Retail Strategy and Analytics: Task 2

Jose A. Maldonado

###Transferring data to R

filepath = "C:/Users/Running Turtle/Desktop/Virtual Internships/Quantium/"  
data = fread(paste0(filepath, "QVI\_mData.csv"))  
theme\_set(theme\_bw())  
theme\_update(plot.title = element\_text(hjust = 0.5))

###Selecting controlled stores

#Stores 77, 86, 88 are trial stores.   
data[, YEARMONTH := year(DATE)\*100 + month(DATE)] #Adding a new month ID column in the data.  
data #Checking the data to verify

## LYLTY\_CARD\_NBR DATE STORE\_NBR TXN\_ID PROD\_NBR  
## 1: 1000 2018-10-17 1 1 5  
## 2: 1002 2018-09-16 1 2 58  
## 3: 1003 2019-03-07 1 3 52  
## 4: 1003 2019-03-08 1 4 106  
## 5: 1004 2018-11-02 1 5 96  
## ---   
## 246736: 2370651 2018-08-03 88 240350 4  
## 246737: 2370701 2018-12-08 88 240378 24  
## 246738: 2370751 2018-10-01 88 240394 60  
## 246739: 2370961 2018-10-24 88 240480 70  
## 246740: 2373711 2018-12-14 88 241815 16  
## PROD\_NAME PROD\_QTY TOT\_SALES PACK\_SIZE  
## 1: Natural Chip Compny SeaSalt175g 2 6.0 175  
## 2: Red Rock Deli Chikn&Garlic Aioli 150g 1 2.7 150  
## 3: Grain Waves Sour Cream&Chives 210G 1 3.6 210  
## 4: Natural ChipCo Hony Soy Chckn175g 1 3.0 175  
## 5: WW Original Stacked Chips 160g 1 1.9 160  
## ---   
## 246736: Dorito Corn Chp Supreme 380g 2 13.0 380  
## 246737: Grain Waves Sweet Chilli 210g 2 7.2 210  
## 246738: Kettle Tortilla ChpsFeta&Garlic 150g 2 9.2 150  
## 246739: Tyrrells Crisps Lightly Salted 165g 2 8.4 165  
## 246740: Smiths Crinkle Chips Salt & Vinegar 330g 2 11.4 330  
## BRAND LIFESTAGE PREMIUM\_CUSTOMER YEARMONTH  
## 1: NATURAL YOUNG SINGLES/COUPLES Premium 201810  
## 2: RRD YOUNG SINGLES/COUPLES Mainstream 201809  
## 3: GRNWVES YOUNG FAMILIES Budget 201903  
## 4: NATURAL YOUNG FAMILIES Budget 201903  
## 5: WOOLWORTHS OLDER SINGLES/COUPLES Mainstream 201811  
## ---   
## 246736: DORITOS MIDAGE SINGLES/COUPLES Mainstream 201808  
## 246737: GRNWVES YOUNG FAMILIES Mainstream 201812  
## 246738: KETTLE YOUNG FAMILIES Premium 201810  
## 246739: TYRRELLS OLDER FAMILIES Budget 201810  
## 246740: SMITHS YOUNG SINGLES/COUPLES Mainstream 201812

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 = c("STORE\_NBR", "YEARMONTH")][order(STORE\_NBR, YEARMONTH)] #Measure calculations that we will use during the analysis  
storesWithfullObs = unique(measureOverTime[, .N, STORE\_NBR][N == 12, STORE\_NBR])  
preTrialMeasures = measureOverTime[YEARMONTH < 201902 & STORE\_NBR %in% storesWithfullObs, ] #Filter to the pre-trial period and stores with full obs periods  
  
calculateCorrelation = function(inputTable, metricCol, storeComparison) {  
 calcCorrTable = data.table(Store1 = numeric(),  
 Store2 = numeric(),  
 corr\_measure = numeric())  
 storeNumbers = unique(inputTable[, STORE\_NBR])  
 for (i in storeNumbers) {  
 calculatedMeasure = data.table("Store1" = storeComparison,  
 "Store2" = i,  
 "corr\_measure" = cor(inputTable[STORE\_NBR == storeComparison,  
 eval(metricCol)], inputTable[STORE\_NBR == i,  
 eval(metricCol)]))  
 calcCorrTable = rbind(calcCorrTable, calculatedMeasure)  
 }  
return(calcCorrTable)  
} #Function to calculate correlation for a measure looping through each control store.  
  
calculateMagnitudeDistance = function(inputTable, metricCol, storeComparison) {  
 calcDistTable = data.table(Store1 = numeric(), Store2 = numeric(), YEARMONTH = numeric(), measure = numeric())  
 storeNumbers = unique(inputTable[, STORE\_NBR])  
 for (i in storeNumbers) {  
 calculatedMeasure = data.table("Store1" = storeComparison,  
 "Store2" = i,  
 "YEARMONTH" = inputTable[STORE\_NBR == storeComparison, YEARMONTH],  
 "measure" = abs(inputTable[STORE\_NBR == storeComparison, eval(metricCol)] - inputTable[STORE\_NBR == i, eval(metricCol)]))  
 calcDistTable = rbind(calcDistTable, calculatedMeasure)} #Function to calculate a standardised magnitude distance for a measure  
 minMaxDist = calcDistTable[, .(minDist = min(measure), maxDist = max(measure)), by = c("Store1", "YEARMONTH")]  
 disTable = merge(calcDistTable, minMaxDist, by = c("Store1", "YEARMONTH"))  
 disTable[, magnitudeMeasure := 1 - (measure - minDist)/(maxDist - minDist)]  
 finalDistTable = disTable[, .(mag\_measure = mean(magnitudeMeasure)), by = .(Store1, Store2)]  
 return(finalDistTable)  
} #Standardise the magnitude distance so that the measure ranges from 0 - 1  
  
trial\_store = 77  
corr\_nSales = calculateCorrelation(preTrialMeasures, quote(totSales), trial\_store)   
corr\_nSales[order(-corr\_measure)] #Calculated correlations vs store 77 based on total sales

