Towards Context-Aware Recommendation Systems in Music Streaming Platforms

A comparison of context related passive listening behavior among the population of Copenhagen and Lisbon

Master's thesis

João André Mafra Tenera

Aalborg University
Sound and Music Computing

Sound and Music Computing
Aalborg University
http://www.aau.dk

AALBORG UNIVERSITY
STUDENT REPORT

Title:
Towards Context-Aware Recommendation Systems in Music Streaming Platforms: A comparison of context related passive listening behavior among the population of Copenhagen and Lisbon

Theme:
Music listening behavior

Project Period:
Spring Semester 2021

Project Group:
-

Participant(s):
João André Mafra Tenera

Supervisor(s):
Daniel Overholt

Copies: 1

Page Numbers: 75

Date of Completion:
October 9, 2021

Abstract:

This research focuses on the study of music listening behavior and context-aware recommendation systems. A conducted survey assumes the playback method of music (e.g. playlist or folder, radio) influences the activity context (working and not working). The target group of the survey is residents of Copenhagen and Lisbon. The results reveal a significant difference between the two groups on the overall preference for the playback method when working and not. Future work considers broadening this same research onto intention-aware and emotion recognition to complement its findings.

Contents

Preface

The research presented in this thesis aims to find meaningful information on music listening behavior among the population of two different cities and use it to improve context-aware music recommendation systems. The writer of this thesis would like to acknowledge the involvement of Daniel Overholt (supervisor), George Palamas, Prithvi Katan, and Sofia Dahl from the Department of Architecture, Design and Media Technology at Aalborg University Copenhagen. Also, all the participants of the research survey, my friends and family.

João André Mafra Tenera

João André Mafra Tenera
<jtener19@student.aau.dk>

Chapter 1

Initial Research

1.1 Recommendation systems

Recommendation systems are part of the field of information filtering system, a field focusing on the prediction of raking and preference of users in concerning an item. It does so by using an automated method to classify and remove redundant information from a stream of information before such information is presented to a user. Such a method allows for the managing of overloaded information. The user's profile in service provides characteristics for the automated system to decide what to present to the user. The system will access the information extracted from the user's profile, the content-based approach, or the social network of the user, the collaborative filtering approach. Both approaches can be used separately or in a hybrid setup. The collaborative filtering approach uses a model from a user's past behavior (rating of items and consumer choices) as well as other users with similar behavior. Such information is used for the prediction and presentation of item suggestions. The content-based filtering approach uses discrete information, such as tags, to recommend additional items that look similar by sharing these same properties. It is common for recommendation systems to apply both systems in a hybrid setup. The difference between these two methods can be exemplified by comparing the service of Last.FM and Pandora. Last.FM originates a music station through analysis of artists and individual tracks that are part of the user's listening history and then compares it to other users with similar listening behavior. Where Pandora applies to signal analysis techniques to identify songs' properties to find and present a music station with compatible properties and compliments the suggestion considering the user feedback over the songs being played.

1.1.1 Recommendation system techniques

Collaborative filtering

Has as a starting point the assumption that the agreement of people in the past is to be kept in the future, so if two people liked the same item in the past then they must both like a similar item in the future. This method does not require an understanding of its items online and how the users relate to them. For example, recommending a movie with a system that only knows the poster of the movie. It is therefore important when building a model from a user's behavior that a distinction is made between implicit and explicit data collection. Requesting users to rate, search, provide a single preference between multiple choices, and like or dislike a list of items is considered explicit data where the observation of a user's search history, consumption history, and social network is considered implicit data. Three main situations affect the performance of collaborative filtering. When a new item does not possess enough data, the accuracy for this item's recommendation drops, this is referred to as a cold start. Scalability might be an issue that refers to the large amount of computational power need for calculating the recommendations. Sparsity on the general ratings of a database such as having popular items with few ratings in comparison to other items of the database.

Content-base filtering

The go-to choice when dealing with a scenario where there's data about the item but not about the user. The recommendation is therefore addressed as a user-specific classification problem where the classifier learns from the user's rating based on features of the items. Keywords are used for the description of items and user profiles to indicate the preference of users. The recommendation takes place through the comparison of various new items, candidates for a new suggestion, with items rated previously by the user resulting in a suggestion based on the best match of both. A model of user preferences and a history of interactions with the recommendation system are the two most common pieces of information used to create a user profile. A vector space algorithm is used in the system to address a content profile of users, based on a weight vector of the features of an item. These weights represent the relevance that each feature has for the user. It is common practice to use average values from the vector of the rated item although more sophisticated methods can also be applied, this involves the use of machine learning techniques for providing a probability of the user preferences. It might be an issue of degrading the accuracy of the recommendation when the system is limited to the recommendation of content that is the same as the one the user already uses. Then the value of the recommendation decreases as other content types from other services can be recommended. The content-based recommendation system might include text mining, information retrieval, and sentiment analysis techniques to aid its recommendations, such methods are part of the so-called opinion-based recommendation system [26]. Deep learning might be applied here.

Knowledge-based system

A system that attempts to represent knowledge explicitly and allows for deriving into new knowledge. The system is therefore divided into two features: a knowledge base and an inference engine. The knowledge base represents facts about the world (frames, conceptual graphs, and logical assertions) and the inference engine allows for the inference of new knowledge (usually it is formed of if-then statements aggregated to forward or backward chaining processes) [69].

Hybrid recommendation system

These systems can be implemented for making collaborative and content-based predictions which then combine both, by merging both approaches into a single model. The main advantage is that such methods can be the solution to the shortcoming issues present in collaborative and content-based filtering. There are many techniques available when applying a hybrid recommending system such as combining scores from different recommendations numerically (weighted), choosing from different recommendations methods (switching), presentation of different recommendations together (mixed), using features gathered from different systems, and combining them into a single recommendation system (feature combination), computing a set of features and apply them as the input for a next technique (feature augmentation), avoid destruction of ties in the scores (cascade) and input a produced mode into a next technique (meta-level [12]).

Session-based recommendation system

This system uses the data of the user's interactions during a single session. It might be relevant to use when the user's history is not made available or simply not relevant for the session taking place. Uses mostly recurrent neural networks and other deep learning approaches as a base to its own generative sequential models.

