

Spotify Project: SQL Queries through Python

```
In [ ]: # Import the required Libraries
import pymysql
import warnings
warnings.filterwarnings("ignore") # Ignore the deprecated warning
from sqlalchemy import create_engine, text
import getpass # To get the password without showing the input
password = getpass.getpass()
```

Lets connect to our SQL database:

```
In [ ]: # Database name
database_name = 'spotify_project'

# Define the server connection string
server_connection_string = f'mysql+pymysql://root:{password}@localhost/{database_name}'
```

```
In [ ]: # Create an engine and connect to the MySQL server
engine = create_engine(server_connection_string)
```

```
In [ ]: # List all tables in the database
query = "SHOW TABLES;"
tables = pd.read_sql(query, engine)
print(tables)
```

```
Tables_in_spotify_project
0      artist_genres
1          artists
2          tracks
```

Lets run some queries!

Top 10 Artists by Number of Tracks in the Top 50 Charts

```
In [ ]: query = """
SELECT artist_name, COUNT(*) AS num_tracks
FROM tracks
GROUP BY artist_name
ORDER BY num_tracks DESC
LIMIT 10;
"""

top_artists_by_tracks = pd.read_sql(query, engine)
top_artists_by_tracks
```

```
Out[ ]:   artist_name  num_tracks
0          Feid        72
1      KAROL G        65
2    Myke Towers        33
3     FloyyMenor        28
4  Grupo Frontera        24
5         Jimin        20
6       Trueno        20
7      Blessd         19
8  Rauw Alejandro        19
9    Bad Bunny         19
```

Top 10 Tracks by Popularity

```
In [ ]: query = """
SELECT DISTINCT track_name, artist_name, popularity
FROM tracks
ORDER BY popularity DESC
LIMIT 10;
"""

top_tracks_by_popularity = pd.read_sql(query, engine)
top_tracks_by_popularity
```

Out[]:

	track_name	artist_name	popularity
0	BIRDS OF A FEATHER	Billie Eilish	99
1	Espresso	Sabrina Carpenter	99
2	Please Please Please	Sabrina Carpenter	98
3	Gata Only	FloyyMenor	96
4	Si Antes Te Hubiera Conocido	KAROL G	94
5	REAL GANGSTA LOVE	Trueno	92
6	LUNA	Feid	91
7	One Of The Girls (with JENNIE, Lily Rose Depp)	The Weeknd	91
8	ADIVINO	Myke Towers	90
9	FRI(END)S	V	90

Average Popularity of Tracks by Country

In []:

```
query = """
SELECT country, AVG(popularity) AS avg_popularity
FROM tracks
GROUP BY country
ORDER BY avg_popularity DESC;
"""

avg_popularity_by_country = pd.read_sql(query, engine)
avg_popularity_by_country
```

Out[]:

	country	avg_popularity
0	Guatemala	84.14
1	Mexico	83.54
2	Costa Rica	83.52
3	Ecuador	83.46
4	El Salvador	83.44
5	Honduras	83.28
6	Peru	83.16
7	Nicaragua	83.02
8	Paraguay	82.44
9	Bolivia	81.54
10	Colombia	80.44
11	Spain	78.94
12	Venezuela	78.86
13	Argentina	77.66
14	Chile	77.32
15	Panama	75.64
16	Uruguay	73.98
17	Dominican Republic	72.00

Genres with the Most Tracks in the Top 50 Charts

In []:

```
query = """
SELECT genre, COUNT(*) AS num_tracks
FROM artist_genres
WHERE genre <> ''
```

```
GROUP BY genre
ORDER BY num_tracks DESC
LIMIT 10;
"""
top_genres = pd.read_sql(query, engine)
top_genres
```

Out[]:

	genre	num_tracks
0	urbano latino	56
1	trap latino	43
2	reggaeton	30
3	latin pop	15
4	sierreno	10
5	corrido	10
6	trap argentino	10
7	sad sierreno	9
8	musica mexicana	8
9	urbano chileno	8

Top 10 Artists by Followers

In []:

```
query = """
SELECT artist_name, followers
FROM artists
ORDER BY followers DESC
LIMIT 10;
"""
top_artists_by_followers = pd.read_sql(query, engine)
top_artists_by_followers
```

