Recommendation System Using Spotify API

JUSTIN GRISANTI

Objectives – 5 Step Model

OBTAIN BUSINESS UNDERSTANDING

UNDERSTAND THE DATA

PREPARE DATA FOR MODELING

CREATE MODELS

GENERATE RESULTS

Business Understanding

Business Problem

- Easier to access music
- Millions of available songs
- Paradox of Choice

Business Understanding

Stakeholders

- Spotify Users
- DJs
- Other Music-Related professionals

Solution

- Content-Based Recommendation System
- Helps a user find similar music

Data Understanding

- **Spotify Million Playlist Dataset**
  - 1m Playlists
  - 2.2m Unique Songs
  - 734k Unique Albums
  - 295k Unique Artists
- Selected top 100 playlists (number of followers) from 200k playlists

Data Understanding

- Spotify Library: Spotipy
  - A lightweight Python library for the Spotify Web API. With Spotipy you get full access to all the music data provided by the Spotify platform.
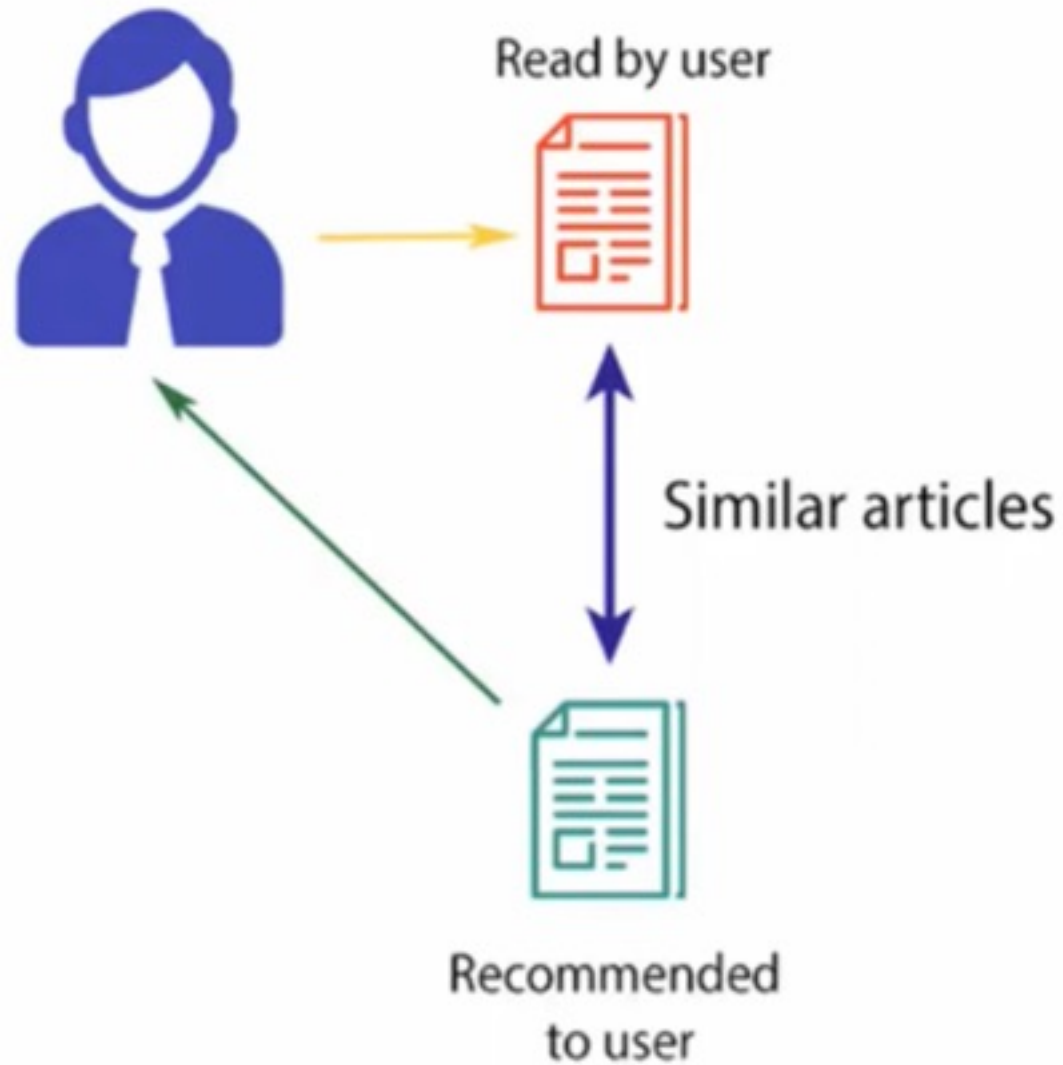
Data Preparation

