**Data Science Methods for Business Improvement**

Thank you for devoting your time to this task. We greatly appreciate it. With this task, we aim to understand:

* Your approach to data analysis
* Your ability to understand the monetization in subscription business
* Your strengths and potential areas for improvement

Most important of all, these tasks will help us create a conversation space where we can exchange our thoughts and expectations on a deeper level for this educational journey.

We assure you that we’ll thoroughly review your solutions, but if you advance to the next round, we won't go over every detail of your presentation. So please make sure to include any important written or visual information in your presentation.

In this case, we have 4 questions for you to see you in the following areas better:

* A/B Testing & Product Monetization
* Data Reliability & Debugging
* Business Intelligence & Data Visualization
* Marketing Analytics

For each of the given questions, you will use a different dataset attached to the email you received along with this task. You can find more detailed information about each dataset on the notion sheet you have delivered.

We will use BigQuery as the data warehouse, and you can upload the data files to BigQuery and use it for your SQL queries.

Please feel free to choose how you deliver your results: a written report, a spreadsheet, a presentation, or anything with which you feel comfortable. Please explain your logic shortly and add your queries to the corresponding answers.

**Try your best to answer questions as detailed as possible, and embrace a ‘challenge accepted’ attitude for all questions.**

**Q1 : Experiment**

As a future data scientist, you have to validate decisions with data, especially when making significant ones. Therefore, you will conduct thousands of A/B tests to date.

For this part of the task, you will be working with a dataset named dataset\_experiment. This dataset contains real data from an important monetization A/B test conducted. Here is the information you will need for your analysis:

● The monetization strategy of this app is to maximize the user's revenue within their d365 (the user's own 365-day lifetime).

● As you will see when you explore the data, the users included in the test were those who first joined between October 2, 2022, and October 22, 2022. The entire dataset covers data from October 2, 2022, to December 30, 2022. We can assume that the analysis is being conducted at the end of the day on December 30, 2022.

Using the above information, we want you to conduct analyses for the following questions:

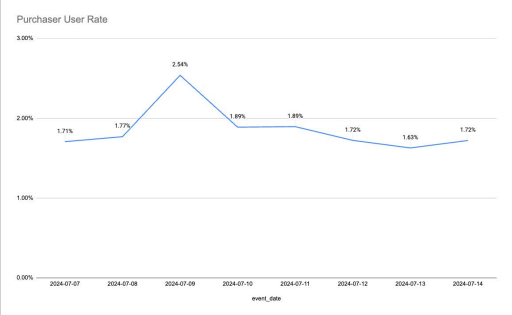
a) Which variant is winning in LTV? Why? How would you interpret the difference between the variants?

b) (Bonus) What would be your additional comments, if we change our strategy to d366 from d365?

**Q2 : Debugging**

As a future data scientist, you will use data in almost every decision-making process, and when you notice a change in the data you can't explain, you should

investigate it because the data might be incorrect.



As seen in the chart above, there is a spike in the purchaser user rate metric on

2024-07-09. The Product and Marketing teams have asked whether this is due to a data-related issue since they confirmed no different actions were taken.

You can use the data\_reliability.csv dataset in your analyses for this part of the task. Here is the information you will need for your analysis:

● Active User Count: In the given table, if a user has triggered any event on the

relevant day, they can be considered an active user.

● Purchaser User Rate = (Purchase Event Count) / (Daily Active User Count)

Using the above information, we want you to investigate the following questions:

a) Could you please investigate this request from the teams? Is there any problem? Can we be confident in the data?

b) If there is a data-related issue, what percentage of users has been affected? What might the problem be? Can you identify any common patterns among the affected users?”

**Q3 : Dashboard**

You should be able to generate dozens, hundreds, or even thousands of metrics from any dataset. While some of these metrics provide important insights, others may be entirely meaningless. Well-defined metrics, combined with clear visualization, can show the whole story of the current state of business in a snap.

We have a new app called X. The Product team needs a dashboard to regularly monitor the metrics of the X app. We have shared some of the data from the X app in the file named dashboard\_design.csv.

Using the above information, can you create a dashboard using this data? You can use Looker Studio or Google Sheets to visualize the dashboard.

**Q4 : Marketing**

We are down to the last question, and this time we're working with marketing. The Marketing team conducts its analyses by combining data from various sources, and in this question, we are expecting a similar analysis from you.

There are two datasets for this part of the case, marketing\_events and marketing\_users.

a. Please find the total revenue of users from Facebook, Apple Search Ads, and Organic platforms. We expect you to present your query results in the format below:



b. Assuming the CPI for Apple is $3.1, and for Facebook is $1.3, please compare the performances of both platforms and share your insights.