Master's in Computer Engineering

Database 2

Team Project 1 : RDF

MARCO ALECCI , LUCA MARTINELLI, ELIA ZIROLDO

ACADEMIC YEAR 2021/2022

Data

❏ Available on Kaggle

❏ Spotify's Daily Top 200 charts

❏ 35 countries + global chart

❏ Over 3 years (2017-2020)

❏ 1.53 GB of data

Country	Date
Position	Track
Track URI	Artist

features in the dataset

Dataset Reduction

We reduced the initial dataset combining two different strategies:

❑ selecting only the first *x positions* of the charts

❑ sampling charts every *k days*

1.53 GB ➤➤ 109 MB

x = 100 positions, k = 7 days

Spotify Web API

To enrich the dataset with more interesting information we decided to use the **Spotify Web API**

Using the API, we added :

❏ **Albums**

❏ **Genres**

❏ **Available Markets**

❏ **More data about Tracks**

❏ **More data about Artists**

Get Track

🔒 OAuth 2.0

Get Spotify catalog information for a single track identified by its unique Spotify ID.

Request

∨ GET /tracks/{id}

id string   required

The Spotify ID for the track.

Example value: "11dFghVXANMIKmJXsNCbNI"

An example about how to retrieve data about a specific track

WikiData and MusicBrainz APIs

To add personal information about the artists, we decided to rely on **WikiData** and **MusicBrainz:**

1 We search for the artist on *MusicBrainz*

2 We take the first result and check if a *WikiData* link is present

3 Using the *WikiData* link, we retrieve the artist personal data (different approach if single artist or a group)

4 If the link is not present, we perform a query on *WikiData* and we perform the same actions of point 3
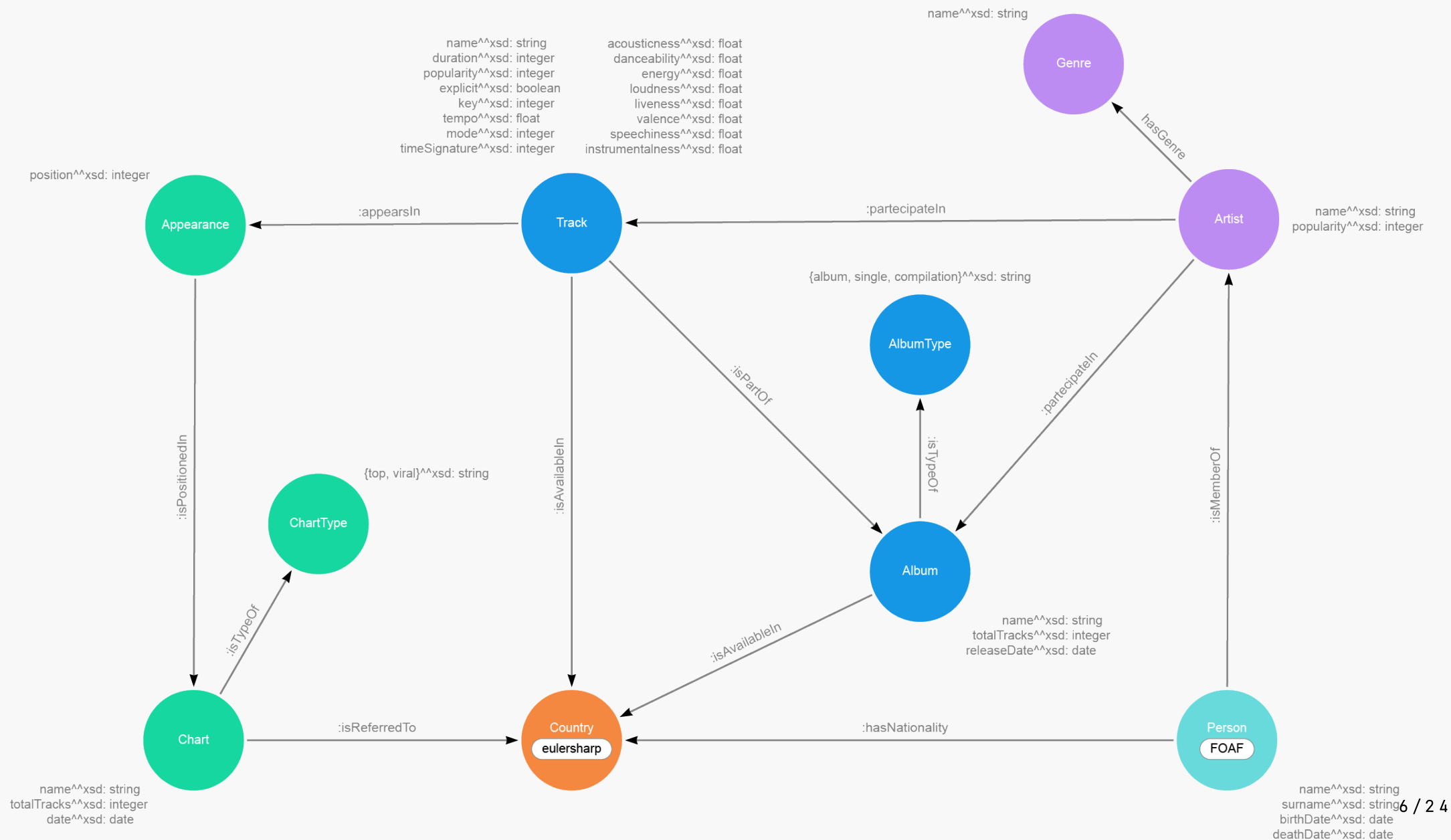
5 If the link is not present and the query was unsuccessful or incomplete, we search for data on *MusicBrainz*
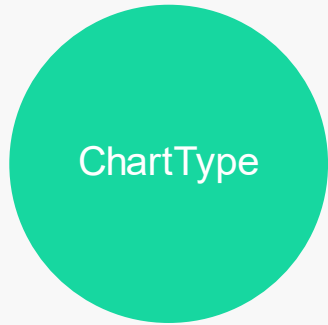
6 Finally, we remove duplicates, and we obtain a dataset with: *Name, Surname, BirthDate, DeathDate, Nationality, ArtistName*

