DSA 0603 Data Handling and Visualization for Histogram Analysis

**List of Programs in R and Tableau**

1. **Monthly Sales Data**

|  |  |
| --- | --- |
| **Month** | **Sales (in $)** |
| January | 15000 |
| February | 18000 |
| March | 22000 |
| April | 20000 |
| May | 23000 |

1. Using R Create a line chart to visualize the monthly sales. Label the axes and tit the chart appropriately.
2. Using R Generate a bar chart to display the top-selling products for the year. Label the chart elements.
3. Using R Develop a scatter plot to explore the relationship between advertising budget and monthly sales. Explain the insights drawn from the scatter plot.
4. Using Tableau Build an interactive dashboard combining the line chart and bar chart to allow users to explore sales data interactively.

Ans: # a. Line Chart

library(ggplot2)

monthly\_sales <- data.frame(

"Month" = c("January", "February", "March", "April", "May"),

"Sales" = c(15000, 18000, 22000, 20000, 23000)

)

ggplot(monthly\_sales, aes(x = Month, y = Sales)) +

geom\_line(color = "blue", size = 2) +

labs(title = "Monthly Sales",

x = "Month",

y = "Sales (in $)")

# b. Bar Chart

monthly\_sales <- data.frame(

"Month" = c("January", "February", "March", "April", "May"),

"Sales" = c(15000, 18000, 22000, 20000, 23000)

)

ordered\_sales <- monthly\_sales[order(monthly\_sales$Sales, decreasing = TRUE),]

ggplot(ordered\_sales, aes(x = Month, y = Sales, fill = Month)) +

geom\_bar(stat = "identity") +

labs(title = "Top-Selling Products",

x = "Month",

y = "Sales (in $)")

# c. Scatter Plot

advertising\_budget <- c(5000, 6000, 8000, 7000, 9000)

scatter\_data <- data.frame(

"Month" = monthly\_sales$Month,

"Sales" = monthly\_sales$Sales,

"Advertising Budget" = advertising\_budget

)

ggplot(scatter\_data, aes(x = Advertising.Budget, y = Sales)) +

geom\_point(color = "green") +

labs(title = "Relationship between Advertising Budget and Monthly Sales",

x = "Advertising Budget",

y = "Sales (in $)")

1. **Customer Feedback Analysis**

|  |  |  |
| --- | --- | --- |
| **Customer ID** | **Age** | **Satisfaction Score** |
| 1 | 25 | 4 |
| 2 | 30 | 5 |
| 3 | 35 | 3 |
| 4 | 28 | 4 |
| 5 | 40 | 5 |

1. Using R Create a histogram to represent the distribution of customer ages. Label the axes and title the chart.
2. Using R Generate a pie chart to display the overall distribution of customer satisfaction scores. Include labels.
3. Using R Build a stacked bar chart to visualize the distribution of customer satisfaction scores by age group.
4. In Tableau, develop a word cloud from open-ended customer feedback to identify prevalent customer sentiments.

Ans: #a.histogram

customer\_data <- data.frame(

"Customer ID" = c(1, 2, 3, 4, 5),

"Age" = c(25, 30, 35, 28, 40),

"Satisfaction Score" = c(4, 5, 3, 4, 5)

)

ggplot(customer\_data, aes(x = Age)) +

geom\_histogram(binwidth = 5, fill = "yellow", color = "black", alpha = 0.7) +

labs(title = "Distribution of Customer Ages",

x = "Age",

y = "Frequency")

#b.pie chart

satisfaction\_distribution <- table(customer\_data$Satisfaction.Score)

pie(satisfaction\_distribution, labels = names(satisfaction\_distribution), main = "Overall Satisfaction Distribution")

customer\_data <- data.frame(

"Customer ID" = c(1, 2, 3, 4, 5),

"Age" = c(25, 30, 35, 28, 40),

"Satisfaction Score" = c(4, 5, 3, 4, 5)

)

#c.stacked bar chart

customer\_data <- data.frame(

"Customer ID" = c(1, 2, 3, 4, 5),

"Age" = c(25, 30, 35, 28, 40),

"Satisfaction Score" = c(4, 5, 3, 4, 5)

)

ggplot(customer\_data, aes(x = Age, fill = factor(Satisfaction.Score))) +

geom\_bar() +

labs(title = "Distribution of Customer Satisfaction Scores by Age Group",

x = "Age",

y = "Count",

fill = "Satisfaction Score")

1. **Employee Performance Evaluation**

|  |  |  |  |
| --- | --- | --- | --- |
| **Employee ID** | **Department** | **Years of Service** | **Performance Score** |
| 1 | Sales | 5 | 85 |
| 2 | HR | 3 | 92 |
| 3 | Marketing | 7 | 78 |
| 4 | Sales | 4 | 90 |
| 5 | HR | 2 | 76 |

1. Using R Create a line chart to visualize the performance trend of employees over time. Include a legend and labels.
2. Using R Generate a bar chart showing the distribution of employees across different departments. Label the chart elements.
3. Using R Build a scatter plot to analyse the correlation between years of service and performance scores. Explain any insights.
4. Develop a Tableau dashboard with the line chart and bar chart for interactive exploration.

Ans: #a.line

employee\_data <- data.frame(

"Employee ID" = c(1, 2, 3, 4, 5),

"Department" = c("Sales", "HR", "Marketing", "Sales", "HR"),

"Years of Service" = c(5, 3, 7, 4, 2),

"Performance Score" = c(85, 92, 78, 90, 76)

)

ggplot(employee\_data, aes(x = Years.of.Service, y = Performance.Score, color = Department)) +

geom\_line(aes(group = Employee.ID)) +

labs(title = "Employee Performance Trend Over Time",

x = "Years of Service",

y = "Performance Score",

color = "Department") +

theme\_minimal()

#b.bar

employee\_data <- data.frame(

"Employee ID" = c(1, 2, 3, 4, 5),

"Department" = c("Sales", "HR", "Marketing", "Sales", "HR"),

"Years of Service" = c(5, 3, 7, 4, 2),

"Performance Score" = c(85, 92, 78, 90, 76)

)

ggplot(employee\_data, aes(x = Department)) +

geom\_bar(fill = "skyblue", color = "black") +

labs(title = "Distribution of Employees Across Departments",

x = "Department",

y = "Number of Employees")

#c.scatter plot

employee\_data <- data.frame(

"Employee ID" = c(1, 2, 3, 4, 5),

"Department" = c("Sales", "HR", "Marketing", "Sales", "HR"),

"Years of Service" = c(5, 3, 7, 4, 2),

"Performance Score" = c(85, 92, 78, 90, 76)

)

ggplot(employee\_data, aes(x = Years.of.Service, y = Performance.Score)) +

geom\_point() +

labs(title = "Correlation between Years of Service and Performance Scores",

x = "Years of Service",

y = "Performance Score")

1. **Product Inventory Management**

|  |  |  |
| --- | --- | --- |
| **Product ID** | **Product Name** | **Quantity Available** |
| 1 | Product A | 250 |
| 2 | Product B | 175 |
| 3 | Product C | 300 |
| 4 | Product D | 200 |
| 5 | Product E | 220 |

1. Using R Create a bar chart to visualize the quantity of each product in the inventory. Label the axes and title the chart.
2. Using R Generate a stacked bar chart to show the quantity of each product within different product categories.
3. Using R Build a scatter plot to explore the relationship between product price and quantity available. Explain the findings.
4. Develop a Tableau dashboard with the bar chart and stacked bar chart to allow users to interact with the data.

Ans: #a.bar chart

library(ggplot2)

product\_inventory <- data.frame(

"Product ID" = c(1, 2, 3, 4, 5),

"Product Name" = c("Product A", "Product B", "Product C", "Product D", "Product E"),

"Quantity Available" = c(250, 175, 300, 200, 220)

)

ggplot(product\_inventory, aes(x = reorder(Product.Name, Quantity.Available), y = Quantity.Available)) +

geom\_bar(stat = "identity", fill = "skyblue") +

labs(title = "Product Quantity in Inventory",

x = "Product Name",

y = "Quantity Available") +

theme(axis.text.x = element\_text(angle = 45, hjust = 1))

#b.stacked bar

library(ggplot2)

product\_inventory <- data.frame(

"Product ID" = c(1, 2, 3, 4, 5),

"Product Name" = c("Product A", "Product B", "Product C", "Product D", "Product E"),

"Quantity Available" = c(250, 175, 300, 200, 220)

)

ggplot(product\_inventory, aes(x = reorder(Product.Name, Quantity.Available), y = Quantity.Available, fill = Product.Name)) +

geom\_bar(stat = "identity") +

labs(title = "Product Quantity by Category",

x = "Product Name",

y = "Quantity Available") +

theme(axis.text.x = element\_text(angle = 45, hjust = 1))

#c.scatter plot

library(ggplot2)

product\_inventory <- data.frame(

"Product ID" = c(1, 2, 3, 4, 5),

"Product Name" = c("Product A", "Product B", "Product C", "Product D", "Product E"),

"Quantity Available" = c(250, 175, 300, 200, 220),

"Product Price" = c(10, 15, 8, 20, 12)

)

ggplot(product\_inventory, aes(x = Product.Price, y = Quantity.Available)) +

geom\_point() +

labs(title = "Relationship between Product Price and Quantity Available",

x = "Product Price",

y = "Quantity Available")

**5. Website Analytics**

|  |  |  |
| --- | --- | --- |
| **Date** | **Page Views** | **Click-through Rate** |
| 2023-01-01 | 1500 | 2.3% |
| 2023-01-02 | 1600 | 2.7% |
| 2023-01-03 | 1400 | 2.0% |
| 2023-01-04 | 1650 | 2.4% |
| 2023-01-05 | 1800 | 2.6% |

1. Using R Create a line chart to visualize the trend in daily page views over time. Label the axes and title the chart.
2. Using R Generate a bar chart showing the top N days with the highest click-through rates. Label the chart elements.
3. Using R Develop a stacked area chart to display the distribution of user interactions (likes, shares, comments) on a website.
4. Build a Tableau dashboard with the line chart and bar chart for interactive exploration of website traffic data.

Ans: #a.line chart

library(ggplot2)

website\_analytics <- data.frame(

"Date" = as.Date(c("2023-01-01", "2023-01-02", "2023-01-03", "2023-01-04", "2023-01-05")),

"Page Views" = c(1500, 1600, 1400, 1650, 1800),

"Click-through Rate" = c(2.3, 2.7, 2.0, 2.4, 2.6)

)

ggplot(website\_analytics, aes(x = Date, y = Page.Views)) +

geom\_line() +

labs(title = "Daily Page Views Trend",

x = "Date",

y = "Page Views")

#b.bar chart

top\_n\_days <- 2 # You can change this to the desired number of top days

top\_days <- head(web\_data[order(web\_data$Click\_through\_Rate, decreasing = TRUE), ], top\_n\_days)

barplot(top\_days$Click\_through\_Rate, names.arg = as.character(top\_days$Date),

col = "skyblue", main = "Top N Days with Highest Click-through Rates",

xlab = "Date", ylab = "Click-through Rate")

#c.stacked area

library(ggplot2)

library(tidyr)

user\_interactions\_long <- pivot\_longer(user\_interactions, cols = c("Likes", "Shares", "Comments"), names\_to = "Interaction\_Type", values\_to = "Number\_of\_Interactions")

ggplot(user\_interactions\_long, aes(x = Date, y = Number\_of\_Interactions, fill = Interaction\_Type)) +

geom\_area() +

labs(title = "User Interactions Distribution",

x = "Date",

y = "Number of Interactions",

fill = "Interaction Type")

**6. Product Sales Analysis**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Product ID** | **Product Name** | **January Sales** | **February Sales** | **March Sales** |
| 1 | Product A | 2000 | 2200 | 2400 |
| 2 | Product B | 1500 | 1800 | 1600 |
| 3 | Product C | 1200 | 1400 | 1100 |

1. Using R Create a grouped bar chart to visualize the sales of each product for the first quarter. Label the chart elements.
2. Using R Generate a stacked area chart to represent the overall sales trend for all products over the three months.
3. Using R Build a table to show the monthly sales data for each product. Label the table elements.
4. Develop a Tableau dashboard combining the grouped bar chart, stacked area chart, and the table for interactive exploration of sales data.

