

Predicting Apartments Prices.

By Ahmed Aloliwi

Abstract:

In the real estate market in booming cities, there is a great disparity in the price of real estate products for each city. Often times it is not controlled and causes mistrust between the seller and the buyer.

With the rise of deep learning in recent, our reliance on artificial intelligence models and machine learning will contribute to raising the price accuracy of real estate products.

The data will be tried and trained on three algorithms in deep learning GradientBoostingRegressor, RandomForestRegressor and XGBRegressor. We will choose the one that is most accurate in predicting the final property price.

We will train our models using training dataset consisting of 1335 properties and then test them against our testing dataset with 571 properties.

We will check our R2 score for each model and we will select the predictions from the model, which is closest to value of 1. R2 scores range from 0 to 1.

Based on best R2 score we will match actual values with predicted values and see the percentage difference.

Design:

The data was obtained by Kaggle platform and includes 38 column and 1906 rows.

I will depend on major data such as the price of the housing unit, the number of rooms and toilets, as well as the quality of the housing unit to predict the price of Apartments units.

Data:

The dataset contains 1906 with 12 features. A few features highlights the price of unit include the price per Sq2 and the quality of the unit and the number of bedroom and the number of bathroom and latitude/longitude.



Algorithms:

- 1- Converting categorical features to binary dummy variables
- 2- Combining particular dummies and ranges of numeric features to highlight strong signals for predicting the house price identified during EDA.

Models:

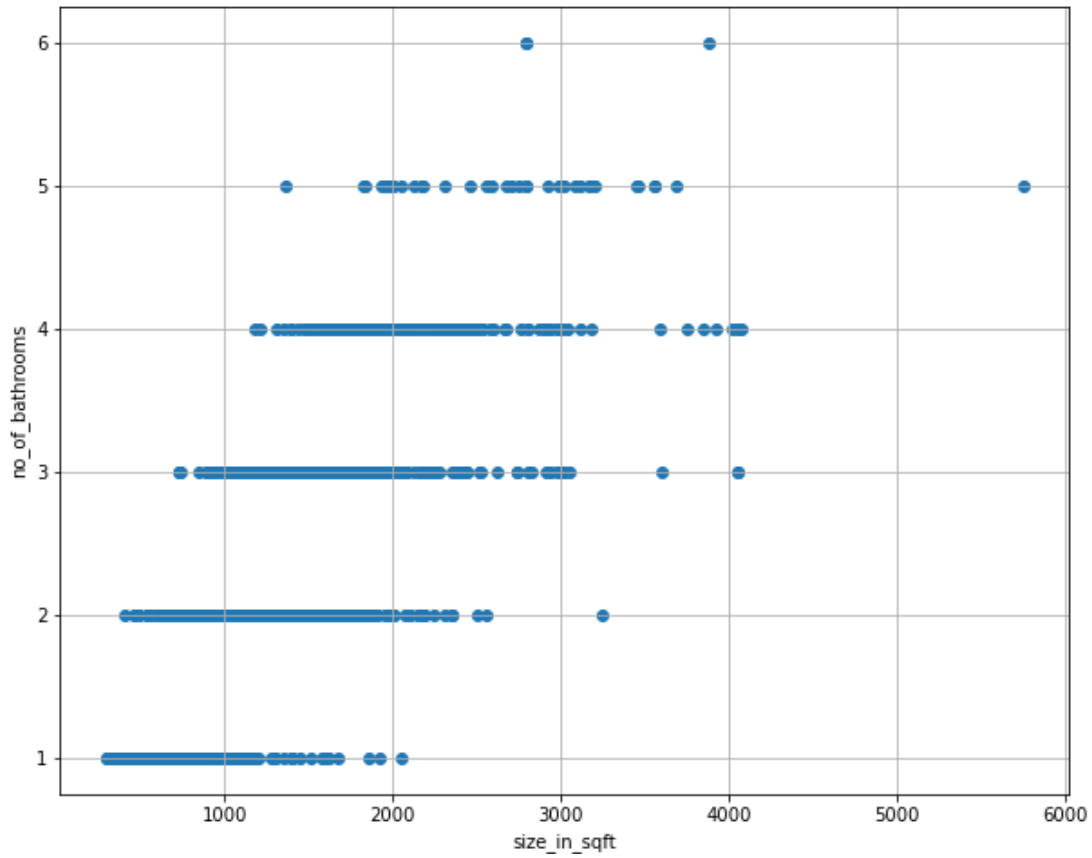
I trained and tested three model and finally I select **Randomforestregressor** because it has high R2 score which was closest to one was 0.9923263147255801 compared to other models.

Tools:

- I used **python packages for data science models** like (sklearn, xgboost, etc.).
- I used Google cloud Colab to write my python script.
- Numpy and Pandas for data manipulation
- Scikit-learn for modeling
- Matplotlib and Seaborn for plotting

Communication:

I will mention in the slide some of the plot that help to find the correlation between some of the features and the target from the model that is predicting the price of the property.



- We have 1-bathroom apartments ranging up to 5 million, most probably because of the locality.
- We have two bathrooms apartments ranging from sub 1 million to up to 5 million.
- We have three bathrooms apartments ranging from 1 million to up to 5 million.