Assignment 2

Fairness in Music Recommender Systems: Mitigating Popularity Bias in Spotify's Algorithm

F. Byrne [0984086], A. Dicks Herrán [1338056], M.L.W. Dielessen [5268125], M. Radbruch [0814709], L. Spedener [6766366]

Utrecht University

Personalisation of Public Media [INFOPPM]

Dr. D. Nguyen

7th April 2023

Word Count: 2525

Universiteit Utrecht

Table of Contents

Introduction

Recommender systems (RSs) provide users with suggestions of what items may be of interest to them, e.g., movies, television shows or songs (Ricci et al., 2022). Music is important in society as a whole, since it has a vast array of uses which can be seen across different cultures (Longhurst, 2007). The distribution of music has changed over the years, and today the most common means is through music streaming services. Spotify is currently the most popular service with over 500 million monthly users (Curry, 2023; Spotify, 2023b). In line with the majority of music streaming services, Spotify makes use of a RS. (Schedl et al., 2018).

There is evidence that there is a streaming inequality across music streaming services, in which popular artists are streamed disproportionately more than other artists (Blake, 2020). This inequality may be due to a popularity bias in music RSs in which lesser-known artists are less likely to be recommended compared to well-known artists (Zhu et al., 2021). This is an issue of fairness in RS design, which is why this public value is chosen as the focus of this paper.

The RSs as provided by music streaming services have multiple stakeholders who have a vested-interest in its workings, such as music listeners, the musical artists and the private organisations who distribute the music. Consequently, the aforementioned streaming inequality affects each of the stakeholders differently. Music listeners may want to be recommended to more lesser-known artists but are not. Private organisations may care about this issue, but do not attempt to mitigate it due to a conflict of interest in which exposing popular artists is more profitable (Amatriain & Basilico, 2015). And lesser-known artists who rely on exposure from such RSs in order to make a living may be negatively affected by this inequality (Schedl et al., 2018).

Consequently, the main objective of the proposed RS is to implement the public value of fairness into Spotify's RS by recommending an equal proportion of lesser-known as better-known artists.

Literature Review

Background

In his article “The Long Tail”*,* Anderson (2004) predicted that the emergence of digital media services would create a more equal marketplace via the provision of on-demand access to millions of songs including niche and independent artists, i.e. the long tail. According to Blake (2020), the opposite has actually happened. Streaming music services have engendered an even steeper long-tail distribution, wherein a small number of popular artists receive most of the total streams. Specifically, they state that the top 10% of artists on streaming platforms take 99.4% of the total streams, rendering the inequality higher on streaming platforms compared to physical album sales. Even though private organisations, such as Spotify, are not public, it can still be argued that these companies should attempt to implement public values (e.g., transparency, diversity, autonomy, and fairness) in their services (Benington, 2011). Spotify acknowledges this and expresses their aim to provide all artists with equal opportunity as part of their mission statement (Spotify, 2023a):

> *“Our mission is to unlock the potential of human creativity—by giving a million creative artists the opportunity to live off their art and billions of fans the opportunity to enjoy and be inspired by it.”*

The aim to provide all artists with equal opportunity to live off their art can be linked to addressing the issue of fairness in the music streaming industry. While Spotify has made some efforts to address this issue (Kopf, 2020) by introducing playlists such as ‘Discover Weekly’ which promote lesser-known artists, these efforts are not enough to mitigate challenges lesser-known artists face on the platform.

Definition of Fairness

The topic of fairness in recommender systems (RSs) has become increasingly researched in recent years (Mehrabi et al., 2021). Nonetheless, the absence of a universally recognized definition of fairness highlights the complexity of the concept. Defining fairness is further complicated by cultural and contextual differences (Saxena, 2019).

Mehrabi et al. (2021) define fairness as “[the] absence of any prejudice or favoritism toward an individual or a group based on their inherent or acquired characteristics”. Hence, an algorithm is considered unfair when its decisions are biased towards a particular group of people.

Definition of Popularity in Streaming Services

Popularity in music streaming services is commonly assessed via streaming metrics, such as the total number of streams of a song. Other metrics are engagement metrics such as, the number of times a song was shared or whether it was added to playlists. Spotify uses its own popularity index, which calculates a popularity score (from 0 to 100) based on streaming metrics (Sciandra & Spera, 2022).

Definition of Popularity Bias

Popularity bias can be defined as when popular items (such as songs from well-known artists) are recommended more than less-popular items (songs from lesser-known artists), even though these items may be of interest to the user (Boratto et al., 2021).

One reason for this could be due to RSs being trained on user preferences, and as popular items are more frequently consumed, there is a non-uniform distribution in relation to the user engagement with items across a platform (Yalcin & Bilge, 2022). Research has shown that state-of-the-art RSs often suffer from this bias leading to a feedback loop that further reinforces the popularity of these items (Turnbull et al., 2022).

Mitigation of popularity bias

Several studies have tried to mitigate popularity-bias through modification of the recommendation algorithm. Miyamoto et al. (2019) created an algorithm which more highly weighted songs that were recommended less frequently to other users but were predicted to be of interest to the user in question. Furthermore, Adamopoulos & Tuzhilin (2014) modified the k-nearest neighbours algorithm by selecting a random sample of items from a large range of neighbours instead of selecting their nearest ones. They did this as more popular items tended to be recommended when they were based only on a small set of a users' nearest neighbours. Both of these techniques showed a reduction of popularity-bias in their results. While popularity bias was reduced in these studies, it could not be eliminated and fairness was not sufficiently achieved.