Context-aware recommendation system

Context-aware recommendation systems try to incorporate context into conventional user-item space. The prediction of a rating can be modeled with the function:

$User \times Item \times Context \rightarrow Rating$

The search space is multi-dimensional and, therefore, computationally expensive. The challenge in front of such systems is to learn user preferences in different contextual situations. Capturing the contextual attributes suitable for the domain under consideration and incorporating them in the recommendation process is a key to develop such systems [47]. Literature on context defines that "the context is any information that characterizes a situation related to the interaction between humans, applications and the surrounding environment" [5]. A context can either be categorized as static, contextual attributes and their structure remains the same over some time, and dynamic, contextual attributes and/or their structure changes over some time (some contextual attributes may become obsolete and thus can be removed from the system) [2]. Also, context can be classified into fully observable, a recommendation system knowing everything about the contextual attributes including their structure, partially observable, only certain aspects of contextual attributes are known, and unobservable, no explicit knowledge about contextual attributes. Another method to include context is contextual preference elicitation and estimation. In this method, recommendation systems attempt to learn user preferences in different contextual situations. Learning depends on implicit or explicit feedback obtained from the users. For learning contextual preferences, the systems either use existing recommendation techniques or use machine learning techniques. There are three basic approaches named contextual pre-filtering, contextual post-filtering, and contextual modeling, all of which perform contextual preference elicitation and estimation [34]. In contextual pre-filtering, values of contextual attributes are used as constraints for the selection of ratings which are then used in conventional user-item space. Contextual post-filtering involves the generation of predicted ratings first and then uses contextual information while ratings are adjusted for every user. In the contextual modeling approach, contextual attributes are used in the process of prediction of the ratings by recommendation systems [47].

1.1.2 Evaluating recommendation performance

Recommendation systems can be evaluated in three ways: employing user studies, online evaluations (A/B tests), and offline evaluations. It is challenging to evaluate a recommendation algorithm due to the impossibility to predict accurate reactions of real users to the recommendations. Therefore, metrics computing the effectiveness of an algorithm are always imprecise. User studies use a small group of users (hundreds) to judge from recommendations (which they think is best) that use different approaches. A/B tests use a higher group of users (thousands) and make them a test on a real product. At least two different recommendation approaches are applied and the effectiveness of them is measured through conversion rates (the number of goal achievements per visitor) and click-through rates (the number of users clicking a link from all those who had access to the same link). Offline evaluations are based on already produced data such as a dataset containing information about ratings to an item. The effectiveness of the approach being tested is measured with the accuracy to predict the rating of the given dataset. When attempting to find the most accurate recommendation algorithm the following factors are considered to be important: diversity of the presented recommendations [83]; persistence by showing again previous suggestions [7]; privacy of user data; demographic data of the user [8]; fraud caused by the users' ability to participate in the recommendation system [46]; surprise in users when presented to a recommendation; trust due to transparency of how the recommendation algorithm works [57]; and labeling influencing the satisfaction of the user on the items (organic or automatic) [6].

Reproducibility

Reproducing the results of the evaluation of a recommendation algorithm is not always possible. This compromises the validity of the results and makes unclear the effectiveness of the algorithm which the performance measurement aims to do. The following suggestions address the issue [9]: a survey of other research fields; a common understanding of what reproducibility is; identification and understanding of what affects reproducibility; make experiments more comprehensive; reform publication practices; improve upon the development of a recommendation evaluation framework; establishment of best-practice guidelines.

1.2 Sentiment analysis

Sentiment analysis combines natural language processing (computer processing of human language), text data mining (deriving information from text), computational linguistics (computational modeling of natural languages), and biometrics (measuring and calculating concerning human characteristics).

1.2.1 Subjective and objective identification

The term objective refers to factual information (something that can be proven). An example of an objective sentence is "to be elected the president of the resident council, a candidate must be a resident for at least six months". In opposition, the term subjective refers to non-factual information, such as personal opinions, judgment and predictions [58]. Three types of attitudes were observed by Liv (2010), they were positive, neutral, and negative opinions [52]. An example of a subjective sentence is "the resident council must elect a president who is mature and can make wise decisions". The detection of subjectivity can be performed with automated learning methods separated into supervised and unsupervised machine learning. When applying these methods, metaphorical expressions [80], the discrepancy in the writing style, context (consideration on previous and following sentences), time (how long until a piece of information becomes outdated), words with less use and the constant growth of data must be considered. To train a classifier using machine learning, a dataset composed of text is used and the text can be either annotated or not. Manually annotated data may suffer from human biases regarding comprehension, errors due to concentration-related issued and required time to complete the task [63]. As an alternative, linguistic patterns can be extracted using bootstrapping methods [64], statistical techniques for estimating quantities about a population by averaging estimated from multiple small data samples.

1.2.2 Feature/aspect-based

Feature-based or aspect-based refers to determining the opinions or sentiments expressed on different features or aspects of entities such as those of a cell phone, a digital camera or a bank [36]. The advantage of feature-based sentiment analysis is the possibility to capture nuances about objects of interest so that different features can generate different sentiment responses [13]. Therefore, it is required to perform identification of relevant entities and extract their features/aspects, as well as evaluating if the opinion expressed for each feature is positive, neutral, or negative [52]. Also, the topic model is a type of statistical model used for discovering abstract topics occurring in a collection of documents and relies on deep learning for the automatic identification of features.

1.2.3 Application in recommendation systems

User-generated text, such as the one found in online social networking and commerce business services, provides a valuable source of the user's sentiment opinions about many items. Text generated by real users can therefore reveal the related feature/aspect of the item as well as the users' sentiments. It can be observed that the item's features/aspects described in the text perform the same rule as the meta-data in content-based filtering. A user may show different sentiments for different items with features in common, and a feature from an item can receive different sentiments from different users. Thus, user's sentiments on the features can be regarded as a multi-dimensional (multiple features/aspects) rating score, reflecting their preferences on the items. A hybrid recommendation system can be constructed based on the sentiments extracted from the user's text and the features/aspects the text comes from [40]. The motivation to recommend an item to a user can be done in different ways. The candidate item can have numerous common features with the user's preferred items [37]. Hence it has been concluded that combining a ranking score of similarity and sentiment rating can be constructed for each candidate item [40].

