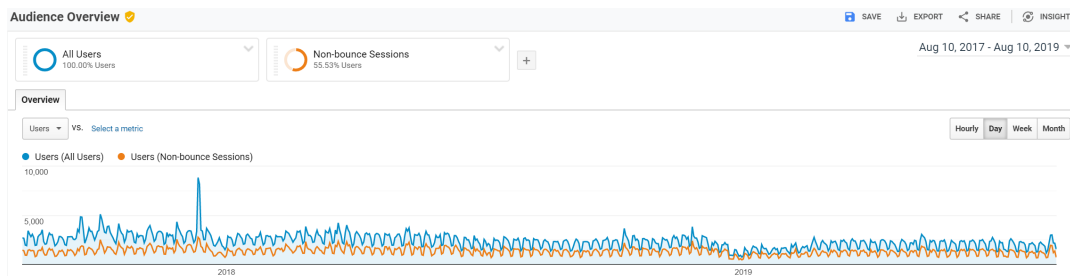


Google Merchandise Store Analysis

Part 1: Find the most valuable acquisition channel

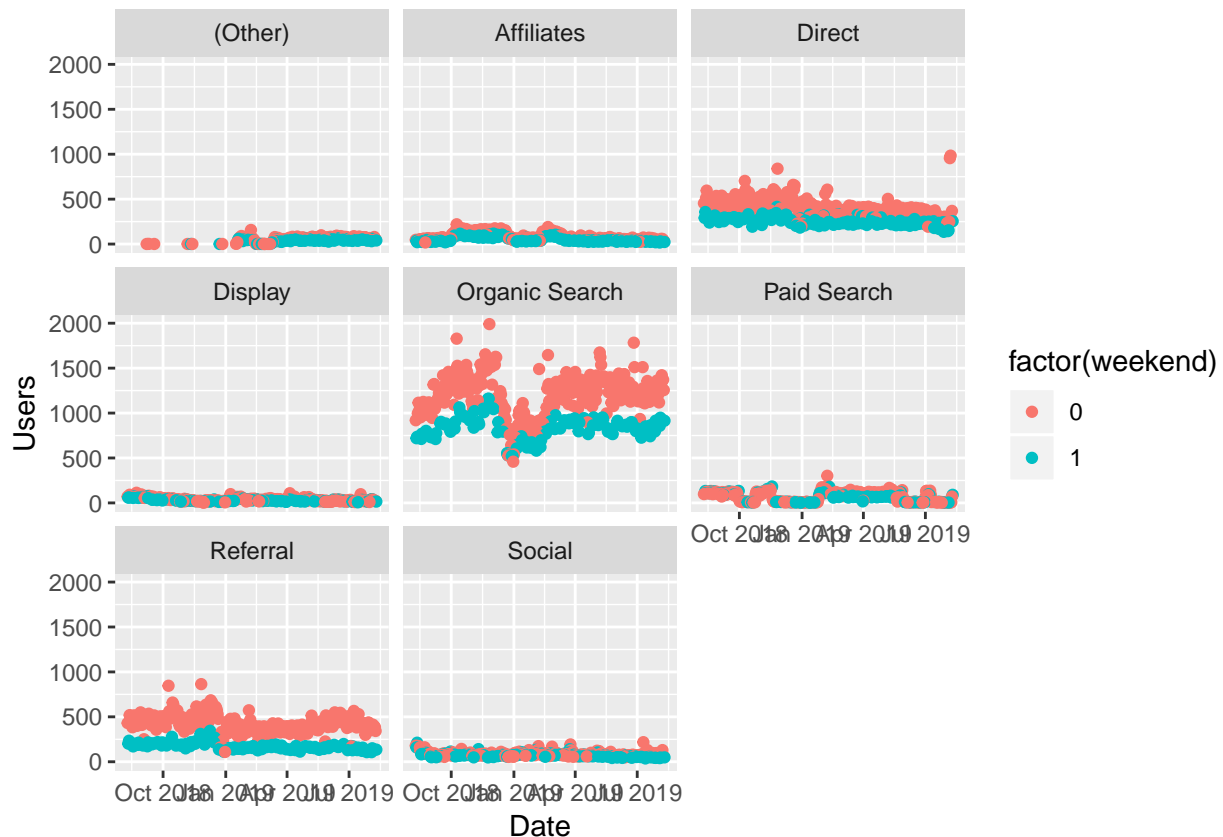
1. Visualize user trends over time on Google Analytics

Look at the users and the non-bounce users of Google Merchandise store in the last 2 years. Although total users have a few spikes, non-bounce users have a stable pattern overtime.