In this paper, popularity-bias will be mitigated (and thus fairness will be implemented) in Spotify's recommender system by generating recommendations based on a user's preferences with an equal proportion of lesser-known as compared to better-known artists.

Method

Survey

Value-sensitive design

To take into account the values of stakeholders when designing the proposed recommender system (RS), a survey was created and distributed to the target group, users of music streaming services. 86 participants completed the survey, and no demographic information was collected as it was not necessary for the project. The questions asked about their music listening behaviour, their current perceptions of RSs, and the extent to which they valued lesser-known artists being recommended to them (see Appendix A). For example, question 8 asked “Would you like your music streaming service to recommend to you...”. Many respondents (41.67%) stated that they would like to be recommended an equal proportion of lesser-known as compared to better-known artists, and a majority agreed (64.41%) or somewhat agreed (20.34%) that lesser-known artists should have an equal chance of being listened to. These results demonstrated that users of music streaming services value fairness in relation to the treatment of lesser-known as compared to better-known artists by the RS.

Personas

Based on the survey responses, three personas were developed. These were derived from participants’ listening behaviour/frequency, favourite genre, and on how important they considered the value of fairness with respect to lesser-known artists. Persona 1 (the person who likes to listen to lesser-known music) listens to a lot of lesser-known artists and wants to be recommended a lot of lesser-known music. Persona 2 (the person who likes to listen to popular music) listens to a lot of better-known artists and wants to be recommended a lot of better-known artists. Persona 3 (the person who wants to find new music) listens to a lot of better-known artists but wants to be recommended more lesser-known artists. Further information on each persona can be observed in Figure 1, 2 and 3 below.

