Task-2-measureimpact

Cristian Torres

2025-12-25

```
 #install.packages("tidyverse")
 #install.packages("readxl")
 #install.packages("lubridate")
 #install.packages("knitr")
 #install.packages("ggplot2")

library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.6
## v forcats   1.0.1     v stringr   1.6.0
## v ggplot2   4.0.1     v tibble    3.3.0
## v lubridate 1.9.4     v tidyr     1.3.2
## v purrr     1.2.0
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(readxl)
library(lubridate)
library(knitr)
library(ggplot2)

setwd("C:/Users/Cristian/Downloads/Project Quantium")
```

Here we segment the data to future use

Separation of seasons (trial and not trial) using vector

```
trial_stores <- c(77, 86, 88)

pre_start <- ymd("2018-02-01")
pre_end <- ymd("2019-01-01")

trial_start <- ymd("2019-02-01")
trial_end <- ymd("2019-04-01")
```

Control in correlation based in start and end trial to sales and customers

Scale control by stores in monthly

```r
make_series <- function(df, trial_store, control_store) {

  t <- df %>% dplyr::filter(STORE_NBR == trial_store)
  c <- df %>% dplyr::filter(STORE_NBR == control_store)

  t_pre <- t %>% dplyr::filter(YEARMONTH >= pre_start, YEARMONTH <= pre_end)
  c_pre <- c %>% dplyr::filter(YEARMONTH >= pre_start, YEARMONTH <= pre_end)


  sf_sales <- sum(t_pre$total_sales, na.rm = TRUE) / sum(c_pre$total_sales, na.rm = TRUE)
  sf_cust  <- sum(t_pre$n_customers, na.rm = TRUE) / sum(c_pre$n_customers, na.rm = TRUE)
  sf_txpc  <- mean(t_pre$txn_per_cust, na.rm = TRUE) / mean(c_pre$txn_per_cust, na.rm = TRUE)

  s <- dplyr::full_join(
    t %>%
      dplyr::select(YEARMONTH, total_sales, n_customers, txn_per_cust) %>%
      dplyr::rename_with(~paste0("trial_", .x), -YEARMONTH),
    c %>%
      dplyr::select(YEARMONTH, total_sales, n_customers, txn_per_cust) %>%
      dplyr::rename_with(~paste0("ctrl_", .x), -YEARMONTH),
    by = "YEARMONTH"
  ) %>%
    dplyr::arrange(YEARMONTH) %>%
    dplyr::mutate(
      scaled_ctrl_sales = ctrl_total_sales * sf_sales,
      scaled_ctrl_cust  = ctrl_n_customers * sf_cust,
      scaled_ctrl_txpc  = ctrl_txn_per_cust * sf_txpc
    )

  s
}

sig_test_ci <- function(series, metric_trial, metric_scaled_ctrl) {
  pre_idx   <- series$YEARMONTH >= pre_start & series$YEARMONTH <= pre_end
  trial_idx <- series$YEARMONTH >= trial_start & series$YEARMONTH <= trial_end

  diff_pre <- (series[[metric_trial]] - series[[metric_scaled_ctrl]])[pre_idx]
  mu <- mean(diff_pre, na.rm = TRUE)
  sd_pre <- sd(diff_pre, na.rm = TRUE)

  upper <- mu + 1.96 * sd_pre
  lower <- mu - 1.96 * sd_pre

  diff_trial <- (series[[metric_trial]] - series[[metric_scaled_ctrl]])[trial_idx]

  tibble::tibble(
    mean_pre_diff = mu,
    sd_pre_diff = sd_pre,
    ci_lower = lower,
    ci_upper = upper,
    trial_months_all_above_upper = all(diff_trial > upper, na.rm = TRUE)
  )
}
```