## Store1 Store2 corr\_measure  
## 1: 77 77 1.0000000  
## 2: 77 233 0.9736429  
## 3: 77 50 0.8977013  
## 4: 77 162 0.8575839  
## 5: 77 71 0.8156345  
## ---   
## 255: 77 244 -0.7293373  
## 256: 77 75 -0.7952057  
## 257: 77 242 -0.8041040  
## 258: 77 9 -0.8132854  
## 259: 77 186 -0.9171306

corr\_nCustomers = calculateCorrelation(preTrialMeasures, quote(nCustomers), trial\_store)  
corr\_nCustomers[order(-corr\_measure)] #Calculated correlations vs store 77 based on customers

## Store1 Store2 corr\_measure  
## 1: 77 77 1.0000000  
## 2: 77 233 0.9656821  
## 3: 77 119 0.9190639  
## 4: 77 113 0.9016299  
## 5: 77 254 0.9016105  
## ---   
## 255: 77 227 -0.7506291  
## 256: 77 186 -0.7668731  
## 257: 77 169 -0.7842412  
## 258: 77 9 -0.8045038  
## 259: 77 54 -0.8314799

magnitude\_nSales = calculateMagnitudeDistance(preTrialMeasures, quote(totSales), trial\_store) #Calculated magnitute based on total sales, store 77  
magnitude\_nCustomers = calculateMagnitudeDistance(preTrialMeasures, quote(nCustomers), trial\_store) #Calcualted magnitute based on customers, store 77  
  
corr\_weight = 0.5  
score\_nSales = merge(corr\_nSales, magnitude\_nSales , by =  
 c("Store1", "Store2"))[, scoreNSales := (corr\_measure + mag\_measure)/2 ] #Combined score of correlation for the sales  
score\_nCustomers = merge(corr\_nCustomers, magnitude\_nCustomers , by = c("Store1", "Store2"))[, scoreNCust := (corr\_measure + mag\_measure)/2] #Combined score of correlation for the customers  
score\_nSales[order(-scoreNSales)] #Checking

## Store1 Store2 corr\_measure mag\_measure scoreNSales  
## 1: 77 77 1.0000000 1.0000000 1.00000000  
## 2: 77 233 0.9736429 0.9858967 0.97976982  
## 3: 77 50 0.8977013 0.9743836 0.93604241  
## 4: 77 41 0.6591279 0.9568926 0.80801022  
## 5: 77 167 0.6546680 0.9546588 0.80466338  
## ---   
## 255: 77 88 -0.3085869 0.1505884 -0.07899929  
## 256: 77 4 -0.3478465 0.1808497 -0.08349841  
## 257: 77 247 -0.7109062 0.5319305 -0.08948786  
## 258: 77 138 -0.6941490 0.5084970 -0.09282599  
## 259: 77 75 -0.7952057 0.3157098 -0.23974799

score\_nCustomers[order(-scoreNCust)] #Checking

## Store1 Store2 corr\_measure mag\_measure scoreNCust  
## 1: 77 77 1.0000000 1.0000000 1.00000000  
## 2: 77 233 0.9656821 0.9909635 0.97832280  
## 3: 77 254 0.9016105 0.9295040 0.91555724  
## 4: 77 35 0.8927414 0.8995606 0.89615102  
## 5: 77 84 0.8515210 0.9230425 0.88728178  
## ---   
## 255: 77 165 -0.3647459 0.1809069 -0.09191945  
## 256: 77 147 -0.7148957 0.5095139 -0.10269086  
## 257: 77 102 -0.6371093 0.4300131 -0.10354806  
## 258: 77 75 -0.5650164 0.3419888 -0.11151380  
## 259: 77 227 -0.7506291 0.4317654 -0.15943183

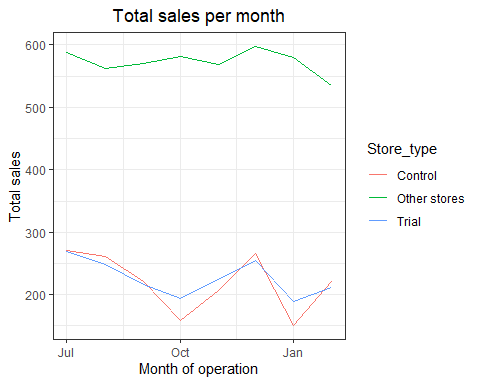
score\_Control = merge(score\_nSales, score\_nCustomers , by = c("Store1", "Store2")) #Combining scores across the drivers  
score\_Control[, finalControlScore := scoreNSales \* 0.5 + scoreNCust \* 0.5]  
score\_Control[order(-finalControlScore)] #Checking

