Understanding Spotify's Recommendation Radio System: A Content and User-Based Analysis.

Link to GitHub repository: https://github.com/n1cffz/personalisation-2024.

The primary aim of the project is to investigate the recommendation algorithms used by Spotify; particularly how personal listening history influences the generated recommendations. This research focuses on analysing the playlists created for the song 'Waiting for You' by Majid Jordan featuring Naomi Sharon. By applying core information retrieval concepts like indexing and similarity measures, this project looks at these systems at macro (generalized song radio playlists) and micro (individual recommendation) levels.

Drawing inspiration from Arun et al.'s (2023) mood-based recommendation system, which combined collaborative and content-based filtering, the latter section in this essay proposes a prototype to enhance the user experience by providing greater control over recommended tracks. Ultimately, the goal is to contribute to explainable artificial intelligence (XAI) models, promoting transparency in recommendation systems.

The research questions include:

1. How does Spotify's recommendation system use personal listening history?
2. What biases or limitations exist in the recommendation algorithm?
3. How do audio features and song popularity influence Spotify's recommendations?

Methodology

Data from Spotify's API, including playlists, top tracks and audio features were extracted. These audio features (danceability, energy, valence, tempo, and loudness) were standardised, and cosine similarity was applied to determine track similarity to 'Waiting For You'. Cosine similarity has been validated by previous research, for instance, Qian et al. (2004) whose approach aligns with the principle that cosine similarity better captures directional similarity in high-dimensional vectors compared to distance-based measures (Ghosh et al., 2006). These techniques enabled a comparative analysis between recommendations for a user account with an established listening history and a 'ghost' account without. Specific statistical techniques (t-test) and machine learning models were used to identify patterns and biases in the system.

Findings

Playlists extracted via Spotify's API for the 'Waiting For You' recommended radio consisted of user curated and personalized Spotify playlists (see **Figure 1**). The song in question also appeared in its own recommended radio on multiple occasions, with a high similarity score (0. 999093) and was categorised under four distinct genres: 'Dutch R&B', 'Canadian hip hop' and 'Canadian contemporary' and 'R&B'.

```
Waiting For You (feat. Naomi Sharon) Radio
Waiting For Tonight Radio
Waiting For (feat. Jamila Woods) Radio
Best of Majid Jordan
Waiting For Never Radio
Waiting for you 💖✈
Waiting for Your Love Radio
Looking out for you vibes
Majid Razavi Mix
Naomi Sharon Radio
```

Figure 1. List of Spotify and user-curated playlists featuring 'Waiting For You'.

In an attempt to understand how my listening activities influence Spotify's recommendations and whether there is a popularity bias (Turnbull et al., 2022) toward charting songs, I reverse-engineered the recommendation system using the API for the analysis of playlist data. (see accompanying Python code). The results from the distribution of the top 10 recommended radio tracks suggested a possible popularity bias favouring mainstream music (e.g., "On My Mama" by Victoria Monet had a recommendation score of 0.817540 and a popularity score of 0.702128). Genres such as R&B, urban contemporary, and alternative R&B exhibited the highest frequency (see **Figure 3**) within these top 10 recommended tracks. This observation raises questions about the extent to which genre is incorporated into Spotify's content-based similarity approach for recommendations, as audio features and popularity appear to be more heavily weighted factors.

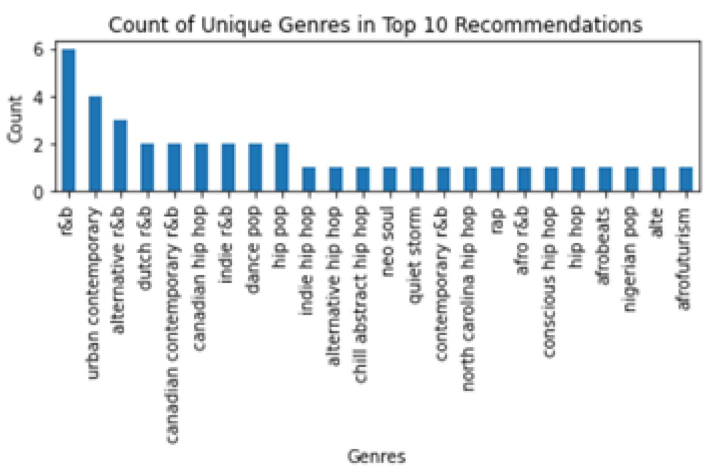


Figure 2. Frequency of unique genres within the top 10 recommended tracks.

