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

Stakeholders

- Spotify Users
- DJs
- Other Music-Related professionals

Business Problem

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

Solution

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

Data Understanding

- Spotify Million Playlist Dataset
  - 1m Playlists
  - 66.3m Tracks
  - 734k Unique Albums
  - 295k Unique Artists
  - 92k Unique Titles
- 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.
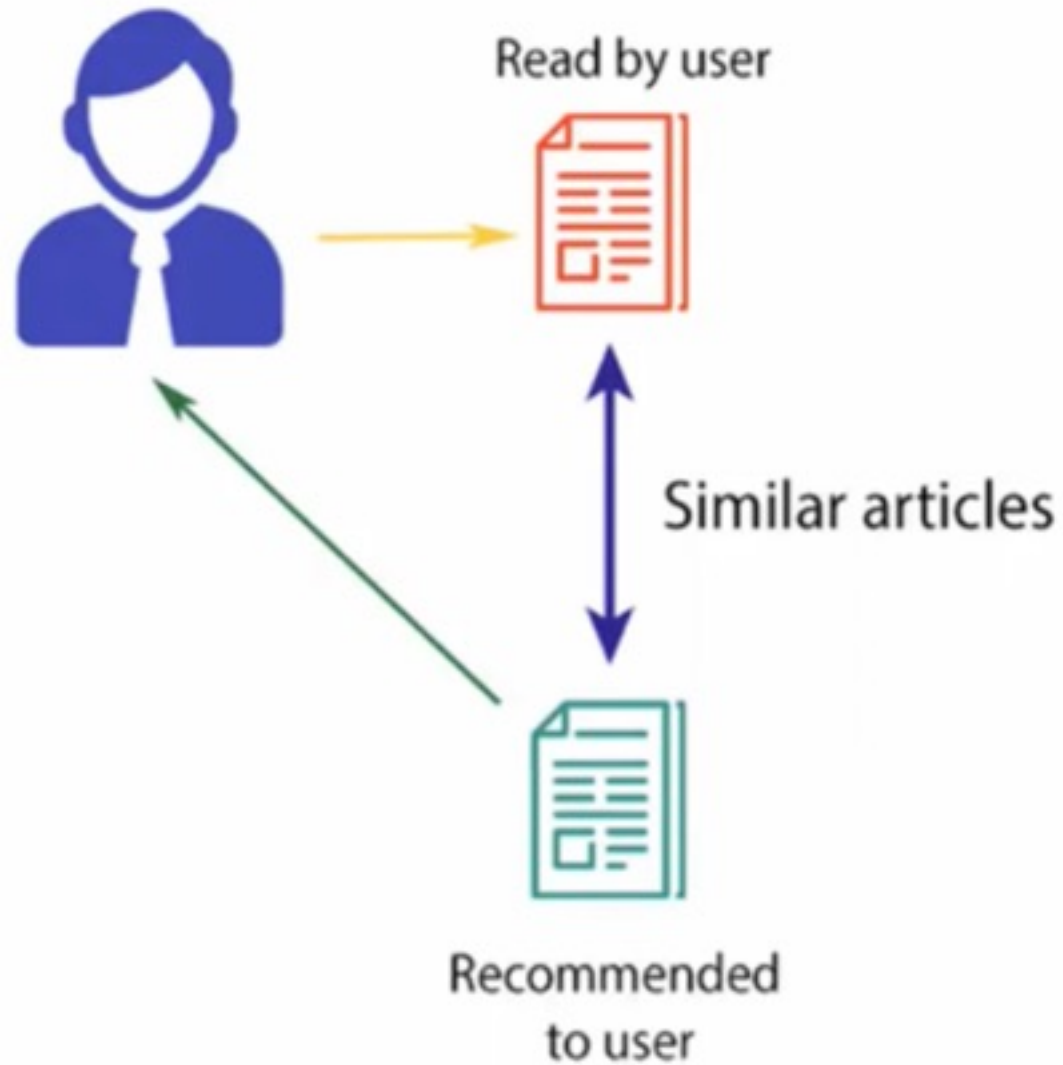
Data Preparation

