IronHack Data Analytics

WEEK 4 | SONG POPULARITY ON SPOTIFY

Agenda

- Project Overview
- Data Acquisition, Enrichment, Examination
- Database Design & Data Transformation
- SQL Insights & Advanced Analysis
- Conclusions & Business Implications
- Major Obstacles & Learnings

Project Overview

- **Original datasets** – all sources from Kaggle
  - Main data set: Spotify Top 50 Tracks 2023
  - Supplemental data set: Top Spotify Hits 2000 – 2019

Business Problem:

Identify the musical attributes that contribute most significantly to a track's popularity on Spotify

Song key features:danceability, energy, valence, tempo, explicitly…

Use Case:

- **Artists & Producers:** Optimize song structure and characteristics based on trends in popular music
- **Record Labels:** Improve song selection for signing new artists and curating successful playlists

Project Overview

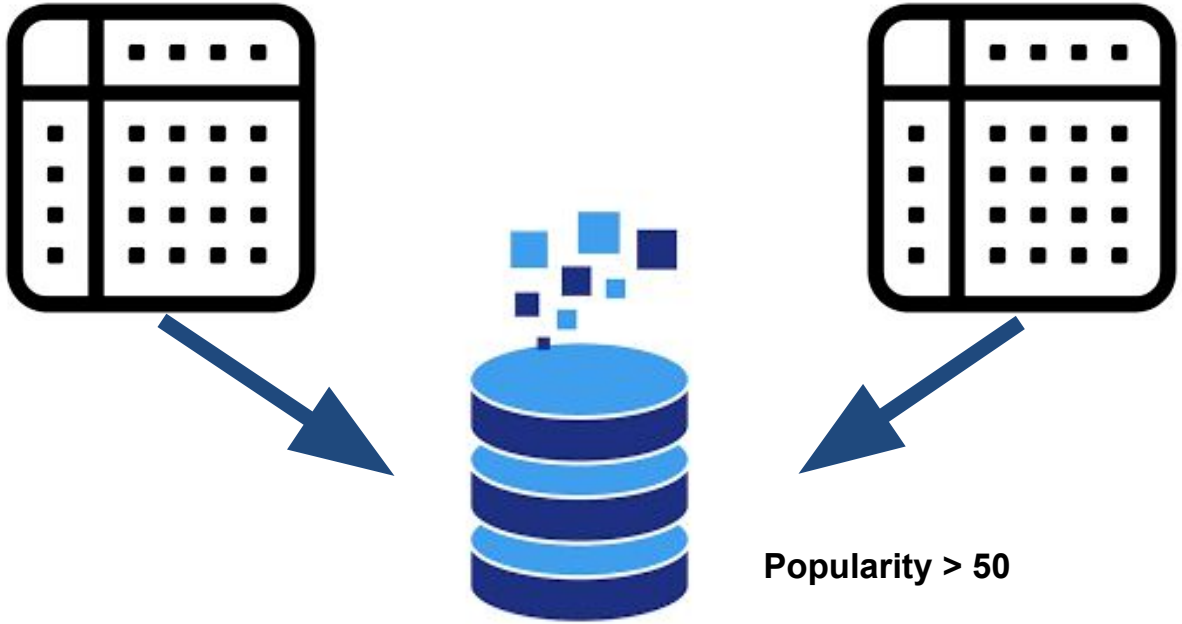
- Hypothesis 1: **Danceability and Energy Positively Impact Popularity**
  - Tracks with higher danceability and energy scores are more likely to have a higher popularity rating

- Hypothesis 2: **Explicit Content is More Popular in Recent Years**
  - Explicit tracks have become more popular over time, particularly in the last decade

