Name of Student: Pushkar Sane		
Roll Number: 45	**Lab Assignment Number: 7**	
Title of Lab Assignment: Implement commands for drawing various Correlation Plots and learn the process of EDA.		
DOP: 14-10-2023	**DOS: 20-10-2023**	
CO Mapped: **CO5**	**PO Mapped:** **PO1, PO2, PO3, PO4, PO5,** **PO7, PO12, PSO1, PSO2**	**Signature:**

Practical No. 7

Aim: Implement commands for drawing various Correlation Plots and learn the process of EDA.

Description:

Exploratory Data Analysis (EDA) is a crucial step in understanding and visualizing your data to uncover insights and patterns. One of the key components of EDA is exploring the relationships between variables, often visualized through correlation plots.

EDA Process:

1. **Load Required Libraries:**
    a. Before you begin, you need to load the necessary libraries.
    b. Example:

```
# Load necessary libraries
library(ggplot2) For data visualization
library(corrplot) For correlation plots
```

2. **Load and Examine Data:**
    a. Load your dataset into R and examine its structure and summary statistics to get an initial understanding of the data.
    b. Example:

```
# Load your dataset
data <- read.csv("your_data.csv")
# Explore the structure and summary statistics
str(data)
summary(data)
```

3. **Correlation Matrix and Plot:**
    a. Create a correlation matrix to understand the relationships between numeric variables. Then, generate a correlation plot.
    b. Example:

```
# Calculate the correlation matrix
correlation_matrix <- cor(data, method = "pearson")
```

```
# Create a correlation plot using corrplot
corrplot(correlation_matrix, method = "color")
```

c. The correlation plot will display the strength and direction of correlations using colors.

4. Scatterplot Matrix:

a. To visualize relationships between pairs of numeric variables, create a scatterplot matrix.

b. Example:

```
# Create a scatterplot matrix using ggplot2
pairs(data)
```

c. This will produce a matrix of scatterplots showing pairwise relationships between numeric variables.

5. Heatmap:

a. If you want to visualize the relationships between variables, including both numeric and categorical, create a heatmap.

b. Example:

```
# Create a heatmap using ggplot2
ggplot(data, aes(x = your_variable1, y = your_variable2, fill =
your_numeric_variable)) + geom_tile()
        + labs(title = "Heatmap", x = "Variable 1", y = "Variable 2", fill = "Numeric
Variable")
```

c. Adjust `your_variable1`, `your_variable2`, and `your_numeric_variable` based on your dataset.

6. Pairwise Scatterplots:

a. To create scatterplots between pairs of numeric variables, you can use the `scatterplot` function from the `car` package.

b. Example:

```
# Install and load the car package
install.packages("car")
library(car)
```

```
#Create pairwise scatterplots
scatterplotMatrix(data)
```

c. This will display a grid of scatterplots for your numeric variables.

Remember that the actual variable names and data may vary depending on your dataset. The EDA process should be tailored to your specific data and research questions. The goal is to gain insights, detect patterns, and identify potential relationships in your data through various exploratory plots and visualizations.

Code (Script):

```
install.packages("ggplot2")
install.packages("corrplot")
install.packages("GGally")
library(GGally)
library(ggplot2)
library(corrplot)

data <- read.csv("company-sales.csv")
data

# Explore the structure and summary statistics
str(data)
summary(data)

# Calculate the correlation matrix
correlation_matrix <- cor(data, method = "pearson")
correlation_matrix

# Create a correlation plot using corrplot
corrplot(correlation_matrix, method = "color")

# Create a scatterplot matrix using ggplot2
ggpairs(data, title = "Scatterplot Matrix")
```

