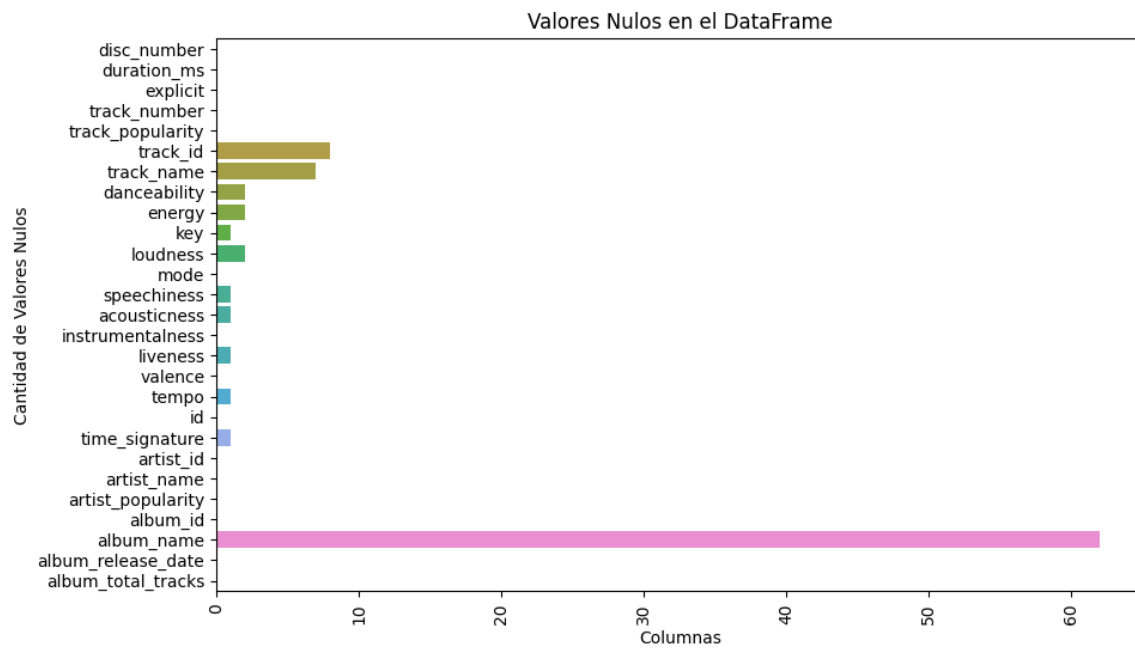


Terminado el análisis de calidad de datos del archivo dataset.csv, podemos visualizar las siguientes detecciones y/o conclusiones:

Visualización de valores faltantes en las distintas variables del dataset:

```
disc_number      0
duration_ms      0
explicit         0
track_number     0
track_popularity 0
track_id         8
track_name       7
danceability     2
energy           2
key              1
loudness         2
mode             0
speechiness      1
acousticness     1
instrumentalness 0
liveness         1
valence          0
tempo            1
id               0
time_signature   1
artist_id        0
artist_name      0
artist_popularity 0
album_id         0
album_name       62
album_release_date 0
album_total_tracks 0
dtype: int64
```



Valores duplicados:

disc_number	duration_ms	explicit	track_number	track_popularity	\
1	234466	True	21	82	
1	170640	False	1	77	
1	221306	False	3	92	
1	190360	False	4	86	
1	211240	False	5	82	
1	173386	False	6	78	
1	234146	False	7	83	
1	222400	False	8	86	
1	287266	False	9	81	
1	198533	False	10	79	
1	190240	False	11	80	
1	201586	False	12	72	
1	200306	False	13	78	
1	171360	False	14	84	
1	223293	False	15	82	
1	193000	False	16	80	
1	150440	False	17	72	
1	293453	False	18	85	

track_id	track_name
3xYJScVfxByb61dYHTwiby	Hits Different
43rA71bccXFGD4C8G0pIIN	I Forgot That You Existed
1dGr1c8CrMLDpV6mPbImSI	Lover
3RauEVgRgj1IuWdJ9fDs70	The Man
3pHkh7d01zM2AlDutz2x37	The Archer
2YWtcWi3a83pdEg3Gif4Pd	I Think He Knows
214nt20w5wOxJnY462k1Lw	Miss Americana & The Heartbreak Prince
4y5bvROuBDPr5fuwXbIBZR	Paper Rings
12M5uqx0ZuwkpLp5rJim1a	Cornelia Street
2dgFqt3w9xIQRjhPtWnk3D	Death By A Thousand Cuts
1LLXZFeAHK9R4xUramtUKw	London Boy
4AYtqFyFbX0Xkc2wtcygTr	Soon You'll Get Better (feat. The Chicks)
5hQ5XkfGbxjZ09uCwd11so	False God
6RRNNciQGZEXnqk8S09yv5	You Need To Calm Down
1SymEzIT3H8UZfibCs3TYi	Afterglow
2Rk4JJ1nc2TPmZe2af99d45	ME! (feat. Brendon Urie of Panic! At The Disco)
1SmiQ65iSAbPto6gPF1BYm	It's Nice To Have A Friend
1fzAuUVbz1hZ11JAx9PtY6	Daylight

[illegible]

Columna: explicit

Número de valores únicos: 4

explicit

False 480

True 54

No 4

Si 1

Name: count, dtype: int64

Columna: track_id

Número de valores únicos: 512

track_id

1fzAuUVbz1hZ11JAx9PtY6 2

2dgFqt3w9xIQRjhPtWnk3D 2

1dGr1c8CrMLDpV6mPbImSI 2

3pHkh7d01zM2A1dUtz2x37 2

2YWtcWi3a83pdEg3Gif4Pd 2

..

01QdEx6kFr78ZejhQtWR5m 1

0tQ9vBYpldCuikPsbG0VKA 1

2m30bD945KvpE5y9A1eUWm 1

3ExweHKZF9B752DPQByRVT 1

1hxLyjC9D9Jpw6EAPKqWv4 1

Name: count, Length: 512, dtype: int64

Columna: track_name

Número de valores únicos: 331

track_name

The Story Of Us 4

Sweet Nothing 3

Lavender Haze 3

Question...? 3

Midnight Rain 3

..

Don't You (Taylor's Version) (From The Vault) 1

Bye Bye Baby (Taylor's Version) (From The Vault) 1

right where you left me - bonus track 1

it's time to go - bonus track 1

Teardrops on My Guitar - Pop Version 1

Name: count, Length: 331, dtype: int64

Columna: instrumentalness

Número de valores únicos: 240

instrumentalness

0 235

2.06e-05 4

2.67e-05 3

1.64e-06 3

2.46e-05 3

...

1.02e-05 1

1.86e-06 1

0.00218 1

9.19e-05 1

0.000807 1

Name: count, Length: 240, dtype: int64

```

-----
Columna: id
Número de valores únicos: 519
id
4y5bvR0uBDPr5fuwXbIBZR      2
1SymEzIT3H8UZfibCs3TYi      2
5hQ5XkFgbxjZ09uCwd11so      2
4AYtqFyFbX0Xkc2wtcygTr      2
1LLXZFeAHK9R4xUramtUKw      2
..
51A8eKvvZz9uydvIZ7xRSV      1
1n2wszmJyVkw6FHqyLnQsY      1
01QdEx6kFr78ZejhQtWR5m      1
0tQ9vBYpldCuikPsbgOVKA      1
1hxLyjC9D9Jpw6EAPKqWv4      1
Name: count, Length: 519, dtype: int64
-----
Columna: artist_id
Número de valores únicos: 1
artist_id
06HL4z0CvFAxyc27GX          539
Name: count, dtype: int64
-----
Columna: artist_name
Número de valores únicos: 1
artist_name
Taylor Swift                  539
Name: count, dtype: int64
-----

-----
Columna: album_id
Número de valores únicos: 26
album_id
1MPAXuTVL2Ej5x0JHiSPq8      46
1NAmidJ1EaVgA3MpcPFYGq      36
0PZ71Aru5FDFHuirTkWe9Z      34
6kZ42qRrzov54LcAk4onW9      30
4hDok00AJd57SGIT8xuWJH      26
1fnJ7k0b11NfL1kVdNVW1A      24
6S6JQWzUrJVcJLK4fi74Fw      22
1KVKqWeRuXsJDLTW0VuD29      22
1o59UpKw81iHR0HPiSkJR0      22
5AEDGb1iTTfj0B8Tsm1sxt      22
64LU4c1nfjz1t4VnGhagcg      21
31S1y25WAhcqJDATJK70Mq      20
1yGbN0tRIgdIiGHOEBaZWf      19
2gP2LMVcIFgVczSJqn340t      19
1pzvBxYgT60VwJLthKrdQK      17
6AORtDjduMM3bupSWzbTSG      17
2fenSS68JI1h4Fo296JfGr      16
6DEjYfKNZh67HP7R9PSZvv      16
08CWGiv27MVQhYpuTtvx83      16
6fyR4wBPwLHKcRtxgd4sGh      16
2Xoteh7uEpea4TohMxjtaq      15
5eyZZoQEFQWRHkV2xgAeBw      15
6Ar2o9KCqcyYF9J0aQP3au      14
151w1FgRZfnKZA9FEcg9Z3      13
5fy0X0JmZRZnVa2UEicIO1      13
1ycoesYxIFymXWebfmz828      8
Name: count, dtype: int64
-----

-----
Columna: album_name
Número de valores únicos: 24
album_name
Lover                        36
folklore: the long pond studio sessions (from the Disney+ special) [deluxe edition] 34
Red (Taylor's Version)      30
Fearless (Taylor's Version) 26
Midnights (The Til Dawn Edition) 24
1989 (Taylor's Version) [Deluxe] 22
Red (Deluxe Edition)        22
Speak Now (Taylor's Version) 22
Speak Now (Deluxe Package)   22
1989 (Taylor's Version)      21
Midnights (3am Edition)      20
Fearless (Platinum Edition)  19
1989 (Deluxe)                19
evermore (deluxe version)     17
folklore (deluxe version)     17
Fearless (International Version) 16
folklore                     16
reputation                   16
evermore                     15
Taylor Swift                  15
Speak Now                     14
1989                         13
Midnights                     13
Live From Clear Channel Stripped 2008 8
Name: count, dtype: int64
-----