Ans: #a.bar chart

library(ggplot2)

library(tidyr)

product\_sales <- data.frame(

"Product ID" = c(1, 2, 3),

"Product Name" = c("Product A", "Product B", "Product C"),

"January Sales" = c(2000, 1500, 1200),

"February Sales" = c(2200, 1800, 1400),

"March Sales" = c(2400, 1600, 1100)

)

product\_sales\_long <- pivot\_longer(product\_sales, cols = starts\_with("January"), names\_to = "Month", values\_to = "Sales")

ggplot(product\_sales\_long, aes(x = factor(Product.ID), y = Sales, fill = Month)) +

geom\_bar(stat = "identity", position = "dodge") +

labs(title = "Product Sales for the First Quarter",

x = "Product",

y = "Sales",

fill = "Month")

#b.stacked area chart

library(ggplot2)

library(tidyr)

product\_sales <- data.frame(

"Product ID" = c(1, 2, 3),

"Product Name" = c("Product A", "Product B", "Product C"),

"January Sales" = c(2000, 1500, 1200),

"February Sales" = c(2200, 1800, 1400),

"March Sales" = c(2400, 1600, 1100)

)

product\_sales\_long <- pivot\_longer(product\_sales, cols = starts\_with("January"), names\_to = "Month", values\_to = "Sales")

ggplot(product\_sales\_long, aes(x = factor(Product.ID), y = Sales, fill = Month)) +

geom\_area(position = "stack") +

labs(title = "Overall Sales Trend for All Products",

x = "Product",

y = "Sales",

fill = "Month")

#c.table

library(DT)

product\_sales <- data.frame(

"Product ID" = c(1, 2, 3),

"Product Name" = c("Product A", "Product B", "Product C"),

"January Sales" = c(2000, 1500, 1200),

"February Sales" = c(2200, 1800, 1400),

"March Sales" = c(2400, 1600, 1100)

)

datatable(product\_sales, options = list(lengthMenu = c(5, 10, 15), pageLength = 5))

7. **Customer Demographics Analysis**

|  |  |  |  |
| --- | --- | --- | --- |
| **Customer ID** | **Age** | **Gender** | **Income (in $)** |
| 1 | 28 | Female | 50000 |
| 2 | 35 | Male | 60000 |
| 3 | 42 | Female | 75000 |

1. Using R Create a bar chart to visualize the distribution of customer ages. Label the axes and title the chart.
2. Using R Generate a pie chart to represent the distribution of customers by gender.
3. Using R Build a table to show the demographic information for each customer. Label the table elements.
4. Develop a Tableau dashboard combining the bar chart, pie chart, and the table for interactive exploration of customer demographics.

Ans: #a.bar chat

library(ggplot2)

customer\_demographics <- data.frame(

"Customer ID" = c(1, 2, 3),

"Age" = c(28, 35, 42),

"Gender" = c("Female", "Male", "Female"),

"Income" = c(50000, 60000, 75000)

)

ggplot(customer\_demographics, aes(x = factor(Customer.ID), y = Age, fill = Gender)) +

geom\_bar(stat = "identity", position = "dodge") +

labs(title = "Distribution of Customer Ages",

x = "Customer ID",

y = "Age",

fill = "Gender")

#b.pie chart

library(ggplot2)

ggplot(customer\_demographics, aes(x = "", fill = Gender)) +

geom\_bar(width = 1, stat = "count") +

coord\_polar(theta = "y") +

labs(title = "Distribution of Customers by Gender",

fill = "Gender") +

theme\_void()

#c.TD

library(DT)

datatable(customer\_demographics, options = list(lengthMenu = c(5, 10, 15), pageLength = 5))

**8 .** **Employee Performance Analysis**

|  |  |  |  |
| --- | --- | --- | --- |
| **Employee ID** | **Department** | **Years of Service** | **Performance Score** |
| 1 | Sales | 5 | 85 |
| 2 | HR | 3 | 92 |
| 3 | Marketing | 7 | 78 |

1. Using R Create a line chart to visualize the performance trend of employees over time. Label the axes and title the chart.
2. Using R Generate a bar chart showing the distribution of employees across different departments. Label the chart elements.
3. Using R Build a table to display the performance data for each employee. Label the table elements.
4. Develop a Tableau dashboard combining the line chart, bar chart, and the table for interactive exploration of employee performance data.

Ans: # a. Line Chart

library(ggplot2)

employee\_performance <- data.frame(

"Employee ID" = c(1, 2, 3),

"Department" = c("Sales", "HR", "Marketing"),

"Years of Service" = c(5, 3, 7),

"Performance Score" = c(85, 92, 78)

)

ggplot(employee\_performance, aes(x = Employee.ID, y = Performance.Score, group = 1)) +

geom\_line(color = "blue", size = 2) +

labs(title = "Employee Performance Trend",

x = "Employee ID",

y = "Performance Score")

# b. Bar Chart

ggplot(employee\_performance, aes(x = Department)) +

geom\_bar(stat = "count", fill = "skyblue") +

labs(title = "Distribution of Employees Across Departments",

x = "Department",

y = "Number of Employees")

# c. Table

library(DT)

datatable(employee\_performance, options = list(lengthMenu = c(5, 10, 15), pageLength = 5))

9. **Product Inventory Management**

|  |  |  |  |
| --- | --- | --- | --- |
| **Product ID** | **Product Name** | **Quantity Available** | **Price (in $)** |
| 1 | Product A | 250 | 20 |
| 2 | Product B | 175 | 15 |
| 3 | Product C | 300 | 18 |

1. Using R Create a bar chart to visualize the quantity of each product in the inventory. Label the axes and title the chart.
2. Using R Generate a stacked bar chart to show the quantity of each product within different product categories.
3. Using R Build a table to show the inventory data for each product. Label the table elements.
4. Develop a Tableau dashboard combining the bar chart, stacked bar chart, and the table for interactive exploration of inventory data.

Ans: # Product Inventory Data

product\_inventory <- data.frame(

"Product ID" = c(1, 2, 3),

"Product Name" = c("Product A", "Product B", "Product C"),

"Quantity Available" = c(250, 175, 300),

"Price" = c(20, 15, 18)

)

# a. Bar Chart

library(ggplot2)

ggplot(product\_inventory, aes(x = factor(Product.ID), y = Quantity.Available, fill = Product.Name)) +

geom\_bar(stat = "identity") +

labs(title = "Quantity of Each Product in Inventory",

x = "Product ID",

y = "Quantity Available",

fill = "Product Name")

# b. Stacked Bar Chart

ggplot(product\_inventory, aes(x = factor(Product.ID), y = Quantity.Available, fill = factor(Product.ID))) +

geom\_bar(stat = "identity") +

labs(title = "Quantity of Each Product within Different Product Categories",

x = "Product ID",

y = "Quantity Available",

fill = "Product ID")

# c. Table

library(DT)

datatable(product\_inventory, options = list(lengthMenu = c(5, 10, 15), pageLength = 5))

10. **Survey Responses Analysis**

|  |  |  |  |
| --- | --- | --- | --- |
| **Survey ID** | **Question 1** | **Question 2** | **Question 3** |
| 1 | A | B | C |
| 2 | B | A | D |
| 3 | C | A | B |

1. Using R Create a grouped bar chart to visualize the distribution of answers for Question 1. Label the chart elements.
2. Using R Generate a stacked bar chart to represent the overall distribution of responses for all three questions.
3. Using R Build a table to show the survey response data for each survey. Label the table elements.
4. Develop a Tableau dashboard combining the grouped bar chart, stacked bar chart, and the table for interactive exploration of survey responses.

Ans: # a. Grouped Bar Chart for Question 1

library(ggplot2)

survey\_responses <- data.frame(

"Survey ID" = c(1, 2, 3),

"Question\_1" = c("A", "B", "C"),

"Question\_2" = c("B", "A", "A"),

"Question\_3" = c("C", "D", "B")

)

# Create and print the grouped bar chart

p <- ggplot(survey\_responses, aes(x = factor(`Survey ID`), fill = Question\_1)) +

geom\_bar(position = "dodge") +

labs(title = "Distribution of Answers for Question 1",

x = "Survey ID",

y = "Count",

fill = "Answer")

# b. Stacked Bar Chart for Overall Distribution

library(tidyr)

library(ggplot2)

survey\_responses <- data.frame(

"Survey ID" = c(1, 2, 3),

"Question\_1" = c("A", "B", "C"),

"Question\_2" = c("B", "A", "A"),

"Question\_3" = c("C", "D", "B")

)

survey\_responses\_long <- pivot\_longer(survey\_responses, cols = starts\_with("Question"), names\_to = "Question", values\_to = "Response")

ggplot(survey\_responses\_long, aes(x = factor(`Survey ID`), fill = Response)) +

geom\_bar() +

labs(title = "Overall Distribution of Survey Responses",

x = "Survey ID",

y = "Count",

fill = "Response")

# c. Table

library(DT)

datatable(survey\_responses, options = list(lengthMenu = c(5, 10, 15), pageLength = 5))

**11.Product Category Analysis**

|  |  |
| --- | --- |
| **Category** | **Sales (in $)** |
| Electronics | 50000 |
| Clothing | 35000 |
| Appliances | 40000 |

1. Using R Create a pie chart to represent the distribution of sales across product categories. Include labels.
2. Using R Generate a funnel chart to analyze the sales conversion process for each product category. Label the stages and title the chart.
3. Using R Build a table to display the sales data for each product category. Label the table elements.
4. Develop a Tableau dashboard combining the pie chart, funnel chart, and the table for interactive exploration of product category data.

