**ML Assignment 1 / Deadline**

You are given house\_price.csv which contains property prices in the city of Bangalore. You need to examine price per square feet do the following:

**Q1**. Perform basic EDA (**Score:1**)

**Q2**. Detect the outliers using following methods and remove it using methods like trimming / capping/ imputation using mean or median (**Score: 4**)

a) Mean and Standard deviation

b)Percentile method

c) IQR(Inter quartile range method)

d) Z Score method

**Q3**. Create a box plot and use this to determine which method seems to work best to remove outliers for this data? (**Score:1**)

**Q4**. Draw histplot to check the normality of the column(price per sqft column) and perform transformations if needed. Check the skewness and kurtosis before and after the transformation. (**Score:1**)

**Q5**. Check the correlation between all the numerical columns and plot heatmap. (**Score:1**)

**Q6**. Draw Scatter plot between the variables to check the correlation between them. (**Score:1**)

**Timely Submission** (**Score:1**)

**Total Marks : 10**

**Dataset :** https://drive.google.com/file/d/1UlWRYU0UglE2ex3iFse0J6eCLEU8g98P/view?usp=sharing

**Reference** :

https://youtu.be/Cw2IvmWRcXs?si=KYjsr1vThZNNrDUr

Learners, have a look into this video. It may be helpful for you to find outliers in a dataset.

For the percentile method, you can consider less than 5% and greater than 95% .