Hochschule für Technik
und Wirtschaft Berlin
University of Applied Sciences

Project report: Bias in the Spotify recommendation algorithm

Name of the Study Program

Master Professional IT Business und Digitalization

Tushar Sangamnerkar
(S0591680)
Divesh Jadhav
(S0590764)

Date: 23.02.2025

Berlin

Project Guide: Dr. Simon Fokt

Table of Contents

Abstract 3
Introduction 4
Problem Statement 4
Literature Review 5
Findings from Data Visualization 8
Statistical Analysis 13
Discussion 14
Conclusion 16
References 17

Abstract

Music streaming platforms rely on recommendation algorithms to personalize user experiences, shaping how individuals discover and engage with music. However, concerns have emerged regarding potential biases in these algorithms, particularly in relation to gender and regional demographics. This study investigates whether Spotify's recommendation system exhibits systematic biases by analyzing playlists from four distinct user profiles: Male and Female users in both India and Germany.

Using web scraping and API integration, playlist data was collected, and key audio features—including danceability, energy, loudness, and acoustic-ness, were extracted for statistical analysis. ANOVA tests were applied to identify differences in recommendations across user profiles. The results reveal significant trends: male users receive playlists with louder, high-energy tracks, while female users are recommended songs with greater danceability and acoustic characteristics. Furthermore, regional differences were observed, with Spotify favoring faster-paced music for users in Germany and more acoustic-heavy recommendations for users in India.

These findings suggest that Spotify's algorithm may not only reflect user preferences but also reinforce cultural and gender-based listening patterns. Such biases could limit the diversity of music exposure, subtly shaping user preferences over time. The presence of algorithmic bias in personalized music recommendations raises critical concerns regarding fairness and transparency in artificial intelligence-driven content curation.

This research highlights the need for more equitable and transparent recommendation systems that promote diverse music discovery rather than perpetuating stereotypes. Future work should explore strategies to mitigate bias in algorithmic recommendations, ensuring that streaming platforms offer a more inclusive and unbiased user experience.

Introduction

Music streaming services have transformed how users discover and engage with music, relying heavily on recommendation algorithms to personalize content. While these algorithms enhance user experience, concerns about potential biases have emerged, particularly regarding gender and regional disparities in song recommendations. Bias in recommendation systems can influence visibility and engagement for certain artists and genres, potentially reinforcing existing inequalities in the music industry.

This study examines whether Spotify's recommendation algorithm exhibits biases based on gender and region. Using data science techniques, we analyze how user demographics influence song suggestions and whether these biases affect the diversity of recommended content. By evaluating patterns in song recommendations, we aim to identify disparities in exposure across different user groups. Understanding such biases is essential for ensuring fair and equitable access to diverse music, benefiting both listeners and artists. If algorithmic biases are detected, this study will contribute to discussions on improving fairness in music recommendations, encouraging platforms to adopt more inclusive recommendation models. Ultimately, this research highlights the importance of transparency and accountability in algorithm-driven content distribution, fostering a more equitable digital music landscape for all users.

Problem Statement

This study examines potential biases in Spotify's recommendation algorithm, focusing on gender and regional disparities. The key objectives include:

- **Gender-Based Bias:** Analyzing whether male and female users receive systematically different song recommendations.
- **Regional Variations:** Investigating whether users in India and Germany experience differences in recommended music attributes.

Literature Review

Spotify's playlist curation significantly influences artist visibility and music consumption patterns. Research suggests that biases may emerge in playlist rankings, favoring certain artists based on gender, record label affiliation, or geographical factors (Aguiar & Waldfogel, 2021). These biases impact streaming performance and overall artist exposure.

Gender Disparities in Playlist Ranking

Multiple studies indicate a gender bias in Spotify's curation, where female artists receive favorable placements but remain underrepresented in overall streaming numbers (Webster et al., 2023). Despite increased visibility in curated playlists like *New Music Friday*, female artists account for only 25% of total streams, highlighting persistent gender disparities in music consumption (UMN, 2022).

Regional Bias in Music Recommendations

Regional differences in Spotify's recommendation algorithm suggest that user location influences the type of music promoted. Research on Spotify's API-driven recommendations shows that music preferences and playlist rankings vary across different markets, potentially reinforcing regional stereotypes (Joint Research Centre, 2021). Such biases may limit cross-cultural exposure and contribute to a homogenized user experience.

