In [ ]:
```
pip install spotipy
```

In [ ]:
```
pip install pandas
```

In [ ]:
```
pip install mysql-connector-python
```

In [ ]:
```
pip install mysql-connector-python sqlalchemy
```

In [ ]:
```

```

In [28]:
```
import mysql.connector
import re
from spotipy.oauth2 import SpotifyClientCredentials
import spotipy
```

In [29]:
```
# Set up Spotify API.
sp = spotipy.Spotify(auth_manager=SpotifyClientCredentials(
    client_id= 'YOUR_CLIENT_ID' ,
    client_secret=  'YOUR_CLIENT_SECRET'
))
```

In [30]:
```
# MySQL DataBase Connection.
db_config = {
    'host':'localhost',
    'user':'root',
    'password':'root',
    'database':'SPOTIFY'
}
```

In [31]:
```
# Connect to the Database.
connection = mysql.connector.connect(**db_config)
cursor = connection.cursor()
```

In [32]:
```
# Read Track URL from file.
file_path = 'My_Tracks.txt'
with open(file_path, 'r') as file:
    my_track_url = file.readlines()
```

In [33]:
```
# Process each URL
for track_url in my_track_url:
    track_url = track_url.strip()

    try:
        # Extract Track ID
        match = re.search(r"track/([a-zA-Z0-9]+)", track_url)
        if not match:
            print(f"Invalid URL: {track_url}")
            continue

        track_id = match.group(1)

        # Fetch track data
        track = sp.track(track_id)

        # -------- Artist Data --------
        artist = track["artists"][0]
        artist_id = artist["id"]
        artist_info = sp.artist(artist_id)

        cursor.execute("""
            INSERT IGNORE INTO artists (artist_id, artist_name, popularity, followers, genres)
            VALUES (%s, %s, %s, %s, %s)
        """, (
            artist_id,
            artist_info["name"],
            artist_info["popularity"],
            artist_info["followers"]["total"],
            ", ".join(artist_info["genres"])
        ))

        # -------- Album Data --------
        album = track["album"]
        album_id = album["id"]
```

```
        cursor.execute("""
            INSERT IGNORE INTO albums (album_id, album_name, artist_id, release_date, total_tracks, album_type)
            VALUES (%s, %s, %s, %s, %s, %s)
        """, (
            album_id,
            album["name"],
            artist_id,
            album["release_date"],
            album["total_tracks"],
            album["album_type"]
        ))

        # -------- Track Data --------
        cursor.execute("""
            INSERT IGNORE INTO tracks (
                track_id, track_name, artist_id, album_id,
                duration_sec, explicit, popularity
            )
            VALUES (%s, %s, %s, %s, %s, %s, %s)
        """, (
            track["id"],
            track["name"],
            artist_id,
            album_id,
            track["duration_ms"] / 60000,
            track["explicit"],
            track["popularity"]
        ))

        connection.commit()
        print(f"Inserted: {track['name']}")

    except Exception as e:
        print(f"Error processing URL: {track_url}")
        print(e)

# Close DB connection
cursor.close()
connection.close()
```

```
print("All tracks have been processed and inserted successfully.")
```

```
Inserted: Oru Manam (From "Dhruva Natchathiram")
Inserted: APT.
Inserted: Happier
Inserted: Left and Right (Feat. Jung Kook of BTS)
Inserted: Glad You Exist
Inserted: Attention
Inserted: Something Just Like This
Inserted: Leave Before You Love Me (with Jonas Brothers)
Inserted: Night Changes
Inserted: We Don't Talk Anymore (feat. Selena Gomez)
All tracks have been processed and inserted successfully.
```

In [ ]:

In [ ]:

In [1]:
```
import pandas as pd
from sqlalchemy import create_engine

# Create engine
engine = create_engine(
    "mysql+mysqlconnector://root:root@localhost/SPOTIFY"
)
```

In [2]:
```
# Query
query_artists = "SELECT * FROM artists"

# Read data
df_artists = pd.read_sql(query_artists, engine)
df_artists.to_csv("artists.csv", index=False)
df_artists
```

Out[2]:

	artist_id	artist_name	popularity	followers	genres
0	29aw5YCdIw2FEXYyAJZI8l	Harris Jayaraj	76	10803813	tamil pop, kollywood, tamil dance
1	3eVa5w3URK5duf6eyVDbu9	ROSÉ	81	13831488	k-pop
2	4AK6F7OLvEQ5QYCBNiQWHq	One Direction	85	40527276	
3	64KEffDW9EtZ1y2vBYgq8T	Marshmello	82	34597914	edm
4	69GGBxA162lTqCwzJG5jLp	The Chainsmokers	82	21502646	
5	6VuMaDnrHyPL1p4EHjYLi7	Charlie Puth	82	25619568	soft pop, pop
6	7z5WFjZAIYejWy0NI5lv4T	Dan + Shay	73	3214958	country

