International Conference on Informatics Engineering, Science & Technology

Session-Based Segmentation of Music Listening Behaviour on Digital Platforms Using K-Means Clustering

Dionisius Salvavictori Wanggur[1*], Rahma Wahdiniwaty[2], M. Yani Syafei[2] Agung Prayoga[1]

[1] Master of Information Systems, Universitas Komputer Indonesia, Bandung, Indonesia
[2] Universitas Komputer Indonesia, Bandung, Indonesia

[*]Corresponding E-mail: dionisius.75124009@mahasiswa.unikom.ac.id

Abstract. Gaining insight into digital music listening behaviour is essential for optimizing personalization and enhancing user engagement. Instead of using traditional aggregated data, we employed a session-based methodology to reveal detailed interaction dynamics. A "session" was simply defined as a set of tracks played sequentially within a 30-minute window. By processing playback data into features and applying K-Means clustering, we successfully delineated four unique listener profiles. These profiles differ significantly across three dimensions: session length, skipping frequency, and listening time patterns. The practical relevance of these findings is high, directly informing platform recommender systems, feature strategy, and marketing efforts, while also contributing to the academic case for session-based user modelling.

Keywords: Clustering, K-Means, Session-based Analysis, Music Listening Behaviour, User Segmentation, Personalization, Recommender Systems.

1. Introduction

Music streaming platforms generate an enormous volume of interaction data that is highly valuable to both industry and academia. For platforms, analysing these behavioural traces enables the design of more accurate recommendation systems, improved user experience, and enhanced retention. For scholars, it yields deeper insights into digital consumption patterns, with applications even extending to psychological well-being, such as pain reduction and emotional regulation in healthcare (Richard-Lalonde et al., 2020).

Existing research on music listening is generally categorized into three areas: content-based analysis (e.g., audio and genre characteristics, such as Billboard chart studies by (Pooransingh & Dhoray, 2021)), while machine learning approaches have utilized attention-based neural networks for weakly labelled audio tagging (Kong et al., 2019). behavioural segmentation (clustering users based on long-term preferences and demographics, as seen in (Zhang & Chang, 2022)). While these studies are useful, they commonly rely on aggregate-level statistics, a limitation that causes them to overlook the short-term and contextual variations in listening behaviour.

The notion of a session—a sequence of listening interactions separated by short inactivity intervals—provides a promising alternative perspective. Session-based modelling has been widely adopted in recommender systems to capture short-term user intent (Bonnin & Jannach, 2014; Hidasi et al., 2016; Liu et al., 2018; Quadrana et al., 2018).

However, most of these studies concentrate primarily on improving algorithmic prediction accuracy, with little attention paid to leveraging session structures for behavioural segmentation and interpretability. Consequently, our understanding of how sequential interactions aggregate into distinct listener profiles remains limited.

To address this gap, we introduce a session-based clustering framework for music listening behaviour. We segment raw playback histories into session units, enrich them with behavioural and temporal features, and apply the K-Means algorithm for clustering. Our research pursues three specific objectives:

1. To successfully establishes listening sessions as meaningful analytical units derived from raw log data.
2. To extract and engineer features that effectively characterize listening patterns at the session level.
3. To segment users into interpretable listener profiles that generate actionable insights for personalization and platform design.

Ultimately, the contributions of this study include: (1) demonstrating that session-based clustering captures contextualized listening patterns that aggregate analyses miss, (2) providing a replicable framework for turning raw data into usable user profiles, and (3) advancing the academic discussion on context-aware personalization in digital media.

2. Literature Review

2.1. Music Listener Segmentation

Music listener segmentation typically employs unsupervised methods like clustering to identify user groups. Prior studies, such as (Zhang & Chang, 2022), which clustered users by frequency and genre, and (Barata & Coelho, 2021), which examined retention drivers, provided useful broad categorizations. However, by relying on static, aggregate data, these approaches fail to capture the contextual and temporal variations that characterize listening behaviour within short interaction periods.

2.2. *Session-Based Recommendation and Behaviour Modelling*

Session-based approaches have become highly favoured in recommender systems because of their ability to capture short-term user intent. Research confirms the superiority of this dynamic modelling; for example, (Ludewig et al., 2021) demonstrated that session-aware algorithms significantly outperform static methods that rely only on aggregated historical data. In this domain, a session is typically defined as consecutive interactions separated by a set inactivity gap (e.g., 30 minutes), a definition commonly used across key studies (Bonnin & Jannach, 2014; Hidasi et al., 2016; Liu et al., 2018; Quadrana et al., 2018).

