**Multiple Linear Regression**

**Assignment**

**BACKGROUND:** The data for analysis contains information of Selling price of the house in million Rs. It contains Carpet area in square feet, Distance from nearest railway station, Number of schools within 2 km distance. The data has 198 rows and 5 columns.

**QUESTIONS-**

1. Import House Price Data. Check the structure of the data.
2. Split the data into Training (80%) and Testing (20%)
3. Build a regression model on training data to estimate average price of a House.
4. List down significant variables and interpret their regression coefficients.
5. What is the R2 and adjusted R2 of the model? Give interpretation.
6. Is there a multicollinearity problem? If yes, do the necessary steps to remove it.
7. Are there any influential observations in the data?
8. Can we assume that errors follow ‘Normal’ distribution?
9. Is there a Heteroscedasticity problem? If yes, do the necessary corrections before running the model again with corrections.
10. Run the corrected model and calculate the RMSE for the Training and Testing data.