

```
import pandas as pd
import json

# Cargar el archivo JSON
with open('D:/Profesion/taylor_swift_spotify.json') as file:
    data = json.load(file)

# Normalizar los datos JSON y convertirlo en DataFrame
df = pd.json_normalize(data, record_path=['albums', 'tracks'],
    meta=['artist_id', 'artist_name', 'artist_popularity',
        ['albums', 'album_id'], ['albums', 'album_name'], ['albums', 'album_release_date'], ['albums', 'album_total_tracks']]

# Guardar como archivo CSV
df.to_csv('data.csv', index=False)

df_spotify=df

df_spotify.describe()
```

	disc_number	duration_ms	track_number	track_popularity	audio_features.energy
count	539.000000	539.000000	539.000000	539.000000	537.000000
mean	1.031540	236003.725417	11.280148	62.918367	0.573061
std	0.174934	55019.871010	7.965621	22.498757	0.192301
min	1.000000	-223093.000000	1.000000	-92.000000	0.118000
25%	1.000000	209486.500000	5.000000	51.000000	0.436000
50%	1.000000	233626.000000	10.000000	69.000000	0.589000
75%	1.000000	259045.500000	15.000000	77.000000	0.729000
max	2.000000	613026.000000	46.000000	152.000000	0.949000

▼ error en calidad de datos

Campo duration_ms con valores menores a 0

```
df_spotify_duration=df_spotify['duration_ms']<0

df_spotify[df_spotify_duration]
```

	disc_number	duration_ms	explicit	track_number	track_popularity	track_id	track_name	audio_features.danceability
392	1	-107133	False	18	0	4eTXfpHxhxVofrBUjAhPMg	I Wish You Would - Voice Memo	0.71
408	1	-223093	False	2	59	7gJtmLyPTwKzhGzMBXtuXH		0.61

2 rows × 27 columns

▼ error calidad de datos con valores nulos

Campos con valores nulos *track_id *track_name *audio_features.danceability *audio_features.energy *audio_features.key *audio_features.loudness *audio_features.speechiness *audio_features.acousticness *audio_features.liveness *audio_features.tempo *audio_features.time_signature *albums.album_name

```
registros_con_nulos = df_spotify[df_spotify.isnull().any(axis=1)]

# Mostrar los registros con valores nulos
```

```
print(registros_con_nulos)
```

	disc_number	duration_ms	explicit	track_number	track_popularity	\
77	1	191039	False	13	68	
91	1	200690	False	3	71	
104	1	196258	False	16	72	
329	1	329160	False	1	42	
330	1	220440	False	2	45	
331	1	246306	False	3	36	
332	1	223093	False	4	39	
333	1	202960	False	5	36	
334	1	204706	False	6	38	
335	1	203440	False	7	43	
336	1	232120	False	8	41	
337	1	219720	False	9	42	
338	1	247160	False	10	85	
339	1	238693	False	11	39	
340	1	247533	False	12	38	
341	1	227906	False	13	59	
342	1	295320	False	14	36	
343	1	231866	False	15	40	
344	1	262920	False	16	37	
345	1	295720	False	17	34	
346	1	245413	False	18	36	
347	1	292920	False	19	35	
348	1	242760	False	20	33	
349	1	215106	False	21	35	
350	1	212600	False	22	37	
351	1	242973	False	23	37	
352	1	353253	False	24	47	
353	1	281053	False	25	33	
354	1	239346	False	26	36	
355	1	235800	False	27	39	
356	1	223920	False	28	34	
357	1	202080	False	29	33	
358	1	250093	False	30	34	
359	1	240266	False	31	33	
360	1	205853	No	32	33	
361	1	264946	False	33	36	
362	1	263613	False	34	37	
363	1	238093	False	35	32	
364	1	239426	False	36	33	
365	1	251213	False	37	34	
366	1	267653	False	38	35	
367	1	226293	False	39	33	
368	1	256040	False	40	32	
369	1	243346	False	41	36	
370	1	245560	False	42	40	
371	1	235400	False	43	33	
372	1	215600	False	44	35	
373	1	220600	False	45	33	
374	1	195706	False	46	36	
375	1	212600	False	1	60	
379	1	193293	False	5	60	
382	1	211933	False	8	61	

	track_id	track_name	\
77	1Q0ii3pa5m8MEda0nbkjfw	None	
91	0Z2kkf2zMkwRGQjZ7T4p8f	None	
104	7712gjoih4QoDbXpljEk21	None	
---	--	--	

