Master in Sound and Music Computing
Universitat Pompeu Fabra

Generating Classical Music Playlists using Radio Broadcast Programming

Anna Eliane Barletta

Supervisors: Emilia Gomèz, Lorenzo Porcaro

August 2022

Contents

Chapter 1

Introduction

Recommender systems are algorithms that aim to suggest relevant items to users, such as movies to watch, products to buy, or in the case of our interest, music to listen to. In recent decades, with the rise of many web services, they have increasingly taken hold in our lives.
Current music recommendation systems trace back, historically, to the work of music programmers and disk jockeys in commercial and institutional radios in the XX century.
These insiders ranged from self-taught youth to highly distinguished musicologists and music critics and provided an extreme diversity of musical propositions.
Of course, their proposals were all indistinctly biased because they were the product of human minds and knowledge. And of course, many (but not all) of such biases were driven by explicit and/or implicit record labels' strong thrusts in the form of bribes, money, benefits etc.
However, this scenario also provided a small cohort of culturally and politically motivated people who were committed to some form of cultural enhancement. This allowed for a small niche diversity in the general picture.

Nowadays, music recommendation systems have almost completely superseded the roles of music programmers in webcasting listening platforms. Such entities are less human–driven and more machine–driven, and they satisfy impatient consumers

who have become used to enjoying vast amounts of music promptly available at their fingertips. [1]
All in all, a century of such recommendation "systems" has produced two widely observable effects:

- the unprecedented pervasive extension of music listening
- the degradation of music listening from a cultural activity to a day–to–day shallow hearing of musical information.

The second item has generated the diffuse intimate perception that music listening is a comforting entertaining activity rather than a profound cultural appropriation. We are far away from the late latin *Quadrivium* which included music in a quartet of very speculative activities [2], and our familiarity with the idea that music is a mind–developing cultural endeavor has crumbled down to zero.

The role attributed to music has changed over time, being reduced to a mere form of entertainment. Dusting off the words of T.W. Adorno: "Music is linked to cognitive habits, mode of consciousness and historical development" [3].
We should therefore make the most of music's ability to foster critical consciousness. By preserving dissonance instead of offering musical resolution and progression, music has the power to challenge cognitive, emotional and perceptual habits [4].
If the findings of neuroscience are correct – positing the condition of familiarity with musical repertoires (among other) for the development of specific parts of the human brain and therefore of the intellect [5] - it is crucial to emphasize the importance and role that recommendation algorithms play in the potential cultural growth of our society [6] [7].

Western classical music has always been treated as a kind of cultural form belonging only to a social elite with a suitable educational background and musical literacy.
Transforming classical music into *common culture* [4] is not just about making it accessible to everyone. Instead, such a process involves an educational system that gives all citizens the opportunity to understand and learn it, until it becomes a

'familiar and ordinary' aspect of culture.

Furthermore, classical music can be preserved in promoting cultural diversity and cultural complexity [8]. If cultural products are left on the market as mere economic items, the concept of cultural diversity may disappear, either because it does not conform to the laws of the market, or because it simply does not fit.

Despite the enormous number of composers and works from the Middle Ages to the present day, Western classical music remains a vastly underrepresented genre in the current music scene of recommendation systems.

The reasons, as already argued, are varied and are certainly to be found in the social-economic aspects of today's society, in the role that music plays in it but also in the profound changes that new technologies have caused culturally and socially.

Recommendation algorithms play a key role in this kind of musical genre "discrimination", so the initial *research questions* that led this project were:

- Is it possible to build a prototype of a free *culturally–motivated* recommender system algorithm, tailored to Western Classical Music, that best represents the vastness of this genre's repertoire in terms of different eras, sub-genres, instrumentation, composers for both newcomers and experts?

- Through the search of comprehensive metadata, derived from reliable sources, by taking advantage of these newly collected details, is it possible to create more comprehensive and customized playlists, following various specifiable criteria?

However, during the process of data collection for the creation of the source dataset, significant issues emerged, some of which affected the scope of the work. These *systemic* issues showed the urgency of turning the spotlight on the obstacles presented by the retrieval, partitioning and cataloguing of classical music data.

This work intends to tackle, at least in a prototypical way, these problems and provide tentative models and solutions to them.

To give a possible answers to the research questions, several steps were followed; the are organized as described below: after analyzing and commenting on the typical structure of classical music data and its current role in today's recommender systems (Chapter II, State of the Art), it will be covered how the data were collected to create the database from three different sources and how various types of playlists were generated (Chapter III, Methodology) then the results obtained and general statistics for understanding them (Chapter IV). This will be succeeded by an explanation and discussion of the outcomes (Chapter V, Discussion) and limitations of the project, and finally, possible future implementations (Chapter VI, Conclusions).

Chapter 2

State of the Art

The algorithms behind current music recommendation systems are clearly designed to favor other music genres such as pop, dance and rock, which share characteristics that often do not fit with classical works.

Scrolling through the proposals of the most widely used music recommendation systems, the western classical music compositions are generally completely dissociated from their original function and context, or reduced to a mere soundtrack and background for other activities.
It is not unusual indeed to find the track *Lacrimosa* from W.A.Mozart's Requiem K626 in the 'Music for Sleeping' playlist on Spotify or the *Swan Lake* from P.I. Tchaikovsky in the 'Music for Reading' one.
This has led over time to the emergence of genre-specific apps, such as IDAGIO [9] which tries to overcome some of the shortcomings present in other platforms at least featuring optimized meta-data and complemented by, "curation by music-lovers for music-lovers".

Since the first research, it is also evident that what is shown, proposed and played of Western Classical Music is only a tiny fraction of the repertoire composed over the last 1,500 years of history. It is worth noting that even in theatres and concert halls, the works and Operas on offer belong to the same trite repertoire. There are several papers and articles that give proof of this assumption [10][11][12] with empirical evidence. This is one of the factors that is leading, over time, to a progressive crisis in the music and theater industry.

According to [13], the most visible evidence of this critical state is the steadily declining numbers of people attending concerts in some countries, sometimes linked with the socio-demographic changes in the audience profile.

In Botstein's view [14] the causes adduced for the decline of interest in classical music range from aesthetic critiques, to the impact of the detrimental effects of modernity and precipitous drop in cultural standards and taste in 20th-century society mass society. In addition, the interactions within segments of society that encouraged audiences in the past to gather and socialize at public occasions of high-art have weakened, making concert attendance also lose its social allure.
The outcame is an inexorable erosion of the audience, with increasingly higher age distribution. Given the decline of music education and promotion in schools and homes, recruiting successive generations of adult audiences will remain an unsolved problem in the absence of an effective surrogate. Recommendation algorithms could thus help prevent classical music from being condemned to a geriatric activity.

Regarding the narrow musical offerings of today's theaters, an immediate return is given by the website Bachtrack.com [15], which every year lists more than 30 thousand performances and live concerts, collects streaming trends from all around the world, and publishes statistics on the webpage. Bachtrack is not a canonical academic database in the sense that it is not set out to be systematic, homogeneous, and complete but, as far as it's known, is the largest one in this field. [10]

The available sources show how much concert programming in theatre has become increasingly conservative, tending to obsess over just a limited handful of composers, around 1-2% out of the total number, whose pieces are played *ad nauseam.* In fact, it seems clear that programmers' choices usually fall on a disproportionate and hegemonic very small group of composers, who are presented in regular associative patterns, which rarely includes works by a group of diverse (and very often living) ones. This highly conservative programming conception can be a significant agent in the apparent crisis of the classical concert and audience decline.

Leaving aside the serious and pervasive cultural impoverishment that this causes, what is important to emphasize for the purpose of this thesis is that the trend is sadly reflected on music recommendation platforms. [10]

As far as it is known, there is no public, structured and comprehensive dataset of Western classical music playlists or a dataset of Western Classical Music for music recommendations.
In the field of Music information retrieval there are several open access datasets [16], in which there are numerically virtuous examples such as Million Song Dataset that counts more than 1 million files. Unfortunately, those related purely to classical music are often reduced to only modest collections of the sounds of particular typically acoustic instruments (violin, piano) or specific collections of a set of songs (such as all of Beethoven's sonatas or Mozart's String Quartets).

If for audio material this is somehow understandable, since the problem of licenses and copyrights afflicting cultural (and other) resources gives huge limitations, this is less obvious for metadata.

The main difficulty, besides the retrieval of the data itself, is certainly the intrinsic structure they have.
More than for other genres, the information concerning even one classical music performance, is much and detailed. In addition to the composer, there is the performer(s), which in most (but not all) cases can be a soloist or an orchestra conducted by a particular director. Of course it is essential the title of the work, which is often divided into movements, is accompanied by at least a number, a sub-title and an abbreviation.

Since execution can vary greatly depending on who is playing, enthusiasts search for a particular performance or even listen to several versions of the same piece interpreted by different musicians or recorded under different record labels. This makes the retrieval and cataloguing of data very complex and is in fact, up to now, reduced to some academic initiatives or foundations connected to theatres (like the virtuous example of the Fundación Juan March of Madrid [17], which was involved in this work, or the Metropolitan Opera House in New York [18]) that collect them for internal statistics.

A further aspect to investigate in order to answer the search questions is the generation of playlists.
They were originally created around the beginning of 20th century, to obviate the tedious task of manually searching for the artists and songs we like most; in a playlists are instead collected in a single list to be scrolled through and listened to all in one go (shuffle mode) or jumping from one track to another (random mode). A clear definition could be "ordered sequence of songs meant to be listened to as a group".

In broad terms, there are three methods of playlist generation: manual (individual song selection), through the use of assisted techniques (guided and visual creation) and automatic (without human intervention), which is what will be developed in this project, even if they are based on playlists created by experienced humans in radio.
The groundbreaking work of Pauws and Eggen [19], or others methods based on different aspects and listening habits of users, such as daily activities [20], time context [21] are just some examples [22]. Despite using different data and techniquess, their actual goal nowadays remains the same: using the created playlists to capture, track and encode users' preferences and habits, to recommend new songs.

According to the literature, there is a considerable number of procedures to automatically generate playlists developed during the last decade. Authors like Bonnin [23] and Sneha[24] provide an in-depth overview of playlist generation and characterization, highlighting its merits and drawbacks, far too detailed for the purpose of this work. Instead, it is significant to mention at least the two main types which are certainly Similarity based algorithm and Collaborative Filtering. For knowledge and completeness, the other categories with their pros and cons, are included in the summary table made by Dias [22] (see 1).

Pros and Cons of the different techniques for automatic playlists generation		
Technique	Pros	Cons
Similarity-based	Scale to large music collections; creates homogenous playlist	Promotes little diversity and song discovery
Collaborative filtering	Extensive research on collaborative filtering methods; Adapts to past preferences	Requires many data to perform accurately
Frequent pattern mining	Generated playlists can implicitly reproduce the observed characteristics of manually defined playlist	Quality of the playlists generated depends on the number and quality of the playlists used for pattern mining
Statistical method	Plenty of algorithms for optimizing the playlist generation	Learning process of these algorithms can be time consuming
case-based reasoning	Low computational complexity when a limited number of cases is used	Do not scale well for repositories with many cases
Discrete Optimization	Generation can satisfy most of the target characteristics, when background knowledge is accurate	Most solutions are computational complex and expensive; scalability issues
Hybrid techniques	Combination of different techniques to overcome individual limitations; Can adapt to different contexts	Can be more expensive and time consuming than a sub-optimal solution

Table 1: Playlists generation techniques

Similarity-based approaches are based on the similarity or closeness between tracks, the distance between which is measured by mathematical functions (e.g., Euclidean) Different features of the audio signal or its metadata can be used as input to the function.