Ans: #a.Pie Chart for Sales Distribution

library(ggplot2)

sales\_data <- data.frame(

Category = c("Electronics", "Clothing", "Appliances"),

Sales = c(50000, 35000, 40000)

)

pie\_chart <- ggplot(sales\_data, aes(x = "", y = Sales, fill = Category)) +

geom\_bar(stat = "identity", width = 1, color = "white") +

coord\_polar("y") +

labs(title = "Sales Distribution Across Product Categories",

fill = "Category") +

theme\_minimal()

print(pie\_chart)

#b.Funnel Chart for Sales Conversion Process

funnel\_chart <- ggplot(sales\_data, aes(x = reorder(Category, -Sales), y = Sales)) +

geom\_bar(stat = "identity", fill = "#69b3a2") +

labs(title = "Sales Conversion Process",

x = "Category",

y = "Sales",

fill = "Category") +

theme\_minimal()

print(funnel\_chart)

#c.Table to display the sales data for each product category

library(ggplot2)

library(plotly)

library(dplyr)

product\_category\_sales <- data.frame(

Category = c("Electronics", "Clothing", "Appliances"),

Sales = c(50000, 35000, 40000)

)

sales\_table <- product\_category\_sales %>%

rename("Product Category" = Category,

"Sales (in $)" = Sales)

print(sales\_table)

**12. Website Traffic**

|  |  |  |
| --- | --- | --- |
| **Date** | **Page Views** | **Click-through Rate** |
| 2023-01-01 | 1500 | 2.3% |
| 2023-01-02 | 1600 | 2.7% |
| 2023-01-03 | 1400 | 2.0% |

1. Using R Create a line chart to visualize the trend in daily page views over time. Label the axes and title the chart.
2. Using R Generate a bar chart showing the top N days with the highest click-through rates. Label the chart elements.
3. Using R Build a table to show the website traffic data for each day. Label the table elements.
4. Develop a Tableau dashboard combining the line chart, bar chart, and the table for interactive exploration of website traffic data.

Ans: # Load necessary packages

install.packages(c("ggplot2", "dplyr", "DT"))

library(ggplot2)

library(dplyr)

library(DT)

# a. Line chart to visualize the trend in daily page views

website\_traffic <- data.frame(

Date = as.Date(c("2023-01-01", "2023-01-02", "2023-01-03")),

Page\_Views = c(1500, 1600, 1400),

Click\_through\_Rate = c(2.3, 2.7, 2.0)

)

line\_chart <- ggplot(website\_traffic, aes(x = Date, y = Page\_Views)) +

geom\_line() +

labs(title = "Daily Page Views Trend",

x = "Date",

y = "Page Views")

# b. Bar chart showing the top N days with the highest click-through rates

N <- 2

top\_days <- website\_traffic[order(website\_traffic$Click\_through\_Rate, decreasing = TRUE), ][1:N, ]

bar\_chart <- ggplot(top\_days, aes(x = Date, y = Click\_through\_Rate)) +

geom\_bar(stat = "identity", fill = "skyblue") +

labs(title = paste("Top", N, "Days with Highest Click-through Rates"),

x = "Date",

y = "Click-through Rate")

# c. Table to show the website traffic data for each day

traffic\_table <- datatable(website\_traffic, options = list(lengthMenu = c(5, 10, 15), pageLength = 5))

print(line\_chart)

print(bar\_chart)

print(traffic\_table)

**13. Geographic Data**

|  |  |  |  |
| --- | --- | --- | --- |
| **City** | **Population** | **Avg. Temperature** | **Elevation** |
| City A | 500000 | 75 | 1000 |
| City B | 700000 | 68 | 800 |
| City C | 600000 | 80 | 1200 |

1. Using R Create a map chart to visualize the distribution of cities on a geographic map. Label the map elements.
2. Using R Generate a scatter plot to explore the relationship between average temperature and population size. Explain any insights.
3. Using R Build a table to display the geographic data for each city. Label the table elements.
4. Develop a Tableau dashboard combining the map chart, scatter plot, and the table for interactive exploration of geographic data.

Ans: library(ggplot2)

library(dplyr)

library(leaflet)

library(DT)

# a. Map chart to visualize the distribution of cities

geographic\_data <- data.frame(

City = c("City A", "City B", "City C"),

Population = c(500000, 700000, 600000),

Avg\_Temperature = c(75, 68, 80),

Elevation = c(1000, 800, 1200)

)

map\_chart <- leaflet(geographic\_data) %>%

addTiles() %>%

addMarkers(

lng = ~Elevation,

lat = ~Avg\_Temperature,

popup = ~City

) %>%

addProviderTiles("Esri.WorldImagery")

# b. Scatter plot to explore the relationship between average temperature and population size

scatter\_plot <- ggplot(geographic\_data, aes(x = Avg\_Temperature, y = Population)) +

geom\_point() +

labs(title = "Relationship between Average Temperature and Population Size",

x = "Average Temperature",

y = "Population")

# c. Table to display geographic data for each city

geographic\_table <- datatable(geographic\_data, options = list(lengthMenu = c(5, 10, 15), pageLength = 5))

print(map\_chart)

print(scatter\_plot)

print(geographic\_table)

**14. Survey Results**

|  |  |  |  |
| --- | --- | --- | --- |
| **Respondent** | **Question 1** | **Question 2** | **Question 3** |
| 1 | A | B | C |
| 2 | B | A | D |
| 3 | C | A | B |

1. Using R Create a stacked bar chart to display the distribution of answers for Question 1. Label the chart elements.
2. Using R Generate a radar chart to visualize the overall pattern of responses across all three questions.
3. Using R Build a table to show the survey response data for each respondent. Label the table elements.
4. Develop a Tableau dashboard combining the stacked bar chart, radar chart, and the table for interactive exploration of survey responses.

Ans: library(ggplot2)

library(dplyr)

library(fmsb)

library(DT)

# Create survey response data

survey\_data <- data.frame(

Respondent = c(1, 2, 3),

Question\_1 = c("A", "B", "C"),

Question\_2 = c("B", "A", "A"),

Question\_3 = c("C", "D", "B")

)

# a. Stacked bar chart to display the distribution of answers for Question 1

stacked\_bar\_chart\_q1 <- ggplot(survey\_data, aes(x = factor(Respondent), fill = Question\_1)) +

geom\_bar() +

labs(title = "Distribution of Answers for Question 1",

x = "Respondent",

y = "Count",

fill = "Answer")

# b. Radar chart to visualize the overall pattern of responses

radar\_chart\_data <- survey\_data %>%

select(-Respondent) %>%

as.data.frame() %>%

mutate\_all(factor)

# Function to create radar chart

radarchart <- function(df, title) {

df <- rbind(rep(1, ncol(df)), df)

colnames(df) <- c("Question\_1", "Question\_2", "Question\_3")

res <- fmsb::radarchart(df, title = title)

return(res)

}

# Create radar chart

radar\_chart <- radarchart(radar\_chart\_data, title = "Overall Pattern of Responses")

# c. Table to show the survey response data for each respondent

respondent\_table <- datatable(survey\_data, options = list(lengthMenu = c(5, 10, 15), pageLength = 5))

# Print the outputs

print(stacked\_bar\_chart\_q1)

print(radar\_chart)

print(respondent\_table)

**15. Customer Satisfaction**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Employee ID** | **Name** | | **Department** | **Years of Service** | | **Performance Score** |
| 1 | John Smith | | Sales | 5 | | 85 |
| 2 | Jane Doe | | HR | 3 | | 92 |
| 3 | Robert Brown | | Marketing | 7 | | 78 |
| 4 | Sarah White | | Sales | 4 | | 90 |
| 5 | Michael Lee | | HR | 2 | | 76 |
| ... | | ... | | | ... | |

1. In R, create a histogram to visualize the distribution of customer ages. Label the axes and title the chart.
2. In R, generate a pie chart to represent the distribution of overall customer satisfaction scores. Include labels.
3. In Tableau, build a stacked bar chart to visualize the distribution of customer satisfaction scores by age group.
4. In R, perform sentiment analysis on open-ended customer feedback and create a word cloud visualization.

Ans: #a.histogram

library(ggplot2)

# Create a data frame with the given customer ages

customer\_ages <- data.frame(

Age = c(28, 35, 42, 25, 31, 29, 38, 45, 27, 33)

)

ggplot(customer\_ages, aes(x = Age)) +

geom\_histogram(binwidth = 5, color = "black", fill = "lightblue") +

labs(title = "Distribution of Customer Ages", x = "Age", y = "Count")

#b.customer satisfaction scores

library(ggplot2)

customer\_satisfaction <- data.frame(

Score = c(85, 92, 78, 90, 76)

)

ggplot(customer\_satisfaction, aes(x = "", fill = Score)) +

geom\_bar(width = 1) +

coord\_polar(theta = "y") +

labs(title = "Distribution of Customer Satisfaction Scores", fill = "Score")

#c.customer feedback

library(tm)

library(SnowballC)

library(wordcloud)

customer\_feedback <- data.frame(

Feedback = c("I love this product!", "This product is terrible.", "It's okay, I guess.", "I'm not sure how I feel about this product.", "This product is amazing!")

)

corpus <- Corpus(VectorSource(customer\_feedback$Feedback))

corpus <- tm\_map(corpus, content\_transformer(tolower))

corpus <- tm\_map(corpus, removeNumbers)

corpus <- tm\_map(corpus, removePunctuation)

corpus <- tm\_map(corpus, stripWhitespace)

corpus <- tm\_map(corpus, removeWords, stopwords("english"))

corpus <- tm\_map(corpus, stemDocument)

tdm <- TermDocumentMatrix(corpus)

m <- as.matrix(tdm)

word\_freq <- sort(rowSums(m), decreasing = TRUE)

wordcloud(names(word\_freq), word\_freq, min.freq = 1, colors = brewer.pal(8, "Dark2"))

**16. Product Sales Analysis**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Product ID** | **Product Name** | **January Sales** | **February Sales** | **March Sales** |
| 1 | Product A | 2000 | 2200 | 2400 |
| 2 | Product B | 1500 | 1800 | 1600 |
| 3 | Product C | 1200 | 1400 | 1100 |

* 1. Using R Create a grouped bar chart to visualize the sales of each product for the first quarter. Label the chart elements.
  2. Using R Generate a stacked area chart to represent the overall sales trend for all products over the three months.
  3. Using R Build a table to show the monthly sales data for each product. Label the table elements.
  4. Develop a Tableau dashboard combining the grouped bar chart, stacked area chart, and the table for interactive exploration of sales data.

Ans: # a.bar chart

library(ggplot2)

# Create a data frame with the given product sales

product\_sales <- data.frame(

Product\_ID = c(1, 2, 3),

Product\_Name = c("Product A", "Product B", "Product C"),

January\_Sales = c(2000, 1500, 1200),

February\_Sales = c(2200, 1800, 1400),

March\_Sales = c(2400, 1600, 1100)

)

product\_sales\_long <- tidyr::pivot\_longer(product\_sales, cols = c(January\_Sales, February\_Sales, March\_Sales), names\_to = "Month", values\_to = "Sales")

ggplot(product\_sales\_long, aes(x = Product\_Name, y = Sales, fill = Month)) +

geom\_bar(stat = "identity", position = "dodge") +

labs(title = "Sales of Each Product for the First Quarter", x = "Product Name", y = "Sales", fill = "Month")

# b. Stacked area chart to represent the overall sales trend for all products over the three months

library(ggplot2)

library(tidyr)

product\_sales <- data.frame(

Product\_ID = c(1, 2, 3),

Product\_Name = c("Product A", "Product B", "Product C"),

January\_Sales = c(2000, 1500, 1200),

February\_Sales = c(2200, 1800, 1400),

March\_Sales = c(2400, 1600, 1100)

)

product\_sales\_long <- pivot\_longer(product\_sales, cols = c(January\_Sales, February\_Sales, March\_Sales), names\_to = "Month", values\_to = "Sales")

ggplot(product\_sales\_long, aes(x = Month, y = Sales, fill = Product\_Name)) +

geom\_area() +

labs(title = "Overall Sales Trend for All Products Over the Three Months", x = "Month", y = "Sales", fill = "Product Name")

# c. Table to show the monthly sales data for each product

product\_sales <- data.frame(

Product\_ID = c(1, 2, 3),

Product\_Name = c("Product A", "Product B", "Product C"),

January\_Sales = c(2000, 1500, 1200),

February\_Sales = c(2200, 1800, 1400),

March\_Sales = c(2400, 1600, 1100)

)

print(product\_sales)

**17. Customer Demographics Analysis**

|  |  |  |  |
| --- | --- | --- | --- |
| **Customer ID** | **Age** | **Gender** | **Income (in $)** |
| 1 | 28 | Female | 50000 |
| 2 | 35 | Male | 60000 |
| 3 | 42 | Female | 75000 |

1. Using R Create a bar chart to visualize the distribution of customer ages. Label the axes and title the chart.
2. Using R Generate a pie chart to represent the distribution of customers by gender.
3. Using R Build a table to show the demographic information for each customer. Label the table elements.
4. Develop a Tableau dashboard combining the bar chart, pie chart, and the table for interactive exploration of customer demographics.

