TrainSample EDA

```
── Attaching packages ──────────────────────────── tidyverse 1.2.1 ──
✓ ggplot2 3.0.0.9000      ✓ purrr   0.2.5
✓ tibble  1.4.2           ✓ dplyr   0.7.6
✓ tidyr   0.8.1           ✓ stringr 1.3.1
✓ readr   1.2.0           ✓ forcats 0.3.0
── Conflicts ───────────────────────────── tidyverse_conflicts() ──
✗ dplyr::filter() masks stats::filter()
✗ dplyr::lag()    masks stats::lag()
Attaching package: 'magrittr'
The following object is masked from 'package:purrr':


    set_names
The following object is masked from 'package:tidyr':


    extract
Attaching package: 'jsonlite'
The following object is masked from 'package:purrr':


    flatten
Attaching package: 'data.table'
The following objects are masked from 'package:dplyr':


    between, first, last
The following object is masked from 'package:purrr':


    transpose
Attaching package: 'lubridate'
The following objects are masked from 'package:data.table':


    hour, isoweek, mday, minute, month, quarter, second, wday, week,
    yday, year
The following object is masked from 'package:base':


    date
Attaching package: 'scales'
```

```
The following object is masked from 'package:purrr':

    discard
The following object is masked from 'package:readr':

    col_factor
Attaching package: 'zoo'
The following objects are masked from 'package:base':

    as.Date, as.Date.numeric
Loading required package: lattice
Attaching package: 'mice'
The following object is masked from 'package:tidyr':

    complete
The following objects are masked from 'package:base':

    cbind, rbind
Highcharts (www.highcharts.com) is a Highsoft software product which is
not free for commercial and Governmental use
Attaching package: 'GGally'
The following object is masked from 'package:dplyr':

    nasa
```

```
[1] "train_v1.csv"
Parsed with column specification:
cols(
  channelGrouping = col_character(),
  date = col_double(),
  device = col_character(),
  fullVisitorId = col_character(),
  geoNetwork = col_character(),
  sessionId = col_character(),
  socialEngagementType = col_character(),
  totals = col_character(),
  trafficSource = col_character(),
  visitId = col_double(),
  visitNumber = col_double(),
  visitStartTime = col_double()
)
Classes 'tbl_df', 'tbl' and 'data.frame':    903653 obs. of  12 variables:
 $ channelGrouping     : chr  "Organic Search" "Organic Search" "Organic Search" "Organic Search" ...
 $ date                : num  20160902 20160902 20160902 20160902 20160902 ...
 $ device              : chr  "{\"browser\": \"Chrome\", \"browserVersion\": \"not available in demo dataset\", \"browserSize\": \"not availab| __truncated__ "{\"browser\": \"Firefox\", \
 $ fullVisitorId       : chr  "1131660440785968503" "377306020877927890" "3895546263509774583" "4763447161404445595" ...
 $ geoNetwork          : chr  "{\"continent\": \"Asia\", \"subContinent\": \"Western Asia\", \"country\": \"Turkey\", \"region\": \"Izmir\", \"| __truncated__ "{\"continent\": \"Oceania\",
 $ sessionId           : chr  "1131660440785968503_1472830385" "377306020877927890_1472880147" "3895546263509774583_1472865386" "4763447161404445595_1472881213" ...
 $ socialEngagementType: chr  "Not Socially Engaged" "Not Socially Engaged" "Not Socially Engaged" "Not Socially Engaged" ...
 $ totals              : chr  "{\"visits\": \"1\", \"hits\": \"1\", \"pageviews\": \"1\", \"bounces\": \"1\", \"newVisits\": \"1\"}" "{\"visits\": \"1\", \"hits\": \"1\", \"pageviews\": \"
 $ trafficSource       : chr  "{\"campaign\": \"(not set)\", \"source\": \"google\", \"medium\": \"organic\", \"keyword\": \"(not provided)\","| __truncated__ "{\"campaign\": \"(not set)\"
 $ visitId             : num  1.47e+09 1.47e+09 1.47e+09 1.47e+09 1.47e+09 ...
 $ visitNumber         : num  1 1 1 1 2 1 1 1 1 1 ...
 $ visitStartTime      : num  1.47e+09 1.47e+09 1.47e+09 1.47e+09 1.47e+09 ...
 - attr(*, "spec")=
  .. cols(
  ..   channelGrouping = col_character(),
  ..   date = col_double(),
  ..   device = col_character(),
```

```
    ..    fullVisitorId = col_character(),
    ..    geoNetwork = col_character(),
    ..    sessionId = col_character(),
    ..    socialEngagementType = col_character(),
    ..    totals = col_character(),
    ..    trafficSource = col_character(),
    ..    visitId = col_double(),
    ..    visitNumber = col_double(),
    ..    visitStartTime = col_double()
    .. )
[1] "DEVICE"
[1] "continent"       "subContinent"    "country"          "region"
 [5] "metro"           "city"            "cityId"           "networkDomain"
 [9] "latitude"        "longitude"       "networkLocation"
[1] "GEO NETWORK"
[1] "continent"       "subContinent"    "country"          "region"
 [5] "metro"           "city"            "cityId"           "networkDomain"
 [9] "latitude"        "longitude"       "networkLocation"
[1] "TOTALS"
[1] "visits"          "hits"              "pageviews"
[4] "bounces"         "newVisits"         "transactionRevenue"
[1] "TRAFFIC SOURCE"
[1] "campaign"                              "source"
 [3] "medium"                               "keyword"
 [5] "isTrueDirect"                         "referralPath"
 [7] "adContent"                            "campaignCode"
 [9] "adwordsClickInfo.criteriaParameters"  "adwordsClickInfo.page"
[11] "adwordsClickInfo.slot"                "adwordsClickInfo.gclId"
[13] "adwordsClickInfo.adNetworkType"       "adwordsClickInfo.isVideoAd"
[1] "TOTALS"
'data.frame':    903653 obs. of  6 variables:
 $ visits          : chr  "1" "1" "1" "1" ...
 $ totalhits       : chr  "1" "1" "1" "1" ...
 $ pageviews       : chr  "1" "1" "1" "1" ...
 $ bounces         : chr  "1" "1" "1" "1" ...
```

```
[1] "TRAFFIC SOURCE"
[1] "campaign"                           "source"
 [3] "medium"                             "keyword"
 [5] "isTrueDirect"                       "referralPath"
 [7] "adContent"                          "campaignCode"
 [9] "adwordsClickInfo.criteriaParameters" "adwordsClickInfo.page"
[11] "adwordsClickInfo.slot"              "adwordsClickInfo.gclId"
[13] "adwordsClickInfo.adNetworkType"     "adwordsClickInfo.isVideoAd"
[1] "TOTALS"
'data.frame':   903653 obs. of  6 variables:
 $ visits           : chr  "1" "1" "1" "1" ...
