Assignment2

Niharika D

2022-10-26

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

Online\_Retail<-read.csv("./Online\_Retail.csv")  
head(Online\_Retail)

## 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 12/1/2010 8:26 2.55 17850 United Kingdom  
## 2 12/1/2010 8:26 3.39 17850 United Kingdom  
## 3 12/1/2010 8:26 2.75 17850 United Kingdom  
## 4 12/1/2010 8:26 3.39 17850 United Kingdom  
## 5 12/1/2010 8:26 3.39 17850 United Kingdom  
## 6 12/1/2010 8:26 7.65 17850 United Kingdom

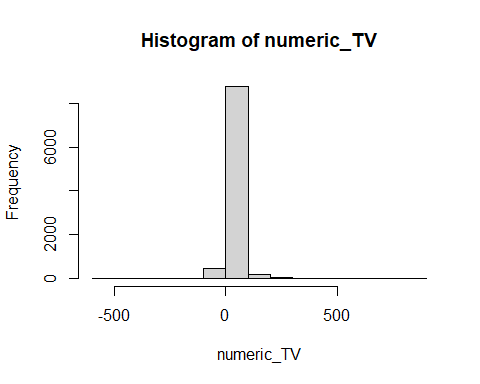
#1  
trans\_countries<-Online\_Retail %>% group\_by(Country) %>% summarise(cnt = n()) %>% mutate(perc =round((cnt/sum(cnt))\*100,5)) %>% filter(perc>1)  
View(trans\_countries)

#2  
TransactionValue<-Online\_Retail$Quantity\*Online\_Retail$UnitPrice  
  
#creating a dataframe  
  
Online\_Retail\_new<-data.frame(InvoiceNo=Online\_Retail$InvoiceNo,StockCode= Online\_Retail$StockCode,Description=Online\_Retail$Description,Quantity=Online\_Retail$Quantity, InvoiceDate=Online\_Retail$InvoiceDate,UnitPrice=Online\_Retail$UnitPrice, CustomerID=Online\_Retail$CustomerID,Country=Online\_Retail$Country, TransactionValue=TransactionValue)  
  
  
head(Online\_Retail\_new)

## 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 TransactionValue  
## 1 12/1/2010 8:26 2.55 17850 United Kingdom 15.30  
## 2 12/1/2010 8:26 3.39 17850 United Kingdom 20.34  
## 3 12/1/2010 8:26 2.75 17850 United Kingdom 22.00  
## 4 12/1/2010 8:26 3.39 17850 United Kingdom 20.34  
## 5 12/1/2010 8:26 3.39 17850 United Kingdom 20.34  
## 6 12/1/2010 8:26 7.65 17850 United Kingdom 15.30

#3  
Trans\_value\_countries<- Online\_Retail\_new %>% group\_by(Country) %>% summarise(sum\_TransactionValue = sum(TransactionValue)) %>% filter(sum\_TransactionValue>130000)  
  
View(Trans\_value\_countries)

##5  
  
  
histogram<- Online\_Retail\_new %>% filter(Country == 'Germany')   
numeric\_TV<- as.integer(histogram$TransactionValue)  
hist(numeric\_TV)



##6  
  
cust\_count<-Online\_Retail\_new %>% group\_by(CustomerID) %>% summarise(cntt = n()) %>% arrange(desc(cntt))  
View(cust\_count)  
  
#Customer 17841 has the highest number of transactions.  
  
cust\_sum<-Online\_Retail\_new %>% group\_by(CustomerID) %>% summarise(sum\_cnt =sum(TransactionValue)) %>% arrange(desc(sum\_cnt))  
  
View(cust\_sum)  
  
#Customer 14646 is the most valuable

#7  
  
missing\_values<- (colMeans(is.na(Online\_Retail\_new))\*100)  
  
View(missing\_values)

#8  
missing<-Online\_Retail\_new %>% filter(is.na(CustomerID)) %>% group\_by(Country) %>% summarise(Countries = n())  
  
View(missing)

#9

#10  
return\_rate <- Online\_Retail\_new %>% filter(Country=='France')  
cancelled\_customers <- nrow(subset(return\_rate,TransactionValue<0))  
View(cancelled\_customers)  
  
return\_rate\_french\_cust=(cancelled\_customers/8556)  
  
View(return\_rate\_french\_cust)

##11  
item\_sum<-Online\_Retail\_new %>% group\_by(Description) %>% summarise(sum\_cnt = sum(TransactionValue)) %>% arrange(desc(sum\_cnt))  
View(item\_sum)  
  
#DOTCOM POSTAGE generates highest revenue for the retailer

##12  
  
xx<- Online\_Retail\_new %>% distinct(CustomerID) %>% summarise(ncount = n())  
View(xx)