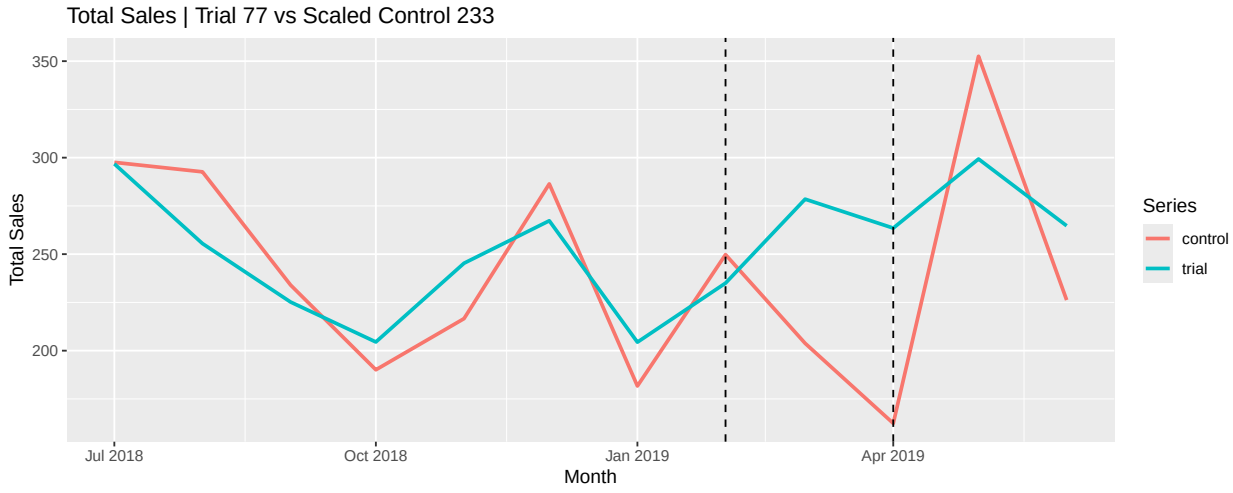
```
uplift_pct <- function(series, metric_trial, metric_scaled_ctrl) {
  trial_idx <- series$YEARMONTH >= trial_start & series$YEARMONTH <= trial_end

  t_sum <- sum(series[[metric_trial]][trial_idx], na.rm = TRUE)
  c_sum <- sum(series[[metric_scaled_ctrl]][trial_idx], na.rm = TRUE)

  (t_sum - c_sum) / c_sum
}

plot_metric <- function(series, title, ylab, trial_col, ctrl_col) {
  series %>%
    dplyr::select(YEARMONTH, trial = dplyr::all_of(trial_col), control = dplyr::all_of(ctrl_col)) %>%
    tidyr::pivot_longer(-YEARMONTH, names_to = "Series", values_to = "Value") %>%
    ggplot2::ggplot(ggplot2::aes(YEARMONTH, Value, color = Series)) +
    ggplot2::geom_line(linewidth = 1) +
    ggplot2::geom_vline(xintercept = trial_start, linetype = "dashed") +
    ggplot2::geom_vline(xintercept = trial_end, linetype = "dashed") +
    ggplot2::labs(title = title, x = "Month", y = ylab)
}
results <- purrr::pmap_dfr(
controls %>% dplyr::select(trial_store, control_store),
function(trial_store, control_store) {

s <- make_series(dat_m, trial_store, control_store)

sig_sales <- sig_test_ci(s, "trial_total_sales", "scaled_ctrl_sales")
sig_cust  <- sig_test_ci(s, "trial_n_customers", "scaled_ctrl_cust")
sig_txpc  <- sig_test_ci(s, "trial_txn_per_cust", "scaled_ctrl_txpc")

tibble::tibble(
  trial_store   = trial_store,
  control_store = control_store,

  sales_uplift_pct = uplift_pct(s, "trial_total_sales", "scaled_ctrl_sales"),
  cust_uplift_pct  = uplift_pct(s, "trial_n_customers", "scaled_ctrl_cust"),
  txpc_uplift_pct  = uplift_pct(s, "trial_txn_per_cust", "scaled_ctrl_txpc"),

  sales_sig = sig_sales$trial_months_all_above_upper,
  cust_sig  = sig_cust$trial_months_all_above_upper,
  txpc_sig  = sig_txpc$trial_months_all_above_upper
)

}
)

results
```

```
## # A tibble: 3 x 8
##   trial_store control_store sales_uplift_pct cust_uplift_pct txpc_uplift_pct
##         <dbl>         <dbl>            <dbl>           <dbl>           <dbl>
## 1          77           233            0.262          0.231         -0.0203
## 2          86           155            0.132          0.135          0.00664
## 3          88           237            0.121          0.0635         0.0641
```

