



Insights Analyst Main Assessment

2023

Part I – SQL/Data Handling

Requirements

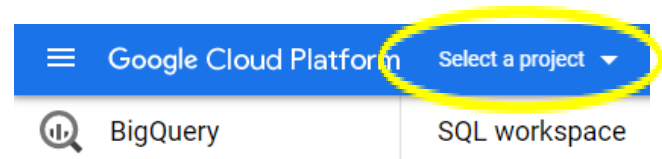
- Please note that you need a gmail (or **google-based**) **account** to complete the assessment, so you have to create if not use an existing one.
- **Download** the [orders.csv](#) file. This is a dummy dataset with basic information regarding efood orders of January 2022.

Set up BigQuery (~15min)

We are mainly using the Standard SQL of BigQuery. Once you have logged in to your google-based account, you can access the BigQuery platform following the [Link to BQ](#).

1. Create a **project** named **efood2022**

Click on "Select a project" and Create New.



2. Add a **dataset** named **main_assessment**

Click on the 3 bullets next to efood2022 name and select create dataset.

Window like screenshot appears where you set the dataset name and location in Europe (eu) and press on the blue button.

Create dataset

Project ID
efood2022 [CHANGE](#)

Dataset ID *
main_assessment

Letters, numbers, and underscores allowed

Data location
eu (multiple regions in European Union) ▼ ?

Default table expiration

☐ Enable table expiration ?

Default maximum table age Days

Encryption

☐ Use a customer-managed encryption key (CMEK)

Leave unchecked to use the default configured for your organization or project

3. Add a **table** named **orders**

Click on the 3 bullets next to main_assessment and select Create table.

Window appears like in the screenshot where you choose to **upload** the [orders.csv](#) you downloaded before.

Create table

Source

Create table from
Upload

Select file *
orders.csv

File format
CSV

Destination

Project *
efood2022

Dataset *
main_assessment

Table *
orders
Unicode letters, marks, numbers, connectors, dashes or spaces allowed.

Table type
Native table

Choose to **edit as text** and paste the schema included in the following gray pane, as shown in the screenshot.

Schema

☐ Auto detect

☒ Edit as text

```
1  order_id:INTEGER,  
2  user_id:INTEGER,  
3  order_timestamp:TIMESTAMP,  
4  city:STRING,  
5  cuisine:STRING,  
6  paid_cash:BOOLEAN,  
7  amount:FLOAT
```

```
order_id:INTEGER,  
user_id:INTEGER,  
order_timestamp:TIMESTAMP,  
city:STRING,  
cuisine:STRING,  
paid_cash:BOOLEAN,  
amount:FLOAT
```

When clicking on Edit as text again you see the table schema as in screenshot

Schema

☐ Auto detect

☒ Edit as text

Field name	Type	Mode	Description	
order_id	INTEGER ▼	NULLAB... ▼		
user_id	INTEGER ▼	NULLAB... ▼		
order_timestamp	TIMEST... ▼	NULLAB... ▼		
city	STRING ▼	NULLAB... ▼		Max length
cuisine	STRING ▼	NULLAB... ▼		Max length
paid_cash	BOOLEAN ▼	NULLAB... ▼		
amount	FLOAT ▼	NULLAB... ▼		
+				

Finally, in Advanced options we set Header Rows to Skip to 1, as CSV has a header that doesn't follow table's schema and that couldn't allow its creation.

Advanced options

Write preference

Write if empty ▼

Number of errors allowed

0

☐ Unknown values ?

Field delimiter

Comma ▼ ?

Header rows to skip

1

☐ Quoted newlines ?

☐ Jagged rows ?

Encryption

☐ Use a customer-managed encryption key (CMEK)

Leave unchecked to use the default configured for your organization or project

- Once setup finished you are able to query on the table created (ie. using the following query or pressing query button from the UI)

```
SELECT *  
  
FROM `efood2022.main_assessment.orders`
```

Query #1 (-35min)

We want to focus on Breakfast cuisine in smaller cities of Greece.

For every city that exceeds the 1,000 orders we would like to compare "**Breakfast**" cuisine **versus** the **total efood**, creating the metrics:

- **Basket** := Amount / Order
- **Frequency** := Orders / Users
- **%Users that exceed 3 orders** := (Users of Frequency >3) / Users

Order **all** your findings, showing us the **5 cities** with the most **Breakfast Orders** and explain them in a few words.

Your end result could have namings of your preference, additional columns and formats but should look like the table below:

Row	city	breakfast_basket	efood_basket	breakfast_freq	efood_freq	breakfast_users3freq_perc	efood_users3freq_perc
1	Βόλος	5.12778625116582	8.084162184472904	3.732746262543518	4.430481741813818	0.3064714314970305	0.38847934167666726
2	Λάρισα	5.40447655719148	8.906952073743598	3.0960729312762973	3.845662803332126	0.2391304347826087	0.3396414342629482
3	Ξάνθη	4.887138754672236	7.232082548936313	4.271417445482866	5.35396120082367	0.34559968847352024	0.4311260431342798
4	Ιωάννινα	5.755508066011544	9.117966368504588	3.1150902527075814	4.031314868631754	0.2571841155234657	0.3643318901607259
5	Ρόδος	6.248684741851732	10.312610486891403	3.715173596228033	4.547736441057822	0.30561508786969566	0.391304347826087

Query #2

What percentage of Orders do the **top 10 users of each city** contribute to their city?

Submitting a solution

Please, create a public **Github** repo and store your SQL queries.

Part II - Analyze Users

- We need to segment existing customers based on their *frequency* and *order* value.
- Ordering Breakfast via efood is a quite new habit that Marketing thinks could create more loyal customers. Which segment could be a valuable target group for a Marketing campaign about “Breakfast” cuisine_parent?

Use the provided csv or the table you’ve created, create user segments, answer the business question, and prepare a presentation to share your findings/insights with the BI team. Your findings could be presented in a Visualization tool of your choice, or through PowerPoint or Excel etc.

Submitting a solution

Please, create a new file in the public **Github** repo you created for **Part I**, that will contain:

- *all the code* and
- *the instructions to run it*

Within efood (and Delivery Hero in general) we use **Python** (3.x), **R** and **SQL** so we suggest preferring them before using another programming stack; still, you can use **whatever** suits you but due to our knowledge limitations we will only be able to verify if it works 100% or it fails.

Reply via email sharing your repo link and presentation/visuals/comments.

Contact

You can send an email to d.kouvari@e-food.gr, gmouratos@e-food.gr with any questions you have. It goes without saying that this assessment is handed over as a real case so you are expected to work as you would in your normal work life (eg search the Nets, reference books etc); as always keep in mind to **cite** anything that is considered a “loan”.