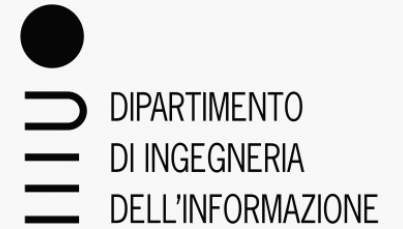




UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Master's in Computer Engineering

Database 2



Team Project 2 : Property Graphs



MARCO ALECCI , LUCA MARTINELLI, ELIA ZIROLDO

ACADEMIC YEAR 2021/2022

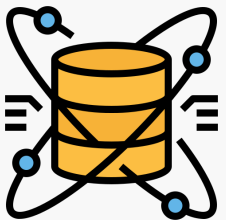
Data Recap



Spotify's Monthly Top 100 charts



35 countries from 2017-2020



Data enriched using WikiData and MusicBrainz APIs

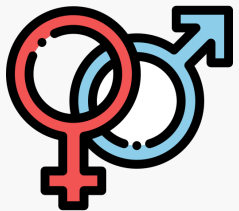
Additional Data



Record Labels which have contracts with artists

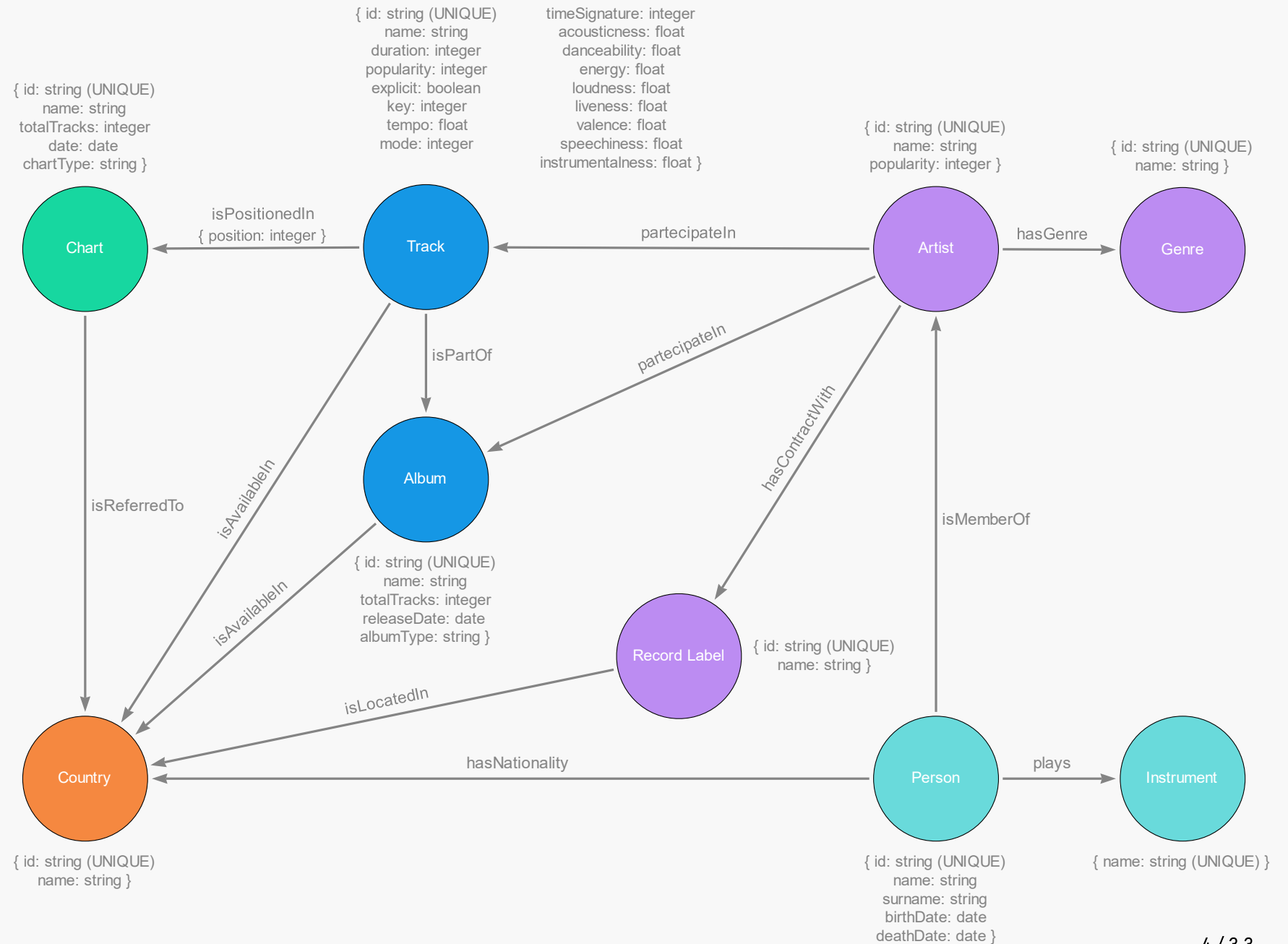


Instruments played by artists



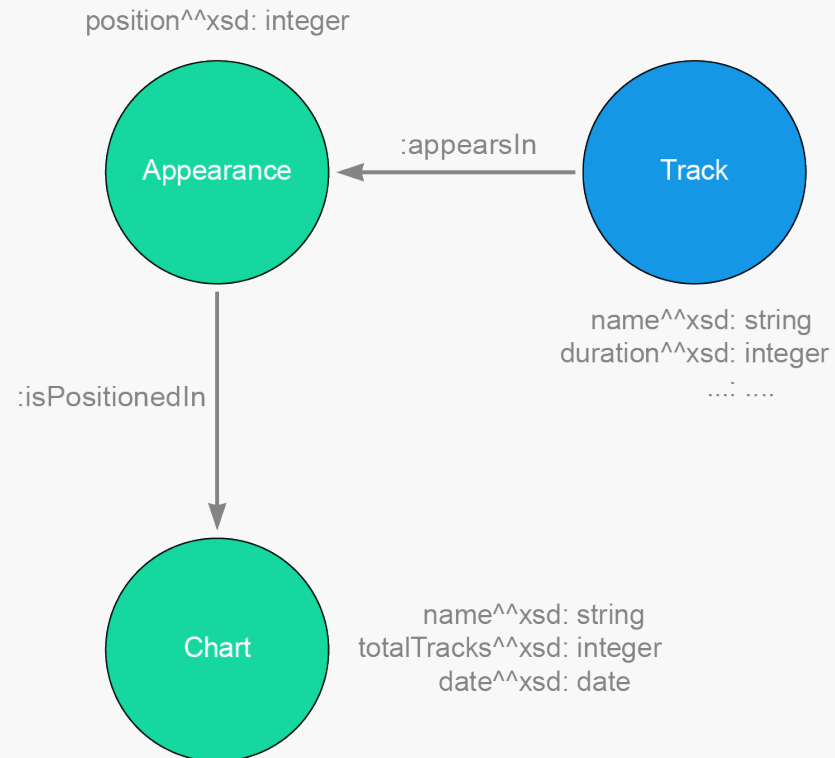
Gender of the artists

Model

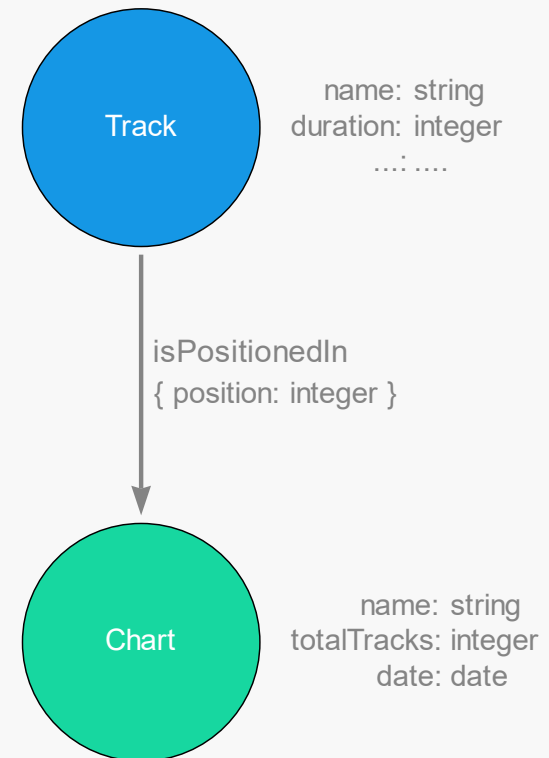


Model : Focus 1

