Quatium_task2

KOBENAN_ZAKARI

2025-07-09

Goal: Assessing whether trial stores improved more than control stores after a marketing intervention, and provide actionable recommendations per location.

```
 #Load library
library(readr)
library(data.table)
library(dplyr)
```

```
##
## Attachement du package : 'dplyr'

## Les objets suivants sont masqués depuis 'package:data.table':
##
##     between, first, last

## Les objets suivants sont masqués depuis 'package:stats':
##
##     filter, lag

## Les objets suivants sont masqués depuis 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(ggplot2)
library(tidyr)
library(stringr)
library(ggmosaic)
```

```
                    ### Select control (matching store (witness) before the trial)
```

```
QVI_data_store <- read.csv ("QVI_data_store.csv", sep = ",")
  # Label period "YEARMONTH" column
 data <- QVI_data_store  %>%
   mutate( YEARMONTH = as.integer(format(as.Date(DATE), "%y%m")))

 #  Measure of KPs as a metric over time
 kpsovertime <- data  %>%
   group_by(STORE_NBR, YEARMONTH) %>%
   summarise(REVENUE = sum(TOT_SALES),
```

```
            AVG_PRICE = sum (TOT_SALES)/ sum(PROD_QTY),
            NUMBER_CUSTOMERS = n_distinct (TXN_ID),
            nTxperCust =  n()/ n_distinct (TXN_ID),
            nPQYpertxn = sum(PROD_QTY)/ n(),
            .groups = "drop")
 head (kpsovertime)
```

```
## # A tibble: 6 x 7
##   STORE_NBR YEARMONTH REVENUE AVG_PRICE NUMBER_CUSTOMERS nTxperCust nPQYpertxn
##       <int>     <int>   <dbl>     <dbl>            <int>      <dbl>      <dbl>
## 1         1      1807    207.      3.34               52          1       1.19
## 2         1      1808    176.      3.26               43          1       1.26
## 3         1      1809    279.      3.72               62          1       1.21
## 4         1      1810    188.      3.24               45          1       1.29
## 5         1      1811    193.      3.38               47          1       1.21
## 6         1      1812    190.      3.33               47          1       1.21
```

```
     # filter to pre_trial stores with full observation
 storewithfullobs <-  kpsovertime %>%
   group_by(STORE_NBR) %>%
   summarise(nMonthsoveryear =n()) %>%
    filter (nMonthsoveryear == 12) %>%
      pull (STORE_NBR)
```

```
#the trial had started on 02 February 2019
 preTrialMeasures <- kpsovertime %>%
   filter (YEARMONTH < 201902, STORE_NBR %in% storewithfullobs)
head(preTrialMeasures)
```

```
## # A tibble: 6 x 7
##   STORE_NBR YEARMONTH REVENUE AVG_PRICE NUMBER_CUSTOMERS nTxperCust nPQYpertxn
##       <int>     <int>   <dbl>     <dbl>            <int>      <dbl>      <dbl>
## 1         1      1807    207.      3.34               52          1       1.19
## 2         1      1808    176.      3.26               43          1       1.26
## 3         1      1809    279.      3.72               62          1       1.21
## 4         1      1810    188.      3.24               45          1       1.29
## 5         1      1811    193.      3.38               47          1       1.21
## 6         1      1812    190.      3.33               47          1       1.21
```

```
 ### correlation function

 calulatecorrelation <- function (inpuTable, metricol, store_comparison) {
   trial_valuecorrelation <- inpuTable %>%
     filter (STORE_NBR == store_comparison) %>%
     select (YEARMONTH, trial_cor_metric = {{metricol}})

    others_store <- inpuTable %>%
      filter (STORE_NBR != store_comparison) %>%
      select (STORE_NBR,YEARMONTH, control_cor_metric = {{metricol}}) %>%
      inner_join (trial_valuecorrelation, by = "YEARMONTH") %>%
      group_by (STORE_NBR) %>%
      summarise (corr_measure = cor(control_cor_metric, trial_cor_metric, use = "complete.obs"),
```

```
                .groups = "drop") %>%
      mutate(STORE1 = store_comparison, STORE2 = STORE_NBR) %>%
      select (STORE1, STORE2, corr_measure)

    }
```

```
  ### Magnitude distance function
 calculateMagnitudeDistance <- function(inputTable, metricCol, store_comparison) {
      # Step 1: Trial store metric values
   trial_values <- inputTable %>%
     filter(STORE_NBR == store_comparison) %>%
     select(YEARMONTH, trial_metric = {{metricCol}})

       # Step 2: Control stores
   other_stores <- inputTable %>%
     filter(STORE_NBR != store_comparison) %>%
     select(STORE_NBR, YEARMONTH, control_metric = {{metricCol}}) %>%
     inner_join(trial_values, by = "YEARMONTH") %>%
     mutate(diff = abs(control_metric - trial_metric)) %>%
     group_by(YEARMONTH) %>%
     mutate(minDist = min(diff), maxDist = max(diff)) %>%
     ungroup() %>%
     mutate(MagnitudeMeasure = 1 - (diff - minDist) / (maxDist - minDist)) %>%
     group_by(STORE1 = store_comparison, STORE2 = STORE_NBR) %>%
     summarise(mag_measure = mean(MagnitudeMeasure), .groups = "drop")

   return(other_stores)
 }
```

```
  ## Applying functions on  store 77
```

```
  trial_store77 <- 77
corr_nsales <- calulatecorrelation(preTrialMeasures,REVENUE, trial_store77)
print(corr_nsales)
```

```
## # A tibble: 259 x 3
##    STORE1 STORE2 corr_measure
##     <dbl>  <int>        <dbl>
##  1     77      1       0.0644
##  2     77      2       0.262
##  3     77      3       0.163
##  4     77      4      -0.295
##  5     77      5      -0.285
##  6     77      6      -0.0213
##  7     77      7      -0.159
##  8     77      8      -0.0939
##  9     77      9      -0.329
## 10     77     10      -0.424
## # i 249 more rows
```

```
corr_ncustomers <- calulatecorrelation(preTrialMeasures,NUMBER_CUSTOMERS, trial_store77)
print(corr_ncustomers)
```

```
## # A tibble: 259 x 3
##    STORE1 STORE2 corr_measure
##     <dbl>  <int>        <dbl>
##  1     77      1        0.342
##  2     77      2        0.311
##  3     77      3        0.439
##  4     77      4       -0.205
##  5     77      5       -0.339
##  6     77      6        0.208
##  7     77      7        0.150
##  8     77      8        0.155
##  9     77      9       -0.264
## 10     77     10       -0.218
## # i 249 more rows
```

```
magnitude_nsales <- calculateMagnitudeDistance(preTrialMeasures,REVENUE, trial_store77)
print(magnitude_nsales)
```

```
## # A tibble: 259 x 3
##    STORE1 STORE2 mag_measure
##     <dbl>  <int>       <dbl>
##  1     77      1       0.952
##  2     77      2       0.934
##  3     77      3       0.358
##  4     77      4       0.239
##  5     77      5       0.577
##  6     77      6       0.965
##  7     77      7       0.377
##  8     77      8       0.906
##  9     77      9       0.938
## 10     77     10       0.524
## # i 249 more rows
```

```
magnitude_ncustomers <- calculateMagnitudeDistance(preTrialMeasures, NUMBER_CUSTOMERS, trial_store77)
print(magnitude_ncustomers)
```

```
## # A tibble: 259 x 3
##    STORE1 STORE2 mag_measure
##     <dbl>  <int>       <dbl>
##  1     77      1       0.956
##  2     77      2       0.946
##  3     77      3       0.388
##  4     77      4       0.274
##  5     77      5       0.478
##  6     77      6       0.949
##  7     77      7       0.399
##  8     77      8       0.943
##  9     77      9       0.927
## 10     77     10       0.413
## # i 249 more rows
```

```
 ### composite score
corr_weight <- 0.5

score_nSales <- corr_nsales %>%
  inner_join(magnitude_nsales, by = c("STORE1", "STORE2")) %>%
  mutate(scoreNsales = corr_weight * corr_measure + (1- corr_weight) * mag_measure )
score_ncustomers <- corr_ncustomers %>%
  inner_join(magnitude_ncustomers, by = c("STORE1", "STORE2")) %>%
  mutate(scoreNcust = corr_weight * corr_measure + (1 - corr_weight) * mag_measure)
```

```
 ### Final control score
score_control <- score_nSales %>%
  select(STORE1, STORE2, scoreNsales) %>%
  inner_join(score_ncustomers, by = c("STORE1", "STORE2")) %>%
  mutate(finalcontrolscore = 0.5 * scoreNsales + 0.5 * scoreNcust)
print(score_control)
```

