Music Popularity Analysis

•••

Jonathan Sada

DataSet Details

Top Spotify Songs in 73 Countries

This dataset contains the Daily top 50 songs on Spotify. The data is updated daily and includes various features such as song duration, artist details, album information, and song popularity.
https://www.kaggle.com/datasets/asaniczka/top-spotify-songs-in-73-countries-daily-updated

Country and Continent Codes List

This dataset contains a relation between continents and country ISO Codes and names.
https://gist.github.com/stevewithington/20a69c0b6d2ff846ea5d35e5fc47f26c

Final Dataset

- The data set has a total of 2.052.083 rows and 26 characteristics

- 23.898 unique songs (1.16%)

- 17 numerical features

- 9 categorical features

- 455.7 MB

	spotify_id	name	artists	daily_rank	daily_movement	weekly_movement	country	snapshot_date	popularity	is_explicit	...
0	2RkZ5LkEzeHGRsmDqKwmaJ	Ordinary	Alex Warren	1	0	1	NaN	2025-05-18	93	False	...
1	7so0lgd0zP2Sbgs2d7a1SZ	Die With A Smile	Lady Gaga, Bruno Mars	2	0	-1	NaN	2025-05-18	92	False	...
2	6dOtVTDdiauQNBQEDOtlAB	BIRDS OF A FEATHER	Billie Eilish	3	0	0	NaN	2025-05-18	100	False	...
3	4wJ5Qq0jBN4ajy7ouZIV1c	APT.	ROSÉ, Bruno Mars	4	2	0	NaN	2025-05-18	90	False	...
4	6iOndD4OFo7GkaDypWQIou	La Plena - W Sound 05	W Sound, Beéle, Ovy On The Drums	5	-1	1	NaN	2025-05-18	94	True	...
...	...	...	...	...	...	...	...	...	...	...	...
2023914	0AYt6NMyyLd0rLuvr0UkMH	Slime You Out (feat. SZA)	Drake, SZA	46	4	0	AE	2023-10-18	84	True	...
2023915	2Gk6fi0dqt91NKvlzGsmm7	SAY MY GRACE (feat. Travis Scott)	Offset, Travis Scott	47	3	0	AE	2023-10-18	80	True	...
2023916	26b3oVLrRUaaybJulow9kz	People	Libianca	48	2	0	AE	2023-10-18	88	False	...
2023917	5ydjxBSUIDn26MFzU3asP4	Rainy Days	V	49	1	0	AE	2023-10-18	88	False	...
2023918	59NraMJsLaMCVtwXTSia8i	Prada	cassö, RAYE, D-Block Europe	50	0	0	AE	2023-10-18	94	True	...

2023919 rows × 25 columns

Analysis

Analysis Overview - Univariate Analysis*

Categorical Values

- is_explicit
- continent
- mode
- popularity
- key
- time_signature

Numerical Values

- duration
- danceability
- energy
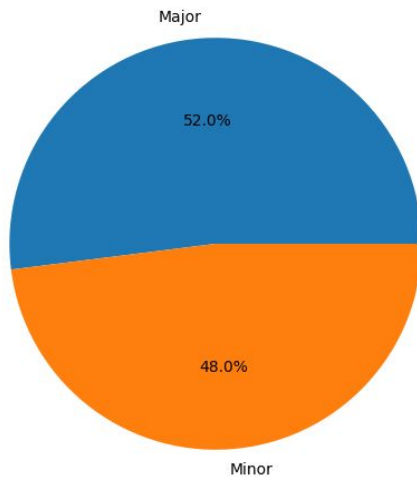
Not Finished

- loudness
- speechiness
- acousticness
- instrumentalness
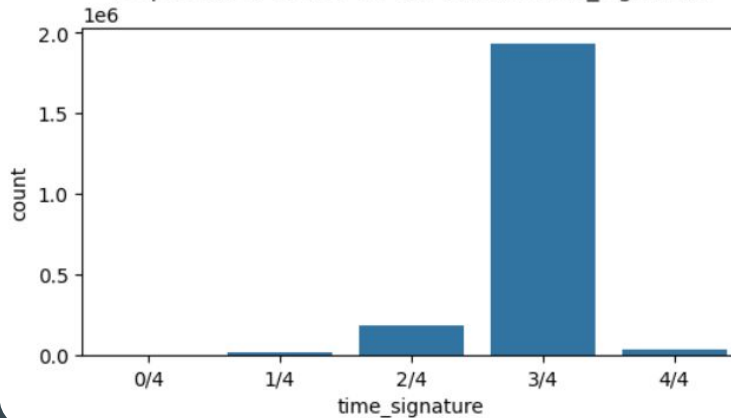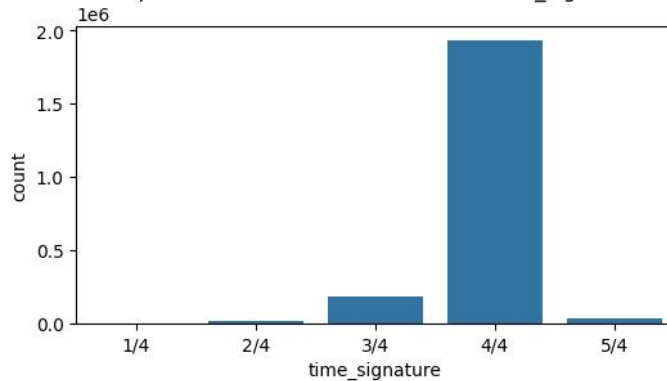- liveness
- valence
- tempo

Univariate Analysis
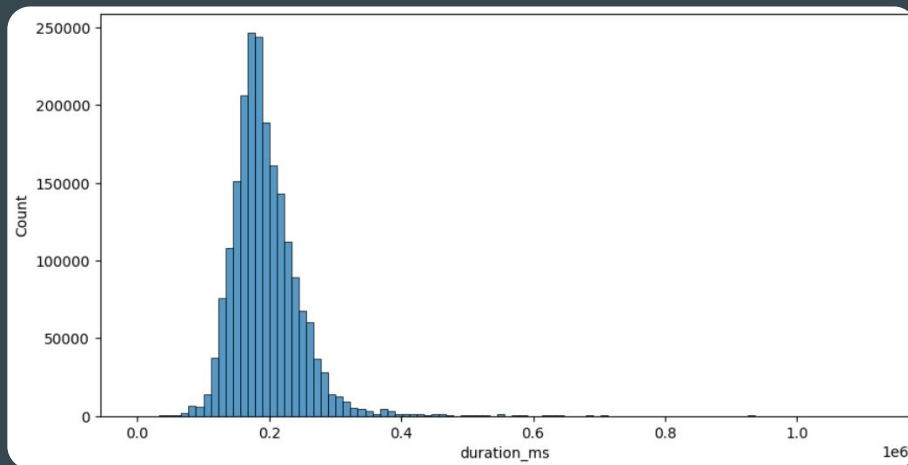

Proportion of values for the feature mode

Proportion of values for the feature time_signature

Proportion of values for the feature time_signature

Univariate Analysis

- Majority are in between 2.7min (162018 ms) and 3.64min (218423.0 ms)

- The most repeated is 3.5min (210373 ms)

- Longest song is 18.57min (1114155 ms) so 25% of the values goes from 3.64 to 18.57min.

