SPOTIFY TRACKS

Full-stack Business Intelligence solution

SPOTIFY TRACKS BUSINESS INTELLIGENCE SOLUTION

Kyrellos El Naggar
23/7/2025

Table of Contents

Contents

1. Data Overview

The project is based on a comprehensive Spotify Tracks Dataset comprising 232,725 tracks distributed evenly across 26 music genres (~10,000 tracks per genre). The dataset is used to build a full-stack Business Intelligence solution using SSIS for ETL, SSAS Tabular for analytics, and Power BI for interactive reporting.

Data Source

- **Input Format:** Excel file
- **Attributes:**
  - `track_id, track_name, artist_name, genre`
  - Audio features: `popularity, danceability, energy, key, mode, speechiness, acousticness, instrumentalness, liveness, valence, tempo, duration_ms, time_signature`

◆ *Identification & Classification Attributes*

Attribute	Description
`track_id`	A unique identifier assigned to each track by Spotify (e.g., `4uLU6hMCjMI75M1A2tKUQC`). Used as the primary key.
`track_name`	The title of the track. May be multilingual or contain special characters.
`artist_name`	Name of the performing artist or band associated with the track.
`genre`	High-level musical genre classification (e.g., Jazz, Rock, Ska). Used for grouping, filtering, and summarization.

◆ *Popularity & Engagement Metrics*

Attribute	Description
`popularity`	A score from **0 to 100** provided by Spotify based on play counts, recency, and user interactions. Higher values indicate higher popularity.

◆ *Audio Feature Metrics (All values between 0.0 and 1.0 unless specified otherwise)*

Attribute	Description
`danceability`	Describes how suitable a track is for dancing based on tempo, rhythm stability, and beat strength. Higher means more danceable.
`energy`	Reflects the intensity and activity level of a track (e.g., fast, loud, and noisy tracks have higher energy).
`speechiness`	Measures the presence of spoken words in a track. Higher values indicate more spoken content (e.g., podcasts, rap).
`acousticness`	Confidence score that a track is acoustic (i.e., not synthesized). A value of **1.0** means it is likely fully acoustic.
`instrumentalness`	Predicts whether a track contains no vocals. Higher scores indicate less vocal content (approaching 1.0 = purely instrumental).
`liveness`	Detects the likelihood of an audience presence. High values (closer to 1.0) suggest the track is live-recorded.

Attribute	Description
`valence`	Describes the musical "positiveness" or emotional tone of a track. A score close to 1.0 indicates a happy, cheerful track; closer to 0.0 indicates a sad or melancholic track.

◆ *Musical Theory Attributes*

Attribute	Description
`key`	The musical key of the track (e.g., C, D#, F). A discrete value representing pitch class (12 total).
`mode`	Indicates the modality: **Major** (1) or **Minor** (0), reflecting the general harmonic characteristics of the track.
`tempo`	Estimated tempo in **beats per minute (BPM)**. Used for rhythm analysis and beat matching.
`time_signature`	Notation that specifies how many beats are in each bar (e.g., 4/4, 3/4). Represented as an integer from **0 to 5**, corresponding to standard musical signatures (e.g., 0 = unknown, 4 = 4/4).

◆ *Duration Attribute*

Attribute	Description
`duration_ms`	Length of the track in **milliseconds**. Used to calculate playback time. Converted to `duration_min` during ETL for easier reporting and filtering in dashboards.

Data Quality Handling

- `key, mode, time_signature` corrected for type issues.
- Null values handled across key fields (`genre, artist_name, track_name, etc.`).
- `duration_ms` converted to minutes for better interpretability.

2. Conceptual & Logical & Physical Model

This solution implements a **star schema** design in the Data Warehouse layer for high performance and ease of analysis.

