Name:	Chaitya Arun Dobariya
UID:	2021600017
Course:	Advanced Data Visualization

Experiment 5

Aim:	Create advanced charts using R programming language on the dataset - Housing data
	Advanced - Word chart, Box and whisker plot, Violin plot, Regression plot (linear and nonlinear), 3D chart, Jitter
	Write observations from each chart

1. **Dataset**

You can find the dataset [here](#).

Description

The dataset contains information on housing properties, including the price, area, number of bedrooms, number of stories, and the furnishing status. The data helps analyze the relationship between these features and how they affect housing prices.

Fields:

- **Price** (Integer):
  The selling price of the house in the local currency (e.g., ₹13,300,000). This represents the monetary value of the house based on its characteristics.
- **Area** (Integer):
  The total area of the house measured in square feet (e.g., 7420 sq. ft.). It reflects the overall size of the house, which typically impacts its value.
- **Bedrooms** (Integer):
  The number of bedrooms available in the house (e.g., 4). Houses with more bedrooms tend to have higher prices, depending on the market.

- **Bathrooms** (Integer):
  The number of bathrooms in the house (e.g., 2). Similar to bedrooms, more bathrooms generally increase the desirability and value of the house.
- **Stories** (Integer):
  The number of stories (floors) in the house (e.g., 3). Multi-story houses may offer more living space and can impact house pricing.
- **Mainroad** (String, "yes" or "no"):
  Indicates whether the house has direct access to a main road (e.g., "yes"). Houses located near main roads may be more accessible but can also face more noise and pollution.
- **Guestroom** (String, "yes" or "no"):
  Indicates whether the house has a guest room (e.g., "no"). A house with a guest room may be more appealing for larger families or for hosting guests.
- **Basement** (String, "yes" or "no"):
  Indicates whether the house has a basement (e.g., "no"). Basements can add to the usable area of the house and may be a desirable feature for storage or extra living space.
- **Hotwaterheating** (String, "yes" or "no"):
  Indicates whether the house has hot water heating (e.g., "no"). This feature provides additional comfort, especially in colder climates.
- **Airconditioning** (String, "yes" or "no"):
  Indicates whether the house is equipped with air conditioning (e.g., "yes"). Air conditioning is a desirable feature in warmer climates.
- **Parking** (Integer):
  The number of parking spaces available with the house (e.g., 2). More parking spaces may increase the value, especially in areas where parking is limited.
- **Prefarea** (String, "yes" or "no"):
  Indicates whether the house is located in a preferred area (e.g., "yes"). Houses in preferred areas typically have higher demand and value due to better neighborhood facilities or location.
- **Furnishingstatus** (String, "furnished", "semi-furnished", or "unfurnished"):
  The level of furnishing in the house (e.g., "furnished"). Furnished houses may command higher prices as they are ready to move in with minimal additional expenses for the buyer.

Importing the libraries and data:

```
1
2 ▾ # Load necessary libraries
3  install.packages("ggplot2")
4  install.packages("plotly")
5  install.packages("wordcloud")
6  install.packages("RColorBrewer")
7  install.packages("scatterplot3d")
8  install.packages("dplyr")
9
10  library(ggplot2)
11  library(plotly)
12  library(wordcloud)
13  library(RColorBrewer)
14  library(scatterplot3d)
15  library(dplyr)
1C
```

Reading the dataset:

```
16
17  getwd()
18
19  setwd("C:\\Users\\Admin\\Desktop\\SPIT\\Sem 7\\ADV\\Exp5")
20  getwd()
21
22 ▾ # Load the dataset
23  housing_data <- read.csv("housing_data.csv")
24
25 ▾ # Preview the data
26  head(housing_data)
27
28 ▾ #Word Chart
```

housing_data.csv

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	price	area	bedrooms	bathrooms	stories	mainroad	guestroom	basement	hotwaterh	airconditic	parking	prefarea	furnishingstatus	
2	13300000	7420	4	2	3	yes	no	no	no	yes	2	yes	furnished	
3	12250000	8960	4	4	4	yes	no	no	no	yes	3	no	furnished	
4	12250000	9960	3	2	2	yes	no	yes	no	no	2	yes	semi-furnished	
5	12215000	7500	4	2	2	yes	no	yes	no	yes	3	yes	furnished	
6	11410000	7420	4	1	2	yes	yes	yes	no	yes	2	no	furnished	

2. Charts & Plots

3.1 Word Chart:

Observation:

The word chart signifies that unfurnished houses are more common than semi-furnished or furnished ones. This indicates a trend where buyers might prefer unfurnished houses due to lower initial costs, giving them flexibility in furnishing the house to their taste.

