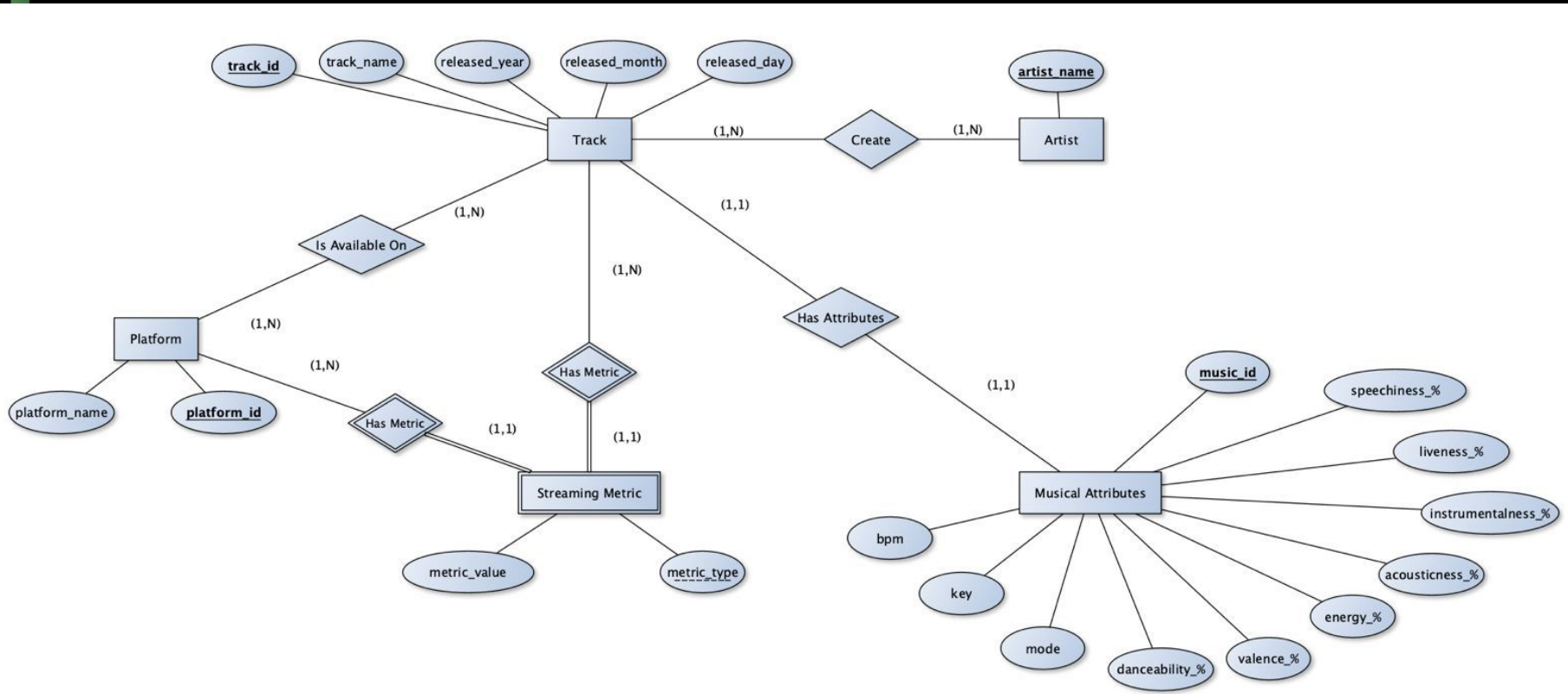
Databases and Big Data: Spotify

Group 8

01

ER Diagram


track_id
track_name
released_year
released_month
released_day
Track
(1,N)
Create
(1,N)
artist_name
Artist
(1,N)
Is Available On
(1,N)
Platform
(1,N)
platform_name
platform_id
(1,N)
Has Metric
(1,1)
Streaming Metric
metric_value
metric_type
(1,N)
Has Metric
(1,1)
(1,1)
Has Attributes
(1,1)
Musical Attributes
music_id
speechiness_%
liveness_%
instrumentalness_%
acousticness_%
energy_%
valence_%
danceability_%
mode
key
bpm

Track Entity

Purpose

The Track entity represents the core entity of songs in the database. It connects related data such as streaming metrics, platforms, artists, and musical attributes.