As described by Cruz[11] the majority of the mainstream recommendation systems use *collaborative filtering* which help predict the users' music preferences based on their past preferences and preferences of similar users.

However using this general recommendation method is not ideal for less mainstream music like classical music; it would benefit much more from the use of a *content-based* recommendation system.

Chapter 3

Methodology

This chapter will explain the methodology by which the database was created and the statistical analyses concerning it and the different playlists it generated.

3.1 Database

Retrieving and especially dividing the data proved particularly arduous.
The sources chosen for the creation of the database were three European 'genre-specific' radio stations (from now on *providers*):

- RAI Radio Classica from Italy
- Radio Clásica from Spain
- BBC Radio 3 from United Kingdom

While in the case of the first two the data were grouped in pdf files, (judging by the typos) manually written, and obtained by explicit request, the BBC data, on the other hand, were retrieved originally from the web.

This led to the creation of three different parser algorithms to extract essential data, such as track title, duration, composer's name and time of transmission. As will be explained later, Wikidata [25] was then used to standardize the composers' name.

The creation of this database is an opportunity to shine a light on the poor digitization of metadata regarding classical music. Most of the problems stem from the fact that the providers themselves are not consistent in their writing, which makes both their reading and cataloging counter-intuitive and unnecessarily tangled.

All the data collected are organized into a source csv file, with a field dedicated to each feature. The database then created is written in SQLite3 [26]. Since it is a relational database, by setting up queries it is possible to know the links of the various features of each entry (See Fig. 1).

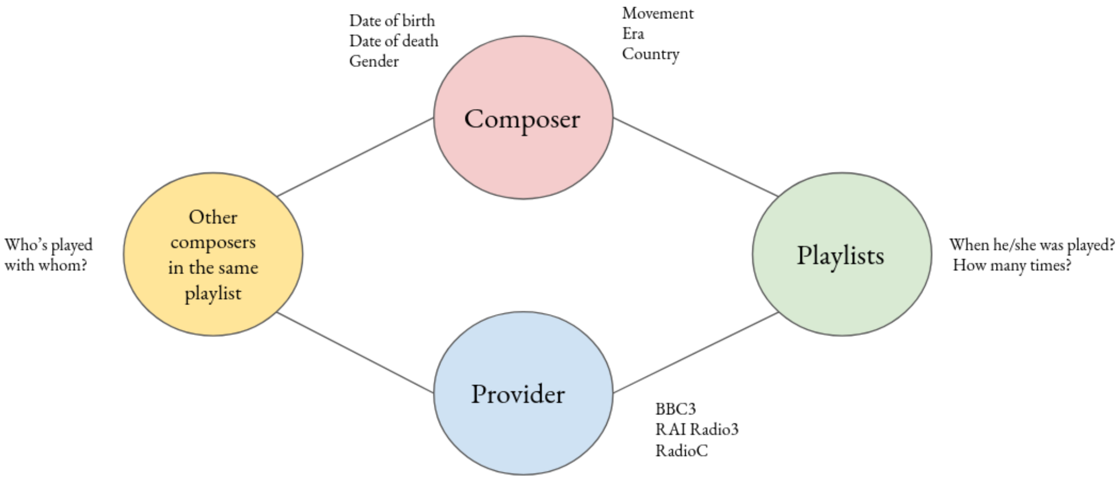

Figure 1: Database structure

3.1.1 RAI Radio Classica

RAI is the Italian state radio and television broadcaster [27]. By contacting the editorial team, it was possible to be included in the mailing list that shares the bi-weekly schedules of the classical music channel [28]. The page contains the time schedule and the name of the author in bold, usually followed by the title, the interpreter,the duration and the music label (See Fig. 2.

As can be seen in the image the structure mentioned above is frequently rehashed to contain other information such as instrumentation or the alias of single movements. This created enormous difficulties in the construction of the parser, as there are no

johann sebastian bach
toccata e fuga in re minore per organo BWV 565
bernard foccroulle, organo
durata 8.55
ricercar ric-276

franz joseph haydn
cinque canti scozzesi per voce, violino, violoncello e pianoforte
dear silvia - the tither morn - I do confess thou art sae fair - the bonnie wee thing - the widow
lorna anderson, soprano; jamie mac dougall, tenore; trio haydn di eisenstadt: harald kosik, pianoforte; verena stourzh, violino; hannes gradwohl, violoncello
durata 7.10
brilliant classics 93782-77

arthur honegger
preludio arioso e fughetta sul nome di bach, per orchestra d'archi
preludio - arioso - fughetta
i musici di montreal diretti da yuli turovsky
durata 6.55

charles gounod
ave maria - rielaborazione del preludio n. 1 in do maggiore dal vol. I del "clavicembalo ben temperato" di johann sebastian bach
jessye norman, soprano; fabrice pierre, arpa
durata 3.30
philips 432-731-2

Figure 2: A fragment of a program file from RAI

fixed elements that can be taken as a reference point to properly subdivide the title from the rest.

It should be emphasized that this provider furnishes the most complete and detailed information; the records are therefore uniquely identified by the catalog number of the label and the exact duration.

Again this suggests that there is a database from which this information is drawn but then transcribed in a non-standardized way.

3.1.2 Radio Clásica

Radio Clásica is part of RTVE, "Corporación de Radio y Televisión Española"[29]. The structure of the data, except for graphic details, resembles that of Italian radio

RAI Classica.

The files were obtained through the cooperation of the editorial staff, which has been making pdf files of monthly programming available for more than 10 years. Through a few simple lines of code on python it was possible to download all the pdfs from 2013 to the present, in order to have an adequate amount of data to compare with the BBC.

However on this radio, time dedicated to programs with speakers and anchors prevails on time dedicated to music pieces; for the same number of years, the musical metadata collected is significantly less.

Unlike the other providers, in the Spanish radio files only the composers' surnames are known; in rare cases (i.e., "C.F.E. Bach") the initials of the first name are present. This has caused quite a few scraps in database creation, as it is sometimes impossible to unambiguously identify a composer without having at least his or her full name.

3.1.3 BBC Radio 3

The BBC, the famous British radio and television channel, has dedicated station 3 of its radio station to the broadcasting of classical music. It has a bot on Twitter [30], *NowPlayingBot* [31], which posts a tweet every time a new piece is played.
The original idea was to download and unpack these tweets. The standard structure of the tweet can be seen in the Fig. 3.

After the "Now Playing" there's always the name of the composer, after the comma the interpreter followed by a dash and the title of the track, with further information such as the number of the opera or the movement. Then the tweet always ends with a couple of hashtags, with the names of the label or the player.

Tweepy [32], a python library for accessing the Twitter API, was used to collect tweets. One of the limitations of the this method is that there's an API limit of downloadable posts set at 3200, corresponding to approximately one month of

Figure 3: Now Playing Bot on Twitter

programming.

One other thing to note is that each tweet is made by, maximum, 140 characters. Unfortunately often that number of letters is reached before the metadata are fully described; consequentially, since the BBC doesn't skimp on the hashtags, the majority of the titles the last part of the post, are incomplete.

To overcome these problems, the BBC Radio 3 official web page was used [28], where the programming is reported everyday. Using the source code, written in HTML, the classes in which the crucial information is contained were identified.
A Python library called 'BeautifulSoup' [33] was used to scrap and parse the html file. It was possible to trace back the schedules from 2013 to now, which fortunately have maintained somehow a fairly stable structure over time. The only missing field for this provider is the duration of the songs.

3.1.4 Titles

Title parsing certainly represents the most delicate part to separate, because they do not have a standard format.
Since, as mentioned earlier, provider programs are not consistent in the way they report metadata, many of the results therefore also have extraneous elements (such as performers or arrangers) in the title field.

An explanatory example is given by the following figure (fig. 4 of a fragment of a music program file, in which defining a rule to differentiate and separate the various elements is, to use an euphemism, challenging.

16.00►Programa de mano
Concierto celebrado en la sala Markus Sittikus de Hohenems, el 6 de septiembre de 2016. Grabación de la ORF, Austria. Schubertiade.
SCHUBERT: *Auf der Donau. Der Zwerg. Meeres Stille.* IBERT: *Chanson du départ. Chanson à Dulcinée. Chanson du duc. Chanson de la mort.* SCHUBERT: *Gruppe aus dem Tartarus. Fahrt zum Hades. Prometheus. Der Tod und das Mädchen. An den Tod. Der Jüngling und der Tod.* SCHUMANN: *Muttertraum. Märzveilchen. Der Soldat. Der Spielmann. Totengräbers Heimwehe. Erlkönig. Nachtgesang.* Tareq Nazmi (bajo), Gerold Huber (p.).

18.00►Grandes ciclos
Max Reger
REGER: *3 Gesange*, Op. 111b (selec.) (2'36"). Coro Monteverdi de Hamburgo. Dir.: J. Juergens. *Prólogo sinfónico a una tragedia*, Op. 108 (32'12"). Orq. Sinf. de Norrkoping. Dir.: L. Segerstam. *3 motetes*, Op. 110 (selec.) (15'46"). Conjunto Vocal de la SWR Stuttgart. Dir.: F. Bernius.

19.00►Fundación Juan March
Transmisión directa desde Madrid.
Vuelta al orden: clasicismos y neoclasicismos. *1780 Clasicismo vienés.*
HAYDN: *Variaciones en fa menor,* Hob. XVII / 6. MOZART: *Sonata nº 13 en Si bemol mayor,* KV 333. HAYDN: *Sonata en Mi bemol mayor,* Hob. XVI / 52. BEETHOVEN: *6 Variaciones en Sol mayor sobre el duetto* "Nel cor più non mi sento". MOZART: *Variaciones sobre el aria* "Salve tu, Domine" *Fantasía en Do menor,* KV 475. BEETHOVEN: *Sonata nº 8 en Do menor,* Op. 13 "Patética". A. Guijarro (p.).

22.00►Paisaje nocturno
MANFREDINI: *Sinfonía en Sol mayor*, Op. 2 nº 8 (7'01"). Capricornius Consort Basel. Dir.: P. Barczi. CAROLAN: *Piezas* (7'20"). J. Christopher (guit.), T. Kain (guit.). SCHUMANN: *Sonata para piano en Sol mayor*, Op. 118 nº 1 (9'20"). J. Lee (p.).

Figure 4: A fragment of a music program file

Each entry is different from the previous and the next. In the first program there are only the titles of various works by the same composer, separated by periods; it is assumed that the musicians are always the same, given at the end. In the second program there is the title (or more than one), flanked by the opera number with the corresponding duration, always followed by the performers. In the third program the works are written differently ('Hob.'), using Roman numerals, and there is again only one performer.

