

# Songs\_Extractor\_Spotify

**Autor: VChangB**

## Installing the necessary libraries

In [1]: `!pip install spotipy pandas`

```
Requirement already satisfied: spotipy in c:\users\acer\anaconda3\lib\site-packages (2.25.1)
Requirement already satisfied: pandas in c:\users\acer\anaconda3\lib\site-packages (2.2.3)
Requirement already satisfied: redis>=3.5.3 in c:\users\acer\anaconda3\lib\site-packages (from spotipy) (6.2.0)
Requirement already satisfied: requests>=2.25.0 in c:\users\acer\anaconda3\lib\site-packages (from spotipy) (2.32.3)
Requirement already satisfied: urllib3>=1.26.0 in c:\users\acer\anaconda3\lib\site-packages (from spotipy) (2.3.0)
Requirement already satisfied: numpy>=1.26.0 in c:\users\acer\anaconda3\lib\site-packages (from pandas) (2.1.3)
Requirement already satisfied: python-dateutil>=2.8.2 in c:\users\acer\anaconda3\lib\site-packages (from pandas) (2.9.0.post0)
Requirement already satisfied: pytz>=2020.1 in c:\users\acer\anaconda3\lib\site-packages (from pandas) (2024.1)
Requirement already satisfied: tzdata>=2022.7 in c:\users\acer\anaconda3\lib\site-packages (from pandas) (2025.2)
Requirement already satisfied: six>=1.5 in c:\users\acer\anaconda3\lib\site-packages (from python-dateutil>=2.8.2->pandas) (1.17.0)
Requirement already satisfied: charset-normalizer<4,>=2 in c:\users\acer\anaconda3\lib\site-packages (from requests>=2.25.0->spotipy) (3.3.2)
Requirement already satisfied: idna<4,>=2.5 in c:\users\acer\anaconda3\lib\site-packages (from requests>=2.25.0->spotipy) (3.7)
Requirement already satisfied: certifi>=2017.4.17 in c:\users\acer\anaconda3\lib\site-packages (from requests>=2.25.0->spotipy) (2025.4.26)
```

In [2]: `import spotipy  
from spotipy.oauth2 import SpotifyOAuth  
import pandas as pd`

## Process

In [3]: `import spotipy  
from spotipy.oauth2 import SpotifyOAuth  
import pandas as pd  
import requests  
import time  
  
# --- Put Your information ---  
CLIENT_ID = "83485f1eb00e4abab0576f4aa1d648ec"  
CLIENT_SECRET = "27d78043a4b749f6aaa7a5fc8f1f35b3"  
REDIRECT_URI = "http://127.0.0.1:8080/callback"  
  
# --- Authentication ---  
scope = "playlist-read-private playlist-read-collaborative"`

```

sp = spotipy.Spotify(auth_manager=SpotifyOAuth(client_id=CLIENT_ID,
                                               client_secret=CLIENT_SECRET,
                                               redirect_uri=REDIRECT_URI,
                                               scope=scope))

# --- ID OF PLAYLIST ---
playlist_link = "https://open.spotify.com/playlist/3Zi2PREAkA0WZWYfF7qrDt?si=a6c
playlist_id = playlist_link.split('/')[-1].split('?')[0]

# --- Function to get ALL tracks in a playlist ---
def get_all_tracks(playlist_id, sp):
    tracks = []
    results = sp.playlist_items(playlist_id, offset=0)
    tracks.extend(results['items'])
    while results['next']:
        results = sp.next(results)
        tracks.extend(results['items'])
    return tracks

# --- Get all tracks ---
tracks = get_all_tracks(playlist_id, sp)

# --- Function to get country from MusicBrainz ---
def get_artist_country(artist_name):
    try:
        url = f"https://musicbrainz.org/ws/2/artist/?query=artist:{artist_name}&
        response = requests.get(url, headers={"User-Agent": "SpotifyCountryCheck"})
        data = response.json()
        for artist in data.get("artists", []):
            country = artist.get("country")
            if country:
                return country
        return None
    except Exception as e:
        print(f"Error with {artist_name}: {e}")
        return None

# --- Extract info for each track ---
data = []
for item in tracks:
    track = item['track']
    if track is None:
        continue # Skip if track doesn't exist

    name = track['name']
    duration_ms = track['duration_ms']
    duration_sec = round(duration_ms / 1000, 2)
    artists = [a['name'] for a in track['artists']]

    # Make sure there is at least one artist with a valid ID
    if not track['artists'] or not track['artists'][0].get('id'):
        genres = []
    else:
        artist_id = track['artists'][0]['id']
        artist_info = sp.artist(artist_id)
        genres = artist_info.get('genres', [])

    # Get country of first artist
    first_artist = artists[0] if artists else ""
    country = get_artist_country(first_artist)