## Store1 Store2 corr\_measure.x mag\_measure.x scoreNSales corr\_measure.y  
## 1: 77 77 1.0000000 1.0000000 1.0000000000 1.0000000  
## 2: 77 233 0.9736429 0.9858967 0.9797698229 0.9656821  
## 3: 77 50 0.8977013 0.9743836 0.9360424099 0.7093977  
## 4: 77 35 0.6910897 0.9076712 0.7993804422 0.8927414  
## 5: 77 254 0.5848729 0.9209658 0.7529193791 0.9016105  
## ---   
## 255: 77 4 -0.3478465 0.1808497 -0.0834984086 -0.3054117  
## 256: 77 147 -0.6640142 0.5661961 -0.0489090421 -0.7148957  
## 257: 77 227 -0.5040822 0.5057102 0.0008139983 -0.7506291  
## 258: 77 138 -0.6941490 0.5084970 -0.0928259921 -0.5483439  
## 259: 77 75 -0.7952057 0.3157098 -0.2397479861 -0.5650164  
## mag\_measure.y scoreNCust finalControlScore  
## 1: 1.0000000 1.00000000 1.00000000  
## 2: 0.9909635 0.97832280 0.97904631  
## 3: 0.9319895 0.82069361 0.87836801  
## 4: 0.8995606 0.89615102 0.84776573  
## 5: 0.9295040 0.91555724 0.83423831  
## ---   
## 255: 0.2022603 -0.05157567 -0.06753704  
## 256: 0.5095139 -0.10269086 -0.07579995  
## 257: 0.4317654 -0.15943183 -0.07930892  
## 258: 0.4165489 -0.06589749 -0.07936174  
## 259: 0.3419888 -0.11151380 -0.17563089

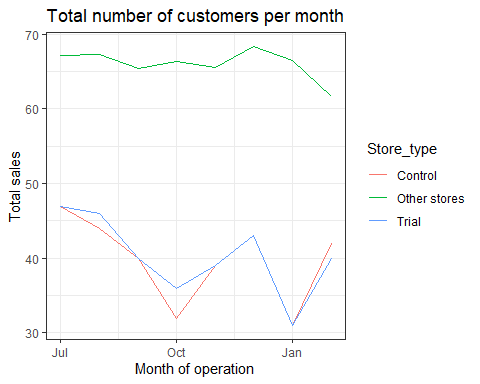
control\_store = score\_Control[Store1 == trial\_store, ][order(-finalControlScore)][2, Store2] #Selects a control store based on the highest matching store  
control\_store #Store 233 was selected

## [1] 233

measureOverTimeSales = measureOverTime  
pastSales = measureOverTimeSales[, Store\_type := ifelse(STORE\_NBR == trial\_store, "Trial",  
 ifelse(STORE\_NBR == control\_store, "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 < 201903 , ]  
ggplot(pastSales, aes(TransactionMonth, totSales, color = Store\_type)) +  
 geom\_line() +  
 labs(x = "Month of operation", y = "Total sales", title = "Total sales per month") #Time plot to check on trend based on the drivers



measureOverTimeCusts = measureOverTime  
pastCustomers = measureOverTimeCusts[, Store\_type := ifelse(STORE\_NBR == trial\_store, "Trial",  
 ifelse(STORE\_NBR == control\_store, "Control", "Other stores"))  
][, numberCustomers := mean(nCustomers), by = c("YEARMONTH", "Store\_type") ][, TransactionMonth := as.Date(paste(YEARMONTH %/% 100, YEARMONTH %% 100, 1, sep = "-"), "%Y-%m-%d")  
][YEARMONTH < 201903 , ]  
ggplot(pastCustomers, aes(TransactionMonth, numberCustomers, color = Store\_type)) + geom\_line() +  
 labs(x = "Month of operation", y = "Total sales", title = "Total number of customers per month") #Time plot to check customer count trends by comparing trial store



###Assesing the trial

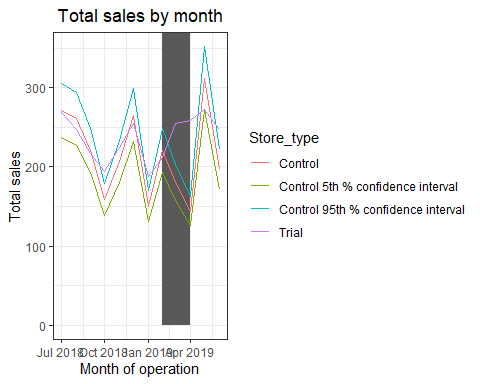
scalingFactorForControlSales = preTrialMeasures[STORE\_NBR == trial\_store & YEARMONTH < 201902, sum(totSales)]/preTrialMeasures[STORE\_NBR == control\_store & YEARMONTH < 201902, sum(totSales)] #Scaling the pre-trial control sales to match pre-trial store sales  
measureOverTimeSales = measureOverTime  
scaledControlSales = measureOverTimeSales[STORE\_NBR == control\_store, ][ , controlSales := totSales \* scalingFactorForControlSales] #Applying scaling factors  
percentageDiff = merge(scaledControlSales[, c("YEARMONTH", "controlSales")],  
 measureOverTime[STORE\_NBR == trial\_store, c("totSales", "YEARMONTH")],  
 by = "YEARMONTH")[, percentageDiff := abs(controlSales - totSales)/controlSales] #Percent diff between scaled control and trial sales  
percentageDiff #Checking  
  
stdDev = sd(percentageDiff[YEARMONTH < 201902 , percentageDiff])  
degreesOfFreedom = 7 #There are 8 months in the pre-trial period DF = 8 -1 = 7  
percentageDiff[, tValue := (percentageDiff - 0)/stdDev  
 ][, TransactionMonth := as.Date(paste(YEARMONTH %/% 100, YEARMONTH %% 100, 1,  
 sep = "-"), "%Y-%m-%d")  
][YEARMONTH < 201905 & YEARMONTH > 201901, .(TransactionMonth, tValue)]

