**Objective**:

The objective of the project is to predict the house prices in Bangalore.

**Outcome**:

To predict the house prices in Bangalore using Linear Regression.

**Tools Used**:

1. Python
2. NumPy and Pandas for data cleaning
3. Matplotlib for data visualization
4. sklearn for model building
5. Jupyter notebook as IDE

**Execution**:

The data is loaded from Kaggle dataset.  
  
Cleaned the data base to

* remove some unwanted columns based on the business knowledge
* remove columns with zero data
* remove outliers - if the property price is too high
* remove properties with more bathrooms than BHKs – looks data may not be correct

Feature Engineering is done to

* split the data and retain only the BHK info - for ex from 2 BHK retain only 2
* in 'total\_sqft' column there are price range like 2100 - 2850 - insert a mean value here and remove the rest
* remove locations with less than 10 properties
* remove residential properties which have less than 300 sq ft per bed room
* calculate price per sq ft and remove those rows which have extreme values
* remove those properties which have prices less for example 3 BHK costing less than 2 BHK - using z score to remove outliers
* convert the location column from string to number

Finally applied the Linear Regression model to predict the house price.  
  
Linear regression has its advantages, it may not capture all the complexities of house price prediction, especially if the relationship between features and house prices is highly nonlinear or if interactions between features are important. In such cases, more advanced techniques like decision trees, random forests, or gradient boosting may be more appropriate.