```

```

-----
Columna: album_release_date
Número de valores únicos: 23
album_release_date
2017-11-09    46
2019-08-23    36
2008-11-11    35
2020-11-25    34
2014-01-01    32
2021-11-12    30
2010-10-25    30
2021-04-09    26
2027-05-26    24
2023-10-27    22
2023-07-07    22
2010-01-01    22
2012-10-22    22
2023-10-26    21
2022-10-22    20
2020-08-18    17
2021-01-07    17
2017-11-10    16
2020-07-24    16
2020-12-11    15
1989-10-24    15
2022-10-21    13
2008-06-28     8
Name: count, dtype: int64
-----

-----
Columna: album_total_tracks
Número de valores únicos: 17
album_total_tracks
22      88
34      64
16      48
46      46
19      38
18      36
17      34
13      26
26      26
24      24
21      21
20      20
15      16
10      15
Thirteen 15
14      14
8        8
Name: count, dtype: int64
-----

```

Tipos de variables presentes en el dataframe:

```
·   disc_number          int64
    duration_ms         int64
    explicit             object
    track_number        int64
    track_popularity     int64
    track_id            object
    track_name          object
    danceability         float64
    energy              float64
    key                 float64
    loudness            float64
    mode               int64
    speechiness         float64
    acousticness        float64
    instrumentalness    object
    liveness           float64
    valence             float64
    tempo              float64
    id                 object
    time_signature     float64
    artist_id          object
    artist_name        object
    artist_popularity   int64
    album_id           object
    album_name         object
    album_release_date  object
    album_total_tracks  object
    dtype: object
```

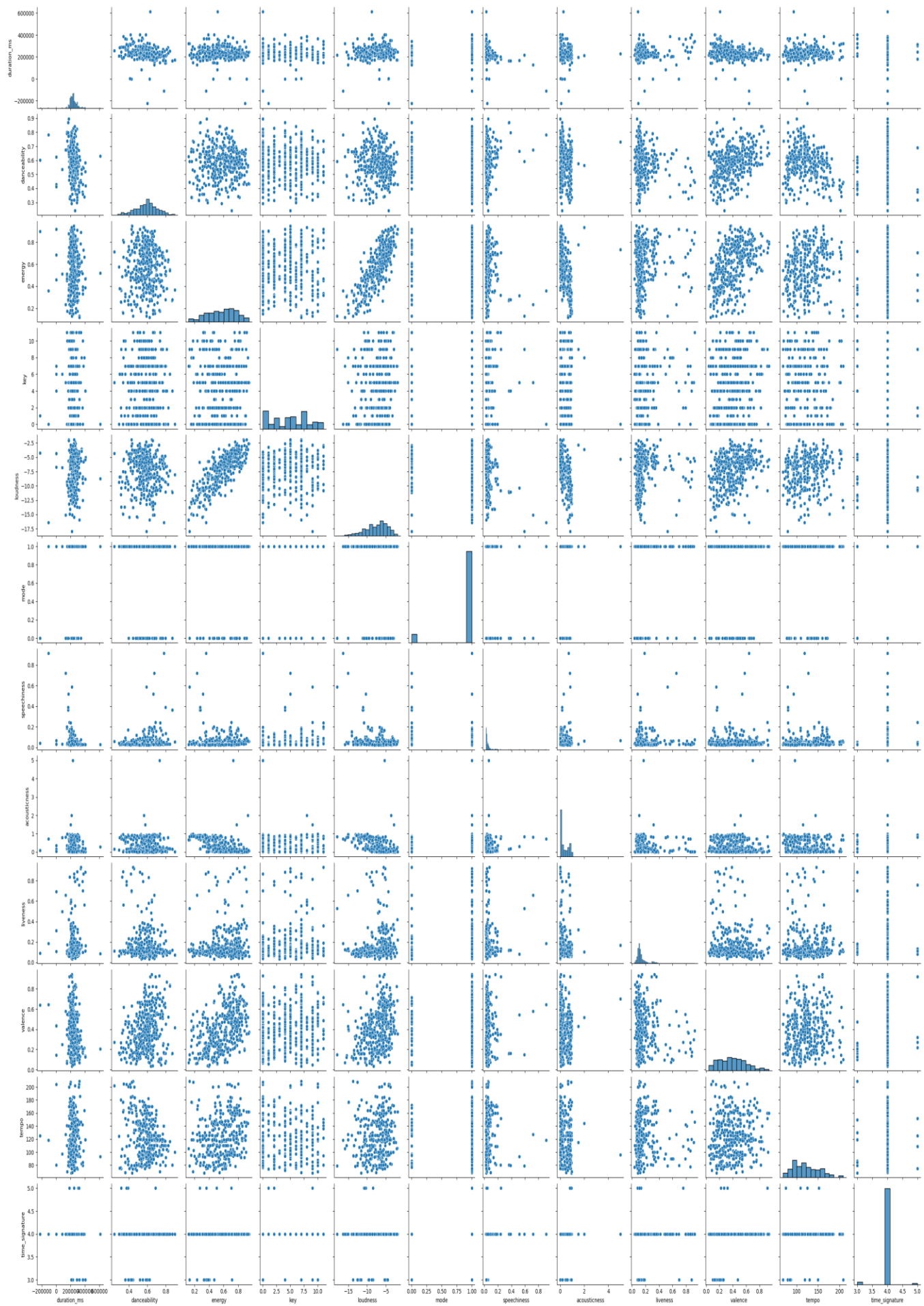
'album_release_date' tendria que cambiar su formato a tipo datetime

Estadísticas descriptivas:

	count	mean	std	min \
duration_ms	539.0	236003.725417	55019.871010	-223093.00000
danceability	537.0	0.587242	0.116858	0.24300
energy	537.0	0.573065	0.192309	0.11800
key	538.0	4.587361	3.246082	0.00000
loudness	537.0	-7.520639	2.933158	-17.93200
mode	539.0	0.912801	0.282388	0.00000
speechiness	538.0	0.057709	0.073146	0.02310
acousticness	538.0	0.337806	0.395292	-0.00354
liveness	538.0	0.163308	0.141800	0.03350
valence	539.0	0.398410	0.199409	0.03740
tempo	538.0	122.362639	30.485522	68.09700
time_signature	538.0	3.986989	0.197323	3.00000

	25%	50%	75%	max
duration_ms	209486.500000	233626.00000	259045.50000	613026.000
danceability	0.517000	0.59500	0.66100	0.897
energy	0.436000	0.58900	0.72900	0.949
key	2.000000	5.00000	7.00000	11.000
loudness	-9.287000	-6.94200	-5.37600	-1.909
mode	1.000000	1.00000	1.00000	1.000
speechiness	0.030525	0.03775	0.05640	0.912
acousticness	0.036250	0.16750	0.66375	5.000
liveness	0.096500	0.11500	0.16225	0.931
valence	0.230000	0.38600	0.53500	0.943
tempo	96.684500	119.00050	143.93900	208.918
time_signature	4.000000	4.00000	4.00000	5.000

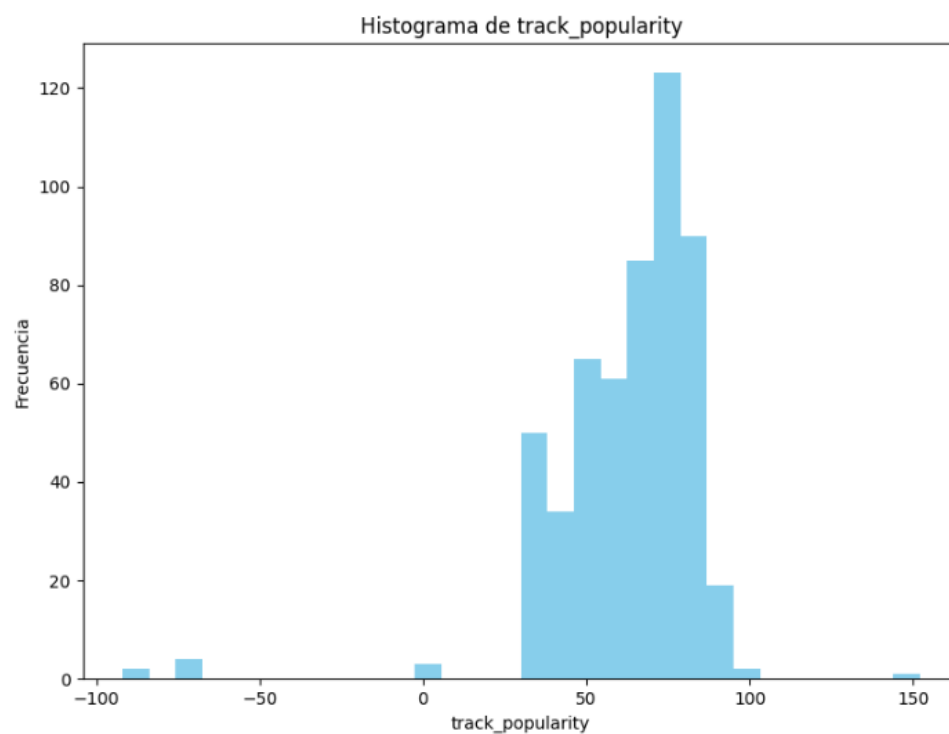
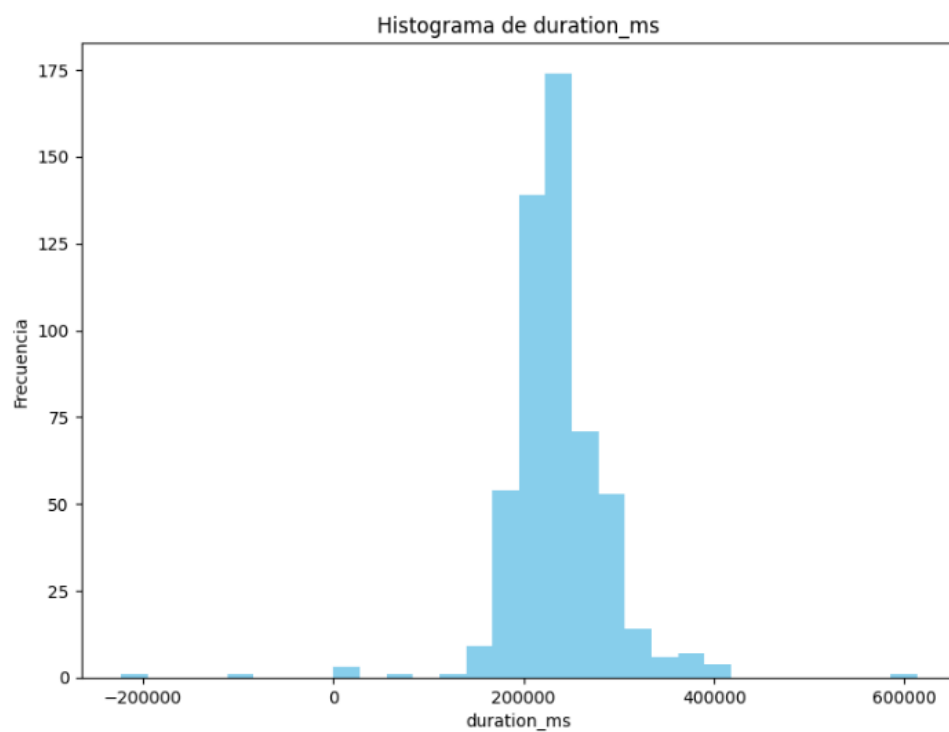
Visualicemos un pairplot con las distintas relaciones entre variables:

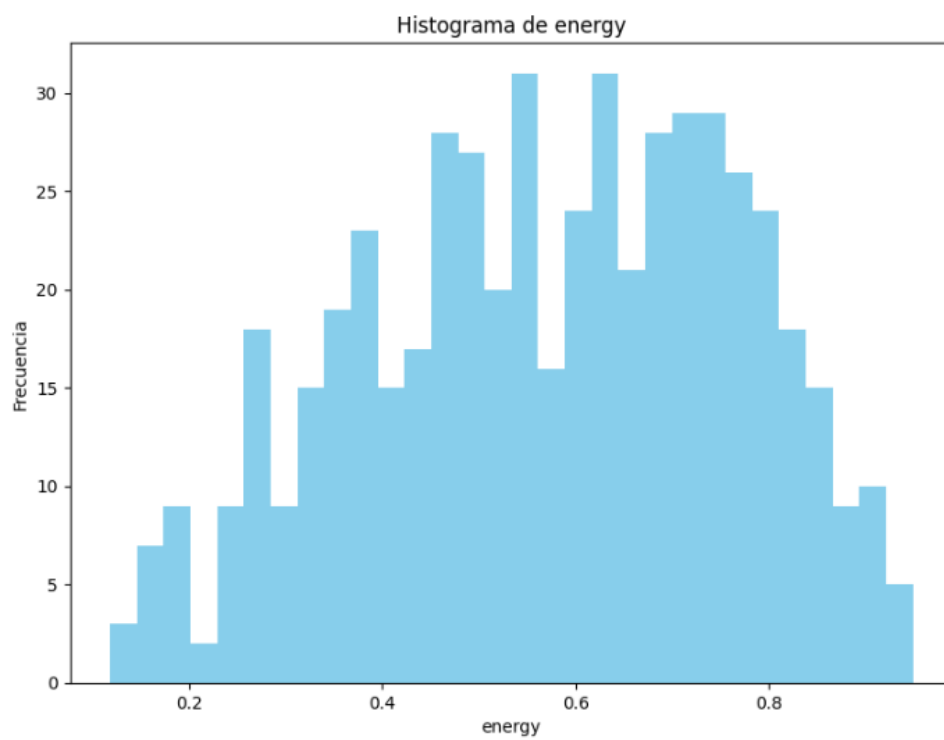
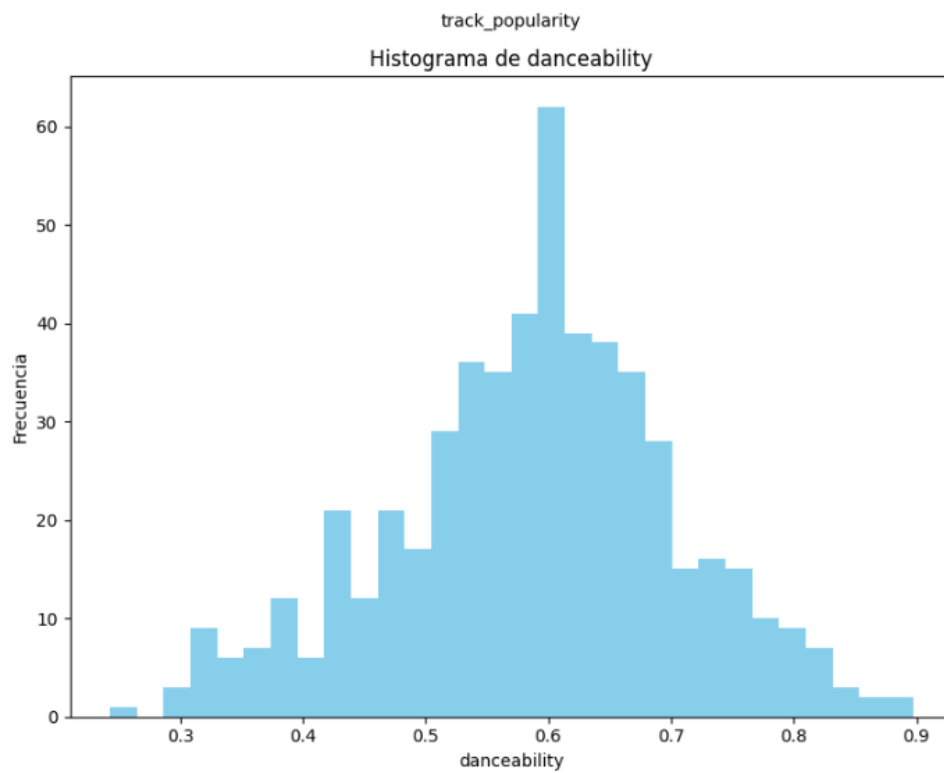


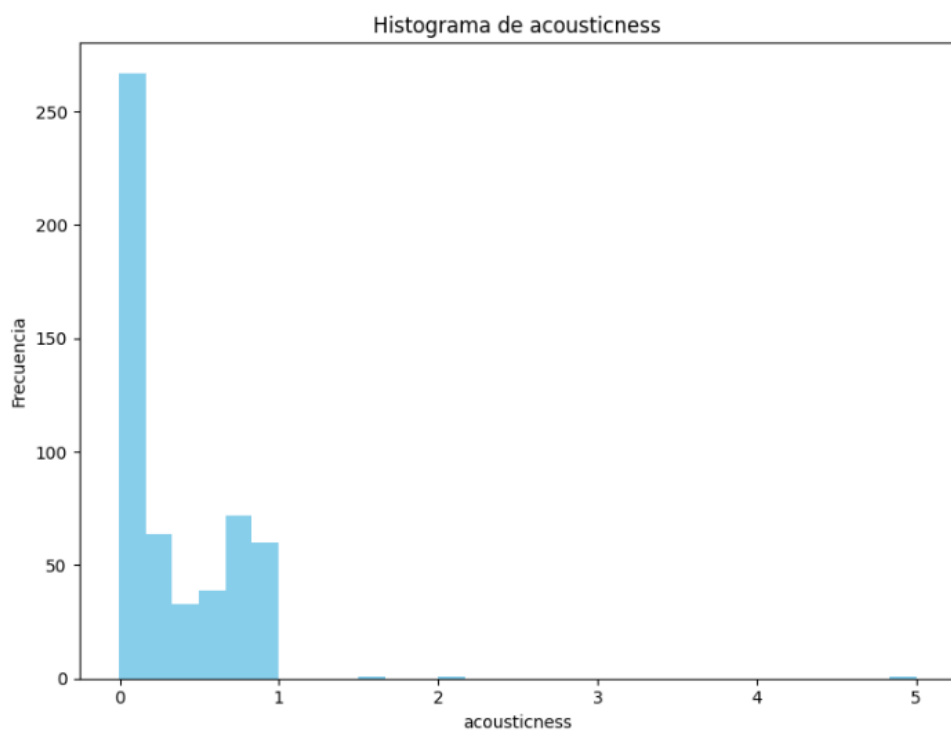
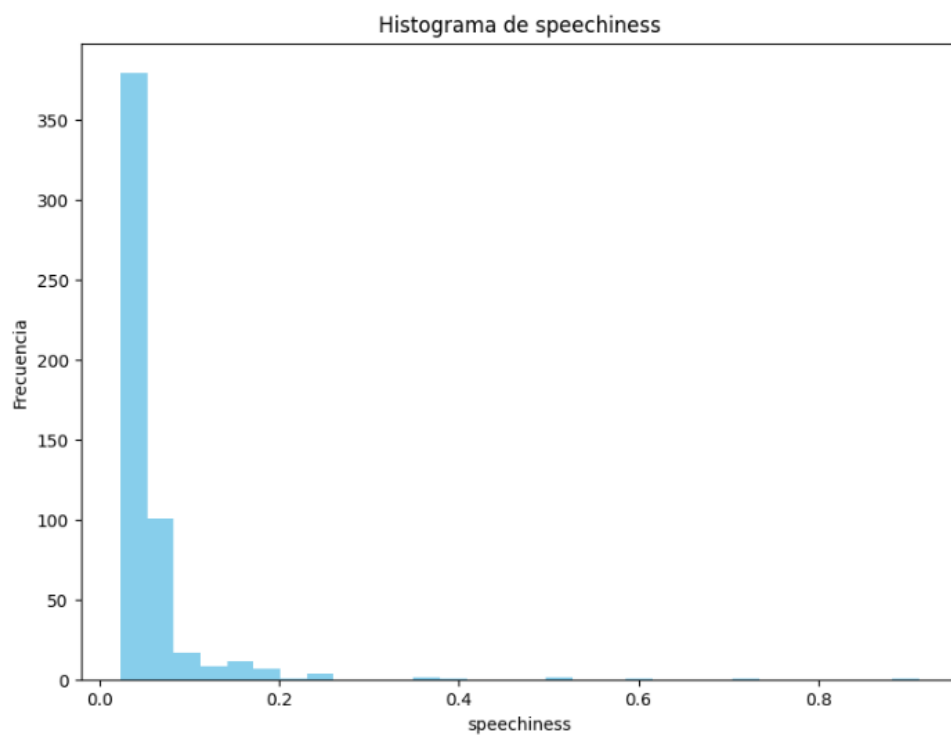
A partir del grafico pairplot anterior podemos denotar las siguientes conclusiones:

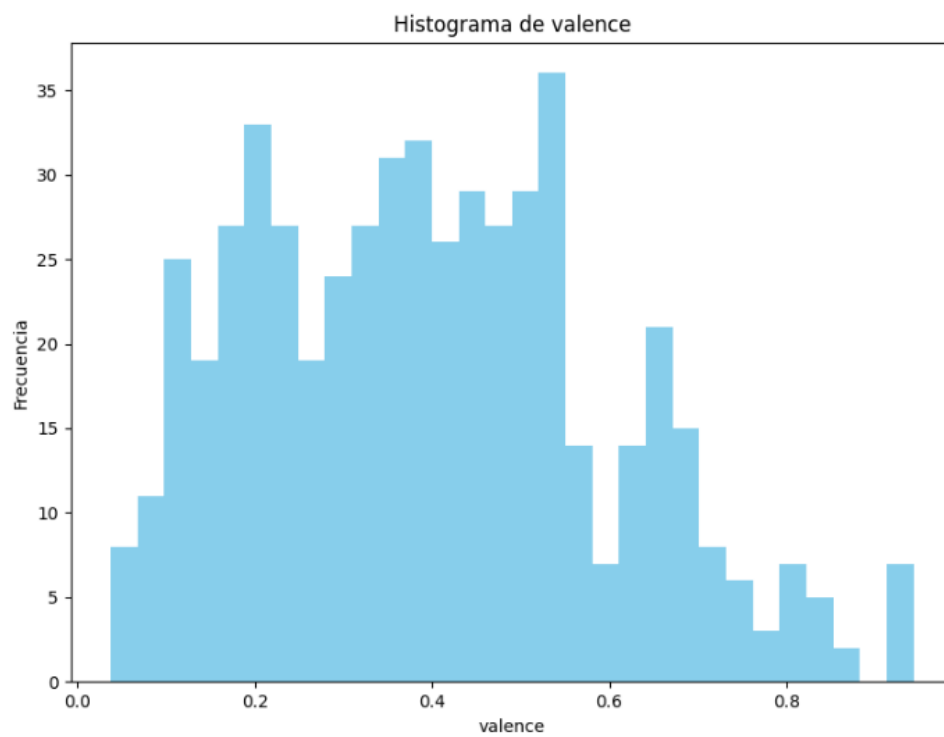
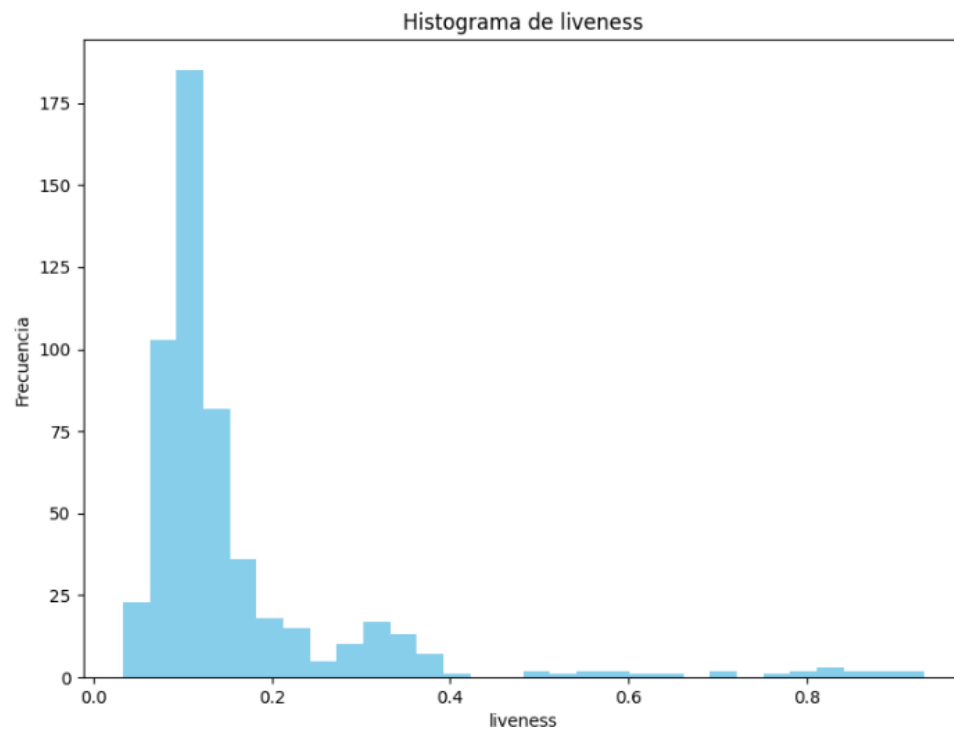
- La variable loudness tiene una correlación negativa con acousticness, lo que significa que a mayor sonoridad, menor acústica.
- La variable energy tiene una correlación positiva con loudness, lo que significa que a mayor energía, mayor sonoridad.
- La variable duration_ms tiene una correlación positiva con liveness, lo que significa que a mayor duración, mayor probabilidad de que la canción sea en vivo.
- La variable valence tiene una correlación positiva con danceability, lo que significa que a mayor valencia, mayor capacidad de baile.
- La variable tempo no tiene una correlación clara con ninguna otra variable, lo que significa que el tempo es independiente de las demás características de las canciones.
- La variable key tiene una distribución uniforme, lo que significa que todas las tonalidades son igualmente probables.
- La variable mode tiene una distribución binomial, lo que significa que solo hay dos modos posibles: mayor o menor.

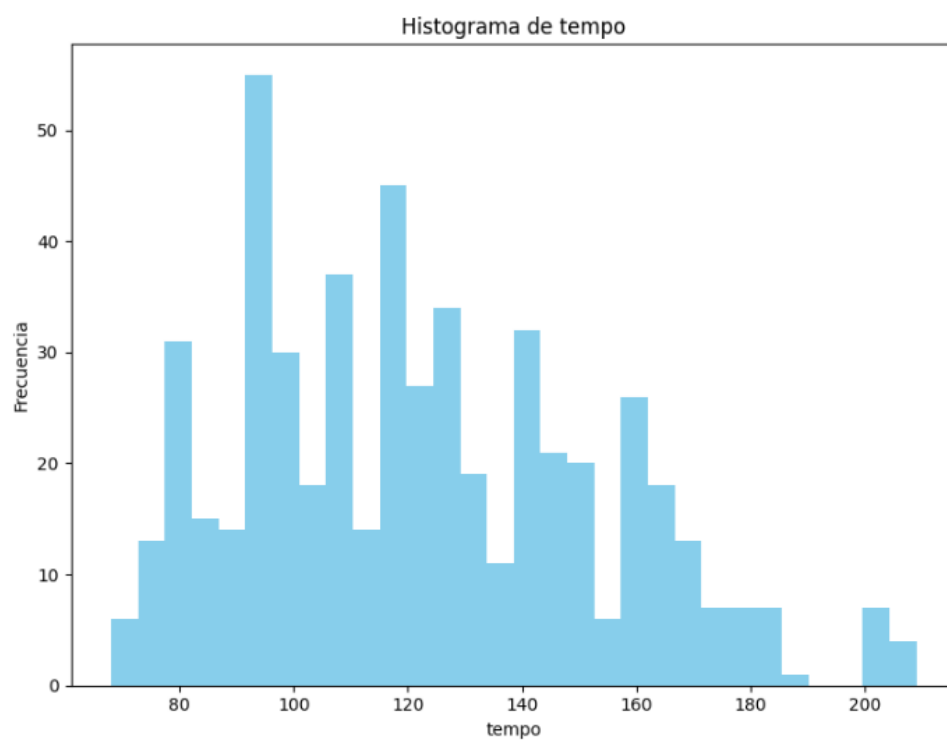
Visualizamos la distribucion de diversas variables de interés:



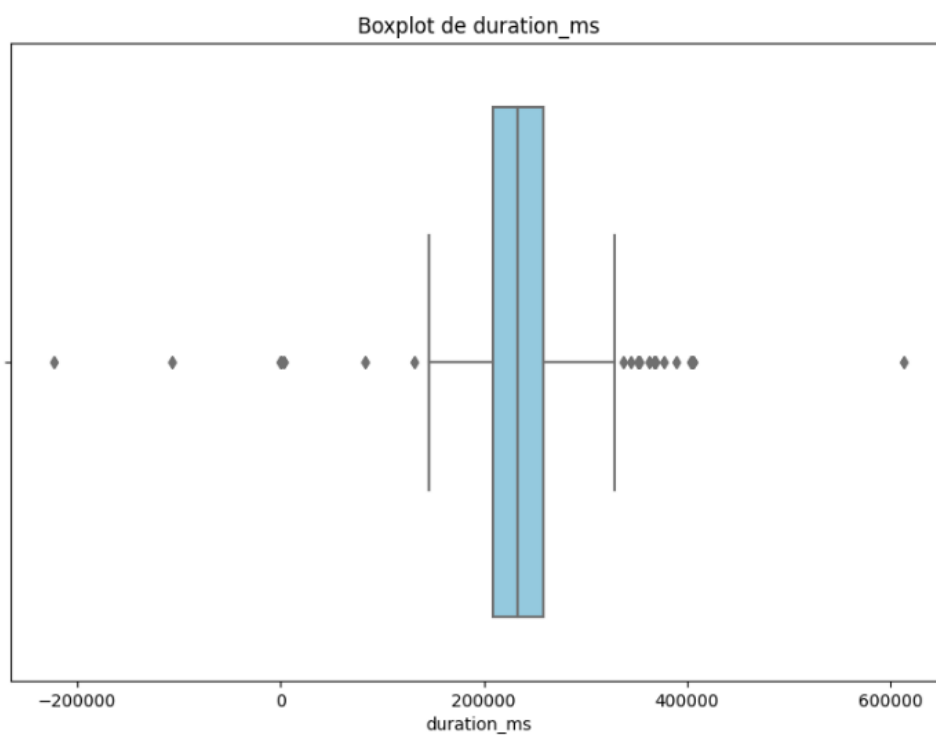


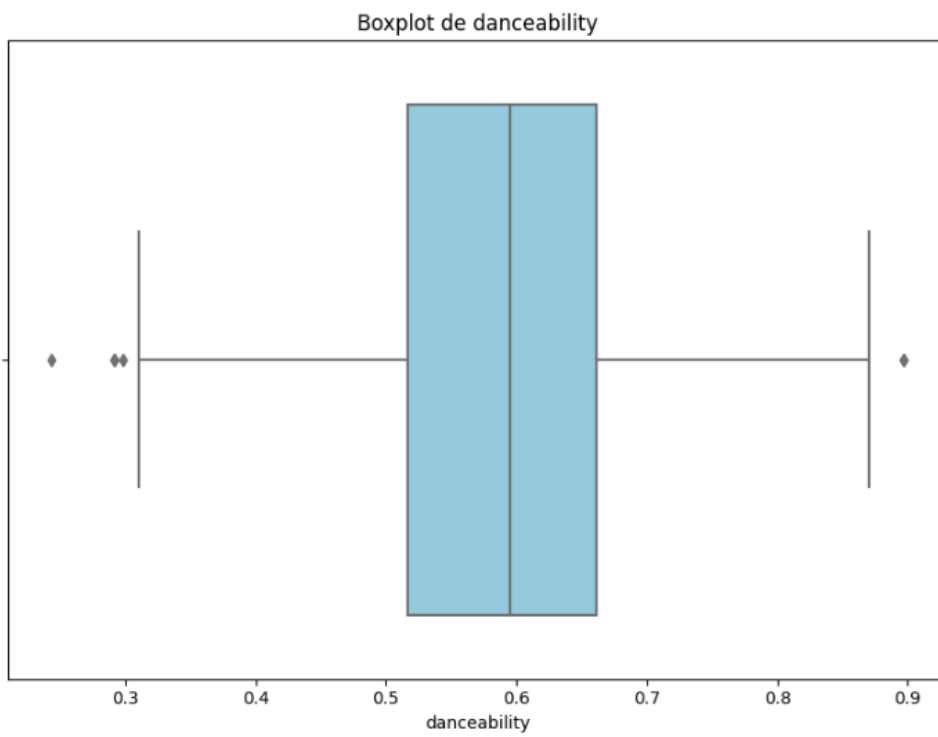
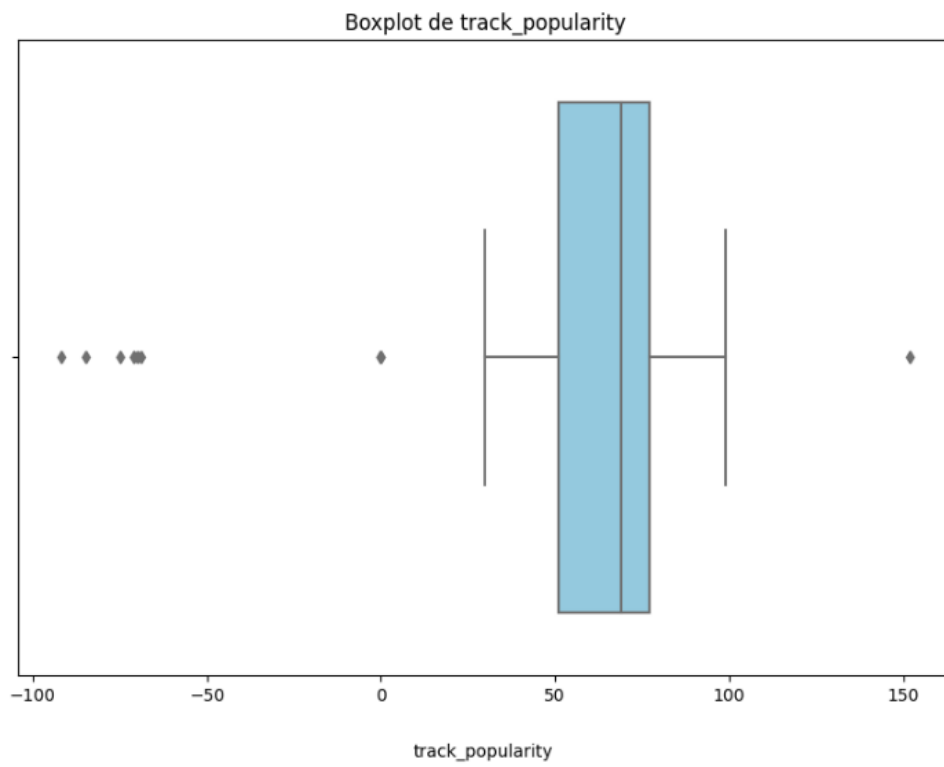




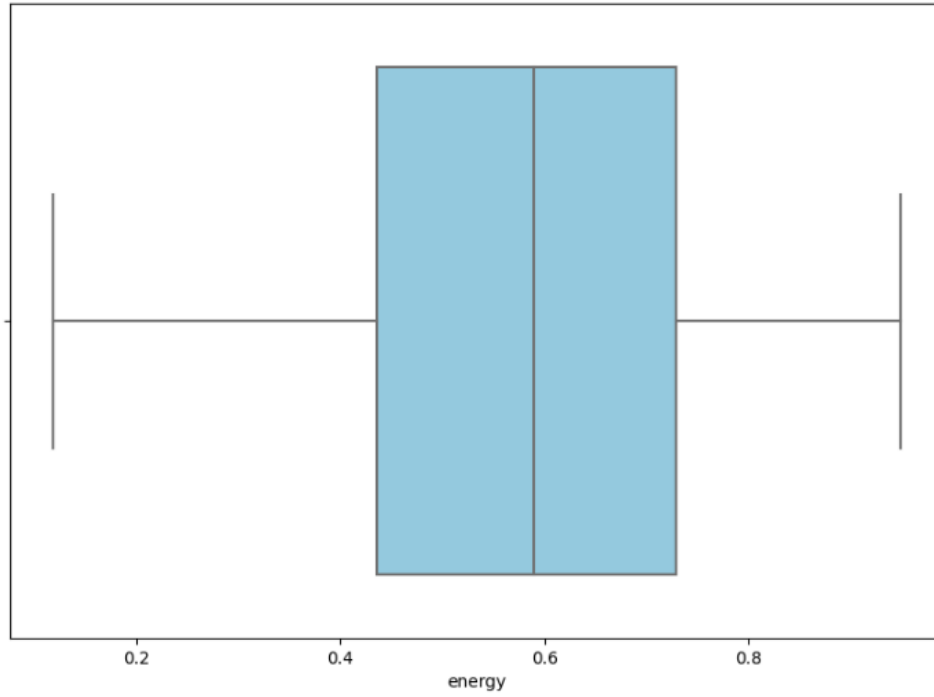


Identifiquemos valores outliers:

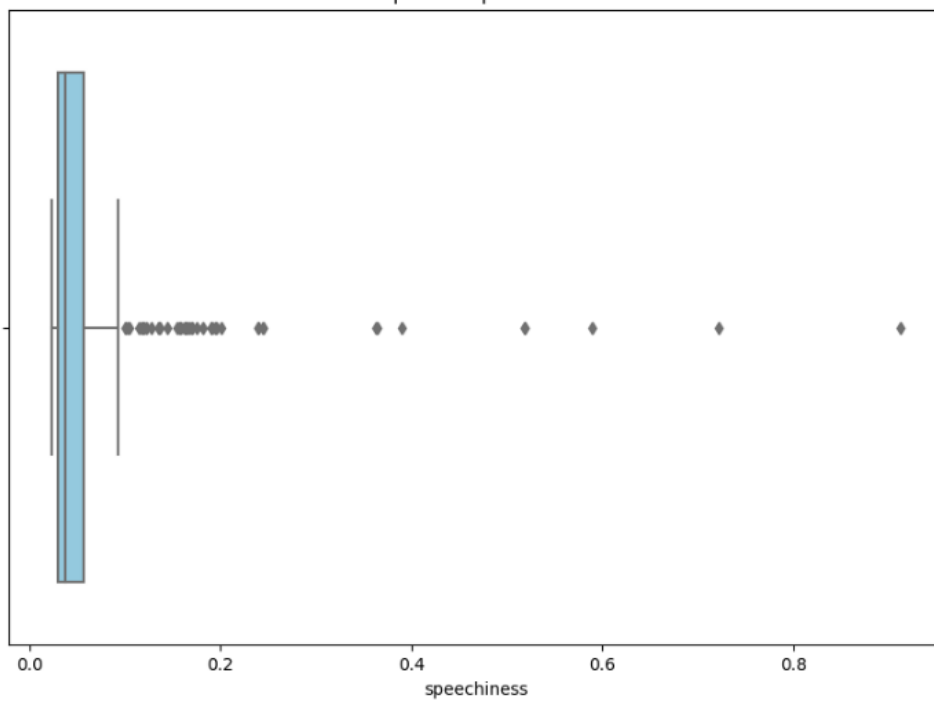




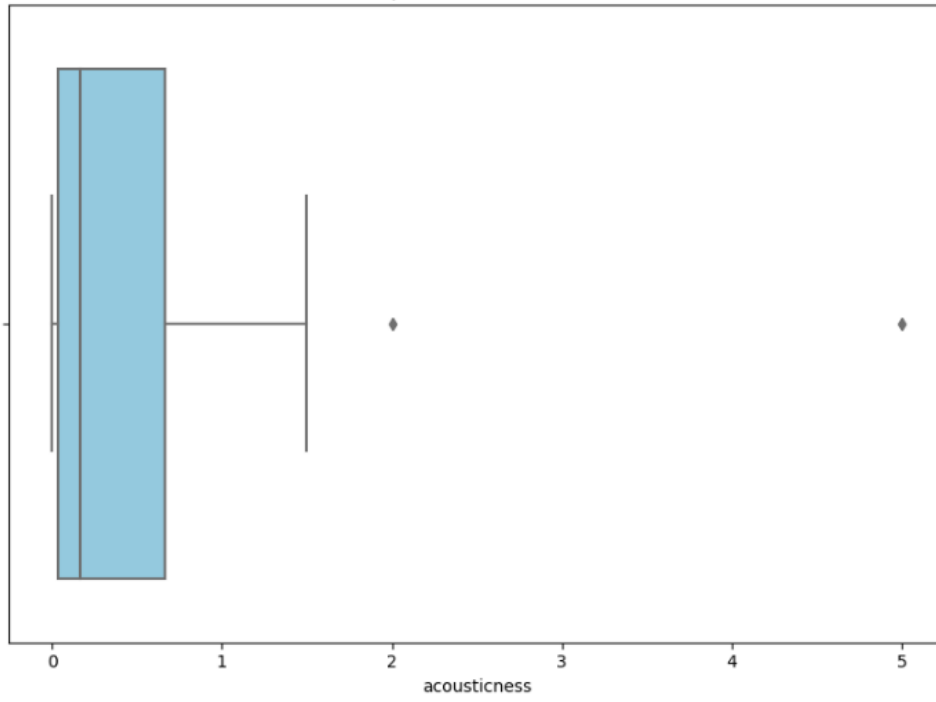
Boxplot de energy



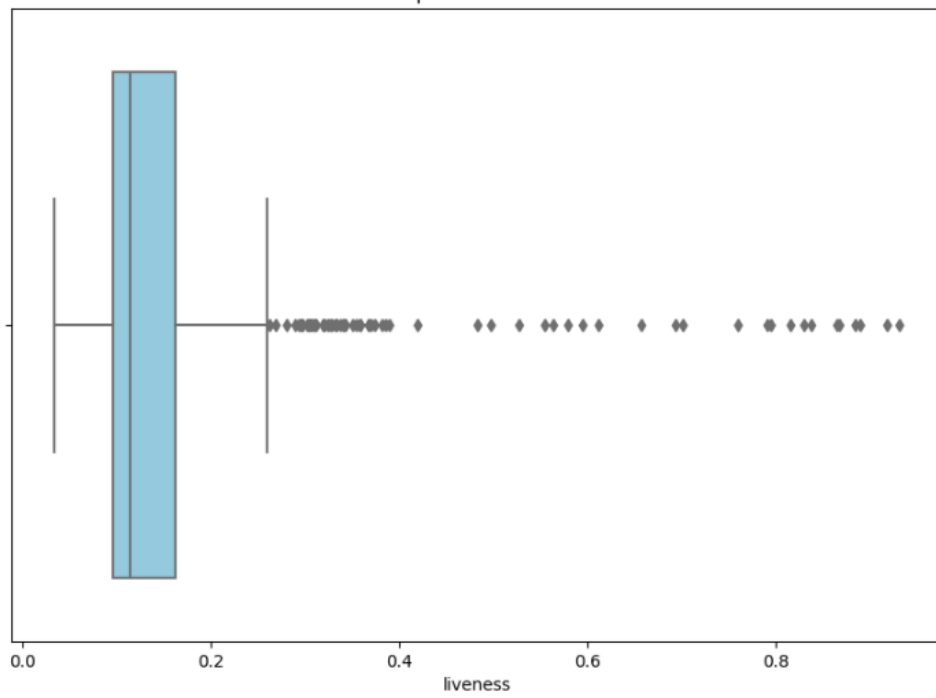
Boxplot de speechiness

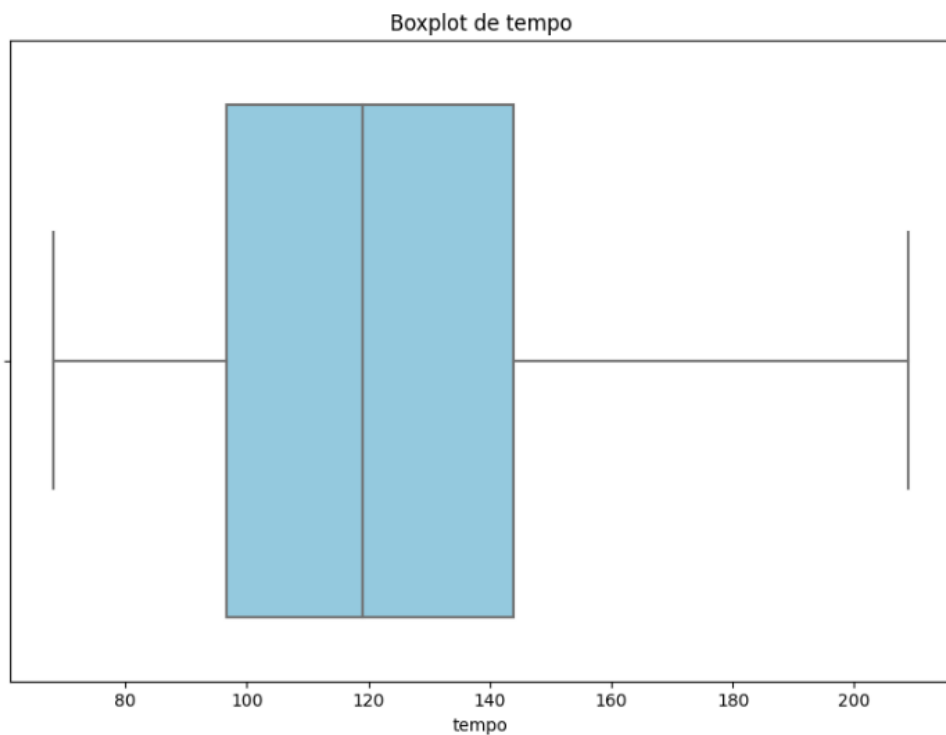
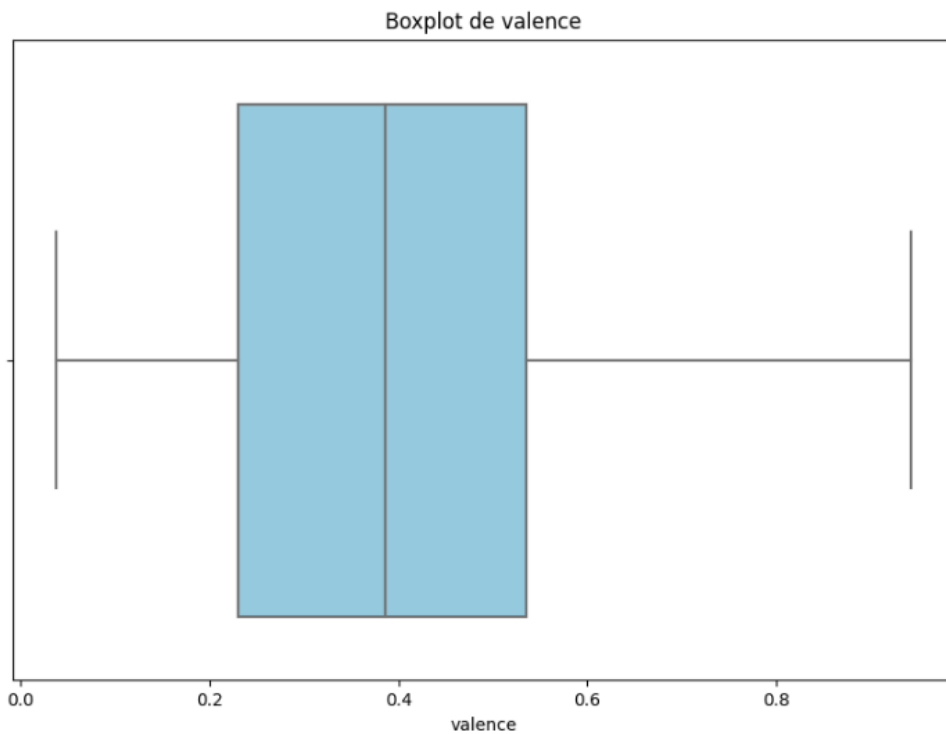


Boxplot de acoustiness



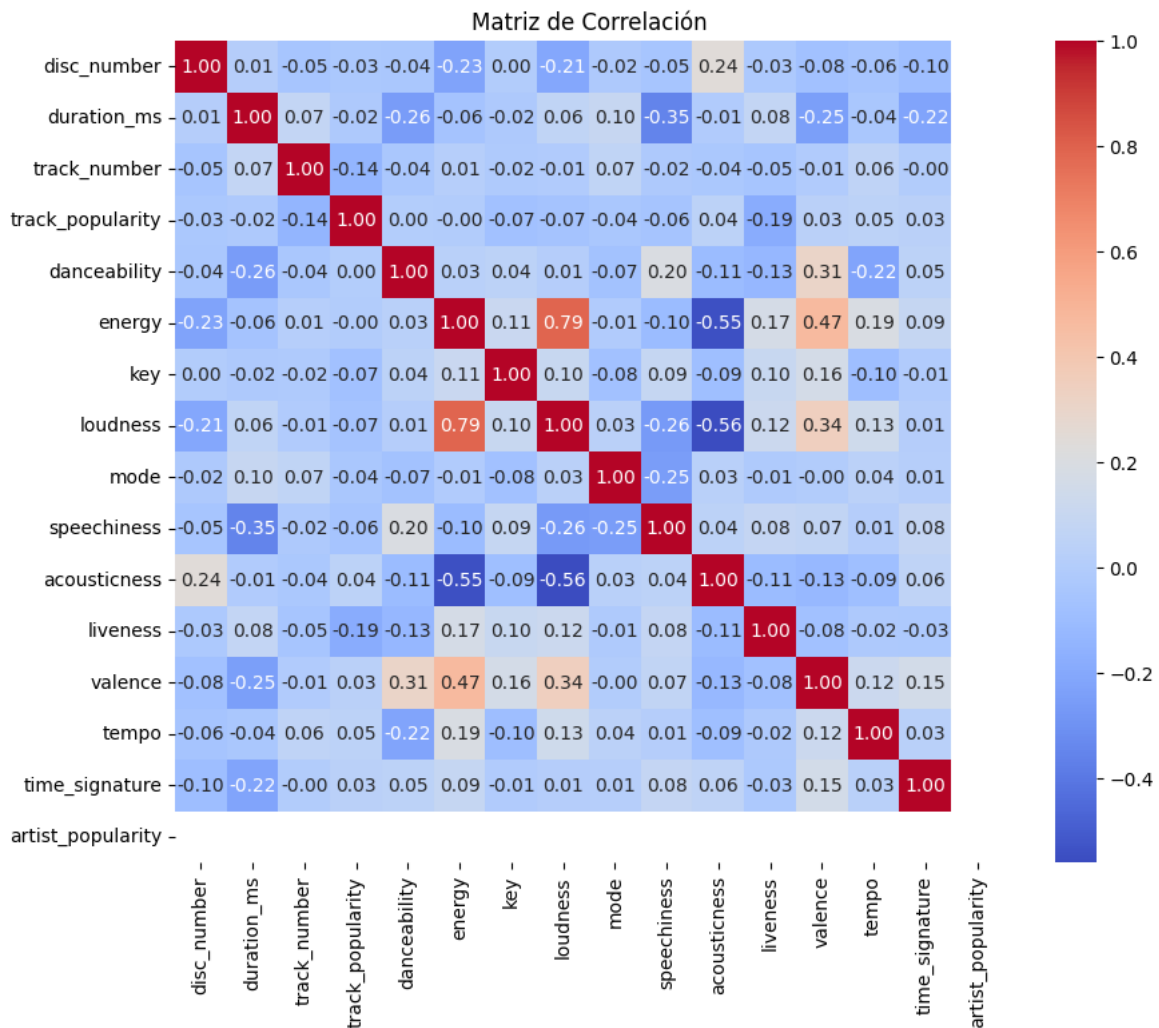
Boxplot de liveness





Podemos visualizar claramente como las variables 'duration_ms', 'track_popularity' y 'danceability' tienen valores outliers alejados de la caja central, mientras que 'speechiness', 'acousticness' y 'liveness' presentan valores outliers alejándose en un extremo derecho de la caja central, por último, las variables que no presentan outliers son 'energy', 'valence' y 'tempo'.

