Data Science Assignment

Internship Program 2025

We are delighted to start the recruitment process with you. Now it is your time to shine. We've prepared a task for you so that you can prove your technical and analytical skills and to let us understand how you are approaching solving problems. Good luck!

The assignment consists of two tasks:

- **First Task:** You'll solve a SQL challenge, please make sure to share your SQL queries in a text file, Google doc or any other easily accessible method
- **Second Task:** You'll analyze a dataset to showcase your data exploration skills and ability to extract actionable business insights. Please make sure to share both your code/notebook and a short presentation to summarise your findings.

First Task (SQL)

Create SQL queries for the following tasks using the tables below with the following columns

Tables:

1. **Users**: This table contains details about Wolt users
    - **User_id**: Unique ID of the user (primary key)
    - **User_country**: Country of the user
    - **User_device_id**: Unique ID of the user's device
    - **User_registration_timestamp_utc**: Timestamp in UTC when the user registered on the platform
    - **User_first_purchase_timestamp_utc**: Timestamp in UTC when the user made the first purchase on the platform
2. **Sales**: This table contains data about the sales for different venues.
    - **Purchase_id**: Unique ID of the purchase (primary key)
    - **User_id**: Unique ID of the user
    - **Venue_id**: Unique ID of the venue
    - **Timestamp_utc**: Timestamp in UTC when the purchase was made
    - **Total_number_units**: The number of units sold within this order
    - **Value_eur**: The total value of the order in Euro
3. **Purchases**: Contains details about the purchases.
    - **Purchase_id**: Unique ID of the purchase
    - **Product_id**: Unique ID of the purchased product

- **Price**: Price of the product for the customer
- **Quantity**: Number of units sold for a specific product

Tasks:

1. **Query 1:** Calculate **number of users registered** in **Finland** in the **last 30 days.**
2. **Query 2:** Count the **number of users** who have **made at least one purchase** in the **past 30 days**, where their **purchases include more than one unique product.**
3. **Query 3:** Retrieve the **most recent price for each purchased product**. Bonus points if you can provide **two different methods** to achieve this.

Second Task (Data Analysis)

Attached to this message you'll find a **CSV file** with some machine-generated user data of a service like Wolt. Every row of the dataset represents **one customer** who has registered to use the service during September 2019 and either has or hasn't made orders during the time from then and October 2020. Please familiarize yourself with the dataset, show us your excellent exploratory data analysis skills and analyse the provided dataset to generate insights.

Tips:

- **Slice & dice the data** to create **insightful charts and interesting observations.** Ensure each insight is **clearly explained**, highlighting its relevance and **value to a business like Wolt**
- What would **3-4 interesting metrics to define and track** regarding this type of service and why?
- Make sure to add **comments to your code** to make it more readable

At Wolt we use Python or R for notebooks and we expect you to return the rendered notebook alongside the code used for the analysis. Make sure we can run and view your solution on MacOS and Linux.

In addition to the notebook used to solve the second task, please include a separate supporting presentation that highlights the key insights and recommended metrics. Please make sure that the presentation should be self-sufficient for us to follow your thought process and shouldn't include more than 10 slides in total.