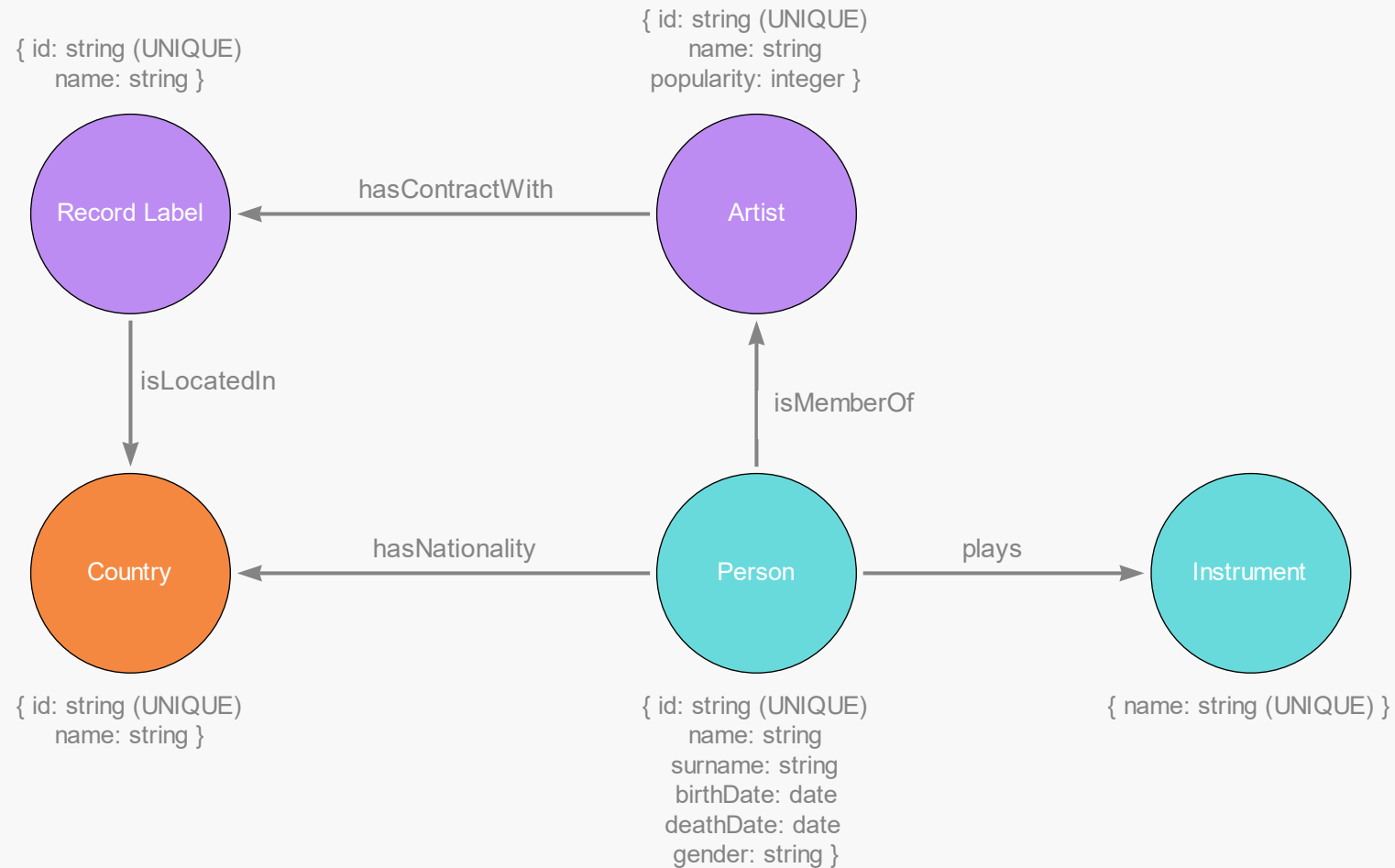
RDF



Property Graphs



Model : Focus 2

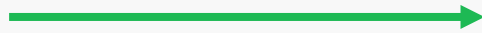




Ingestion Phase

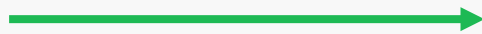
Ingestion Phase

MERGE



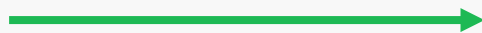
Check if already exists

CREATE



Not uniqueness check

CREATE +
UNIQUE CONSTRAINTS



Indexes created for the IDs

LOAD_CSV
with customized CSV files



Less control, but faster

Ingestion Phase Analysis

Tested with Intel i5 5200U CPU @ 2.20GHz, 2195 Mhz and 8GB of RAM

Method	Genres	Countries	Artists	Albums	Tracks	Charts	People
MERGE	16s	7s	3m 21s	650m	24H+	/	/
CREATE	16s	6s	3m 01s	516m	24H+	/	/
CREATE + UNIQUE CONSTRAINTS	15s	4s	2m 39s	183m	317m	77m	2m 41s
LOAD_CSV	4s	5s	8s	1m5s	4m35s	53s	9s