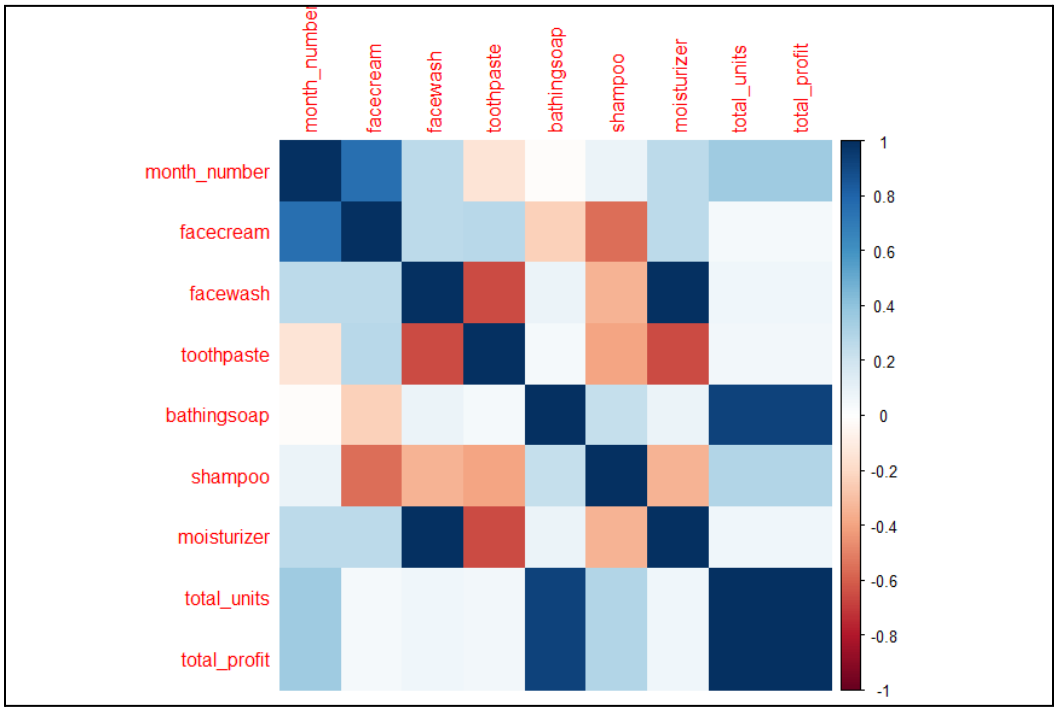
```
# Create boxplots or violin plots for numeric variables
ggplot(data, aes(x = "Group", y = facewash)) + geom_boxplot(fill = "blue") +
  labs(title = "Boxplot or Violin Plot", x = "Categorical Variable", y = "Numeric Variable")

# Create histograms
ggplot(data, aes(x = facewash)) + geom_histogram(binwidth = 40, fill = "blue") +
  labs(title = "Histogram", x = "Numeric Variable")

# Create a heatmap using ggplot2
ggplot(data, aes(x = total_units, y = bathingsoap)) + geom_tile(aes(fill = total_profit)) +
  labs(title = "Heatmap", x = "Categorical Variable 1", y = "Categorical Variable 2", fill =
       "Numeric Variable")
```

Output:

