Data analysis project

SPOTIFY DATA

SABRINA BERMÚDEZ

Data analysis portfolio - 2025

OBJECTIVE - SCOPE - LIMITATIONS

Objectives

- Finding and analyzing different insights that help understand trends and behaviours within the dataset

Scope

- The project aims to showcase data analysis skills using different tools and data-driven decision making and analysis

Limitations

- The data does not have timestamps or stream counts.
- The genre classification can be inconsistent

DATA DESCRIPTION

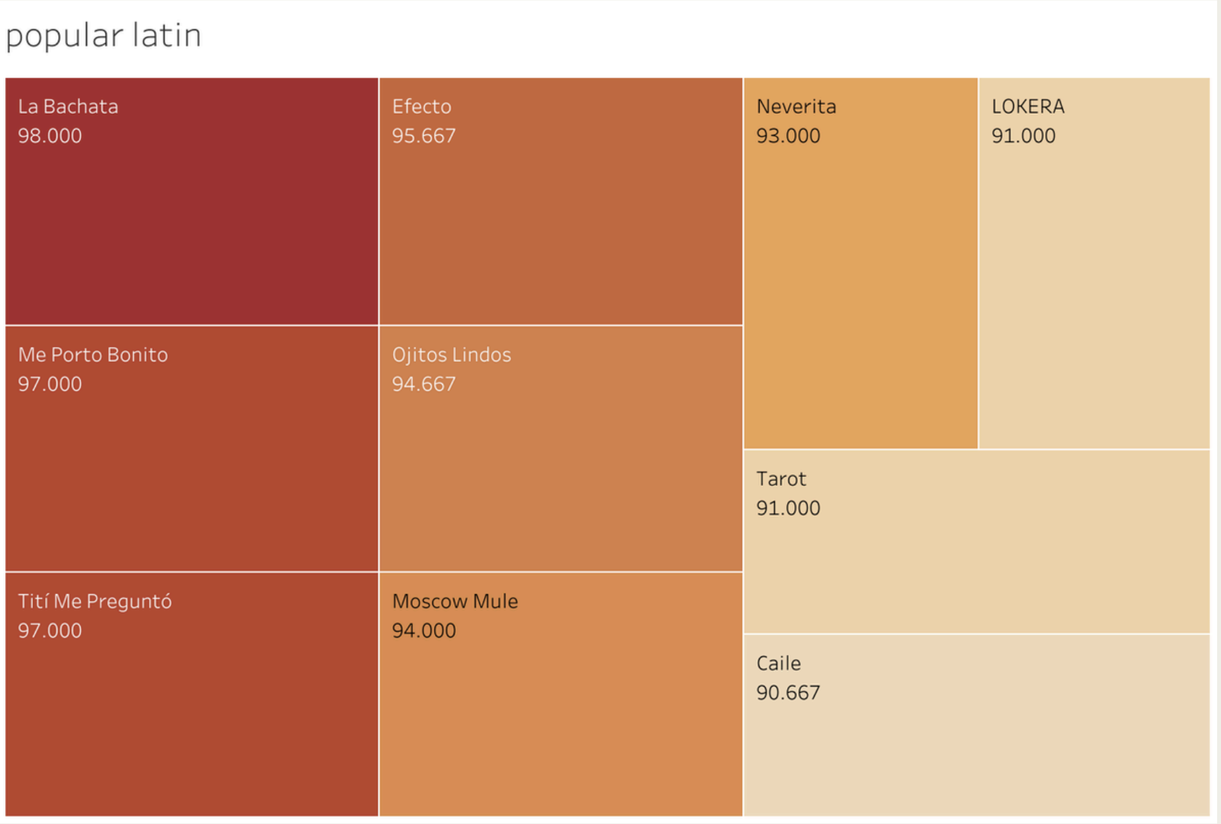
The data consists of one csv file with 21 columns presented as long data. The data provides information like track name, artist, duration in miliseconds and follows up with song attributes, such as genre, tempo, danceability, energy, loudness, liveness, valence and more.

