Project-2

(1) EDA each grap conclusion:

a) central tendency for latitude - 24.97
central tendency for longitude - 121.545

b) Inverse / hyperbolic relationship b/w response & x_3 (distance to nearest metro station). Non linear relat^n

c) So as the distance from nearest Metro station increases the price of house decreases.

d) x_4 is discrete r.v & now we obtain CT's & Std deviat^n if we fix locat^n & vary the BHK.

e) Now, obtain a regression plot b/w age of house & price. We observe that as age ↑, there is a decline in prices. It can be well estimated by regression plot.

f) In again we consider the effect of presence of convenience stores. So for each value of age if we have convenience stores price values tends to be high.

g) Again we can consider the effect for the same locat^n obtained from CT of latitude & longitude.

h) We obtain correlation heatmap to see the response & predictors are associated & to check how predictors are associated. We observe strong negative correlation b/w distance from Metro stat^n and no. of convenient stores.

Now multicollinearity does not exist, if it had existed then it shall affect linear kernel. So accordingly some action would have been taken To avoid multicollinearity

If I remove "distance from metro Stat" then accuracy was reduced indicates loss of inf. ∴ it was better to keep both. Further they both have different relat^n with response.

one is +ive & other is negative, so they both provide diff^t inf.

SVM → kernel = linear
= rbf
= polynomial

tuning hyperparameter

Conclusion :

There is certain other significant features missing which can help to improve accuracy.