

IronHack Data Analytics

WEEK 4 | SPOTIFY DATASET

Group 2 - Ricardo | Rebecca

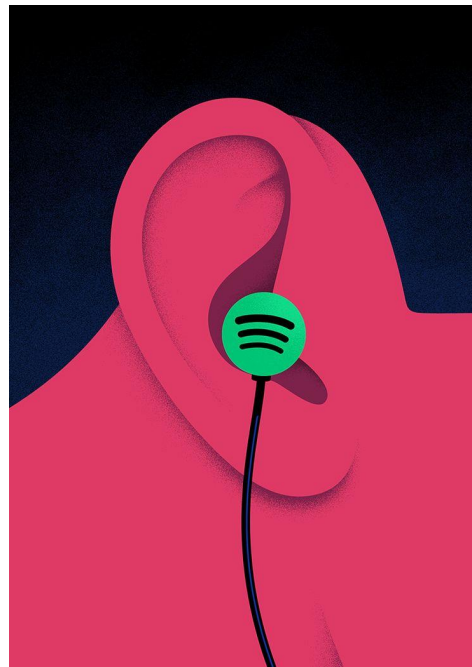
Agenda

- **Project Overview**
- **Business Opportunity**
- **Datasets Overview**
- **Database Design & Transformation**
- **SQL Insights**
- **Exploratory Data Analysis**
- **Major Obstacles & Lessons Learned**
- **Conclusion and Insights**

Project Overview

The project aims to analyze **Spotify streaming data** to identify **music trends across Europe** and examine these insights in relation to major **music festivals**.

By understanding **user preferences**, **genre popularity**, and **artist performance**, the study will provide valuable insights for festival organizers, music marketers, and artists.



Business Opportunity

Use Spotify data to analyze music trends and identify the **most popular artists and tracks** for an upcoming music festival in Europe.

Investigate music preferences regarding **danceability or energy level** across countries.

Analyse on the **economic impact** of music festivals to maximize audience engagement, ticket sales, and sponsorship opportunities.

Identify **genres that are trending but underrepresented** in existing 2024 festivals.

Assess audience preferences to determine the **ideal country for hosting the festival**.

Hypothesis: UK, Belgium, Germany are the ideal countries for hosting a new music festival.



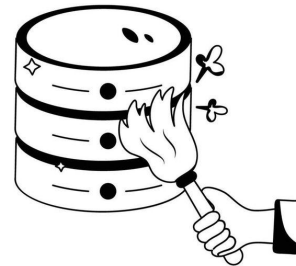
Top Spotify Songs Dataset

Top Spotify Songs in 73 Countries

- **Source:** [Kaggle](#)
- **Shape:** 1,674,294 rows × 25 columns

Top 50 songs for each country is updated daily to provide the most up-to-date information on the popularity of songs in the world.

spotify_id, name, artists, daily_rank, daily_movement, weekly_movement, country, snapshot_date, popularity, is_explicit, album_name, album_release_date, danceability, energy, loudness etc.



Data Wrangling and Cleaning

- **Standardized column names** for consistency
- **Removed null values** from key columns
- **Converted date** format & extracted month/year
- Filtered data from the last 3 months
- Selected only top 20 ranked records
- **Mapped country codes** to full names & continents
- *Dropped irrelevant columns* for analysis
- Renamed 'country_full' to 'country'
- Filtered records to include **only Europe**



```
rename(columns=...)
dropna(subset=...)
to_datetime(...)
apply(lambda x: ...)
df[df['snapshot_date'] >= '2024-09-01']
df[df['daily_rank'] <= 20]
map(lambda x: ...)
df[[...]]
df[df['continent'] == 'Europe']
```

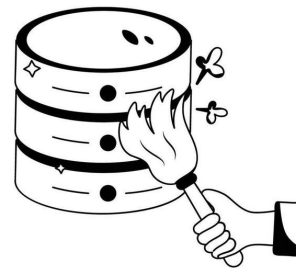
Music Festivals 2024 Dataset

European and International Music Festivals 2024

- Source: [Kaggle](#)
- 205 rows × 7 columns

Provides details about the festival's name, location, attendance, visitor demographics, economic impact, and the music genres featured.

