Name	Gitanjali Gangurde
UID	2021300034
Batch	G
Exp	5 - Create Advanced charts using R for a housing dataset.

Aim : To create Advanced Graphs of Logistic and Linear regression for the given housing dataset.

Code :

```r
install.packages("caTools")
library(caTools)
install.packages("tidyverse")
# Load required libraries
library(tidyverse)
library(caTools)

# Load the dataset housing =
read_csv("/content/housing.csv")

# View the first few rows
head(housing)

# Create new variables based on the dataset's columns
housing$mean_bedrooms = housing$`Avg. Area Number of Bedrooms` /
housing$`Area Population`
housing$mean_rooms = housing$`Avg. Area Number of Rooms` / housing$`Area
Population`

# Prepare data for the logistic regression and linear regression analysis
# Normalize the numeric features drops =
c('Address', 'Price') # Columns to drop
housing_num = housing[, !(names(housing) %in% drops)]
scaled_housing_num = scale(housing_num)
# Recombine cleaned dataset
```

```r
cleaned_housing = cbind(scaled_housing_num, Price = housing$Price)
cleaned_housing = as.data.frame(cleaned_housing) # Ensure it is a
dataframe

# Linear Regression Plot plot(cleaned_housing$Price,
cleaned_housing$`Avg. Area Income`, xlab = "House Price",
     ylab = "Income of the Residents",
     main = "Scatter Plot of Median Resident Income Vs House Prices")

# Fit linear model
lm_model <- lm(Price ~ `Avg. Area Income`, data = cleaned_housing)

# Add the regression line to the plot
abline(lm_model, col = "blue")

# Plot using ggplot for better visualization
ggplot(cleaned_housing, aes(x = `Avg. Area Income`, y = Price)) +
  geom_point(colour = 'red') +
  geom_smooth(method = 'lm', colour = 'blue')
  + ggtitle('Income Vs House Price') +
  xlab('Income') + ylab('House Price')

# Prepare data for logistic regression
# Define the expensive category (price greater than 250,000)
cleaned_housing$expensive = ifelse(cleaned_housing$Price > 250000, 1, 0)

# Logistic regression model
logistic_model <- glm(expensive ~ `Avg. Area Income`, data =
cleaned_housing, family = "binomial")

# Plot logistic regression results
ggplot(cleaned_housing, aes(x = `Avg. Area Income`, y = expensive)) +
  geom_point(colour = 'green') +
```
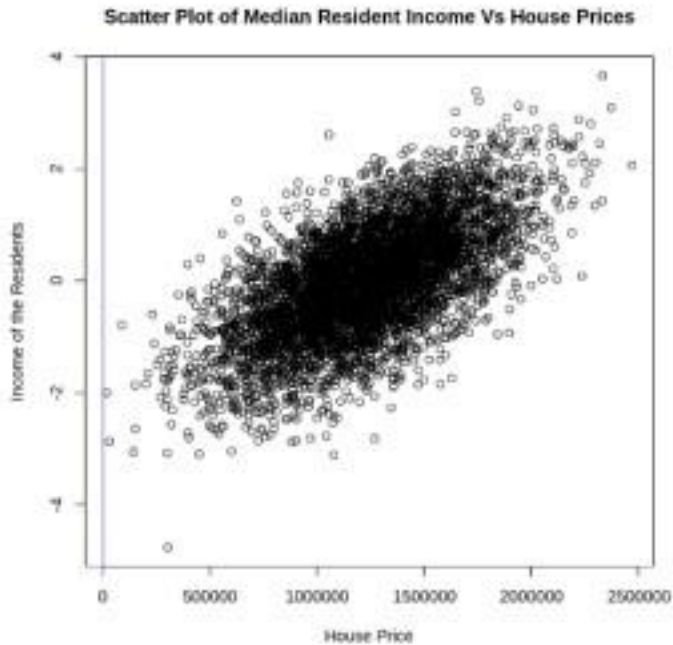
```
    geom_smooth(method = 'glm', method.args = list(family = "binomial"),
 colour = 'purple') + ggtitle('Logistic Regression: Income Vs Expensive
                          House') + xlab('Income') +
    ylab('Expensive House (1 = Yes, 0 = No)')
```
Graphs :

Scatter Plot of Median Resident Income Vs House Prices

This plot visualizes the relationship between **Median Resident Income** (on the x-axis) and **House Prices** (on the y-axis) without any fitted regression line.

- Each **black point** represents an individual house, where its position indicates both the price of the house and the income of the residents in that area.
- The distribution of the points provides a visual indication of the relationship between resident income and house prices:
    - **Clusters of points** can show where most of the houses lie in terms of income and price ranges.
    - A **positive correlation** between income and house prices may be evident if the points tend to form an upward slope, meaning that higher incomes are generally associated with more expensive houses.
    - If the points are more scattered or show no discernible pattern, this suggests a weaker or no relationship between income and house prices.