## TransactionMonth tValue  
## 1: 2019-02-01 1.223912  
## 2: 2019-03-01 5.633494  
## 3: 2019-04-01 11.336505

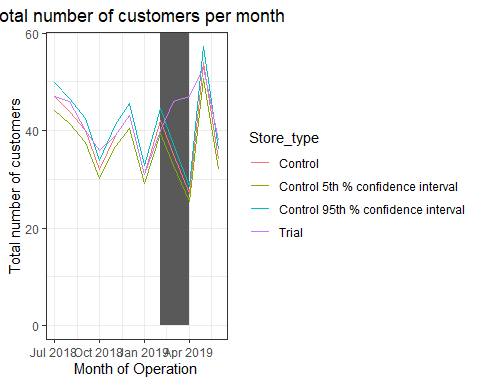
#Our null hypothesis is that the trial period is the same as the pre-trial period.   
#We will test with a null hypothesis of there being 0 difference between trial and control stores.  
qt(0.95, df = degreesOfFreedom)

## [1] 1.894579

#We see that the t-value is much larger than the 95th percentile value of the distribution for March and April. This means the increase in sales in the trial store in march and April is statistically greater than in the control store.   
  
measureOverTimeSales = measureOverTime  
pastSales = measureOverTimeSales[, Store\_type := ifelse(STORE\_NBR == trial\_store, "Trial", ifelse(STORE\_NBR == control\_store, "Control", "Other\_stores"))  
][, totSales := mean(totSales), by = c("YEARMONTH", "Store\_type")  
][, TransactionMonth := as.Date(paste(YEARMONTH %/% 100, YEARMONTH %% 100, 1, sep = "-"), "%Y-%m-%d")  
][Store\_type %in% c("Trial", "Control"), ] #Creating new variables storetype, totsales, and trancationmonth in the data table  
pastSales\_Controls95 = pastSales[Store\_type == "Control",  
][, totSales := totSales \* (1 + stdDev \* 2)  
][, Store\_type := "Control 95th % confidence interval"] #Control store 95th percentile  
pastSales\_Controls5 = pastSales[Store\_type == "Control",  
][, totSales := totSales \* (1 - stdDev \* 2) ][, Store\_type := "Control 5th % confidence interval"] #Control store 5th percentile  
trialAssessment = rbind(pastSales, pastSales\_Controls95, pastSales\_Controls5)  
  
ggplot(trialAssessment, aes(TransactionMonth, totSales, color = Store\_type)) +  
 geom\_rect(data = trialAssessment[ YEARMONTH < 201905 & YEARMONTH > 201901 ,], 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")



#The graph shows that the trial in store 77 is significantly different to its control store in the trail period as the trial store performance lies outside the 5% to 95% confidence interval of the control store in two of the three trial months.  
scalingFactorForControlCust = preTrialMeasures[STORE\_NBR == trial\_store & YEARMONTH < 201902, sum(nCustomers)] / preTrialMeasures[STORE\_NBR == control\_store & YEARMONTH < 201902, sum(nCustomers)]  
measureOverTimeCusts = measureOverTime  
scaledControlCustomers = measureOverTimeCusts[STORE\_NBR == control\_store, ][, controlCustomers := nCustomers \* scalingFactorForControlCust ][, Store\_type := ifelse(STORE\_NBR == trial\_store, "Trial", ifelse(STORE\_NBR == control\_store, "Control", "Other stores"))]  
percentageDiff = merge(scaledControlCustomers[, c("YEARMONTH", "controlCustomers")], measureOverTimeCusts[STORE\_NBR == trial\_store, c("nCustomers", "YEARMONTH")], by = "YEARMONTH"  
)[, percentageDiff := abs(controlCustomers-nCustomers)/controlCustomers]  
  
stdDev = sd(percentageDiff[YEARMONTH < 201902 , percentageDiff])  
degreesOfFreedom = 7 #There are 8 months in the pre-trial period DF = 8 -1 = 7  
pastCustomers = measureOverTimeCusts[, nCusts := mean(nCustomers), by = c("YEARMONTH", "Store\_type")  
 ][Store\_type %in% c("Trial", "Control"), ] #Trial and control store number of customers  
pastCustomers\_Controls95 = pastCustomers[Store\_type == "Control",  
 ][, nCusts := nCusts \* (1 + stdDev \* 2)  
 ][, Store\_type := "Control 95th % confidence interval"] #Control store 95th percentile  
pastCustomers\_Controls5 = pastCustomers[Store\_type == "Control",  
 ][, nCusts := nCusts \* (1 - stdDev \* 2)  
 ][, Store\_type := "Control 5th % confidence interval"] #Control store 5th percentile  
trialAssessment = rbind(pastCustomers, pastCustomers\_Controls95, pastCustomers\_Controls5)  
ggplot(trialAssessment, aes(TransactionMonth, nCusts, color = Store\_type)) +  
 geom\_rect(data = trialAssessment[YEARMONTH < 201905 & YEARMONTH >201901,], 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 number of customers", title = "Total number of customers per month")

