Max Branca - 2061432

Data Warehouse Project

A multidimensional data warehouse to explore trends in audio features, genres, and commercial success across decades.

Project Objectives

This project aims to design and implement a unified data warehouse that brings together diverse aspects of the music industry, including audio features, artist profiles, genre classifications, and chart performance. The goal is to create a flexible analytical platform capable of revealing long-term trends and correlations in popular music and how musical styles, popularity, and artist success evolve over time.

Tools & Technologies

This project leveraged three main tools across different stages of the workflow:

Kaggle

Used as the primary source for collecting the datasets. Kaggle offers a wide range of open datasets, including those related to music, artists, and song popularity.

Python

Employed during the entire ETL process. Python was used to extract, clean, merge, and transform the data into a structured format ready for analysis.

Tableau

Utilized for the analytical and visualization phase. Tableau allowed the creation of charts, graphs and tables to explore trends and insights.

Dataset Selection

For this project, I selected **four datasets from Kaggle** to build a complete picture of the music world:

- Two datasets focus on song features: the **Song Features Dataset** and the **Spotify 1.2M+ Songs Dataset** provide information like song name, artist name and audio characteristics.
- One dataset covers artist information: the **Worldwide Music Artists Dataset** includes artist names, genres, and countries.
- One dataset tracks song rankings: the **Billboard Hot 100 Songs Dataset** contains weekly chart positions of the most popular songs in the U.S. since 1958.

ETL Process Overview

The ETL (Extract, Transform, Load) process followed a structured approach to prepare the datasets for integration:

Cleaning

Removed duplicates, handled missing values, and normalized text fields to ensure consistency across all datasets.

Merging Song Features

The two datasets containing song-level audio features were combined into a single, unified dataset with consistent structure and attributes.

Data Standardization

Standardized key fields such as artist_id and song_id across the merged datasets, the artist dataset, and the chart dataset. This ensured reliable joins between tables.

ETL Results

charts.csv

Tipo	Nome campo
Abc	Song Id (Charts Final With Artist Id.Csv)
📅	Date
#	Rank
Abc	Song
Abc	Artist
#	Peak-Rank
#	Weeks-On-Board
Abc	Artist Id (Charts Final With Artist Id.Csv)

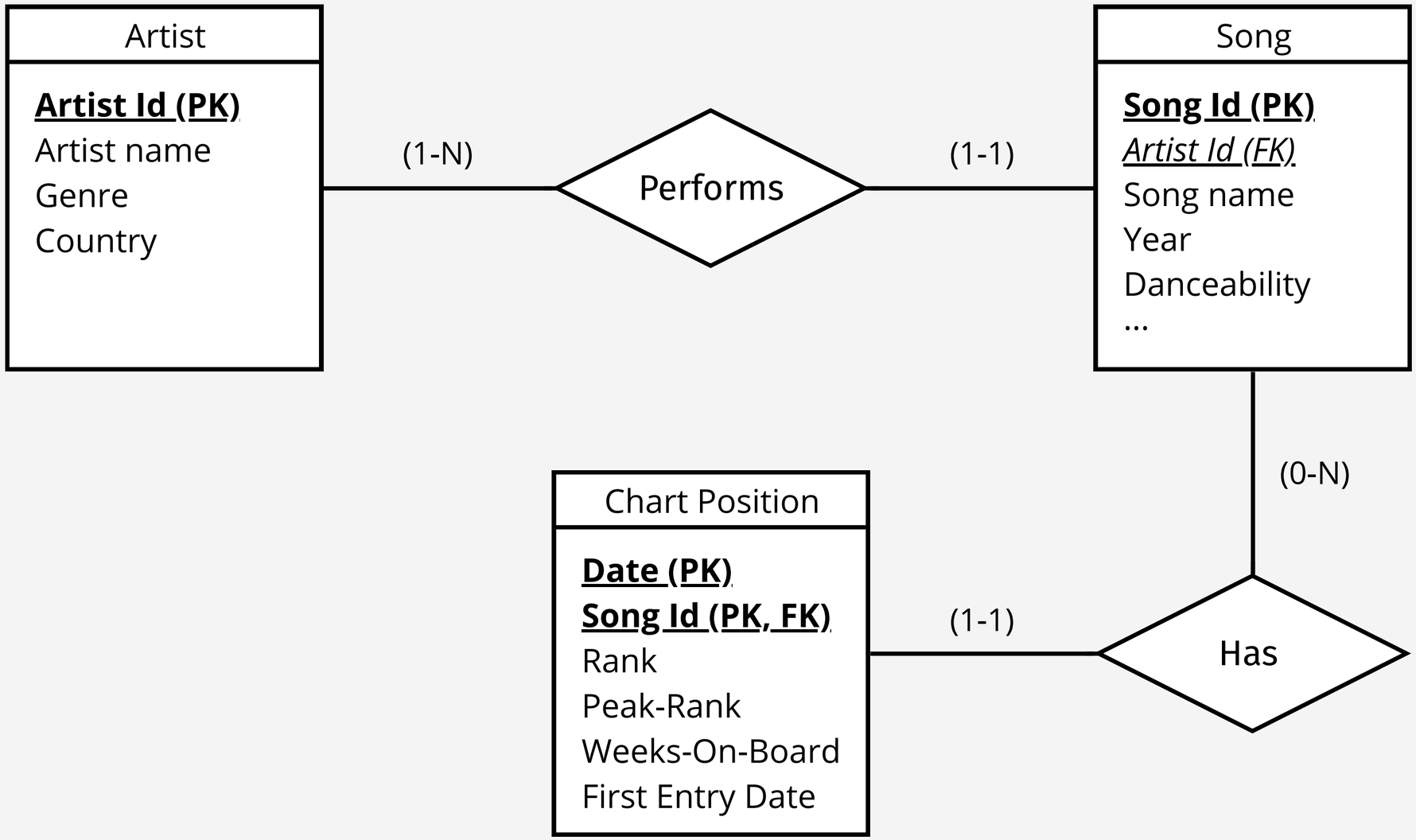
tracks.csv

Tipo	Nome campo
Abc	Song Id
Abc	Name
Abc	Artists
#	Danceability
#	Energy
#	Loudness
#	Acousticness
#	Liveness
#	Valence
#	Tempo
#	Year
Abc	Artist Id
=#	Decade

artists.csv

Tipo	Nome campo
Abc	Artist Id (Artists Final No Img.Csv)
Abc	Artist Name
Abc	Artist Genre
🌐	Country
=Abc	Main Genre

ER Schema - Reconcilied Data Layer



From Reconcilied Data Layer to DW

```sql
-- Song Dimension
CREATE TABLE dim_song (
  song_id TEXT PRIMARY KEY,
  artist_id TEXT,
  name TEXT,
  year INT,
  danceability FLOAT,
  acousticness FLOAT,
  energy FLOAT,
  liveness FLOAT,
  loudness FLOAT,
  valence FLOAT,
  tempo FLOAT
);

COPY dim_song("song_id", "artist_id", "name", "year", "danceability",
"acousticness", "energy", "liveness", "loudness", "valence", "tempo")
FROM 'tracks.csv'
DELIMITER ','
CSV HEADER;
```

```sql
-- Artist Dimension
CREATE TABLE dim_artist (
  artist_id TEXT PRIMARY KEY,
  artist_name TEXT,
  artist_genre TEXT,
  country TEXT
);

COPY dim_artist("artist_id", "artist_name",
"artist_genre", "country")
FROM 'artists.csv'
DELIMITER ','
CSV HEADER;
```

