Spotify Recommender

Dylan Cavaliere

- Agenda
- Problem Statement
- Data Overview
- Preferences
- Feature Explore
- Modeling
- Next Steps
- App Demo
- Thank You

Agenda

Problem Statement

Data Overview & EDA

Model Results

Conclusion & App Demo

Problem Statement

Build a recommendation engine

I am a Data Scientist working on the Spotify Team. My job entails surfacing recommendations after a user searches and listens to a specific song.

Process Overview

Collect
Spotify API

Prepare
Clean and prep the data for modeling

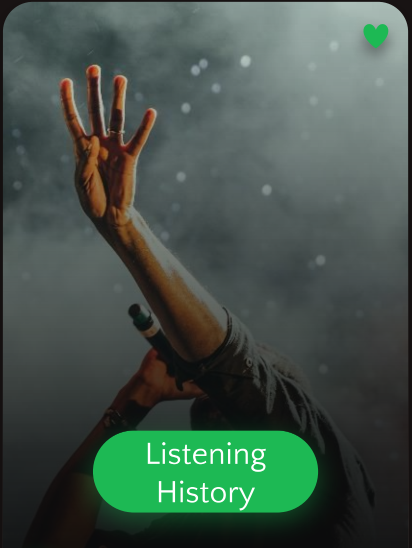
Analyze
Review Features & LIstening History

Model
Build, Tune & Evaluate Recommendations

Data Overview
& EDA

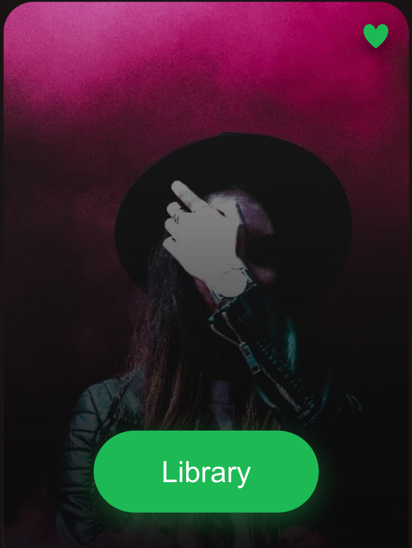
- Agenda
- Problem Statement
- **Data Overview**
- Preferences
- Feature Explore
- Modeling
- Next Steps
- App Demo
- Thank You

Data Summary



Listening History

- Songs: 12,323
- Artists: 2,975
- Genres: 724



Library

- Songs: 170,653
- Artists: 20,632
- Genres: 1,767

Data Sources

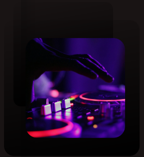
Kaggle –
List of Songs CSV

Spotify Extended Listening History JSON

Spotify API

- Agenda
- Problem Statement
- Data Overview
- **Preferences**
- Feature Explore
- Modeling
- Next Steps
- App Demo
- Thank You

Listening Preferences

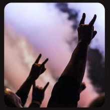


2022

Top Artist: Beyonce
Top Song: CRP- Summer Walker

2021

Top Artist: Giveon
Top Song: Favorite Mistake - Giveon



2020

Top Artist: Pinegrove
Top Song: The Alarmist - Pinegrove

2019

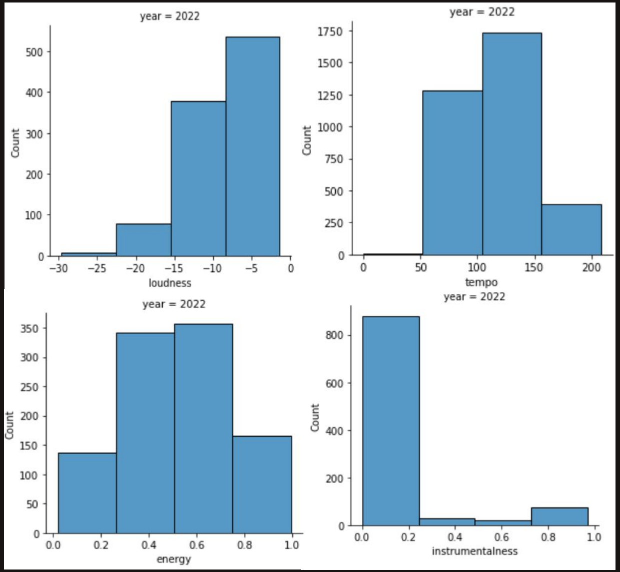
Top Artist: Daniel Caesar
Top Song: Omen - Sam Smith

Genre	Count
Rock	35,182
Pop	32,930
Hip-Hop	24,854
Blues	19,171
Soul	15,106
Indie	11,207
Rap	9,480
Contemporary	7,446
Alternative	6,657

Exploring Features

Artist:
Genre, Popularity, Followers

Track:
Energy, Danceability, Acousticness, Liveness, Valence, Tempo, Loudness, Instrumentalness, Mode, Key, Speechiness



While I had a wide variation in artists and songs across years, the distribution of track features remained fairly consistent YoY

Most songs I listen to are mid-energy, non-instrumental, and relatively loud.

Modeling

Model Overview

Two Models

Model 1 – Cosine Similarity – Eleven Track Feature + Artist Popularity & Followers.

Model 2 – Cosine Similarity – Same Features as Model 1 + Countvectorizer on Artist Genre

*Both Models filter out songs from the same artist of the input song

Model Evaluation

1) Selected samples from my top played tracks
2) Generated top 5 recommendations from each model
3) Listen to recommendations and rank best matches

Model 1

Song	Genre	Dance	Acoustic	Energy	Loudness	Liveness	Tempo
*CPR	r&b	.693	.701	.437	-11.56	.123	83.17
Fall In Love To Easy (5)	lofi	.662	.637	.349	-14.02	.124	141.95
Scared (4th)	hawaiian rap	.597	.652	.418	-10.03	.120	129.95
Where this flower bloom (6th)	hip-hop	.686	.710	.429	-10.40	.166	130.09

Model 2

Song	Genre	Dance	Acoustic	Energy	Loudness	Liveness	Tempo
*CPR	r&b	.693	.701	.437	-11.56	.123	83.17
Yanghwa BRDG (3rd)	k-pop	.538	.726	.436	-8.73	.151	87.11
I Hate U (1st)	r&b	.535	.507	.388	-9.80	.11	106.70
Drew Barrymore (2nd)	r&b	.577	.491	.523	-5.53	.147	134.93

Model 1

Song	Genre	Dance	Acoustic	Energy	Loudness	Liveness	Tempo
*Lose Control	edm	.598	.129	.526	-8.69	.14	123.93
Spellbinder	rock	.603	.137	.529	-7.393	.11	188.52
Pointblank	rock	.479	.120	.423	-13.126	.10	116.06
Die For You	pop	.586	.111	.525	-7.163	.13	133.629

Model 2

Song	Genre	Dance	Acoustic	Energy	Loudness	Liveness	Tempo
*Lose Control	edm	.598	.129	.526	-8.69	.14	123.93
More Than You Know (2)	edm	.646	.0275	.741	-4.97	.31	123.07
Ride It (1)	edm	.880	.177	.751	-4.25	.10	117.94
Leave the World Behind (3)	edm	.525	.002	.849	-7.72	.32	128.05

Conclusion & Next Steps

Model 2 – incorporating genres results in better recommendations

The recommendation engine can successfully take in an artist and song name and return fairly useful recommendations however with additional time more improvements can be made.

Next Steps

KNN
More Models

Expand the
Library

New Features
Audio

Genre
Likeness

User
Feedback

Streamlit Demo

Thank You