Figure 1

Persona 1

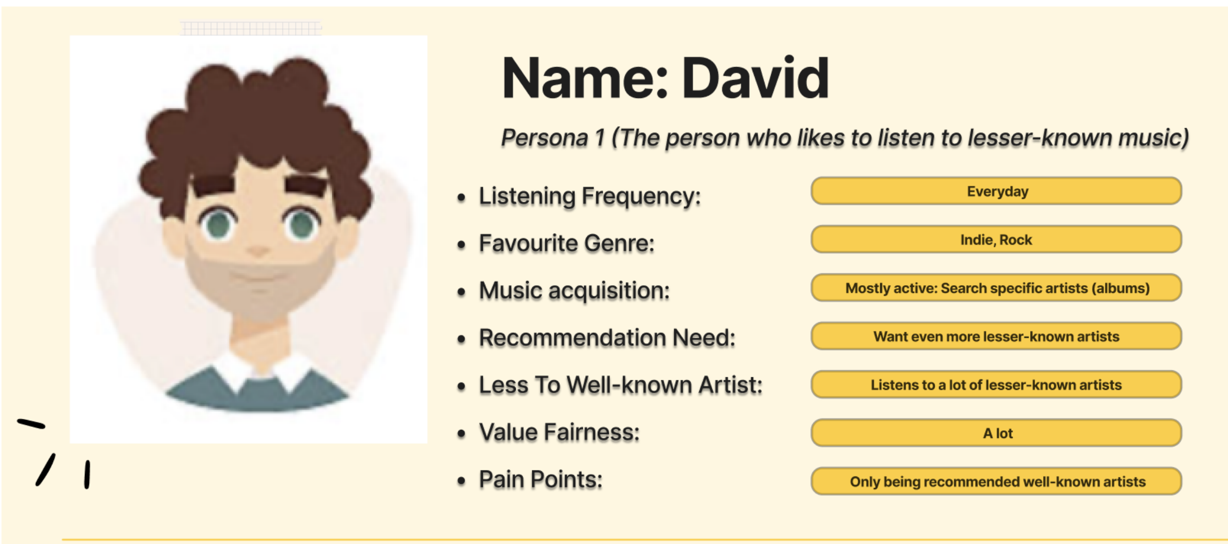


Figure 2

Persona 2

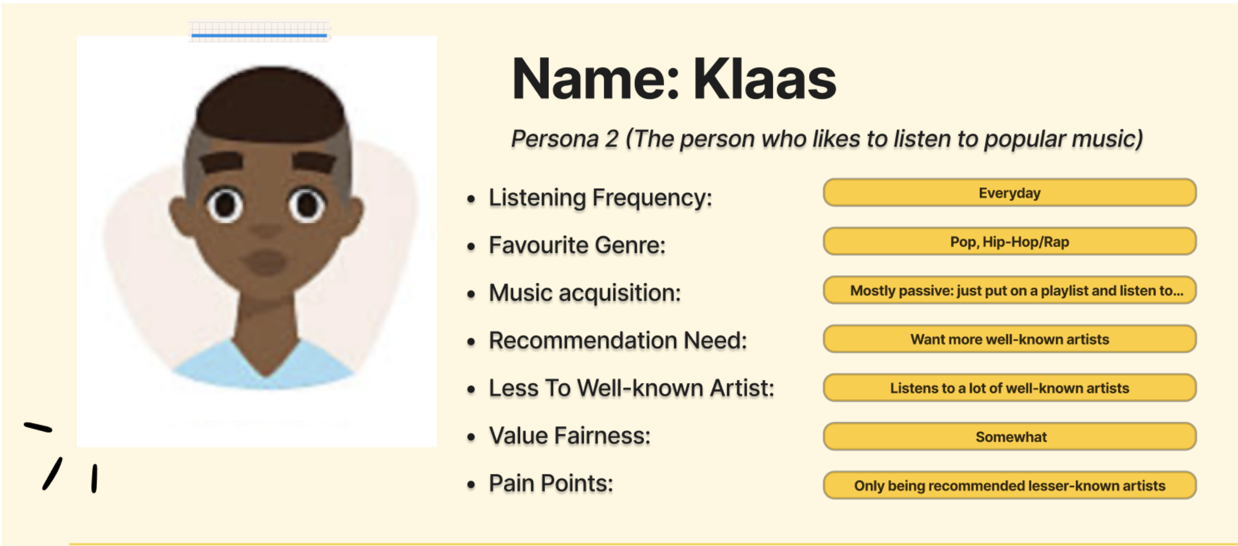
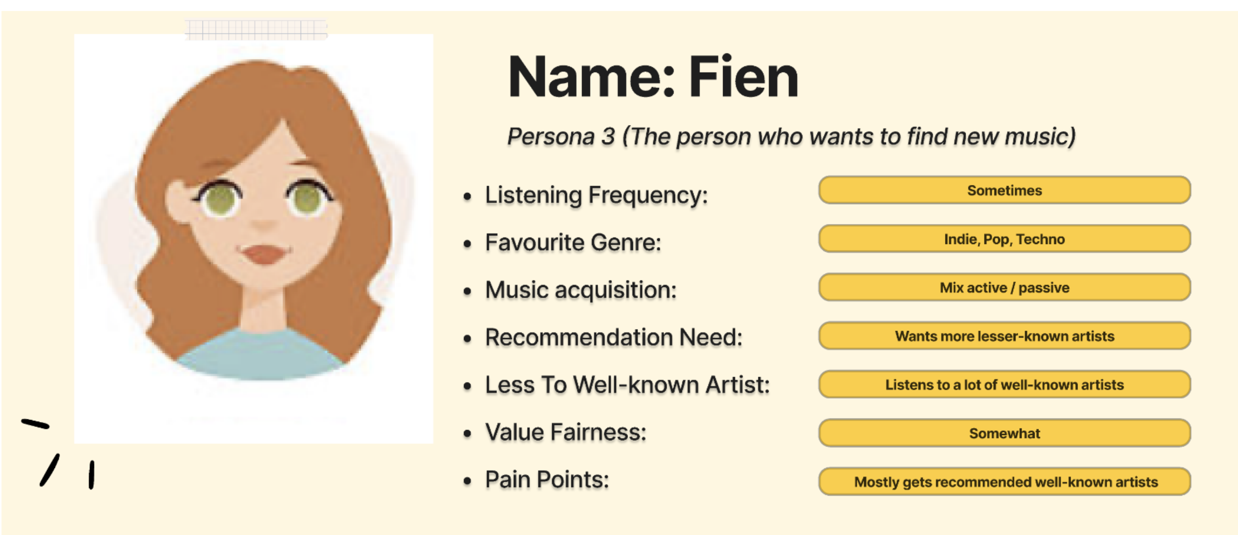


Figure 3

Persona 3



Data collection

Content data

Content data from Spotify was scraped via the Spotify Developer API. This data consisted of artist name, track name, the artist's Spotify popularity index score, and genre. An artist list was created, which consisted of 10 better-known and 10 lesser-known artists from each of the five genres: Pop, Indie, Hip-Hop/Rap, Rock and Techno. These genres were the most frequently listened to according to the survey (Appendix B). From each of the 100 artists, their five top tracks were scraped resulting in data from 500 songs being collected.

Popularity was measured using the artist's Spotify's popularity index. This metric, which runs from 0 to 100 (0 being not popular, and 100 being extremely popular) is calculated for each song based on its total streams, the age of the song and its total streams per listener ratio. An artists' popularity index is then derived from each of their songs' popularity. Lesser-known artists were defined as those having a popularity index of less or equal to 30 (which corresponded to 50,000 monthly listeners), as these were expected to be most in need of increased exposure. Better-known artists were defined as those having a popularity index of greater than 30.