Out[]:

	artist_name	followers
0	Billie Eilish	96266004
1	The Weeknd	86456471
2	Bad Bunny	83750445
3	KAROL G	48499948
4	Ozuna	36856823
5	J Balvin	36703992
6	Anuel AA	36474013
7	Daddy Yankee	34160123
8	Maluma	33999148
9	Marshmello	33709079

Top 10 Tracks with the Longest Duration

In []:

```
query = """
SELECT DISTINCT track_name, artist_name, duration_ms / 1000 AS duration_secs
FROM tracks
ORDER BY duration_secs DESC
LIMIT 10;
"""

longest_tracks = pd.read_sql(query, engine)
longest_tracks
```

Out[]:

	track_name	artist_name	duration_secs
0	WYA REMIX BLACK AND YELLOW (feat. iZaak & Pirlo)	J Abdiel	390.545
1	Potra Salvaje - Hard Remix	Isabel Aaiún	342.000
2	WYA REMIX RED	J Abdiel	321.741
3	Tiroteo - Remix	Marc Seguí	321.555
4	Whisky	Samuel G	305.265
5	Enrolar - Fantastic Four Remix	Kidd Voodoo	303.582
6	Besitos	BARBEL	301.276
7	La Niña Consentida	BARBEL	299.650
8	HOLLYWOOD	Peso Pluma	284.864
9	BESAME (feat. Tiago PZK, Khea & Neo Pista) - ...	Bhavi	284.308

Top 10 Tracks with the Shortest Duration

In []:

```
query = """
SELECT DISTINCT track_name, artist_name, duration_ms / 1000 AS duration_secs
FROM tracks
ORDER BY duration_secs ASC
LIMIT 10;
"""

shortest_tracks = pd.read_sql(query, engine)
shortest_tracks
```

Out[]:

	track_name	artist_name	duration_secs
0	HAY LUPITA	Lomiiel	105.836
1	A 30	Jezzy	115.432
2	3 PECADOS DESPUES...	Milo j	118.558
3	PIKY	Jezzy	118.998
4	IMAN (Two Of Us)	Maria Becerra	122.405
5	Una Foto	Mesita	122.500
6	Ellos Vienen Ellos Van - Remix	La Perversa	124.028
7	Hoy Si	La Perversa	124.285
8	SAN TURRONA	La Joaqui	124.689
9	Ante	NTG	125.087

Count of Explicit Tracks by Country

In []:

```
query = """
SELECT country, COUNT(*) AS explicit_tracks
FROM tracks
WHERE explicit = TRUE
GROUP BY country
ORDER BY explicit_tracks DESC;
"""

explicit_tracks_by_country = pd.read_sql(query, engine)
explicit_tracks_by_country
```

Out[]:

	country	explicit_tracks
0	Chile	34
1	Dominican Republic	32
2	Panama	24
3	Colombia	23
4	Mexico	23
5	Peru	22
6	Venezuela	21
7	Ecuador	21
8	Costa Rica	21
9	El Salvador	21
10	Spain	21
11	Honduras	20
12	Paraguay	20
13	Guatemala	18
14	Bolivia	17
15	Nicaragua	15
16	Uruguay	15
17	Argentina	14

Artists with the Most Genres

In []:

```
query = """
SELECT a.artist_name, COUNT(ag.genre) AS num_genres
FROM artist_genres ag
JOIN artists a ON ag.artist_id = a.artist_id
```

```
GROUP BY a.artist_name
ORDER BY num_genres DESC
LIMIT 10;
"""
artists_with_most_genres = pd.read_sql(query, engine)
artists_with_most_genres
```

Out[]:

	artist_name	num_genres
0	Los Enanitos Verdes	7
1	Aventura	6
2	Becky G	6
3	Manuel Turizo	6
4	Feid	6
5	Sebastian Yatra	5
6	Julieta Venegas	5
7	Maluma	5
8	Luis R Conriquez	5
9	Natanael Cano	5

Top 10 Most Popular Tracks in Each Country

In []:

```
query = """
SELECT country, track_name, artist_name, popularity
FROM tracks t1
WHERE (
    SELECT COUNT(*)
    FROM tracks t2
    WHERE t2.country = t1.country
    AND t2.popularity > t1.popularity
) < 10
ORDER BY country, popularity DESC;
"""
```

```
top_tracks_per_country = pd.read_sql(query, engine)
top_tracks_per_country
```