Furthermore, recent work is pushing sessions beyond mere prediction accuracy: (Tran et al., 2024) showed how to incorporate recurrent behaviours into session modelling, while (Mo & Wang, 2025)improved robustness by balancing item popularity with personalized preferences. These advancements underscore the dual value of session analysis: boosting accuracy *and* revealing behavioural structures essential for personalization.

2.3. *Feature and Dimensionality Challenges*

Playback logs are high-dimensional and noisy. Raw datasets include timestamps, track IDs, user IDs, and playback durations, which are not directly interpretable for behavioural analysis. Feature engineering—such as extracting skip behaviour, session duration, and temporal usage patterns—is therefore necessary to transform low-level logs into meaningful variables. (Ramírez-Gallego et al., 2017) emphasized the importance of preprocessing and dimensionality reduction in improving the robustness of clustering outcomes. While prior studies often employ advanced dimensionality techniques like PCA, such methods can make results less interpretable for applied behavioural analysis.

(Alfaifi, 2024) reviewed recommender systems comprehensively, identifying feature extraction and explainability as ongoing challenges. This supports the need for interpretable session-level features, as employed in this study, to ensure that segmentation outcomes are both statistically valid and practically meaningful.

2.4. *Behavior-Informed Personalization*

Combining behavioural segmentation with personalization strategies improves both recommendation quality and user satisfaction. (Lu & Wu, 2025) demonstrated that integrating user interaction patterns with audio content features enhances personalization accuracy. Similarly, (Pooransingh & Dhoray, 2021) showed that genre similarities can be exploited for recommendation purposes. However, very few studies have directly connected session-level behaviour with practical personalization strategies, such as adaptive playlist design or marketing interventions.

2.5. *Research Gap*

Two major gaps emerge from the reviewed literature:

1. Most segmentation studies are aggregate-based, clustering users through long-term historical data while missing the contextual information embedded in short listening sessions.
2. Few works incorporate session-based behavioural features for segmentation. Existing session-oriented studies primarily focus on predictive modelling rather than identifying interpretable clusters of user behaviour.

This study fills these gaps by proposing a session-based segmentation framework that converts playback data into session-level features and applies K-Means clustering. Unlike prior work, this approach emphasizes interpretability and contributes both theoretically to recommender system literature and practically to platform design and personalization strategies.

3. Method

This study adopts a quantitative approach with unsupervised learning to identify music listening behaviour profiles on a digital platform. Prior research has shown that user interaction patterns can be analysed from various data sources (Chung et al., 2022). Session-based recommendation studies demonstrate their ability to capture short-term intent and contextual preferences (Ludewig et al., 2021), supporting analytical approach used in this work.

3.1. Data Collection

The dataset was obtained from the "**Top Spotify Songs in Countries**" dataset on Kaggle (Shaw, 2023), comprising 149,000 entries containing track metadata, playback behaviour, and user interaction data. The main features used include playback duration (`ms_played`), track and artist identifiers (`track_name` and `artist_name`), and skip behaviour (`skipped`). This dataset was selected due to its richness in behavioural traces and its coverage across multiple regions, making it suitable for cross-contextual behavioural segmentation.

3.2. Data Preprocessing

In preparation for session-based feature engineering and adhering to best practices in data analysis (Ramírez-Gallego et al., 2017), the raw dataset required several essential preprocessing steps to maximize its reliability. Specifically, we performed four key actions: standardizing timestamps for **temporal feature extraction**, filling missing data with the **median** (a robust method against outliers), ordering all records sequentially to maintain **session continuity**, and translating *skip* action into a **binary numerical format** for quantitative modelling. This sequence of steps confirmed the data's integrity and consistency.

3.3. Session Definition and Feature Engineering

A session is defined as a sequence of consecutive track plays with less than a 30-minute gap between plays (Bonnin & Jannach, 2014; Hidasi et al., 2016; Liu et al., 2018; Quadrana et al., 2018).

From 8,881 sessions, nine features were extracted:

- **Session duration** : total, mean, and standard deviation of playback time (`ms_played_sum`, `ms_played_mean`, `ms_played_std`)
- **Skip behaviour** : total and average number of skipped tracks (`skipped_sum`, `skipped_mean`).
- **Temporal patterns** : mode of listening hour, day of week, and month (`mode_hour`, `mode_day`, `mode_month`), and
- **Unique track count** : count of unique tracks per session (`track_name_count`).

