#Read UK Bank Customers Data

Bank = read\_xls("UK-Bank-Customers.xls",sheet = 1)

#Checked table top data

head(Bank)

#structure of data

str(Bank)

#View of data

View(Bank)

#Changed Column Names

colnames(Bank) = c("CustomerID","Name","Surname","Gender","Age",

"Region","JobClassification","DateJoined","Balance")

colnames(Bank)

#Conversion of date

Bank$DateJoined = as.Date(Bank$DateJoined,"%d.%b.%y")

Bank$DateJoined = format(Bank$DateJoined,"20%y-%b-%d")

#Summary

summary(Bank)

#Applying factor to variables

Bank$Gender = factor(Bank$Gender)

Bank$Region = factor(Bank$Region)

Bank$JobClassification = factor(Bank$JobClassification)

Bank$DateJoined = factor(Bank$DateJoined)

#Splitting date column into Year,month and day using tidyr package

DateSplit = separate(Bank,DateJoined,c("Year","Month","Day"),sep = "-")

View(DateSplit)

#Applying factor to Datesplit variables

DateSplit$Year = factor(DateSplit$Year)

?month.abb

DateSplit$Month = factor(DateSplit$Month,levels = month.abb)

DateSplit$Day = factor(DateSplit$Day)

#plotting customers joined on various months in 2015

ggplot(data = DateSplit,aes(Month))+geom\_bar(fill="Blue")+

ggtitle("Customers Joined on Various months")+

xlab("Month")+ylab("No of customers")+

scale\_y\_continuous(breaks = seq.int(0,600,50))+

theme(plot.title = element\_text(colour ="Black",size = 20),

axis.title.x = element\_text(colour = "Green",size = 15),

axis.title.y = element\_text(colour = "Green",size = 15))

#Datesplit Summary

summary(DateSplit$Month)

#write summary of customer joined details to excel file

a = summary(DateSplit$Month)

View(a)

a = as.data.frame(a)

colnames(a) = "No of Customers"

write.xlsx(a,"UK-Customer-Analysis.xlsx",

sheetName = "Month vs Customers Joined")

#Plotting customers count from various regions

ggplot(DateSplit,aes(Region))+geom\_bar(fill="Green",width = 0.5)+

ggtitle("Region vs No of Customers")+

xlab("Region")+ylab("No of Customers")+

scale\_y\_continuous(breaks = seq.int(0,2250,250))+

theme(plot.title = element\_text(colour = "Red",size = 20),

axis.title.x = element\_text(colour = "Violet",size = 15),

axis.title.y = element\_text(colour = "Blue",size = 15),

axis.text.x = element\_text(size = 10),

axis.text.y = element\_text(size = 10))

#summary of Region

b = summary(DateSplit$Region)

b = as.data.frame(b)

View(b)

colnames(b) = "No of Customers"

#write values to excel file

write.xlsx(b,"UK-Customer-Analysis.xlsx",

sheetName = "Region vs No of Customers",append = TRUE)

#write summary of month vs region

d = table(DateSplit$Month,DateSplit$Region)

View(d)

d = as.data.frame(d)

e = spread(d,key = Var2,value = Freq)

colnames(e)[1] = "Month"

colnames(e)

View(e)

e$Total = rowSums(e[2:5])

r = c("Jan",colSums(e[2:5]))

levels(e) = c(month.abb,"Total")

y = rbind(e,r)

View(y)

colnames(r)

View(r)

write.xlsx(y,"UK-Customer-Analysis.xlsx",

sheetName = "Month vs Region",append = TRUE)

#Plot customer joined vs month by region

ggplot(DateSplit,aes(Month))+geom\_bar(aes(fill=Region))+facet\_grid(Region~.,scales = "free\_y")+

ggtitle("Region wise customers joining details")+

xlab("Month")+ylab("No of Customers")+

theme(plot.title = element\_text(colour = "Red",size = 20),

axis.title.x = element\_text(colour = "DarkBlue",size = 15),

axis.title.y = element\_text(colour = "DarkGreen",size = 15))

#Region wise gender details

g = table(DateSplit$Region,DateSplit$Gender,DateSplit$JobClassification)

colnames(g) = c("Region","Gender","JobClassification","No of Customers")

View(g)

g = as.data.frame(g)

g = arrange(g,Region,Gender,JobClassification)

#write region wise gender details to excel

write.xlsx(g[,1:4],"UK-Customer-Analysis.xlsx",

sheetName = "Region wise Gender and Job statistics",append = TRUE,

row.names = F )

#Plot Gender against Regions

ggplot(DateSplit,aes(Gender,fill=Gender))+

geom\_bar(width = 0.5)+

facet\_grid(Region~JobClassification)+

ggtitle("Gender vs Job Classification")+

xlab("Gender")+ylab("No of Customers")+

theme(plot.title = element\_text(colour = "Blue",size = 20),

axis.title.x = element\_text(colour = "Green",size = 15),

axis.title.y = element\_text(colour = "Red",size = 15))

#Summary of Age

summary(DateSplit$Age)

#Histogram plot for Age

ggplot(DateSplit,aes(Age,fill=Gender))+geom\_histogram(binwidth = 3)+

scale\_x\_continuous(breaks = seq.int(10,70,5))+

scale\_y\_continuous(breaks = seq.int(0,400,100))+facet\_grid(Region~Gender)+

ggtitle("Age vs Gender")+ylab("No of Customers")+

theme(plot.title = element\_text(colour = "DarkBlue",size = 20))

