

ENTREGABLE 1 - EXTRAER DATOS DE API DE SPOTIFY

Para obtener credenciales y usar la API de Spotify, necesitas registrarte como desarrollador en la plataforma de Spotify y crear una aplicación. Aquí están los pasos generales que debes seguir:

Crea una cuenta de Spotify: Si no tienes una cuenta de Spotify, crea una en el sitio web de Spotify.

Inicia sesión en el Panel de Desarrolladores de Spotify: Visita el Panel de Desarrolladores de Spotify en <https://developer.spotify.com/dashboard/login>.

Crea una nueva aplicación: Una vez que hayas iniciado sesión, puedes crear una nueva aplicación. Proporciona un nombre y una descripción para tu aplicación.

Configura los ajustes de la aplicación:

Especifica la descripción y los detalles de la aplicación. Indica si tu aplicación es comercial o personal. Proporciona la dirección de correo electrónico de contacto. Selecciona las capacidades que tu aplicación usará (por ejemplo, acceder a datos de usuario, controlar dispositivos, etc.). Aceptar los términos y condiciones: Asegúrate de leer y aceptar los términos y condiciones de la plataforma de desarrolladores de Spotify.

Obtén las credenciales de API:

Después de crear la aplicación, se te proporcionarán las credenciales de API, que generalmente incluyen un ID de cliente (client ID) y una clave secreta (client secret). Estas credenciales serán necesarias para autenticarte y realizar solicitudes a la API de Spotify. Configura redireccionamientos de URI: En la configuración de la aplicación, es posible que necesites especificar los redireccionamientos de URI permitidos, que se utilizan en el flujo de autenticación. Esto depende del tipo de autenticación que utilices.

Utiliza las credenciales en tu aplicación: Ahora puedes usar las credenciales de API en tu aplicación para autenticarte y realizar solicitudes a la API de Spotify.

Es importante tener en cuenta que la API de Spotify puede requerir diferentes tipos de autenticación según tus necesidades. Puedes encontrar más información sobre cómo autenticarte y cómo realizar solicitudes específicas en la documentación oficial de la API de Spotify.