Playlist Data

- Playlist Average Audio Features
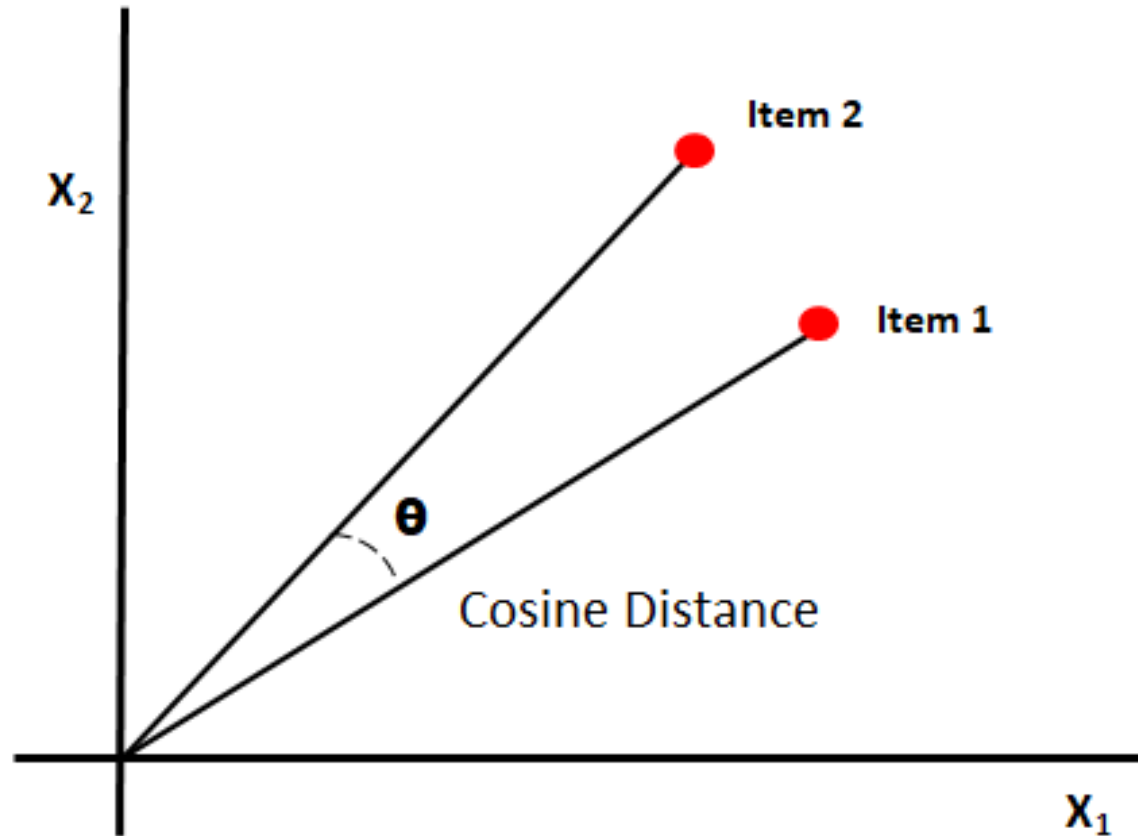- Playlist Average Popularity
- Playlist Genre

