Assignment for BI Developer Intern

We are delighted to see you applying. Now is your time to shine. We've prepared a task for you so that you can show your skills and allow us to understand how you approach problems in general.

In this exercise you get to work with purchase data, build a derived table, and calculate KPIs based on the datasets provided.

Good luck!

Introduction and Data

Managing the cost of sales and gross margin is important for Wolt's first-party retail business. The goal of this task is to provide a dataset that can be used for both calculation of KPIs and to guide decision making by giving users of the data visibility into profitability of certain purchases, venues, and items.

The data consists of three csv files:

- purchase_data.csv
- purchase_item_data.csv
- item_data.csv

The data is an artificial dataset that could be produced by our purchase process, where a user makes a purchase from Wolt Market. In "purchase_data.csv" you will find information about the purchase itself, and you'll find exactly which items were purchased in "purchase_item_data.csv". The "item_data.csv" has item-level information, including the costs for the items that Wolt pays for the vendors.

Fields in datasets

purchase_data.csv

- PURCHASE_ID: The unique identifier of the purchase
- TIME_RECEIVED: The time when the customer made the order
- TIME_DELIVERED: The time when the order was delivered to the customer
- CURRENCY: Local currency the purchase was made in
- COUNTRY: The country of the venue where the purchase was made
- VENUE_ID: The unique identifier of the venue

purchase_item_data.csv

- PRODUCT_ID: Identifier for the item used at Wolt
- PURCHASE_ID: The unique identifier of the purchase
- COUNT: How many items of a certain type were purchased
- VENUE_ID: The unique identifier of the venue
- BASEPRICE: The price of an item in local currency including value-added tax (VAT)
- VAT_PERCENTAGE: Value-added tax (VAT) percentage

item_data.csv

- VENUE_ID: The unique identifier of the venue
- TIMESTAMP: Time when the item became available
- BRAND: Name of the brand of a product
- MANUFACTURER: Name of the manufacturer of a product
- COST_PER_UNIT: Cost of single unit of item in local currency excluding value-added tax (VAT)
- COST_PER_UNIT_EUR: Cost of single unit of item in Euro excluding value-added tax (VAT)
- CURRENCY: Currency for the item price
- APPLICABLE_TAX_PERC: Applicable tax percentage of the item
- PRODUCT_ID: Identifier for the item used at Wolt
- ITEM_IDENTIFIER: Global Trade Item Number (GTIN)
- EXTERNAL_ID: External ID of the item used by the eternal vendor

Tasks

Task 1

Using SQL, create a query that produces a table where we can see the profitability of each purchase. Think about what granularity is best for this, which fields should be included, and how the solution works so that it can also handle any small errors there might be in the dataset and if any further data would be introduced.

Task 2

Calculate and visualize the following KPIs:

- Top 10 venues with the highest margins, and plot the relationship between the margin and average order size
- Top 5 countries for each of the following metrics:
    - average order size
    - average order value
    - order volume
- Plot the monthly cumulative Woltwide (i.e. global) margin

Additionally, we'd like you to think about and also answer the following questions:

- What assumptions about the data did you make when creating your gross margin calculations?
- Did you encounter any problems in the data? And if you did, how did you solve them?
- Which additional data sources or business information do you think could be used to improve your solution?

As we'd like to see some of your SQL skills in action, please solve this task using SQL. You are free to use any additional tools, for example, SQL package in a Jupyter Notebook, or you can go

with PostgreSQL with pgAdmin. Please also share your code, specifying the SQL dialect you've used.

Regarding the visualizations, opt for any tool you wish to use in order to visualize your results in a suitable way. Choose chart types, color palettes and layouts that convey the messages best.

Deliverables to share with us

Please return your solution in English in one PDF file. The PDF should include all your answers and visualizations as well as the SQL queries and any other scripts that you used in order to achieve your final results.

When your solution is ready, share it with us via Google Drive, Dropbox, OneDrive (or similar). Remember to check permissions! If we cannot access the file, we cannot review your solution.

Please don't store your solution in a public GitHub repository during the application period.