1.2.4 VADER (Valence Aware Dictionary and Sentiment Reasoner)

VADER [16] is a lexicon and rule-based sentiment analysis tool that is directed to sentiments expressed in social media. Examples of typical use cases for sentiment analysis with VADER include typical negations ("not good"), contractions as negations ("wasn't very good"), punctuation to signal increased sentiment intensity ("good!"), word-shape to signal emphasis (using ALL CAPS for words or phrases), degree modifiers to alter sentiment intensity (boosters such as "very" and intensity dampeners such as "kind of"), sentiment-laden slang words ("sux"), slang words as modifiers ("uber", "frigging", "kinda"), emoticons (":)"), utf-8 encoded emojis ("heart"), initialism and acronyms ("lol"). VADER contains examples of tricky sentences that may confuse other sentiment analysis tools. The NLTK (Natural Language Tool Kit) [10] to do sentiment analysis on longer texts (articles, novels, etc) to decompose paragraphs into the sentence-level analysis. The NLTK is a platform for building Python programs that work with human language data. It provided a suite of text processing libraries for classification, tokenization, stemming, tagging, parsing, and semantic reasoning. VADER also contains a concept for assessing the sentiment of images, video, and a variety of tagged multimedia content. Consequently, a translating service combined with VADER can provide a multilingual analysis setup.

VADER's score

The most common type of score used for sentiment analysis is the compound score. It is computed by summing the valence score of each word in the lexicon and then normalized to be between -1 (most extreme negative) and +1 (most extreme positive). The threshold values are: positive sentiment: compound score >= 0.05; neutral sentiment: compound score > -0.05 and compound score < 0.05; negative sentiment: compound score <= -0.05.

1.3 Music listening behavior

1.3.1 A survey on music listening and management behavior [42]

Major findings in this research show that elements like the familiarity of songs, how distracting they are, how much they match the listener's mood, and the desire of changing the mood within one listening session, are all affected by the activity during which music is listened to. While people want to have options for manipulating the above elements, at the same time, they prefer a minimal amount of interaction in general. Active listening, commuting, exercising, work and housework were the activities aimed for study and for the playback of music methods the following were taken into consideration: song after song; artist, album or genre; playlist or folder; shuffle on collection; online recommendation; radio; surfing the internet; playing video games. Mentions to other studies present [11] that the most popular places for listening to music were the car, followed by the living room and work and that the distinction between activities during which music is listened to can be the amount of attention the activities need. The study hypothesizes the having to pay or not pay attention to the activity would affect four aspects. Those aspects were the familiarity of songs, pickiness of the listener (importance of song), preference for constant or various moods, and what is the amount of desired interaction. It was concluded that familiar songs are preferred in general for both attention and non-attention activities, and in the case of attention activities, participants strongly preferred familiar songs with nearly 0% preferring new ones. Along with the findings, it is concluded that mood, genre, and artist are the most important elements for creating playlists.

1.3.2 Quantitative study of music listening behavior in a social and affective context [81]

This study initially states that a scientific understanding of emotional experience requires information on the contexts in which the emotion is induced and that one of music's primary functions is to regulate the listener's mood. Hence, an individual's short-term music preference may reveal the emotional state of the individual. The study attempts at presenting a computational model of the latent structure underlying music listening and mood regulation. It is noted that the study work focused on the emotional state of an individual before listening to music instead of the emotion evoked as a result of music listening. MER (Music Emotion Recognition) is a field focusing on tagging songs with emotional labels that listeners perceive when listening to the song. In this case, it is assumed that the tag is not biased by the prior emotional state of the listener. Therefore, it is concluded that to create an effective emotion-based music recommendation system, an understanding of the interplay between music emotion and user mood is needed. UMR (User Mood Recognition) focuses on the emotion an individual feels in response to a stimulus (not necessarily music). The service AllMusic is addressed at this point for its 190 emotional tags for music available. The study assumes the user data must be recorded spontaneously and social media is selected as the best place with access to such data. Blogs are also mentioned as a source of spontaneous information of the everyday context of people. In continuation, a tripartite relationship between music emotion, used mood and music listening behavior is modeled.