3.2 Wikidata

Wikidata [25] is a collaboratively edited multilingual knowledge graph; it's the common source of open data that Wikimedia projects such as Wikipedia, and anyone else, can use under the public domain license Each item listed on Wikidata is associated with a unique identification number (Q######) and has a set of properties associated with it. The name of each composer was automatically searched on Wikidata by a query written in SPARQL using the query service [34]. Specifically, the properties extracted for each composer are: name, sex or gender, date of birth, date of death (if any), era, city of citizenship and its location (latitude and longitude).

Wikidata was therefore used as an authority to normalize and standardize composers' names, especially those containing special characters and accents that are often misspelled or translated.

The search is performed through the above query, whose search limit has been conventionally set to the first 30 entries.
In the case of homonyms, which are quite frequent especially between English names, the item description is decisive and must contain relevant words such as "musician or composer".

Extensive as it is, Wikidata is also not complete. Some items have been completed manually (changing also the item's page on Wikidata), while others are unfortunately discarded due to lack of data.
In particular, the discrimination is made on the date of birth, which, if absent, does not allow the composer to be placed on the timeline.

To make up for the often missing era data, a timeline was created with the 7 main historical periods (Medieval, Renaissance, Baroque, Classical, Romantic, Modernism, Contemporary) with their corresponding time intervals.
By knowing the composer's date of birth and death, it was possible to place them manually.

At the same time, to get a map plot of the geographic origin of the composers, Geopy [35] was used, which derives the spatial latitude and longitude coordinates of a given place, indicated on Wikidata. In the case of an empty field, "unknown" is put for each feature.

3.3 Playlists

Having gathered more details about composers, it is possible to customize playlists and choose one or more criteria for their creation from the choices made by radio stations.

Item b and in general internal sorting in the matrices shows large color differences that are grouped into so-called zones. Zone 0 includes the 30 most famous composers which, not coincidentally, also have the largest number of intersections, while, following an exponential trend, the last one zone 9, includes those broadcast a few times. The zones, accordingly, represent the *popularity* of the composers.

Every playlist consists of 10 composers' names and songs' titles, and belongs to one of these main types, or their blends:

- Random
- Zone
- Path
- Movement

The rationale is to use the features collected during database creation to customize and personalize the lists and generate them according to established principles.

"Random mode" creates a playlist of 10 names, randomly drawing them from the database.

A *"Zone playlist"* consists of a list of composers, each belonging to a different popularity zone, in order, starting from Zone 0.

A *"Movement playlist"* contains 10 composers belonging to the same era. Unfortunately they do not have an even representation within the database; medieval composers, for example, are a minuscule fraction of the total.

"Path Playlists" are the very essence of the concept behind this work.
The algorithm, starting with the first composer, generates the next one, making sure that it belongs to the next zone and that the "distance" (the number of times it has been paired with the previous, normalized between 0 and 1) is within a certain specified range. Thus, the result is a playlist with decreasing popularity consisting of titles and related artists that are more or less distant from each other. The variation in this range is a strong determinant.

These are just the main categories of possible playlists, but it is perhaps more intriguing to combine them. Thus one can create a list of Renaissance composers very close to each others, or one consisting of composers who rarely crossed over etc.

What happens to the titles?
In list creation, in path mode, once the composer generated by the previous name is identified, the algorithm chooses a title that was actually played in a playlist of radio stations where both composers were present. In other modes, however, where the intersection between the composers may never have occurred, the title to be associated is chosen from all possible.

Of curse it is important to be aware that the choice of indicators for creating them greatly influences the results. Indeed, specifying a very small distance range (i.e., 0.25-0.30) greatly narrows the span of possible matches that precisely satisfy the condition. Equally, since composers are more likely to have crossed a limited number of times and not to be always matched (distance about 1) choosing a very high interval (i.e. 0.8 -1.0) could generate incomplete lists due to lack of options.

To overcome this problem an algorithm has been implemented that in case it cannot find any composer that deviates the specified measure, it widens the range by 0.1 by searching again for a suitable name. If even this expedient does not suffice perhaps one has ended up in a "dead end", given by an artist who has no further crossings except with the previous one that generated it.

At this point the code is programmed to take a step back, return to the previous composer and bring about another name. Obviously, to avoid recursive and ringing processes, the draw is without repetition.

Chapter 4

Results

The results achieved are noteworthy. The database is fairly substantial and consists, for the most part, of composers with very low "popularity indexes." This suggests a wide variety of names and the potential unseen relationships between them.
This chapter diverse statistics will shown that will help to understand their distribution and characteristics.

4.1 Heatmaps

Through the data contained in a counter present for each composer, heatmaps were plotted. They are basically matrices of occurrences in which the color of each box is a measure of the number of times the two elements (composers) crossing at that point were broadcasted in the same program, by each provider. They could be called graphical popularity matrices, symmetrical with respect to the diagonal.

Two basic numbers are associated with each composer:

- a number, between 0 and 1, for each column of the matrix, representing the number of normalized crossings it has had with all the various composers. (i.e. obviously 0 will mean that they were never included in the same playlist by any provider)

- the total absolute number, sums of all the above, of occurrences, i.e., how many times a composer has been broadcasted.

Item a represents a measure of *distance* between composers. Musicians who are often associated in playlists and will consequently be considered similar and therefore close; conversely, they will be different and more distant if they have never met each other. This measurement is calculated based on the selections made by those who write radio programs, who are reputed experts in the field. Their work is then considered ground truth, assuming that solid and reliable criteria are behind the choices made.

Arrangements have been taken to mitigate the somewhat hierarchical and polarized layout of the maps, plotted based on the numbers in item b. Indeed, there are composers present in one-third of the playlists generated by radio stations compared with others that are broadcasted twice in almost a decade.

The number of occurrences was normalized between 0 and 1, transposed to logarithmic scale base 10, and re-scaled according to the *Gini coefficient* [36].
The Gini-factor takes into account not only the number of intersections but especially their distribution. This does not alter the data, simply the order is somewhat re-scaled by favoring a composer that has been broadcasted many times but more importantly has been mixed with many others. This implies that a more versatile composer is more valuable than one who is simply popular. Sometimes, things can coincide.

Shown below is the global heatmap (Fig. 5) that takes into account all the intersections and those of individual providers compared.

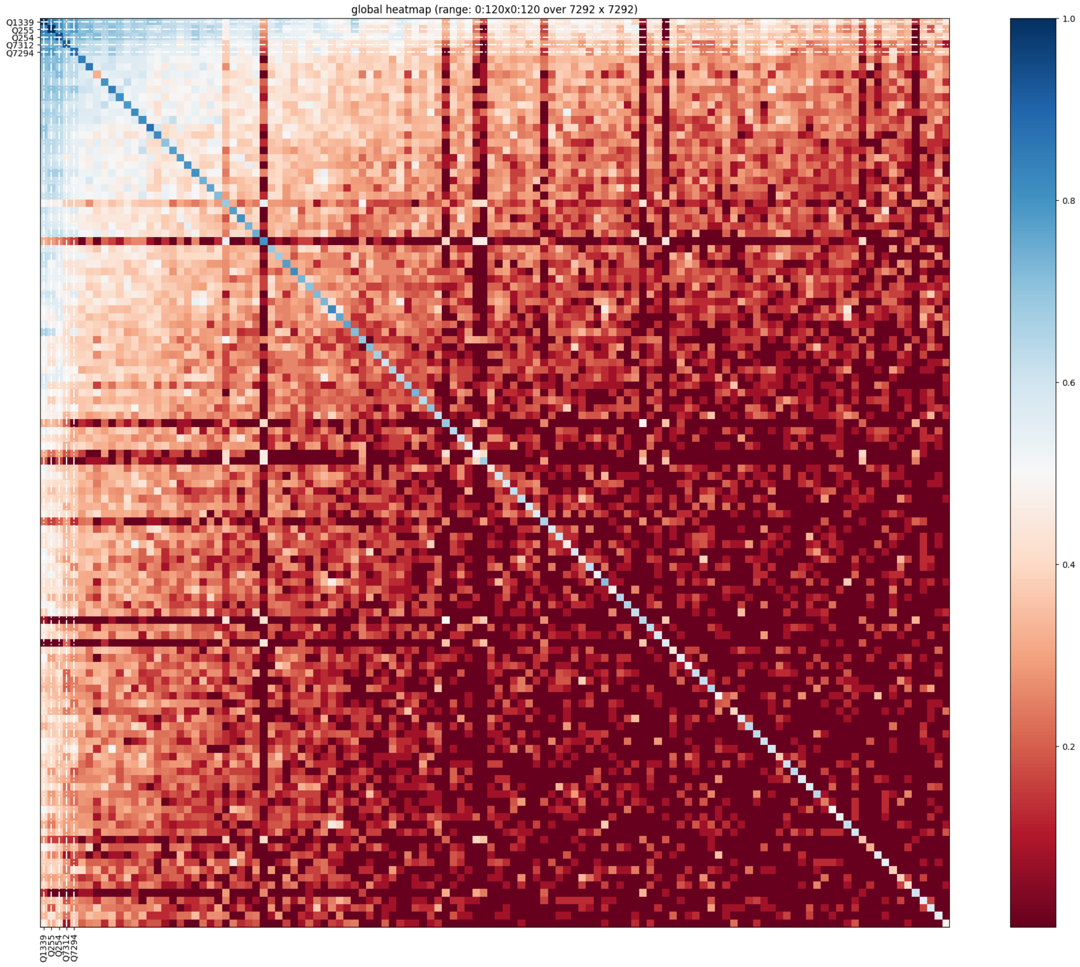

Figure 5: Global Heatmap - First 120 elements

In Fig.5 the crossings for the first 120 composers out of more than 7000, in the other one the detail of the zone of highest popularity, formed by the first 50 names (Fig. 6). Notice how the zones of highest occurrence, distinguishable by colors ranging from white to blue, are highly concentrated are in the upper left corner.

Highlighted number ids (Q#####) are those whose total number of occurrences exceed a certain threshold, arbitrarily set at 0.7.

Specifically they are, in the order of appearance: J.S. Bach, W.A. Mozart, L.W. Beethoven, F. Schubert, J. Brahms.