In [3]:
```
# Query
query_albums = "SELECT * FROM albums"

# Read data
df_albums = pd.read_sql(query_albums, engine)
df_albums.to_csv("albums.csv", index=False)

df_albums
```

Out[3]:

	album_id	album_name	artist_id	release_date	total_tracks	album_type
0	0mZIUXje90JtHxPNzWsJNR	Voicenotes	6VuMaDnrHyPL1p4EHjYLi7	2018-05-11	13	album
1	2IYQwwgxgOln7t3iF6ufFD	APT.	3eVa5w3URK5duf6eyVDbu9	2024-10-18	1	single
2	2u4Yp2ADTKYPwFSBFL4ffa	Happier	64KEffDW9EtZ1y2vBYgq8T	2018-08-17	1	single
3	4gCNyS7pidfK3rKWhB3JOY	FOUR (Deluxe)	4AK6F7OLvEQ5QYCBNiQWHq	2014-11-17	16	album
4	4JPguzRps3kuWDD5GS6oXr	Memories...Do Not Open	69GGBxA162lTqCwzJG5jLp	2017-04-07	12	album
5	4z2eVLzV0UxLc2O4VZMIjV	Glad You Exist	7z5WFjZAIYejWy0NI5lv4T	2021-02-05	1	single
6	5J4SS8wTmXdyIEVYjmHzpZ	Nine Track Mind (Special Edition)	6VuMaDnrHyPL1p4EHjYLi7	2015-11-05	20	album
7	5Jk4Eg7pxYhDrWJCVVzmMt	CHARLIE	6VuMaDnrHyPL1p4EHjYLi7	2022-10-06	12	album
8	66JuK41D3LpkbX3HCTGcQk	Leave Before You Love Me	64KEffDW9EtZ1y2vBYgq8T	2021-05-21	1	single
9	69MQlshE2Cveetu5T2b2JT	Oru Manam (From "Dhruva Natchathiram")	29aw5YCdIw2FEXYyAJZI8l	2020-10-09	1	single

In [4]:
```
# Query
query_tracks = "SELECT * FROM tracks"

# Read data
df_tracks = pd.read_sql(query_tracks, engine)
df_tracks.to_csv("tracks.csv", index=False)
df_tracks
```

Out[4]:

	track_id	track_name	artist_id	album_id	duration_sec	explicit	popularity
0	472vIK1ldetTxRxG3ovaiY	Glad You Exist	7z5WFjZAIYejWy0NI5lv4T	4z2eVLzV0UxLc2O4VZMIjV	2.408883	0	54
1	4qu63nuBpdn0qHUHuObEj1	Leave Before You Love Me (with Jonas Brothers)	64KEffDW9EtZ1y2vBYgq8T	66JuK41D3LpkbX3HCTGcQk	2.583050	0	82
2	54PbBpquVfhfrwRwvjSXbI	We Don't Talk Anymore (feat. Selena Gomez)	6VuMaDnrHyPL1p4EHjYLi7	5J4SS8wTmXdyIEVYjmHzpZ	3.628433	0	83
3	55mEN7iiqTNzWfYjF8QbQn	Oru Manam (From "Dhruva Natchathiram")	29aw5YCdIw2FEXYyAJZI8l	69MQlshE2Cveetu5T2b2JT	5.670900	0	56
4	5cF0dROlMOK5uNZtivgu50	Attention	6VuMaDnrHyPL1p4EHjYLi7	0mZIUXje90JtHxPNzWsJNR	3.479767	0	88
5	5O2P9iiztwhomNh8xkR9lJ	Night Changes	4AK6F7OLvEQ5QYCBNiQWHq	4gCNyS7pidfK3rKWhB3JOY	3.776617	0	88
6	5Odq8ohlgIbQKMZivbWkEo	Left and Right (Feat. Jung Kook of BTS)	6VuMaDnrHyPL1p4EHjYLi7	5Jk4Eg7pxYhDrWJCVVzmMt	2.574767	0	74
7	5vNRhkKd0yEAg8suGBpjeY	APT.	3eVa5w3URK5duf6eyVDbu9	2IYQwwgxgOIn7t3iF6ufFD	2.831950	0	88
8	6RUKPb4LETWmmr3iAEQktW	Something Just Like This	69GGBxA162lTqCwzJG5jLp	4JPguzRps3kuWDD5GS6oXr	4.119333	0	87
9	7BqHUALzNBTanL6OvsqmC1	Happier	64KEffDW9EtZ1y2vBYgq8T	2u4Yp2ADTKYPwFSBFL4ffa	3.571483	0	82

In [ ]:

In [ ]:

1. Artist-Level Analytics (df_artists)

Top Artists by Popularity

In [5]:
```
df_artists[['artist_name','popularity']] \
    .sort_values(by='popularity', ascending=False)
```

Out[5]:

	artist_name	popularity
2	One Direction	85
4	The Chainsmokers	82
3	Marshmello	82
5	Charlie Puth	82
1	ROSÉ	81
0	Harris Jayaraj	76
6	Dan + Shay	73