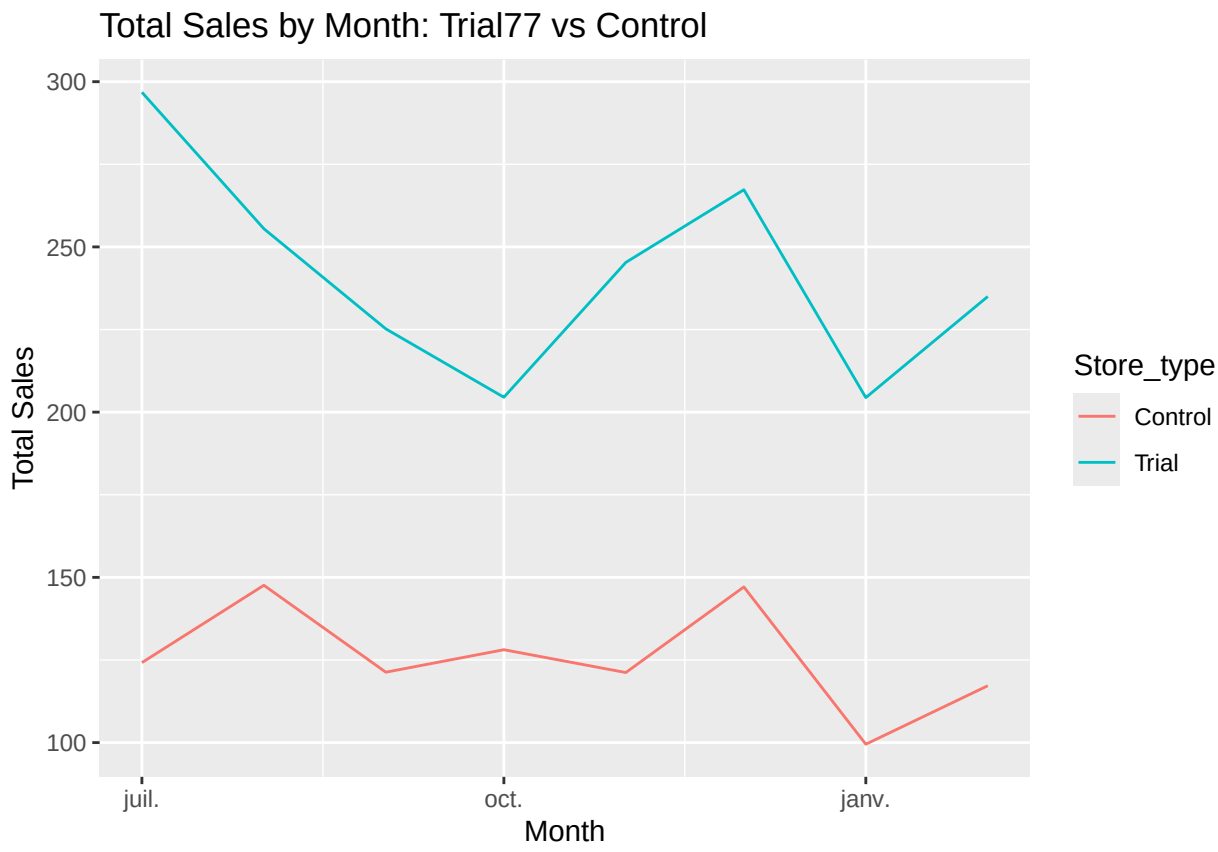
```
## # A tibble: 259 x 7
##    STORE1 STORE2 scoreNsales corr_measure mag_measure scoreNcust
##     <dbl>  <int>       <dbl>        <dbl>       <dbl>      <dbl>
##  1     77      1      0.508         0.342       0.956     0.649
##  2     77      2      0.598         0.311       0.946     0.628
##  3     77      3      0.260         0.439       0.388     0.414
##  4     77      4     -0.0278       -0.205       0.274     0.0345
##  5     77      5      0.146        -0.339       0.478     0.0695
##  6     77      6      0.472         0.208       0.949     0.578
##  7     77      7      0.109         0.150       0.399     0.274
##  8     77      8      0.406         0.155       0.943     0.549
##  9     77      9      0.304        -0.264       0.927     0.332
## 10     77     10      0.0495       -0.218       0.413     0.0975
## # i 249 more rows
## # i 1 more variable: finalcontrolscore <dbl>
```

```
###select best matching control store
control_store <- score_control %>%
  arrange(desc (finalcontrolscore)) %>%
  filter (STORE2 != trial_store77) %>%
  slice(1) %>%
  pull (STORE2)
print(control_store)
```

```
## [1] 35
```

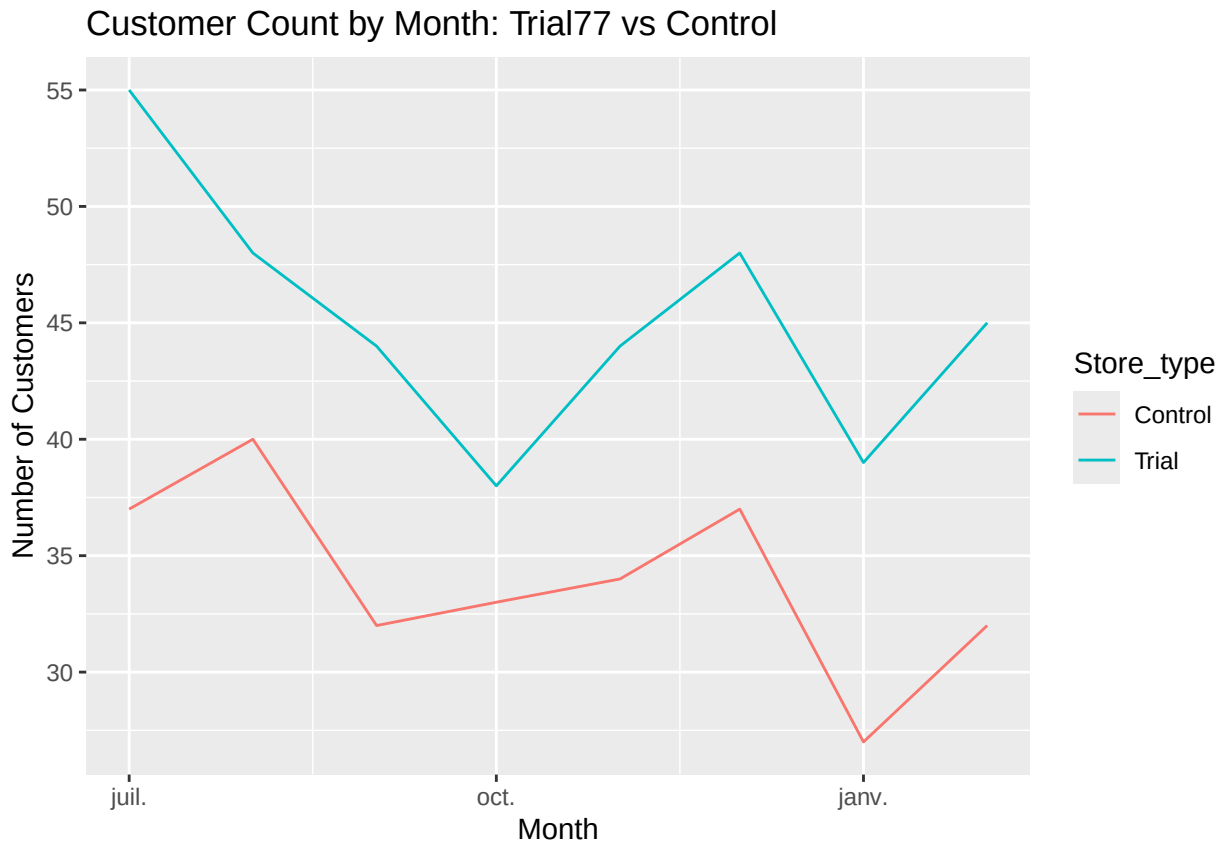
```
### visual check REVENUE

measureOverTimeSales <- kpsovertime %>%
  mutate(Store_type = case_when(
    STORE_NBR == trial_store77 ~ "Trial",
    STORE_NBR == control_store ~ "Control",
    TRUE ~ "Other"
  )) %>%
  filter(Store_type %in% c("Trial", "Control")) %>%
```

```
  group_by(YEARMONTH, Store_type) %>%
  summarise(totSales = mean(REVENUE), .groups = "drop") %>%
  mutate(TransactionMonth = as.Date(paste0("20", YEARMONTH, "01"), format = "%Y%m%d")) %>%
  filter(YEARMONTH < 1903)

ggplot(measureOverTimeSales, aes(TransactionMonth, totSales, color = Store_type)) +
  geom_line() +
  labs(x = "Month", y = "Total Sales", title = "Total Sales by Month: Trial77 vs Control")
```




```
### visual check nCustoment

measureOverTimeCusts <-  kpsovertime %>%
  mutate(Store_type = case_when(
    STORE_NBR == trial_store77 ~ "Trial",
    STORE_NBR == control_store ~ "Control",
    TRUE ~ "Other"
  )) %>%
  filter(Store_type %in% c("Trial", "Control")) %>%
  group_by(YEARMONTH, Store_type) %>%
  summarise(nCustomers = mean( NUMBER_CUSTOMERS), .groups = "drop") %>%
  mutate(
    TransactionMonth = as.Date(paste0("20", YEARMONTH, "01"), format = "%Y%m%d")
  ) %>%
  filter(YEARMONTH < 1903)
```

```
ggplot(measureOverTimeCusts, aes(TransactionMonth, nCustomers, color = Store_type)) +
  geom_line() +
  labs(x = "Month", y = "Number of Customers", title = "Customer Count by Month: Trial77 vs Control")
```



