

Graph Databases A.Y. 2024-2025
Task 4 – SPARQL Queries And Analytics
Master Degree in Computer Engineering

Deadline: 10 January, 2025

Group Acronym	MELODY	
Github	https://github.com/ludovicodimartino/MELODY	
Last Name	First Name	Badge Number
Di Martino	Ludovico	2104292
Galli	Filippo	2120826
Rigobello	Manuel	2103374

1 Queries

In the following sections, we describe what we believe are the most interesting queries we performed on our MELODY graph database. Through these queries, our objective was to uncover some insights from the data and gather statistics that could be useful, for instance, to song producers and people working in the music industry.

1.1 Top 10 Songs

This query extracts the most popular songs in the whole database. The popularity of a song is based on the number of times it appears in the first 10 positions of a Billboard.

```
1 PREFIX mel: <http://www.dei.unipd.it/~gdb/ontology/melody#>
2 PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
3
4 SELECT ?title ?artistName (COUNT(?position) AS ?totalAppearancesInTheFirst10Pos) WHERE {
5     ?song a mel:Song ;
6         mel:name ?title ;
7         mel:classified ?membership ;
8         mel:sungBy/mel:name ?artistName .
9
10    ?membership a mel:Membership ;
11        mel:position ?position ;
12        mel:classifiedIn ?billboard .
13
14    FILTER(?position <= "10"^^xsd:positiveInteger) .
15 }
16 GROUP BY ?title ?artistName
17 ORDER BY DESC (?totalAppearancesInTheFirst10Pos)
18 LIMIT 10
```

Song Title	Artist Name	Appearances
Shape of You	Ed Sheeran	33
How Do I Live	LeAnn Rimes	32
Smooth (As Made Famous By Santana Featuring Rob Thomas)	The Karaoke Crew	30
That's What I Like	Bruno Mars	28
Perfect	Ed Sheeran	27
God's Plan	Drake	26
Truly Madly Deeply	Savage Garden	26
Apologize	OneRepublic	25
Counting Stars	OneRepublic	25
Trap Queen	Fetty Wap	25

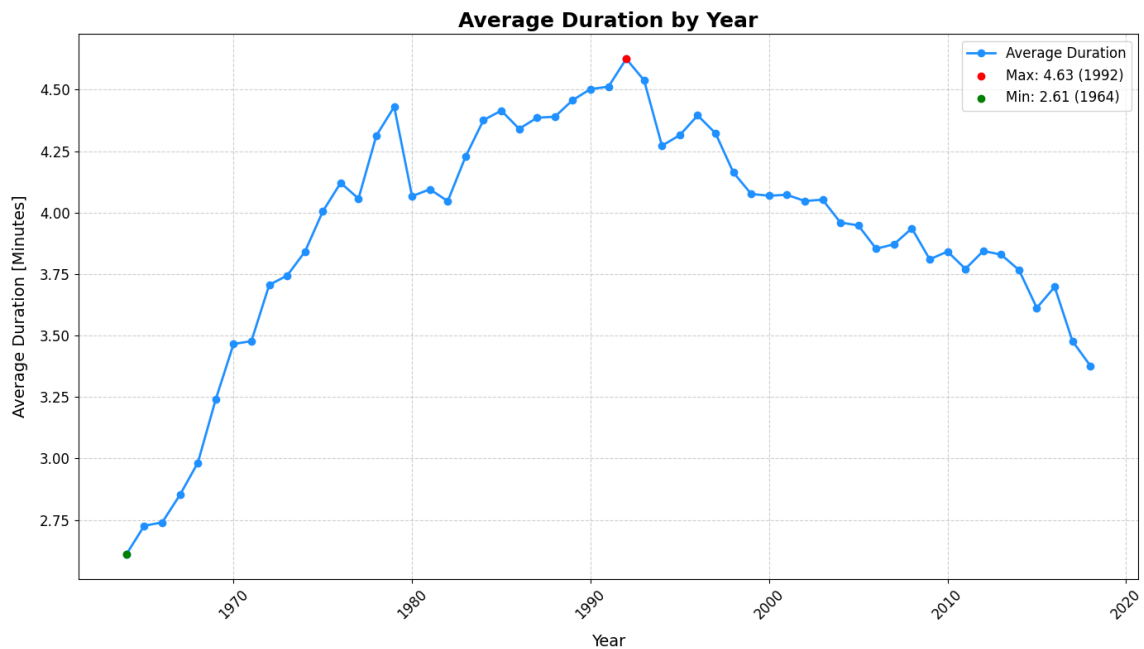
Table 2: Top 10 songs and their number of appearances in the first 10 positions of the Billboard charts.

