



Bhartiya Vidya Bhavan's  
Sardar Patel Institute of Technology, Mumbai-400058  
Department of Computer Science and Engineering

## **ADVANCE DATA VISUALIZATION(EXP 5)**

**Name:** Darshit Bhagtani

**UID:** 2021700006

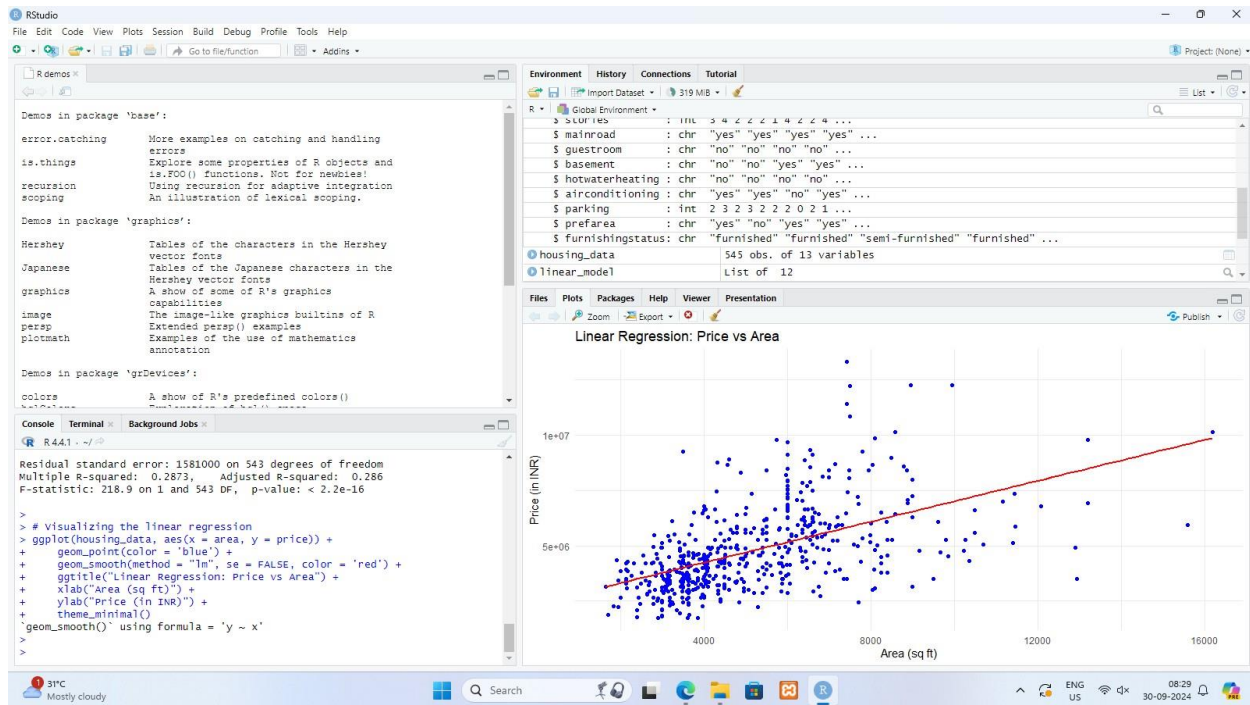
**Branch:**4th year CSE-DS

**AIM:** Create advance chart charts(linear regression,logistic regression)  
using R on dataset Housingdata set

**DATA:** The Housing dataset contains information about house pricing and various related features. It includes columns such as price, area, bedrooms, bathrooms, stories, and several binary variables like mainroad (whether the house is located on a main road), guestroom, basement, hotwaterheating, and airconditioning. Additional features include parking, prefarea (preferred area), and furnishingstatus (whether the house is furnished, semi-furnished, or unfurnished). The dataset provides a comprehensive set of attributes that influence housing prices, making it suitable for regression analysis to model relationships between

house characteristics and price, as well as classification tasks like determining whether a house is located on a main road based on its attributes.