Festival_Name, Location, Attendance_Numbers, Visitor_Demographics, Economic_Impact, Music_Genre



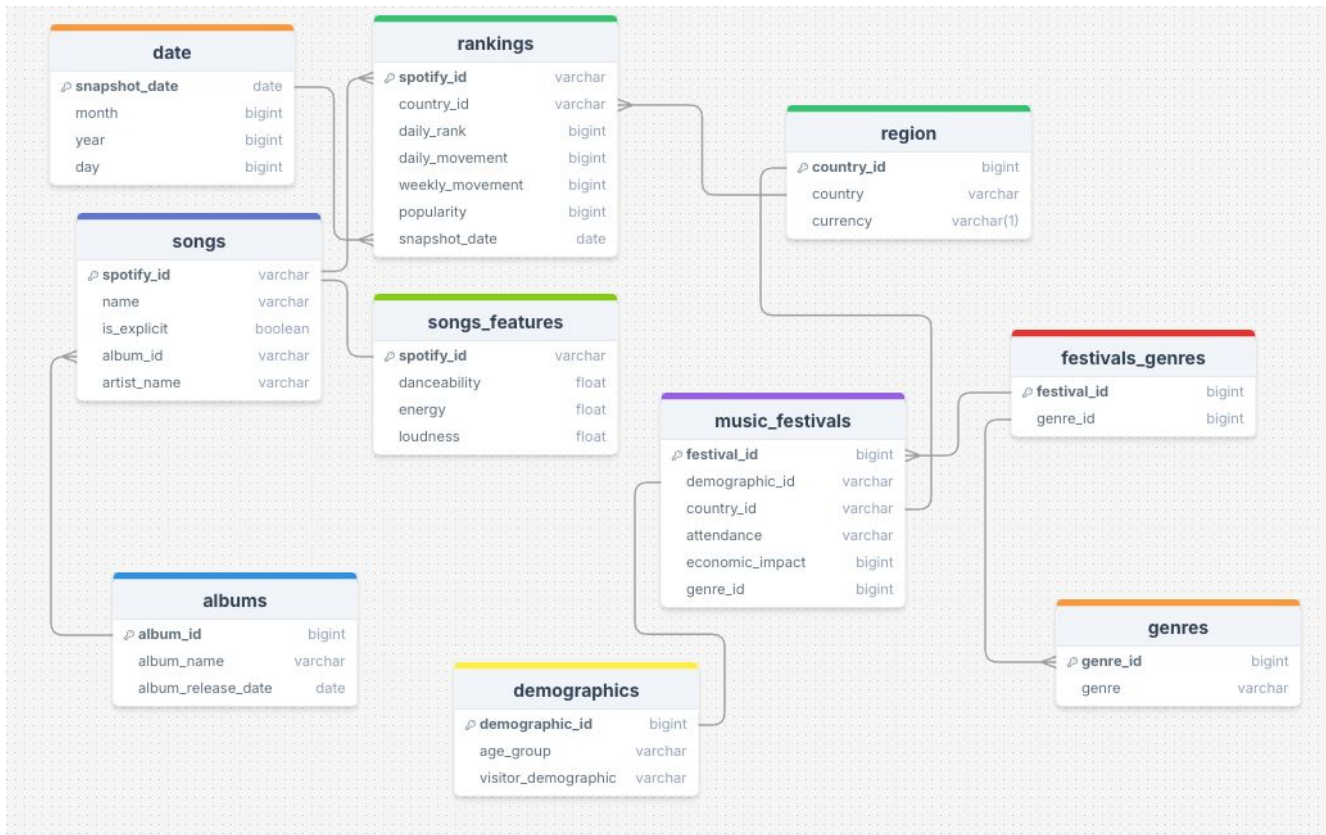
Data Wrangling and Cleaning

- Standardized column names
- Extracted **currency symbols** & cleaned 'economic_impact'
- Filtered records with (£, €) currency
- **Renamed key columns** for consistency
- Converted 'attendance' & 'economic_impact' to **numeric**
- Cleaned 'country' values & replaced 'UK' with 'United Kingdom'
- Assigned **unique festival IDs**
- Dropped irrelevant columns



```
rename(columns=...)
astype(str).str[0]
apply(lambda x: ...)
df[df['currency'].isin(['£', '€'])]
astype(float)
reset_index()
drop(columns=...)
replace('UK', 'United Kingdom')
```