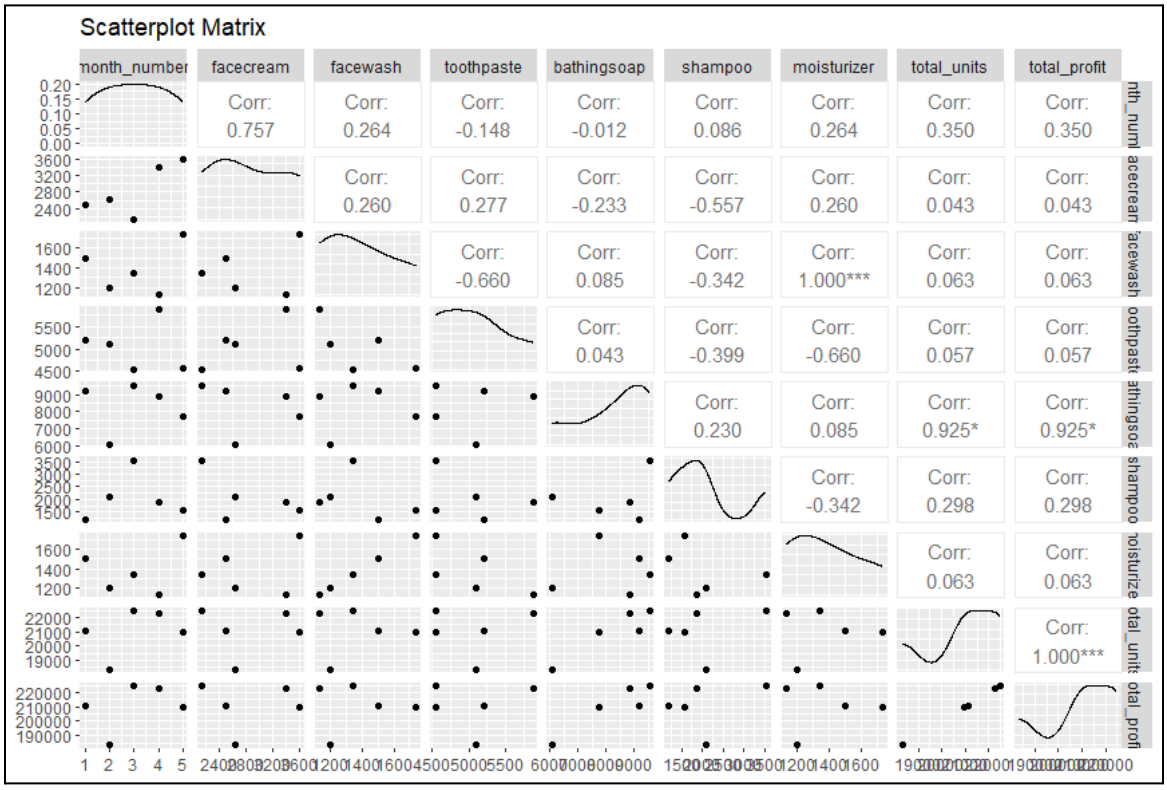
```
> library(GGally)
> library(ggplot2)
> library(corrplot)
> data <- read.csv("F:/Pushkar/MCA/Sem-1/DAR/SalesData1.csv")
> data
  month_number facecream facewash toothpaste bathingsoap shampoo moisturizer
total_units total_profit
1            1      2500     1500       5200        9200    1200        1500
21100       211000
2            2      2630     1200       5100        6100    2100        1200
18330       183300
3            3      2140     1340       4550        9550    3550        1340
22470       224700
4            4      3400     1130       5870        8870    1870        1130
22270       222700
5            5      3600     1740       4560        7760    1560        1740
20960       209600
> # Explore the structure and summary statistics
> str(data)
'data.frame':     5 obs. of  9 variables:
 $ month_number: int  1 2 3 4 5
 $ facecream   : int  2500 2630 2140 3400 3600
 $ facewash    : int  1500 1200 1340 1130 1740
 $ toothpaste  : int  5200 5100 4550 5870 4560
 $ bathingsoap : int  9200 6100 9550 8870 7760
 $ shampoo     : int  1200 2100 3550 1870 1560
 $ moisturizer : int  1500 1200 1340 1130 1740
 $ total_units : int  21100 18330 22470 22270 20960
 $ total_profit: int  211000 183300 224700 222700 209600
> summary(data)
  month_number    facecream       facewash       toothpaste     bathingsoap
shampoo      moisturizer    total_units
 Min.   :1      Min.   :2140   Min.   :1130   Min.   :4550   Min.   :6100
Min.   :1200   Min.   :1130   Min.   :18330
 1st Qu.:2     1st Qu.:2500   1st Qu.:1200   1st Qu.:4560   1st Qu.:7760   1st
Qu.:1560   1st Qu.:1200   1st Qu.:20960
 Median :3      Median :2630   Median :1340   Median :5100   Median :8870
Median :1870   Median :1340   Median :21100
 Mean   :3      Mean   :2854   Mean   :1382   Mean   :5056   Mean   :8296
Mean   :2056   Mean   :1382   Mean   :21026
 3rd Qu.:4     3rd Qu.:3400   3rd Qu.:1500   3rd Qu.:5200   3rd Qu.:9200   3rd
Qu.:2100   3rd Qu.:1500   3rd Qu.:22270
 Max.   :5      Max.   :3600   Max.   :1740   Max.   :5870   Max.   :9550
Max.   :3550   Max.   :1740   Max.   :22470
  total_profit
 Min.   :183300
```

```
 1st Qu.:209600
 Median :211000
 Mean   :210260
 3rd Qu.:222700
 Max.   :224700
> # Calculate the correlation matrix
> correlation_matrix <- cor(data, method = "pearson")
> correlation_matrix
               month_number    facecream     facewash   toothpaste bathingsoap
shampoo moisturizer total_units
month_number     1.00000000   0.75684574   0.26438960  -0.14800837  -0.01243202
0.08598715  0.26438960  0.35038895
facecream        0.75684574   1.00000000   0.26039407   0.27724218  -0.23326099
-0.55680046  0.26039407  0.04310301
facewash         0.26438960   0.26039407   1.00000000  -0.65960883   0.08537187
-0.34226770  1.00000000  0.06274722
toothpaste      -0.14800837   0.27724218  -0.65960883   1.00000000   0.04333450
-0.39860046 -0.65960883  0.05743385
bathingsoap     -0.01243202  -0.23326099   0.08537187   0.04333450   1.00000000
0.23048226  0.08537187  0.92482277
shampoo          0.08598715  -0.55680046  -0.34226770  -0.39860046   0.23048226
1.00000000 -0.34226770  0.29848723
moisturizer      0.26438960   0.26039407   1.00000000  -0.65960883   0.08537187
-0.34226770  1.00000000  0.06274722
total_units      0.35038895   0.04310301   0.06274722   0.05743385   0.92482277
0.29848723  0.06274722  1.00000000
total_profit     0.35038895   0.04310301   0.06274722   0.05743385   0.92482277
0.29848723  0.06274722  1.00000000
             total_profit
month_number   0.35038895
facecream      0.04310301
facewash       0.06274722
toothpaste     0.05743385
bathingsoap    0.92482277
shampoo        0.29848723
moisturizer    0.06274722
total_units    1.00000000
total_profit   1.00000000
```