Another interesting aspect is given by the diagonal, with significant numbers of intersections for practically all of them. This means that often a composer is broadcasted

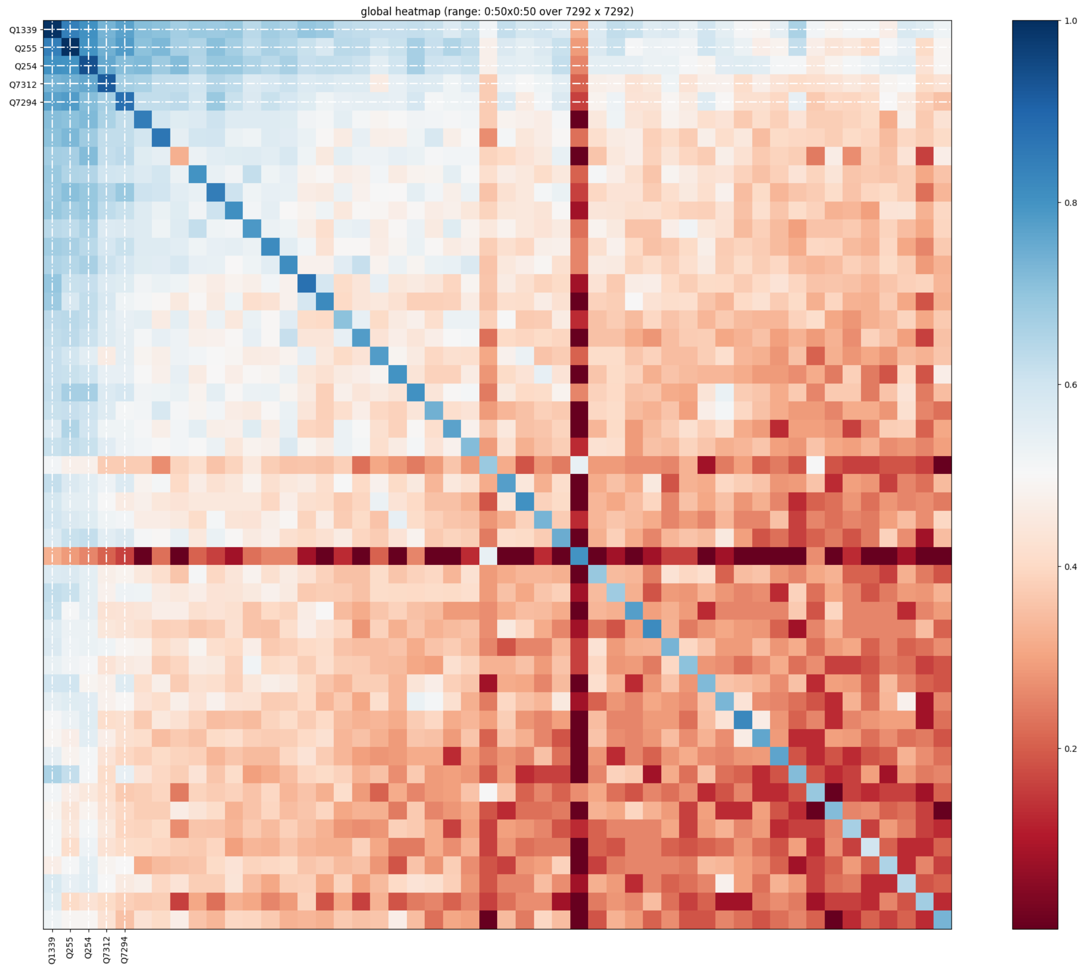

Figure 6: Global Heatmap - Top right corner

more than once in the same playlist.

Following the same principles, heatmaps of the eras were also created (Fig. 7). The medieval era has the fewest intersections, followed by the Renaissance. As will be evident later, this is also given by the fact that they are significantly less represented than the other historical periods. Again, the diagonal has mainly colors indicating a level of occurrence above 0.7, even touching the maximum for the Romantic and Modern Era.
This suggests some stylistic and historical consistency of choices made by providers in creating playlists.

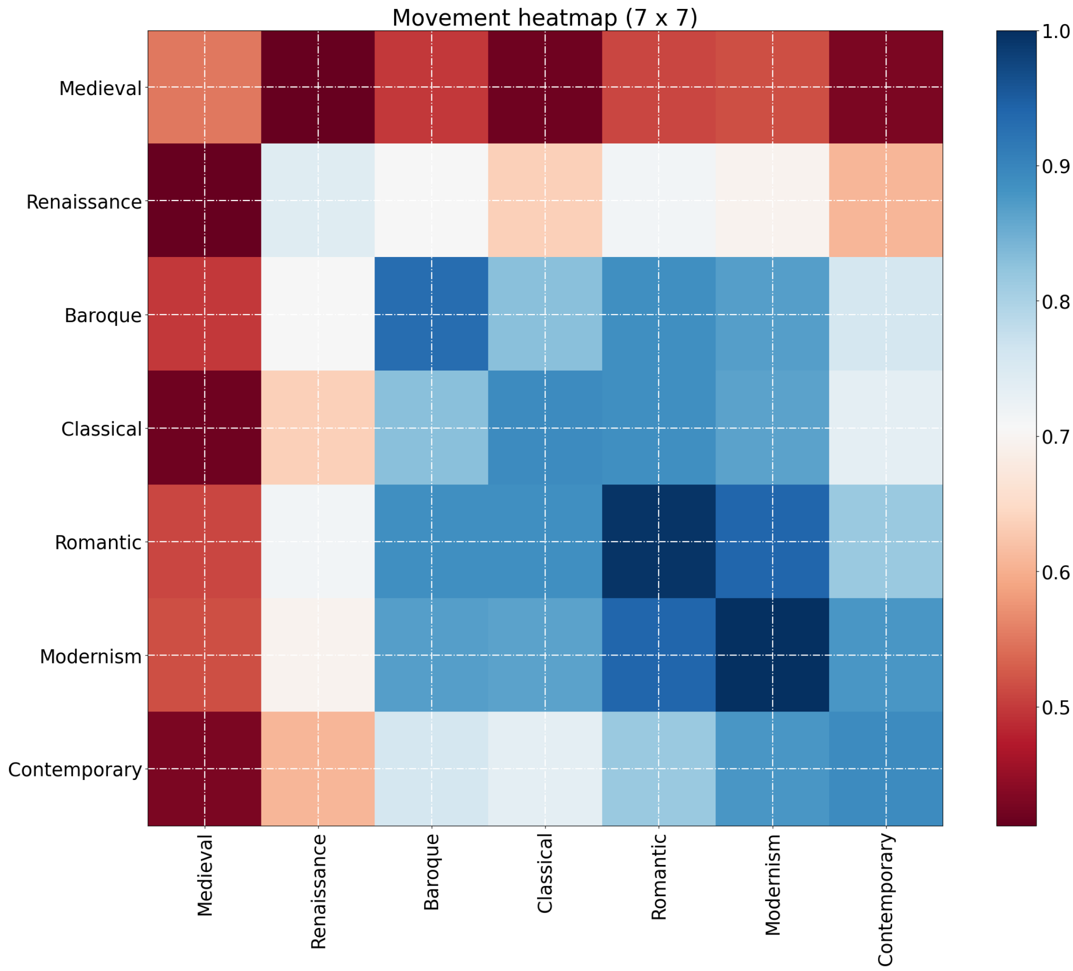

Figure 7: Eras Heatmap

4.2 Playlists

Examples of playlists are in the images 8 9 10 below. In the columns, the various details are highlighted: the Q-id number assigned by Wikidata, the era to which the composer belongs, the popularity zone in which it is placed, the normalized absolute popularity value and the number of crossings he/she has in total, the normalized logarithmic distance, the linear distance he/she has with the composer that follows and finally the title. The number "-4.80046e-016" (or similar) which sometimes appears, is evidently a very small one; it is a minimum offset threshold to avoid the argument being null in logarithmic operations. It is considerable equal to 0, or very close to it.

name	nid	movement	zone	pop_value	crossings	log_dist	lin_dist	title
"Robert Schumann"	"Q7351"	Romantic	0	0.423983	4318	0	0	"Symphony No. 2 in C major Op.61"
"Carl Ruggles"	"Q958763"	Modernism	8	0.00038963	4	0.08043	2	""Men and Mountains""
"Joseph Lanner"	"Q44927"	Romantic	6	0.00861491	52	0.00115	1	"Styrian Dances"
"Dmitri Kabalevsky"	"Q153776"	Modernism	4	0.0220588	91	0.00115	1	"Concierto para piano y orquesta nº 3 en Re mayor Op. 50 : mov. 3 º Presto "
"Rodion Shchedrin"	"Q165706"	Modernism	5	0.0131296	57	0.00115	1	"concerto n. 1 per orchestra (naughty limeriks)"
"Eugen Suchoň"	"Q46100"	Modernism	6	0.00878571	32	0.00115	1	"Concertino for Clarinet and Orchestra"
"Charles Gounod"	"Q180278"	Romantic	2	0.0699432	363	0.08043	2	"Meditation sur le premier prelude de Bach (Ave Maria)"
"Jean Sibelius"	"Q45682"	Modernism	0	0.310125	2560	0.08043	2	"Symphony No. 4 in A minor Op.63"
"Julian Anderson"	"Q2746982"	Contemporary	4	0.0270083	119	0.12748	3	"Bell Mass"
"Laurence Crane"	"Q6500554"	Contemporary	7	0.0017891	28	0.00115	1	"John White in Berlin"
"Christoph Graupner"	"Q115067"	Baroque	5	0.018086	81	0.00115	1	"Bassoon Concerto in C major GWV 301"

Figure 8: Zone Path Playlist with range 0.0-0.3

name	nid	movement	zone	pop_value	crossings	log_d	lin_d	title
"Vincenzo Righini"	"Q1348369"	Classical	8	-4.80E-16	1	0	0	"La selva incantata "
"João de Sousa Carvalho"	"Q2247998"	Classical	8	0.00038963	2	0	0	"Organ Sonata in D major"
"Maria Teresa Agnesi Pinottini"	"Q447804"	Classical	9	-4.80E-16	1	0	0	"Alessandro nell' Indie Acto III Aria de Erissena ' Son confusa pastorella' "
"Julije Bajamonti"	"Q3504283"	Classical	7	0.0017891	6	0	0	"Symphony in C major"
"Maddalena Laura Sirmen"	"Q454061"	Classical	7	0.00391072	14	0	0	"Duet No 6 in C for Two Violins"
"Maria Antonia of Bavaria"	"Q61649"	Classical	9	-4.80E-16	4	0	0	"Talestri regina delle Amazzoni (Sinfonia)"
"Josef Fiala"	"Q742748"	Classical	7	0.0021783	12	0	0	"Cor anglais concerto in E flat major"
"Jiří Antonín Benda"	"Q213535"	Classical	7	0.00311402	16	0	0	"Symphony no.5 in G major (1st mvt)"
"Johann Wilhelm Wilms"	"Q540689"	Classical	6	0.00497417	18	0	0	"concerto in mi maggiore per pianoforte e orchestra op. 3"
"Jean-Baptiste Krumpholz"	"Q697663"	Classical	9	-4.80E-16	3	0	0	"Concerto no.6 pour Harpe et Orchestre: 1st mvt; Allegro Moderato)"

Figure 9: Movement Playlist (Classical) with no intersection

name	nid	movement	zone	pop_value	crossings	log_d	lin_d	
"Wolfgang Amadeus Mozart"	"Q254"	Classical	0	0.799969	10312	0	0	"Piano Concerto No. 19 in F major K.459"
"Giuseppe Verdi"	"Q7317"	Romantic	1	0.197232	1717	0	0	"sinfonia in do"
"Gerald Finzi"	"Q713025"	Modernism	2	0.0663746	288	0	0	"God is gone up (3 Anthems Op.27)"
"Grażyna Bacewicz"	"Q230978"	Modernism	3	0.041244	233	0	0	"Overture"
"Wynton Marsalis"	"Q273076"	Contempor	4	0.0194617	74	0	0	"Hymn / Scripture "
"Hans Huber"	"Q673938"	Romantic	5	0.0120637	41	0	0	"Eine Lustspielouverture op 50"
"Mikalojus Konstantinas Čiurlionis"	"Q297137"	Romantic	6	0.00466394	25	0	0	"En el bosque "
"Paloma fata"	"Q240324"	Contempor	7	0.0016427	9	0	0	"Mouth to Mouth"
"Denis Matsuev"	"Q739505"	Contempor	8	0.00038963	1	0	0	"Jazz Improvisation"
"Martijn Padding"	"Q768085"	Contempor	9	-4.80E-16	1	0	0	"Softly Bouncing"

Figure 10: Zone Playlist

4.3 Statistics

This section shows the statistical analyses performed to understand the characteristics of the data collected by the 3 providers. For many of them, the overall result and the result for the individual radio will be shown.

