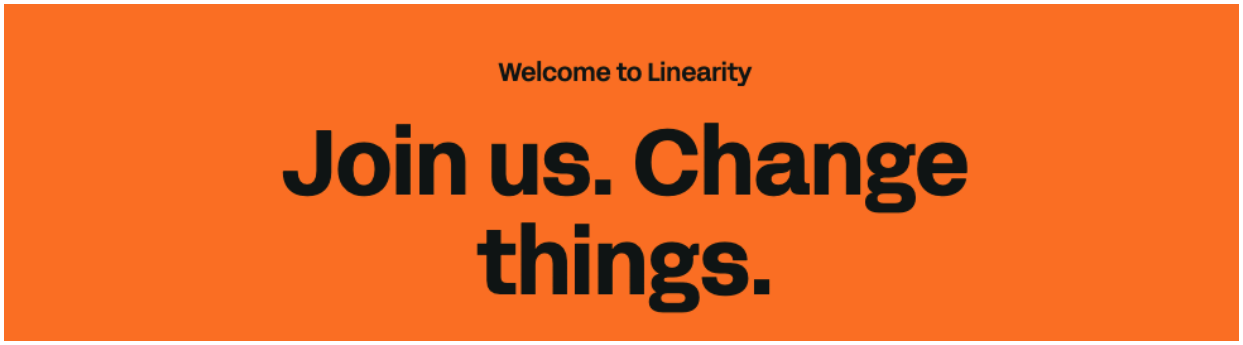

Lead Analyst Home Assignment

Context

Thousands of users use our application on a daily basis to create high quality graphic designs at-the-office or on-the-go. As of now, our application is free to use for everyone, all you need to do is download it from the app store. As soon as a user opens the application and starts using it, we receive that data in the form of live events. Every click inside the application, use of tools & different features are stored as an event in our database. Using this application data, we create a profile for each of our users with their activity timestamps & usage of most common features.

With continuous improvements and launch of new features, we decided to introduce signups (account creation) for our users. However, initially we decided not to keep it mandatory. On opening of the application, users see a pop-up to signup which they can dismiss or continue. If continued, the signup flow uses a web based backend system integrated inside the app-client. The same signup flow can also be accessed directly on our website as well where the user doesn't necessarily have to install the application first & instead create their account & then download the application. Like all other events, the account creation event is also captured by us but since it's based on the web, the account creation information is collected separately and stored in a S3 bucket in chunks of CSVs.

For every user, we have a unique identifier called ***user_id***. For every account created we have an ***id*** as the unique identifier. For existing users of the application user_id already exists and thus the same is assigned as their id for account creation. For new users who decide to signup first, id is created on account creation and then the same is assigned as a user_id as soon as they start using the application. This is done so that we have a common identifier to link the application usage with account creation.

Provided Data Sources

1. We have taken out a sample of user profiles data and stored it in a table in postgres database for you under "public" schema. It consists of users who used our application during a period of time and might or not have created an account with (signup).
   The table also has various other information about the users such as their last activity on the application, numbers of documents (graphic designs) created by the user, number of documents exported (sharing of document), etc.

Connection Parameters

Host	linearity-postgres.ci4darskkd34.eu-central-1.rds.amazonaws.com
Port	5432
database	tempdb
Username	linearity-home-assignment-user
Password	e0a1bf24bb7a81de7ac81834343ad6169711366863e0f6d66f71f2280e314668

2. A sample csv dump in a S3 bucket consisting of data of users' account creation (signup). The bucket has open access to read objects.

Bucket Name	data-temp-bucket
Region	eu-central-1
Filename	account_users.csv

Terminologies:

- Event - Any event sent via the app client, can consist of marketing events as well such opening of a notification.
- Active event - An event is considered active only if it involves user actions or activities such as drawing something or creating a new document. Marketing events are filtered out from these.
- Onboarding - A step by step quick tour offered to new users to make them familiar with the application.
- Export - When a user exports (share) documents in any offered format such as png, svg. (North Star metric)
- Activation - As soon as a user exports their first document, they are considered activated.
- Tools - Different features that help users build such pen tool, pencil tool, etc.
- First_7d - Stands for the first 7 days. Any action done in the first 7 days.
- Last_7d - Stands for the last 7 days. Any action done in the last 7 days.
- First or last _4w - Any action done in the first or last 4 weeks.
- First or last _12w - Any action done in the first or last 12 weeks.

Tasks

1. Write a python code to access the given postgres database & copy the user data in a Google sheet/csv.
2. Write a python code to access the accounts data csv from S3 bucket & copy its contents in Google sheet/csv.

3. Upload the obtained CSVs in your preferred BI solution. If you don't have one, we can recommend Metabase.
4. Basis the above two data tables, answer the following using SQL queries:
    a. What is the number of users & accounts created on a daily basis?
    b. How many users might have interacted with a marketing campaign after their last active session?
    c. How does the app usage looks amongst the users who signed-up vs who did not sign-up?
    d. What percentage of users activated within the first 7 days overall & by split between signed-up & not signed-up?
    e. Is there correlation between onboarding completion & activation?
    f. Users from which country are creating the most number of documents & what's their exporting percentage?
    g. What percentage of users are on the latest version of the application?
    h. How many users downloaded the application first and then signed-up vs how many signed-up first & later downloaded the app?
    i. Which os type has better number of exports?
    j. How many users never returned after the first week of usage? Also, share by split of signed-up vs not signed-up

Bonus:

1. Any other interesting insights you would like to share?