Grover BI Engineer Assignment

Thank you for participating in this home assignment. There are three exercises included, which should be read carefully:

- Data Modelling + ETL pipeline
- SQL code
- Python

1) Data Modelling + ETL pipeline

There are 4 csv files that contain the following information about orders from across 3 countries with the following schema definitions

Order : Contains the details of orders placed at an e-commerce company.

Column Name	Column Definition
order_id	The unique id of the order
creation_date	The date of order creation
order_value	The amount of the order placed
country_id	The country from where the order was placed
status_code	The status in which the order currently is
category_id	Category of the device ordered

Country: The country in which the order was placed

Column Name	Column Definition
id	The id of the country
Country	Name of the country

Order_status: The status codes of an order

Column Name	Column Definition
status_code	The status code of the order
status	APPROVED | When the order is approved successfully by the Company DECLINED | When the order is rejected by the company CANCELLED | When the order is cancelled from the customers end

category: The status codes of an order

Column Name	Column Definition
id	The id of the category
category	Category name

Requirements

- Use any combination of ETL tool + Database you feel comfortable with – preferably Python + PostgreSQL. The only condition is that you need to make the code available for review (SQL + ETL);

- Create and load the tables (please consider the attached files). Provide the DDL statements as well – preferably using PostgreSQL syntax.

- Create fact and dimension tables that can be used for End of month reports:
  - Number of orders and total order value;
  - be able to aggregate based on order status, country or category.

Important: please populate the model using procedures or functions.

- Deliver your solution as a git repository structure inside a zip file, including folders separating the different steps – Sample files, ETL and SQL.

2) SQL code

Write some SQL code that generates one output with the following fields:

- Order_ID
- Status
- Country
- List of Categories (list separated by "|") in case an order is assigned to more than 1 category
- Number of days between order and today (create custom function to calculate this

value)
- Order amount
- Country_order_value_index – this represents the position in terms of order value for every country. For example, if this order has the highest order value for its **Country**, then it should show the value **1**, if it's the second highest then should show the value **2**. As there are three countries, there should be three order ids with the value **1**.

Please provide the query as part of the package above.

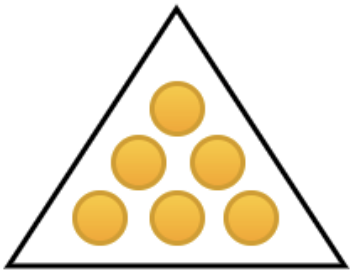
3) Python

The purpose of this exercise is to create a Python function that calculates how many rows will be filled with a specific number of balls if you have a triangle shaped box, based on an input variable.

For example:

- 1 ball will fill one row. Function_X(1) outputs 1
- 2 means will still fill only one row. Function_X(2) outputs 1
- 3 balls will fill two rows. Function_X(3) outputs 2

You can find a visual representation below:



Please provide the Python code in a .py file. You can include it as part of the package created above or separated.