```
#Libraries Loading
library(data.table)
library(ggplot2)
library(tidyr)

#Data Loading
filepath = "~/Data Analysis Projects/"
Data <-  fread(paste0(filepath,"QVI_data.csv"))

View(Data)

#Year-Month Column
Data[,YEARMONTH := format(Data$DATE, '%Y%m')]
str(Data)

#Measure to Calculate total sales, No. of Customers, Transactions per Customer,
#Chips per Customer and Average price unit monthly for each store
measureovertime <- Data[, .(totSales = sum(TOT_SALES),
                  nCustomers = uniqueN(LYLTY_CARD_NBR),
                  nTxnPerCUst = uniqueN(TXN_ID)/uniqueN(LYLTY_CARD_NBR),
                  nChipsPerTxn = sum(PROD_QTY)/uniqueN(TXN_ID),
                  avgPricePerUnit = sum(TOT_SALES)/sum(PROD_QTY)
                 ), by = .(STORE_NBR,YEARMONTH)
                 ][order(STORE_NBR,YEARMONTH)]
View(measureovertime)
measureovertime[,YEARMONTH := as.numeric(YEARMONTH)]

#Identify the stores with full observations(12 months) and extract the pre trial records
storewithFullObs <- unique(measureovertime[,.N,STORE_NBR][N == 12, STORE_NBR])
preTrialMeasures <- measureovertime[YEARMONTH < 201902 & STORE_NBR %in%
storewithFullObs,]
View(preTrialMeasures)

##Methods to Identify the suitable control store for each trial store
#Trend Analysis
#Function to assess the correlation performance between trial and control stores
calculateCorrelation <- function(inputTable,metriCol,storeComparison) {
 calCorrTable <- data.table(Store1 = numeric(), Store2 = numeric(), corr_measure =
numeric())
 storenumbers <- unique(inputTable$STORE_NBR)
 for (i in storenumbers) {
  dataStore1 <- inputTable[STORE_NBR == storeComparison, get(metriCol)]
  dataStore2 <- inputTable[STORE_NBR == i, get(metriCol)]
  correlation_value = cor(dataStore1,dataStore2)
```

```
   calMeasureTable <- data.table(
    "Store1" = storeComparison,
    "Store2" = i,
    "corr_measure" = correlation_value)
   calCorrTable <- rbind(calCorrTable,calMeasureTable)
  }
  return(calCorrTable)
}

#Similarity Analysis
#Function to calculate the absolute difference between trial and control stores
calculateMagnitude <- function(inputTable,metriCol,storeComparison) {
  calDistTable <- data.table(Store1 = numeric(),Store2 = numeric(), YEARMONTH =
numeric(),measure=numeric())
  storenumbers <- unique(inputTable$STORE_NBR)
  for (i in storenumbers) {
   calculateMeasure <- data.table(
    "Store1" = storeComparison,
    "Store2" = i,
    "YEARMONTH" = inputTable[STORE_NBR == storeComparison, YEARMONTH],
    "measure" = abs(inputTable[STORE_NBR == storeComparison, eval(metriCol)] -
     inputTable[STORE_NBR == i, eval(metriCol)])
   )
   calDistTable <- rbind(calDistTable,calculateMeasure)
  }

  minMaxDist <- calDistTable[, .(minDist = min(measure), maxDist = max(measure)),
                   by = c("Store1", "YEARMONTH")]
  distTable <- merge(calDistTable, minMaxDist, by = c("Store1", "YEARMONTH"))
  distTable[, magnitudeMeasure := 1 - (measure - minDist)/(maxDist - minDist)]

  finalDistTable <- distTable[, .(mag_measure = mean(magnitudeMeasure)), by =
                    .(Store1, Store2)]
  return(finalDistTable)
}

####Trial Store - 77
#Correlation between Store 77 and Other Stores for Total Sales and No. of Customers
corr_nSales_77 <- calculateCorrelation(preTrialMeasures, quote(totSales), 77)
corr_nCustomers_77 <- calculateCorrelation(preTrialMeasures, quote(nCustomers),77)
View(corr_nSales_77)
View(corr_nCustomers_77)
```

```
#Magnitude Distance between Store 77 and Other Stores for Total Sales and No. of
Customers
magnitude_nSales_77 <- calculateMagnitude(preTrialMeasures, quote(totSales), 77)
magnitude_nCustomers_77 <- calculateMagnitude(preTrialMeasures, quote(nCustomers),
77)
View(magnitude_nSales_77)
View(magnitude_nCustomers_77)

corr_weight <- 0.5
