

GRAPH DATABASES PROJECT

MELODY PROJECT

Di Martino Ludovico
Galli Filippo
Rigobello Manuel

NEXT



PROJECT AGENDA

1. OBJECTIVES
AND DOMAIN

2. DATASET

3. GENRES
AND CHATGPT

4. ONTOLOGY
CREATION

5. MAPPING
RDF DATASET

6. SHACL

7. QUERIES &
ANALYTICS

8. CONCLUSION
AND FUTURE
WORK



ANALYSIS OF POPULAR SONGS FROM 1962 TO 2018

OBJECTIVES AND DOMAIN

Project Goal: Build a graph database containing information about the most popular songs, album and artists, their evolution and correlations with Grammy Awards

Domain: Evolution of the characteristics of the most popular songs belonging to the Billboard Hot 100 between 1962 and 2018

Focus: Investigate trends in music popularity, attributes, and artist career



DATASET



GRAMMY DATASET

Contains all the Grammy Awards and categories from 1959 to 2019

BILLBOARD TOP 100 DATASET

Includes data on the top 100 songs ranked weekly, with details such as the Spotify song ID, the rank and the week

MUSICOSET DATASET

Provides information about songs, albums, and artists featured in the Billboard Top 100.

For each **song** the dataset includes attributes such as key, liveness, loudness, tempo

For **artists**, it provides data on their followers, genres, and popularity.

For **albums**, the dataset contains information on the album's name, the total number of tracks, and the relationships between the album and the songs it contains



SONG CHART.CSV

song_id	rank_score	week
3zkWCteF82vJwv0hRLba76	71	2016-05-07
3nnG7AM9QopHVPEuLX3Khk	70	2016-05-07
43PuMrRfbyyuz4QpZ3oAwN	69	2016-05-07
00fzPUxPzxpI6ndQ0TXbdg	67	2016-05-07
3CJvmtWw2bJsdubAC5uCQk	66	2016-05-07
4sPm07WMQUAf45kwM0tONw	65	2016-05-07
3pXF1nA74528Edde4of9CC	64	2016-05-07

TRACKS.CSV

song_id	album_id	track_number
1TE4S9kqvm19o8JpPR8kBF	3QtcoigGc0KKzE6mVAkFFY	6
6tMIiFeZMw10B3aPEF14Z2	7ndqT6LH9o6YQ5uA7aT0St	5
1dXYwIIbdDaW5tKFZUeuPb	6NS8uFJqbcKqLUu4VzxpD4	6
61gnKQj1YZm79YZ50ausHT	12WlibQIazyVIlduAqBB15	6
4sPm07WMQUAf45kwM0tONw	0K4pIO0sfJ91K80jrZfxzd	11
5oRCENOCYft0g0PdS0MnEG	2DdbTYjNxtex2id1tmPw08C	7
0QpZ69CTjRbJ3ZVsntaGfa	0gC5d1rbDIOAx40R5pG4MW	19

ALBUMS.CSV

ALBUM_ID	NAME	BILLBOARD_ARTISTS	POPULARITY	TOTAL_TRACKS	ALBUM_TYPE

ACOUSTIC FEATURE.CSV

SONG_ID DURATION_MS KEY
MODE TIME_SIGNATURE
ACOUSTICNESS DANCEABILITY
ENERGY INSTRUMENTALNESS
LIVENESS LOUDNESS
SPEECHINESS VALENCE TEMPO