- **Centralized Data:** The entity organizes and identifies tracks, linking them to key entities like platforms and musical attributes.
- **Scalability:** It provides a structured, efficient framework for expanding the database with new tracks or associated data.

Cardinalities

Relationship with Streaming Metrics:

- 1-to-Many: A single track can have multiple streaming metrics because it can be streamed on different platforms and have multiple types of metrics.
- 1-to-1: Each streaming metric entry belongs to only one track.

Relationship with Platform:

1-to-Many:

- A single platform (e.g. Spotify) can host multiple tracks.
- A track can be hosted on one or more platforms.

Relationship with Artist:

1-to-Many:

- An artist can have multiple tracks.
- Each track is associated to one or more artists.

Relationship with Musical Attributes:

1-to-1:

- A track has exactly one set of musical attributes.
- Each set of attributes belongs to one and only one track.

Musical Attributes Entity

Purpose

Provides a detailed description of each track's musical profile, such as:

- Musical notations like **`bpm`**, **`key`**, and **`mode`**
- Percentages like **`danceability`**, **`valence (positivity)`**, **`energy`**, **`acousticness`**, **`instrumentalness`**, **`liveness`**, and **`speechiness`**

Cardinalities

- The Musical Attributes entity relates directly to the Track Entity

1-to-1 relationship:

- A track has exactly one set of musical attributes
- Each set of Musical Attributes corresponds to one and only one track

Artist Entity

Purpose

Provides a centralized representation of artists which consists of the name of the artists that were involved in the production.

It contains one and only one attribute that serves as the identifier of this entity:

- `artist_name`

Cardinalities

- The Artist entity relates directly to the Track Entity

1-to-Many relationship:

- An artist can create multiple tracks
- Each track is associated with one or more artists

Platform Entity

Purpose

- Describes the tracks that are available on various platforms.
- It contains **`platform_id`** and **`platform_name`**.
- It also serves as a form of metric supplied by the platform via its own statistics.

Cardinalities

Relationship with Track:
1-to-Many:
- A single platform can host multiple tracks.
- Each track can be hosted on multiple platforms.

Relationship with Streaming Metrics:
1-to-Many :
- A single platform has multiple associated streaming metrics.

Streaming Metric Entity

Purpose

- Streaming Metric entity is a weak entity designed to capture detailed performance metrics for tracks on various platforms
- It takes **platform_id** and **track_id** as foreign keys and has a partial key called **metric_type**
- **metric_value** is the numeric value associated with each metric

Cardinalities

- The Streaming Metric entity relates directly to the Track Entity

Relationship with Track:

- 1-to-Many: A Track can have multiple associated Streaming Metric entries
- 1-to-1: Each Streaming Metric entry is linked to exactly one specific Track

Relationship with Platform (Total Participation):

- 1-to-Many: A Platform has multiple Streaming Metric entries for various tracks
- 1-to-1: Each Streaming Metric entry is tied to exactly one Platform

02

SQL and Python Application

Preparation

To start the analysis, we imported key libraries:

- **`mysql.connector`** for database connections
- **`pandas`** for data manipulation and preprocessing
- **`rich.console`** and **`rich.table`** for visually enhanced outputs

The goal here is:

1. Import necessary libraries for database interaction and data processing
2. Set up the credentials for connecting to a MySQL database
3. Create a function (**`connect_db`**) that establishes the connection to the MySQL database
4. Specify the file path to the CSV for future data manipulation

→ The code sets up the foundation for the subsequent operations.

Setting up the Database and Tables

Database Setup	Tables Setup
• `CREATE DATABASE IF NOT EXISTS spotify_db`: Created “spotify_db” database if not already present. • Switched to ”spotify_db” context for table creation.	• “Platform”, “Artist”, “Track”, “MusicalAttributes”, “SteamingMetric” with primary keys and constraints. • Tables for relationships: “TrackArtist”, “TrackMusicalAttributes”

Populating the Tables

- The data was populated from a CSV file containing details on tracks, artists, and attributes
- Missing or NaN values were handled by replacing them with “None” (Interpreted as NULL on SQL)
- Entries were inserted uniquely into the appropriate tables while ensuring data integrity through constraints like PRIMARY KEY and FOREIGN KEY

Query 1: Comparison of Track Attribute Analytics by Season of Release

This query looks at how musical attributes vary across seasons, based on the month songs were released.

