A1

“ADITI”

2024-02-06

#Dataset - <https://www.kaggle.com/datasets/aungpyaeap/supermarket-sales>

knitr::opts\_chunk$set(echo = TRUE)  
library(readr)  
Supermarket <- read\_csv("C:/Users/tiwar/OneDrive/Desktop/FML ASSIGNMENT/supermarket\_sales - Sheet1.csv",show\_col\_types = FALSE)

View(Supermarket)  
summary(Supermarket)

## Invoice ID Branch City Customer type   
## Length:1000 Length:1000 Length:1000 Length:1000   
## Class :character Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character Mode :character   
##   
##   
##   
## Gender Product line Unit price Quantity   
## Length:1000 Length:1000 Min. :10.08 Min. : 1.00   
## Class :character Class :character 1st Qu.:32.88 1st Qu.: 3.00   
## Mode :character Mode :character Median :55.23 Median : 5.00   
## Mean :55.67 Mean : 5.51   
## 3rd Qu.:77.94 3rd Qu.: 8.00   
## Max. :99.96 Max. :10.00   
## Tax 5% Total Date Time   
## Min. : 0.5085 Min. : 10.68 Length:1000 Length:1000   
## 1st Qu.: 5.9249 1st Qu.: 124.42 Class :character Class1:hms   
## Median :12.0880 Median : 253.85 Mode :character Class2:difftime   
## Mean :15.3794 Mean : 322.97 Mode :numeric   
## 3rd Qu.:22.4453 3rd Qu.: 471.35   
## Max. :49.6500 Max. :1042.65   
## Payment cogs gross margin percentage gross income   
## Length:1000 Min. : 10.17 Min. :4.762 Min. : 0.5085   
## Class :character 1st Qu.:118.50 1st Qu.:4.762 1st Qu.: 5.9249   
## Mode :character Median :241.76 Median :4.762 Median :12.0880   
## Mean :307.59 Mean :4.762 Mean :15.3794   
## 3rd Qu.:448.90 3rd Qu.:4.762 3rd Qu.:22.4453   
## Max. :993.00 Max. :4.762 Max. :49.6500   
## Rating   
## Min. : 4.000   
## 1st Qu.: 5.500   
## Median : 7.000   
## Mean : 6.973   
## 3rd Qu.: 8.500   
## Max. :10.000

head(Supermarket,5)#read the first 5 rows of the dataset

## # A tibble: 5 × 17  
## `Invoice ID` Branch City `Customer type` Gender `Product line` `Unit price`  
## <chr> <chr> <chr> <chr> <chr> <chr> <dbl>  
## 1 750-67-8428 A Yangon Member Female Health and be… 74.7  
## 2 226-31-3081 C Naypyi… Normal Female Electronic ac… 15.3  
## 3 631-41-3108 A Yangon Normal Male Home and life… 46.3  
## 4 123-19-1176 A Yangon Member Male Health and be… 58.2  
## 5 373-73-7910 A Yangon Normal Male Sports and tr… 86.3  
## # ℹ 10 more variables: Quantity <dbl>, `Tax 5%` <dbl>, Total <dbl>, Date <chr>,  
## # Time <time>, Payment <chr>, cogs <dbl>, `gross margin percentage` <dbl>,  
## # `gross income` <dbl>, Rating <dbl>

numeric\_columns <- Supermarket[sapply(Supermarket, is.numeric)]  
print(numeric\_columns)

## # A tibble: 1,000 × 8  
## `Unit price` Quantity `Tax 5%` Total cogs `gross margin percentage`  
## <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 74.7 7 26.1 549. 523. 4.76  
## 2 15.3 5 3.82 80.2 76.4 4.76  
## 3 46.3 7 16.2 341. 324. 4.76  
## 4 58.2 8 23.3 489. 466. 4.76  
## 5 86.3 7 30.2 634. 604. 4.76  
## 6 85.4 7 29.9 628. 598. 4.76  
## 7 68.8 6 20.7 434. 413. 4.76  
## 8 73.6 10 36.8 772. 736. 4.76  
## 9 36.3 2 3.63 76.1 72.5 4.76  
## 10 54.8 3 8.23 173. 165. 4.76  
## # ℹ 990 more rows  
## # ℹ 2 more variables: `gross income` <dbl>, Rating <dbl>

dim(Supermarket)#It gives the dimension of the dataset

## [1] 1000 17

sum(is.na(Supermarket))#To check the null values in a dataset

## [1] 0

library(dplyr)

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

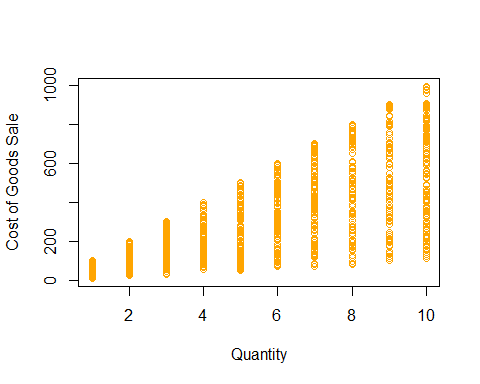
## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

By\_total = group\_by(Supermarket, Gender)  
summarise(By\_total, Avg\_total = mean(Total), qty = n())#Here the overview of sales according to the gender is evaluated. As well as average total is calculatd.

## # A tibble: 2 × 3  
## Gender Avg\_total qty  
## <chr> <dbl> <int>  
## 1 Female 335. 501  
## 2 Male 311. 499

## Including Plots

x <- Supermarket$Quantity  
y <- Supermarket$cogs  
plot(x, y, col = "orange",xlab="Quantity", ylab="Cost of Goods Sale")



x <- Supermarket$'Unit price'  
y <- Supermarket$'Total'  
plot(x, y, col = "red", main="Unit Price Vs Total",  
 xlab="Unit price", ylab="Total")