3.4. Feature Scaling and Clustering

Pearson correlation analysis was performed to identify relationships between features (Schober et al., 2018) (Figure 1) revealed:

- a strong positive correlation between `ms_played_sum` and `track_name_count` (0.71).
- a moderate negative correlation between `ms_played_mean` and `skipped_mean` (-0.22).
- low correlations among temporal features, suggesting independence from behavioural metrics.

The MinMaxScaler standardized feature ranges to the [0, 1] interval. The optimal number of clusters was determined using the Elbow Method(Kodinariya & Makwana, 2013) and validated with the KneeLocator algorithm (Satopaa et al., 2011). Both methods indicated an optimal K = 4. The K-Means algorithm was applied with `n_clusters = 4`, `n_init = 10`, and `random_state = 42`.

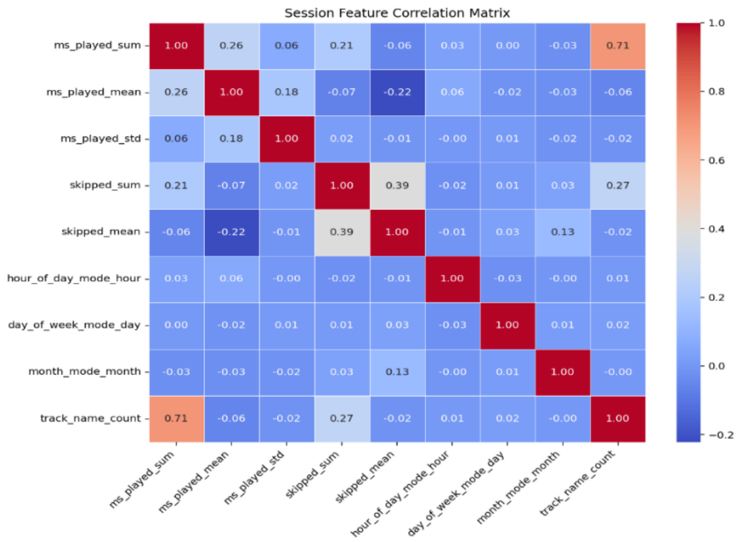


Figure 1. Correlation matrix of engineered features.

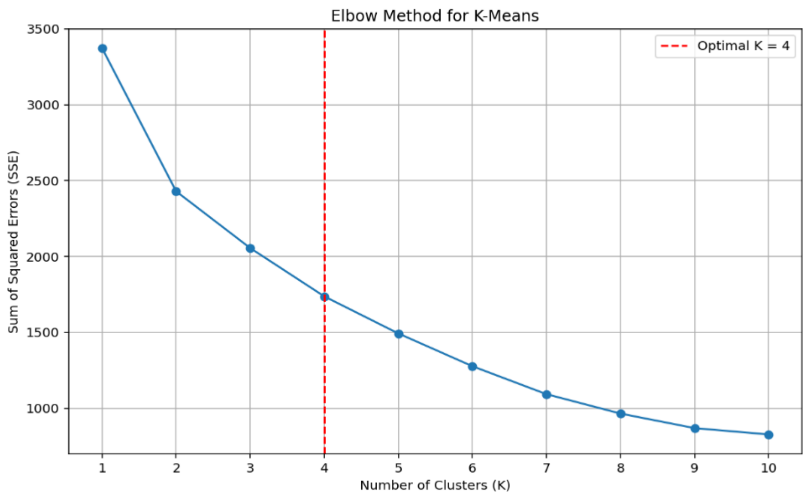


Figure 2. : **Elbow Method curve for determining optimal cluster count.**

3.5. Validation and Interpretation

Cluster validation was carried out using two approaches:

1. **Statistical testing** – An **ANOVA** was conducted to assess whether mean differences across clusters were statistically significant for each feature.
2. **Interpretive labelling** – Cluster labels were assigned based on dominant behavioural and temporal patterns (e.g., *Late-Night Enthusiast, Daily Routine User*).

This two-step process ensured that clusters were not only statistically valid but also semantically interpretable, aligning with the study's goal of producing actionable insights for personalization and platform design.