Database Design - Entity Relationship Diagram



Combined Datasets

Data Wrangling and Cleaning

- Standardized column names
- Extracted **currency symbols** & cleaned 'economic_impact'
- Filtered records with (£, €) currency
- **Renamed key columns** for consistency
- Converted 'attendance' & 'economic_impact' to **numeric**
- Cleaned 'country' values & replaced 'UK' with 'United Kingdom'
- Assigned **unique festival IDs**
- Dropped irrelevant columns



Top Spotify Songs in 73 Countries

- **Source:** [Kaggle](#)
- **Shape:** 1,674,294 rows × 25 columns

European and International Music Festivals 2024

- **Source:** [Kaggle](#)
- **205 rows × 7 columns**

Analysis: Economic Impact of Music Festival Market in Europe

Economic impact | Genre | Number of Attendance share

- The highest **economic impact**
- 20 festivals, competitive landscape
- The UK market is already **saturated**
- Home to **Tomorrowland (EDM)**
- well-established, **limiting opportunities for new players**
- Popular genres like **Rock, Metal** align with Spotify streaming preferences
- **Established Festival Markets** that serve the audience



SQL Queries & Analysis

Most economically impactful festivals by country and currency:

- Tools used:
 - JOIN, SUBQUERY, SUM, GROUP BY, ORDER BY*

country	formatted_impact	currency
United Kingdom	561,000,000	£
Belgium	365,000,000	€
Germany	275,000,000	€
Spain	274,000,000	€
Hungary	157,000,000	€
Netherlands	55,000,000	€
France	50,000,000	€
Serbia	50,000,000	€
Croatia	37,000,000	€
Portugal	30,000,000	€

Top 5 most economically impactful festivals by country and genre:

- Tools used:
 - JOIN, SUBQUERY, SUM, GROUP BY, ORDER BY*

country	genre	total_economic_imp...
United Kingdom	Various	172,000,000.00
Belgium	EDM	150,000,000.00
Hungary	Various	130,000,000.00
Spain	Indie/Electronic	88,000,000.00
Belgium	Rock	85,000,000.00

SQL Queries & Analysis

Top most economically impactful **genres**:

- Tools used:
 - JOIN, SUBQUERY, SUM, GROUP BY, ORDER BY, CASE, LIMIT, WHERE*

genre	country	currency	total_impact	total_attendan...	number_of_festivals
Various	United Kingdom	£	172,000,000.00	335,000.00	4
EDM	Belgium	€	150,000,000.00	400,000.00	1
Various	Hungary	€	130,000,000.00	500,000.00	1
Indie/Electronic	Spain	€	88,000,000.00	235,000.00	2
Rock	Belgium	€	85,000,000.00	150,000.00	1

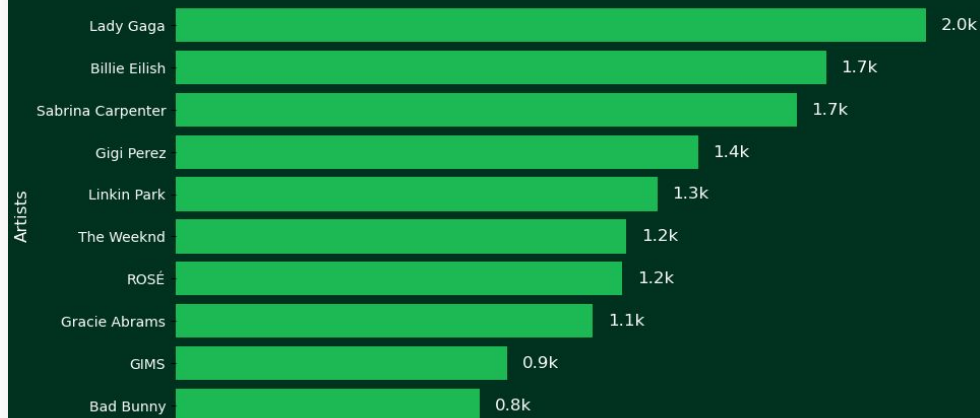
genre	country	currency	total_impact	total_attendan...	number_of_festivals
Various	Hungary	€	130,000,000.00	500,000.00	1
Indie/Electronic	Spain	€	88,000,000.00	235,000.00	2
Rock/Alternative	Spain	€	45,000,000.00	80,000.00	1
Electronic	Spain	€	35,000,000.00	100,000.00	1
Pop/Rock	Spain	€	30,000,000.00	150,000.00	1
Pop/Indie	Spain	€	27,000,000.00	70,000.00	2