It can be seen that the last two in the ranking are always the Medieval and Renaissance eras, which are present only in small percentages. It is necessary to point out the disparity in the amount of data collected by the providers, who therefore contribute differently to the global top chart, which sees the Modernism era in the lead.

4.3.1 Top Eras

The following plots show the frequency hierarchy of of musical eras.

In fact, the x-axis represents not the number of composers but the number of performances, which are the actual executions of their pieces.

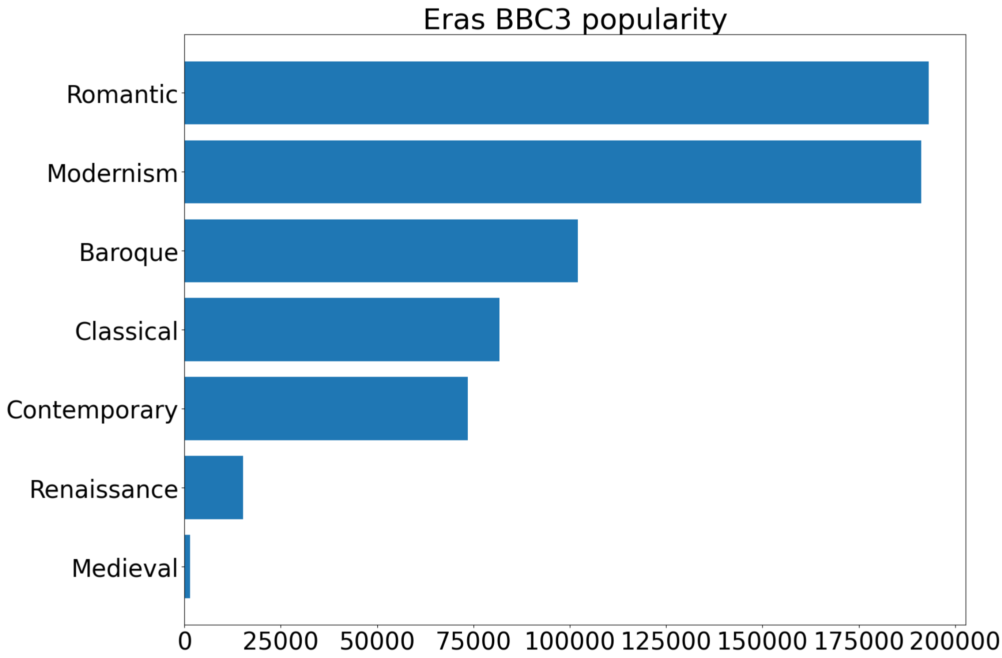

Figure 11: Top eras BBC3

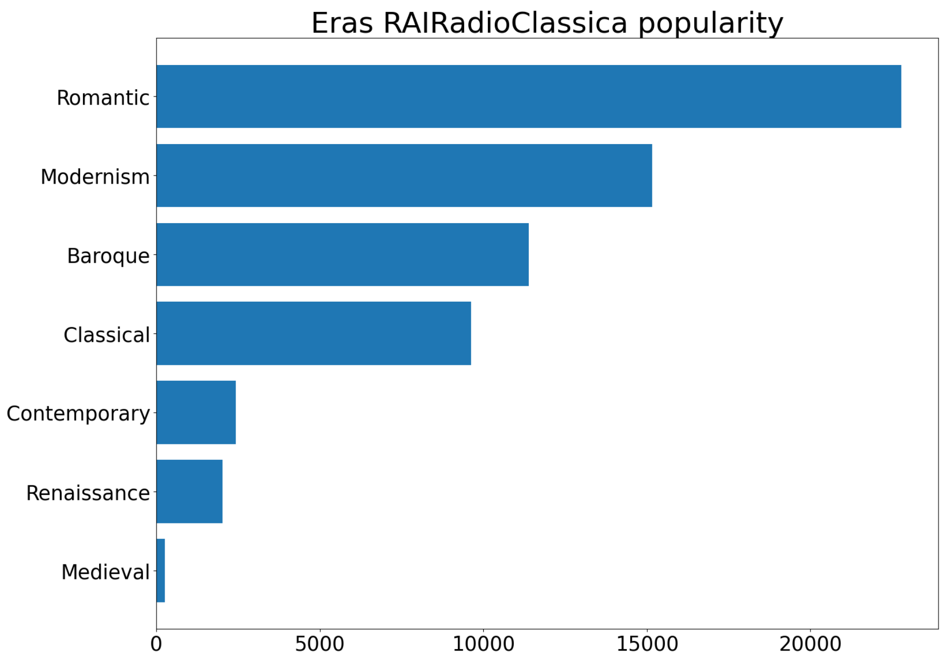

Figure 12: Top eras RAI

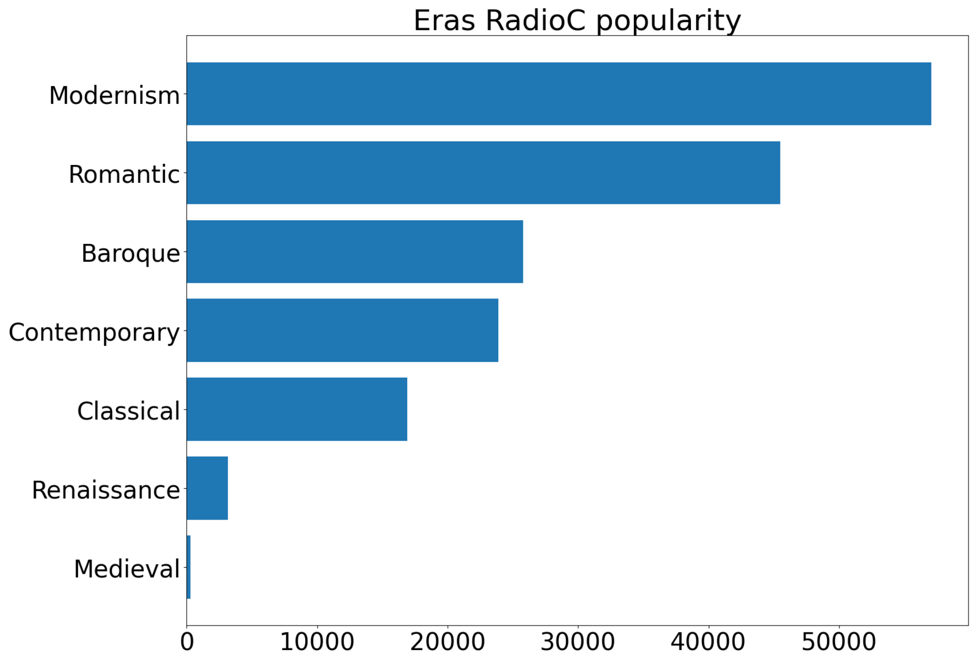

Figure 13: Top eras Radio Clasica

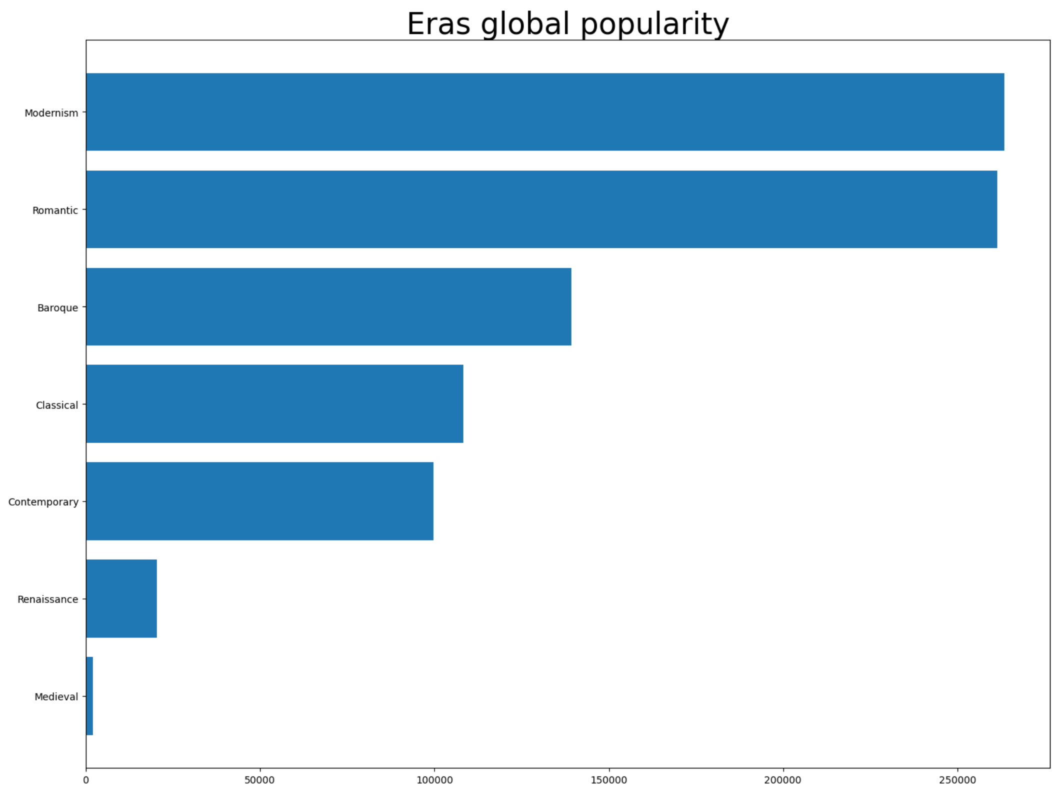

Figure 14: Top eras global

4.3.2 Top Composer

These charts (Figs: 15, 16, 17, 18) contain the top 20 composers, sorted by popularity. The number on the x-axis again represents how many times they were broadcasted.
As the heatmaps had already suggested there is a fairly small group of them that appear numerous times in the playlists offered by radio stations. These names always occupy the first locations of the ranking with small differences in position. The difference between the popularity index of the first and subsequent ones is very marked; in general, the trend follows a decreasing exponential curve.