Model

name^^xsd: string

acousticness^^xsd: float
danceability^^xsd: float
energy^^xsd: float
loudness^^xsd: float
liveness^^xsd: float
valence^^xsd: float
speechiness^^xsd: float
instrumentalness^^xsd: float

name^^xsd: string
duration^^xsd: integer
popularity^^xsd: integer
explicit^^xsd: boolean
key^^xsd: integer
tempo^^xsd: float
mode^^xsd: integer
timeSignature^^xsd: integer

position^^xsd: integer

Genre

Artist

name^^xsd: string
popularity^^xsd: integer

Appearance

:appearsIn

Track

:partecipateIn

:hasGenre

{album, single, compilation}^^xsd: string

AlbumType

:isPartOf

:isAvailableIn

:isPositionedIn

{top, viral}^^xsd: string

ChartType

:isTypeOf

:isTypeOf

:partecipateIn

Album

:isMemberOf

name^^xsd: string
totalTracks^^xsd: integer
releaseDate^^xsd: date

Chart

:isReferredTo

Country

eulersharp

:isAvailableIn

:hasNationality

Person

FOAF

name^^xsd: string
totalTracks^^xsd: integer
date^^xsd: date

name^^xsd: string
surname^^xsd: string
birthDate^^xsd: date
deathDate^^xsd: date

Model : Focus 1

ChartType

{top, viral}^^xsd: string

ENUMERATION

AlbumType

{album, single, compilation}^^xsd: string

ENUMERATION

Genre

name^^xsd: string

NOT AN
ENUMERATION

Model : Focus 2

position^^xsd: integer

Appearance

:appearsIn

Track

name^^xsd: string
duration^^xsd: integer
...: ...

dummy node to represent
the positioning of a track in a
specific Chart

:isPositionedIn
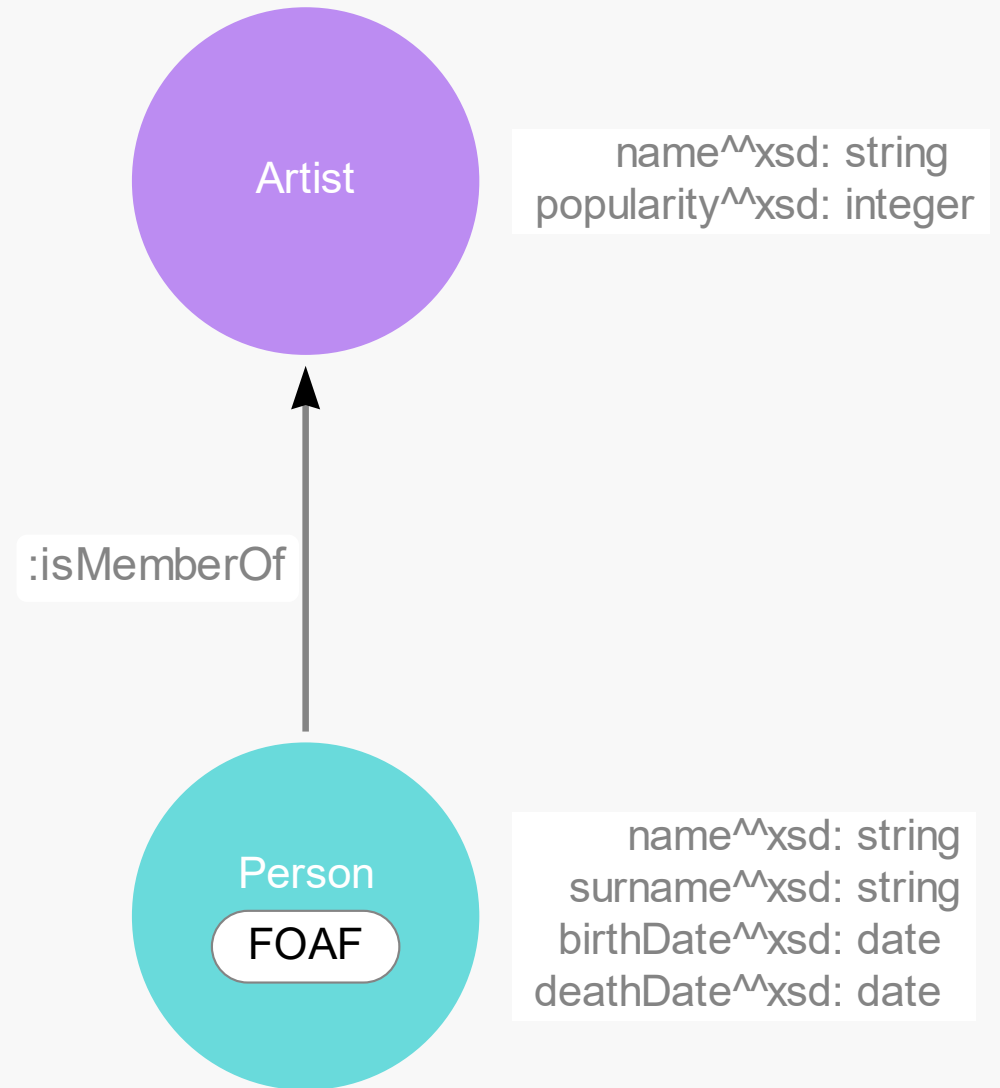
Contains all the information
related to a specific chart