Analysis Overview - Bivariate Analysis*

Categorical vs Categorical

- continent - is_explicit
- key - mode
- key - time_signature
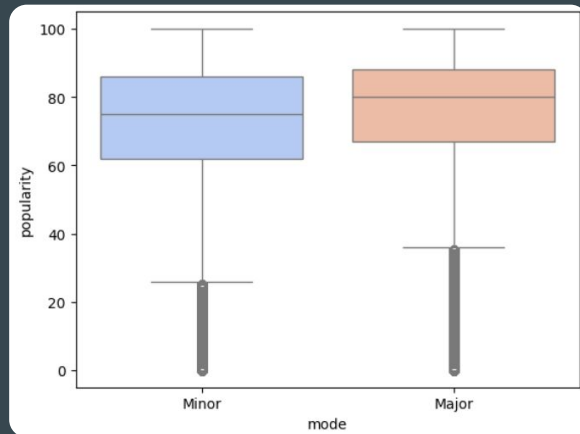- mode - time-signature
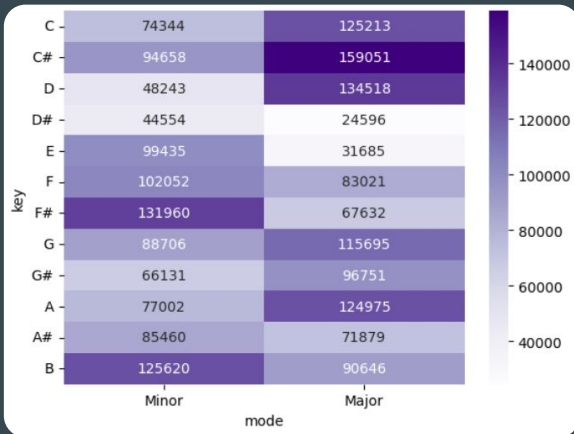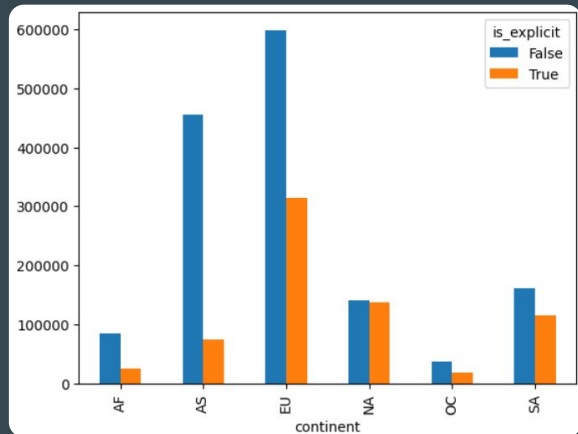
Categorical vs Continuous

- popularity - continent
- popularity - mode
- popularity - key

Categorical vs Continuous

- popularity - duration_ms
- popularity - danceability
- popularity - energy
- popularity - loudness
- popularity - speechiness
- popularity - instrumentales
- popularity - liveness
- popularity - valence
- popularity - tempo

Bivariate Analysis

Questions

Questions*

1. How do song properties affects their popularity?

2. How popular are acoustic songs compared to the AVG?

3. How popular are instrumental songs compared to the AVG?

4. How popular are live songs compared to the AVG?

5. What are the most popular songs?

6. What are the 5 most popular songs in Europe

7. What are the most popular songs songs per continent?

8. What are the 5 most popular artist?

9. What are the properties of the songs of the most popular artists?

10. How does song properties evolve across the time? (properties vs release date)

11. How does the time impact popularity? (popularity vs release date)

12. How many albums were released by the most popular artists?

13. When was the release of the first album of the most popular artist?

How do song properties affect its popularity?

- **duration_ms**
  (parson: 0.03 / spearman 0.07)

- **danceability**
  (parson: -0.05 / spearman -0.05)

- **energy**
  (parson: -0.01 / spearman -0.02)

- **loudness**
  (parson: 0.02 / spearman 0.11)

- **speechiness**
  (parson: -0.15 / spearman -0.17)

- **instrumentales**
  (parson: 0.04 / spearman 0.06)

- **liveness**
  (parson: -0.04 / spearman -0.03)

- **valence**
  (parson: -0.02 / spearman -0.03)

