BHARATIYA VIDYA BHAVAN'S
SARDAR PATEL INSTITUTE OF TECHNOLOGY
Bhavan's Campus, Munshi Nagar, Andheri (West), Mumbai – 400093-India
(Autonomous College Affiliated to University of Mumbai)

Department of Computer Science and Engineering

Course – Advanced Data Visualization (ADV)

Name	Tejal Subhash Komb
UID	2021600037
Batch	A
Lab no	5

Aim :- Create advanced charts using R programming language on the dataset - Housing data.

Objectives :-

1. To visualize the distribution and relationship between various features in the housing dataset.
2. To identify potential outliers and understand the spread of the data.
3. To explore the relationship between independent variables and the target variable (e.g., house prices).
4. To create informative visualizations that can guide decision-making in the housing market.

Dataset :- Housing Price Dataset:-

Dataset contains information about various properties, including their size, number of bedrooms and bathrooms, neighborhood type, year built, and price. Here's a brief description of each column:

- SquareFeet: The total area of the property in square feet.
- Bedrooms: The number of bedrooms in the property.
- Bathrooms: The number of bathrooms in the property.
- Neighborhood: The type of neighborhood where the property is located, categorized as Rural, Suburb, or Urban.
- YearBuilt: The year the property was built.
- Price: The price of the property in dollars.

BHARATIYA VIDYA BHAVAN'S
SARDAR PATEL INSTITUTE OF TECHNOLOGY
Bhavan's Campus, Munshi Nagar, Andheri (West), Mumbai – 400093-India
(Autonomous College Affiliated to University of Mumbai)

Department of Computer Science and Engineering

A	B	C	D	E	F
SquareFee	Bedrooms	Bathroom	Neighborh	YearBuilt	Price
2126	4	1	Rural	1969	215355.3
2459	3	2	Rural	1980	195014.2
1860	2	1	Suburb	1970	306891
2294	2	1	Urban	1996	206786.8
2130	5	2	Suburb	2001	272436.2
2095	2	3	Suburb	2020	198208.8
2724	2	1	Suburb	1993	343429.3
2044	4	3	Rural	1957	184992.3
2638	4	3	Urban	1959	377998.6
1121	5	2	Urban	2004	95961.93
1466	5	3	Suburb	1951	191113.8
2238	3	3	Suburb	1987	253358.6
1330	2	2	Suburb	1992	132172.4
2482	4	3	Suburb	1989	231157
1087	4	1	Urban	1976	118393.8
2396	2	2	Suburb	1993	267377.4
2123	5	2	Rural	1956	190773.1
1871	4	2	Suburb	1977	172989.8
2687	5	1	Urban	1979	239222.7
1130	4	3	Rural	1962	143050.2
2685	4	3	Urban	1999	405523.8
2332	3	3	Rural	1978	263954.2
1769	4	1	Suburb	1963	148310.6
1343	2	3	Rural	2013	151733.9
2515	2	3	Urban	2013	307961.1
2437	2	1	Rural	2004	276162.9
1805	5	1	Rural	2002	243985.2
1385	4	3	Rural	1996	88030.54
2215	5	1	Urban	1952	282909

BHARATIYA VIDYA BHAVAN'S
SARDAR PATEL INSTITUTE OF TECHNOLOGY
Bhavan's Campus, Munshi Nagar, Andheri (West), Mumbai – 400093-India
(Autonomous College Affiliated to University of Mumbai)

