Uncovering User Behavior in Digital Music Marketplace

A Cluster Analysis of Bandcamp Sales

Nguyen Nguyen | Github:
Pace University, Seidenberg School of Computer Science and Information Systems | Advisor: Dr. Christelle Scharff

Abstract

Motivated by the dynamics of the digital music distribution landscape beyond streaming services, this study leverages clustering to analyze a dataset of one million sales on Bandcamp, identifying four distinct transaction clusters that highlight unique music preferences and spending habits. Findings demonstrate Bandcamp's potential in supporting artist autonomy and user engagement and the need for further improved clustering and enhanced handling of 'tagless' items. The study also highlights the need for further processing of text data beyond music tags.

Dataset

The dataset [9] used comprises 1,000,000 sales spanning a one-month period (9/9/2020 - 10/2/2020) collected from Bandcamp's frontpage feed. To enhance the analysis, keyword tag features were incorporated by merging with a separate dataset [9] based on shared URL information.

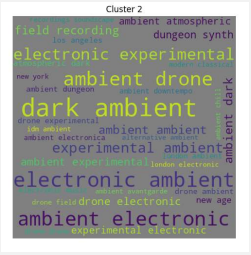
Related Works

Digital music distribution, particularly within the do-it-yourself (DIY) landscape, remains an understudied domain in e-commerce research. Existing studies on clustering and segmentation have focused on sentiment analysis of text data [7] and traditional numerical data [3,4,5], including item descriptions for inventory categorization [6,8]. The widely adopted K-means algorithm has been preferred for its simplicity and usability in these studies..

Methodology

Utilizing K-means clustering, this study initially clusters item tags on Bandcamp and subsequently applies clustering to the full transaction dataset. NLP techniques, including TF-IDF and Wordcloud visualization [6,8], are employed on music keyword text to generate a new feature representing music genre. This feature is integrated into the processed data alongside other relevant features such as pricing (converted to USD) and item format. The optimal number of clusters (K) for both clustering stages is determined using the elbow method [5] and validated using silhouette score and Davies-Bouldin index.

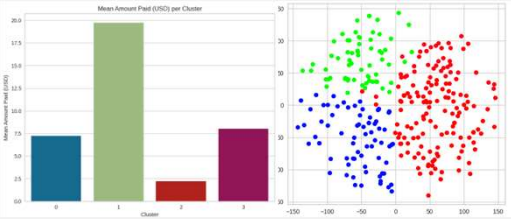
Genre Cluster WordCloud

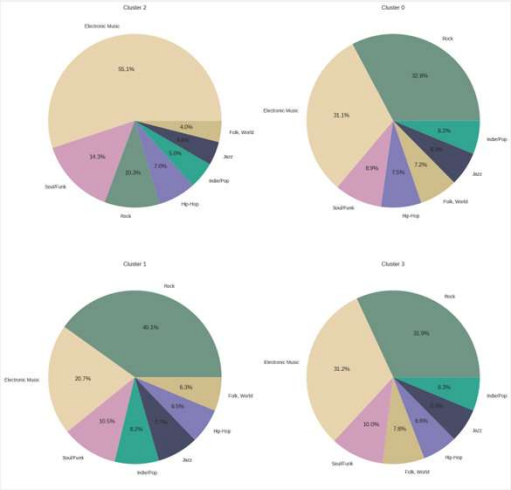


Tags' tf-idf scores

	tfidf
suny	11.341995
free french	11.341995
blues deep	11.341995
blues dixieland	11.341995
free drum	11.341995
punk nwobhm	11.341995
noise florence	11.341995
noise free	11.341995
blues electroacoustic	11.341995

TSNE plot of transaction data sample





Results

4 groups of buying patterns were formed:

Group 0: Eclectic listeners with diverse music preferences, including electronic and rock genres, with a particular interest in ambient music – 35% of dataset

Group 1: Enthusiasts of physical merchandise, with the highest mean payment at $19.72 USD. – 24% of dataset

Group 2: Dance music aficionados and DJs, favors electronic music, often opting for individual tracks. – 26% of dataset

Group 3: Generous supporters mirroring Cluster 0's music preferences, but distinguished by their heightened generosity – 14% of dataset

Sales Distribution by Format Table

	Album	Tracks	Physical	Discog
Group 0	72.6%	X	X	70.5%
Group 1	X	X	100%	6%
Group 2	X	100%	X	0.1%
Group 3	27.4%	X	X	23.4%

Conclusion and Future Work

This analysis hints at a variety of distinct user segments within the Bandcamp community.Despite these insights, the analysis presents several limitations such as the silhouette score (0.135) of music tag clusters and Davies-Bouldin index (0.826) indicates scope for improving cluster separation and compactness.

Further refinement of 'tagless' items and a more granular exploration of text description and genre categorization are recommended for future research to uncover additional dimensions and nuances in user preferences and purchasing behaviors.

References

1. Hesmondhalgh, D., Jones, E., & Rauh, A. (2019)
2. Garima, H. Gulati and P. K. Singh (2015)
3. D. V. S. Shalini, M. Shashi and A. M. Sowjanya (2011)
4. R. S. Bahri, I. D. Sudirman, I. D. Utama and R. H. Susanto (2023)
5. Phan Duy Hung, Nguyen Thi Thuy Lien,Nguyen Duc Ngoc. (2019)
6. Boyu Shen (2021)
7. Jorge Fresneda, Jeremy Hui, and Chelsey Hill (2022)
8. Jingyi ye (2021)
9. Andrew Thompson (2021)