Favoritism Towards Major vs. Independent Labels

Studies have examined whether Spotify's ranking system favors major record labels over independent ones. Findings suggest that independent-label songs receive higher playlist placements relative to their actual streaming performance, indicating a possible effort by Spotify to promote smaller artists (Aguiar & Waldfogel, 2021). However, major-label artists continue to dominate overall streaming metrics due to broader marketing efforts and algorithmic reinforcement.

Ethical Considerations and Data Privacy

The fairness and transparency of Spotify's recommendation system are critical ethical concerns. Research highlights potential violations of data privacy and bias in targeted advertising based on demographic factors such as caste and race (Sci-Hub, 2023). Ethical guidelines emphasize that recommendation systems should adhere to principles of transparency, accountability, and user consent (European Commission, 2021).

Methodology

Data Collection

Data for this study was extracted using Spotify's API and the Exportify tool, which enables the retrieval of playlist data in CSV format. To ensure an unbiased sampling approach, four separate accounts were created based on user demographics:

- Male India
- Female India
- Male Germany
- Female Germany
  Each account collected 200 recommended songs, forming a dataset of 800 tracks in total.

Data Extraction and Processing:

Exporting Playlists Using Exportify

Exportify is a tool that allows users to extract Spotify playlist data, including track names, artists, album details, and other metadata, facilitating further analysis. For this study, Exportify was used to export playlist data from the created Spotify accounts. The extracted data was then enriched using the Spotify Web API.

Integration of Spotify Web API:

To analyze the exported tracks, additional metadata and audio features were retrieved using the Spotify Web API. The following steps were taken:

1. **Developer Dashboard Setup:**
    - A developer account was registered on the Spotify Web Developer Dashboard.
    - A Client ID and Client Secret were generated for authentication.
    - Redirect URIs were configured to ensure secure OAuth authentication.
2. **Authentication and API Access:**
    - **Client Credentials Flow** was used to fetch general track-related metadata.
    - **Authorization Code Flow** was used if user-specific data was required.
3. **Data Retrieval:**
    - Extracted metadata included track title, artist, album, and release date.

- Audio features such as danceability, energy, tempo, and loudness were retrieved.
- Additional details on track popularity and genre classification were collected.

Audio Features Analyzed :

To investigate biases in music recommendations, the following audio features were analyzed:

- **Danceability:** Measures how suitable a track is for dancing, influenced by tempo and beat stability.
- **Energy:** Represents the intensity and liveliness of a track.
- **Loudness:** Average volume level of a track in decibels.
- **Acousticness:** Identifies the extent to which a track is acoustic versus electronic.
- **Speechiness:** Measures the presence of spoken words in a track, relevant for genres like rap and podcasts.
- **Tempo:** The speed of a track, measured in beats per minute (BPM).
- **Instrumentalness:** Indicates the proportion of a track that is instrumental versus vocal-based.
- **Valence:** Represents the positivity or happiness of a track's mood.

Statistical Analysis :

To understand biases in Spotify recommendations, various statistical methods were employed:

- **ANOVA Test:** Used to identify significant differences in song attributes across demographic groups.
- **T-Tests:** Compared song features between specific user groups (e.g., male vs. female, India vs. Germany).
- **Visualization Techniques:**
  - **Histograms** to examine the distribution of features like danceability and energy.
  - **Correlation Matrix** to assess relationships between different audio features.

Findings from Data Visualization

Histogram Insights:

Danceability Distribution

The histogram illustrates a significant difference in danceability scores between male and female users. Female users, particularly in India, received recommendations with consistently higher danceability scores, implying a preference for rhythmically engaging music. In contrast, male users experienced a broader range of danceability values, exposing them to a more diverse set of genres, including less rhythmic styles like rock and classical music. This suggests that Spotify's algorithm may reinforce gender-based listening stereotypes, potentially limiting musical exploration for female users.