Number of songs in **top 20 spotify by artist and country**

artist_name	country	song_count
Marracash	Italy	13
Mirella	Finland	13
Bad Bunny	Spain	12
Shiva	Italy	11
Linkin Park	Luxembourg	11
Plutonio	Portugal	11
Linkin Park	Germany	10
Saul	Slovakia	10
Quevedo	Spain	10
Lazza	Italy	10

artist_name	country	song_count
Bad Bunny	Spain	12
Quevedo	Spain	10
Manuel	Hungary	8
Allah, Yorghaki	Spain	3
Pogány Induló	Hungary	3
DESH	Hungary	3
Azahriah, DESH	Hungary	3
DESH, Young Fly	Hungary	3
Rauw Alejandro	Spain	3
Dzsúdló	Hungary	2

Top 10 Popular Artists Streamed



Top 10 Popular Tracks Streamed



Analysis: Most Popular Artists and Tracks

- Mainstream pop, alternative, electronic dominate, with artists like **Lady Gaga, Billie Eilish, Sabrina Carpenter**
- **Bad Bunny and GIMS** signal the rising influence of Latin and Francophone music
- **ROSÉ** is rising star who can boost festival attendance, especially for Gen Z and K-pop fans
- Maximize Ticket Sales by **booking top-ranked artists** and **trending genres**, festivals can cater to a **wide European audience**

Analysis: Song Danceability



Danceability in Music

- Hungarian audiences prefer **rhythmic, upbeat music** that encourages dancing.
- **DESH**: most popular artist (Hip hop / Rap)
- Set up stages/zones with **immersive lighting and sound**

artist_name	country	count_rank
DESH	Hungary	101
Majka	Hungary	11
BSW	Hungary	9
Manuel	Hungary	8
T. Danny	Hungary	5
Pogány Induló	Hungary	2

Analysis: Song Energy



genre	country	currency	total_impact	total_attendan...	number_of_festivals
Electronic/Rock	Romania	€	25,000,000.00	50,000.00	1

Energy in Music

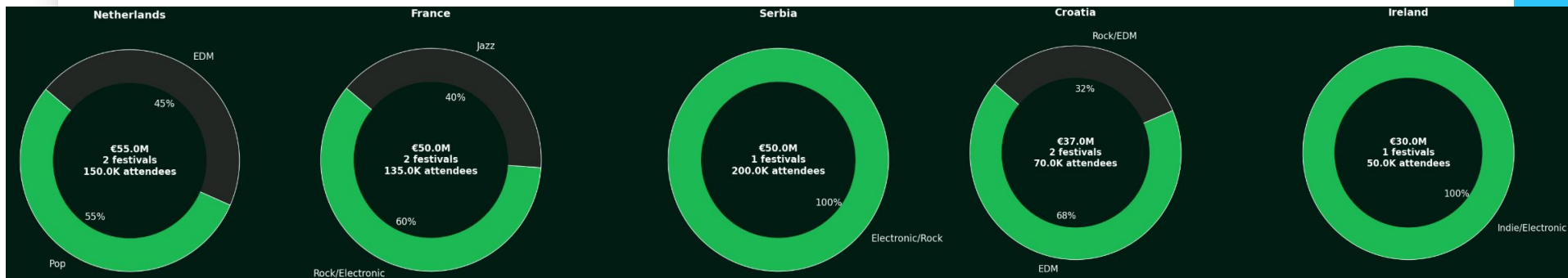
- **Festival Potential:** Romania could be an ideal location for **new electronic or high-energy music festivals**, targeting fans of EDM, Techno, and Hip Hop.
- High-energy artists lineups

artist_name	country	count_rank
Dani Mocanu	Romania	51
Jimin	Romania	43
Theo Rose	Romania	28
Mgk666	Romania	5
Babasha, Ministerul Manelelor	Romania	3
V	Romania	3