Department of Computer Science and Engineering

Implementation :- Implementation in R programming

adv5.R*

Source on Save | Run

```r
library(ggplot2)
library(dplyr)
library(scales)
library(tidyr)
library(car)
library(reshape2)
library(rgl)

house_data <- read.csv("D:/Blog/R programming/housing_price.csv")

# 1. Word Chart (For this example, we'll use the neighborhood column)
word_chart <- ggplot(house_data, aes(x = Neighborhood)) +
  geom_bar(fill = "blue") +
  ggtitle("Word Chart of Neighborhoods") +
  xlab("Neighborhood") +
  ylab("Count") +
  theme_minimal()

# Display the word chart
print(word_chart)

# 2. Box and Whisker Plot (Price by Neighborhood)
box_plot <- ggplot(house_data, aes(x = Neighborhood, y = Price)) +
  geom_boxplot(fill = "orange") +
  ggtitle("Box and Whisker Plot of Price by Neighborhood") +
  xlab("Neighborhood") +
  ylab("Price") +
  scale_y_continuous(labels = comma) +
  theme_minimal()

# Display the box plot
print(box_plot)

# 3. Violin Plot (Price by Neighborhood)
violin_plot <- ggplot(house_data, aes(x = Neighborhood, y = Price)) +
  geom_violin(fill = "lightblue") +
  ggtitle("Violin Plot of Price by Neighborhood") +
  xlab("Neighborhood") +
  ylab("Price") +
  scale_y_continuous(labels = comma) +
  theme_minimal()

# Display the violin plot
print(violin_plot)

```

BHARATIYA VIDYA BHAVAN'S
SARDAR PATEL INSTITUTE OF TECHNOLOGY
Bhavan's Campus, Munshi Nagar, Andheri (West), Mumbai – 400093-India
(Autonomous College Affiliated to University of Mumbai)

Department of Computer Science and Engineering

adv5.R*

Source on Save | Run

```r
# 4. Regression Plot (Linear and Nonlinear) - Price vs SquareFeet
# Linear
linear_plot <- ggplot(house_data, aes(x = SquareFeet, y = Price)) +
  geom_point(color = "blue") +
  geom_smooth(method = "lm", color = "red") +
  ggtitle("Linear Regression Plot: Price vs SquareFeet") +
  xlab("Square Feet") +
  ylab("Price") +
  scale_y_continuous(labels = comma) +
  theme_minimal()

# Display the linear regression plot
print(linear_plot)

# Select the top 20 rows based on Price
top_values <- house_data[order(house_data$Price, decreasing = TRUE), ][1:20, ]

# Linear Regression Plot using the filtered dataset
linear_plot <- ggplot(top_values, aes(x = SquareFeet, y = Price)) +
  geom_point(color = "blue") +
  geom_smooth(method = "lm", color = "red") +
  ggtitle("Linear Regression Plot (Top 20 Prices): Price vs SquareFeet") +
  xlab("Square Feet") +
  ylab("Price") +
  scale_y_continuous(labels = comma) +
  theme_minimal()

# Display the linear regression plot
print(linear_plot)

# Non-linear (Loess)
nonlinear_plot <- ggplot(house_data, aes(x = SquareFeet, y = Price)) +
  geom_point(color = "green") +
  geom_smooth(method = "loess", color = "purple", se = FALSE) +
  ggtitle("Non-Linear Regression Plot: Price vs SquareFeet (LOESS)") +
  xlab("Square Feet") +
  ylab("Price") +
  scale_y_continuous(labels = comma) +
  theme_minimal()

# Display the non-linear regression plot
print(nonlinear_plot)

# Check for any missing values and remove them (if needed)
house_data <- na.omit(house_data)
# Take a subset of 50 random rows from the dataset
set.seed(123)  # For reproducibility
house_data_subset <- house_data[sample(nrow(house_data), 50), ]
# Create the 3D scatter plot with the subset data
open3d()
options(scipen = 999)

scatter3d(house_data_subset$SquareFeet, house_data_subset$Price, house_data_subset$YearBuilt,
          xlab = "Square Feet", ylab = "Price", zlab = "Year Built",
          surface = FALSE, grid = TRUE, point.col = "blue")

# To interact with the plot in RStudio, you can use:
rglwidget()

# 6. Jitter Plot (Bedrooms vs Price by Neighborhood)
jitter_plot <- ggplot(house_data, aes(x = Bedrooms, y = Price, color = Neighborhood)) +
  geom_jitter() +
  ggtitle("Jitter Plot: Bedrooms vs Price by Neighborhood") +
  xlab("Bedrooms") +
  ylab("Price") +
  scale_y_continuous(labels = comma) +
  theme_minimal()
# Display the jitter plot
print(jitter_plot)