Chart

name^^xsd: string
totalTracks^^xsd: integer
date^^xsd: date
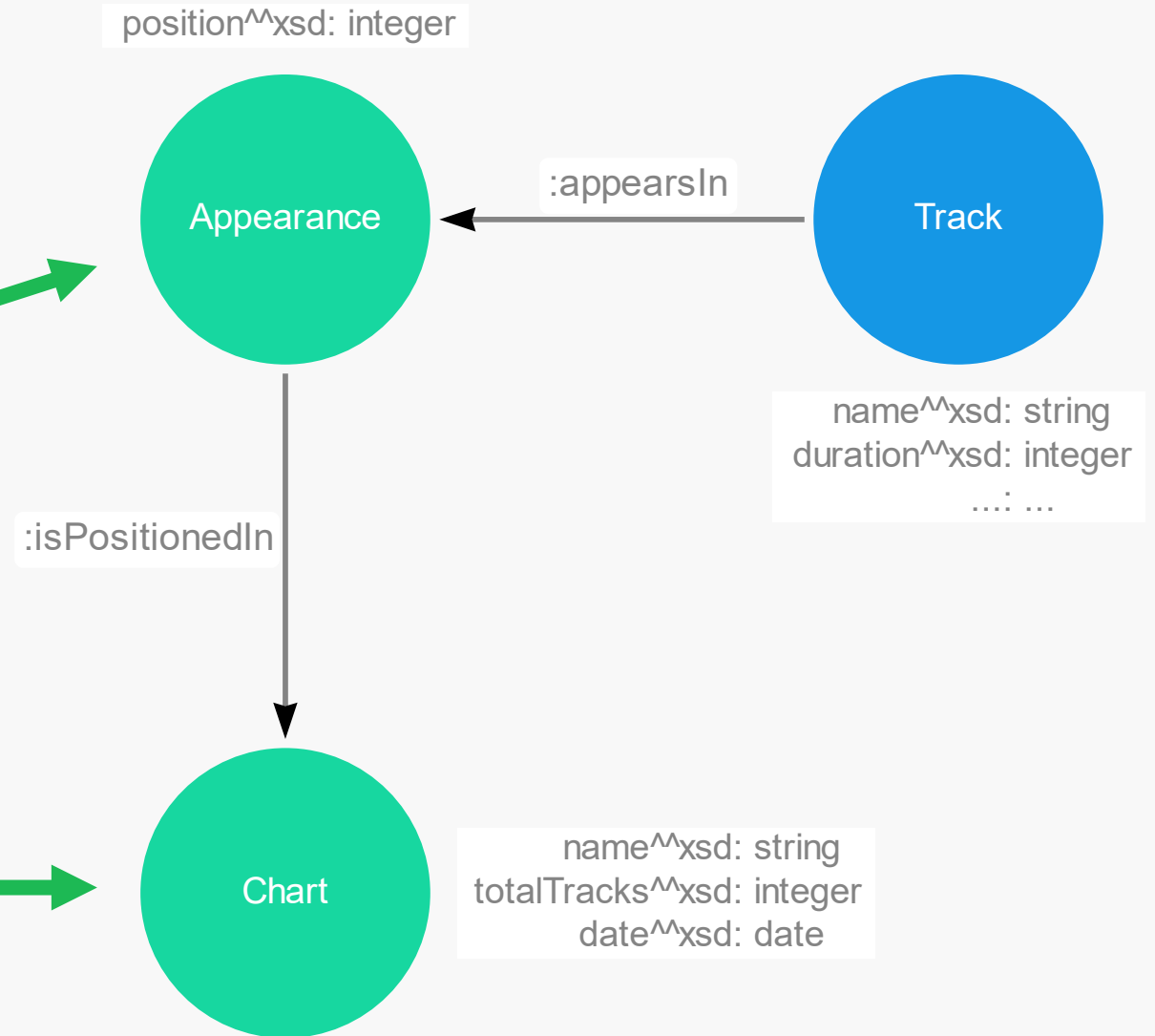
Model : Focus 3

The **Artist node** is used to represent both:

❑ Group

❑ Single artist

Example:

❑ *Sting* and *The Police* are both represented as artists

❑ *Gordon Sumner* (Sting's real name) is both linked to *Sting* and *The Police*

Artist

name^^xsd: string
popularity^^xsd: integer

:isMemberOf

Person

FOAF

name^^xsd: string
surname^^xsd: string
birthDate^^xsd: date
deathDate^^xsd: date

Serialization

Load the CSV file and iterate through it

↓

Retrieve the unique ID used by Spotify or create a unique ID by ourself

↓

Add the node to the graph using the unique ID

↓

Add all the data properties

↓

Add all the object properties

↓

Serialize the data and save them into a TTL file

❑ Output: **turtle format** (TTL)

