

ASSIGNMENT 2 BA

Rachana

2022-10-31

```
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("C://Users//kurra//Downloads//Online_Retail.csv")
```

1

Breakdown of the number of transactions by countries in both percentage and count, and showing the countries accounting for more than 1% of the total transactions

```
data<-Online_Retail %>% group_by(Country) %>% summarise(cnt = n()) %>% mutate(perc =round((cnt/sum(cnt))
head(data)

## # A tibble: 4 x 3
##   Country      cnt  perc
##   <chr>     <int> <dbl>
## 1 EIRE        8196    2
## 2 France      8557    2
## 3 Germany     9495    2
## 4 United Kingdom 495478  91
```

##2. Creating a new variable TransactionValue and adding it to the dataframe:

```
Online_Retail$TransactionValue<-c(Online_Retail$Quantity*Online_Retail$UnitPrice)
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 TransactionValue
## 1 12/1/2010     2.55       17850 United Kingdom      15.30
## 2 12/1/2010     3.39       17850 United Kingdom      20.34
## 3 12/1/2010     2.75       17850 United Kingdom      22.00
## 4 12/1/2010     3.39       17850 United Kingdom      20.34
## 5 12/1/2010     3.39       17850 United Kingdom      20.34
## 6 12/1/2010     7.65       17850 United Kingdom      15.30
```

##3.Transaction value breakdown by countries

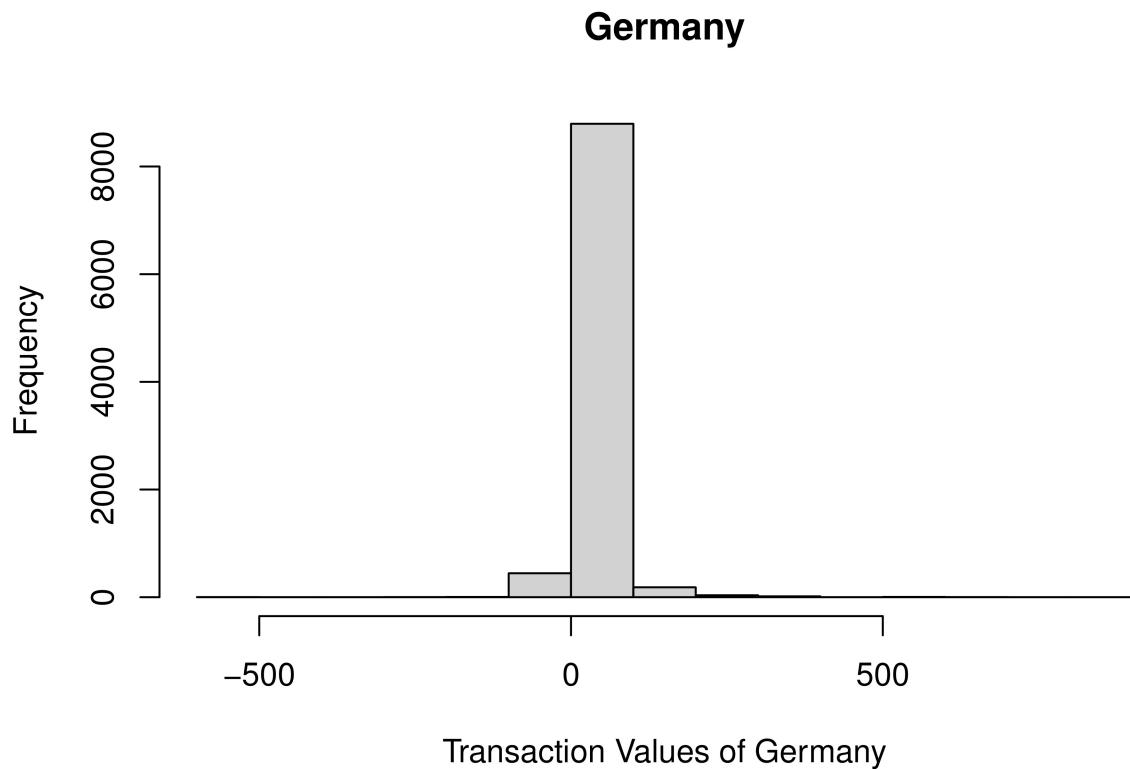
```
breakdown<-Online_Retail %>% group_by(Country) %>% summarise(sum_TransactionValue = sum(TransactionValue))

head(breakdown)
```

```
## # A tibble: 6 x 2
##   Country      sum_TransactionValue
##   <chr>            <dbl>
## 1 Australia      137077.
## 2 Belgium        40911.
## 3 Channel Islands 20086.
## 4 Denmark        18768.
## 5 EIRE           263277.
## 6 Finland        22327.
```