Recuerda también que es fundamental cumplir con las políticas de uso y términos de servicio de Spotify al desarrollar aplicaciones que interactúen con su plataforma.

```
In [82]: !pip install spotipy
!pip install wheel
```

```
!pip install pandas
!pip install psycpg2
```

```
[notice] A new release of pip is available: 23.2.1 -> 23.3.1
[notice] To update, run: python.exe -m pip install --upgrade pip
Requirement already satisfied: spotipy in c:\users\cnieto1\anaconda3\envs\dhdsblend2021\lib\site-packages (2.23.0)
Requirement already satisfied: redis>=3.5.3 in c:\users\cnieto1\anaconda3\envs\dhdsblend2021\lib\site-packages (from spotipy) (5.0.0)
Requirement already satisfied: requests>=2.25.0 in c:\users\cnieto1\anaconda3\envs\dhdsblend2021\lib\site-packages (from spotipy) (2.27.1)
Requirement already satisfied: six>=1.15.0 in c:\users\cnieto1\anaconda3\envs\dhdsblend2021\lib\site-packages (from spotipy) (1.16.0)
Requirement already satisfied: urllib3>=1.26.0 in c:\users\cnieto1\anaconda3\envs\dhdsblend2021\lib\site-packages (from spotipy) (1.26.9)
Requirement already satisfied: async-timeout>=4.0.2 in c:\users\cnieto1\anaconda3\envs\dhdsblend2021\lib\site-packages (from redis>=3.5.3->spotipy) (4.0.2)
Requirement already satisfied: certifi>=2017.4.17 in c:\users\cnieto1\anaconda3\envs\dhdsblend2021\lib\site-packages (from requests>=2.25.0->spotipy) (2022.6.15)
Requirement already satisfied: charset-normalizer~=2.0.0 in c:\users\cnieto1\anaconda3\envs\dhdsblend2021\lib\site-packages (from requests>=2.25.0->spotipy) (2.0.4)
Requirement already satisfied: idna<4,>=2.5 in c:\users\cnieto1\anaconda3\envs\dhdsblend2021\lib\site-packages (from requests>=2.25.0->spotipy) (3.3)
Requirement already satisfied: wheel in c:\users\cnieto1\anaconda3\envs\dhdsblend2021\lib\site-packages (0.37.1)
```

```
[notice] A new release of pip is available: 23.2.1 -> 23.3.1
[notice] To update, run: python.exe -m pip install --upgrade pip
Requirement already satisfied: pandas in c:\users\cnieto1\anaconda3\envs\dhdsblend2021\lib\site-packages (1.1.5)
```

```
[notice] A new release of pip is available: 23.2.1 -> 23.3.1
[notice] To update, run: python.exe -m pip install --upgrade pip
Requirement already satisfied: python-dateutil>=2.7.3 in c:\users\cnieto1\anaconda3\envs\dhdsblend2021\lib\site-packages (from pandas) (2.8.2)
Requirement already satisfied: pytz>=2017.2 in c:\users\cnieto1\anaconda3\envs\dhdsblend2021\lib\site-packages (from pandas) (2021.3)
Requirement already satisfied: numpy>=1.15.4 in c:\users\cnieto1\anaconda3\envs\dhdsblend2021\lib\site-packages (from pandas) (1.19.2)
Requirement already satisfied: six>=1.5 in c:\users\cnieto1\anaconda3\envs\dhdsblend2021\lib\site-packages (from python-dateutil>=2.7.3->pandas) (1.16.0)
Requirement already satisfied: psycpg2 in c:\users\cnieto1\anaconda3\envs\dhdsblend2021\lib\site-packages (2.9.7)
```

```
[notice] A new release of pip is available: 23.2.1 -> 23.3.1
[notice] To update, run: python.exe -m pip install --upgrade pip
```

```
In [83]: #results = {
#         'tracks': {
#             'items': [
#                 # Lista de pistas (cada pista es un diccionario)
#                 {
#                     'id': '...',
#                     'name': '...',
#                     'artists': [...],
#                     'album': {...},
#                     'duration_ms': ...,
#                     'popularity': ...
#                 }
#                 # Otros campos específicos de la pista
#             },
#             # Más pistas...
#         }
```

```
#    }
#}
```

```
In [84]: # Guardo mi contraseña "client_secret" de la API de Spotify en un archivo .txt por cue
with open("C:/Users/cnieto1/Desktop/Curso Data Engineering - Coderhouse/Clases/Entrega
pwd= f.read()
```

```
In [85]: #Chequeo si la variable pwd tomó bien el archivo txt:
import os

file_path = "C:/Users/cnieto1/Desktop/Curso Data Engineering - Coderhouse/Clases/Entrega
if os.path.exists(file_path):
    with open(file_path, 'r') as f:
        #pwd = 'cbc24622177244b7a7a7db138da28ae7'
        pwd = f.read()
else:
    print("El archivo no existe en la ruta especificada.")
```

```
In [86]: #Ingreso mis credenciales de la API de Spotify:
import pandas as pd
import spotipy
from spotipy.oauth2 import SpotifyClientCredentials
import datetime
import time

client_id = 'dbe61651ee31461681339d9d1780f672'
client_secret = pwd

sp = spotipy.Spotify(auth_manager=SpotifyClientCredentials(client_id, client_secret))
```

```
In [87]: #Consultando Los datos de La API de Spotify:
import pandas as pd
results = sp.search(q='year:2023', type='track', limit=50)
data = {'Id': [], 'Artista': [], 'Cancion': [], 'Duracion_ms': [], 'Genero': [], 'Album': []}
for track in results['tracks']['items']:
    id = track['id']
    artist_name = track['artists'][0]['name']
    artist_id = track['artists'][0]['id']
    track_name = track['name']
    duration_ms = track['duration_ms']
    track_id = track['id']
    album_group = track['album']['name']
    album_img = track['album']['images'][0]['url'] #imagen de album
    album_cont = track['album']['total_tracks']
    track_genre = sp.artist(artist_id)['genres']
    track_popularity = track['popularity']
    track_year = track['album']['release_date']
    preview_url = track['preview_url']
    #Quitar las comillas
    track_name = track_name.replace('"', '')
    album_group = album_group.replace('"', '')
    #Separar el género por coma
    track_genre = ', '.join(track_genre)

    data['Id'].append(id)
    data['Artista'].append(artist_name)
    data['Cancion'].append(track_name)
    data['Duracion_ms'].append(duration_ms)
```

```

data['Album'].append(album_group)
data['Album_img'].append(album_img)
data['Total_canciones_album'].append(album_cont)
data['Genero'].append(track_genre)
data['Popularidad'].append(track_popularity)
data['fecha_lanzamiento'].append(track_year)
data['Audio_preview'].append(preview_url)

df = pd.DataFrame(data)
#Evitar que haya canciones duplicadas
df.drop_duplicates(subset=['Artista', 'Cancion', 'Album'], keep='first', inplace=True)
#Reemplazar valores nulos o vacios en el campo Género por Desconocido
df['Genero'].fillna('Desconocido', inplace=True)
df.loc[df['Genero'] == '', 'Genero'] = 'Desconocido'
#Evitar que se cargue una canción con duración 0 ms
df = df[df['Duracion_ms'] != 0]
#Verificar que la fecha se muestre en formato fecha
df['fecha_lanzamiento'] = pd.to_datetime(df['fecha_lanzamiento'], format='%Y-%m-%d')
#display(df)
df.head()