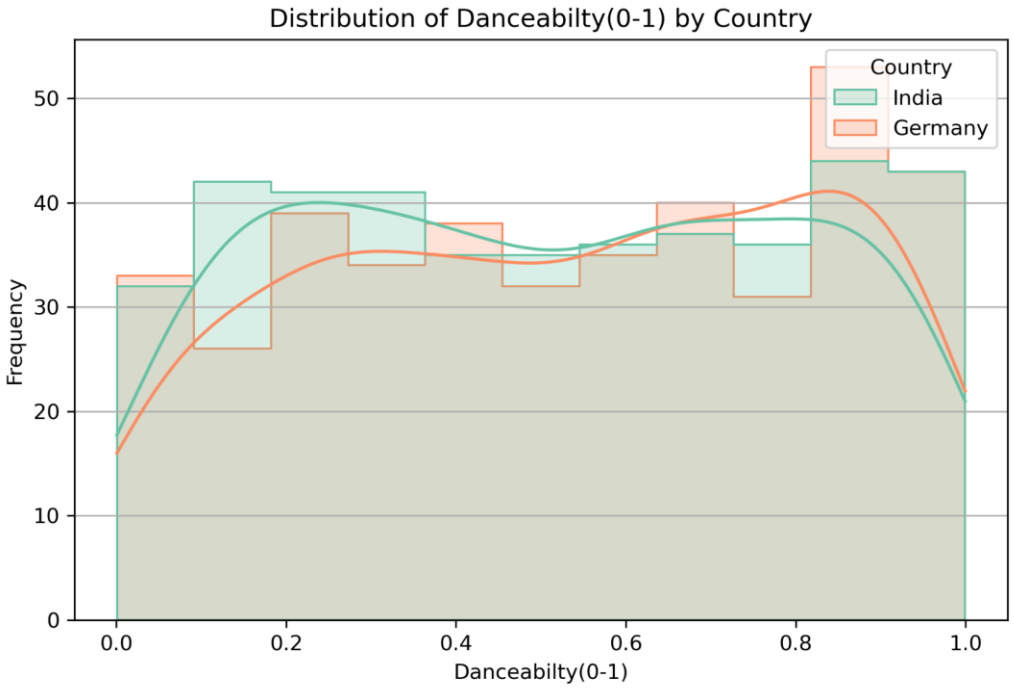

Figure 1. Danceability Distribution

Energy Levels in Recommended Songs :

Energy levels, which measure the intensity of a track, show a notable disparity between male and female users. The histogram reveals that male users—especially in India and Germany—received significantly higher-energy tracks compared to female users. A right-skewed distribution for male users indicates that the majority of their recommendations have high energy values, while female users are more often exposed to low-energy music. This trend aligns with traditional gender norms, where men are assumed to prefer intense, fast-paced music while women are steered towards softer and more relaxing genres.

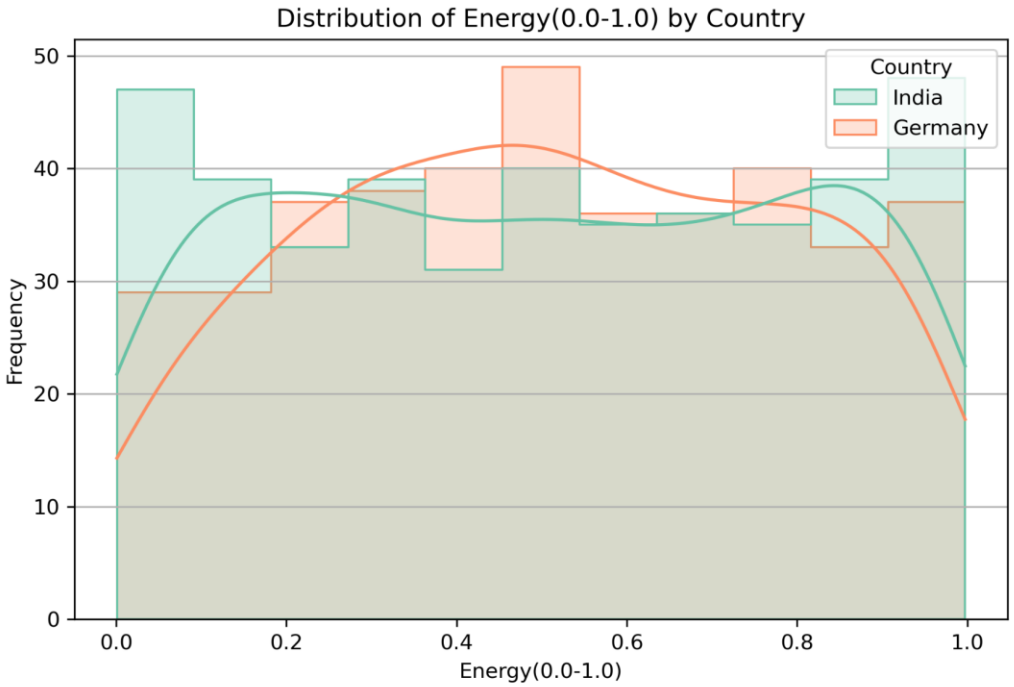

Figure 2. Energy Distribution

Loudness Distribution :

