Predict real estate SMP using regression Models.

Omar Aloufi
SDAIA T5-Data Science Bootcamp

Abstract:

The Goal from this project was to use a regression model to predict real Estate Square per meter price (SPM) in Madinah, based on data that available on MOJ open data.to help real estate byer to find the suitable Price. Three new features were added to study how the Price of real estate Affected by its location from AL masjid AL Nabawi.

Data:

11076 sample with 10 features about deals executed in Madinah in 2019, obtained from Ministry of Justise open data.

Exploratory Eata Analysis (EDA):

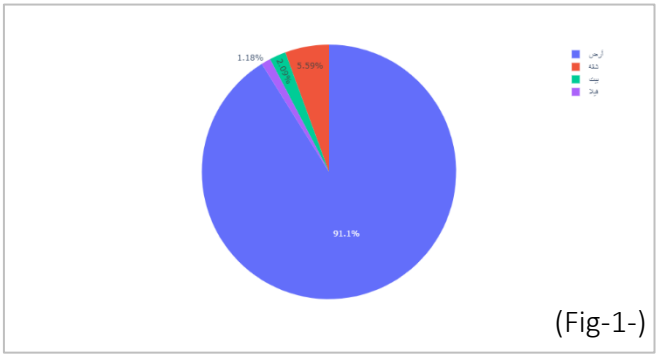
We Explore the data and find that 91% (Fig-1-) of the deals is Land purchases, furthermore, we find that more than 40% of deals is real estate Located East of Madinah (Fig-2-). On top of that we found the nearest the real estate from AL masjid AL Nabawi SMP increase dramatically (Fig-3).

Modeling:

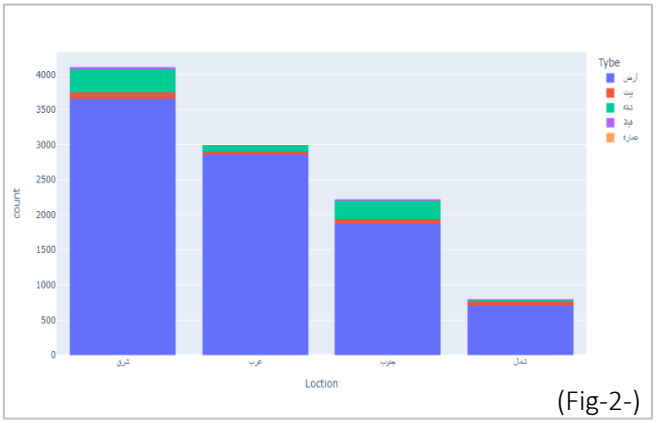
We used K-fold cross validation with five regression model (1. Leaner, 2. Ridge, 3. Polynomial, 4. Random Forest Regressor and 5. Xgboost Regressor) by using, R^2 to score models number (1,2,3) and MSE, RMSE to score models number (4,5). Another point is to select model with best score to make prediction about square meter price (SMP). We couldn't get high accuracy on the models because the data has a high variance, and inconstant. Highest score was 0.5 with Xgboost Regressor (Fig-4-).

Conclusion:

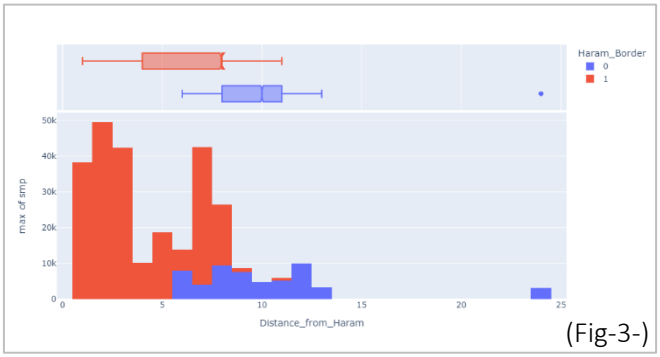
The price of real estate price effected by many factors, so we need to search for new features such as (Utility, services, attraction,). Also, we will research for best modeling and fitting parameters to give us a better result.



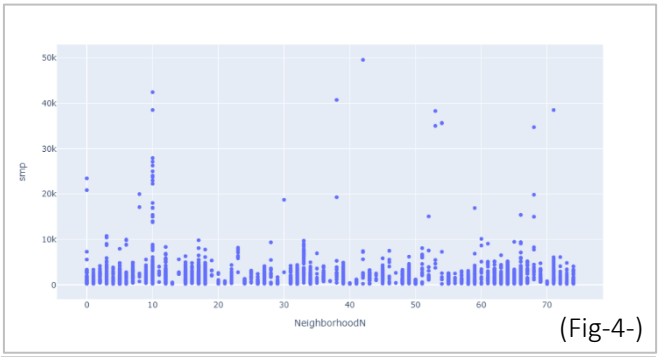
(Fig-1-)



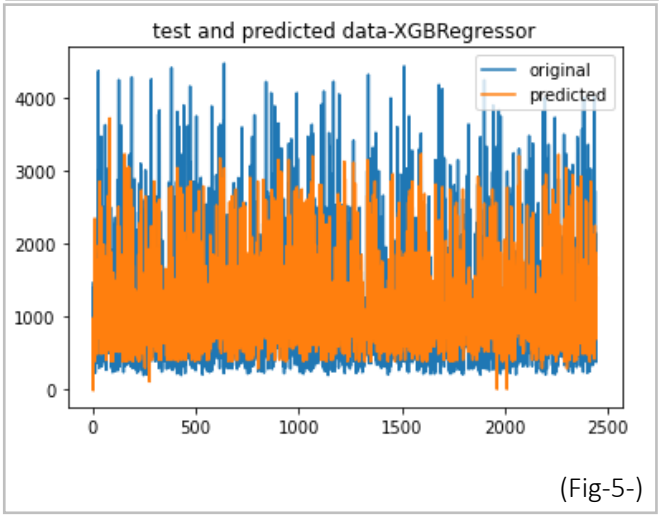
(Fig-2-)



(Fig-3-)



(Fig-4-)



(Fig-5-)