

ENTREGABLE 1 - EXTRAER DATOS DE API DE YOUTUBE E INSERTAR LA DATA EN UNA TABLA DE AWS REDSHIFT.

Explicación del Proyecto:

El script siguiente obtiene los 10 videos más populares (con más vistas) de YouTube en el momento de la ejecución del script. La API realiza una búsqueda de videos, ordenándolos por la cantidad de vistas (ordenados en orden descendente). Luego, obtiene los detalles de esos videos y crea un DataFrame con la información relevante, incluida la cantidad de vistas, y lo imprime.

Además, en la tabla resultante de la API le incluyo la columna "Insert_Date", que contiene el día de ejecución del script, para identificar en qué día se obtuvieron los 10 registros con los 10 videos más vistos.

Por último, esos 10 registros son insertados en la tabla "videos" dentro de mi Base de Datos de Redshift, haciendo previamente la conexión a dicha Base de Datos (y creando la tabla en primera instancia).

```
In [1]: # Guardo mi contraseña "api_key" de La API de Youtuber en un archivo .txt por cuestion
with open("C:/Users/cnieto1/Desktop/Curso Data Engineering - Coderhouse/Clases/Entrega
pwd= f.read()
```

```
In [2]: import pandas as pd
from googleapiclient.discovery import build
import datetime

# Definir tu clave de API de YouTube
API_KEY = pwd

# Crear una instancia del servicio de La API de YouTube
youtube = build('youtube', 'v3', developerKey=API_KEY)

# Función para obtener el nombre de La categoría a partir del ID
def get_category_name(youtube, category_id):
    categories_response = youtube.videoCategories().list(
        part='snippet',
        id=category_id
    ).execute()
    if 'items' in categories_response:
        return categories_response['items'][0]['snippet']['title']
    else:
        return 'Desconocida'

# Función para convertir La duración en formato "PT11M13S" a segundos
def convert_duration_to_seconds(duration):
    parts = duration[2:].split('T')[-1].split('H')
    hours = int(parts[0]) if len(parts) > 1 else 0
```

```

minutes_parts = parts[-1].split('M')
minutes = int(minutes_parts[0]) if len(minutes_parts) > 1 else 0
seconds_parts = minutes_parts[-1].split('S')
seconds = int(seconds_parts[0]) if len(seconds_parts) > 1 else 0
total_seconds = hours * 3600 + minutes * 60 + seconds
return total_seconds

# Obtener la fecha actual en el formato requerido por la API de YouTube
current_date = datetime.datetime.now().strftime('%Y-%m-%dT00:00:00Z')

# Realizar la búsqueda de videos ordenados por vistas y limitar a 10 resultados
search_response = youtube.search().list(
    part='id',
    maxResults=10,
    order='viewCount',
    type='video'
).execute()

# Extraer los IDs de los videos obtenidos en la búsqueda
video_ids = [item['id']['videoId'] for item in search_response['items']]

# Obtener detalles de los videos
videos_response = youtube.videos().list(
    part='snippet,statistics,contentDetails',
    id=', '.join(video_ids)
).execute()

# Crear una lista de diccionarios con la información de los videos
video_data = []
for video in videos_response['items']:
    video_id = video['id']
    video_info = {
        "ID_del_Video": video_id,
        "Título": video['snippet']['title'],
        "Descripción": video['snippet']['description'],
        "Canal_Propietario": video['snippet']['channelTitle'],
        "Fecha_de_Publicación": video['snippet']['publishedAt'],
        "Categoría_ID": video['snippet']['categoryId'],
        "Categoría": get_category_name(youtube, video['snippet']['categoryId']),
        "Duración_segundos": convert_duration_to_seconds(video['contentDetails']['duration'], video_id),
        "URL_del_Video": f"https://www.youtube.com/watch?v={video_id}",
        "Vistas": video['statistics']['viewCount'],
        "Likes": video['statistics'].get('likeCount', 0),
        "Dislikes": video['statistics'].get('dislikeCount', 0),
        "Favorite_Count": video['statistics'].get('favoriteCount', 0),
        "Comment_Count": video['statistics'].get('commentCount', 0),
        "Insert_Date": current_date
    }
    video_data.append(video_info)

# Crear un DataFrame a partir de la lista de diccionarios
df = pd.DataFrame(video_data)

# Mostrar el DataFrame
print(df)