2.1 Conceptual Model

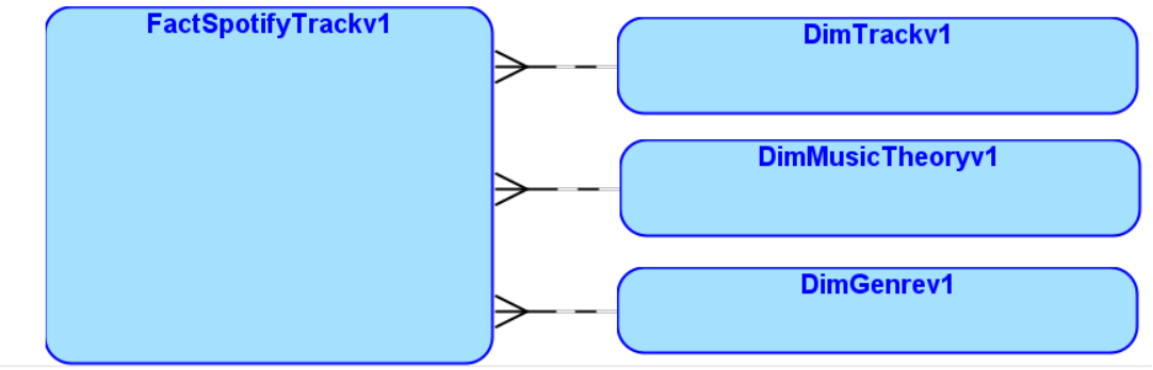
The conceptual model defines the **core entities**, their **roles**, and **relationships** within the business context—abstracted away from technical implementation. It describes **what data exists (Tables)** and **how it's related**, without getting into physical storage or data types.

The conceptual model consists of **one central fact entity** and **three surrounding dimension entities**, forming a **classic star schema** structure.

Entities (Blocks):

- **FactSpotifyTrack** *(Fact Table)*
- **DimTrack** *(Dimension)*
- **DimGenre** *(Dimension)*
- **DimMusicalTheory** *(Dimension)*

All dimensions related to fact in 1:M relationship.



2.2 Logical Model

The **logical model** is defined using the following entities:

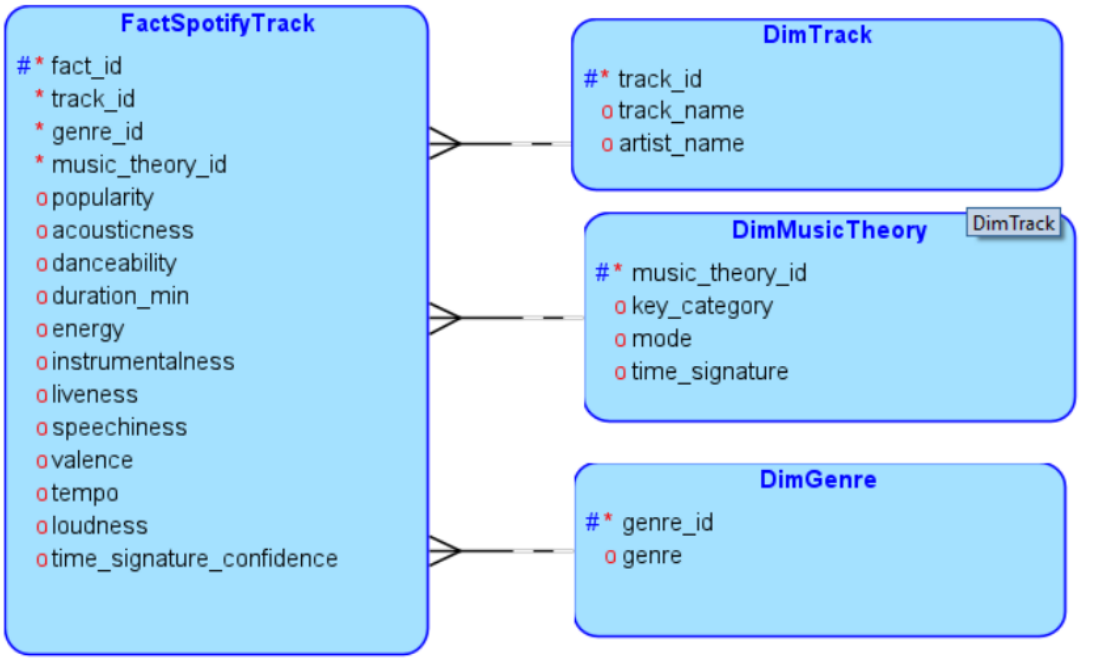
Dimensions:

- DimTrack: Track metadata including track_id, track_name, artist_name
- DimGenre: Genre classifications
- DimMusicalTheory: Musical attributes like key, mode, time_signature

Fact Table:

- FactSpotifyTrack: Fact table referencing all dimensions and housing measurable audio features and calculated durations.

