Analysis of User Data at Augie Studios

Authors: Ayush Gupta, Joaquin Wang Zheng, Martin Cheong | Advisor: Zach Branson | Client: Coleman Isner

Background

Introduction & Research Questions

Augie Studios integrates AI into video creation and editing. Using data about the videos users have created from October 2024 to February 2025, we focus on two main research questions:

- How do we measure **engagement** with Augie Studio softwares, and what does that look like?
- What **features/behaviors** distinguishes an **engaged user** from a **non-engaged user**?

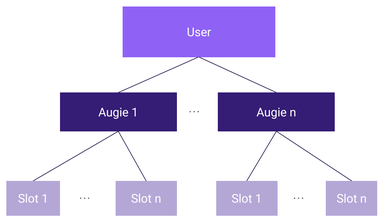
Answering these questions will help the company assess demand for their product over time and understand user behavior. This will help the company prioritize certain software features and marketing strategies.

Data

Description

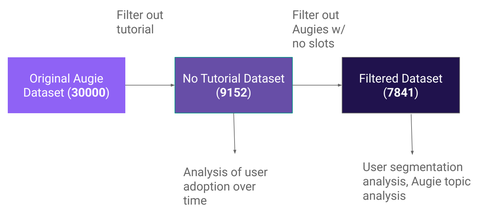
We primarily worked with **3 datasets** that Augie provided: User data, Augie data, and Slot data. Each user can make some number of **Augies (videos)**, and each Augie can contain some number of **slots (clips)**. All three datasets were linked by unique IDs. Examples of variables we focused on:

- **User Search Text:** For topic analysis of each Augie/User
- **User Summary Statistics:** total number of Augies, slots, video length, etc. created by each user to measure engagement



Preprocessing

Data provided included the tutorial video that came pre-loaded with each account, which we filtered out to focus on user-generated content in all our analysis. Additionally, some Augies did not map to any slots in the accompanying dataset, which we removed for the user and topic analysis.

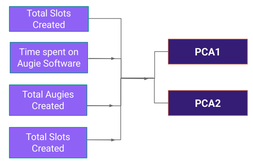


Methods

LDA / Topic Modelling

We performed Topic Modeling on users' search text using Latent Dirichlet Allocation (LDA). The text data was preprocessed by removing stop words (eg. a, the, etc.) then applying LDA to find the main topics to characterize and group search behavior into interpretable themes.

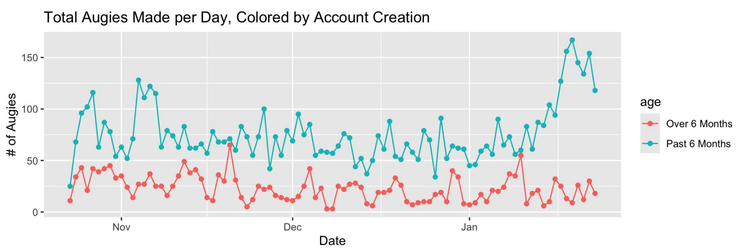
PCA Analysis/K-Means Clustering



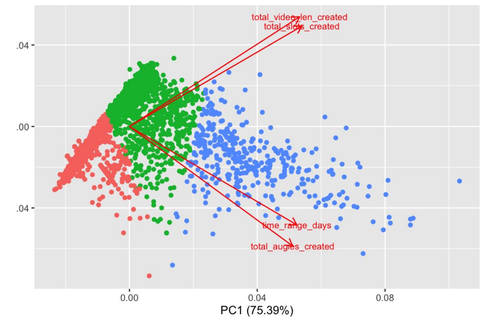
- PCA Analysis was used to convert four user summary statistics into two principal components
- K-means clustering was performed to divide Augie Studio users into three clusters based on principal components 1 and 2
- 3 clusters were formed (elbow plot was used to determine number of clusters)
- This allowed us to understand unique segments of users that engage differently with Augie Studios

Results

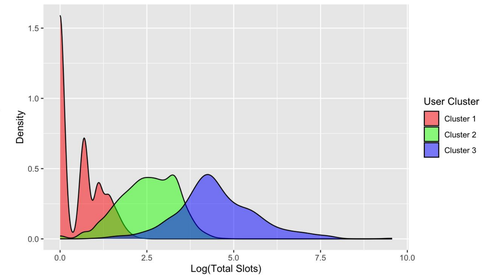
Augie creation rates remain level for older users but growth is being primarily driven by newer users, or those who created their accounts within the last 6 months.



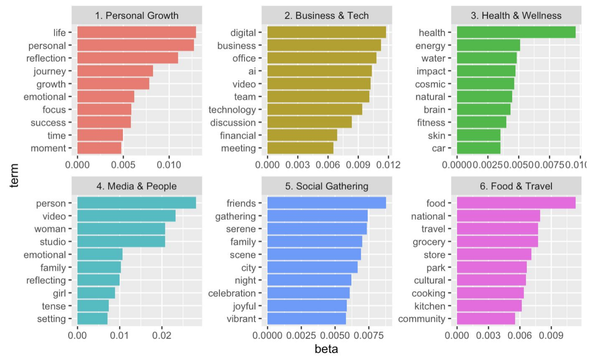
PCA Loading Plot + K-means clustering on user statistics. PC1 seems to be a measure of user engagement, and 3 clusters were formed.



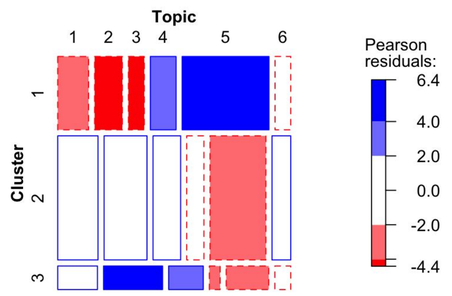
Distribution of log(total slots) for each user cluster. From least slots -> most slots is red -> green -> blue (similar pattern shown for other variables)



LDA revealed 6 topics in the user search behavior. The topics range from personal growth, business & tech to food & travel.



Mosaic plot showing distributions of topics by each user cluster. Cluster 1 (less engaged users) talk more frequently about Media & People / Social Gathering, whereas Cluster 3 (active users) talk more about business & tech / health & wellness.



Comparison word cloud revealed that user search text differs between highly active users and not engaged users where engaged users more frequently search about life, personal, growth, and emotional.

Conclusion

- User growth is slightly upward and newer users are engaging more with the product, driving content creation more than older users.
- Three different types of users were identified: those highly engaged with the product (10%), those not very active (39%), and those who spent some amount of time (51%)
- We noticed that less active users tended to favour topics related to social gathering, while more active users favored Business & Tech / Health & Wellness.