score_nSales_77 <- merge(corr_nSales_77,magnitude_nSales_77, by =
c("Store1","Store2"))
score_nSales_77[,scoreNSales := corr_weight * corr_measure + (1-corr_weight) *
mag_measure]
score_nCustomers_77 <- merge(corr_nCustomers_77,magnitude_nCustomers_77,by =
c("Store1","Store2"))
score_nCustomers_77[,scoreNCustomers :=  corr_weight * corr_measure + (1-corr_weight)
* mag_measure]
score_Control_77 <- merge(score_nCustomers_77,score_nSales_77,
by=c("Store1","Store2"))
score_Control_77[,finalcontrolscore := 0.5 * scoreNSales + 0.5 * scoreNCustomers]
control_store_77 <- score_Control_77[Store1 != Store2][order (-
finalcontrolscore)][1,Store2]
control_store_77
#Store 233 with score: 0.9680425 is the most suitable to
#consider as the Control Store for Trial Store 77

#Visual Comparison of Sales in between Store 77, Store 233(Control Store) and Other
Stores
measureOverTimeSales <- data.table::copy(measureovertime)
pastSales <- measureOverTimeSales[, Store_type := ifelse(STORE_NBR == 77,
                                    "Trial",
                                    ifelse(STORE_NBR == control_store_77,
                                        "Control", "Other stores"))
][, totSales := mean(totSales), by = c("YEARMONTH",
                          "Store_type")
][, TransactionMonth := as.Date(paste(YEARMONTH %/%
                           100, YEARMONTH %% 100, 1, sep = "-"), "%Y-%m-%d")
][YEARMONTH < 201902 , ]

View
ggplot(pastSales, aes(TransactionMonth, totSales, color = Store_type)) +
 geom_line() +
 labs(x = "Month of operation", y = "Total sales", title = "Total sales by month")
```

```
#Visual Comparison of Total Customers in between Store 77, Store 233(Control Store) and Other Stores
measureOverTimeCustomers <-  data.table::copy(measureovertime)
pastCustomers <- measureOverTimeCustomers [,Store_type:= ifelse(STORE_NBR == 77, "Trial",
                                        ifelse(STORE_NBR == control_store_77,"Control","Other Stores"))
][, nCustomers := mean(nCustomers), by = c("YEARMONTH","Store_type")
  ][,TransactionMonth := as.Date(paste(YEARMONTH%/%100,YEARMONTH %% 100, 1, sep = "-"), "%Y-%m-%d")
  ][YEARMONTH < 201902, ]

ggplot(pastCustomers, aes (TransactionMonth, nCustomers, color = Store_type)) +
  geom_line() +
  labs(x = "Month of operation", y = "Total Customers", title = "Total Customers by month")

#Trial Period Assessment
#Hypothesis Testing - Total Sales
#H0: There is no significant difference in Total sales during the trial period
#H1: There is significant difference in Total sales during the trial period
#Scaling
scalingFactorforControlSales_77 <- preTrialMeasures[STORE_NBR == 77 & YEARMONTH < 201902,sum(totSales)]/
  preTrialMeasures[STORE_NBR == control_store_77 & YEARMONTH < 201902,sum(totSales)]
measureOverTimeSales <-  data.table::copy(measureovertime)
SclaedControlSales_77 <- measureOverTimeSales[STORE_NBR == control_store_77,
                                      ][,ControlSales := totSales * scalingFactorforControlSales_77]
trialSales_77 <- measureovertime[STORE_NBR == 77, .(YEARMONTH, totSales)]
scaledControlSales <- SclaedControlSales_77[, .(YEARMONTH, ControlSales)]

percentageDiff_77 <- merge(trialSales_77,scaledControlSales, by = "YEARMONTH")
percentageDiff_77[,percentageDiff := (totSales - ControlSales)/ControlSales]
View(percentageDiff_77)

stdDev <- sd(percentageDiff_77[YEARMONTH < 201902 , percentageDiff])
View(stdDev)

degreeoffreedom <- 7 #Length of the pre trial period is 8 months. degree = len(df) - 1

percentageDiff_77[YEARMONTH >= 201902 & YEARMONTH <= 201904,
          tvalue := (percentageDiff - 0)/stdDev]
criticalTValue_77 <- qt(0.95,df=degreeoffreedom)
```

```
View(percentageDiff_77)
View(criticalTValue_77)
#The Critical Value is 1.89 and T values for trial months is -0.5,3.6,6.2
#With higher t values it can be statistically proven that the trial has increased the sales
significantly

#Visual Representation of the Total sales by month
View(measureOverTimeSales)
measureOverTimeSales <- data.table::copy(measureovertime)
pastSales <- measureOverTimeSales[, Store_type := ifelse(STORE_NBR == 77, "Trial",
                                    ifelse(STORE_NBR == control_store_77, "Control", "Other"))
                    ][, totSales := totSales
                    ][, TransactionMonth := as.Date(paste(YEARMONTH%/%100,YEARMONTH
%% 100, 1, sep = "-"), "%Y-%m-%d")
                    ][Store_type %in% c("Trial", "Control"), ]


pastSales_Controls95 <- pastSales[Store_type == "Control",
                    ][, totSales := totSales * (1 + stdDev * 2)
                    ][, Store_type := "Control 95th % confidence interval"]

pastSales_Controls05 <- pastSales[Store_type == "Control",
                    ][, totSales := totSales * (1 - stdDev * 2)
                    ][, Store_type := "Control 5th % confidence interval"]

trialAssessment <- rbind(pastSales, pastSales_Controls95, pastSales_Controls05)

View(trialAssessment)

ggplot(trialAssessment, aes(TransactionMonth, totSales, color = Store_type)) +
 geom_rect(data = trialAssessment[ YEARMONTH >= 201902 & YEARMONTH <= 201904,],
       aes(xmin = min(TransactionMonth), xmax = max(TransactionMonth), ymin = 0 , ymax =
           Inf, color = NULL), show.legend = FALSE) +
 geom_line() +
 labs(x = "Month of operation", y = "Total sales", title = "Total sales by month")

#Hypothesis Testing - No. Of Customers
#H0: There is no significant difference in Total customers during the trial period
#H1: There is significant difference in Total customers during the trial period
#Scaling
scalingFactorforControlCustomers_77 <- preTrialMeasures[STORE_NBR == 77 &
YEARMONTH < 201902,sum(nCustomers)]/
 preTrialMeasures[STORE_NBR == control_store_77 & YEARMONTH <
201902,sum(nCustomers)]
```

```
measureOverTimeCustomers <- data.table::copy(measureovertime)
scaledControlCustomers_77 <- measureOverTimeCustomers[STORE_NBR ==
control_store_77,][,ScaledCustomers := nCustomers
*scalingFactorforControlCustomers_77]
trialCustomers_77 <- measureovertime[STORE_NBR == 77, .(YEARMONTH, nCustomers)]
scaledControlCustomers_s77 <- scaledControlCustomers_77[, .(YEARMONTH,
ScaledCustomers)]

custPercentageDiff_77 <- merge(trialCustomers_77,scaledControlCustomers_s77, by =
"YEARMONTH")
custPercentageDiff_77[,percentageDiff := (nCustomers -
ScaledCustomers)/ScaledCustomers]
View(custPercentageDiff_77)

stdDev <- sd(custPercentageDiff_77[YEARMONTH < 201902 , percentageDiff])
View(stdDev)

custPercentageDiff_77[YEARMONTH >= 201902 & YEARMONTH <= 201904,
          tvalue := (percentageDiff - 0)/stdDev]

View(custPercentageDiff_77)
View(criticalTValue_77)

#The Critical Value is 1.89 and T values for trial months is -0.12,8.9,20.4
#With higher t values it can be statistically proven that the trial has increased the No. Of
Customers significantly

#Visual Representation of the Total Customers by month
measureOverTimeCustomers <- data.table::copy(measureovertime)
pastCustomers <- measureOverTimeCustomers[, Store_type := ifelse(STORE_NBR == 77,
"Trial",
                                    ifelse(STORE_NBR == control_store_77, "Control", "Other"))
                        ][, nCust := mean(nCustomers), by = c("YEARMONTH","Store_type")
                        ][, TransactionMonth :=
as.Date(paste(YEARMONTH%/%100,YEARMONTH %% 100, 1, sep = "-"), "%Y-%m-%d")
                        ][Store_type %in% c("Trial", "Control"), ]


pastCustomers_Controls95 <- pastCustomers[Store_type == "Control",
                        ][, nCust := nCust * (1 + stdDev * 2)
