**AIR- BUSINESS ANALYTICS INTERNSHIP**

**Name : Hari Krishna P**

**University : Vellore Institute Of Technology - AP**

**Libraries Used:**

library(tidyr)

library("dplyr")

library(stringr)

library(ggplot2)

library(arules)

library(arulesViz)

library(datasets)

library(RColorBrewer)

**Analyzing Data And Cleaning it (DATA-PREProcessing)**

dim(BreadBasket\_DMS)

Clean\_Data <- drop\_na(BreadBasket\_DMS)

head(Clean\_Data)

str(Clean\_Data)

summary(Clean\_Data)

#Clean\_Data$Date <- as.character(Clean\_Data$Date)

#Clean\_Data$Time <- as.character(Clean\_Data$Time)

#Clean\_Data$Transaction <- as.character(Clean\_Data$Transaction)

#Clean\_Data$Item <- as.character(Clean\_Data$Item)

#Clean\_Data %>%

# mutate\_if(is.character, str\_trim) -> Clean\_Data

# Clean\_Data$Date <- as.factor(Clean\_Data$Date)

# Clean\_Data$Time <- as.factor(Clean\_Data$Time)

# Clean\_Data$Transaction <- as.factor(Clean\_Data$Transaction)

# Clean\_Data$Item <- as.factor(Clean\_Data$Item)

sum(is.na(Clean\_Data))

mean(Clean\_Data$Transaction)

mean(Clean\_Data$Item)

mean(Clean\_Data$Time)

mean(Clean\_Data$Date)

median(Clean\_Data$Transaction)

median(Clean\_Data$Item)

median(Clean\_Data$Time)

sd(Clean\_Data$Transaction)

mode(Clean\_Data$Date)

mode(Clean\_Data$Time)

mode(Clean\_Data$Transaction)

mode(Clean\_Data$Item)

**Pie Plot For The Whole DATA SET**

ggplot(data = BreadBasket\_DMS,mapping = aes(x=Item,y=..prop..,fill=cut))+

geom\_bar(width = 1)+coord\_polar(theta = "x")

**#Finding top 2**

max(Clean\_Data$Item)

max(Clean\_Data$Item[Clean\_Data$Item != max(Clean\_Data$Item)])

**#Comparing**

newdata <- Clean\_Data[Clean\_Data$Item =="Victorian Sponge",, drop=FALSE]

View(newdata)

newdata1 <- Clean\_Data[Clean\_Data$Item =="Vegan mincepie",, drop=FALSE]

View(newdata1)

**#PLOTTING SALES OF TOP 2**

barplot(height = newdata$Transaction,xlab = " Victorian Sponge ",main = "Sales Of Victorian Sponge",col = 'red')

Chart, bar chart

Description automatically generated

barplot(height = newdata1$Transaction,xlab = "Vegan mincepie",main = "Sales Of Vegan mincepie",col = 'green')

Chart, histogram

Description automatically generated

Here According to my knowledge afternoon time is the most sales presenting and at this time we can combine Victorian Sponge and Vegan mincepie so sales could be more

# Q2; Here I take the top 50 Transactions first

top50 = as.data.frame(head(Clean\_Data[order(Clean\_Data$Item, decreasing = TRUE), 1:4], n = 50))

top50

**Histogram**

hist(top50$Transaction,col = rainbow(top50$Transaction))

Chart, histogram

Description automatically generated

barplot(top50$Transaction,

main = "Each Item",

xlab = "Class",

col = c("red","green")

)

legend("topleft",

c("Victorian Sponge","Vegan mincepie"),

fill = c("red","green")

)

Chart, histogram

Description automatically generated

#TOP 10 Products Top 100 sales According to transaction

top100 = as.data.frame(head(Clean\_Data[order(Clean\_Data$Item, decreasing = TRUE), 1:4], n = 100))

view(top100)

Plotting Top 100 sales Item

ggplot(top100, aes(x = Item ,y = Transaction,fill = "red")) +

geom\_col() + coord\_flip() + xlab("Items")

Chart, bar chart, funnel chart

Description automatically generated

Q3: find the least sold product and remove It

r = min(Clean\_Data$Item[Clean\_Data$Item != min(Clean\_Data$Item)])

r

q3 <- Clean\_Data[which(Clean\_Data$Item=='Afternoon with the baker'),]

View(q3)

q3 <- q3[c(-4)] #removes the item