SONGS.CSV

song_id	song_name	billboard	artists	popularity
1mea3bSkSGXuIRvnydlB5b	Viva La Vida	('Viva La Vida', 'Coldplay')	{'4gzpq5DPGxSnKTe4SA8HAU': 'Coldplay'}	77
7iN1s7xHE4ifF5povM6A48	Let It Be - Remastered 2009	('Let It Be', 'The Beatles')	{'3WrFJ7ztbogyGnTHbHJF12': 'The Beatles'}	77
40ri0y7x9W7GXjyGp4pjAv	Hotel California - 2013 Remaster	('Hotel California', 'Eagles')	{'0ECwFtbIWEVNwjlrfc6xoL': 'Eagles'}	81
7x8dCjCr0x6x21XKujYD34	The Pretender	('The Pretender', 'Foo Fighters')	{'7jy3rLJdDQY210gRLCZ9sD': 'Foo Fighters'}	75
61gnKQj1YZm79YZ50ausHT	Another Brick in the Wall	('Another Brick In The Wall', 'Pink Floyd')	{'0k17h0D3J5VfsdmQ1iztE9': 'Pink Floyd'}	78
4sPm07WMQUAf45kwM0tONw	Hello	('Hello', 'Adele')	{'4dpARuHxo51G3z768sgnrY': 'Adele'}	69
			False	Solo

THE GRAMMY AWARDS.CSV

year	category	nominee	workers	winner
1965	Best New Artist	The Beatles (artist)	,	True
1981	Best Engineered Album Non Classical	The Wall.	"Pink Floyd (artists), James Guthrie (engineer)"	, True
1989	Song of the Year	"Don't Worry, Be Happy."	Bobby McFerrin (artist)	, True
1989	Song of the Year	Fast Car.	Tracy Chapman	, False
2014	Best Dance/electronic Album	Random Access Memories.	Daft Punk	, True
2017	Song of the Year	Hello	Adele, Greg Kurstin	, True

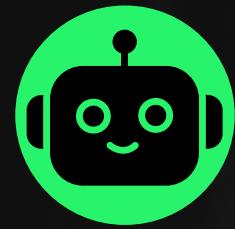
ARTISTS.CSV

ARTIST_ID	NAME	FOLLOWERS	POPULARITY	ARTIST_TYPE	MAIN_GENRE	GENRES

GENRES AND CHATGPT



approximately 1500 unique genres classified into 32 different macro-genres



Category discrepancies compared to official DBpedia categories

ChatGPT to map CSV categories to official categories and grouped them into 13 macro-categories with official names

