

# Google Analytics Sample - Analyse des Parcours des Navigation

Tung Anh

2025-07-30

## Introduction

Après une première phase d'exploration réalisée à partir des données téléchargées en local, j'ai poursuivi l'analyse des performances du site web en approfondissant les parcours de navigation des utilisateurs — au-delà des indicateurs classiques tels que le taux de rebond, la durée de session ou les pages vues. L'objectif principal est d'identifier les schémas de navigation, les points de sortie ainsi que les comportements récurrents ou inhabituels.

Cependant, pour analyser les parcours à partir des colonnes `hit`, il aurait été nécessaire de re-télécharger l'ensemble des données dans leur version complète, ce qui dépasse rapidement les quotas gratuits de Google Cloud Platform. Pour éviter cela, j'ai opté dans cette seconde partie pour une extraction ciblée directement depuis BigQuery, en ne récupérant que le sous-ensemble de données pertinent, incluant les variables manquantes, et ce, sans surcharger le stockage local.

## Configuration

```
# Charger les bibliothèques nécessaires
library(bigrquery)
library(DBI)
library(dplyr)
library(tidyr)
library(readr)
library(lubridate)
library(arrow)
library(glue)
library(here)
library(stringr)
```

```
bq_auth() # Authentication via un compte Google
```

## Plan d'analyse

La première partie de mon analyse a déjà mis en évidence un comportement anormal entre octobre et décembre 2016. Dans cette seconde phase, j'étends l'intervalle d'étude d'août 2016 à février 2017. L'objectif est non seulement d'analyser les comportements de navigation mois par mois, mais aussi de comparer les tendances observées avant octobre 2016 et après décembre 2016, afin de mieux cerner les écarts ou ruptures dans les parcours utilisateurs.

```

project_id <- "bigdata-382912"

# Requête SQL
query1 <- "
SELECT
  PARSE_DATE('%Y%m%d', _TABLE_SUFFIX) AS date,
  fullVisitorId,
  visitId,
  visitNumber,
  hits.hitNumber AS hitNumber,
  hits.page.pagePath AS pagePath,
  --dimension
  channelGrouping,
  totals.newVisits AS total_newVisits,
  trafficSource.source AS trafficSource_source,
  device.deviceCategory AS device_deviceCategory,
  geoNetwork.country AS geoNetwork_country
FROM
  `bigquery-public-data.google_analytics_sample.ga_sessions_*`,
  UNNEST(hits) AS hits
WHERE
  _TABLE_SUFFIX BETWEEN '20160801' AND '20170228'
  AND
  hits.type = 'PAGE'
ORDER BY
  fullVisitorId,
  visitId,
  visitNumber,
  hitNumber
"

# Charger l'output de query sur un df
df1 <- bq_project_query(project_id, query1) %>%
  bq_table_download()

# Vérifier le résultat
head(df1)

```

```

## # A tibble: 6 x 11
##   date      fullVisitorId      visitId visitNumber hitNumber pagePath
##   <date>      <chr>          <int>      <int>      <int> <chr>
## 1 2016-10-20 0000010278554503158 1477029466         1         1 /home
## 2 2016-10-20 0000010278554503158 1477029466         1         2 /google+redes~
## 3 2016-10-20 0000010278554503158 1477029466         1         3 /google+redes~
## 4 2016-10-20 0000010278554503158 1477029466         1         4 /google+redes~
## 5 2016-10-20 0000010278554503158 1477029466         1         6 /google+redes~
## 6 2016-10-20 0000010278554503158 1477029466         1         9 /google+redes~
## # i 5 more variables: channelGrouping <chr>, total_newVisits <int>,
## #   trafficSource_source <chr>, device_deviceCategory <chr>,
## #   geoNetwork_country <chr>