```

```
time.sleep(1) # To respect MusicBrainz API limits

data.append({
    "Song Name": name,
    "Artists": ", ".join(artists),
    "Genres": ", ".join(genres),
    "Duration (sec)": duration_sec,
    "Country": country
})

# --- Create DataFrame and export to Excel ---
df = pd.DataFrame(data)
df.to_excel("playlist_songs_with_country.xlsx", index=False)
print("¡Excel exportado exitosamente!")
df.head()
```

Error with Natanael Cano: ('Connection aborted.', ConnectionResetError(10054, 'Se ha forzado la interrupción de una conexión existente por el host remoto', None, 10054, None))  
Error with Kim Carnes: ('Connection aborted.', ConnectionResetError(10054, 'Se ha forzado la interrupción de una conexión existente por el host remoto', None, 10054, None))  
Error with Rod Stewart: ('Connection aborted.', ConnectionResetError(10054, 'Se ha forzado la interrupción de una conexión existente por el host remoto', None, 10054, None))  
Error with Sidewalks and Skeletons: ('Connection aborted.', ConnectionResetError(10054, 'Se ha forzado la interrupción de una conexión existente por el host remoto', None, 10054, None))  
Error with Grupo Frontera: ('Connection aborted.', ConnectionResetError(10054, 'Se ha forzado la interrupción de una conexión existente por el host remoto', None, 10054, None))  
Error with Romeo Santos: ('Connection aborted.', ConnectionResetError(10054, 'Se ha forzado la interrupción de una conexión existente por el host remoto', None, 10054, None))  
Error with N'Samble: ('Connection aborted.', ConnectionResetError(10054, 'Se ha forzado la interrupción de una conexión existente por el host remoto', None, 10054, None))  
Error with New West: ('Connection aborted.', ConnectionResetError(10054, 'Se ha forzado la interrupción de una conexión existente por el host remoto', None, 10054, None))  
Error with Los Ronisch: ('Connection aborted.', ConnectionResetError(10054, 'Se ha forzado la interrupción de una conexión existente por el host remoto', None, 10054, None))  
Error with Dino: ('Connection aborted.', ConnectionResetError(10054, 'Se ha forzado la interrupción de una conexión existente por el host remoto', None, 10054, None))  
Error with Luis Enrique: ('Connection aborted.', ConnectionResetError(10054, 'Se ha forzado la interrupción de una conexión existente por el host remoto', None, 10054, None))  
Error with Armonía 10: ('Connection aborted.', ConnectionResetError(10054, 'Se ha forzado la interrupción de una conexión existente por el host remoto', None, 10054, None))  
Error with Sebastian Yatra: ('Connection aborted.', ConnectionResetError(10054, 'Se ha forzado la interrupción de una conexión existente por el host remoto', None, 10054, None))  
Error with Empire Of The Sun: ('Connection aborted.', ConnectionResetError(10054, 'Se ha forzado la interrupción de una conexión existente por el host remoto', None, 10054, None))  
Error with Bryan Adams: ('Connection aborted.', ConnectionResetError(10054, 'Se ha forzado la interrupción de una conexión existente por el host remoto', None, 10054, None))  
Error with Soda Stereo: ('Connection aborted.', ConnectionResetError(10054, 'Se ha forzado la interrupción de una conexión existente por el host remoto', None, 10054, None))  
Error with ZWE1HVNDXR: ('Connection aborted.', ConnectionResetError(10054, 'Se ha forzado la interrupción de una conexión existente por el host remoto', None, 10054, None))  
Error with Jerry Rivera: ('Connection aborted.', ConnectionResetError(10054, 'Se ha forzado la interrupción de una conexión existente por el host remoto', None, 10054, None))  
Error with Internacional Mallanep: ('Connection aborted.', ConnectionResetError(10054, 'Se ha forzado la interrupción de una conexión existente por el host remoto', None, 10054, None))  
Error with Gorillaz: ('Connection aborted.', ConnectionResetError(10054, 'Se ha forzado la interrupción de una conexión existente por el host remoto', None, 10054, None))

