Investigating Content Engagement on TikTok among Wellesley College Students

NAIMA ABDIRAHMAN, BERNADETTE HARGROVE, FRIDAH NTIKA, and RACHEL XU, Wellesley College, USA

This study investigates whether Wellesley College students engage with similar content on TikTok. Leveraging metadata from Project 2, various analytical methods such as cosine similarity, K clustering, and hypothesis testing are employed for analysis. The findings aim to provide insights into the content preferences and trends among Wellesley students on the popular social media platform TikTok.

ACM Reference Format:
Naima Abdirahman, Bernadette Hargrove, Fridah Ntika, and Rachel Xu. 2018. Investigating Content Engagement on TikTok among Wellesley College Students. 1, 1 (April 2018), 1 page. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

In recent years, social media platforms have become integral parts of daily life, facilitating communication, entertainment, and information dissemination on a global scale. Among these platforms, TikTok has emerged as one of the most popular, particularly among younger demographics due to its short-form video format and algorithm-driven content recommendation. What users consume on TikTok and how they engage with content sheds light on their interest in certain aspects of their life and the world - whether it's sports, dance, or baking tutorials. This study focuses on a specific demographic: students at Wellesley College. By examining whether Wellesley College students engage with similar content on TikTok, we aim to understand their content preferences and trends within this unique cohort. In previous projects, we created personas and set up bots to like TikTok videos that have relevant hashtags and investigated the bots' video recommendations given their distinct interest. We found that TikTok has a relevantly good grasp of the bots' interest in videos and will suggest videos on their For You page to enhance engagement. However, the study's result varies on each persona - whether they are too niche or too broad, and was only run in a short period. Therefore, in this study, we aim to look at the metadata of already-existing users, and analyze whether TikTok feeds them content based on their "persona" or interest.

2 LITERATURE REVIEW

We incorporate insights from Boeker and Urman's study, "An Empirical Investigation of Personalization Factors on TikTok." Their meticulous analysis of user interactions and metadata reveals crucial factors shaping content personalization on the platform. Specifically, they find that certain factors exert a stronger influence on TikTok's recommendation algorithm than others. They note, "The order of the most influential factor to the least among those that were tested is the following: (1) following specific content creators, (2) watching certain videos for a longer period, and finally (3) liking specific posts." Furthermore, Boeker and Urman's research underscores the complex interplay between user preferences and platform algorithms. They emphasize, "Through our analysis, we identify significant correlations between factors such as user persona, video view rate, and content similarity." These findings highlight the role of personalized content delivery for the TikTok user experience. By integrating insights from Boeker and Urman's study, we aim to deepen our understanding of TikTok's recommendation system and its impact on user browsing habits among Wellesley College students.

3 DATA AND METHODS

3.1 Data

To analyze whether TikTok's algorithm recommended similar content to people in the same area, we collected data from fellow students at Wellesley College. We received 9 CSV files and 1 JSON file. The JSON file contained the video browsing history dates and video URLs. The csv files contained metadata on the content of each video, such as video description, author and video IDs, suggested keywords, shares, and likes, among other variables that had been scraped from their respective JSON files using a Python package, Pyktok. We ran Pyktok on the donated JSON file to obtain the metadata needed for our study, totaling the number of CSV files to 10.

3.2 Methods

3.2.1 Topic Modeling. One way we investigated our research question on whether Wellesley students view similar content on TikTok was through topic modeling. From each csv file, we extracted all video descriptions into one file, the text corpus, and each user (csv file) was regarded as a document. We used the Latent Dirichlet Allocation (LDA) technique to choose topics from the distribution of text in each file and words that led the model to assign a certain topic. Once we had done this for all the documents, we combined all the topics (columns) and users (rows) into one dataframe. The cells represented the probability of a topic appearing in a given document.

4 CITATIONS AND BIBLIOGRAPHIES

Authors' address: Naima Abdirahman; Bernadette Hargrove; Fridah Ntika; Rachel Xu, Wellesley College, 106 Wellesley College Road, Wellesley, Massachusetts, USA, 02481.

XXXX-XXXX/2018/4-ART $15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn