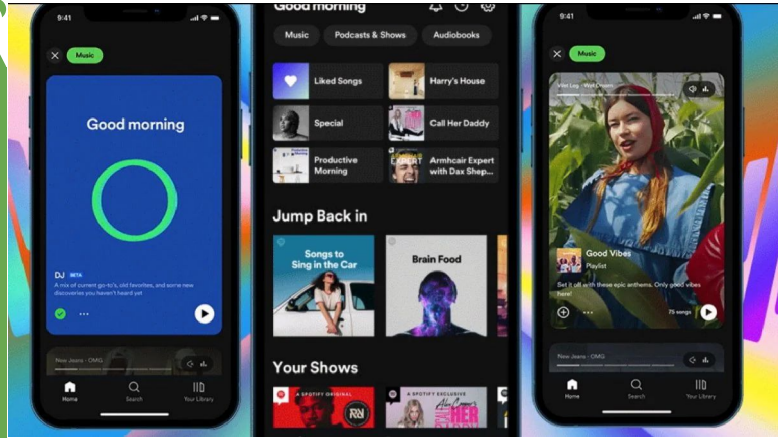
Spotify Data Mining

Kira Luo, Linda Liang, Rohit Sharma, Sameera Boppana

Table of contents



Project Objective and Dataset Selection

- This dataset contains all songs available on Spotify from 1921 - 2020

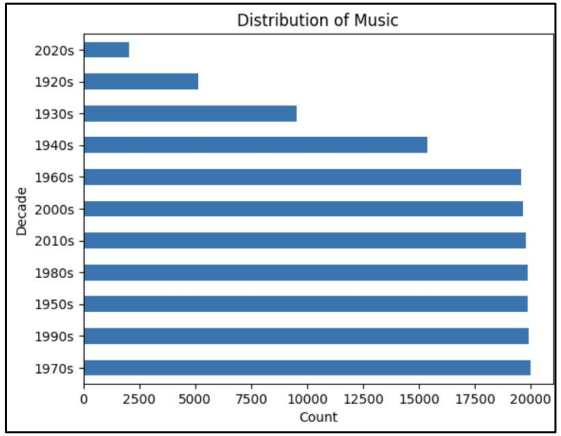
- Individual Song Information:
  - Artists
  - Genres
  - Year
  - Song Characteristics (Acousticness, Energy, Valence etc)

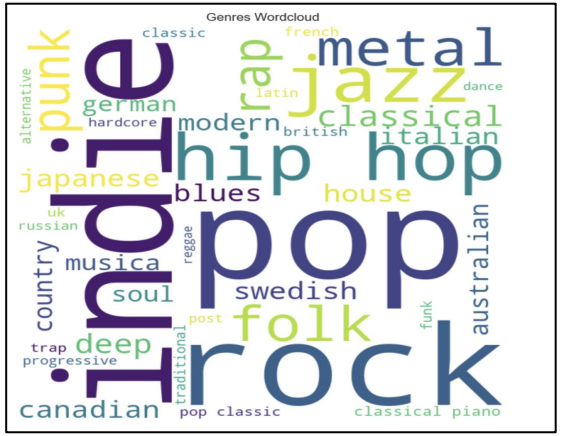
- Aggregated Data
  - Information regarding music characteristics grouped by artists, genre and year

- Want to explore songs and characteristics that are similar to each other and be able to recommend similar songs and artists

- This project aims to enhance user experience, increase engagement and loyalty of Spotify users

EDA



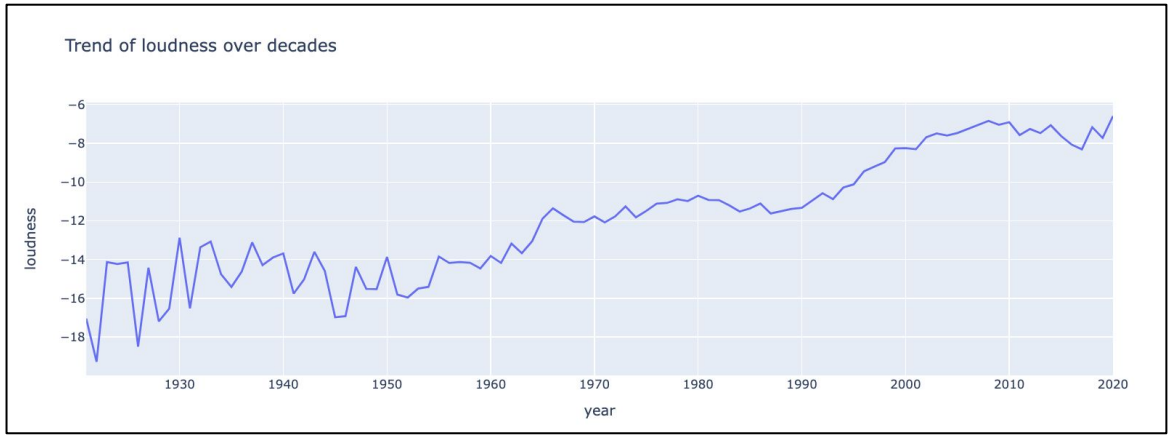


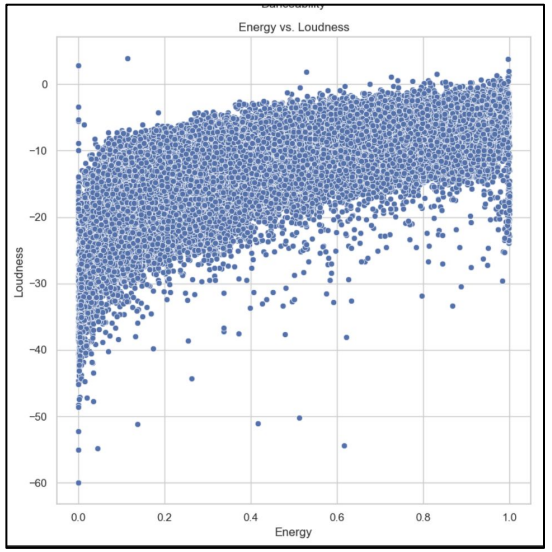
170,653 Songs

28,680 Artists

2,973 Genres

EDA





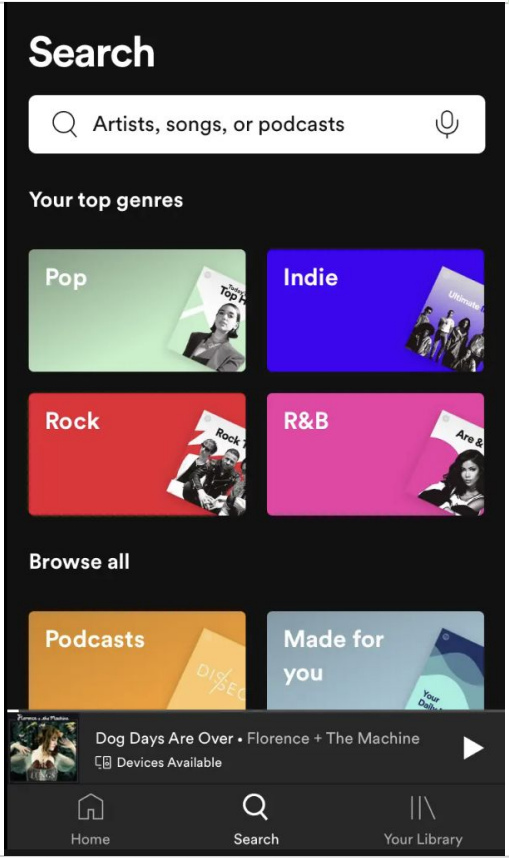
Over the years, there has been a significant increase in song loudness, correlating with higher energy levels in music

Business Problems

1. How can we categorize our user base into distinct segments based on their listening habits to tailor marketing strategies?

2. Which artists or songs serve as bridges in the listening habits of our users, connecting different music genres or communities?

3. How can we use content-based recommender method to to enhance user satisfaction?



Solution 1: K-Means Clustering

1. We plan to perform K- means clustering on **Genre** so we could better know the characteristics of each genre and recommend users based on user's genre preferences.

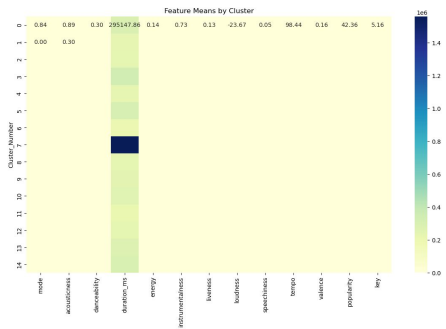
2. Dimensional Reduction using T-SNE and UMAP

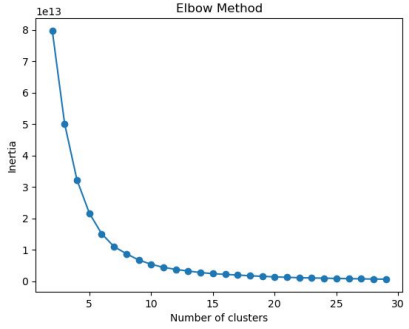
3. Select k=15 using Elbow Method

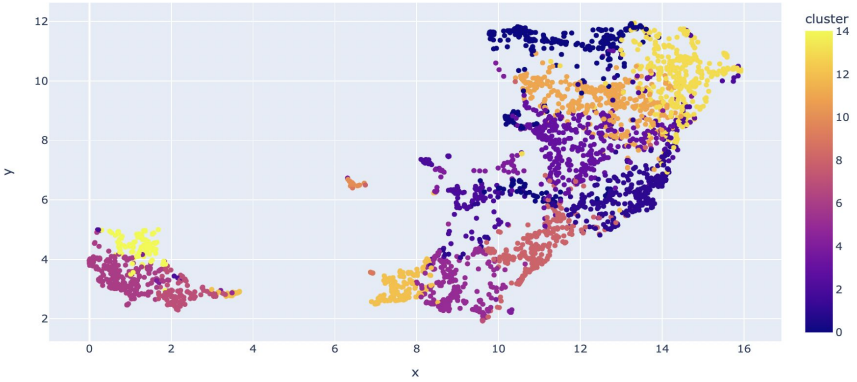
4. Visualize the dataset (UMAP has a better visualization)

5. Explore characteristics for each genre cluster and find most popular genres according to popularity scores

Key Takeaway: Top 5 popular clusters are cluster 0, 5, 8, 9, and 6 and their common characteristics are **high energy and high tempo.**



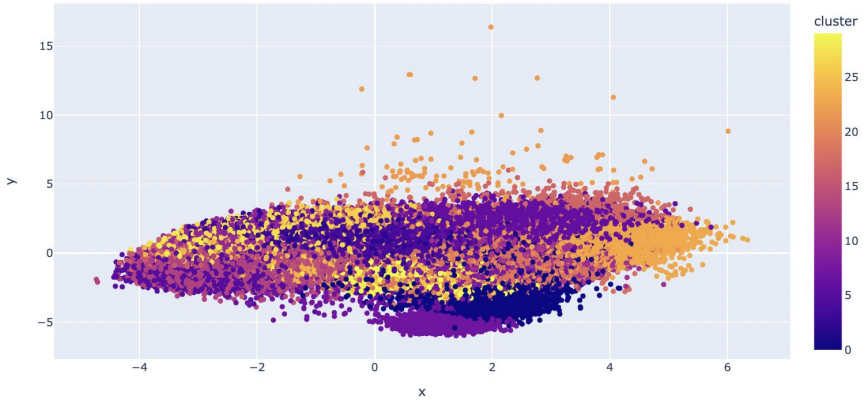




Solution 1: K-Means Clustering

We performed PCA on the **Song** dataset and performed K-means clustering with 30 clusters.

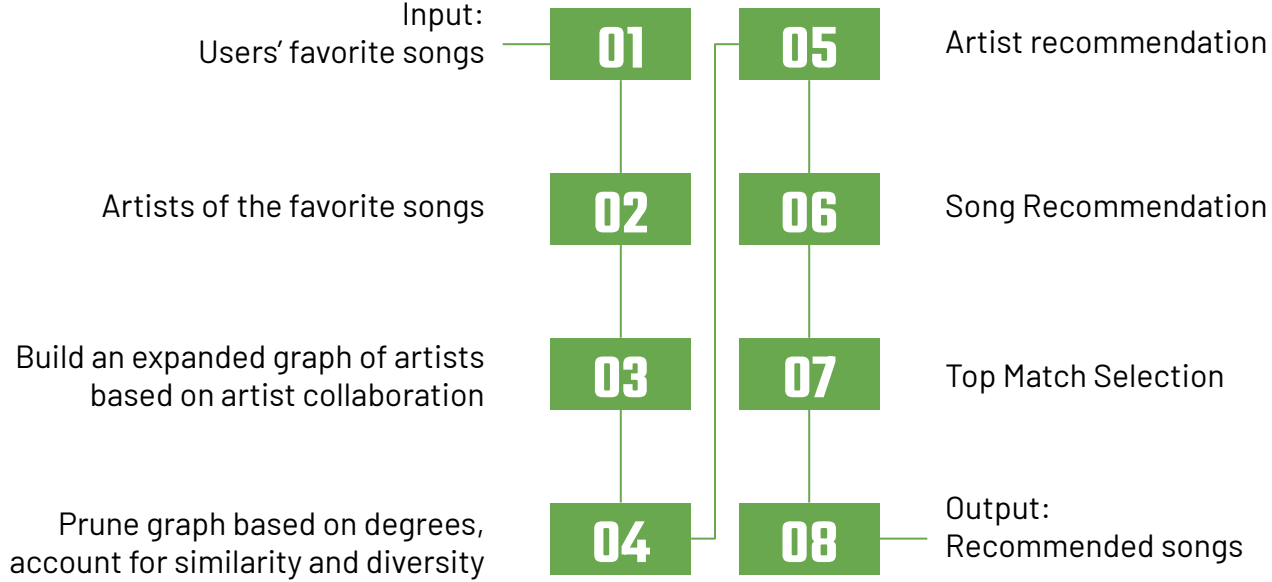
Within each cluster, we found the most popular song and artists, which means that we possibly could match users with each song cluster and **recommend users with the most popular song or artist in that cluster**.



	Cluster_Number	Average_Popularity	Name
0	9	55.542766	Lonely
1	3	54.369233	Intro
2	18	53.164604	Limbo
3	27	47.561060	Without You
4	11	47.489499	Come As You Are
5	22	46.475501	Eye of the Tiger
6	1	44.454705	Have Yourself a Merry Little Christmas
7	20	44.390575	Hello
8	17	44.300917	Main Title
9	2	43.443308	A Ella
10	24	42.574241	Jingle Bell Rock

	Cluster_Number	Average_Popularity	artists
0	9	55.542766	['Eminem']
1	3	54.369233	['Eminem']
2	18	53.164604	['Taylor Swift']
3	27	47.561060	['Queen']
4	11	47.489499	['Metallica']
5	22	46.475501	['BTS']
6	1	44.454705	['Vicente Fernández']
7	20	44.390575	['Bob Marley & The Wailers']
8	17	44.300917	['John Williams']
9	2	43.443308	['Ramones']
10	24	42.574241	['The Rolling Stones']