#Histogram Plot for Balance

ggplot(DateSplit,aes(Balance,fill=Gender))+geom\_histogram(binwidth = 5000)+

scale\_x\_continuous(breaks = seq.int(0,200000,20000))+

scale\_y\_continuous(breaks = seq.int(0,325,25))+facet\_grid(Region~Gender)+

ggtitle("Balance range vs No of Customers")+

xlab("Dollars")+ylab("No of Customers")+

theme(plot.title = element\_text(colour = "Black",size = 20),

axis.title.x = element\_text(colour = "Blue",size=15),

axis.title.y = element\_text(colour = "Blue",size = 15),

axis.text.x = element\_text(colour = "Red",size = 10),

axis.text.y =element\_text(colour = "Red",size = 10))

#Histogram plot for Jobclassification

ggplot(DateSplit,aes(Balance,fill=JobClassification))+geom\_histogram(binwidth = 5000)+

scale\_x\_continuous(breaks = seq.int(0,200000,30000))+

scale\_y\_continuous(breaks = seq.int(0,325,25))+facet\_grid(Region~JobClassification)+

ggtitle("Balance Range and Job Details")+

xlab("Dollars")+ylab("No of Customers")+

theme(plot.title = element\_text(colour = "Black",size = 20),

axis.title.x = element\_text(colour = "Blue",size=15),

axis.title.y = element\_text(colour = "Blue",size = 15),

axis.text.x = element\_text(colour = "Red",size = 10),

axis.text.y =element\_text(colour = "Red",size = 10))

#Subset Male date

Male = filter(DateSplit,Gender == "Male")

View(Male)

#Plot Balance and Job details for male data

ggplot(Male,aes(Balance,fill=JobClassification))+geom\_histogram(binwidth = 5000)+

scale\_x\_continuous(breaks = seq.int(0,200000,30000))+

scale\_y\_continuous(breaks = seq.int(0,325,25))+facet\_grid(Region~JobClassification)+

ggtitle("Balance Range and Job Details(Male)")+

xlab("Dollars")+ylab("No of Customers")+

theme(plot.title = element\_text(colour = "Green",size = 40),

axis.title.x = element\_text(colour = "Violet",size=15),

axis.title.y = element\_text(colour = "Blue",size = 15),

axis.text.x = element\_text(colour = "Red",size = 10),

axis.text.y =element\_text(colour = "Red",size = 10))

#Subset Female data

Female = filter(DateSplit,Gender=="Female")

View(Female)

#Plot Balance and Job details for female data

ggplot(Female,aes(Balance,fill=JobClassification))+geom\_histogram(binwidth = 5000)+

scale\_x\_continuous(breaks = seq.int(0,200000,30000))+

scale\_y\_continuous(breaks = seq.int(0,325,25))+facet\_grid(Region~JobClassification)+

ggtitle("Balance Range and Job Details(Female)")+

xlab("Dollars")+ylab("No of Customers")+

theme(plot.title = element\_text(colour = "Green",size = 40),

axis.title.x = element\_text(colour = "Violet",size=25),

axis.title.y = element\_text(colour = "Blue",size = 25),

axis.text.x = element\_text(colour = "Red",size = 10),

axis.text.y =element\_text(colour = "Red",size = 10))

#Plot Bar chart for day variable

ggplot(DateSplit,aes(Day,fill=Gender))+geom\_bar()+facet\_wrap(~Month,nrow = 6,ncol = 2)+

ggtitle("Day of Month vs No of Customers")+

xlab("Day of Month")+ylab("No of Customers")+

theme(plot.title = element\_text(colour = "Green",size = 40),

axis.title.x = element\_text(colour = "Violet",size=25),

axis.title.y = element\_text(colour = "Blue",size = 25),

axis.text.x = element\_text(colour = "Red",size = 10),

axis.text.y =element\_text(colour = "Red",size = 10))

#Filter top 5 wealthy customers

for (row in 1:nrow(DateSplit)) {

if(DateSplit$Gender[row] == "Male"){

DateSplit$CustomerName[row] = paste("Mr.",DateSplit$Name[row],DateSplit$Surname[row])

}else if(DateSplit$Gender[row] == "Female"){

DateSplit$CustomerName[row] = paste("Ms.",DateSplit$Name[row],DateSplit$Surname[row])

}

}

DateSplit = DateSplit[,c(1:3,12,4:11)]

Comb = unite(DateSplit,col = "Date",Year,Month,Day,sep = "-")

Topfive = Comb %>% arrange(desc(Balance)) %>% top\_n(n=5) %>% select(c(1,4,6:10))

#write to excel file

write.xlsx(Topfive,"UK-Customer-Analysis.xlsx",sheetName = "Top five Customers",append = T)

#Region wise top 5 customers

Wealth = Comb %>% arrange(desc(Balance)) %>%

group\_by(Region) %>% top\_n(n = 5) %>%

select(c(1,4,6:10))

View(Wealth)

Eng = filter(Wealth,Region == "England")

View(Eng)

length(Eng)

nrow(Eng)

attributes(Eng)$class = c("data.frame")

write.xlsx(Eng,"UK-Customer-Analysis.xlsx",sheetName = "England",

append = TRUE,row.names = FALSE)

North = filter(Wealth,Region == "Northern Ireland")

attributes(North)$class = c("data.frame")

write.xlsx(North,"UK-Customer-Analysis.xlsx",sheetName = "Northern Ireland",

append = TRUE,row.names = FALSE)

#group based on Gender

anal = Comb %>% arrange(desc(Balance)) %>%

group\_by(Gender) %>% top\_n(n = 5) %>%

select(c(1,4:10))

View(anal)

View(Comb)

F = filter(anal,Gender=="Female")

View(F)

attributes(F)$class = c("data.frame")

write.xlsx(F,"UK-Customer-Analysis.xlsx",sheetName = "Female",

append = TRUE,row.names = FALSE)