A/B Test

Nidhi

Abstract

This is a glance at A/B testing by visualization alone. I have plotted graphs of various parameters and speculated on the final inference. These claims can be confirmed or denied by conducting a hypothesis test.

Introduction

In modern data analytics, deciding whether two numerical samples come from the same underlying distribution is called *A/B testing*. The name refers to the labels of the two samples, A and B.

A/B testing, also known as split testing, refers to a randomized experimentation process wherein two or more versions of a variable (web page, page element, etc.) are shown to different segments of website visitors at the same time to determine which version leaves the maximum impact and drives business metrics. Essentially, A/B testing eliminates all the guesswork out of website optimization and enables experience optimizers to make data-backed decisions. In A/B testing, A refers to 'control' or the original testing variable. Whereas B refers to 'variation' or a new version of the original testing variable.

A/B testing is one of the components of the overarching process of Conversion Rate Optimization (CRO), using which we can gather both qualitative and quantitative user insights. We can further use this collected data to understand user behavior, engagement rate, pain points, and even satisfaction with website features, including new features, revamped page sections, etc.

Dataset Description

A company recently introduced a new bidding type, "average bidding", as an alternative to its exisiting bidding type, called "maximum bidding". One of our clients, ...com, has decided to test this new feature and wants to conduct an A/B test to understand if average bidding brings more conversions than maximum bidding.

The A/B test has run for 1 month and ....com now expects you to analyze and present the results of this A/B test.

Details about columns in Dataset:

- CampaignName : Name of the type of data(control or test)
- Date : Date
- SpendUSD : Money spent on each day
- Impressions : It is a variable for the user to see an ad.
- Reach : The number of unique people who saw an ad.

- WebsiteClicks : It is the variable related to the user clicking the website link in the advertisement.
- Searches : It is the variable related to the user performing a search on the website.
- ViewContent : It is the variable related to the user viewing the details of a product.
- AddToCart : It is the variable related to the user adding the product to the cart.
- Purchases : It is the variable related to the user's purchase of the product.

Importing Dataset

```
control = read.csv("control_group.csv", sep = ";")
test =  read.csv("test_group.csv", sep = ";")
head(test)
```

```
##   Campaign.Name      Date Spend..USD. X..of.Impressions Reach
## 1 Test Campaign 1.08.2019        3008             39550 35820
## 2 Test Campaign 2.08.2019        2542            100719 91236
## 3 Test Campaign 3.08.2019        2365             70263 45198
## 4 Test Campaign 4.08.2019        2710             78451 25937
## 5 Test Campaign 5.08.2019        2297            114295 95138
## 6 Test Campaign 6.08.2019        2458             42684 31489
##   X..of.Website.Clicks X..of.Searches X..of.View.Content X..of.Add.to.Cart
## 1                 3038           1946               1069               894
## 2                 4657           2359               1548               879
## 3                 7885           2572               2367              1268
## 4                 4216           2216               1437               566
## 5                 5863           2106                858               956
## 6                 7488           1854               1073               882
##   X..of.Purchase
## 1            255
## 2            677
## 3            578
## 4            340
## 5            768
## 6            488
```

```
head(control)
```

```
##      Campaign.Name      Date Spend..USD. X..of.Impressions  Reach
## 1 Control Campaign 1.08.2019        2280             82702  56930
## 2 Control Campaign 2.08.2019        1757            121040 102513
## 3 Control Campaign 3.08.2019        2343            131711 110862
## 4 Control Campaign 4.08.2019        1940             72878  61235
## 5 Control Campaign 5.08.2019        1835                NA     NA
## 6 Control Campaign 6.08.2019        3083            109076  87998
##   X..of.Website.Clicks X..of.Searches X..of.View.Content X..of.Add.to.Cart
## 1                 7016           2290               2159              1819
## 2                 8110           2033               1841              1219
## 3                 6508           1737               1549              1134
## 4                 3065           1042                982              1183
## 5                   NA             NA                 NA                NA
## 6                 4028           1709               1249               784
##   X..of.Purchase
## 1            618
## 2            511
## 3            372
## 4            340
## 5             NA
## 6            764
```