Assessment of trial on customers

```
# pretrial data
### sum pre-trial sales for the trial store
trial_customers_total <-  preTrialMeasures %>%
  filter (STORE_NBR == trial_store77, YEARMONTH < 201902) %>%
  summarise (trial_CUSTOMERS = sum(NUMBER_CUSTOMERS)) %>%
  pull (trial_CUSTOMERS)

### sum pre-trial sales for the control store
control_sales_total <-  preTrialMeasures %>%
  filter (STORE_NBR == control_store, YEARMONTH < 201902) %>%
  summarise (control_CUSTOMERS = sum(NUMBER_CUSTOMERS))%>%
  pull (control_CUSTOMERS)
```

```
# calculate scaling factore for custumers
scalingFactorForCustomer <-  trial_customers_total / control_sales_total
# Application of the scaling factor to the control stores customers data

measureOverTimeCusts <-  kpsovertime
scaledcontroleCusts <- measureOverTimeCusts %>%
  filter(STORE_NBR == control_store) %>%
```

```
  mutate(controlCusts = NUMBER_CUSTOMERS*scalingFactorForCustomer)
print(measureOverTimeCusts)
```

```
## # A tibble: 3,169 x 7
##    STORE_NBR YEARMONTH REVENUE AVG_PRICE NUMBER_CUSTOMERS nTxperCust nPQYpertxn
##        <int>     <int>   <dbl>     <dbl>            <int>      <dbl>      <dbl>
##  1         1      1807    207.      3.34               52          1       1.19
##  2         1      1808    176.      3.26               43          1       1.26
##  3         1      1809    279.      3.72               62          1       1.21
##  4         1      1810    188.      3.24               45          1       1.29
##  5         1      1811    193.      3.38               47          1       1.21
##  6         1      1812    190.      3.33               47          1       1.21
##  7         1      1901    155.      3.69               36          1       1.17
##  8         1      1902    225.      3.47               55          1       1.18
##  9         1      1903    193.      3.33               49          1       1.18
## 10         1      1904    193.      3.38               43          1       1.33
## # i 3,159 more rows
```

```
### Calculate percentage difference during trial

percentagediff <- scaledcontroleCusts  %>%
  select (YEARMONTH, controlCusts) %>%
  inner_join(
    kpsovertime %>%
      filter(STORE_NBR == trial_store77) %>%
      select (YEARMONTH, trial_CUSTOMERS = NUMBER_CUSTOMERS),
    by = "YEARMONTH") %>%
  mutate(percentagediff = (trial_CUSTOMERS - controlCusts)/controlCusts)
print(percentagediff)
```

```
## # A tibble: 12 x 4
##    YEARMONTH controlCusts trial_CUSTOMERS percentagediff
##        <int>        <dbl>           <int>          <dbl>
##  1      1807         49.0              55        0.121
##  2      1808         53.0              48       -0.0947
##  3      1809         42.4              44        0.0374
##  4      1810         43.7              38       -0.131
##  5      1811         45.1              44       -0.0237
##  6      1812         49.0              48       -0.0213
##  7      1901         35.8              39        0.0898
##  8      1902         42.4              45        0.0609
##  9      1903         51.7              55        0.0640
## 10      1904         53.0              48       -0.0947
## 11      1905         55.7              56        0.00593
## 12      1906         41.1              42        0.0222
```

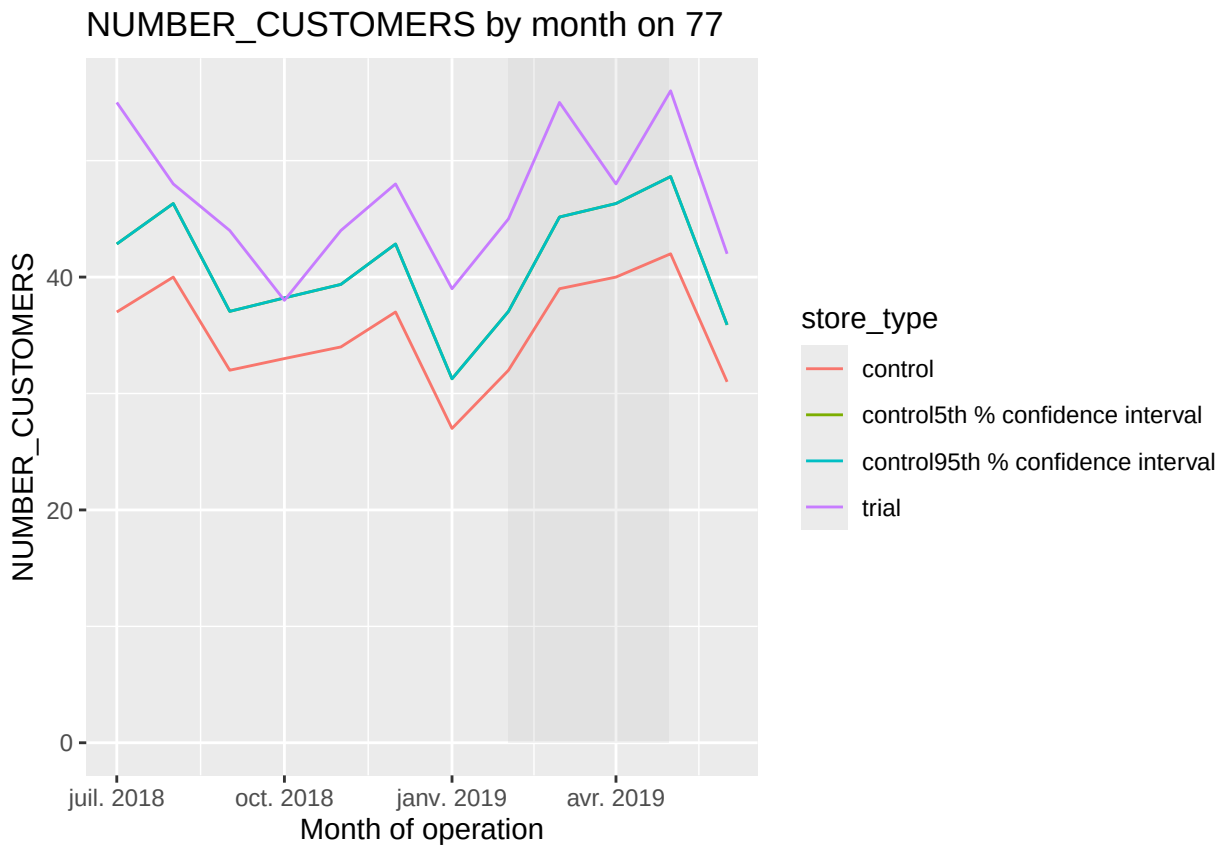
```
### Calculate t-values for trial month
stdDev <-  percentagediff %>%
  filter (YEARMONTH < 201902) %>%
  pull ( percentagediff) %>%
  sd()
```

```r
degreesofFreedom <- 7

percentagediff77 <- percentagediff  %>%
  mutate(tvalue= percentagediff /stdDev,
         TransactionMonth = as.Date(paste0("20", YEARMONTH, "01"), format = "%Y%m%d"))
print (percentagediff77)
```

```
## # A tibble: 12 x 6
##    YEARMONTH controlCusts trial_CUSTOMERS percentagediff  tvalue
##        <int>        <dbl>           <int>          <dbl>   <dbl>
##  1      1807         49.0              55         0.121    1.54
##  2      1808         53.0              48        -0.0947  -1.20
##  3      1809         42.4              44         0.0374   0.473
##  4      1810         43.7              38        -0.131   -1.66
##  5      1811         45.1              44        -0.0237  -0.300
##  6      1812         49.0              48        -0.0213  -0.269
##  7      1901         35.8              39         0.0898   1.14
##  8      1902         42.4              45         0.0609   0.772
##  9      1903         51.7              55         0.0640   0.810
## 10      1904         53.0              48        -0.0947  -1.20
## 11      1905         55.7              56         0.00593  0.0751
## 12      1906         41.1              42         0.0222   0.281
## # i 1 more variable: TransactionMonth <date>
```

```r
### create plot : total sales and confidence intervals

measureOverTimeCusts <- kpsovertime

pastCusts <- measureOverTimeCusts %>%
  mutate(
    store_type = case_when(
      STORE_NBR == trial_store77 ~ "trial",
      STORE_NBR == control_store ~ "control",
      TRUE ~ "other"),
    TransactionMonth = as.Date(paste0("20", YEARMONTH, "01"), format = "%Y%m%d")
  ) %>%
  filter(store_type %in% c("trial", "control"))

#95th percentile : upper limit of what we expect if there's no true effect
pastCusts_controls95 <- pastCusts %>%
  filter ( store_type == "control") %>%
  mutate ( NUMBER_CUSTOMERS = NUMBER_CUSTOMERS * (1 + stdDev *2),
           store_type = "control95th % confidence interval")


#5th percentile : upper limit of what we expect if there's no true effect
pastCusts_controls5 <-pastCusts %>%
  filter ( store_type == "control") %>%
  mutate (NUMBER_CUSTOMERS = NUMBER_CUSTOMERS * (1 + stdDev *2),
           store_type = "control5th % confidence interval")
```

```r
trialAssessment <- bind_rows(pastCusts,pastCusts_controls95, pastCusts_controls5) %>%
   filter(!is.na(NUMBER_CUSTOMERS))


trial_period <- data.frame(
   xmin = as.Date("2019-02-01"),
   xmax = as.Date("2019-04-30"),
   ymin = 0,
   ymax = Inf
 )

 ggplot(trialAssessment, aes(TransactionMonth, NUMBER_CUSTOMERS, color = store_type)) +
   geom_rect(
     data = trial_period,
     aes(xmin = xmin, xmax = xmax, ymin = ymin, ymax = ymax),
     fill = "grey", alpha = 0.2, inherit.aes = FALSE
   ) +
   geom_line() +
   labs(x = "Month of operation", y = "NUMBER_CUSTOMERS", title = "NUMBER_CUSTOMERS by month on 77 ")
```