Ingestion Phase Reference

Analysis of different importing methods



Jiuyuan Wang

Geologist | Data-driven scientist
| Bilingual storyteller

📍 New Haven, CT

🌐 Website

✉ Email

🌐 LinkedIn

🐙 Github

👤 Google Scholar

	CREATE	LOAD CSV	Batch Inserter	Batch Importer	Neo4j-import
When to use?	1 - 10,000 nodes	10,000 - 100,000 nodes	> 1,000,000 nodes	> 1,000,000 nodes	> 1,000,000 nodes
Speed	very slow (1000 nodes/s)	slow (5000 nodes/s)	fast (10,000 nodes/s)	fast (> 10,000 nodes/s)	fast (> 10,000 nodes/s)
Advantages	convenient, real time import	real-time import, pre-load local/server CSV	fast	Based on Batch Inserter, execute complied JAR, direct import data from database	official release, cost less resources than Batch Inserter
Disadvantages	extremely slow	slow, need to convert to CSV	use only in JAVA, need to convert to CSV first, have to stop neo4j while importing	convert to CSV first, not real-time	convert to CSV first, not real-time, can only import into new database (NO for existing database)

Ref: <https://julianwangnwu.github.io/posts/2019/06/import-neo4j/>



Queries

Queries

1

Example Queries

Show how it is possible to use the new added information such as *record label* and *instruments*

2

Italian Tracks and Italian Artists

Study Italian tracks and artists present in the TOP 100 Italy

3

Italian Music Abroad

Discover if Italian tracks are listened also outside Italy



Queries Part 1: Example Queries

Top 3 record labels according to the ratio: $\frac{\text{average position reached}}{\text{number of tracks released}}$



Average Position: 50th
Number of Tracks: 2255



Average Position: 47th
Number of Tracks: 1217



Average Position: 51st
Number of Tracks: 1186

Most common instruments played by members of rock groups



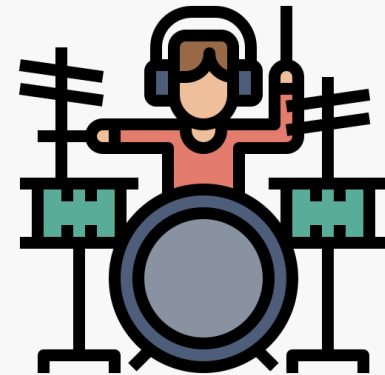
Guitar

159 artists



Voice

127 artists



Drum Kit

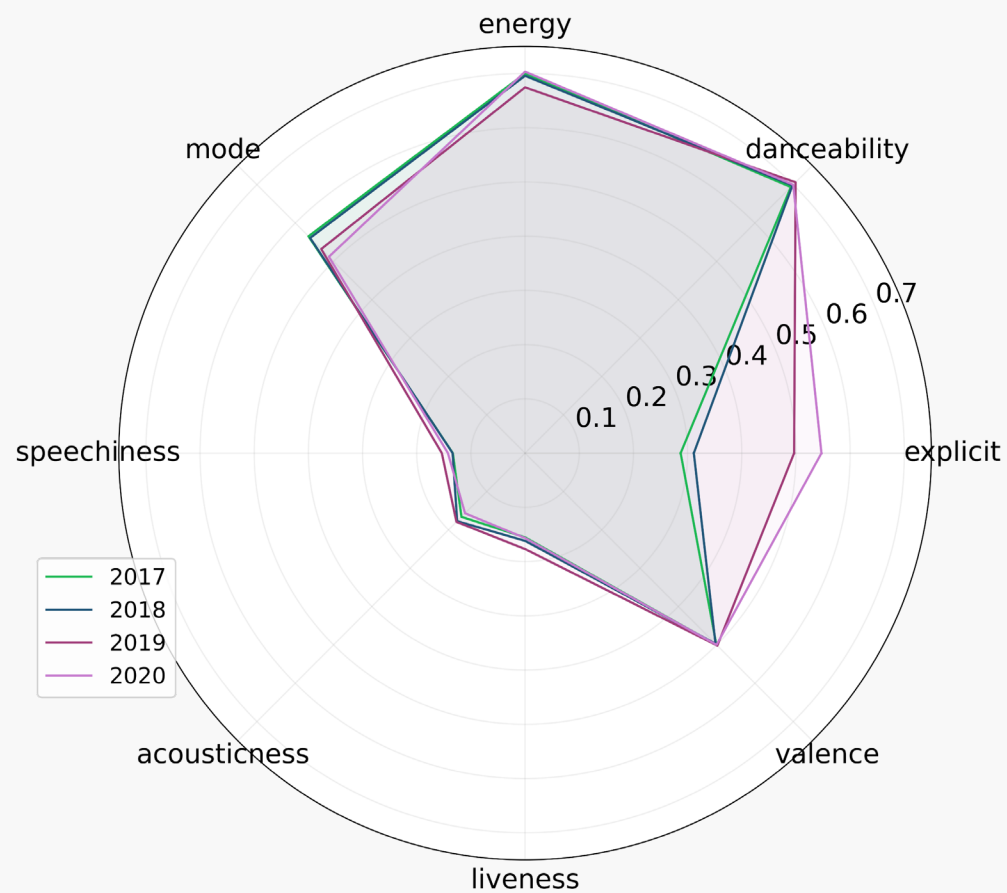
67 artists



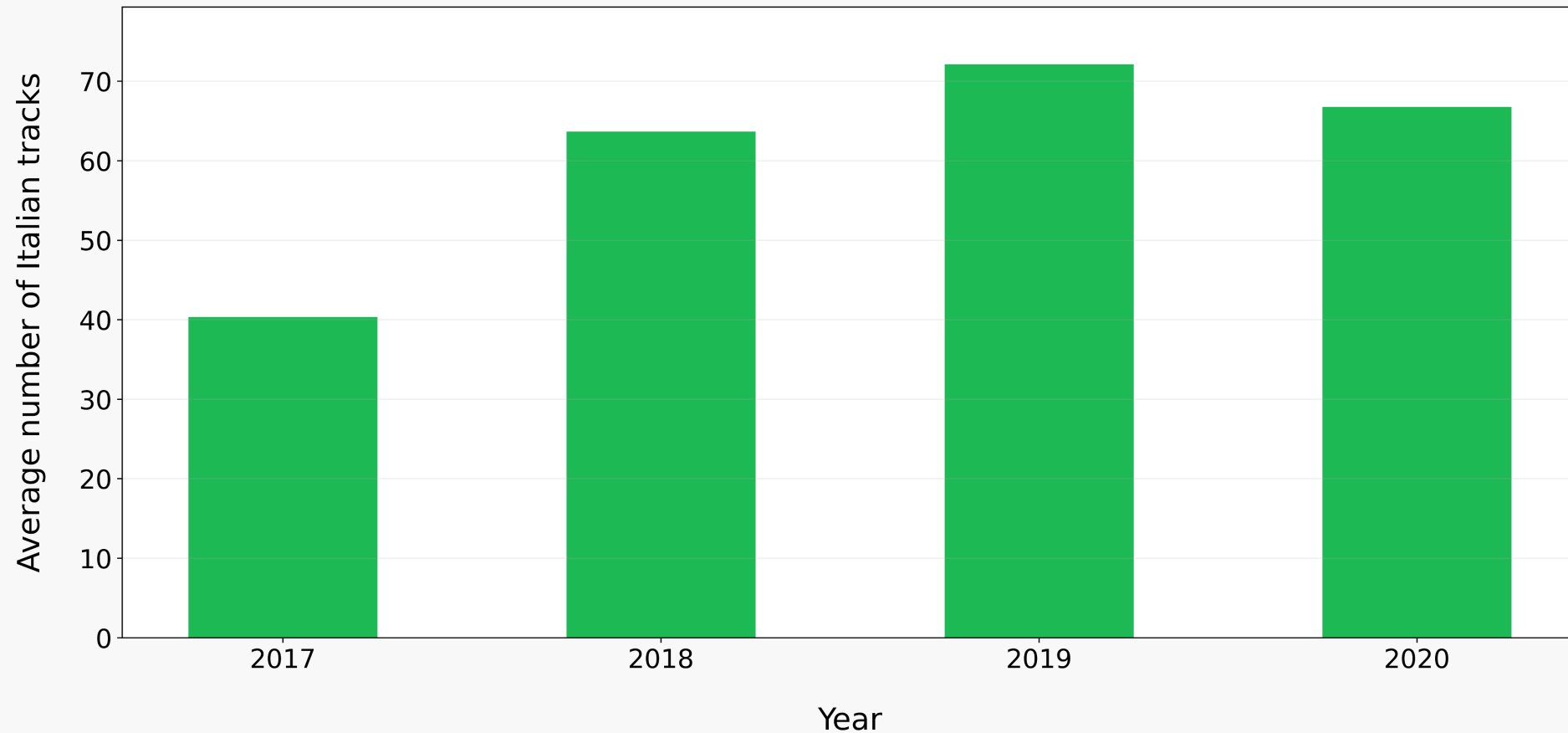
Queries Part 2:

Italian Tracks and Italian Artists

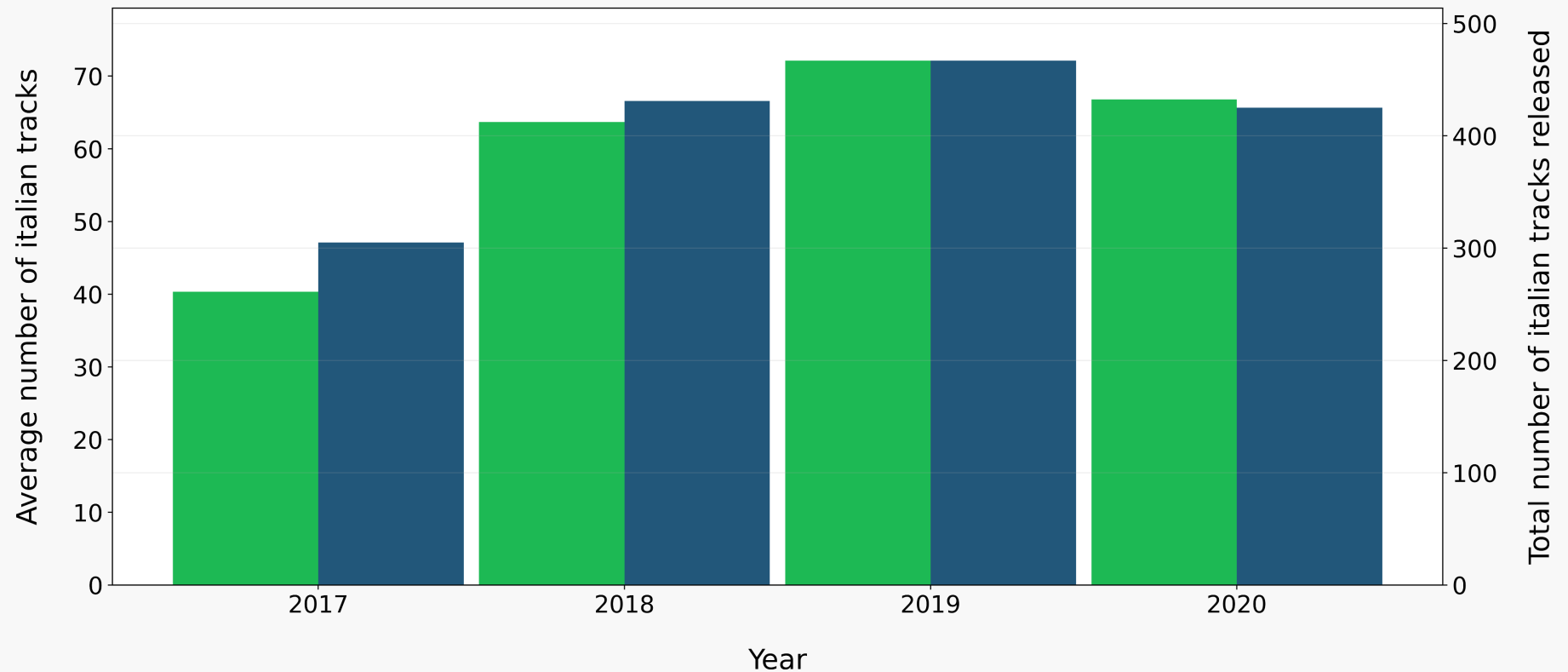
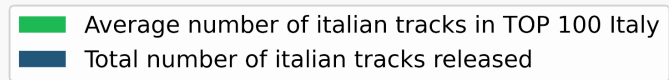
Technical features of Italian Tracks through the years 2017-2020



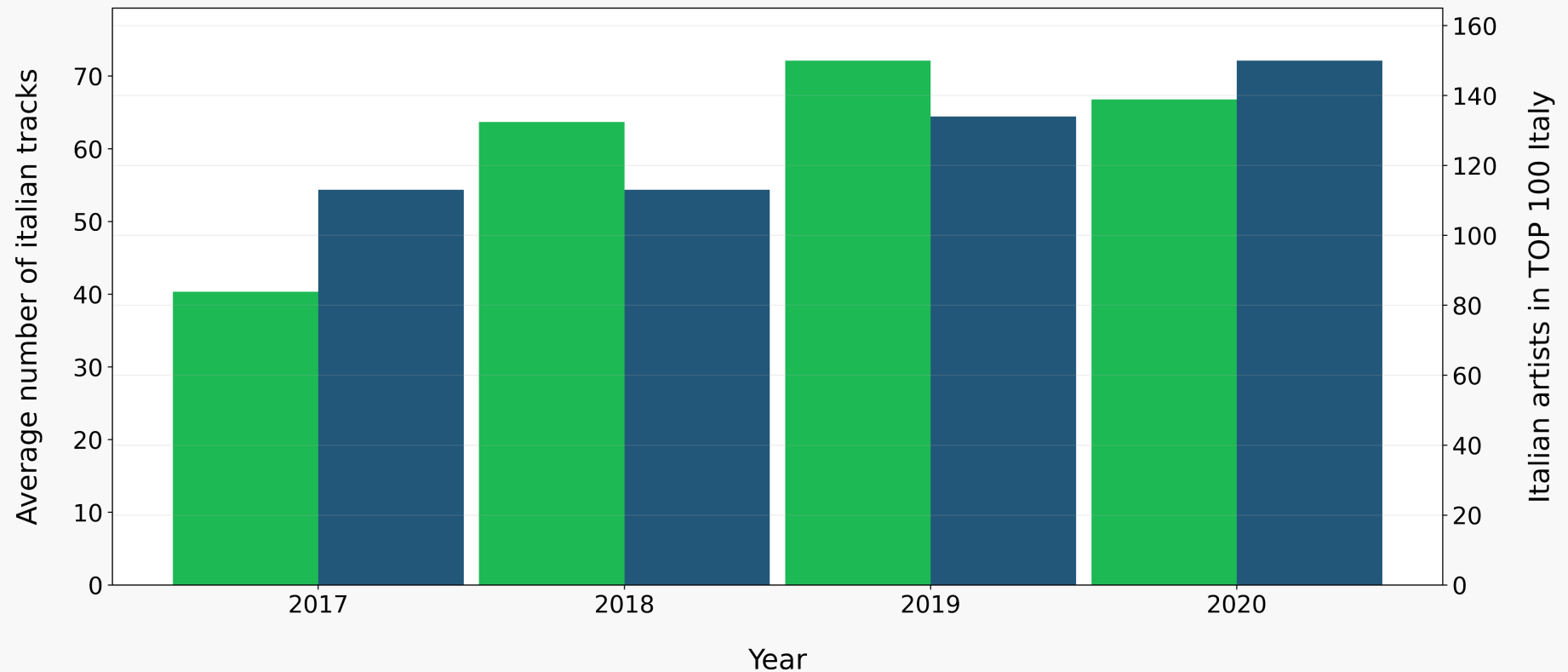
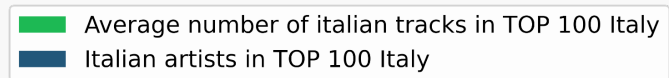
Average number of Italian tracks present in Top 100 Italy for each year