 $ totalhits        : chr  "1" "1" "1" "1" ...
 $ pageviews        : chr  "1" "1" "1" "1" ...
 $ bounces          : chr  "1" "1" "1" "1" ...
 $ newVisits        : chr  "1" "1" "1" "1" ...
 $ transactionRevenue: chr  NA NA NA NA ...

Observations: 903,653
Variables: 55
$ channelGrouping           <chr> "Organic Search", "Organic Sear...
$ date                      <dbl> 20160902, 20160902, 20160902, 2...
$ fullVisitorId             <chr> "1131660440785968503", "3773060...
$ sessionId                 <chr> "1131660440785968503_1472830385...
$ socialEngagementType      <chr> "Not Socially Engaged", "Not So...
$ visitId                   <dbl> 1472830385, 1472880147, 1472865...
$ visitNumber               <dbl> 1, 1, 1, 1, 2, 1, 1, 1, 1, 1, 1...
$ visitStartTime            <dbl> 1472830385, 1472880147, 1472865...
$ browser                   <chr> "Chrome", "Firefox", "Chrome", ...
$ browserVersion            <chr> "not available in demo dataset"...
$ browserSize               <chr> "not available in demo dataset"...
$ operatingSystem           <chr> "Windows", "Macintosh", "Window...
$ operatingSystemVersion    <chr> "not available in demo dataset"...
$ isMobile                  <lgl> FALSE, FALSE, FALSE, FALSE, TRU...
$ mobileDeviceBranding      <chr> "not available in demo dataset"...
$ mobileDeviceModel         <chr> "not available in demo dataset"...
$ mobileInputSelector       <chr> "not available in demo dataset"...
```

```
$ mobileDeviceInfo                   <chr> "not available in demo dataset"...
$ mobileDeviceMarketingName          <chr> "not available in demo dataset"...
$ flashVersion                       <chr> "not available in demo dataset"...
$ language                           <chr> "not available in demo dataset"...
$ screenColors                       <chr> "not available in demo dataset"...
$ screenResolution                   <chr> "not available in demo dataset"...
$ deviceCategory                     <chr> "desktop", "desktop", "desktop"...
$ continent                          <chr> "Asia", "Oceania", "Europe", "A...
$ subContinent                       <chr> "Western Asia", "Australasia", ...
$ country                            <chr> "Turkey", "Australia", "Spain",...
$ region                             <chr> "Izmir", "not available in demo...
$ metro                              <chr> "(not set)", "not available in ...
$ city                               <chr> "Izmir", "not available in demo...
$ cityId                             <chr> "not available in demo dataset"...
$ networkDomain                      <chr> "ttnet.com.tr", "dodo.net.au", ...
$ latitude                           <chr> "not available in demo dataset"...
$ longitude                          <chr> "not available in demo dataset"...
$ networkLocation                    <chr> "not available in demo dataset"...
$ visits                             <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1...
$ totalhits                          <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1...
$ pageviews                          <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1...
$ bounces                            <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1...
$ newVisits                          <dbl> 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1...
$ transactionRevenue                 <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0...
$ campaign                           <chr> "(not set)", "(not set)", "(not...
$ source                             <chr> "google", "google", "google", "...
$ medium                             <chr> "organic", "organic", "organic"...
$ keyword                            <chr> "(not provided)", "(not provide...
$ isTrueDirect                       <lgl> NA, NA, NA, NA, TRUE, NA, NA, N...
$ referralPath                       <chr> NA, NA, NA, NA, NA, NA, NA, NA,...
$ adContent                          <chr> NA, NA, NA, NA, NA, NA, NA, NA,...
$ campaignCode                       <chr> NA, NA, NA, NA, NA, NA, NA, NA,...
$ adwordsClickInfo.criteriaParameters <chr> "not available in demo dataset"...
$ adwordsClickInfo.page              <chr> NA, NA, NA, NA, NA, NA, NA, NA,...
$ adwordsClickInfo.slot              <chr> NA, NA, NA, NA, NA, NA, NA, NA,...
$ adwordsClickInfo.gclId             <chr> NA, NA, NA, NA, NA, NA, NA, NA,...
```

```
$ adwordsClickInfo.gclId          <chr> NA, NA, NA, NA, NA, NA, NA, NA,...
$ adwordsClickInfo.adNetworkType   <chr> NA, NA, NA, NA, NA, NA, NA, NA,...
$ adwordsClickInfo.isVideoAd       <lgl> NA, NA, NA, NA, NA, NA, NA, NA,...

[1] "CUSTOMIZE DATASET: creating additional dataset sample from whole dataset: shrinking size down to ~40K"
[1] "whole training dataset: % of non=zero revenue data= 1.27"
[1] "sample training dataset: % of non=zero revenue data= 34.05"
```



```
[1] "Missing (75% or more) train data features identified and removed: adwordsClickInfo.criteriaParameters, networkLocation, latitude, longitude, cityId, screenResolution, screenColors, la
[1] "mobileDeviceMarketingName, mobileDeviceInfo, mobileInputSelector, mobileDeviceModel, mobileDeviceBranding, operatingSystemVersion, browserSize, browserVersion, campaignCode, metro,"
[1] "adwordsClickInfo.page, adwordsClickInfo.slot, adwordsClickInfo.gclId, adwordsClickInfo.adNetworkType, adwordsClickInfo.isVideoAd, keyword, campaign,adContent"
        channelGrouping              date        fullVisitorId
                   8               366               31417
   socialEngagementType          visitId         visitNumber
                   1             33792                 152
       visitStartTime           browser      operatingSystem
               33792                26                  14
             isMobile     deviceCategory           continent
                   2                 3                   6
          subContinent          country              region
                  22               160                 261
                 city      networkDomain              visits
                 388              3220                   1
             totalhits         pageviews             bounces
                 217               170                   2
            newVisits  transactionRevenue             source
                   2              5333                 105
```

```
             medium          isTrueDirect
                  6                     2
```



```
[1] "Feature variance displayed."
[1] "socialEngagementType & visits removed from train dataset"
'data.frame':   33818 obs. of  24 variables:
 $ channelGrouping   : chr  "Direct" "Organic Search" "Referral" "Referral" ...
 $ date              : num  20160902 20160902 20160902 20160902 20160902 ...
 $ fullVisitorId     : chr  "6194193421514403509" "5327166854580374902" "8885051388942907862" "0185467632009737931" ...
