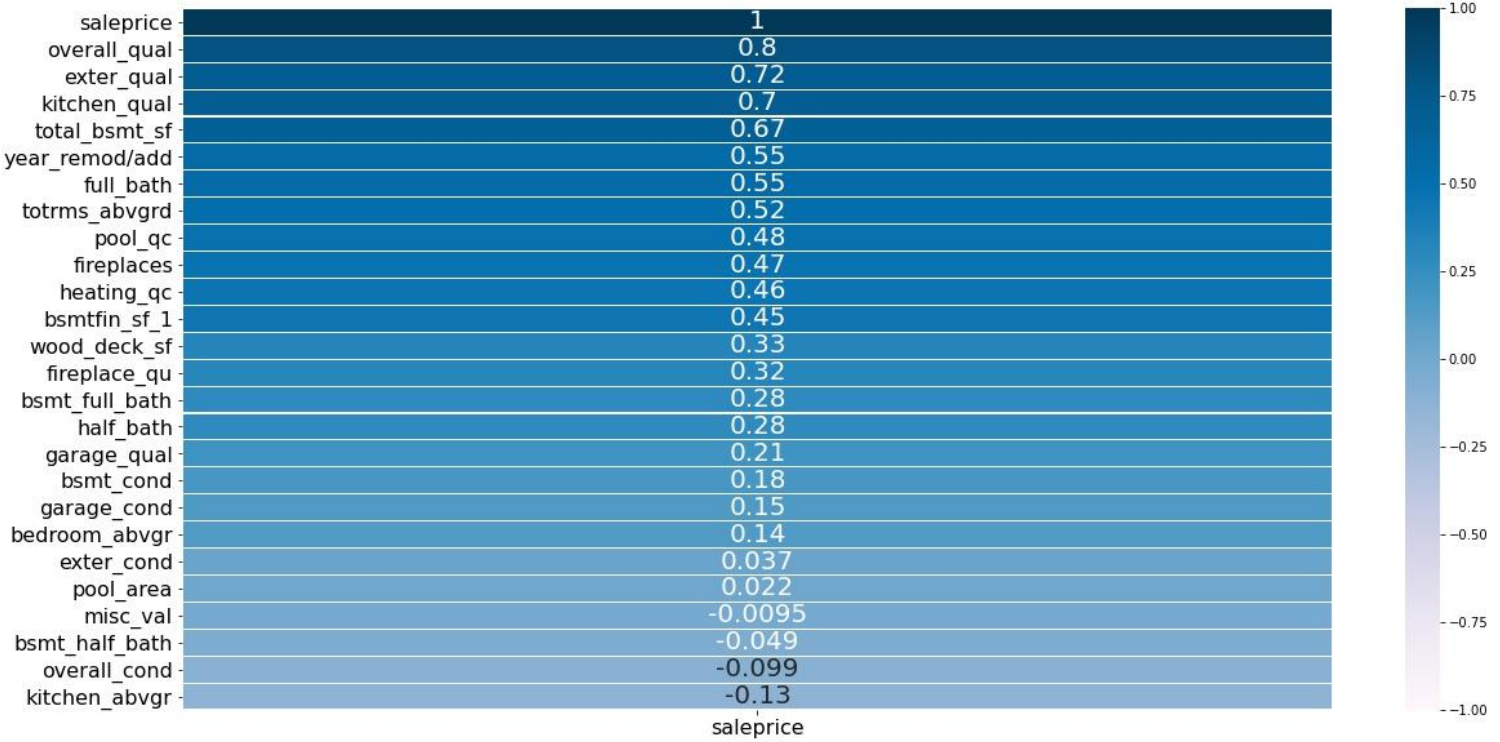
- The quality of the material on the exterior
- The type of electrical / heating / A/C systems
- Whether there's a pool in the garden or not

Step 1 - Selecting features:

By contrast, features that do not depend (so easily) by the homeowners were excluded, for instance:

- Length of street connected to property
- Type of zone in which the property is located (residential, industrial, etc.)
- 1st or 2nd floor square footage
- Number of stories in the house

Correlations of various house features with Sale Price



Step 2 - Dealing with outliers:

Outliers in different features were identified visually using scatter plots.

Overall, we **removed 10 data points** categorized as outliers, out of 2044 observations in our data set.