```

ID_del_Video	Título \
0 0aZ7lPQ5EXs	El Gallo y la Pata - Canciones de la Granja de...
1 eNLjdPI9zdE	La Vaca Lola - Canciones de La Granja de Zenón 2
2 ebVVuJN1WFM	Bartolito - La Granja de Zenón 3
3 bX3S-_jUauc	Paulo Londra ft Lenny Tavarez - Nena Maldicion...
4 aSjflT_J0Xo	Paulo Londra - Adan y Eva (Official Video)
5 wPNQw8naE2Q	El Pollito Pío 3D - Canciones de la Granja de ...
6 NPpELzyP4rw	Paulo Londra - Tal Vez (Official Video)
7 KATq-Ws3xtM	Patitos Cua Cua Cua - Canciones y clásicos inf...
8 uTK_7MOFV4s	El Marinero Baila - Paco El Marinero El Rein...
9 4ShOpJPHRxA	Bartolito - La Granja de Zenón 3

Descripción	Canal_Propietario \
0 🎁 En estas Navidades, encuentra los productos ...	El Reino Infantil
1 🎁 En estas Navidades, encuentra los productos ...	El Reino Infantil
2 🎁 En estas Navidades, encuentra los productos ...	La Granja de Zenón
3 Paulo Londra "Nena Maldicion"\n\n(Apple Music)...	Paulo Londra
4 Paulo Londra "Adan y Eva"\n\n(Apple) ▶ https:...	Paulo Londra
5 Encuentra las Marionetas de la Granja de Zenón...	El Reino Infantil
6 Paulo Londra "Tal Vez"\n\n(Apple) ▶ https://appl...	Paulo Londra
7 🎁 ¡Bienvenido a El Reino Infantil en Español! A...	El Reino Infantil
8 ¡ATENCIÓN AMIGOS!\n¡Llegó la App de El Reino I...	El Reino Infantil
9 🎁 En estas Navidades, encuentra los productos ...	El Reino Infantil





Fecha_de_Publicación	Categoría_ID	Categoría	Duración_segundos \
0 2014-03-15T14:30:00Z	10	Music	144
1 2014-04-12T14:30:01Z	10	Music	144
2 2015-11-14T13:00:01Z	10	Music	148
3 2018-01-29T22:00:05Z	10	Music	232
4 2018-11-05T18:00:04Z	10	Music	261
5 2014-03-22T14:00:03Z	10	Music	188
6 2019-04-03T22:00:02Z	10	Music	272
7 2014-12-06T12:30:03Z	10	Music	117
8 2014-11-29T12:00:09Z	10	Music	172
9 2015-09-12T12:00:00Z	10	Music	148

URL_del_Video	Vistas	Likes	Dislikes \
0 https://www.youtube.com/watch?v=0aZ7lPQ5EXs	1917903293	4560800	0
1 https://www.youtube.com/watch?v=eNLjdPI9zdE	1825885871	5113775	0
2 https://www.youtube.com/watch?v=ebVVuJN1WFM	1272138375	2880199	0
3 https://www.youtube.com/watch?v=bX3S-_jUauc	1239792434	6595530	0
4 https://www.youtube.com/watch?v=aSjflT_J0Xo	1221520448	7286498	0
5 https://www.youtube.com/watch?v=wPNQw8naE2Q	1210640702	2548502	0
6 https://www.youtube.com/watch?v=NPpELzyP4rw	1080513090	6968824	0
7 https://www.youtube.com/watch?v=KATq-Ws3xtM	1024496522	1643467	0
8 https://www.youtube.com/watch?v=uTK_7MOFV4s	941788194	1483172	0
9 https://www.youtube.com/watch?v=4ShOpJPHRxA	924459950	2973876	0

Favorite_Count	Comment_Count	Insert_Date
0 0	0	2023-08-29T00:00:00Z
1 0	0	2023-08-29T00:00:00Z
2 0	0	2023-08-29T00:00:00Z
3 0	250802	2023-08-29T00:00:00Z
4 0	279538	2023-08-29T00:00:00Z
5 0	0	2023-08-29T00:00:00Z
6 0	212381	2023-08-29T00:00:00Z
7 0	0	2023-08-29T00:00:00Z
8 0	0	2023-08-29T00:00:00Z
9 0	0	2023-08-29T00:00:00Z





In [3]: df.head()

Out[3]:

	ID_del_Video	Título	Descripción	Canal_Propietario	Fecha_de_Publicación	Categoría
0	0aZ7IPQ5EXs	El Gallo y la Pata - Canciones de la Granja de...	 En estas Navidades, encuentra los productos ...	El Reino Infantil	2014-03-15T14:30:00Z	
1	eNLjdPI9zdE	La Vaca Lola - Canciones de La Granja de Zenón 2	 En estas Navidades, encuentra los productos ...	El Reino Infantil	2014-04-12T14:30:01Z	
2	ebVVuJN1WFM	Bartolito - La Granja de Zenón 3	 En estas Navidades, encuentra los productos ...	La Granja de Zenón	2015-11-14T13:00:01Z	
3	bX3S-_jUauc	Paulo Londra ft Lenny Tavarez - Nena Maldicion...	Paulo Londra "Nena Maldicion"\n\n(Apple Music)...	Paulo Londra	2018-01-29T22:00:05Z	
4	aSjflT_J0Xo	Paulo Londra - Adan y Eva (Official Video)	Paulo Londra "Adan y Eva"\n\n(Apple  https:...	Paulo Londra	2018-11-05T18:00:04Z	

In [4]: df

Out[4]:

	ID_del_Video	Título	Descripción	Canal_Propietario	Fecha_de_Publicación	Categoría
0	0aZ7IPQ5EXs	El Gallo y la Pata - Canciones de la Granja de...	 En estas Navidades, encuentra los productos ...	El Reino Infantil	2014-03-15T14:30:00Z	
1	eNLjdPI9zdE	La Vaca Lola - Canciones de La Granja de Zenón 2	 En estas Navidades, encuentra los productos ...	El Reino Infantil	2014-04-12T14:30:01Z	
2	ebVVuJN1WFM	Bartolito - La Granja de Zenón 3	 En estas Navidades, encuentra los productos ...	La Granja de Zenón	2015-11-14T13:00:01Z	
3	bX3S-_jUauc	Paulo Londra ft Lenny Tavarez - Nena Maldicion...	Paulo Londra "Nena Maldicion"\n\n(Apple Music)...	Paulo Londra	2018-01-29T22:00:05Z	
4	aSjflT_J0Xo	Paulo Londra - Adan y Eva (Official Video)	Paulo Londra "Adan y Eva"\n\n(Apple)  https://...	Paulo Londra	2018-11-05T18:00:04Z	
5	wPNQw8naE2Q	El Pollito Pío 3D - Canciones de la Granja de ...	Encuentra las Marionetas de la Granja de Zenón...	El Reino Infantil	2014-03-22T14:00:03Z	
6	NPpELzyP4rw	Paulo Londra - Tal Vez (Official Video)	Paulo Londra "Tal Vez"\n\n(Apple)  https://appl...	Paulo Londra	2019-04-03T22:00:02Z	
7	KATq-Ws3xtM	Patitos Cua Cua Cua - Canciones y clásicos inf...	 ¡Bienvenido a El Reino Infantil en Español! A...	El Reino Infantil	2014-12-06T12:30:03Z	
8	uTK_7MOFV4s	El Marinero Baila - Paco El Marinero El Rein...	¡ATENCIÓN AMIGOS!\n¡Llegó la App de El Reino I...	El Reino Infantil	2014-11-29T12:00:09Z	
9	4ShOpJPHRxA	Bartolito - La Granja	 En estas Navidades, encuentra los productos ...	El Reino Infantil	2015-09-12T12:00:00Z	

ID_del_Video	Título	Descripción	Canal_Propietario	Fecha_de_Publicación	Categoría
de Zenón					

In [5]: *#Obtenemos un Dataframe de 10 registros y 15 columnas:*
`df.shape`

Out[5]: (10, 15)

In [6]: *#Hago las siguientes transformaciones a las columnas del Dataframe df:*

```

# Recortar la columna "Descripción" y "Título" a 301 caracteres:
df['Descripción'] = df['Descripción'].str[:301]
df['Título'] = df['Título'].str[:301]

import pandas as pd
from datetime import datetime

# Convertir la columna "Fecha de Publicación" y "Insert Date" a objetos datetime:
df['Fecha_de_Publicación'] = pd.to_datetime(df['Fecha_de_Publicación'])
df['Insert_Date'] = pd.to_datetime(df['Insert_Date'])

# Formatear la columna "Fecha de Publicación" y "Insert Date" en el formato deseado:
df['Fecha_de_Publicación'] = df['Fecha_de_Publicación'].dt.strftime('%Y-%m-%d')
df['Insert_Date'] = df['Insert_Date'].dt.strftime('%Y-%m-%d')

df.head()
```

Out[6]:

	ID_del_Video	Título	Descripción	Canal_Propietario	Fecha_de_Publicación	Categoría
0	0aZ7IPQ5EXs	El Gallo y la Pata - Canciones de la Granja de...	 En estas Navidades, encuentra los productos ...	El Reino Infantil	2014-03-15	
1	eNLjdPI9zdE	La Vaca Lola - Canciones de La Granja de Zenón 2	 En estas Navidades, encuentra los productos ...	El Reino Infantil	2014-04-12	
2	ebVVuJN1WFM	Bartolito - La Granja de Zenón 3	 En estas Navidades, encuentra los productos ...	La Granja de Zenón	2015-11-14	
3	bX3S-_jUauc	Paulo Londra ft Lenny Tavarez - Nena Maldicion...	Paulo Londra "Nena Maldicion"\n\n(Apple Music)...	Paulo Londra	2018-01-29	
4	aSjflT_J0Xo	Paulo Londra - Adan y Eva (Official Video)	Paulo Londra "Adan y Eva"\n\n(Apple  https:...	Paulo Londra	2018-11-05	

```
In [7]: # 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 [8]: # 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= 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 [9]: # Código para hacer Drop Table de La tabla "canciones". USARLO SOLO en caso de que La
# RECORDAR: Antes de correr este código, correr primero el código anterior (Creando La

# Crear un cursor:
#cur = conn.cursor()

# Ejecutar La sentencia DROP TABLE:
#cur.execute("DROP TABLE IF EXISTS videos")

# Hacer commit para aplicar los cambios:
#conn.commit()
```

```
In [10]: #Crear la tabla si no existe:
with conn.cursor() as cur:
    cur.execute("""
        CREATE TABLE IF NOT EXISTS videos
        (
            Id_del_Video VARCHAR(50) primary key
            ,Título VARCHAR(350)
            ,Descripción VARCHAR(350)
            ,Canal_Propietario VARCHAR(255)
            ,Fecha_de_Publicación date
            ,Categoría_ID VARCHAR(50)
            ,Categoría VARCHAR(100)
            ,Duración_segundos INTEGER
            ,URL_del_Video NVARCHAR(500)
            ,Vistas INTEGER
            ,Likes INTEGER
            ,Dislikes INTEGER
            ,Favorite_Count INTEGER
            ,Comment_Count INTEGER
            ,Insert_Date date

        )
    """)
    conn.commit()
```

```
In [11]: # Comento este paso, para que cada vez que corra el script de La API, que me vaya inse
#Vaciar la tabla para evitar duplicados o inconsistencias:
#with conn.cursor() as cur:
#    cur.execute("Truncate table canciones")
#    count = cur.rowcount
#    count
```

```
In [12]: #consultando La tabla canciones:
cur = conn.cursor()
cur.execute("SELECT * FROM videos")
results = cur.fetchall()
#results
```

```
In [13]: #Insertando Los datos en Redsfhif:
from psycopg2.extras import execute_values
with conn.cursor() as cur:
    execute_values(
        cur,
        """
        INSERT INTO videos (ID_del_Video, Título, Descripción, Canal_Propietario, Fech
        VALUES %s
```







```
    '''
    [tuple(row) for row in df.values],
    page_size=len(df)
)
conn.commit()
```

```
In [14]: # Veo cómo quedó la tabla en Redshift luego de hacer los Insert:
#consultando la tabla
cur = conn.cursor()
cur.execute("SELECT * FROM videos")
results = cur.fetchall()
```

```
In [15]: # Veo cómo quedó la tabla "canciones" en Redshift. Convierto "results" al DataFrame "c"
column_names=['Video_ID', 'Title', 'Description', 'Channel', 'Publication_Date', 'Cate
df_redshift= pd.DataFrame(results, columns=column_names)
df_redshift.head()
```

Out[15]:

	Video_ID	Title	Description	Channel	Publication_Date	Category_ID	Category
0	0aZ7IPQ5EXs	El Gallo y la Pata - Canciones de la Granja de...	 En estas Navidades, encuentra los productos ...	El Reino Infantil	2014-03-15	10	Music
1	eNLjdPI9zdE	La Vaca Lola - Canciones de La Granja de Zenón 2	 En estas Navidades, encuentra los productos ...	El Reino Infantil	2014-04-12	10	Music
2	ebVVuJN1WFM	Bartolito - La Granja de Zenón 3	 En estas Navidades, encuentra los productos ...	La Granja de Zenón	2015-11-14	10	Music
3	bX3S-_jUauc	Paulo Londra ft Lenny Tavarez - Nena Maldicion...	Paulo Londra "Nena Maldicion"\n\n(Apple Music)...	Paulo Londra	2018-01-29	10	Music
4	aSjflT_J0Xo	Paulo Londra - Adan y Eva (Official Video)	Paulo Londra "Adan y Eva"\n\n(Apple  https:...	Paulo Londra	2018-11-05	10	Music

```
In [16]: # Cierro tanto el cursor como la conexión a la base de datos:
cur.close()
conn.close()
```