Each dimension is designed to support one-to-many relationships with the fact table.



2.3 Physical Model

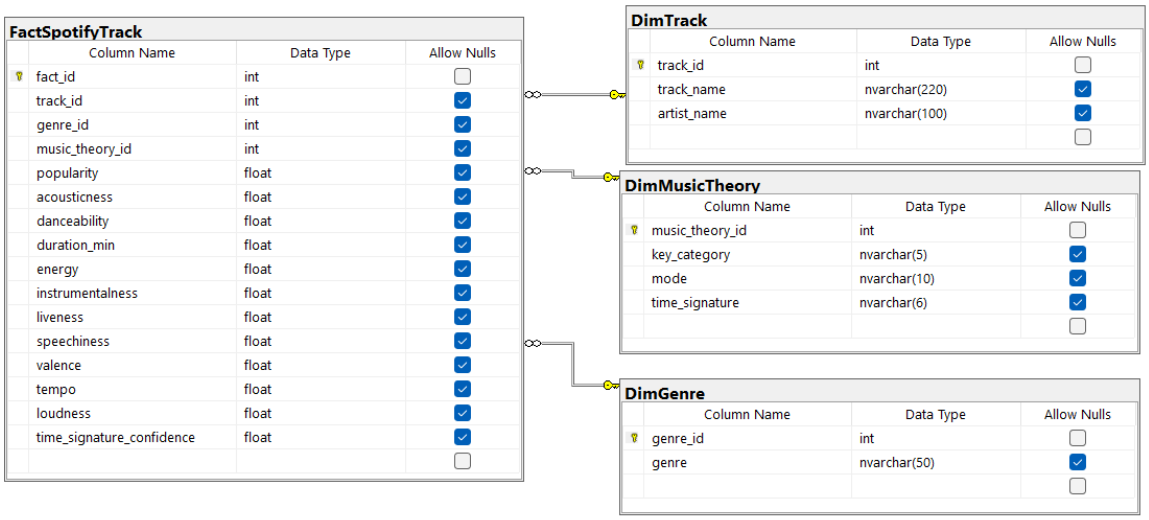
The physical model implements the **logical star schema** within a SQL Server-based Data Warehouse. It focuses on **how data is stored** in the database, **table structures**, and **physical organization**—not on how data gets there.

Databases Involved:

- **DWH_Spotify** → Final Data Warehouse containing structured data for analytics
- **Tables Stored Physically:**
  - **FactSpotifyTrack** *(Fact table storing measurable metrics)*
  - **DimTrack** *(Descriptive details of tracks and artists)*
  - **DimGenre** *(Genre classifications)*
  - **DimMusicalTheory** *(Musical key, mode, time signature)*

Structure:

- Surrogate keys used for all dimension tables
- Foreign key relationships enforced between fact and dimension tables
- Star schema optimized for analytical querying and SSAS consumption
- All text fields normalized and typecast to avoid BLOB issues in reporting tools



3. SSIS Overview

The SSIS (SQL Server Integration Services) layer orchestrates the **ETL (Extract, Transform, Load)** process in three clearly defined stages, aligned with the architecture's layered design:

ETL Database Layers:

1. **ODS_Spotify** – Raw operational data store
2. **STG_Spotify** – Cleaned and conformed staging layer
3. **DWH_Spotify** – Final dimensional warehouse

A. ODS Package – Raw Data Load:

Objective: Ingest raw Spotify Excel data and apply basic cleaning and conversions before loading it into the ODS layer.

