

Opera Software - BI Intern Competence Test

Data

A sample of daily aggregated data of one of Opera's products.

Schema (field name - data type)

- **partition_date** - DATE - table partition creation date (new data is appended to the database daily and stored in a corresponding partition)
- **timeperiod** - DATE - user cohort date (note: `timeperiod` >= `partition_date`)
- **country** - STRING - 2-letter country code
- **full_version** - STRING - full product version
- **news_locale** - STRING - newsfeed locale {country code}:{language code}
- **channel** - STRING - user acquisition channel
- **source** - STRING - user acquisition referrer source
- **impressions** - INTEGER - count of impressions a user cohort generated
- **clicks** - INTEGER - count of clicks a user cohort generated
- **active_users** - INTEGER - count of unique active user IDs on cohort date
- **new_users** - INTEGER - count of unique new user IDs on cohort date
- **retained_01** - INTEGER - count of unique new user IDs on cohort date that appeared active on cohort date + 1 day
- **retained_07** - INTEGER - count of unique new user IDs on cohort date that appeared active on cohort date + 7 days
- **retained_30** - INTEGER - count of unique new user IDs on cohort date that appeared active on cohort date + 30 days

Tools

Our team uses Google BigQuery (and a few other SQL databases), Jupyter Notebooks, Python, Tableau, Google Sheets, Excel. We encourage you to set up a project in BigQuery and upload the test data there. BigQuery is easy to start with and free for data below 10GB ([More info](#)). Otherwise, you can use any other software you're most familiar with.

Assignment

1. Load the provided data into a database.
2. Get familiar with the data structure (hint: notice the relationship between `partition_date` and `timeperiod`). Find out what time range the data covers, what are the possible values for the fields.
3. Calculate absolute average daily new user run-rate per country code (in BigQuery use Standard SQL instead of Legacy SQL).
4. Display the average daily active user in July 2018 for top 10 sources by total new user count in June 2018.
5. Display top 5 countries by the average daily R30 retention ratio ($\text{retained_30} / \text{new_users}$) in June 2018.
6. Open question. Analyze user acquisition trends. Which distribution sources bring good volume of the best quality users? Consider users' retention rates and clicks/impressions in relation to new user growth. Prepare a report with relevant tables and plots. Can be in a format of an interactive dashboard with comments or a spreadsheet.

Assessment criteria

- **Correctness of the calculated metrics.** The basic requirement is to get the correct numbers for the KPIs in question.
- **Effectiveness of the report.** Aim to deliver a to-the-point report that is easy to read (remember visual clarity and data-ink ratio).
- **Bonus points** for having a well-documented solution, writing efficient queries and code.