Error with Ed Maverick: ('Connection aborted.', ConnectionResetError(10054, 'Se ha forzado la interrupción de una conexión existente por el host remoto', None, 10054, None))  
 Error with Survivor: ('Connection aborted.', ConnectionResetError(10054, 'Se ha forzado la interrupción de una conexión existente por el host remoto', None, 10054, None))  
 Error with Richard Ashcroft: ('Connection aborted.', ConnectionResetError(10054, 'Se ha forzado la interrupción de una conexión existente por el host remoto', None, 10054, None))  
 Error with The Clash: ('Connection aborted.', ConnectionResetError(10054, 'Se ha forzado la interrupción de una conexión existente por el host remoto', None, 10054, None))  
 Error with Cuco: ('Connection aborted.', ConnectionResetError(10054, 'Se ha forzado la interrupción de una conexión existente por el host remoto', None, 10054, None))  
 Error with Charly García: ('Connection aborted.', ConnectionResetError(10054, 'Se ha forzado la interrupción de una conexión existente por el host remoto', None, 10054, None))  
 Error with Hard-Fi: ('Connection aborted.', ConnectionResetError(10054, 'Se ha forzado la interrupción de una conexión existente por el host remoto', None, 10054, None))  
 Error with Rio: ('Connection aborted.', ConnectionResetError(10054, 'Se ha forzado la interrupción de una conexión existente por el host remoto', None, 10054, None))  
 ¡Excel exportado exitosamente!

Out[3]:

	Song Name	Artists	Genres	Duration (sec)	Country
0	La Carretera	Prince Royce	bachata, latin	238.71	US
1	Cada Vez Que Digo Adiós	Los Enanitos Verdes	rock en español, latin rock, argentine rock, m...	232.08	AR
2	Ebrio De Amor	Valentín Elizalde	banda, norteño, corrido, grupera	167.53	MX
3	Amor Tumbado	Natanael Cano	corridos tumbados, corrido, corridos bélicos, ...	221.55	None
4	Sunflower - Spider-Man: Into the Spider-Verse	Post Malone, Swae Lee		157.56	US

In [4]: df.shape

Out[4]: (2053, 5)

In [5]: pip install requests

```
Requirement already satisfied: requests in c:\users\acer\anaconda3\lib\site-packages (2.32.3)
Requirement already satisfied: charset-normalizer<4,>=2 in c:\users\acer\anaconda3\lib\site-packages (from requests) (3.3.2)
Requirement already satisfied: idna<4,>=2.5 in c:\users\acer\anaconda3\lib\site-packages (from requests) (3.7)
Requirement already satisfied: urllib3<3,>=1.21.1 in c:\users\acer\anaconda3\lib\site-packages (from requests) (2.3.0)
Requirement already satisfied: certifi>=2017.4.17 in c:\users\acer\anaconda3\lib\site-packages (from requests) (2025.4.26)
```

```
In [6]: import requests
import time

def get_artist_country(artist_name):
    try:
        url = f"https://musicbrainz.org/ws/2/artist/?query=artist:{artist_name}&
response = requests.get(url, headers={"User-Agent": "SpotifyCountryCheck"})
data = response.json()
# Busca el primer artista con country disponible
for artist in data.get("artists", []):
    country = artist.get("country")
    if country:
        return country
return None
except Exception as e:
    print(f"Error with {artist_name}: {e}")
    return None
```

```
In [8]: df["Country"].isnull().sum()
```

```
Out[8]: np.int64(58)
```

```
In [9]: df.to_excel("C:/Users/Acer/Documents/Portafolio/playlist_songs_with_country.xlsx")
```

```
In [10]: df
```

Out[10]:

	Song Name	Artists	Genres	Duration (sec)	Country
0	La Carretera	Prince Royce	bachata, latin	238.71	US
1	Cada Vez Que Digo Adiós	Los Enanitos Verdes	rock en español, latin rock, argentine rock, m...	232.08	AR
2	Ebrio De Amor	Valentín Elizalde	banda, norteño, corrido, grupera	167.53	MX
3	Amor Tumbado	Natanael Cano	corridos tumbados, corrido, corridos bélicos, ...	221.55	None
4	Sunflower - Spider-Man: Into the Spider-Verse	Post Malone, Swae Lee		157.56	US
...	...	...	...	...	...
2048	Leila	Indochine	variété française, chanson, french pop	233.40	FR
2049	I Took A Pill In Ibiza - Seeb Remix	Mike Posner, Seeb		199.23	US
2050	Stole the Show	Kygo, Parson James	tropical house	222.80	NO
2051	Fast Car	Jonas Blue, Dakota	tropical house	212.42	GB
2052	Firestone	Kygo, Conrad Sewell	tropical house	273.68	NO