                        ][, Store_type := "Control 95th % confidence interval"]

pastCustomers_Controls05 <- pastCustomers[Store_type == "Control",
                        ][, nCust := nCust * (1 - stdDev * 2)
```

```
][, Store_type := "Control 5th % confidence interval"]

trialAssessment_cust <- rbind(pastCustomers, pastCustomers_Controls95,
pastCustomers_Controls05)

View(trialAssessment_cust)

ggplot(trialAssessment_cust, aes(TransactionMonth, nCust, color = Store_type)) +
 geom_rect(data = trialAssessment_cust[ YEARMONTH >= 201902 & YEARMONTH <=
201904,],
        aes(xmin = min(TransactionMonth), xmax = max(TransactionMonth), ymin = 0 , ymax =
            Inf, color = NULL), show.legend = FALSE) +
 geom_line() +
 labs(x = "Month of operation", y = "Total Customers", title = "Total Customers by month")

####Trial Store - 86
#Correlation between Store 86 and Other Stores for Total Sales and No. of Customers
View(preTrialMeasures)
corr_nSales_86 <- calculateCorrelation(preTrialMeasures, quote(totSales), 86)
corr_nCustomers_86 <- calculateCorrelation(preTrialMeasures, quote(nCustomers),86)
View(corr_nSales_86)
View(corr_nCustomers_86)

#Magnitude Distance between Store 86 and Other Stores for Total Sales and No. of
Customers
magnitude_nSales_86 <- calculateMagnitude(preTrialMeasures, quote(totSales), 86)
magnitude_nCustomers_86 <- calculateMagnitude(preTrialMeasures, quote(nCustomers),
86)
View(magnitude_nSales_86)
View(magnitude_nCustomers_86)

corr_weight <- 0.5
score_nSales_86 <- merge(corr_nSales_86,magnitude_nSales_86, by =
c("Store1","Store2"))
score_nSales_86[,scoreNSales := corr_weight * corr_measure + (1-corr_weight) *
mag_measure]
score_nCustomers_86 <- merge(corr_nCustomers_86,magnitude_nCustomers_86,by =
c("Store1","Store2"))
score_nCustomers_86[,scoreNCustomers :=  corr_weight * corr_measure + (1-corr_weight)
* mag_measure]
score_Control_86 <- merge(score_nCustomers_86,score_nSales_86,
by=c("Store1","Store2"))
score_Control_86[,finalcontrolscore := 0.5 * scoreNSales + 0.5 * scoreNCustomers]
```

```
control_store_86 <- score_Control_86[Store1 != Store2][order (-
finalcontrolscore)][1,Store2]
control_store_86
View(score_Control_86)
#Store 155 with score: 0.9421896 is the most suitable to
#consider as the Control Store for Trial Store 86

#Visual Comparison of Sales in between Store 86, Store 155(Control Store) and Other
Stores
measureOverTimeSales <- data.table::copy(measureovertime)
pastSales <- measureOverTimeSales[, Store_type := ifelse(STORE_NBR == 86,
                                    "Trial",
                                    ifelse(STORE_NBR == control_store_86,
                                        "Control", "Other stores"))
][, totSales := mean(totSales), by = c("YEARMONTH",
                          "Store_type")
][, TransactionMonth := as.Date(paste(YEARMONTH %/%
                          100, YEARMONTH %% 100, 1, sep = "-"), "%Y-%m-%d")
][YEARMONTH < 201902 , ]

ggplot(pastSales, aes(TransactionMonth, totSales, color = Store_type)) +
 geom_line() +
 labs(x = "Month of operation", y = "Total sales", title = "Total sales by month")

#Visual Comparison of Total Customers in between Store 86, Store 155(Control Store) and
Other Stores
measureOverTimeCustomers <-  data.table::copy(measureovertime)
pastCustomers <- measureOverTimeCustomers [,Store_type:= ifelse(STORE_NBR == 86,
"Trial",
                                    ifelse(STORE_NBR == control_store_86,"Control","Other
Stores"))
                          ][, nCustomers := mean(nCustomers), by =
c("YEARMONTH","Store_type")
                          ][,TransactionMonth :=
as.Date(paste(YEARMONTH%/%100,YEARMONTH %% 100, 1, sep = "-"), "%Y-%m-%d")
                          ][YEARMONTH < 201902, ]