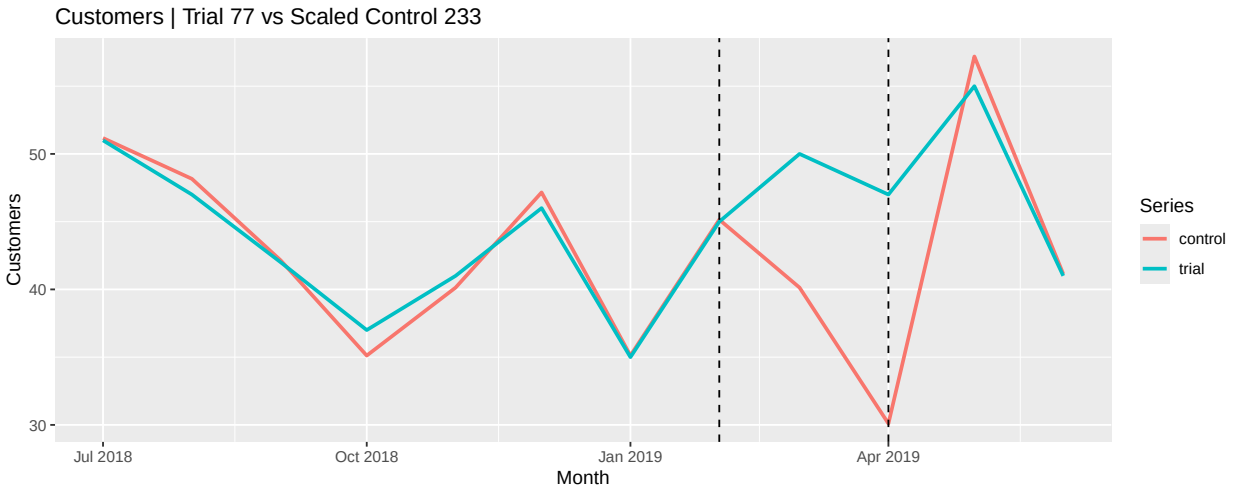
```
## # i 3 more variables: sales_sig <lgl>, cust_sig <lgl>, txpc_sig <lgl>
```

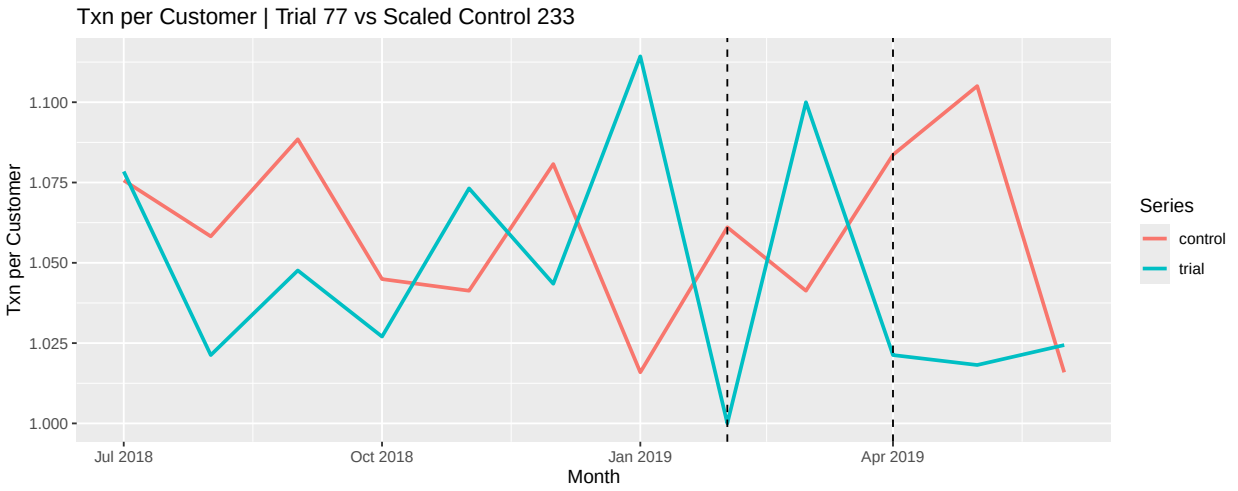
Visualization supports labels and functions