4. Results

4.1. *Cluster Distribution*

The K-Means algorithm grouped the listening sessions into four distinct clusters. These profiles differed in session duration, skip rate, and temporal preferences, confirming that session-based segmentation yields interpretable behavioural insights. Following the application of K-Means clustering, listening sessions were grouped into four distinct clusters. The distribution of sessions per cluster is shown in Table 1, with its visualization in Figure 3.

Table 1. Distribution of 8,881 listening sessions across four clusters.

Cluster Profile Name	Number of Sessions	Proportion (%)
Daily Routine User	2385	26.86
Late-Night Music Enthusiast	2632	29.64
Long-Duration Focus Session	2041	22.98
Relaxed Nighttime Listener	1823	20.53

The “Late-Night Music Enthusiast” profile dominates (29.64%), followed by “Daily Routine User” (26.86%) and “Long-Duration Focus Session” (22.98%). “Relaxed Nighttime Listener” is least frequent (20.53%).

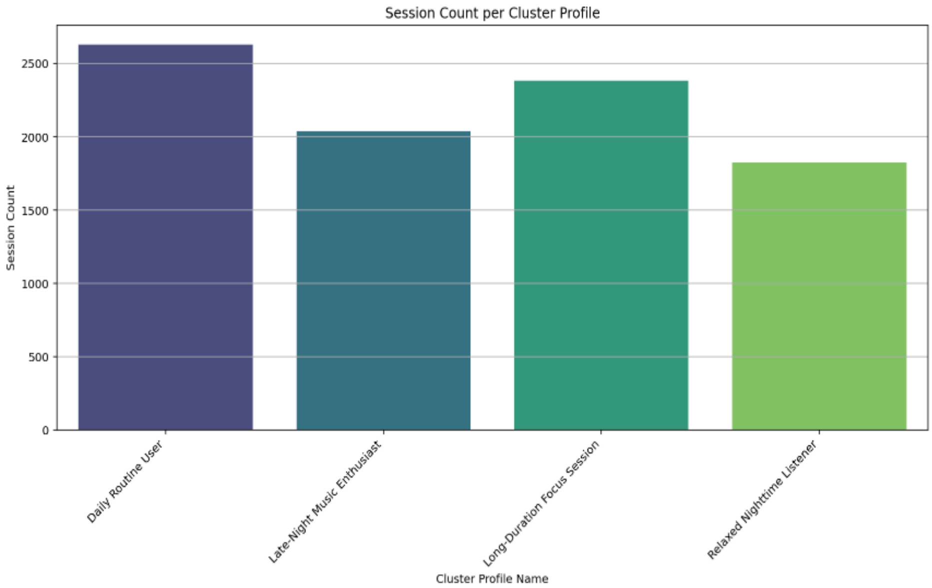


Figure 3. Distribution of session duration across the four listener clusters.

4.2. *Cluster Profile Analysis*

Table 2 presents the mean, median, and mode of key aggregated features for each cluster. Descriptive names were assigned based on feature interpretation and content preference analysis.

Table 2. Behavioural and temporal characteristics of the four listener clusters.

Feature	Cluster 0	Cluster 1	Cluster 2	Cluster 3
ms played sum (ms)	2200300	2000000	2482506	1904300
skipped mean	0.0193	0.0175	0.0164	0.0185
track name count	17.7104	15.7304	16.1579	18.1324
mode hour (0–23)	0	0	20	20
mode day (0–6)	4	1	2	2
mode month (1–12)	9	8	1	11

As summarized in Table 2, the four clusters reveal distinct behavioural patterns. **Daily Routine User** (Cluster 0) engage in sessions of around 37 minutes with low skip rates, often mixing modern rock and classics. **Late-Night Enthusiasts** (Cluster 1) listen for about 33 minutes with minimal skipping, mostly during late-night or early-morning hours. **Long-Duration Focus Sessions** (Cluster 2) feature the longest average duration (≈41 minutes), with high track counts, low skip rates, and a preference for film scores and folk/rock. Meanwhile, **Relaxed Nighttime Listeners** (Cluster 3) show shorter sessions, moderate skip rates, and a tendency towards complex or ambient tracks.