ggplot(pastCustomers, aes (TransactionMonth, nCustomers, color = Store_type)) +
 geom_line() +
 labs(x = "Month of operation", y = "Total Customers", title = "Total Customers by month")

#Trial Period Assessment
#Hypothesis Testing - Total Sales
#H0: There is no significant difference in Total sales during the trial period
```

```
#H1: There is significant difference in Total sales during the trial period
#Scaling
scalingFactorforControlSales_86 <- preTrialMeasures[STORE_NBR == 86 & YEARMONTH <
201902,sum(totSales)]/
  preTrialMeasures[STORE_NBR == control_store_86 & YEARMONTH <
201902,sum(totSales)]
measureOverTimeSales <-  data.table::copy(measureovertime)
SclaedControlSales_86 <- measureOverTimeSales[STORE_NBR == control_store_86,
][,ControlSales := totSales * scalingFactorforControlSales_86]
trialSales_86 <- measureovertime[STORE_NBR == 86, .(YEARMONTH, totSales)]
scaledControlSales <- SclaedControlSales_86[, .(YEARMONTH, ControlSales)]

percentageDiff_86 <- merge(trialSales_86,scaledControlSales, by = "YEARMONTH")
percentageDiff_86[,percentageDiff := (totSales - ControlSales)/ControlSales]
View(percentageDiff_86)

stdDev <- sd(percentageDiff_86[YEARMONTH < 201902 , percentageDiff])
View(stdDev)

degreeoffreedom <- 7 #Length of the pre trial period is 8 months. degree = len(df) - 1

percentageDiff_86[YEARMONTH >= 201902 & YEARMONTH <= 201904,
            tvalue := (percentageDiff - 0)/stdDev]
criticalTValue_86 <- qt(0.95,df=degreeoffreedom)

View(percentageDiff_86)
View(criticalTValue_86)
#The Critical Value is 1.89 and T values for trial months is 1.4,8.3,0.9
#With lower t values it can be statistically proven that the trial has no impact on the sales

#Visual Representation of the Total sales by month
measureOverTimeSales <- data.table::copy(measureovertime)
pastSales <- measureOverTimeSales[, Store_type := ifelse(STORE_NBR == 86, "Trial",
                                    ifelse(STORE_NBR == control_store_86, "Control", "Other"))
                    ][, totSales := totSales
                    ][, TransactionMonth := as.Date(paste(YEARMONTH%/%100,YEARMONTH
%% 100, 1, sep = "-"), "%Y-%m-%d")
                    ][Store_type %in% c("Trial", "Control"), ]


pastSales_Controls95 <- pastSales[Store_type == "Control",
                    ][, totSales := totSales * (1 + stdDev * 2)
                    ][, Store_type := "Control 95th % confidence interval"]
```

```
pastSales_Controls05 <- pastSales[Store_type == "Control",
                    ][, totSales := totSales * (1 - stdDev * 2)
                    ][, Store_type := "Control 5th % confidence interval"]

trialAssessment <- rbind(pastSales, pastSales_Controls95, pastSales_Controls05)

View(trialAssessment)

ggplot(trialAssessment, aes(TransactionMonth, totSales, color = Store_type)) +
 geom_rect(data = trialAssessment[ YEARMONTH >= 201902 & YEARMONTH <= 201904,],
       aes(xmin = min(TransactionMonth), xmax = max(TransactionMonth), ymin = 0 , ymax =
           Inf, color = NULL), show.legend = FALSE) +
 geom_line() +
 labs(x = "Month of operation", y = "Total sales", title = "Total sales by month")

#Hypothesis Testing - No. Of Customers
#H0: There is no significant difference in Total customers during the trial period
#H1: There is significant difference in Total customers during the trial period
#Scaling
scalingFactorforControlCustomers_86 <- preTrialMeasures[STORE_NBR == 86 &
YEARMONTH < 201902,sum(nCustomers)]/
 preTrialMeasures[STORE_NBR == control_store_86 & YEARMONTH <
201902,sum(nCustomers)]
measureOverTimeCustomers <- data.table::copy(measureovertime)
scaledControlCustomers_86 <- measureOverTimeCustomers[STORE_NBR ==
control_store_86,][,ScaledCustomers := nCustomers
*scalingFactorforControlCustomers_86]
trialCustomers_86 <- measureovertime[STORE_NBR == 86, .(YEARMONTH, nCustomers)]