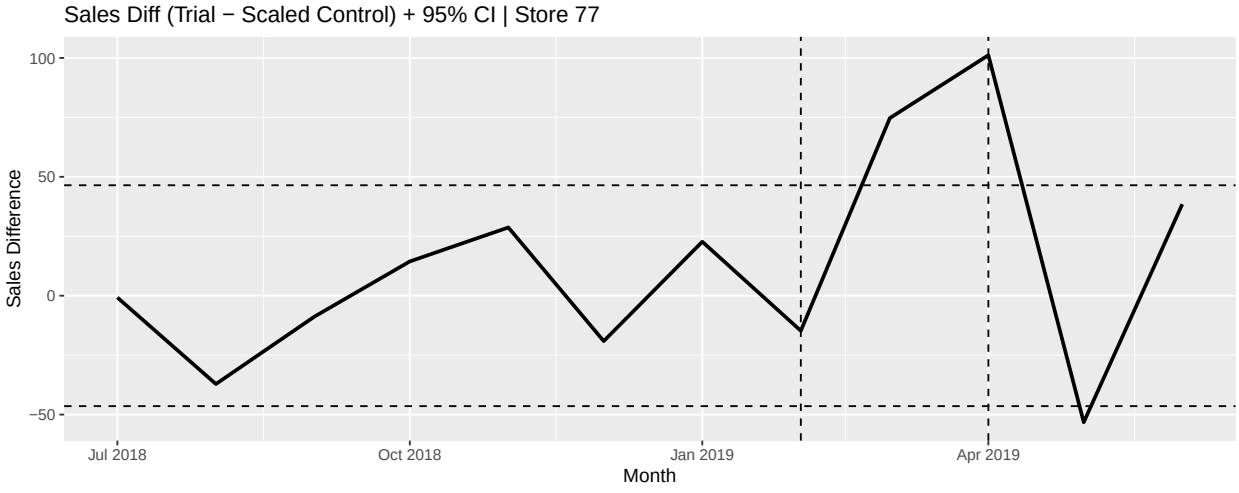
```
plot_metric <- function(series, title, ylab, trial_col, ctrl_col) {
  series %>%
    dplyr::select(YEARMONTH, trial = dplyr::all_of(trial_col), control = dplyr::all_of(ctrl_col)) %>%
    tidyr::pivot_longer(-YEARMONTH, names_to = "Series", values_to = "Value") %>%
    ggplot2::ggplot(ggplot2::aes(YEARMONTH, Value, color = Series)) +
    ggplot2::geom_line(linewidth = 1) +
    ggplot2::geom_vline(xintercept = trial_start, linetype = "dashed") +
    ggplot2::geom_vline(xintercept = trial_end, linetype = "dashed") +
    ggplot2::labs(title = title, x = "Month", y = ylab)
}

plot_diff_ci <- function(series, metric_trial, metric_scaled_ctrl, title, ylab) {
  pre_idx   <- series$YEARMONTH >= pre_start & series$YEARMONTH <= pre_end
  diff_pre  <- (series[[metric_trial]] - series[[metric_scaled_ctrl]])[pre_idx]
  mu <- mean(diff_pre, na.rm = TRUE)
  sd_pre <- sd(diff_pre, na.rm = TRUE)
  upper <- mu + 1.96 * sd_pre
  lower <- mu - 1.96 * sd_pre

  dfp <- tibble::tibble(
    YEARMONTH = series$YEARMONTH,
    diff = series[[metric_trial]] - series[[metric_scaled_ctrl]],
    ci_lower = lower,
    ci_upper = upper
  )

  ggplot2::ggplot(dfp, ggplot2::aes(YEARMONTH, diff)) +
    ggplot2::geom_line(linewidth = 1) +
    ggplot2::geom_hline(yintercept = upper, linetype = "dashed") +
    ggplot2::geom_hline(yintercept = lower, linetype = "dashed") +
    ggplot2::geom_vline(xintercept = trial_start, linetype = "dashed") +
    ggplot2::geom_vline(xintercept = trial_end, linetype = "dashed") +
    ggplot2::labs(title = title, x = "Month", y = ylab)
}
```