Income Vs House Price
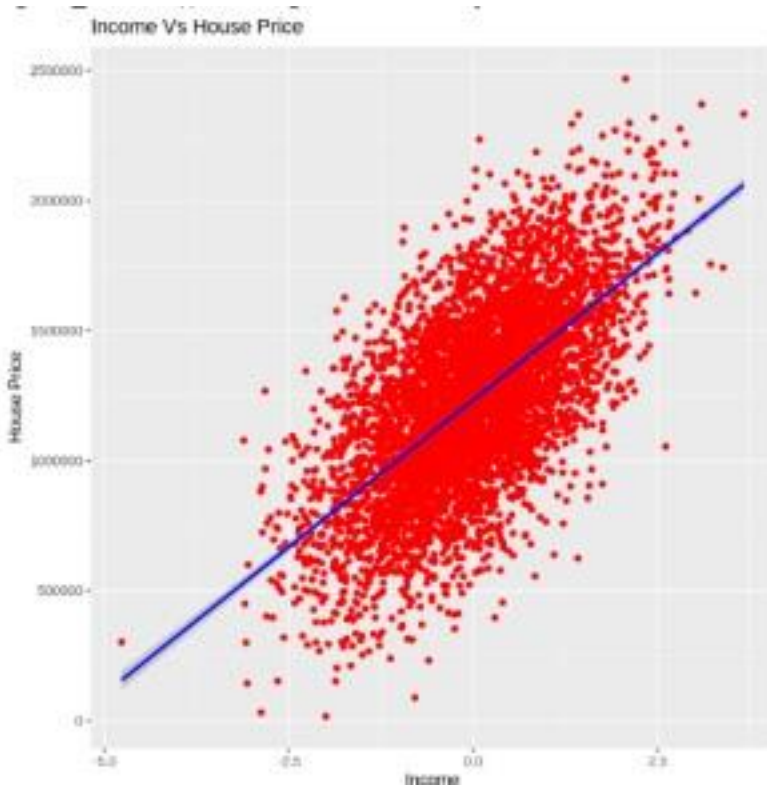
This graph plots **Median Income** on the x-axis and **House Price** on the y-axis. Each red point on the graph represents a data point for a house, showing the relationship between the income of residents in a given area and the price of the houses in that area.

- The blue line represents the **linear regression fit**, which is an attempt to model the relationship between income and house price using a straight line.
- The slope of this line suggests whether there is a positive or negative correlation between income and house price:
  - If the line slopes upwards, it indicates a **positive correlation**—as income increases, house prices also increase.
  - If the line slopes downwards, it indicates a **negative correlation**—as income increases, house prices decrease.

This graph helps visualize whether income is a strong predictor of house price.

Logistic Regression: Income Vs Expensive House
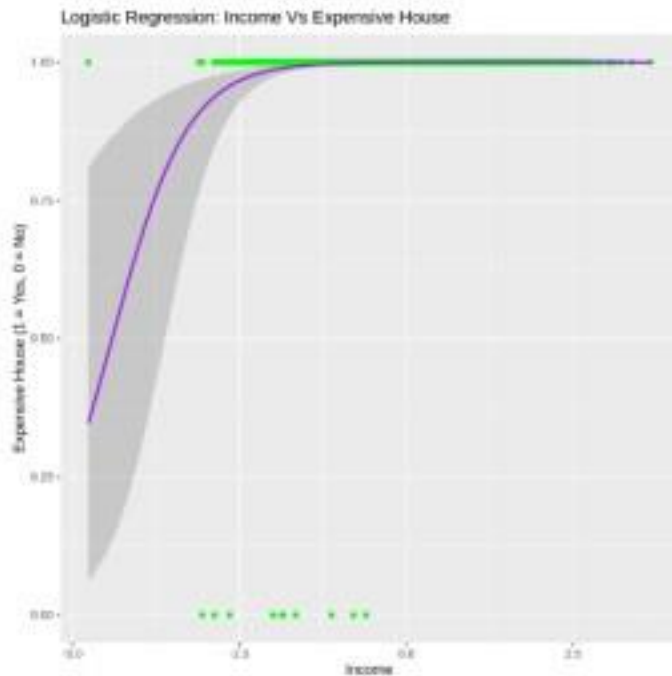
This graph uses **Median Income** on the x-axis and a binary indicator (0 or 1) on the y-axis to represent whether a house is considered **expensive** (1 if the price is above $250,000, and 0 otherwise).

- The green points represent individual houses and their classification as either expensive or not based on the income of the area.
- The purple curve represents the **logistic regression fit**, which models the probability that a house is expensive based on the income in the area.
  - The curve shows the probability of a house being classified as expensive as income increases.
  - The curve typically has an S-shape (sigmoid curve), which shows that at lower income levels, the probability of a house being expensive is low, but as income rises, the likelihood increases.

Conclusion : The analysis demonstrates a positive correlation between median resident income and house prices. The linear regression line confirms this relationship, while the logistic regression shows increasing likelihood of houses being expensive as income rises. These insights help predict housing affordability based on income levels.