```

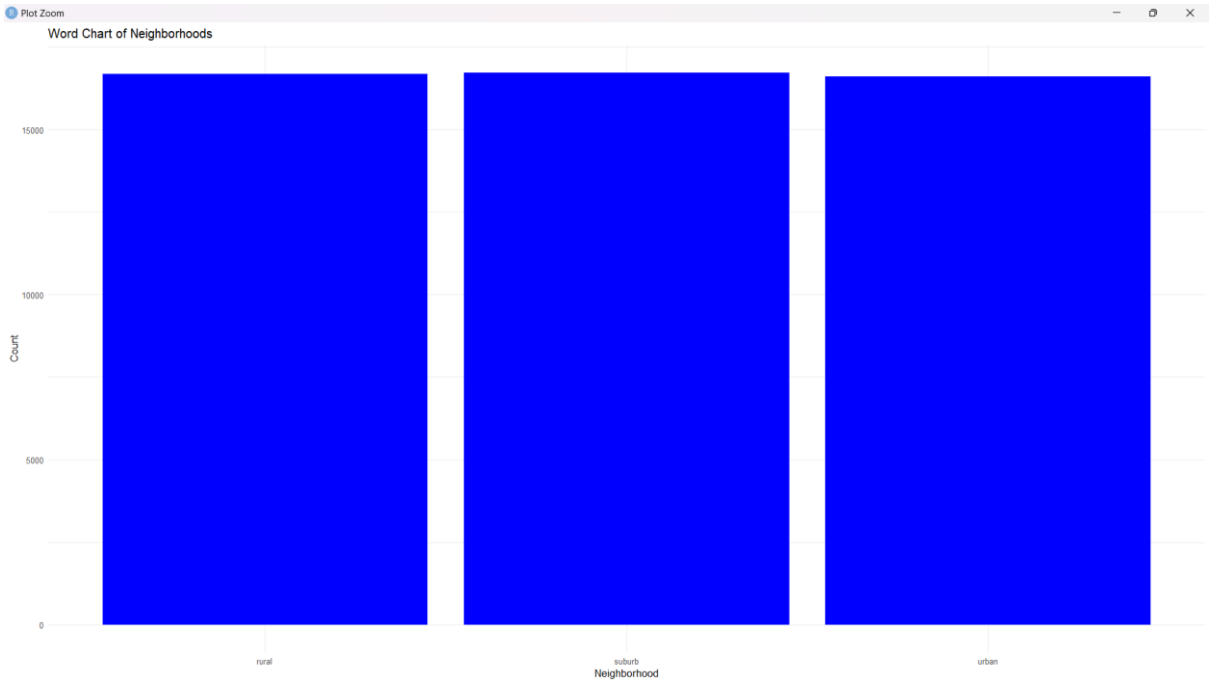
112:18 (Top Level)

BHARATIYA VIDYA BHAVAN'S
SARDAR PATEL INSTITUTE OF TECHNOLOGY
Bhavan's Campus, Munshi Nagar, Andheri (West), Mumbai – 400093-India
(Autonomous College Affiliated to University of Mumbai)