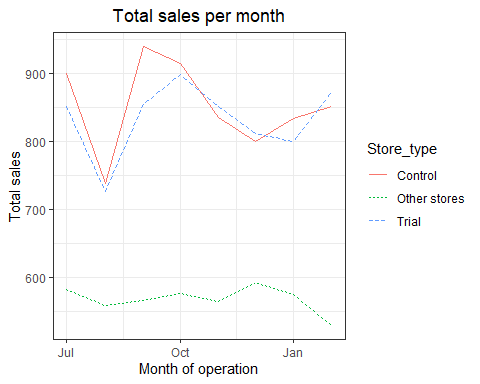


###Store 86

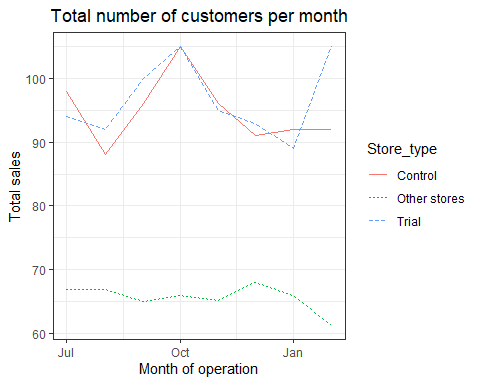
measureOverTime = data[, .(totSales = sum(TOT\_SALES),  
 nCustomers = uniqueN(LYLTY\_CARD\_NBR),  
 nTxnPerCust = uniqueN(TXN\_ID)/uniqueN(LYLTY\_CARD\_NBR),  
 nChipsPerTxn = sum(PROD\_QTY)/uniqueN(TXN\_ID),  
 avgPricePerUnit = sum(TOT\_SALES)/sum(PROD\_QTY)),  
 by = c("STORE\_NBR", "YEARMONTH")][order(STORE\_NBR, YEARMONTH)]  
  
trial\_store = 86  
  
corr\_nSales = calculateCorrelation(preTrialMeasures, quote(totSales), trial\_store)  
  
corr\_nCustomers = calculateCorrelation(preTrialMeasures, quote(nCustomers), trial\_store)  
  
magnitude\_nSales = calculateMagnitudeDistance(preTrialMeasures, quote(totSales), trial\_store)  
  
magnitude\_nCustomers = calculateMagnitudeDistance(preTrialMeasures, quote(nCustomers), trial\_store)  
  
corr\_weight = 0.5  
score\_nSales = merge(corr\_nSales, magnitude\_nSales , by =  
 c("Store1", "Store2"))[, scoreNSales := (corr\_measure + mag\_measure)/2 ]  
score\_nCustomers = merge(corr\_nCustomers, magnitude\_nCustomers , by = c("Store1", "Store2"))[, scoreNCust := (corr\_measure + mag\_measure)/2]  
score\_Control = merge(score\_nSales, score\_nCustomers , by = c("Store1", "Store2"))  
score\_Control[, finalControlScore := scoreNSales \* 0.5 + scoreNCust \* 0.5]  
control\_store = score\_Control[Store1 == trial\_store, ][order(-finalControlScore)][2, Store2]  
control\_store #Store 155 was selected

## [1] 155

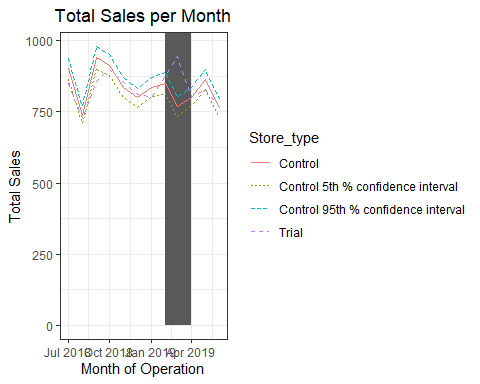
measureOverTimeSales = measureOverTime  
pastSales = measureOverTimeSales[, Store\_type := ifelse(STORE\_NBR == trial\_store, "Trial",  
 ifelse(STORE\_NBR == control\_store, "Control", "Other stores"))  
][, totSales := mean(totSales), by = c("YEARMONTH", "Store\_type")  
][, TransactionMonth := as.Date(paste(YEARMONTH %/% 100, YEARMONTH %% 100, 1, sep = "-"), "%Y-%m-%d")  
][YEARMONTH < 201903 , ]  
ggplot(pastSales, aes(TransactionMonth, totSales, color = Store\_type)) +  
 geom\_line(aes(linetype = Store\_type)) +  
 labs(x = "Month of operation", y = "Total sales", title = "Total sales per month")