In the sample of 100 tracks from my top tracks for this year , R&B also scored the highest frequency (see **Figure 3**), which negates Maheshwari's (2023) assertion that audio analysis within these systems enables explorative recommendations beyond a user's comfort zone.

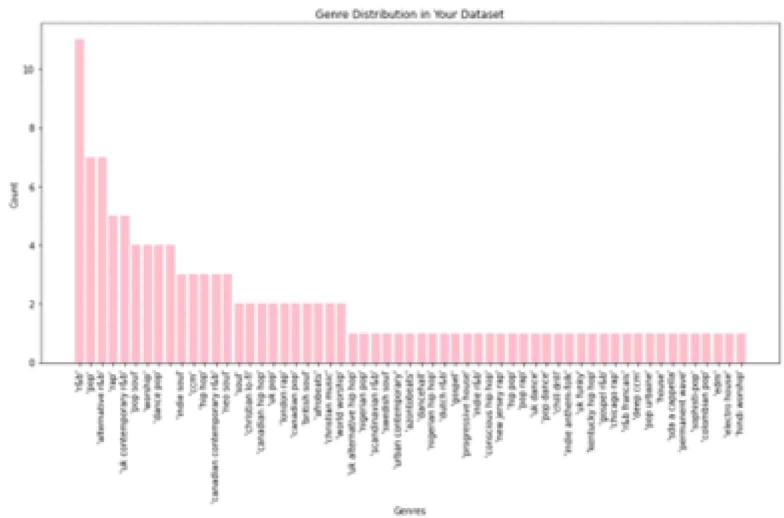


Figure 3. Distribution of genres in the personal dataset based on the count of each genre from my top artists.

Genre classification can also be subjective, not confined to a singular label, which may explain major differences observed such as the recommendation of "Free Fall (feat. J. Cole)," whose genres (alte, Nigerian pop, afrobeat, and hip hop) differ substantially from the initial song. Agrawal et al. (2023) emphasize the importance of genre labels in recommendation systems, particularly for movies and customer expectations. They propose the concept of "micro-genres" (2023:76), which could aid in grouping similar songs more effectively, thereby enhancing personalization for each radio station.

Further studies, for example, Sah et al. (2024:2092) assert the capabilities of large language models (LLMs) to tailor recommendations which reduce instances of bias and resonate with the user's unique tastes and characteristics. Spotify holds a diverse global user base of 422 million monthly active users (Gallant, 2022), primarily young adults in Europe, which fits my demographic. While access is limited to personal data obtained through the Spotify Web API, rendering the results non-representative and non-replicable, these exploratory analyses focused on song-song similarity can still provide valuable insights. This personal user data was transformed into a CSV file ('user_item_data.csv') from which audio values features were extracted for song-song similarity.

Considering the Five Factor Model (Dhelim et al., 2023), which suggests that personality-aware recommendation systems could explain song recommendations based on the connection between personality traits and musical preferences, Moscato et al. (2020) fail to address the challenges of cold-start accounts, a significant limitation in personalization. The results for the 'ghost' account (no previous listening history) show notable differences compared to the personal account, visually (see **Figures 4** and **5**).