The number of customers in trial store (77) showed similar trends to the control store before the trial. During the trial period (Fed-Apr 2019), the number of customers in the trial increased and remained within the confidence interval. The control suggesting not a potentially significant impact of the store layout. The customer numbers fell after trial that proved a no lasting effect.

```r
### Applying functions on  store 86
```

```r
trial_store86 <- 86
corr_nsales <- calulatecorrelation(preTrialMeasures,REVENUE, trial_store86)
print(corr_nsales)
```

```
## # A tibble: 259 x 3
##    STORE1 STORE2 corr_measure
##     <dbl>  <int>        <dbl>
##  1     86      1      0.344
##  2     86      2      0.00110
##  3     86      3      0.178
##  4     86      4     -0.179
##  5     86      5     -0.322
##  6     86      6      0.211
##  7     86      7      0.347
##  8     86      8      0.220
##  9     86      9      0.212
## 10     86     10      0.184
## # i 249 more rows
```

```r
corr_ncustomers <- calulatecorrelation(preTrialMeasures,NUMBER_CUSTOMERS, trial_store86)
print(corr_ncustomers)
```

```
## # A tibble: 259 x 3
##    STORE1 STORE2 corr_measure
##     <dbl>  <int>        <dbl>
##  1     86      1       0.305
##  2     86      2      -0.0445
##  3     86      3       0.0262
##  4     86      4      -0.239
##  5     86      5      -0.576
##  6     86      6      -0.0101
##  7     86      7       0.0920
##  8     86      8       0.298
##  9     86      9      -0.0710
## 10     86     10       0.0762
## # i 249 more rows
```

```r
magnitude_nsales <- calculateMagnitudeDistance(preTrialMeasures,REVENUE, trial_store86)
print (magnitude_nsales)
```

```
## # A tibble: 259 x 3
##    STORE1 STORE2 mag_measure
##     <dbl>  <int>       <dbl>
##  1     86      1       0.220
##  2     86      2       0.184
##  3     86      3       0.791
##  4     86      4       0.606
##  5     86      5       0.878
##  6     86      6       0.248
##  7     86      7       0.821
```

```
##  8     86      8       0.421
##  9     86      9       0.374
## 10     86     10       0.917
## # i 249 more rows
```

```
 magnitude_ncustomers <- calculateMagnitudeDistance(preTrialMeasures, NUMBER_CUSTOMERS, trial_store86)
print (magnitude_ncustomers)
```

```
## # A tibble: 259 x 3
##    STORE1 STORE2 mag_measure
##     <dbl>  <int>       <dbl>
##  1     86      1       0.367
##  2     86      2       0.323
##  3     86      3       0.916
##  4     86      4       0.845
##  5     86      5       0.874
##  6     86      6       0.340
##  7     86      7       0.918
##  8     86      8       0.332
##  9     86      9       0.426
## 10     86     10       0.902
## # i 249 more rows
```

```
### composite score
 corr_weight <- 0.5

 score_nSales <- corr_nsales %>%
   inner_join(magnitude_nsales, by = c("STORE1", "STORE2")) %>%
   mutate(scoreNsales = corr_weight * corr_measure + (1- corr_weight) * mag_measure )

 score_ncustomers <- corr_ncustomers %>%
   inner_join(magnitude_ncustomers, by = c("STORE1", "STORE2")) %>%
   mutate(scoreNcust = corr_weight * corr_measure + (1 - corr_weight) * mag_measure)
```

```
 ### Final control score
 score_control <- score_nSales %>%
   select(STORE1, STORE2, scoreNsales) %>%
   inner_join(score_ncustomers, by = c("STORE1", "STORE2")) %>%
   mutate(finalcontrolscore = 0.5 * scoreNsales + 0.5 * scoreNcust)
print(score_control)
```

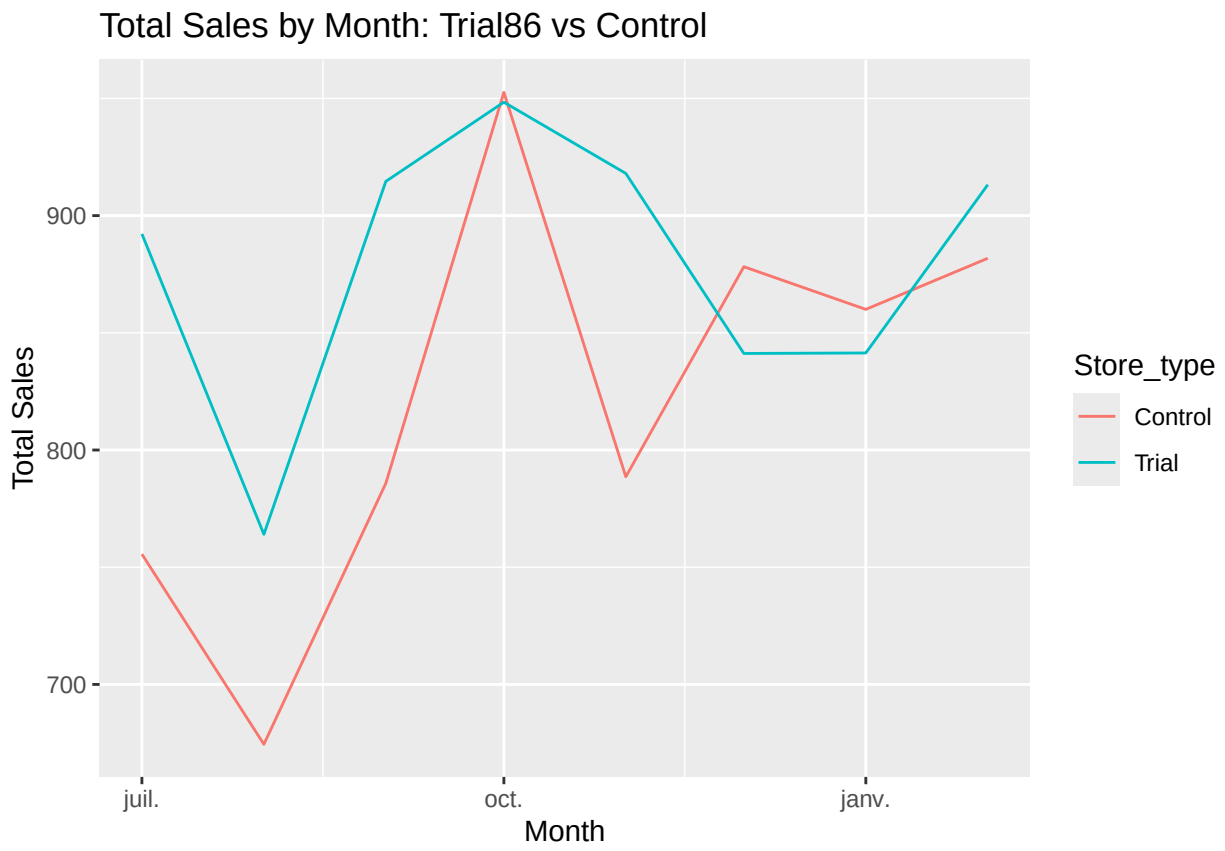
```
## # A tibble: 259 x 7
##    STORE1 STORE2 scoreNsales corr_measure mag_measure scoreNcust
##     <dbl>  <int>       <dbl>        <dbl>       <dbl>      <dbl>
##  1     86      1      0.282         0.305       0.367      0.336
##  2     86      2      0.0925       -0.0445      0.323      0.139
##  3     86      3      0.484         0.0262      0.916      0.471
##  4     86      4      0.213        -0.239       0.845      0.303
##  5     86      5      0.278        -0.576       0.874      0.149
##  6     86      6      0.229        -0.0101      0.340      0.165
##  7     86      7      0.584         0.0920      0.918      0.505
##  8     86      8      0.320         0.298       0.332      0.315
```

```
##  9     86      9      0.293       -0.0710       0.426      0.178
## 10     86     10      0.550        0.0762       0.902      0.489
## # i 249 more rows
## # i 1 more variable: finalcontrolscore <dbl>
```