Out[]:

	country	track_name	artist_name	popularity
0	Argentina	Gata Only	FloyyMenor	96
1	Argentina	Si Antes Te Hubiera Conocido	KAROL G	94
2	Argentina	REAL GANGSTA LOVE	Trueno	92
3	Argentina	LUNA	Feid	91
4	Argentina	SI NO ES CONTIGO	Cris Mj	89
...
218	Venezuela	Mirame	Blessd	87
219	Venezuela	Who	Jimin	87
220	Venezuela	AMARGURA	KAROL G	87
221	Venezuela	QLONA	KAROL G	87
222	Venezuela	PERRO NEGRO	Bad Bunny	87

223 rows × 4 columns

Average Duration of Tracks by Country

In []:

```
query = """
SELECT country, AVG(duration_ms / 1000) AS avg_duration_secs
FROM tracks
GROUP BY country
ORDER BY avg_duration_secs DESC;
"""

avg_duration_by_country = pd.read_sql(query, engine)
avg_duration_by_country
```

Out[]:

	country	avg_duration_secs
0	Peru	195.95638
1	Spain	195.26864
2	Panama	195.01994
3	Honduras	193.80934
4	El Salvador	193.61708
5	Nicaragua	193.01260
6	Colombia	192.57694
7	Mexico	190.85126
8	Venezuela	190.54982
9	Ecuador	190.51888
10	Costa Rica	188.42440
11	Paraguay	188.03746
12	Chile	186.69068
13	Guatemala	184.57216
14	Bolivia	181.49732
15	Argentina	179.08752
16	Uruguay	178.50222
17	Dominican Republic	172.72420

Top 10 Longest Tracks by Country

In []:

```
query = """
SELECT country, track_name, artist_name, duration_ms / 1000 AS duration_secs
FROM tracks t1
WHERE (
```

```

SELECT COUNT(*)
FROM tracks t2
WHERE t2.country = t1.country
AND t2.duration_ms > t1.duration_ms
) < 10
ORDER BY country, duration_secs DESC;
"""

longest_tracks_per_country = pd.read_sql(query, engine)
longest_tracks_per_country

```

Out[]:

	country	track_name	artist_name	duration_secs
0	Argentina	WYA REMIX RED	J Abdiel	321.741
1	Argentina	BESAME (feat. Tiago PZK, Khea & Neo Pistea) - ...	Bhavi	284.308
2	Argentina	BBY BOO - REMIX	iZaak	275.862
3	Argentina	Una Foto Remix (feat. Emilia)	Mesita	243.898
4	Argentina	Gata Only	FloyyMenor	222.000
...
175	Venezuela	Orion	Boza	240.964
176	Venezuela	LA RANGER (feat. Myke Towers) The Academy: Segunda Misión	229.738	
177	Venezuela	CARACAS EN EL 2000	ELENA ROSE	227.784
178	Venezuela	Closer Than This	Jimin	223.301
179	Venezuela	Gata Only	FloyyMenor	222.000

180 rows × 4 columns

Top 10 Artists by Average Popularity

In []:

```

query = """
SELECT artist_name, AVG(popularity) AS avg_popularity
FROM tracks
GROUP BY artist_name
ORDER BY avg_popularity DESC

```

```
LIMIT 10;  
"""  
top_artists_by_avg_popularity = pd.read_sql(query, engine)  
top_artists_by_avg_popularity
```

Out[]:

	artist_name	avg_popularity
0	Billie Eilish	99.0000
1	Sabrina Carpenter	98.2000
2	The Weeknd	91.0000
3	FloyyMenor	90.9643
4	Trueno	90.3500
5	Sevdaliza	89.0000
6	V	88.5000
7	KAROL G	88.1385
8	Myke Towers	87.5758
9	Luis R Conriquez	86.5714

Top 10 Most Followed Artists by Genre

```
In [ ]: query = """  
SELECT g.genre, a.artist_name, a.followers  
FROM artist_genres g  
JOIN artists a ON g.artist_id = a.artist_id  
WHERE g.genre <> ''  
AND (g.artist_id, g.genre) IN (  
    SELECT ranked.artist_id, ranked.genre  
    FROM (  
        SELECT ag.artist_id, ag.genre,  
            ROW_NUMBER() OVER (PARTITION BY ag.genre ORDER BY ar.followers DESC) as rn  
        FROM artist_genres ag  
        JOIN artists ar ON ag.artist_id = ar.artist_id  
    ) AS ranked  
    WHERE ranked.rn = 1
```

```

)
ORDER BY g.genre;
"""

top_followed_artists_by_genre = pd.read_sql(query, engine)
top_followed_artists_by_genre

