Dear CleverCogs team,

*xxxxxxxxxxxx. (Email subject and greeting stuff)*

After the initial data cleaning, we found some problems with the data that we’d like to check with your team. We would appreciate it if you could explain them to us, or we could also arrange a meeting with your technical staff to discuss them.

1. Users that are likely to be machines or internal testers
2. **Age too young**

From the platform’s regulation, we got to know that CleverCogs is only for users that are older than 18. But the data shows that there are a few users who do not meet this requirement.

1. **External ID missing**

In the database given, the external id is used to connect different datasets. However, some pieces of data lack external id, failing to gather further information from other datasets, which possibly makes the analysis inaccurate.

1. **Gender unexplained**

We noticed that there are three categories under the gender column in the database, that is female, male, and unspecified. We were wondering does “unspecified” mean non-binary, or not willing to say, or not a real user. As you know, gender can be a very important factor when it comes to medical conditions. Also, we tried to clean the data that fails to meet the above two conditions, and after that, there are no longer users with unspecified gender.

**Our assumption** is that these “users” are all machines, not real people, therefore should be deleted from the database and not be considered anymore. We’d like to hear your team’s opinions before moving to the next step.

1. Categories xxx (Muyu’s part)

*Xxxxxxxxxxxx (Ending words)*