

# FML\_Assignment

Namrah

2023-09-10

## This dataset is downloaded from Kaggle

```
library(openxlsx)
library(ggplot2)

# Read the Excel file
shipment_dataset <- read.xlsx("C:/Users/Namrah/Desktop/Shipment_data.xlsx")

# To display the first few rows of the dataset
head(shipment_dataset)
```

##	ID	Warehouse_block	Mode_of_Shipment	Customer_care_calls	Customer_rating
## 1	1	D	Flight	4	2
## 2	2	F	Flight	4	5
## 3	3	A	Flight	2	2
## 4	4	B	Flight	3	3
## 5	5	C	Flight	2	2
## 6	6	F	Flight	3	1

##	Cost_of_the_Product	Prior_purchases	Product_importance	Gender
## 1	177	3	low	F
## 2	216	2	low	M
## 3	183	4	low	M
## 4	176	4	medium	M
## 5	184	3	medium	F
## 6	162	3	medium	F

##	Discount_offered	Weight_in_gms	Reached.on.Time_Y.N
## 1	44	1233	1
## 2	59	3088	1
## 3	48	3374	1
## 4	10	1177	1
## 5	46	2484	1
## 6	12	1417	1

```
# Selecting quantitative variables
quantitative_vars <- c("Customer_care_calls", "Customer_rating",
"Cost_of_the_Product", "Prior_purchases", "Discount_offered",
"Weight_in_gms")

# Selecting categorical variables
categorical_vars <- c("Warehouse_block", "Mode_of_Shipment",
"Product_importance", "Gender", "Reached.on.Time_Y.N")
```

```

# Printing descriptive statistics for quantitative variables
quantitative_summary <- summary(shipment_dataset[, quantitative_vars])
print("Descriptive Statistics for Quantitative Variables:")

## [1] "Descriptive Statistics for Quantitative Variables:"

print(quantitative_summary)

## Customer_care_calls Customer_rating Cost_of_the_Product Prior_purchases
## Min. :2.000 Min. :1.000 Min. : 96.0 Min. : 2.000
## 1st Qu.:3.000 1st Qu.:2.000 1st Qu.:169.0 1st Qu.: 3.000
## Median :4.000 Median :3.000 Median :214.0 Median : 3.000
## Mean :4.054 Mean :2.991 Mean :210.2 Mean : 3.568
## 3rd Qu.:5.000 3rd Qu.:4.000 3rd Qu.:251.0 3rd Qu.: 4.000
## Max. :7.000 Max. :5.000 Max. :310.0 Max. :10.000
## Discount_offered Weight_in_gms
## Min. : 1.00 Min. :1001
## 1st Qu.: 4.00 1st Qu.:1840
## Median : 7.00 Median :4149
## Mean :13.37 Mean :3634
## 3rd Qu.:10.00 3rd Qu.:5050
## Max. :65.00 Max. :7846

# Printing frequency tables for categorical variables
print("\nDescriptive Statistics for Categorical Variables:")

## [1] "\nDescriptive Statistics for Categorical Variables:"

for (var in categorical_vars) {
  cat(paste(var, ":\n"))
  print(table(shipment_dataset[, var]))
  cat("\n")
}

## Warehouse_block :
##
## A B C D F
## 1833 1833 1833 1834 3666
##
## Mode_of_Shipment :
##
## Flight Road Ship
## 1777 1760 7462
##
## Product_importance :
##
## high low medium
## 948 5297 4754
##
## Gender :

```

```
##
##      F      M
## 5545 5454
##
## Reached.on.Time_Y.N :
##
##      0      1
## 4436 6563

# Transform the 'Weight_in_gms' variable and create a new variable
# 'sqrt_Weight_in_gms'
shipment_dataset$sqrt_Weight_in_gms <- sqrt(shipment_dataset$Weight_in_gms)
head(shipment_dataset)
```

