WHAT ARE YOUR NEXT SONGS TO PLAY?

Yidan   (Cathy)   WANG

Kexian (Kay)   WU

Zhiyi   (Judy)   ZHU

Agenda

- Business Question & Objective

- EDAs

- Recommendation Methods

  - Clustering - PCA + KMeans

  - Popularity based recommendation for a song

  - Content based recommendation for a song and a playlist

  - Hybrid model for a playlist

- Experimentations

  - Rating

- Conclusion / Next Steps

- Q & A

- Appendix and references

# Business Question and Objective

- In this project, we aim to construct **recommender systems** for personalizing music suggestions based on the the **popularity, content of music, and user preferences**.
- **Clustering** algorithms analyze the traits of popular music, enabling us to recommend it more accurately to users.

Data Sources

- Analyzing 10,000 top songs (1960-2023) from ARIA and Billboard for **popularity and content-based filtering**.
  - 9,999 rows and 35 columns (track information and audio features)
- Using Music Playlist Data for **artist preference analysis**.
  - 12,891,680 rows and 4 columns (user id, artist, track name, playlist name)
  - Filtering the playlist dataset based on the unique artist names appearing in the top 10,000 songs.
- Fetching Cathy's Spotify playlist via the Spotify API, containing 16 songs.
- Fetching Judy's Spotify playlist via the Spotify API, including 17 songs.
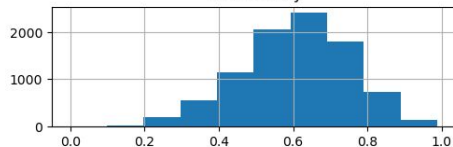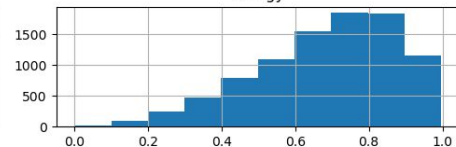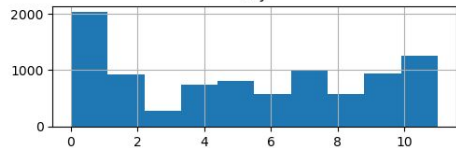
Exploratory Data Analysis
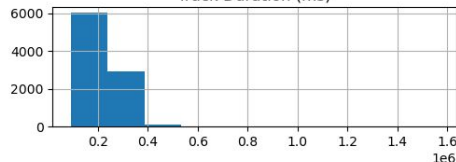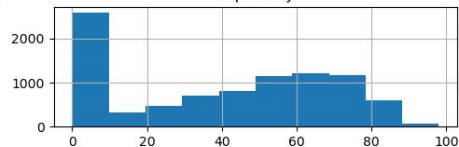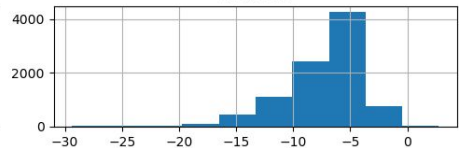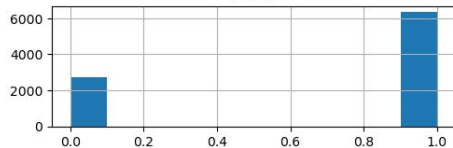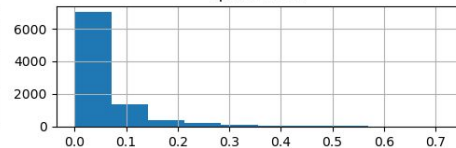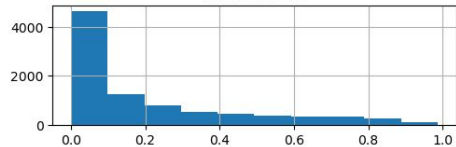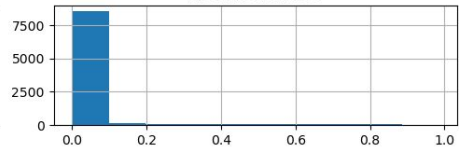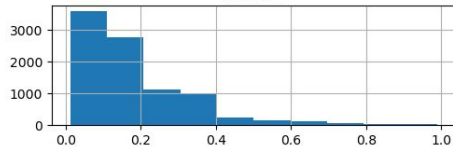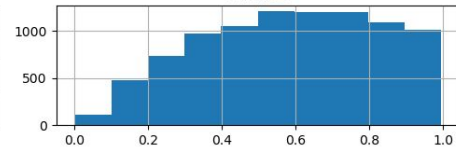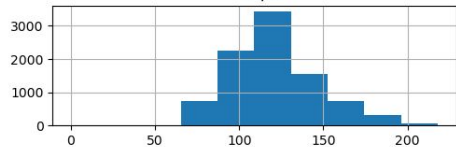
Exploratory Data Analysis
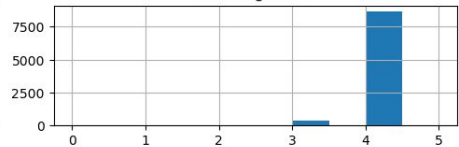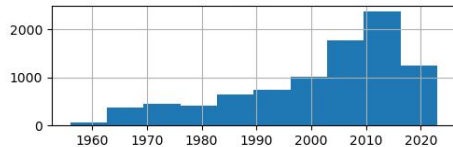
Correlation Heatmap of Columns
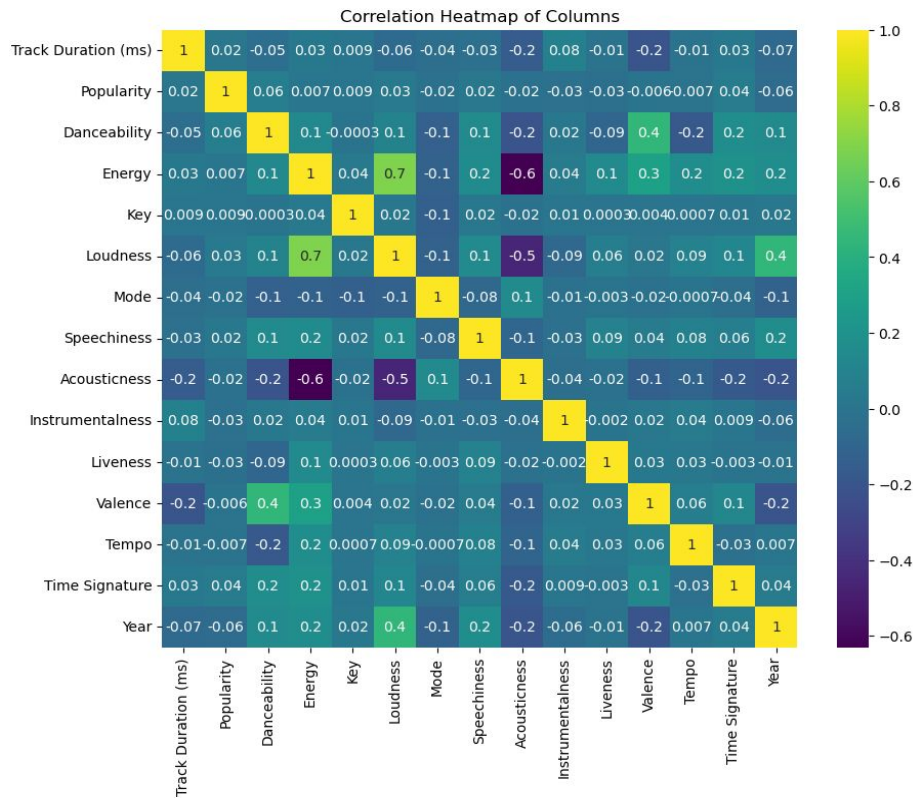
- **Energy** and **Loudness** have **high positive correlation**, as do **Acousticness** and **Instrumentalness**
- **Danceability** and **Valence** have **moderate positive correlation** (dance tracks perceived as happier)
- **Acousticness** has **high negative correlation** w/ **Energy**
- Several features (ex. Track Duration) have very low or no correlation with most other variables
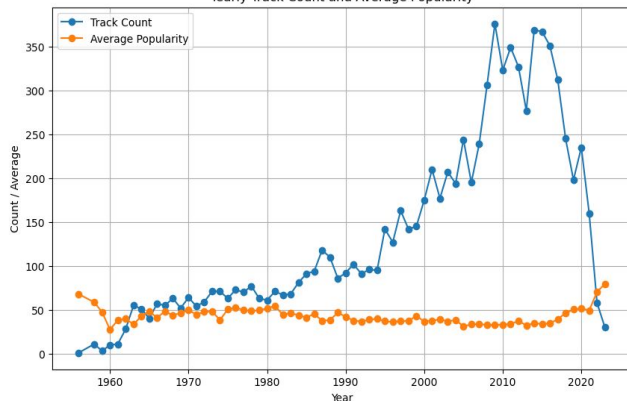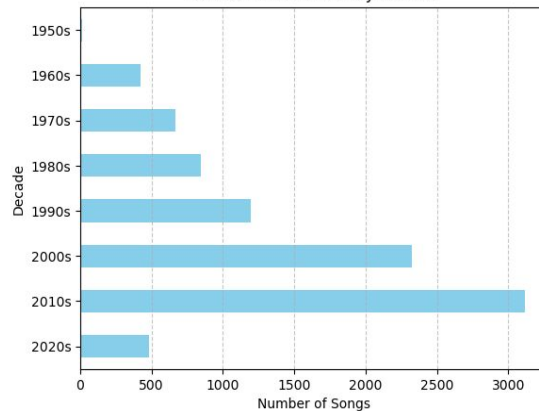
Exploratory Data Analysis
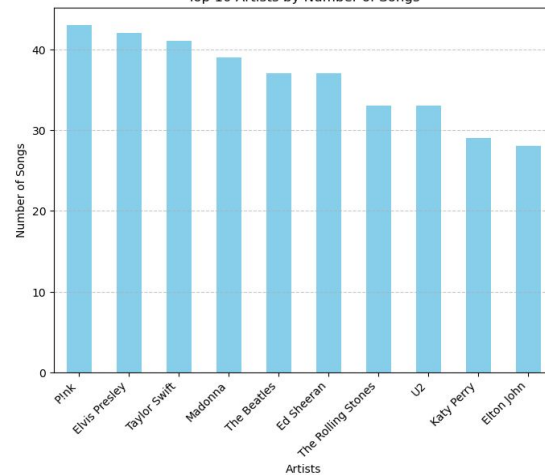
Refining Artist Genre