It groups the results into seasons (using CASE statement on SQL), providing details like the average beats per minute (BPM), the percentage of songs in (major and minor) keys, and averages for its other characteristics.

This offers valuable insights into seasonal trends for industry decision-making.

Q1 Result

Query 1: Comparison of Track Analytics by Season of Release.

Season	Avg BPM	Major Key %	Minor Key %	Avg Valence	Avg Energy	Avg Danceability	Avg Acousticness	Avg Liveness	Avg Speechiness
Winter	123.637	54	46	56.0	65.0	66.0	29.0	19.0	9.0
Spring	123.5179	58	43	51.0	64.0	68.0	28.0	18.0	11.0
Summer	120.8196	58	42	51.0	66.0	69.0	24.0	18.0	10.0
Fall	121.4115	63	37	46.0	62.0	65.0	27.0	18.0	10.0

1. BPM (Tempo)

- Highest in Winter (123.637)
- Lowest in Summer (120.8196)

2. Major and Minor Key Percentages

- Major Key: Peaks in Fall (63%)
- Minor Key: Highest in Winter (46%)

3. Valence (Musical Positivity)

- Highest in Winter (56.0)
- Lowest in Fall (46.0)

4. Energy and Danceability

- Summer Leads in Both (Energy: 66.0, Danceability: 69.0)
- Lowest in Fall (Energy: 62.0, Danceability: 65.0)

5. Acousticness

- Highest in Winter (29.0)
- Lowest in Summer (24.0)

6. Liveness

- Consistent across seasons (~18-19), with a slight peak in Winter (19.0).

7. Speechiness

- Highest in Spring (11.0)
- Lowest in Winter (9.0)

Query 2: Top 10 Artists Based on Weighted Streaming Metrics Across Platforms

This query ranks the top 10 artists across platforms (Spotify, Apple, Deezer, and Shazam) by calculating a weighted score (using STDev as the criterion).

It adjusts metrics like playlist appearances, chart rankings, and Spotify streams based on each platform's averages and variations, creating a comprehensive ranking of artist performance.

Q2 Result

Query 2: Top Artists Based on Weighted Streaming Metrics Across Platforms

Artist	Weighted Score
Harry Styles	457.38
Bad Bunny	402.73
Taylor Swift	353.01
The Weeknd	317.39
Kendrick Lamar	169.21
Drake	166.37
Quevedo	164.32
Feid	162.55
Metro Boomin	148.03
Anuel Aa	137.71

- **Harry Styles** leads with a weighted score of **457.38**, reflecting his dominance across streaming platforms through strong playlist presence, chart, and streams.
- **Bad Bunny** follows closely with **402.73**, highlighting his significance in the with consistent streaming and platform appearances.

Query 3: Analyzing Playlist Popularity by Danceability with Significant Deviations

This query looks at how a track's danceability affects its popularity on playlists.

It starts by analyzing the songs, finding a correlation between danceability (expressed in ranges) and appearance in playlists across platforms, thus creating an entry-level prediction system.

Then it analyzes the songs once again to find outliers (using a z-score of 2.5 to classify them as such), then outputs them by deviation in descending order (showing most extreme outliers first)

Q3 Result

Query 3: Analyzing Playlist Popularity by Danceability with Significant Deviations