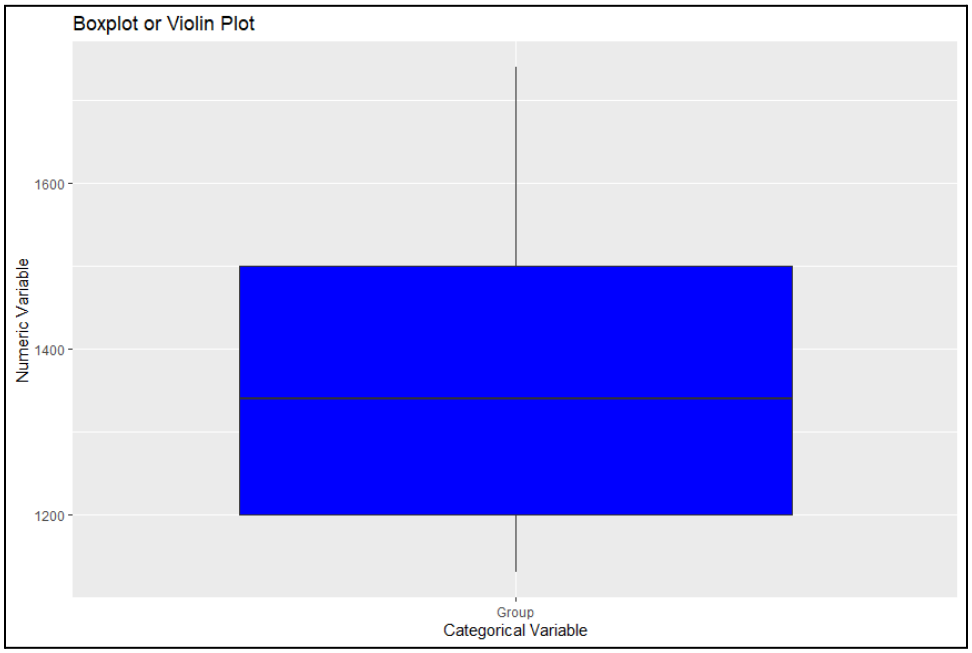
```
> # Create a correlation plot using corrplot
> corrplot(correlation_matrix, method = "color")
```



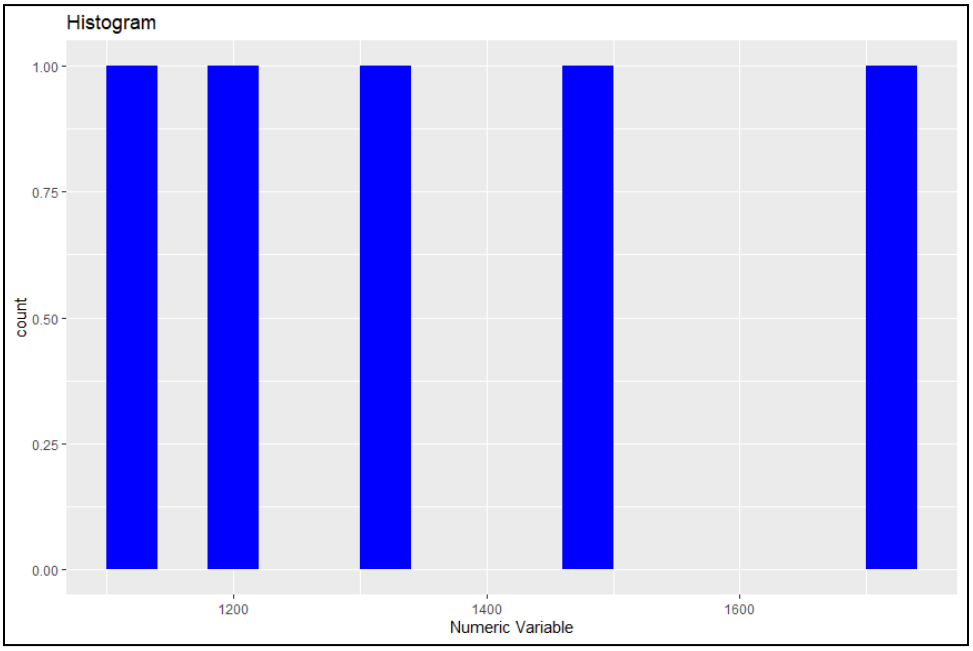

```
> # Create a scatterplot matrix using ggplot2
> ggpairs(data, title = "Scatterplot Matrix")
```



