



PREDICTING REAL-ESTATE PRICE IN RIYADH (MVP)

PREPARED BY

Hayat Aldhahri
Muneera Alshunaifi

DATE

11 Oct 2021

Table of Contents

Objective	2
MVP	2

Table of Figures

Figure 1 Heatmap of the features	2
Figure 2 Pairplot of corr between features.....	3
Figure 3 Heatmap after treating outliers in Price.....	3

Objective

Project's main objective is to create a model that predicts the real-estate prices using one of the most commonly machine learning method which is Linear Regression.

MVP

The real-estate data were scraped and collected from AQAR, Propertyfinder websites which are a property search websites. The data were properly cleaned, and EDA were performed to further analyze and explore patterns and correlations between the variables/features.

Here is some of the results we have explored :

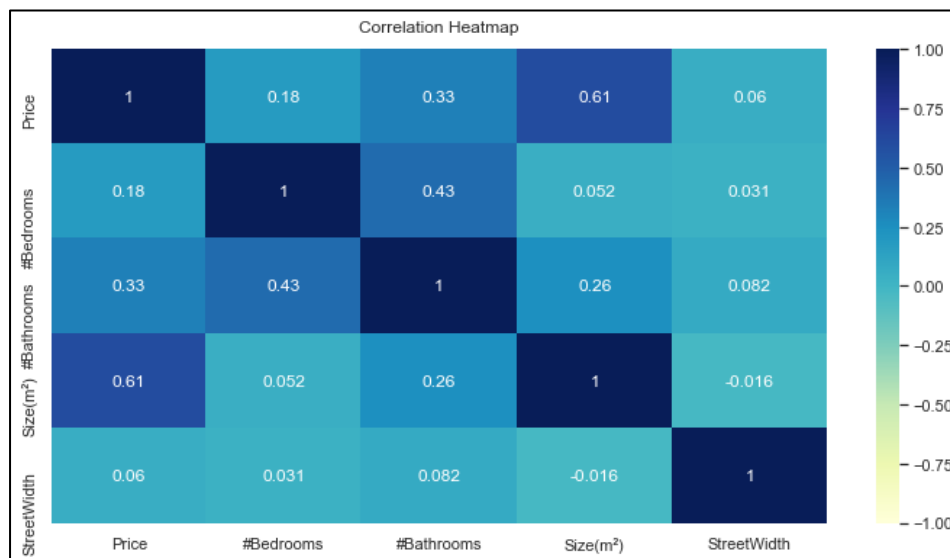


Figure 1 Heatmap of the features

The above figures show the correlations between the five features, the main insights are:

- The "Size" feature has high positive correlation 0.61 with our target "Price"
- The "Bathrooms" feature has positive correlation 0.33 with our target "Price"
- The "Bedrooms" feature has low positive correlation 0.18 with our target "Price".

Before feeding the data to the model, outliers were removed from “Price” feature, and since the “District” feature is an object type of data, it was encoded to numeric to be able to perform modeling.