Track	Danceability %	Danceability Range	Playlists	Avg Playlists for Dancebility	Playlists w.r.t. Standard Deviation	Z-Score Deviation
Get Lucky - Radio Edit	79.0	61-80%	52898	4758	7017	6.86
One Dance	77.0	61-80%	43257	4758	7017	5.49
Somebody That I Used To Know	86.0	81-100%	42798	4758	7017	5.42
Wake Me Up - Radio Edit	53.0	51-60%	50887	5780	8794	5.13
Smells Like Teen Spirit - Remastered 2021	52.0	51-60%	49991	5780	8794	5.03
We Found Love	73.0	61-80%	36843	4758	7017	4.57
Mr. Brightside	35.0	31-50%	51979	7244	9987	4.48
Take On Me	57.0	51-60%	44927	5780	8794	4.45
Rolling in the Deep	73.0	61-80%	35684	4758	7017	4.41
Still D.R.E.	81.0	81-100%	33966	4758	7017	4.16
Everybody Wants To Rule The World	64.0	61-80%	41751	5780	8794	4.09
Thinking Out Loud	78.0	61-80%	33032	4758	7017	4.03
HUMBLE.	91.0	81-100%	33206	4331	7210	4.00

1. Danceability

- Lowest: Mr. Brightside (35.0)
- Highest: The Next Episode (92.0)

2. Danceability Range

- 31-50%: 5 songs
- 51-60%: 6 songs
- 61-80%: 15 songs
- 81-100%: 9 songs

3. Playlists

- Highest: Get Lucky - Radio Edit (52898)
- Lowest: Every Breath You Take - Remastered 2003 (22439)

4. Average Playlists for Danceability (for each range)

- 31-50%:7244
- 51-60%: 5780
- 61-80%: 4758
- 81-100%: 4758

5. Playlists with Regard to Stdv.

- 31-50%: 9987
- 51-60%: 8794
- 61-80%: 7017
- 81-100%: 7017

6. Z-score Deviation

- Highest: Get Lucky - Radio Edit (6.86)
- Lowest: Every Breath You Take - Remastered 2003 (2.52)

Query 4: High-Intensity Workout Playlist With Increasing BPM

This query generates a workout playlist by selecting tracks with a BPM over 130, high energy (above 70), positive mood (valence above 50), and high danceability (above 60), without prioritizing live recordings, acousticness, and spoken words.

Tracks are sorted by BPM in ascending order with increasing tempo (with ties broken by valence, energy, and danceability).

The result is a playlist designed for progressively intense workouts.

Q4 Result

Query 4: High-Intensity Workout Playlist With Increasing BPM

Track Name	Artist(s)	BPM
Super Freaky Girl	Nicki Minaj	133.0
Boy's a liar Pt. 2	Ice Spice, PinkPantheress	133.0
Boy's a liar	PinkPantheress	133.0
Forgot About Dre	Eminem, Dr. Dre	134.0
Beggin	M?ne	134.0
Malvada	Z? Fe	135.0
AMG	Peso Pluma, Gabito Ballesteros, Natanael Cano	136.0
Nxde	(G)I-DLE	136.0
B.O.T.A. (Baddest Of Them All) - Edit	Interplanetary Criminal, Eliza Rose	137.0
PRC	Peso Pluma, Natanael Cano	138.0
El Gordo Trae El Mando	Chino Pacas	140.0
LADY GAGA	Peso Pluma, Gabito Ballesteros, Junior H	140.0
I Ain't Worried	OneRepublic	140.0
Sky	Playboi Carti	140.0
Shivers	Ed Sheeran	141.0
One Thing At A Time	Morgan Wallen	142.0

The result is a high-intensity workout playlist with increasing BPM values varying from 130 to 196, consisting of 40 tracks.

Query 5: Top Charting Tracks Across Platforms

This query finds the tracks on the top 10 charts over all platforms (Spotify, Apple Music, and Deezer).

For each track, it displays a combined list of its artists, key, and musical mode.

It shows the chart position for the track on each platform. The results are sorted by their rankings, offering a clear picture of the musical traits of successful hits across multiple platforms.

