A. Problem Reflection:

1. The core problem appears to be the drop-off when the excitement of using a new app wears off and the initial exploration phase ends. Users are no longer intrigued by what the app has to offer, and once they're out of the app, they tend to forget about it altogether. The app is failing to create a ***habit loop*** or ***provide*** the ***sense of instant gratification*** that other high-engagement apps successfully deliver.

2. There are a few assumptions I would make given the limited context and they are as follows:
    a. The app is likely ***not capturing key components of user behavior*** that could help create a more personalized and engaging experience (e.g. their ***preference*** for ***content type*** or ***length of content***).
    b. Another assumption is that ***no push notifications*** or ***reminder emails*** are being ***sent to users*** (e.g. to inform them they're losing a challenge streak or that new content has been added.)
    c. Users may ***not*** be ***receiving*** the kind of ***instant rewards*** or ***gratification*** that they've come to expect from other apps.
    d. The last assumption would be that there are ***no feedback loops*** in the app and the users are not able to actively tweak the content based on their likes or dislikes.

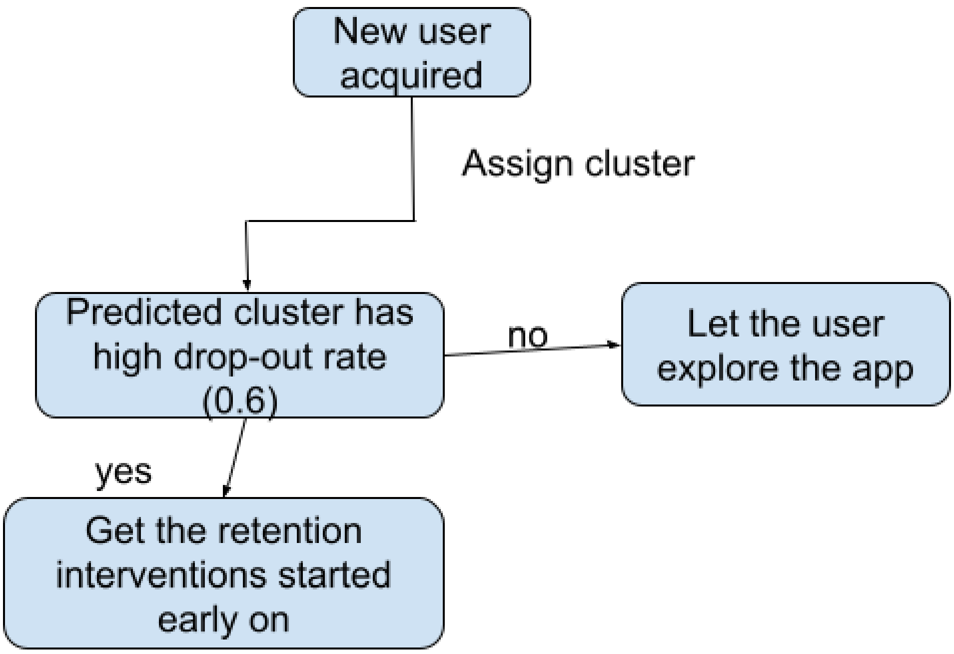
3. The signals which may help early user engagement are as follows:
    a. **User Goals:** If the users have a defined goal like weight reduction or eating better, then the content could be tailored accordingly.
    b. **Interactions with the app:** how is a user interacting with the app, which areas of the app or which features the user is finding most engaging or is interacting with the most, it could even include the content type or format.
    c. User Demographics: Understanding the underlying demographic differences could make the app experience (seem) more personalized in the initial days, and could aid engagement in the early days of app usage.
    d. Content Delivery time: Some scroll social media during lunch breaks, others do not get active on their phones until they are back from work in the evening. If this difference can be leveraged then it could have an impact on user engagement.

