**INTRODUCTION:**

I am doing project on real-estate data of Bengaluru which is a biggest and busiest city in India. By using this model, I can suggest a person to invest or to buy a property (Land, House) where he can get maximum profit with minimum investment and also predict the growth of the property in future by using statistical analysis technics.

**DATA:**

I have collected data from various sources. for instance, I got data related to price of the land from **www. Kaggel.com** which is a free and trustworthy data source to analysis. But it doesn’t have data related to house prices, so I found a web site called **99acers.com** in online who acts as a online mediator , they dint provided separate data set to download so I web scraped the data by using python beautiful soup and stored data in to csv .

**PROCESS:**

Clean and combine data(pythone)

Statistical

Technics (R)

Land data

House data

So as of now I have gathered all the data which require for my statistical analysis to reach my objective but still there are lot of anomalies in my data like missing values and unrelated data or missing data. so, I star my project initially by cleaning data and I also drop some columns of unnecessary data from it. I use python padas to perform these operations and I use RStudio’s for statistical analysis (ANOVA, REGRESSION) .