Visualizing the data in R

Setting things up by including required packages

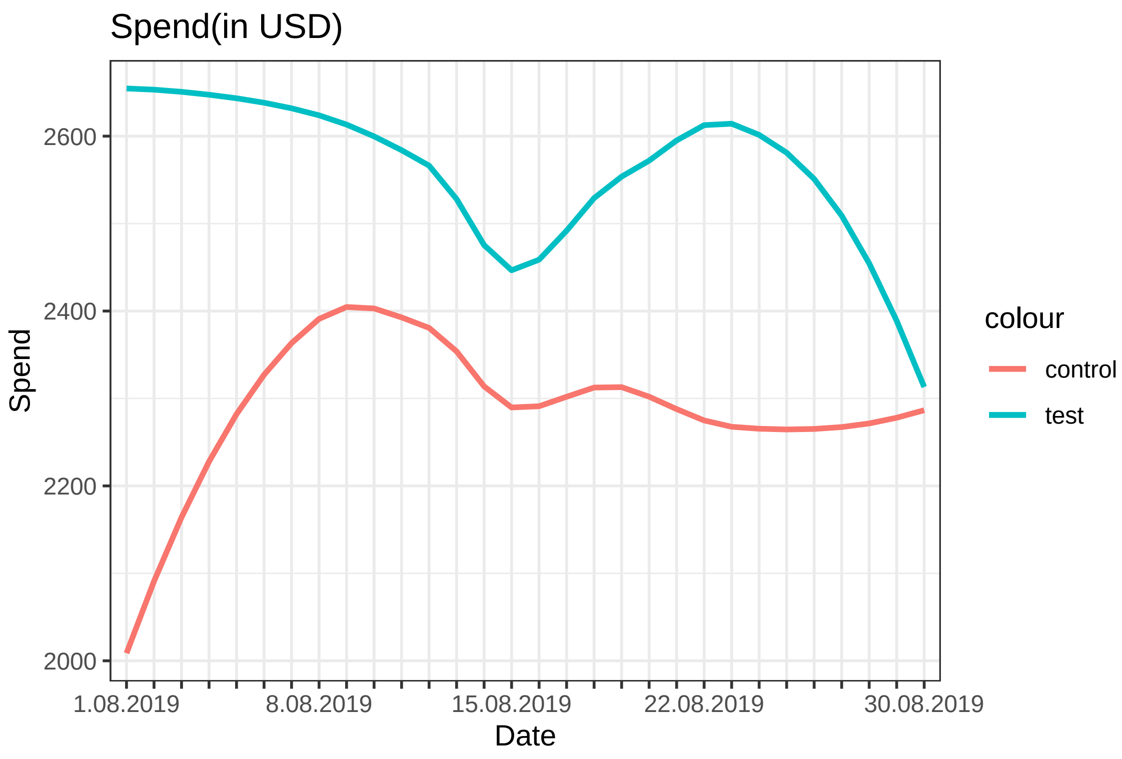
```
library(tidyverse)
library(dplyr)
library(scales)
```

Cleaning Dataset by changing names of columns.