Ans: #a.Bar chart

library(ggplot2)

# Create a data frame with the given customer ages

customer\_ages <- data.frame(

Age = c(28, 35, 42)

)

# Create a bar chart

ggplot(customer\_ages, aes(x = Age)) +

geom\_bar() +

labs(title = "Distribution of Customer Ages", x = "Age", y = "Count")

#b.pie chart

library(ggplot2)

# Create a data frame with the given customer genders

customer\_genders <- data.frame(

Gender = c("Female", "Male", "Female")

)

# Create a pie chart

ggplot(customer\_genders, aes(x = "", fill = Gender)) +

geom\_bar(width = 1) +

coord\_polar(theta = "y") +

labs(title = "Distribution of Customers by Gender", fill = "Gender")

# Create a data frame with the given customer demographics

customer\_demographics <- data.frame(

Customer\_ID = c(1, 2, 3),

Age = c(28, 35, 42),

Gender = c("Female", "Male", "Female"),

Income = c(50000, 60000, 75000)

)

# Print the table

print(customer\_demographics)

**18. Employee Performance Analysis**

|  |  |  |  |
| --- | --- | --- | --- |
| **Employee ID** | **Department** | **Years of Service** | **Performance Score** |
| 1 | Sales | 5 | 85 |
| 2 | HR | 3 | 92 |
| 3 | Marketing | 7 | 78 |

1. Using R Create a line chart to visualize the performance trend of employees over time. Label the axes and title the chart.
2. Using R Generate a bar chart showing the distribution of employees across different departments. Label the chart elements.
3. Using R Build a table to display the performance data for each employee. Label the table elements.
4. Develop a Tableau dashboard combining the line chart, bar chart, and the table for interactive exploration of employee performance data.

Ans: #a.line chart

library(ggplot2)

df <- data.frame(

Employee\_ID = c(1, 2, 3),

Department = c("Sales", "HR", "Marketing"),

Years\_of\_Service = c(5, 3, 7),

Performance\_Score = c(85, 92, 78)

)

ggplot(df, aes(x = Years\_of\_Service, y = Performance\_Score, group = Employee\_ID, color = Department)) +

geom\_line() +

labs(title = "Employee Performance Trend Over Time", x = "Years of Service", y = "Performance Score")

#b.bar chart

library(ggplot2)

df <- data.frame(

Employee\_ID = c(1, 2, 3),

Department = c("Sales", "HR", "Marketing"),

Years\_of\_Service = c(5, 3, 7),

Performance\_Score = c(85, 92, 78)

)

ggplot(df, aes(x = Department)) +

geom\_bar() +

labs(title = "Distribution of Employees Across Departments", x = "Department", y = "Count")

#c.table

library(knitr)

df <- data.frame(

Employee\_ID = c(1, 2, 3),

Department = c("Sales", "HR", "Marketing"),

Years\_of\_Service = c(5, 3, 7),

Performance\_Score = c(85, 92, 78)

)

kable(df, col.names = c("Employee ID", "Department", "Years of Service", "Performance Score"))

**19. Product Inventory Management**

|  |  |  |  |
| --- | --- | --- | --- |
| **Product ID** | **Product Name** | **Quantity Available** | **Price (in $)** |
| 1 | Product A | 250 | 20 |
| 2 | Product B | 175 | 15 |
| 3 | Product C | 300 | 18 |

1. Using R Create a bar chart to visualize the quantity of each product in the inventory. Label the axes and title the chart.
2. Using R Generate a stacked bar chart to show the quantity of each product within different product categories.
3. Using R Build a table to show the inventory data for each product. Label the table elements.
4. Develop a Tableau dashboard combining the bar chart, stacked bar chart, and the table for interactive exploration of inventory data.

Ans: #a.bar chart

library(ggplot2)

df <- data.frame(

Product\_ID = c(1, 2, 3),

Product\_Name = c("Product A", "Product B", "Product C"),

Quantity\_Available = c(250, 175, 300),

Price = c(20, 15, 18)

)

ggplot(df, aes(x = Product\_Name, y = Quantity\_Available)) +

geom\_bar(stat = "identity", fill = "blue") +

labs(title = "Quantity of Each Product in Inventory", x = "Product Name", y = "Quantity Available")

library(ggplot2)

#b.stacked bar chart

df <- data.frame(

Product\_ID = c(1, 2, 3),

Product\_Name = c("Product A", "Product B", "Product C"),

Quantity\_Available = c(250, 175, 300),

Price = c(20, 15, 18)

)

ggplot(df, aes(x = "", y = Quantity\_Available, fill = Product\_Name)) +

geom\_bar(width = 1, stat = "identity") +

coord\_polar(theta = "y") +

labs(title = "Quantity of Each Product in Inventory by Product Category", x = NULL, y = NULL) +

scale\_fill\_manual(values = c("blue", "green", "red"))

library(knitr)

#c.table

df <- data.frame(

Product\_ID = c(1, 2, 3),

Product\_Name = c("Product A", "Product B", "Product C"),

Quantity\_Available = c(250, 175, 300),

Price = c(20, 15, 18)

)

kable(df, col.names = c("Product ID", "Product Name", "Quantity Available", "Price ($)"))

**20. Survey Responses Analysis**

|  |  |  |  |
| --- | --- | --- | --- |
| **Survey ID** | **Question 1** | **Question 2** | **Question 3** |
| 1 | A | B | C |
| 2 | B | A | D |
| 3 | C | A | B |

1. Using R Create a grouped bar chart to visualize the distribution of answers for Question 1. Label the chart elements.
2. Using R Generate a stacked bar chart to represent the overall distribution of responses for all three questions.
3. Using R Build a table to show the survey response data for each survey. Label the table elements.
4. Develop a Tableau dashboard combining the grouped bar chart, stacked bar chart, and the table for interactive exploration of survey responses.

Ans: #a.grouped bar chart

library(ggplot2)

df <- data.frame(

Survey\_ID = c(1, 2, 3),

Question\_1 = c("A", "B", "C"),

Question\_2 = c("B", "A", "A"),

Question\_3 = c("C", "D", "B")

)

ggplot(df, aes(x = Question\_1, fill = Survey\_ID)) +

geom\_bar() +

labs(title = "Distribution of Answers for Question 1", x = "Question 1", y = "Count", fill = "Survey ID")

#b.stacked bar chart

library(ggplot2)

df <- data.frame(

Survey\_ID = c(1, 2, 3),

Question\_1 = c("A", "B", "C"),

Question\_2 = c("B", "A", "A"),

Question\_3 = c("C", "D", "B")

)

# Reshape the data to long format

df\_long <- reshape2::melt(df, id.vars = "Survey\_ID")

ggplot(df\_long, aes(x = variable, fill = value)) +

geom\_bar() +

facet\_wrap(~Survey\_ID) +

labs(title = "Overall Distribution of Responses for All Three Questions", x = NULL, y = "Count", fill = NULL)

#c.table

library(knitr)

df <- data.frame(

Survey\_ID = c(1, 2, 3),

Question\_1 = c("A", "B", "C"),

Question\_2 = c("B", "A", "A"),

Question\_3 = c("C", "D", "B")

)

kable(df, col.names = c("Survey ID", "Question 1", "Question 2", "Question 3"))

**21. Stock Analysis**

|  |  |  |  |
| --- | --- | --- | --- |
| **Date** | **Stock A** | **Stock B** | **Stock C** |
| 2023-01-01 | 100 | 150 | 120 |
| 2023-01-02 | 105 | 152 | 118 |
| 2023-01-03 | 110 | 148 | 122 |

1. Using R Create a line chart to visualize the stock prices of three companies (Stock A, Stock B, and Stock C) over a specific time period. Label the axes and title the chart.
2. Using R Generate a bar chart showing the daily percentage change in stock prices for Stock A. Label the chart elements.
3. Using R Build a table to display the stock price data for each company over the given period. Label the table elements.
4. Develop a Tableau dashboard combining the line chart, bar chart, and the table for interactive exploration of stock prices.

Ans: #a.line chart

library(ggplot2)

df <- data.frame(

Date = c("2023-01-01", "2023-01-02", "2023-01-03"),

Stock\_A = c(100, 105, 110),

Stock\_B = c(150, 152, 148),

Stock\_C = c(120, 118, 122)

)

df$Date <- as.Date(df$Date)

ggplot(df, aes(x = Date)) +

geom\_line(aes(y = Stock\_A, color = "Stock A")) +

geom\_line(aes(y = Stock\_B, color = "Stock B")) +

geom\_line(aes(y = Stock\_C, color = "Stock C")) +

labs(title = "Stock Prices of Three Companies Over Time", x = "Date", y = "Stock Price", color = "Company")

#b.bar chart

stock\_data <- data.frame(

"Date" = c("2023-01-01", "2023-01-02", "2023-01-03"),

"Stock\_A" = c(100, 105, 110)

)

# Calculate the daily percentage change for Stock A

stock\_data$Stock\_A\_Percentage\_Change <- c(0, diff(stock\_data$Stock\_A) / lag(stock\_data$Stock\_A) \* 100)

# Generate a bar chart using ggplot2

library(ggplot2)

ggplot(stock\_data, aes(x = Date, y = Stock\_A\_Percentage\_Change)) +

geom\_bar(stat = "identity", fill = "blue", alpha = 0.7) +

labs(title = "Daily % Change in Stock A Prices", x = "Date", y = "% Change")

#c.table

library(knitr)

df <- data.frame(

Date = c("2023-01-01", "2023-01-02", "2023-01-03"),

Stock\_A = c(100, 105, 110),

Stock\_B = c(150, 152, 148),

Stock\_C = c(120, 118, 122)

)

df$Date <- as.Date(df$Date)

kable(df, col.names = c("Date", "Stock A", "Stock B", "Stock C"))

**22. Sales Data**

|  |  |  |
| --- | --- | --- |
| **Customer ID** | **Age** | **Satisfaction Score** |
| 1 | 28 | 4 |
| 2 | 35 | 5 |
| 3 | 42 | 3 |
| 4 | 30 | 4 |
| 5 | 45 | 5 |

1. a.In R, create a histogram to visualize the distribution of customer ages. Label the axes and title the chart.
2. In R, generate a pie chart to represent the distribution of overall customer satisfaction scores. Include labels.
3. In Tableau, build a stacked bar chart to visualize the distribution of customer satisfaction scores by age group.
4. In R, perform sentiment analysis on open-ended customer feedback and create a word cloud visualization.