Algorithm

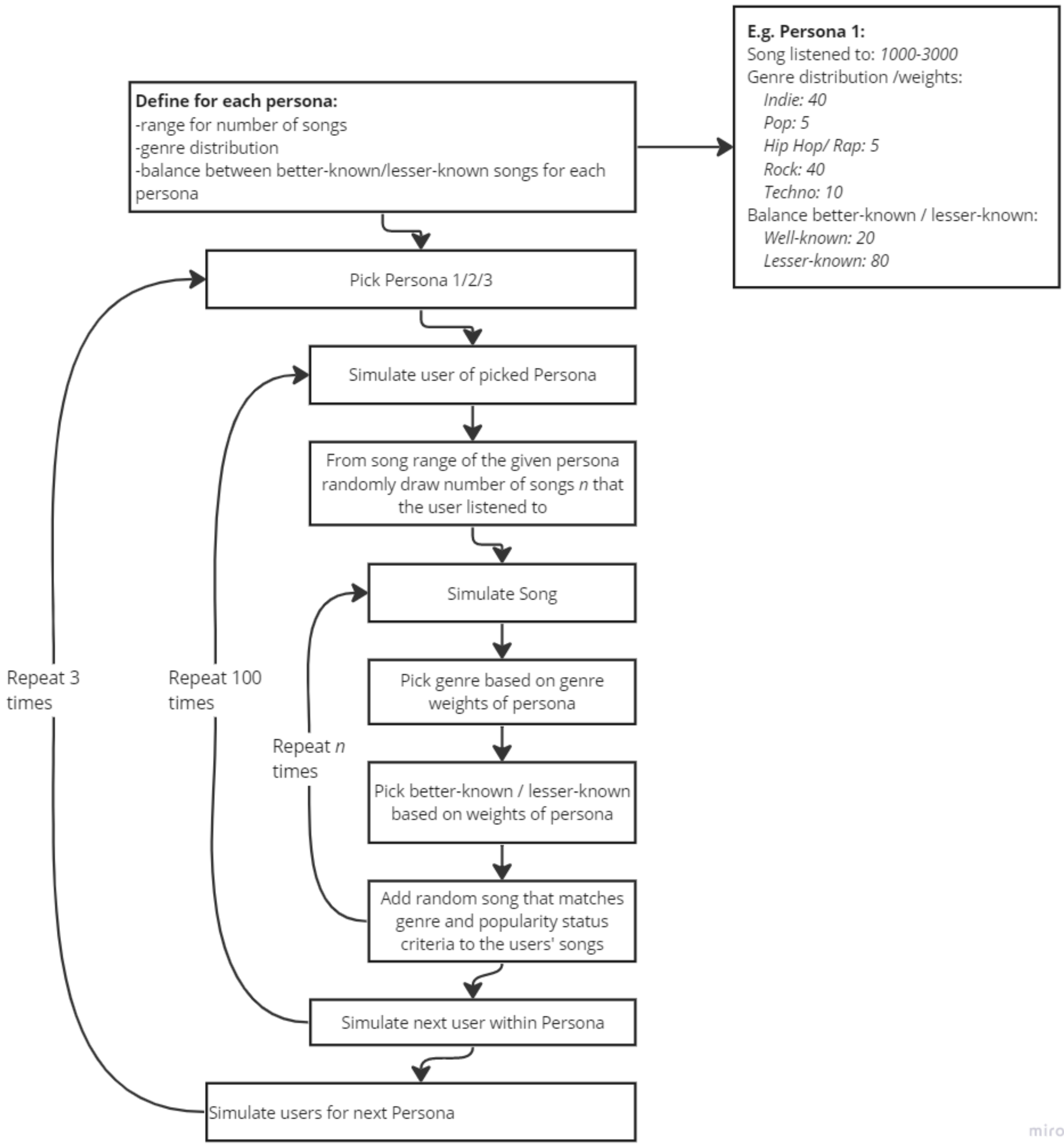
Synthetic data

For each persona, parameters were defined with regards to how much the persona listens to each genre, their distribution of better-known and lesser-known songs, and how many songs they listen to. With these parameters, each song each user 'listened to' was simulated using

Python. Listening data of a total of 300 users (100 per persona) was simulated. A detailed illustration of the synthesising process can be found in Fig. 4.