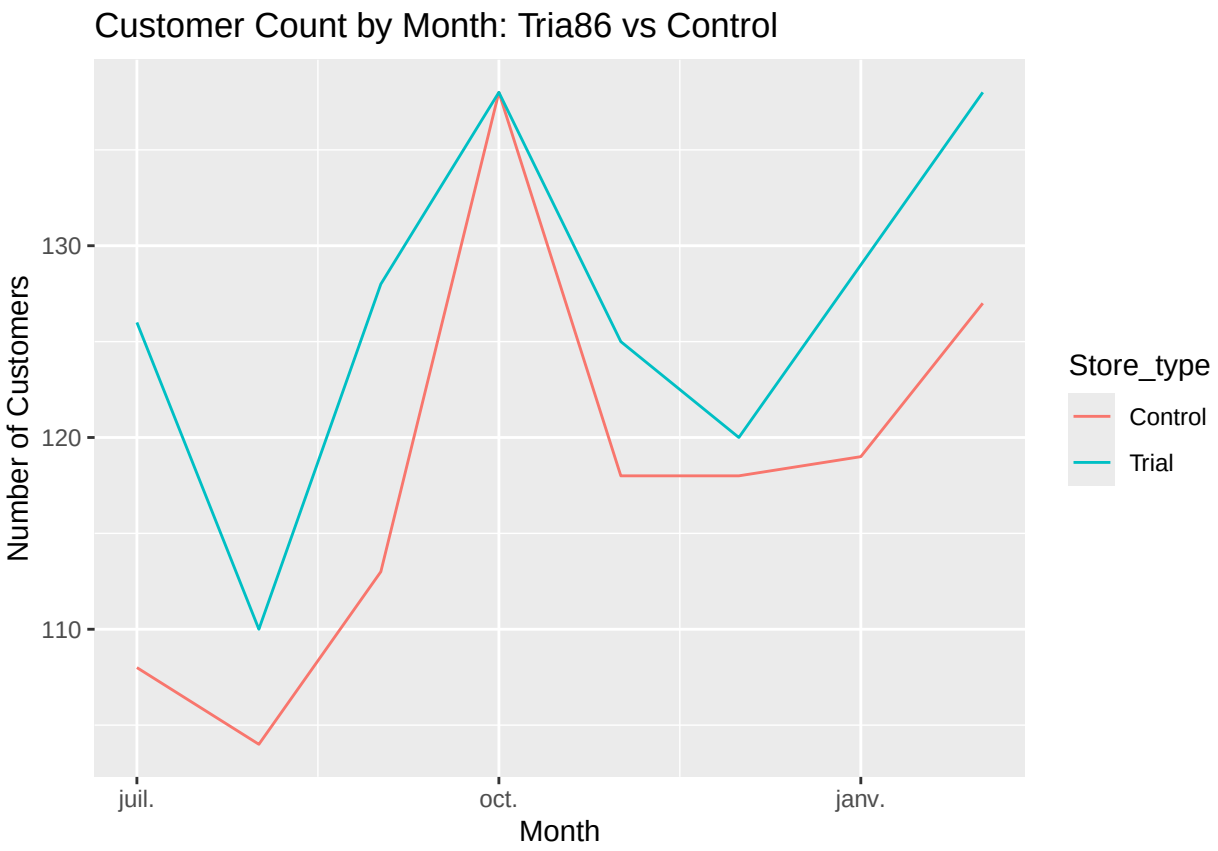
```
###select best matching control store
 control_store <- score_control %>%
   arrange(desc (finalcontrolscore)) %>%
   filter (STORE2 != trial_store86) %>%
   slice(1) %>%
   pull (STORE2)
print(control_store)
```

```
## [1] 147
```

```
 ### visual check REVENUE

 measureOverTimeSales <- kpsovertime %>%
   mutate(Store_type = case_when(
     STORE_NBR == trial_store86 ~ "Trial",
     STORE_NBR == control_store ~ "Control",
     TRUE ~ "Other"
   )) %>%
   filter(Store_type %in% c("Trial", "Control")) %>%
   group_by(YEARMONTH, Store_type) %>%
   summarise(totSales = mean(REVENUE), .groups = "drop") %>%
   mutate(TransactionMonth = as.Date(paste0("20", YEARMONTH, "01"), format = "%Y%m%d")) %>%
   filter(YEARMONTH < 1903)

 ggplot(measureOverTimeSales, aes(TransactionMonth, totSales, color = Store_type)) +
   geom_line() +
   labs(x = "Month", y = "Total Sales", title = "Total Sales by Month: Trial86 vs Control")
```




```
### visual check nCustoment

measureOverTimeCusts <-  kpsovertime %>%
  mutate(Store_type = case_when(
    STORE_NBR == trial_store86 ~ "Trial",
    STORE_NBR == control_store ~ "Control",
    TRUE ~ "Other"
  )) %>%
  filter(Store_type %in% c("Trial", "Control")) %>%
  group_by(YEARMONTH, Store_type) %>%
  summarise(nCustomers = mean( NUMBER_CUSTOMERS), .groups = "drop") %>%
  mutate(
    TransactionMonth = as.Date(paste0("20", YEARMONTH, "01"), format = "%Y%m%d")
  ) %>%
  filter(YEARMONTH < 1903)

ggplot(measureOverTimeCusts, aes(TransactionMonth, nCustomers, color = Store_type)) +
  geom_line() +
  labs(x = "Month", y = "Number of Customers", title = "Customer Count by Month: Tria86 vs Control")
```




```
## Assessment of trial on customers

 # pretrial data
 ### sum pre-trial sales for the trial store
 trial_customers_total <-  preTrialMeasures %>%
   filter (STORE_NBR == trial_store86, YEARMONTH < 201902) %>%
   summarise (trial_CUSTOMERS = sum(NUMBER_CUSTOMERS)) %>%
   pull (trial_CUSTOMERS)

 ### sum pre-trial sales for the control store
 control_sales_total <-  preTrialMeasures %>%
   filter (STORE_NBR == control_store, YEARMONTH < 201902) %>%
   summarise (control_CUSTOMERS = sum(NUMBER_CUSTOMERS))%>%
   pull (control_CUSTOMERS)
```