Ans: #a.histogram

library(ggplot2)

df <- data.frame(

Age = c(28, 35, 42, 30, 45),

stringsAsFactors = FALSE

)

ggplot(df, aes(x = Age)) +

geom\_histogram(binwidth = 5, fill = "blue", color = "white") +

labs(title = "Distribution of Customer Ages", x = "Age", y = "Count")

#b.pie chart

df <- data.frame(

Satisfaction\_Score = c(4, 5, 3),

stringsAsFactors = FALSE

)

pie(df$Satisfaction\_Score, labels = df$Satisfaction\_Score, col = rainbow(nrow(df)), main = "Distribution of Overall Customer Satisfaction Scores")

#c.open-ended customer feedback

feedback\_data <- data.frame(

"Customer ID" = c(1, 2, 3, 4, 5),

"Feedback" = c("Great service!", "Very satisfied with the product.", "Could be better.", "Excellent experience.", "Outstanding service!")

)

library(tm)

corpus <- Corpus(VectorSource(feedback\_data$Feedback))

corpus <- tm\_map(corpus, content\_transformer(tolower))

corpus <- tm\_map(corpus, removePunctuation)

corpus <- tm\_map(corpus, removeNumbers)

corpus <- tm\_map(corpus, removeWords, stopwords("english"))

corpus <- tm\_map(corpus, stripWhitespace)

dtm <- DocumentTermMatrix(corpus)

library(wordcloud)

wordcloud(words = names(sort(colSums(as.matrix(dtm)))), freq = colSums(as.matrix(dtm)), scale = c(3, 0.5), colors = brewer.pal(8, "Dark2"))

**23. Time Series Analysis**

|  |  |
| --- | --- |
| **Month** | **Sales (in $)** |
| January | 15000 |
| February | 18000 |
| March | 22000 |
| April | 20000 |
| May | 23000 |

1. In R, create a time series line chart to visualize the trend in monthly sales. Label the axes and title the chart.
2. In R, generate a scatter plot to analyse the relationship between advertising budget and monthly sales. Explain any insights.
3. In Tableau, build a dashboard combining a line chart showing sales trend and a pie chart displaying the distribution of sales across products.
4. In R, create an autocorrelation plot to identify seasonality in the time series data.

Ans: #a.Create a time series object

df <- data.frame(

Month = c("January", "February", "March", "April", "May"),

Sales = c(15000, 18000, 22000, 20000, 23000)

)

ts\_data <- ts(df$Sales, start = c(2023, 1), frequency = 12)

# Generate a time series line chart

plot(ts\_data, main = "Monthly Sales", xlab = "Month", ylab = "Sales")

#b.scatter plot

advertising\_data <- data.frame(

"Month" = c("January", "February", "March", "April", "May"),

"Advertising\_Budget" = c(5000, 6000, 7000, 5500, 8000),

"Sales" = c(15000, 18000, 22000, 20000, 23000)

)

library(ggplot2)

scatter\_plot <- ggplot(advertising\_data, aes(x = Advertising\_Budget, y = Sales)) +

geom\_point() +

labs(title = "Relationship Between Advertising Budget and Monthly Sales",

x = "Advertising Budget",

y = "Monthly Sales")

print(scatter\_plot)

#c.Create a time series object

df <- data.frame(

Month = c("January", "February", "March", "April", "May"),

Sales = c(15000, 18000, 22000, 20000, 23000)

)

ts\_data <- ts(df$Sales, start = c(2023, 1), frequency = 12)

acf(ts\_data, main = "Autocorrelation Plot for Monthly Sales")

**24. Employee Performance Analysis**

|  |  |  |  |
| --- | --- | --- | --- |
| **Employee ID** | **Department** | **Years of Service** | **Performance Score** |
| 1 | Sales | 5 | 85 |
| 2 | HR | 3 | 92 |
| 3 | Marketing | 7 | 78 |
| 4 | Sales | 4 | 90 |
| 5 | HR | 2 | 76 |

1. In R, create a bar chart to visualize the distribution of employees across different departments. Label the chart elements.
2. In R, generate a line chart to visualize the performance trend of employees over time. Label the axes and title the chart.
3. In Tableau, build a dashboard combining a bar chart showing department distribution and a scatter plot displaying the relationship between years of service and performance scores.
4. In R, develop a table showing the performance data for each employee. Label the table elements.

Ans: #a.bar chart

library(ggplot2)

df <- data.frame(

Employee\_ID = c(1, 2, 3, 4, 5),

Department = c("Sales", "HR", "Marketing", "Sales", "HR"),

Years\_of\_Service = c(5, 3, 7, 4, 2),

Performance\_Score = c(85, 92, 78, 90, 76)

)

ggplot(df, aes(x = Department)) +

geom\_bar(fill = "steelblue") +

ggtitle("Distribution of Employees Across Departments") +

xlab("Department") +

ylab("Number of Employees")

#b.line chart

library(ggplot2)

df <- data.frame(

Month = c("January", "February", "March", "April", "May"),

Performance\_Score = c(85, 92, 78, 90, 76)

)

ggplot(df, aes(x = Month, y = Performance\_Score)) +

geom\_line(color = "steelblue") +

ggtitle("Performance Trend of Employees Over Time") +

xlab("Month") +

ylab("Performance Score")

#c.table

library(knitr)

df <- data.frame(

Employee\_ID = c(1, 2, 3, 4, 5),

Department = c("Sales", "HR", "Marketing", "Sales", "HR"),

Years\_of\_Service = c(5, 3, 7, 4, 2),

Performance\_Score = c(85, 92, 78, 90, 76)

)

kable(df, caption = "Performance Data for Each Employee")

**25. Product Inventory Management**

|  |  |  |
| --- | --- | --- |
| **Product ID** | **Product Name** | **Quantity Available** |
| 1 | Product A | 250 |
| 2 | Product B | 175 |
| 3 | Product C | 300 |
| 4 | Product D | 200 |
| 5 | Product E | 220 |

1. In R, create a bar chart to visualize the quantity of each product in the inventory. Label the axes and title the chart.
2. In R, generate a stacked bar chart to show the quantity of each product within different product categories.
3. In Tableau, build a dashboard combining the bar chart and stacked bar chart for interactive exploration of inventory data.
4. In R, create a scatter plot to explore the relationship between product price and quantity available. Explain any insights.

Ans: #a.bar chart

library(ggplot2)

df <- data.frame(

Product\_ID = c(1, 2, 3, 4, 5),

Product\_Name = c("Product A", "Product B", "Product C", "Product D", "Product E"),

Quantity\_Available = c(250, 175, 300, 200, 220)

)

ggplot(df, aes(x = Product\_Name, y = Quantity\_Available)) +

geom\_bar(stat = "identity", fill = "steelblue") +

ggtitle("Quantity of Each Product in the Inventory") +

xlab("Product Name") +

ylab("Quantity Available")

#b.stacked bar chart

library(ggplot2)

df <- data.frame(

Product\_ID = c(1, 2, 3, 4, 5),

Product\_Name = c("Product A", "Product B", "Product C", "Product D", "Product E"),

Quantity\_Available = c(250, 175, 300, 200, 220),

Product\_Category = c("Category 1", "Category 2", "Category 1", "Category 2", "Category 1")

)

ggplot(df, aes(x = Product\_Category, y = Quantity\_Available, fill = Product\_Name)) +

geom\_bar(stat = "identity") +

ggtitle("Quantity of Each Product Within Different Product Categories") +

xlab("Product Category") +

ylab("Quantity Available")

#c.scatter plot

product\_inventory <- data.frame(

"Product ID" = c(1, 2, 3, 4, 5),

"Product Name" = c("Product A", "Product B", "Product C", "Product D", "Product E"),

"QuantityAvailable" = c(250, 175, 300, 200, 220)

)

set.seed(123)

product\_inventory$Product\_Price <- runif(n = nrow(product\_inventory), min = 10, max = 100)

library(ggplot2)

scatter\_plot <- ggplot(product\_inventory, aes(x = Product\_Price, y = QuantityAvailable)) +

geom\_point() +

labs(title = "Relationship Between Product Price and Quantity Available",

x = "Product Price",

y = "Quantity Available")

print(scatter\_plot)

**26. Website Traffic Analysis**

|  |  |  |
| --- | --- | --- |
| **Date** | **Page Views** | **Click-through Rate** |
| 2023-01-01 | 1500 | 2.3% |
| 2023-01-02 | 1600 | 2.7% |
| 2023-01-03 | 1400 | 2.0% |
| 2023-01-04 | 1650 | 2.4% |
| 2023-01-05 | 1800 | 2.6% |

1. In Tableau, create a line chart to visualize the trend in page views over time. Label the axes and title the chart.
2. In R, generate a bar chart to show the top N days with the highest click-through rates. Label the chart elements.
3. In R, build a stacked area chart to display the distribution of user interactions (likes, shares, comments) on the website.
4. In Tableau, create a dashboard with an interactive map showing traffic sources and a bar chart displaying page views by source.

Ans: # a.Bar chart

library(ggplot2)

df <- data.frame(

Date = c("2023-01-01", "2023-01-02", "2023-01-03", "2023-01-04", "2023-01-05"),

Page\_Views = c(1500, 1600, 1400, 1650, 1800),

Click\_through\_Rate = c(0.023, 0.027, 0.02, 0.024, 0.026)

)

df <- df[order(-df$Click\_through\_Rate),]

ggplot(df[1:3,], aes(x = Date, y = Click\_through\_Rate)) +

geom\_bar(stat = "identity", fill = "steelblue") +

ggtitle("Top 3 Days with the Highest Click-through Rates") +

xlab("Date") +

ylab("Click-through Rate")

# b.Stacked area chart

website\_traffic <- data.frame(

"Date" = as.Date(c("2023-01-01", "2023-01-02", "2023-01-03", "2023-01-04", "2023-01-05")),

"Page\_Views" = c(1500, 1600, 1400, 1650, 1800),

"Click\_through\_Rate" = c(2.3, 2.7, 2.0, 2.4, 2.6)

)

user\_interactions <- data.frame(

"Date" = as.Date(c("2023-01-01", "2023-01-02", "2023-01-03", "2023-01-04", "2023-01-05")),

"Likes" = c(500, 600, 400, 550, 700),

"Shares" = c(200, 250, 180, 220, 300),

"Comments" = c(100, 120, 90, 110, 150)

)

library(ggplot2)

library(tidyr)

user\_interactions\_long <- pivot\_longer(user\_interactions, cols = c("Likes", "Shares", "Comments"), names\_to = "Interaction\_Type", values\_to = "Number\_of\_Interactions")

merged\_data <- merge(website\_traffic, user\_interactions\_long, by = "Date")

ggplot(merged\_data, aes(x = Date, y = Number\_of\_Interactions, fill = Interaction\_Type)) +

geom\_area() +

labs(title = "User Interactions Distribution",

x = "Date",

y = "Number of Interactions",

fill = "Interaction Type")

**27. Employee Sales Data**

|  |  |
| --- | --- |
| Employee | Sales (in dollars) |
| John | 5000 |
| Alice | 6200 |
| Bob | 4500 |
| Sarah | 7400 |

* 1. Using R Create a line chart to visualize the monthly sales. Label the axes and tit the chart appropriately.
  2. Using R Generate a bar chart to display the top-selling products for the year. Label the chart elements.
  3. Using R Develop a scatter plot to explore the relationship between advertising budget and monthly sales. Explain the insights drawn from the scatter plot.
  4. Using Tableau Build an interactive dashboard combining the line chart and bar chart to allow users to explore sales data interactively.