From Reconcilied Data Layer to DW

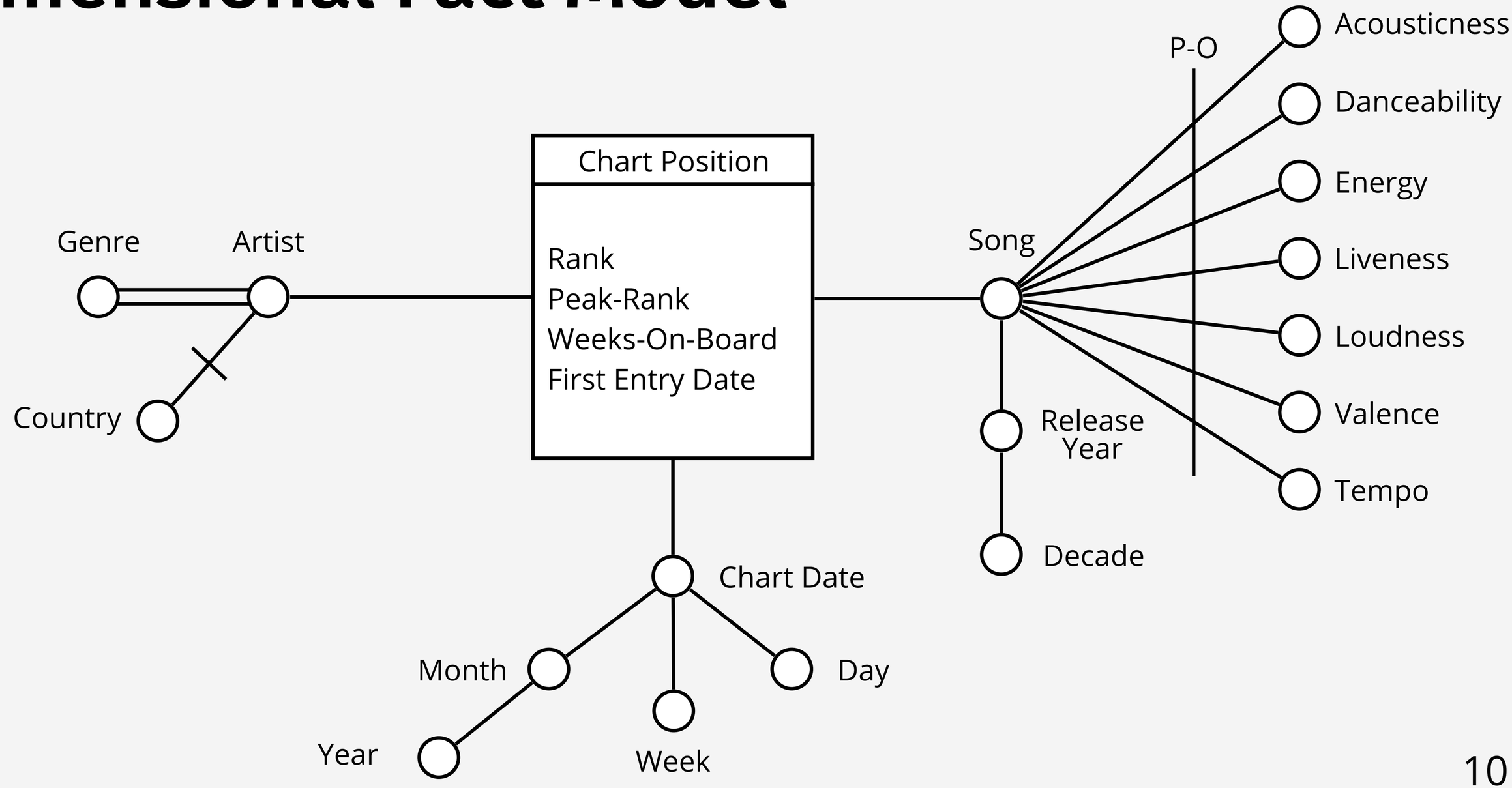
```sql
-- Temp table (used for Date dim and Fact Table)
CREATE TABLE temp_chart (
  song_id TEXT,
  date DATE,
  rank INT,
  peak_rank INT,
  weeks_on_board INT
);

COPY temp_chart("song_id", "date", "rank",
"peak_rank", "weeks_on_board")
FROM 'charts.csv'
DELIMITER ','
CSV HEADER;
```

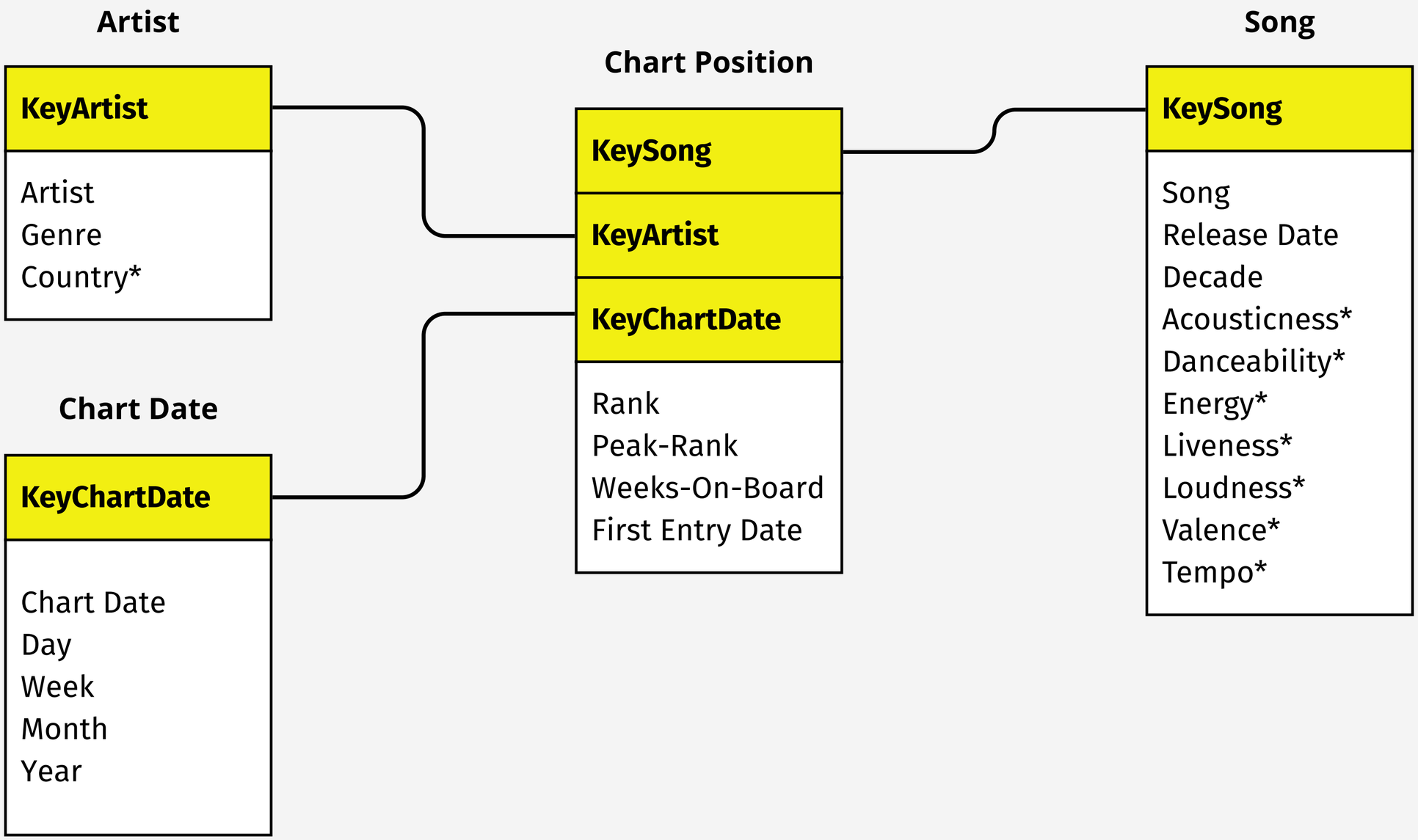
```sql
-- Date Dimension
CREATE TABLE dim_date AS
SELECT DISTINCT
  date AS date_key,
  EXTRACT(YEAR FROM date) AS year,
  EXTRACT(MONTH FROM date) AS month,
  EXTRACT(DAY FROM date) AS day,
  EXTRACT(WEEK FROM date) AS week
FROM temp_chart;

-- FACT TABLE
CREATE TABLE fact_chart_position AS
SELECT
  c.song_id AS song_key,
  s.artist_id AS artist_key,
  c.date AS date_key,
  c.rank,
  c.peak_rank,
  c.weeks_on_board,
  c.first_entry_date
FROM temp_chart_raw c
JOIN dim_song s ON c.song_id = s.song_id;
```