ETL Flow:

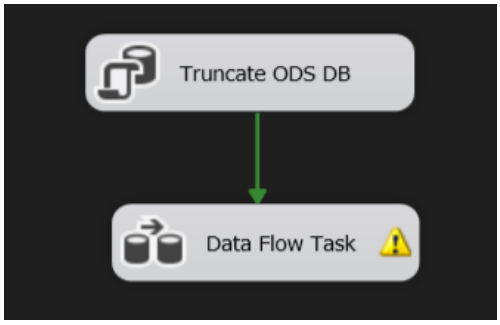
- **Source:** Excel file with 232,725 tracks
- **Destination:** ODS_Spotify database

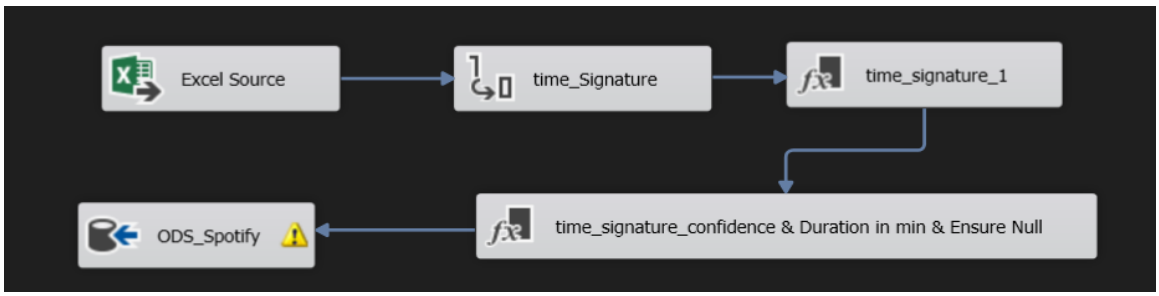
Key Tasks:

- Truncate existing data (Execute SQL Task)
- Convert time_signature from date to string (DT_WSTR)
- Replace NULL values in key descriptive fields:
  - track_name, artist_name, genre, key, mode, time_signature
- Derive duration_min from duration_ms for usability in analytics

SSIS Components Used:

- Excel Source
- Data Conversion
- Derived Column
- Data Flow Task
- OLE DB Destination





B. STG Package – Staging and Transformation

Objective: Prepare and organize cleaned data into dimensions and fact tables for further warehousing.

ETL Flow:

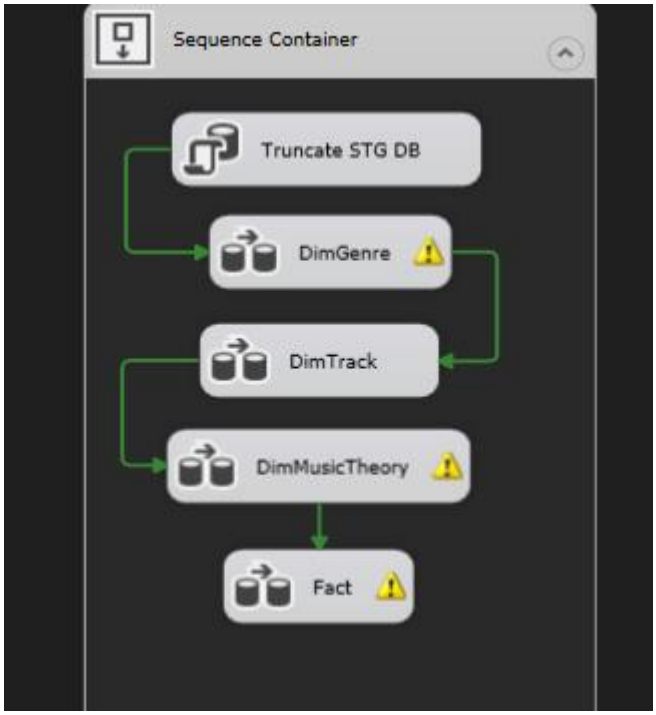
- **Source:** ODS_Spotify database
- **Destination:** STG_Spotify database

Key Tasks:

- Truncate STG dimension and fact tables
- Store and structure data into:
    - DimGenre
    - DimTrack → with validation of track_name, artist_name (null & length checks)
    - DimMusicalTheory → validated time_signature, key, and mode
    - Fact_Spotify → includes references and prep for lookup in DWH

SSIS Components Used:

- Sequence Container
- Data Flow Tasks
- Derived Column
- OLE DB Destination



C. DWH Package – Warehouse Load & Lookup

Objective: Populate the dimensional model in DWH_Spotify with clean, fully integrated data.