Ans: # a. Line chart for monthly sales

library(ggplot2)

df <- data.frame(

Employee = c("John", "Alice", "Bob", "Sarah"),

Sales = c(5000, 6200, 4500, 7400)

)

ggplot(df, aes(x = Employee, y = Sales)) +

geom\_line(color = "steelblue") +

ggtitle("Monthly Sales") +

xlab("Employee") +

ylab("Sales (in dollars)")

# b. Bar chart for top-selling employees

employee\_sales <- data.frame(

"Employee" = c("John", "Alice", "Bob", "Sarah"),

"Sales" = c(5000, 6200, 4500, 7400)

)

bar\_chart <- ggplot(employee\_sales, aes(x = Employee, y = Sales)) +

geom\_bar(stat = "identity", fill = "skyblue") +

labs(title = "Top-Selling Employees",

x = "Employee",

y = "Sales")

print(bar\_chart)

# c. Scatter plot to explore the relationship between advertising budget and monthly sales

employee\_data <- data.frame(

"Employee" = c("John", "Alice", "Bob", "Sarah"),

"Sales" = c(5000, 6200, 4500, 7400),

"Advertising\_Budget" = c(2000, 3000, 1500, 4000)

)

scatter\_plot <- ggplot(employee\_data, aes(x = Advertising\_Budget, y = Sales)) +

geom\_point() +

labs(title = "Relationship Between Advertising Budget and Monthly Sales",

x = "Advertising Budget",

y = "Sales")

# Display the scatter plot

print(scatter\_plot)

**28. Weather Data**

|  |  |  |
| --- | --- | --- |
| Date | Temperature (°C) | Precipitation (mm) |
| 2023-01-01 | 12 | 0.5 |
| 2023-01-02 | 9 | 2.0 |
| 2023-01-03 | 15 | 0.0 |
| 2023-01-04 | 8 | 5.0 |

1. Using R Create a map chart to visualize the distribution of temperature on a geographic map. Label the map elements.
2. Using R Generate a scatter plot to explore the relationship between average temperature and precipitation. Explain any insights.
3. Using R Build a table to display the weather data for each date. Label the table elements.
4. Develop a Tableau dashboard combining the map chart, scatter plot, and the table for interactive exploration of weather data.

Ans: #a.map chart

library(ggplot2)

df <- data.frame(

Date = c("2023-01-01", "2023-01-02", "2023-01-03", "2023-01-04"),

Temperature = c(12, 9, 15, 8),

Precipitation = c(0.5, 2.0, 0.0, 5.0)

)

ggplot(df, aes(x = 1, y = 1, fill = Temperature)) +

geom\_tile() +

scale\_fill\_gradient(low = "blue", high = "red") +

ggtitle("Distribution of Temperature on a Geographic Map") +

theme\_void()

#b.scatter plot

library(ggplot2)

avg\_temp <- mean(df$Temperature)

avg\_precip <- mean(df$Precipitation)

ggplot(df, aes(x = Temperature, y = Precipitation)) +

geom\_point(color = "steelblue") +

ggtitle("Relationship between Average Temperature and Precipitation") +

xlab("Average Temperature (°C)") +

ylab("Precipitation (mm)") +

geom\_vline(xintercept = avg\_temp, color = "red", linetype = "dashed") +

geom\_hline(yintercept = avg\_precip, color = "red", linetype = "dashed")

#c.table

library(knitr)

df <- data.frame(

Date = c("2023-01-01", "2023-01-02", "2023-01-03", "2023-01-04"),

Temperature = c(12, 9, 15, 8),

Precipitation = c(0.5, 2.0, 0.0, 5.0)

)

kable(df, caption = "Weather Data for Each Date")

**29. Product Inventory**

|  |  |  |
| --- | --- | --- |
| Product | Quantity | Price (USD) |
| Widget A | 100 | 10.00 |
| Widget B | 75 | 15.00 |
| Widget C | 120 | 8.50 |
| Widget D | 50 | 20.00 |

1. Using R Create a bar chart to visualize the quantity of each product in the inventory. Label the axes and title the chart.
2. Using R Generate a stacked bar chart to show the quantity of each product within different product categories.
3. Using R Build a table to show the inventory data for each product. Label the table elements.
4. Develop a Tableau dashboard combining the bar chart, stacked bar chart, and the table for interactive exploration of inventory data.

Ans: #a.bar chart

library(ggplot2)

df <- data.frame(

Product = c("Widget A", "Widget B", "Widget C", "Widget D"),

Quantity = c(100, 75, 120, 50),

Price = c(10.00, 15.00, 8.50, 20.00)

)

ggplot(df, aes(x = Product, y = Quantity)) +

geom\_bar(stat = "identity", fill = "steelblue") +

ggtitle("Quantity of Each Product in the Inventory") +

xlab("Product") +

ylab("Quantity")

# b.stacked bar chart

library(ggplot2)

df <- data.frame(

Product = c("Widget A", "Widget B", "Widget C", "Widget D"),

Quantity = c(100, 75, 120, 50),

Category = c("Category 1", "Category 2", "Category 1", "Category 2")

)

ggplot(df, aes(x = Category, y = Quantity, fill = Product)) +

geom\_bar(stat = "identity") +

ggtitle("Quantity of Each Product Within Different Product Categories") +

xlab("Product Category") +

ylab("Quantity")

# c.table

product\_inventory <- data.frame(

"Product" = c("Widget A", "Widget B", "Widget C", "Widget D"),

"Quantity" = c(100, 75, 120, 50),

"Price (USD)" = c(10.00, 15.00, 8.50, 20.00)

)

inventory\_table <- DT::datatable(product\_inventory, options = list(lengthMenu = c(5, 10, 15), pageLength = 5),

colnames = c("Product", "Quantity", "Price (USD)"),

caption = "Product Inventory Data")

print(inventory\_table)

**30. Student Exam Scores**

|  |  |  |  |
| --- | --- | --- | --- |
| Student | Math | Science | History |
| Student A | 85 | 92 | 78 |
| Student B | 76 | 88 | 89 |
| Student C | 92 | 79 | 87 |
| Student D | 88 | 95 | 82 |

1. Using R Create a bar chart to visualize the marks of each student.
2. Using R Generate a stacked bar chart to show different subject status.
3. Using R Build a scatter plot to explore the relationship between maths subject and history subject
4. Develop a Tableau dashboard with the pie chart to visualize the distribution of students by grade level.

Ans: # a.Bar chart

student\_scores <- data.frame(

"Student" = c("Student A", "Student B", "Student C", "Student D"),

"Math" = c(85, 76, 92, 88),

"Science" = c(92, 88, 79, 95),

"History" = c(78, 89, 87, 82)

)

barplot(as.matrix(student\_scores[,2:4]), beside = TRUE, col = c("red", "green", "blue"),

main = "Student Exam Scores", xlab = "Students", ylab = "Scores")

legend("topright", inset = 0.05, title = "Subjects", c("Math", "Science", "History"), fill = c("red", "green", "blue"))

# b.Stacked bar chart

student\_scores <- data.frame(

"Student" = c("Student A", "Student B", "Student C", "Student D"),

"Math" = c(85, 76, 92, 88),

"Science" = c(92, 88, 79, 95),

"History" = c(78, 89, 87, 82)

)

student\_scores\_long <- reshape2::melt(student\_scores, id.vars = "Student")

ggplot(student\_scores\_long, aes(x = Student, y = value, fill = variable)) +

geom\_bar(stat = "identity") +

labs(title = "Student Exam Scores", x = "Students", y = "Scores") +

scale\_fill\_manual(values = c("red", "green", "blue"), name = "Subjects")

# C.Scatter plot

student\_scores <- data.frame(

"Student" = c("Student A", "Student B", "Student C", "Student D"),

"Math" = c(85, 76, 92, 88),

"Science" = c(92, 88, 79, 95),

"History" = c(78, 89, 87, 82)

)

ggplot(student\_scores, aes(x = Math, y = History)) +

geom\_point() +

labs(title = "Math vs History", x = "Math", y = "History")

**31. Customer Feedback**

|  |  |  |
| --- | --- | --- |
| Customer | Rating | Feedback |
| Customer 1 | 4 | "Great service!" |
| Customer 2 | 3 | "Decent but could improve." |
| Customer 3 | 5 | "Excellent experience!" |
| Customer 4 | 2 | "Not satisfied at all." |

1. Using R Create a histogram to represent the distribution of customer ratings. Label the axes and title the chart.
2. Using R Generate a pie chart to display the overall distribution of customer rating. Include labels.
3. Using R Build a stacked bar chart to visualize the distribution of feedback by customer rating
4. In Tableau, develop a word cloud from open-ended customer feedback to identify prevalent customer sentiments.

Ans: #a.histogram

customer\_feedback <- data.frame(

"Customer" = c("Customer 1", "Customer 2", "Customer 3", "Customer 4"),

"Rating" = c(4, 3, 5, 2),

"Feedback" = c("Great service!", "Decent but could improve.", "Excellent experience!", "Not satisfied at all.")

)

hist(customer\_feedback$Rating, main = "Distribution of Customer Ratings", xlab = "Rating", ylab = "Frequency")

#b.pie chart

customer\_feedback <- data.frame(

"Customer" = c("Customer 1", "Customer 2", "Customer 3", "Customer 4"),

"Rating" = c(4, 3, 5, 2),

"Feedback" = c("Great service!", "Decent but could improve.", "Excellent experience!", "Not satisfied at all.")

)

pie(table(customer\_feedback$Rating), main = "Overall Distribution of Customer Ratings",

labels = c("2", "3", "4", "5"), col = rainbow(4))

#c.stacked bar chart

library(ggplot2)

data <- data.frame(

Customer = c("Customer 1", "Customer 2", "Customer 3", "Customer 4"),

Rating = c(4, 3, 5, 2),

Feedback = c("Great service!", "Decent but could improve.", "Excellent experience!", "Not satisfied at all.")

)

agg\_data <- aggregate(Feedback ~ Rating, data = data, FUN = length)

ggplot(agg\_data, aes(x = Rating, y = Feedback, fill = Feedback)) +

geom\_bar(stat = "identity") +

labs(title = "Distribution of Feedback by Customer Rating", x = "Rating", y = "Count") +

theme(legend.position = "none")

**32. Product Inventory**

|  |  |  |  |
| --- | --- | --- | --- |
| Product | Category | Quantity | Price (USD) |
| Laptop A | Electronics | 50 | 800.00 |
| Phone B | Electronics | 75 | 500.00 |
| Chair C | Furniture | 120 | 50.00 |
| Book D | Books | 200 | 10.00 |

1. Using R Create a bar chart to visualize the quantity of each product in the inventory. Label the axes and title the chart.
2. Using R Generate a stacked bar chart to show the quantity of each product within different product categories.
3. Using R Build a scatter plot to explore the relationship between product price and quantity available. Explain the findings.
4. Develop a Tableau dashboard with the bar chart and stacked bar chart to allow users to interact with the data.