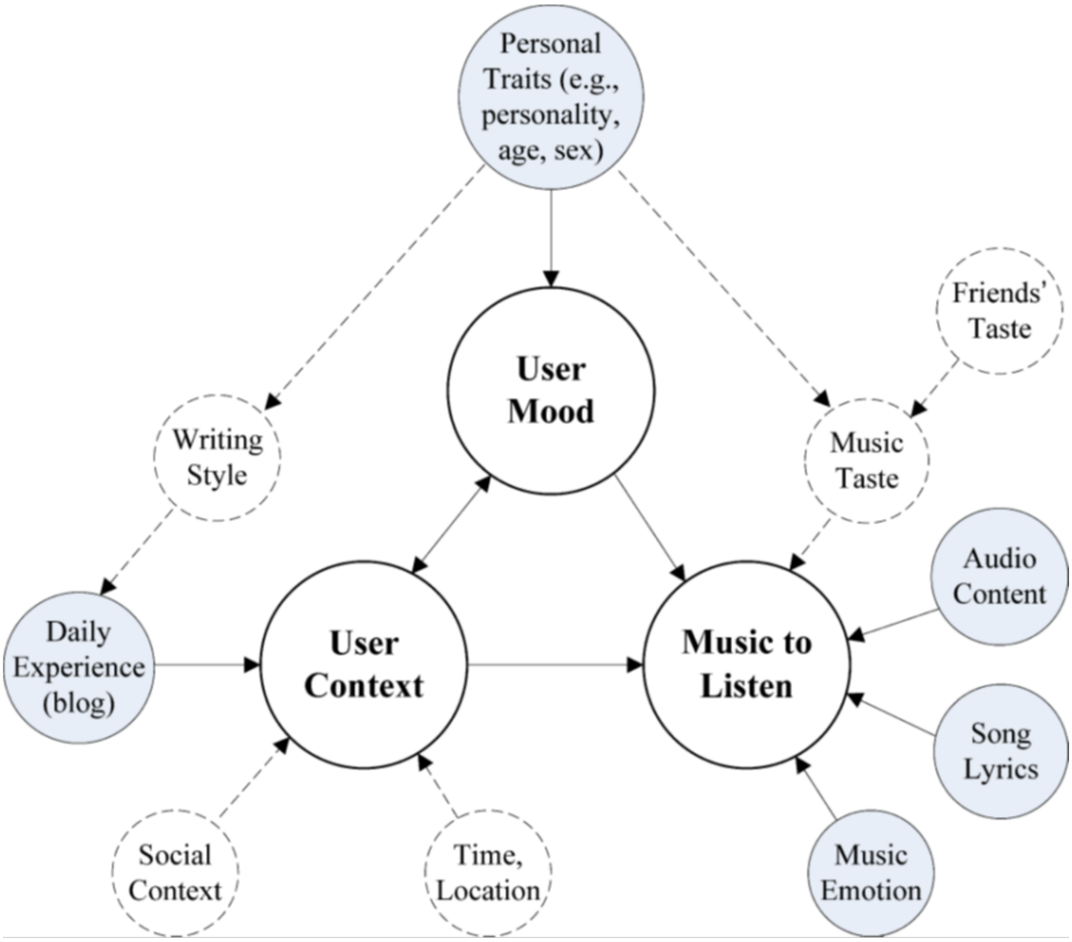


Figure 1.1: Tripartite relationship between music emotion, user mood and music listening behavior from [81]

Figure 1.1 is a graphical model of the musical, personal, and situational factors considered in the study. The shaded nodes represent the observed data in our study, whereas the dashed ones are not considered and left as future work. The three factors central to the graph are the user mood, user context, and the music the user listens to. The user's mood is influenced by the user's context and personal traits. The user context is determined by factors such as the daily experience, social factors (listening alone or with friends), time, and location. The music listening behavior is influenced by the user's mood and user context but is also conditioned on the individual's music taste and the audio, lyrics, and effective content of the music. Conclusions derived from this model provide observation of how, in general, people listen to mood-congruent music when being in a positive mood but tend to listen to mood-incongruent music when being in a negative mood. Also, people with different personalities prefer different music even when being in the same mood. Energy, rhythm, and timbre are content that always improves the accuracy of prediction whether audio-base UMR or MER are applied. To finalize, it is suggested that an individual mood can be inferred from social media-generated text and used to recommend either mood-congruent or incongruent music, depending of course on the user mood and personality traits.

1.4 Culture and demographics

1.4.1 Culture in music cognition

Culture in music cognition is the field encompassing the impact that an individual's culture has upon their music cognition. This includes their preferences, emotion recognition, and musical memory. Musical preferences start with cultural and familiar music traditions present since infancy. In adulthood, emotion's classification of a musical piece will depend on specific cultural features as well as universal structural features [73]. Therefore, it can be reasoned that culture is the main influence of music cognition. The effect of culture and musical experience are important factors that determine preferences. As for emotion recognition, psychophysical cues (tempo, loudness, and timbre), culturally bound cues (musical traditions), cue-redundancy model (exposure to individual's own culture), STEM (stereotype theory of emotion in music), complexity, repetition, and methodological limitations are important factors to have in consideration. Memory plays an important role in culture in music cognition. Long-term and working memory systems are a dependency for the comprehension and appreciation of music.
Enculturation is the name given to the process of learning the dynamics of culture surrounding and individual as well as the acquisition of values and norms appropriate to the related culture [30]. Still considering memory, enculturation has cleared an influence on a person's memory of music which also considers the person's development since childhood and their biases toward music expectations later through the. Although, enculturation has limits that can be encountered, for example, when a person is raised with more than one influential culture.

1.4.2 Demographics of music

Observing the music consumption study "The Overall Music Landscape" (2018) [3], the most common data points captured are gender, age, region, income, and profession. The intersection of data happens between preferred media sources (music, online video streaming, social media content, etc) by age and average time spent on each media source. The different sources referring to music listening are considered: AM/FM radio, on-demand streaming, digital download/files, other internet radio, CDs, satellite radio, AM/FM radio stations streamed online (e.g.

Radio Garden), vinyl. The on-demand modality of music listening can then be divided by different service providers (Youtube, Spotify, Apple Music) and be intersected by the age of groups by listeners. The referred study in this paragraph extends its research to the impact of the listening experience considering the listening devices (smartphones, smart speakers, etc). The study concludes that FM/AM radio leads music discovery, followed by Youtube. The targets of this study are citizens (3000) from the US.

1.5 Music streaming services

1.5.1 Pandora's Music Genome

Used by the music streaming service Pandora [59], the Music Genome Project is an effort to "capture the essence of music". It does so using 450 attributes to represent songs and then organize with an algorithm. There are five subcategories on which songs are divided: pop/rock, hip-hop/electronica, jazz, world music, classic. Each given song gets a representation from a vector that contains 450 genes (an analogy to the determining traits genes possess) where each gene is a data point representing a characteristic of the music such as lead singer gender or level of distortion in the guitar. Then, a matching algorithm is used to find similar songs recurring to vectors of these genes. Apart from that, each song is analyzed by a musician (20 to 30 minutes process per song), and 10% of songs are analyzed by more than one as an effort to ensure statistical reliability. This model is limited to the users of Pandora radio service, exclusive to the US, and the full list of attributes is not available to the public.

1.5.2 Spotify's music genres

Glenn McDonald is a "data alchemist" at Spotify [70]. He developed Every Noise at Once [55] which contains more than 1700 labels for music genres. It started as a debugging tool, later on, these labels found themselves useful to understand the content of a song by complementing the information given by other attributes (happy, sad, danceable, etc). Some of the label's names for music genres are created by the author Glenn McDonald, who visualizes new musical ideas and directions taken by artists working within already-popular styles. This work supports Spotify features Daily Mix and Discover Weekly, while attempting to be more in tune with users' listening habits. Some "genres" are not named yet.

1.5.3 Last.fm's scrobbling

In Last.fm [48], a scrobble is a piece of data that tells people what you listened to and when. It normally consists of the track name, artist and time stamp as well as album name and album artist. In that sense, scrobbling is the process of sending Last.fm our listening data (our scrobbles). It can be achieved through the official app or third-party applications that use Last.fm's API. For example, it is possible to scrobble what we listened to on Spotify. Millions of tracks are scrobbled every day and there are ways available to scrobble vinyl records and music played on the radio. When a track is scrobbled successfully, it appears on the profile page and is stored in the library. Scrobbles and scrobbling allow for Last.fm to generate charts as well as recommending new music based on our listening data. In a similar fashion to Spotify, it will also generate personalized radio stations.