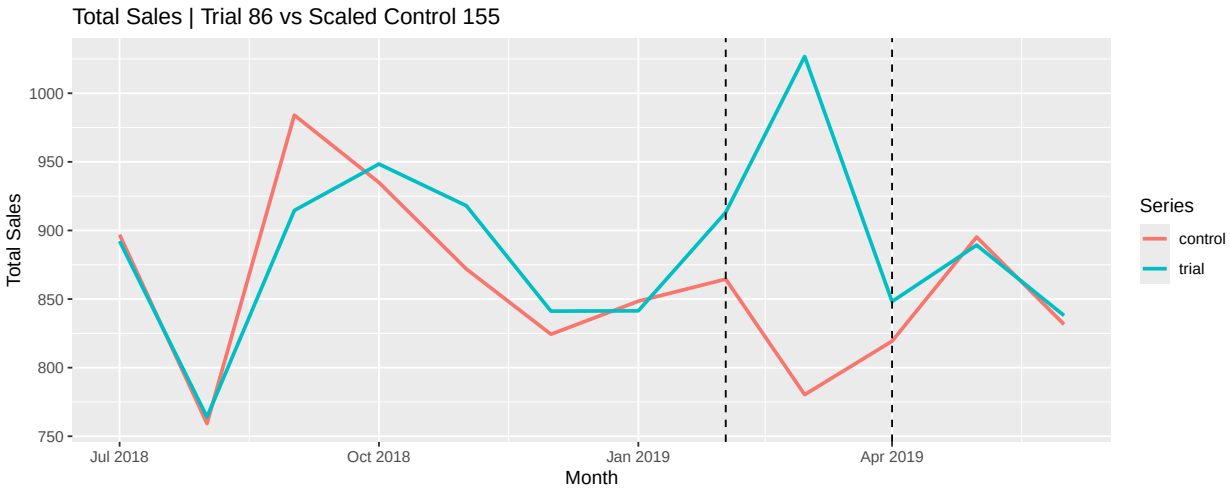


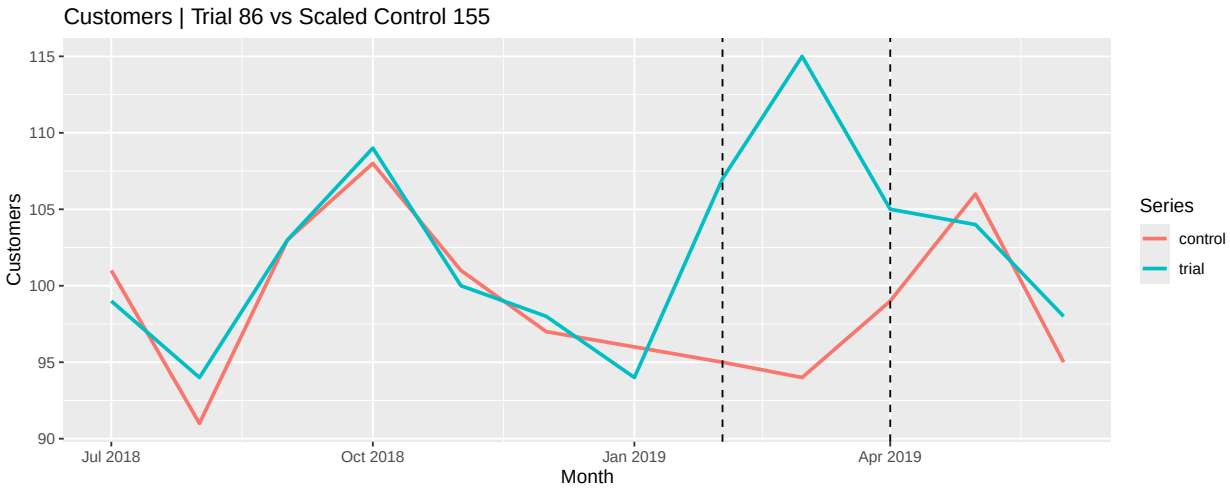
Viz deployed

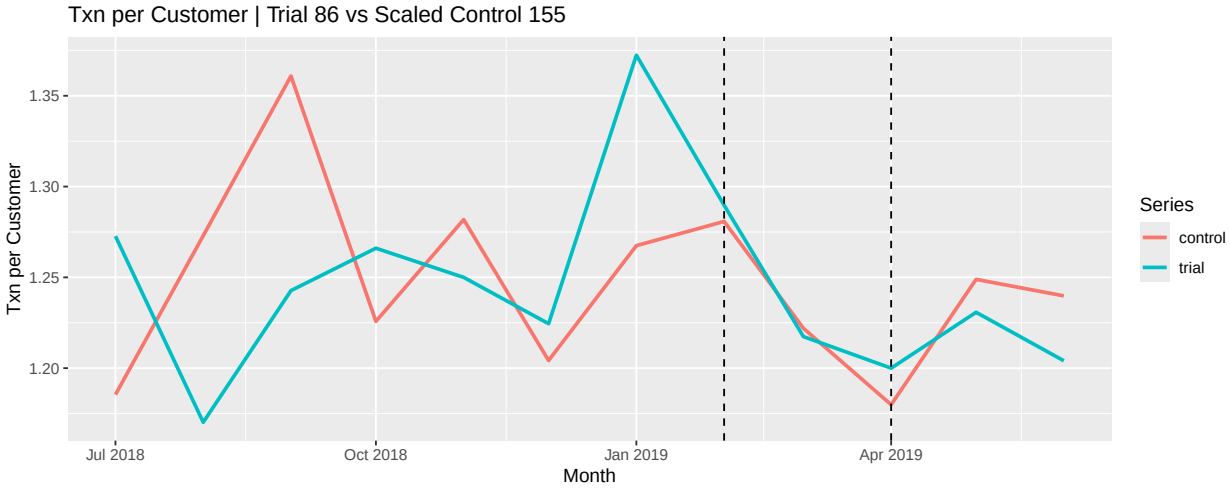

Customers | Trial 77 vs Scaled Control 233
Customers
50
40
30
Jul 2018
Oct 2018
Jan 2019
Apr 2019
