Hi there!

Please open, make a copy, review, and complete it ASAP.  
  
Once done, pls send to [Vladyslav Dodonov](mailto:vladyslav.dodonov@zubale.com)  
[vladyslav.dodonov@zubale.com](mailto:vladyslav.dodonov@zubale.com) and CC [Maria Bravo](mailto:maria.bravo@zubale.com)  
  
All the best!

#### **Challenge 1**:

We have [two CSV files](https://drive.google.com/drive/folders/1P7v03V__0O-44UUz-M_EBTpS5m7b6Pe1): one for products and the other for orders, with the following structure.  
products.csv:

* id
* name
* category
* price

orders.csv:

* id
* product\_id
* quantity
* created\_date

Please note that every order can only contain one product type.

We need to combine both files to consolidate the information in a single CSV file in the following columns:

* order\_created\_date
* order\_id
* product\_name
* quantity
* total\_price

This data has to be stored in a newly created file named “order\_full\_information.csv”.

#### **Challenge 2**:

Due to an error in one of our systems, the price information from orders.csv came in Brazilian currency (BRL), and we need to convert that value to US dollars.

Get the latest currency data from: <https://app.freecurrencyapi.com/>  
You will need to get an API key from the website.

Include a step in your code to get this information, and use the data needed to get the total price in the desired currency. Persist the results in a new file “fixed\_order\_full\_information.csv”, with the following columns:

* order\_created\_date
* order\_id
* product\_name
* quantity
* total\_price\_br
* total\_price\_us

Now we want to explore a little with our data. Use Python to find the following information:

1. Date where we create the max amount of orders.
2. Most demanded product and the total sell price.
3. The top 3 most demanded categories.

Store the results in a single CSV file named: kpi\_product\_orders.csv

**Challenge 3**:

Instead of using the CSV files, create a PostgreSQL database and create the tables using these files as data sources.

Programmatically do this step (using a script).  
Use SQL to get the information for each of the previous points:

1. The date with max amount of orders
2. The most demanded product
3. The top 3 most demanded categories

(Note: query efficiency will be taken into account, regardless of the amount of data these tables have, imagine that the productive environment has millions of records)

**Challenge 4**:

* What other insights do you think would add value to the business that can be extracted using at least one of these tables?

Pick up to three, and explain why they might be useful and how we can get them. (If more tables are needed, list them.)

* What ETL/ELT tool would you use to extract this data and insert it into BigQuery? Explain the steps of creating this type of pipeline.
* What AI-based pipeline could you add to this pipeline? Describe it.

**Important Information**

What we expect to get from you:

* Python code that writes all the output files
* SQL script used in Challenge 3
* Queries for analyzing the data in the DB
* A technical document that explains what the code is doing and any considerations that are relevant to you.
* You can share it directly as attachments, github repo, or any other way you consider.