ETL Flow:

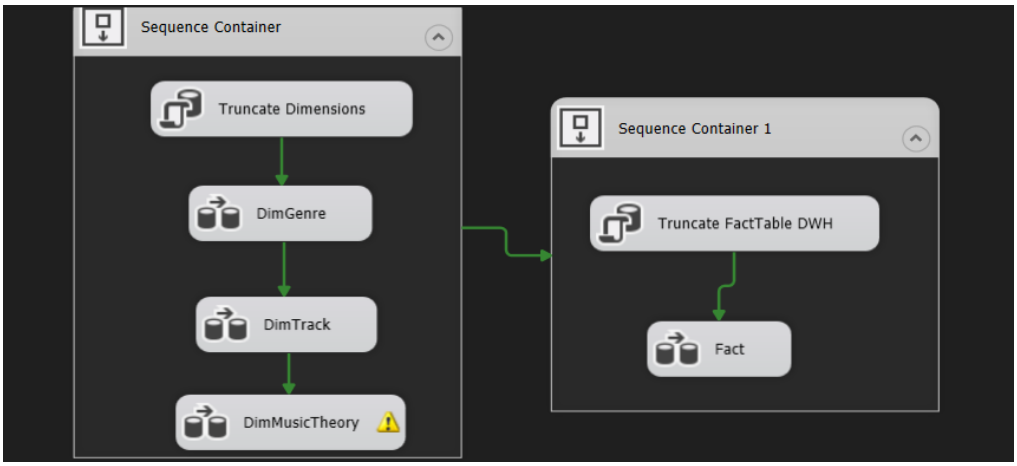
- **Source:** STG_Spotify database
- **Destination:** DWH_Spotify database

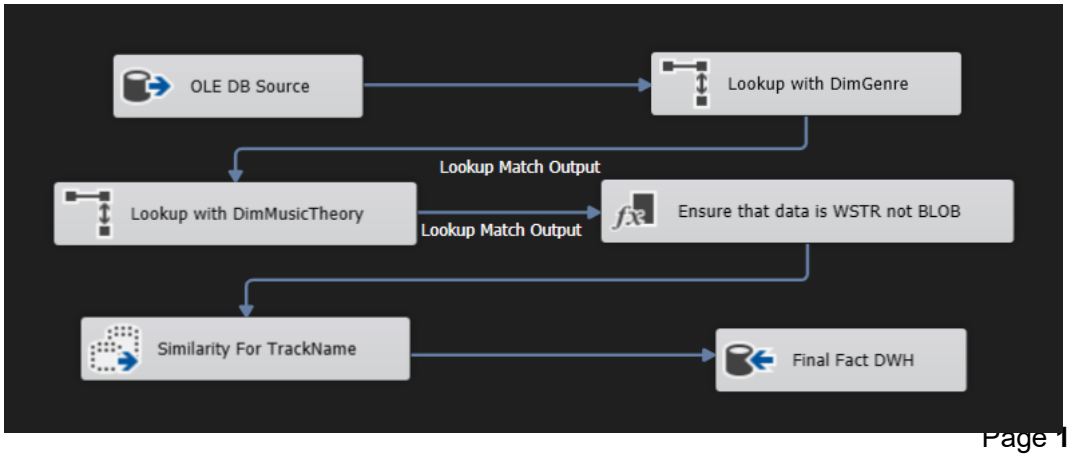
Key Tasks:

- Truncate final dimension and fact tables
- Load dimensions: DimTrack, DimGenre, DimMusicalTheory
- Populate FactSpotifyTrack with:
    - **Lookups:** Genre and musical theory IDs
    - **Fuzzy Lookup:** DimTrack based on track_name and artist_name (similarity threshold = 0.98) to handle multilingual/inconsistent entries
    - **Data Type Enforcement:** Ensures string fields are compatible with SSAS and Power BI (avoids BLOB output)

SSIS Components Used:

- Two Sequence Containers (Dimensions / Fact)
- Lookup
- Fuzzy Lookup
- Derived Column
- Data Flow Tasks
- OLE DB Destination