1.5.4 Online radio communities

Online radios are services that stream digital audio through the internet. They present listeners with a continuous stream of audio that cannot be paused (might be possible with the use of middleware) or replayed (possible through later access to shows' archives). Accessible assets to set up an online radio station have made it easy for individuals to carry out their station. Online radios are examples of how online communities are bound together from different physical locations. Although music streaming in online radio can be aided by a recommendation system, it is also expected that the music is curated manually by one or more individuals. NTS [1] is an online radio community divided into two continuous streaming stations, physically located in London and Los Angeles, with hosts from all over the world playing their music every one or two hours (the website offers a chat room where the listeners and host can interact during the streaming). In Lisbon, "underground music" radio stations have bloomed in recent years and one of the most popular radios is Radio Quântica [4], which streams more than a hundred author's shows and hosts live events for the listeners and supporter's of the radio and defends a do-it-yourself culture approach. Smaller online radios can also be found, such as Rádio Paranóia [35], launched in March 2021, gathers a small (30 people) community of listeners and hosts with a shared chat room. In Copenhagen, Absalon Radio is a new radio build upon the already established associative building, Folkehuset Absalon [28], where dancing, yoga classes, and sports matches take place. Christiania's Radio 90.4 [20] is an FM radio station streaming from Freetown Christiania in Copenhagen, it merges the radio activities with the cultural activities of Freetown Christiania.

Chapter 2

Development

2.1 Final research

2.1.1 Context-aware recommendation system

As mentioned in the previous chapter, recommendation systems face three major issues. Cold start, whenever a new user or new item is introduced in the recommendation system, there is insufficient information available about such user or item. If it is a new user, it is difficult to learn about the preferences because of the lack of data available. In case of a new item, having less number of users who have used that item, incorporation of it in the recommendation process takes time [50]. Data sparsity, when feedback given by costumers is insufficient for the recommendation system models to work on. It happens due to a cold start or to the tendency of users not giving enough feedback, which therefore impacts the range and quality of the recommendation [77]. Scalability, as the number of users and items increases, the computation cost also increases which may affect the performance and response time of the recommendation system [39].
To measure the performance of a context-aware recommendation system, evaluation criteria have been specified in the literature and separates into two types, offline and online evaluation. Offline evaluation is performed when the dataset is collected before the design of the system and then the system operates on this data and predicts the preference of the users on the items. This performance measure is evaluating the accuracy of the system in terms of its capacity to predict preferences [67]. Offline evaluation is a faster way to evaluate although it has the disadvantage that the system won't be able to track changes in user preferences in real-time. There are different metrics for offline evaluation, these are root mean squared error (RMSE), mean absolute error (MAE), precision, recall, f-measure, area under the curve (AUC), normalized discounted cumulative gain (NDCG), mean average precision (MAP), hit rate and perplexity. RMSE is the square root of the sum of the

squares of the difference between predicted values and corresponding actual value specified number of observations, it fives the standard deviation of the prediction errors [67]. MAE is the arithmetic average of the absolute difference between predicted and actual values, precision is the number of true positives over the sum of true positive and true negatives, it reveals the proportion of data that the model predicts relevant is relevant [32]. A recall is the number of true positives over the sum of true positives and false negatives, it enables the model to find all the relevant data from the dataset. F-measure is the harmonic mean of precision and recall [67]. AUC is the capability of the model to identify different classes in the dataset. NDCG is a measure related to the ranking of the items a recommendation system returns, where every item in the list has a relevance score associated with it, called gain. The summation of all gains is named the cumulative gain [67] and each of the gain values is divided by the logarithm of the position of the item, a process named discounting [32]. MAP is the mean of the average precision values over the ranks in the relevant recommended items [32]. The hit rate is the number of items in the test set that were also present in the recommended items given by the system for each user (the number of hits over the total number of users is named hit rate). [22]). Perplexity is a metric to evaluate topical models, it measures the quality of topics extracted by the topical model using a training document which allows predicting the occurrence of the words in testing documents [76]. On the other hand, online evaluation is used when the experiment is done in real-time and therefore evaluates the real-time feedback of the users. In comparison to offline evaluation, it takes more time to evaluate online metrics as it has to be monitored over a longer period. Online evaluation helps to understand user interaction behavior with the system which also is an important consideration while assessing the quality of the recommendations [67][32][74]. There are two common online evaluation metrics used in context-aware recommendation systems, click-through rate (CTR) and bounce rate. CTR is the count of recommendations that are clicked by the user, measures real-time feedback of users regarding their preferences. Bounce rate is the percentage of users who have seen the list of recommendations given by the system but instead of exploring those recommendations further chose to exit the recommendation system [74][33].

Some other evaluation criteria are coverage, confidence, trust, novelty, diversity, adaptivity, and user satisfaction. Coverage is either the percentage of total available items a system can recommend or the percentage of total available users for which a system can recommend [32]. Confidence is the system's trust in the prediction of the recommended items, usually, measure in terms of the probability of correctness of the predicted value [32]. Trust defines how much of it the users have on the recommendations give by the system. Novelty is the percentage of items recommended out of the total recommended items that are unknown to the user [32]. Diversity measures how many recommended items belong to a diverse group of items and helps to understand the interest of users in different types of items [32]. Adaptivity measures the system's capacity to adapt to the changes in the item space or changing trends of the user preferences [32]. User satisfaction measures the satisfaction level of the user after going through the recommendation list. It can be measured using implicit or explicit feedback from the users [32]. Also, an important aspect in context-aware recommendation systems is the availability of the datasets having context information, this factor plays an important role especially in offline experimentation. [67][32].