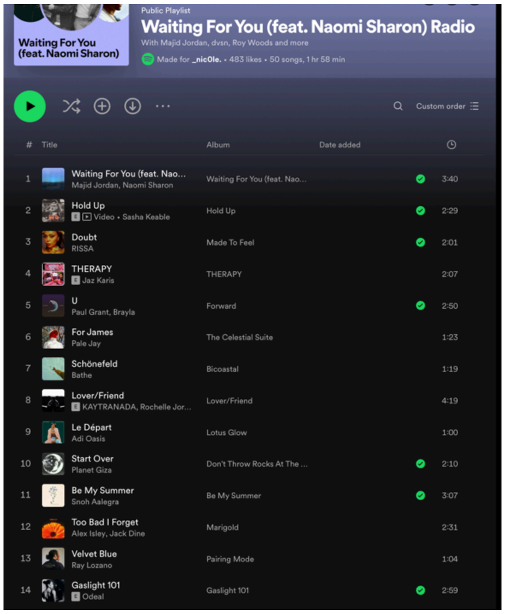


Figure 4. Current recommended radio for the track 'Waiting For You', features already saved songs on my personal Spotify account.

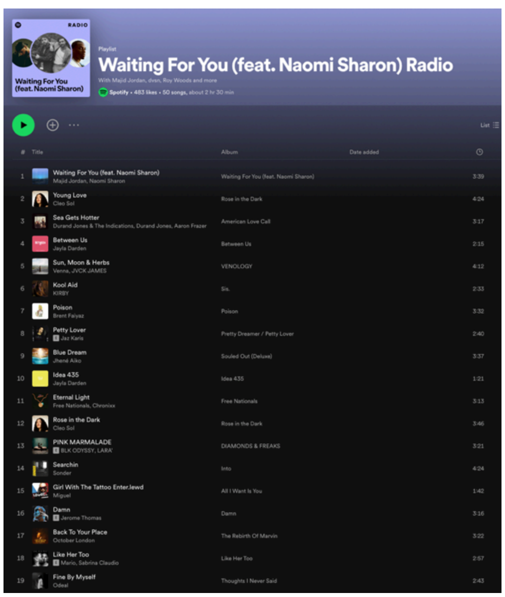


Figure 5. Song radio for the 'ghost' Spotify account.

A statistical analysis was conducted to determine if there are significant differences in the means of audio features between the "ghost" account (data from 'ghost_playlist_audio_features.csv') and the personal account (data from 'updated_track_data.csv'). The results of the T-tests for the selected audio features show that the audio features 'energy', and 'loudness' had significant differences (<0.05) in the mean values between the two accounts (see **Figure 6**). Thus, suggesting that listening history and persona building in tandem, play crucial roles in the influence of song recommendations for Spotify’s radio algorithm.

```
Audio Feature Index - Ghost: 0, Personal: 25
T-Statistic: 0.9582254233130664
P-Value: 0.34053460849309514
There is no statistically significant difference in the means of the selected audio feature.

Audio Feature Index - Ghost: 1, Personal: 26
T-Statistic: -4.880138067250321
P-Value: 4.2124970166153365e-06
There is a statistically significant difference in the means of the selected audio feature.

Audio Feature Index - Ghost: 2, Personal: 27
T-Statistic: -1.0978026171527107
P-Value: 0.2749941138129153
There is no statistically significant difference in the means of the selected audio feature.

Audio Feature Index - Ghost: 3, Personal: 28
T-Statistic: -3.505675297919728
P-Value: 0.000688905818359639
There is a statistically significant difference in the means of the selected audio feature.

Audio Feature Index - Ghost: 4, Personal: 29
T-Statistic: -0.6100425384628332
P-Value: 0.5432467234371245
There is no statistically significant difference in the means of the selected audio feature.
...
T-Statistic: 2.7525408567027556
P-Value: 0.0070500564430520584
There is a statistically significant difference in the means of the selected audio feature.
```

Figure 6. Results of Statistical Analysis (T-Test).

Feature importance analysis was also conducted using Principal Component Analysis (PCA) and examining K-means cluster centroids for the generalised playlist audio features by comparison to my personal recommended playlist for ‘Waiting For You’. This proved crucial for understanding key factors influencing track grouping and recommendations. It was determined that five clusters effectively captured differential patterns in the datasets (Kodinariya et al., 2013).