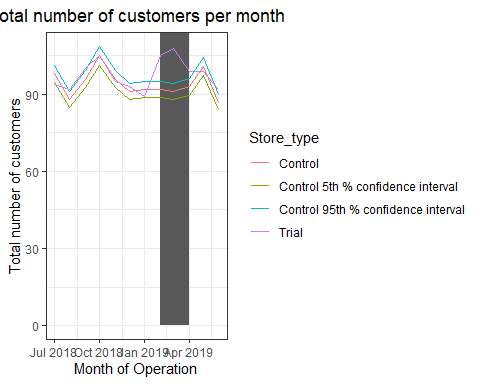
measureOverTimeCusts = measureOverTime  
pastCustomers = measureOverTimeCusts[, Store\_type := ifelse(STORE\_NBR == trial\_store, "Trial",  
 ifelse(STORE\_NBR == control\_store, "Control", "Other stores"))  
][, numberCustomers := mean(nCustomers), by = c("YEARMONTH", "Store\_type")  
][, TransactionMonth := as.Date(paste(YEARMONTH %/% 100, YEARMONTH %% 100, 1, sep = "-"), "%Y-%m-%d")  
][YEARMONTH < 201903 , ]  
ggplot(pastCustomers, aes(TransactionMonth, numberCustomers, color = Store\_type)) +  
 geom\_line(aes(linetype = Store\_type)) +  
 labs(x = "Month of operation", y = "Total sales", title = "Total number of customers per month")



scalingFactorForControlSales = preTrialMeasures[STORE\_NBR == trial\_store & YEARMONTH < 201902, sum(totSales)]/preTrialMeasures[STORE\_NBR == control\_store & YEARMONTH < 201902, sum(totSales)]  
measureOverTimeSales = measureOverTime  
scaledControlSales = measureOverTimeSales[STORE\_NBR == control\_store, ][ , controlSales := totSales \* scalingFactorForControlSales]  
percentageDiff = merge(scaledControlSales[, c("YEARMONTH", "controlSales")], measureOverTime[STORE\_NBR == trial\_store, c("totSales", "YEARMONTH")],  
 by = "YEARMONTH")[, percentageDiff := abs(controlSales - totSales)/controlSales]  
stdDev = sd(percentageDiff[YEARMONTH < 201902, percentageDiff])  
degreesOfFreedom = 7  
measureOverTimeSales = measureOverTime  
pastSales = measureOverTimeSales[, Store\_type := ifelse(STORE\_NBR == trial\_store, "Trial", ifelse(STORE\_NBR == control\_store, "Control", "Other stores"))  
][, totSales := mean(totSales), by = c("YEARMONTH", "Store\_type")  
][, 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\_Controls5 = pastSales[Store\_type == "Control",  
 ][, totSales := totSales \* (1 - stdDev \* 2)  
 ][, Store\_type := "Control 5th % confidence interval"]  
trialAssessment = rbind(pastSales, pastSales\_Controls95, pastSales\_Controls5)  
ggplot(trialAssessment, aes(TransactionMonth, totSales, color = Store\_type)) +  
 geom\_rect(data = trialAssessment[YEARMONTH < 201905 & YEARMONTH >201901,], aes(xmin = min(TransactionMonth), xmax = max(TransactionMonth), ymin = 0, ymax = Inf, color = NULL ), show.legend = FALSE) +  
 geom\_line(aes(linetype = Store\_type)) + labs(x = "Month of Operation", y = "Total Sales", title = "Total Sales per Month")



#The results show that the trial in store 86 is not significantly different to its control store in the trial period as the trial store performance lies inside the 5% - 95% confidence interval of the control store in two of the three trial months.  
  
scalingFactorForControlCust = preTrialMeasures[STORE\_NBR == trial\_store & YEARMONTH < 201902, sum(nCustomers)] / preTrialMeasures[STORE\_NBR == trial\_store & YEARMONTH < 201902, sum(nCustomers)]  
measureOverTimeCusts = measureOverTime  
scaledControlCustomers = measureOverTimeCusts[STORE\_NBR == control\_store, ][, controlCustomers := nCustomers \* scalingFactorForControlCust ][, Store\_type := ifelse(STORE\_NBR == trial\_store, "Trial", ifelse(STORE\_NBR == control\_store, "Control", "Other stores"))]  
percentageDiff = merge(scaledControlCustomers[, c("YEARMONTH", "controlCustomers")],  
 measureOverTime[STORE\_NBR == trial\_store, c("nCustomers", "YEARMONTH")],  
 by = "YEARMONTH"  
 )[, percentageDiff := abs(controlCustomers-nCustomers)/controlCustomers]  
  
stdDev = sd(percentageDiff[YEARMONTH < 201902 , percentageDiff])  
degreesOfFreedom = 7  
pastCustomers = measureOverTimeCusts[, nCusts := mean(nCustomers), by = c("YEARMONTH", "Store\_type")  
 ][Store\_type %in% c("Trial", "Control"), ]  
pastCustomers\_Controls95 = pastCustomers[Store\_type == "Control",  
 ][, nCusts := nCusts \* (1 + stdDev \* 2)  
 ][, Store\_type := "Control 95th % confidence interval"]  
pastCustomers\_Controls5 = pastCustomers[Store\_type == "Control",  
 ][, nCusts := nCusts \* (1 - stdDev \* 2)  
 ][, Store\_type := "Control 5th % confidence interval"]  
trialAssessment = rbind(pastCustomers, pastCustomers\_Controls95, pastCustomers\_Controls5)  
ggplot(trialAssessment, aes(TransactionMonth, nCusts, color = Store\_type)) +  
 geom\_rect(data = trialAssessment[YEARMONTH < 201905 & YEARMONTH >201901,], 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 number of customers", title = "Total number of customers per month")



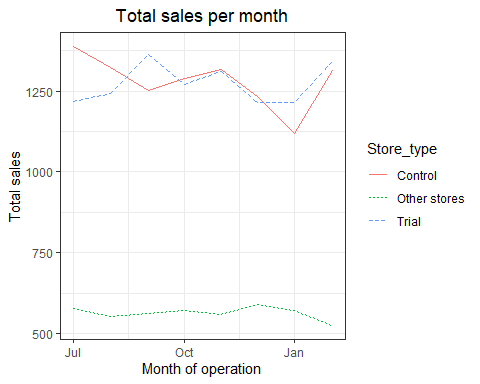
#It looks like the number of customers is significantly higher in all of the three months. This seems to indicate that the trial had a significant impact on increasing the number of customers in trial store 86 but as we saw sales were not significantly higher. It is recommended to check with the category manager if there were special deals in the trial store that may have resulted in lower prices, impacting the results.