Data Acquisition, Enrichment and Examination

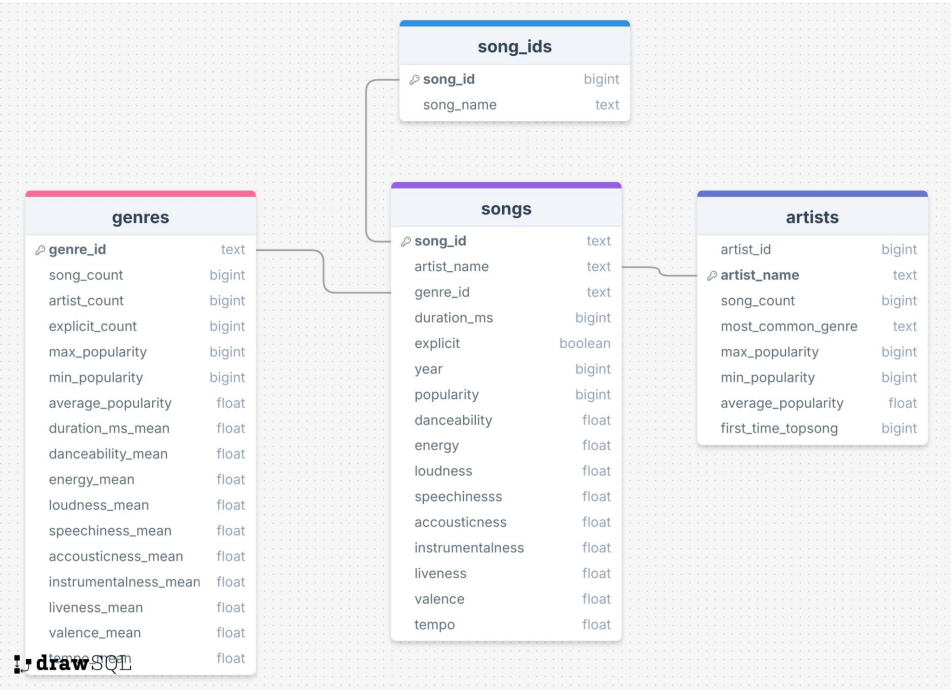
Top Spotify Songs 2020 – 2019

Spotify Top Tracks 2023



Popularity > 50

Database Design & Data Transformation



Key Challenges:

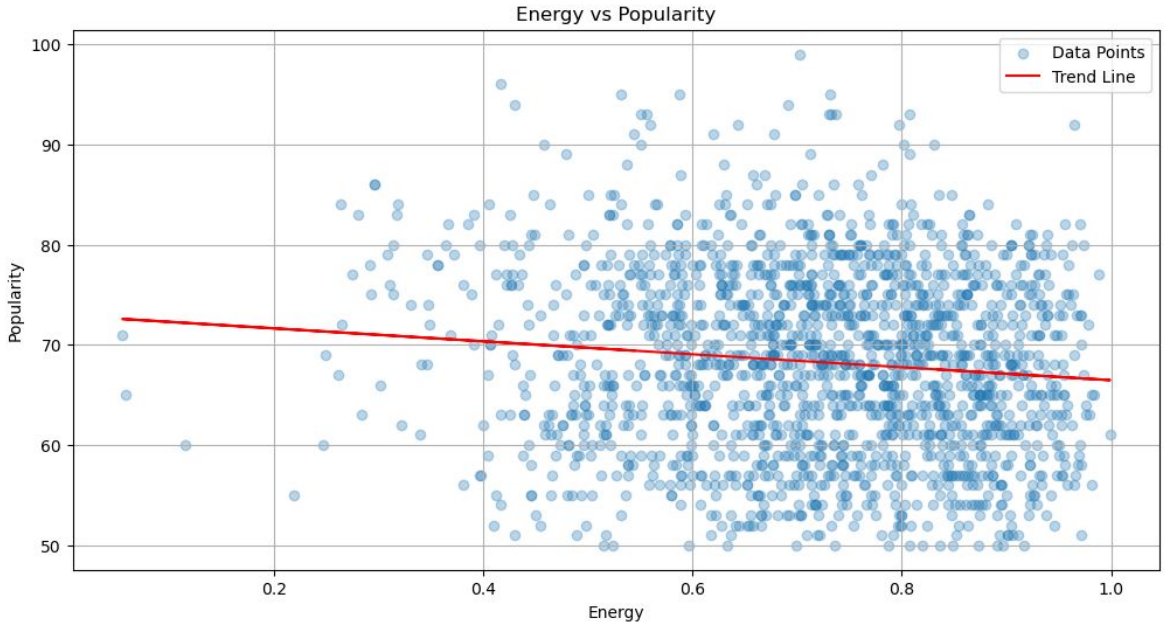
- OS issue while loading the database with data via CSV import on Mac
- Defining complete song titles as main key generated errors when loading data on database
- **Solution**: new table named **song_ids** to transform **song_id** (int) onto **song_name** titles

