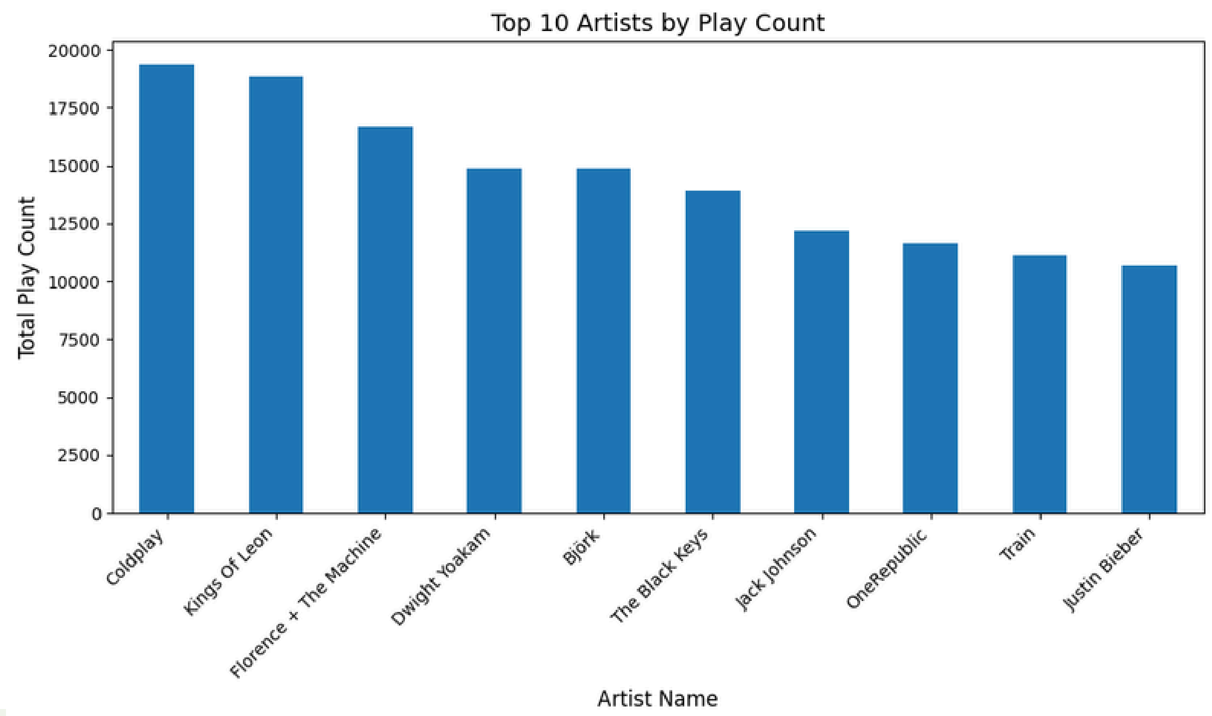
Music recommendation system

Univeristy of Tartu
Introduction to data science
2024/2025
Group F6
.

Data

For our project, we used data found on the Million Song Dataset website. The primary file we utilized was the Taste Profile Subset, which provides user_id-song_id-play_count triplets. To enhance the dataset, we merged the Taste Profile Subset with another file containing metadata about the songs, such as title and artist. Merging the two files made it more intuitive and easier to understand the data