B. Solution proposal:
The AI/ML based approach that I have proposed here is a two step process:

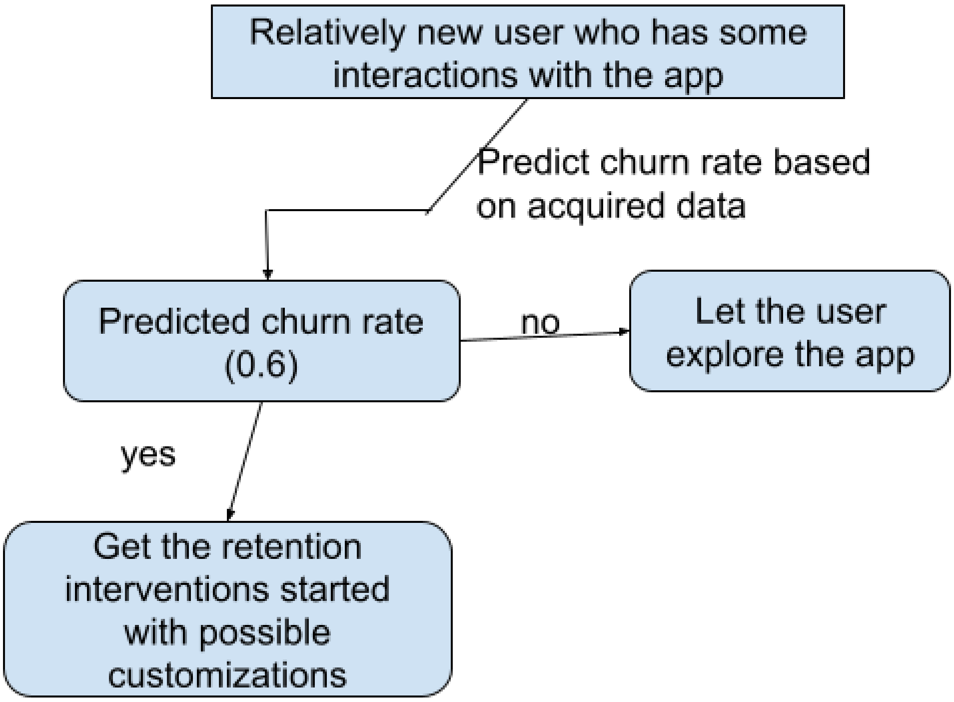
1. **Clustering users:** As soon as a new user enters the system, based on their on-boarding questionnaire they are assigned to a cluster. This cluster assignment is based on an already trained K-Prototypes clustering model, which was trained on existing user data. The output of this not only tells us which cluster the user currently belongs to but also tells us the churn rate of users from that cluster. This could be an alarm already for us and if the average churn rate of the users from the cluster is high then we need to keep special attention for the user.
2. **Churn Prediction:** This model is implemented for the users who have interacted with the app a bit and based on their interaction the probability of them dropping-out of the app is predicted. This could help us monitor early-on who are the users who may drop-off.

Fit in User journey:

1. **Clustering users:**



2. **Churn Prediction:**



Idea for further improvement: To further improve this process the content consumed by the users who didn't drop-out of that cluster can be presented to the new users as this behavioural similarity may give the users what they are looking for and may help in user retention.

C. Trade-offs and risks:

1. **Explainability:** AI/ML methods do create a black-box specially the random forest in the current implementation may prove to be a bit difficult to explain. However the visualzition for Random Forest may aid us in understanding the decisions better.
2. **Fairness:** For this assignment relatively sensitive feature like gender is used but in real world data it may introduce some bias.
3. **Scalability:** The ML models used here (K-Mod and Random Forest) may work for smaller dataset but might have some potential issues when we have larger datasets.
4. **Feature engineering:** More relevant features would potentially help in prediction improvements.
5. **Data:** The prototype currently uses synthetic data but working with real-data wouldn't be this straight forward.
6. **Cold Start:** The synthetic data needs to exist before the clustering can take place and also some sessions need to be present for the churn prediction
7. **Latency:** currently with this small dataset the predictions and data creation are relatively faster but this would be bigger issue when deployed to production.

D. **If you had just 2 days in a discovery sprint:** If I had just two days in discovery sprint then I would have focused on sticking to on one hypothesis that probably the users of a certain demographic are more likely to drop-off and then would have created some experiments supporting it and then could have figured out if the hypothesis was correct or not. This hypothesis stems from the fact that there may be some inherent bias in the content which is generated.