 $ visitId           : num  1.47e+09 1.47e+09 1.47e+09 1.47e+09 1.47e+09 ...
 $ visitNumber       : num  1 3 7 6 4 2 1 13 1 7 ...
 $ visitStartTime    : num  1.47e+09 1.47e+09 1.47e+09 1.47e+09 1.47e+09 ...
 $ browser           : chr  "Chrome" "Chrome" "Chrome" "Chrome" ...
 $ operatingSystem   : chr  "Linux" "Macintosh" "Linux" "Windows" ...
 $ isMobile          : logi  FALSE FALSE FALSE FALSE FALSE FALSE ...
 $ deviceCategory    : chr  "desktop" "desktop" "desktop" "desktop" ...
 $ continent         : chr  "Americas" "Americas" "Americas" "Americas" ...
 $ subContinent      : chr  "Northern America" "Northern America" "Northern America" "Northern America" ...
 $ country           : chr  "United States" "United States" "United States" "United States" ...
 $ region            : chr  "Michigan" "New York" "New York" "California" ...
 $ city              : chr  "Ann Arbor" "New York" "New York" "Mountain View" ...
 $ networkDomain     : chr  NA NA NA NA ...
 $ totalhits         : num  11 11 13 13 17 17 19 20 22 22 ...
 $ pageviews         : num  11 10 11 12 14 17 16 15 16 20 ...
 $ bounces           : num  0 0 0 0 0 0 0 0 0 0 ...
 $ newVisits         : num  1 0 0 0 0 1 0 1 0 0 ...
 $ transactionRevenue: num  3.79e+07 3.07e+08 6.80e+07 2.62e+07 5.74e+08 ...
 $ source            : chr  "(direct)" "google" "mall.googleplex.com" "mall.googleplex.com" ...
 $ medium            : chr  NA "organic" "referral" "referral" ...
 $ isTrueDirect      : logi  TRUE TRUE TRUE TRUE TRUE TRUE ...
```

```
[1] "2016-08-01" "2017-08-01"
```



```
channelGrouping              date     fullVisitorId            visitId
              0                 0                 0                  0
    visitNumber           browser   operatingSystem           isMobile
              0                 0               109                  0
 deviceCategory         continent       subContinent            country
              0                36                36                 36
         region              city     networkDomain          totalhits
          17996             18207             16621                  0
       pageviews           bounces         newVisits transactionRevenue
              0                 0                 0                  0
         source            medium        isTrueDirect              year
              0              5724                 0                  0
          month              hour              wday               yday
              0                 0                 0                  0
             YM
              0
[1] "There are no numeric missing values."
[1] "Categorical missing values for browser (8), source (69), continent (1468), subContinent (1468), country (1468) and operating system (4695) range from 0.01% - 0.52%."
[1] "Features with 5% or less missing data will be imputed. All others will be assigned the values missing"
[1] "*****************       Review of data - traget feature     ***************"
 Min.   1st Qu.    Median      Mean   3rd Qu.       Max.
0.000e+00 0.000e+00 0.000e+00 4.554e+07 2.590e+07 2.313e+10
[1] "*****************       distribution is right skewed; mean is greater than median     *************"
 Min.   1st Qu.   Median     Mean  3rd Qu.     Max.
  0.00     0.00     0.00    45.54   25.90 23129.50
[1] "*****************       distribution is right skewed; mean is greater than median     *************"
```

```
[1] FALSE
```
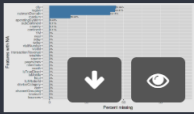


```
[1] "Distribution tending towards normal when using the natural log of the target variable"
[1] "transaction revenue skewness actual vs log vs log1p"
[1] 41.39671
[1] NaN
[1] 0.6950806
[1] "transaction revenue skewness for purchases only actual vs log vs log1p"
[1] 25.716
[1] 0.3374775
[1] 0.3374791
[1] "Base Observations"
[1] "Total number of observations/visits = 33818"
[1] "Total unique visitors = 31417"
[1] "Visitors who made at least one purchase = 9996"
[1] "Total number of revenue generating visits =  11515"
[1] "Only 34.05 % of visits generated revenue"
[1] "Total new visits = 21994"
[1] "Total revenue generated (/10e6) = 1540071.24"
[1] "Total page views = 404456"
[1] "Total Hits = 510043"
[1] "Total bounces = 11329"
```

```
channelGrouping        date              fullVisitorId         visitId
 Length:11515     Min.   :2016-08-01     Length:11515     Min.   :1.470e+09
 Class :character  1st Qu.:2016-11-01    Class :character  1st Qu.:1.478e+09
 Mode  :character  Median :2017-01-27    Mode  :character  Median :1.486e+09
                   Mean   :2017-01-30                      Mean   :1.486e+09
                   3rd Qu.:2017-05-04                      3rd Qu.:1.494e+09
                   Max.   :2017-08-01                      Max.   :1.502e+09


  visitNumber        browser          operatingSystem        isMobile
 Min.   :  1.000  Length:11515      Length:11515      Min.   :0.00000
 1st Qu.:  1.000  Class :character  Class :character  1st Qu.:0.00000
 Median :  2.000  Mode  :character  Mode  :character  Median :0.00000
 Mean   :  4.083                                      Mean   :0.08858
 3rd Qu.:  4.000                                      3rd Qu.:0.00000
 Max.   :315.000                                      Max.   :1.00000


 deviceCategory       continent          subContinent          country
 Length:11515      Length:11515      Length:11515      Length:11515
 Class :character  Class :character  Class :character  Class :character
 Mode  :character  Mode  :character  Mode  :character  Mode  :character




    region              city            networkDomain         totalhits
 Length:11515      Length:11515      Length:11515      Min.   :  2.00
 Class :character  Class :character  Class :character  1st Qu.: 19.00
 Mode  :character  Mode  :character  Mode  :character  Median : 28.00
                                                       Mean   : 36.27
                                                       3rd Qu.: 44.00
                                                       Max.   :500.00
```

```
    pageviews          bounces       newVisits       transactionRevenue
 Min.   :  2.00    Min.   :0     Min.   :0.0000    Min.   :1.000e+04
 1st Qu.: 16.00    1st Qu.:0     1st Qu.:0.0000    1st Qu.:2.493e+07
 Median : 23.00    Median :0     Median :0.0000    Median :4.945e+07
 Mean   : 28.33    Mean   :0     Mean   :0.3878    Mean   :1.337e+08
 3rd Qu.: 34.00    3rd Qu.:0     3rd Qu.:1.0000    3rd Qu.:1.077e+08
 Max.   :469.00    Max.   :0     Max.   :1.0000    Max.   :2.313e+10

    source             medium          isTrueDirect         year
 Length:11515      Length:11515      Min.   :0.0000    Min.   :2016