4.3. *Statistical Analysis of Cluster Differences*

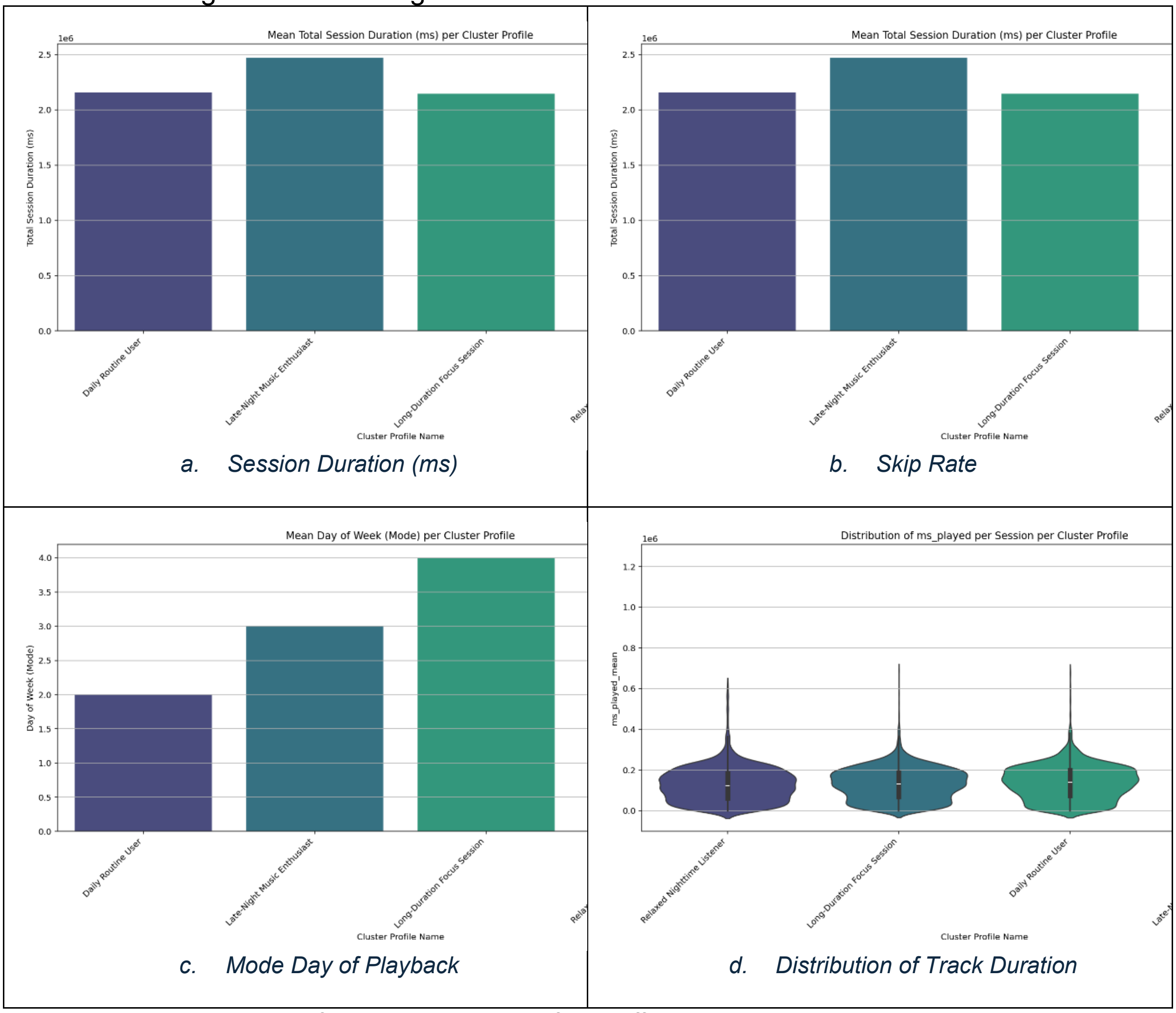
The results of the one-way ANOVA confirm significant differences across clusters for most behavioural features. **Session duration** differed significantly ($F = 9.57$, $p < 0.001$), with the Long-Duration Focus Session group recording longer sessions compared to others. **Skip rate** also varied substantially ($F = 40.62$, $p < 0.001$), where the Long-Duration Focus Session cluster showed the lowest skipping behaviour. In terms of **average track duration**, differences were again significant ($F = 24.64$, $p < 0.001$), with both the Long-Duration Focus Session and Daily Routine User clusters listening to longer tracks. By contrast, **track count** did not show statistically significant differences across groups ($F = 2.03$, $p = 0.108$), suggesting that diversity of tracks played per session is relatively similar among clusters.