```

## La tendance mensuel de parcours de navigation

On analyse d'abord, les 10 parcours de navigation les plus fréquent de chaque mois:

```
# Fonction pour tronquer les chemins longs en gardant la fin visible
truncate_path <- function(path, max_length = 50) {
  if (str_length(path) > max_length) {
    return(paste0("...", str_sub(path, -max_length, -1)))
  } else {
    return(path)
  }
}

df1$date <- as.Date(df1$date)

# Top 10 parcours le plus fréquent par mois
top_paths_by_month <- df1 %>%
  mutate(month = format(date, "%Y-%m")) %>%
  arrange(fullVisitorId, visitId, hitNumber) %>%
  group_by(fullVisitorId, visitId, month) %>%
  summarise(path = paste(pagePath, collapse = " -> "), .groups = 'drop') %>%
  count(month, path, name = "frequency") %>%
  group_by(month) %>%
  slice_max(frequency, n = 10) %>%
  arrange(month, desc(frequency)) %>%
  mutate(path_display = sapply(path, truncate_path))

# Affichage mois par mois avec chemin tronqué pour la visibilité (le parcours complète est toujours sur
for(m in unique(top_paths_by_month$month)) {
  cat("\n== ", m, " ==\n")
  print(top_paths_by_month %>% filter(month == m) %>% select(month, path_display, frequency))
}
```

```
##
## == 2016-08 ==
## # A tibble: 10 x 3
## # Groups:   month [1]
##   month   path_display      frequency
##   <chr>   <chr>             <int>
## 1 2016-08 /home                30484
## 2 2016-08 /home -> /home          5416
## 3 2016-08 /google+redesign/apparel/men++s/men++s+t+shirts 1285
## 4 2016-08 /home -> /signin.html    639
## 5 2016-08 /home -> /home -> /home   604
## 6 2016-08 /google+redesign/drinkware 560
## 7 2016-08 ...-> /google+redesign/apparel/men++s/men++s+t+shirts 442
## 8 2016-08 /home -> /google+redesign/drinkware 404
## 9 2016-08 /google+redesign/bags     354
## 10 2016-08 /home -> /home -> /signin.html 265
##
## == 2016-09 ==
## # A tibble: 10 x 3
## # Groups:   month [1]
```

```

##   month   path_display           frequency
##   <chr>   <chr>             <int>
## 1 2016-09 /home                28690
## 2 2016-09 /home -> /home        5263
## 3 2016-09 /google+redesign/apparel/men++s/men++s+t+shirts 1106
## 4 2016-09 /home -> /signin.html 1101
## 5 2016-09 /home -> /home -> /home 620
## 6 2016-09 /home -> /home -> /signin.html 517
## 7 2016-09 /google+redesign/brands/youtube/home 380
## 8 2016-09 ...-> /google+redesign/apparel/men++s/men++s+t+shirts 358
## 9 2016-09 /home -> /google+redesign/shop+by+brand/youtube 350
## 10 2016-09 /store.html          295
##
## == 2016-10 ==
## # A tibble: 10 x 3
## # Groups:   month [1]
##   month   path_display           frequency
##   <chr>   <chr>             <int>
## 1 2016-10 /home                45016
## 2 2016-10 /home -> /home        9099
## 3 2016-10 /home -> /signin.html 1948
## 4 2016-10 /google+redesign/apparel/men++s/men++s+t+shirts 1339
## 5 2016-10 /home -> /home -> /home 1032
## 6 2016-10 /home -> /home -> /signin.html 983
## 7 2016-10 /google+redesign/shop+by+brand/youtube 901
## 8 2016-10 /home -> /google+redesign/shop+by+brand/youtube 461
## 9 2016-10 /home -> /signin.html -> /signin.html 440
## 10 2016-10 /google+redesign/brands/youtube/home 429
##
## == 2016-11 ==
## # A tibble: 10 x 3
## # Groups:   month [1]
##   month   path_display           frequency
##   <chr>   <chr>             <int>
## 1 2016-11 /home                48490
## 2 2016-11 /home -> /home        15342
## 3 2016-11 /home -> /google+redesign/shop+by+brand/youtube 5766
## 4 2016-11 /home -> /home -> /signin.html 2340
## 5 2016-11 /google+redesign/shop+by+brand/youtube 2103
## 6 2016-11 /google+redesign/apparel/men++s/men++s+t+shirts 1466
## 7 2016-11 /home -> /signin.html 1099
## 8 2016-11 /home -> /home -> /home 909
## 9 2016-11 /home -> /home -> /signin.html -> /signin.html 403
## 10 2016-11 /google+redesign/drinkware 380
##
## == 2016-12 ==
## # A tibble: 10 x 3
## # Groups:   month [1]
##   month   path_display           frequency
##   <chr>   <chr>             <int>
## 1 2016-12 /home                23019
## 2 2016-12 /home -> /home        5936
## 3 2016-12 /google+redesign/shop+by+brand/youtube 3436
## 4 2016-12 /home -> /google+redesign/shop+by+brand/youtube 1556