2. Import data from Google Analytics and explore data

For most acquisition channel, user in weekdays and weekend have significant difference, showing that weekend is a strong feature that needs to be controlled.



3. Regression

Run regression to explore the relationship between acquisition channel and: a. Bounce Rate b. Average Session Duration c. E-commerce Conversion Rate d. Revenue

The variables we need to control to avoid omitted variable bias: 1. time trends 2. the number of new users 3. whether or not it is a weekend

4. Conclusion

Referral is the only channel that has significant advantage in all the 4 metrics (low bounce rate, high session duration, conversion rate, and revenue), which means that it is the most valuable acquisition channel

```
## [1] "Bounce Rate"
```

```
##               Estimate   Std. Error
## (Intercept)    -1.924308e+00 2.801827e-01
## `Default Channel Grouping`Direct    -2.158435e-01 7.537428e-03
## `Default Channel Grouping`Display     2.177773e-01 6.171936e-03
## `Default Channel Grouping`Organic Search -1.256708e-01 1.669668e-02
## `Default Channel Grouping`Paid Search  -1.528559e-01 6.155385e-03
## `Default Channel Grouping`Referral    -2.940235e-01 6.581139e-03
## `Default Channel Grouping`Social      4.852832e-02 6.158561e-03
## Date          1.361028e-04 1.561267e-05
## `New Users`    8.281603e-05 1.884161e-05
## weekend         1.992145e-02 3.941328e-03
##               t value   Pr(>|t|)
## (Intercept)    -6.868046 8.146057e-12
## `Default Channel Grouping`Direct    -28.636230 1.243451e-156
## `Default Channel Grouping`Display     35.285080 1.771722e-222
## `Default Channel Grouping`Organic Search -7.526694 7.171828e-14
## `Default Channel Grouping`Paid Search  -24.832872 4.107405e-122
## `Default Channel Grouping`Referral    -44.676687 1.482197e-322
## `Default Channel Grouping`Social      7.879815 4.811614e-15
## Date          8.717459 5.027668e-18
## `New Users`    4.395380 1.151015e-05
## weekend         5.054501 4.620776e-07
```

```
## [1] "Session Duration"
```

```
##               Estimate   Std. Error
## (Intercept)    -261.07144070 194.66831869
## `Default Channel Grouping`Direct      81.80597055 5.23693350
## `Default Channel Grouping`Display    -87.79534005 4.28820276
## `Default Channel Grouping`Organic Search  50.95385045 11.60069396
## `Default Channel Grouping`Paid Search   43.84729081 4.27670312
## `Default Channel Grouping`Referral    125.04826671 4.57251300
## `Default Channel Grouping`Social    -52.74753175 4.27891003
## Date          0.02279022 0.01084753
## `New Users`   -0.05097864 0.01309097
## weekend         5.51154768 2.73839753
##               t value   Pr(>|t|)
## (Intercept)    -1.341109 1.800044e-01
## `Default Channel Grouping`Direct     15.620968 1.323325e-52
## `Default Channel Grouping`Display    -20.473691 2.221050e-86
```

```

## `Default Channel Grouping`Organic Search    4.392311  1.167265e-05
## `Default Channel Grouping`Paid Search      10.252592  3.371230e-24
## `Default Channel Grouping`Referral         27.347821  1.237062e-144
## `Default Channel Grouping`Social           -12.327329  5.855377e-34
## Date                                       2.100959  3.574243e-02
## `New Users`                               -3.894184  1.010585e-04
## weekend                                    2.012691  4.425202e-02

## [1] "Conversion Rate"

##                                     Estimate   Std. Error
## (Intercept)                       4.575498e-01 2.387047e-02
## `Default Channel Grouping`Direct    2.433670e-03 6.421593e-04
## `Default Channel Grouping`Display    9.781501e-04 5.258248e-04
## `Default Channel Grouping`Organic Search 5.170374e-04 1.422492e-03
## `Default Channel Grouping`Paid Search 4.839262e-03 5.244147e-04
## `Default Channel Grouping`Referral    5.349186e-03 5.606872e-04
## `Default Channel Grouping`Social     3.742946e-05 5.246853e-04
## Date                               -2.549214e-05 1.330138e-06
## `New Users`                         2.633979e-06 1.605231e-06
## weekend                             -3.720016e-04 3.357857e-04
##                                     t value   Pr(>|t|)
## (Intercept)                       19.16802527 1.288092e-76
## `Default Channel Grouping`Direct    3.78982268 1.542435e-04
## `Default Channel Grouping`Display    1.86022075 6.296926e-02
## `Default Channel Grouping`Organic Search 0.36347312 7.162816e-01
## `Default Channel Grouping`Paid Search 9.22793025 5.619105e-20
## `Default Channel Grouping`Referral    9.54040996 3.198743e-21
## `Default Channel Grouping`Social     0.07133698 9.431352e-01
## Date                               -19.16503444 1.354467e-76
## `New Users`                         1.64087178 1.009473e-01
## weekend                             -1.10785419 2.680292e-01

## [1] "Revenue"

##                                     Estimate   Std. Error
## (Intercept)                       24855.9179305 2354.8571039
## `Default Channel Grouping`Direct    -40.0498593  63.3499592
## `Default Channel Grouping`Display    30.5215298  51.8733854
## `Default Channel Grouping`Organic Search -495.1702399 140.3308806
## `Default Channel Grouping`Paid Search  49.3575025  51.7342769
## `Default Channel Grouping`Referral    263.7239189  55.3126199
## `Default Channel Grouping`Social     -16.1411289  51.7609734
## Date                               -1.3878873   0.1312201
## `New Users`                         0.9846104   0.1583584
## weekend                             -47.3432696  33.1257543
##                                     t value   Pr(>|t|)
## (Intercept)                       10.5551704 1.600944e-25
## `Default Channel Grouping`Direct    -0.6322002 5.273127e-01
## `Default Channel Grouping`Display    0.5883852 5.563259e-01
## `Default Channel Grouping`Organic Search -3.5285907 4.251596e-04
## `Default Channel Grouping`Paid Search  0.9540580 3.401446e-01
## `Default Channel Grouping`Referral    4.7678797 1.966399e-06
## `Default Channel Grouping`Social     -0.3118397 7.551878e-01
## Date                               -10.5767928 1.283744e-25
## `New Users`                         6.2176069 5.878836e-10

