Meegrow Case Study – Online Retail Data

Please find attached following R code for the analysis.

#Invoking Library

```{r}
library(ggplot2)
library(coefplot)
library(corrplot)
library(psych)
library(tidyverse)
library(MASS)
library(lattice)
library(DataExplorer)
library(ModelMetrics)
library(lmtest)
library(caret)
library(recipes)
library(Amelia)
library(caTools)
library(lubridate)
library(dplyr)
library(readxl)
```

```{r}
#Importing Dataset
setwd("P:/Meegrow_CaseStudy")
getwd()

library(readr)
Casedata <- read.csv("Online Retail_Case data.csv")
str(Casedata)
```

summary(Casedata)

dim(Casedata)

names(Casedata)

attach(Casedata)

```

```{r}

#Check for missing values and treating them

anyNA(Casedata)

plot_missing(Casedata)

missmap(Casedata)

sum(is.na(Casedata))

```

```{r}

Data_omit <- na.omit(Casedata)

Data_omit

```

```{r}

#Reducing redundant variable

Casedata.New = Data_omit[,-c(2)]


```

```{r}

# Univariate Analysis of numeric variables

colnames(Casedata.New[,sapply(Casedata.New, is.numeric)])

ggplot(Casedata.New, aes(x = UnitPrice)) + geom_histogram(bins = 30, fill = "darkgreen", col = "lightgreen")

ggplot(Casedata.New, aes(x = Quantity)) + geom_histogram(bins = 30, fill = "orange", col = "lightgreen")

ggplot(Casedata.New, aes(x = Final.Sale)) + geom_histogram(bins = 30, fill = "turquoise", col = "blue")


# Univariate analysis of caterogical variables

```
ggplot(Casedata.New, aes(x = Country, fill = Country)) + geom_bar()

#Bivariate Analysis

ggplot(Casedata.New, aes(x = UnitPrice, y = Quantity)) + geom_point()

ggplot(Casedata.New, aes(x = Final.Sale, y = Quantity)) + geom_line()
```

```{r}
#RFM Analysis of the data

library(rfm)

data = Casedata.New

str(data)

summary(data)


analysis.date = lubridate::as_date('2011-12-09', tz = 'UTC')

result = rfm_table_customer(Casedata.New, Description, Quantity, UnitPrice, Final.Sale,
analysis.date)

result



#Visualization

rfm_heatmap(result)

rfm_bar_chart(result)

rfm_histograms(result)

rfm_order_dist(result)

rfm_rm_plot(result)

rfm_fm_plot(result)

rfm_rf_plot(result)
```