Department of Computer Science and Engineering

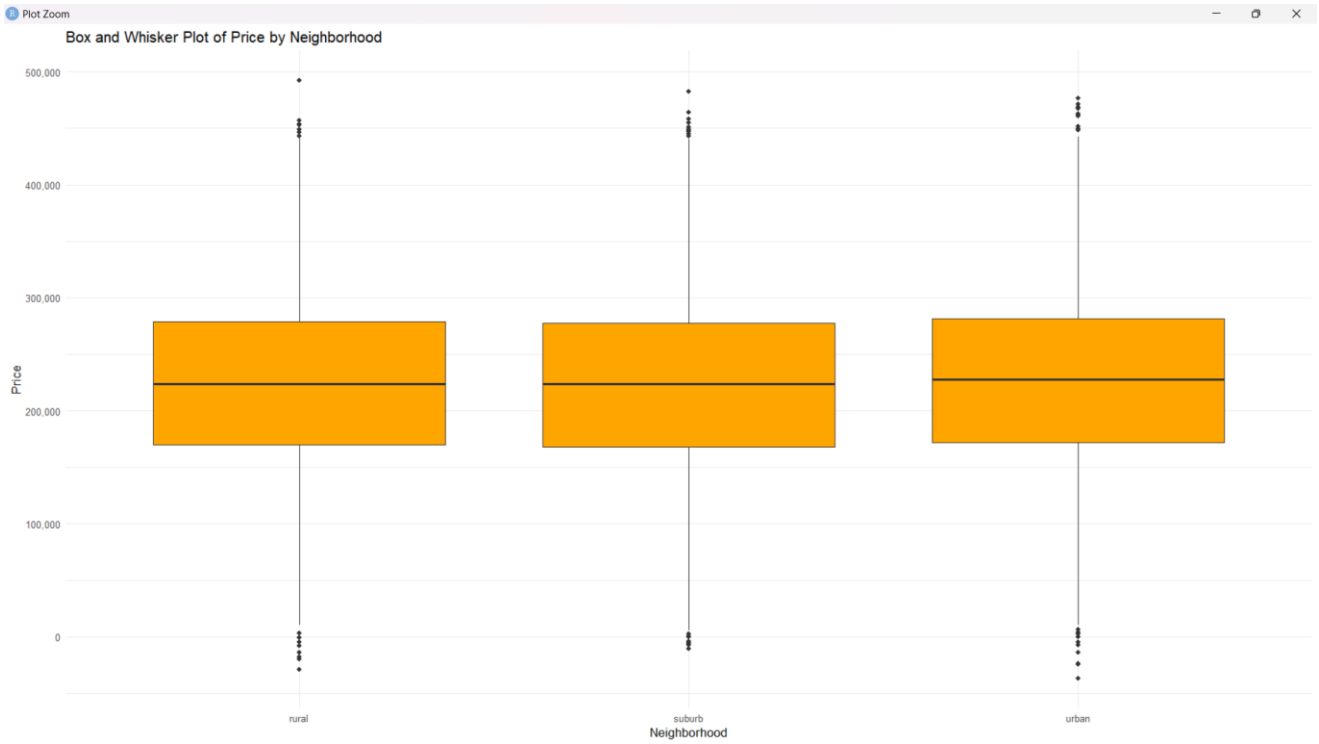
Output :-

1) **Word Chart:** Word Chart of Neighborhoods



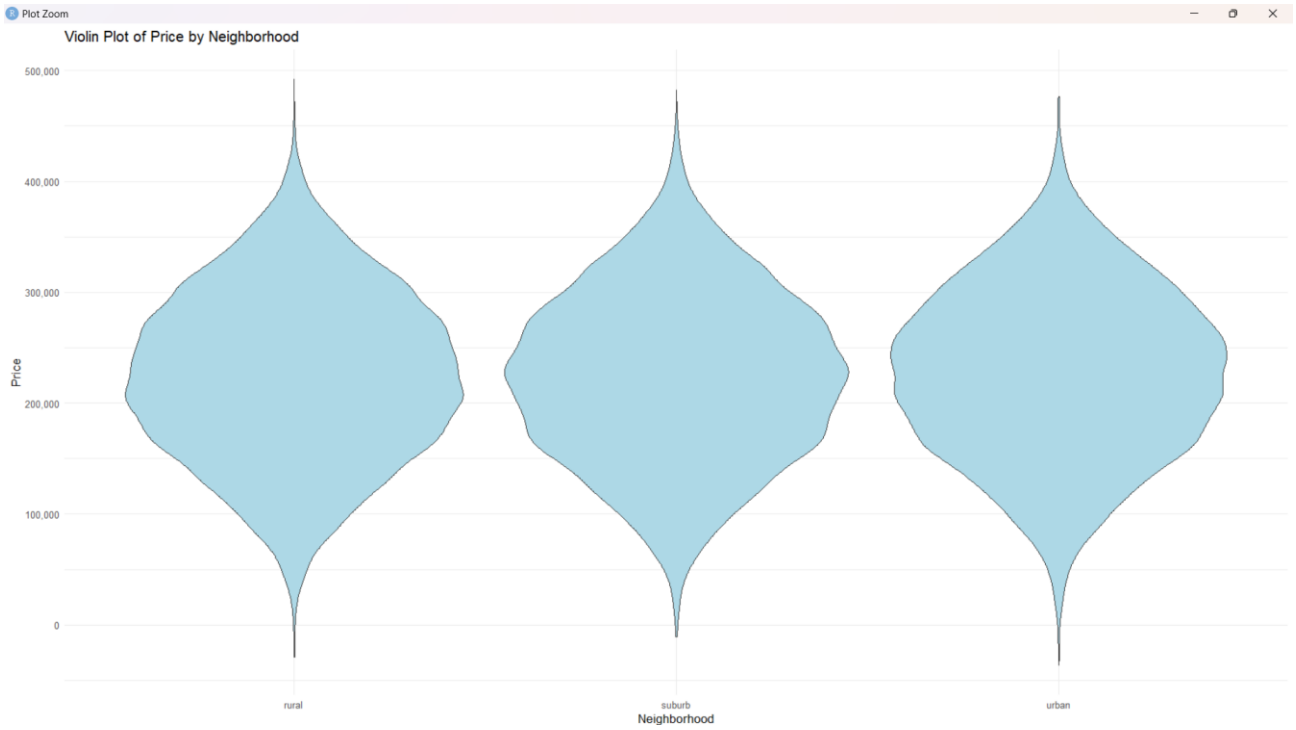
Observation: The chart shows the count of properties in each neighborhood type (Rural, Suburb, Urban). It helps in understanding the distribution of properties across different neighborhood types