```

Out[87]:

			Id	Artista	Cancion	Duracion_ms	Genero	Album	
0	4KULAYmBBJcPRpk1yO4dOG	Zach Bryan			Remember Everything (feat. Kacey Musgraves)	227195	classic oklahoma country, modern country pop	Zach Bryan	https://i.scdn
1	4rXLjWdF2ZZpXCVTfWcshS	Gunna			fukumean	125040	atl hip hop, melodic rap, rap, trap	a Gift & a Curse	https://i.scdn
2	2IGMVunlBsBLtEQyoI1Mu7	Doja Cat			Paint The Town Red	231750	dance pop, pop	Paint The Town Red	https://i.scd
3	2YSzYUF3jWqb9YP9VXmpjE	Drake			IDGAF (feat. Yeat)	260111	canadian hip hop, canadian pop, hip hop, pop r...	For All The Dogs	https://i.scdn
4	3vkCueOmm7xQDoJ17W1Pm3	Mitski			My Love Mine All Mine	137773	brooklyn indie, pov: indie	The Land Is Inhospitable and So Are We	https://i.scd

```

In [88]: # Construyo mi Dataframe con las columnas que considero más importantes y con el order
#df = {'Id': [], 'Artista': [], 'Cancion': [], 'Duracion_ms': [], 'Album': [], 'Album_img
#data = pd.DataFrame()

#Agrego la columna de 'Insert_date' a mi dataframe "data":
from datetime import datetime
current_date = datetime.now().strftime('%Y-%m-%d')

```

```
df['Insert_date'] = current_date

df.head()
```

Out[88]:

		Id	Artista	Cancion	Duracion_ms	Genero	Album	
0	4KULAYmBBJcPRpk1yO4dOG		Zach Bryan	Remember Everything (feat. Kacey Musgraves)	227195	classic oklahoma country, modern country pop	Zach Bryan	https://i.scdn
1	4rXLjWdF2ZZpXCVTfWcshS		Gunna	fukumean	125040	atl hip hop, melodic rap, rap, trap	a Gift & a Curse	https://i.scdn
2	2lGMVunlBsBLtEQyo1Mu7		Doja Cat	Paint The Town Red	231750	dance pop, pop	Paint The Town Red	https://i.scd
3	2YSzYUF3jWqb9YP9VXmpjE		Drake	IDGAF (feat. Yeat)	260111	canadian hip hop, canadian pop, hip hop, pop r...	For All The Dogs	https://i.scdn
4	3vkCueOmm7xQDoJ17W1Pm3		Mitski	My Love Mine All Mine	137773	brooklyn indie, pov: indie	The Land Is Inhospitable and So Are We	https://i.scd

```
In [89]: # Guardo mi contraseña "pwd_redshift" de mi cuenta de Redshift en un archivo .txt por
with open("C:/Users/cnieto1/Desktop/Curso Data Engineering - Coderhouse/Clases/Entrega
pwd_redshift= f.read()
```

```
In [90]: # Creando la conexión a Redshift
import psycopg2
url="data-engineer-cluster.cyhh5bfevlmn.us-east-1.redshift.amazonaws.com"
data_base="data-engineer-database"
user="christian_r_coderhouse"

try:
    conn = psycopg2.connect(
        host='data-engineer-cluster.cyhh5bfevlmn.us-east-1.redshift.amazonaws.com',
        dbname=data_base,
        user=user,
        #password='3b4LjN1aLG',
        password=pwd_redshift,
        port='5439'
    )
    print("Conectado a Redshift con éxito!")

except Exception as e:
```

```
print("No es posible conectar a Redshift")
print(e)
```