1.2 Evolution of songs duration over the years

This query examines the evolution of song duration over the years. Specifically, for each year, we calculated the average song duration of the songs featured in the Billboard Hot 100. Since the exact release year of each song is not in the database (see the Ontology design diagram), we used the year in which the song first appeared on the Billboard chart as a proxy for its release year.

```
1 PREFIX mel: <http://www.dei.unipd.it/~gdb/ontology/melody#>
2
3 SELECT ?year ((AVG(?duration)/1000)/60 AS ?avgDurationMinutes) WHERE {
4   {
5     SELECT ?song (MIN(?billboardYear) AS ?year) WHERE {
6       ?song a mel:Song ;
7         mel:name ?songTitle ;
8         mel:classified ?membership .
9
10      ?membership a mel:Membership ;
11        mel:classifiedIn ?billboard .
12
13      ?billboard a mel:BillboardHot100 ;
14        mel:date ?date .
15
16      BIND(YEAR(?date) AS ?billboardYear) .
17    }
18    GROUP BY ?song
19  }
20
21  ?song mel:duration ?duration .
22 }
23 GROUP BY ?year
24 ORDER BY DESC (?year)
```

With the data acquired from this query we could plot a graph that visually shows the evolution of songs duration over the years.



1.3 Songs by key

This query counts the number of songs in each musical key. The results are quite interesting, as we observe a significant number of songs in the key of C and very few in D# minor.

```
1 PREFIX mel: <http://www.dei.unipd.it/~gdb/ontology/melody#>
2
3 SELECT ?key ?count (((?count*100)/?tot_songs) AS ?percentage) WHERE {
4   {
5     SELECT ?key (COUNT(*) AS ?count) WHERE {
6       ?song a mel:Song ;
7         mel:key ?key .
8     }
9     GROUP BY ?key
10  }
11  {
12    SELECT (COUNT(*) AS ?tot_songs) WHERE {
13      ?song a mel:Song .
14    }
15  }
16 }
17 ORDER BY DESC (?count)
```