###Store 88

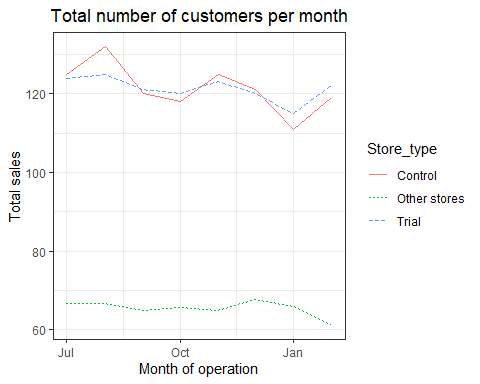
measureOverTime = data[, .(totSales = sum(TOT\_SALES),  
 nCustomers = uniqueN(LYLTY\_CARD\_NBR),  
 nTxnPerCust = uniqueN(TXN\_ID)/uniqueN(LYLTY\_CARD\_NBR),  
 nChipsPerTxn = sum(PROD\_QTY)/uniqueN(TXN\_ID),  
 avgPricePerUnit = sum(TOT\_SALES)/sum(PROD\_QTY)),  
 by = c("STORE\_NBR", "YEARMONTH")][order(STORE\_NBR, YEARMONTH)]  
trial\_store = 88  
corr\_nSales = calculateCorrelation(preTrialMeasures, quote(totSales), trial\_store)  
corr\_nCustomers = calculateCorrelation(preTrialMeasures, quote(nCustomers), trial\_store)  
  
magnitude\_nSales = calculateMagnitudeDistance(preTrialMeasures, quote(totSales), trial\_store)  
magnitude\_nCustomers = calculateMagnitudeDistance(preTrialMeasures, quote(nCustomers), trial\_store)  
  
corr\_weight = 0.5  
score\_nSales = merge(corr\_nSales, magnitude\_nSales , by =  
 c("Store1", "Store2"))[, scoreNSales := (corr\_measure + mag\_measure)/2 ]  
score\_nCustomers = merge(corr\_nCustomers, magnitude\_nCustomers , by = c("Store1", "Store2"))[, scoreNCust := (corr\_measure + mag\_measure)/2]  
score\_Control = merge(score\_nSales, score\_nCustomers , by = c("Store1", "Store2"))  
score\_Control[, finalControlScore := scoreNSales \* 0.5 + scoreNCust \* 0.5]  
control\_store = score\_Control[Store1 == trial\_store, ][order(-finalControlScore)][2, Store2]  
control\_store #Store 237 was selected

## [1] 237

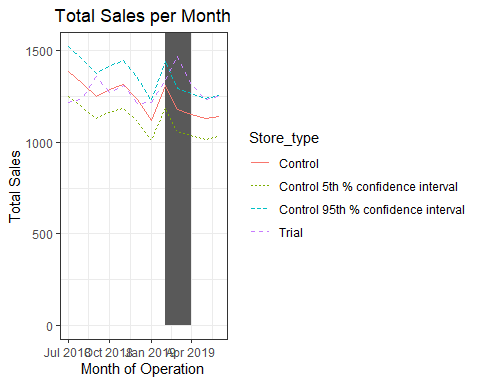
measureOverTimeSales = measureOverTime  
pastSales = measureOverTimeSales[, Store\_type := ifelse(STORE\_NBR == trial\_store, "Trial",  
 ifelse(STORE\_NBR == control\_store, "Control", "Other stores"))  
][, totSales := mean(totSales), by = c("YEARMONTH", "Store\_type")  
][, TransactionMonth := as.Date(paste(YEARMONTH %/% 100, YEARMONTH %% 100, 1, sep = "-"), "%Y-%m-%d")  
][YEARMONTH < 201903 , ]  
ggplot(pastSales, aes(TransactionMonth, totSales, color = Store\_type)) +  
 geom\_line(aes(linetype = Store\_type)) +  
 labs(x = "Month of operation", y = "Total sales", title = "Total sales per month")



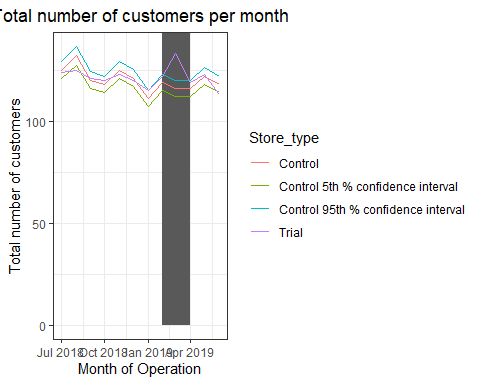
measureOverTimeCusts = measureOverTime  
pastCustomers = measureOverTimeCusts[, Store\_type := ifelse(STORE\_NBR == trial\_store, "Trial",  
 ifelse(STORE\_NBR == control\_store, "Control", "Other stores"))  
][, numberCustomers := mean(nCustomers), by = c("YEARMONTH", "Store\_type")  
][, TransactionMonth := as.Date(paste(YEARMONTH %/% 100, YEARMONTH %% 100, 1, sep = "-"), "%Y-%m-%d")  
][YEARMONTH < 201903 , ]  
ggplot(pastCustomers, aes(TransactionMonth, numberCustomers, color = Store\_type)) +  
 geom\_line(aes(linetype = Store\_type)) +  
 labs(x = "Month of operation", y = "Total sales", title = "Total number of customers per month")