2) Box and Whisker Plot: Box and Whisker Plot of Price by Neighborhood



Observation: The box plot gives insights into the price distribution across different neighborhoods. It shows the median, quartiles, and potential outliers. Suburb neighborhoods, for instance, might have a wider range of prices. Urban neighborhoods tend to have higher median prices compared to Rural and Suburb neighborhoods.

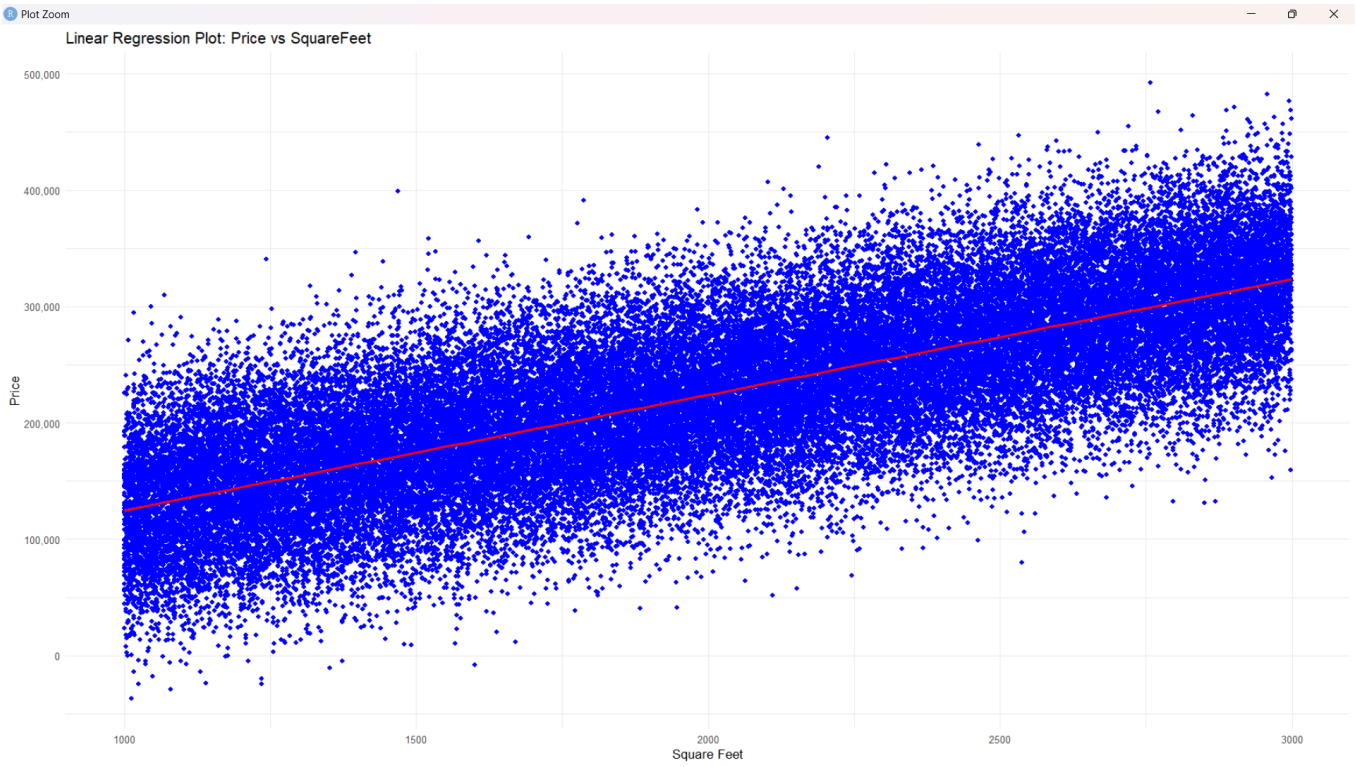
3) Violin Plot: Violin Plot of Price by Neighborhood