Track Data

- Unique Song Audio Features
- Unique Song Popularity
- Unique Song Genre

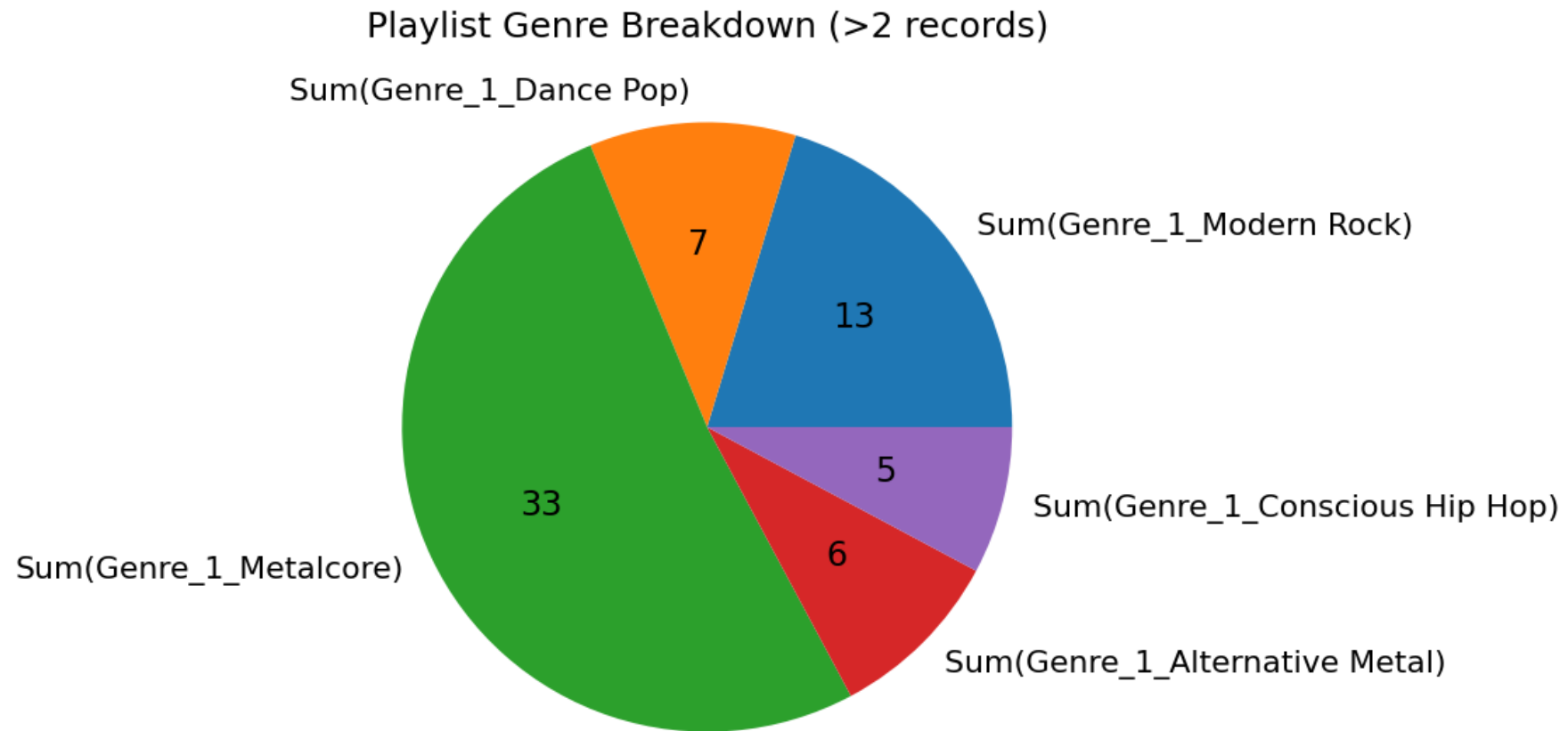Modeling: Content-Based Filtering

Cosine Distance/Similarity

Metric: Cosine Similarity

Modeling: Playlist #102585

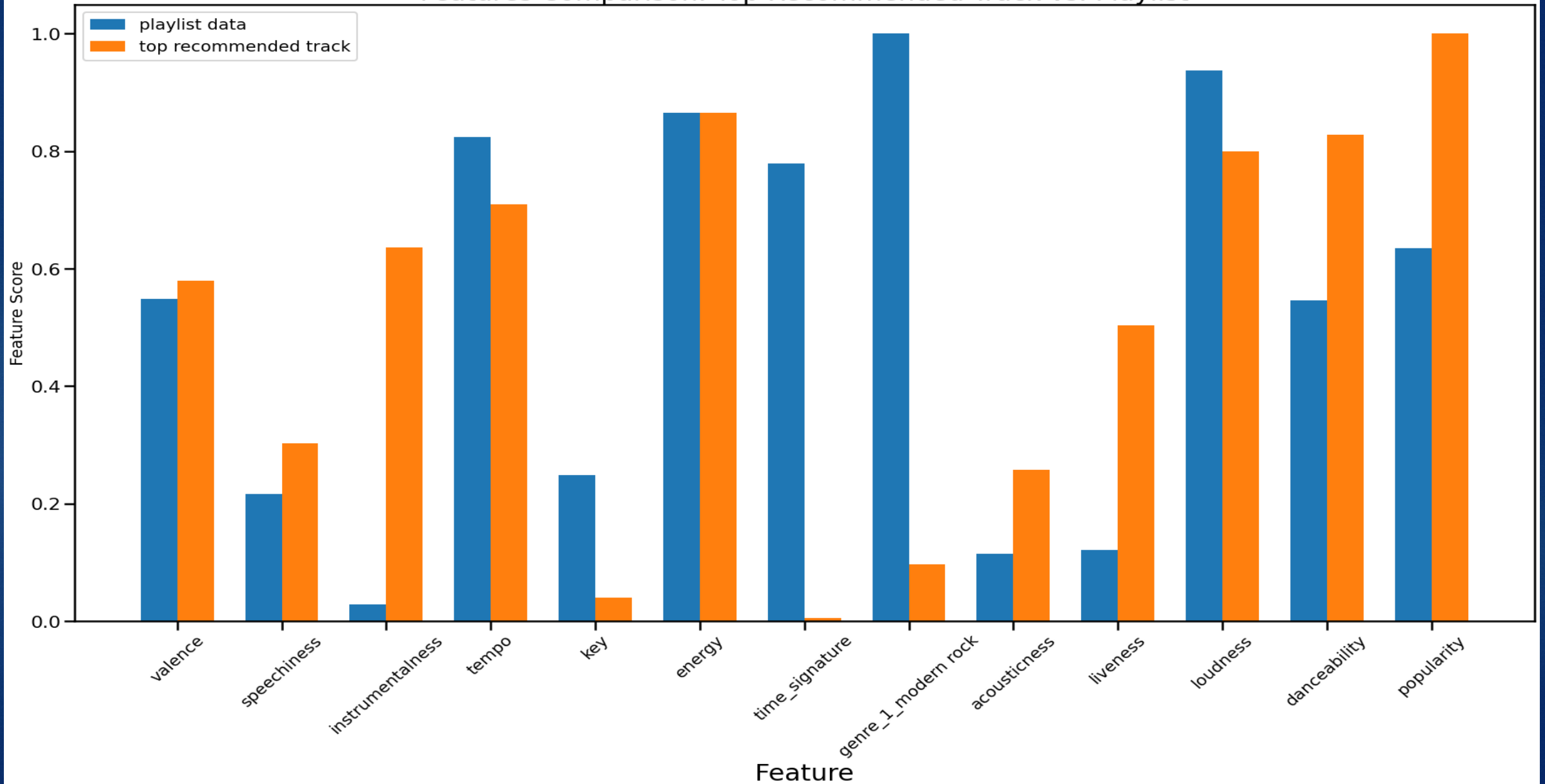| pos	track_uri	track_name	artist_name	album_name
0	67WTwafOMgegV6ABnBQxcE	Some Nights	fun.	Some Nights
1	6Ep6BzlOB9tz3P4sWqiiAB	Radioactive	Imagine Dragons	Night Visions
2	4wCmqSrbyCgxEXROQE6vtV	Somebody That I Used To Know	Gotye	Making Mirrors
3	5j9iuo3tMmQIfnEEQOOjxh	Best Day Of My Life	American Authors	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

Playlist Genre Breakdown (>2 records)

Modeling: Playlist #102585

	track_uri	track_name	artist_name	album_name
0	5zT5cMnMKoyruPj13TQXGx	I Found	Amber Run	5AM (Deluxe)
1	4GlTtbZtRCQXhWLMXrWXHt	Roots	Imagine Dragons	Roots
2	673WCjn0SxKJD4qRKczaCk	She Had The World	Panic! At The Disco	Pretty. Odd.
3	0z8yrlXSjnI29Rv30RssNl	Shots - Broiler Remix	Imagine Dragons	Smoke + Mirrors
4	2AObzKd3JYlWQqQ067Z0Yl	Release	Imagine Dragons	Smoke + Mirrors

Modeling: Playlist #102585

Features Comparison: Top Recommended Track vs. Playlist

Results

- Sample variance low for our top recommended song

- Recommendations appear to be mainly driven by genre

- Audio Features and Popularity then provide differentiation once the genre is 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