Step 3 - Preparing features:

Categorical features were coded with *Backward Difference Encoding* (from the ‘category_encoder’ module).

This contrast coding system compares each category in the feature to the previous one - features values re-ordered according to the comparisons we want to make.

Particularly suitable for cases in which the categories are ordered increasingly or decreasingly.

Step 3 - Preparing features:

For example:

The feature “**basement_condition**”:

Rating of the overall condition of the basement - 1 (worst) to 5 (best).

0 or NA = no basement.

Step 3 - Preparing features:

Thus, for the feature categories : **‘no_basement’, ‘1’, ‘2’, ‘3’, ‘4’, ‘5’** -

Backward Difference Encoding would create 5 comparisons:

1 - no_basement
2 - 1
3 - 2
4 - 3
5 - 4

With which we get a sense of the effect as related to the internal ordering of the feature categories.

Step 3 - Preparing features:

Numeric features were scaled (z-score) using *Standard Scaler*, in order to put all of them on the same scale.

Step 4 - Modeling:

We started out including 31 variables, resulting in 70 slope coefficient parameters.

We split the data into a training set (2/3) and a test set.

A Linear Regression Model resulted as overfit. Hence, we used Lasso regularization.

Step 4 - Modeling:

With the cross-validation Lasso model, some coefficients (not many) were zeroed out.

Moreover, we removed from the model other features that, in a re-assessment of their suitability for the model, were suspected of being **collinear / redundant** with other features.

Step 4 - Modeling:

The Lasso model on which our recommendations are based included 50 slope parameter coefficients.

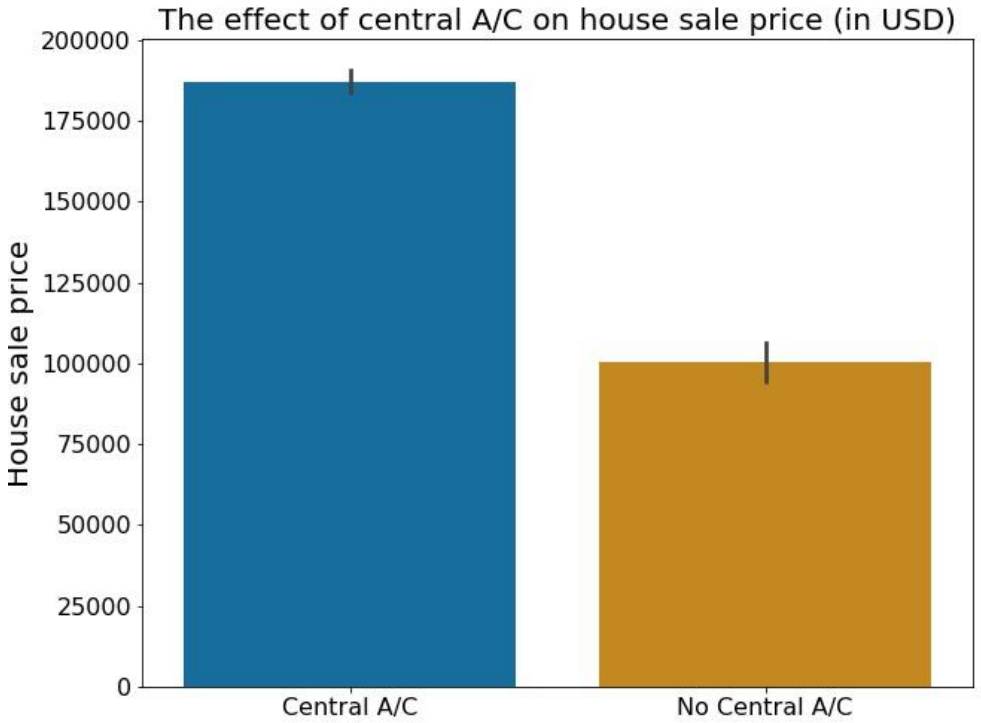
R-squared **training** set: **0.86**
R-squared **testing** set: **0.83**

RMSE training set: $30,438
RMSE testing set: $31,619

We're still working on ways to reduce its variance.