Playlist/Track Vectors

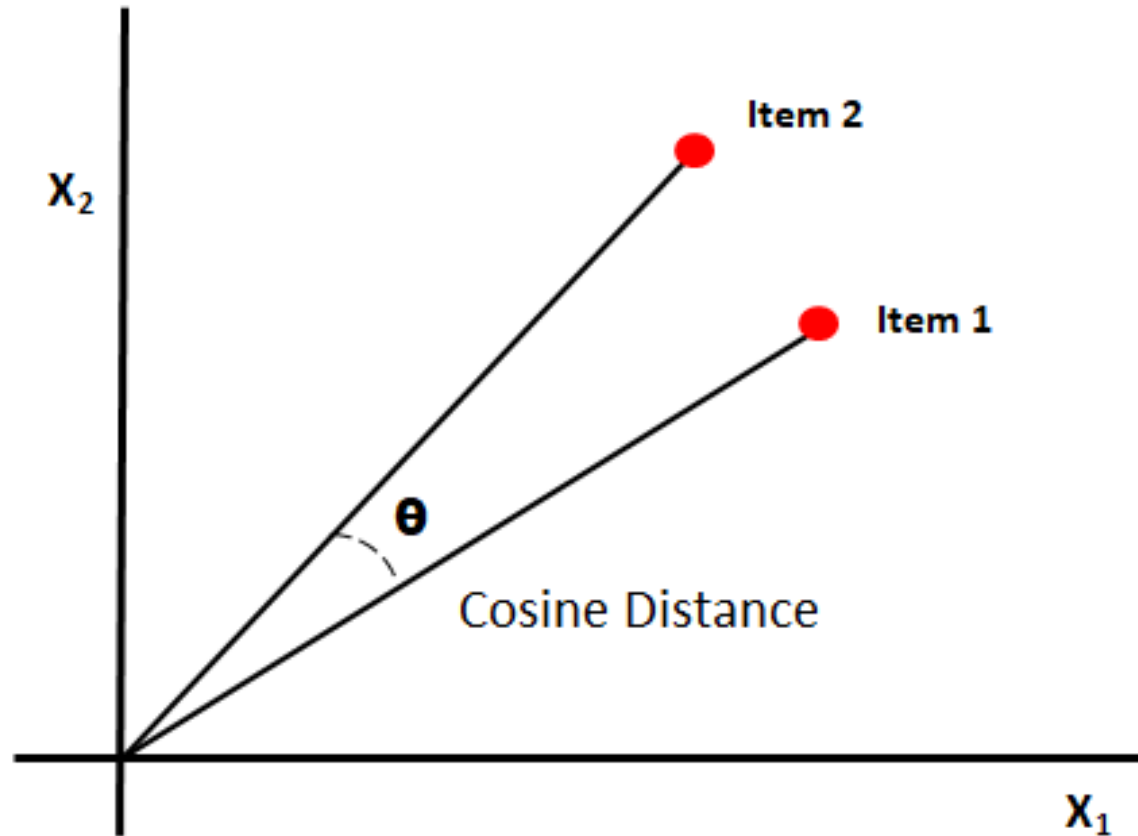- Audio Features
- Popularity
- Genre

Data Preparation

Audio Features

- Valence (Positive)
- Instrumentalness
- Energy
- Etc.

Modeling: Content-Based Filtering

Cosine Distance/Similarity

Item 2

Item 1

X_2

θ

Cosine Distance

X_1

Metric: Cosine Similarity

Modeling: Playlist #102585

pos	track_uri	track_name	artist_name	album_name
0	67WTwafOMgegV6ABnBQxcE	Some Nights	fun.	Some Nights
1	6Ep6BzlOB9tz3P4sWqiiAB	Radioactive	Imagine Dragons	Night Visions
2	4wCmqSrbyCgxEXROQE6vtV	Somebody That I Used To Know	Gotye	Making Mirrors
3	5j9iuo3tMmQlfnEEQOOjxh	Best Day Of My Life	American Authors	Oh, What A Life
4	6cpk00i5TxCqSeqNi2Hule	One More Night	Maroon 5	Overexposed Track By Track
5	5edBgVtRD0fvWk140Sl21T	Counting Stars	OneRepublic	Native
6	1jdNcAD8lr58RlsdGjJJdx	Ho Hey	The Lumineers	The Lumineers
7	7zWj09xkFgA9tcV6YhfU6q	Drive By	Train	California 37
8	6g1NlCpW7fgqDnWbCCDrHl	Wake Me Up - Radio Edit	Avicii	True
9	5rgy6ghBq1eRApCkeUdJXf	We Are Young (feat. Janelle Monáe) - feat. Jan...	fun.	Some Nights