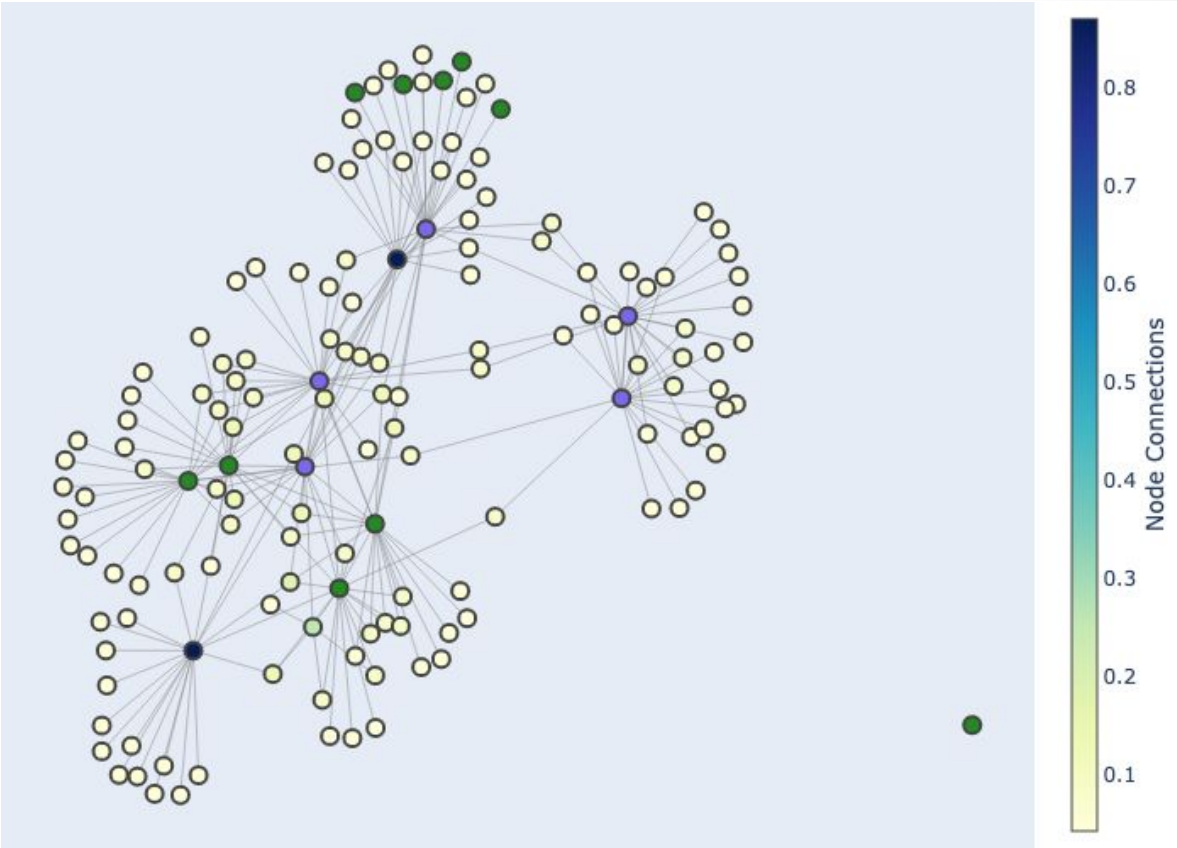
Workflow


Input:
Users' favorite songs
01
Artists of the favorite songs
02
Build an expanded graph of artists based on artist collaboration
03
Prune graph based on degrees, account for similarity and diversity
04
05
Artist recommendation
06
Song Recommendation
07
Top Match Selection
08
Output:
Recommended songs

Solution 2: Graph Mining

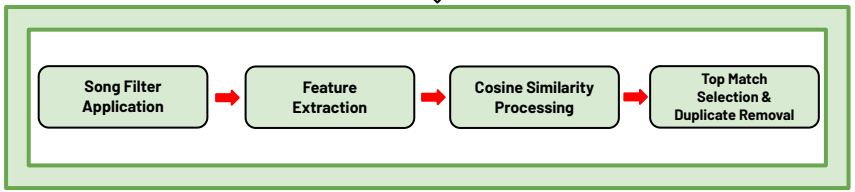
Olivia's Favorite Songs:
'You Broke Me First' by Tate McRae
'Watermelon Sugar' by Harry Styles
'Heather' by Conan Gray
'Blinding Lights' by The Weekend
'Circles' by Post Malone

Recommended Artists:
Anna Clendening
Louis Tomlinson
Niall Horan
Liam Payne
One Direction
ZAYN
Madison Beer
Alec Benjamin
Sabrina Carpenter
Gracie Abrams



Solution 3: Song Recommender System

```
Recommended Artists: ['Anna Clendening', 'Louis Tomlinson', 'Niall Horan', 'Liam Payne', 'One Direction', 'ZAYN',
'Madison Beer', 'Alec Benjamin', 'Sabrina Carpenter', 'Gracie Abrams']
```



	name	artists	year
0	The Code (feat. Polo G)	King Von	2020
1	No Se Me Quita (feat. Ricky Martin)	Maluma	2019
2	Solo Quédate En Silencio	Maite Perroni	2004
3	Headlines	Drake	2011
4	BROWN SKIN GIRL	SAINt JHN	2019

RECOMMENDED ARTISTS

RECOMMENDER SYSTEM

FINAL SONG RECOMMENDATIONS

Conclusions

Improved Customer Segmentation

K-means Clustering helps in identifying distinct groups of users based on their music preferences. This segmentation enables targeted marketing strategies, personalized user experiences, and efficient resource allocation

Improving Personalization

A hybrid system combines multiple data sources and recommendation techniques to provide highly personalized music suggestions. This addresses the challenge of meeting diverse user preferences and adapting to their evolving tastes.

Discovering Influential Content

By analyzing the graph of user interactions with songs and artists, Spotify can identify trending music, emerging artists, and viral content early on. This enables Spotify to promote content that is likely to perform well, optimizing its catalog's appeal

Enhancing Customer Satisfaction

Overall, these techniques provide a superior listening experience through personalized, diverse, and socially connected features encourages user loyalty, which is crucial for long-term success in the competitive streaming market.

Thank you for listening