Month
Series
control
trial

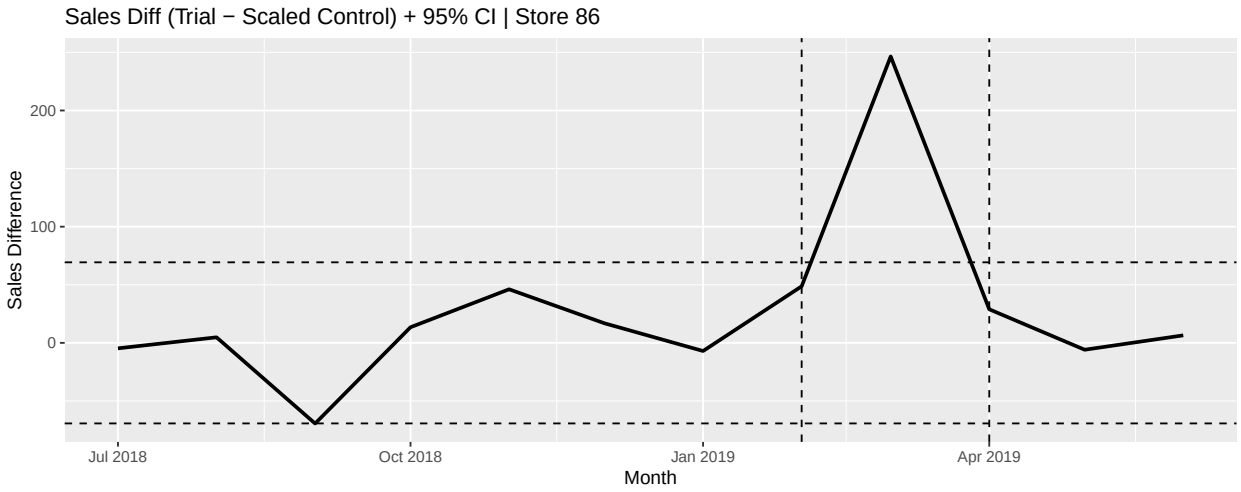

Txn per Customer | Trial 77 vs Scaled Control 233
Txn per Customer
1.100
1.075
1.050
1.025
1.000
Jul 2018
Oct 2018
Jan 2019
Apr 2019
Month
Series
control
trial

