

Passo 1: Criação de dataframe com os dados

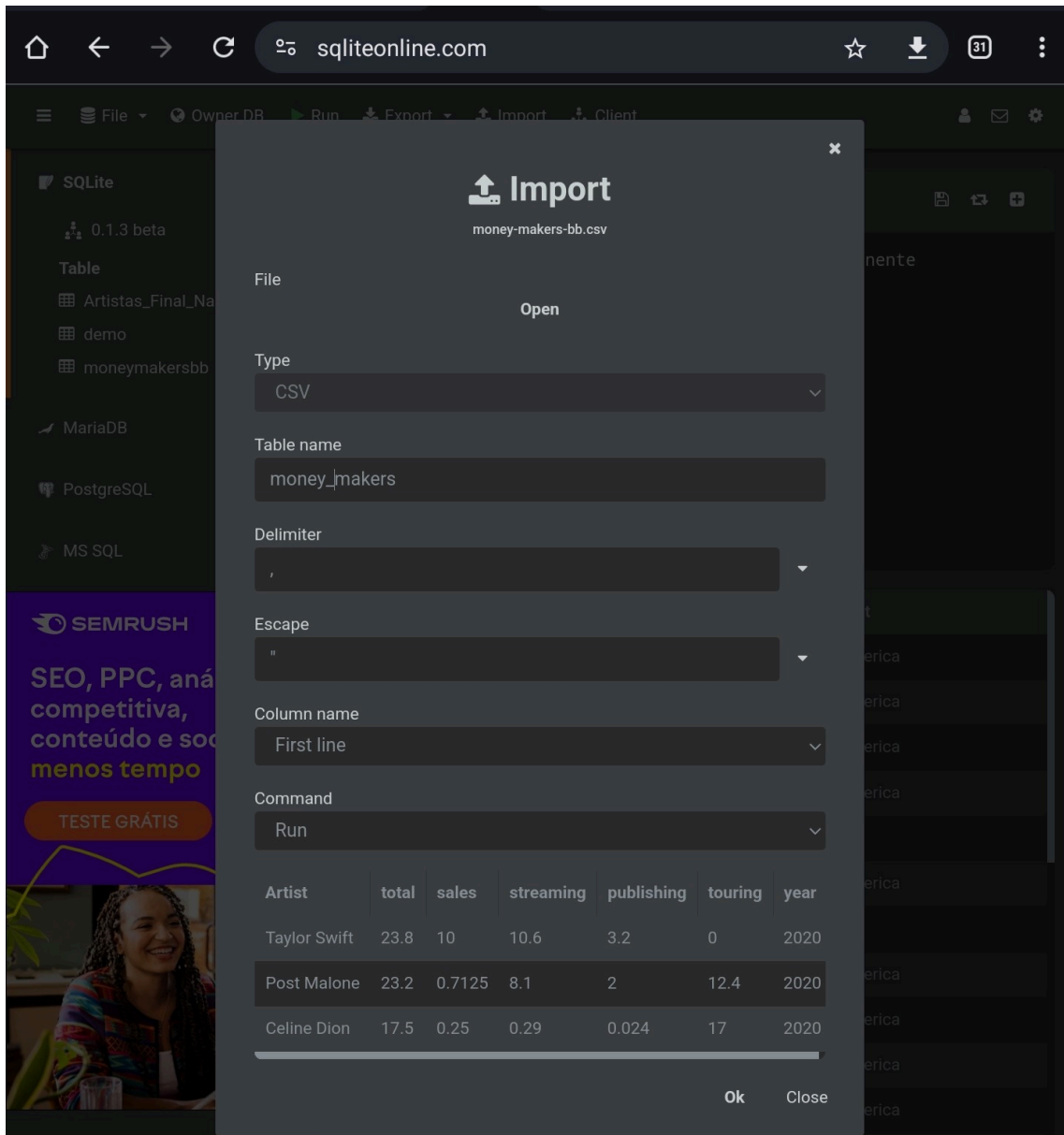
```
In [ ]: import pandas as pd
import sqlite3

# Criar uma conexão com um banco de dados SQLite
conn = sqlite3.connect('money_makers.db')

# Criar um objeto cursor
cursor = conn.cursor()

def executa_sql(conexao, comando_sql):
    try:
        cursor = conexao.cursor()
        cursor.execute(comando_sql)
        conexao.commit()
    except sqlite3.Error as erro:
        print("Erro ao inserir dados no banco de dados:", erro)
```