❑ Relying on **RDFLib** library

❑ Easy linking thanks to **Spotify unique ID**

Queries

Query 1

On average how many artist of a specific nationality are in the top 100 of the same country?
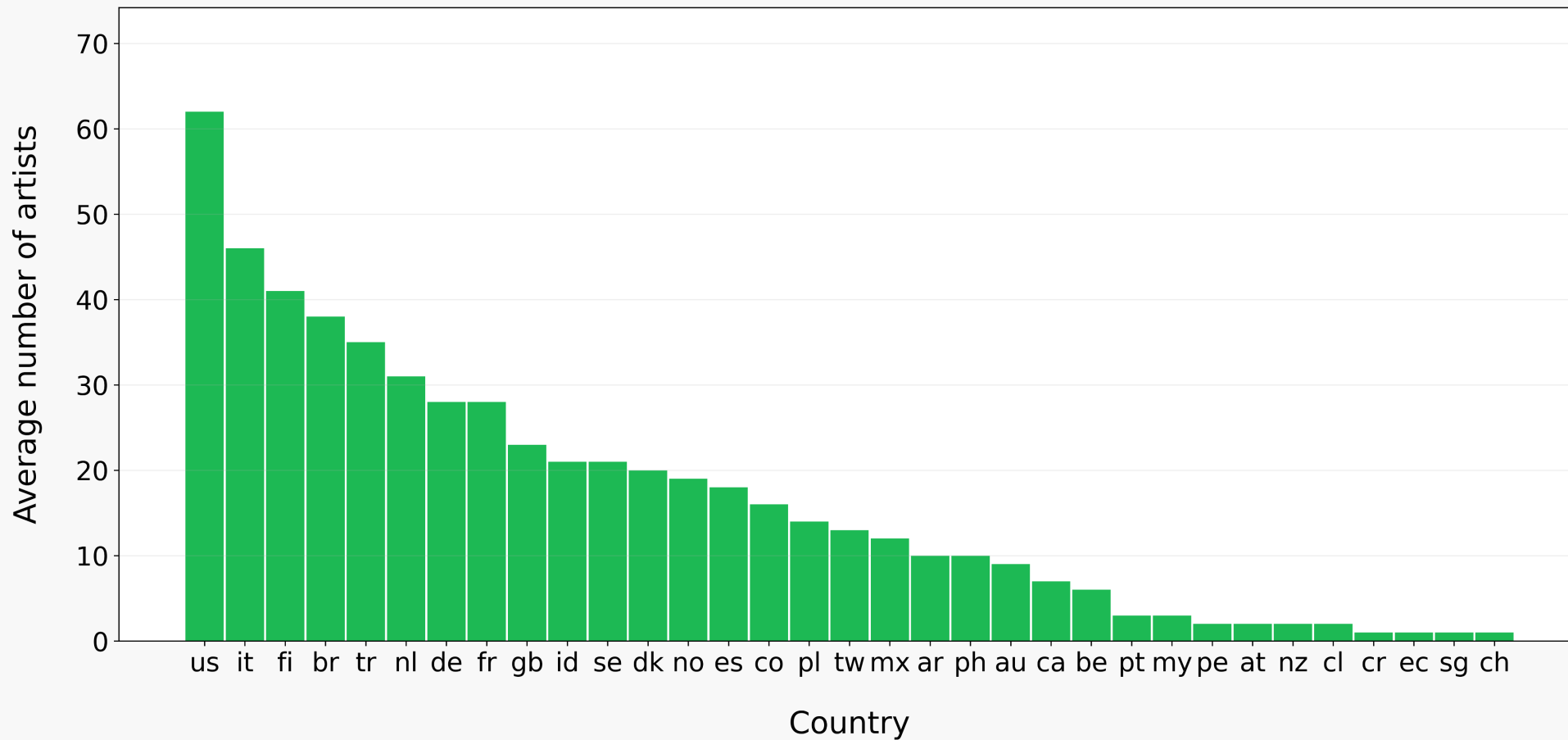
Query 2

How many times tracks made by an artist born after 2000 were present in top 100 Italy in 2020?

Track Name	Artist Name	Artist Real Name	BirthDate	numAppereances
blun7 a swishland	tha Supreme	Davide Mattei	2001-03-17	44
Cin Cin	Yanomi	Andrea De Filippi	2000-08-22	38
Roses – Imanbek Remix	Imanbek	Imanbek	2000-10-21	37
Dance Monkey	Tones And I	Toni Watson	2000-08-13	32
fuck 3x	tha Supreme	Davide Mattei	2001-03-17	30
Nena (feat. Geolier & Andry The Hitmaker)	Geolier	Geolier	2000-03-23	28
SUPREME – L'ego (feat. tha Supreme & Sfera Ebb…)	tha Supreme	Davide Mattei	2001-03-17	28
TesTa TrA Le NuVoLE, pT. 2 – prod. Yanomi	Yanomi	Andrea De Filippi	2000-08-22	24
death bed (coffee for your head)	beabadoobee	Beatrice Kristi Laus	2000-06-03	24
M' Manc (con Geolier & Sfera Ebbasta)	Geolier	Geolier	2000-03-23	21