NOT CORRELATED

How popular are acoustic/instrumental/live songs?

```sql
-- 2. How popular are acoustic songs?
SELECT (SELECT AVG(popularity) as popularity FROM songs
WHERE acousticness >= 0.9) AS acoustic_popularity,(SELECT
AVG(popularity) as popularity FROM songs) AS avg_popularity;

-- 3. How popular are instrumental songs?
SELECT (SELECT AVG(popularity) as popularity FROM songs
WHERE instrumentalness >= 0.9) AS instrumental_popularity ,
(SELECT AVG(popularity) as popularity FROM songs) AS
avg_popularity;

-- 4. How popular are live songs?
SELECT (SELECT AVG(popularity) as popularity FROM songs
WHERE liveness >= 0.9) AS live_popularity , (SELECT
AVG(popularity) as popularity FROM songs) AS avg_popularity;
```
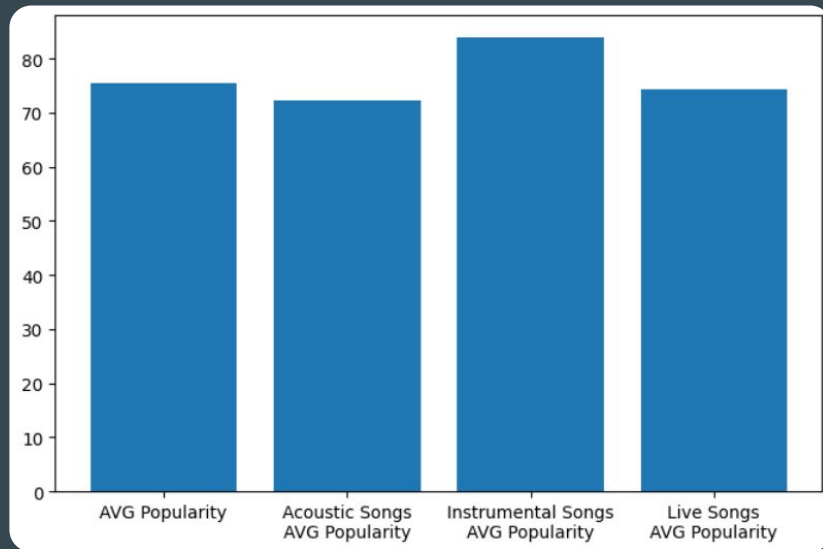
What's the most popular song per continent?

```sql
SELECT s.continent, s.name, s.artists, MAX(s.avg_popularity)as popularity
FROM (
        SELECT s1.spotify_id, s1.name, s1.artists, AVG(s1.popularity) as avg_popularity, s1.continent
        FROM songs as s1
        GROUP BY s1.spotify_id, s1.continent) as s
GROUP BY s.continent;
```

	continent	name	artists	popularity
1	AF	greedy	Tate McRae	99.1404494382023
2	AS	Beautiful Things	Benson Boone	98.9597855227882
3	EU	Die With A Smile	Lady Gaga, Bruno Mars	98.4424254016832
4	NA	Die With A Smile	Lady Gaga, Bruno Mars	98.4401098901099
5	OC	Die With A Smile	Lady Gaga, Bruno Mars	98.2984293193717
6	SA	Beautiful Things	Benson Boone	99.9874213836478

# What are the most popular artist songs' characteristics?

```sql
SELECT ("duration_ms" / 60000) as duration_min, "is_explicit", "danceability", "energy", "key",
"loudness", "mode", "speechiness", "acousticness", "instrumentalness", "liveness", "valence", "tempo",
"time_signature"
FROM songs WHERE artists = (
    SELECT s2.artists
    FROM songs as s2
    GROUP BY s2.artists
    ORDER BY AVG(s2.popularity) DESC
    LIMIT 1
) GROUP BY spotify_id;
```

|	duration_min	is_explicit	danceability	energy	key	loudness	mode	speechiness	acousticness	instrumentalness	liveness	valence	tempo	time_signature
1	2	0	0.613	0.822	9	-4.087	1	0.026	0.0092	0.506	0.429	0.567	105.997	4
2	2	0	0.504	0.308	9	-14.958	1	0.0321	0.868	0.135	0.158	0.121	113.95	4