We refined the Artist Genres column to 25 main music genres, referring to
https://www.damvibes.com/music-theory/25-types-of-music-genres/ for guidance.

	Artist Genres	Artist Main Genres
0	pop	Pop
1	k-pop girl group	Pop
2	pop	Pop
3	pop	Pop
4	dance pop,pop,uk pop	Pop
5	big room,dance pop,edm,pop,pop dance,dance pop,pop	Pop
6	piano rock,pop	Pop
7	hip hop,rap,canadian contemporary r&b,canadian pop,pop,atl hip hop,hip hop,rap,trap	Hip Hop
8	canadian contemporary r&b,canadian pop,pop,electro,filter house,rock	Pop
9	afrobeats,nigerian pop,pop,post-teen pop	Pop

Top 10 Most Common Artist Genres
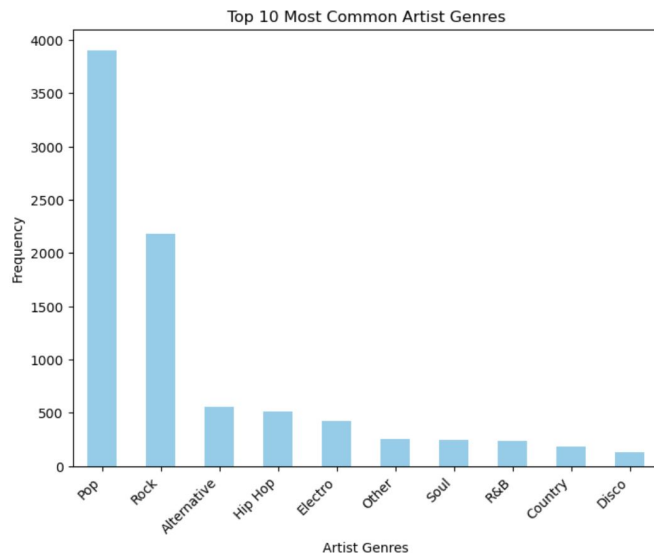
Exploratory Data Analysis
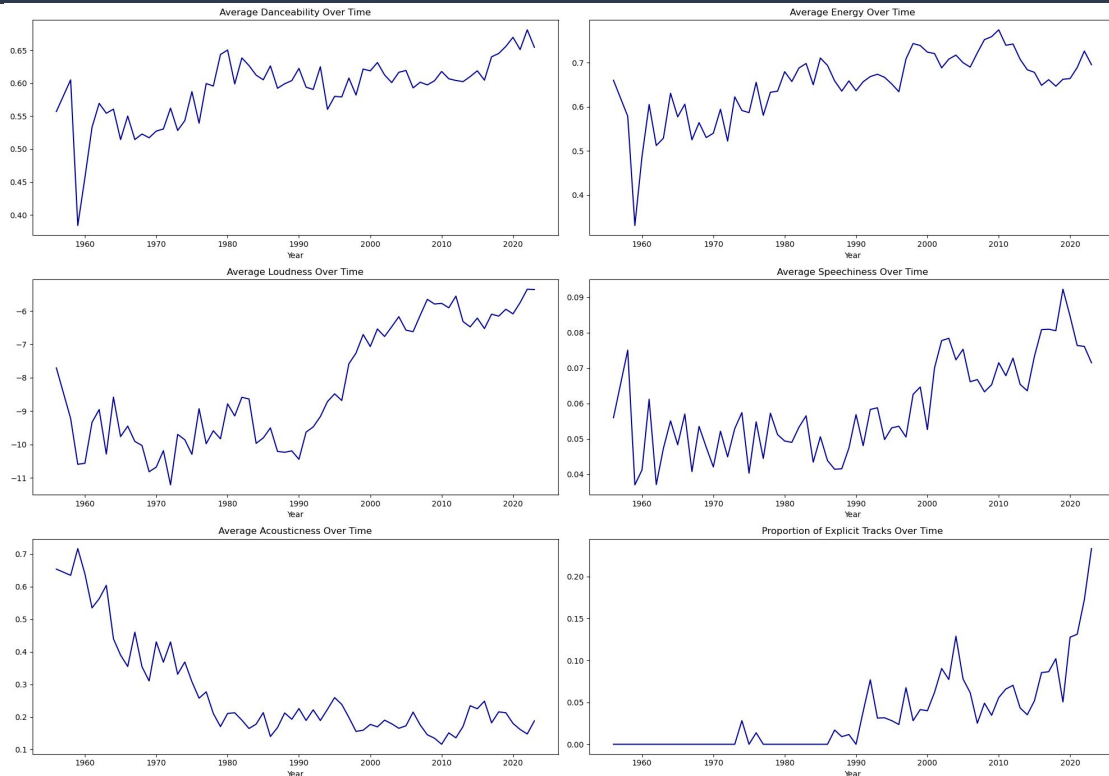
- General upward trend, especially from 1960s for **Danceability**
- General upward trend, with some fluctuations for **Energy**
- Clear increase, especially from the 1990s onwards for **Loudness**
- Varied significantly over time, peak in early 1960s, mid 1990s, and early 2020s for **Speechiness**
- Marked decline (shift towards electronic or processed sounds in music) for **Acousticness**
- Dramatic increase particularly from 2000s, peaked in 2020 for **Proportion of Explicit Tracks**

Word Map for 10,000 top songs (1960–2023)