The k-means clustering analysis revealed distinct patterns in the distribution of clusters for the recommended radio playlist on the personal account and the ‘ghost’ account. In the personal account's recommended playlist, cluster 0 tended to have higher values on the right side of the scatter plot, deviating significantly from the other clusters (1, 2, 3, and 4). This contrast was particularly pronounced between clusters 0 and 1, as shown in **Figure 8**. On the other hand, the clustering results for the new account's recommended playlist displayed a tighter distribution, with cluster 0 having only two main outliers skewed towards higher values on the right side, while the central tendency remained more concentrated (see **Figure 7**). The application of such techniques aligns with industry best practices, as evidenced by their widespread adoption in fields like healthcare, where they have proven instrumental in providing transparent insights into the recommendation generation process (Alyousef et al., 2018). The assumption that Spotify’s recommendation system considers audio characteristics is thus supported and showcased as though the clusters vary in distribution, the similar shapes in the k-means are displayed.

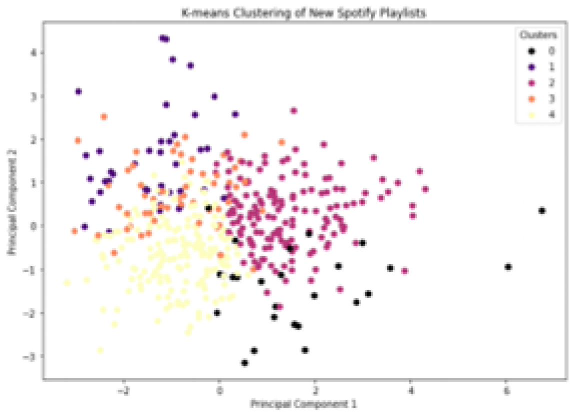


Figure 7. Principal Component Analysis (PCA) Visualization of K-Means Clustering for new Spotify account playlists.

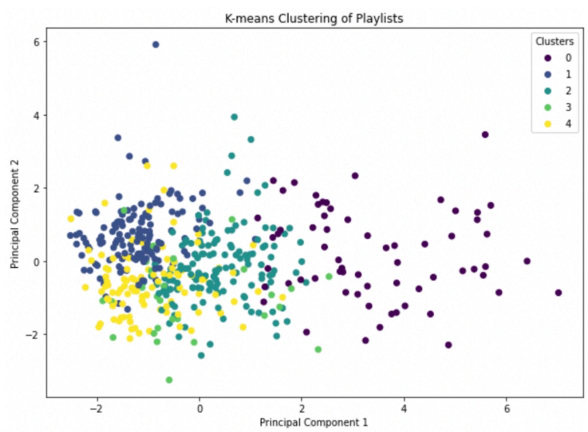


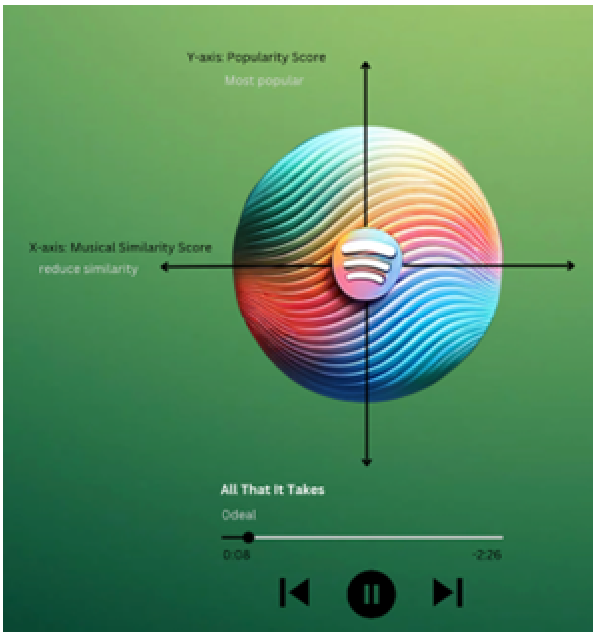
Figure 8. Principal Component Analysis (PCA) Visualization of K-Means Clustering for Personalized Spotify Song Radio.