2.1.2 Music listening behavior: context and playback method

According to the study "A survey on music listening and management behaviors" (2012) [42] several contexts for listening to music and playback methods are identified. The identified contexts are active listening, commuting, exercising, work, and housework. And the identified playback methods are song after song, artist, album or genre, playlist or folder, shuffle on collection, online recommendation, and radio. Figure 2.1. shows the relation between preferred methods of playback

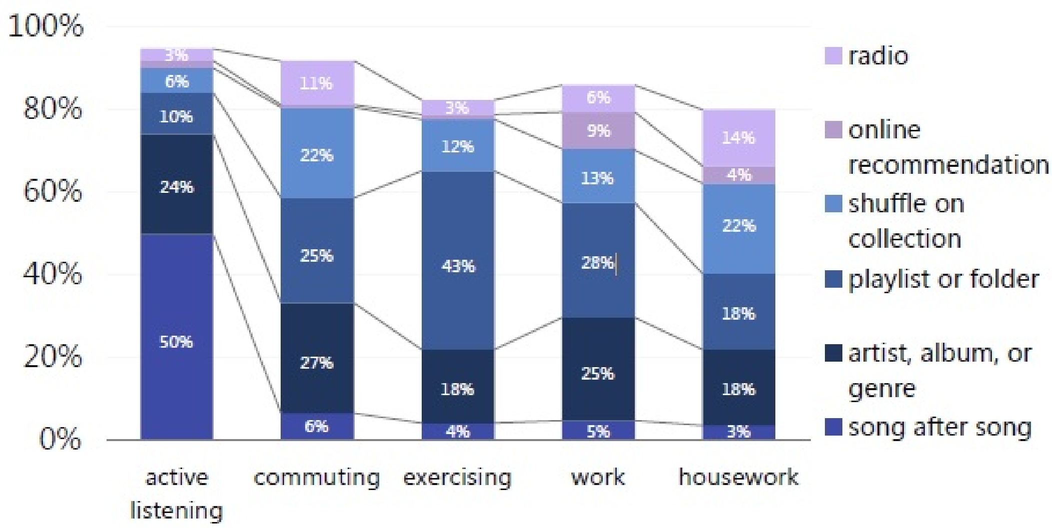


Figure 2.1: Preferred methods of playback for different listening contexts from [42]

for different listening contexts. The questionnaire was conducted by 222 participants. Song after song is clear to be a popular choice when active listening while it becomes the least popular in any other context. Playing music by artist, album or genre, has a stable percentage, from 18% to 27% of all answers, for each listening context, being the highest for commuting. Playing music from a playlist or folder is most popular when exercising, 43% of all answers, and is almost as popular as listening to artist, album or genre for commuting, work, and housework but is not popular, 10% of all answers, for active listening. Shuffle-on collection is most popular for commuting and housework, 22% of all answers both. It is the third most popular playback method for exercising and work, 12% and 13% of all answers respectively, and 6% of all answers for active listening. The online recommendation is the least popular playback method with less than 1% of all answers for active listening, commuting, and exercising contexts, although it increases to 9% for work and 4% for housework contexts. Although not the most popular answer, radio has a listening percentage from 3% to 14% across all contexts, being commuting and housework the ones with a higher percentage and active listening and exercising the ones with the least. The limited popularity of answers for the online recommendation was considered surprising by the authors of the study as for the consideration that participants had most technical education backgrounds and young ages.

2.1.3 Copenhangen population and Lisbon population

Copenhagen has a total population of 638678 [71], from which 474074 (74%) have danish nationality. The other 164604 (26%) have other nationalities. Lisbon has a total population of 508368 [29], from which 247865 (49%) has portuguese nationality. The other 260503 (51%) have other nationalities. There is a total of 197 different nationalities (excluding Denmark) in the population of Copenhagen and 182 different nationalities (excluding Portugal) in the population of Lisbon. There is a total population with portuguese nationality of 1336, 0.2% of the total population of Copenhagen and a total population with danish nationality of 653, 0.12% of the total population of Lisbon. Pakistan, Turkey, Iraq, Germany, Poland, Somalia, Sweden, Morocco, United Kingdom, and Lebanon, in order, represent the highest number of other nationalities among the population of Copenhagen apart from Denmark. And the highest number of other nationalities among the population of Lisbon apart from Portugal are represented, in order, by Brazil, Cape Verde, China, Guinea Bissau, Angola, Italy, Romania, France, Nepal, and Ukraine. Figures A.29 and A.30 in the appendix contain in detail the numbers for all nationalities in both cities. Denmark and Portugal are considered to have different work cultures. Therefore, the present research considers the possibility of difference among the two populations on how the activity context influences on their general music listening behavior when performing work activity and not. A resume of the total population for each city is presented in Figure 2.2 and 2.3.

2.1.4 Music streaming and media players

According to Google Store's music & audio category and Apple Store's music category of apps available on store, a comparison between the most downloaded and used (Google Store only) in Denmark and Portugal (not Copenhagen and Lisbon),

		Total	Men	Women
1	Danish	474074	233121	240953
2	Other nationality	164604	82078	82526
3	Total population	638678	315199	323479

Figure 2.2: Total population of Copenhagen

		Total	Men	Women
1	Portuguese	247865	101940	145925
2	Other nationality	260503	130900	129603
3	Total population	508368	232840	275528

Figure 2.3: Total population of Lisbon

free and paid. SimilarWeb [68] was used to access both app store's data, which was collected from the update on SimilarWeb's website on May 21, 2021. Apple Store's most popular paid applications in the category of music apps are music production tools both in Denmark and Portugal. And free, Spotify [70] is the most downloaded in both countries. Apart from one music production app, both Denmark and Portugal have music media players in their five most downloaded paid apps in Google Store. Music media player apps make part of the ten most-used apps in both countries, music streaming apps (Spotify, etc) do not. Both countries have Spotify as the most downloaded free app from Google Store. In Denmark, the most used apps are music streaming apps and in Portugal, the most used apps are music media players. Spotify is probably the most popular music streaming app in both Denmark and Portugal. An observable difference from Google Store's apps shows that in Portugal the most popular free music & audio apps are music media players, while in Denmark are the music streaming platforms. This indicates that users in Portugal have storage of music files in their mobile devices, which doesn't seem to be so much the case in Denmark (from the most used apps in Google Store point of view).