 Class :character  Class :character  1st Qu.:0.0000    1st Qu.:2016
 Mode  :character  Mode  :character  Median :1.0000    Median :2017
                                     Mean   :0.6083    Mean   :2017
                                     3rd Qu.:1.0000    3rd Qu.:2017
                                     Max.   :1.0000    Max.   :2017

    month             hour             wday             yday
 Min.   : 1.000   Min.   : 0.00   Min.   :1.000    Min.   :  1.0
 1st Qu.: 4.000   1st Qu.: 7.00   1st Qu.:3.000    1st Qu.:110.0
 Median : 7.000   Median :17.00   Median :4.000    Median :198.0
 Mean   : 6.912   Mean   :14.13   Mean   :4.009    Mean   :195.1
 3rd Qu.:10.000   3rd Qu.:20.00   3rd Qu.:5.000    3rd Qu.:285.0
 Max.   :12.000   Max.   :23.00   Max.   :7.000    Max.   :366.0

       YM            revenue         transactionRevenue1 transactionRevenue2
 Dec 2016:1409   Min.   :    0.01   Min.   : 9.21        Min.   : 9.21
 May 2017:1118   1st Qu.:   24.93   1st Qu.:17.03        1st Qu.:17.03
 Aug 2016:1117   Median :   49.45   Median :17.72        Median :17.72
 Jul 2017:1024   Mean   :  133.75   Mean   :17.82        Mean   :17.82
 Jun 2017: 947   3rd Qu.:  107.66   3rd Qu.:18.49        3rd Qu.:18.49
 Apr 2017: 921   Max.   :23129.50   Max.   :23.86        Max.   :23.86
 (Other) :4979
```

```
   purchaseYN
 Min.    :1
 1st Qu.:1
 Median :1
 Mean    :1
 3rd Qu.:1
 Max.    :1


channelGrouping          date              fullVisitorId           visitId
 Length:22303       Min.    :2016-08-01    Length:22303       Min.    :1.470e+09
 Class :character   1st Qu.:2016-10-26     Class :character   1st Qu.:1.477e+09
 Mode  :character   Median :2017-01-08     Mode  :character   Median :1.484e+09
                    Mean    :2017-01-19                       Mean    :1.485e+09
                    3rd Qu.:2017-04-19                        3rd Qu.:1.493e+09
                    Max.    :2017-08-01                       Max.    :1.502e+09


  visitNumber          browser            operatingSystem         isMobile
 Min.    :  1.000   Length:22303       Length:22303        Min.    :0.0000
 1st Qu.:  1.000    Class :character   Class :character    1st Qu.:0.0000
 Median :  1.000    Mode  :character   Mode  :character    Median :0.0000
 Mean    :  2.211                                          Mean    :0.2679
 3rd Qu.:  1.000                                           3rd Qu.:1.0000
 Max.    :317.000                                          Max.    :1.0000


 deviceCategory       continent          subContinent            country
 Length:22303       Length:22303       Length:22303        Length:22303
 Class :character   Class :character   Class :character    Class :character
 Mode  :character   Mode  :character   Mode  :character    Mode  :character
```

```
     region                 city              networkDomain                totalhits
 Length:22303          Length:22303          Length:22303          Min.    :  1.00
 Class :character      Class :character      Class :character      1st Qu.:  1.00
 Mode  :character      Mode  :character      Mode  :character      Median :  1.00
                                                                   Mean    :  4.14
                                                                   3rd Qu.:  4.00
                                                                   Max.    :213.00


    pageviews              bounces              newVisits           transactionRevenue
 Min.    :  0.000      Min.    :0.000       Min.    :0.0000       Min.    :0
 1st Qu.:  1.000       1st Qu.:0.000        1st Qu.:1.0000        1st Qu.:0
 Median :  1.000       Median :1.000        Median :1.0000        Median :0
 Mean    :  3.508      Mean    :0.508       Mean    :0.7859       Mean    :0
 3rd Qu.:  3.000       3rd Qu.:1.000        3rd Qu.:1.0000        3rd Qu.:0
 Max.    :145.000      Max.    :1.000       Max.    :1.0000       Max.    :0


     source                medium              isTrueDirect               year
 Length:22303          Length:22303          Min.    :0.0000       Min.    :2016
 Class :character      Class :character      1st Qu.:0.0000        1st Qu.:2016
 Mode  :character      Mode  :character      Median :0.0000        Median :2017
                                             Mean    :0.3027       Mean    :2017
                                             3rd Qu.:1.0000        3rd Qu.:2017
                                             Max.    :1.0000       Max.    :2017


     month                  hour                 wday                   yday
 Min.    : 1.000       Min.    : 0.00       Min.    :1.000        Min.    :  1.0
 1st Qu.: 4.000        1st Qu.: 7.00        1st Qu.:3.000         1st Qu.:102.0
 Median : 7.000        Median :14.00        Median :4.000         Median :207.0
 Mean    : 6.951       Mean    :12.53       Mean    :4.029        Mean    :196.5
 3rd Qu.:10.000        3rd Qu.:18.00        3rd Qu.:6.000         3rd Qu.:296.0
 Max.    :12.000       Max.    :23.00       Max.    :7.000        Max.    :366.0
```

```
       YM              revenue    transactionRevenue1 transactionRevenue2
Nov 2016:2811    Min.    :0    Min.    :-Inf     Min.    :0
Oct 2016:2349    1st Qu.:0    1st Qu.:-Inf     1st Qu.:0
Sep 2016:1868    Median :0    Median :-Inf     Median :0
Dec 2016:1864    Mean    :0    Mean    :-Inf     Mean    :0
Aug 2016:1861    3rd Qu.:0    3rd Qu.:-Inf     3rd Qu.:0
Mar 2017:1781    Max.    :0    Max.    :-Inf     Max.    :0
(Other) :9769
   purchaseYN
Min.    :0
1st Qu.:0
Median :0
Mean    :0
3rd Qu.:0
Max.    :0
```

```
   Visits    Current_visit Next_visit retention
1   1 to 2      21994          4716       0.21
2   2 to 3       4716          2268       0.48
3   3 to 4       2268          1333       0.59
4   4 to 5       1333           807       0.61
5   5 to 6        807           550       0.68
6   6 to 7        550           383       0.70
7   7 to 8        383           276       0.72
8   8 to 9        276           231       0.84
9   9 to 10       231           185       0.80
10 10 to 11       185           117       0.63
11 11 to 12       117           106       0.91
12 12 to 13       106            84       0.79
13 13 to 14        84            81       0.96
14 14 to 15        81            50       0.62
15 15 to 16        50            52       1.04
16 16 to 17        52            48       0.92
17 17 to 18        48            35       0.73
18 18 to 19        35            30       0.86
19 19 to 20        30            24       0.80
20 20 to NA        24            NA         NA
```

Warning message:
"Removed 56 rows containing missing values (geom_point)."