Figure 4

Data Synthesis Process

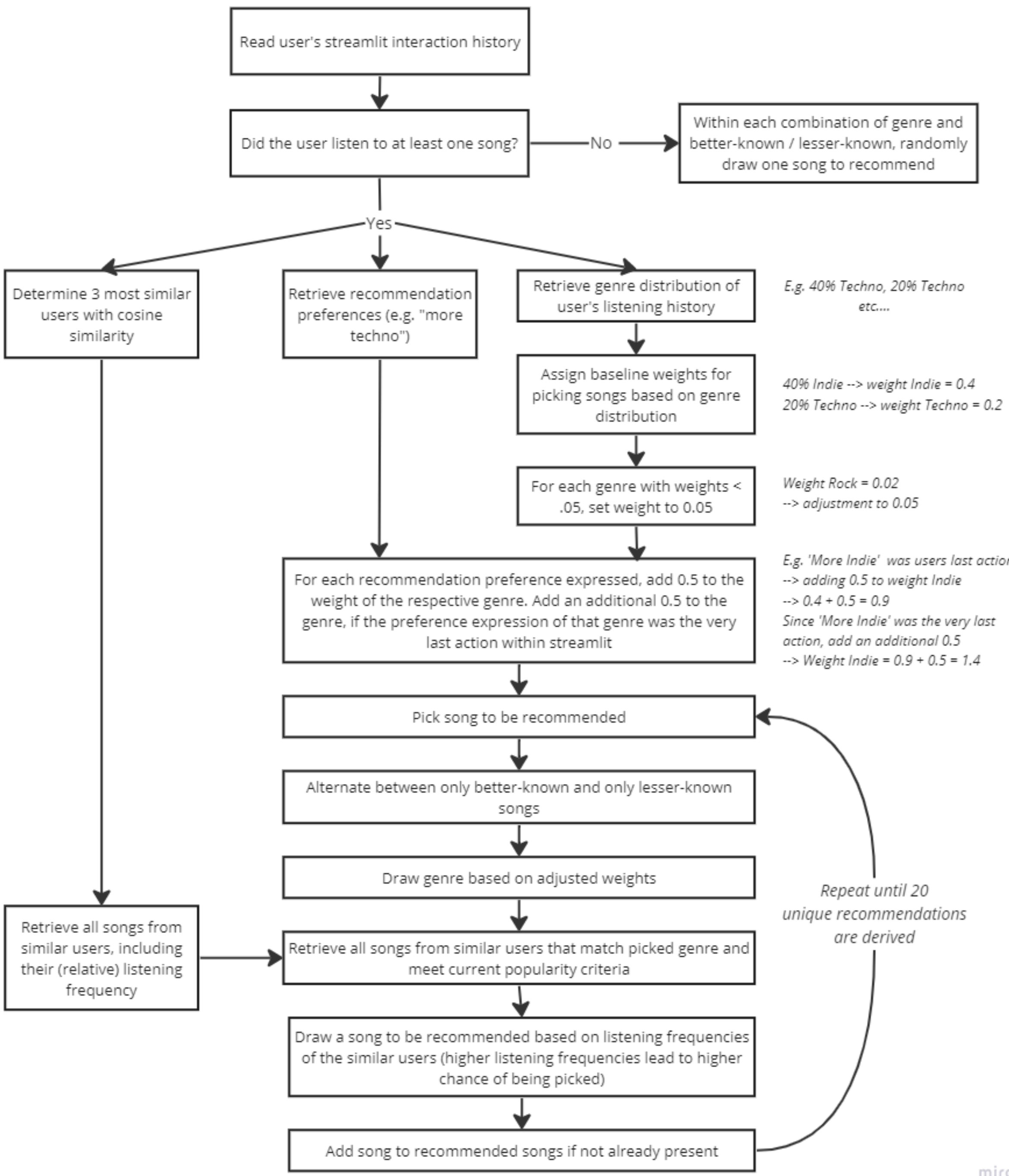


Recommender system

Recommendations are derived from a combination of listening behaviour of similar users (collaborative filtering), the active user's genre preferences (content-based filtering), the preferences expressed by the user in the interface (e.g. 'more Rock'), and an adaptation to ensure equal distribution between better-known and lesser-known songs. The similar users were found using cosine similarity based on what songs they listened to and to what frequency. This method was chosen over other similarity measures because it is considered to be effective when working with sparse data, which is often the case with real-world user listening data (Kirişci, 2023). The listening frequency was used to measure the likeability of a song for a given user since it was found to be a good predictor. (Madison & Schiölde, 2017). Nevertheless, other metrics such as the number of times a song was shared and how often it was saved to playlists could serve as an extension of this measure. A detailed illustration of the algorithm logic is displayed in Fig. 5.

Figure 5

Recommendation Algorithm



The recommendations were re-derived after each interaction the user made. All the files and code for the RS can be found at:

https://github.com/anedicksh/Fairness-in-music-recommender-systems

Measurement of fairness

To quantify fairness in user's recommendations, one approach is to calculate the disparate impact (DI) ratio (see Formula 1). This metric quantifies the difference between probabilities of positive outcomes (e.g. being recommended) for individuals that belong to the favoured group (e.g. better-known artists) and those who belong to the disfavoured group (e.g. lesser-known artists) (L. Cardoso et al., 2019).

$$disparate\ impact\ (DI)\ ratio = \frac{P(y^{+} \mid s = \bar{g})}{P(y^{+} \mid s = g)} \quad (1)$$

If the ratio is equal to 1, it indicates a fair balance between a user's recommendations for both groups. On the contrary, a ratio lower or higher than 1 suggests potential popularity bias. In order to measure fairness, the list of songs recommended to each user along with the artist's level of popularity are needed. The obtained DI ratio for the proposed RS was 1, as this was programmed in the algorithm to achieve fairness.

