**Part 2: project report**

The project consisted in an exploratory data analysis of the user data provided.

For that purpose, three datasets have been used:

* User dataset
* WatchedVideo dataset
* Video dataset

Here are the steps and the analysis that have been performed.

1. **Data cleaning and analysis details**
   1. **User dataset**

I first started analyzing the user datasets alone.

Interestingly, the 5 classes of users are quite balanced, with a slight predominance of ‘Gamer’ and ‘Super Gamer’. This might indicate that the app is more popular on the game side.

For what concerns the created accounts over time, I first checked the time range covered and found out that the sampling has been entirely performed in 2021. I then plotted a bar chart indicating the number of accounts created over the year and found a trend that peaks in July and has its lowest value in December.

For what concerns the spatial distribution of the accounts, I plotted them on a world map using the provided coordinates and found a quite uniform distribution.

I’ve also plotted the geographical data over time, to see where the first downloads have been performed, and where the most recent users are located.

The app seems to be quite popular in the US, in Europe and in India. On the other hand, China, Africa, and South America have registered a lower number of downloads.

* 1. **User + watchedVideo datasets**

After this first analysis, I merged the user and the watchedVideo datasets to have more information regarding the activity of the users.

A first interesting finding is that there is not a preferred OS: users download and watch the video both on iOS and Android.

There is instead a preference on the length of the videos: users tend to watch them for and average duration of 3.41 x (it is not clear if the duration is expressed in minutes or seconds. Through some exploration of the app, I’ve assumed seconds).

After that, I’ve performed some analysis on what class the most active users belong to and where they are located (assuming the activity is determined by how many videos a user has interacted with): interestingly, although in the previous section it has been showed that the majority of users are located in US, Europe and India, it seems that many of them are ‘inactive’. The class activity hasn’t revealed any interesting finding.

A last interesting analysis I decided to operate in this section concerns the time of the day users are more likely to watch videos. It seems that there’s a preference for morning time.

* 1. **User + watchedVideo + video datasets**

Finally, I’ve performed an analysis on the user video preferences, merging the previous two datasets with the videos one.

In this section, in particular, I’ve calculated an index of ‘popularity’ for each video, by multiplying the number of visualizations (i.e. how many people watched the video) by the number of shares. Based on that, I was able to determine the characteristics that make a video more popular: without entering the details, at a first look it seems that the 11th attribute\_id (which I mapped as ‘K’) and the 1st and 3rd game\_id (which I mapped as G1, G3) are the most significant indexes of popularity.

Also, the AR videos have registered more views and shares than the non-AR ones.

* 1. **Limitations**

The analysis performed could serve as a starting point for a deeper and more accurate one.

Some of the limitations / aspects that I have figured out are:

* Since the app is pretty recent, there’s a limited amount of data.
* There’s not enough information concerning the ‘Gaming aspect’ of the app, which is one of the strengths of it.

1. **Product recommendations**

From what I could experience by analyzing the data and downloading / playing the app for some time, I came out with the following considerations:

* I would introduce themes for major events / celebrations. For example, previously we discussed how in December it has been registered lower number of downloads: the introduction of a ‘Christmas theme’ could have engaged the users more.
* Since the users are more attracted to certain video features (for example the attribute ‘K’), I would recommend to further explore those and come out with even more engaging ones.
* The app is popular in India. On the other hand, in the US and in Europe, it seems that the users are less active. Therefore, I would suggest to further analyze these users and come out with a strategy to make them more involved.
* It would also be a good idea to realize a survey (simply rate the app) for the actual users.

1. **Next steps**

The major upgrades which I would suggest for the app are:

* Introduction of multi-player games / multi-avatar videos (with the possibility to interact with each other): in this way, the users would be even more engaged.
* Introduction of competitive games / tournaments (maybe a dance competition!).
* Introduction of animal avatars.
* Improvement of the AR and introduction of features related to the place one is in: ‘Pokemon GO’ has become really popular because it involved a virtual infrastructure on top of the physical world. Krikey could introduce a similar network, where for example users would need to complete challenges in specific places.