Analysis: Economic Impact of Music Festival Market in Europe

Economic impact | Genre | Number of Attendance share

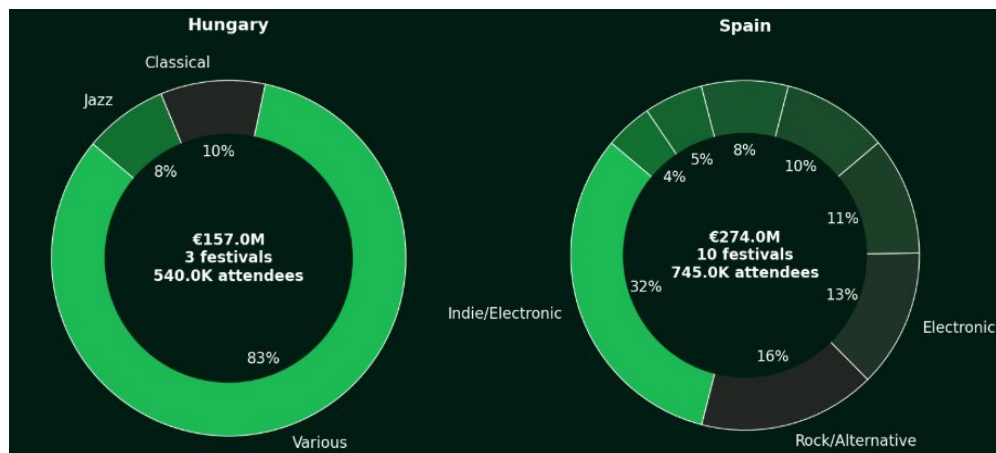
- Smaller Markets with Potential
- **France** (€50M, 2 festivals) – A mix of Rock/Electronic, but **room to expand into indie and pop festivals**
- **Serbia & Croatia** (€50M combined, 3 festivals) – Emerging markets with **Electronic/Rock & EDM**, suitable for **smaller, high-energy festivals**



Analysis: Economic Impact of Music Festival Market in Europe

Economic impact | Genre | Number of Attendance share

- **Hungary and Spain** has a **gap for Latin trap and reggaeton**, making Bad Bunny and similar artists ideal for filling this demand
- Hungary's largest and most iconic music festival Sziget 83% of economic impact
- Expanding Hip Hop, Rap, Techno, and high-energy dance music could further enhance the festival's attractiveness
- **Best market fit for new music festivals**



Obstacles & Lessons Learned

Obstacle 1

Missing values and duplicate records affected the accuracy of results.

How We Overcame It:

- We applied imputation strategies for missing data and removed duplicates using `pandas.drop_duplicates()`.

Lessons Learned:

- Validating data quality early on reduces downstream errors and processing time.



Obstacles & Lessons Learned

Obstacle 2:

CSV files with special characters (such as accented letters) caused errors during import into MySQL.

How We Overcame It:

- Loaded each csv file (or table) using SQL commands.

Lessons Learned:

- Ensuring **consistent character** encoding prevents data corruption and import errors.



Conclusions & Next Steps

- **Most-streamed tracks** feature a mix of **pop, electronic, and indie**, showing **strong audience demand** for diverse genres

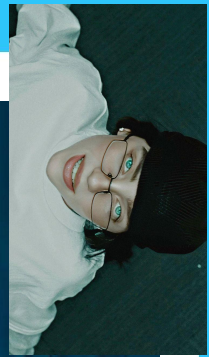
❌ **Hypothesis:** UK, Belgium, Germany are the ideal countries for hosting a new music festival.

📌 Ideal Location for a New Festival: **Hungary and Spain**

- Top choices due to economic potential, genre gaps, and audience demand.
- Spain's Latin music gap presents an opportunity for a dedicated **Latin & urban festival**.
- Hungary's high-energy music demand suggests potential for a **dance-focused festival** (EDM, Latin, and hip-hop fusion)

Next steps:

- Promotion & Marketing Campaign: Utilize influencers and social media platforms (TikTok, Instagram, etc.) to build excitement and anticipation leading up to the festival
- Ticketing and pricing strategy: VIP, general admission, early bird offers



SPOTIFY DATASET ANALYSIS



THANKS !

Ricardo, Rebecca