4.4. *Analysis of Track Characteristics per Cluster*

The analysis of representative tracks highlights the distinctive musical preferences of each cluster. **Daily Routine Users** frequently listen to contemporary and classic rock, with popular tracks such as *Married with Children* (Oasis) and *You Sexy Thing* (Hot Chocolate). **Late-Night Music Enthusiasts** favour alternative rock and folk-inspired songs, including *Dying Breed* (The Killers) and *Reminder* (Mumford & Sons), which align with their nocturnal listening habits. **Long-Duration Focus Session** listeners gravitate towards immersive and reflective tracks, such as *Crucify Your Mind* (Rodríguez) and *Caution* (The Killers), suggesting a preference for concentrated engagement. Finally, **Relaxed Nighttime Listeners** lean towards ambient and melodic content, with tracks like *Concerning Hobbits* (Howard Shore) and *I Will* (The Beatles), indicating a tendency for relaxed evening listening. Overall, the profiles differ not only in session duration and skip behaviour but also in musical content, underscoring the interpretive labelling of clusters.

4.5. Visualizations of Cluster Characteristics

Figures 4a–4d show clear differences in session duration and skip rate across clusters, indicating distinct listening behaviours.



a. Session Duration (ms) *b. Skip Rate* *c. Mode Day of Playback* *d. Distribution of Track Duration*

Figure 4. Key characteristics of clusters showing significant differences in session duration, skip rate, playback day, and track duration distribution. Variables include skip rates, playback duration, and dominant temporal patterns (hour, day, month). Significant differences ($p < 0.001$) were confirmed by ANOVA tests.

5. Discussion

5.1. Comparison with Prior Studies

This research advances previous listener segmentation studies (Barata & Coelho, 2021; Zhang & Chang, 2022) by implementing a session-based behavioural modelling approach instead of relying solely on aggregated data. Our findings confirm that incorporating temporal and interaction features yields more nuanced listener profiles. While recent work focuses on algorithmic refinements like handling repeat patterns (Tran et al., 2024) or stratifying content popularity (Mo & Wang, 2025), our study intentionally prioritizes **interpretability**. By using clustering to uncover distinct, explainable segments, we contribute to the vital academic discussion on balancing predictive accuracy with model **explainability** in recommender systems.

5.2. Practical Implications

The identified listener profiles provide **actionable insights** for digital platforms. Personalization can be refined for each segment—for example, by creating uninterrupted playlists for focused users or diverse, short playlists for daily routine listeners. Platforms can also adjust their interface (e.g., implementing a late-night theme) to align with specific temporal usage patterns. Furthermore, these clear segments enable more effective and targeted marketing strategies.

In line with the call from (Alfaifi, 2024) for real-world recommender systems to prioritize **user trust**, this research enhances transparency. By labelling session-based clusters in plain behavioural terms (e.g., "Daily Routine User," "Late-Night Enthusiast"), we contribute to greater acceptance of personalization strategies among users.

5.3. Academic Contributions

This work demonstrates that relatively **simple algorithms** like **K-Means** can produce insights comparable in value to those from more complex neural network models. By focusing on session-level interpretability, our study complements the latest algorithmic advances (Mo & Wang, 2025; Tran et al., 2024) and directly supports the necessity of developing **explainable, user-centric** recommender systems, as advocated by (Alfaifi, 2024).

5.4. Limitations and Future Work

The study acknowledges several limitations: The dataset is limited to Spotify and might not generalize to other platforms or cultural settings. Additionally, only K-Means clustering was applied; future work should explore alternative clustering techniques (e.g., DBSCAN or hierarchical methods) to identify other structural patterns. Future research can also strengthen the model by incorporating content-based audio features or advanced deep learning architectures (like recurrent or transformer models) to capture sequential behaviour more effectively.

6. Conclusion

In conclusion, this research successfully segmented music listening behaviour into four distinct profiles using session-based K-Means clustering. These profiles revealed significant variations in session length, skipping behaviour, and temporal usage patterns. The findings bridge the gap between aggregate-level analysis and session-based modelling, providing both practical value for personalized service design and an academic contribution to contextualized recommendation research.