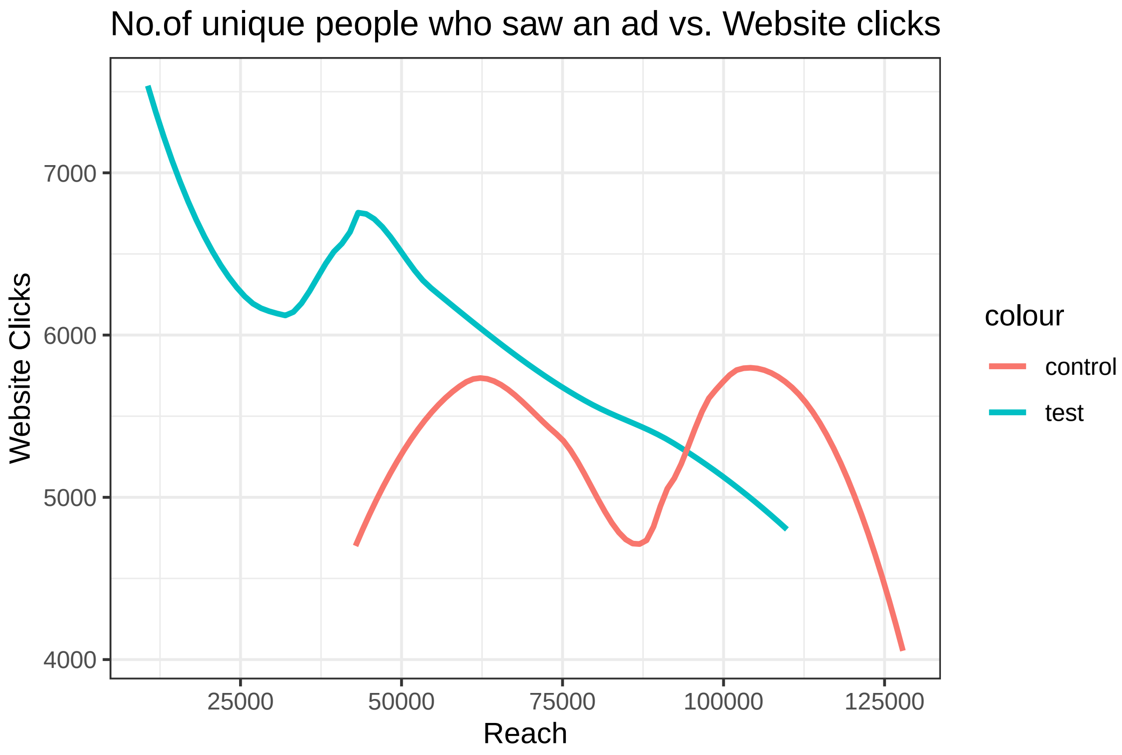
```
names(control) = c("CampaignName", "Date", "SpendUSD", "Impressions", "Reach",
"WebsiteClicks", "Searches", "ViewContent", "AddToCart", "Purchases")

names(test) = c("CampaignName", "Date", "SpendUSD", "Impressions", "Reach",
"WebsiteClicks", "Searches", "ViewContent", "AddToCart", "Purchases")
```

```
print(ggplot() +
   geom_smooth(test, mapping = aes(x =factor(Date, levels = unique(Date)),
   y =SpendUSD, group=1, colour = "test"), se=F) + theme_bw() +
   geom_smooth(control, mapping = aes(x =factor(Date, levels = unique(Date)),
   y =SpendUSD , group=1, color = "control"), se=F) + theme_bw() +
   scale_x_discrete(guide = guide_axis(check.overlap = TRUE))+
   labs(x = "Date", y = "Spend", title ="Spend(in USD)"))


 print(ggplot() +
   geom_smooth(test, mapping = aes(x = Reach, y =WebsiteClicks, group=1,
   colour = "test"), se=F)+theme_bw()+
   geom_smooth(control, mapping = aes(x = Reach, y =WebsiteClicks, group=1,
   color="control"), se=F)+theme_bw()+labs(x= "Reach", y = "Website Clicks",
   title = "No.of unique people who saw an ad vs. Website clicks"))

print(ggplot() +
   geom_smooth(test, mapping = aes(x =WebsiteClicks, y =Searches, group=1,
   colour = "test"), se=F) + theme_bw() +
   geom_smooth(control, mapping = aes(x =WebsiteClicks, y=Searches, group=1,
   color = "control"), se=F) + theme_bw()+labs(x="Website Clicks", y="Search",
   title = "No.of website clicks vs. User performing a search on the website"))
```

```
print(ggplot() +
    geom_smooth(test, mapping = aes(x =Searches, y =ViewContent, group=1,
    colour = "test"), se=F) + theme_bw() +
    geom_smooth(control, mapping = aes(x =Searches, y =ViewContent, group=1,
    color="control"),se=F) + theme_bw()+labs(x="Search", y="Content View",
    title = "No. of user performing a search on the website vs.
    User viewing the details of a product"))

print(ggplot() +