Modeling: Playlist #102585

Song Genre in Playlist (Most Common)
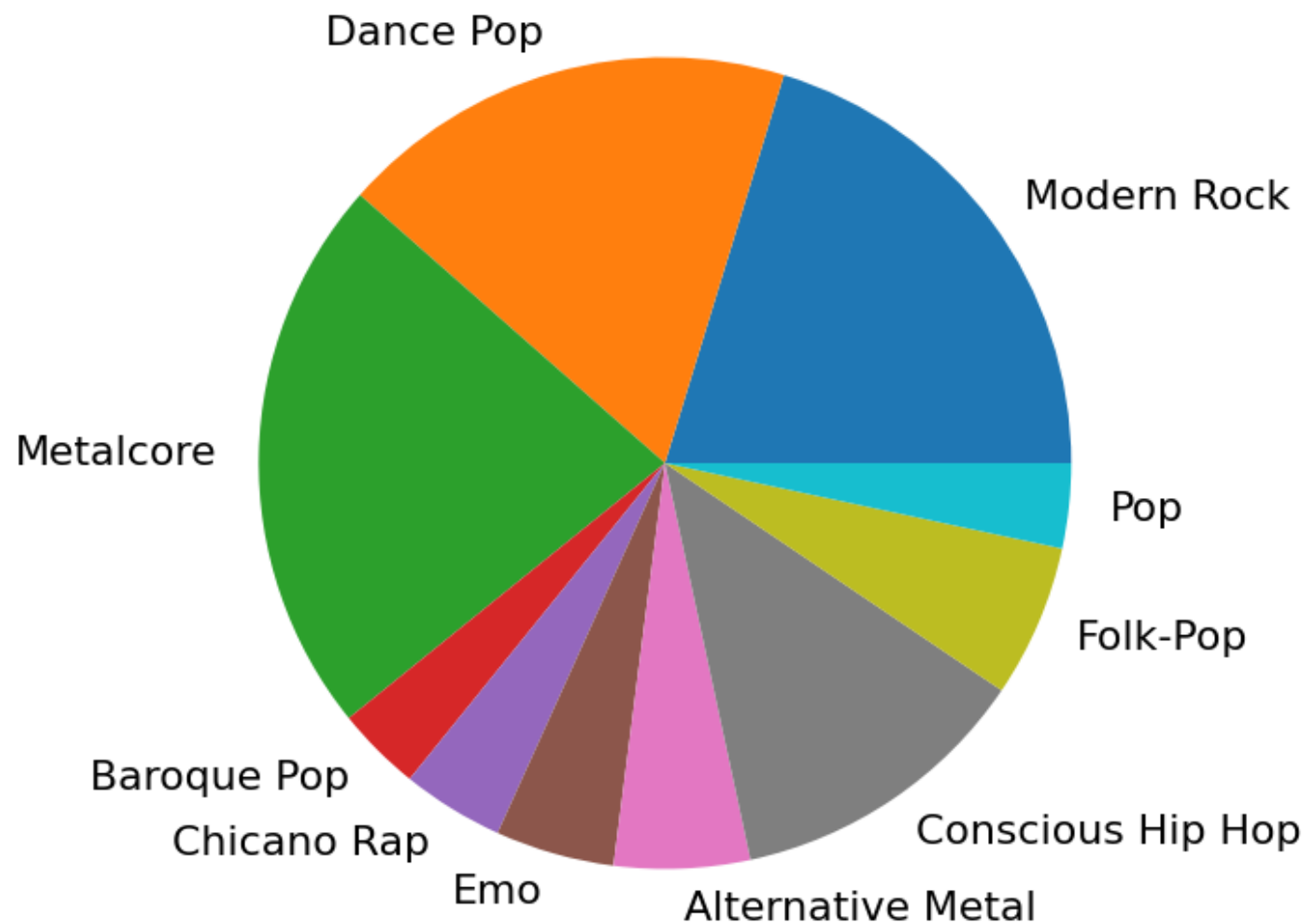
Modeling: Playlist #102585

	track_uri	track_name	artist_name	album_name
0	3s5ogvexUgA6XjNj37zpnP	Piledriver waltz	Alex Turner	Submarine - Original Songs From The Film By Al...
1	57LTVY8oPSGy2a7c8SzJpD	What Goes Around…	Alesana	Punk Goes Pop, Vol. 2
2	0fiB5cRA2IngpF06X1Ou4u	Bang Bang You're Dead	Dirty Pretty Things	Waterloo To Anywhere
3	5lDriBxJd22IhOH9zTcFrV	Dirty Little Secret	The All-American Rejects	Move Along
4	0X0Lz7LwpiIWcdGqVWaxXD	Mess Around	Cage The Elephant	Tell Me I'm Pretty

