

Project: Part 1

1a) What kind of data will you be working with?

Ans- I am going to be working on data available on Spotify. For instance, the songs, artists, albums and charts available on Spotify so that I can make any kind of conclusions about the trends of songs on Spotify. 1b) Why are these data interesting?

Ans- I think this data is interesting because I can use it to derive relationships between different types of artists, or the the type of genre an artist has an affinity towards, and much more. I chose it because it's a versatile dataset and a lot of information can be generated from this dataset. 1c) What kinds of information would you like to generate from the database you will build? You can provide a couple of example questions here.

Ans- This data can be used to create a database that can derive the following information

1. I would like to generate information about which genre of music is the most popular.
2. Which artist prefers which genre of music.
3. What genre of music has the most number of artists.

And much more.

2a) Description of the dataset:

1. What types of information are contained within the proposed database?

Ans- The proposed database is going to contain information available in the Spotify library. There are going to be 4 entities in the database, and they are going to store information about the songs on Spotify, the songs that charted to the top on Spotify in the year 2022, the artists on Spotify and the various albums available on Spotify. The goal is to derive any relationships between the entities to come up with conclusions about the data in the dataset.

3. How many entities and attributes in each table?

There are going to be at least 4 tables/ entities in the database with at least 5 attributes.

a) SONG - which is going to be a table containing data about the various songs available on Spotify.

b) CHART- which is a table that contains data about the top songs of the year 2022 on Spotify.

c) ARTIST- which is an entity that is going to contain information about the various artists on Spotify.

d) ALBUM- this entity is going to contain information about albums released by various artists.

4. What are the logical connections between the tables?

Logical connections between SONG and CHART:

SONG and CHART have a one to many cardinality because an track artist can have multiple tracks in CHART.

SONG and ARTIST also have a one to many cardinality because an artist in ARTIST can have multiple tracks in SONG . Same goes for CHART and ARTIST because the same chart can have multiple instances of the same ARTIST.

SONG and ALBUM have a many to many cardinality because the artist in ALBUM can have multiple tracks in SONG and the same album can appear multiple times in SONG.

ALBUM and CHART have a one to many cardinality because an the artist of an album can show up multiple times in CHART.

Similarly, ARTIST and ALBUM also have a one to many cardinality because an ARTIST can have multiple albums in ALBUM.

Dataset generation:

- The dataset from SONG and CHART was derived from datasets available on Kaggle. The size of these dataset was reduced by selecting the first 200 entries from these datasets.
- For the next step I reduced the number of attributes by selecting the columns that were relevant for my DBMS and removing any irrelevant data and entries that were not deemed to be important for the database.
- On the other hand, the data for ARTIST & ALBUMS was generated using the Spotify web API for developers, the datasets available on Kaggle, and information available on google. The first step was to extract a list of artists and albums from a dataset or google.
- After that, I created a new project on the Spotify webAPI in order to obtain a client ID and client secret which is used to access the Spotify dataset.
- The next step is to create a project on pyCharm, creating a (.env) file to store the client ID & client secret
- write a program in python that is used to obtain information from the Spotify API, by first obtaining the authentication token that allows a developer the access to the data in Spotify.
- Then writing functions to obtain specific information about artists like their top album, song and popularity. Whereas, for Albums its used to generate information like artist name, album rank album genre etc.

Resources:

1. Sveta151. "Spotify Top Chart Songs 2022." *Kaggle*, 5 Sept. 2022, www.kaggle.com/datasets/sveta151/spotify-top-chart-songs-2022.
2. Arvidsson, Joakim. "30000 Spotify Songs." *Kaggle*, 1 Nov. 2023, www.kaggle.com/datasets/joebeachcapital/30000-spotify-songs.
3. Anam, Adnan. "Spotify Artist Stats." *Kaggle*, 22 Sept. 2022, www.kaggle.com/datasets/adnananam/spotify-artist-stats.

4. “Web Api.” *Web API / Spotify for Developers*, developer.spotify.com/documentation/web-api. Accessed 7 Feb. 2024.
5. Developer, Akamai. “How to Use Spotify’s API with Python: Write a Program to Display Artist, Tracks, and More.” *YouTube*, 7 Dec. 2022, youtu.be/WAmEZBEeNmg?si=A-evl6J-6aMVA64O.