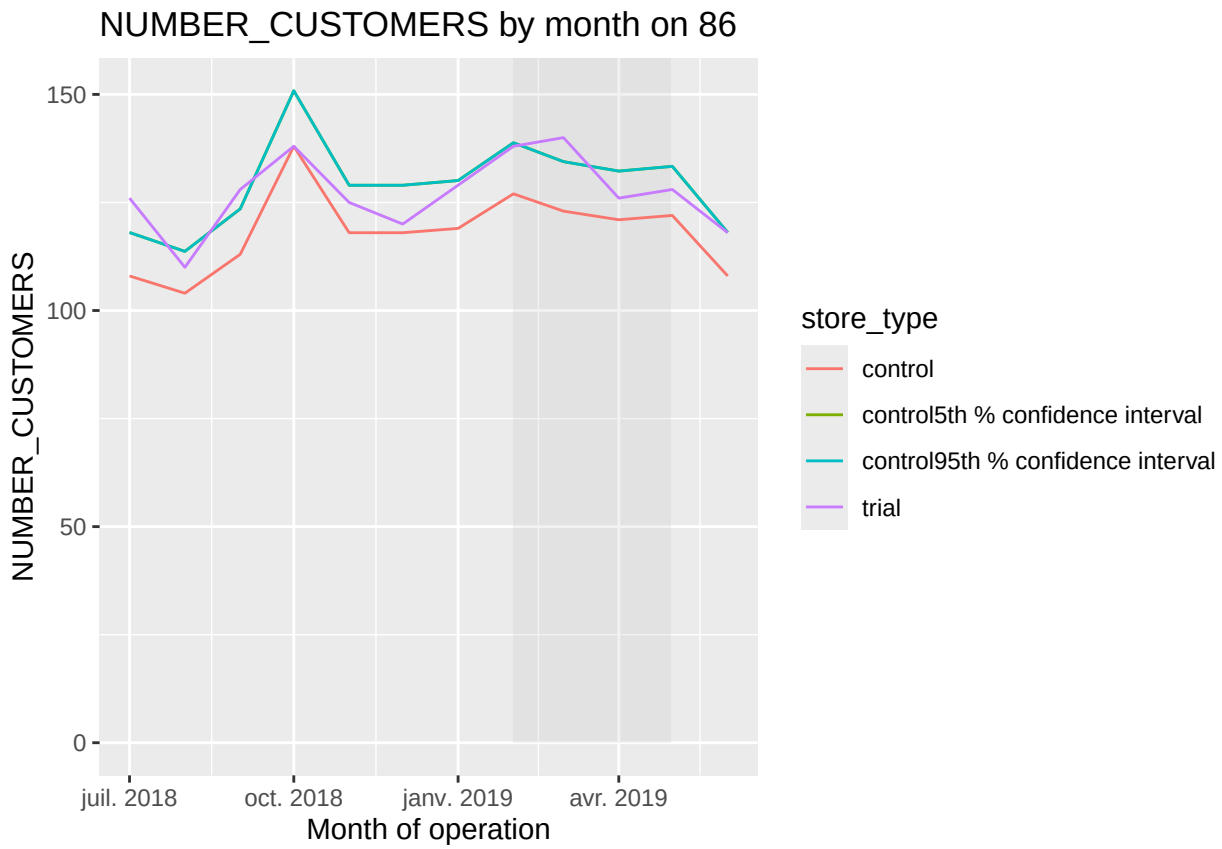
```
# calculate scaling factore for custumers
 scalingFactorForCustomer <-  trial_customers_total / control_sales_total
 # Application of thescaling factor to the control stores sales data

 measureOverTimeCusts <-  kpsovertime
 scaledcontroleCusts <- measureOverTimeCusts %>%
   filter(STORE_NBR == control_store) %>%
   mutate(controlCusts = NUMBER_CUSTOMERS* scalingFactorForCustomer)
```

```
### Calculate percentage difference during trial

percentagediff86 <- scaledcontroleCusts  %>%
  select (YEARMONTH, controlCusts) %>%
  inner_join(
    kpsovertime %>%
      filter(STORE_NBR == trial_store86) %>%
      select (YEARMONTH, trial_CUSTOMERS = NUMBER_CUSTOMERS),
    by = "YEARMONTH") %>%
  mutate(percentagediff86 = (trial_CUSTOMERS - controlCusts)/controlCusts)
```

```
### Calculate t-values for trial month
stdDev <-  percentagediff86 %>%
  filter (YEARMONTH < 201902) %>%
  pull ( percentagediff86) %>%
  sd()

degreesofFreedom <- 7

percentagediff86 <- percentagediff86  %>%
  mutate(tvalue= percentagediff86 /stdDev,
         TransactionMonth = as.Date(paste0("20", YEARMONTH, "01"), format = "%Y%m%d"))
```

```
### create plot : total sales and confidence intervals

measureOverTimeCusts <- kpsovertime

pastCusts <- measureOverTimeCusts %>%
  mutate(
    store_type = case_when(
      STORE_NBR == trial_store86 ~ "trial",
      STORE_NBR == control_store ~ "control",
      TRUE ~ "other"),
    TransactionMonth = as.Date(paste0("20", YEARMONTH, "01"), format = "%Y%m%d")
  ) %>%
  filter(store_type %in% c("trial", "control"))

#95th percentile : upper limit of what we expect if there's no true effect
pastCusts_controls95 <- pastCusts %>%
  filter ( store_type == "control") %>%
  mutate ( NUMBER_CUSTOMERS = NUMBER_CUSTOMERS * (1 + stdDev *2),
           store_type = "control95th % confidence interval")
#5th percentile : upper limit of what we expect if there's no true effect
pastCusts_controls5 <- pastCusts %>%
  filter ( store_type == "control") %>%
  mutate (NUMBER_CUSTOMERS = NUMBER_CUSTOMERS * (1 + stdDev *2),
          store_type = "control5th % confidence interval")

trialAssessment <- bind_rows(pastCusts,pastCusts_controls95, pastCusts_controls5) %>%
  filter(!is.na(NUMBER_CUSTOMERS))

trial_period <- data.frame(
  xmin = as.Date("2019-02-01"),
```

```
  xmax = as.Date("2019-04-30"),
  ymin = 0,
  ymax = Inf
)

ggplot(trialAssessment, aes(TransactionMonth, NUMBER_CUSTOMERS, color = store_type)) +
  geom_rect(
    data = trial_period,
    aes(xmin = xmin, xmax = xmax, ymin = ymin, ymax = ymax),
    fill = "grey", alpha = 0.2, inherit.aes = FALSE
  ) +
  geom_line() +
  labs(x = "Month of operation", y = "NUMBER_CUSTOMERS", title = "NUMBER_CUSTOMERS by month on 86")
```



The number of customers in trial store (86) showed similar trends to the control store before the trial.During the trial period (Feb-April 2019), the number of customers in the trial store increased. this increase appears to be above the 95% confidence interval of the control, suggesting a potentially significant impact of the new store layout.However, the trend trial store goes down after the trial period and indicating a no lasting effect.

```
### Applying functions on  store 88
```

```
trial_store88 <- 88
 corr_nsales <- calulatecorrelation(preTrialMeasures,REVENUE, trial_store88)
 print(corr_nsales)
```

```
## # A tibble: 259 x 3
##    STORE1 STORE2 corr_measure
##     <dbl>  <int>        <dbl>
##  1     88      1       0.371
##  2     88      2       0.284
##  3     88      3      -0.208
##  4     88      4      -0.320
##  5     88      5      -0.325
##  6     88      6      -0.311
##  7     88      7       0.444
##  8     88      8      -0.195
##  9     88      9      -0.140
## 10     88     10      -0.0938
## # i 249 more rows
```

```
corr_ncustomers <- calulatecorrelation(preTrialMeasures,NUMBER_CUSTOMERS, trial_store88)
 print(corr_ncustomers)
```

```
## # A tibble: 259 x 3
##    STORE1 STORE2 corr_measure
##     <dbl>  <int>        <dbl>
##  1     88      1       0.293
##  2     88      2       0.296
##  3     88      3      -0.0628
##  4     88      4      -0.186
##  5     88      5      -0.638
##  6     88      6       0.0368
##  7     88      7       0.483
##  8     88      8       0.200
##  9     88      9      -0.224
## 10     88     10      -0.136
## # i 249 more rows
```

```
magnitude_nsales <- calculateMagnitudeDistance(preTrialMeasures,REVENUE, trial_store88)
 print(magnitude_nsales)
```

```
## # A tibble: 259 x 3
##    STORE1 STORE2 mag_measure
##     <dbl>  <int>       <dbl>
##  1     88      1       0.144
##  2     88      2       0.120
##  3     88      3       0.794
##  4     88      4       0.864
##  5     88      5       0.587
##  6     88      6       0.163
##  7     88      7       0.773
##  8     88      8       0.276
##  9     88      9       0.245
## 10     88     10       0.637
## # i 249 more rows
```

```r
magnitude_ncustomers <- calculateMagnitudeDistance(preTrialMeasures, NUMBER_CUSTOMERS, trial_store88)
print(magnitude_ncustomers)
```

```
## # A tibble: 259 x 3
##    STORE1 STORE2 mag_measure
##     <dbl>  <int>       <dbl>
##  1     88      1       0.302
##  2     88      2       0.265
##  3     88      3       0.806
##  4     88      4       0.872
##  5     88      5       0.736
##  6     88      6       0.279
##  7     88      7       0.796
##  8     88      8       0.273
##  9     88      9       0.351
## 10     88     10       0.788
## # i 249 more rows
```

```r
### composite score
 corr_weight <- 0.5

 score_nSales <- corr_nsales %>%
   inner_join(magnitude_nsales, by = c("STORE1", "STORE2")) %>%
   mutate(scoreNsales = corr_weight * corr_measure + (1- corr_weight) * mag_measure )

 score_ncustomers <- corr_ncustomers %>%
   inner_join(magnitude_ncustomers, by = c("STORE1", "STORE2")) %>%
   mutate(scoreNcust = corr_weight * corr_measure + (1 - corr_weight) * mag_measure)
```