scaledControlCustomers_s86 <- scaledControlCustomers_86[, .(YEARMONTH,
ScaledCustomers)]

custPercentageDiff_86 <- merge(trialCustomers_86,scaledControlCustomers_s86, by =
"YEARMONTH")
custPercentageDiff_86[,percentageDiff := (nCustomers -
ScaledCustomers)/ScaledCustomers]
View(custPercentageDiff_86)

stdDev <- sd(custPercentageDiff_86[YEARMONTH < 201902 , percentageDiff])
View(stdDev)

custPercentageDiff_86[YEARMONTH >= 201902 & YEARMONTH <= 201904,
            tvalue := (percentageDiff - 0)/stdDev]

View(custPercentageDiff_86)
```

```
View(criticalTValue_86)

#The Critical Value is 1.89 and T values for trial months is 6.5,11.9,3.1
#With higher t values it can be statistically proven that the trial has increased the No. Of Customers significantly

#Visual Representation of the Total Customers by month
measureOverTimeCustomers <- data.table::copy(measureovertime)
pastCustomers <- measureOverTimeCustomers[, Store_type := ifelse(STORE_NBR == 86, "Trial",
                                        ifelse(STORE_NBR == control_store_86, "Control", "Other"))
                        ][, nCust := mean(nCustomers), by = c("YEARMONTH","Store_type")
                        ][, TransactionMonth :=
as.Date(paste(YEARMONTH%/%100,YEARMONTH %% 100, 1, sep = "-"), "%Y-%m-%d")
                        ][Store_type %in% c("Trial", "Control"), ]


pastCustomers_Controls95 <- pastCustomers[Store_type == "Control",
                        ][, nCust := nCust * (1 + stdDev * 2)
                        ][, Store_type := "Control 95th % confidence interval"]

pastCustomers_Controls05 <- pastCustomers[Store_type == "Control",
                        ][, nCust := nCust * (1 - stdDev * 2)
                        ][, Store_type := "Control 5th % confidence interval"]

trialAssessment_cust <- rbind(pastCustomers, pastCustomers_Controls95,
pastCustomers_Controls05)

View(trialAssessment_cust)

ggplot(trialAssessment_cust, aes(TransactionMonth, nCust, color = Store_type)) +
 geom_rect(data = trialAssessment_cust[ YEARMONTH >= 201902 & YEARMONTH <=
201904,],
        aes(xmin = min(TransactionMonth), xmax = max(TransactionMonth), ymin = 0 , ymax =
            Inf, color = NULL), show.legend = FALSE) +
 geom_line() +
 labs(x = "Month of operation", y = "Total Customers", title = "Total Customers by month")


####Trial Store - 88
#Correlation between Store 88 and Other Stores for Total Sales and No. of Customers
corr_nSales_88 <- calculateCorrelation(preTrialMeasures, quote(totSales), 88)
corr_nCustomers_88 <- calculateCorrelation(preTrialMeasures, quote(nCustomers),88)
View(corr_nSales_88)
```

```
View(corr_nCustomers_88)

#Magnitude Distance between Store 88 and Other Stores for Total Sales and No. of
Customers
magnitude_nSales_88 <- calculateMagnitude(preTrialMeasures, quote(totSales), 88)
magnitude_nCustomers_88 <- calculateMagnitude(preTrialMeasures, quote(nCustomers),
88)
View(magnitude_nSales_88)
View(magnitude_nCustomers_88)

corr_weight <- 0.5
score_nSales_88 <- merge(corr_nSales_88,magnitude_nSales_88, by =
c("Store1","Store2"))
score_nSales_88[,scoreNSales := corr_weight * corr_measure + (1-corr_weight) *
mag_measure]
score_nCustomers_88 <- merge(corr_nCustomers_88,magnitude_nCustomers_88,by =
c("Store1","Store2"))
score_nCustomers_88[,scoreNCustomers :=  corr_weight * corr_measure + (1-corr_weight)
* mag_measure]
score_Control_88 <- merge(score_nCustomers_88,score_nSales_88,
by=c("Store1","Store2"))
score_Control_88[,finalcontrolscore := 0.5 * scoreNSales + 0.5 * scoreNCustomers]
control_store_88 <- score_Control_88[Store1 != Store2][order (-
finalcontrolscore)][1,Store2]
control_store_88
View(score_Control_88)
#Store 237 with score: 0.7998667 is the most suitable to
#consider as the Control Store for Trial Store 88