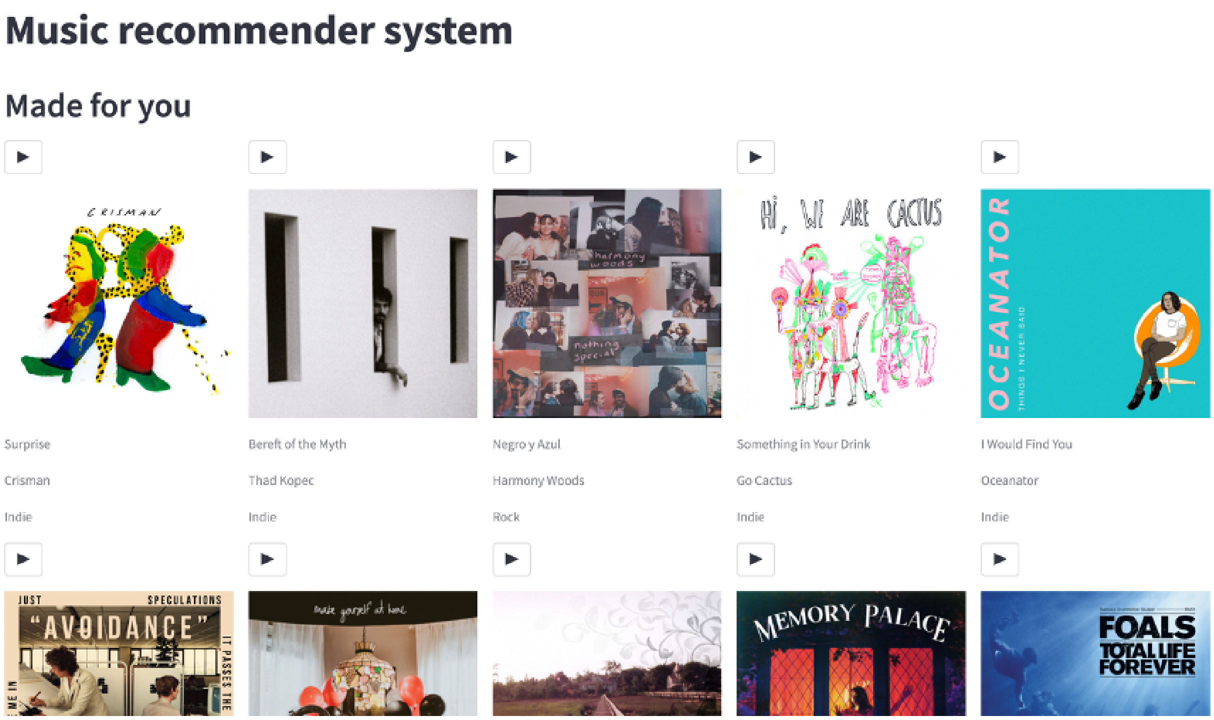
Note that this formula assumes that the target distribution is an equal proportion of well-known and lesser-known artists for all users, regardless of their preference.

Interface

Streamlit was used to create the user interface. The interface includes a ‘Made for you’ section that displays the personalised music recommendations. Additionally, it incorporates a button that the user can press to play the desired song (see Figure 6).

Figure 6

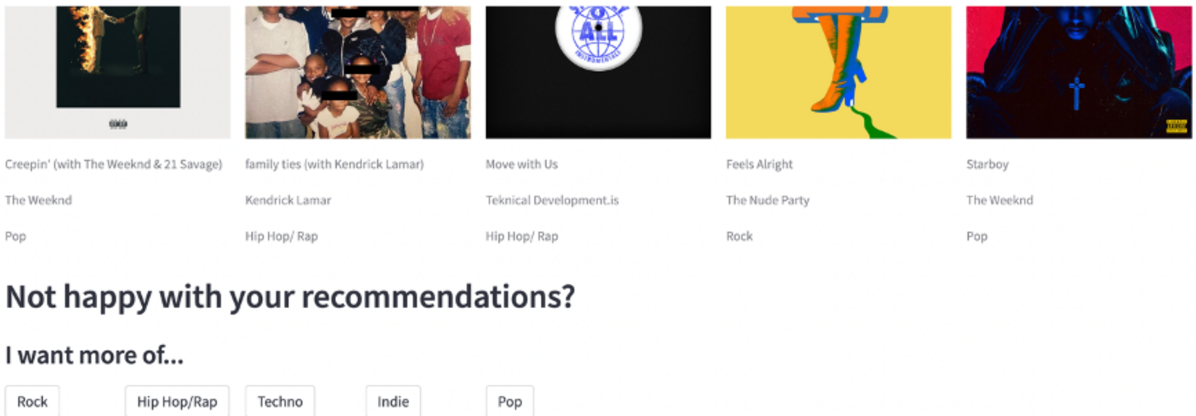
‘Made for you’ section



Moreover, the user has the option to indicate their preferred music genre by making use of the ‘I want more of…’ section and choosing the desired genre (see Figure 7).

Figure 7

'I want more of…' section



Discussion

This report provided a framework for implementing fairness into Spotify's recommender system (RS) through the mitigation of popularity-bias. The prototype RS recommended an equal proportion of lesser-known to better-known artists in order to reduce the inequality observed across music streaming services (Blake, 2020). The RS achieved this as the disparate impact ratio was one (as programmed by the recommendation algorithm), which meant that both groups were equally exposed through the system.

A significant limitation of the developed RS was that the operationalisation of popularity (i.e., the distinction between lesser-known and better-known artists) was binary, whereas popularity is a continuous construct. A clear-cut criteria for lesser-known artists having a Spotify popularity index of less than 30 was used (Sciandra & Spera, 2022). However, artists who had a popularity index of between 31 and 100 were treated the same by the RS. An operationalisation of popularity which takes into account the different levels of popularity, similar to that employed by Miyamoto et al. (2019), might be beneficial. Future research should attempt to incorporate a more nuanced operationalisation of popularity.