Dimensional Fact Model



Star Schema

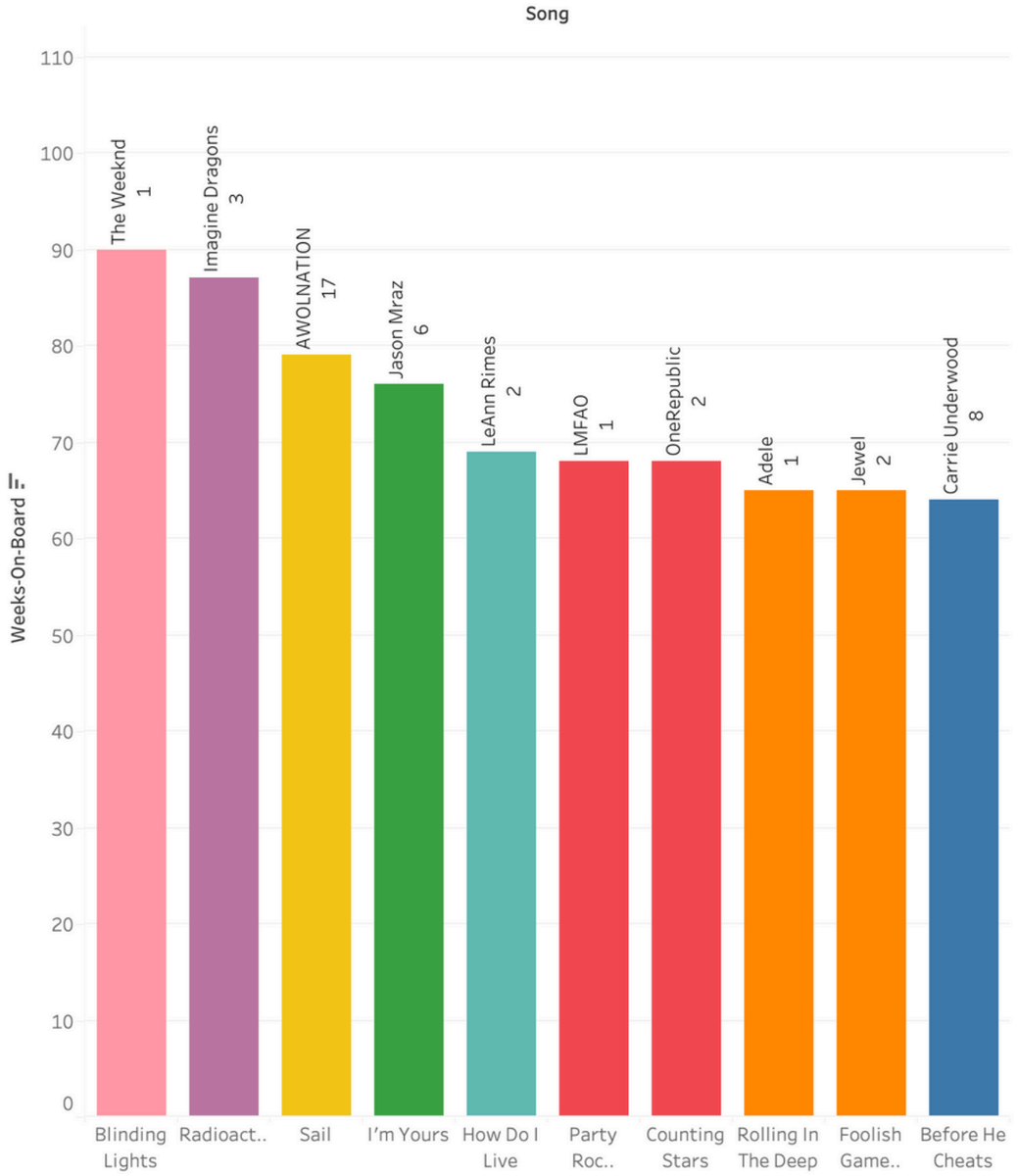

Artist
KeyArtist
Artist
Genre
Country*
Chart Date
KeyChartDate
Chart Date
Day
Week
Month
Year
Chart Position
KeySong
KeyArtist
KeyChartDate
Rank
Peak-Rank
Weeks-On-Board
First Entry Date
Song
KeySong
Song
Release Date
Decade
Acousticness*
Danceability*
Energy*
Liveness*
Loudness*
Valence*
Tempo*

Top 10 Longest Charting Songs on Billboard Hot 100

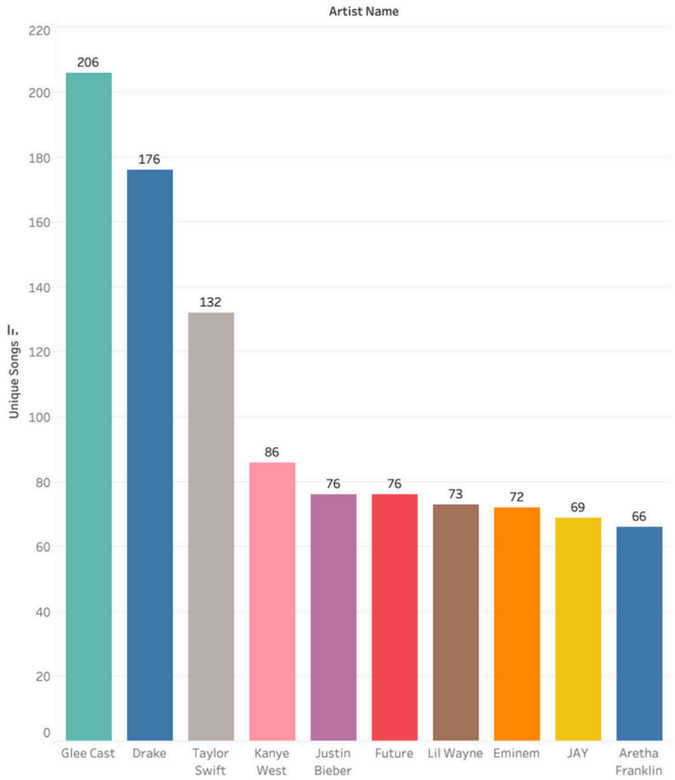
This bar chart displays the **Top 10 songs with the highest number of weeks on the Billboard Hot 100**, highlighting their lasting popularity over time.

- **"Blinding Lights" by The Weeknd** leads with **90 weeks**, making it the longest-charting song in Billboard history.
- Other notable long-runners include:
  - "Radioactive" by Imagine Dragons (87 weeks)
  - "Sail" by AWOLNATION (79 weeks)
  - "I'm Yours" by Jason Mraz (76 weeks)

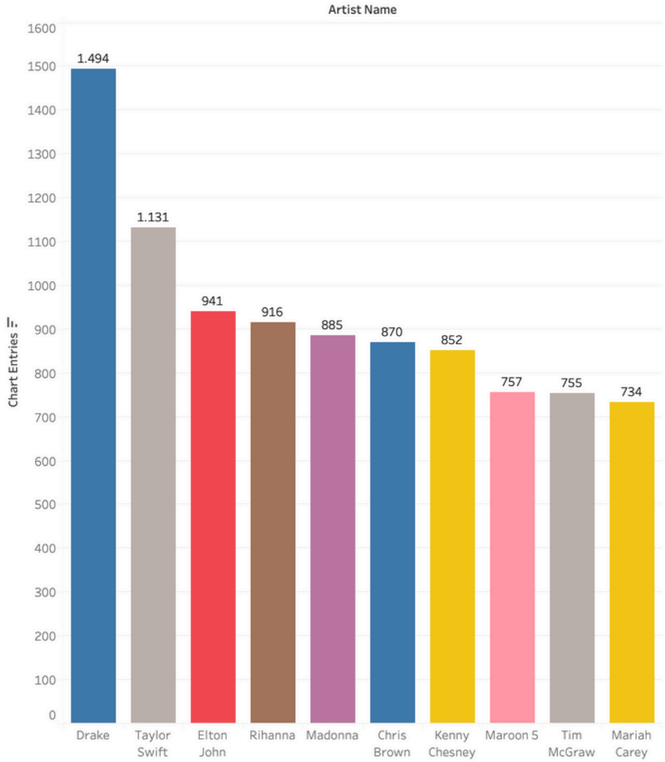
These results showcase how certain songs maintain public attention and commercial relevance well beyond their initial release, offering insight into enduring musical appeal.



Other Bar Chart Examples



Top 10 Artists with most Unique Songs on Billboard Hot 100



Top 10 Artists by Chart Entries on Billboard Hot 100

Top 10 Songs with Longest Consecutive Weeks at #1

This table shows the **Top 10 songs that held the #1 position on the Billboard Hot 100 for the most consecutive weeks**:

Song	Artist Name	
Old Town Road	Lil Nas X	19
Despacito	Luis Fonsi	16
One Sweet Day	Mariah Carey	16
Candle In The Wind 1997/Something About The Way You Look Tonight	Elton John	14
I Gotta Feeling	The Black Eyed Peas	14
I Will Always Love You	Whitney Houston	14
I'll Make Love To You	Boyz II Men	14
Macarena (Bayside Boys Mix)	Los Del Rio	14
Uptown Funk!	Mark Ronson	14
We Belong Together	Mariah Carey	14