4. SSAS Overview

The **SSAS Tabular model** is built on top of DWH_Spotify using the following:

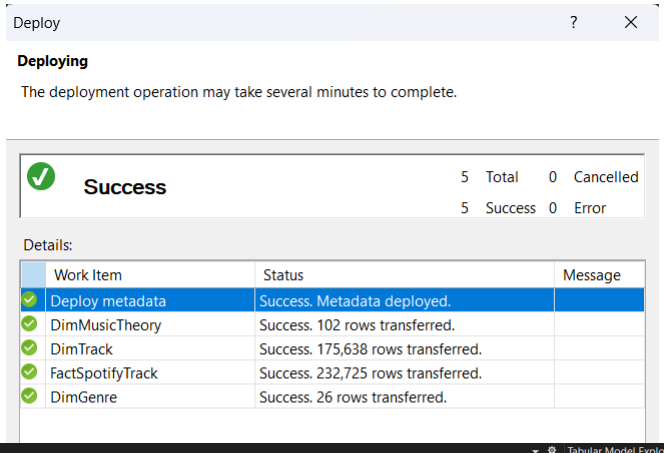
Model Structure:

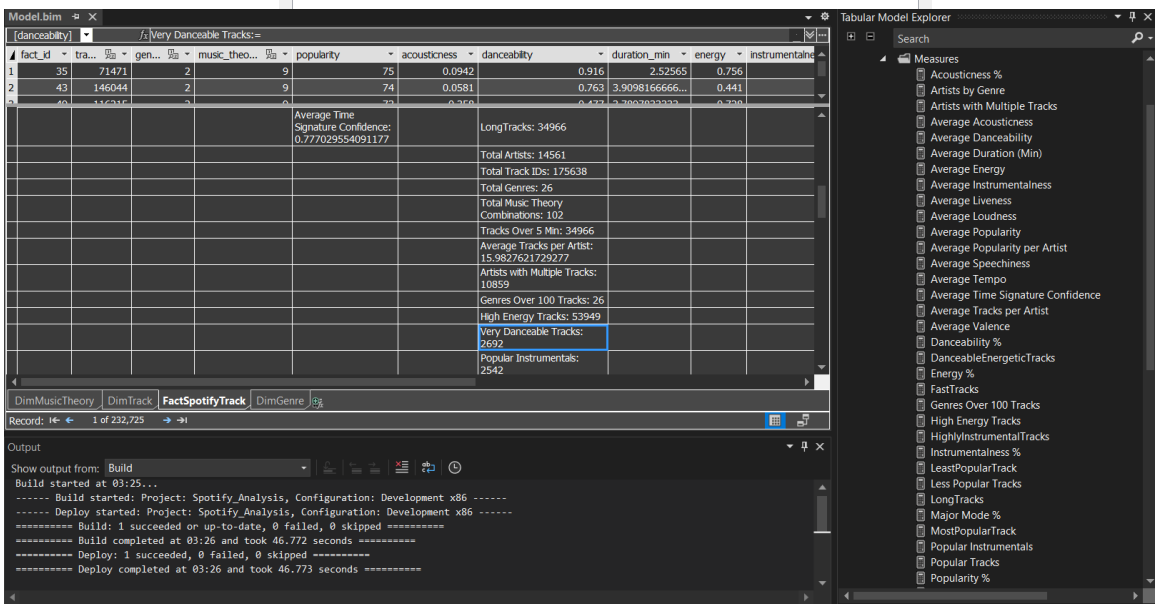
- Tables: DimGenre, DimMusicalTheory, DimTrack, FactSpotifyTrack
- Relationships: Standard star schema relationships between fact and dimensions

Measures Implemented (Sample):

Measure Type	Example Measures
Aggregates	Total Tracks, Total Artists, Average Popularity
Conditional Counts	DanceableEnergeticTracks, FastTracks, HighlyInstrumentalTracks
Percentages	Major Mode %, Danceability %, Energy %, Popularity %
Custom Metrics	Average Tracks per Artist, Popular Instrumentals