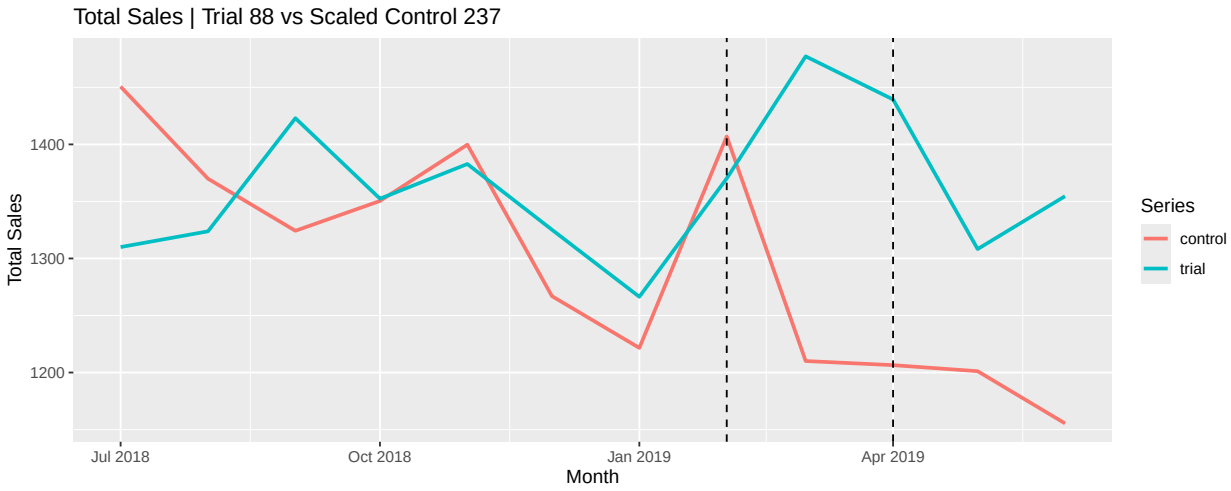

Sales Diff (Trial − Scaled Control) + 95% CI | Store 77
Sales Difference
100
50
0
−50
Jul 2018
Oct 2018
Jan 2019
Apr 2019
Month

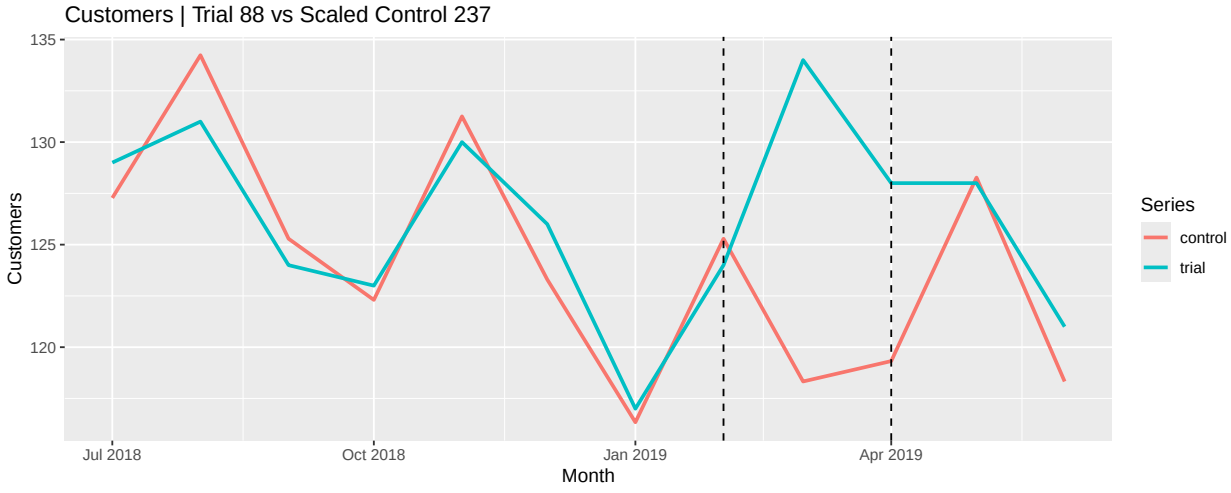

Total Sales | Trial 86 vs Scaled Control 155
Total Sales
1000
950
900
850
800
750
Jul 2018
Oct 2018
Jan 2019
Apr 2019
Month
Series
control
trial