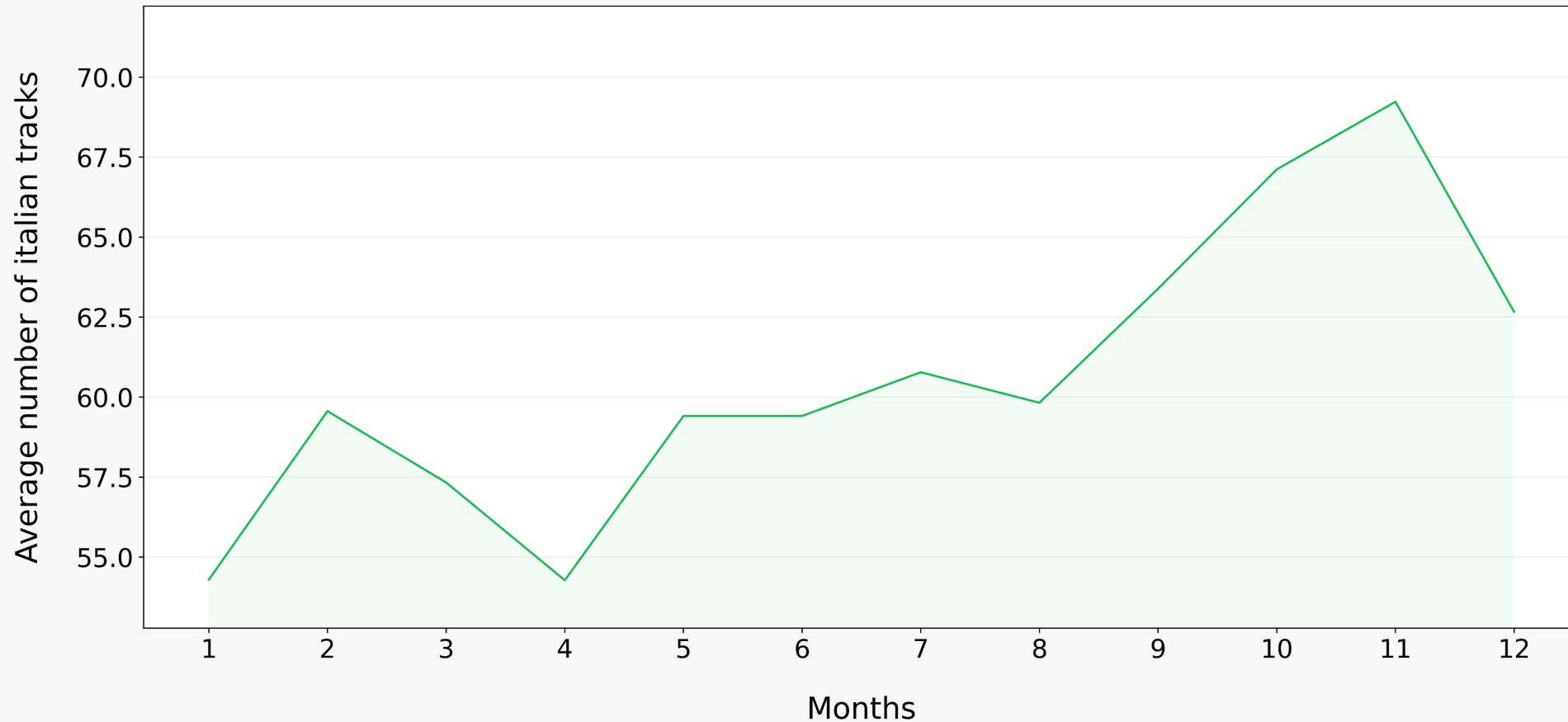
Italian tracks released every year from 2017 to 2020