OBJECTIVE

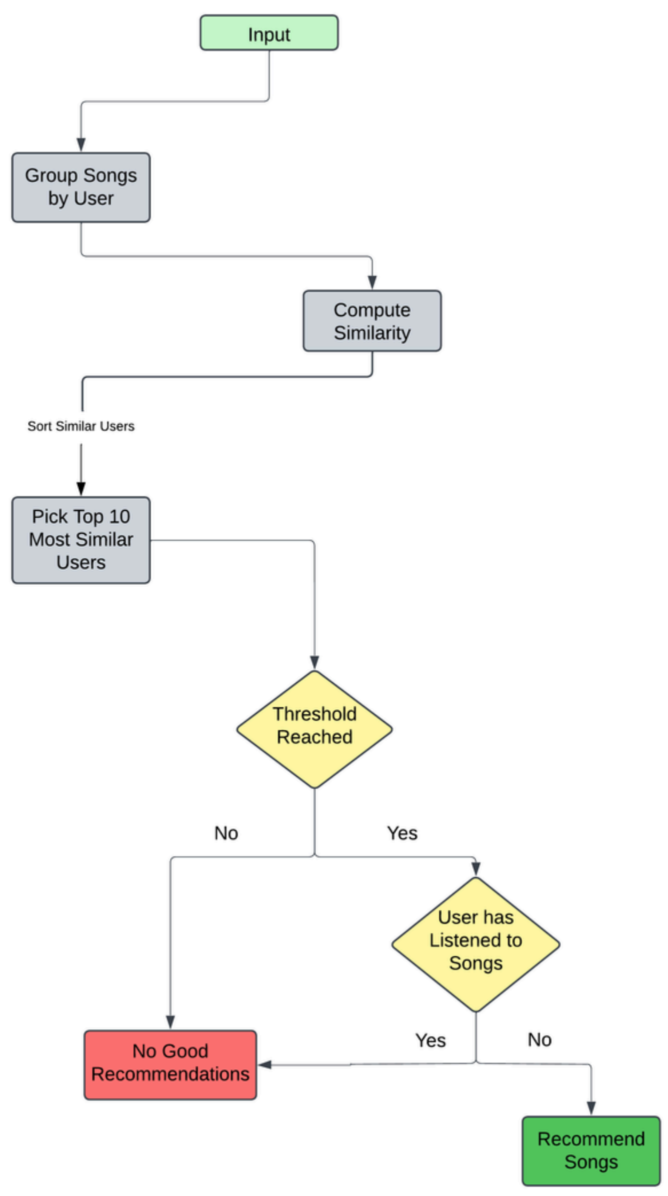
Our goal was to create a simple, educational and user-friendly music recommendation system that demonstrates the mechanics of recommendation algorithms. Additionally to make the recommendation system more easily understandable, we aimed to provide accurate song recommendations for users already in the dataset and introduce an interactive feature that allows new users to receive tailored suggestions based on their input.

METHODOLOGY

We chose collaborative filtering because it is a widely used approach in recommendation systems, aligning perfectly with our goal of creating a simple, educational, and user-friendly project. This approach not only ensured accurate and personalized recommendations but also demonstrated the mechanics of recommendation algorithms effectively, making the system accessible and understandable for others.

RESULTS

The project successfully achieved its goals of providing recommendations for both existing users in the dataset and new users via interactive input. Collaborative filtering proved effective in identifying users with similar music preferences and ensuring recommendations were tailored to shared tastes. The project offers insights into how analyzing user data can produce music recommendations, demonstrating the power of shared preferences in creating impactful results.



Procsess

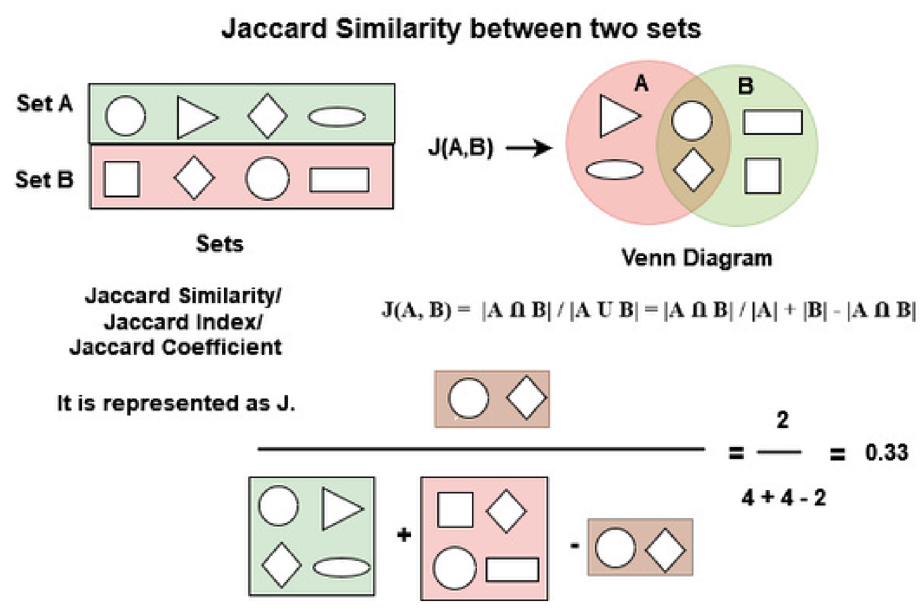
- The process began by calculating the Jaccard similarity between the input user's listening history and the histories of all other users in the dataset.
- Then we ranked users by how closely their song preferences matched the input user.
- From these rankings, we selected the n (usually we set n=10) most similar users and aggregated their listening histories.
- Next, we counted how many of these n users had listened to each song.
- To qualify as a recommendation, a song needed to have been played by at least n% of the most similar users, ensuring that the suggestions were popular within this group.
- Furthermore, we filtered out any songs that the input user had already listened to, ensuring that the recommendations were genuinely new and personalized.
- The final list of recommended songs was sorted by popularity and limited to the top 10 results for relevance and simplicity. This method ensured that the recommendations were both meaningful and diverse, drawing from the collective preferences of similar users.