```

```
## weekend -1.4291982 1.530697e-01
```

Part 2: Analyze what make customers add products to their carts

1. Import data from Google Analytics

Create the segmentation that users who added a product to their cart

2. Data manipulation

3. Run regression on features to analyze what makes users more or less likely to add products to their carts

The variables we need to control to avoid omitted variable bias: 1. Time trend 2. acquisition channel 3. number of all users

4. Conclusion

The two significant variables are new users and average session duration. New users are less likely to add products to their cart. People spend more time on the page per session are more likely to add products to carts

```
##
## Call:
## lm(formula = cart_prob ~ all_users + new_users + Avg_Session_Duration +
##      others + Affiliates + Direct + Display + organic_search +
##      paid_search + Referral + Social + Date, data = df2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.023237 -0.005658 -0.000771  0.004921  0.026274
##
## Coefficients: (1 not defined because of singularities)
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    2.037e+00  8.207e-01   2.482   0.0151 *
## all_users       1.615e-05  7.118e-05   0.227   0.8211
## new_users      -1.184e-04  4.178e-05  -2.834   0.0058 **
## Avg_Session_Duration 1.687e-04  7.695e-05   2.193   0.0312 *
## others          4.977e-05  1.144e-04   0.435   0.6647
## Affiliates       1.733e-04  1.350e-04   1.284   0.2029
## Direct           8.057e-05  5.698e-05   1.414   0.1612
## Display          5.520e-05  9.720e-05   0.568   0.5717
## organic_search    4.975e-05  5.477e-05   0.908   0.3664
## paid_search      -4.180e-05  6.289e-05  -0.665   0.5082
## Referral         1.586e-04  6.939e-05   2.286   0.0248 *
## Social              NA          NA      NA      NA
## Date             -1.081e-04  4.523e-05  -2.391   0.0191 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.009355 on 81 degrees of freedom
## Multiple R-squared:  0.7738, Adjusted R-squared:  0.7431
```

F-statistic: 25.19 on 11 and 81 DF, p-value: < 2.2e-16