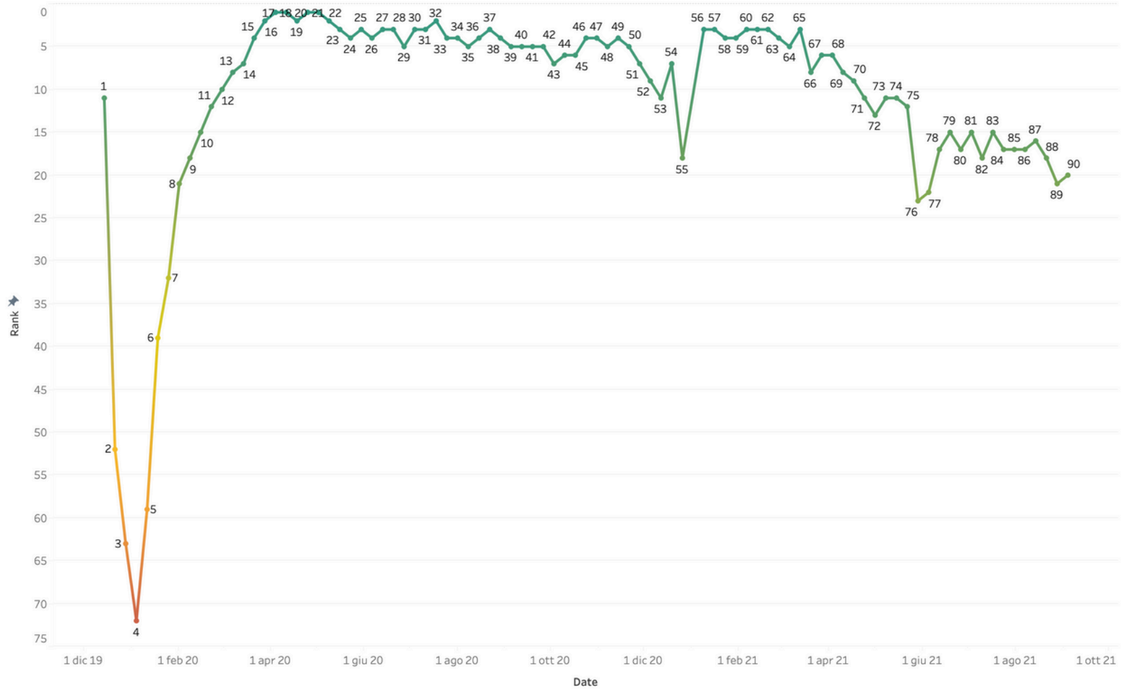
- **"Old Town Road" by Lil Nas X** stands out with **19** consecutive weeks at #1 — a record-breaking run.
- "Despacito" (Luis Fonsi) and "One Sweet Day" (Mariah Carey) follow with 16 weeks each.
- Several iconic tracks stayed at the top for 14 consecutive weeks, including:
  - "I Gotta Feeling" – The Black Eyed Peas
  - "Uptown Funk!" – Mark Ronson
  - "We Belong Together" – Mariah Carey

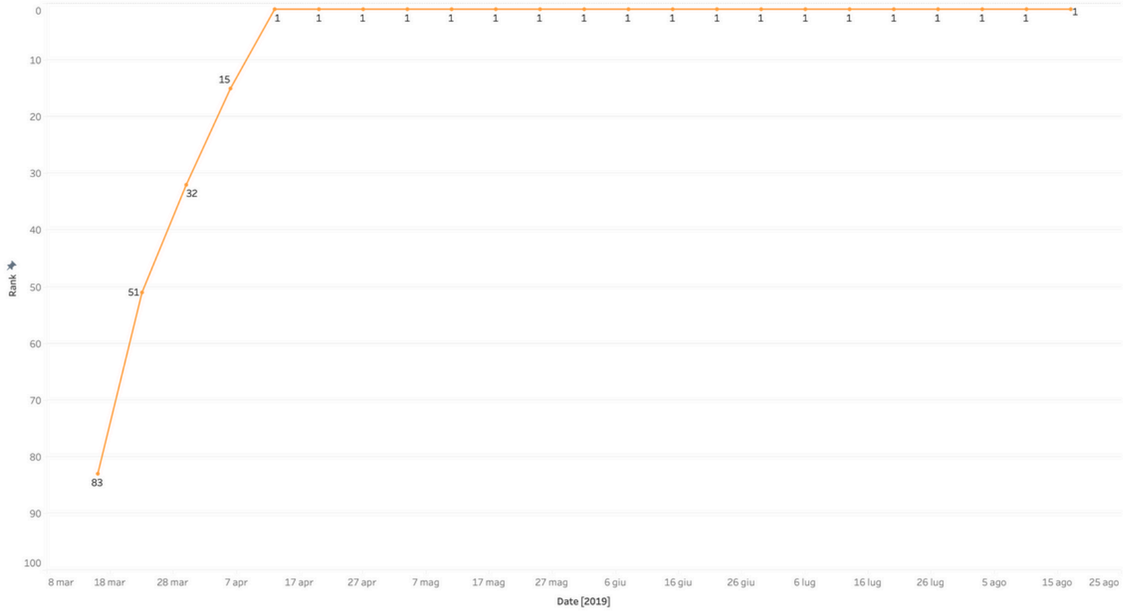
These songs achieved rare levels of popularity, dominating public attention over long periods.

Chart Trajectories of Blinding Lights and Old Town Road

These line charts illustrate the distinct Billboard Hot 100 journeys of two hits:

- **“Blinding Lights” (The Weeknd)**, which debuted outside the top 10, quickly rose to #1, and maintained a dominant position for a remarkably extended period (**90-week total run**).
- **“Old Town Road” (Lil Nas X)**, which entered the chart at position #83, reached the **#1 spot**, and remained there **for 19 consecutive weeks**, setting the all-time Billboard Hot 100 record for the **longest run at #1**.





Instant Hits:
Songs That Debuted at #1

This table lists **songs that entered the Billboard Hot 100 directly at the #1 position**, marking an immediate and powerful impact on release.

- Examples of instant chart-toppers include:
  - **"7 Rings" – Ariana Grande** (2019)
  - **"Drivers License" – Olivia Rodrigo** (2021)
  - **"God's Plan" – Drake** (2018)
- In addition to the song title and artist name, the table also displays the total number of weeks each song remained on the Billboard chart and the week when the song first charted.

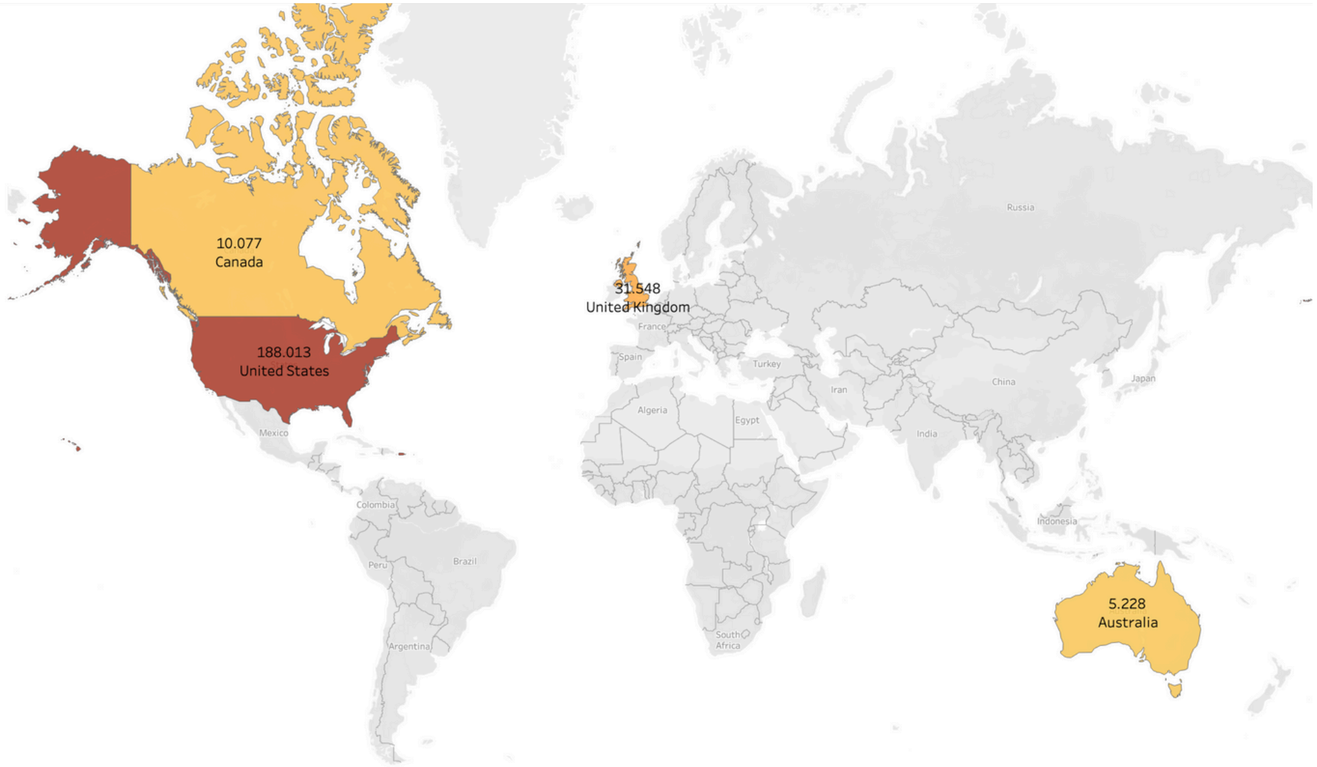
This phenomenon reflects the influence of digital releases, strong fanbases, and media anticipation in driving songs to debut at the very top of the charts.