    geom_smooth(test, mapping = aes(x =AddToCart, y =Purchases, group=1,
    colour = "test"), se=F) + theme_bw() +
    geom_smooth(control, mapping = aes(x =AddToCart, y =Purchases, group=1,
    color = "control"), se=F) + theme_bw()+labs(x="Add to cart", y="Purchase",
    title="User adding the product to the cart vs.User's purchase of the \nproduct"))

print(ggplot() +
    geom_smooth(test, mapping = aes(x =factor(Date, levels = unique(Date)),
    y =AddToCart, group=1, colour = "test"), se=F) + theme_bw() +
    geom_smooth(control, mapping = aes(x =factor(Date, levels = unique(Date)),
    y =AddToCart, group=1, color="control"), se=F) + theme_bw()+
    scale_x_discrete(guide = guide_axis(check.overlap = TRUE))+labs(x="Date",
    y="Add to cart", title = "No. of products added to the cart on each day"))



test <- test %>%
  mutate(CRT = Purchases * 100/AddToCart,
       IMRT = WebsiteClicks*100/Impressions
       )

control <- control %>%
  mutate(CRT = Purchases * 100/AddToCart,
       IMRT = WebsiteClicks*100/Impressions
)

print(ggplot()+
  geom_smooth(test, mapping = aes(x=factor(Date, levels = unique(Date)),
  y = CRT, group = 1, colour = "test"), se = F ) + theme_bw()+
  geom_smooth(control, mapping = aes(x=factor(Date, levels = unique(Date)),
  y = CRT, group = 1, color="control"), se = F ) +theme_bw()+
  scale_x_discrete(guide = guide_axis(check.overlap = TRUE))+labs(x="Date",
  y="Conversion Rate(%)", title = "Conversion Rate(%) for each day"))


print(ggplot()+
  geom_smooth(test, mapping = aes(x=factor(Date, levels = unique(Date)),
  y = IMRT, group = 1, colour = "test"),se=F ) + theme_bw()+
  geom_smooth(control, mapping = aes(x=factor(Date, levels = unique(Date)),
  y = IMRT, group = 1, color = "control"), se = F ) +theme_bw()+
  scale_x_discrete(guide = guide_axis(check.overlap = TRUE))+labs(x="Date",
  y="Impression Rate(%)", title = "Impression Rate(%) for each day"))
```