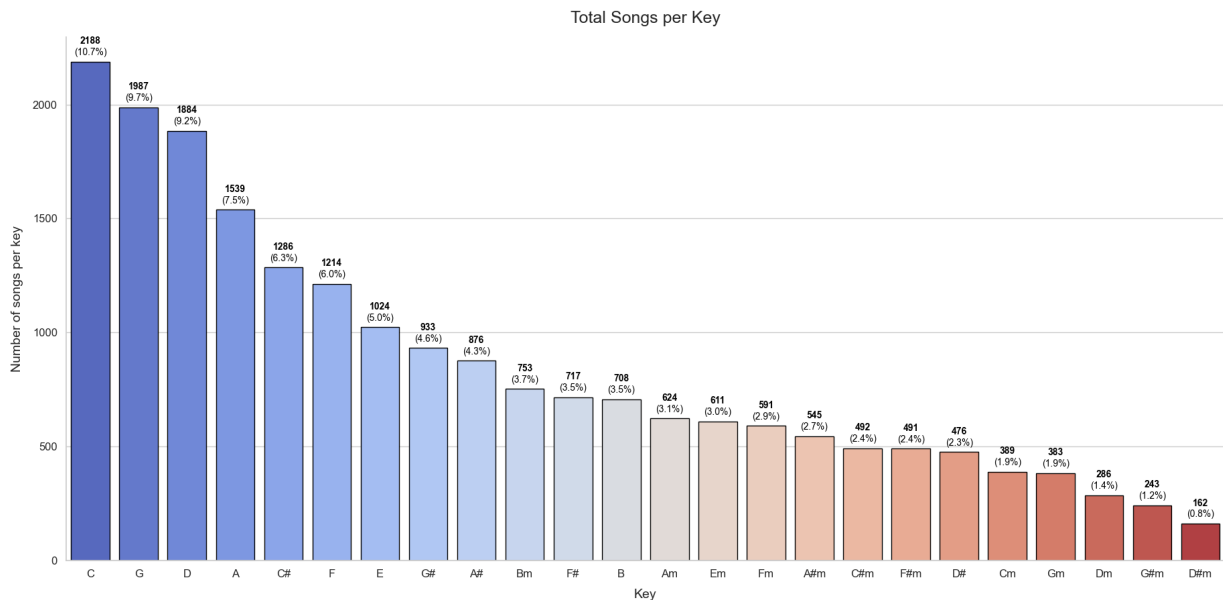


Figure 1: Bar plot showing the number of songs in each key, arranged in descending order. The number of songs and their percentage of the total are displayed above each bar.

Furthermore, as we observed from the ASK query below, the number of major key songs is more than twice the number of minor key songs. This might suggest that songs in the major scale are more likely to appear on the Billboard charts. However, this is purely speculative, since we do not have information about songs not in the Billboard Hot 100 charts.

```

1 PREFIX mel: <http://www.dei.unipd.it/~gdb/ontology/melody#>
2
3 ASK WHERE {
4   {
5     SELECT (COUNT(*) AS ?minorKeySongs) WHERE {
6       ?song a mel:Song ;
7       FILTER(Regex(?key, "m$")) .
8     }
9   }
10  {
11    SELECT (COUNT(*) AS ?majorKeySongs) WHERE {
12      ?song a mel:Song ;
13      mel:key ?key .
14      FILTER(Regex(?key, "^(?!.*m$)")) .
15    }
16  }
17  FILTER(?majorKeySongs > ?minorKeySongs*2) .
18 }
19
20
21 # ANSWER: YES

```

1.4 Song titles word-cloud

To analyze the most frequently occurring words in song titles, we wrote a query that concatenates all the song titles into a single string, converting uppercase letters to lowercase.

```
1 PREFIX mel: <http://www.dei.unipd.it/~gdb/ontology/melody#>
2
3 SELECT (GROUP_CONCAT(LCASE(?songTitle); SEPARATOR=" ") AS ?allSongTitles) WHERE {
4   ?song a mel:Song ;
5         mel:name ?songTitle .
6 }
```

Using the resulting text, we generated a word cloud with the wordcloud Python module. To avoid considering words such as *version* or *featuring* we defined a custom stopwords list to remove them.



Figure 2: Wordcloud of the song titles

As we could have imagined, the most recurring word is "love" with 1643 occurrences.

```
1 PREFIX mel: <http://www.dei.unipd.it/~gdb/ontology/melody#>
2
3 SELECT (COUNT(*) AS ?songWithLOVEInTheTitle) WHERE {
4     ?song a mel:Song ;
5         mel:name ?songTitle .
6
7     FILTER(REGEX(LCASE(?songTitle), "\\blove\\b")) .
8 }
```

