**Walmart Sales Prediction with R Programming**

Project by Pravat Kumar Sahu

Mentor: Deepti Gupta

Simplilearn Batch: PGP DA FEB 2021 Cohort 1

May 13, 2021

**R PROGAMMING CODES**

#Walmart Sales Prediction with R Programming

**#Introduction**

#Problem: To predict sales

#Solution: To Build Multiple linear regression model

**#Step 1: Identify dependent and Independent variables**

#Dependent Variable: Weekly Sales

#Independent variable: All other variables

**#Step2: Data Understanding**

#There are 6435 observations and 8 variables in the data. The variables are Store number, Date, Weekly Sales, Holiday flag which are marked as 0 if date falling on non-holiday and 1 if the date falling on holiday, temperature for week, fuel price for the week, prevailing consumer price index for the week, and lastly the unemployment rate for the week.

**#Step3: Import data to Lab**

walmart\_data <- read.csv("Walmart\_Store\_sales.csv", header= TRUE,sep=",")

View(walmart\_data)

#Step4: Data exploration

str(walmart\_data)

head(walmart\_data)

tail(walmart\_data)

nrow(walmart\_data)

ncol(walmart\_data)

dim(walmart\_data)

class(walmart\_data)

table(walmart\_data$Store)

summary(walmart\_data)

#To change the data variable to Date Format from Character Format.

walmart\_data$Date <- as.Date(walmart\_data$Date, format = "%d-%m-%Y")

str(walmart\_data)

**#Step5: Dividing data into Training (70%) and Testing (30%) ratio.**

library(caTools) #load package

#set the seed to freeze the sample

set.seed(1)

sample <- sample.split(walmart\_data$Weekly\_Sales, SplitRatio=0.7)

sample

train\_data <- subset(walmart\_data, sample== TRUE )

test\_data<- subset(walmart\_data,sample==FALSE)

**#Step6: Build the Model using Training data**

model <- lm(Weekly\_Sales ~ ., data= train\_data)

summary(model)

**#Step 7: Look at the Coefficient Table to determine the significant variables.**

**Text

Description automatically generated with medium confidence**

**#Step 8: Rerun the model using significant independent variables:**

model1 <- lm(Weekly\_Sales ~ Store + Holiday\_Flag + Temperature + CPI + Unemployment , data= train\_data)

summary(model1)

A picture containing text, receipt, screenshot

Description automatically generated

**#Step 9: Rerun the model using significant independent variables:**

model2 <- lm(Weekly\_Sales ~ Store + Holiday\_Flag + CPI + Unemployment , data= train\_data)

summary(model2)

Text

Description automatically generated

**#Step 10: prediction on testing data set**

predtest<- predict(model2,test\_data)

predtest

#attach it with the dataframe

predtest1<-data.frame(predtest)

#to bind the predicted data set with test data set by cbind function

final\_data<- cbind(test\_data,predtest1)

View(final\_data)

**#cal rmse**

sqrt(mean((final\_data$Weekly\_Sales - final\_data$predtest)^2))

Graphical user interface, text, application, email

Description automatically generated

write.csv(final\_data,"walmart\_sale\_prediction\_output.csv")

View(walmart\_sale\_prediction\_output)

sales\_prediction\_10 <- head(walmart\_sale\_prediction\_output,10)

View(sales\_prediction\_10)

Table

Description automatically generated