Observation: The violin plot provides a kernel density estimate of the price distribution in each neighborhood. It allows for a better understanding of where most prices are concentrated in each area. Urban neighborhoods show a wider spread of prices, indicating more variability.

4) Regression Plot (Linear and Non-linear):

1] Linear Regression Plot: Price vs SquareFeet

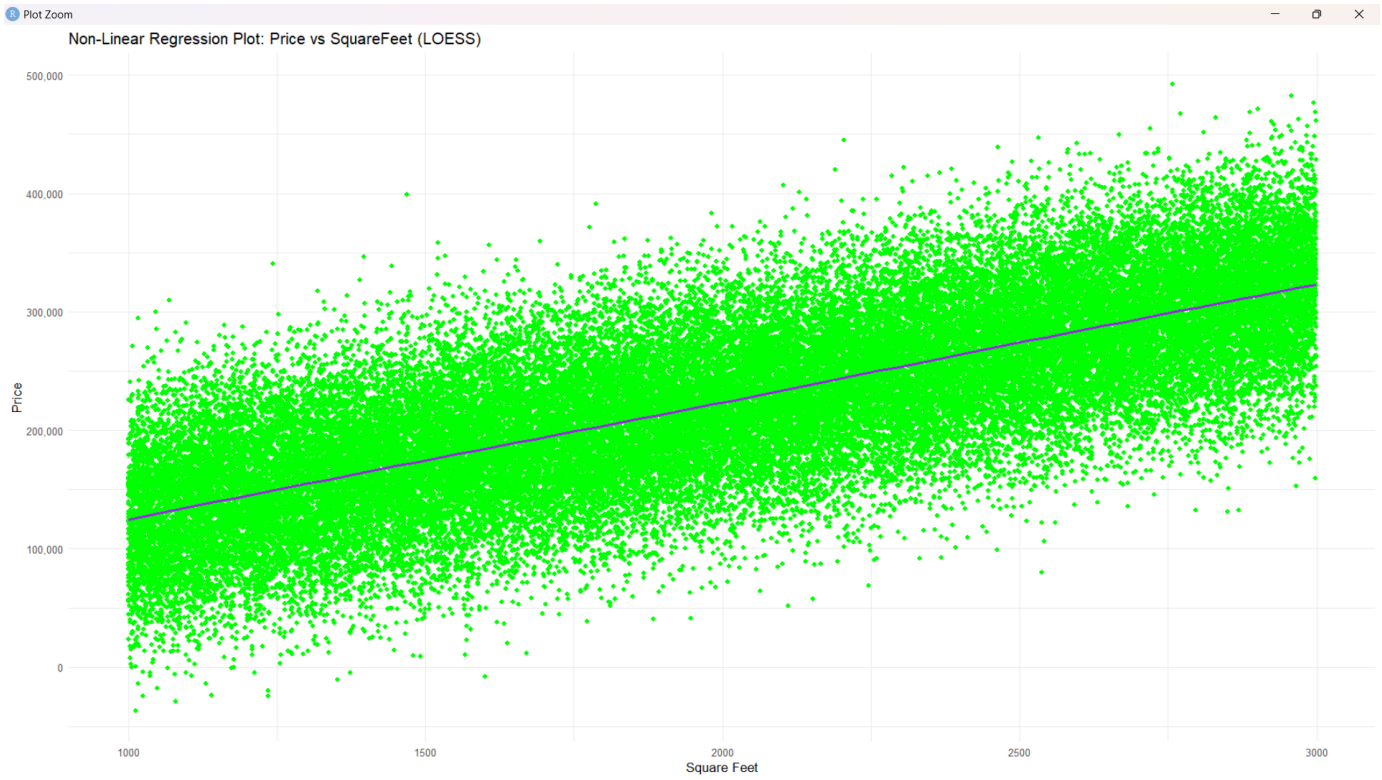


Observation (Linear): Plot shows the relationship between the size of the property (SquareFeet) and its price. The red line represents the linear regression model, indicating a positive correlation between property size and price. Larger properties generally have higher prices.

