Exploratory Analysis

KRM

2024-07-15

EXPLAROTARY ANALYSIS (EDA)

# Load packages

# Load the data

library("readxl")   
retail\_data<-read\_excel("C:/Users/maniv/Desktop/TATA/Online Retail.xlsx")

# Number of rows of the data

nrow(retail\_data)

## [1] 541909

# Number of Variables or columns of the data

ncol(retail\_data)

## [1] 8

# Variables or column names of the data

colnames(retail\_data)

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

# Variables or column names of the data

head(retail\_data)

## # A tibble: 6 × 8  
## InvoiceNo StockCode Description Quantity InvoiceDate UnitPrice  
## <chr> <chr> <chr> <dbl> <dttm> <dbl>  
## 1 536365 85123A WHITE HANGING HEAR… 6 2010-12-01 08:26:00 2.55  
## 2 536365 71053 WHITE METAL LANTERN 6 2010-12-01 08:26:00 3.39  
## 3 536365 84406B CREAM CUPID HEARTS… 8 2010-12-01 08:26:00 2.75  
## 4 536365 84029G KNITTED UNION FLAG… 6 2010-12-01 08:26:00 3.39  
## 5 536365 84029E RED WOOLLY HOTTIE … 6 2010-12-01 08:26:00 3.39  
## 6 536365 22752 SET 7 BABUSHKA NES… 2 2010-12-01 08:26:00 7.65  
## # ℹ 2 more variables: CustomerID <dbl>, Country <chr>

# View the sample data

View(retail\_data)

# Structure of the data

str(retail\_data)

## tibble [541,909 × 8] (S3: tbl\_df/tbl/data.frame)  
## $ InvoiceNo : chr [1:541909] "536365" "536365" "536365" "536365" ...  
## $ StockCode : chr [1:541909] "85123A" "71053" "84406B" "84029G" ...  
## $ Description: chr [1:541909] "WHITE HANGING HEART T-LIGHT HOLDER" "WHITE METAL LANTERN" "CREAM CUPID HEARTS COAT HANGER" "KNITTED UNION FLAG HOT WATER BOTTLE" ...  
## $ Quantity : num [1:541909] 6 6 8 6 6 2 6 6 6 32 ...  
## $ InvoiceDate: POSIXct[1:541909], format: "2010-12-01 08:26:00" "2010-12-01 08:26:00" ...  
## $ UnitPrice : num [1:541909] 2.55 3.39 2.75 3.39 3.39 7.65 4.25 1.85 1.85 1.69 ...  
## $ CustomerID : num [1:541909] 17850 17850 17850 17850 17850 ...  
## $ Country : chr [1:541909] "United Kingdom" "United Kingdom" "United Kingdom" "United Kingdom" ...

# summary of the data statistics

summary(retail\_data)

## InvoiceNo StockCode Description Quantity   
## Length:541909 Length:541909 Length:541909 Min. :-80995.00   
## Class :character Class :character Class :character 1st Qu.: 1.00   
## Mode :character Mode :character Mode :character Median : 3.00   
## Mean : 9.55   
## 3rd Qu.: 10.00   
## Max. : 80995.00   
##   
## InvoiceDate UnitPrice CustomerID   
## Min. :2010-12-01 08:26:00.00 Min. :-11062.06 Min. :12346   
## 1st Qu.:2011-03-28 11:34:00.00 1st Qu.: 1.25 1st Qu.:13953   
## Median :2011-07-19 17:17:00.00 Median : 2.08 Median :15152   
## Mean :2011-07-04 13:34:57.16 Mean : 4.61 Mean :15288   
## 3rd Qu.:2011-10-19 11:27:00.00 3rd Qu.: 4.13 3rd Qu.:16791   
## Max. :2011-12-09 12:50:00.00 Max. : 38970.00 Max. :18287   
## NA's :135080   
## Country   
## Length:541909   
## Class :character   
## Mode :character   
##   
##   
##   
##

variable.names(retail\_data)

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

UNIQE value of Variable country verified

unique(retail\_data$Country)

## [1] "United Kingdom" "France" "Australia"   
## [4] "Netherlands" "Germany" "Norway"   
## [7] "EIRE" "Switzerland" "Spain"   
## [10] "Poland" "Portugal" "Italy"   
## [13] "Belgium" "Lithuania" "Japan"   
## [16] "Iceland" "Channel Islands" "Denmark"   
## [19] "Cyprus" "Sweden" "Austria"   
## [22] "Israel" "Finland" "Bahrain"   
## [25] "Greece" "Hong Kong" "Singapore"   
## [28] "Lebanon" "United Arab Emirates" "Saudi Arabia"   
## [31] "Czech Republic" "Canada" "Unspecified"   
## [34] "Brazil" "USA" "European Community"   
## [37] "Malta" "RSA"