On\_Line

Dinesh\_Kumar

2025-06-05

# Load libraries  
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

library(ggplot2)  
library(lubridate)

##   
## Attaching package: 'lubridate'

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

library(cluster)  
library(forecast)

## Registered S3 method overwritten by 'quantmod':  
## method from  
## as.zoo.data.frame zoo

library(tidyr)

setwd("C:/Users/me/Desktop/Online\_Retail")

df = read.csv("Online Retail.csv")

head(df)

## InvoiceNo StockCode Description Quantity  
## 1 536365 85123A WHITE HANGING HEART T-LIGHT HOLDER 6  
## 2 536365 71053 WHITE METAL LANTERN 6  
## 3 536365 84406B CREAM CUPID HEARTS COAT HANGER 8  
## 4 536365 84029G KNITTED UNION FLAG HOT WATER BOTTLE 6  
## 5 536365 84029E RED WOOLLY HOTTIE WHITE HEART. 6  
## 6 536365 22752 SET 7 BABUSHKA NESTING BOXES 2  
## InvoiceDate UnitPrice CustomerID Country  
## 1 2010-12-01 08:26:00 2.55 17850 United Kingdom  
## 2 2010-12-01 08:26:00 3.39 17850 United Kingdom  
## 3 2010-12-01 08:26:00 2.75 17850 United Kingdom  
## 4 2010-12-01 08:26:00 3.39 17850 United Kingdom  
## 5 2010-12-01 08:26:00 3.39 17850 United Kingdom  
## 6 2010-12-01 08:26:00 7.65 17850 United Kingdom

str(df)

## 'data.frame': 541909 obs. of 8 variables:  
## $ InvoiceNo : chr "536365" "536365" "536365" "536365" ...  
## $ StockCode : chr "85123A" "71053" "84406B" "84029G" ...  
## $ Description: chr "WHITE HANGING HEART T-LIGHT HOLDER" "WHITE METAL LANTERN" "CREAM CUPID HEARTS COAT HANGER" "KNITTED UNION FLAG HOT WATER BOTTLE" ...  
## $ Quantity : int 6 6 8 6 6 2 6 6 6 32 ...  
## $ InvoiceDate: chr "2010-12-01 08:26:00" "2010-12-01 08:26:00" "2010-12-01 08:26:00" "2010-12-01 08:26:00" ...  
## $ UnitPrice : num 2.55 3.39 2.75 3.39 3.39 7.65 4.25 1.85 1.85 1.69 ...  
## $ CustomerID : int 17850 17850 17850 17850 17850 17850 17850 17850 17850 13047 ...  
## $ Country : chr "United Kingdom" "United Kingdom" "United Kingdom" "United Kingdom" ...

# Convert date column

df$InvoiceDate <- as.Date(df$InvoiceDate, format = "%y-%m-%d")

#Check columns name

colnames(df)

## [1] "InvoiceNo" "StockCode" "Description" "Quantity" "InvoiceDate"  
## [6] "UnitPrice" "CustomerID" "Country"

#Unit Price column edit

colnames(df)[6] <- "Unit\_Price"

# Handle missing values

df <- df %>% drop\_na()

#Total Monthly Sales

df <- df %>%  
 mutate(  
 year = year(InvoiceDate),  
 month = month(InvoiceDate, label = TRUE),  
 total\_price = Unit\_Price\* Quantity  
   
 )

Monthly\_sales <- df %>%  
 group\_by(year, month) %>%  
 summarise(TotalSales = sum(total\_price), .groups = "drop")