BHARATIYA VIDYA BHAVAN'S
SARDAR PATEL INSTITUTE OF TECHNOLOGY
Bhavan's Campus, Munshi Nagar, Andheri (West), Mumbai – 400093-India
(Autonomous College Affiliated to University of Mumbai)

Department of Computer Science and Engineering

2] Non-Linear Regression Plot: Price vs SquareFeet (LOESS) :-

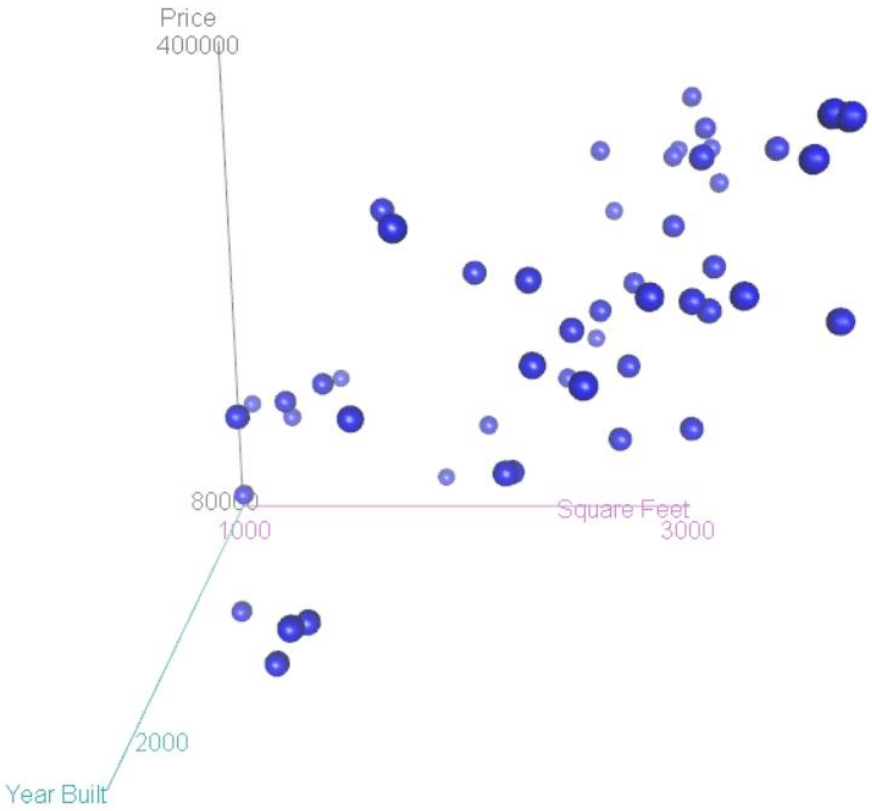


Observation (Non-linear): This plot uses a LOESS (Locally Estimated Scatterplot Smoothing) method to fit a non-linear regression line. The loess non-linear plot may capture more complex relationships, indicating deviations from a purely linear relationship, especially in the middle ranges of square footage.
It captures more complex relationships between property size and price, showing potential non-linear trends.