Matriz de correlación de variables numéricas:



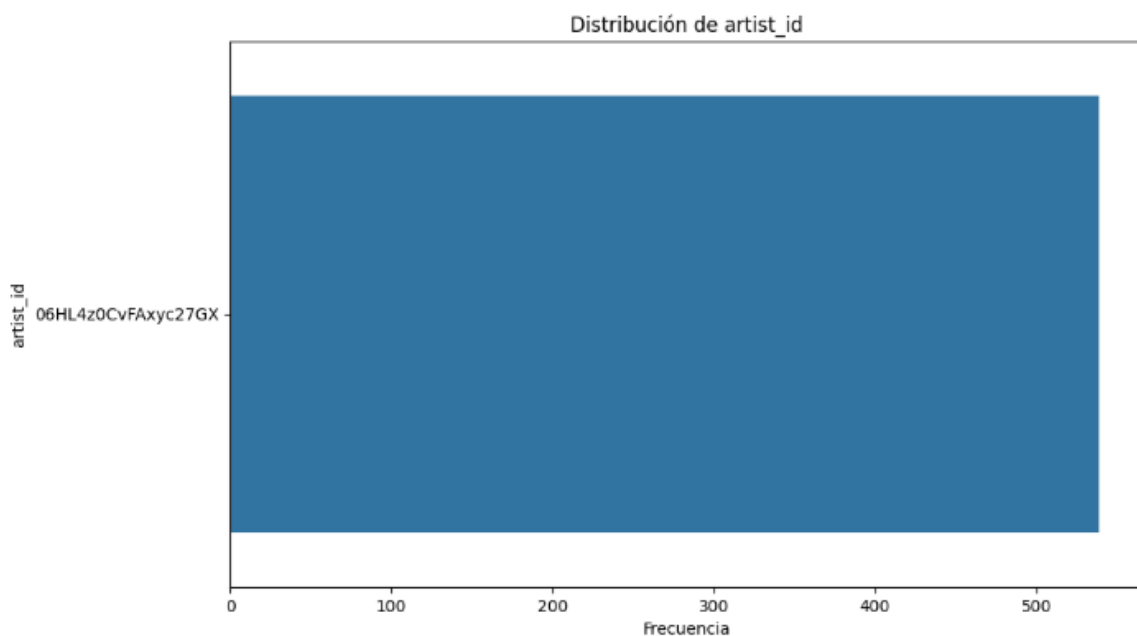
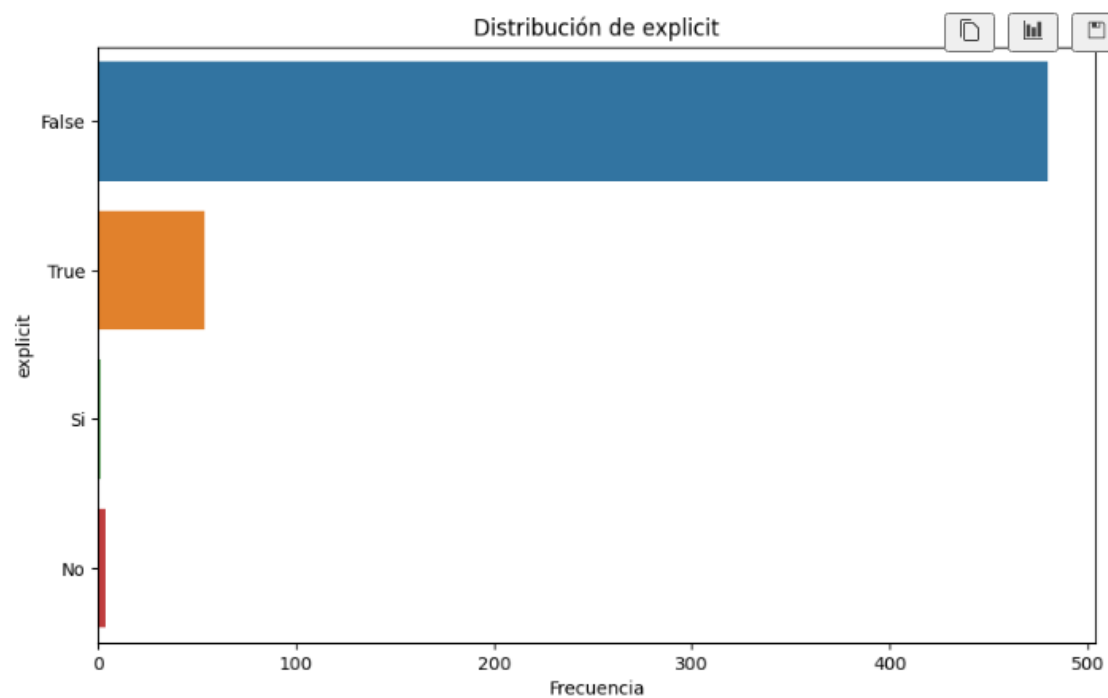
Podemos identificar diversas conclusiones en esta matriz de correlación ya que hay varias correlaciones fuertes y otras débiles.

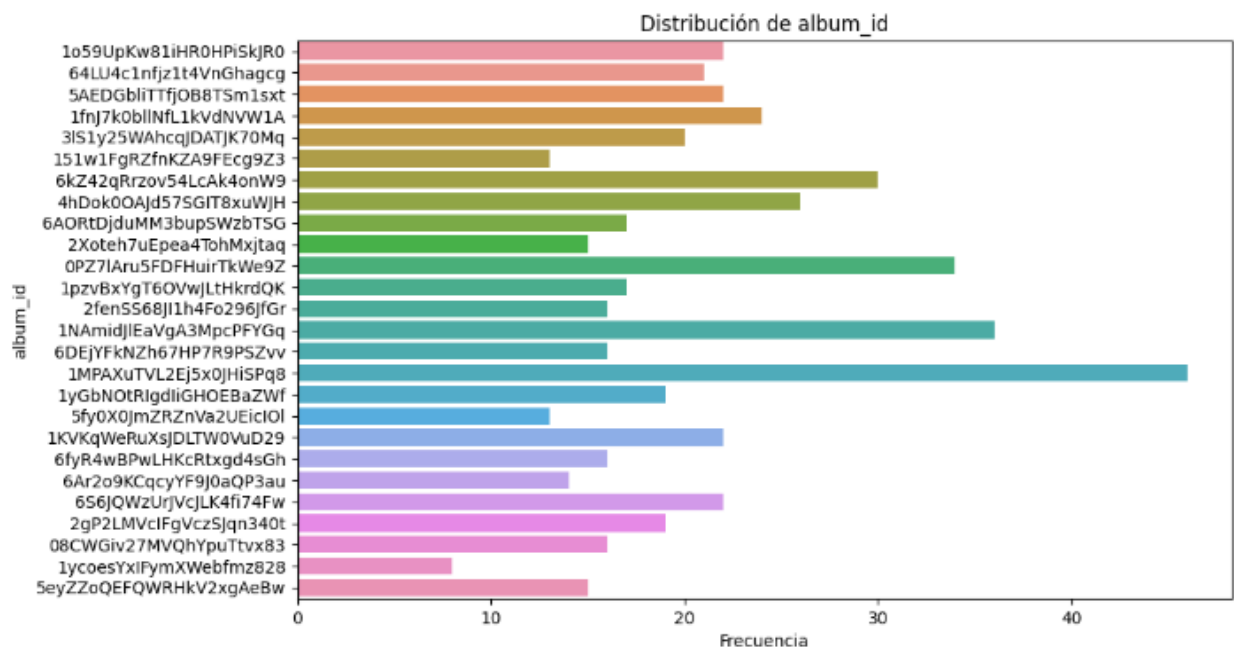
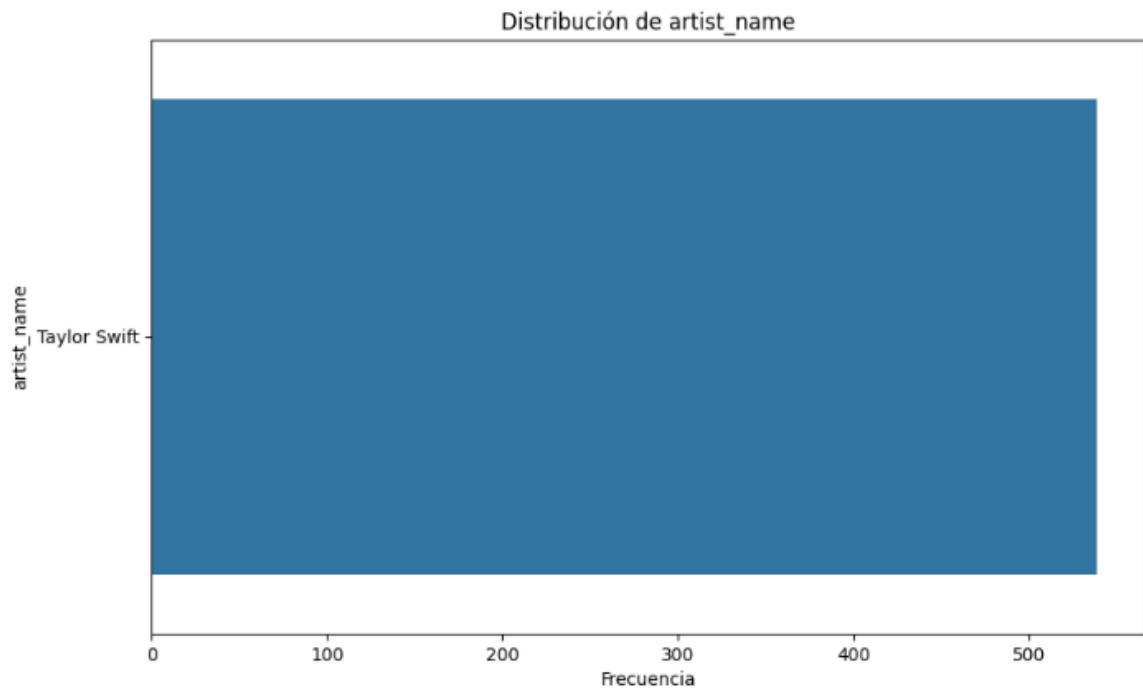
****Energía y Loudness (0.79)**:** Esta alta correlación positiva indica que las canciones más energéticas tienden a ser más altas. Esto tiene sentido, ya que la energía mide el nivel de intensidad y actividad de una pista, y la sonoridad mide el nivel de volumen percibido por el oído humano.