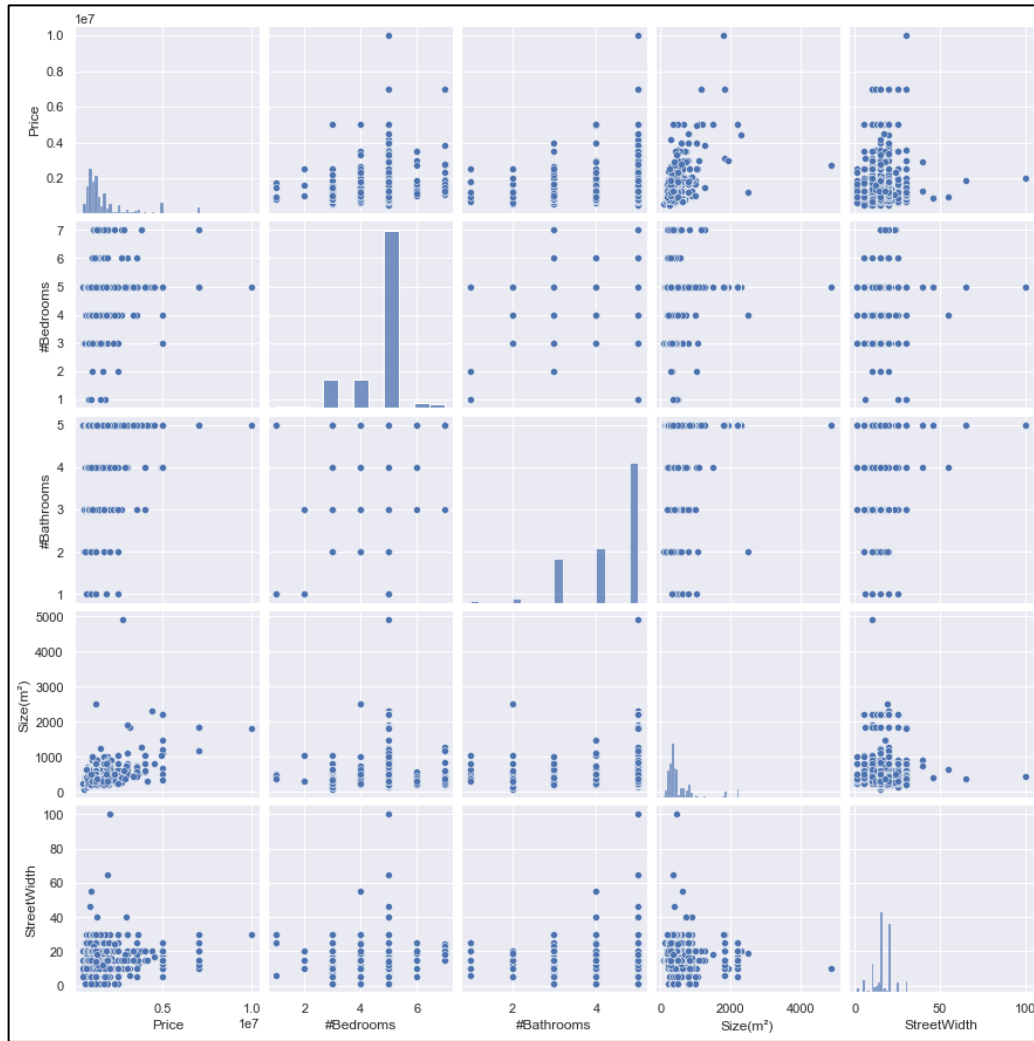


Figure 2 Pairplot of corr between features

The figure above shows the correlation between features, we can ensure that Size has properly linear correlation with our target “Price” .

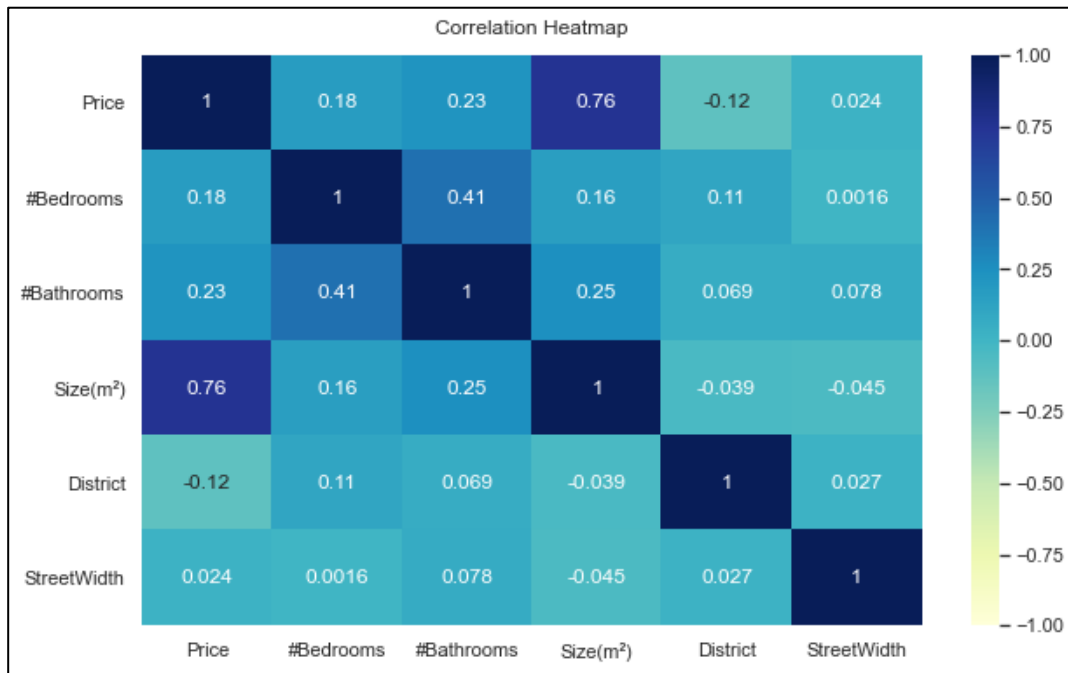


Figure 3 Heatmap after treating outliers in Price

Heatmap of the features shown above illustrates the correlations after removing the outliers and performing data pre-processing. It clearly emphasizes on the positive correlation between the features “Price” and “Size”.

For the linear regression model, the data were split into train and target to perform a 80/20 train/validation split. The reported results are shown below:

- Linear Regression Test (R^2): 0.632
- Cross Validation for Linear Model Score: 0.463

Currently, we need to do more analysis in order to decrease the R^2 and increase the cross validation score, also, we need to do more plots to see the fit of the model. Furthermore, for the last step we want to test the data on real user input.