scalingFactorForControlSales = preTrialMeasures[STORE\_NBR == trial\_store & YEARMONTH < 201902, sum(totSales)]/preTrialMeasures[STORE\_NBR == control\_store & YEARMONTH < 201902, sum(totSales)]  
measureOverTimeSales = measureOverTime  
scaledControlSales = measureOverTimeSales[STORE\_NBR == control\_store, ][ , controlSales := totSales \* scalingFactorForControlSales]  
percentageDiff = merge(scaledControlSales[, c("YEARMONTH", "controlSales")], measureOverTime[STORE\_NBR == trial\_store, c("totSales", "YEARMONTH")],  
 by = "YEARMONTH")[, percentageDiff := abs(controlSales - totSales)/controlSales]  
stdDev = sd(percentageDiff[YEARMONTH < 201902, percentageDiff])  
degreesOfFreedom = 7  
measureOverTimeSales = measureOverTime  
pastSales = measureOverTimeSales[, Store\_type := ifelse(STORE\_NBR == trial\_store, "Trial", ifelse(STORE\_NBR == control\_store, "Control", "Other stores"))  
][, totSales := mean(totSales), by = c("YEARMONTH", "Store\_type")  
][, 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\_Controls5 = pastSales[Store\_type == "Control",  
 ][, totSales := totSales \* (1 - stdDev \* 2)  
 ][, Store\_type := "Control 5th % confidence interval"]  
trialAssessment = rbind(pastSales, pastSales\_Controls95, pastSales\_Controls5)  
ggplot(trialAssessment, aes(TransactionMonth, totSales, color = Store\_type)) +  
 geom\_rect(data = trialAssessment[YEARMONTH < 201905 & YEARMONTH >201901,], aes(xmin = min(TransactionMonth), xmax = max(TransactionMonth), ymin = 0, ymax = Inf, color = NULL ), show.legend = FALSE) +  
 geom\_line(aes(linetype = Store\_type)) + labs(x = "Month of Operation", y = "Total Sales", title = "Total Sales per Month")



#From the results we can see that the trial in store 88 is significantly different to its control store in the trial period as the trial store performance lies outside of the 5% - 95% C.I of the control store in two of the three trial months.   
  
scalingFactorForControlCust = preTrialMeasures[STORE\_NBR == trial\_store & YEARMONTH < 201902, sum(nCustomers)] / preTrialMeasures[STORE\_NBR == trial\_store & YEARMONTH < 201902, sum(nCustomers)]  
measureOverTimeCusts = measureOverTime  
scaledControlCustomers = measureOverTimeCusts[STORE\_NBR == control\_store, ][, controlCustomers := nCustomers \* scalingFactorForControlCust ][, Store\_type := ifelse(STORE\_NBR == trial\_store, "Trial", ifelse(STORE\_NBR == control\_store, "Control", "Other stores"))]  
percentageDiff = merge(scaledControlCustomers[, c("YEARMONTH", "controlCustomers")],  
 measureOverTime[STORE\_NBR == trial\_store, c("nCustomers", "YEARMONTH")],  
 by = "YEARMONTH"  
 )[, percentageDiff := abs(controlCustomers-nCustomers)/controlCustomers]  
  
stdDev = sd(percentageDiff[YEARMONTH < 201902 , percentageDiff])  
degreesOfFreedom = 7  
pastCustomers = measureOverTimeCusts[, nCusts := mean(nCustomers), by = c("YEARMONTH", "Store\_type")  
 ][Store\_type %in% c("Trial", "Control"), ]  
pastCustomers\_Controls95 = pastCustomers[Store\_type == "Control",  
 ][, nCusts := nCusts \* (1 + stdDev \* 2)  
 ][, Store\_type := "Control 95th % confidence interval"]  
pastCustomers\_Controls5 = pastCustomers[Store\_type == "Control",  
 ][, nCusts := nCusts \* (1 - stdDev \* 2)  
 ][, Store\_type := "Control 5th % confidence interval"]  
trialAssessment = rbind(pastCustomers, pastCustomers\_Controls95, pastCustomers\_Controls5)  
ggplot(trialAssessment, aes(TransactionMonth, nCusts, color = Store\_type)) +  
 geom\_rect(data = trialAssessment[YEARMONTH < 201905 & YEARMONTH >201901,], 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 number of customers", title = "Total number of customers per month")



#The total number of customers in the trail period for the trail store is significantly higher than the control store for two out of the three months which indicates a positive trail effect.

###Conclusion: ####We’ve found control stores 233, 155, 237 for trail stores 77, 86, and 88 respectively. Te results fro trial stores 77 and 88 during the trial period show a significant difference in at least wo of the three trial months but this is not the case for trial store 86. We can check with the client if the implementation of the trial was different in trial store 86. Overall the trial shows a significant increase in sales.