Modeling: Playlist #102585

Features Comparison: Playlist vs. Recommended Track Average
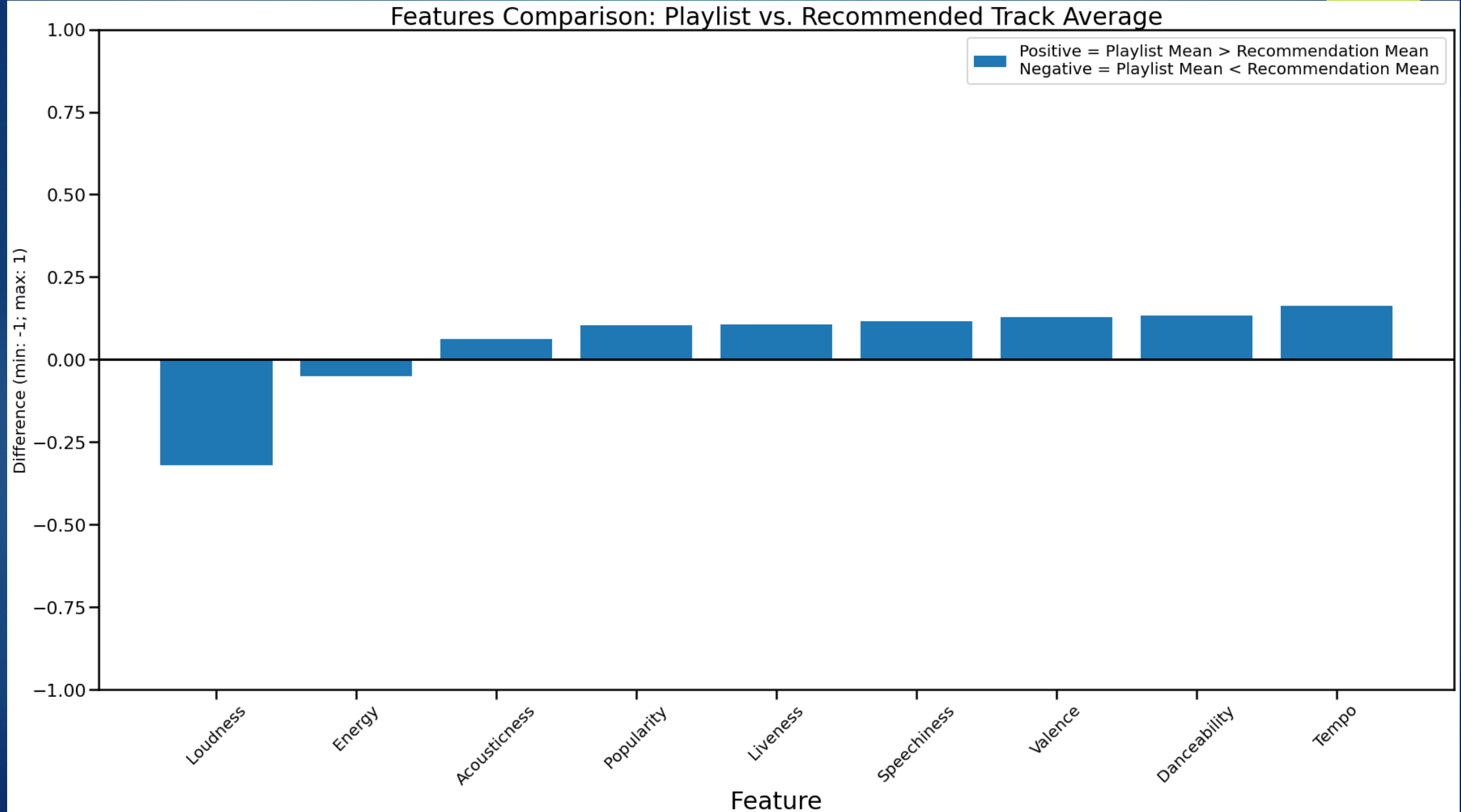
Results

- Aggregate difference low for our recommendations

- Recommendations appear to be mainly driven by genre

- Audio Features and Popularity then provide differentiation once the main genres are selected

- Drawbacks:

  - Suggesting songs from the same basket that the machine is learning from may lead to bias

  - There are few metrics to assess content-based filtering, which makes it harder to determine the broader successfulness of the model

  - Subjective analysis can lead to bias

Next Steps

- Easier Genre Export from Spotipy
- Release-Date Data
- User-review Data -> Collaborative-Based Filtering
- Research clustering
- Use a more powerful computer to use all Spotify data

Thank you For Your Time!

[LinkedIn](#)
[Project Link](#)
[GitHub Profile](#)