```

Out[]:

	genre	artist_name	followers
0	alternative r&b	Sevdaliza	588718
1	argentine hip hop	Bizarrap	15687314
2	argentine rock	Los Enanitos Verdes	4201238
3	art pop	Billie Eilish	96266004
4	bachata	Aventura	8186978
...
95	urbano espanol	Rels B	7398309
96	urbano latino	Bad Bunny	83750445
97	urbano mexicano	Kevin AMF	491168
98	venezuelan hip hop	Rawayana	696141
99	venezuelan rock	Rawayana	696141

100 rows × 3 columns

Average Duration of Tracks for Each Genre

In []:

```

query = """
SELECT ag.genre, AVG(t.duration_ms / 1000) AS avg_duration_secs
FROM tracks t
JOIN artist_genres ag ON t.artist_id = ag.artist_id
WHERE ag.genre <> ''
GROUP BY ag.genre
ORDER BY avg_duration_secs DESC;
"""

```

```
avg_duration_per_genre = pd.read_sql(query, engine)
avg_duration_per_genre
```

Out[]:

	genre	avg_duration_secs
0	spanish pop	279.283667
1	canadian contemporary r&b	244.684000
2	canadian pop	244.684000
3	drill chileno	237.900000
4	classic venezuelan pop	228.554000
...
95	r&b argentino	142.410000
96	pop peruano	141.115000
97	trap venezolano	135.000000
98	trap dominicano	128.913333
99	dembow dominicano	123.255458

100 rows × 2 columns

Total Number of Unique Artists per Country's Top 50 Chart

In []:

```
query = """
SELECT t.country, COUNT(DISTINCT t.artist_id) AS num_artists
FROM tracks t
GROUP BY t.country
ORDER BY num_artists DESC;
"""

tracks_and_artists_per_country = pd.read_sql(query, engine)
tracks_and_artists_per_country
```

Out[]:

	country	num_artists
0	Honduras	38
1	Nicaragua	37
2	Peru	37
3	Argentina	36
4	Costa Rica	36
5	Ecuador	36
6	Guatemala	35
7	Panama	34
8	Paraguay	34
9	Uruguay	34
10	Dominican Republic	33
11	El Salvador	33
12	Spain	33
13	Bolivia	32
14	Mexico	32
15	Venezuela	32
16	Colombia	31
17	Chile	29

Top 10 Artists with the Most Explicit Tracks

In []:

```
query = """
SELECT a.artist_name, COUNT(t.track_id) AS num_explicit_tracks
FROM tracks t
JOIN artists a ON t.artist_id = a.artist_id
```

```
WHERE t.explicit = 1
GROUP BY a.artist_name
ORDER BY num_explicit_tracks DESC
LIMIT 10;
"""

artists_with_most_explicit_tracks = pd.read_sql(query, engine)
artists_with_most_explicit_tracks
```

Out[]:

	artist_name	num_explicit_tracks
0	Feid	42
1	Myke Towers	33
2	KAROL G	32
3	FloyyMenor	28
4	Blessd	19
5	Rauw Alejandro	19
6	Bad Bunny	18
7	Trueno	17
8	Cris Mj	16
9	J Abdiel	12

Average Followers and Popularity of Artists by Genre

In []:

```
query = """
SELECT ag.genre, AVG(a.followers) AS avg_followers, AVG(a.popularity) AS avg_popularity
FROM artists a
JOIN artist_genres ag ON a.artist_id = ag.artist_id
WHERE ag.genre <> ''
GROUP BY ag.genre
ORDER BY avg_followers DESC;
"""

avg_followers_and_popularity_per_genre = pd.read_sql(query, engine)
avg_followers_and_popularity_per_genre
```