2053 rows × 5 columns

In [11]:

```
duplicates = df[df.duplicated(subset=["Song Name", "Artists"], keep=False)]
print(duplicates)
```

	Song Name	Artists	\
18	My Way	Calvin Harris	
29	Si Antes Te Hubiera Conocido	KAROL G	
30	Stolen Dance	Milky Chance	
70	Si Antes Te Hubiera Conocido	KAROL G	
71	Mil Noches	Orquesta Caribeños de Guadalupe	
...	...	...	
2021	Me Resfrié en Brasil	Arena Hash	
2024	Bésame	Arena Hash	
2025	El túnel del bus	Arena Hash	
2036	My Way	Calvin Harris	
2044	Maniac	Michael Sembello	
	Genres	Duration (sec)	Country
18	edm	219.16	US
29	reggaeton, latin, urbano latino	195.82	DE
30		313.68	DE
70	reggaeton, latin, urbano latino	195.82	DE
71	cumbia, chicha, cumbia norteña, huayno	269.67	PE
...	...	...	...
2021	rock en español	227.31	PE
2024	rock en español	221.24	PE
2025	rock en español	275.11	PE
2036	edm	219.16	US
2044		244.71	US

[102 rows x 5 columns]

```
In [12]: # Eliminar duplicados donde Song Name y Artists sean idénticos
df = df.drop_duplicates(subset=["Song Name", "Artists"], keep="first").reset_index()

In [13]: duplicates = df[df.duplicated(subset=["Song Name", "Artists"], keep=False)]
print(duplicates)

Empty DataFrame
Columns: [Song Name, Artists, Genres, Duration (sec), Country]
Index: []

In [14]: df.shape

Out[14]: (2002, 5)

In [15]: df["Country"].isnull().sum()

Out[15]: np.int64(56)

In [16]: filas_vacias = df[df["Country"].isna()]
print(filas_vacias)
```

	Song Name \
3	Amor Tumbado
7	Bette Davis Eyes
8	Young Turks - 2008 Remaster
14	Beanie
84	Passagem so de ida
129	Once Upon A Time ( Remix )
136	MTG - RITMO ZELDA - Super Slowed
137	SPOOKY
139	FUNK DO BOUNCE (Slowed)
151	Goth
215	NO SE VA - EN VIVO
218	Montagem Lunar Celestia 1.0 (SLOWED)
219	MAZKO
220	BUFÓN!
238	Esto Es Ser Adulto
373	Un nuevo amor
409	Imitadora
422	Dream Girl
426	Hoy He Vuelto a Ver al Amor
427	Those Eyes
462	Dirty Harry (feat. Bootie Brown)
498	Are You Bored Yet? (feat. Clairo)
506	Five More Hours
527	Ya No Te Puedo Perdonar
551	Mil Años
554	Yo No Sé Mañana
559	Mil Botellas
561	Traicionera
570	Walking On A Dream
591	The Automotivo Infernal 1.0 - Purple
665	Heaven
674	FRESH
677	Signos - Remasterizado 2007
683	LOVELY BASTARDS
720	MONTAGEM - PR FUNK
902	The Perfect Girl
903	After Dark
933	De Sol A Sol
951	Un Amor Como El Nuestro
961	Dónde Está El Amor
1020	Loco
1021	On Melancholy Hill
1107	Karpe Diem
1111	Red Eyes
1112	Eye of the Tiger
1168	Born to Be Strangers
1173	Should I Stay or Should I Go - Remastered
1234	Alexa Chung
1242	Summer Time High Time
1506	Feels Good
1716	Disculpa los Malos Pensamientos
1717	Narcisista por Excelencia
1718	Procedimientos para Llegar a un Común Acuerdo
1719	Los Malaventurados No Lloran
1773	Huir de Ti
1974	Magdalena