Song	Artist Name	Max. Week..	
3	Britney Spears	20	24/10/2009
7 Rings	Ariana Grande	33	02/02/2019
Born This Way	Lady Gaga	20	26/02/2011
Butter	BTS	20	05/06/2021
Can't Stop The Feeling!	Justin Timberlake	52	28/05/2016
Candle In The Wind 1997/Somethin..	Elton John	42	11/10/1997
Cardigan	Taylor Swift	14	08/08/2020
Do I Make You Proud	Taylor Hicks	8	01/07/2006
Doo Wop (That Thing)	Lauryn Hill	21	14/11/1998
Drivers License	Olivia Rodrigo	28	23/01/2021
Dynamite	BTS	32	05/09/2020
Exhale (Shoop Shoop) (From "Waiti..	Whitney Houston	21	25/11/1995
Fantasy	Mariah Carey	25	30/09/1995
Franchise	Travis Scott	9	10/10/2020
God's Plan	Drake	36	03/02/2018
Good 4 U	Olivia Rodrigo	24	29/05/2021
Harlem Shake	Baauer	20	02/03/2013
HIGHEST IN THE ROOM	Travis Scott	22	19/10/2019
Hold It Against Me	Britney Spears	17	29/01/2011
I Believe	Fantasia	10	10/07/2004
I Don't Want To Miss A Thing	Aerosmith	20	05/09/1998
I'll Be Missing You	Puff Daddy	33	14/06/1997
I'm The One	DJ Khaled	22	20/05/2017
I'm Your Angel	R. Kelly	18	05/12/1998
Inside Your Heaven	Carrie Underwood	12	02/07/2005
Life Goes On	BTS	3	05/12/2020
Montero (Call Me By Your Name)	Lil Nas X	31	10/04/2021
My Heart Will Go On	Celine Dion	20	28/02/1998
My Universe	Coldplay	5	09/10/2021
Nice For What	Drake	25	21/04/2018
Not Afraid	Eminem	25	22/05/2010
One Sweet Day	Mariah Carey	27	02/12/1995
Part Of Me	Katy Perry	22	03/03/2012
Peaches	Justin Bieber	30	03/04/2021
Permission To Dance	BTS	7	24/07/2021
Pillowtalk	ZAYN	24	20/02/2016
Poor Little Fool	Ricky Nelson	11	04/08/1958

Top 4 Contributing Countries to the Billboard Hot 100

This map highlights the top four countries by the total number of songs their artists have contributed to the Billboard Hot 100.

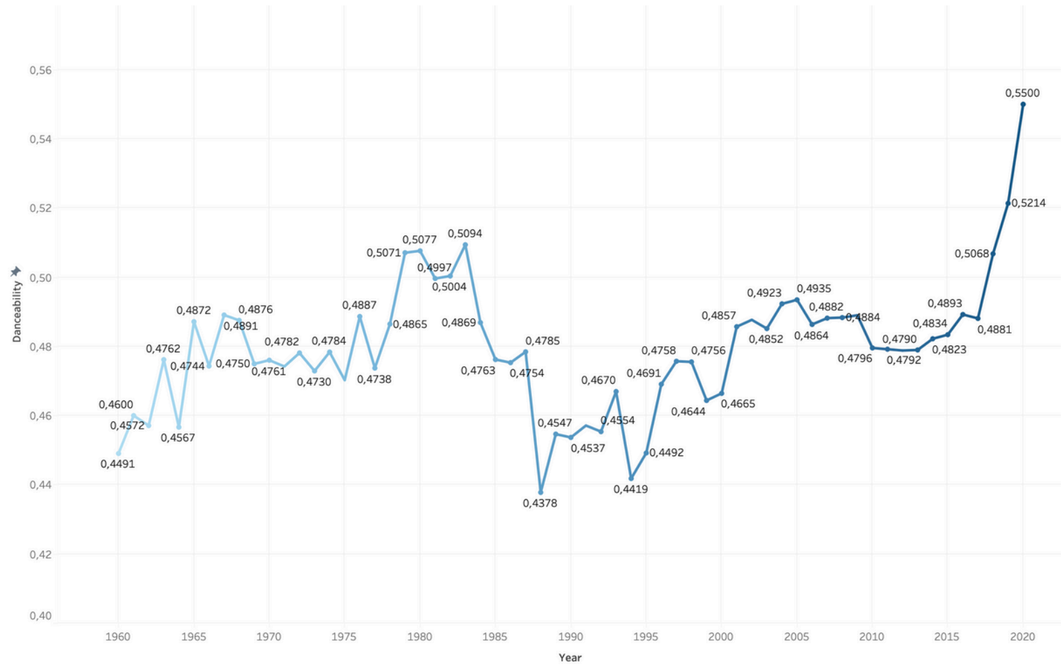
- **United States is by far the dominant contributor with 188,013 songs**, reflecting its central role in the global music industry.
- **United Kingdom follows with 31,548 songs**, known for consistently exporting successful artists and genres.
- **Canada contributed 10,077 songs**, boosted by artists like Drake, The Weeknd, and Justin Bieber.
- **Australia adds 5,228 songs**, underscoring its steady presence in international pop and rock scenes.

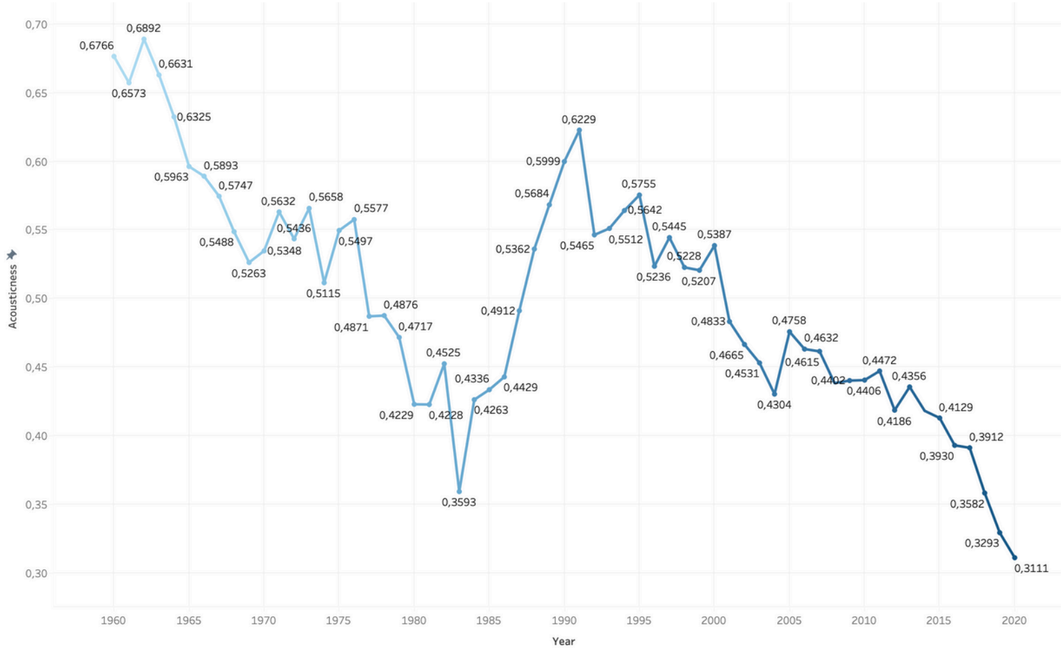


Average Danceability and Acousticness (1960-2020)

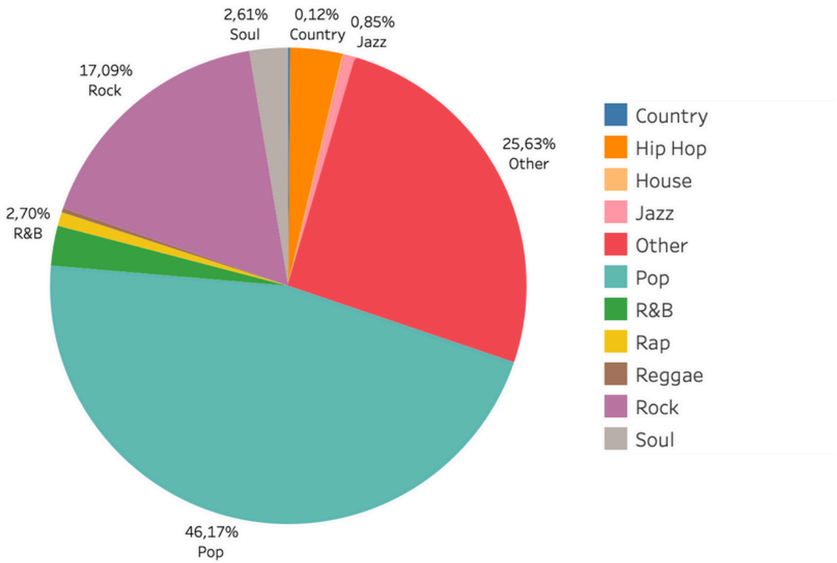
This line chart tracks the **yearly average acousticness of songs from 1960 to 2020** across the entire dataset:

- **Danceability** remained relatively **stable for decades**, reflecting consistent rhythmic appeal. **From 2015 onward, there's a marked increase**, which suggests that **modern songs are becoming more danceable**, due to the influence of hip-hop and global pop styles.
- **Acousticness was highest in the 1960s and early '70s**, indicating a dominance of organic, instrument-based music. A steady **decline begins in the late 1970s,** coinciding **with the rise of digitally produced music**. Since 2000, acousticness has dropped significantly, marking a clear shift toward electronic and digitally enhanced soundscapes.