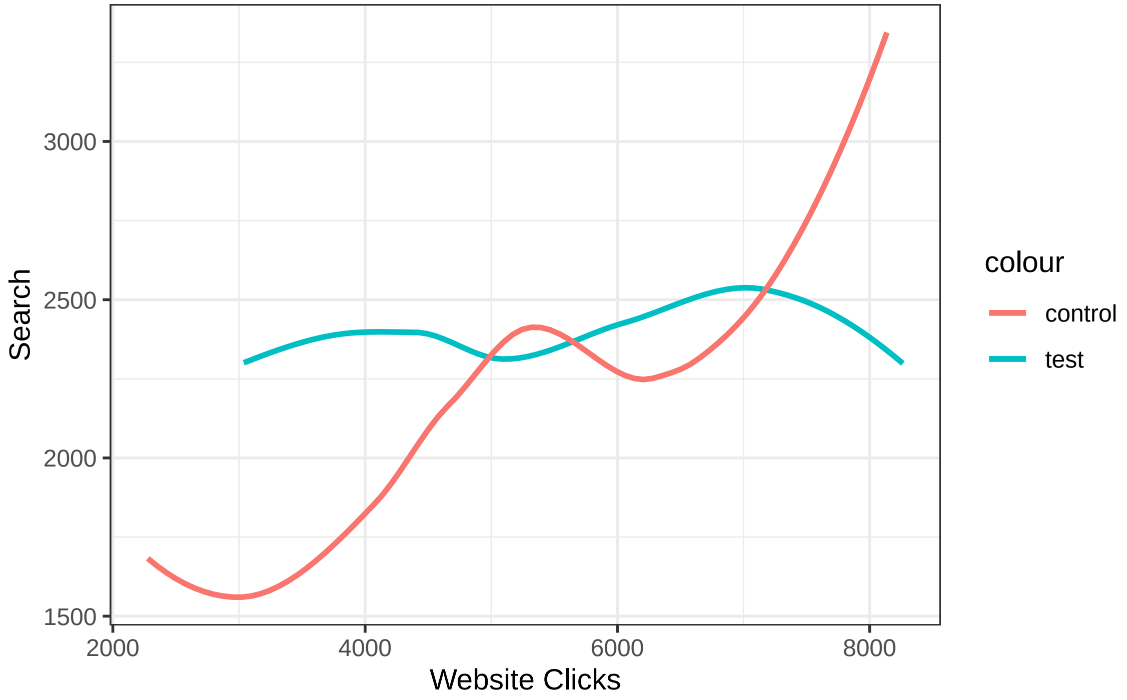


Spend(in USD)
Spend
2600
2400
2200
2000
1.08.2019
8.08.2019
15.08.2019
22.08.2019
30.08.2019
Date
colour
control
test