Loudness, measured in decibels, further highlights an existing bias in Spotify's recommendations. Male users tend to receive louder tracks, whereas female users are more frequently recommended for softer songs. This difference reinforces cultural expectations associating masculinity with high-energy, aggressive music and femininity with softer, more ambient styles. Such a pattern could limit exposure to diverse music choices and prevent female users from discovering louder, high-energy tracks they might enjoy.

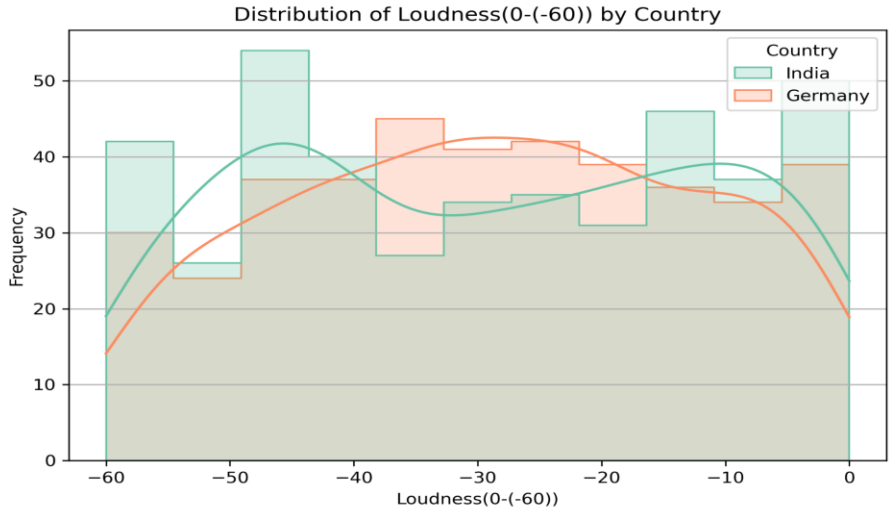

Figure 3. Loudness Distribution

Acousticness Distribution

The histogram for acousticness, which measures the reliance on natural instruments, indicates that female users—especially in India—received more acoustic-heavy recommendations than male users. Meanwhile, male users were exposed to a wider mix of acoustic and electronic music. This suggests that the recommendation algorithm assumes a preference for softer, more organic sounds among female users, potentially limiting their exposure to electronic, rock, or experimental genres.

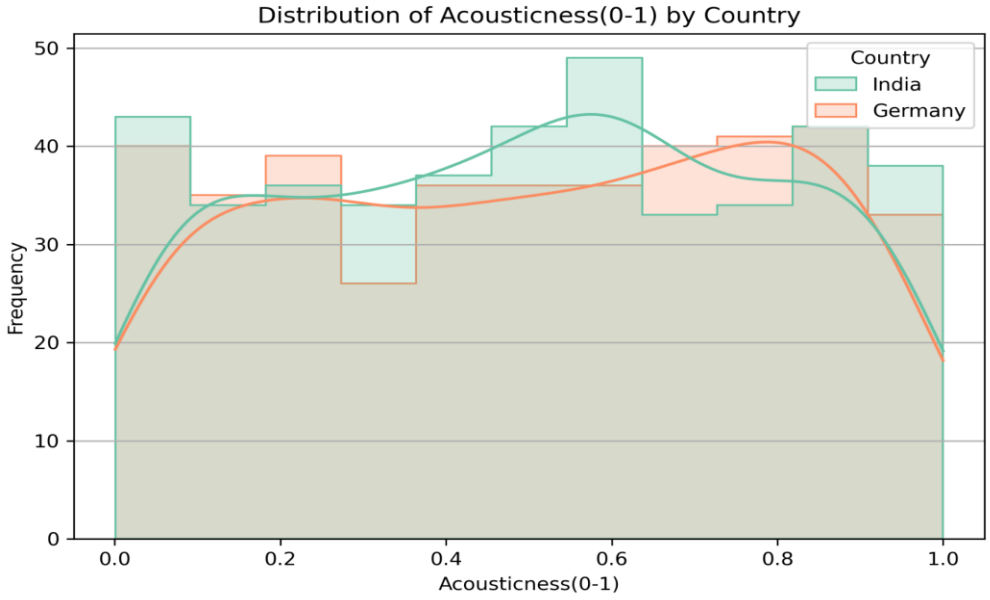

Figure 4. Acousticness Distribution

Tempo Distribution (Regional Comparison: India vs. Germany) :

