Down Data Analyst Home Assignment

Exposure Dashboard

Our company has experienced a significant increase in new sign-ups, making it essential to monitor and optimize how these new users interact within the app. To achieve this, we need to develop an Exposure Dashboard that **tracks and analyzes user interactions**, with a **focus on new user exposure**.

An **exposure** happens when a user's profile gets viewed by another user, who can choose to send a like, or skip that profile. From the perspective of the **user being viewed, we call it an exposure**. From the perspective of the **user who is viewing, we call it a profile view.** Both are important, but for this assignment we will focus on exposures.

Dataset

1. **agg_profile_views.csv**: Contains profile view interactions.
    - ds : The date on which the profile views occurred.
    - user_id : The unique identifier of the user who viewed another user's profile.
    - viewed_user_id : The unique identifier of the user whose profile was viewed.
    - filter_type : The type of filter applied when the user viewed the profile
    - cnt : The count of profile views that match the specified criteria on the given date.
2. **dim_user table**: Contains user demographic and status information.
    - created_at : The timestamp of user creation
    - gender : gender
    - user_id : The unique identifier of the user
3. **bad_actor table**: Contains users flagged as bad actors, meaning they have been banned for violating our policies.
    - user_id: The unique identifier of the user

Considerations

- The table dim_user contains duplicates, this can happen because a user deletes and opens the account several times or even because the user installs and uninstalls the app several times.

   For this reason, I will consider only the first created_at (first time that a user creates an account), and the most common gender (if the user creates the account 4 times, 3 as men and one as women, the user will be classified as men).
   Finally, I created a field called reactivated_user that is True if the user has more than one different created_at different dates and False if the user only has one created_at date.

With duplicates the table contains 230,768 rows, without duplicates, the table contains 174,197 unique users.

● Users flagged as bad actors are not being considered in our analysis, because it can be a non real exposure.
● Cnt: after some investigations, I understand the cnt as the number of times that a user A views a user B in one day using one filter type. The following table shows an example:

| Row	ds ↑	user_id	viewed_user_id	filter_type	cnt
1	2024-08-02	eUGMGDNP+sug1jJ0/7+Q2d6...	+RjXurKChswfGXO24P2Q3pP9...	nearby	2
2	2024-08-03	eUGMGDNP+sug1jJ0/7+Q2d6...	+RjXurKChswfGXO24P2Q3pP9...	nearby	7
3	2024-08-04	eUGMGDNP+sug1jJ0/7+Q2d6...	+RjXurKChswfGXO24P2Q3pP9...	nearby	2
4	2024-08-05	eUGMGDNP+sug1jJ0/7+Q2d6...	+RjXurKChswfGXO24P2Q3pP9...	nearby	11
5	2024-08-06	eUGMGDNP+sug1jJ0/7+Q2d6...	+RjXurKChswfGXO24P2Q3pP9...	nearby	11
6	2024-08-07	eUGMGDNP+sug1jJ0/7+Q2d6...	+RjXurKChswfGXO24P2Q3pP9...	nearby	10
7	2024-08-08	eUGMGDNP+sug1jJ0/7+Q2d6...	+RjXurKChswfGXO24P2Q3pP9...	nearby	3
8	2024-08-09	eUGMGDNP+sug1jJ0/7+Q2d6...	+RjXurKChswfGXO24P2Q3pP9...	nearby	2
9	2024-08-10	eUGMGDNP+sug1jJ0/7+Q2d6...	+RjXurKChswfGXO24P2Q3pP9...	nearby	5

● Table profile views contain views only for August 2024. Thus, we are not going to consider users created after 1st of september.
● The user uses mostly the nearby option to view other profiles (86.8%), then 10% of users are looking for a 3some (exposed user needs to be open to 3somes), and around 2% are looking for users in specific locations (not nearby, i think that this is a premium feature).

filter_type	unique_users	table_rows	ratio	rows_perc	rows_cummulative_p	total_rows
nearby	173960	1527554	8.78	86.8	86.8	1760277
3some	56213	176604	3.14	10.0	96.8	1760277
hot	8448	21003	2.49	1.2	98.0	1760277
new_york_ny	4145	17352	4.19	1.0	99.0	1760277
los_angeles_ca	6091	15651	2.57	0.9	99.9	1760277
chicago_il	868	1490	1.72	0.1	100.0	1760277
dallas_tx	177	389	2.2	0.0	100.0	1760277
null	23	193	8.39	0.0	100.0	1760277
asia	35	35	1.0	0.0	100.0	1760277
visited_you	6	6	1.0	0.0	100.0	1760277

Filters were reclassified in 4 categories: nearby, 3some, hot, location and others.

● Another important topic is the definition of a new user. To address this topic I consider the sign-ups of 1st or 2nd of August and I study the number of active users every day since the sign up. And I saw that the percentage of active users drops from almost 60% of the users who sign-up to only 20%. Then after 1 week, the percentage of users drops to 10%.
For this reason, I consider a user in the category **New** if the user has less than 7 days since the sign up because of the significant drop, and because every user has one week to use the app (considering that user uses the app more on the weekends).

