Data analyst tech challenge

The Swissborg app allows users to easily buy and trade cryptos. Users can make fiat or crypto deposits and withdrawals, trade between any two currencies directly, and even earn additional yield on some tokens.

Your task will be to explore some key aspects of our user acquisition and conversion from the first steps of onboarding all the way to making deposits, carrying out exchanges or subscribing to a yield wallet. The goal of this exercise is to highlight pain points or areas of opportunity for improvement, and support the goal setting process by providing clear and measurable KRs (key results).

Datasets

user_identity table (download here, ~159MB)

This table contains key events related to the registration/onboarding process of our users. If you have tried out the Swissborg app, you already know that users need to answer some questions and submit some documents to verify their identity. This table contains events related to these steps, starting from the initial phone registration until their account is verified. Note that the table contains events from users who started their registration between July 1 2021 and December 31 2021.

transaction_summary table (download here, ~63MB)

This table contains daily aggregation of deposits/withdrawal amounts; exchange volume; yield subscriptions / redemptions; various fees that the user generated; and finally their premium tier. It contains entries only from users who started their registration between July 1 2021 and December 31 2021.

acquisition_cost table (download here, ~8.3MB)

This table contains the acquisition cost for each user who started their registration between July 1 2021 and December 31 2021.

Tasks

I. Overview of 2021 H2, preparation for the follow-up tasks

Explore our acquisition during 2021 H2, put together a few slides that provide a good overview on the key elements. Present overall metrics on the number of users, conversion rates to verification/deposit/yield, as well as comparison of acquisition cost and revenues. What are the main bottlenecks that cause the biggest drop-offs? What were the most successful or most problematic/underperforming segments. Your investigation here will help you address the follow-up tasks, so check those first and plan your exploration accordingly.

II. Targets for 2022 H1

To help improve our business, the executive team decided to create two squads, one that would focus on improving our acquisition flow (up until user verification), and one that would focus on product discovery, so that more users engage with the smart exchange and smart yield, and so we generate more revenues. The product managers for these two squads need your help in selecting quantitative metrics that would serve as key results (KRs) to reflect the squads' performance/success at the end of the semester. Your job is to suggest 2-3 KRs for each squad, with 3 stretch goal targets for each KR. One target should be in line with past performance, one target should reflect an overall performance similar to what was observed in the most successful segment in the past, and one target that should correspond to a breakthrough success. Make sure that the KRs are clearly defined, and show how these numbers look for 2021 H2.

Bonus: Think outside the box. Feel free to use external data if it could complement the data we provided and would help give a more complete picture. Even if you can't actually find such data, feel free to bring ideas on what information could make your analysis more complete.

III. A/B experiment

The product manager for product discovery wants to improve the conversion rate from verification to yield subscription. Bring a hypothesis on how we could improve the current conversion, and formalize the details of an AB experiment setup: what is proposed experiment, what should be the target cohort, what would be the target metric, and how long the experiment should run based on the data we provided.

Note: If you suggested different KRs for the product discovery squad, feel free to do this exercise for one of those instead.

IV. SQL challenges:

Assuming you have the above tables in a database, explicitly write SQL queries for each of the following:

1. Write a query that gives the funnel of the user journey with conversion rates between each step as well as the overall conversion rate from the start of the user journey;
2. Write a query that gives the average acquisition cos per user, and average revenues per user within the first 3 months after they are verified, broken down by country;
3. Write a query that gives the percentage of verified users who have generated more fees than their acquisition cost within X months after their onboarding. The query should return this percentage for X between 1 and 9.

Deliverables:

Please provide the scripts/notebook of your analysis, and create a presentation summarizing your findings and recommendations that you could present to stakeholders (product/marketing managers).
For the questions about the SQL queries, please provide the queries and also create the plots that you would show to the stakeholders using the data we provided. Tip: You can use the pandasql library to test the queries.

Good luck!