Audio features like energy, danceability, and valence, and popularity metrics and collaborative filtering signals like total plays and user listening history, significantly influenced cluster formation, track grouping, and recommendations. These findings align with observed potential popularity bias and the influence of audio features, corroborating the need for incorporating personality traits. These results, however, should be interpreted with the understanding that the analysis was based on a limited dataset due to the request rate limitations imposed by the Spotify API (the data collection was restricted to fewer than 600 data points at any given time). While this sample size may not provide a comprehensive representation of the entire population, it nevertheless yielded some intriguing insights.

<u>Proposed Interactive Music Discovery Interface</u>

The proposed UX enhancement for Spotify's radio recommendation system aims to address potential biases through user feedback and interactivity at a micro, content-based level. The gamified concept features a circular XY chart displaying the Spotify logo (see **Figure 9**)., which users can manipulate along the x-axis to control the musical similarity of recommended songs. The x-axis is calculated using the cosine similarity between the audio features (e.g., danceability, energy, valence, tempo) of the current and candidate songs, ranging from -1 (very low similarity) to 1 (very high similarity). While the y-axis (0 to 1) represents the recommendation's relevance or popularity based on factors like overall popularity, listening history, or collaborative filtering signals.

Consequently, this approach facilitates dynamic, personalized music discovery by providing real-time feedback on the next queued song based on the user's adjustments, fostering engagement and tailoring the exploration process to individual preferences.



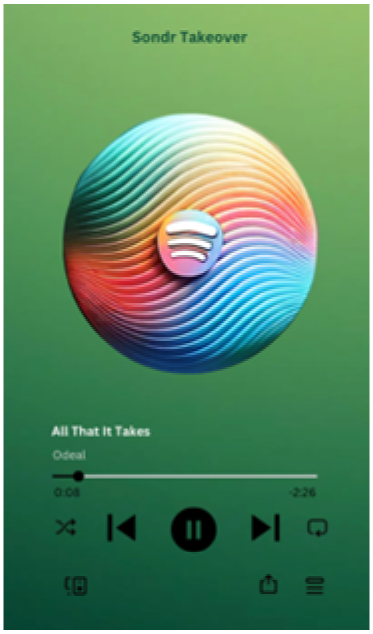


Figure 9. Interactive Music Discovery Interface Concept Mock-up: 'Sondr Takeover'.

To critique, the interface, the reliance on pre-computed audio similarities presents notable challenges regarding computational resources and scalability. Constantly updating audio in real-time to accommodate evolving user playlists could be computationally and temporally expensive. In addition, calculating audio similarity scores for every possible playlist state becomes infeasible for extensive music catalogues.

To address this issue, the use of Locality Sensitive Hashing (LDAC) emerges as a promising solution, leveraging the difference of convolution property of audio features to enable efficient similarity detection even in high-dimensional music feature spaces (Rabbani et al., 2023). This thus ensures a real-time experience for users exploring songs like their current audio context while considering the emotional flow of user-curated playlists could enhance recommendations, an area Spotify has yet to perfect. The prototype critically examines the potential for improving track matching based on musicality and incorporates real-time feedback, allowing for dynamic adjustments to the playlist queue based on individual preferences.

Conclusion

Overall, the employed exploratory methods and song-song similarity approach applied to identify the closest neighbours for "Waiting For You" revealed Spotify's reliance on personal listening history and collaborative filtering signals such as the total number of plays and user listening history, to generate personalised recommendations. Results indicate a significant dependence on previously engaged content which the system falls short by re-recommending featured tracks already in rotation from the user's liked songs or curated playlists. Ultimately, these findings solidify existing knowledge and serve as a contribution to the broader goal of enhancing the transparency and explainability of music recommendation systems.