1.5 One-Hit Wonders (The Musical Comets)

This query aims to identify artists who experienced a short moment of glory in the Top 50 of the Billboard charts. The query searches for artists who had at least one song reach the Top 50 but never appeared in the Billboard charts again with any other song, even outside the Top 50.

```
1
2 PREFIX mel: <http://www.dei.unipd.it/~gdb/ontology/melody#>
3
4 SELECT ?artistName ?songName ?position WHERE {
5   ?song mel:name ?songName ;
6     mel:sungBy/mel:name ?artistName ;
7     mel:classified ?membership .
8   ?membership mel:position ?position .
9   FILTER (?position < 51)
10
11  FILTER NOT EXISTS {
12    ?song2 mel:sungBy/mel:name ?artistName ;
13      mel:classified ?membership2 .
14    FILTER (?song2 != ?song)
15  }
16
17  FILTER NOT EXISTS {
18    ?song mel:name ?songName ;
19      mel:sungBy/mel:name ?artistName ;
20      mel:classified ?membership3 .
21    ?membership3 mel:position ?position2 .
22    FILTER (?membership3 != ?membership)
23  }
24 }
25 GROUP BY ?artistName ?songName ?position
26 ORDER BY(?position)
```

Song Title	Artist Name(s)	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 & John Scott Trotter & His Orchestra & Bing Crosby	48
Christmas (Baby Please Come Home)	Darlene Love	50

Table 3: Artists with a Single Top 50 Appearance on the Billboard Charts

1.6 Comeback Songs

This query explores the phenomenon of "comeback" songs: tracks that initially appeared on the Billboard charts, disappeared, and then made a reappearance at least a decade later. The query identifies the earliest year a song charted and then looks for subsequent song entries for the same song that occurred at least 10 years after that initial appearance. By calculating the difference between the return year and the earliest year of charting, the query highlights the songs that have demonstrated a remarkable resurgence in popularity over a significant time period.

```

1
2 PREFIX mel: <http://www.dei.unipd.it/~gdb/ontology/melody#>
3
4 SELECT DISTINCT ?songName ?artist ?earliestYear (YEAR(?date2) AS ?returnYear) (?returnYear - ?
   earliestYear AS ?distance)
5 WHERE {
6   ?song mel:name ?songName ;
7       mel:classified ?membership2 .
8
9   ?membership2 a mel:Membership ;
10              mel:classifiedIn ?billboard2 .
11   ?billboard2 mel:date ?date2 .
12
13   {
14     SELECT ?songName ?artist (MIN(YEAR(?date1)) AS ?earliestYear)
15     WHERE {
16       ?song mel:name ?songName ;
17           mel:classified ?membership1 ;
18           mel:sungBy ?art.
19       ?art mel:name ?artist.
20       ?membership1 a mel:Membership ;
21                   mel:classifiedIn ?billboard1 .
22       ?billboard1 mel:date ?date1 .
23     }
24     GROUP BY ?songName ?artist
25   }
26
27   FILTER (YEAR(?date2) > ?earliestYear)
28   FILTER (YEAR(?date2) - ?earliestYear >= 10)
29 }
30 ORDER BY DESC(?distance) ?songName ?artist ?earliestYear ?returnYear

```

Song Name	Artist	Earliest Year	Return Year	Distance
Money	The Kingsmen	1964	2018	54
Stay	Frankie Valli & The Four Seasons	1964	2018	54
Tequila	Bill Black's Combo	1964	2018	54
Try Me	Jimmy Hughes	1964	2018	54
Alone	Frankie Valli & The Four Seasons	1964	2017	53
Everybody	Tommy Roe	1964	2017	53
This Is It	Jim Reeves	1965	2018	53
Today	The New Christy Minstrels	1964	2017	53
Alone	Frankie Valli & The Four Seasons	1964	2016	52
Love Me Now	Brook Benton	1965	2017	52