Query 3

Show the 20 most popular under-30 Italian artist who published at least a track in 2018
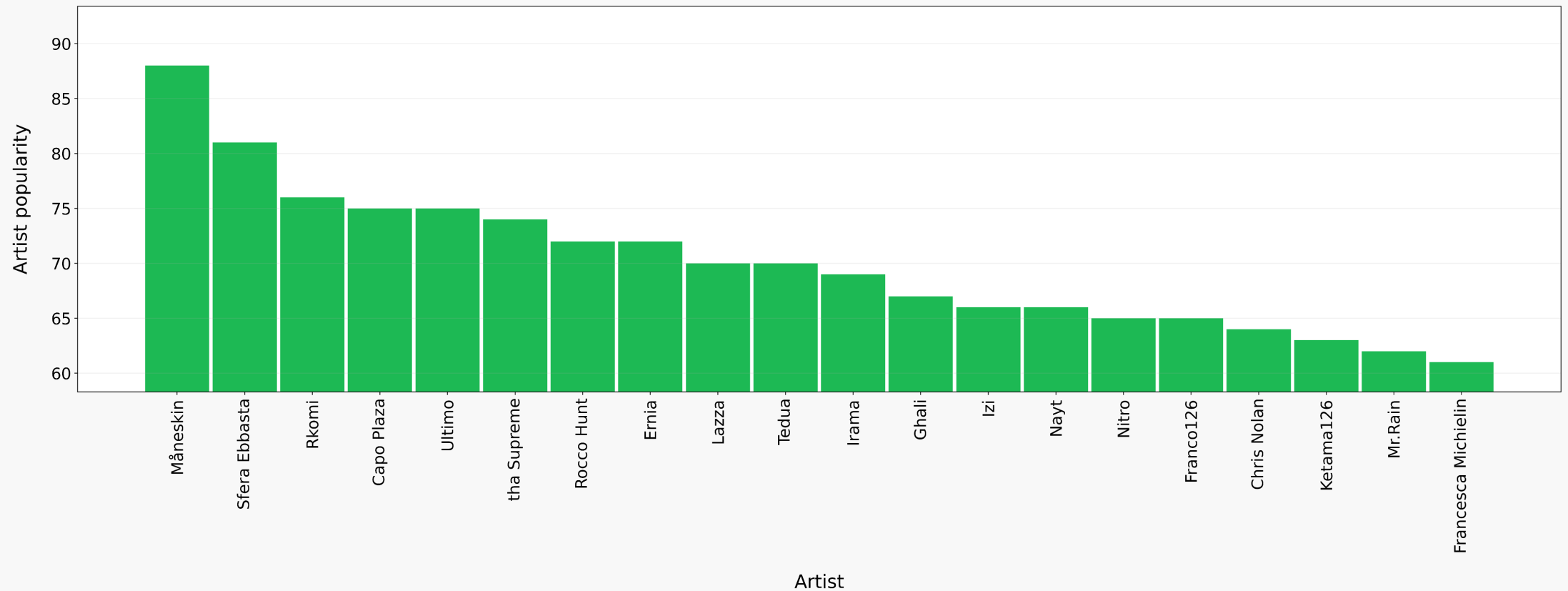
Query 4

In 2019 there was more distinct singles in top 100 Argentina rather than in top 100 Italy?

Query 4

In 2019 there was more distinct singles in top 100 Argentina rather than in top 100 Italy?
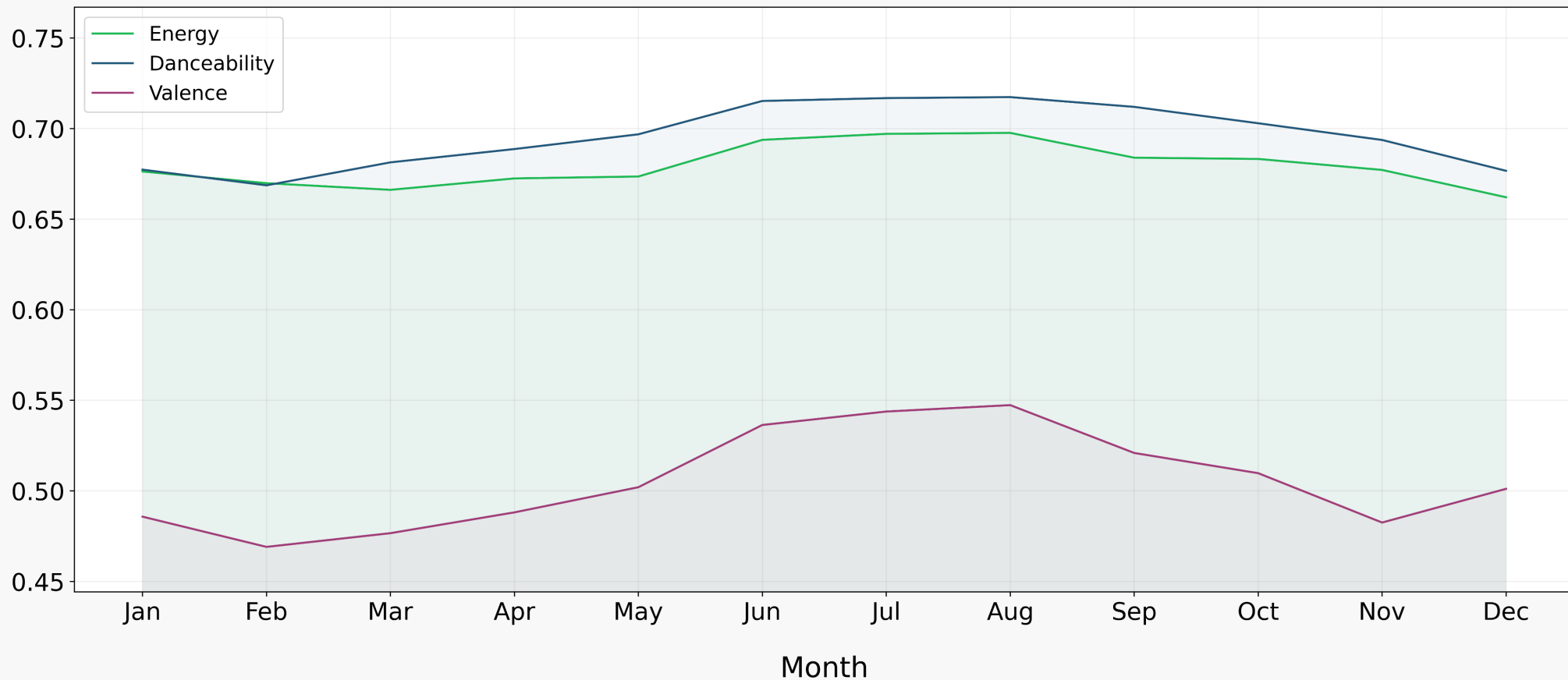
YES

