Analysis of User Behavior on a Children's Video Book Platform

By: Michelle Chen, Aadi Menon, Hannah Hayes | Advisor: Zach Branson

INTRODUCTION

- Vooks is a video book platform targeted at children ages 2-8
- We were provided data on video views, book information, and channel information
- Channels are playlists of books: when watching through a channel, books autoplay in a predetermined order
- Research Questions:
 - What factors affect **viewer watch time** and **retention** on the platform? • What factors are associated with **high engagement**?
- By answering these questions, we aim to help Vooks understand their user base and make improvements to the platform and/or their marketing tactics

DATA

Provided Data:

- View Dataset: 500,000 observations
 - Unique video view attempts made by users
 - 444,900 observations were from subscribed users
 - 66,000 unique users
- Book Dataset: 550 observations
 - Video book metadata
- Channel Dataset: 3300 observations
 - Timeline of updates to the channels on the platform
 - Linked to the book dataset via IDs

Feature Engineering:

- Eliminating redundant variables
- Creation of time zone variable from latitude and longitude data
- Creation of time lag difference variables
- Channel Version: constructed using the timestamps of the update



METHODS





View distribution across time in UTC

Flatten the datasets into one final dataset, labeling rows by the timestamp

> IF THERE IS A MATCH: Classify the matching subsequence as a channel view

OTHERWISE: Classify the matching subsequence as not a channel view

Distribution of views based on whether or not the book was viewed in a channel



PCA Biplot of k-Means clustered users and variables



Group 1 has a strong association with high number of unique watches, Group 2 with high distances between watches, Group 4 with one-time watches.

CONCLUSION

- No definite results on whether or not Vooks should increase promotion of their channels
- Channels may improve popularity of lesser known books
- Vooks may want to experiment with tailoring experiences based on time of watch (ie: boost educational videos during school hours and boost casual videos outside of school hours)
- Users can be characterized by totality and consistency of views
- Next steps:
 - Generalize findings on the full dataset
 - Improve channel classification pipeline with more insights on how channels function

ANALYSIS + RESULTS

Top 10 books with the highest proportion of their views from being in a channel



Hourly local time of Vooks usage of the top **10** countries

Most countries tend to follow a spike in the morning-afternoon, with a level-off in the evening. Some have a more bimodal distribution, and some spike in the evening.



• Watch behavior is heavily related to the country that the user is watching from

REFERENCES

Correlation Matrix : A Quick Start Guide to Analyze, Format and Visualize a Correlation Matrix Using R Software - Easy Guides - Wiki - STHDA https://www.sthda.com/english/wiki/correlation-matrix-a-quick-start-guide-to-analyze-format-a nd-visualize-a-correlation-matrix-using-r-software. Accessed 27 Apr. 2025.

"How to Build a Segmentation with K-Means Clustering and PCA in R." Step By Step DSci, https://www.stepbystepdatascience.com/segmentation-with-kmeans-and-pca. Accessed 27 Apr. 2025.

Pipis, George. How to Determine the Number of Clusters for K-Means in R | R-Bloggers. 17 May 2020,

https://www.r-bloggers.com/2020/05/how-to-determine-the-number-of-clusters-for-k-means-i n-r/

Plotting PCA (Principal Component Analysis) https://cran.r-project.org/web/packages/ggfortify/vignettes/plot_pca.html. Accessed 27 Apr. 2025.

Preacher, Kristopher J., and Robert C. MacCallum. "Repairing Tom Swift's Electric Factor Analysis Machine." Understanding Statistics, vol. 2, no. 1, Feb. 2003, pp. 13–43. DOI.org (Crossref), https://doi.org/10.1207/S15328031US0201_02.

RPubs - Clustering on PCA Results. https://rpubs.com/Bury/ClusteringOnPcaResults. Accessed 27 Apr. 2025.

Carnegie Vlellon **University**

While channels may not be the preferred method of watching content on Vooks, channels appear to boost the popularity of books in foreign languages.