```

```
## 5 2016-12 /signin.html 1252
## 6 2016-12 /google+redesign/apparel/men++s/men++s+t+shirts 1192
## 7 2016-12 /home -> /home -> /home 867
## 8 2016-12 /basket.html 522
## 9 2016-12 /google+redesign/apparel/men++s/men++s+outerwear 508
## 10 2016-12 /home -> /home -> /signin.html 500
##
## == 2017-01 ==
## # A tibble: 10 x 3
## # Groups:   month [1]
##   month path_display frequency
##   <chr> <chr> <int>
## 1 2017-01 /home 19417
## 2 2017-01 /google+redesign/shop+by+brand/youtube 3742
## 3 2017-01 /home -> /home 3430
## 4 2017-01 /signin.html 1112
## 5 2017-01 /google+redesign/apparel/men++s/men++s+t+shirts 972
## 6 2017-01 /home -> /home -> /home 799
## 7 2017-01 /google+redesign/apparel/men++s/men++s+outerwear 568
## 8 2017-01 /basket.html 524
## 9 2017-01 ...=vieworder -> /myaccount.html?mode=vieworderdetail 394
## 10 2017-01 ...=vieworder -> /myaccount.html?mode=vieworderdetail 361
##
## == 2017-02 ==
## # A tibble: 10 x 3
## # Groups:   month [1]
##   month path_display frequency
##   <chr> <chr> <int>
## 1 2017-02 /home 19027
## 2 2017-02 /google+redesign/shop+by+brand/youtube 4198
## 3 2017-02 /home -> /home 3296
## 4 2017-02 /google+redesign/apparel/men++s/men++s+t+shirts 1058
## 5 2017-02 ...redesign/apparel/men++s+outerwear/blm+sweatshirt.axd 838
## 6 2017-02 /home -> /home -> /home 633
## 7 2017-02 /signin.html 533
## 8 2017-02 /google+redesign/drinkware 434
## 9 2017-02 /google+redesign/apparel/men++s/men++s+outerwear 404
## 10 2017-02 /home -> /google+redesign/shop+by+brand/youtube 365
```

À côté des parcours avec répétition de page /home et le parcours de type /home/signin.html qui sont les plus souvent utilisés pour l'action de connecter au compte d'utilisateur, on observe que: - **Avant octobre 2016 (août-septembre 2016)**: Les utilisateurs visitent majoritairement des pages produits spécifiques, notamment la catégorie «apparel men's t-shirts» (/google+redesign/apparel/men++s/men++s+t+shirts) et des sections comme «drinkware» ou «bags» - **Entre octobre et décembre 2016**: On constate une montée en fréquence des visites vers des pages de marques (/google+redesign/shop+by+brand/youtube), alors que «apparel men's t-shirts» est toujours le produit spécifiquement intéressés. Cette augmentation est bien cohérente avec le source de trafic de youtube.com qu'on trouve sur la première partie d'analyse. - **Après décembre 2016 (janvier-février 2017)**: La tendance de naviguer les produits par la marque youtube se confirme avec un maintien élevé des visites terminant sur cette page. En plus, des parcours vers des pages panier (/basket.html) et des pages de vérifier la commande (/myaccount.html?mode=vieworderdetail) apparaissent plus fréquemment, traduisant une évolution vers des actions d'achat. Les séquences sur les pages produit deviennent plus variées, avec des parcours incluant différentes catégories (ex. «apparel men's outerwear»). Cela indique un comportement utilisateur plus engagé, avec un cheminement vers la conversion.