Q5 Result

Query 5: Top Charting Tracks Across Platforms

Track Name	Artist(s)	Key	Mode	Spotify Rank	Apple Music Rank	Deezer Rank
Adore You	Harry Styles	G#	Major	1	5	2
DAN?A	Mc Pedrinho, Pedro Sampaio	A	Minor	2	1	1
With you	Jimin, HA SUNG WOON	D#	Major	2	1	1
Sin Se?	Quevedo, Ovy On The Drums	B	Minor	2	1	1
Envolver	Anitta	E	Minor	2	4	1
Down Under (feat. Colin Hay)	Luude, Colin Hay	B	Minor	2	7	1
Se?o	Shawn Mendes, Camila Cabello	A	Minor	2	8	1
Antidepresan	Mabel Matiz, Mert Demir	B	Minor	2	9	1
Numb	Marshmello, Khalid	None	Minor	2	9	1
Good Days	SZA	C#	Minor	2	10	1

Popular Keys: A Minor, and B Minor (suggesting that they resonate well with listeners across platforms)

Standout:

“Adore you” (Harry Styles)

- Consistently top-ranked across platforms (#1 on Spotify, top 5 on Apple Music and Deezer).
- Upbeat and universally appealing with a G# Major key.

Query 6: High-Energy, Low-Speechiness Tracks on Spotify’s Top 20

This query focuses on high-energy tracks with low speechiness that are ranked in Spotify's top 20 charts.

For each track it displays its energy and speechiness percentages, and Spotify chart position.

Only tracks with energy above 70% and speechiness below 10% are included (as an intuitive measure).

The results are sorted to show the most energetic songs first, followed by the best chart rankings, offering a glimpse at Spotify’s most successful focus-inducing tracks.

Q6 Result

Insights for Chart Success

- **High Energy (Above 90.0):**

Tracks in this range are best for keeping a steady tempo while in a focused state, as they are engaging and dynamic.

- **Speechiness Range (4.0–7.0):**

Tracks with moderate speechiness maintain melodic appeal without overemphasis on lyrics or spoken content to not intervene with the focused state.

Well-suited for programming, gaming, intense study sessions and more.

Query 6: High-Energy, Low-Speechiness Tracks on Spotify's Top 20

Track Name	Artists	Energy	Speechiness	Spotify Chart Rank
T?	dennis, MC Kevin o Chris	96.0	5.0	15
Bombonzinho – Ao Vivo	Ana Castela, Israel & Rodolffo	95.0	5.0	6
Freaks	Surf Curse	94.0	5.0	3
?	YOASOBI	94.0	9.0	16
Mr. Brightside	The Killers	93.0	8.0	15
Layla	Sch?rze, DJ R	92.0	7.0	3
Seu Brilho Sumiu – Ao Vivo	Israel & Rodolffo, Mari	92.0	5.0	5

Example Outputs

Top 3 in charts:

Freaks - Surf Curse

- Energy: 94.0
- Speechiness: 5.0

Top 3 in charts:

Layla - Sch?rze, DJ R

- Energy: 92.0
- Speechiness: 7.0

Query 7: Uplifting and Danceable Tracks in Spotify's Top 25

This query highlights tracks in Spotify's top charts that are exceptionally danceable and uplifting to create a happiness-inducing effect.

Only tracks with danceability and valence scores above 80 are included (as an intuitive measure).

This offers a look at some of the most feel-good hits on Spotify. The result can be interpreted as:

- Pseudo-Caffeine! Maintaining focus and awake state while driving
- Getting over a broken heart with the mood-enhancing qualities
- Maximizing sales by inducing a happier mindset in customers

Q7 Result

Query 7:Uplifting and Danceable Tracks in Spotify's Top 25

Track	Artist(s)	Danceability %	Valence &	Spotify Rank
Ai Preto	Bianca, DJ Biel do Furduncinho, L7nnon	95.0	83.0	4
If We Ever Broke Up	Mae Stephens	90.0	96.0	4
ANTIFRAGILE	LE SSERAFIM	88.0	82.0	12
Monoton?	Ozuna, Shakira	87.0	82.0	15
Me Arrepent?	Ak4:20, Cris Mj, Pailita	86.0	91.0	2
A Tu Merced	Bad Bunny	86.0	89.0	11
Mi?n	Maria Becerra, Tini	85.0	92.0	16
MERCHO	LiL CaKe, Migrantes, Nico Valdi	84.0	96.0	20
Made You Look	Meghan Trainor	84.0	88.0	6
Shape of You	Ed Sheeran	83.0	93.0	10

Danceability:

- Highest - Ai Preto → 95.0
- Lowest - MERCHO & Made you look → 84.0

Valence:

- Highest - If We Ever Broke Up & MERCHO → 96.0

Query 8: Dynamic Music Playlist Generator

This music recommendation system gathers user preferences—such as mood, danceability, lyric preference, and acoustic style—to create personalized playlists.