BHARATIYA VIDYA BHAVAN'S
SARDAR PATEL INSTITUTE OF TECHNOLOGY
Bhavan's Campus, Munshi Nagar, Andheri (West), Mumbai – 400093-India
(Autonomous College Affiliated to University of Mumbai)

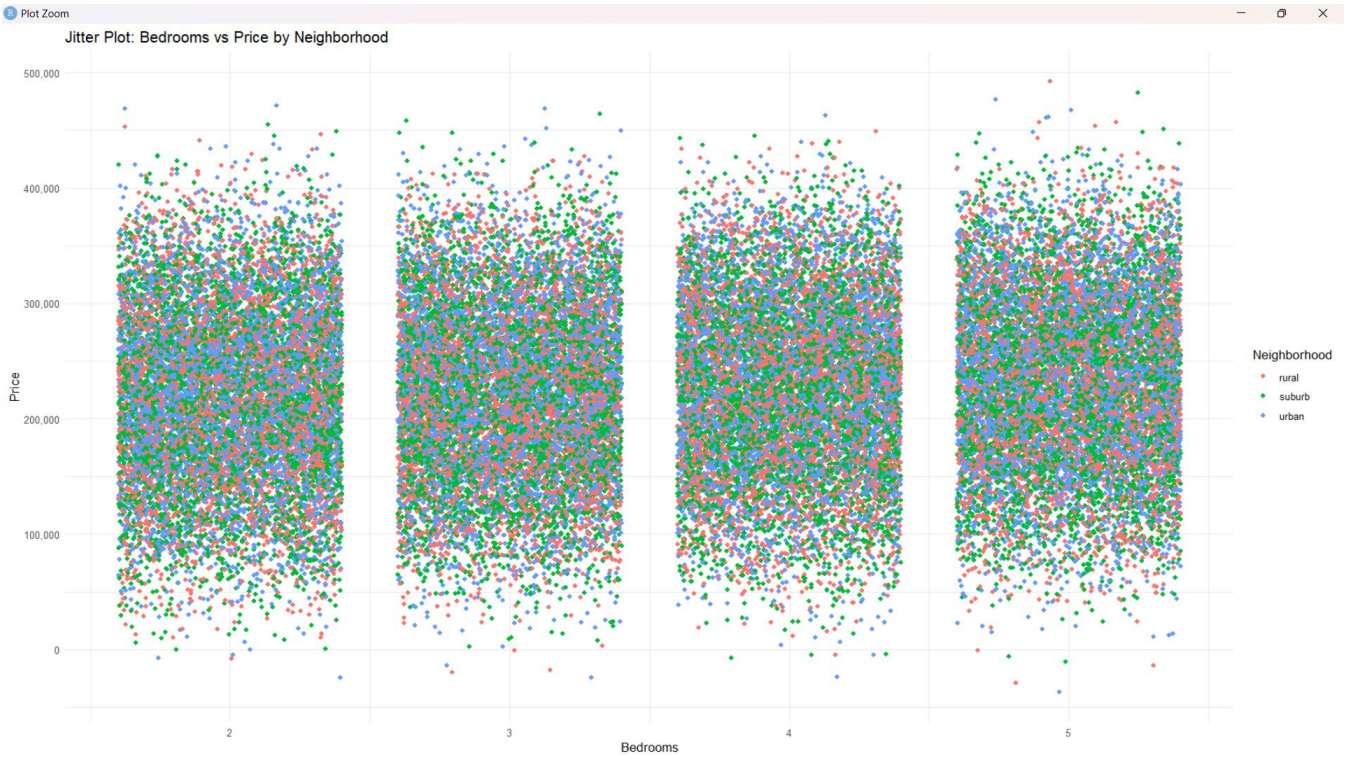
Department of Computer Science and Engineering

5) 3D Chart: SquareFeet vs Price vs YearBuilt



Observation: The 3D chart visualizes the relationship between square footage, price, and the year the property was built. It can reveal patterns like newer properties being larger and more expensive.
This interactive plot visualizes the relationship between property size, price, and the year the property was built.

6) Jitter Plot:



Observation: The jitter plot shows how prices are spread out depending on the number of bedrooms, with some clustering visible based on the neighborhood, allowing us to spot trends within specific areas.
This plot shows the relationship between the number of bedrooms and property price, with points jittered to avoid overlap. It is color-coded by neighborhood type, allowing comparison across different neighborhoods. Properties with more bedrooms generally have higher prices, but there is variability within each neighborhood type.

Conclusion :- In this experiment, we aimed to visualize housing data using R programming. The results showed that urban properties generally have higher prices and a wide range of values. Larger properties and those with more bedrooms tend to be more expensive. The relationship between property size and price is mostly positive, though not strictly linear. Newer properties might have different price trends compared to older ones, likely due to modern amenities and construction quality.