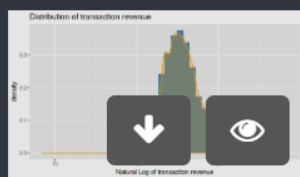
`geom_smooth()` using method = 'loess' and formula 'y ~ x'

```
`geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



```
[1] "Features with 5% or less missing data will be imputed. All other categorical NAs will be replaced with the word Missing"
   channelGrouping          fullVisitorId                visitId
                 0                      0                      0
       visitNumber                browser        operatingSystem
                 0                      0                      0
          isMobile         deviceCategory              continent
                 0                      0                      0
       subContinent                country                 region
                 0                      0                      0
              city          networkDomain               totalhits
                 0                      0                      0
         pageviews                bounces              newVisits
                 0                      0                      0
 transactionRevenue                 source                 medium
                 0                      0                      0
       isTrueDirect                   year                  month
                 0                      0                      0
              hour                   wday                   yday
                 0                      0                      0
                YM    transactionRevenueLog
                 0                      0
 [1] "channelGrouping" "browser"          "operatingSystem" "deviceCategory"
 [5] "continent"       "subContinent"     "country"          "region"
 [9] "city"            "networkDomain"    "source"           "medium"
[13] "YM"              "browser_dev"      "browser_os"       "browser_chan"
[17] "chan_os"         "country_medium"   "country_source"   "dev_chan"
```

```
'data.frame':   33818 obs. of  34 variables:
 $ channelGrouping      : Factor w/ 8 levels "(Other)","Affiliates",..: 3 5 7 7 7 7 3 7 5 7 ...
 $ visitNumber          : num  1 3 7 6 4 2 1 13 1 7 ...
 $ browser              : Factor w/ 26 levels "Amazon Silk",..: 6 6 6 6 6 6 19 6 6 6 ...
 $ operatingSystem      : Factor w/ 14 levels "Android","BlackBerry",..: 6 7 6 12 7 6 7 6 1 3 ...
 $ isMobile             : int  0 0 0 0 0 0 0 0 1 0 ...
 $ deviceCategory       : Factor w/ 3 levels "desktop","mobile",..: 1 1 1 1 1 1 1 1 2 1 ...
 $ continent            : Factor w/ 6 levels "Africa","Americas",..: 2 2 2 2 2 2 2 2 2 2 ...
 $ subContinent         : Factor w/ 22 levels "Australasia",..: 12 12 12 12 12 12 12 12 12 12 ...
 $ country              : Factor w/ 160 levels "Afghanistan",..: 154 154 154 154 154 154 154 154 154 154 ...
 $ region               : Factor w/ 261 levels "Abu Dhabi","Adana",..: 139 154 154 42 142 42 42 252 42 42 ...
 $ city                 : Factor w/ 388 levels "Aalborg","Abu Dhabi",..: 15 238 238 221 211 221 313 326 221 221 ...
 $ networkDomain        : Factor w/ 3220 levels "012.net.il","018.net.il",..: 1678 1678 1678 1678 1678 1678 593 1678 750 1678 ...
 $ totalhits            : num  11 11 13 13 17 17 19 20 22 22 ...
 $ pageviews            : num  11 10 11 12 14 17 16 15 16 20 ...
 $ bounces              : num  0 0 0 0 0 0 0 0 0 0 ...
 $ newVisits            : num  1 0 0 0 0 1 0 1 0 0 ...
 $ transactionRevenue   : num  3.79e+07 3.07e+08 6.80e+07 2.62e+07 5.74e+08 ...
 $ source               : Factor w/ 105 levels "(direct)","0.shared.bow.cat2.ads-bow.yw.borg.google.com:9895",..: 1 31 66 66 66 66 1 66 31 92 ...
 $ medium               : Factor w/ 6 levels "affiliate","cpc",..: 4 5 6 6 6 6 4 6 5 6 ...
 $ isTrueDirect         : int  1 1 1 1 1 1 1 1 0 1 ...
 $ year                 : num  2016 2016 2016 2016 2016 ...
 $ month                : num  9 9 9 9 9 9 9 9 9 9 ...
 $ hour                 : int  19 19 14 19 14 22 0 17 21 22 ...
 $ wday                 : num  6 6 6 6 6 6 7 6 6 6 ...
 $ yday                 : num  246 246 246 246 246 246 247 246 246 246 ...
 $ YM                   : Factor w/ 13 levels "Aug 2016","Sep 2016",..: 2 2 2 2 2 2 2 2 2 2 ...
 $ transactionRevenueLog: num  17.4 19.5 18 17.1 20.2 ...
 $ browser_dev          : Factor w/ 45 levels "Amazon Silk_tablet",..: 7 7 7 7 7 7 34 7 8 7 ...
 $ browser_os           : Factor w/ 54 levels "Amazon Silk_Android",..: 9 10 9 12 10 9 43 9 6 7 ...
 $ browser_chan         : Factor w/ 93 levels "Amazon Silk_Direct",..: 18 20 22 22 22 22 76 22 20 22 ...
 $ chan_os              : Factor w/ 62 levels "Android_Affiliates",..: 29 38 33 60 40 33 36 33 4 15 ...
 $ country_medium       : Factor w/ 474 levels "Afghanistan_organic",..: 455 456 457 457 457 457 455 457 456 457 ...
 $ country_source       : Factor w/ 864 levels "Afghanistan_google",..: 765 789 805 805 805 805 765 805 789 826 ...
 $ dev_chan             : Factor w/ 22 levels "desktop_Affiliates",..: 2 4 6 6 6 6 2 6 12 6 ...
```

```
Observations: 33,818
Variables: 33
$ channelGrouping       <int> 3, 5, 7, 7, 7, 7, 3, 7, 5, 7, 5, 7, 7, 5, 7, ...
$ visitNumber           <dbl> 1, 3, 7, 6, 4, 2, 1, 13, 1, 7, 6, 3, 2, 1, 1,...
$ browser               <int> 6, 6, 6, 6, 6, 6, 19, 6, 6, 6, 6, 6, 6, 19, 6...
$ operatingSystem       <int> 6, 7, 6, 12, 7, 6, 7, 6, 1, 3, 12, 7, 6, 5, 7...