Query 4

In 2019 there was more distinct singles in top 100 Argentina rather than in top 100 Italy?
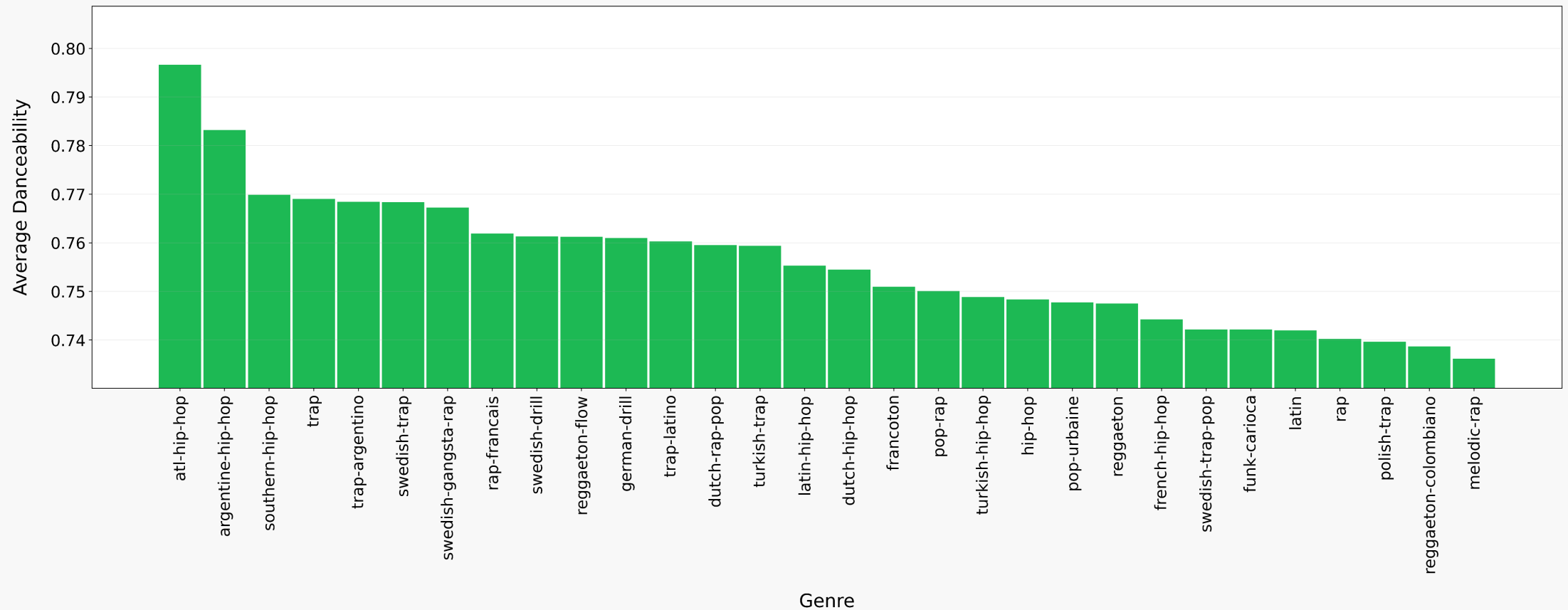
YES

❑ Argentina : 302 singles

❑ *Italy : 272 singles*

Query 5

Show average of different features of tracks through the months of the year in Italy

Query 6

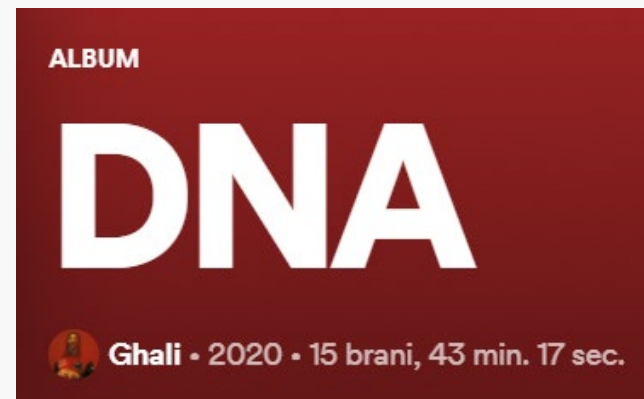Show the 30 genres with most danceable songs relative to songs released in 2019
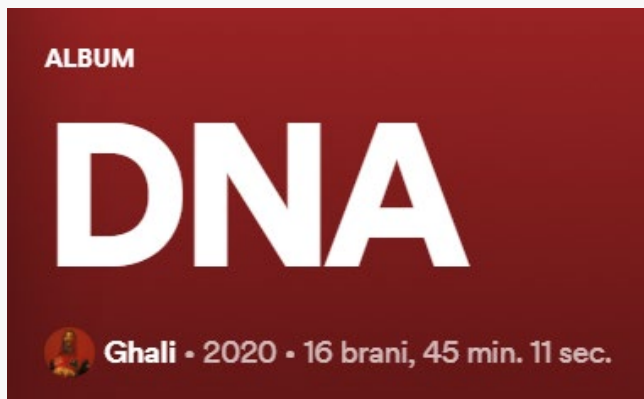
Query 7

How many albums (at least 2) contains the same track?

Track Name	Albums	albumCount
Boogieman (feat. Salmo) (Ghali, Salmo)	DNA, Boogieman (feat. Salmo)	6
Believer (Imagine Dragons)	Evolve, Believer	6
Thunder (Imagine Dragons)	Evolve, Thunder	6
Feelings (Lauv)	~how i'm feeling~, Feelings	6
i'm so tired... (Troye Sivan, Lauv)	~how i'm feeling~, i'm so tired...	6
Verloren (Tabitha)	Beste Zangers Seizoen 2020 (Tabitha)	5
Good Times (Ghali)	DNA	5
Whatever It Takes (Imagine Dragons)	Evolve, Whatever It Takes	5
Sunflower – Spider-Man: Into the Spider-Verse ...	Hollywood's Bleeding, Sunflower (Spider-Man: I...	5
My Only Wish (This Year)	Pop & Wave - Christmas Edition, Five Songs Of ...	5

Query 7

How many albums (at least 2) contains the same track?