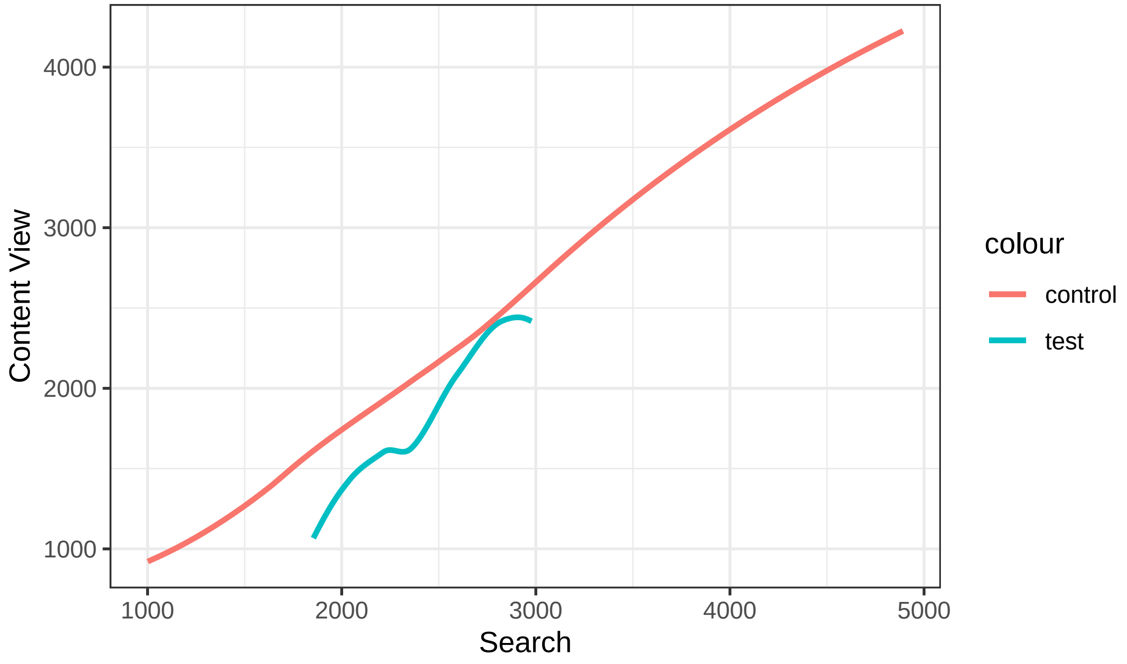


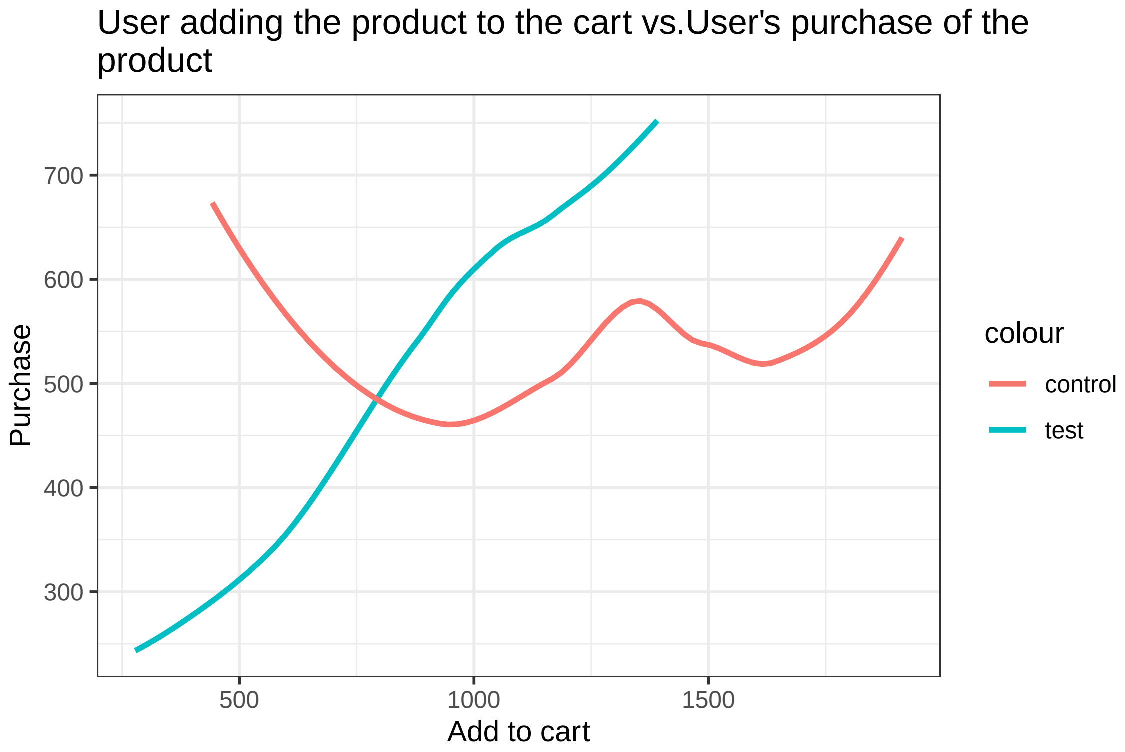
No.of unique people who saw an ad vs. Website clicks
Website Clicks
7000
6000
5000