Songs
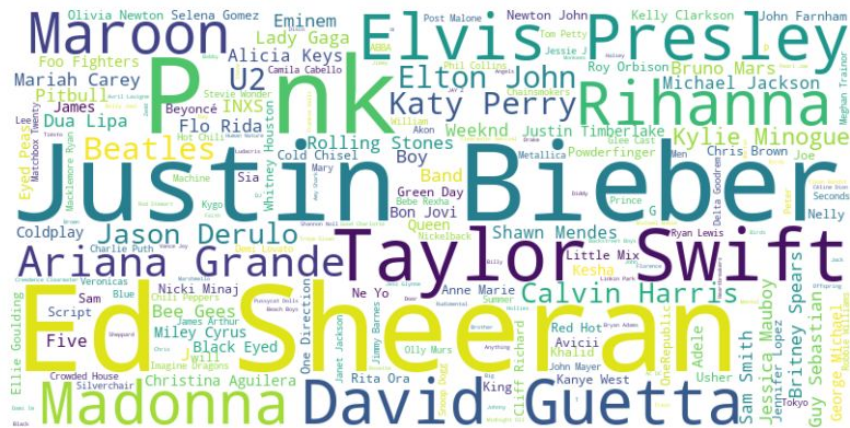
Artists

Recommendation Methods

Clustering: PCA + KMeans

Distortion Score Elbow for KMeans Clustering

2D Clustering of Categorical Data using t-SNE
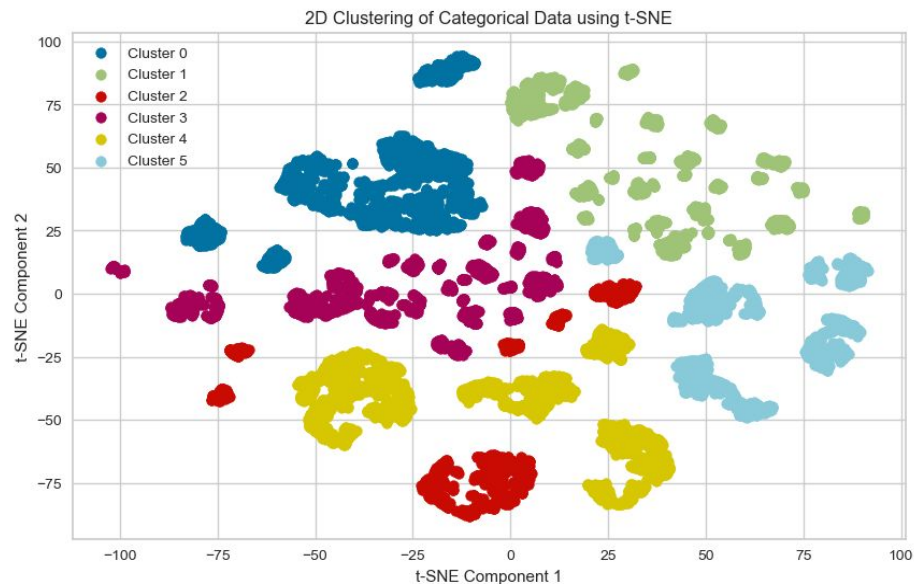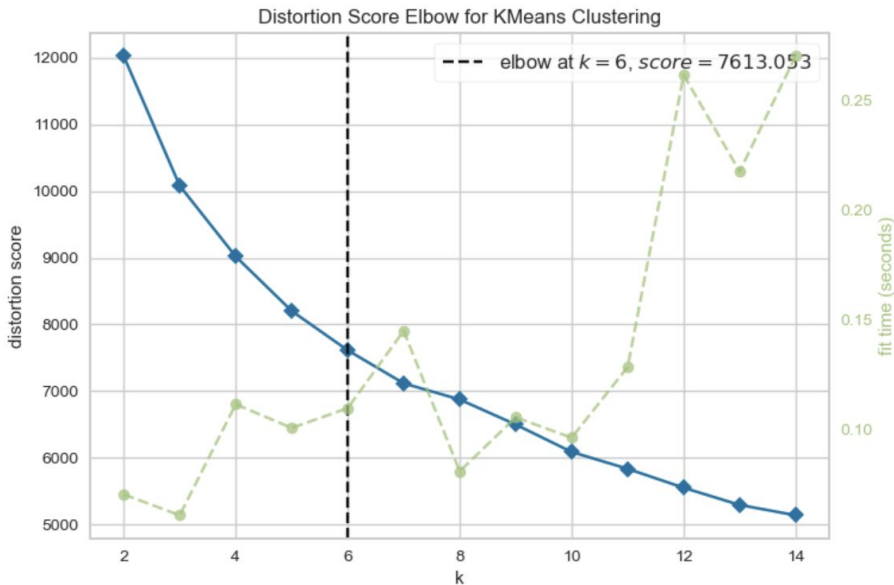
- Elbow Method: k=6

- A silhouette score near 0.3 indicates vague clusters

- But in music, genre labels may not always capture song characteristics accurately, leading to ambiguous clusters.

Clustering: PCA + KMeans

Cluster Characteristics
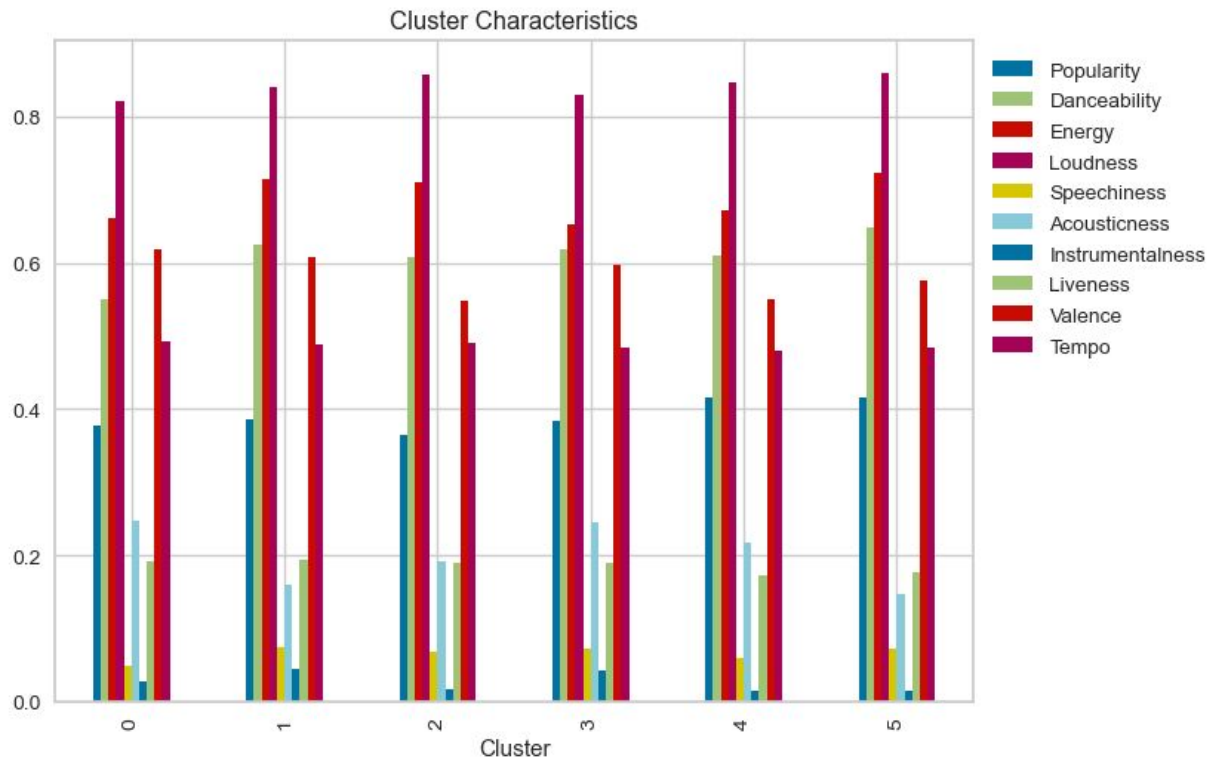
Similar patterns throughout all clusters