$ isMobile              <int> 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, ...
$ deviceCategory        <int> 1, 1, 1, 1, 1, 1, 1, 1, 2, 1, 1, 1, 1, 2, 1, ...
$ continent             <int> 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, ...
$ subContinent          <int> 12, 12, 12, 12, 12, 12, 12, 12, 12, 12, 12, 1...
$ country               <int> 154, 154, 154, 154, 154, 154, 154, 154, 154, ...
$ region                <int> 139, 154, 154, 42, 142, 42, 42, 252, 42, 42, ...
$ city                  <int> 15, 238, 238, 221, 211, 221, 313, 326, 221, 2...
$ totalhits             <dbl> 11, 11, 13, 13, 17, 17, 19, 20, 22, 22, 22, 2...
$ pageviews             <dbl> 11, 10, 11, 12, 14, 17, 16, 15, 16, 20, 16, 2...
$ bounces               <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ...
$ newVisits             <dbl> 1, 0, 0, 0, 0, 0, 1, 0, 1, 0, 0, 0, 0, 1, 1, ...
$ transactionRevenue    <dbl> 37860000, 306670000, 68030000, 26250000, 5741...
$ source                <int> 1, 31, 66, 66, 66, 66, 1, 66, 31, 92, 31, 66,...
$ medium                <int> 4, 5, 6, 6, 6, 6, 4, 6, 5, 6, 5, 6, 6, 5, 6, ...
$ isTrueDirect          <int> 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 0, 0, ...
$ year                  <dbl> 2016, 2016, 2016, 2016, 2016, 2016, 2016, 201...
$ month                 <dbl> 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, ...
$ hour                  <int> 19, 19, 14, 19, 14, 22, 0, 17, 21, 22, 14, 17...
$ wday                  <dbl> 6, 6, 6, 6, 6, 6, 7, 6, 6, 6, 6, 6, 6, 6, 7, ...
$ yday                  <dbl> 246, 246, 246, 246, 246, 246, 247, 246, 246, ...
$ YM                    <int> 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, ...
$ transactionRevenueLog <dbl> 17.44941, 19.54128, 18.03546, 17.08318, 20.16...
$ browser_dev           <int> 7, 7, 7, 7, 7, 7, 34, 7, 8, 7, 7, 7, 7, 35, 7...
$ browser_os            <int> 9, 10, 9, 12, 10, 9, 43, 9, 6, 7, 12, 10, 9, ...
$ browser_chan          <int> 18, 20, 22, 22, 22, 22, 76, 22, 20, 22, 20, 2...
$ chan_os               <int> 29, 38, 33, 60, 40, 33, 36, 33, 4, 15, 58, 40...
$ country_medium        <int> 455, 456, 457, 457, 457, 457, 455, 457, 456, ...
$ country_source        <int> 765, 789, 805, 805, 805, 805, 765, 805, 789, ...
$ dev_chan              <int> 2, 4, 6, 6, 6, 6, 2, 6, 12, 6, 4, 6, 6, 12, 6...
```

```
[1] "traintest.coef"


channelGrouping visitNumber browser operatingSystem
channelGrouping          1.00         -0.04      0.04      0.13
visitNumber             -0.04          1.00     -0.05      0.05
browser                  0.04         -0.05      1.00     -0.14
operatingSystem          0.13          0.05     -0.14      1.00
isMobile                -0.22         -0.06      0.33     -0.64
deviceCategory          -0.21         -0.06      0.34     -0.59
continent                0.02         -0.08      0.12      0.12
subContinent             0.06         -0.03      0.05      0.03
country                 -0.03          0.07     -0.08     -0.11
region                  -0.06          0.03      0.06      0.08
city                    -0.04          0.02     -0.05     -0.02
totalhits               -0.01          0.08     -0.16     -0.04
pageviews               -0.02          0.08     -0.17     -0.04
bounces                  0.05         -0.06      0.16      0.08
newVisits                0.00         -0.25      0.18      0.07
transactionRevenue      -0.03          0.24     -0.06      0.00
source                   0.87         -0.05      0.09      0.17
medium                   0.75         -0.04     -0.01      0.09
isTrueDirect            -0.38          0.19     -0.14     -0.11
year                    -0.17          0.02     -0.07     -0.08
month                    0.15         -0.02      0.07      0.05
hour                    -0.01          0.04     -0.07     -0.01
wday                     0.00          0.00      0.01     -0.02
yday                     0.15         -0.02      0.07      0.05
YM                      -0.15          0.02     -0.05     -0.09
transactionRevenueLog    0.00          0.10     -0.22     -0.06
browser_dev              0.03         -0.05      1.00     -0.16
browser_os               0.07         -0.04      0.99      0.01
browser_chan             0.12         -0.05      0.99     -0.12
chan_os                  0.24          0.04     -0.13      0.99
country_medium          -0.03          0.07     -0.08     -0.12
country_source           0.01          0.07     -0.09     -0.12
dev_chan                 0.24         -0.08      0.36     -0.54
```

	isMobile	deviceCategory	continent	subContinent	country
channelGrouping	-0.22	-0.21	0.02	0.06	-0.03
visitNumber	-0.06	-0.06	-0.08	-0.03	0.07
browser	0.33	0.34	0.12	0.05	-0.08
operatingSystem	-0.64	-0.59	0.12	0.03	-0.11