Usando a importação pelo SGBD 'SQLITEONLINE'



Usando a importação de dados pelo Python

```
In [ ]: # Nome do arquivo CSV Local
csv_file = 'money-makers-bb.csv'

# Lendo o arquivo CSV
df = pd.read_csv(csv_file)

## Remover a tabela se ela já existir
cursor.execute("DROP TABLE IF EXISTS money_makers")

## Criar a tabela com a estrutura apropriada com base no arquivo CSV
create_table_query = """
    CREATE TABLE money_makers (
        artist TEXT,
        total NUM,
        sales NUM,
        streaming NUM,
        publishing NUM,
        touring NUM,
        year INT
    )
"""
cursor.execute(create_table_query)

### Confirmar a transação
conn.commit()

dados = ""
def executa_sql(conexao, comando_sql):
    try:
        cursor = conexao.cursor()
        cursor.execute(comando_sql)
        conexao.commit()
    except sqlite3.Error as erro:
        print("Erro ao inserir dados no banco de dados:", erro)
for index, row in df.iterrows():
    comando = """
        INSERT INTO money_makers (Artist, Total, Sales, Streaming, Publishing, Touring, Year)
        VALUES (
            """+str(row[0])+"""+""",
            """+str(row[1])+""",
            """+str(row[2])+""",
            """+str(row[3])+""",
            """+str(row[4])+""",
            """+str(row[5])+""",
            """+str(row[6])+""",
        )
    """
    executa_sql(conn, comando)

validacao = """SELECT * FROM money_makers"""
print("<--Teste de verificação dos dados -->")
print(pd.read_sql_query(validacao, conn))
print("<--Final do teste de verificação-->")
```

```
<--Teste de verificação dos dados -->
      artist  total    sales  streaming  publishing  \
0      Taylor Swift  23.8  10.0000      10.60      3.200
1      Post Malone  23.2   0.7125       8.10      2.000
2      Celine Dion  17.5   0.2500       0.29      0.024
3      Eagles      16.3   0.9700       2.70      1.200
4      Bellie Eilish 14.7   2.1000       5.90      5.700
..      ...      ...      ...      ...      ...
115  YoungBoy Never Broke Again  10.4   0.1200       8.90      1.400
116      Thomas Rhett  10.3   0.1700       2.40      0.410
117      Zac Brown Band  10.3   0.2300       1.60      1.000
118      Maluma      10.2   0.0020       1.30      0.810
119      The Weeknd   10.1   1.3000       7.30      1.600

      touring  year
0           0.0  2020
1          12.4  2020
2          17.0  2020
3          11.4  2020
4           1.0  2020
..      ...      ...
115         0.0  2021
116         7.4  2021
117         7.4  2021
118         8.1  2021
119         0.0  2021
```

[120 rows x 7 columns]

<--Final do teste de verificação-->

Questão 1: 0,5 Ponto

- Quantos artistas constam na base?

```
In [ ]: comando = "Select distinct artist from money_makers"
        artistas = pd.read_sql_query(comando, conn)
        qtd = artistas.count() - 1
        print("Constam ", qtd["artist"], " artistas na base de dados, considerando variações no n
```

Constam 89 artistas na base de dados, considerando variações no nome de artistas ano a ano, como por exemplo, a banda que teve o nome `The Beatles` em 2020 e `Beatles` em 2021

- Quantos registros essa base possui?

```
In [ ]: comando = "Select * from money_makers"
        quantidade = pd.read_sql_query(comando, conn)
        valor = quantidade.count()
        print("Constam ", valor["total"], " registros na base de dados")
```

Constam 120 registros na base de dados

- Quantas colunas essa base possui?

```
In [ ]: comando = """SELECT *
        FROM PRAGMA_table_info('money_makers');"""
        valor = pd.read_sql_query(comando, conn)
        quantidade = valor.count()
        print("Existem ", quantidade["name"], " colunas na tabela")
```

Existem 7 colunas na tabela

Questão 2: 0,5 Ponto

- Qual artista teve o maior número de visualizações no total?

```
In [ ]: comando = """Select artist,
sum(streaming) qt
from money_makers
group by artist
order by qt desc"""
valor = pd.read_sql_query(comando, conn)
dados = valor[0:1]
mens = "O artista com a maior quantidade de visualizações foi"
quantidade = dados['qt'][0]
artista = dados['artist'][0]
print(mens, artista, 'com', quantidade , 'mil visualizações')
```

