2020

GOOGLE ANALYTICS QUERIES USING BIGQUERY

BY **AMR ALFAYOUMY**

NP CYCLE #3

THE DATASET

Google Analytics Sample

- Type: Datasets
- Last updated: 9/20/19
- Category: Advertising, Analytics
- Dataset source: Google Merchandise Store
- Cloud service: BigQuery
- Update frequency: Historical

The dataset provides 12 months (August 2016 to August 2017) of obfuscated Google Analytics 360 data from the Google Merchandise Store , a real ecommerce store that sells Google-branded merchandise, in BigQuery. It's a great way analyze business data and learn the benefits of using BigQuery to analyze Analytics 360 data.

QUERY#1:
ALL THE TRANSACTIONS BY VISITORS IN JULY 2017

```sql
SELECT
  h.TRANSACTION.transactionid,
  fullvisitorid,
  TIMESTAMP_SECONDS(visitstarttime) as visit_time,
FROM
  `bigquery-public-data.google_analytics_sample.ga_sessions_*`,
  UNNEST(hits) AS h
WHERE
  _TABLE_SUFFIX BETWEEN '20170701'
  AND '20170731'
  AND h.TRANSACTION.transactionid IS NOT NULL
GROUP BY
  h.TRANSACTION.transactionid,
  fullvisitorid,
  visit_time,
  date
ORDER BY visit_time ASC
```

Query complete (1.3 sec elapsed, 2.7 MB processed)

Job information | Results | JSON | Execution details

Row	transactionid	fullvisitorid	visit_time
1	ORD201707011581	2024172620398136134	2017-07-01 19:32:21 UTC
2	ORD201707011643	3641847080616677863	2017-07-01 21:46:26 UTC
3	ORD201707011628	4855673784399652593	2017-07-02 03:12:38 UTC
4	ORD201707021563	8197879643797712877	2017-07-02 11:54:36 UTC
5	ORD201707021560	3117858606721790013	2017-07-02 13:32:02 UTC
6	ORD201707021536	5569275433551045400	2017-07-02 15:04:51 UTC
7	ORD201707021551	9460311742865907976	2017-07-02 16:26:56 UTC
8	ORD201707021449	2760610370970936430	2017-07-02 19:00:42 UTC
9	ORD201707021442	3694234028523165868	2017-07-02 20:07:52 UTC
10	ORD201707021552	9344315745283537381	2017-07-03 01:22:29 UTC

Rows per page: 10 1 - 10 of 1031 First page

QUERY#2:
FREQUENCY OF EACH CONTENT GROUP PER COUNTRY

```sql
CREATE TEMP FUNCTION
  GetNamesAndCounts(elements ARRAY<STRING>) AS ( ARRAY(
    SELECT
      AS STRUCT elem AS name,
      COUNT(*) AS freq
    FROM
      UNNEST(elements) AS elem
    GROUP BY
      elem
    ORDER BY
      freq DESC ) );
SELECT
  geoNetwork.country AS country,
  GetNamesAndCounts(ARRAY_AGG(h.contentGroup.contentGroup2)) AS Category
FROM
  `bigquery-public-data.google_analytics_sample.ga_sessions_*`,
  UNNEST(hits) AS h
WHERE
  geoNetwork.country != '(not set)'
  AND h.contentGroup.contentGroup2 != '(not set)'
GROUP BY
  country
ORDER BY
  country ASC
```

Query complete (8.2 sec elapsed, 50 MB processed)

Job information | Results | JSON | Execution details

Row	country	Category.name	Category.freq
1	Afghanistan	Brands	62
		Apparel	10
		Accessories	5
		Electronics	2
2	Albania	Brands	138
		Apparel	90
		Electronics	28
		Bags	16
		Drinkware	15
		Accessories	13
		Office	9

Rows per page: 10 | 1 - 10 of 214 | First page

58	Egypt	Brands	928
		Apparel	666
		Bags	316
		Electronics	257
		Office	194
		Drinkware	173
		Accessories	96
		Lifestyle	16
		Nest	2

QUERY#3:
HOW MANY CLICKS ARE THERE PER LANDING PAGE?

```sql
SELECT
  hits.page.pagePath AS LandingPage,
  COUNT(totals.hits) AS Clicks
FROM
  `bigquery-public-data.google_analytics_sample.ga_sessions_*`,
  UNNEST(hits) AS hits
WHERE
  hits.type="PAGE"
  AND hits.hitNumber IS NOT NULL
GROUP BY
  LandingPage
ORDER BY
  Clicks DESC
```

Query complete (9.4 sec elapsed, 178.4 MB processed)