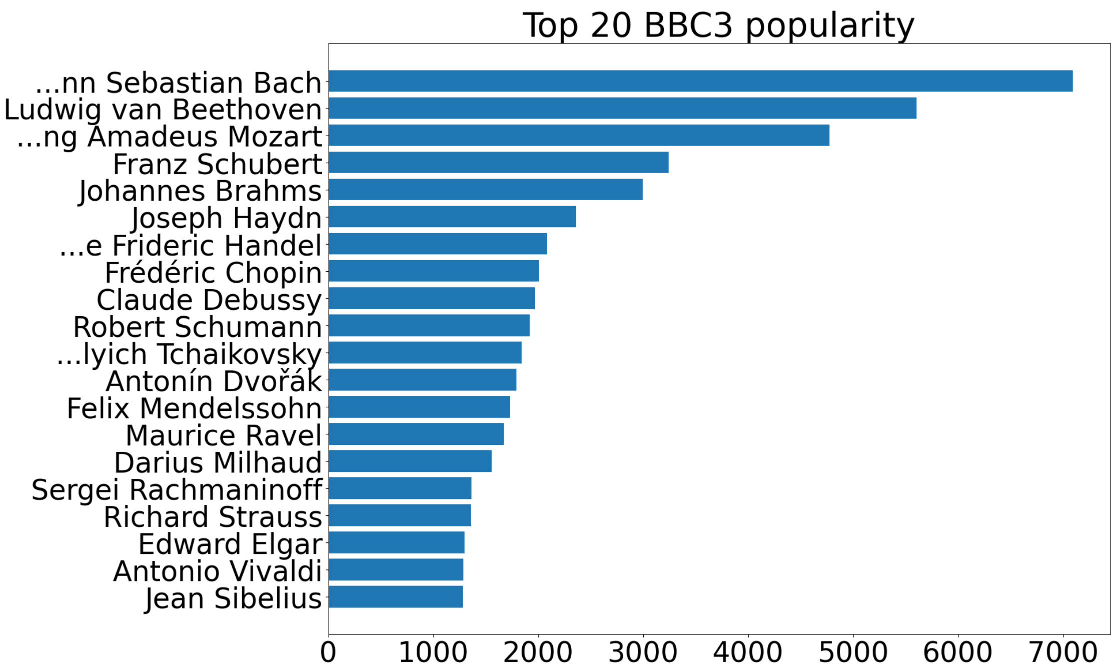

Figure 15: Top 20 composers BBC3 relative to the number of performances

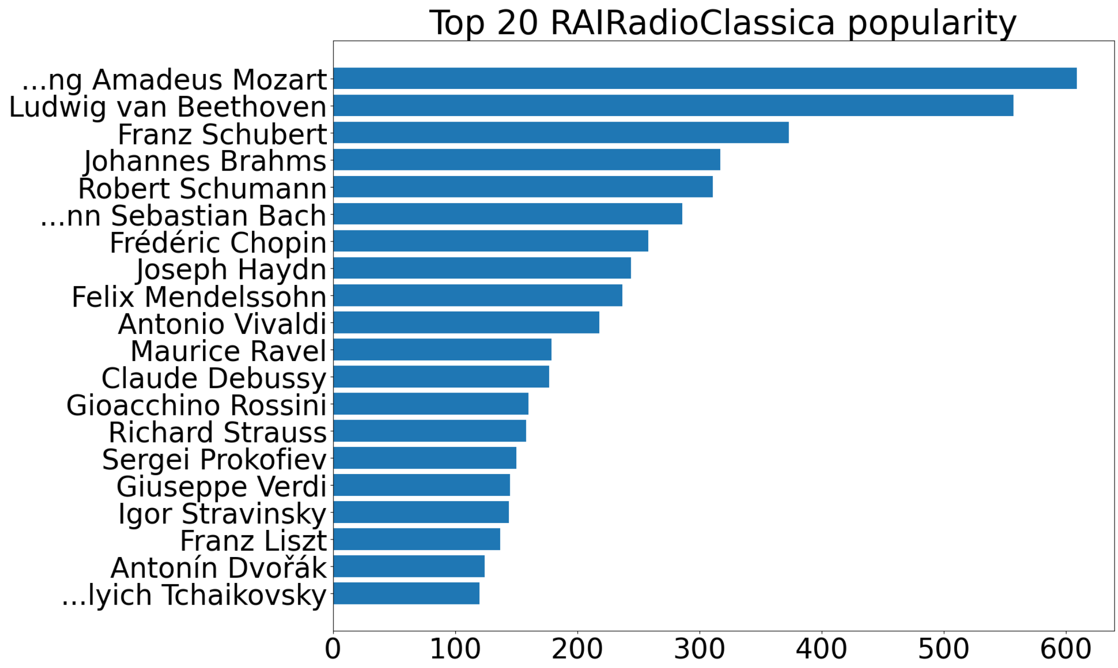

Figure 16: Top 20 composers RAI relative to the number of performances

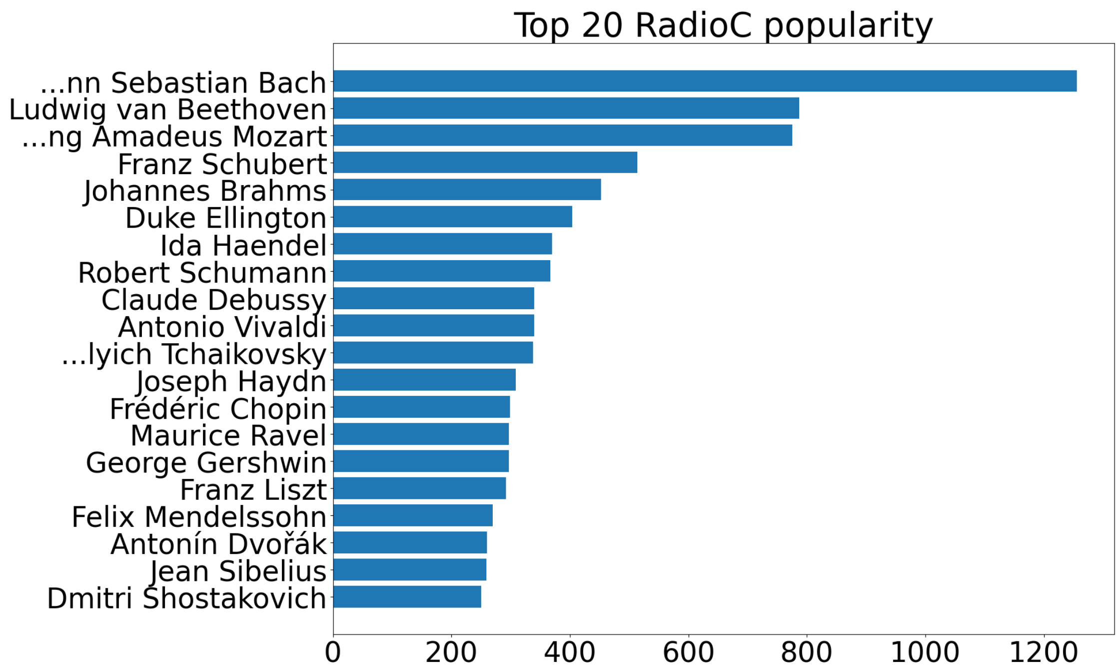

Figure 17: Top 20 composers Radio Clasica relative to the number of performances

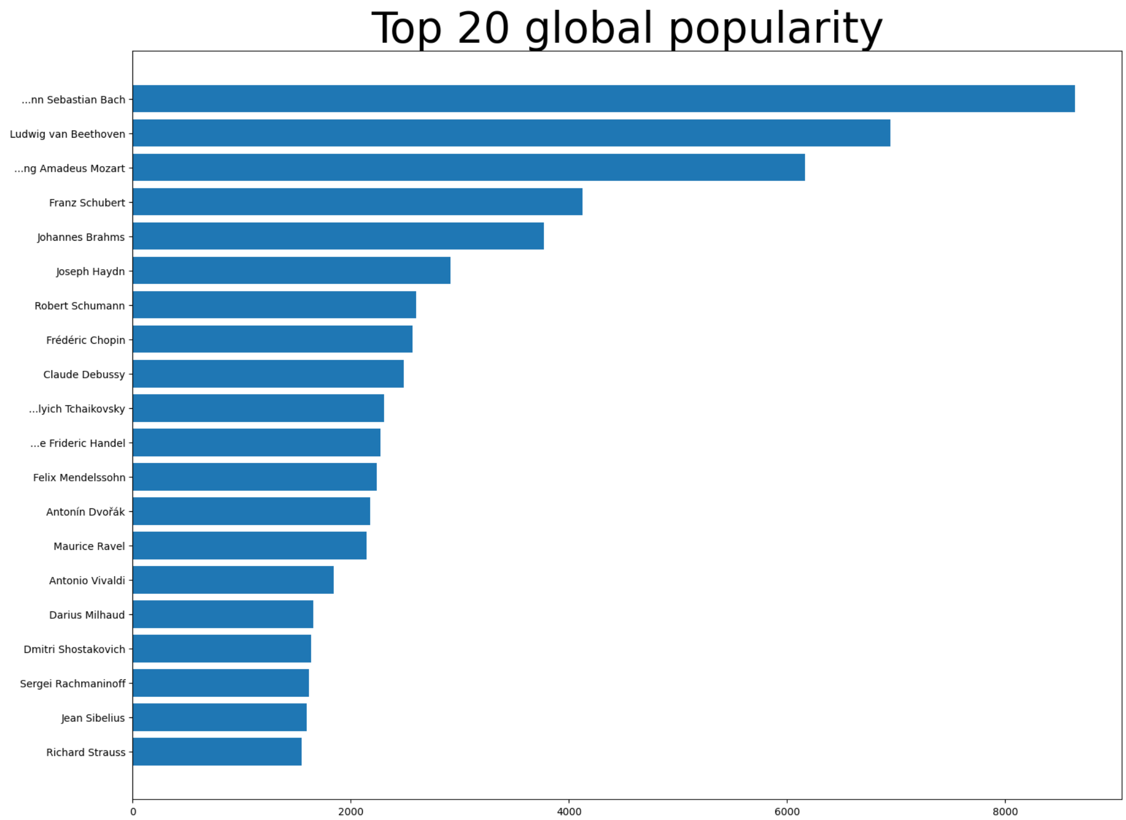

Figure 18: Top 20 composers

4.3.3 Country distribution

Taking advantage of the property extracted from the composers wikidata page, the geode map with the geographical distribution of the composers in the Dataset is shown below.

It is a list of 7300 rows, containing latitude and longitude of the exact location. The color intensity of the bubbles varies according to the concentration of data in that particular location. is evidently highly concentrated in 'Middle Europe'; the detail of the continent is given below.
This distribution reflects where providers come from, so they are quite nationalistic in their choices.