Ans: #a. Bar Chart for Quantity of Each Product

product\_inventory <- data.frame(

"Product" = c("Laptop A", "Phone B", "Chair C", "Book D"),

"Quantity" = c(50, 75, 120, 200)

)

library(ggplot2)

bar\_chart <- ggplot(product\_inventory, aes(x = Product, y = Quantity)) +

geom\_bar(stat = "identity", fill = "skyblue") +

labs(title = "Quantity of Each Product in Inventory",

x = "Product",

y = "Quantity")

print(bar\_chart)

#b.Stacked Bar Chart for Quantity Within Different Categories

product\_inventory <- data.frame(

"Product" = c("Laptop A", "Phone B", "Chair C", "Book D"),

"Category" = c("Electronics", "Electronics", "Furniture", "Books"),

"Quantity" = c(50, 75, 120, 200)

)

library(ggplot2)

stacked\_bar\_chart <- ggplot(product\_inventory, aes(x = Category, y = Quantity, fill = Product)) +

geom\_bar(stat = "identity") +

labs(title = "Quantity of Each Product Within Categories",

x = "Product Category",

y = "Quantity")

print(stacked\_bar\_chart)

#c.Scatter Plot for Relationship Between Product Price and Quantity

product\_inventory <- data.frame(

"Product" = c("Laptop A", "Phone B", "Chair C", "Book D"),

"Quantity" = c(50, 75, 120, 200),

"Price" = c(800.00, 500.00, 50.00, 10.00)

)

library(ggplot2)

scatter\_plot <- ggplot(product\_inventory, aes(x = Quantity, y = Price, label = Product)) +

geom\_point() +

geom\_text(hjust = 1, vjust = 1) +

labs(title = "Relationship Between Product Price and Quantity",

x = "Quantity",

y = "Price")

print(scatter\_plot)

**33. Employee Demographics**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Employee | Age | Gender | Department | Salary (USD) |
| John | 32 | Male | HR | 60000 |
| Alice | 28 | Female | Sales | 70000 |
| Bob | 35 | Male | IT | 75000 |
| Sarah | 29 | Female | Finance | 68000 |

1. Using R Create a bar chart to visualize the distribution of customer ages. Label the axes and title the chart.
2. Using R Generate a pie chart to represent the distribution of customers by gender.
3. Using R Build a table to show the demographic information for each customer. Label the table elements.
4. Develop a Tableau dashboard combining the bar chart, pie chart, and the table for interactive exploration of customer demographics.

Ans: #a.Bar Chart for Distribution of Employee Ages

employee\_data <- data.frame(

"Employee" = c("John", "Alice", "Bob", "Sarah"),

"Age" = c(32, 28, 35, 29)

)

library(ggplot2)

bar\_chart <- ggplot(employee\_data, aes(x = Employee, y = Age)) +

geom\_bar(stat = "identity", fill = "lightblue") +

labs(title = "Distribution of Employee Ages",

x = "Employee",

y = "Age")

print(bar\_chart)

#b. Pie Chart for Distribution of Employees by Gender

employee\_data <- data.frame(

Employee = c("John", "Alice", "Bob", "Sarah"),

Age = c(32, 28, 35, 29),

Gender = c("Male", "Female", "Male", "Female"),

Department = c("HR", "Sales", "IT", "Finance"),

Salary = c(60000, 70000, 75000, 68000)

)

library(ggplot2)

ggplot(employee\_data, aes(x = "", fill = Gender)) +

geom\_bar(width = 1) +

coord\_polar(theta = "y") +

labs(title = "Distribution of Employees by Gender", fill = "Gender") +

theme\_void()

#c.Table for Demographic Information

employee\_data <- data.frame(

"Employee" = c("John", "Alice", "Bob", "Sarah"),

"Age" = c(32, 28, 35, 29),

"Gender" = c("Male", "Female", "Male", "Female"),

"Department" = c("HR", "Sales", "IT", "Finance"),

"Salary" = c(60000, 70000, 75000, 68000)

)

print(employee\_data)

**34. Customer Orders**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Order ID | Customer Name | Product | Quantity | Price |
| 2001 | John | Widget X | 10 | $5 |
| 2002 | Alice | Widget Y | 5 | $8 |
| 2003 | Bob | Widget X | 8 | $5 |
| 2004 | Sarah | Widget Z | 12 | $7 |

1. In R, create a histogram to visualize the distribution of quantity of products. Label the axes and title the chart.
2. In R, generate a pie chart to represent the distribution of overall quantity and price Include labels.
3. In Tableau, build a stacked bar chart to visualize the distribution of quantity by price.
4. In R, perform sentiment analysis on open-ended customer feedback and create a word cloud visualization.

Ans: #a.Histogram for Distribution of Quantity of Products

customer\_orders <- data.frame(

"Order ID" = c(2001, 2002, 2003, 2004),

"Customer Name" = c("John", "Alice", "Bob", "Sarah"),

"Product" = c("Widget X", "Widget Y", "Widget X", "Widget Z"),

"Quantity" = c(10, 5, 8, 12),

"Price" = c(5, 8, 5, 7)

)

library(ggplot2)

histogram <- ggplot(customer\_orders, aes(x = Quantity)) +

geom\_histogram(binwidth = 1, fill = "lightblue", color = "black") +

labs(title = "Distribution of Quantity of Products",

x = "Quantity",

y = "Frequency")

print(histogram)

#b.Pie Chart for Distribution of Overall Quantity and Price

library(ggplot2)

pie\_chart <- ggplot(customer\_orders, aes(x = "", y = Quantity, fill = as.factor(Quantity))) +

geom\_bar(width = 1, stat = "identity") +

coord\_polar(theta = "y") +

labs(title = "Distribution of Overall Quantity and Price",

fill = "Quantity") +

theme\_void()

print(pie\_chart)

#c.R Sentiment Analysis and Word Cloud Visualization

customer\_feedback <- data.frame(

"Order ID" = c(2001, 2002, 2003, 2004),

"Customer Name" = c("John", "Alice", "Bob", "Sarah"),

"Feedback" = c("Great service and fast delivery!", "Product quality is excellent.", "Satisfied with the purchase.", "Had issues with the order."),

stringsAsFactors = FALSE

)

library(tm)

library(wordcloud)

corpus <- Corpus(VectorSource(customer\_feedback$Feedback))

corpus <- tm\_map(corpus, content\_transformer(tolower))

corpus <- tm\_map(corpus, removePunctuation)

corpus <- tm\_map(corpus, removeNumbers)

corpus <- tm\_map(corpus, removeWords, stopwords("english"))

corpus <- tm\_map(corpus, stripWhitespace)

wordcloud(words = corpus, min.freq = 1, scale = c(3, 0.5), colors = brewer.pal(8, "Dark2"))

**35. Student Attendance**

|  |  |  |
| --- | --- | --- |
| Student | Class Date | Attendance |
| Student A | 2023-01-01 | Present |
| Student B | 2023-01-01 | Absent |
| Student C | 2023-01-02 | Present |
| Student D | 2023-01-02 | Present |

1. Using R Create a bar chart to visualize the attendance of each student.
2. Using R Generate a stacked bar chart to show most recent attendance status.
3. Using R Build a scatter plot to explore the relationship between class date and attendance
4. Develop a Tableau dashboard with the pie chart to visualize the distribution of students by grade level.

Ans: #a.histogram

library(ggplot2)

data <- data.frame(

Customer = c("Customer A", "Customer B", "Customer A", "Customer C"),

Product = c("Widget X", "Widget Y", "Widget Z", "Widget X"),

Quantity = c(5, 3, 2, 4),

Price = c(10.00, 15.00, 8.50, 9.00)

)

ggplot(data, aes(x = Quantity)) +

geom\_histogram(binwidth = 1, fill = "blue", color = "black") +

labs(title = "Distribution of Quantity", x = "Quantity", y = "Count")

#b.pie chart

library(ggplot2)

data <- data.frame(

Customer = c("Customer A", "Customer B", "Customer A", "Customer C"),

Product = c("Widget X", "Widget Y", "Widget Z", "Widget X"),

Quantity = c(5, 3, 2, 4),

Price = c(10.00, 15.00, 8.50, 9.00)

)

agg\_data <- aggregate(cbind(Quantity, Price) ~ Customer, data = data, FUN = sum)

ggplot(agg\_data, aes(x = "", y = Quantity, fill = Price)) +

geom\_bar(stat = "identity", width = 1) +

coord\_polar("y", start = 0) +

labs(title = "Distribution of Quantity and Price", fill = "Price", x = NULL, y = NULL) +

theme\_void() +

geom\_text(aes(label = paste0(Quantity, "\n", Price)), position = position\_stack(vjust = 0.5))

#c.open-ended customer feedback

# Assuming you have customer purchases data in a data frame named customer\_purchases

customer\_purchases <- data.frame(

"Customer" = c("Customer A", "Customer B", "Customer A", "Customer C"),

"Product" = c("Widget X", "Widget Y", "Widget Z", "Widget X"),

"Quantity" = c(5, 3, 2, 4),

"Price" = c(10.00, 15.00, 8.50, 9.00),

"Feedback" = c("Great product!", "Not satisfied", "Good value for money", "Fast delivery")

)

library(wordcloud)

library(tm)

# Preprocessing for sentiment analysis

corpus <- Corpus(VectorSource(customer\_purchases$Feedback))

corpus <- tm\_map(corpus, content\_transformer(tolower))

corpus <- tm\_map(corpus, removePunctuation)

corpus <- tm\_map(corpus, removeNumbers)

corpus <- tm\_map(corpus, removeWords, stopwords("en"))

corpus <- tm\_map(corpus, stripWhitespace)

# Create a document-term matrix

dtm <- DocumentTermMatrix(corpus)

# Convert to a data frame

feedback\_matrix <- as.data.frame(as.matrix(dtm))

# Perform sentiment analysis (assuming binary sentiment)

sentiment <- colSums(feedback\_matrix)

# Create a word cloud

wordcloud(names(sentiment), sentiment, scale = c(3, 0.5), colors = brewer.pal(8, "Dark2"), main = "Sentiment Word Cloud")

**36. Sales by Region**

|  |  |  |
| --- | --- | --- |
| Region | Month | Sales (USD) |
| North | Jan | 5000 |
| North | Feb | 6200 |
| South | Jan | 4800 |
| South | Feb | 5600 |

1. Using R Create a grouped bar chart to visualize the sales of each product for the first quarter. Label the chart elements.
2. Using R Generate a stacked area chart to represent the overall sales trend for all products over the three months.
3. Using R Build a table to show the monthly sales data for each product. Label the table elements.
4. Develop a Tableau dashboard combining the grouped bar chart, stacked area chart, and the table for interactive exploration of sales data.

ans: # a.bar chart to visualize the sales of each region for the first quarter

library(ggplot2)

sales\_data <- data.frame(

Region = c("North", "North", "South", "South"),

Month = c("Jan", "Feb", "Jan", "Feb"),

Sales = c(5000, 6200, 4800, 5600)

)

ggplot(sales\_data, aes(x = Region, y = Sales, fill = Month)) +

geom\_bar(stat = "identity", position = "dodge") +

labs(title = "Sales by Region for Q1", x = "Region", y = "Sales (USD)", fill = "Month")

# b. Stacked area chart to represent the overall sales trend for all regions over the three months

library(ggplot2)

sales\_data <- data.frame(

Region = c("North", "North", "South", "South"),

Month = c("Jan", "Feb", "Jan", "Feb"),

Sales = c(5000, 6200, 4800, 5600)

)

ggplot(sales\_data, aes(x = Month, y = Sales, fill = Region)) +

geom\_area() +

labs(title = "Overall Sales Trend for Q1", x = "Month", y = "Sales (USD)", fill = "Region")

# c. Table to show the monthly sales data for each region

library(knitr)

sales\_data <- data.frame(

Region = c("North", "North", "South", "South"),

Month = c("Jan", "Feb", "Jan", "Feb"),

Sales = c(5000, 6200, 4800, 5600)

)

kable(sales\_data, caption = "Monthly Sales Data for Q1")

**37. Customer Purchases**

|  |  |  |  |
| --- | --- | --- | --- |
| Customer | Product | Quantity | Price (USD) |
| Customer A | Widget X | 5 | 10.00 |
| Customer B | Widget Y | 3 | 15.00 |
| Customer A | Widget Z | 2 | 8.50 |
| Customer C | Widget X | 4 | 9.00 |

1. In R, create a histogram to visualize the distribution of Quantity. Label the axes and title the chart.
2. In R, generate a pie chart to represent the distribution of overall quantity and price. Include labels.
3. In Tableau, build a stacked bar chart to visualize the distribution overall quantity and price
4. In R, perform sentiment analysis on open-ended customer feedback and create a word cloud visualization.