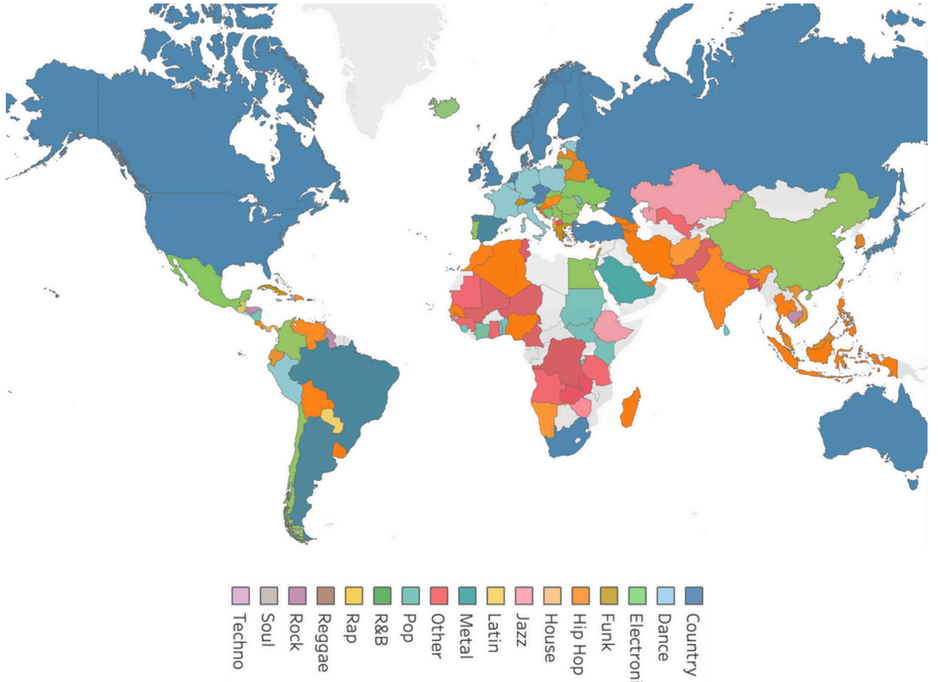




Genre Insights



Genre Distribution of #1 Billboard Hot 100 Hits

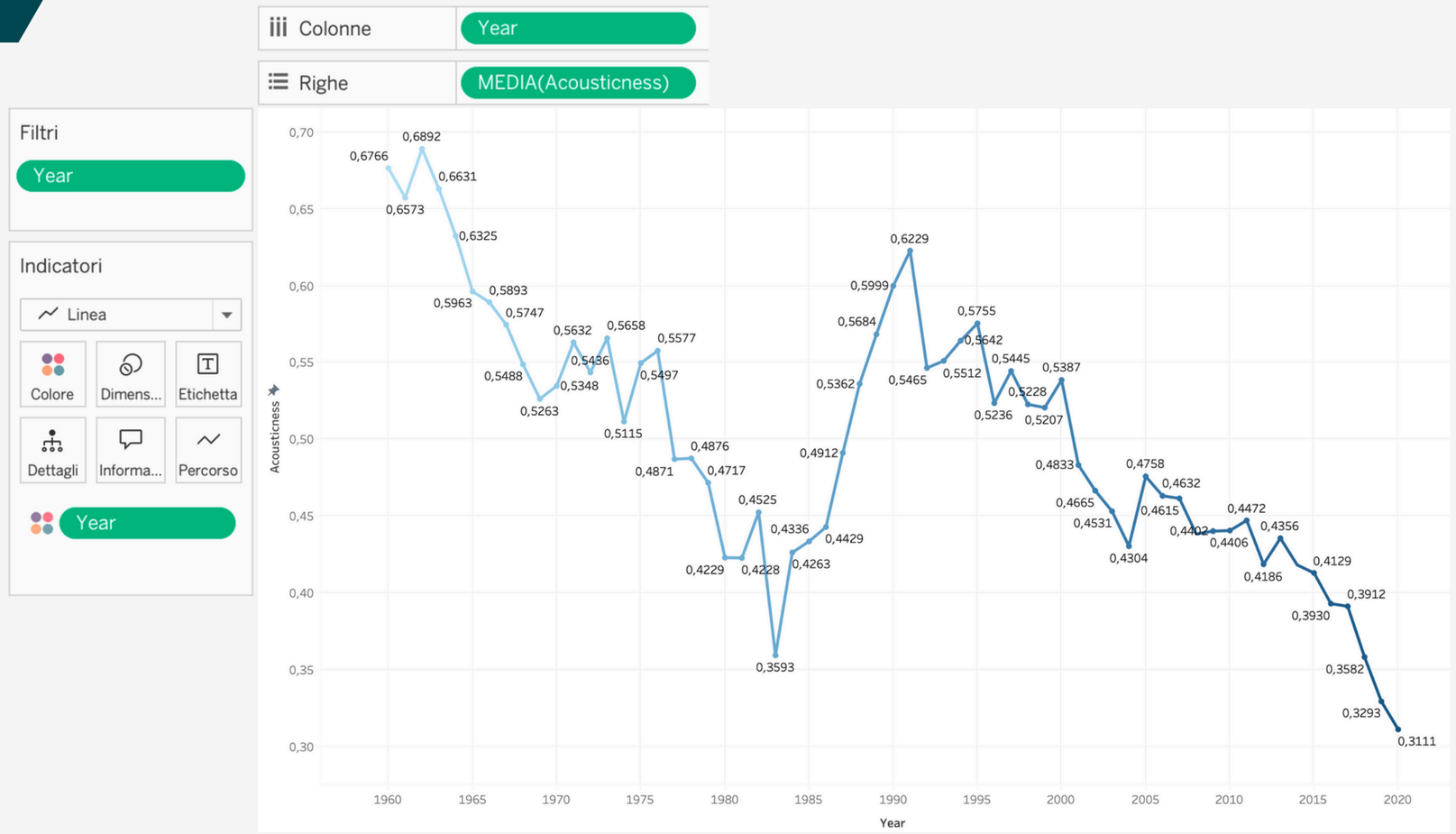


Most Represented Genre per Country (by Artist Count)

OLAP Session

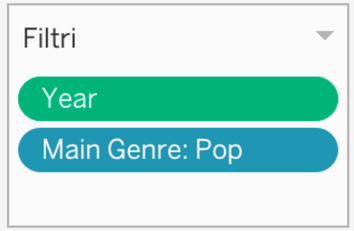
OLAP Operation - Slice

To demonstrate the **slice operation**, I took the "Average Acousticness of Songs (1960–2020)".

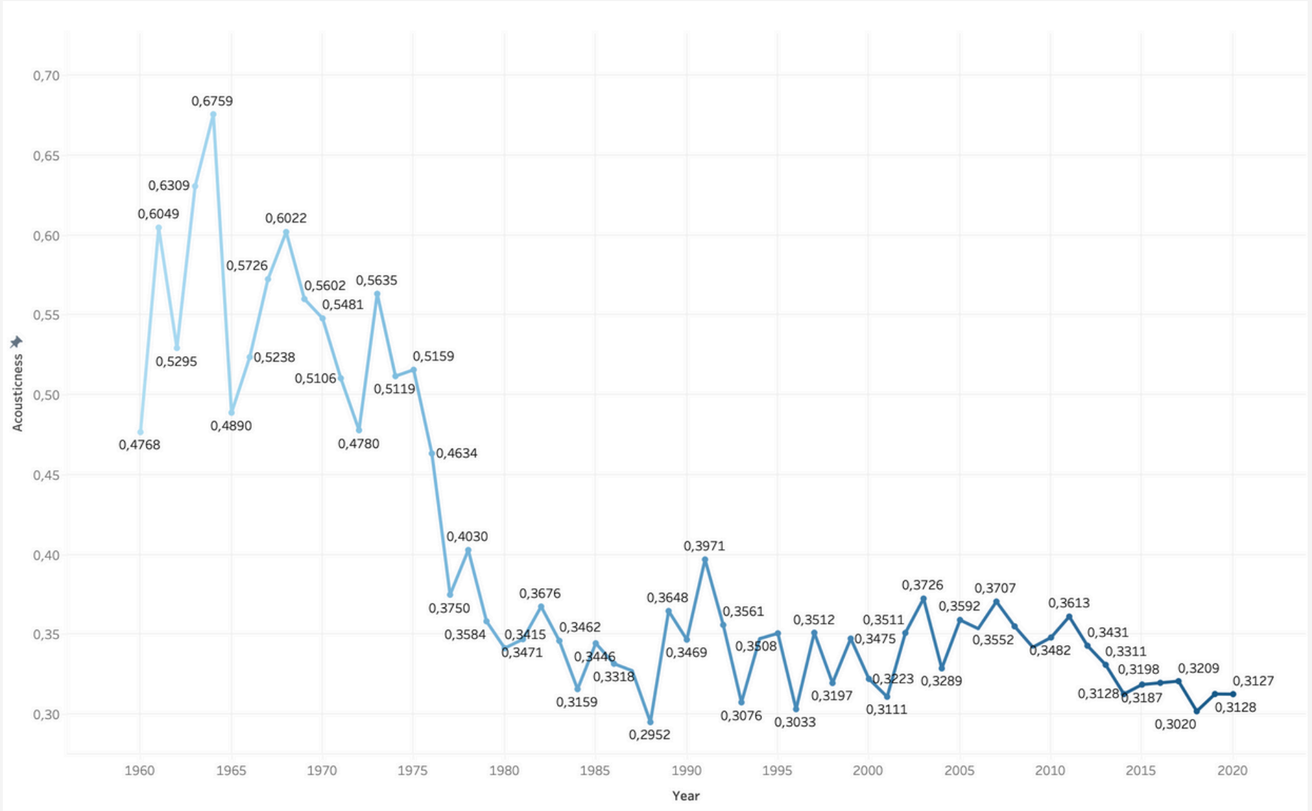


OLAP Operation - Slice

I selected only songs classified as **Pop**, filtering out all other genres. This allowed a focused analysis on how the acousticness of Pop music evolved over the decades.