## La tendance mensuel de point de sortie

```
# Requête SQL
query2 <- "
SELECT
  PARSE_DATE('%Y%m%d', _TABLE_SUFFIX) AS date,
  fullVisitorId,
  visitId,
  visitNumber,
  hits.hitNumber AS hitNumber,
  hits.page.pagePath AS pagePath,
  --dimension
  channelGrouping,
  totals.newVisits AS total_newVisits,
  trafficSource.source AS trafficSource_source,
  device.deviceCategory AS device_deviceCategory,
  geoNetwork.country AS geoNetwork_country
FROM
  `bigquery-public-data.google_analytics_sample.ga_sessions_*`,
  UNNEST(hits) AS hits
WHERE
  _TABLE_SUFFIX BETWEEN '20160801' AND '20170228'
  AND
  hits.type = 'PAGE'
ORDER BY
  fullVisitorId,
  visitId,
  visitNumber,
  hitNumber
"

# Charger l'output de query sur un df
df2 <- bq_project_query(project_id, query2) %>%
  bq_table_download()
```

```
df2$date <- as.Date(df2$date)

# Extraire le dernier hit (point de sortie) par visite
exit_points <- df2 %>%
  group_by(fullVisitorId, visitId) %>%
  filter(hitNumber == max(hitNumber)) %>%
  ungroup()

# Ajouter une colonne mois au format "YYYY-MM"
exit_points <- exit_points %>%
  mutate(month = format(date, "%Y-%m"))

# Compter la fréquence des pages de sortie par mois
top_exit_points_by_month <- exit_points %>%
  group_by(month, pagePath) %>%
  summarise(frequency = n(), .groups = "drop") %>%
  group_by(month) %>%
  slice_max(order_by = frequency, n = 10) %>%
```

```

arrange(month, desc(frequency))

# Affichage mois par mois
for(m in unique(top_exit_points_by_month$month)) {
  print(top_exit_points_by_month %>% filter(month == m))
}

## # A tibble: 10 x 3
## # Groups:   month [1]
##   month   pagePath      frequency
##   <chr>   <chr>         <int>
## 1 2016-08 /home          39777
## 2 2016-08 /google+redesign/apparel/men++s/men++s+t+shirts 2756
## 3 2016-08 /google+redesign/drinkware      2085
## 4 2016-08 /basket.html      1925
## 5 2016-08 /google+redesign/office        1808
## 6 2016-08 /signin.html       1703
## 7 2016-08 /store.html        1663
## 8 2016-08 /google+redesign/electronics    1637
## 9 2016-08 /asearch.html       1465
## 10 2016-08 /google+redesign/bags          1464
## # A tibble: 10 x 3
## # Groups:   month [1]
##   month   pagePath      frequency
##   <chr>   <chr>         <int>
## 1 2016-09 /home          38040
## 2 2016-09 /signin.html     2539
## 3 2016-09 /google+redesign/apparel/men++s/men++s+t+shirts 2360
## 4 2016-09 /store.html      2217
## 5 2016-09 /basket.html     1609
## 6 2016-09 /google+redesign/shop+by+brand/youtube    1218
## 7 2016-09 /asearch.html    1170
## 8 2016-09 /google+redesign/office    1148
## 9 2016-09 /google+redesign/drinkware     985
## 10 2016-09 /google+redesign/bags       972
## # A tibble: 11 x 3
## # Groups:   month [1]
##   month   pagePath      frequency
##   <chr>   <chr>         <int>
## 1 2016-10 /home          59398
## 2 2016-10 /signin.html     4204
## 3 2016-10 /google+redesign/apparel/men++s/men++s+t+shirts 2808
## 4 2016-10 /google+redesign/shop+by+brand/youtube    2128
## 5 2016-10 /basket.html     1773
## 6 2016-10 /asearch.html    1482
## 7 2016-10 /google+redesign/apparel      1255
## 8 2016-10 /google+redesign/shop+by+brand/google    1242
## 9 2016-10 /store.html       971
## 10 2016-10 /google+redesign/apparel/men++s/men++s+outerwear 876
## 11 2016-10 /google+redesign/drinkware      876
## # A tibble: 10 x 3
## # Groups:   month [1]
##   month   pagePath      frequency