Step 5 - Interpretation of coefficients & recommendations:

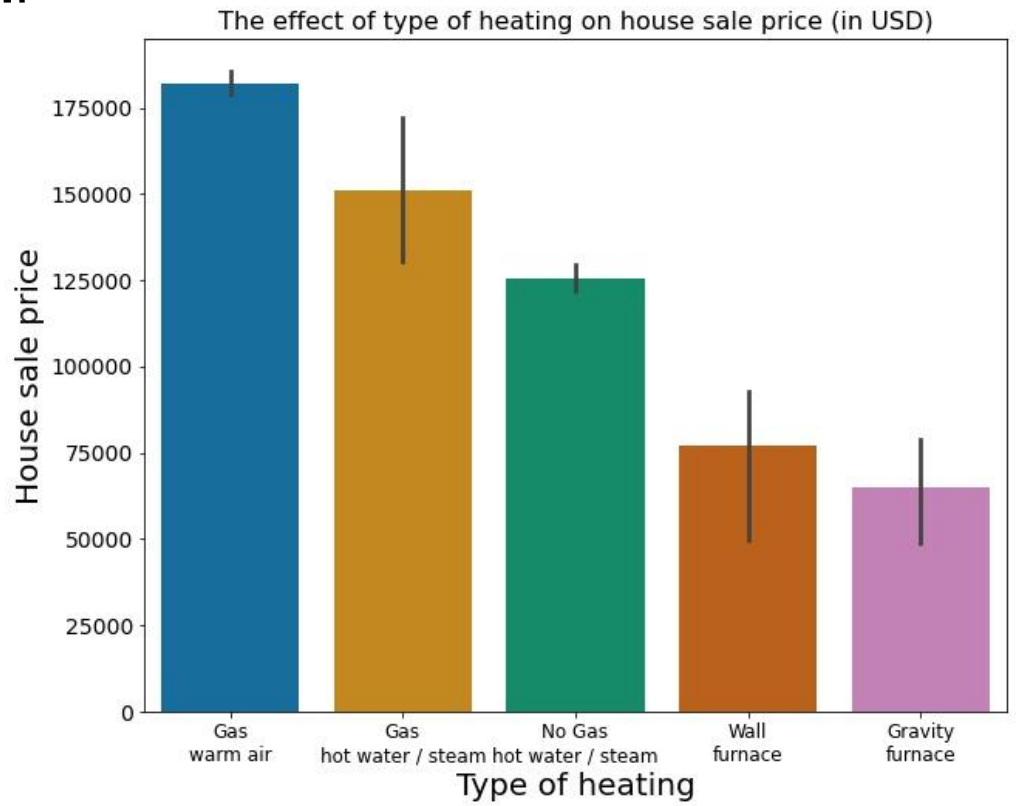
The effect of **Central A/C** (+ $2,796)



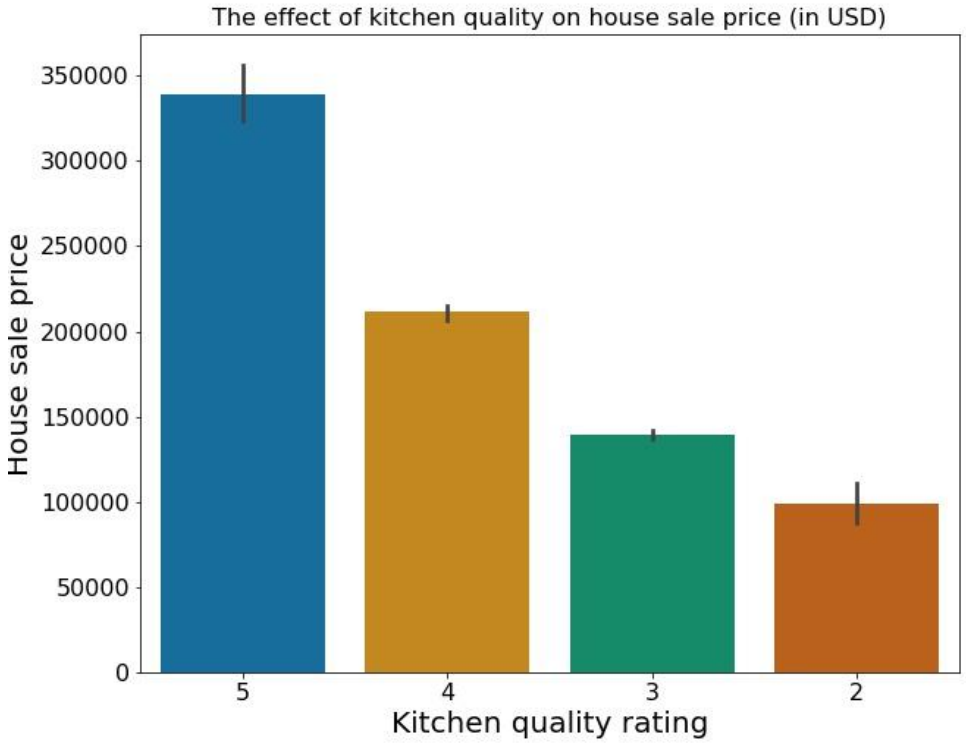
The effect of **Heating System**:

Wall vs. Gravity (+ $16,184)

No Gas vs. Wall (- $41,459)

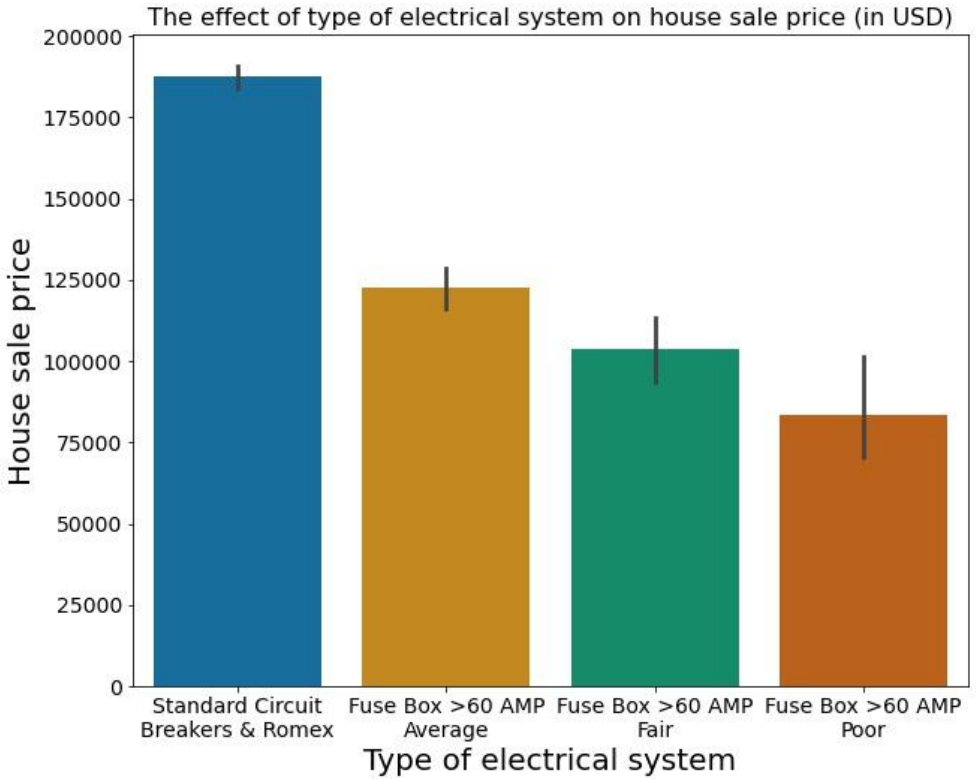


The effect of **Kitchen Quality**:
(+ $10,093)