4000
25000
50000
75000
100000
125000
Reach
colour
control
test

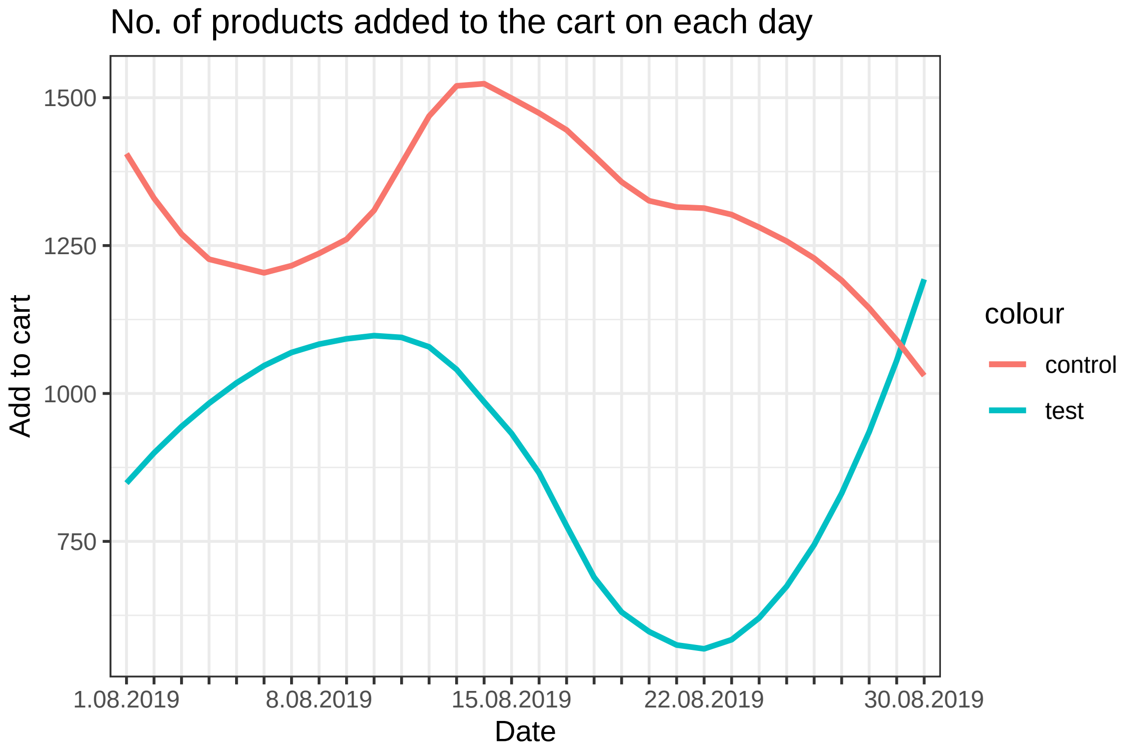
No.of website clicks vs. User performing a search on the website