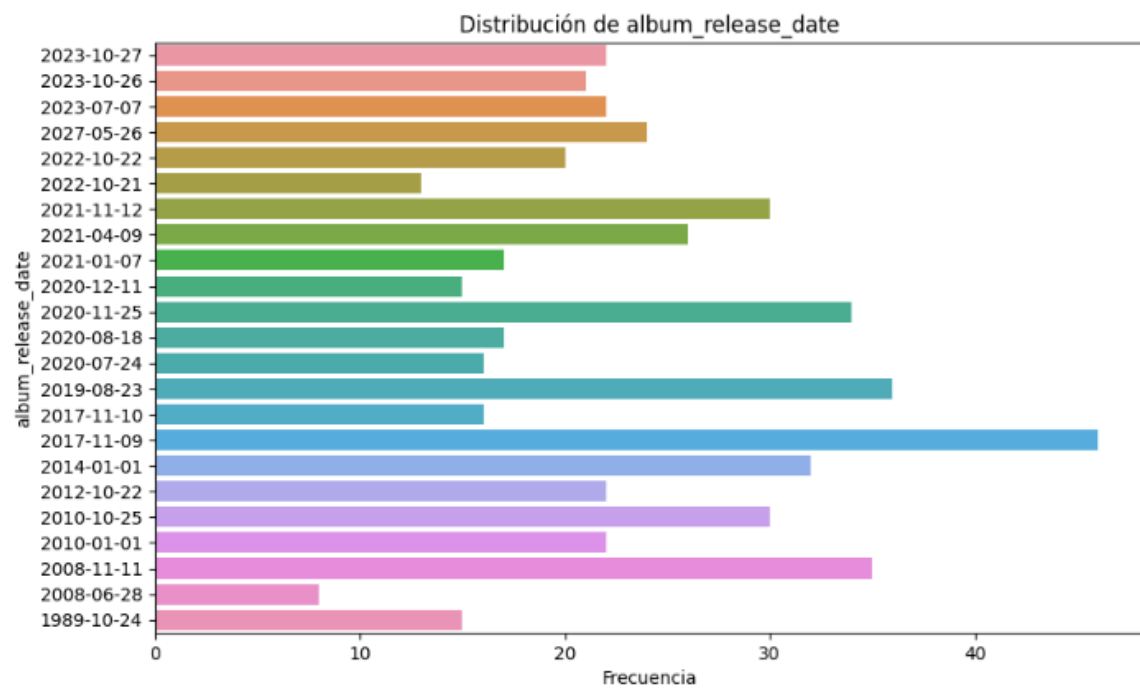
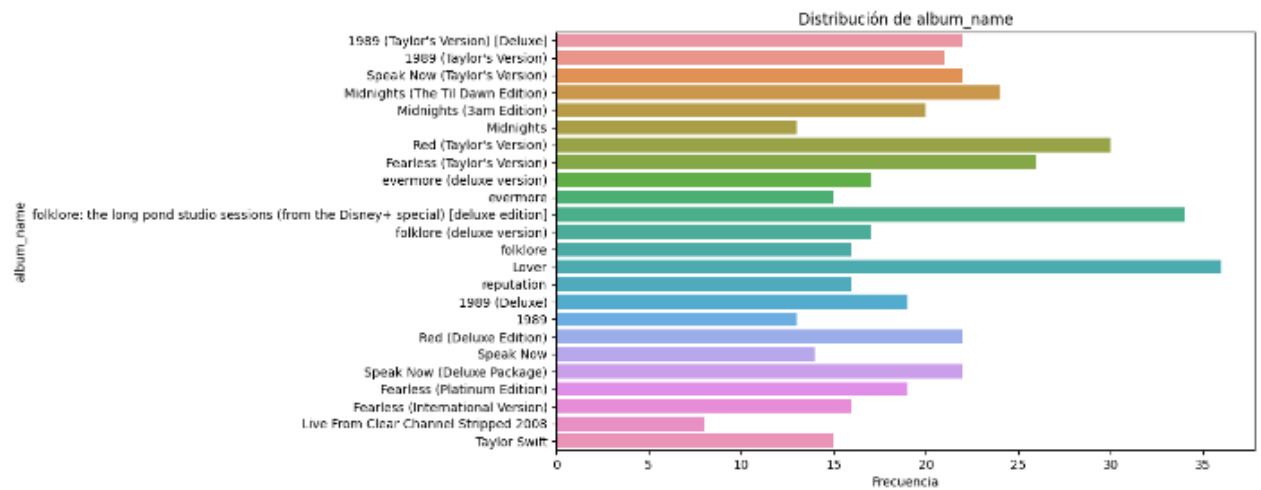
****Acousticness y Energy (-0.55)**:** Esta correlación negativa moderada sugiere que las canciones acústicas tienden a tener menos energía. Esto también tiene sentido, ya que la acústica mide el nivel de presencia de instrumentos acústicos en una pista, y estos suelen ser más suaves y tranquilos que los instrumentos eléctricos o electrónicos.

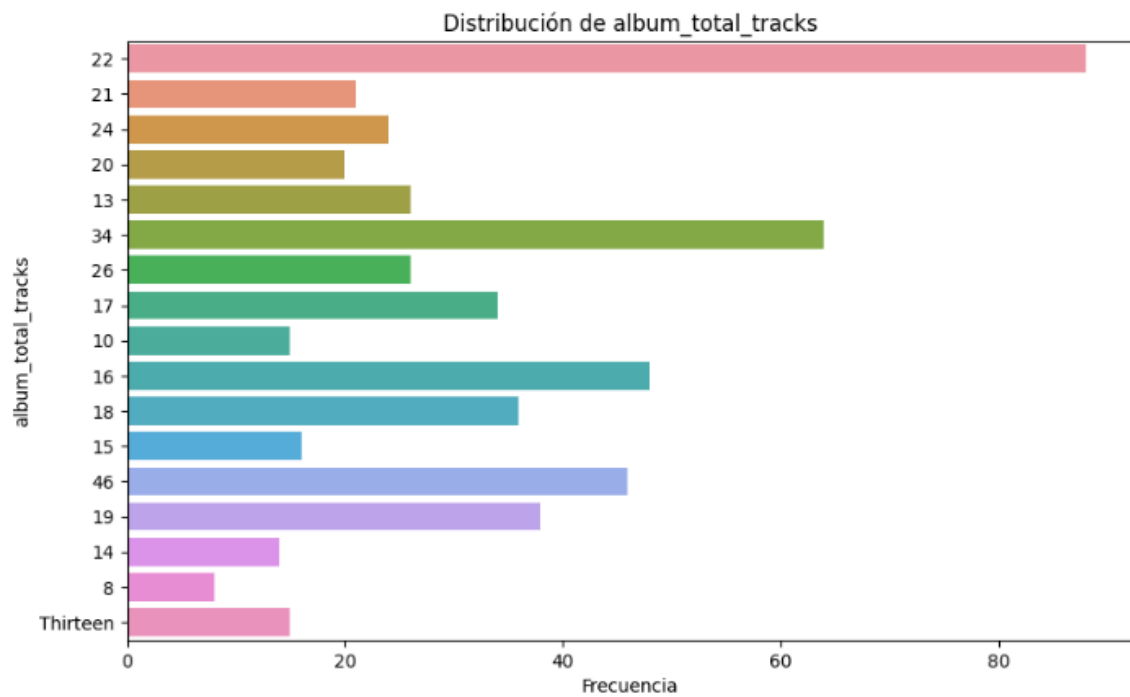
****Valence y Danceability (0.34)**:** Esta correlación positiva moderada indica que las canciones con un tono más positivo o feliz son más propensas a ser bailables. Esto puede reflejar el hecho de que las canciones bailables suelen tener ritmos más rápidos y contagiosos, y letras más optimistas y divertidas.

Conteo de frecuencia para las variables categóricas:

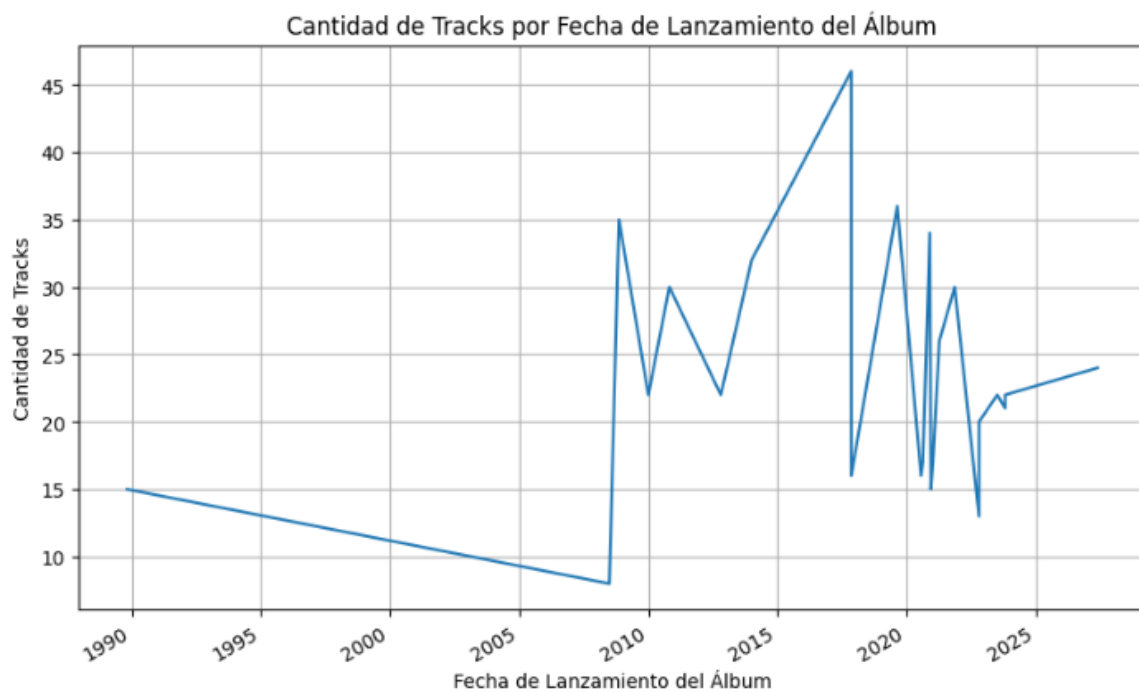








Análisis de tendencias temporales a lo largo del tiempo:



- La gráfica muestra la evolución de la cantidad de pistas en los álbumes lanzados a lo largo del tiempo, desde 1990 hasta el futuro cercano.

- Se puede observar que hay una tendencia creciente hasta alrededor del año 2015, donde se alcanza el máximo de casi 45 pistas por álbum. Esto podría deberse a que los artistas aprovecharon las ventajas de la tecnología digital y el streaming para producir y distribuir más música.
- Sin embargo, después del año 2015, se produce una caída significativa en la cantidad de pistas por álbum, llegando a menos de 15 en el año 2020. Esto podría deberse a que los artistas se adaptaron a los cambios en el mercado y las preferencias de los consumidores, optando por lanzar álbumes más cortos y frecuentes, o incluso singles individuales.
- Finalmente, se observa un ligero aumento en la cantidad de pistas por álbum hacia el año 2025, lo que podría indicar una recuperación o una nueva tendencia en la industria musical.