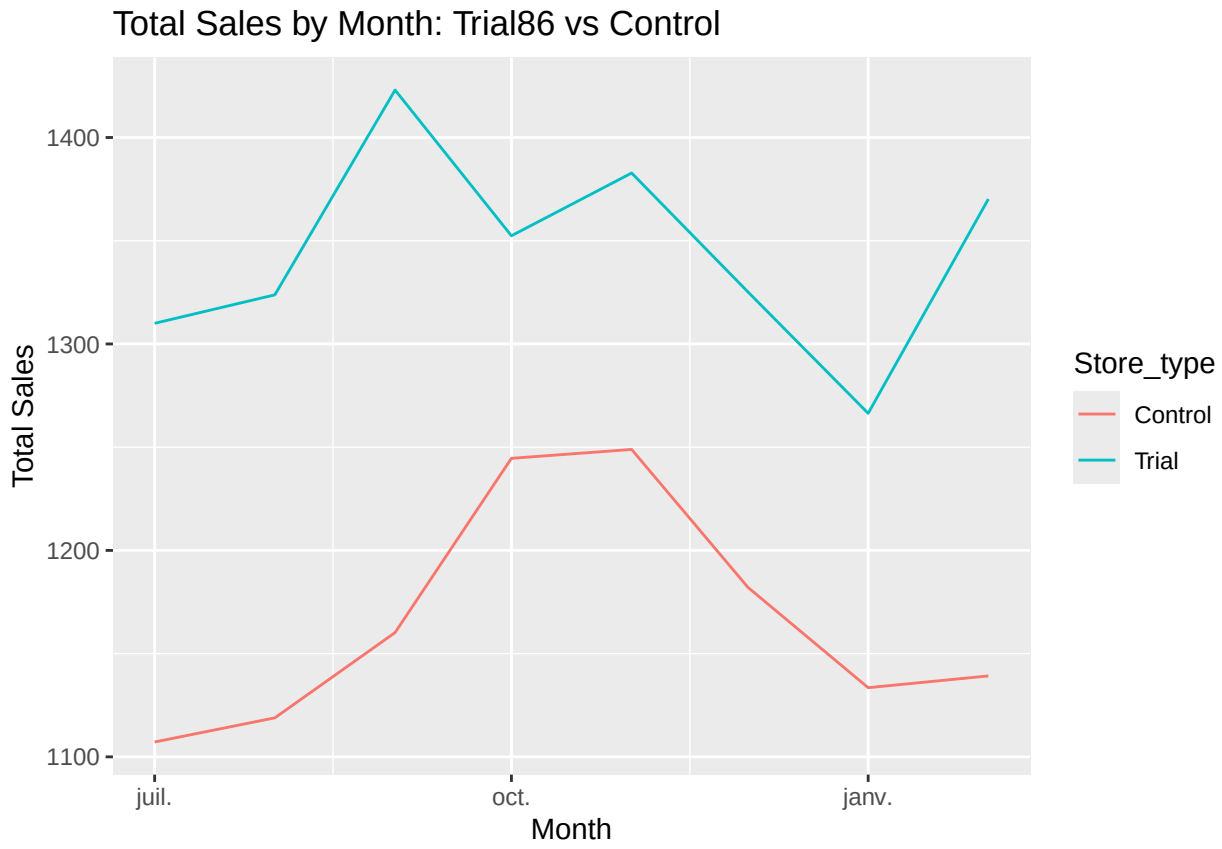
```r
 ### Final control score
 score_control <- score_nSales %>%
   select(STORE1, STORE2, scoreNsales) %>%
   inner_join(score_ncustomers, by = c("STORE1", "STORE2")) %>%
   mutate(finalcontrolscore = 0.5 * scoreNsales + 0.5 * scoreNcust)
```

```r
 ###select best matching control store
 control_store <- score_control %>%
   arrange(desc (finalcontrolscore)) %>%
   filter (STORE2 != trial_store88) %>%
   slice(1) %>%
   pull (STORE2)
print(control_store)
```

```
## [1] 201
```

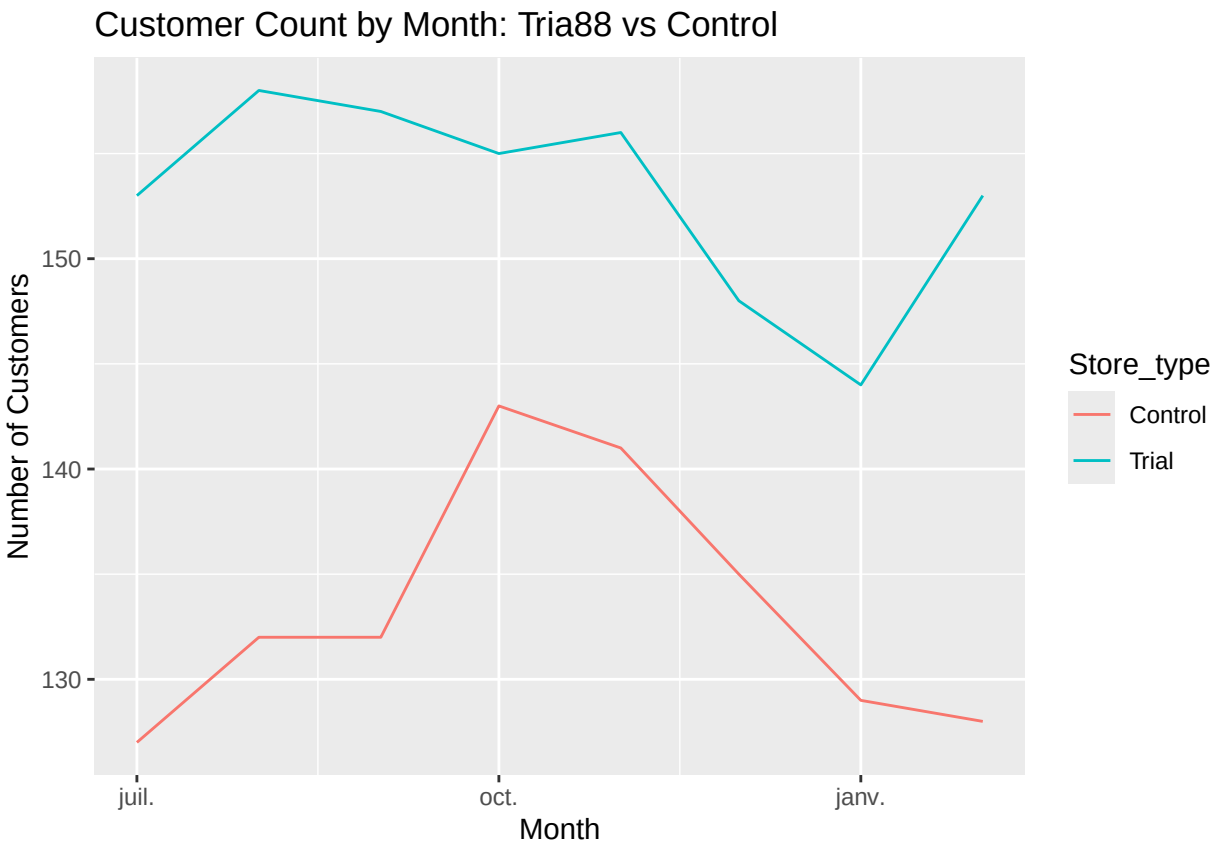
```r
### visual check REVENUE

 measureOverTimeSales <- kpsovertime %>%
   mutate(Store_type = case_when(
     STORE_NBR == trial_store88 ~ "Trial",
     STORE_NBR == control_store ~ "Control",
     TRUE ~ "Other"
```

```
  )) %>%
  filter(Store_type %in% c("Trial", "Control")) %>%
  group_by(YEARMONTH, Store_type) %>%
  summarise(totSales = mean(REVENUE), .groups = "drop") %>%
  mutate(TransactionMonth = as.Date(paste0("20", YEARMONTH, "01"), format = "%Y%m%d")) %>%
  filter(YEARMONTH < 1903)

ggplot(measureOverTimeSales, aes(TransactionMonth, totSales, color = Store_type)) +
  geom_line() +
  labs(x = "Month", y = "Total Sales", title = "Total Sales by Month: Trial86 vs Control")
```




```
### visual check nCustoment

measureOverTimeCusts <-  kpsovertime %>%
  mutate(Store_type = case_when(
    STORE_NBR == trial_store88 ~ "Trial",
    STORE_NBR == control_store ~ "Control",
    TRUE ~ "Other"
  )) %>%
  filter(Store_type %in% c("Trial", "Control")) %>%
  group_by(YEARMONTH, Store_type) %>%
  summarise(nCustomers = mean( NUMBER_CUSTOMERS), .groups = "drop") %>%
  mutate(
    TransactionMonth = as.Date(paste0("20", YEARMONTH, "01"), format = "%Y%m%d")
  ) %>%
```

```
  filter(YEARMONTH < 1903)

ggplot(measureOverTimeCusts, aes(TransactionMonth, nCustomers, color = Store_type)) +
  geom_line() +
  labs(x = "Month", y = "Number of Customers", title = "Customer Count by Month: Tria88 vs Control")
```




```
## Assessment of trial on customers

 # pretrial data
 ### sum pre-trial sales for the trial store
 trial_customers_total <-  preTrialMeasures %>%
   filter (STORE_NBR == trial_store88, YEARMONTH < 201902) %>%
   summarise (trial_CUSTOMERS = sum(NUMBER_CUSTOMERS)) %>%
   pull (trial_CUSTOMERS)

 ### sum pre-trial sales for the control store
 control_sales_total <-  preTrialMeasures %>%
   filter (STORE_NBR == control_store, YEARMONTH < 201902) %>%
   summarise (control_CUSTOMERS = sum(NUMBER_CUSTOMERS))%>%
   pull (control_CUSTOMERS)
```

```
# calculate scaling factore for custumers
 scalingFactorForCustomer <-  trial_customers_total / control_sales_total
 # Application of the scaling factor to the control stores customer data
```

```
measureOverTimeCusts <-  kpsovertime
scaledcontroleCusts <- measureOverTimeCusts %>%
  filter( STORE_NBR == control_store) %>%
  mutate(controlCusts = NUMBER_CUSTOMERS*scalingFactorForCustomer)
print(measureOverTimeCusts)
```

```
## # A tibble: 3,169 x 7
##    STORE_NBR YEARMONTH REVENUE AVG_PRICE NUMBER_CUSTOMERS nTxperCust nPQYpertxn
##        <int>     <int>   <dbl>     <dbl>            <int>      <dbl>      <dbl>
##  1         1      1807    207.      3.34               52          1       1.19
##  2         1      1808    176.      3.26               43          1       1.26
##  3         1      1809    279.      3.72               62          1       1.21
##  4         1      1810    188.      3.24               45          1       1.29
##  5         1      1811    193.      3.38               47          1       1.21
##  6         1      1812    190.      3.33               47          1       1.21
##  7         1      1901    155.      3.69               36          1       1.17
##  8         1      1902    225.      3.47               55          1       1.18
##  9         1      1903    193.      3.33               49          1       1.18
## 10         1      1904    193.      3.38               43          1       1.33
## # i 3,159 more rows
```