```

```

##      <chr>      <chr>                                     <int>
## 1 2016-11 /home                                           68594
## 2 2016-11 /google+redesign/shop+by+brand/youtube        8954
## 3 2016-11 /signin.html                                   4979
## 4 2016-11 /google+redesign/apparel/men++s/men++s+t+shirts 2831
## 5 2016-11 /basket.html                                  1985
## 6 2016-11 /asearch.html                                 1641
## 7 2016-11 /google+redesign/apparel                      1240
## 8 2016-11 /google+redesign/shop+by+brand/google         1162
## 9 2016-11 /store.html                                   980
## 10 2016-11 /google+redesign/apparel/men++s/men++s+outerwear 912
## # A tibble: 10 x 3
## # Groups:   month [1]
##   month   pagePath      frequency
##   <chr>   <chr>         <int>
## 1 2016-12 /home           33666
## 2 2016-12 /google+redesign/shop+by+brand/youtube        6200
## 3 2016-12 /signin.html    2828
## 4 2016-12 /basket.html    2380
## 5 2016-12 /google+redesign/apparel/men++s/men++s+t+shirts 2347
## 6 2016-12 /asearch.html   2098
## 7 2016-12 /myaccount.html?mode=vieworderdetail          1863
## 8 2016-12 /google+redesign/apparel/men++s/men++s+outerwear 1393
## 9 2016-12 /google+redesign/electronics                   1160
## 10 2016-12 /ordercompleted.html                          1159
## # A tibble: 10 x 3
## # Groups:   month [1]
##   month   pagePath      frequency
##   <chr>   <chr>         <int>
## 1 2017-01 /home           27165
## 2 2017-01 /google+redesign/shop+by+brand/youtube        5086
## 3 2017-01 /basket.html    2094
## 4 2017-01 /signin.html    1868
## 5 2017-01 /asearch.html   1751
## 6 2017-01 /google+redesign/apparel/men++s/men++s+t+shirts 1730
## 7 2017-01 /google+redesign/apparel/men++s/men++s+outerwear 1693
## 8 2017-01 /myaccount.html?mode=vieworderdetail          1473
## 9 2017-01 /google+redesign/electronics                    826
## 10 2017-01 /google+redesign/shop+by+brand/google          764
## # A tibble: 10 x 3
## # Groups:   month [1]
##   month   pagePath      frequency
##   <chr>   <chr>         <int>
## 1 2017-02 /home           26431
## 2 2017-02 /google+redesign/shop+by+brand/youtube        5924
## 3 2017-02 /google+redesign/apparel/men++s/men++s+t+shirts 1942
## 4 2017-02 /basket.html    1707
## 5 2017-02 /asearch.html   1558
## 6 2017-02 /google+redesign/shop+by+brand/google         1297
## 7 2017-02 /signin.html    1202
## 8 2017-02 /google+redesign/apparel/men+s+outerwear/blm+sweatshirt.axd 1029
## 9 2017-02 /google+redesign/apparel/men++s/men++s+outerwear 974
## 10 2017-02 /google+redesign/electronics                   698

```