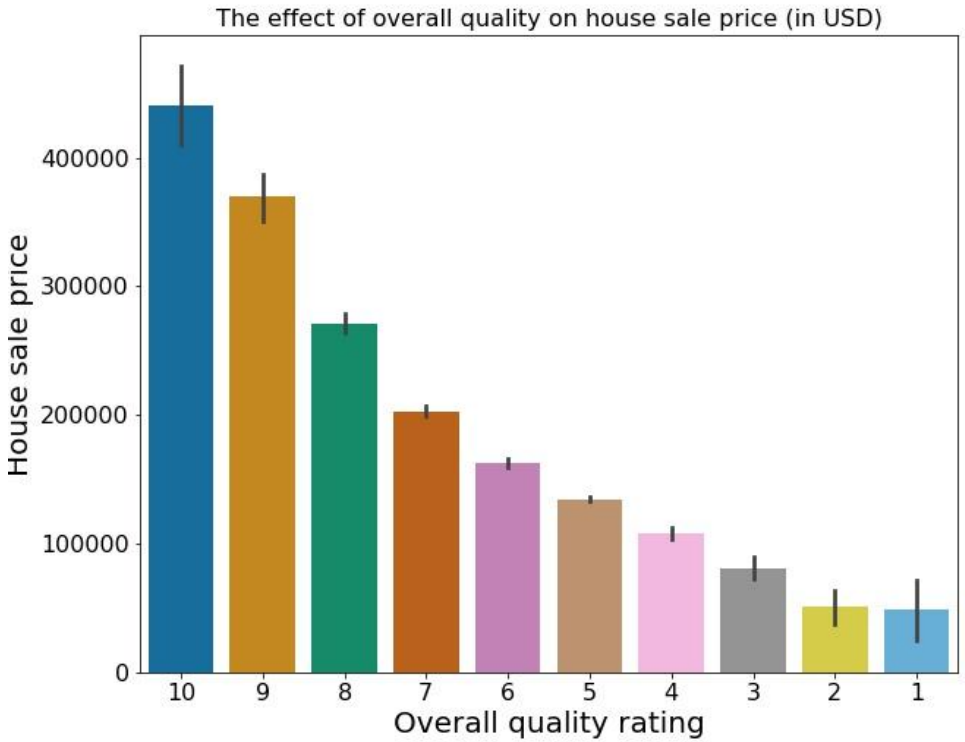


The effect of **Electrical System**:

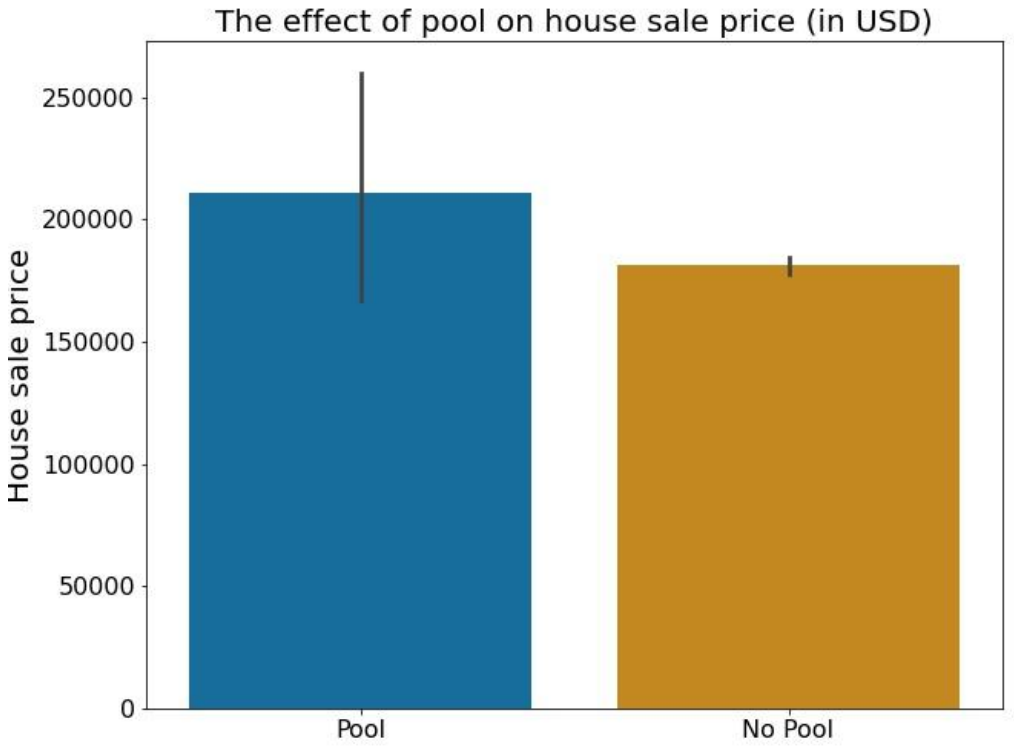
Fuse Poor vs. Fuse Fair (+ $13,126)



The effect of **Overall Quality Rating** (+ $18,556)



The effect of **Swimming Pool** (+ $23,728)



In sum, you will be receiving from us the full report with the list of recommendations.

If you have any questions, please don’t hesitate to contact us.

Good luck!