#Visual Comparison of Sales in between Store 88, Store 237(Control Store) and Other
Stores
measureOverTimeSales <- data.table::copy(measureovertime)
pastSales <- measureOverTimeSales[, Store_type := ifelse(STORE_NBR == 88,
                                    "Trial",
                                    ifelse(STORE_NBR == control_store_88,
                                        "Control", "Other stores"))
][, totSales := mean(totSales), by = c("YEARMONTH",
                        "Store_type")
][, TransactionMonth := as.Date(paste(YEARMONTH %/%
                         100, YEARMONTH %% 100, 1, sep = "-"), "%Y-%m-%d")
][YEARMONTH < 201902 , ]

ggplot(pastSales, aes(TransactionMonth, totSales, color = Store_type)) +
  geom_line() +
```

```
 labs(x = "Month of operation", y = "Total sales", title = "Total sales by month")

#Visual Comparison of Total Customers in between Store 88, Store 237(Control Store) and Other Stores
measureOverTimeCustomers <-  data.table::copy(measureovertime)
pastCustomers <- measureOverTimeCustomers [,Store_type:= ifelse(STORE_NBR == 88, "Trial",
                                          ifelse(STORE_NBR == control_store_88,"Control","Other Stores"))
][, nCustomers := mean(nCustomers), by = c("YEARMONTH","Store_type")
][,TransactionMonth := as.Date(paste(YEARMONTH%/%100,YEARMONTH %% 100, 1, sep = "-"), "%Y-%m-%d")
][YEARMONTH < 201902, ]

ggplot(pastCustomers, aes (TransactionMonth, nCustomers, color = Store_type)) +
 geom_line() +
 labs(x = "Month of operation", y = "Total Customers", title = "Total Customers by month")

#Trial Period Assessment
#Hypothesis Testing - Total Sales
#H0: There is no significant difference in Total sales during the trial period
#H1: There is significant difference in Total sales during the trial period
#Scaling
scalingFactorforControlSales_88 <- preTrialMeasures[STORE_NBR == 88 & YEARMONTH < 201902,sum(totSales)]/
 preTrialMeasures[STORE_NBR == control_store_88 & YEARMONTH < 201902,sum(totSales)]
measureOverTimeSales <-  data.table::copy(measureovertime)
SclaedControlSales_88 <- measureOverTimeSales[STORE_NBR == control_store_88,
][,ControlSales := totSales * scalingFactorforControlSales_88]
trialSales_88 <- measureovertime[STORE_NBR == 88, .(YEARMONTH, totSales)]
scaledControlSales <- SclaedControlSales_88[, .(YEARMONTH, ControlSales)]

percentageDiff_88 <- merge(trialSales_88,scaledControlSales, by = "YEARMONTH")
percentageDiff_88[,percentageDiff := (totSales - ControlSales)/ControlSales]
View(percentageDiff_88)

stdDev <- sd(percentageDiff_88[YEARMONTH < 201902 , percentageDiff])
View(stdDev)

degreeoffreedom <- 7 #Length of the pre trial period is 8 months. degree = len(df) - 1

percentageDiff_88[YEARMONTH >= 201902 & YEARMONTH <= 201904,
            tvalue := (percentageDiff - 0)/stdDev]
```

```
criticalTValue_88 <- qt(0.95,df=degreeoffreedom)

View(percentageDiff_88)
View(criticalTValue_88)
#The Critical Value is 1.89 and T values for trial months is -0.4,3.8,3.3
#With higher t values it can be statistically proven that the trial has increased the sales
significantly

#Visual Representation of the Total sales by month
measureOverTimeSales <- data.table::copy(measureovertime)
pastSales <- measureOverTimeSales[, Store_type := ifelse(STORE_NBR == 88, "Trial",
                                    ifelse(STORE_NBR == control_store_88, "Control", "Other"))
                    ][, totSales := totSales
                    ][, TransactionMonth := as.Date(paste(YEARMONTH%/%100,YEARMONTH
%% 100, 1, sep = "-"), "%Y-%m-%d")
                    ][Store_type %in% c("Trial", "Control"), ]


pastSales_Controls95 <- pastSales[Store_type == "Control",
                    ][, totSales := totSales * (1 + stdDev * 2)