##	ID	Warehouse_block	Mode_of_Shipment	Customer_care_calls	Customer_rating
## 1	1		D Flight	4	2
## 2	2		F Flight	4	5
## 3	3		A Flight	2	2
## 4	4		B Flight	3	3
## 5	5		C Flight	2	2
## 6	6		F Flight	3	1

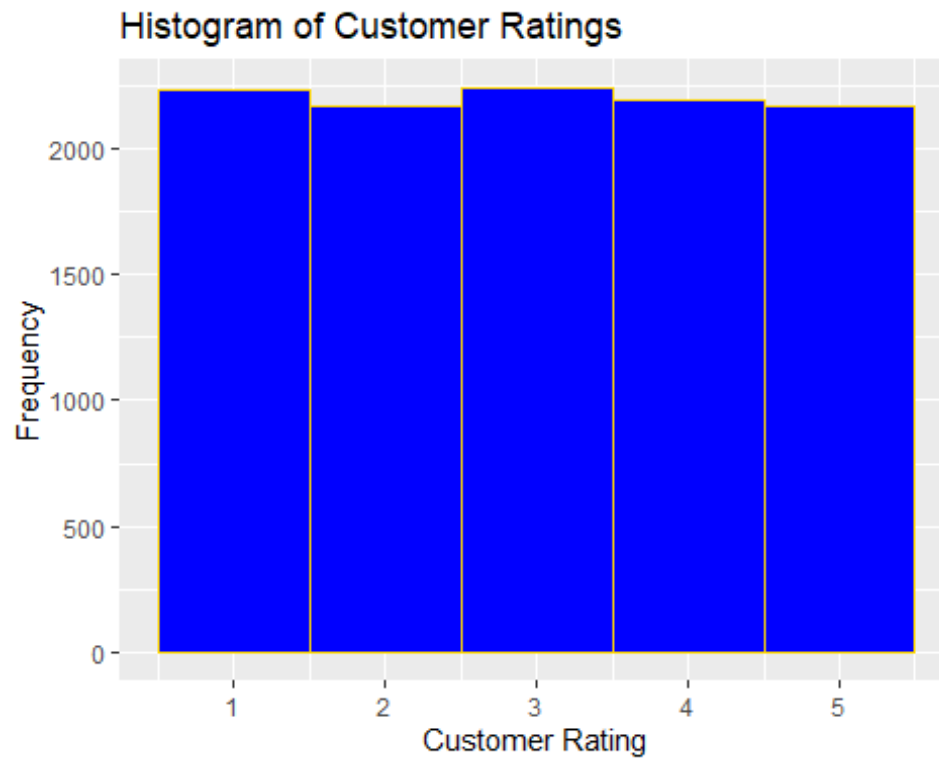
  

##	Cost_of_the_Product	Prior_purchases	Product_importance	Gender
## 1	177	3	low	F
## 2	216	2	low	M
## 3	183	4	low	M
## 4	176	4	medium	M
## 5	184	3	medium	F
## 6	162	3	medium	F

##	Discount_offered	Weight_in_gms	Reached.on.Time_Y.N	sqrt_Weight_in_gms
## 1	44	1233	1	35.11410
## 2	59	3088	1	55.56978
## 3	48	3374	1	58.08614
## 4	10	1177	1	34.30743
## 5	46	2484	1	49.83974
## 6	12	1417	1	37.64306

```
# Creating a histogram for the 'Customer_rating' variable
ggplot(shipment_dataset, aes(x = Customer_rating)) +
  geom_histogram(binwidth = 1, fill = "blue", color = "gold") +
  labs(title = "Histogram of Customer Ratings",
       x = "Customer Rating",
       y = "Frequency")
```



```
# Creating a scatterplot between 'Prior_purchases' and 'Cost_of_the_Product'
plot(shipment_dataset$Prior_purchases, shipment_dataset$Cost_of_the_Product,
     xlab = "Prior Purchases", ylab = "Cost of the Product",
     main = "Scatterplot of Prior Purchases vs. Cost of the Product",
     col = "purple", pch = 20) # Customize colors and point shapes
```

## Scatterplot of Prior Purchases vs. Cost of the Product