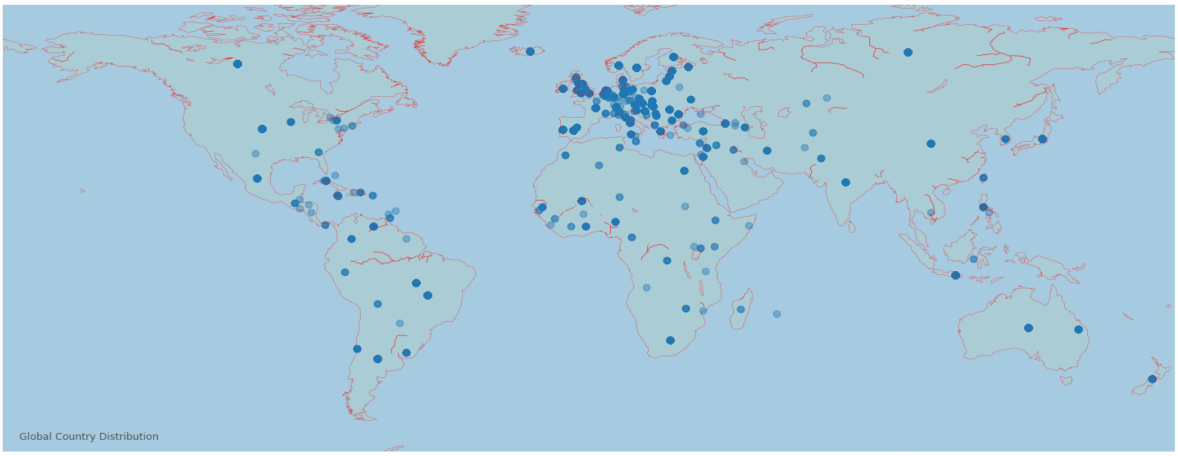

Figure 19: Geographical Distribution

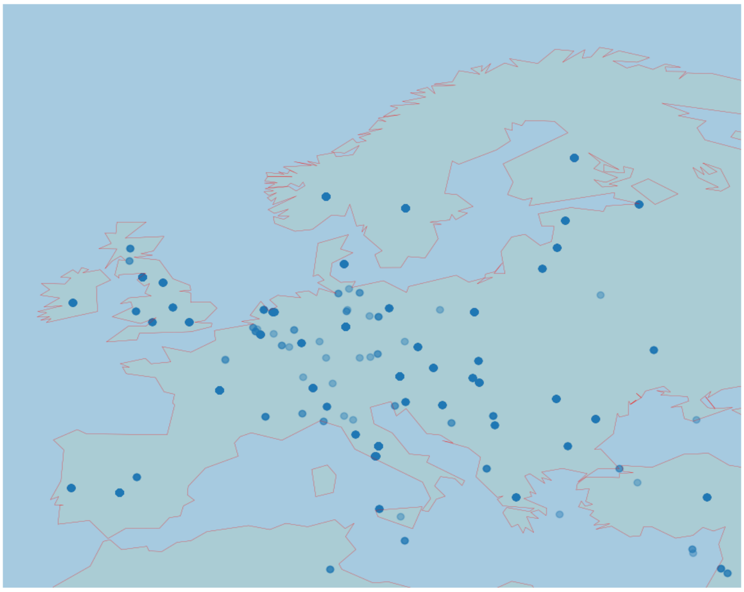

Figure 20: European Distribution

4.3.4 Gender Distribution

Gender distribution was also examined.

The trend is very clear and consistent: men far outnumber any other gender. Observing the source data there is a clear disproportion in the gender distribution over the total number of composers. The fact is greatly accentuated by looking at performances i.e., how much male composers are broadcasted compared to women, non-binary and others. So not only was the initial situation highly imbalanced, but over time it has only worsened.

The data was collected for each individual provider, so the results must be calibrated against the number of inserts of each. Also, this feature, more than others, is obviously affected by what is reported on Wikidata (which is why there are "unknown").

Total Gender Distribution	
Sex or Gender	Number
Male	6413
Female	863
Transgender	5
Non-Binary	2
Unknown	8

Table 2: Total gender distribution

Gender Distribution with respect to the performances			
Sex or Gender	Rai	Radio Clasica	BBC
Male	12617	33324	124399
Female	115	1162	7242
Transgender	0	0	33
Non-Binary	0	0	10
Unknown	4	35	8

Table 3: Gender Distribution for each provider

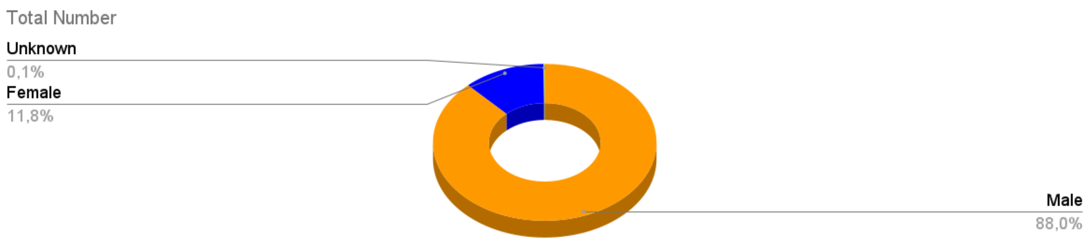

Figure 21: Gender distribution out of ~7300 composers

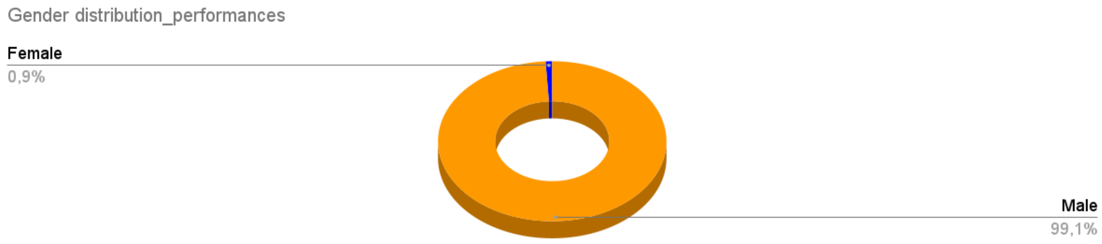

Figure 22: Gender distribution counting the performances

4.3.5 Playlists' Statistics

To make qualitative statistics on the obtained playlists, an algorithm was built that generates 10000 of them, of each expected type Random, Zone, Zone-Path, Movement, Movement-Path with variable parameters.
As for composers, rankings of Top composers, Top era were then collected and in addition average popularity values were extracted for each mode.

Considering the Zone path playlists the true algorithmic core of this project, it is interesting to observe its characteristics (See Fig. 23). It is noticeable that some names among the top 40 composers in this list are different from those in the previous statistics. The x-axis represents their presence as a percentage of the 10000 playlists generated; so the first, Cole Porter appeared about 1.3% of the time.

Moreover, the exponential trend of the curve highlighting their popularity is much less pronounced. Somehow then, the fame has been redistributed, thanks to the zoning and the Gini coefficient re-scaling the positions. This is the first step toward greater variety and fairer representation.

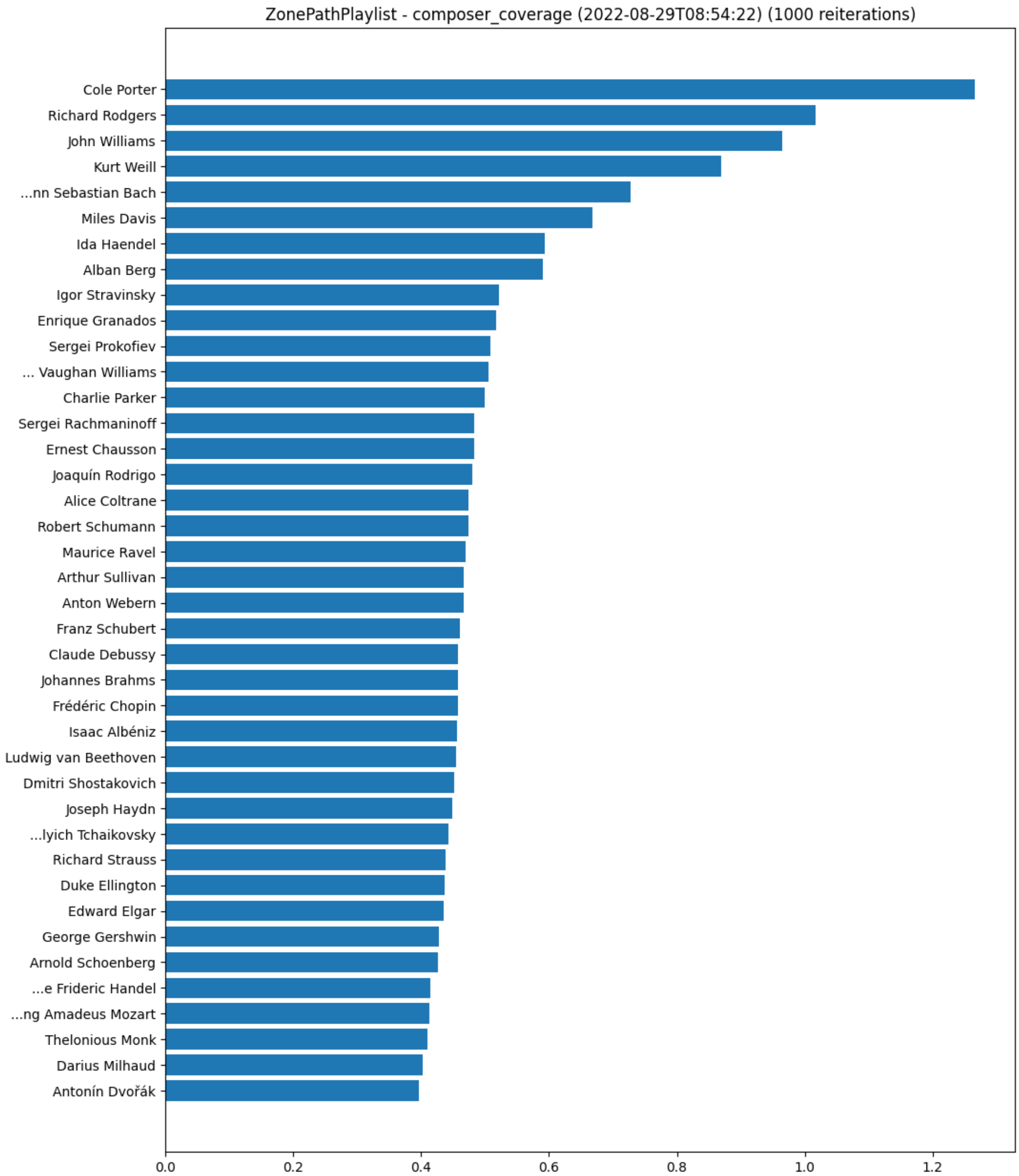

Figure 23: Top 40 Composers, 10k playlists generated

Chapter 5

Discussion

Recalling the research questions that guided the project, it can be said at this point that a new recommendation algorithm model for Classical Music is possible.
The laborious construction of a classical music database complete with titles, composers with related general information proved to be a very arduous task. Nevertheless, it turned out to be a rich source of novel and unexpected information.

They reveal not only that there are systemic problems with this genre of music and its representation but also that it lends itself well to being a cue for new conscious processes of cultural approach or reinforcement.
The tool by which this is implemented are the playlists generated from the collection of metadata. The arrangements made have led to the actual and potential creation of new lists of very varied and to most unknown names and titles.

5.1 Limitations

Despite all the measures taken, this thesis work is not immune to limitations. Some arise from the sources themselves, others from the methodology applied, which inevitably affects the outcomes. There is no doubt that many steps of this process generate biases.

A possible aspect to notice regards the origin of the data to fill the database. The providers used for data collection were three central European radio stations. This certainly somewhat limited the "geographic coverage" of the data of interest and thus cannot claim to be a fully representative dataset of classical music (at least at the European level).

Most of the partialities affected more or less directly the generation of the playlists. For instance, the normalization of composers' names and the subsequent search for further information is done in the first 30 results of Wikidata but lesser-known artists, who could potentially provide much variety in playlists, are often not present on it and their name may be beyond the search range. This generates a *popularity bias*, which is difficult to avoid.
This is even more true for titles, which do not even have a dedicated page, except for the great masterpieces of the most famous musicians and are unfortunately often discarded in the database entry.