- High Loudness, Energy, Valence, Danceability
- Moderate Temp
- Lower Acousticness, Speechiness, Liveness, and Instrumentalness

Popularity Based Recommendation
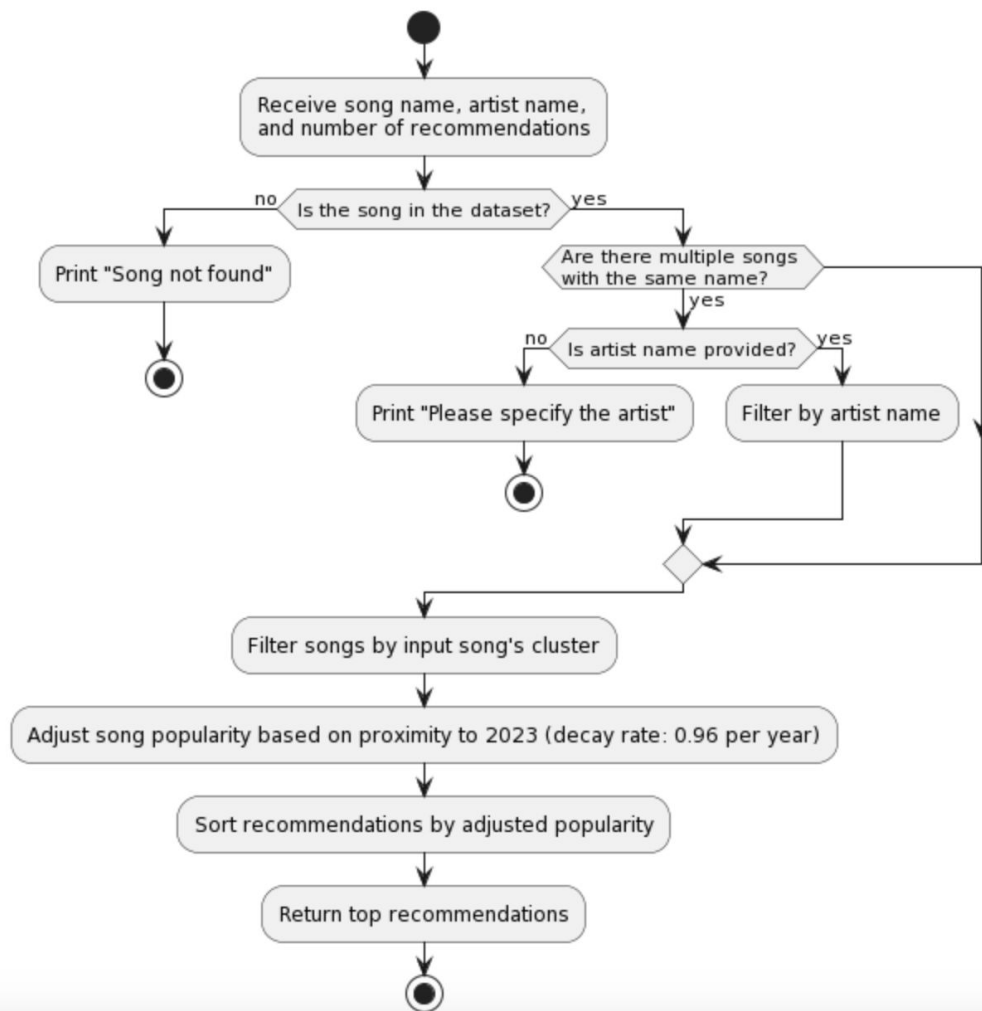
Popularity Based Recommendation for a Song

```
popularity_based_recommendations('Super Freaky Girl', num_recommendations=10)
```

	Track Name	Artist Name(s)	Album Name	Album Release Date	Popularity	Track URI	Cluster	Adjusted Popularity
0	vampire	Olivia Rodrigo	vampire	2023-06-30	0.98	https://open.spotify.com/track/3k79jB4aGmMDUQzEwa46Rz	4	0.9800
2	Flowers	Miley Cyrus	Flowers	2023-01-13	0.96	https://open.spotify.com/track/0yLdNVWF3Srea0uzk55zFn	4	0.9600
10	Baby Don't Hurt Me	David Guetta, Anne-Marie, Coi Leray	Baby Don't Hurt Me	2023-04-06	0.94	https://open.spotify.com/track/3BKD1PwArikchz2Zrlp1qi	4	0.9400
15	Miracle (with Ellie Goulding)	Calvin Harris, Ellie Goulding	Miracle (with Ellie Goulding)	2023-03-10	0.92	https://open.spotify.com/track/5eTaQYBE1yrActixMAeLcZ	4	0.9200
6	I Ain't Worried	OneRepublic	I Ain't Worried (Music From The Motion Picture "Top Gun: Maverick")	2022-05-13	0.95	https://open.spotify.com/track/4h9wh7iOZ0GGn8QVp4RAOB	4	0.9120
3	Anti-Hero	Taylor Swift	Midnights	2022-10-21	0.95	https://open.spotify.com/track/0V3wPSX9ygBnCm8psDlegu	4	0.9120
9	Calm Down (with Selena Gomez)	Rema, Selena Gomez	Calm Down (with Selena Gomez)	2022-08-25	0.94	https://open.spotify.com/track/0WtM2NBVQNNJLh6scP13H8	4	0.9024
39	Flowers	Miley Cyrus	Endless Summer Vacation	2023-03-10	0.90	https://open.spotify.com/track/4DHcnVTT87F0zZhRPYmZ3B	4	0.9000
13	Until I Found You (with Em Beihold) - Em Beihold Version	Stephen Sanchez, Em Beihold	Until I Found You (Em Beihold Version)	2022-04-22	0.93	https://open.spotify.com/track/1Y3LN4zO1Edc2EluIoSPJN	4	0.8928
33	Unholy (feat. Kim Petras)	Sam Smith, Kim Petras	Unholy (feat. Kim Petras)	2022-09-22	0.90	https://open.spotify.com/track/3nqQXoyQOWXiESFLlDF1hG	4	0.8640