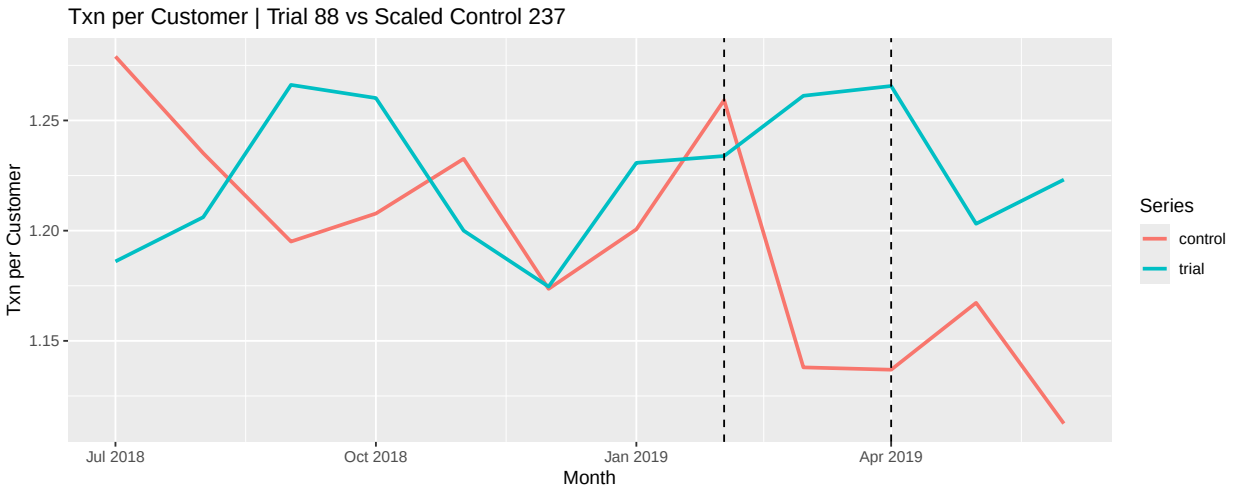

Customers | Trial 86 vs Scaled Control 155
Customers
115
110
105
100
95
90
Jul 2018
Oct 2018
Jan 2019
Apr 2019
Month
Series
control
trial

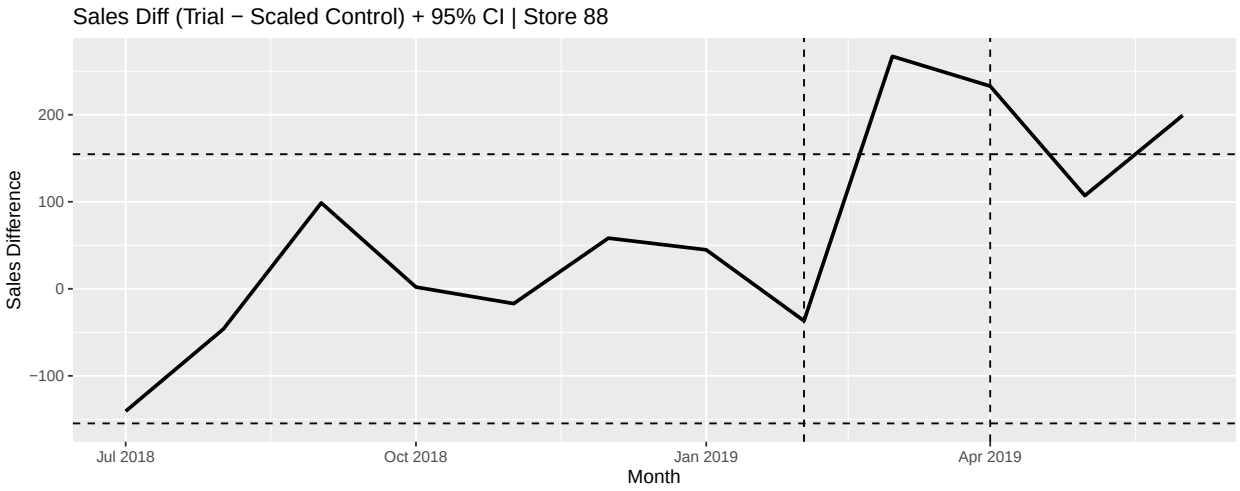

Txn per Customer | Trial 86 vs Scaled Control 155
Txn per Customer
1.35
1.30
1.25
1.20
Jul 2018
Oct 2018
Jan 2019
Apr 2019
Month
Series
control
trial


Sales Diff (Trial − Scaled Control) + 95% CI | Store 86
Sales Difference
200
100
0
Jul 2018
Oct 2018
Jan 2019
Apr 2019
Month


Total Sales | Trial 88 vs Scaled Control 237
Total Sales
1400
1300
1200
Jul 2018
Oct 2018
Jan 2019
Apr 2019
Month
Series
control
trial


Customers | Trial 88 vs Scaled Control 237
Customers
135
130
125
120
Jul 2018
Oct 2018
Jan 2019
Apr 2019
Month
Series
control
trial


Txn per Customer | Trial 88 vs Scaled Control 237
Txn per Customer
1.25
1.20
1.15
Jul 2018
Oct 2018
Jan 2019
Apr 2019
Month
Series
control
trial


Sales Diff (Trial − Scaled Control) + 95% CI | Store 88
Sales Difference
200
100
0
−100
Jul 2018
Oct 2018
Jan 2019
Apr 2019
Month