2.1.5 Ethnography

Ethnography is a research method central to knowing the world from the standpoint of its social relations [79]. The discipline relies on the employment of qualitative and quantitative methods for data acquisition. In the virtual domain, digital ethnography uses social media platforms data to study human interaction and behaviors. On the other end, relational ethnography is applied in specific places and employs observation over targeted instances that relate to a certain topic. Ethnography research considers its findings of context and situation and, therefore, through uncovering relationships and then use the resultant data to test and explain empirical assumptions that allow the integration of ethnography in a quantitative research [23]. In recent years, the streaming platform Spotify has been a subjet of an ethnographic study [78]. The current research relies on the principles of ethnography to deduce human listening behavior over the influence of activity context, and eventual differences between working individuals of two different cities.

2.2 Introduction to experiment design [21]

Variables

A fundamental element we need to consider when designing an experiment is the variables. The variables take at least two values and are useful to understand the answer to a research question, e.g. "Can sound effects help players solve a puzzle game more efficiently?". Variables may relate to one another in a way where variable A has an effect over variable B, this relationship can be confirmed through analysis of acquired data on variables. In order to confirm this, an experiment where A is manipulated and B is measured. Confirmation is concluded when B changes due to nothing else but A getting altered. This is the essence of an experiment: manipulating only one variable, keeping everything else constant, and then measuring the effect. In this scenario, A (manipulated variable) is called an independent variable and B (measured variable) a dependent variable.

Measurement scales

Different scales can be used to measure the dependent variable. The data can be either numerical or categorical and different types of measurement scales require analysis methods in accordance. It depends on the research question which measurement scale shall be used. A nominal scale implies that there is no natural ordering of the categories, as is the case with gender, race, and religion. An ordinal scale refers to categories that can be ordered such as sizes (small, medium, large, etc.), still the difference between small and medium may not be the same as medium and large. An interval scale implies meaningful numbers, a scale where all values are equidistant. And a ratio scale is where zero is a meaningful value such as 0kg or 0 degrees Celsius.

Sensitivity

The sensitivity indicates the way a question is asked such as "Do you like this product?" and be presented with a yes/no answer choice or "To what degree did you like this product?" and be given a scale from "Not at all" to "Very much" and asked to choose where to place across. The first example is categorical and the second is continuous. Therefore, different measurement scales were applied by different measures. In this case, the second example would provide more detailed information on the success of the product.

Confounding variables

Confounding variables may originate from the participant's different prior knowledge, experience, and preferences. The place where the testing takes place or by whom and in what way the instructions are presented. To access the right conclusions on the effect of the independent variable, the experiment needs control for all confounding variables to avoid any systematic differences that can influence the interpretation of the manipulated variables. Therefore, all the possible confounding factors need to be kept either constant or random. Constant means that instructions and testing set up are done the same way to all participants and random means that test conditions are randomly assigned to avoid any systematic effects due to the order of the presentation.

Control groups

Due to the confounding variables, it is required to compare the intervention to a situation without any intervention. This can be achieved through the manipulation of the independent variable. It is important to consider the variability on personal traits of each individual, the confounding variables when foreseen may allow better control of acquired data.

Between-group design

The between-group randomly assigns participants to take part in one of the experimental conditions. This way, any systematic differences between the groups that might affect data are avoided. Sample sizes can sometimes be approximately equal. This allows for the monitoring of how many get assigned to one condition and reduces the probability of this condition coming up for the remaining participants.

Within-group design

Letting participants take part in both experimental and control conditions minimizes the risk of having the individual variability hiding any possible effect. Fewer participants are needed to spot an effect in the case of reduced individual differences in regards to the variability. No systematic effect of order as well as consideration for how participants are assigned to the experiment is required to not contaminate the results. Tiredness, boredom, and acquisition of skills while taking the experiment can affect the accuracy of the acquired data.

Operational definition

Defining a musician as someone who can earn money by playing music is an example of an operational definition. In this example, if this quality (earn money by playing music) is sought after in the experiment, then it is the target of measurement and should be made clear. Therefore, an operational definition of concepts is performed.

Reliability and validity

A good experiment should collect measurements that are reliable and valid. Reliability means that the experiment if reproduced will achieve the same results. Validity means that what is measured is indeed what is meant to be measured.

Measurement error and reliability

When something is measured, the error of measurement is taken into account. The consistency of measurement is achieved through measuring multiple times (repeating the same task on the same test occasion in an unnoticed fashion to the test taker). It can also be achieved through a comparison of ratings from several people. The similarity will indicate a low measurement error.

Internal and external validity

The intended population for the experiment must not be systematically different, as the results do not generalize to the larger population. Data collected from an experiment would likely be invalid if systematic differences occur between the experimental and control groups.

Likert items and scales

Likert items and scales are useful to assure reliability and validity as they combine multiple ratings of statements that contribute toward one construct of what is intended to measure and, therefore, reduce the measuring error.

Populations and samples

A population is the full set of individuals of interest in a research study. Measuring a whole population can be unpractical. A sample is meant to represent the target population studied. For trustworthiness, the sample group should be formed in a fashion where there is an equal of each individual to be selected. This is of most importance if bias and unreliable data are to be avoided.

Central tendency

The acquired sample must be collected, classified, summarized and the data presented for interpretation. Descriptive statistics are used for this task. Depending on how the data looks and which measurement scales are used, the summarized data will differ. The measure of central tendencies that is based on all scores is named the mean value. The point below which half of the scores fall is named the median, it is less sensitive to extreme values. The most commonly occurring score in the data, and therefore a score always present in the data, is named the mode.

Variability

Information about variability or spread in it are often necessary to get a meaningful conclusion from the reported mean values. Giving the range of the data, the smallest and largest values, is one simple form of reporting the variability. Variance and standard deviation are more useful to provide an idea of the variability and where most values can be found. The standard deviation can be thought of as a measure of distance from the mean value.

Frequency distributions

The shape of the distribution can be observed through histograms. The expected shapes to be observed are normal distributions (symmetrical and bell-shaped), skewed distributions (a longer "tail" than the other), bimodal distributions (two main peaks), and uniform distributions (flat). The normal distributions, also named Gaussian distributions, are the most commonly occurring results and relate to different biological processes.

Sample mean and sample size

A large sample is needed for the sample mean to be a good estimate of the true population mean. Therefore, the larger possible sample is preferred.

Standard error of the mean

The estimate of the true population mean varies with each sample. Most sample means will be close to the true population mean when the sample size is large enough. The standard deviation of the distribution, named standard error of the mean, provides a number of the precision of our estimation of the true mean.

