# Predicting Indian House Prices

# Introduction

* The Housing dataset contains information about different houses in India which in turn could be used for accurately predicting house prices. The data was originally a part of [House Price Prediction Challenge](https://www.machinehack.com/hackathons/house_price_prediction_beat_the_benchmark/overview) conducted by Machine Hack in Sept’20.
* This dataset has been collected over various property aggregators across India and each of the entries represents aggregate information about various features of homes located in various suburbs across India.
* The dataset provides 11 influencing factors which can be manipulated and trained to fit on a ML model to predict the house prices as accurately as possible.

# Objectives

* Since the buyers are just not concerned about the size (square feet) of the house and there are various other factors that play a key role to decide the price of a house/property.
* It can be extremely difficult to figure out the right set of attributes that are contributing to understanding the buyer's behavior.

# Data Description

* Train.csv - 29451 rows x 12 columns
* Test.csv - 68720 rows x 11 columns

# Attributes Description

|  |  |
| --- | --- |
| **Attribute** | **Description** |
| POSTED\_BY | Category Marking Who Has Listed the Property |
| UNDER\_CONSTRUCTION | Under Construction or Not |
| RERA | RERA (Real Estate Regulatory Authority) Approved Or Not |
| BHK\_NO | Number of Rooms |
| BHK\_OR\_RK | Type of Property |
| SQUARE\_FT | Total Area of The House in Square Feet |
| READY\_TO\_MOVE | Category Marking Ready to Move or Not |
| RESALE | Category Marking Resale or Not |
| ADDRESS | Address of the Property |
| LONGITUDE | Longitude of the Property |
| LATITUDE | Latitude of the Property |
| TARGET(PRICE\_IN\_LACS) | Sale Value of the Property |

# STTM

|  |  |  |
| --- | --- | --- |
| Target Column Name | Transformation | Source Column Name |
| POSTED\_BY | Owner : 0 Dealer : 1 Builder : 2 | POSTED\_BY |
| UNDER\_CONSTRUCTION | Both columns are giving  same information. | UNDER\_CONSTRUCTION |
| READY\_TO\_MOVE |
| RERA | Flow through Source | RERA |
| BHK\_NO | Flow through Source | BHK\_NO |
| BHK\_OR\_RK | BHK : 0 RK : 1 | BHK\_OR\_RK |
| SQUARE\_FT | log(SQUARE\_FT) | SQUARE\_FT |
| SQ\_PER\_ROOM | SQUARE\_FT/BHK\_NO | SQUARE\_FT |
| BHK\_NO |
| RESALE | Flow through Source | RESALE |
| CITY | ADDRESS.split(',').get(-1) | ADDRESS |
| LOCALITY | ADDRESS.split(',').get(-2) | ADDRESS |
| LONGITUDE | Linear Scaling | LONGITUDE |
| LATITUDE | Linear Scaling | LATITUDE |
| TARGET(PRICE\_IN\_LACS) | log(TARGET(PRICE\_IN\_LACS)) | TARGET(PRICE\_IN\_LACS) |