isMobile	1.00	0.95	0.07	0.02	-0.07
deviceCategory	0.95	1.00	0.08	0.01	-0.06
continent	0.07	0.08	1.00	0.18	-0.46
subContinent	0.02	0.01	0.18	1.00	-0.17
country	-0.07	-0.06	-0.46	-0.17	1.00
region	0.06	0.07	0.12	-0.03	-0.21
city	-0.01	-0.01	-0.08	-0.12	0.01
totalhits	-0.15	-0.13	-0.28	-0.12	0.25
pageviews	-0.15	-0.14	-0.29	-0.12	0.26
bounces	0.12	0.10	0.33	0.14	-0.29
newVisits	0.12	0.12	0.24	0.11	-0.22
transactionRevenue	-0.07	-0.07	-0.10	-0.04	0.08
source	-0.23	-0.21	0.07	0.09	-0.08
medium	-0.21	-0.20	0.02	0.05	-0.03
isTrueDirect	-0.02	-0.02	-0.21	-0.10	0.18
year	0.11	0.09	-0.01	-0.03	0.00
month	-0.08	-0.07	-0.01	0.02	0.00
hour	-0.05	-0.05	-0.07	0.01	0.08
wday	0.02	0.02	-0.01	0.00	0.00
yday	-0.08	-0.07	-0.01	0.02	0.00
YM	0.10	0.09	-0.02	-0.03	0.01
transactionRevenueLog	-0.21	-0.20	-0.40	-0.16	0.37
browser_dev	0.36	0.38	0.12	0.05	-0.08
browser_os	0.22	0.24	0.13	0.06	-0.10
browser_chan	0.30	0.31	0.12	0.05	-0.08
chan_os	-0.70	-0.65	0.09	0.03	-0.09
country_medium	-0.07	-0.06	-0.47	-0.18	1.00
country_source	-0.08	-0.07	-0.49	-0.18	1.00
dev_chan	0.85	0.90	0.09	0.04	-0.07

	region	⋯	yday	YM	transactionRevenueLog	browser_dev
channelGrouping	-0.06	⋯	0.15	-0.15	0.00	0.03
visitNumber	0.03	⋯	-0.02	0.02	0.10	-0.05
browser	0.06	⋯	0.07	-0.05	-0.22	1.00
operatingSystem	0.08	⋯	0.05	-0.09	-0.06	-0.16
isMobile	0.06	⋯	-0.08	0.10	-0.21	0.36
deviceCategory	0.07	⋯	-0.07	0.09	-0.20	0.38
continent	0.12	⋯	-0.01	-0.02	-0.40	0.12
subContinent	-0.03	⋯	0.02	-0.03	-0.16	0.05
country	-0.21	⋯	0.00	0.01	0.37	-0.08
region	1.00	⋯	-0.02	-0.01	-0.10	0.06
city	-0.09	⋯	-0.03	0.03	0.06	-0.05
totalhits	-0.05	⋯	0.02	-0.04	0.64	-0.17
pageviews	-0.05	⋯	0.02	-0.04	0.67	-0.17
bounces	0.09	⋯	0.00	-0.03	-0.51	0.16
newVisits	0.09	⋯	0.02	-0.04	-0.40	0.18
transactionRevenue	-0.01	⋯	-0.01	0.01	0.27	-0.06
source	-0.04	⋯	0.17	-0.19	-0.07	0.08
medium	-0.07	⋯	0.10	-0.10	0.05	-0.02
isTrueDirect	-0.05	⋯	-0.06	0.05	0.30	-0.14
year	0.00	⋯	-0.86	0.87	0.03	-0.06
month	-0.02	⋯	1.00	-0.49	-0.01	0.06
hour	-0.04	⋯	-0.03	0.01	0.11	-0.07
wday	-0.01	⋯	0.00	0.01	-0.01	0.01
yday	-0.02	⋯	1.00	-0.50	-0.01	0.06
YM	-0.01	⋯	-0.50	1.00	0.05	-0.05
transactionRevenueLog	-0.10	⋯	-0.01	0.05	1.00	-0.22
browser_dev	0.06	⋯	0.06	-0.05	-0.22	1.00
browser_os	0.07	⋯	0.08	-0.07	-0.22	0.98
browser_chan	0.06	⋯	0.08	-0.06	-0.22	0.99
chan_os	0.05	⋯	0.07	-0.10	-0.03	-0.15
country_medium	-0.21	⋯	0.00	0.01	0.38	-0.08
country_source	-0.22	⋯	0.01	0.01	0.39	-0.09
dev_chan	0.04	⋯	0.00	0.02	-0.20	0.39

	browser_os	browser_chan	chan_os	country
channelGrouping	0.07	0.12	0.24	-0.03
visitNumber	-0.04	-0.05	0.04	0.07
browser	0.99	0.99	-0.13	-0.08
operatingSystem	0.01	-0.12	0.99	-0.12
isMobile	0.22	0.30	-0.70	-0.07
deviceCategory	0.24	0.31	-0.65	-0.06
continent	0.13	0.12	0.09	-0.47
subContinent	0.06	0.05	0.03	-0.18
country	-0.10	-0.08	-0.09	1.00
region	0.07	0.06	0.05	-0.21
city	-0.06	-0.06	-0.02	0.02
totalhits	-0.17	-0.16	-0.02	0.25
pageviews	-0.17	-0.17	-0.02	0.26
bounces	0.17	0.16	0.06	-0.29
newVisits	0.19	0.18	0.05	-0.23
transactionRevenue	-0.06	-0.06	0.01	0.09
source	0.12	0.15	0.27	-0.08
medium	0.00	0.04	0.18	-0.02
isTrueDirect	-0.16	-0.17	-0.14	0.18
year	-0.08	-0.08	-0.10	0.00
month	0.08	0.08	0.07	0.00
hour	-0.07	-0.07	-0.01	0.08
wday	0.01	0.01	-0.02	0.00
yday	0.08	0.08	0.07	0.00
YM	-0.07	-0.06	-0.10	0.01
transactionRevenueLog	-0.22	-0.22	-0.03	0.38
browser_dev	0.98	0.99	-0.15	-0.08
browser_os	1.00	0.99	0.02	-0.10
browser_chan	0.99	1.00	-0.10	-0.09
chan_os	0.02	-0.10	1.00	-0.09
country_medium	-0.10	-0.09	-0.09	1.00
country_source	-0.10	-0.09	-0.09	1.00
dev_chan	0.27	0.37	-0.55	-0.07

	country_source	dev_chan
channelGrouping	0.01	0.24
visitNumber	0.07	-0.08
browser	-0.09	0.36
operatingSystem	-0.12	-0.54
isMobile	-0.08	0.85
deviceCategory	-0.07	0.90
continent	-0.49	0.09
subContinent	-0.18	0.04
country	1.00	-0.07
region	-0.22	0.04
city	0.02	-0.02
totalhits	0.26	-0.14
pageviews	0.27	-0.14
bounces	-0.30	0.13
newVisits	-0.23	0.12
transactionRevenue	0.09	-0.08
source	-0.05	0.17
medium	0.00	0.14
isTrueDirect	0.17	-0.19
year	0.00	0.02
month	0.01	0.00
hour	0.08	-0.06
wday	0.00	0.02
yday	0.01	0.00
YM	0.01	0.02
transactionRevenueLog	0.39	-0.20
browser_dev	-0.09	0.39
browser_os	-0.10	0.27
browser_chan	-0.09	0.37
chan_os	-0.09	-0.55
country_medium	1.00	-0.07
country_source	1.00	-0.07
dev_chan	-0.07	1.00

```
[1] "matrix"

operatingSystem isMobile deviceCategory continent
channelGrouping           0.13          -0.22        -0.21         0.02
visitNumber               0.05          -0.06        -0.06        -0.08
browser                  -0.14           0.33         0.34         0.12
operatingSystem           1.00          -0.64        -0.59         0.12
isMobile                 -0.64           1.00         0.95         0.07
deviceCategory           -0.59           0.95         1.00         0.08