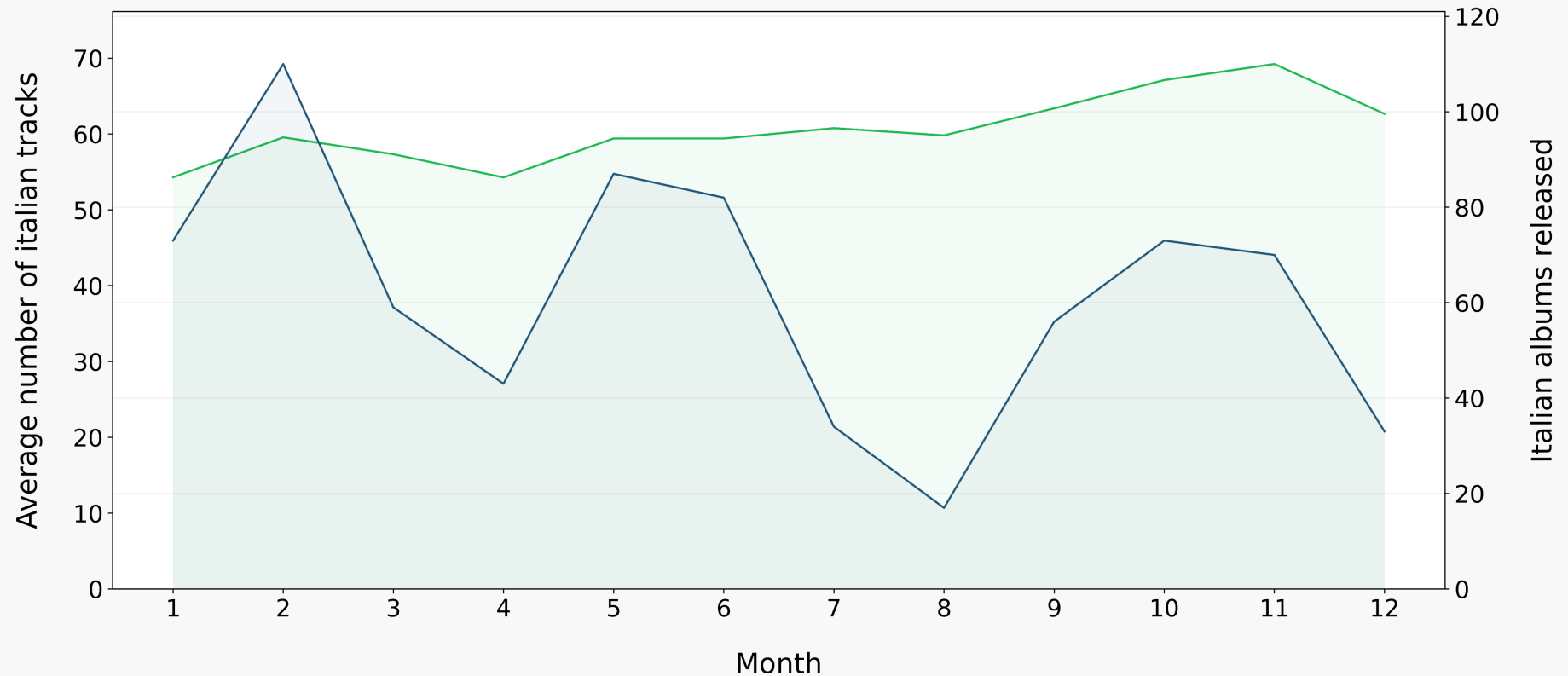
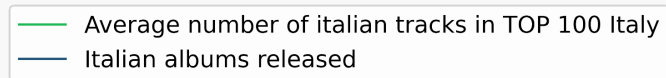
Italian artist entered in Top 100 Italy for each year



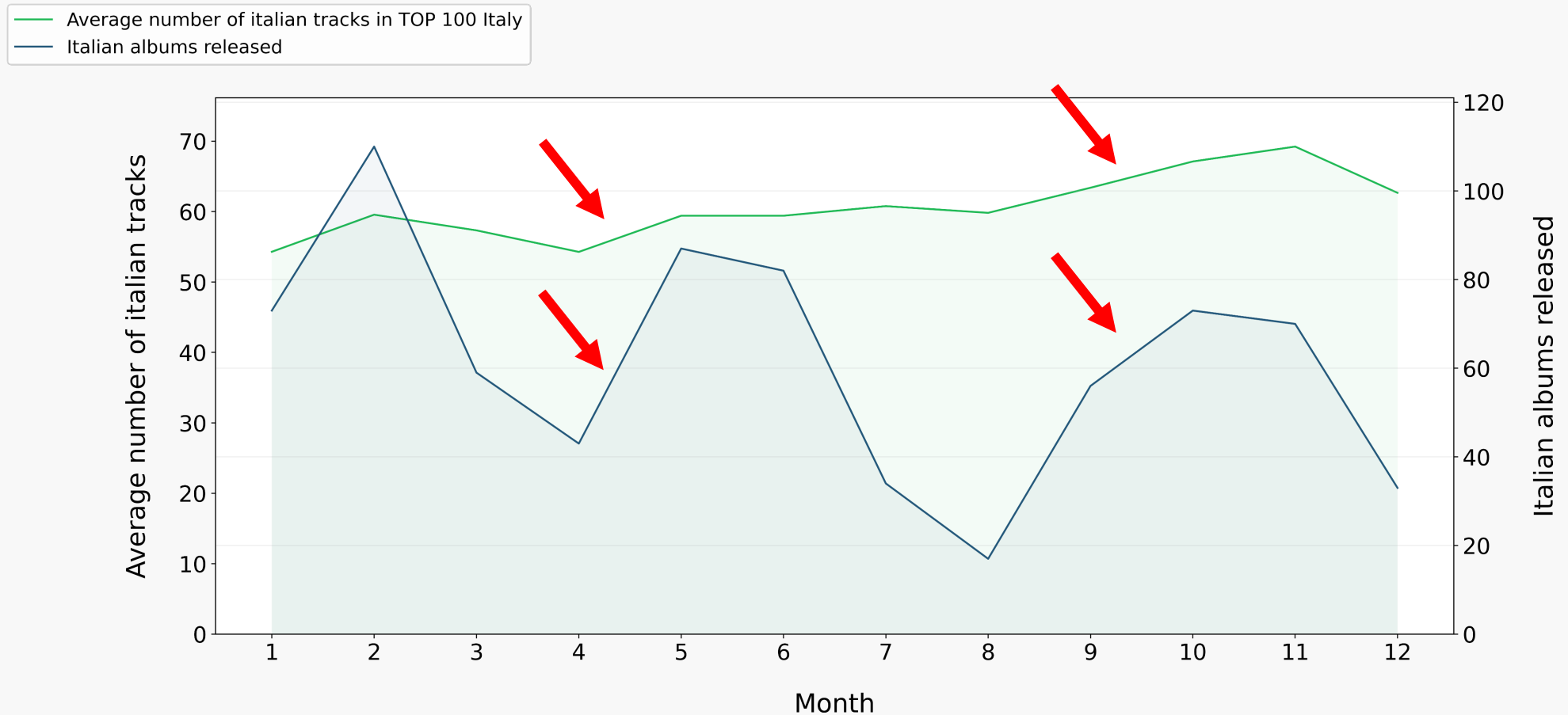
Average number of Italian tracks in Top 100 Italy through months of the year



Albums from Italian artists released through the different months of the year

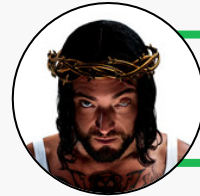


Albums from Italian artists released through the different months of the year



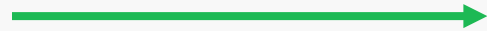
Italian artists with the highest number of tracks in Top 100 Italy for each year