Another limitation is that there are several value-tensions between the multiple stakeholders of the RS. This project's operationalisation of fairness was focused on the treatment of unpopular musical artists in general. However, there are other characteristics other than popularity which might account for unfairness, such as gender (Mansoury et al., 2020) and race (Yao & Huang, 2017). Additionally, not all music listeners will want lesser-known artists to be recommended to them. While the user can influence the genre proportion in the developed RS, the user has no control over the ratio of popular and less-popular songs. Therefore, fairness clashes with the value of autonomy, which is the capacity of a user to make his own informed decisions (Tiribelli, 2023). In addition, this conflicts with the value of transparency, i.e., being open and honest (Ball, 2009), as the RS does not make it apparent to users that this implementation has occurred. Moreover, the private organisations which run these streaming services (e.g., Spotify) may value the accumulation of wealth over fairness. As they are a private company, they may ignore public values as they are less responsible in abiding to such values (van Dijck, 2020). Consequently, a balance must be found in which all the stakeholders' needs are taken into account. Future research may attempt to do this by implementing transparency and autonomy into the proposed RS through being explicit about the popularity bias mitigation procedure (which can be easily fixed) and by giving users a choice in this matter (which is more controversial as this may reduce fairness in the RS).

References

Adamopoulos, P., & Tuzhilin, A. (2014). On over-specialization and concentration bias of recommendations: Probabilistic neighborhood selection in collaborative filtering systems. *Proceedings of the 8th ACM Conference on Recommender Systems*, 153–160. https://doi.org/10.1145/2645710.2645752

Amatriain, X., & Basilico, J. (2015). Recommender Systems in Industry: A Netflix Case Study. In F. Ricci, L. Rokach, & B. Shapira (Eds.), *Recommender Systems Handbook* (pp. 385–419). Springer US. https://doi.org/10.1007/978-1-4899-7637-6_11

Anderson, C. (2004). *The Long Tail*. https://www.wired.com/2004/10/tail/

Ball, C. (2009). What is transparency? *Public Integrity*, *11*(4), 293–308.

Blake, E. (2020, September 9). Data Shows 90 Percent of Streams Go to the Top 1 Percent of Artists. *Rolling Stone*. https://www.rollingstone.com/pro/news/top-1-percent-streaming-1055005/

Boratto, L., Fenu, G., & Marras, M. (2021). Connecting user and item perspectives in popularity debiasing for collaborative recommendation. *Information Processing & Management*, *58*(1), 102387.

Curry, D. (2023). *Music Streaming App Revenue and Usage Statistics (2023)*. Business of Apps. https://www.businessofapps.com/data/music-streaming-market/

Kirişci, M. (2023). New cosine similarity and distance measures for Fermatean fuzzy sets and TOPSIS approach. *Knowledge and Information Systems*, *65*(2), 855–868. https://doi.org/10.1007/s10115-022-01776-4

L. Cardoso, R., Meira Jr., W., Almeida, V., & J. Zaki, M. (2019). A Framework for Benchmarking Discrimination-Aware Models in Machine Learning. *Proceedings of the*

2019 AAAI/ACM Conference on AI, Ethics, and Society, 437–444. https://doi.org/10.1145/3306618.3314262

Longhurst, B. (2007). *Popular Music and Society*. Polity.

Madison, G., & Schiölde, G. (2017). Repeated Listening Increases the Liking for Music Regardless of Its Complexity: Implications for the Appreciation and Aesthetics of Music. *Frontiers in Neuroscience*, *11*, 147. https://doi.org/10.3389/fnins.2017.00147

Mansoury, M., Abdollahpouri, H., Smith, J., Dehpanah, A., Pechenizkiy, M., & Mobasher, B. (2020). *Investigating Potential Factors Associated with Gender Discrimination in Collaborative Recommender Systems* (arXiv:2002.07786). arXiv. https://doi.org/10.48550/arXiv.2002.07786

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A Survey on Bias and Fairness in Machine Learning. *ACM Computing Surveys*, *54*(6), 115:1-115:35. https://doi.org/10.1145/3457607

Miyamoto, S., Zamami, T., & Yamana, H. (2019). Improving Recommendation Diversity Across Users by Reducing Frequently Recommended Items: 2018 IEEE International Conference on Big Data, Big Data 2018. *Proceedings - 2018 IEEE International Conference on Big Data, Big Data 2018*, 5392–5394. https://doi.org/10.1109/BigData.2018.8622314

Ricci, F., Rokach, L., & Shapira, B. (2022). Recommender Systems: Techniques, Applications, and Challenges. In F. Ricci, L. Rokach, & B. Shapira (Eds.), *Recommender Systems Handbook* (pp. 1–35). Springer US. https://doi.org/10.1007/978-1-0716-2197-4_1

Saxena, N. A. (2019). Perceptions of Fairness. *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*, 537–538. https://doi.org/10.1145/3306618.3314314

Schedl, M., Zamani, H., Chen, C.-W., Deldjoo, Y., & Elahi, M. (2018). Current challenges and visions in music recommender systems research. *International Journal of Multimedia Information Retrieval*, *7*(2), 95–116. https://doi.org/10.1007/s13735-018-0154-2

Sciandra, M., & Spera, I. C. (2022). A model-based approach to Spotify data analysis: A Beta GLMM. *Journal of Applied Statistics*, *49*(1), 214–229. https://doi.org/10.1080/02664763.2020.1803810

Spotify. (2023a). *Spotify Reveals More Opportunities and Features for Creators During Stream On*. Spotify. https://newsroom.spotify.com/2023-03-08/stream-on-announcements-new-features-updates-2023-event/

Spotify. (2023b). *Spotify—About Spotify*. Spotify. https://newsroom.spotify.com/company-info/

