```
###############################################################################
#Analysis of Sales Report of a clothes Manufacturing Outlet . We will find current
#trends,and attributes affecting sales.
#*********************************SECTION 1***************************#
#In order to automate the process of recommendations, the store needs to analyze
#the given attributes of the product like style, season etc. and come up with a
#model to predict the recommendation of products in binary output(0 or 1)
#*********************************SECTION 1***************************#
#Created by : SHAHROO AKHTAR
###############################################################################

#Installing packages and libraries importing data.

library(dplyr)
library(readxl)

###############################################################################

#Importing the data file from the saved location

Data <- read.csv("/Users/apple/Desktop/DressAttributes.csv", header=T)
#Preparing the data for analysis
head(Data)
#Following inferences are noted by looking at the data :
#***Dependent Variable: Recommendation
#***Independent Variables: Style, Price, Season, SleeveLength, Material, Fabric
#***                       Type,Decoration , waiseline, NeckLine and Pattern

###############################################################################

#Getting all the column names
colnames(Data)
#change the column name "Pattern Type" to Pattern
colnames(Data)[13] <- "Pattern"

###############################################################################

#Removing missing values
Data[Data ==" "] <- NA
#Replacing Null with NA
Data[Data =="null"] <- NA
View(Data)

###############################################################################

#check for the count of missing values in each columns
colSums(is.na(Data))
#It shows that FabricType= 265, Decoration= 235, Material= 127, Pattern= 108
#We will only use the columns with no missing values for our analysis
MyData <- select(Data,Style, Price, Season,Size,NeckLine,SleeveLength, Recommendation)

###############################################################################

#Lets view the structure of the data
str(MyData)
head(MyData)
#Converting the catagorical data into factors for analysis
newdata <- mutate_if(MyData,is.character,as.factor)
str(newdata)
#Lets check the no. of recommendations
table(newdata$Recommendation)
#It shows that 210 product were recommended and 290 were not recommended.

###############################################################################

#*********************************APPLYING THE MODEL***********************#

#Lets apply the Logit regression model
mymodel <- glm(Recommendation~.,family = binomial(link = 'logit'),
```

```
            data = newdata)
#Predict the Model
Pmymodel <- predict(mymodel,newdata)

#***************************APPLYING CONFUSION MATRIX*********************#

tab <- table(Pmymodel > 0.5, newdata$Recommendation)
tab
#Checking the predicted accuracy
sum(diag(tab))/sum(tab)*100
#The Predicted accuracy is 69.6

#Now lets check the actual accuracy
table(newdata$Recommendation)
290/(290+210)*100
#The actual accuracy is 58.0

#RESULT: Since our Predicted Accuracy is higher that is 69.6 than Actual
#accuracy we conclude that our Prediction Model is better.

##########################################################################
```