Table 4: Artists with a Single Top 50 Appearance on the Billboard Charts

1.7 Album Excellence

This query identifies albums with a significant concentration of Grammy-nominated or winning songs, calculating the percentage of tracks within albums that contain more than two songs and that feature at least two Grammy-recognized songs. Interestingly, the resulting percentages typically remain below 20%. This observation likely reflects the structure of the Grammy Awards, where dedicated "Best Album" categories often recognize the artistic merit of an entire album rather than solely focusing on individual tracks. While exceptional and critically acclaimed songs within these albums may receive individual nominations or wins, the primary recognition for the entire body of work often comes through the "Best Album" awards, potentially limiting the percentage of individually recognized tracks.

```
1
2 PREFIX mel: <http://www.dei.unipd.it/~gdb/ontology/melody#>
3
4 SELECT distinct ?albumName ?totTracks ?artistName
5     (COUNT(distinct ?song) as ?songPerAlbum)
6     (((COUNT(distinct ?song) / ?totTracks) * 100 AS ?grammyPercentage)
7     (GROUP_CONCAT(distinct ?songName; SEPARATOR=", ") AS ?grammySongs)
8 WHERE {
9     ?song a mel:Song;
10     mel:winner|mel:candidated ?grammy;
11     mel:name ?songName;
12     mel:sungBy ?artist.
13     ?artist mel:name ?artistName.
14     ?album mel:containsSong ?song;
15     mel:name ?albumName;
16     mel:totalTracks ?totTracks.
17     FILTER(?totTracks>2) .
18 }
19 GROUP BY ?albumName ?artistName ?totTracks
20 HAVING(?songPerAlbum >1)
21 ORDER BY DESC(?grammyPercentage)
```

Album Name	Total Tracks	Artist Name	Songs Per Album	Grammy Percentage	Grammy Songs
Hotel California (2013 Remaster)	9	Eagles	2	22.2%	Hotel California - 2013 Remaster, New Kid in Town - 2013 Remaster
24K Magic	9	Bruno Mars	2	22.2%	That's What I Like, 24K Magic
From A Room: Volume 1	9	Chris Stapleton	2	22.2%	Broken Halos, Either Way
Clapton Chronicles: The Best of Eric Clapton	14	Eric Clapton	3	21.4%	Bad Love, Change the World, Tears in Heaven
Viva La Vida or Death and All His Friends	10	Coldplay	2	20.0%	Viva La Vida, Violet Hill
True Colors	10	Cyndi Lauper	2	20.0%	What's Going On, Change of Heart
Talking Book	10	Stevie Wonder	2	20.0%	You Are The Sunshine Of My Life, Superstition
Soul Provider	10	Michael Bolton	2	20.0%	How Am I Supposed to Live Without You, Georgia On My Mind
Minute By Minute	10	The Doobie Brothers	2	20.0%	What a Fool Believes, Minute by Minute
Monkey Business	16	The Black Eyed Peas	3	18.7%	My Humps, Don't Phunk With My Heart, Don't Lie

Table 5: Album Information

1.8 Top 10 Grammy-winning artists, including songs and albums

Our ontology divides the Grammy awards according to whether the category refers to a song, an album, or an artist. Knowing that a song is sung by an artist and that the artist has released an album, this query shows the artist with the highest number of Grammy wins, also considering the songs and the albums. Note that the query result does not precisely reflect the reality due to a dataset without well-defined fields, which leads our team to do some simplification to the matching logic between the Grammy and the winner.

```
1 PREFIX mel: <http://www.dei.unipd.it/~gdb/ontology/melody#>
2
3 SELECT ?name (SUM(?wins) AS ?totalWins)
4 WHERE {
5 {
6     SELECT ?name (COUNT(?albumGrammy) AS ?wins)
7     WHERE {
8         ?a a mel:Artist ;
9         mel:name ?name ;
10        mel:releasedAlbum ?album .
11        ?album mel:winner ?albumGrammy .
12    }
13    GROUP BY ?name
14 }
15 UNION
16 {
17     SELECT ?name (COUNT(?artistGrammy) AS ?wins)
18     WHERE {
19         ?a a mel:Artist ;
20         mel:name ?name ;
21         mel:winner ?artistGrammy .
22     }
23     GROUP BY ?name
24 }
25 UNION
26 {
27     SELECT ?name (COUNT(?songGrammy) AS ?wins)
28     WHERE {
29         ?artist a mel:Artist ;
30         mel:name ?name ;
31         mel:sing ?song .
32         ?song mel:winner ?songGrammy .
33     }
34     GROUP BY ?name
35 }
36 }
37 GROUP BY ?name
38 ORDER BY DESC(?totalWins)
39 LIMIT 10
```

Artist Name	Total Wins
Stevie Wonder	19
Eminem	14
Adele	11
Bruno Mars	11
Aretha Franklin	11
Michael Jackson	10
Dixie Chicks	10
Alicia Keys	10
U2	9
Foo Fighters	9

Table 6: Top 10 Grammy-winning artists, including songs and albums.