Job information | Results | JSON | Execution details

Row	LandingPage	Clicks
1	/home	981285
2	/basket.html	209360
3	/google+redesign/shop+by+brand/youtube	145026
4	/signin.html	101299
5	/store.html	93551
6	/google+redesign/apparel/men++s/men++s+t+shirts	67471
7	/asearch.html	62380
8	/google+redesign/electronics	56839
9	/google+redesign/apparel	56552
10	/google+redesign/bags	53686

Rows per page: 10 1 - 10 of 2571 First page

QUERY#4:
HOW MANY SESSIONS ARE THERE PER MONTH?

```sql
SELECT EXTRACT(MONTH FROM PARSE_DATE ("%Y%m%d", date)) as Month,
  SUM(totals.visits) AS sessions
FROM
  `bigquery-public-data.google_analytics_sample.ga_sessions_*`
  WHERE
    _TABLE_SUFFIX BETWEEN '20160801'
    AND '20170731'
GROUP BY Month
ORDER BY sessions DESC
```

Query complete (8.7 sec elapsed, 15.5 MB processed)

Job information | Results | JSON | Execution details

Row	Month	sessions
1	11	113972
2	10	97506
3	12	79124
4	8	74759
5	7	71812
6	9	71032
7	3	69931
8	4	67126
9	5	65371
10	1	64694
11	6	63578

QUERY#5:
HOW MANY VISITS ARE THERE PER USER?

```sql
SELECT
  fullVisitorId,
  COUNT(visitNumber) as total_visits_per_user
FROM
  `bigquery-public-data.google_analytics_sample.ga_sessions_*`
WHERE
  visitNumber IS NOT NULL
GROUP BY
  fullVisitorId
ORDER BY
  total_visits_per_user DESC
```

Query complete (12.6 sec elapsed, 24.9 MB processed)

Job information | Results | JSON | Execution details

Row	fullVisitorId	total_visits_per_user
1	1957458976293878100	278
2	0824839726118485274	255
3	3608475193341679870	201
4	1856749147915772585	199
5	3269834865385146569	155
6	0720311197761340948	153
7	7634897085866546110	148
8	4038076683036146727	138
9	0232377434237234751	135
10	3694234028523165868	129

Rows per page: 10 1 - 10 of 714167 First page

QUERY#6:
PERCENTAGE OF VISITS PER OPERATING SYSTEM

```sql
SELECT
  device.operatingSystem AS OS,
  COUNT(*) AS visits_per_OS,
  (COUNT(*) / SUM(COUNT(*)) OVER ())*100 AS Percentage
FROM
  `bigquery-public-data.google_analytics_sample.ga_sessions_*`
GROUP BY
  OS
ORDER BY
  Percentage DESC
```

Query complete (7.4 sec elapsed, 7.8 MB processed)

Job information | Results | JSON | Execution details

Row	OS	visits_per_OS	Percentage
1	Windows	350072	38.739648958173106
2	Macintosh	253938	28.101273387019134
3	Android	123892	13.710129883926683
4	iOS	107665	11.914418477003894
5	Linux	35034	3.8769306359852735
6	Chrome OS	26337	2.9145036867027496
7	(not set)	4695	0.5195578391263018
8	Windows Phone	1216	0.13456492702397935
9	Samsung	280	0.030985345038416292
10	BlackBerry	218	0.024124304351338402

Rows per page: 10 1 - 10 of 20 First page

QUERY#7:
PERCENTAGE OF VISITS PER COUNTRY


```sql
SELECT
  geoNetwork.country AS country,
  COUNT(*) AS visits_per_country,
  (COUNT(*) / SUM(COUNT(*)) OVER ())*100 AS Percentage
FROM
  `bigquery-public-data.google_analytics_sample.ga_sessions_*`
GROUP BY
  country
ORDER BY
  Percentage DESC
```

Query complete (9.1 sec elapsed, 10.1 MB processed)

Job information | Results | JSON | Execution details

Row	country	visits_per_country	Percentage
1	United States	364744	40.363281038186116
2	India	51140	5.659251947373605
3	United Kingdom	37393	4.137982167933931
4	Canada	25869	2.8627138957099683
5	Vietnam	24598	2.7220625616248713
6	Turkey	20522	2.2710044674227827
7	Thailand	20123	2.2268503507430393
8	Germany	19980	2.2110256923841343
9	Brazil	19783	2.1892252889106767
10	Japan	19731	2.1834708676892567

Rows per page: 10 1 - 10 of 222 First page

THANK YOU!

ANY QUESTIONS?

AMR ALFAYOUMY
a.alfayoumy@outlook.com
linkedin.com/in/alfayoumy/
github.com/alfayoumy/