Data Analysis Test

General instructions

The goal of this test is to see how “data literate” you are, especially in the context of turning raw, granular data into useful insights.
We will look at the quality of the resulting analysis (structure, presentation, accuracy of findings), as well as your ability for understanding and manipulating data.

You are free to use whatever tools you want, and present your analysis in the format you think is most appropriate for the use case decribed below. We want to see what you come up with!

A bit of context

The data provided is usage data for a mobile application. The product owner, in charge of running the development team for the app, is interested in better understanding user behaviour and checking if the high-level KPIs for the application are improving with each release of a new version.
Some (but not all!) of the KPIs commonly used for mobile applications are: total number of users, monthly active users, retention curve, time in application, and more…

One way to perform this analysis would be to present a summarised overview of these KPIs and how they differ for the different versions of the application. You could then dig deeper into the metric you think will be most interesting for the product owner, and present your findings.

Data dictionnary

Users dataset

In this dataset, each row describes a unique user of the application.

- `user_id` : the user’s unique ID
- `install_date` : the date the user first installed the application
- `app_version` : the version the user downloaded. Only 3 versions of the app were released (1, 2 and 3)

For this exericse we will assume the user can't upgrade their application, so the version they have downloaded is the one they will always have.

Events dataset

The event dataset, each row describes an event that was triggered by a user performing a specific action. For this application, an event is triggered every time a user exits one of the 3 areas of the application (or closes the application).
It is as follows:

- `user_id` : the unique ID for the user who triggered the event
- `user_session_id` : the ID of the session, which is unique for every user for each session. Each session starts when the user opens or resumes the application, and ends when the user leaves or minimises the application
- `event_timestamp` : the time and date the event was triggered at
- `app` the area of the application the user was just in. The 3 areas for this application are: `build` , `draw` , and `learn` .
- `time_seconds` : the amount of time, in seconds, the user spent in that area of the app