The dataset I used for this project is "Spotify Tracks Dataset (Kaggle, Maharshi Pandya)", which is openly available on Kaggle. The available tools available for this project are BigQuery, Tableau Public, Excel, RStudio (free/online versions)

INSIGHTS AND VISUALS

MOST POPULAR LATIN SONGS

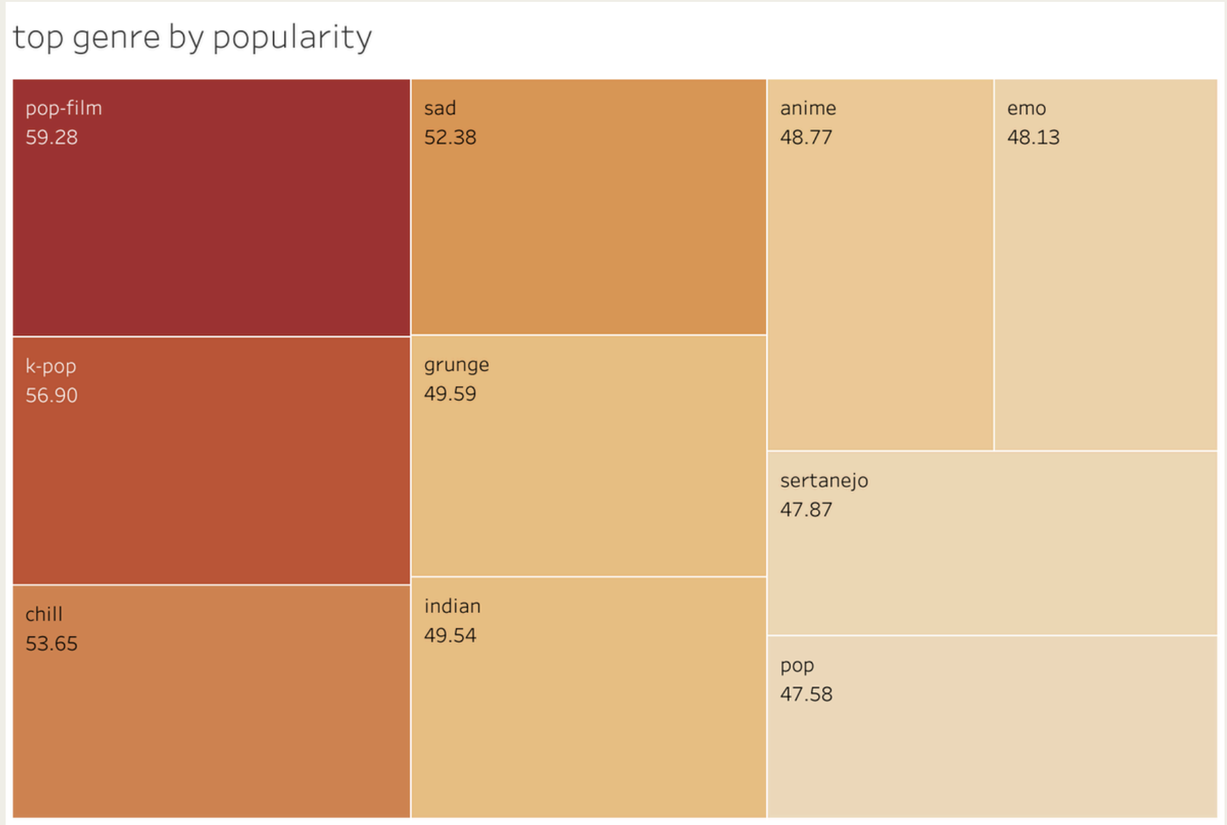
By filtering the genre to only show “latin”, “latino”, “spanish” and “reggaeton” and sorting from most to least popular, I obtained the top 10 most popular songs for this genre in Tableau Public.



INSIGHTS AND VISUALS

TOP GENRE BY POPULARITY

By sorting genre by average song duration from longest to shortest, I obtained the top 10 genres with longest songs, on average, in Tableau Public.