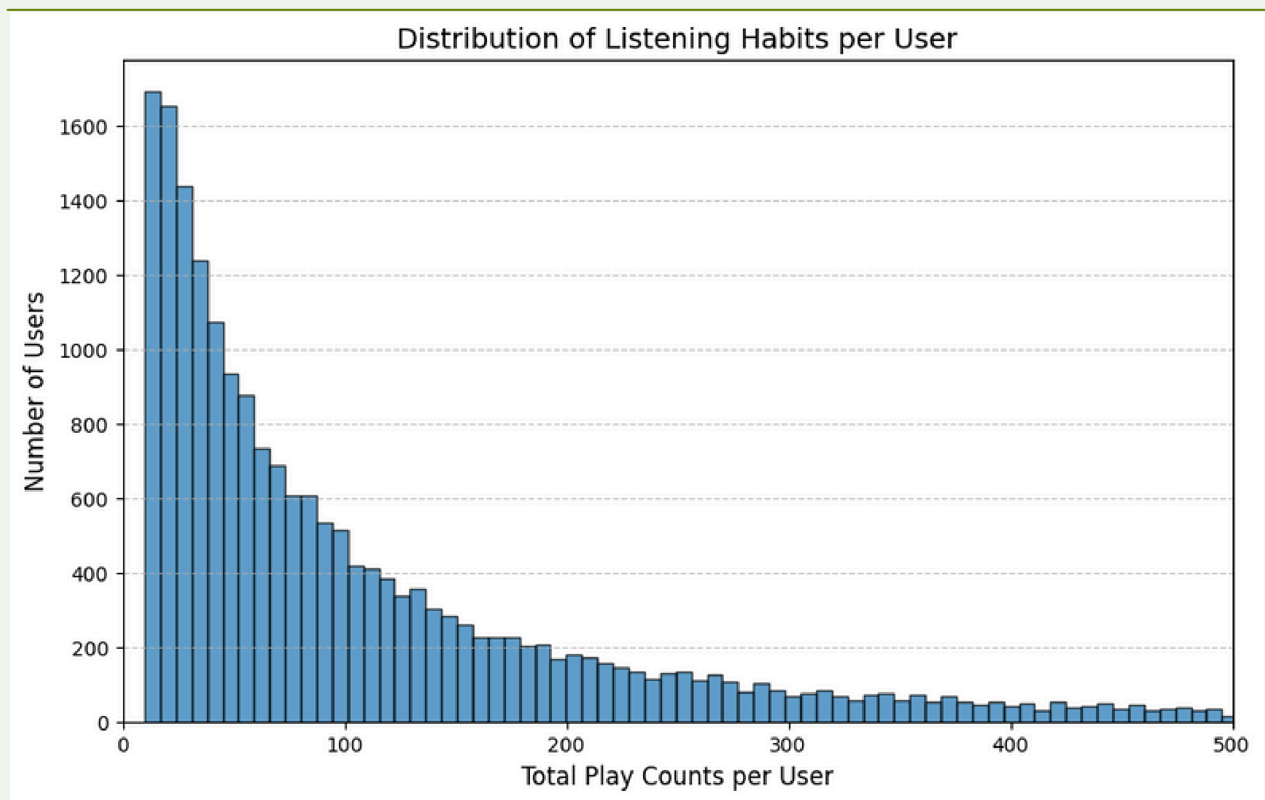
Jaccard similarity coefficient

The Jaccard similarity coefficient is a metric used to measure the similarity between two sets by dividing the size of their intersection by the size of their union.

This formula quantifies how much two sets overlap, with values ranging from 0 (no overlap) to 1 (complete overlap).

In our project, the Jaccard similarity coefficient was crucial for identifying users with the most similar listening histories to the input user.





Conclusion

In conclusion, our project successfully implemented a collaborative filtering-based music recommendation system that demonstrates the mechanics of recommendation algorithms.
By leveraging Jaccard similarity to analyze user listening histories, we provided accurate and personalized song suggestions for users within our dataset, while also introducing an interactive feature that allows new users to input their preferences for tailored recommendations. This approach not only reflects real-world listening behaviors but also avoids the complexities of metadata analysis, making the system efficient and educational.
Through comprehensive documentation and a focus on user-friendly design, the project achieves its goal of serving as both a practical tool and a resource for understanding how recommendation systems work. We hope it inspires further exploration into this fascinating area of data science.