#5 plotting histogram

```
germany_histogram <- subset(Online_Retail$TransactionValue, Online_Retail$Country == "Germany")
hist(germany_histogram, xlab = "Transaction Values of Germany", main = "Germany")
```



#6 Finding Which customer had the highest number of transactions? Which customer is most valuable

```
Online_new <- na.omit(Online_Retail)
highest<- summarise(group_by(Online_new, CustomerID), sum2= sum(TransactionValue))
highest[which.max(highest$sum2),]
```

```
## # A tibble: 1 x 2
##   CustomerID     sum2
##       <int>    <dbl>
## 1        14646 279489.
```

#CUSTOMER 14646 IS THE MOST VALUABLE CUSTOMER

```
temp <- table(Online_Retail$CustomerID)
temp<- as.data.frame(temp)
temp1 <- temp[which.max(temp$Freq),]
temp1
```

```
##      Var1 Freq
## 4043 17841 7983
```

#CUSTOMER 17841 IS HAVING HIGHEST NUMBER OF TRANSACTIONS

7 MISSING VALUES IN DATASET

```
Online_Retail %>% is.na() %>% colMeans() * 100
```

```
##           InvoiceNo      StockCode      Description      Quantity
##           0.00000       0.00000       0.00000       0.00000
##      InvoiceDate      UnitPrice      CustomerID      Country
##           0.00000       0.00000      24.92669       0.00000
## TransactionValue
##           0.00000
```

#8 The number of transactions with missing CustomerID records by countries

```
miss <- Online_Retail %>% filter(is.na(CustomerID)) %>% group_by(Country) %>% count()
View(miss)
```

#9 average number of days the customer comeback to the website for next shopping

```
Time= strptime(Online_Retail$InvoiceDate, format = '%m/%d/%Y %H:%M', tz = 'GMT')
```

```
Online_Retail$New_Invoice_Date <- as.Date(Time)
```

```
del<-na.omit(Online_Retail)
del<- subset(del, Quantity > 0)
sub<- del[,c("CustomerID","New_Invoice_Date")]
uni <- distinct(sub)
uni %>% group_by(CustomerID) %>% arrange(New_Invoice_Date) %>% summarise(avg=mean(diff(New_Invoice_Date)))%>%
```

```
## # A tibble: 1 x 1
##   avg_days_between_shopping
##   <dbl>
## 1 78.42025
```

#10 Return rate of France

```
temp<-Online_Retail %>% filter(Country=="France") %>% count()
temp1<-Online_Retail %>% filter(Country=="France" & Quantity<0) %>% summarize(count = n())
rate = (temp1/temp)*100
View(rate)
```

```
#return rate = 1.74%
```

11 The product that has generated the highest revenue

```
TransactionValue<- tapply(Online_Retail$TransactionValue, Online_Retail$StockCode, sum)
TransactionValue[which.max(TransactionValue)]
```

```
##      DOT
## 206245.5
```

```
#The product with stock code DOT has generated the highest revenue
```

```
#12 number of unique customers
```

```
unique_customers <- unique(Online_Retail$CustomerID)
length(unique_customers)
```

```
## [1] 4373
```