Conectado a Redshift con éxito!

```
In [91]: #Crear la tabla si no existe
with conn.cursor() as cur:
    cur.execute("""
        CREATE TABLE IF NOT EXISTS canciones
        (
            id VARCHAR(50) primary key
            ,artista VARCHAR(255)
            ,cancion VARCHAR(255)
            ,duracion_ms INTEGER
            ,genero VARCHAR(300)
            ,album VARCHAR(100)
            ,album_img VARCHAR(300)
            ,total_canciones_album INTEGER
            ,Popularidad INTEGER
            ,fecha_lanzamiento date
            ,Audio_preview NVARCHAR(300)
            ,Insert_date date
        )
    """)
    conn.commit()
```

```
In [92]: #Vaciar la tabla para evitar duplicados o inconsistencias
with conn.cursor() as cur:
    cur.execute("Truncate table canciones")
    count = cur.rowcount
```

```
In [93]: #Eliminar la tabla para evitar duplicados o inconsistencias
#with conn.cursor() as cur:
#    cur.execute("Drop table canciones")
#    count = cur.rowcount
```

```
In [94]: #consultando la tabla canciones
cur = conn.cursor()
cur.execute("SELECT * FROM canciones")
results = cur.fetchall()
results
```

```
Out[94]: []
```

```
In [95]: df.dtypes
```

```
Out[95]: Id                                object
Artista                                object
Cancion                                object
Duracion_ms                           int64
Genero                                object
Album                                 object
Album_img                             object
Total_canciones_album                 int64
Popularidad                           int64
fecha_lanzamiento                     datetime64[ns]
Audio_preview                         object
Insert_date                           object
dtype: object
```

```
In [96]: #Insertando Los datos en Redsfhifit
from psycopg2.extras import execute_values
with conn.cursor() as cur:
    execute_values(
        cur,
        '''
        INSERT INTO canciones (Id, Artista, Cancion, Duracion_ms, Genero, Album, Album
VALUES %s
        ''',
        [tuple(row) for row in df.values],
        page_size=len(df)
    )
    conn.commit()
```

```
In [97]: # Veo como quedó la tabla en Redshift luego de hacer los Insert:
#consultando la tabla
cur = conn.cursor()
cur.execute("SELECT * FROM canciones")
results = cur.fetchall()
```

```
In [98]: # Veo cómo quedó la tabla "canciones" en Redshift:
column_names=['Id', 'Artista', 'Cancion', 'Genero', 'Album', 'Total_canciones_album',
df = pd.DataFrame(results, columns=column_names)
df.head()
```

Out[98]:

			Id	Artista	Cancion	Genero	Album	Total_canciones_album	Po
0	4KULAymBBJcPRpk1yO4dOG			Zach Bryan	Remember Everything (feat. Kacey Musgraves)	classic oklahoma country, modern country pop	Zach Bryan	16	
1	4rXLjWdF2ZZpXCVTfWcshS			Gunna	fukumean	atl hip hop, melodic rap, rap, trap	a Gift & a Curse	15	
2	2IGMVunlBsBLtEQyo11Mu7			Doja Cat	Paint The Town Red	dance pop, pop	Paint The Town Red	1	
3	2YSzYUF3jWqb9YP9VXmpjE			Drake	IDGAF (feat. Yeat)	canadian hip hop, canadian pop, hip hop, pop r...	For All The Dogs	23	
4	3vkCueOmm7xQDoJ17W1Pm3			Mitski	My Love Mine All Mine	brooklyn indie, pov: indie	The Land Is Inhospitable and So Are We	11	

```
In [99]: cur.close()  
         conn.close()
```