SONG OF THE YEAR

SONG_OF_THE_YEAR

BEST NEW ARTIST
OF 1962

BEST_NEW_ARTIST

BEST GOSPEL
PERFORMANCE, MALE

BEST_GOSPEL_VOCAL_
PERFORMANCE_MALE

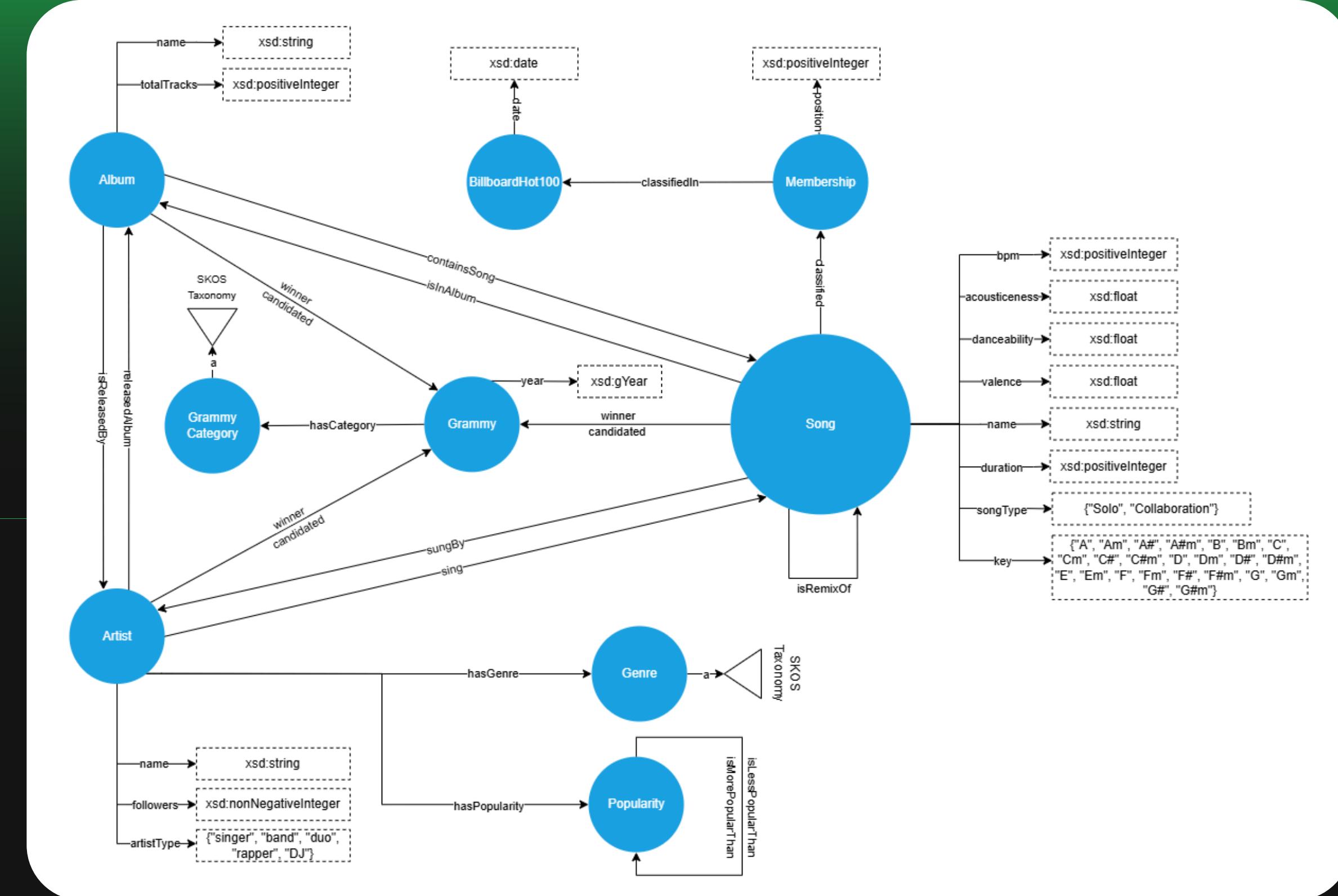
BEST ALBUM
(CLASSICAL)

BEST_CLASSICAL_ALBUM

CLASSICAL ALBUM OF
THE YEAR

BEST_CLASSICAL_ALBUM

ONTOLOGY CREATION



MAPPING OPEN DATA TO RDF



- Find the remixes based on song titles
- Associate the song with nominees of a Grammy



Create triple if Grammy category contains:

- “artist”: check nominee if workers is empty
- “producer”: check nominee if workers is empty
- “performance”: check nominee and workers must be empty



- Exclusion of albums without a Billboard Hot 100 song
- Link to the Grammy if the category includes “album” and check the nominee if worker is not empty
- Address homonymity by comparing workers



Name individual based on hash of year and category



SHACL VALIDATION



Shapes for all the fundamental resources of the ontology

Song · Artist · Album · Grammy · Billboard · Membership



Helpful for identifying dataset errors

We discovered tracks in the song dataset that were not included in the Billboard dataset



```
mel:SongShape
  a sh:NodeShape ;
  sh:targetClass mel:Song ;
  sh:property [
    sh:path mel:name ;
    sh:datatype xsd:string ;
    sh:minCount 1 ;
    sh:maxCount 1 ;
  ];
  sh:property [
    sh:path mel:acousticness ;
    sh:datatype xsd:float ;
    sh:maxCount 1 ;
    sh:minInclusive 0.0 ;
    sh:maxInclusive 1.0 ;
  ];
  sh:property [
    sh:path mel:danceability ;
    sh:datatype xsd:float ;
    sh:maxCount 1 ;
    sh:minInclusive 0.0 ;
    sh:maxInclusive 1.0 ;
  ];
  sh:property [
    sh:path mel:classified ;
    sh:minCount 1 ;
  ];
  {...} # Omitted Code

  sh:closed true ;
  sh:ignoredProperties ( rdf:type owl:topDataProperty
owl:topObjectProperty mel:key mel:isRemixOf mel:isInAlbum mel:winner
mel:candidated ) ;
.
```