- Measures written in **DAX**
- Supports advanced analysis like artist productivity, track mood, genre energy levels



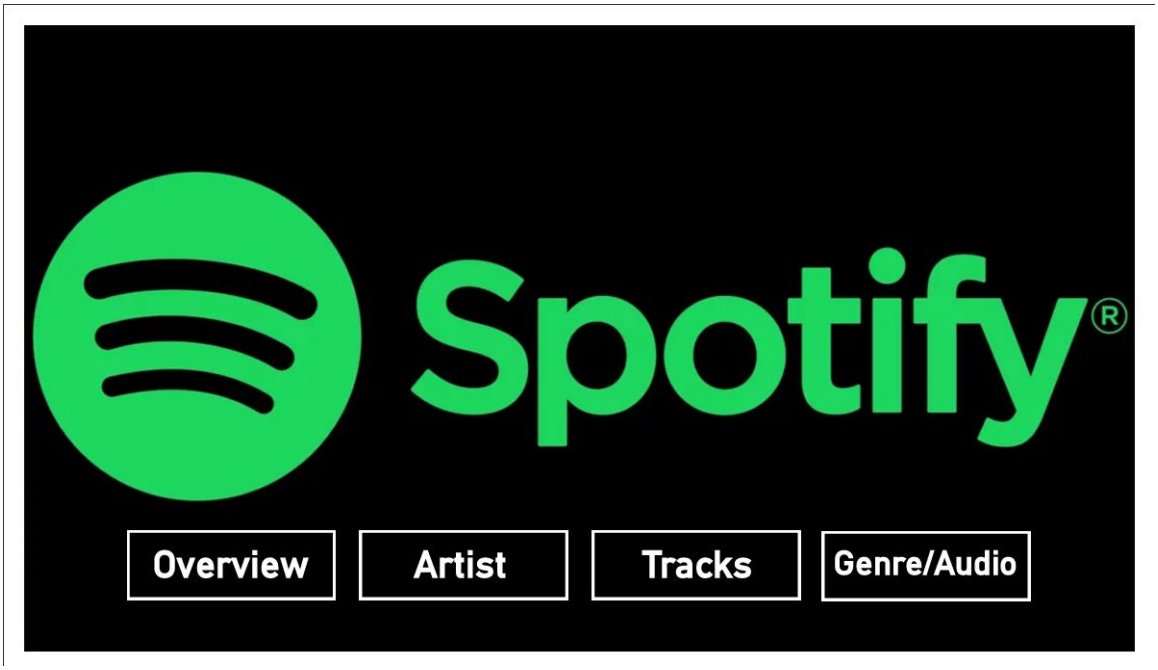


5. Power BI Overview

Dashboard Navigation Structure:

- **Cover Page**: Contains buttons to:
    1. Overview
    2. Artist
    3. Tracks
    4. Genre/Audio

- **Navigation Tools:**
    - Filter bar
    - Home button (Spotify logo)
    - Forward/Backward navigation
    - Clear filters




Overview
Track
Artist Name
Mode
Genre
Musical key
Time Signature
Track/Artist Info
- A BEGINNING
-
!!!
14.561K
232.725K
26
Track
Artist
Artists
Tracks
Genres
Music Info
0.37
0.12
0.45
0.22
0.55
41.13
Acousticness
Speechiness
Valence
Liveness
Danceability
Popularity
Average Tracks per Artist
15.98
53.949K
High Energy Tracks
Total Tracks by mode
Very Danceable Tracks
2.692K
Major
Minor
Artists Vs Tracks
Total Artists
Total Tracks
54.529K
Short Tracks
34.966K
Long Tracks
Popularity
Energy
100.00%
57.15%
Total Tracks by time_signature
4/4
3/4
5/4
1/4
0/0
201K
24K
5K
3K
0K
0K
50K
100K
150K
200K
Musical Key
A
Count of Beats Per Bar by Musical Key
Time Signature
0/0
C
C#
D
F
G#
A
A#
B
D#
E
F#
G