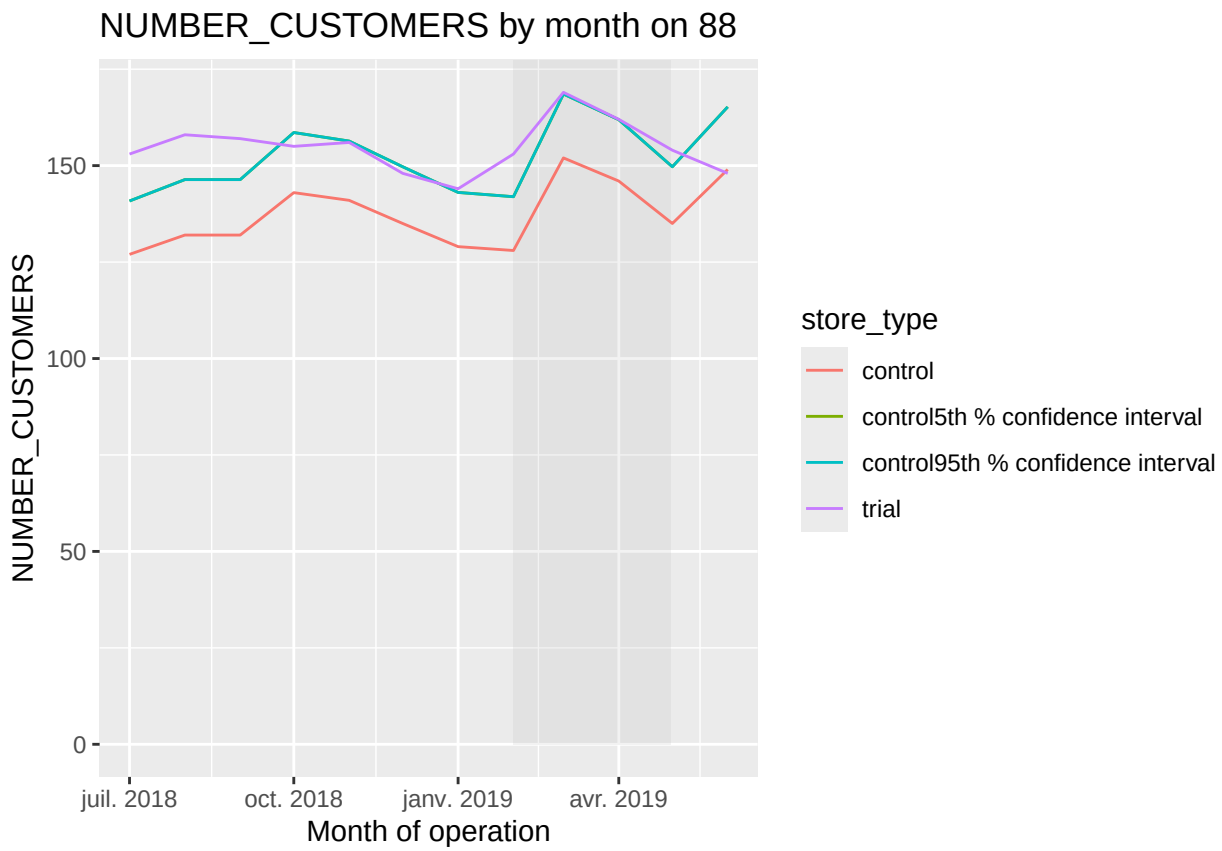
```
### Calculate percentage difference during trial

percentagediff88 <- scaledcontroleCusts  %>%
  select (YEARMONTH, controlCusts) %>%
  inner_join(
    kpsovertime %>%
      filter(STORE_NBR == trial_store88) %>%
      select (YEARMONTH, trial_CUSTOMERS = NUMBER_CUSTOMERS),
    by = "YEARMONTH") %>%
  mutate(percentagediff88 = (trial_CUSTOMERS - controlCusts)/controlCusts)
```

```
### Calculate t-values for trial month
stdDev <-  percentagediff88 %>%
  filter (YEARMONTH < 201902) %>%
  pull ( percentagediff88) %>%
  sd()

degreesofFreedom <- 7

percentagediff88 <- percentagediff88  %>%
  mutate(tvalue= percentagediff88 /stdDev,
         TransactionMonth = as.Date(paste0("20", YEARMONTH, "01"), format = "%Y%m%d"))
print(percentagediff88)
```

```
## # A tibble: 12 x 6
##    YEARMONTH controlCusts trial_CUSTOMERS percentagediff88 tvalue
##        <int>        <dbl>           <int>            <dbl>  <dbl>
##  1      1807         143.             153          0.0698   1.28
##  2      1808         149.             158          0.0629   1.15
##  3      1809         149.             157          0.0562   1.03
##  4      1810         161.             155         -0.0375  -0.688
```

```
##  5      1811         159.             156         -0.0175  -0.322
##  6      1812         152.             148         -0.0265  -0.486
##  7      1901         145.             144         -0.00875 -0.161
##  8      1902         144.             153          0.0614   1.13
##  9      1903         171.             169         -0.0127  -0.233
## 10      1904         164.             162         -0.0147  -0.270
## 11      1905         152.             154          0.0130   0.238
## 12      1906         168.             148         -0.118   -2.17
## # i 1 more variable: TransactionMonth <date>
```

```r
### create plot : total customers and confidence intervals

measureOverTimeCusts <- kpsovertime

pastCusts <- measureOverTimeCusts %>%
  mutate(
    store_type = case_when(
      STORE_NBR == trial_store88 ~ "trial",
      STORE_NBR == control_store ~ "control",
      TRUE ~ "other"),
    TransactionMonth = as.Date(paste0("20", YEARMONTH, "01"), format = "%Y%m%d")
  ) %>%
  filter(store_type %in% c("trial", "control"))

#95th percentile : upper limit of what we expect if there's no true effect
pastCusts_controls95 <- pastCusts %>%
  filter ( store_type == "control") %>%
  mutate ( NUMBER_CUSTOMERS = NUMBER_CUSTOMERS * (1 + stdDev *2),
           store_type = "control95th % confidence interval")
#5th percentile : upper limit of what we expect if there's no true effect
pastCusts_controls5 <- pastCusts %>%
  filter ( store_type == "control") %>%
  mutate (NUMBER_CUSTOMERS = NUMBER_CUSTOMERS * (1 + stdDev *2),
          store_type = "control5th % confidence interval")

trialAssessment <- bind_rows(pastCusts,pastCusts_controls95, pastCusts_controls5) %>%
  filter(!is.na(NUMBER_CUSTOMERS))

trial_period <- data.frame(
  xmin = as.Date("2019-02-01"),
  xmax = as.Date("2019-04-30"),
  ymin = 0,
  ymax = Inf
)

ggplot(trialAssessment, aes(TransactionMonth, NUMBER_CUSTOMERS, color = store_type)) +
  geom_rect(
    data = trial_period,
    aes(xmin = xmin, xmax = xmax, ymin = ymin, ymax = ymax),
    fill = "grey", alpha = 0.2, inherit.aes = FALSE
  ) +
  geom_line() +
  labs(x = "Month of operation", y = "NUMBER_CUSTOMERS", title = "NUMBER_CUSTOMERS by month on 88")
```



The number of customers in the trial store (88) showed similar trends to the control store before the trial Period. During the trial period (Fed-Apr 2019), the number of customers in the trial increased and remained within the confidence interval but it is closed to the upper bound 95% confidence interval . This may suggest a potential increase in customer traffic, the change was not statistically significant at the 95% confidence level. Therefore, we cannot confidently attribute the observed trend to the trial intervention. However, the close proximity to the confidence threshold indicator a potential positive effect that warrants further investigation. The customer numbers go down after the trial period, indicating a no lasting effect.

```
### CONCLUSION
```

The analysis shows the trial store 77 matches to store 115 as control store, trial store 86 to control store 160 and trial store 88 to control store 81. The analysis shows that the number of customers in the trial store (store 77) remained within the 5th and 95 th percentile confidence interval of the control store throughout the trial period (Feb-April 2019). Although the trial store's customer numbers approached the upper bound of this interval, the difference was not statistically significant. Therefore, we cannot confidently conclude that the trial intervention had a measurable impact on customer visits. However, the proximity to the 95% threshold suggests a potential positive effect, and further investigation with additional data or across more stores may help determine whether the trend is meaningful. Based on this result the store trial 88 and 86 need a further investigation due to their behavior after the trial in order to be corrected.