Content Based Recommendation For a Song
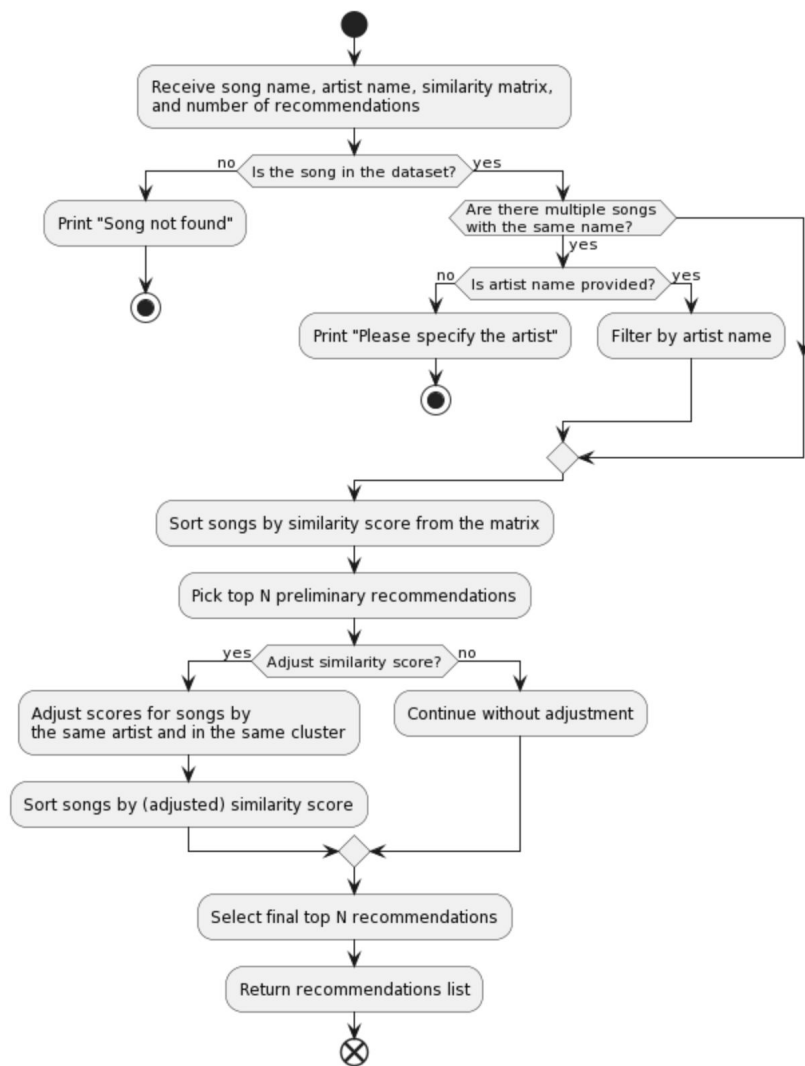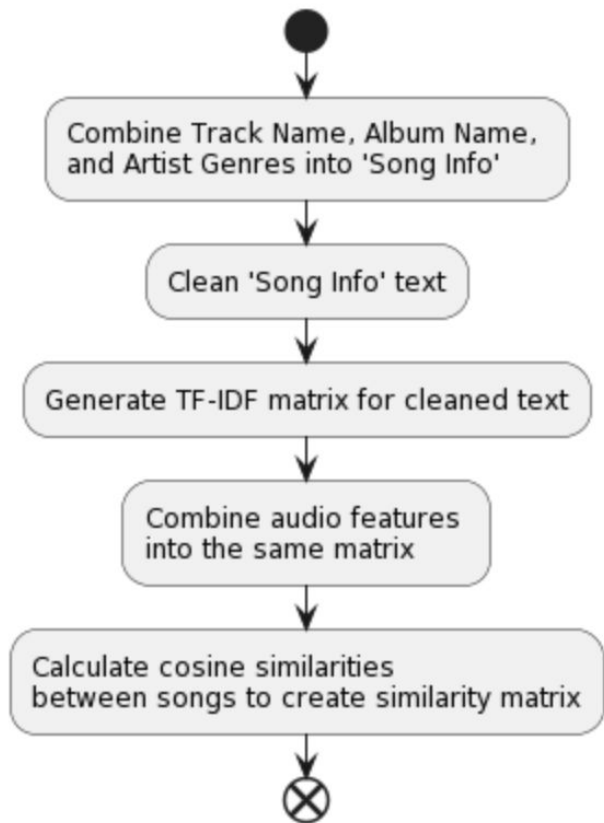
Content Based Recommendation For a Song

Still using *Super Freaky Girl* as an example:

	Track Name	Artist Name(s)	Album Name	Album Release Date	Popularity	Track URI	Artist Main Genres	Cluster	similar_score
1238	Freaky Friday (feat. Chris Brown)	Lil Dicky, Chris Brown	Freaky Friday (feat. Chris Brown)	2018-03-15	0.73	https://open.spotify.com/track/10lgtw8bSDyyFs7KlsKngZ	Hip Hop	3	0.858165
3273	Super Bass	Nicki Minaj	Pink Friday (Deluxe Edition)	2010-11-22	0.56	https://open.spotify.com/track/0tUVnDB0pM7xYGGgIKnyVf	Pop	2	0.852410
680	Mo Money Mo Problems (feat. Puff Daddy & Mase) - 2014 Remaster	The Notorious B.I.G., Mase, Diddy	Life After Death (2014 Remastered Edition)	1997-03-04	0.78	https://open.spotify.com/track/4lNDiWSKvqSKDEu7mh8HFz	Hip Hop	1	0.845333
1235	It's Tricky	Run–D.M.C.	Raising Hell	1986-05-15	0.73	https://open.spotify.com/track/6jBCehpNMkwFVF3dz4nLIW	Hip Hop	1	0.843399
1136	Shoop	Salt-N-Pepa	Very Necessary	1993-10-12	0.74	https://open.spotify.com/track/0Pu71wxadDlB8fJXfjljeJ	Hip Hop	3	0.842166
127	Hey Ya!	Outkast	Speakerboxxx/The Love Below	2003	0.86	https://open.spotify.com/track/2PpruBYCo4H7WOBJ7Q2EwM	Hip Hop	1	0.838101
3744	Always On Time	Ja Rule, Ashanti	Pain Is Love	2001-01-01	0.52	https://open.spotify.com/track/4hrae8atte6cRlSC9a7VCO	Hip Hop	1	0.836810
3535	Always On Time	Ja Rule, Ashanti	Pain Is Love	2001-10-02	0.54	https://open.spotify.com/track/39w4Xgs8FvY2SJmZsHKn2K	Hip Hop	1	0.835853
2658	Holidae In	Chingy, Ludacris, Snoop Dogg	Jackpot	2003-01-01	0.62	https://open.spotify.com/track/0FmCLEHflNGqwlyzvklNqy	Hip Hop	3	0.835521
376	P.I.M.P.	50 Cent	Get Rich Or Die Tryin'	2003-02-06	0.82	https://open.spotify.com/track/3G0yz3DZn3lfraledmBCT0	Hip Hop	1	0.832714

Content Based Playlist Recommendation

Receive playlist data, scaler, PCA, and column order

Create dummy variables for main genres, year ranges, and mode

Scale 'Popularity', 'Loudness', and 'Tempo'

Apply PCA to transform numeric columns to determine song cluster numbers

Calculate playlist numeric column averages

Assign cluster mode labels

Return playlist numeric averages

Receive playlist numeric averages, playlist data, 10,000 top songs, num recommendations

Compute cosine similarity between playlist numeric averages and top 10,000 songs

Sort songs by similarity score

Remove playlist songs from recommendations

Want all recommendations? — yes / no

Keep all recommendations

Select top N recommendations

Return recommendations

Valence	Tempo	...	Metal	Musical	Other	Pop	R&B	Reggae	Rock	Soul	Soundtrack	Cluster
0.520389	0.371377	...	0.0	0.0	0.0	0.777778	0.0	0.0	0.0	0.0	0.0	4

Content Based Recommendation For a Playlist

A playlist created for Judy:

	Track Name	Artist Name(s)	Album Name	Album Release Date	Popularity	Track URI	Artist Main Genres	Cluster	Similarity Score
2080	Girls Like (feat. Zara Larsson)	Tinie Tempah, Zara Larsson	Girls Like (feat. Zara Larsson)	2016-02-26	0.66	https://open.spotify.com/track/0rHVd8h0cFwgLgPFem6S60	Pop	4	0.932210
1236	Supalonely	BENEE, Gus Dapperton	STELLA & STEVE	2019-11-15	0.73	https://open.spotify.com/track/4nK5YrxbMGZstTLbvj6Gxw	Pop	4	0.931997
1567	Good Grief	Bastille	Wild World	2016-09-09	0.70	https://open.spotify.com/track/6IGA5YrFiEr5ANrMqKkR6H	Pop	4	0.931829
1019	Iko Iko (My Bestie)	Justin Wellington, Small Jam	Iko Iko (My Bestie)	2019-06-03	0.75	https://open.spotify.com/track/7MC4XR9M9amdmKQr2iDF7i	Pop	4	0.931634
786	All About That Bass	Meghan Trainor	Title (Deluxe)	2015-01-09	0.77	https://open.spotify.com/track/5jE48hhRu8E6zBDPRSkEq7	Pop	4	0.930724
1928	Rise	Jonas Blue, Jack & Jack	Rise	2018-05-25	0.68	https://open.spotify.com/track/3u1S1OmAUhx5DRlLrXqyp3	Pop	4	0.930542
1714	24K Magic	Bruno Mars	24K Magic	2016-10-07	0.69	https://open.spotify.com/track/2gFvRmQiWg9fN9i74Q0aiw	Pop	4	0.930010
1630	Easy Love	Sigala	Brighter Days	2018-09-28	0.70	https://open.spotify.com/track/5s7xgzXtmY4gMjeSlgisjy	Pop	4	0.929547
665	Better When I'm Dancin'	Meghan Trainor	Better When I'm Dancin'	2015-10-14	0.78	https://open.spotify.com/track/5k5fWendNngd89O8JKoE8L	Pop	4	0.929511
3736	Girls Talk Boys – From "Ghostbusters" Original Motion Picture Soundtrack	5 Seconds of Summer	Girls Talk Boys (From "Ghostbusters" Original Motion Picture Soundtrack)	2016-07-15	0.52	https://open.spotify.com/track/5XWmfpinsT4XbfFsUASSaG	Pop	4	0.929343

Hybrid Playlist Recommendations: Blending Popularity, Content, and User Preferences

Flowchart 1

- **(start)**
- Receive input artists from user's playlist, artist similarity matrix, and number of recommendations
- Loop through artists from user's playlist
- Artist exists in artist similarity matrix? — yes
- Add similarity scores to cumulative scores for each unlistened artist with respect to the artist
- Sort similar artists by cumulative similarity score in descending order
- Return top N most similar unlistened artists as recommendations
- **(end)**

Flowchart 2

- Receive input: playlist numeric averages, processed playlist data, processed 10,000 songs, artist similarity matrix, num recommendations (min 5)
- Generate content-based recommendations using playlist numeric averages
- Adjust song popularity based on proximity to 2023 (decay rate: 0.96 per year)
- Calculate hybrid score using 0.1 * adjusted popularity + 0.9 * content-based similarity score
- Split recommendations into two categories: with similar unlistened artists and others
- Select 80% recommendations from top recommendations by other artists
- Select remaining 20% recommendations from top recommendations by similar unlistened artists
- Return concatenated final recommendations sorted by hybrid score
- **(end)**

Diversity Artists & Songs!!!