Artists \  
Natanael Cano

7	Kim Carnes
8	Rod Stewart
14	Chezile
84	Bigodee
129	Leasis
136	JXM, Polaris
137	fennecxx, DJ ALEX
139	Ariis
151	Sidewalks and Skeletons
215	Grupo Frontera
218	TOKYOPHILE
219	S4nri0, MC LyC4N
220	Scythermane, NXVAMANE, LXKERSON
238	Monkeyvan
373	Tranzas
409	Romeo Santos
422	Crisaunt
426	N'Samble
427	New West
462	Gorillaz, Bootie Brown
498	Wallows, Clairo
506	Deorro, Chris Brown
527	Los Ronisch
551	Dino
554	Luis Enrique
559	Armonía 10
561	Sebastian Yatra
570	Empire Of The Sun
591	MRL, Mc Gw
665	Bryan Adams
674	NXVAMANE
677	Soda Stereo
683	ZWE1HVNDXR, yatashigang
720	S3BZS, Mc Gw, Mc Menor Do Alvorada
902	Mareux
903	Mr.Kitty
933	Salserin
951	Salserin
961	Internacional Mallanep
1020	M2H
1021	Gorillaz
1107	Ed Maverick
1111	Cosmovision
1112	Survivor
1168	Richard Ashcroft
1173	The Clash
1234	Moodhosa
1242	Cuco, Kwe\$t
1506	Hard-Fi
1716	PXNDX
1717	PXNDX
1718	PXNDX
1719	PXNDX
1773	Rio
1974	NSQNSC

	Genres	Duration (sec)	\
3	corridos tumbados, corrido, corridos bélicos, ...	221.55	
7		228.00	
8	soft rock	302.41	

14		132.16
84		145.87
129		138.62
136	phonk, brazilian phonk	107.09
137	phonk, brazilian phonk	96.92
139	phonk, brazilian phonk	107.54
151	witch house, darkwave	207.12
215	cumbia norteña, banda, norteño, música mexican...	193.82
218	phonk, brazilian phonk	97.32
219	phonk, brazilian phonk	75.69
220	phonk, brazilian phonk	71.41
238		168.00
373	latin pop, colombian pop	230.80
409	bachata, latin	234.83
422	bedroom pop	179.29
426	timba, salsa	334.97
427		220.75
462		230.43
498	indie	178.00
506	melbourne bounce, electro house	211.97
527	chicha, electrocumbia, huayno, cumbia	235.05
551	bolero	294.00
554	salsa, merengue	259.85
559	cumbia, chicha, cumbia norteña, huayno, electr...	296.56
561	colombian pop, reggaeton	229.39
570		198.44
591	phonk, brazilian phonk	114.59
665		243.36
674	phonk, brazilian phonk	68.69
677	rock en español, latin rock, argentine rock, l...	315.35
683	phonk, brazilian phonk	116.10
720	phonk, brazilian phonk	61.63
902	darkwave, cold wave	194.42
903	witch house, darkwave	259.15
933	salsa, timba, merengue	297.07
951	salsa, timba, merengue	328.52
961	cumbia, chicha, huayno	238.22
1020		237.42
1021		233.87
1107	mexican indie, latin folk	203.48
1111		218.00
1112		245.64
1168	britpop, madchester	253.16
1173	punk	188.99
1234		191.20
1242	bedroom pop	208.07
1506	britpop	234.01
1716	rock en español, mexican rock, latin rock	238.37
1717	rock en español, mexican rock, latin rock	227.35
1718	rock en español, mexican rock, latin rock	194.16
1719	rock en español, mexican rock, latin rock	207.79
1773	rock en español, latin rock	287.23
1974		160.00

## Country

3	None
7	None
8	None
14	None
84	None

129	None
136	None
137	None
139	None
151	None
215	None
218	None
219	None
220	None
238	None
373	None
409	None
422	None
426	None
427	None
462	None
498	None
506	None
527	None
551	None
554	None
559	None
561	None
570	None
591	None
665	None
674	None
677	None
683	None
720	None
902	None
903	None
933	None
951	None
961	None
1020	None
1021	None
1107	None
1111	None
1112	None
1168	None
1173	None
1234	None
1242	None
1506	None
1716	None
1717	None
1718	None
1719	None
1773	None
1974	None

```
In [17]: df.to_excel("C:/Users/Acer/Documents/Portafolio/playlist_songs_with_country_clea
```

```
In [ ]:
```