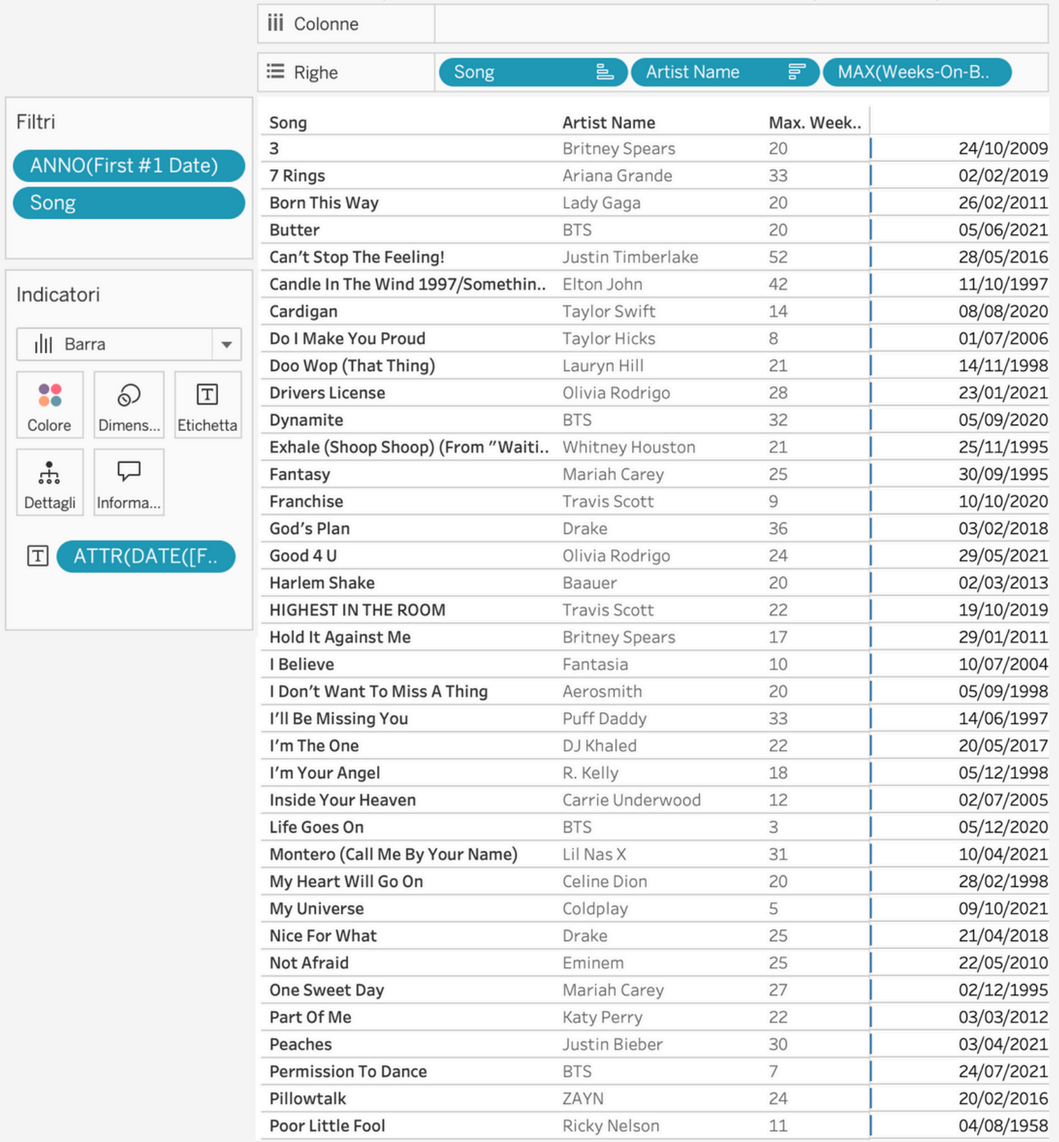


The resulting trend clearly shows a significant **decline in acousticness within Pop songs**, especially after the 1970s. This highlights a strong shift away from organic, acoustic instrumentation toward **digitally produced and electronically enhanced soundscapes**.



OLAP Operation - Dice

To demonstrate the **dice operation**, I took the "Instant Hits: Songs That Debuted at #1" chart.

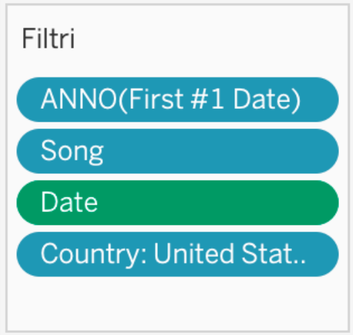


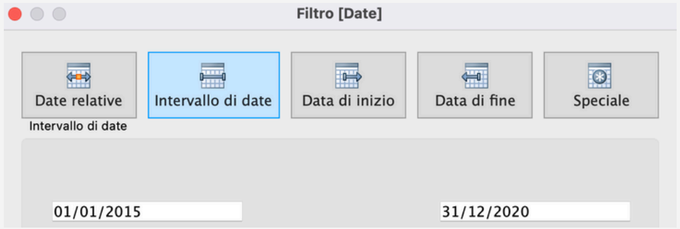
Song	Artist Name	Max. Week..	
3	Britney Spears	20	24/10/2009
7 Rings	Ariana Grande	33	02/02/2019
Born This Way	Lady Gaga	20	26/02/2011
Butter	BTS	20	05/06/2021
Can't Stop The Feeling!	Justin Timberlake	52	28/05/2016
Candle In The Wind 1997/Somethin..	Elton John	42	11/10/1997
Cardigan	Taylor Swift	14	08/08/2020
Do I Make You Proud	Taylor Hicks	8	01/07/2006
Doo Wop (That Thing)	Lauryn Hill	21	14/11/1998
Drivers License	Olivia Rodrigo	28	23/01/2021
Dynamite	BTS	32	05/09/2020
Exhale (Shoop Shoop) (From "Waiti..	Whitney Houston	21	25/11/1995
Fantasy	Mariah Carey	25	30/09/1995
Franchise	Travis Scott	9	10/10/2020
God's Plan	Drake	36	03/02/2018
Good 4 U	Olivia Rodrigo	24	29/05/2021
Harlem Shake	Baauer	20	02/03/2013
HIGHEST IN THE ROOM	Travis Scott	22	19/10/2019
Hold It Against Me	Britney Spears	17	29/01/2011
I Believe	Fantasia	10	10/07/2004
I Don't Want To Miss A Thing	Aerosmith	20	05/09/1998
I'll Be Missing You	Puff Daddy	33	14/06/1997
I'm The One	DJ Khaled	22	20/05/2017
I'm Your Angel	R. Kelly	18	05/12/1998
Inside Your Heaven	Carrie Underwood	12	02/07/2005
Life Goes On	BTS	3	05/12/2020
Montero (Call Me By Your Name)	Lil Nas X	31	10/04/2021
My Heart Will Go On	Celine Dion	20	28/02/1998
My Universe	Coldplay	5	09/10/2021
Nice For What	Drake	25	21/04/2018
Not Afraid	Eminem	25	22/05/2010
One Sweet Day	Mariah Carey	27	02/12/1995
Part Of Me	Katy Perry	22	03/03/2012
Peaches	Justin Bieber	30	03/04/2021
Permission To Dance	BTS	7	24/07/2021
Pillowtalk	ZAYN	24	20/02/2016
Poor Little Fool	Ricky Nelson	11	04/08/1958

OLAP Operation - Dice

I applied a dice operation by filtering based on multiple conditions:

- Debuted at #1
- Released between 2015 and 2020
- Artist from North America



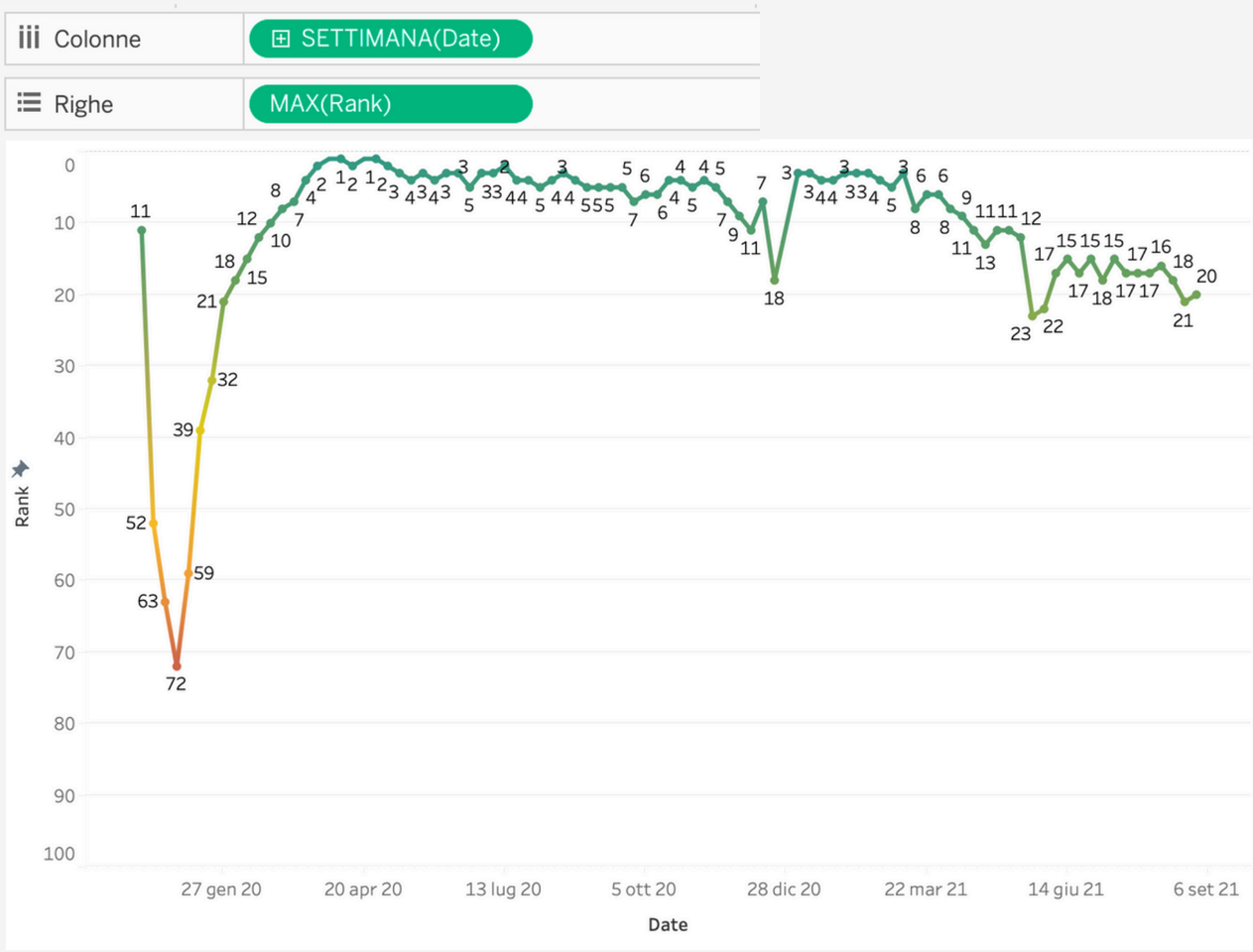


Song	Artist Name	Max. Week..	
7 Rings	Ariana Grande	33	02/02/2019
Can't Stop The Feeling!	Justin Timberlake	52	28/05/2016
Cardigan	Taylor Swift	14	08/08/2020
Dynamite	BTS	17	05/09/2020
I'm The One	DJ Khaled	22	20/05/2017
Life Goes On	BTS	3	05/12/2020
Positions	Ariana Grande	8	07/11/2020
Rain On Me	Lady Gaga	20	06/06/2020
Shake It Off	Taylor Swift	50	06/09/2014
Stuck With U	Ariana Grande	18	23/05/2020
Sucker	Jonas Brothers	47	16/03/2019
Thank U, Next	Ariana Grande	28	17/11/2018
This Is America	Childish Gambino	17	19/05/2018
Trollz	6ix9ine	4	27/06/2020
WAP	Cardi B	19	22/08/2020
Willow	Taylor Swift	1	26/12/2020

This subset highlighted the power of social media and fanbases in pushing North American artists (e.g., Taylor Swift, Drake) to instant chart-topping debuts in recent years.

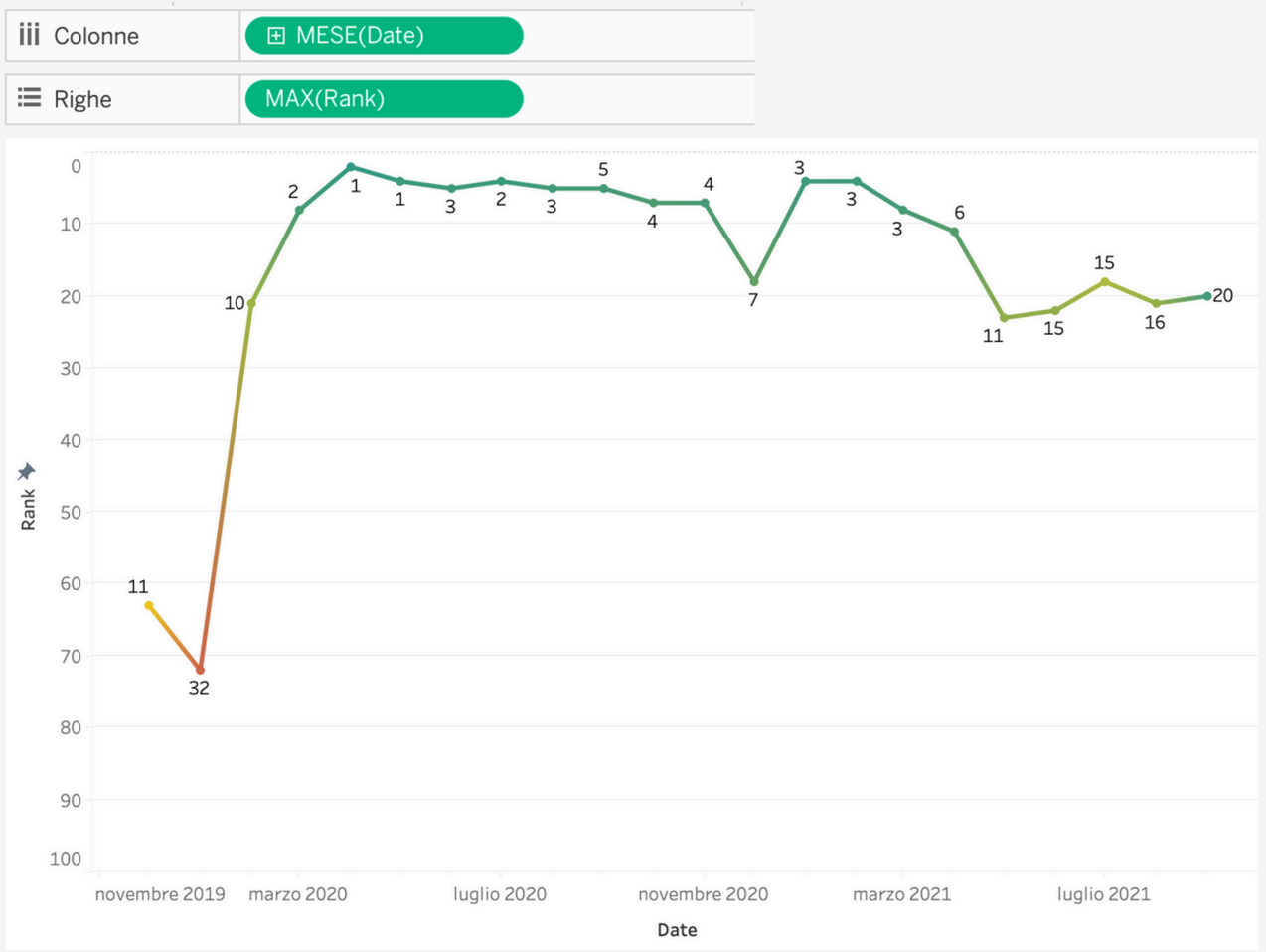
OLAP Operation - Roll-Up

To demonstrate the **roll-up operation**, I took the "Chart Trajectory of Blinding Lights" and visualize the best rank over each week.



OLAP Operation - Roll-Up

I applied a roll-up operation by **aggregating the results on month** (instead of week), in order to visualize the **best rank obtained in each month**.



Thank you for your attention!