**PROJECT**

**Chennai House price prediction**

I was provided with a csv file containing information regarding House prices in chennai. Firstly I checked what all information is present in the csv file. I found aht all columns are there and how many rows and columns are present. How many of these contain numerical values and how many are strings. My first task was to clean the data. Therefore I found the number of missing values and then I replaced all the missing values with the mean as there were very few missing values in the data. Then I found the distribution of data in each column also checked for any spelling or typo errors. I replaced all the spelling and typo errors accordingly. Then I did data visualization in order to know the dependence of sales price on various factors and observe the trend. I tried to observe the linear dependence of price on various factors mentioned in the file. The visualisions i used were scatter plots,bar graphs and line plots. I also formed the heat maps to know the correlation between various factors to understand better. I dropped all the features that were not required in modeling. Then I modeled using linear regression, random forest and xgs boost. I also calculated the scores to compare the models. On that basis I was able to make conclusions about the data and help the buyers and the seller for advice in investing in property in chennai according to the data provided.