O artista com a maior quantidade de visualizações foi Drake com 49.3 mil visualizações

- Qual artista teve menos visualizações no streaming?

```
In [ ]: comando = """
Select artist,
sum(streaming) qt
from money_makers
group by artist
order by qt asc
"""
valor = pd.read_sql_query(comando, conn)
dados = valor[0:1]
mens = "O artista com a menor quantidade de visualizações foi"
quantidade = dados['qt'][0]
artista = dados['artist'][0]
print(mens, artista, 'com', quantidade , 'mil visualizações')
```

O artista com a menor quantidade de visualizações foi Jennifer Lopez com 0.08 mil visualizações

- Qual artista teve mais visualizações em turnês?

```
In [ ]: comando = """
SELECT artist ,
sum(touring) qt
FROM money_makers
GROUP BY artist
order by qt desc
"""
valor = pd.read_sql_query(comando, conn)
dados = valor[0:1]
mens = "O artista com a maior quantidade de visualizações em turnê foi"
quantidade = dados['qt'][0]
artista = dados['artist'][0]
print(mens, artista, 'com', quantidade , 'mil visualizações em turnês')
```

O artista com a maior quantidade de visualizações em turnê foi Taylor Swift com 90.5 mil visualizações em turnês

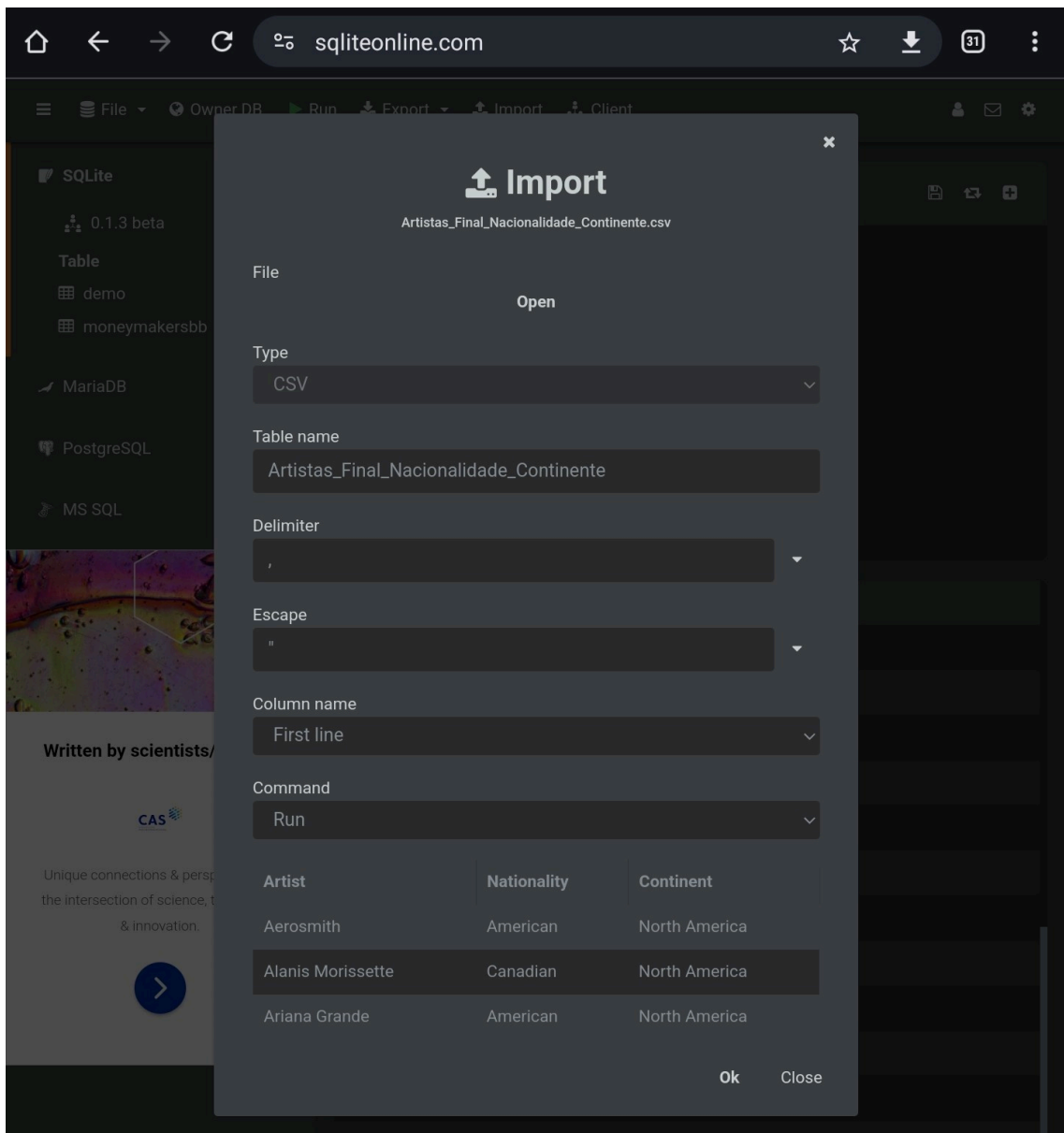
Questão 3: 0,5 Ponto

Crie uma tabela contendo

- nome do artista,
- país e,
- continente de origem,

os artistas listados nessa tabela deverão ser os mesmos do arquivo money makers bb

Usando a importação pelo SGBD 'SQLITEONLINE'



Usando a importação de dados pelo Python

```
In [ ]: # Nome do arquivo CSV com dados dos artistas, nacionalidade e continente
csv_file = 'artist_country_continent.csv'

# Lendo o arquivo CSV
df = pd.read_csv(csv_file)

## Remover a tabela se ela já existir
executa_sql(conn, "DROP TABLE IF EXISTS artist_country_continent")

## Criar a tabela com a estrutura apropriada com base no arquivo CSV
create_table_query = """
    CREATE TABLE artist_country_continent (
        artist TEXT,
        country TEXT,
        continent TEXT
    )
"""
executa_sql(conn, create_table_query)
```

```

for index, row in df.iterrows():
    comando = """
        INSERT INTO artist_country_continent (
            artist,
            country,
            continent
        )
        VALUES (