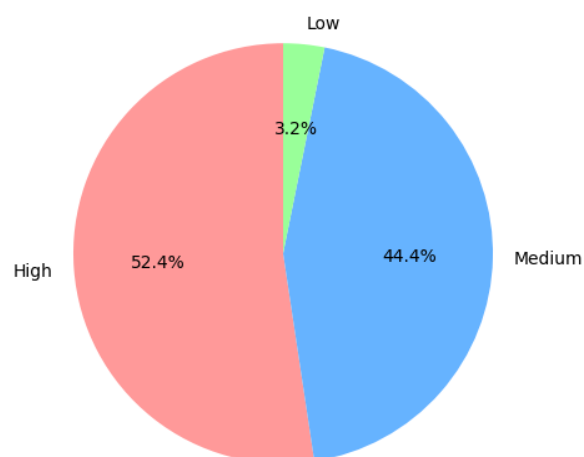
1.9 Total Grammy wins per artist popularity

With this query, we want to check whether, given an artist, their popularity reflects the number of Grammys won. As we can see from the pie chart, this is verified: "High" popularity has a higher number of wins than "Medium" popularity, which, in turn, has a higher number of wins than "Low" popularity.

```

1 PREFIX mel: <http://www.dei.unipd.it/~gdb/ontology/melody#>
2
3 SELECT ?popularity (COUNT(?grammy) as ?totalGrammy) WHERE {
4   ?artist a mel:Artist ;
5           mel:hasPopularity ?popularity ;
6           mel:winner ?grammy .
7
8   {
9     SELECT (COUNT(?grammy) as ?totGrammy) WHERE {
10      ?artist a mel:Artist ;
11             mel:winner ?grammy .
12    }
13  }
14 GROUP BY ?popularity

```



1.10 Average weeks on Billboard Hot 100 for artists with at least 10 songs

The idea of this query is to find out the average number of weeks an artist's song has been on the Billboard Hot 100 chart. In order to eliminate some misleading results in case an artist has one song or, in general, too few songs, the artist must have at least 10 songs that have been included in the Hot 100.

For brevity, in the results, we only show the top 10 artists with the highest average.

```
1 PREFIX mel: <http://www.dei.unipd.it/~gdb/ontology/melody#>
2
3 SELECT ?artistName (COUNT(?song) AS ?totalSongs) (AVG(?totalWeeks) as ?average) WHERE {
4   ?artist a mel:Artist ;
5     mel:name ?artistName ;
6     mel:sing ?song .
7   {
8     SELECT ?song ?songName (COUNT(?billboard) as ?totalWeeks) WHERE {
9       ?song a mel:Song ;
10         mel:name ?songName ;
11         mel:classified ?membership .
12       ?membership mel:classifiedIn ?billboard .
13     }
14     GROUP BY ?song ?songName
15     ORDER BY DESC(?totalWeeks)
16   }
17 }
18 GROUP BY ?artistName
19 HAVING (?totalSongs >= 10)
20 ORDER BY DESC(?average)
21 LIMIT 10
```

Artist Name	Total Songs	Average
Imagine Dragons	13	30.46153
Bruno Mars	14	27.21428
The Karaoke Crew	29	26.96551
3 Doors Down	10	26.8
Matchbox Twenty	10	26.5
OneRepublic	13	24.92307
The Black Eyed Peas	16	24.375
Destiny's Child	13	24.07692
LeAnn Rimes	10	24
Adele	12	23.41666

Table 7: Average weeks on Billboard Hot 100 for artists with at least 10 songs.