As explained earlier, the collection of titles has been one of the most difficult challenges faced in this thesis; the reason is surely the ever-changing way in which providers are used to report them, which does not allow the writing of a robust and consistent algorithm to parse them. Since a snippet of the original file was previously shown, to get an idea of the quality of the designed method (or vice versa its fallibility), by means of a query, the parsed titles and the relative number of times they appear in the database were observed at random. The image perfectly gives the sense of the unpredictability of writing providers and the amount of elements that cause the test code to fail.
The most prominent elements of discordance between titles that are the same but are not recognized as such have been highlighted in red (See Fig. 24).

Composer	Title	#
Zoltán Kodály	Variations on a Hungarian Folk Song (The Peacock)	1
Zoltán Kodály	Variations on a Hungarian Folksong 'The Peacock'	2
Zoltán Kodály	Variations on a Hungarian folksong (The Peacock) for orchestra	1
Zoltán Kodály	Viennese Musical Clock (Hary Janos Suite Op.35a)	1
Zoltán Kodály	Viennese Musical Clock (Hary Janos Suite Op 35a)	1
Wolfgang Amadeus Mozart	Violin Sonata in E minor K 304	1
Wolfgang Amadeus Mozart	Violin Sonata in E minor K 304 arr Oboe	4
Wolfgang Amadeus Mozart	Violin Sonata in E minor K304	1
Wolfgang Amadeus Mozart	Violin Sonata in E minor KV304: 1st movement - Allegro	1
Wolfgang Amadeus Mozart	Violin Sonata in E minor K 304	2
Wolfgang Amadeus Mozart	Violin Sonata in E minor K.304 (1st mvt)	2
Wolfgang Amadeus Mozart	Violin Sonata in E minor K304 (Allegro)	1
Sylvius Leopold Weiss	Sonata in A minor Lbl.29 'L'infidele'	2
Sylvius Leopold Weiss	Sonata in A minor Lbl.29 (L'infidele) for lute	1
Sylvius Leopold Weiss	Sonata in A minor Lbl.29 (L'infidele) for lute [sic]	1
Sylvius Leopold Weiss	Suite L'Infidele	1

Figure 24: Result of title parsing

Another relevant obstacle hindering titles normalization, which constitutes a limitation, is the language.

Each provider is used to translating song titles into the country's official language; this makes it very arduous to match them, avoiding "duplication."

At present, for obvious reasons, an algorithm that succeeds in matching something like "Sinfonía nº 2 en Mi bemol mayor" (ESP), "Symphony no.2 in E-flat Major" (EN) with "Sinfonia n.2 in Mi bemolle Maggiore"(IT), has not been developed.

The internal division of the database also contributes since areas of popularity are not equally populated. As mentioned earlier zone 0 contains only 50 names, versus the more crowded end zones where, however, probably the most interesting and rare names reside. Nevertheless, this implies that, except by explicit request, in the serial generation of hundreds or thousands of playlists, the first zones will be adequately or even totally represented, while it is possible that there may be little-known composers who are never picked out.

Again, this is the cause of a *popularity* bias that could be avoided by limiting name searches to only the areas of lesser-known artists.

Considering, however, not only the computational generation but also the real listening done by users, it is important to emphasize the internal balance that playlists

must have.
In fact according to some studies [22], the preference or familiarity with some songs, are factors that influence the perceived quality of a playlist.
Thinking especially of newcomers, considering that the genre is rather constrained to Western Classical Music with some openings to jazz and bossanova, familiarity depends very much on the popularity zone of extraction.
The assumption is that, in general, composers from zone 0 are, at least by hearsay, more familiar than others.
As a result, a playlist consisting only of lesser-known composers might be too difficult to listen to. A trade-off is therefore necessary, mixing elements already known with elements of novelty to intrigue the listener with more researched elements and unfamiliar titles.

Chapter 6

Conclusions

Starting with the music programming of three European radio stations, through the collection of metadata and their through Wikidata, a fairly large Classical Music Database of 7300 entries was constructed. From it, almost infinite playlists can now be generated, which over time can be increasingly rich and customizable.
This will require further data cleansing, an increase in sources, ideally extendable to the entire world and a massive contribution from users and fans.

It is crucially important that the contributions made so far (and future contributions as well) from this thesis therefore remain free and accessible to all. For this reason, the entire source code of the project is available online at the link **https://github.com/nabarlet/Master_Thesis_WCM** . The written code is duly commented to be easy to understand.
All this work is under Creative commons - NonCommercial - Attribution - Share-Alike license.

6.1 Future Work

As already underlined, the aspiration to build a Western classical music database, as omni-comprehensive as possible, certainly needs more information, which can be obtained in the future by extending the collection to at least one broadcaster per

European country.

Since Wikidata was use to retrieve information it is necessary to note again how it also affected the results, due to potentially incorrect or worse, absent information. Work to be done in the future will necessarily include a *waste reduction* algorithm, and possibly implement statistics on the excluded items as well, to derive useful information for improvement.

This will involve writing and implementing new, source-specific parsers. As widely shown it is likely that title problems will remain unless the equivalent of Wikidata for composers, an authority, is found.

A mild attempt has already been made during the research and development work of this thesis using MusicBrainz [37]. Unfortunately, the results were not satisfactory for several reasons.
The search using the API generated many false positives, that is, other titles containing words belonging to the original phrase, even as first results. This was to be expected, because the source string (due to parsing problems mentioned earlier) contained misleading information.
However, at this point even once matching results are found it is still not clear how to recognize the "right" result from the "wrong" one, the official one from the partial one.
It must also be accepted that MusicBrainz does not aim to be a complete platform for classical repertoire, so it remains deficient in many aspects concerning metadata. Therefore, one of the future goals is to find a way to normalize titles and through that to contribute to the database growth of similar platforms.

Another possible area for improvement concerns playlists. First, new settings for their customization can be entered: this could include some of the parameters already collected (e.g., a playlist of only American composers or only women composers of the 1800s and so on) or imply further collection of metadata.
Afterwards, drawing from the literature, criteria can be established for their evaluation subsequent adjustment by real users: fans, newbies, and listeners of the radio

stations used as sources. Using questionnaires, one could then qualitatively and quantitatively compare the offerings proposed by the broadcasts with those performed by the algorithm.
To accomplish the same evaluation goal, the statistics traced on the results could be compared with those performed on the playlists of major music streaming platforms, such as Spotify, Deezer or Last.fm.

Finally, it is believed that the utilization of advanced Deep Learning technologies could give a major boost to these kinds of projects. With proper training of more complex algorithms, the logistical difficulties of lower-level codes could be overcome, obtaining stunning results that are more easily extended and replicable.

List of Figures

List of Tables

Bibliography

[1] Abdollahpouri, H. & Mansoury, M. Multi-sided exposure bias in recommendation. *arXiv preprint arXiv:2006.15772* (2020).

[2] Pangrazi, T. *Ritorno al cielo - L'estetica musicale in Italia dal trecento al primo novecento.* (Mimesis, 2016).

[3] Adorno, T. Sociology of music. In *The Routledge Reader on the Sociology of Music*, 67–72 (Routledge, 2015).

[4] Wang, J. Classical music: a norm of" common" culture embedded in cultural consumption and cultural diversity. *International Review of the Aesthetics and Sociology of Music* 195–205 (2016).

[5] Freitas, C. *et al.* Neural correlates of familiarity in music listening: A systematic review and a neuroimaging meta-analysis. *Frontiers in neuroscience* **12**, 686 (2018).

[6] Adorno, T. W. & Horkheimer, M. *Dialectic of enlightenment*, vol. 15 (Verso, 1997).

[7] Adorno, T. W., Rognoni, L. & Manzoni, G. *Introduzione alla sociologia della musica* (Einaudi, 1994).

[8] Couldry, N. *Inside culture: Re-imagining the method of cultural studies* (Sage, 2000).

[9] Janczukowicz, T. Idagio. URL `https://www.idagio.com/`.

[10] Marín, M. Á. Challenging the listener: How to change trends in classical music programming. *Números* (2022).

[11] Cruz, A. F. T. & Coronel, A. D. Towards developing a content-based recommendation system for classical music. In *Information Science and Applications*, 451–462 (Springer, 2020).

[12] Agid, P. & Tarondeau, J. *The management of opera: An international comparative study* (Springer, 2010).

[13] Kolb, B. M. The effect of generational change on classical music concert attendance and orchestras' responses in the uk and us. *Cultural Trends* **11**, 1–35 (2001).

[14] Botstein, L. Music of a century: museum culture and. *The Cambridge History of Twentieth-Century Music* 40 (2004).

[15] Bachtrack ltd. URL `https://bachtrack.com/`.

[16] Mir datasets. URL `https://github.com/ismir/mir-datasets/blob/master/outputs/mir-datasets.md`.

[17] Fundación "juan march". URL `https://www.march.es/es`.

[18] Metropolitan opera house (usa). URL `http://archives.metoperafamily.org/archives/frame.htm`.

[19] Pauws, S. & Eggen, B. Pats: Realization and user evaluation of an automatic playlist generator. In *ISMIR*, vol. 2, 222–230 (2002).

[20] Dias, R., Fonseca, M. J. & Cunha, R. A user-centered music recommendation approach for daily activities. In *CBRecSys@ RecSys*, 26–33 (2014).

[21] Herrera, P., Resa, Z. & Sordo, M. Rocking around the clock eight days a week: an exploration of temporal patterns of music listening. In *1st Workshop On Music Recommendation And Discovery (WOMRAD), ACM RecSys, 2010, Barcelona, Spain* (2010).

[22] Dias, R., Gonçalves, D. & Fonseca, M. J. From manual to assisted playlist creation: a survey. *Multimedia Tools and Applications* **76**, 14375–14403 (2017).

[23] Bonnin, G. & Jannach, D. Automated generation of music playlists: Survey and experiments. *ACM Computing Surveys (CSUR)* **47**, 1–35 (2014).

[24] Antony, S. & Jayarajan, J. Survey on playlist generation techniques. *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)* **3** (2014).

[25] Wikidata. URL `https://www.wikidata.org/wiki/Wikidata:Main_Page`.

[26] Sqlite3. URL `https://sqlite.org/version3.html`.

[27] Rai radio classica. URL `https://www.raiplaysound.it/radio3classica`.

[28] Bbc radio 3 website. URL `https://www.bbc.co.uk/sounds/play/live:bbc_radio_three`.

[29] Radio clásica. URL `https://www.rtve.es/radio/radioclasica/`.

[30] Twitter. URL `https://twitter.com/home`.

[31] Bbc radio 3 musicbot. URL `https://twitter.com/BBCR3MusicBot`.

[32] An easy-to-use python library for accessing the twitter api. URL `https://www.tweepy.org/`.

[33] Beautifulsoup. URL `https://beautiful-soup-4.readthedocs.io/en/latest/`.

[34] Wikidata query service. URL `https://www.wikidata.org/wiki/Wikidata:SPARQL_query_service/queries/examples`.

[35] Geopy. URL `https://pypi.org/project/geopy/`.

[36] Gini coefficient. URL `https://en.wikipedia.org/wiki/Gini_coefficient`.

[37] Musicbrainz. URL `https://musicbrainz.org/`.