By delving into statistical differences in audio features, this research also illustrates how significant attributes like energy and loudness shape individual listening experiences and reflect broader patterns within Spotify's recommendation framework. Building on these findings, the proposed gamified prototype for the recommendation radio addresses personalisation through more focused user feedback and interactivity. These insights reinforce the importance of incorporating diverse user data and advanced modelling techniques to create a more inclusive and personalised recommendation system.

Bibliography

Agrawal, S., Trenkle, J., & Kawale, J. (2023). Beyond Labels: Leveraging Deep Learning and LLMs for Content Metadata. *Proceedings of the 17th ACM Conference on Recommender Systems*, 1–1. Available at: https://doi.org/10.1145/3604915.3608883.[Accessed 26 May 2024].

Alyousef, A.A., Nihtyanova, S., Denton, C., Bosoni, P., Bellazzi, R. and Tucker, A., (2018). Nearest consensus clustering classification to identify subclasses and predict disease. *Journal of healthcare informatics research*, *2*, pp.402-422. Available at: https://link.springer.com/article/10.1007/s41666-018-0029-6. [Accessed 18 May 2024].

Arun, A., Soni, M., Choudhary, P. and Arora, S. (2023). *EXPLORE -Explainable Song Recommendation*. Available at: https://arxiv.org/pdf/2401.00353 [Accessed 1 Jun. 2024].

Dhelim, S., Chen, L., Aung, N., Zhang, W. and Ning, H.(2023). A hybrid personality-aware recommendation system based on personality traits and types of models. *Journal of Ambient Intelligence and Humanized Computing*, 14,(9),12775-12788.Available at:
Gallant, D. (2022). *Spotify Target Market Segmentation – User Demographics & Audience Targeting Strategy for 2022 - Start.io - A Sell-side Omnichannel Advertising Platform*. Start.io - A Sell-side Omnichannel Advertising Platform. Available at: https://www.start.io/blog/spotify-target-market-segmentation-user-demographics-audience-targeting-strategy/ [Accessed 31 May 2024].

Ghosh, J. and Strehl, A.(2006.) Similarity-based text clustering: A comparative study. In *Grouping Multidimensional Data: Recent Advances in Clustering*, 73-97.Berlin, Heidelberg: Springer Berlin Heidelberg. Available at: https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=4f6850fcb09925db07a53e54b948b15e4db37031. [Accessed 1 June 2024].

Kodinariya, T.M. and Makwana, P.R. (2013). Review on determining number of Cluster in K-Means Clustering. *International Journal*, 1(6), pp.90-95. Available at: https://www.researchgate.net/publication/313554124 . [Accessed 28 May 2024].

Maheshwari, C. (2023). Music Recommendation on Spotify using Deep Learning. *arXiv. Available at: https://arxiv.org/pdf/2312.10079.* [Accessed 19 May 2024].

Rabbani, T., Bornstein, M. and Huang, F. (2023). Large-Scale Distributed Learning via Private On-Device Locality-Sensitive Hashing. Available at: https://arxiv.org/pdf/2306.02563. [Accessed 28 May 2024].

Sah, C.K., Xiaoli, D.L. and Islam, M.M.(2024). Unveiling bias in fairness evaluations of large language models: A critical literature review of music and movie recommendation systems. *arXiv. Available at: https://arxiv.org/pdf/2401.04057. [Accessed 13 May 2024].*

Spotify.com. (2024). *Web API Reference | Spotify for Developers*. Available at: https://developer.spotify.com/documentation/web-api/reference/get-an-artists-related-artists [Accessed 31 May 2024].

Turnbull, D.R., McQuillan, S., Crabtree, V., Hunter, J. and Zhang, S.(2022). Exploring popularity bias in music recommendation models and commercial steaming services. *arXiv. Available at: https://arxiv.org/pdf/2208.09517.* [Accessed 1 June 2024].