What's the evolution of the music characteristics over time?

```sql
SELECT strftime('%Y', datetime("album_release_date")) as year,(AVG("duration_ms") / 60000) as "avg_duration_min",
AVG("danceability") as "avg_danceability",
AVG("energy") as "avg_energy",
AVG("loudness") as "avg_loudness",
AVG("speechiness") as "avg_speechiness",
AVG("acousticness") as "avg_acousticness",
AVG("instrumentalness") as "avg_instrumentalness",
AVG("liveness") as "avg_liveness",
 AVG("valence") as "avg_valence",
AVG("tempo") as "avg_tempo"
FROM songs
GROUP BY year
ORDER BY year ASC;
```
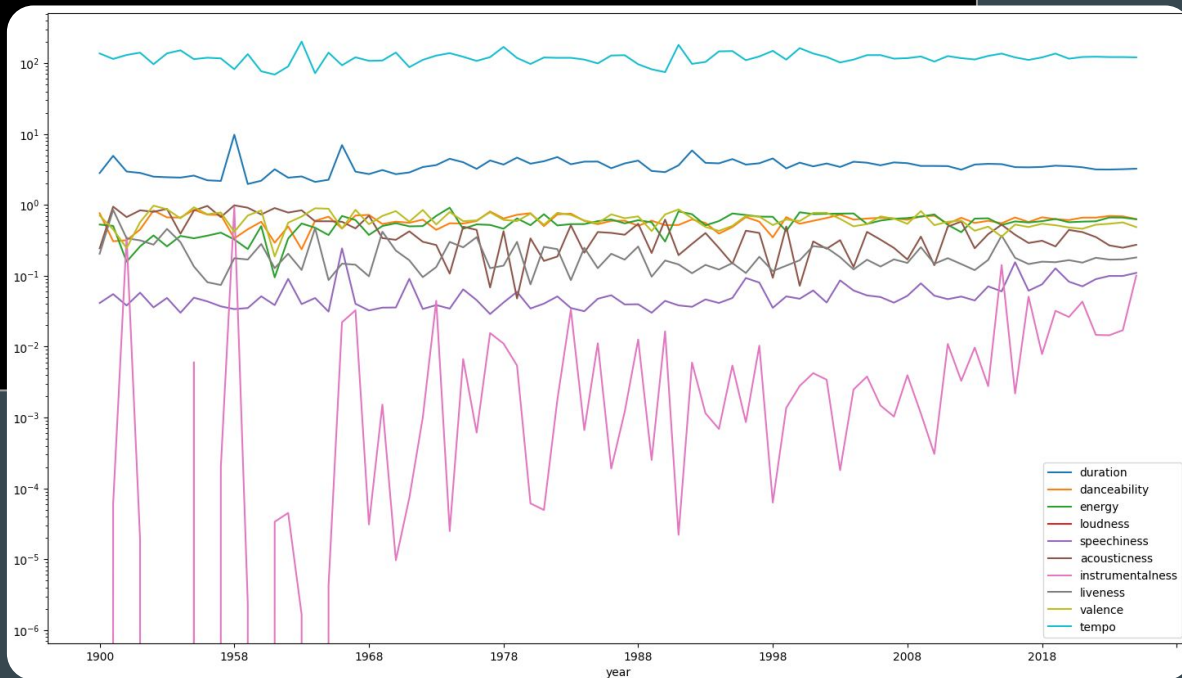
Whats the evolution of the music popularity? (release based)

```sql
SELECT  strftime('%Y',
datetime("snapshot_date")) as pop_year,
strftime('%Y',
datetime("album_release_date")) as
release_year, AVG(popularity) as popularity
FROM songs
GROUP BY pop_year, release_year
ORDER BY pop_year ASC, release_year ASC;
```
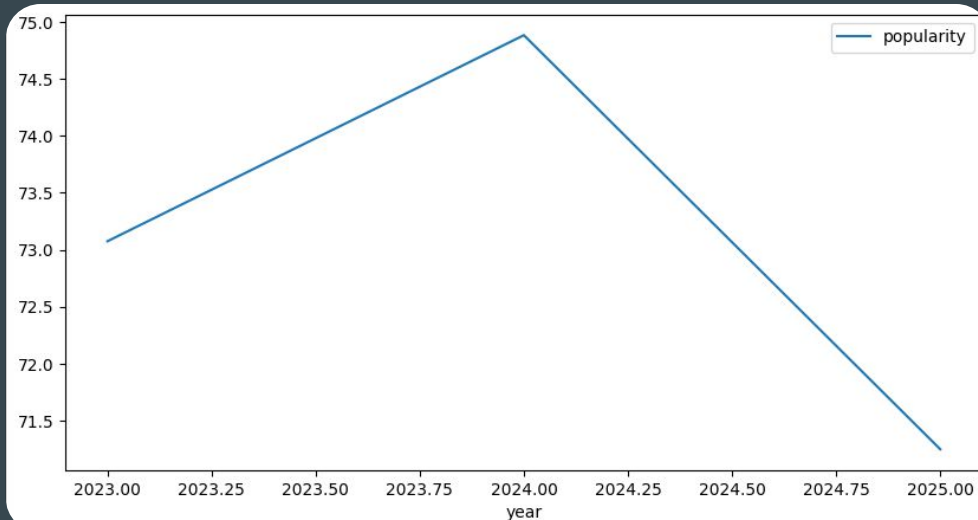

*Music released between 1980 and 1990

Learnings & Next Steps

Learnings & Next Steps

Learnings:

- How pandas import CSV and creates NA values
- Diferenciate Nominative, Ordinal, Discrete and Continuous variables
- Working with Pandas and Databases
- VSCode Outliner

Next Steps:

- Finish analysis left.
- Split dataset (songs features, songs popularity).
- Include genre on the dataset.
- Extracting data from a music service API.

Thank you