Top Artists by Followers

In [6]:
```
df_artists[['artist_name','followers']] \
    .sort_values(by='followers', ascending=False)
```

Out[6]:

	artist_name	followers
2	One Direction	40527276
3	Marshmello	34597914
5	Charlie Puth	25619568
4	The Chainsmokers	21502646
1	ROSÉ	13831488
0	Harris Jayaraj	10803813
6	Dan + Shay	3214958

Genre Distribution

In [7]:
```
df_artists['genres'] \
    .str.split(', ') \
    .explode() \
    .value_counts()
```

Out[7]:
```
genres
               2
tamil pop      1
kollywood      1
tamil dance    1
k-pop          1
edm            1
soft pop       1
pop            1
country        1
Name: count, dtype: int64
```

2. Album-Level Analytics (df_albums)

Album Type Distribution

In [8]:
```
df_albums['album_type'].value_counts()
```

Out[8]:
```
album_type
album     5
single    5
Name: count, dtype: int64
```

Albums Released Per Year

In [9]:
```
df_albums['release_year'] = pd.to_datetime(df_albums['release_date']).dt.year
df_albums['release_year'].value_counts().sort_index()
```

Out[9]:
```
release_year
2014    1
2015    1
2017    1
2018    2
2020    1
2021    2
2022    1
2024    1
Name: count, dtype: int64
```

Albums per Artist

In [21]:
```
df_albums.merge(df_artists, on='artist_id').groupby('artist_name') \
    .size().sort_values(ascending=False)
```

Out[21]:
```
artist_name
Charlie Puth        3
Marshmello          2
Dan + Shay          1
Harris Jayaraj      1
One Direction       1
ROSÉ                1
The Chainsmokers    1
dtype: int64
```

3. Track-Level Analytics (df_tracks)

Top Tracks by Popularity

In [11]:
```
df_tracks[['track_name','popularity']] \
    .sort_values(by='popularity', ascending=False)
```

Out[11]:

	track_name	popularity
4	Attention	88
5	Night Changes	88
7	APT.	88
8	Something Just Like This	87
2	We Don't Talk Anymore (feat. Selena Gomez)	83
1	Leave Before You Love Me (with Jonas Brothers)	82
9	Happier	82
6	Left and Right (Feat. Jung Kook of BTS)	74
3	Oru Manam (From "Dhruva Natchathiram")	56
0	Glad You Exist	54

Average Track Duration

In [12]:
```
df_tracks['duration_sec'].mean()
```

Out[12]:
```
np.float64(3.4645183333333334)
```

4. Cross-Table Analytics

Merge All Data

In [15]:
```
df_full = (
    df_tracks
    .merge(df_artists, on='artist_id')
    .merge(df_albums, on='album_id')
)
```

Most Popular Artist (Avg Track Popularity)

In [16]:
```
df_full.groupby('artist_name')['popularity_x'].mean().sort_values(ascending=False)
```

Out[16]:
```
artist_name
One Direction       88.000000
ROSÉ                88.000000
The Chainsmokers    87.000000
Marshmello          82.000000
Charlie Puth        81.666667
Harris Jayaraj      56.000000
Dan + Shay          54.000000
Name: popularity_x, dtype: float64
```

Album Performance (Avg Track Popularity)

In [17]:
```
df_full.groupby('album_name')['popularity_x'].mean().sort_values(ascending=False)
```

Out[17]:
```
album_name
APT.                                      88.0
FOUR (Deluxe)                             88.0
Voicenotes                                88.0
Memories...Do Not Open                    87.0
Nine Track Mind (Special Edition)         83.0
Happier                                   82.0
Leave Before You Love Me                  82.0
CHARLIE                                   74.0
Oru Manam (From "Dhruva Natchathiram")    56.0
Glad You Exist                            54.0
Name: popularity_x, dtype: float64
```

Genre vs Track Popularity

In [18]:
```
df_full['genre'] = df_full['genres'].str.split(', ')
df_full.explode('genre').groupby('genre')['popularity_x'].mean().sort_values(ascending=False)
```

Out[18]:
```
genre
k-pop          88.000000
               87.500000
edm            82.000000
soft pop       81.666667
pop            81.666667
tamil dance    56.000000
kollywood      56.000000
tamil pop      56.000000
country        54.000000
Name: popularity_x, dtype: float64
```

5. Business Insights

Do Followers Impact Track Popularity?

In [19]:
```
df_full[['followers','popularity_x']].corr()
```

Out[19]:

	followers	popularity_x
followers	1.000000	0.687147
popularity_x	0.687147	1.000000

Singles vs Albums Performance

In [20]:
```
df_full.groupby('album_type')['popularity_x'].mean()
```

Out[20]:
```
album_type
album     84.0
single    72.4
Name: popularity_x, dtype: float64
```

In [ ]:
```

```