References

Alfaifi, Y. H. (2024). Recommender Systems Applications: Data Sources, Features, and Challenges. *Information*, *15*(10). https://doi.org/10.3390/info15100660

Barata, M. L., & Coelho, P. S. (2021). Music streaming services: Understanding the drivers of customer purchase and intention to recommend. *Heliyon*, *7*(8). https://doi.org/10.1016/j.heliyon.2021.e07783

Bonnin, G., & Jannach, D. (2014). Automated Generation of Music Playlists: Survey and Experiments. *Proceedings of the 15th International Society for Music Information Retrieval Conference (ISMIR 2014)*, 627–632. https://ismir2014.ismir.net/proceedings/ismir2014_submission_128.pdf

Chung, J., Lee, J., & Yoon, J. (2022). Understanding music streaming services via text mining of online customer reviews. *Electronic Commerce Research and Applications*, *53*, 101145. https://doi.org/10.1016/j.elerap.2022.101145

Hidasi, B., Karatzoglou, A., Baltrunas, L., & Tikk, D. (2016). Session-based Recommendations with Recurrent Neural Networks. *4th International Conference on Learning Representations (ICLR)*. https://arxiv.org/abs/1511.06939

Kodinariya, T. M., & Makwana, P. R. (2013). Review on determining number of Cluster in K-Means Clustering. *International Journal of Advance Research in Computer Science and Management Studies*, *1*(6), 90–95.

Kong, Q., Yu, C., Xu, Y., Iqbal, T., Wang, W., & Plumbley, M. (2019). Weakly labelled AudioSet tagging with attention neural networks. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, *27*(11), 1791–1802. https://doi.org/10.1109/TASLP.2019.2930913

Liu, Q., Zeng, Y., Mokhosi, R., & Zhang, H. (2018). STAMP: Short-Term Attention/Memory Priority Model for Session-based Recommendation. *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 1831–1839. https://doi.org/10.1145/3219819.3220026

Lu, J., & Wu, M. (2025). Design and application of a music recommendation system based on user behavior and feature recognition. *Systems and Soft Computing*, 200274. https://doi.org/10.1016/j.sasc.2025.200274

Ludewig, M., Mauro, N., Latifi, S., & Jannach, D. (2021). Empirical analysis of session-based recommendation algorithms: M. Ludewig et al. *User Modeling and User-Adapted Interaction*, *31*(1), 149–181. https://doi.org/10.1007/s11257-020-09277-1

Mo, Y., & Wang, H. (2025). Session-Based Recommendation Method Using Popularity-Stratified Preference Modeling. *Mathematics*, *13*(6), 960.

Pooransingh, A., & Dhoray, D. (2021). Similarity analysis of modern genre music based on billboard hits. *IEEE Access*, *9*, 144916–144926. https://doi.org/10.1109/ACCESS.2021.3122386

Quadrana, M., Cremonesi, P., & Jannach, D. (2018). Sequence-Aware Recommender Systems. *ACM Computing Surveys*, *51*(4), 1–36. https://doi.org/10.1145/3190616

Ramírez-Gallego, S., Krawczyk, B., García, S., Woźniak, M., & Herrera, F. (2017). A survey on data preprocessing for data stream mining: Current status and future directions. *Neurocomputing*, *239*, 39–57. https://doi.org/10.1016/j.neucom.2017.01.078

Richard-Lalonde, M., Gelinas, C., Boitor, M., Gosselin, E., Feeley, N., Cossette, S., & Chlan, L. L. (2020). The effect of music on pain in the adult intensive care unit: a systematic review of randomized controlled trials. *Journal of Pain and Symptom Management*, *59*(6), 1304–1319. https://doi.org/10.1016/j.jpainsymman.2019.12.359

Satopaa, V., Albrecht, J., Irwin, D., & Raghavan, B. (2011). Finding a “Kneedle” in a Haystack: Detecting Knee Points in System Behavior. *2011 31st International Conference on Distributed Computing Systems Workshops*, 166–171.

Schober, P., Boer, C., & Schwarte, L. A. (2018). Correlation coefficients: appropriate use and interpretation. *Anesthesia & Analgesia*, *126*(5), 1763–1768.

Shaw, A. (2023). *Top Spotify Songs in Countries*. https://www.kaggle.com/datasets/anandshaw2001/top-spotify-songs-in-countries

Tran, V.-A., Salha-Galvan, G., Sguerra, B., & Hennequin, R. (2024). Transformers meet ACT-R: repeat-aware and sequential listening session recommendation. *Proceedings of the 18th ACM Conference on Recommender Systems*, 486–496.

Zhang, Z., & Chang, J. (2022). Clustering-based categorization of music users through unsupervised learning. *Economics & Management Information*, 1–8. https://doi.org/10.58195/emi.2022.1006