2017



Guè, 30 Tracks

2018



Gemitaiz, 31 Tracks

2019



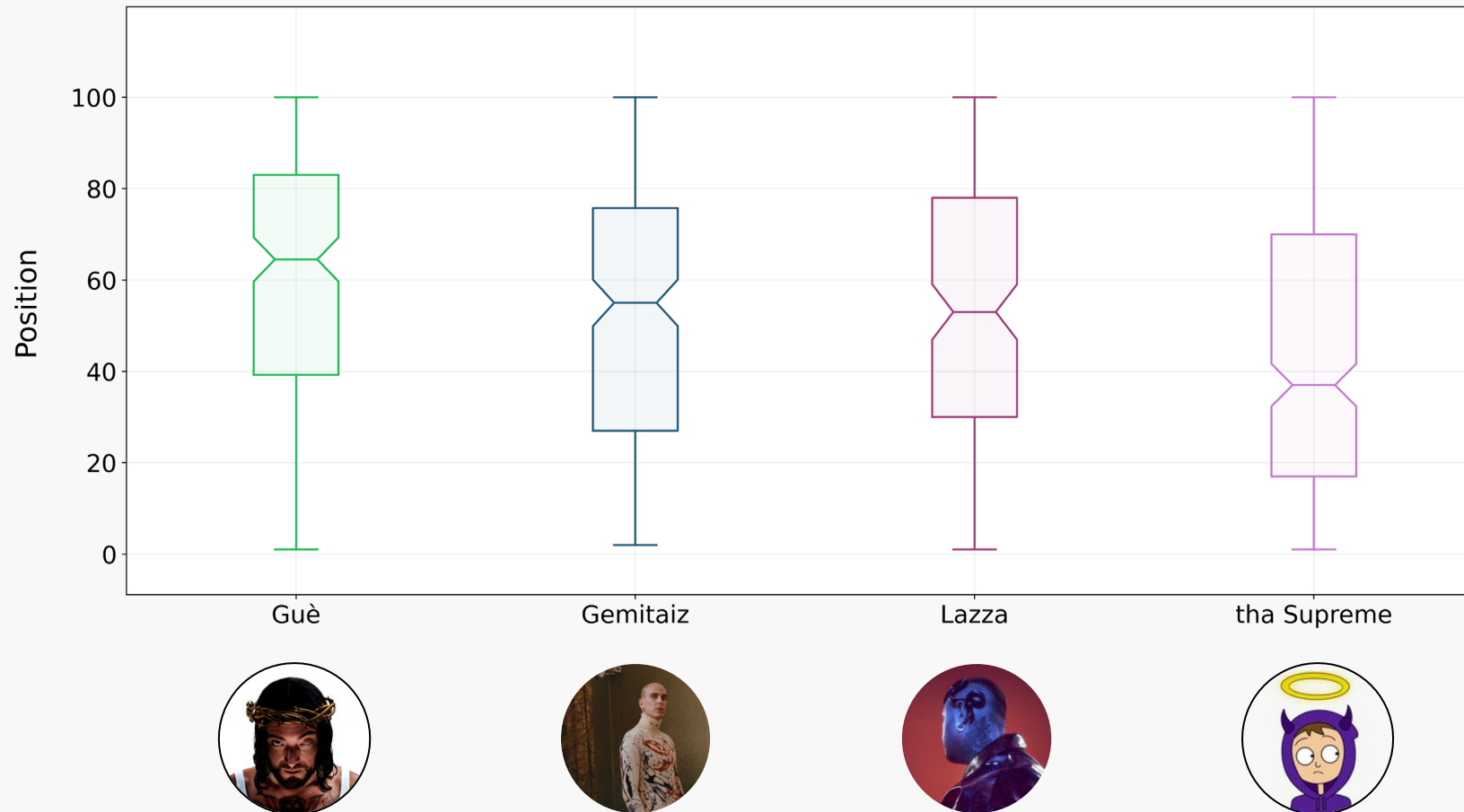
Lazza, 35 Tracks

2020



Tha Supreme, 40 Tracks

Distribution of the positions in the TOP 100 Italy reached by these artists



Top 3 Italian artists with the highest number of tracks in Top 100 Italy at the same time



Tha Supreme
17th Nov 2019
24 tracks



Tедуa
14th June 2020
21 tracks



Marracash
17th Nov 2019
19 tracks



Check if these artists released an album on the same month (or previous one)