Out[]:

	genre	avg_followers	avg_popularity
0	art pop	96266004.0	94.0
1	canadian pop	86456471.0	93.0
2	canadian contemporary r&b	86456471.0	93.0
3	pop	52088442.8	89.2
4	progressive electro house	33709079.0	83.0
...
95	trap venezolano	41060.0	43.0
96	plena uruguaya	39441.0	49.5
97	raspe	26991.0	65.0
98	pop peruano	7931.0	44.0
99	pop boliviano	4687.0	38.0

100 rows × 3 columns

Number of Tracks Released per Year

In []:

```
query = """
SELECT YEAR(t.release_date) AS release_year, COUNT(*) AS num_tracks
FROM tracks t
JOIN artist_genres ag ON t.artist_id = ag.artist_id
WHERE ag.genre <> ''
GROUP BY release_year
ORDER BY release_year DESC;
"""

tracks_released_per_year = pd.read_sql(query, engine)
tracks_released_per_year
```

Out[]:

	release_year	num_tracks
0	2024	1109
1	2023	925
2	2022	168
3	2021	73
4	2020	22
5	2019	56
6	2018	30
7	2016	4
8	2015	40
9	2014	3
10	2013	7
11	2012	12
12	2011	2
13	2010	51
14	2009	14
15	2006	76
16	2005	4
17	2004	10
18	2003	15
19	2002	4
20	1996	3
21	1989	6

release_year	num_tracks
--------------	------------

22	1986	14
----	------	----

Most Popular Song per Release Year

```
In [ ]: query = """
SELECT DISTINCT t1.release_year, t1.track_name, t1.artist_name, t1.popularity, t1.release_date
FROM (
    SELECT
        track_id,
        track_name,
        artist_name,
        release_date,
        popularity,
        YEAR(release_date) AS release_year
    FROM tracks
) t1
JOIN (
    SELECT
        YEAR(release_date) AS release_year,
        MAX(popularity) AS max_popularity
    FROM tracks
    GROUP BY YEAR(release_date)
) t2
ON t1.release_year = t2.release_year AND t1.popularity = t2.max_popularity
ORDER BY t1.release_year DESC;
"""

most_popular_song_per_year = pd.read_sql(query, engine)
most_popular_song_per_year
```

Out[]:

	release_year	track_name	artist_name	popularity	release_date
0	2024	BIRDS OF A FEATHER	Billie Eilish	99	2024-05-17
1	2024	Espresso	Sabrina Carpenter	99	2024-04-12
2	2023	LUNA	Feid	91	2023-12-01
3	2023	One Of The Girls (with JENNIE, Lily Rose Depp)	The Weeknd	91	2023-06-23
4	2022	The Astronaut	JIN	85	2022-10-28
5	2021	Desesperados	Rauw Alejandro	83	2021-12-10
6	2020	La Curiosidad	Jay Wheeler	82	2020-06-12
7	2019	LA CANCIÓN	J Balvin	84	2019-06-28
8	2018	Amanece	Anuel AA	82	2018-12-14
9	2016	Voy Después	Amenazzy	66	2016-08-09
10	2015	Borro Cassette	Maluma	82	2015-10-30
11	2014	Candy	Plan B	71	2014-08-25
12	2013	Hace Mucho Tiempo	Arcángel	72	2013-11-19
13	2012	A las Nueve	No Te Va Gustar	71	2012-10-18
14	2012	Ese Maldito Momento	No Te Va Gustar	71	2012-10-18
15	2011	Dile a Tu Amiga	Dalmata	69	2011-12-13
16	2010	Es un Secreto	Plan B	84	2010-07-20
17	2009	Solos (feat. Plan B)	Tony Dize	76	2009-11-09
18	2006	Todo Cambió	Camila	80	2006-05-09
19	2006	Me Voy	Julieta Venegas	80	2006-05-25
20	2005	Dia de Enero	Shakira	81	2005-06-03
21	2004	Tu Príncipe	Daddy Yankee	77	2004-07-13

release_year		track_name	artist_name	popularity	release_date
22	2004	Tú de Qué Vas	Franco De Vita	77	2004-04-28
23	2003	Dile	Don Omar	82	2003-01-01
24	2002	Es Por Ti	Juanes	79	2002-01-01
25	1996	Llorarás	Dimension Latina	76	1996-01-17
26	1989	Ven Porque Te Necesito	Los Temerarios	76	1989-01-01
27	1986	La Muralla Verde	Los Enanitos Verdes	79	1986-11-20