QUERIES AND ANALYTICS



Average song duration



Songs by key

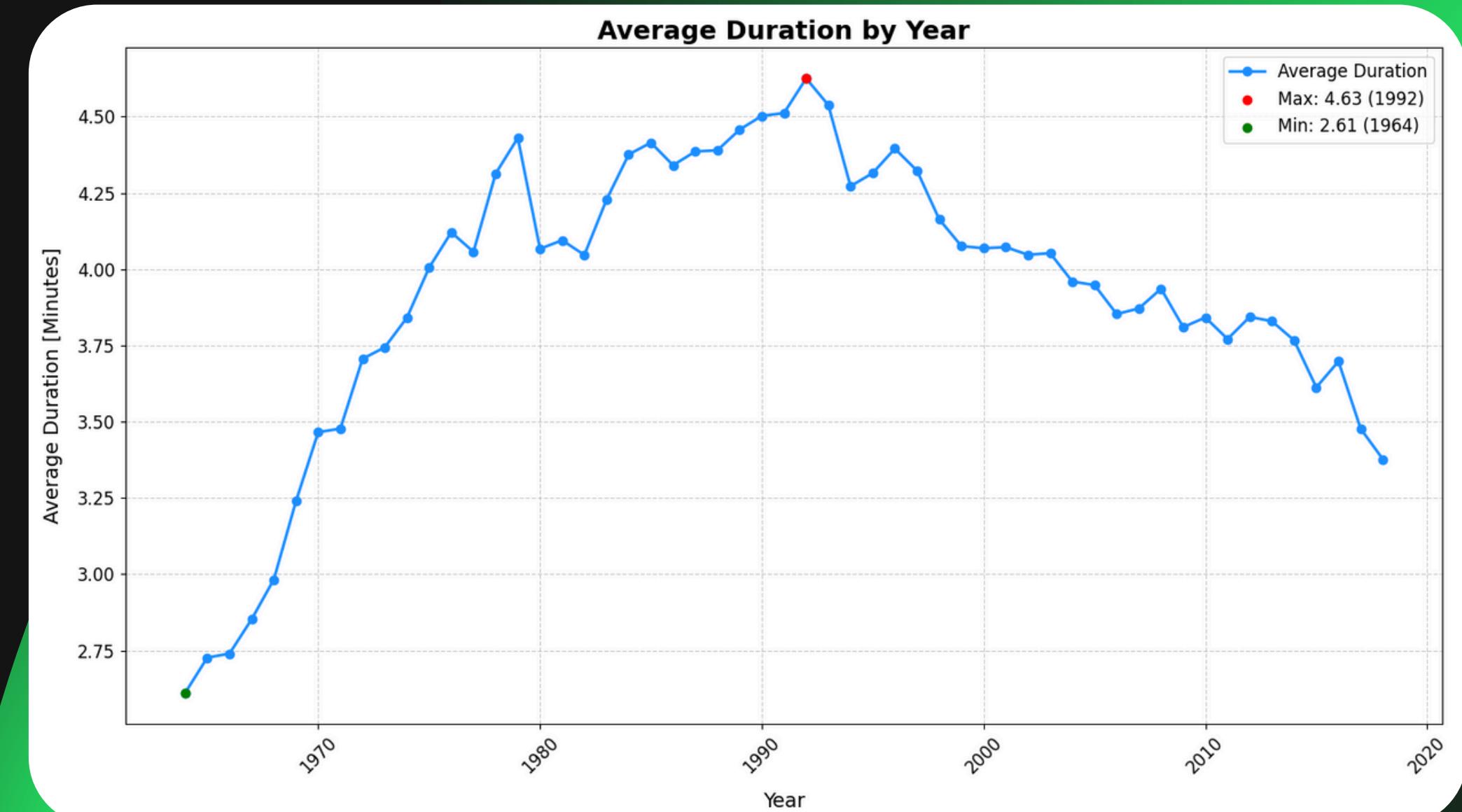


One-Hit Wonders



AVERAGE SONG DURATION

```
● ● ●  
PREFIX mel: <http://www.dei.unipd.it/~gdb/ontology/melody#>  
  
SELECT ?year ((AVG(?duration)/1000)/60 AS ?avgDurationMinutes) WHERE {  
  {  
    SELECT ?song (MIN(?billboardYear) AS ?year) WHERE {  
  
      ?song a mel:Song ;  
             mel:name ?songTitle ;  
             mel:classified ?membership .  
  
      ?membership a mel:Membership ;  
                 mel:classifiedIn ?billboard .  
  
      ?billboard a mel:BillboardHot100 ;  
                 mel:date ?date .  
  
      BIND(YEAR(?date) AS ?billboardYear) .  
  
    } GROUP BY ?song  
  }  
  ?song mel:duration ?duration .  
} GROUP BY ?year  
ORDER BY DESC (?year)
```

