Predicting Houses Prices Using Machine Learning

# About

When coming to buy something anyone wants to buy less as contradictory sell at high price as much as possible. This tendency has common on coming to buying house prices as well. As the middle-income buyer as always suffer because of high volatility on prices and features.

No buyer never looked on previous selling prices and made some analysis on it. Simply, they negotiate basis realtor prices or other realtor prices. This makes high volatility on the prices.

As technology advances this problem can find solution using data science techniques. Total project as based on the India selling house prices data which was licensed on [GPL v2](https://www.gnu.org/licenses/old-licenses/gpl-2.0.en.html). This [dataset](https://www.kaggle.com/anmolkumar/house-price-prediction-challenge) was available on open to use.

## Features / Attributes

Features or columns represents outcome of the result. Each feature contributes and maintain some relationship on other features. By combining all features will result to final result.

### Feature and information

**Posted\_BY**

This feature tells about who posted about this record. As see below picture most of the houses posted by dealer and owner.

Chart, bar chart

Description automatically generated

**UNDER\_CONSTRUCTION**

About feature reveals this house is on under-construction or fully constructed.

Chart, bar chart

Description automatically generated

Mostly houses were already constructed and only 5000 houses were still in construction.

**RERA**

Real estate government approved layout. This act was brought 2016. [For more information link](https://www.up-rera.in/pdf/reraact.pdf).

Chart, bar chart

Description automatically generated

Even most of sites were not RERA approved. Means they have applied or approved by local governments.

**BHK\_NO**

Total bed rooms available on all records. Most of the records two 2,3 and 1 bedrooms.

Chart, bar chart

Description automatically generated

This looks left skewed 90% accuracy is from 1 bed room to 4 bed room. This feature can also reveals high prices in cities.

**BHK\_OR\_RK**

BHK or RK this feature will reveals BHK (Bed,Hall,Kitchen) & RK(Room, Kitchen) of the house.

Chart, bar chart

Description automatically generated

In 29451 records, RK have 24 records and BHK have 29427. This equalent to nill.

**SQUARE\_FT**

Chart, bar chart

Description automatically generated

Here SqFt have divided into bins. Mostly majority sites have 2000 and above. This can relate to cost and cities as well. This can be group feature.

**READY\_TO\_MOVE**

Chart, bar chart

Description automatically generated

This feature directly relate under-construction. If under construction ready to move will be no.

**RESALE**

Chart, bar chart

Description automatically generated

This feature has high correlations with posted by. Most of builder and dealer will sell on if resale value is there.

**Chart, bar chart

Description automatically generated**

This relationship reveals how bedrooms was distributed according to price and construction. Seven bed room was 15 core and it is highest amount on all sites.