                    ][, Store_type := "Control 95th % confidence interval"]

pastSales_Controls05 <- pastSales[Store_type == "Control",
                    ][, totSales := totSales * (1 - stdDev * 2)
                    ][, Store_type := "Control 5th % confidence interval"]

trialAssessment <- rbind(pastSales, pastSales_Controls95, pastSales_Controls05)

View(trialAssessment)

ggplot(trialAssessment, aes(TransactionMonth, totSales, color = Store_type)) +
 geom_rect(data = trialAssessment[ YEARMONTH >= 201902 & YEARMONTH <= 201904,],
       aes(xmin = min(TransactionMonth), xmax = max(TransactionMonth), ymin = 0 , ymax =
           Inf, color = NULL), show.legend = FALSE) +
 geom_line() +
 labs(x = "Month of operation", y = "Total sales", title = "Total sales by month")

#Hypothesis Testing - No. Of Customers
#H0: There is no significant difference in Total customers during the trial period
#H1: There is significant difference in Total customers during the trial period
#Scaling
scalingFactorforControlCustomers_88 <- preTrialMeasures[STORE_NBR == 88 &
YEARMONTH < 201902,sum(nCustomers)]/
```

```
 preTrialMeasures[STORE_NBR == control_store_88 & YEARMONTH <
201902,sum(nCustomers)]
measureOverTimeCustomers <- data.table::copy(measureovertime)
scaledControlCustomers_88 <- measureOverTimeCustomers[STORE_NBR ==
control_store_88,][,ScaledCustomers := nCustomers
*scalingFactorforControlCustomers_88]
trialCustomers_88 <- measureovertime[STORE_NBR == 88, .(YEARMONTH, nCustomers)]
scaledControlCustomers_s88 <- scaledControlCustomers_88[, .(YEARMONTH,
ScaledCustomers)]

custPercentageDiff_88 <- merge(trialCustomers_88,scaledControlCustomers_s88, by =
"YEARMONTH")
custPercentageDiff_88[,percentageDiff := (nCustomers -
ScaledCustomers)/ScaledCustomers]
View(custPercentageDiff_88)

stdDev <- sd(custPercentageDiff_88[YEARMONTH < 201902 , percentageDiff])
View(stdDev)

custPercentageDiff_88[YEARMONTH >= 201902 & YEARMONTH <= 201904,
            tvalue := (percentageDiff - 0)/stdDev]

View(custPercentageDiff_88)
View(criticalTValue_88)

#The Critical Value is 1.89 and T values for trial months is -0.6,8.3,4.5
#With higher t values it can be statistically proven that the trial has increased the No. Of
Customers significantly

#Visual Representation of the Total Customers by month
measureOverTimeCustomers <- data.table::copy(measureovertime)
pastCustomers <- measureOverTimeCustomers[, Store_type := ifelse(STORE_NBR == 88,
"Trial",
                                    ifelse(STORE_NBR == control_store_88, "Control", "Other"))
                        ][, nCust := mean(nCustomers), by = c("YEARMONTH","Store_type")
                        ][, TransactionMonth :=
as.Date(paste(YEARMONTH%/%100,YEARMONTH %% 100, 1, sep = "-"), "%Y-%m-%d")
                        ][Store_type %in% c("Trial", "Control"), ]


pastCustomers_Controls95 <- pastCustomers[Store_type == "Control",
                    ][, nCust := nCust * (1 + stdDev * 2)
                    ][, Store_type := "Control 95th % confidence interval"]
```

```
pastCustomers_Controls05 <- pastCustomers[Store_type == "Control",
                          ][, nCust := nCust * (1 - stdDev * 2)
                          ][, Store_type := "Control 5th % confidence interval"]

trialAssessment_cust <- rbind(pastCustomers, pastCustomers_Controls95,
pastCustomers_Controls05)

View(trialAssessment_cust)

ggplot(trialAssessment_cust, aes(TransactionMonth, nCust, color = Store_type)) +
  geom_rect(data = trialAssessment_cust[ YEARMONTH >= 201902 & YEARMONTH <=
201904,],
          aes(xmin = min(TransactionMonth), xmax = max(TransactionMonth), ymin = 0 , ymax =
              Inf, color = NULL), show.legend = FALSE) +
  geom_line() +
  labs(x = "Month of operation", y = "Total Customers", title = "Total Customers by month")
```