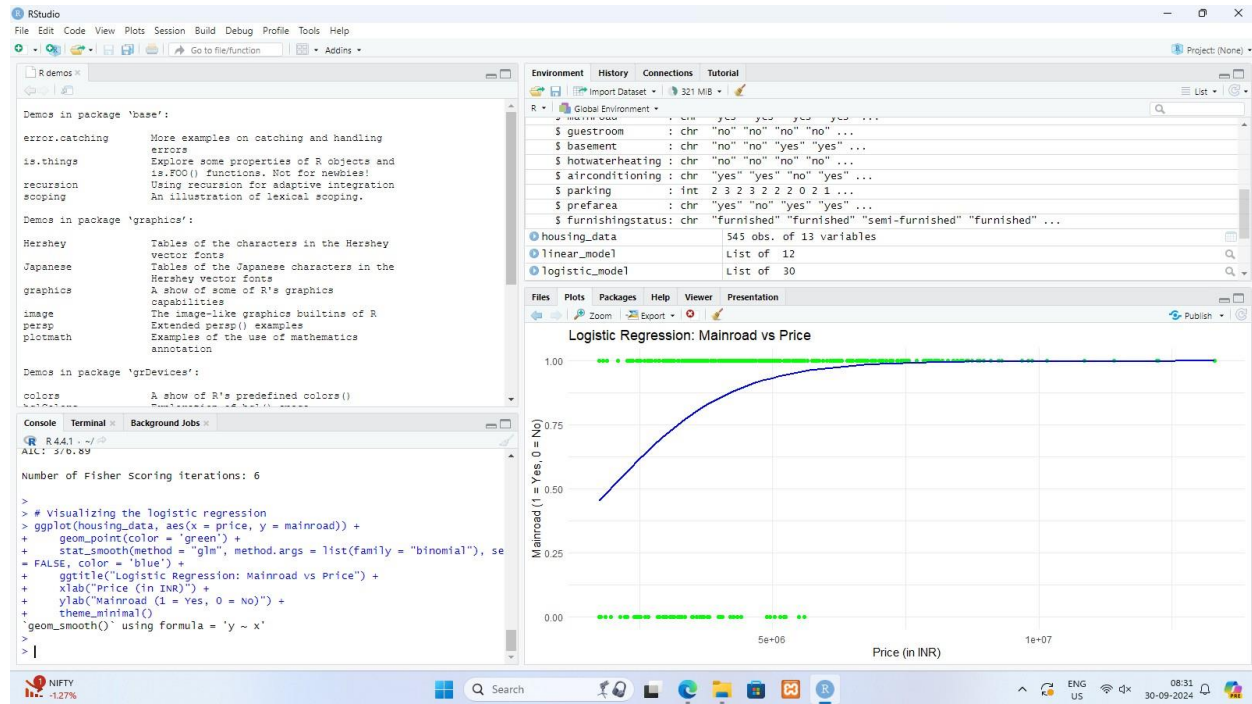
## LINEAR REGRESSION (PRICE VS AREA):



## Observations for Linear Regression: Price vs Area

1. **Positive Correlation:** There is a positive relationship between **price** and **area**, as indicated by the upward trend of the red regression line.
2. **Data Dispersion:** The points are spread widely around the regression line, suggesting a moderate correlation with some variability in the relationship between price and area.
3. **Clustered Data:** A majority of data points are clustered in the lower range of both **price** and **area**, with fewer data points at higher prices and larger areas.
4. **Outliers:** There are a few houses with significantly larger areas and higher prices, but they remain relatively spread out from the main cluster, indicating potential outliers.

## LOGISTIC REGRESSION (MAIN ROAD VS PRICE):



### Observations for Logistic Regression: Mainroad vs Price

1. **Probability Curve:** The logistic regression curve (blue) shows an increasing trend in the probability of being on a main road (`mainroad = 1`) as the `price` increases.
2. **Mainroad Likelihood:** Houses with lower prices have a higher probability of not being on the main road (`mainroad = 0`), as seen in the concentration of green points near 0 for lower price ranges.
3. **Saturation:** As the `price` increases beyond a certain threshold, the probability of being on a main road stabilizes near 1, indicating that most high-priced houses are on the main road.
4. **Outliers:** There are a few data points with high prices where houses are still not on the main road, but they are exceptions.