This query filters tracks based on musical attributes like valence, BPM, energy, and speechiness, ensuring tailored recommendations.

If no matches are found (or not enough to create a satisfactory playlist), the system defaults to popular chart-toppers.

Q8 Result

```
Welcome to Your Personalized Playlist Creator!

Are you feeling happy or sad? (h/s):
Your Choice: happy
Do you feel like dancing? (y/n):
Your Choice: yes
Do you prefer Lyrics, instrumentals or both? (l/i/b):
Your Choice: lyrics
Do you prefer electronic or acoustic music? (e/a):
Your Choice: electronic
Do you like rap music? (y/n):
Your Choice: yes
```

Example Usage:

- Here the user is happy, and feels like dancing! They prefer songs with lyrics, and electronic music. They enjoy rap music.

Here is your personalized Playlist!

Query 8: Dynamic Music Playlist Generator

Track Name	Artist(s) Name
Bamba (feat. Aitch & BIA)	B?, Aitch, Luciano
PERO T?	Quevedo, Karol G
Super Freaky Girl	Nicki Minaj
Sobrio	Maluma
Sky	Playboi Carti

Best viewed with an extended Terminal.

Example Output:

- After the insertion of values into python, the are interpreted as “<=” or “>=” and some thresholds are put into place.
- If very little matches are found, an alternate query is ran, resulting in the top 5 streamed tracks.

The User Interface

```
MAIN MENU

Welcome! Initiate the Program, or Exit.
Best viewed with an extended Terminal.

(i) Initialize
(m) Miscellaneous
(s) Stop
```

In the MAIN MENU the User Can:

- Initialize tries to connect to SQL (try-except error handling.
    - If successful, redirects to the QUERIES MENU.
    - Else, redirects to the credential login, then redirects to the QUERIES MENU.
- View Miscellaneous information and notes.
- Stop running the Program.

```
SQL connection unsuccessful. Please enter credentials below:

Enter Username, or use default "root" (r):
root

Enter Password (Hidden):

Password entered...

Enter host, or choose default "localhost" (l):
localhost

SQL connection established.

Enter Filepath:

Your Choice: /Users/yaseminates/Desktop/Spotify Most Streamed Songs.csv

Filepath correct.
Now showing the QUERIES.
```

The input is stripped and made lowercase for ease of processing.

This process goes on until a valid input is entered, or the program is exited.

The User Interface

```
QUERIES MENU

View All Queries, Explore a Query, or Exit the Program.

    (v) View
    (1-8) Explore
    (b) Back to MAIN MENU
    (s) Stop
```

```
Your choice: 4

        Query 4: High-Intensity Workout Playlist With Increasing BPM

        Would you like to 'Run' the query or 'Explore' it?

            (r) Run
            (e) Explore

Your Choice: e

        This query generates a workout playlist featuring tracks with increasing BPM.
        The tracks are selected based on high valence, energy, and danceability, ensuring
        they are upbeat and rhythmic. The playlist also filters out tracks with low
        acousticness, speechiness, and liveness, favoring synthetic and studio-recorded music.

        The results are ordered by BPM, from the lowest to the highest, for progressively intense workouts.

        Type "Spoti" on MAIN MENU for a surprise.
```

In the QUERIES MENU The User Can:

- View All Queries (called from preset dictionary)
- Explore a Query
- Go back to MAIN MENU
- Stop running the Program

When (e) The User Can:

- Read a description of the query
- Run the Query to see its results (As seen previously)

All these redirections and operations are possible through “while” loops started and ended by pre-set boolean variables, coupled with try-except error handling!

Thank you for listening!

Group 8

Guia Ludovica Basso	296881
Alessio Giannotti	297061
Yasemin Ateş	304011
Elina Yılmaz	305561

A.Y. 2024-2025 for Databases and Big Data