Case Study for Product Analytics Candidates

Instructions

This case study is based on actual data and business problems from Babbel. It is confidential and should not be shared.

This is your opportunity to show off both your **data analysis** and **visualization** skills and your ability to **communicate** your findings to an audience of typical stakeholders in a compelling way.

Please work through the below case study and prepare a presentation (Powerpoint, Keynote, or comparable) to walk a mixed audience through your results during your case study interview. You'll have **20-25 minutes** to present. Afterwards, there will be about 20 minutes of time for questions and discussion. In light of the COVID-19 pandemic, please be prepared to present your slides remotely using Google Meet (formerly Hangouts) and ensure that your browser is set up for screen sharing.

Imagine your audience will be your typical stakeholder group (as well as potential team mates): likes straight to the point, likes actionable, likes easy to read.

We expect the entire challenge to take you 4-5 hours, but you may dedicate as much time as you like. Please note that we expect you to submit your presentation **within one week of receiving this case study**. Feel free to choose whatever tool you like to analyze and visualize the data attached, may it be Excel, Python or anything else. In case you do not manage to work through all questions, just let us know what you'd like to do.

Learning Engagement in the Babbel Ecosystem

Babbel is a quality leader when it comes to learning a language via digital self-study. We build our product for those who are aiming to become conversational in a language and are willing to pay for achieving this goal.

Our products offer a portfolio of Learning Experiences – different features offering learning activities with a stand-alone learning value proposition. Our main learning experiences are:

- **Lessons** – short units that introduce vocabulary and grammar via contextual content, including exercises
- **Review** – vocabulary review to practice previously introduced words and phrases
- **Audio experiences** such as Podcasts
- **Games** – fun and casual ways to practice vocabulary
- **Live Classes** – Digital live classroom (60 min) with a teacher and a small group of students

The respective learning content for these is designed by our team of linguistic experts and language teachers, and specifically tailored but not necessarily available in each learning language.

Task

A product team approaches you to **analyse how Babbel learners using games are engaging with other learning experiences** and identify possible avenues for improvement. The overall goal of the team is to activate learners in their early learning journey.

Please use the attached dataset and the following guiding questions to analyse the learning engagement performance of games, and point out optimisation potential:

1. Given the data you have available, how would you operationalise the team's goal in terms of specific metrics?
2. How do games perform differently across learner segments?
3. How are games being used in conjunction with other learning experiences? How does this behavior relate to the team's overall goal?
4. Assume that the team is planning to build a game aimed at activating new learners.
   To test if the game achieves this goal the team plans to release it as an A/B test but is unsure how and where exactly to expose new subscribers to it. From your experience and insights gathered from the dataset provided here, what would you suggest to the team? Are there any risks they should be aware of?

Beyond the data you have at hand, what kind of metrics would you use to monitor the learning engagement across different learning experiences? What kind of tracking would need to be in place for this?

Data

The dataset “learning_sessions.csv” contains learning session data of subscribers in their first weeks with Babbel. Each row represents a unique learning session.[1]

Column name	**Definitions**
uuid	String, user identifier
subscription_started_at	Timestamp of the start of the subscription (in UTC)
age	String, age bracket of the user
motivation	String, answer to question during registration: What is the main reason you want to learn a new language?
subscription_type	The last node visited by the visitor (Registration funnel consists of 8 nodes.)
geo_area	String, indicates the geographic areas that the subscription was purchased from ● ESM (English speaking other than U.S.A) ● GSM (German speaking) ● FIM (French speaking) ● SPM (Spanish speaking) ● USA
learning_activity	String, indicates the learning activity used in the session
learning_os	String, indicates the operating system of the user’s device
learning_language	String, indicates the language the user is learning
session_ended_as	String, state in which the session was ended (completed, aborted, missing)
session_started_at	Timestamp of learning session start (in UTC)
session_local_started_at	Timestamp of learning session start in user’s local timezone
time_spent	Float, minutes spent learning

[1] The data here is close to what you would be facing in an actual analysis but we have sampled, anonymised it, and changed some details here and there so as not to expose any sensitive information.