**House Price prediction**

**Tasks**

1. **Analyse the data understand the features of data**
2. **Clean the data**
3. **Add new feature(integer) for bhk (Bedrooms Hall Kitchen)**
4. **Explore total_sqft feature**
5. **Add new feature called price per square feet**
6. **Examine locations which is a categorical variable. Apply dimensionality reduction technique here to reduce number of locations**
7. **As a data scientist when you have a conversation with your business manager (who has expertise in real estate), he will tell you that normally square ft per bedroom is 300 (i.e. 2 bhk apartment is minimum 600 sqft. If you have for example 400 sqft apartment with 2 bhk than that seems suspicious and can be removed as an outlier. Remove such outliers by keeping our minimum thresold per bhk to be 300 sqft**
8. **Remove outliers per location using mean and one standard deviation**
9. **check if for a given location how does the 2 BHK and 3 BHK property prices look like also visualize**
10. **also remove properties where for same location, the price of (for example) 3-bedroom apartment is less than 2-bedroom apartment (with same square ft area). What we will do is for a given location**
11. **remove those 2 BHK apartments whose price_per_sqft is less than mean price_per_sqft of 1 BHK apartment**
12. **Plot scatter chart to visualize price_per_sqft for 2 BHK and 3 BHK properties again**
13. **Analyze the bathroom feature**
14. **Again, the business manager has a conversation with you (i.e. a data scientist) that if you have 4-bedroom home and even if you have bathroom in all 4 rooms plus one guest bathroom, you will have total bath = total bed + 1 max. Anything above that is an outlier or a data error and can be removed**
15. **Build a model**
16. **Train the model**
17. **Find the best model with high accuracy**
18. **Predict the price of house using the model**