3.2 Box and Whisker Chart

Boxplot of Price by Bedrooms

Observation:

The box and whisker plot provides insights into the distribution of house prices based on the number of bedrooms. Houses with 3-4 bedrooms exhibit the highest price ranges, with a few outliers present in larger properties. This suggests that while 3-4 bedroom houses are more common, larger homes (5+ bedrooms) may be priced higher due to their rarity and higher demand in certain markets.

3.3 Violin Plot

Violin Plot of Price by Stories

Observation:

The violin plot highlights that houses with more stories generally tend to have higher prices. The price range for 3-story houses is broader, and there is a higher density of expensive houses for properties with 2 or more stories, likely due to the additional living space and better utility of land area.
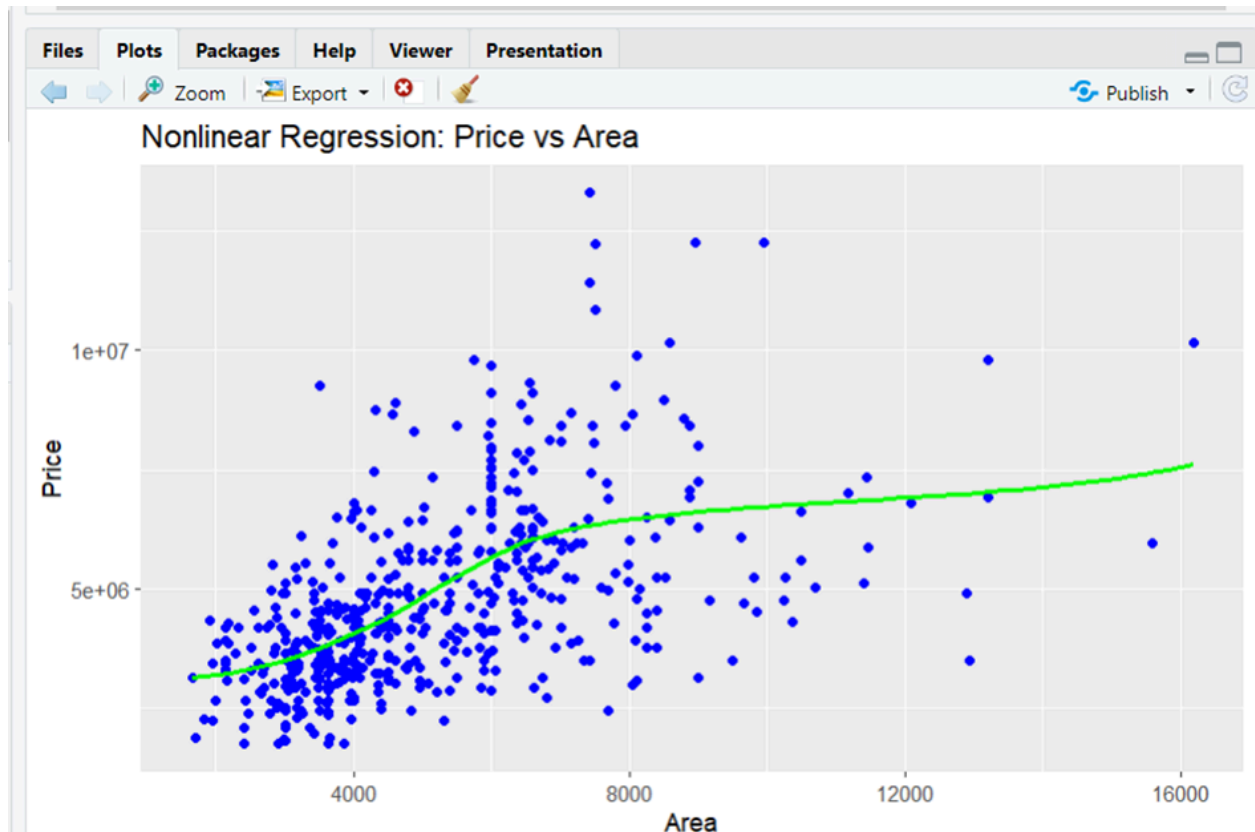
3.4 Linear Regression Plot
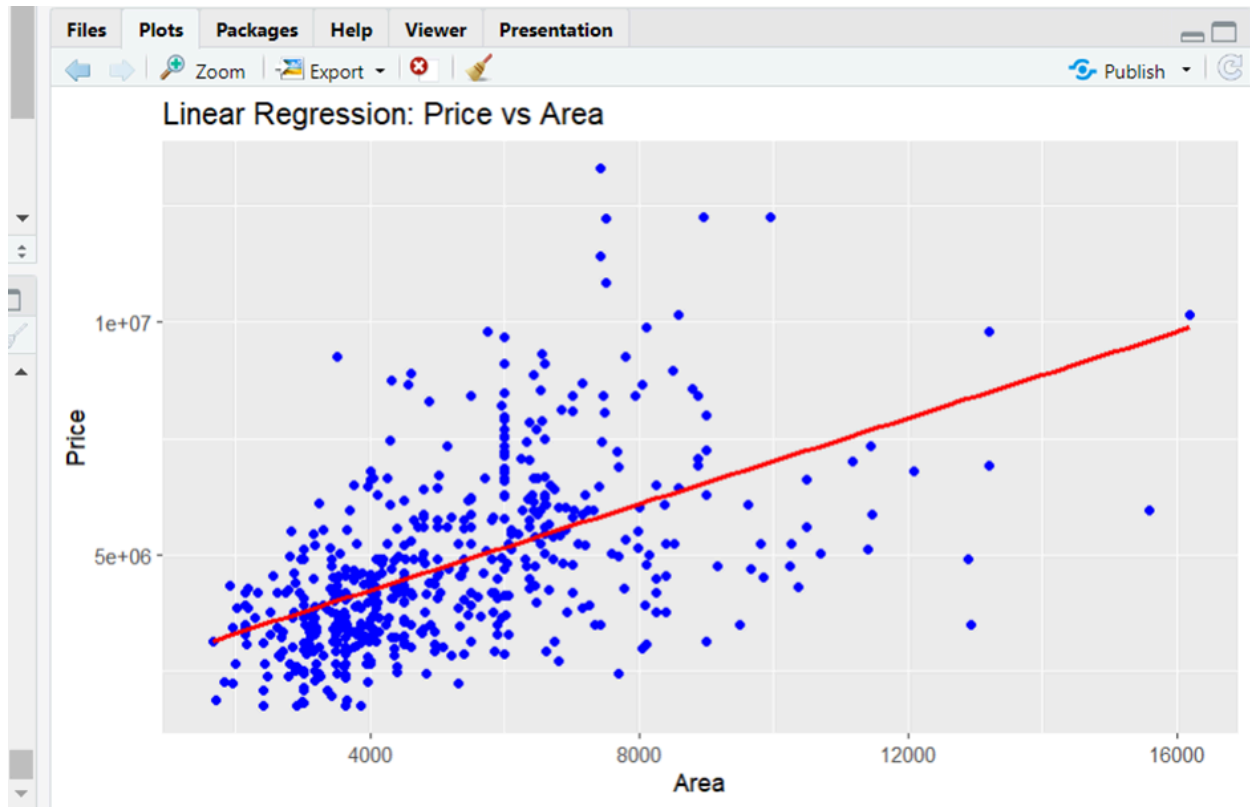
Observation:

The linear regression plot reveals a positive linear relationship between area and price, where larger homes tend to be priced higher. However, the spread of the data points suggests that factors other than just area, such as location and features, might also play a significant role in determining house prices.