Ans: #a.histogram

library(ggplot2)

data <- data.frame(

Customer = c("Customer A", "Customer B", "Customer A", "Customer C"),

Product = c("Widget X", "Widget Y", "Widget Z", "Widget X"),

Quantity = c(5, 3, 2, 4),

Price = c(10.00, 15.00, 8.50, 9.00)

)

ggplot(data, aes(x = Quantity)) +

geom\_histogram(binwidth = 1, fill = "blue", color = "black") +

labs(title = "Distribution of Quantity", x = "Quantity", y = "Count")

#b.pie chart

library(ggplot2)

data <- data.frame(

Customer = c("Customer A", "Customer B", "Customer A", "Customer C"),

Product = c("Widget X", "Widget Y", "Widget Z", "Widget X"),

Quantity = c(5, 3, 2, 4),

Price = c(10.00, 15.00, 8.50, 9.00)

)

agg\_data <- aggregate(cbind(Quantity, Price) ~ Customer, data = data, FUN = sum)

ggplot(agg\_data, aes(x = "", y = Quantity, fill = Price)) +

geom\_bar(stat = "identity", width = 1) +

coord\_polar("y", start = 0) +

labs(title = "Distribution of Quantity and Price", fill = "Price", x = NULL, y = NULL) +

theme\_void() +

geom\_text(aes(label = paste0(Quantity, "\n", Price)), position = position\_stack(vjust = 0.5))

#c.open-ended customer feedback

# Assuming you have customer purchases data in a data frame named customer\_purchases

customer\_purchases <- data.frame(

"Customer" = c("Customer A", "Customer B", "Customer A", "Customer C"),

"Product" = c("Widget X", "Widget Y", "Widget Z", "Widget X"),

"Quantity" = c(5, 3, 2, 4),

"Price" = c(10.00, 15.00, 8.50, 9.00),

"Feedback" = c("Great product!", "Not satisfied", "Good value for money", "Fast delivery")

)

library(wordcloud)

library(tm)

# Preprocessing for sentiment analysis

corpus <- Corpus(VectorSource(customer\_purchases$Feedback))

corpus <- tm\_map(corpus, content\_transformer(tolower))

corpus <- tm\_map(corpus, removePunctuation)

corpus <- tm\_map(corpus, removeNumbers)

corpus <- tm\_map(corpus, removeWords, stopwords("en"))

corpus <- tm\_map(corpus, stripWhitespace)

# Create a document-term matrix

dtm <- DocumentTermMatrix(corpus)

# Convert to a data frame

feedback\_matrix <- as.data.frame(as.matrix(dtm))

# Perform sentiment analysis (assuming binary sentiment)

sentiment <- colSums(feedback\_matrix)

# Create a word cloud

wordcloud(names(sentiment), sentiment, scale = c(3, 0.5), colors = brewer.pal(8, "Dark2"), main = "Sentiment Word Cloud")

**38. Student Enrollment**

|  |  |  |
| --- | --- | --- |
| Student | Grade Level | Enrollment Date |
| Student A | 9th | 2023-08-15 |
| Student B | 10th | 2023-09-02 |
| Student C | 11th | 2023-07-20 |
| Student D | 9th | 2023-08-30 |

1. Using R Create a bar chart to visualize the average grade level of the students.
2. Using R Generate a stacked bar chart to show most recent enrollment date.
3. Using R Build a scatter plot to explore the relationship between grade level and enrolment date.
4. Develop a Tableau dashboard with the pie chart to visualize the distribution of students by grade level.

ans: # Assuming you have student enrollment data in a data frame named student\_enrollment

student\_enrollment <- data.frame(

"Student" = c("Student A", "Student B", "Student C", "Student D"),

"Grade\_Level" = c("9th", "10th", "11th", "9th"),

"Enrollment\_Date" = as.Date(c("2023-08-15", "2023-09-02", "2023-07-20", "2023-08-30"))

)

library(ggplot2)

# a. Bar chart to visualize the average grade level of the students

bar\_chart <- ggplot(student\_enrollment, aes(x = Grade\_Level)) +

geom\_bar(stat = "count", fill = "blue") +

labs(title = "Average Grade Level of Students",

x = "Grade Level",

y = "Count") +

theme\_minimal()

# b. Stacked bar chart to show the most recent enrollment date

student\_enrollment$Grade\_Level <- factor(student\_enrollment$Grade\_Level, levels = unique(student\_enrollment$Grade\_Level))

stacked\_bar\_chart <- ggplot(student\_enrollment, aes(x = Grade\_Level, fill = as.factor(Enrollment\_Date))) +

geom\_bar() +

labs(title = "Most Recent Enrollment Date by Grade Level",

x = "Grade Level",

y = "Count",

fill = "Enrollment Date") +

theme\_minimal()

# c. Scatter plot to explore the relationship between grade level and enrollment date

scatter\_plot <- ggplot(student\_enrollment, aes(x = Grade\_Level, y = Enrollment\_Date, color = Grade\_Level)) +

geom\_point() +

labs(title = "Relationship Between Grade Level and Enrollment Date",

x = "Grade Level",

y = "Enrollment Date",

color = "Grade Level") +

theme\_minimal()

# Print the visualizations

print(bar\_chart)

print(stacked\_bar\_chart)

print(scatter\_plot)

**39. Product Sales by Store**

|  |  |  |  |
| --- | --- | --- | --- |
| Store | Product | Quantity Sold | Revenue (USD) |
| Store A | Widget X | 100 | 500.00 |
| Store B | Widget Y | 75 | 1125.00 |
| Store A | Widget Y | 50 | 750.00 |
| Store C | Widget X | 80 | 400.00 |

1. Using R Create a bar chart to visualize the quantity of each product in the inventory. Label the axes and title the chart.
2. Using R Generate a stacked bar chart to show the quantity of each product within different product categories.
3. Using R Build a scatter plot to explore the relationship between Revenue and quantity available. Explain the findings.
4. Develop a Tableau dashboard with the bar chart and stacked bar chart to allow users to interact with the data.

Ans: # Assuming you have product sales data in a data frame named product\_sales

product\_sales <- data.frame(

"Store" = c("Store A", "Store B", "Store A", "Store C"),

"Product" = c("Widget X", "Widget Y", "Widget Y", "Widget X"),

"Quantity\_Sold" = c(100, 75, 50, 80),

"Revenue\_USD" = c(500.00, 1125.00, 750.00, 400.00)

)

library(ggplot2)

# a. Bar chart to visualize the quantity of each product

bar\_chart <- ggplot(product\_sales, aes(x = Product, y = Quantity\_Sold, fill = Store)) +

geom\_bar(stat = "identity") +

labs(title = "Quantity of Each Product by Store",

x = "Product",

y = "Quantity Sold",

fill = "Store") +

theme\_minimal()

# b. Stacked bar chart to show the quantity of each product within different stores

stacked\_bar\_chart <- ggplot(product\_sales, aes(x = Store, y = Quantity\_Sold, fill = Product)) +

geom\_bar(stat = "identity") +

labs(title = "Quantity of Each Product Within Different Stores",

x = "Store",

y = "Quantity Sold",

fill = "Product") +

theme\_minimal()

# c. Scatter plot to explore the relationship between Revenue and Quantity

scatter\_plot <- ggplot(product\_sales, aes(x = Quantity\_Sold, y = Revenue\_USD, color = Product)) +

geom\_point() +

labs(title = "Relationship Between Revenue and Quantity Sold",

x = "Quantity Sold",

y = "Revenue (USD)",

color = "Product") +

theme\_minimal()

# Print the visualizations

print(bar\_chart)

print(stacked\_bar\_chart)

print(scatter\_plot)

**40. Employee Training Hours**

|  |  |  |
| --- | --- | --- |
| Employee | Department | Training Hours |
| John | HR | 10 |
| Alice | Sales | 8 |
| Bob | IT | 12 |
| Sarah | Finance | 6 |

1. Using R Create a histogram to represent the distribution of Training Hours. Label the axes and title the chart.
2. Using R Generate a pie chart to display the overall employee Training hours. Include labels.
3. Using R Build a stacked bar chart to visualize the distribution of training hours by Employee
4. In Tableau, develop a word cloud from open-ended customer feedback to identify prevalent customer sentiments.

Ans: # Assuming you have employee training hours data in a data frame named employee\_training

employee\_training <- data.frame(

"Employee" = c("John", "Alice", "Bob", "Sarah"),

"Department" = c("HR", "Sales", "IT", "Finance"),

"Training\_Hours" = c(10, 8, 12, 6)

)

library(ggplot2)

library(wordcloud)

library(tm)

# a. Histogram to represent the distribution of Training Hours

histogram <- ggplot(employee\_training, aes(x = Training\_Hours)) +

geom\_histogram(binwidth = 2, fill = "#0073C2FF", color = "#1F3552FF", alpha = 0.7) +

labs(title = "Distribution of Training Hours",

x = "Training Hours",

y = "Frequency")

# b. Pie chart to display the overall employee Training hours

overall\_hours\_pie <- ggplot(employee\_training, aes(x = "", y = Training\_Hours, fill = Employee)) +

geom\_bar(stat = "identity", width = 1, color = "white") +

coord\_polar(theta = "y") +

labs(title = "Overall Employee Training Hours",

fill = "Employee") +

theme\_void()

# c. Stacked bar chart to visualize the distribution of training hours by Employee

stacked\_bar\_chart <- ggplot(employee\_training, aes(x = Employee, y = Training\_Hours, fill = Department)) +

geom\_bar(stat = "identity") +

labs(title = "Distribution of Training Hours by Employee",

x = "Employee",

y = "Training Hours",

fill = "Department") +

theme\_minimal() +

theme(axis.text.x = element\_text(angle = 45, hjust = 1))

# Print the visualizations

print(histogram)

print(overall\_hours\_pie)

print(stacked\_bar\_chart)

Top of Form