continent                 0.12           0.07         0.08         1.00
subContinent              0.03           0.02         0.01         0.18
country                  -0.11          -0.07        -0.06        -0.46
region                    0.08           0.06         0.07         0.12
city                     -0.02          -0.01        -0.01        -0.08
totalhits                -0.04          -0.15        -0.13        -0.28
pageviews                -0.04          -0.15        -0.14        -0.29
bounces                   0.08           0.12         0.10         0.33
newVisits                 0.07           0.12         0.12         0.24
transactionRevenue        0.00          -0.07        -0.07        -0.10
source                    0.17          -0.23        -0.21         0.07
medium                    0.09          -0.21        -0.20         0.02
isTrueDirect             -0.11          -0.02        -0.02        -0.21
year                     -0.08           0.11         0.09        -0.01
month                     0.05          -0.08        -0.07        -0.01
hour                     -0.01          -0.05        -0.05        -0.07
wday                     -0.02           0.02         0.02        -0.01
yday                      0.05          -0.08        -0.07        -0.01
YM                       -0.09           0.10         0.09        -0.02
transactionRevenueLog    -0.06          -0.21        -0.20        -0.40
browser_dev              -0.16           0.36         0.38         0.12
browser_os                0.01           0.22         0.24         0.13
browser_chan             -0.12           0.30         0.31         0.12
chan_os                   0.99          -0.70        -0.65         0.09
country_medium           -0.12          -0.07        -0.06        -0.47
country_source           -0.12          -0.08        -0.07        -0.49
dev_chan                 -0.54           0.85         0.90         0.09
```

	subContinent	country	region	city	totalhits	pagevie
channelGrouping	0.06	-0.03	-0.06	-0.04	-0.01	-0.02
visitNumber	-0.03	0.07	0.03	0.02	0.08	0.08
browser	0.05	-0.08	0.06	-0.05	-0.16	-0.17
operatingSystem	0.03	-0.11	0.08	-0.02	-0.04	-0.04
isMobile	0.02	-0.07	0.06	-0.01	-0.15	-0.15
deviceCategory	0.01	-0.06	0.07	-0.01	-0.13	-0.14
continent	0.18	-0.46	0.12	-0.08	-0.28	-0.29
subContinent	1.00	-0.17	-0.03	-0.12	-0.12	-0.12
country	-0.17	1.00	-0.21	0.01	0.25	0.26
region	-0.03	-0.21	1.00	-0.09	-0.05	-0.05
city	-0.12	0.01	-0.09	1.00	0.04	0.05
totalhits	-0.12	0.25	-0.05	0.04	1.00	0.99
pageviews	-0.12	0.26	-0.05	0.05	0.99	1.00
bounces	0.14	-0.29	0.09	-0.07	-0.41	-0.43
newVisits	0.11	-0.22	0.09	-0.06	-0.26	-0.27
transactionRevenue	-0.04	0.08	-0.01	0.01	0.25	0.24
source	0.09	-0.08	-0.04	-0.06	-0.07	-0.08
medium	0.05	-0.03	-0.07	-0.02	0.02	0.02
isTrueDirect	-0.10	0.18	-0.05	0.06	0.19	0.19
year	-0.03	0.00	0.00	0.04	-0.04	-0.04
month	0.02	0.00	-0.02	-0.03	0.02	0.02
hour	0.01	0.08	-0.04	0.01	0.04	0.04
wday	0.00	0.00	-0.01	0.01	0.00	0.00
yday	0.02	0.00	-0.02	-0.03	0.02	0.02
YM	-0.03	0.01	-0.01	0.03	-0.04	-0.04
transactionRevenueLog	-0.16	0.37	-0.10	0.06	0.64	0.67
browser_dev	0.05	-0.08	0.06	-0.05	-0.17	-0.17
browser_os	0.06	-0.10	0.07	-0.06	-0.17	-0.17
browser_chan	0.05	-0.08	0.06	-0.06	-0.16	-0.17
chan_os	0.03	-0.09	0.05	-0.02	-0.02	-0.02
country_medium	-0.18	1.00	-0.21	0.02	0.25	0.26
country_source	-0.18	1.00	-0.22	0.02	0.26	0.27
dev_chan	0.04	-0.07	0.04	-0.02	-0.14	-0.14

	bounces	newVisits	transactionRevenue	source	medium
channelGrouping	0.05	0.00	-0.03	0.87	0.75
visitNumber	-0.06	-0.25	0.24	-0.05	-0.04
browser	0.16	0.18	-0.06	0.09	-0.01
operatingSystem	0.08	0.07	0.00	0.17	0.09
isMobile	0.12	0.12	-0.07	-0.23	-0.21
deviceCategory	0.10	0.12	-0.07	-0.21	-0.20
continent	0.33	0.24	-0.10	0.07	0.02
subContinent	0.14	0.11	-0.04	0.09	0.05
country	-0.29	-0.22	0.08	-0.08	-0.03
region	0.09	0.09	-0.01	-0.04	-0.07
city	-0.07	-0.06	0.01	-0.06	-0.02
totalhits	-0.41	-0.26	0.25	-0.07	0.02
pageviews	-0.43	-0.27	0.24	-0.08	0.02
bounces	1.00	0.24	-0.12	0.12	0.01
newVisits	0.24	1.00	-0.14	0.08	-0.02
transactionRevenue	-0.12	-0.14	1.00	-0.04	-0.02
source	0.12	0.08	-0.04	1.00	0.60
medium	0.01	-0.02	-0.02	0.60	1.00
isTrueDirect	-0.18	-0.66	0.12	-0.37	-0.20
year	-0.02	-0.04	0.01	-0.21	-0.12
month	0.00	0.02	-0.01	0.17	0.10
hour	-0.07	-0.08	0.04	-0.03	0.02
wday	0.00	0.00	0.00	-0.01	0.00
yday	0.00	0.02	-0.01	0.17	0.10
YM	-0.03	-0.04	0.01	-0.19	-0.10
transactionRevenueLog	-0.51	-0.40	0.27	-0.07	0.05
browser_dev	0.16	0.18	-0.06	0.08	-0.02
browser_os	0.17	0.19	-0.06	0.12	0.00
browser_chan	0.16	0.18	-0.06	0.15	0.04
chan_os	0.06	0.05	0.01	0.27	0.18
country_medium	-0.29	-0.23	0.09	-0.08	-0.02
country_source	-0.30	-0.23	0.09	-0.05	0.00
dev_chan	0.13	0.12	-0.08	0.17	0.14