ALBUM

DNA

Ghali · 2020 · 16 brani, 45 min. 11 sec.

ALBUM

DNA

Ghali · 2020 · 15 brani, 43 min. 17 sec.

ALBUM

DNA

Ghali · 2020 · 17 brani, 47 min. 36 sec.

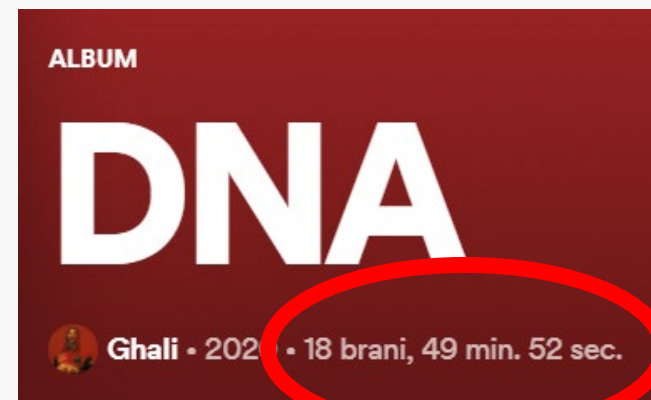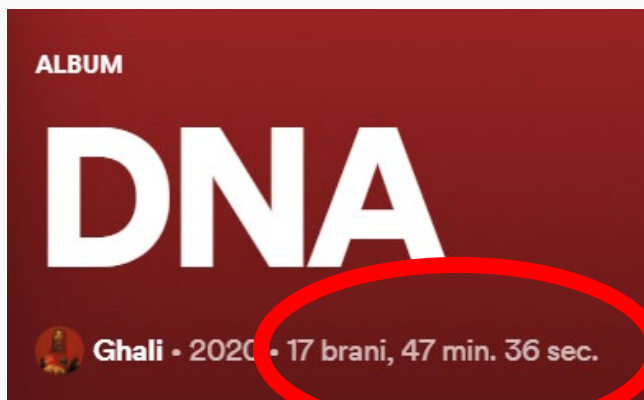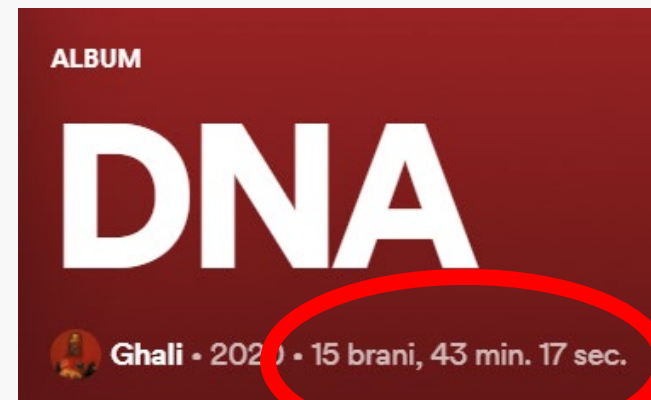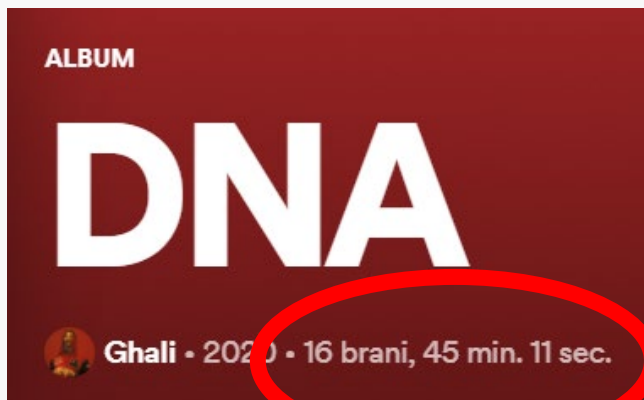ALBUM

DNA

Ghali · 2020 · 18 brani, 49 min. 52 sec.
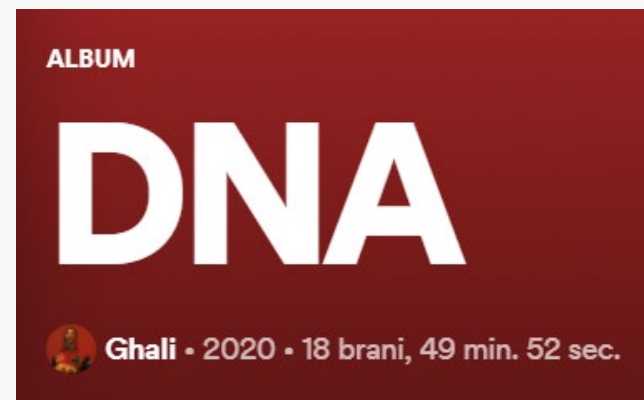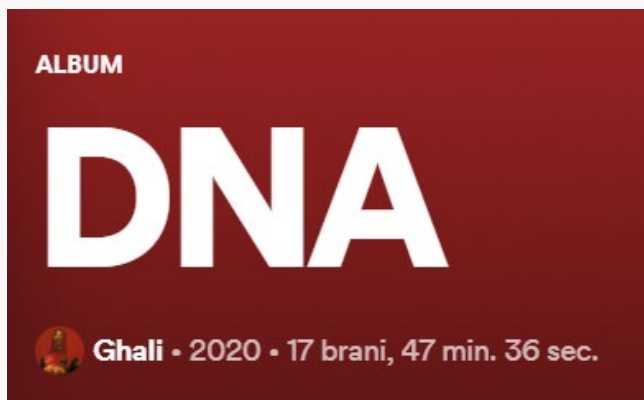
Query 7

How many albums (at least 2) contains the same track?