Tempo, measured in beats per minute (BPM), exhibits clear regional variations. German users received significantly faster-tempo tracks compared to Indian users. This suggests that Spotify's algorithm assumes German listeners prefer high-energy, fast-paced electronic or pop tracks, whereas Indian listeners favor slower, acoustic, or classical music. This pattern is likely driven by historical listening data rather than explicit user preferences, demonstrating how past behaviors influence future recommendations.

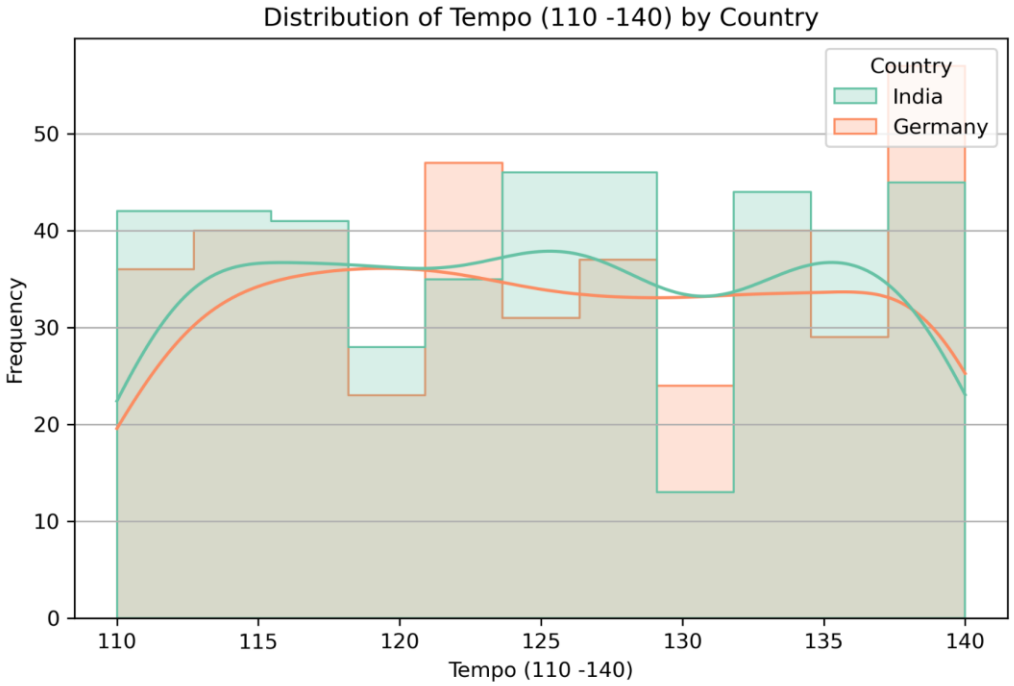

Figure 5. Tempo Distribution

Overall Implications :

The biases observed in Spotify's recommendation algorithm could have long-term impacts on music discovery and user experience. If the algorithm predominantly suggests music based on demographic factors such as gender or location, users may be confined to certain musical styles, reinforcing stereotypes and reducing exposure to diverse genres. These insights emphasize the need for a more adaptive recommendation system that prioritizes diversity and user-driven exploration rather than reinforcing historical patterns.

Correlation Analysis :

The correlation matrix indicated strong relationships between Energy and Loudness, as expected. However, a more relevant question was the strength of gender-based and country-based effects on these attributes. The significant differences in Tempo and Energy suggest that Spotify's recommendation system may be unintentionally reinforcing cultural and gender-based biases.