Playlist Recommendation Experimentations

Testing Models with Manual Song Ratings

Judy's Playlist:

https://open.spotify.com/playlist/32nyghbn3sIw168FA7VKFT

Rating for Content-Based Model: 8.5/10

Rating for Hybrid Model: 9/10

Cathy's Playlist:

https://open.spotify.com/playlist/4mDJRCphx1MUd4Dfejj6tj
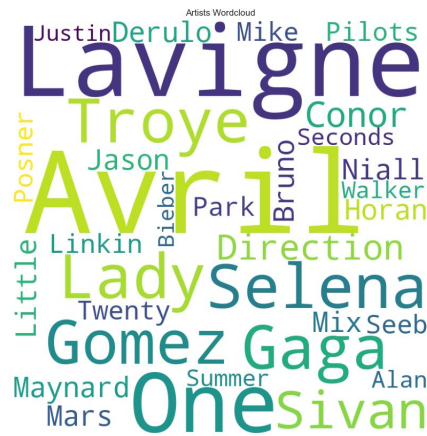
Rating for Content-Based Model: 9/10

Rating for Hybrid Model: 9/10

Top 10 Hybrid Recommendation For Judy

	Track Name	Artist Name(s)	Album Name	Album Release Date	Popularity	Adjusted Popularity	Similarity Score	Artist Main Genres	Cluster	Hybrid Score	Recommendation Type
1019	Iko Iko (My Bestie)	Justin Wellington, Small Jam	Iko Iko (My Bestie)	2019-06-03	0.75	0.637010	0.931634	Pop	4	0.902172	Regular
1236	Supalonely	BENEE, Gus Dapperton	STELLA & STEVE	2019-11-15	0.73	0.620023	0.931997	Pop	4	0.900800	Regular
1630	Easy Love	Sigala	Brighter Days	2018-09-28	0.70	0.570761	0.929547	Pop	4	0.893668	Regular
1398	Never Really Over	Katy Perry	Never Really Over	2019-05-31	0.72	0.611530	0.924552	Pop	4	0.893250	Ad Hoc
786	All About That Bass	Meghan Trainor	Title (Deluxe)	2015-01-09	0.77	0.555470	0.930724	Pop	4	0.893199	Regular
1928	Rise	Jonas Blue, Jack & Jack	Rise	2018-05-25	0.68	0.554453	0.930542	Pop	4	0.892933	Regular
1567	Good Grief	Bastille	Wild World	2016-09-09	0.70	0.526013	0.931829	Pop	4	0.891247	Regular
1714	24K Magic	Bruno Mars	24K Magic	2016-10-07	0.69	0.518499	0.930010	Pop	4	0.888858	Regular
2080	Girls Like (feat. Zara Larsson)	Tinie Tempah, Zara Larsson	Girls Like (feat. Zara Larsson)	2016-02-26	0.66	0.495955	0.932210	Pop	4	0.888585	Regular
3186	Smile	Katy Perry	Smile	2020-07-10	0.57	0.504300	0.906419	Pop	4	0.866207	Ad Hoc

Top 10 Hybrid Recommendation For Cathy

	Track Name	Artist Name(s)	Album Name	Album Release Date	Popularity	Adjusted Popularity	Similarity Score	Artist Main Genres	Cluster	Hybrid Score	Recommendation Type
418	Shotgun	George Ezra	Staying at Tamara's	2018-03-23	0.81	0.660452	0.931633	Pop	4	0.904515	Regular
1398	Never Really Over	Katy Perry	Never Really Over	2019-05-31	0.72	0.611530	0.933540	Pop	4	0.901339	Ad Hoc
1117	Rewrite The Stars	James Arthur, Anne-Marie	The Greatest Showman: Reimagined (Deluxe)	2017-12-08	0.74	0.579241	0.932559	Pop	4	0.897227	Regular
1416	Born To Be Yours	Kygo, Imagine Dragons	Born To Be Yours	2018-06-15	0.72	0.587068	0.931197	Pop	4	0.896784	Regular
665	Better When I'm Dancin'	Meghan Trainor	Better When I'm Dancin'	2015-10-14	0.78	0.562684	0.932853	Pop	4	0.895836	Regular
1928	Rise	Jonas Blue, Jack & Jack	Rise	2018-05-25	0.68	0.554453	0.932630	Pop	4	0.894812	Regular
2423	Used To Love (with Dean Lewis)	Martin Garrix, Dean Lewis	Used To Love (with Dean Lewis)	2019-10-31	0.64	0.543582	0.931431	Pop	4	0.892646	Regular
2080	Girls Like (feat. Zara Larsson)	Tinie Tempah, Zara Larsson	Girls Like (feat. Zara Larsson)	2016-02-26	0.66	0.495955	0.935064	Pop	4	0.891153	Regular
2406	Here for You (feat. Ella Henderson)	Kygo, Ella Henderson	Here for You (feat. Ella Henderson)	2015-09-04	0.64	0.461689	0.931199	Pop	4	0.884248	Regular
1160	Harleys In Hawaii	Katy Perry	Harleys In Hawaii	2019-10-16	0.74	0.628516	0.891562	Pop	5	0.865257	Ad Hoc

Next Steps

- Build an interactive website that collects user preferences

- Expand data sources to include more recent songs and artists

- Collect more recent playlist data

- Build test metrics and fine-tune for better models

Q & A

Appendix

- https://www.kaggle.com/datasets/andrewmvd/spotify-playlists

- https://www.kaggle.com/datasets/joebeachcapital/top-10000-spotify-songs-1960-now/data

- https://www.damvibes.com/music-theory/25-types-of-music-genres/

Reference

Anderson, A. *et al.* (2020) 'Algorithmic effects on the diversity of consumption on Spotify', *Proceedings of The Web Conference 2020* [Preprint]. doi:10.1145/3366423.3380281.