Query 8

How many songs of artists from Japan or South Korea reaches Top 20 in US, Canada or Mexico?

❑ US : 7 tracks

❑ *Canada : 9 track*

❑ *Mexico : 0 tracks*

Query 8

How many songs of artists from Japan or South Korea reaches Top 20 in US, Canada or Mexico?

❑ US : 7 tracks

❑ *Canada : 9 track*

❑ *Mexico : 0 tracks*

Track Name	Artist
Run	Jojo
Happy Xmas (War Is Over) – Remastered 2010	Yoko Ono, The Plastic Ono Band
Dynamite	BTS
Boy With Luv (feat. Halsey)	BTS
Sour Candy (with BLACKPINK)	BLACKPINK
Ice Cream (with Selena Gomez)	BLACKPINK
How You Like That	BLACKPINK
Old Town Road (feat. RM of BTS) – Seoul Town Remix	BTS, RM
Lovesick Girls	BLACKPINK

Query 9

Show how many explicit tracks appears in the top 100 for each country
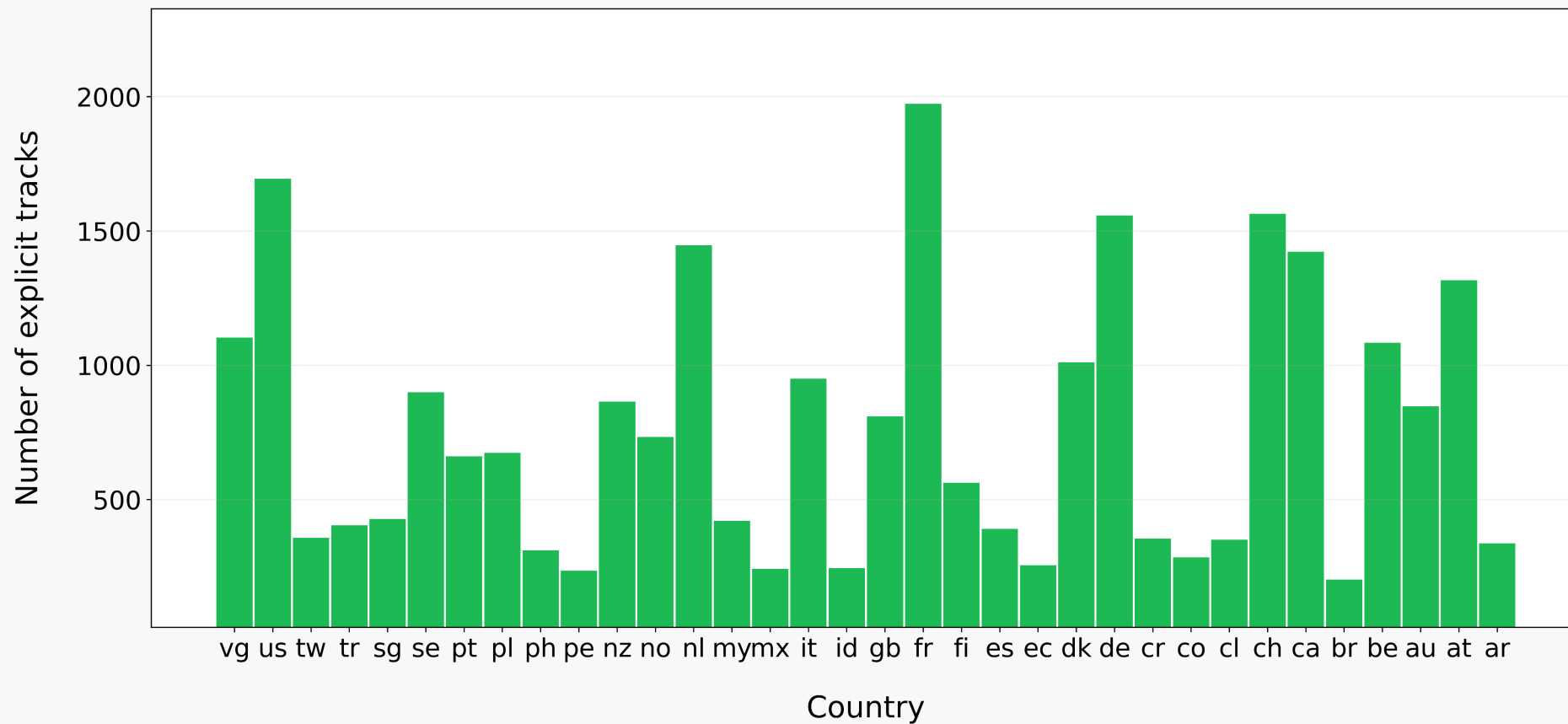
Query 10

Show the 30 albums with the highest number of tracks present in Top 30 at the same time, showing also how many tracks are contained in the album

	Chart	Date	Album Name	numTracks	totalTracks
1	TOP 100 USA	2018-07-01	Scorpion	23	25
2	TOP 100 Canada	2018-07-01	Scorpion	22	25
3	TOP 100 Canada	2017-03-19	More Life	21	22
4	TOP 100 USA	2017-03-19	More Life	21	22
5	TOP 100 Italy	2019-11-17	23 6451	20	20

References

❑ Kaggle Spotify Dataset :
   https://www.kaggle.com/pepepython/spotify-huge-database-daily-charts-over-3-years?select=Database+to+calculate+popularity.csv

❑ Spotify Web API : https://developer.spotify.com/documentation/web-api/

❑ WikiData : https://www.wikidata.org/wiki/Wikidata:Main_Page

❑ MusicBrainz : https://musicbrainz.org/

Thank you for the attention