Artists
Track
Artist Name
Mode
Genre
Musical key
Time Signature
Artist Info
!!!
14.561K
42.60
15.98
Artist
Artists
Popularity
Avg Tracks/Artist
Music Info
0.37
0.12
0.45
0.22
0.55
41.13
Acousticness
Speechiness
Valence
Liveness
Danceability
Popularity
Artists with Multiple Tracks To Total Artists
Artists with Multiple Tracks
Total Artists
10.86K
14.56K
Time signature Confidence
145.38K
Tracks by Artist
0.78
0.00
1.00
Top 10 Artist Based on Avg Danceability vs Valence
Average Valence
Average Danceability
Top 10 Artists by Track Length
Total Tracks
LongTracks
ShortTracks
artist_name
Top 10 Artists by Track
Total Tracks
artist_name
Top 10 Artists of Avg Loudness and Avg Popularity
Average Loudness
Average Popularity
artist_name


Tracks
Track
Artist Name
Mode
Genre
Musical key
Time Signature
Track Info
232.725K
Tracks
26
Genres
15.98
Avg Tracks/Artist
3.92
Avg Duration (Min)
Music Info
0.37
Acousticness
0.12
Speechiness
0.45
Valence
0.22
Liveness
0.55
Danceability
41.13
Popularity
Genres Over 100 Tracks
26
18.54K
Danceable Energetic Tracks
26
Genres Over 100 Tracks
Tracks Over 5 Min
34.966K
Long Vs Short Tracks
Less Popular ...
Popular T...
36.454K
Popular Tracks
196.271K
Less Popular Tracks
Total Tracks by key_category
Tracks by Mode
Major
Minor
Top 10 Popular Tracks
Average Popularity
7 RINGS
BREAK UP WITH YOUR GIRLFRIEND, I'M BORED
WOW.
CON CALMA
CALMA - REMIX
DANCING WITH A STRANGER (WITH NORMANI)
SUNFLOWER - SPIDER-M... INTO THE SPIDER-VE...
SWEET BUT PSYCHO
MIDDLE CHILD
TAKI TAKI (WITH SELENA GOMEZ, OZUNA & ...
track_name
mode
key_category
time_signat...
genre
Total Tracks
232725
Major
151744
Minor
80981
C
21944
G
20029
D
18466
4/4
17447
3/4
2034
5/4
354
Soul
223
Country
213
Reggaeton
199


Genre/Audio
Track
Artist Name
Mode
Genre
Musical key
Time Signature
Genre/Audio Info
232.725K
Tracks
26
Genres
15.98
Avg Tracks/Artist
3.92
Avg Duration (Min)
Music Info
0.37
Acousticness
0.12
Speechiness
0.45
Valence
0.22
Liveness
0.55
Danceability
41.13
Popularity
Tempo %
48.44
0.00
96.88
105.311K
Fast Tracks
Major Mode %
53.949K
High Energy Tracks
0.65
0.00
1.30
genre
Loudness
Liveness
Danceability
Acousticness
Children's Music -5.88 0.21 0.73 0.22
Folk -6.05 0.19 0.64 0.15
Rock -6.17 0.24 0.53 0.11
Classical -6.50 0.18 0.64 0.23
Movie -6.53 0.20 0.54 0.18
Rap -6.54 0.20 0.54 0.18
Dance -6.67 0.20 0.70 0.17
Soundtrack -6.86 0.20 0.72 0.18
Electronic -7.04 0.21 0.62 0.12
Ska -7.29 0.19 0.54 0.20
Alternative -7.34 0.19 0.58 0.27
A Capella -7.52 0.19 0.70 0.19
Total -9.57 0.22 0.55 0.37
Average Valence by genre
A Capella
Opera
Classical
Comedy
Reggae
Alternative
R&B
Pop
Jazz
World
Children's M...
Folk
Rap
Rock
Ska
Movie
Blues
Electronic
Indie
Hip-Hop
Anime
Popular Genre
Average Popularity
genre
genre
Total Tracks
World 9543
Soundtrack 9295
Soul 9646
Ska 9272
Rock 8874
Reggaeton 8280
Reggae 5403
Rap 9263
R&B 8992
Pop 8916
Opera 9023
Total 232725
Genre by Total Tracks and Avg Duration (Min)
Total Tracks
Average Duration (Min)
genre