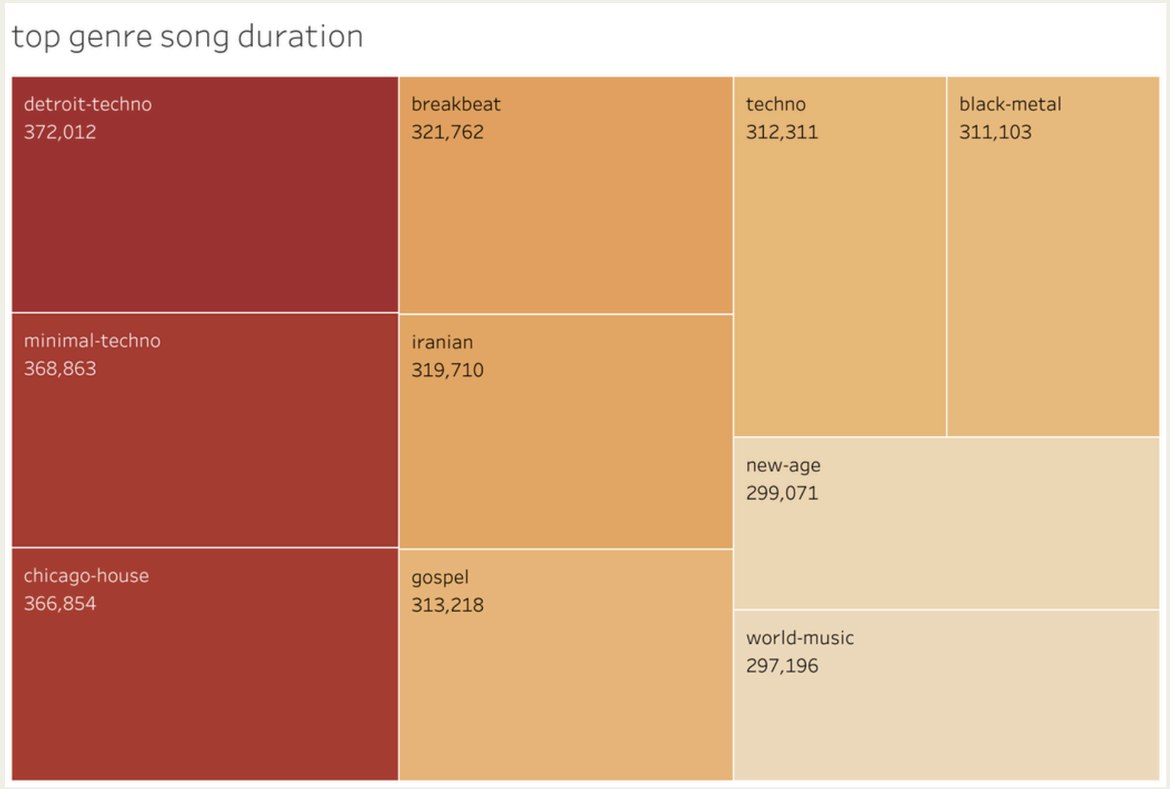
I also made a query in BigQuery, obtaining the same results

Row	track_genre	avg_popularity
1	pop-film	59.28
2	k-pop	56.9
3	chill	53.65
4	sad	52.38
5	grunge	49.59
6	indian	49.54
7	anime	48.77
8	emo	48.13
9	sertanejo	47.87
10	pop	47.58

INSIGHTS AND VISUALS

TOP GENRE BY AVERAGE SONG DURATION

By sorting genre by average song duration from longest to shortest, I obtained the top 10 genres with longest songs, on average, in Tableau Public.



I also made a query in BigQuery, obtaining the same results

Row	track_genre	duration
1	detroit-techno	372012.4
2	minimal-techno	368863.25
3	chicago-house	366853.87
4	breakbeat	321762.22
5	iranian	319709.54
6	gospel	313218.04
7	techno	312311.48
8	black-metal	311103.11
9	new-age	299070.68
10	world-music	297195.62

INSIGHTS AND VISUALS

ATTRIBUTE CORRELATION

I wanted to see if there is a correlation between some song attributes and how popular the songs are, so I made a SQL query that would help me see if there is any relation between these variables. The results are close to zero, which indicate there is no relationship between the variables

Row	danceability_corr	energy_corr	tempo_corr	valence_corr	speechiness_corr	instrumentalness_co
1	-0.06963	-0.08232	0.03509	-0.08772	-0.02165	0.0161

SQL HIGHLIGHTS

For more information about the queries that were used in each section please visit:

https://docs.google.com/document/d/1ZtsyyAkyRAYshdvfmn1MxMkLodAI-9D38cgh5plivbk/edit?usp=sharing

CONCLUSIONS AND NEXT STEPS

While I obtained valuable information of the data we obtained, a dataset that includes attributes related to the stream count and the dates could have provided me valuable insights about the growth of an artist or a specific genre during a certain time period.

For next steps, I would revise the genres that are available and grouping them into more general groups, for example, “latin”, “latino”, “reggaeton” and “spanish” as “latin”.

Some valuable insights I would look into if I had more information would be how much the top 10 artists that fit into the “latin” genre have grown in the last five years.

Also, I would look into more attributes that could be related to the popularity of the songs, such as how many videos have been made with the song audio on TikTok, which is a social media that can increase a song or artists popularity at a fast pace.

Thank you!

SABRINA BERMÚDEZ