Tiribelli, S. (2023). *The AI ethics principle of autonomy in health recommender systems*. https://doi.org/10.14275/2465-2334/20230.TIR

Turnbull, D. R., McQuillan, S., Crabtree, V., Hunter, J., & Zhang, S. (2022). *Exploring Popularity Bias in Music Recommendation Models and Commercial Steaming Services* (arXiv:2208.09517). arXiv. http://arxiv.org/abs/2208.09517

van Dijck, J. (2020). Governing digital societies: Private platforms, public values. *Computer Law & Security Review*, *36*, 105377. https://doi.org/10.1016/j.clsr.2019.105377

Yalcin, E., & Bilge, A. (2022). Evaluating unfairness of popularity bias in recommender systems: A comprehensive user-centric analysis. *Information Processing & Management*, *59*(6), 103100. https://doi.org/10.1016/j.ipm.2022.103100

Yao, S., & Huang, B. (2017). Beyond Parity: Fairness Objectives for Collaborative Filtering. *Advances in Neural Information Processing Systems*, *30*.

https://proceedings.neurips.cc/paper/2017/hash/e6384711491713d29bc63fc5eeb5ba4f-Abstract.html

Zhu, Z., He, Y., Zhao, X., Zhang, Y., Wang, J., & Caverlee, J. (2021). Popularity-Opportunity Bias in Collaborative Filtering. *Proceedings of the 14th ACM International Conference on Web Search and Data Mining*, 85–93. https://doi.org/10.1145/3437963.3441820

Appendix A

Survey

Opinions on Music Recommender Systems Survey

Q1 Do you use a music streaming service (e.g. Spotify, Apple Music, YouTube…)

#	Field		Choice Count
1	Yes	95.00%	76
2	No	5.00%	4
			80

Q2 - What is your preferred music streaming service, and why?

#	Field		Choice Count
1	Spotify	90.00%	54
2	Apple Music	5.00%	3
3	YouTube	3.33%	2
4	Tidal	0.00%	0
5	Deezer	0.00%	0
6	Other	1.67%	1
			60

Q3 - How often do you listen to music on your music streaming service?

#	Field		Choice Count
1	Frequently (everyday)	86.67%	52
2	Occasionally (3-5 days a week)	11.67%	7
3	Rarely (1-2 days a week)	1.67%	1
			60

Q4 - Select the top three music genres that you listen to

#	Field	Choice Count	
1	Pop	18.03%	33
2	Hip-Hop/Rap	15.30%	28
3	Indie	14.21%	26
4	Rock	18.58%	34
5	Jazz	5.46%	10
6	Techno	10.93%	20
7	Metal	4.37%	8
8	Classical	5.46%	10
9	Other	7.65%	14
			183

Q5 - Do you care about the recommendations that your music streaming service provides you with?

#	Field	Choice Count	
1	Yes, they are good	33.33%	20
2	Yes, they are easily accessible	36.67%	22
3	No, I find new music elsewhere	11.67%	7
4	No, they are bad	3.33%	2
5	I don't mind	15.00%	9
			60

Q6 - Do you value less-known artists having as much of a chance to be listened to as well-known artists?

#	Field	Choice Count	
1	Agree	64.41%	38
2	Somewhat agree	20.34%	12
3	Somewhat disagree	13.56%	8
4	Disagree	1.69%	1
			59

Q7 - Which artists are you currently being recommended…

#	Field	Choice Count	
1	Only well-known artists	5.00%	3
2	Mainly well-known + some less-known artists	41.67%	25
3	50/50	23.33%	14
4	Mainly less-known artists + some well-known artists	15.00%	9
5	Only less-known artists	3.33%	2
6	I don't know	11.67%	7
			60

Q8 - Would you like your music streaming service to recommend to you…

#	Field	Choice Count	
1	Only well-known artists	0.00%	0
2	Mainly well-known + some less-known artists	18.33%	11
3	50/50	41.67%	25
4	Mainly less-known artists + some well-known artists	26.67%	16
5	Only less-known artists	5.00%	3
6	I don't mind	8.33%	5
			60

Q9 - Would you consider it unfair if only less-known artists were recommended to you?

#	Field		Choice Count
1	Yes	33.33%	20
2	Maybe	33.33%	20
3	No	33.33%	20
			60

Q10 - How much do you care about being able to adjust the ratio of less-known/well-known artists in your recommendations?

#	Field		Choice Count
1	A lot	30.00%	18
2	Somewhat	45.00%	27
3	Not much	20.00%	12
4	Not at all	5.00%	3
			60

Appendix B

Persona weights and definition for data synthesis

	Persona 1	**Persona 2**	**Persona 3**
Number of songs listened to	1000 - 3000	1000 - 3000	200 - 500
Genre weights distribution	Indie: 40 Pop: 5 Hip Hop/ Rap: 5 Rock: 40 Techno: 10	Indie: 0 Pop: 40 Hip Hop/ Rap: 40 Rock: 10 Techno: 10	Indie: 30 Pop: 30 Hip Hop/ Rap: 5 Rock: 5 Techno: 30
Better-known / lesser-known weights distribution	Better-known: 20 Lesser-known: 80	Better-known: 70 Lesser-known: 30	Better-known: 80 Lesser-known: 20