3.5 Non-Linear Regression Plot

Nonlinear Regression: Price vs Area

Observation:

The non-linear regression plot captures a more complex relationship between area and price. As houses become larger, the rate of price increase becomes more pronounced, especially in high-demand areas, indicating diminishing returns for extremely large houses in specific markets. The non-linearity hints at price spikes for larger luxury homes.

3.6 3D Scatter Chart

3D Scatterplot of Price, Area, and Bedrooms
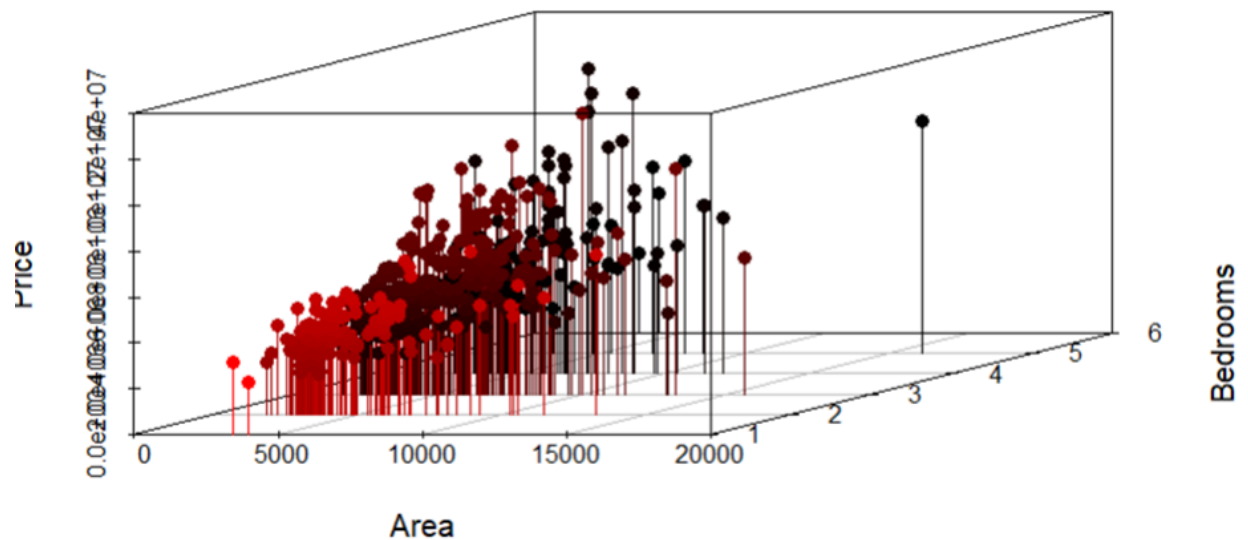
Observation:

In the 3D scatter plot, the relationship between area, price, and number of bedrooms is clearly visualized. Houses with larger areas and more bedrooms tend to be priced higher. The clustering of data points shows that houses with 2-3 bedrooms are the most common, but those with larger areas and more bedrooms have higher outliers in terms of price.

3.7 Jitter Plot

Jitter Plot: Price by Parking Spaces

Observation:

The jitter plot effectively handles overplotting and highlights how houses with direct access to the main road, guest rooms, and other such features are spread across the price range. There seems to be a slightly higher density of higher-priced houses with main road access and guest rooms, indicating these features' importance in increasing property value.

Conclusion

I worked on the analysis of the housing dataset through various advanced data visualization techniques, which provided valuable insights into the trends and distributions of key attributes such as price, area, bedrooms, bathrooms, stories, and additional factors like main road access, guestroom availability, basement, hot water heating, air conditioning, parking spaces, preferred area, and furnishing status. The analysis highlighted how different features contribute to the price of a house.
I explored various advanced plots including the Word chart (for categorical variables like main road access, guestroom, etc.), Box and whisker plot (to explore the distribution of house prices across different numbers of bedrooms and bathrooms), Violin plot (to understand the distribution and density of house prices based on area and number of stories), Regression plots (both linear and nonlinear to study relationships between house prices and area), 3D charts (to visualize the interaction between price, area, and

other factors), and Jitter plots (to handle overplotting and better visualize categorical attributes like main road and guestroom availability). This experiment provided a comprehensive understanding of how various housing features affect the price distribution across different attributes.