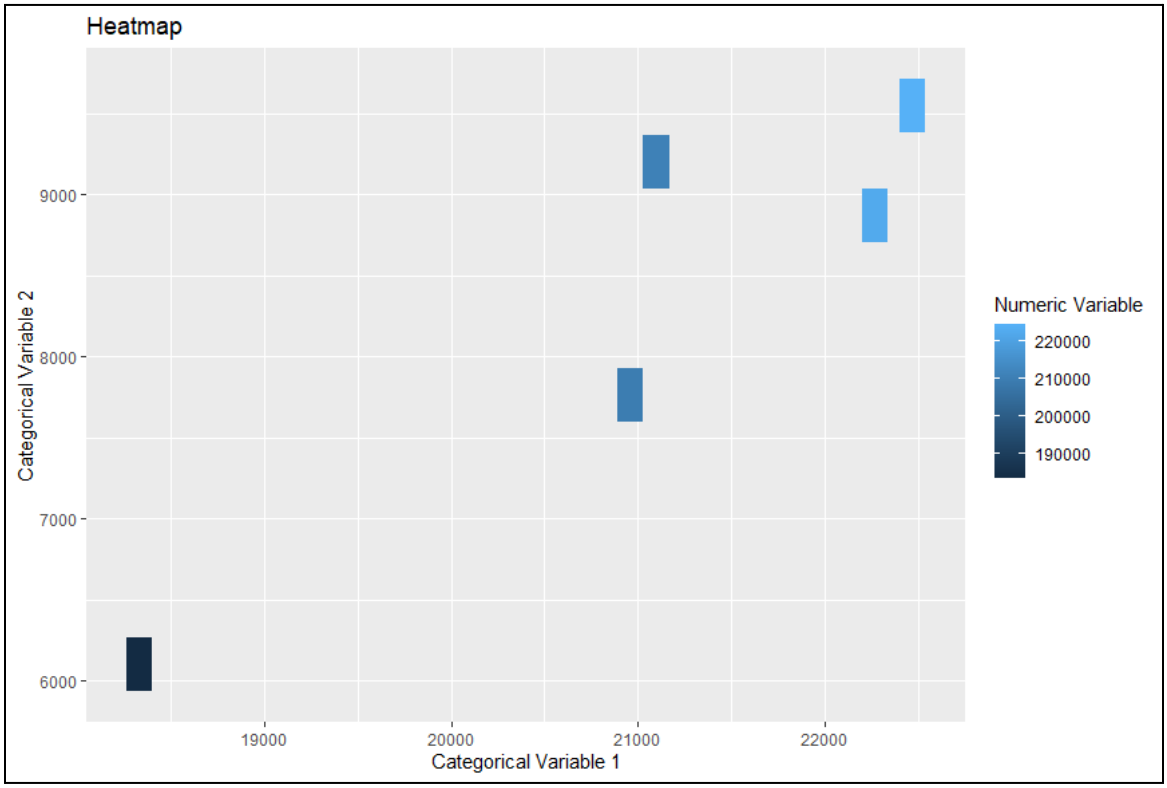
```
> # Create boxplots or violin plots for numeric variables
> ggplot(data, aes(x = "Group", y = facewash)) +
+  geom_boxplot(fill = "blue") +
+  labs(title = "Boxplot or Violin Plot", x = "Categorical Variable", y =
"Numeric Variable")
```




```
> # Create histograms
> ggplot(data, aes(x = facewash)) +
+  geom_histogram(binwidth = 40, fill = "blue") +
+  labs(title = "Histogram", x = "Numeric Variable")
```



```
> # Create a heatmap using ggplot2
> ggplot(data, aes(x = total_units, y = bathingsoap)) +
+  geom_tile(aes(fill = total_profit)) +
+ labs(title = "Heatmap", x = "Categorical Variable 1", y = "Categorical Variable 2", fill = + "Numeric Variable")
```



Conclusion:

In this practical we learned the EDA(Exploratory Data Analytics) process by which we can gain insights, detect patterns, and identify potential relationships in our data through various exploratory plots and visualizations.