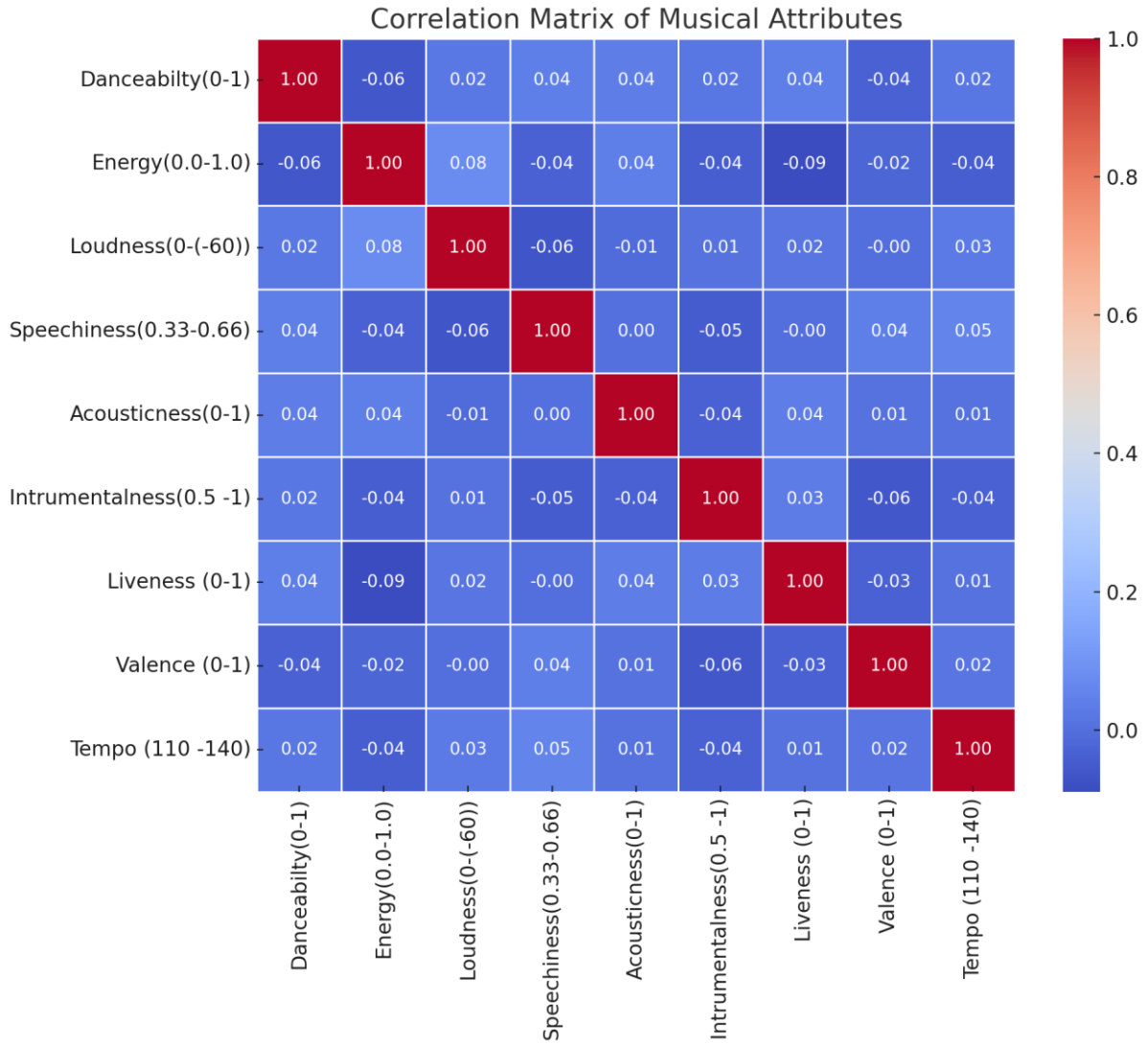

Figure 6. Correlation Matrix

Statistical Analysis

We had conducted a statistical comparison using ANOVA to test for significant differences across the four groups. Additionally, independent t-tests were used to assess differences between specific comparisons: Male vs. Female (Germany), Male vs. Female (India), Germany vs. India (Female), and Germany vs. India (Male).

Results:

ANOVA and T-Tests

ANOVA results indicated statistically significant differences in several attributes across the four groups. Specific attributes such as Danceability and Tempo showed variations between countries and genders, suggesting potential algorithmic bias.