            """+str(row[0])+"""+""",
            """+str(row[1])+"""+""",
            """+str(row[2])+"""+""",
        )
    """
    executa_sql(conn, comando)

print("Criada a tabela 'artist_country_continent'")
print("Abaixo a lista para verificação dos dados")
print(pd.read_sql_query("SELECT * FROM artist_country_continent", conn))

```

Criada a tabela 'artist_country_continent'

Abaixo a lista para verificação dos dados

	artist	country	continent
0	AC DC	American	North America
1	Aerosmith	American	North America
2	Alanis Morissette	Canadian	North America
3	Ariana Grande	American	North America
4	Aventura	Dominican-American	North America
..
84	Weezer	American	North America
85	XXXTentacion	American	North America
86	YoungBoy Never Broke Again	American	North America
87	Zac Brown Band	American	North America
88	Dead & Company	Mexican	North America

[89 rows x 3 columns]

- Utilizando as informações das duas tabelas, liste quais foram os cinco países que tiveram o maior número de visualizações.

```

In [ ]: comando = """
        SELECT t1.country país,
            SUM(T2.streaming) visualizacoes
        FROM artist_country_continent T1
        JOIN money_makers T2
            ON( T1.artist = T2.artist )
        GROUP BY country
        order by visualizacoes DESC
        limit 5
    """

print("5 Países com a maior quantidade de visualizações")
print(pd.read_sql_query(comando, conn))

```

5 Países com a maior quantidade de visualizações

	país	visualizacoes
0	American	324.874
1	British	101.570
2	Canadian	24.800
3	Mexican	8.480
4	South Korean	7.300

Questão 4: 0,5 Ponto

- Utilizando python, conecte ao banco de dados e faça uma consulta para trazer qual artista teve maior faturamento em 2020

```
In [ ]: comando = """
        select artist, sum(sales) vendas, year
        from money_makers
        WHERE year = 2020
        group by artist, year
        order BY vendas desc
        limit 1
        """

        resultado = pd.read_sql_query(comando, conn)
        print("A artista com o maior faturamento em 2020 foi ", resultado['artist'][0], " com ", re
```

A artista com o maior faturamento em 2020 foi Taylor Swift com 10 milhões em vendas no período