% of Active User per Days Since Sign-Up
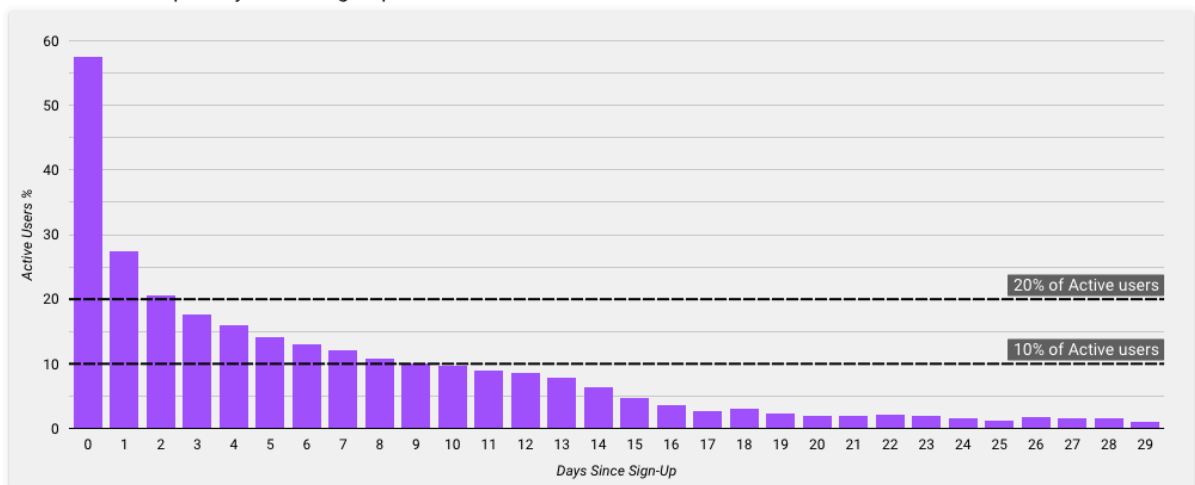
- In table agg_profile_views we have only viewed_user_id created before 14th of August.
- Finally, it is important to consider that not all the users have exposures, over the users that sign-up between 1st and 13th of August (38,121), only 21,3% (8,119) were viewed by someone in August and 95,7% (36,434) view at least one profile, maybe the rest are user who didn't complete their profiles. In this dashboard we are only interested in studying the exposures of the new users.
- In our user database, we have 174,197 users (not included in bad actors), where 93,5% (162,846) are males and 6.5% (11,323) are females. The category others is composed only of 28 users, so I removed them in this analysis.
- I removed the category Other in the type of filter.

Key Metrics

- **# exposures**: How many views a exposed user receives in a period of time. Here we are considering the fact that a user can view another several times. It is a metric which indicates intensity, derived from how engaged are the users with a specific profile. More exposures per user indicates that the exposed user brings more interest from others.

- **# of unique viewers**: How many unique viewers an exposed user had a period of time. This helps to measure global exposure and popularity of each user.

- **Profile views vs Exposure**: This ratio indicates how many profile views an exposed user receives for every view that the user gives to someone else.

- **% of interactions**: This value shows the percentage of unique viewers which the exposed user is having an interaction with (the users view each other, not necessarily the same day). It helps to understand if the exposed user is just being viewed or if the exposed user is engaged with our app and is using it to interact with others. This indicator should be really interesting to create a scatter plot where every point is a new

user and the other dimension is the number of exposures. Because with this scatterplot the users can be classified by its exposure and engagement.

Dashboard

Dashboard link

Explain Business Value

The first important point that is already mentioned in the consideration chapter is that most of the new users are males (93,5%), this ratio male/female end up exposing the new females sign-ups 125 times more than the new males sign-ups (1,250 vs 10 exposures) and in this particular dashboard we are only considering users with exposures, we have a lot of males without one single view.

Moreover, the ratio between profile views vs exposures in females are close to 100 per day, so over 100 views that a female receives, only visits one profile. That can be a little bit overwhelming, provoking a lack of interest in the app by the women and also a lack of interest in the males because of the low exposure. A list of recommendations to improve this situation can be:
- Create a tutorial with tips for males to increase their number of visits in an organic way. For example, we can suggest which photo is the most likable by other people (using AI) and which photos should be removed.
- Encourage users to fill their profiles and create smart filters to suggest females the best profiles for her (according to her profile and what she is looking for).

One interesting fact is that if a male is open to 3somes, he receives much more exposure 25 vs 9 (or 15.8 vs 6.2 unique viewers), this can be because are not a male is a couple that is looking for another person. A recommendation can be to let the user the option "couple" when she or he is creating the profile (not just male or female or trans).

Exposures and unique viewers are decreasing after the sign-up, especially after the 2nd day (day after sign-up). I recommend taking actions to make the decrease smoother.

The second sheet is helpful to export the data and analyze the viewers and exposures per user considering different filters, such as the search filter, gender of the exposed user and if the user potentially interacts with a visitor. With this information, CRM can make marketing campaigns classifying the users by its exposure. One key point is that if CRM wants to get the top popular users, they will need to filter by gender. Because the threshold to consider a user as a popular in males is lower than in females.

Future steps

One interesting aspect that can be added in a second version is to check when the peak of exposures happens, to alert the users when the activity in the app is bigger and increase the user engage with the app.

Some interesting dimensions that can enrich the dashboard are:
- User intention: just fun, short term relationships or long term relationships.
- Location: Country, City, Postal Code.
- Age.
- Premium user vs Freemium (can be premium and new as well).