Important variables

- **Popularity**: The higher the value the more popular the song is (max 100). Our database contains only songs with **popularity higher than 50**
- **Explicit**: The lyrics or content of a song or a music video contain one or more of the criteria offensive or unsuitable for children (**True/False**)
- **Danceability**: Danceability describes how suitable a track is for dancing. A value of 0.0 is least danceable and 1.0 is most danceable
- **Energy**: Energy is a measure from 0.0 to 1.0 and represents a perceptual measure of intensity and activity

SQL Insights & Advanced Analysis:

ENERGY VS. POPULARITY



Trend line shows slight negative slope

→ higher **energy scores** are **weakly associated with** lower **popularity**

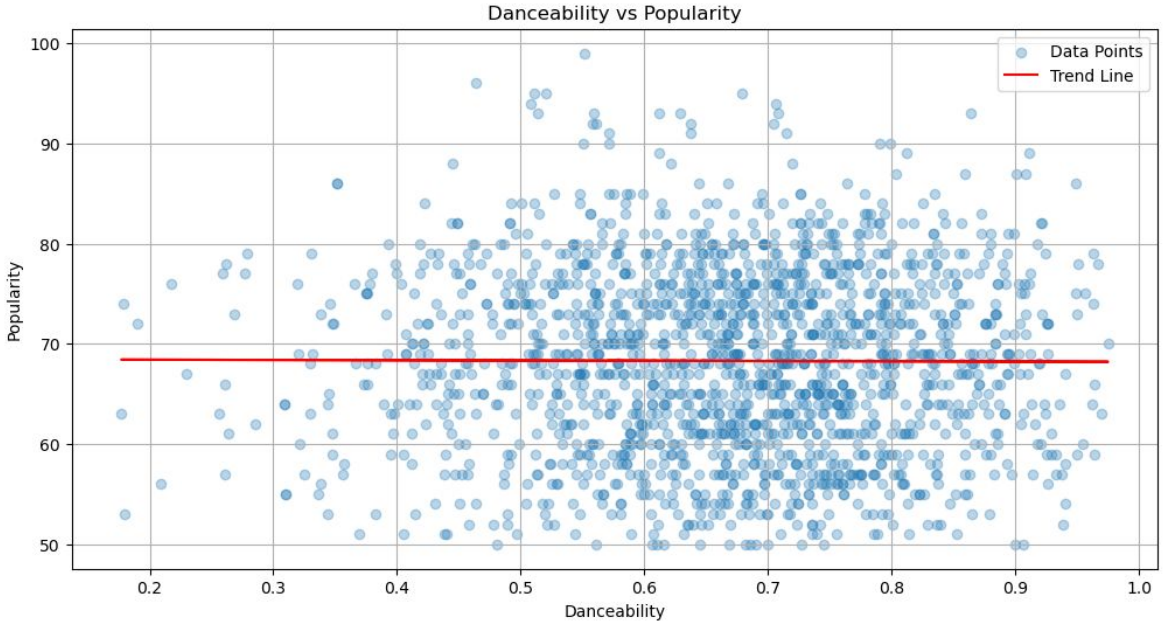
→ directly contradicts the hypothesis!

SQL Insights:

353 songs with energy lower than 0.60

SQL Insights & Advanced Analysis:

DANCEABILITY VS. POPULARITY



Trend line almost flat

→ **no significant correlation** between danceability and popularity

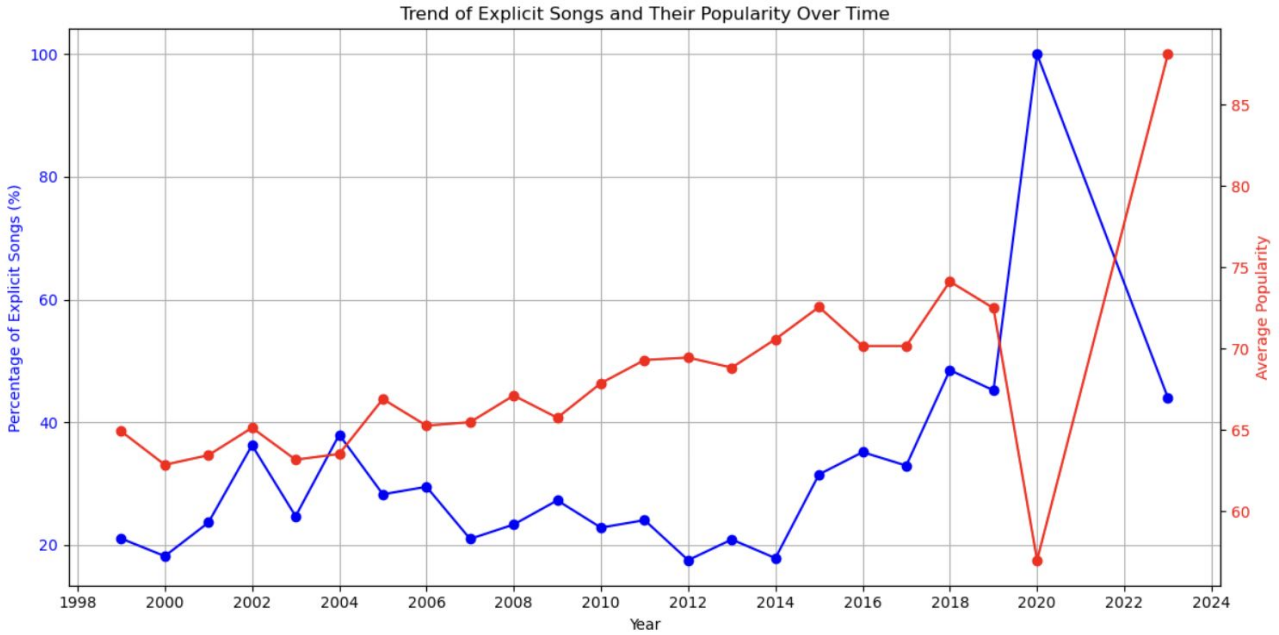
SQL insights:

852 songs have energy higher than average

198 songs with danceability lower than 0.50

SQL Insights & Advanced Analysis:

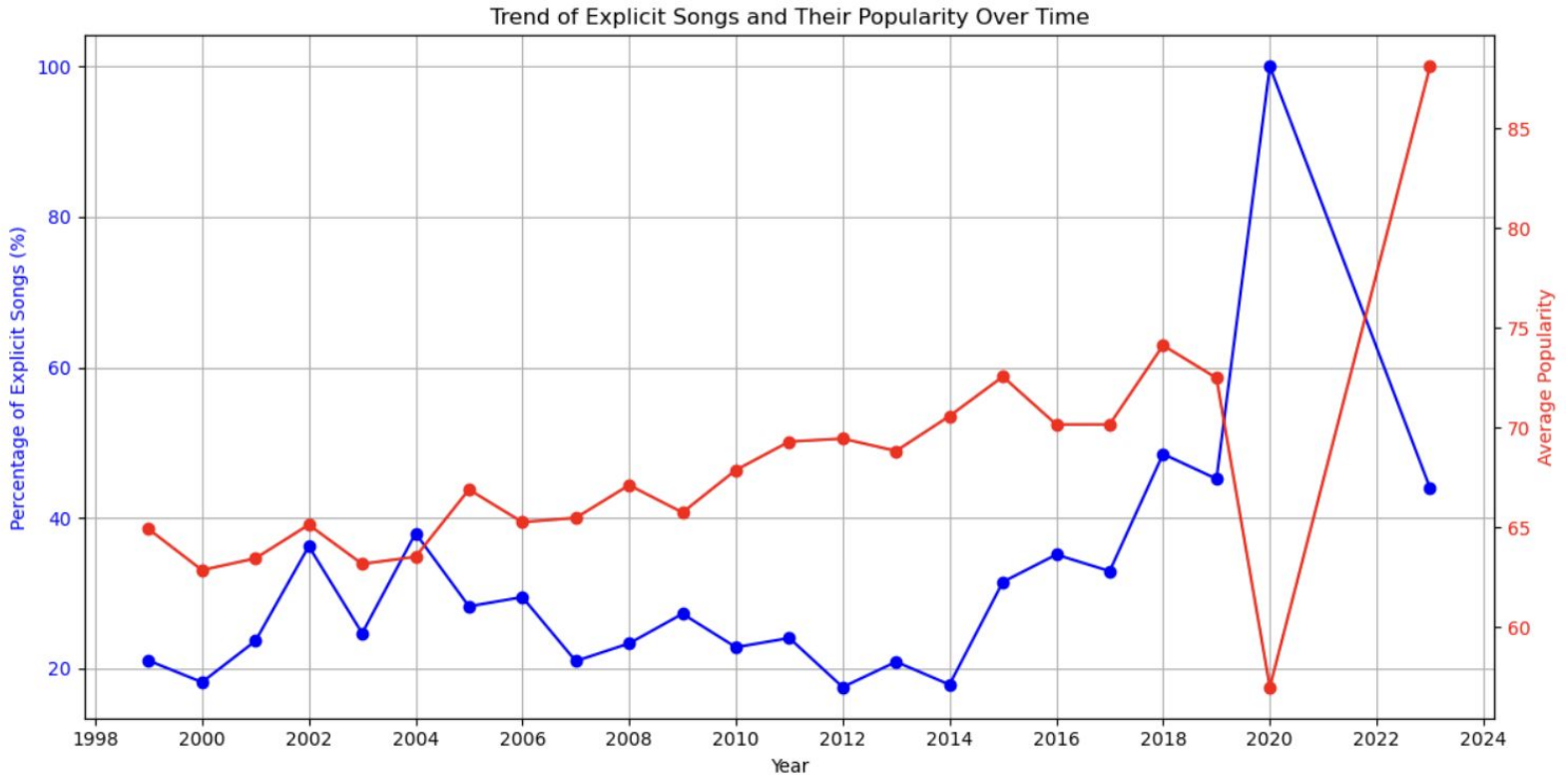
EXPLICIT CONTENT



Year	Explicit songs %
1999	21.05
2000	19.05
2001	24.18
2002	35.53
2003	25.33
2004	38.46
2005	26.74
2006	30.67
2007	21.33
2008	22.89
2009	25.33
2010	24.10
2011	25.33
2012	17.20
2013	20.00
2014	18.18
2015	33.33
2016	33.82
2017	34.57
2018	48.35
2019	46.48

SQL Insights: Explicit songs increasing during last decade

Explicit Content



Explicit Content

Year	Explicit songs %
1999	21.05
2000	19.05
2001	24.18
2002	35.53
2003	25.33
2004	38.46
2005	26.74
2006	30.67
2007	21.33
2008	22.89
2009	25.33
2010	24.10
2011	25.33
2012	17.20
2013	20.00
2014	18.18
2015	33.33
2016	33.82
2017	34.57
2018	48.35
2019	46.48

Hypothesis 2: **Explicit Content is More Popular in Recent Years**

✓ **CORRECT**, explicit songs reached its maximum at mid-early '00

In 2023 the percentage of explicit songs was **44.19%.**

SQL Insights: Explicit songs increasing during last decade

Conclusions & Business Implications

✓ Hypothesis 1: **Danceability and Energy Positively Impact Popularity**

- Tracks with higher danceability and energy scores are more likely to have a higher popularity rating

✓ Hypothesis 2: **Explicit Content is More Popular in Recent Years.**

- Explicit tracks have become more popular over time, particularly in the last decade
- In **2023** the percentage of explicit songs was **44.19%**

Implications:

- Danceability and Energy no must-have ingredients for a song to become popular
- Tracks with explicit content no longer taboo, shock lyrics bode well in terms of popularity

Major Obstacles & Learnings

- **Major obstacles:**
  - Loading the Data Frames into MySQL database causing significant delays
  - Redoing the ERD to account for unique identifier (song_ids)

- **Lessons learnt:**
  - Main key != titles or names, better an integer ID
  - Creating an ID transformation table (songs_ids) is a good practice

PROJECT SONG POPULARITY ON SPOTIFY

THANKS !

Lukas Günther| Víctor Ramírez