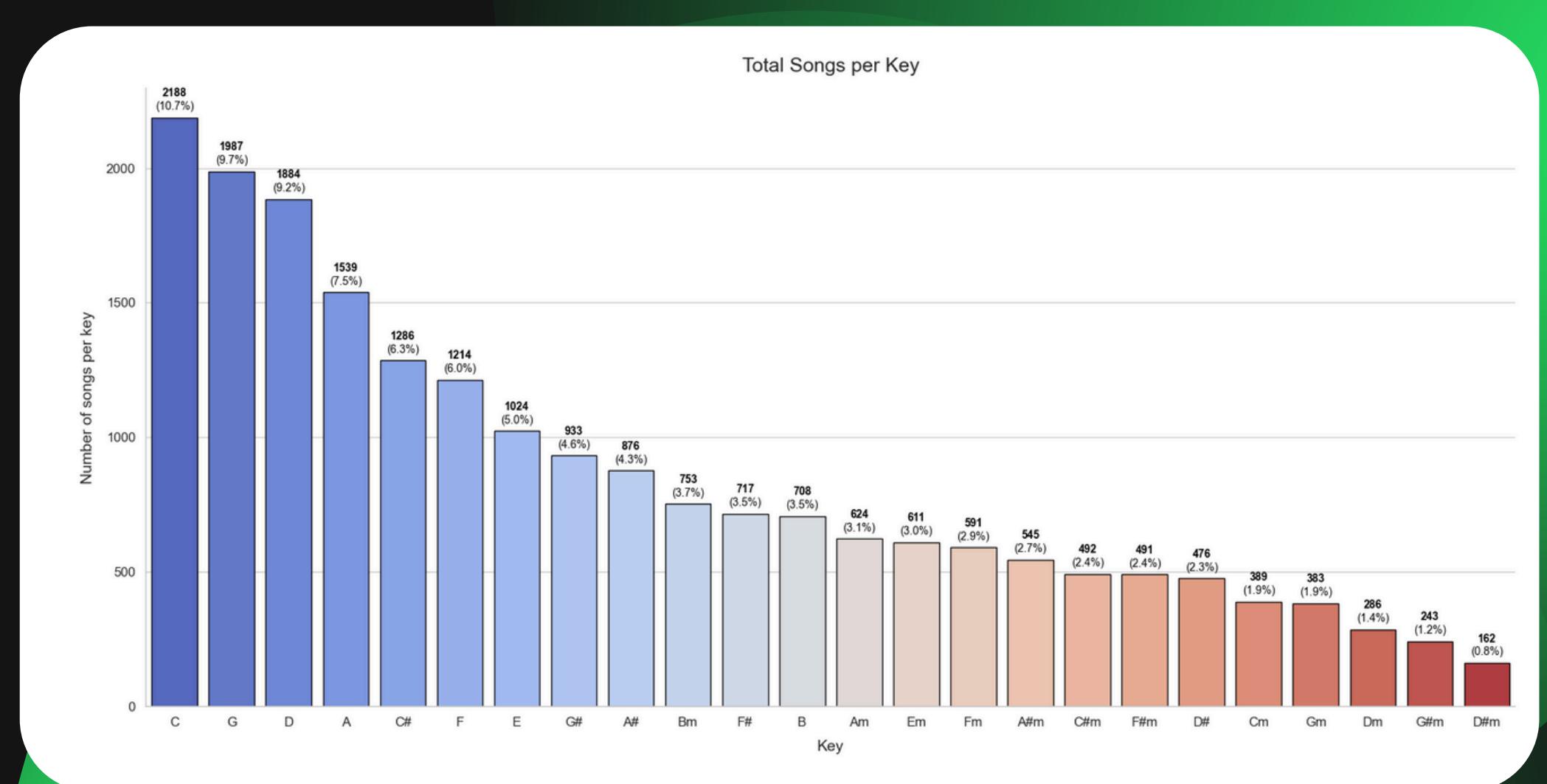


SONGS BY KEY



PREFIX mel: <http://www.dei.unipd.it/~gdb/ontology/melody#>

```
SELECT ?key ?count (((?count*100)/?tot_songs) AS ?percentage) WHERE {
  { SELECT ?key (COUNT(*) AS ?count) WHERE {
    ?song a mel:Song ;
    mel:key ?key .
  }
  GROUP BY ?key
  }
  { SELECT (COUNT(*) AS ?tot_songs) WHERE {
    ?song a mel:Song .
  }
  }
}
ORDER BY DESC (?count)
```



ONE-HIT WONDERS



```
PREFIX mel: <http://www.dei.unipd.it/~gdb/ontology/melody#>

SELECT ?artistName ?songName ?position WHERE {
  ?song mel:name ?songName ;
    mel:sungBy/mel:name ?artistName ;
    mel:classified ?membership .

  ?membership mel:position ?position .

  FILTER (?position < 51) .

  FILTER NOT EXISTS {
    ?song2 mel:sungBy/mel:name ?artistName ;
      mel:classified ?membership2 .

    FILTER (?song2 != ?song) .

  } FILTER NOT EXISTS {
    ?song mel:name ?songName ;
      mel:sungBy/mel:name ?artistName ;
      mel:classified ?membership3 .

    ?membership3 mel:position ?position2 .

    FILTER (?membership3 != ?membership) .
  }
}

GROUP BY ?artistName ?songName ?position
ORDER BY(?position)
```

Song Title	Artist Name	Peak Position
We Might Be Dead By Tomorrow	Soko	9
Wasted Love	Matt McAndrew	14
Hey soul sister	Glee Cast Karaoke's band	29
3AM (as made famous by Eminem)	Radio Killers	32
My Baby's Got A Smile On Her Face	Craig Wayne Boyd	34
(There's No Place Like) Home for the Holidays	Mitchell Ayres & His Orchestra	41
Wonderful Summer	Robin Ward	43
Hold Up My Heart	Brooke White	47
Wonderful Christmastime	Jimmy Fallon	47
White Christmas	Ken Darby Singers & Bing Crosby & John Scott Trotter & His Orchestra	48
Christmas (Baby Please Come Home)	Darlene Love	50



CONCLUSIONS AND FUTURE WORK



MELODY Graph Database

Graph database containing information about the most popular songs, albums, and artists, their evolution and correlations with Grammy Awards



Improve String Matching

The associations between the songs, albums, and artists with the Grammy dataset could be improved



THANK YOU!

Questions?



Di Martino Ludovico
Galli Filippo
Rigobello Manuel