Tha Supreme
17th Nov 2019
24 tracks



"23 6451"
15th Nov 2019



Tedua
14th June 2020
21 tracks



"Vita Vera Mixtape"
5th June 2020



Marracash
17th Nov 2019
19 tracks



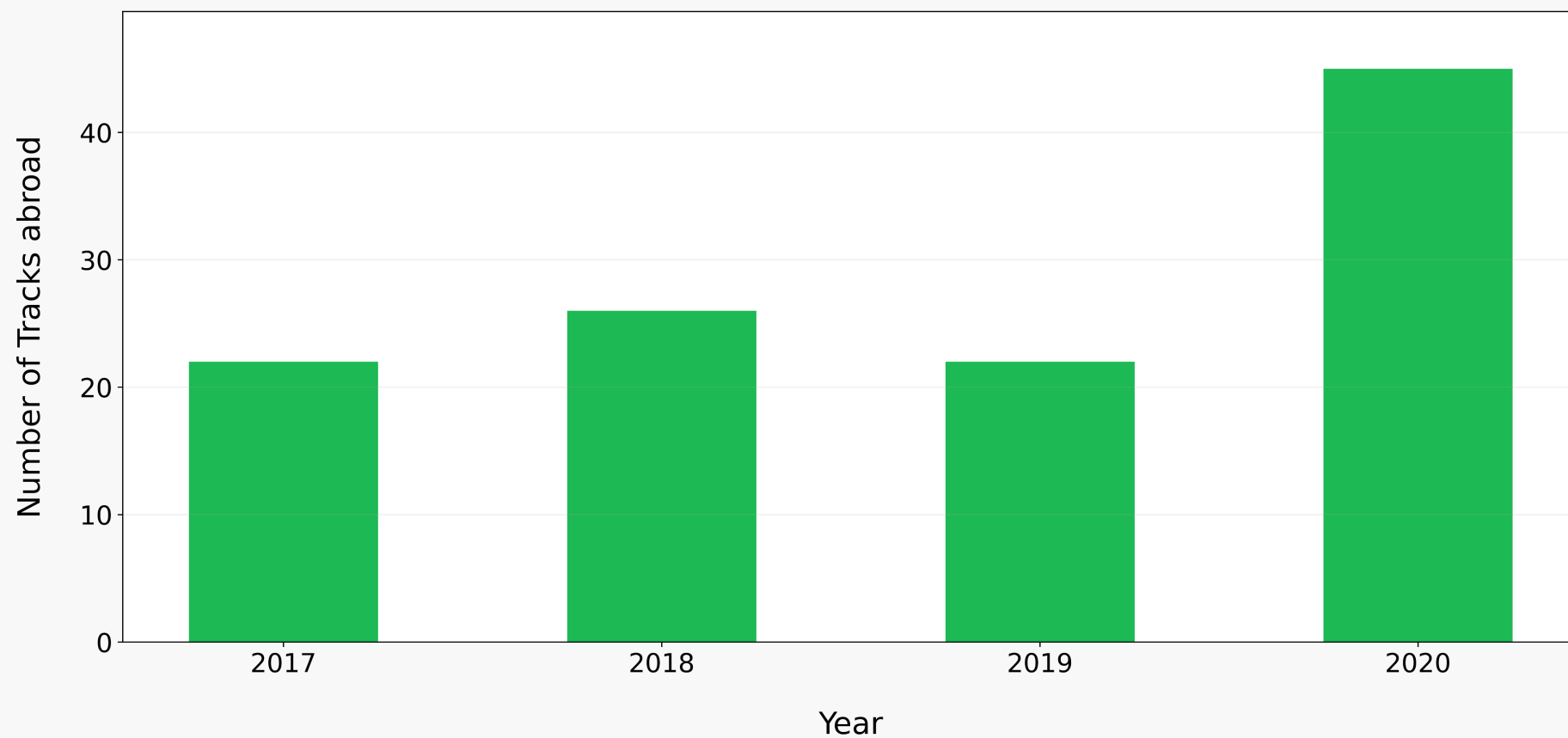
"Persona"
31th Oct 2019



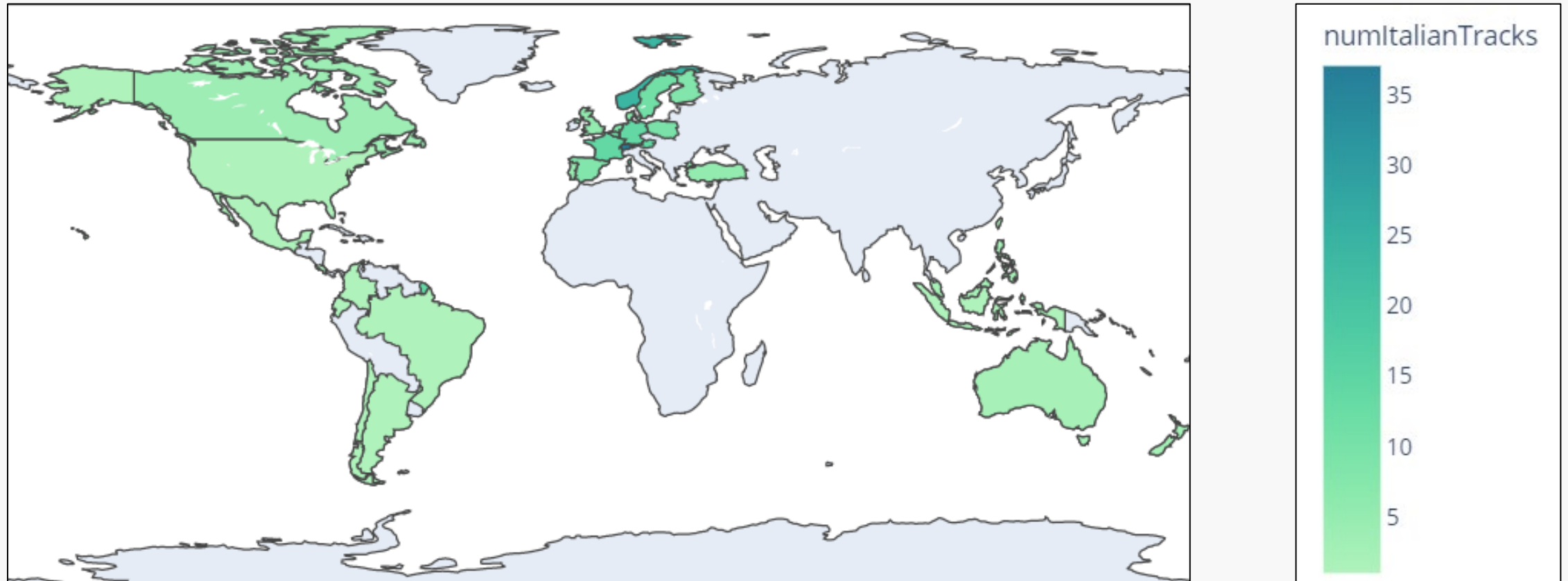
Queries Part 3:

Italian Music Abroad

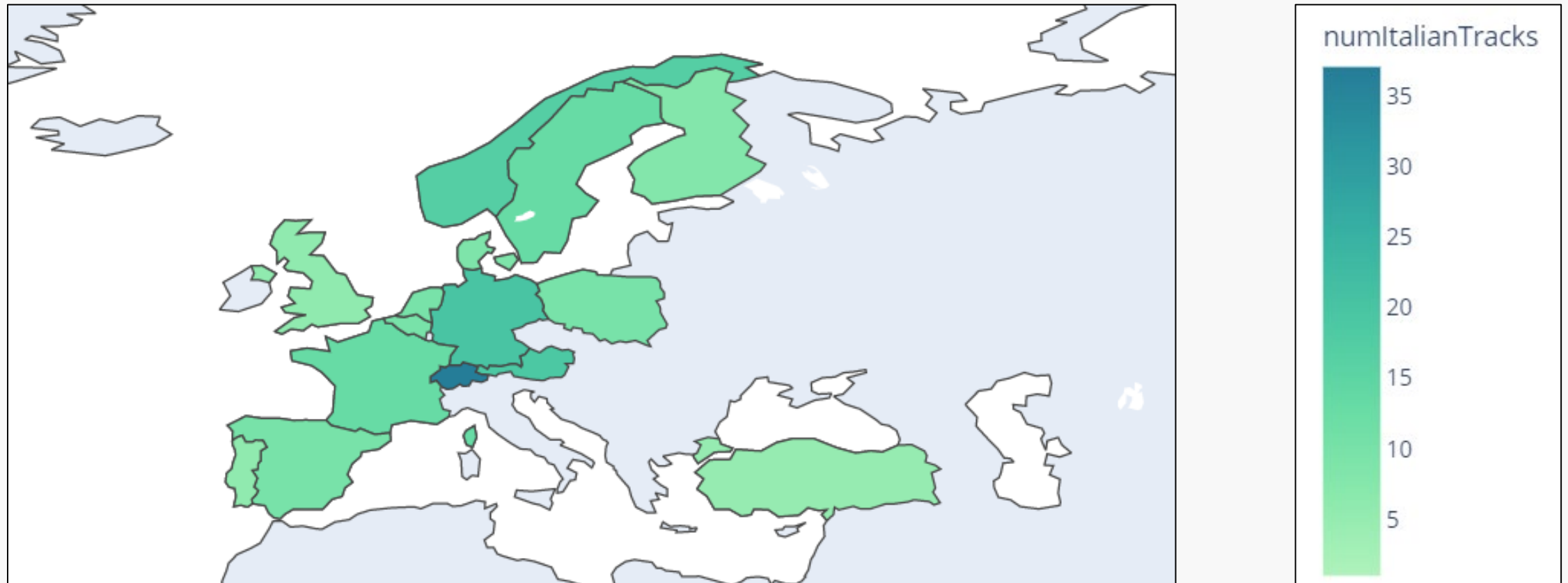
Italian tracks in a Top 100 of a country different from Italy for each year



Countries that listen the most to Italian tracks



Countries that listen the most to Italian tracks: Europe Focus





How to determine the most widespread Italian artist?

❑ 1st Approach:

Count, for each artist, how many different countries he has reached excluding Italy.

❑ 2nd Approach:

Count, for each artist, how many of his tracks have entered in a TOP 100 of a country different from Italy

Most widespread Italian artist abroad

1st Approach:



Gigi D'Agostino

30 different countries excluding Italy

2nd Approach:



Sfera Ebbasta

12 tracks in a TOP 100 of a country different from Italy



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/>
- ❑ Neo4j Ingestion Analysis: <https://julianwangnwu.github.io/posts/2019/06/import-neo4j/>

Thank you for the attention