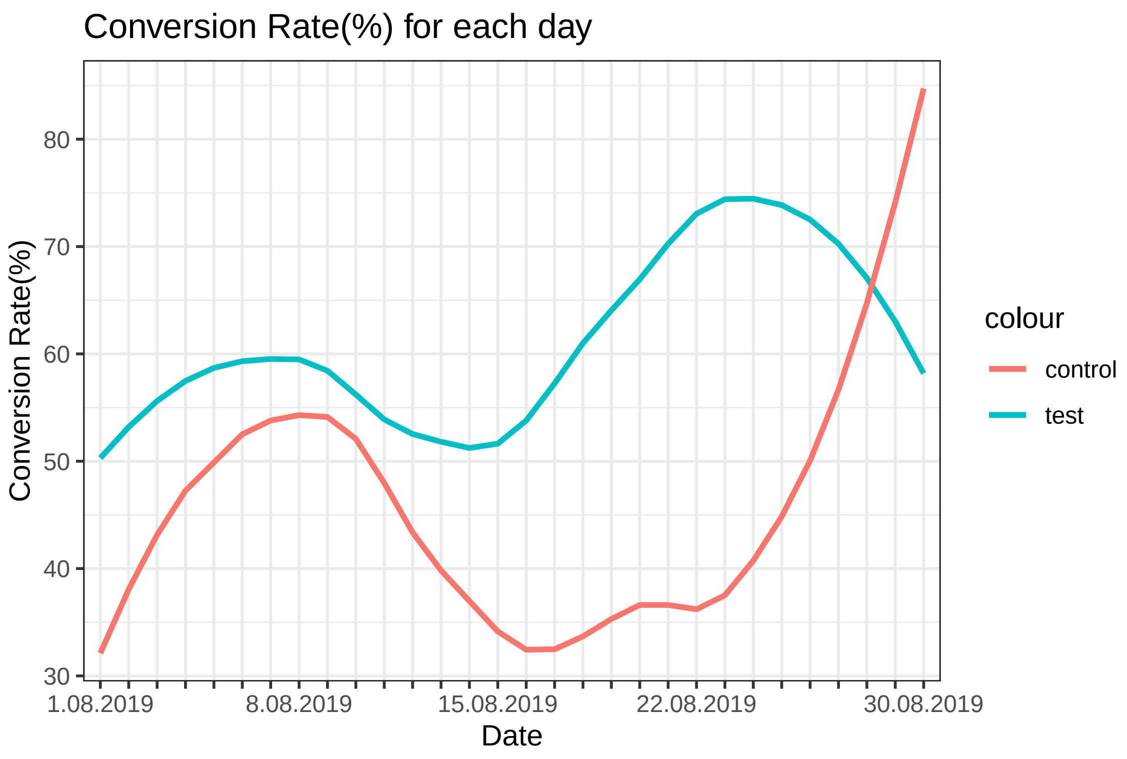


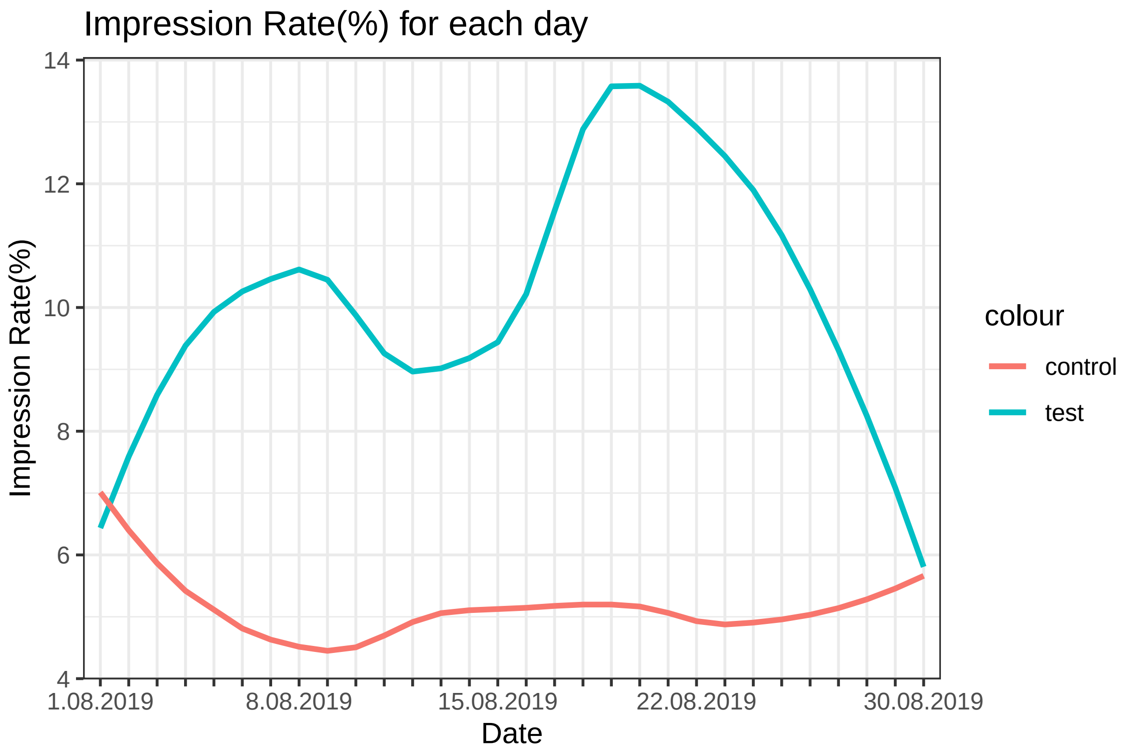
No. of user performing a search on the website vs.
User viewing the details of a product




User adding the product to the cart vs.User's purchase of the product
Purchase
700
600
500
400
300
500
1000
1500
Add to cart
colour
control
test


No. of products added to the cart on each day
Add to cart
1500
1250
1000
750
1.08.2019
8.08.2019
15.08.2019
22.08.2019
30.08.2019
Date
colour
control
test


Conversion Rate(%) for each day
Conversion Rate(%)
80
70
60
50
40
30
1.08.2019
8.08.2019
15.08.2019
22.08.2019
30.08.2019
Date
colour
control
test


Impression Rate(%) for each day
Impression Rate(%)
14
12
10
8
6
4
1.08.2019
8.08.2019
15.08.2019
22.08.2019
30.08.2019
Date
colour
control
test

Summary of Analysis

- The daily spend on test data is higher than the daily spend on control data.
- There is variation in spend pattern for test and control data, however both test and control experience a dip in spend at the middle of the month.
- There is a clear increase in trend in control over the month as website clicks increase, so did unique searches. For the test data however, there is significantly less variation across the month.
- Even though it might look like test has steady increase in purchase vs. add to cart, more people are adding to cart in control than in test.

Conclusion

Initially, there is no significant difference in conversion rate for both test data and control data but for last eight days of the month, there is a significant increase in conversion rate of control data.