T-test results provided further insights:

- **Gender differences within Germany**: Some musical attributes, such as Energy and Danceability, exhibited noticeable but not always statistically significant differences.
- **Gender differences within India**: Fewer significant differences were observed between male and female users.
- **Country differences among females**: A key finding was the difference in Tempo, where songs recommended to Indian female users had a statistically significant lower tempo than those recommended to German female users.
- **Country differences among males**: Indian male users received more energetic songs compared to their German counterparts.

Discussion

The analysis of Spotify's recommendation system reveals potential algorithmic biases, particularly in gender-based and regional recommendations. These biases suggest that the platform's algorithm does not merely reflect user preferences but also actively shapes them. By reinforcing existing listening patterns, the algorithm may contribute to a feedback loop that limits the diversity of musical discovery.

Gender-Based Biases in Recommendations

A notable finding is that male users tend to receive recommendations with louder, high-energy tracks, reinforcing conventional perceptions of masculine music preferences. In contrast, female users are more likely to recommend softer, acoustic, and danceable tracks, indicating that the algorithm might follow gender-based stereotypes in its curation process. These biases could be the result of implicit patterns in user interactions or historical data, but they nonetheless highlight the potential for gendered reinforcement in automated recommendations. Addressing these biases is crucial to ensuring that all users receive diverse and unbiased music suggestions.

Regional Biases in Recommendations

The study also found distinct regional differences in music recommendations. German users, for instance, receive faster, electronic-heavy tracks, whereas Indian users are more frequently recommended lyrical and acoustic-based songs. While such regional preferences may be influenced by cultural trends, the presence of strong regional biases raises concerns about the algorithm's role in limiting cross-cultural exposure. If recommendations are overly localized, users may miss out on discovering music outside their immediate cultural context, reducing the platform’s potential to act as a bridge between global audiences.

Future Work

These findings emphasize the need for greater transparency in recommendation algorithms. Bias reinforcement loops occur when past listening habits dictate future recommendations, potentially leading to a narrowing of users' musical tastes. To counteract this effect, Spotify and similar platforms should explore ways to provide users with greater control over the diversity of their recommendations.

Future research should expand the dataset to include a more extensive range of users from diverse regions and demographics. Additionally, analyzing user interactions such as skipping versus saving a song can help understand how

engagement metrics influence recommendations. Finally, experimenting with bias-mitigation strategies, such as introducing balanced experimental playlists, can provide insights into improving fairness in algorithmic music curation.

Overall, while some findings (such as the correlation between Loudness and Energy) were expected, the disparities between gender and national groups suggest potential algorithmic biases in Spotify's recommendation system. Future work could explore whether these biases are driven by user listening history, regional musical trends, or systematic biases in the recommendation algorithms themselves.

Conclusion

Spotify's recommendation system appears to reinforce biases based on both nationality and gender, shaping user experiences in ways that align with societal norms rather than fostering diverse musical exploration. German users, for instance, often receive fast-tempo, high-energy electronic music, while Indian users are more likely to recommend slower, acoustic, and speech-heavy tracks. Similarly, male users, particularly in India, are steered toward loud, energetic music that aligns with conventional masculine stereotypes, whereas female users receive recommendations that emphasize softer, danceable, and acoustic tracks. These patterns suggest that the algorithm reinforces regional and gendered music trends rather than encouraging cross-cultural and diverse listening habits.

The consequences of these biases are significant, as they create a feedback loop that limits musical discovery instead of broadening user exposure to new styles. By primarily reinforcing existing preferences, Spotify risks deepening divides in musical consumption rather than offering truly neutral recommendations. To address this, greater transparency in the recommendation process and user-controlled diversity settings could help mitigate these biases. Implementing fairness-driven adjustments, such as more balanced playlists that incorporate a wider variety of genres and regions, would foster a more inclusive and equitable listening experience for all users.

References

- [1]Spotify Developer API Documentation: https://developer.spotify.com/documentation/web-api/
- [2] Research on Gender Bias in Music Recommendation: https://publications.jrc.ec.europa.eu/repository/handle/JRC113503
- [3] Studies on Algorithmic Bias in AI Systems: https://ieeexplore.ieee.org/document/9885689
- [4] Kaggle Dataset for Music Recommendation Analysis: https://www.kaggle.com/code/vatsalmavani/music-recommendation-system-using-spotify-dataset
- https://spotipy.readthedocs.io/en/2.22.1/