Relationship between variables

The plot of data against each other is a method to observe if both the independent and dependent variables have some kind of relationship. It is often of interest to study such relationships.

Covariance

Covariance is a measure of the joint variability of two random variables [62]. If the greater values of one variable mainly correspond with the greater values of the other variable, and the same holds for the lesser values (that is, the variables tend to show similar behavior), the covariance is positive [18].

Correlation coefficient

The Pearson Product-Moment Correlation Coefficient is a standardized measure of the linear relationship between two variables. The degree to which the relationship can be explained by a linear model is expressed by the coefficient. By calculating the correlation coefficient for different ratings of the same stimuli, it is possible to get a measure of how reliable the measure is.

Linear regression

In linear regression, a variable y is predicted by x in terms of a linear relationship between the two. The method finds the straight line that best fits the data points and y.

Inferential statistics

Frequency distributions can be related to probabilities. The probability of an event varies between 0 and 1, this probability is defined as the number of outcomes divided by the total number of possible outcomes. Mutually exclusive outcomes can be added. Inferential statistics provide systems and techniques helpful for good decision-making and accurate predictions based on data. They are used to model patterns of data and make inferences about the population studied. Therefore, without testing on the whole population it is made possible to predict something about it and test such predictions. The testing hypothesis is performed in research recurring to such a method. The hypothesis is a prediction, it may or not be based on previous theory or studies. To develop a good hypothesis it is important to be possible to falsify it. The confirmation or rejection of a hypothesis is based on the probability of an outcome. Likewise, the probability is linked to the frequency distribution of a population. Knowing the mean and standard deviation of our study population makes it possible to score how likely it is to belong to such a population. When the probability of obtaining a result by random is lower than a particular criterion, a result is generally seen as being statically significant from that predicted by the null hypothesis. For testing significance, a "statistical significant level" of 0.05 is used. Its purpose is to observe where a score belongs into a normal distribution. It is considered a significant difference if the score has a 5% chance or less of occurring. The effect size is a measure of how much of the variability in the data the effect (between the experimental and control group difference of score) accounts for. There are two types of variance effects, unsystematic and systematic. Unsystematic variance is caused due to random variability in data due to preferences or measurement errors. Systematic variance is caused by a systematic difference in measurements between groups. If the experiment is well-controlled, the systemic variance should only happen due to the manipulation of the independent variable so that a clear interpretation of the results is possible. In inferential statistics, two types of errors can occur, type I and type II errors. Type I falsely states that there is a statistically significant effect when the result occurred by chance. Type II falsely states that there is no statistically significant effect when there is one. With the 0.05 criteria, 5% of times it is likely that the null hypothesis is rejected. More strict significance levels can be selected and this will likely increase one of the error's type chance of occurrence and decrease the other one. In terms of statistical power, the power of a test is the probability of detecting an effect (in case it exists). 80% detection of the actual effect is considered reasonable [27].

Parametric data

Parametric tests assume the use of (arithmetic) mean and variance, the statistics that describe the normal distribution. Therefore, the measures need to be accurate and meaningful enough to have trust in the results. Before using parametric statistics the data should be on an interval or ratio scale, all groups should have approximate variances, and the distribution of the population should be normal. Likert scale items tend to use scales and it can be considered an interval measure when a subscale is based on several combined Likert items. Data must have approximately equal variance since it is related to the precision of the mean. Box plots summarize the majority of data points and whether there are outliers. Likewise, the Levene test tests the null hypothesis that all input samples are from populations with equal variances. Assuming a normal distribution of data is performed by plotting histograms of data and determine the type of distribution of the population that the sample was drawn from, although the histogram shape can vary a lot. QQ plots display the data points against a plot drawn from a population with normal distribution by ordering the values against the theoretical qualities. If the assumptions for parametric data are violated, parametric tests should not be used. For most kinds of tests, there is a corresponding non-parametric test that can be used as an alternative for data that do not fulfill the criteria.

Student's t-test

The t-test is any statistical hypothesis test in which the test statistic follows a Student's t-distribution under the null hypothesis. It can determine if the means of two sets of data are significantly different from each other. A test of whether the mean of a population has a value specified in a null hypothesis is named a one-sample location test. A test of the null hypothesis such that the means of two populations are equal is named a two-sample location test. The name student's t-test applies when it is assumed the variance of the two populations to be equal.

2.3 Research hypotheses

Current times have presented to the world population a large range of options regarding music listening. An individual may store their music files either on a laptop or a portable device and listen to them as well as having music played directly from an online source. Both music media players and online music streaming platforms offer various methods of playback, one can listen to an album or have their whole collection on shuffle. Music streaming services such as Spotify offer recommendations that present people to a continuous and automatic stream of music, similar to the more antique FM radio stations. In that matter, radio stations are more commonly found online at present and can be perceived as an alternative form to the automatic stream of music. Whichever preferences an individual has in terms of the platforms and devices he uses to listen to music, research on music listening and management behaviors [42] present activity context (e.g. workout, communing) to influence the playback method of the music (e.g. shuffle on collection, radio). The aim of the current research is an attempt to verify the variance in music playback preferences, for passive listening, when at work and not. The research considers the population of two cities and their different work cultures, Lisbon and Copenhagen, therefore questioning the possibility of significant differences of passive music listening behavior (playback method of music) when performing work tasks and not. Additionally, due to the higher download of music streaming apps in Copenhagen and of media players in Lisbon, the research questions if this characteristic can be confirmed among the experiment participants.

Moving towards the improvement of recommendation systems in music streaming platforms, the current research expects the acquired data to contain relevant information for practical application in context-aware recommendation systems.

2.4 Experiment design

An online survey was designed and includes a demographic and a music listening behavior section of questions. The demographic section contains questions regarding the city of residence (Copenhagen or Lisbon), years residing in that city, age, nationality, and gender. The section of music listening behavior, questions how the participants passively listen to music when working and when not (two questions), considering the following playback preferences: song after song, artist, album or genre, playlist or folder, shuffle on collection, online recommendation, radio. The music preferences selected are the same had in account by Kamalzadeh et al. [42] in their research on music listening and management behaviors, and rely on a 7 point scale presented linearly. The questionnaire got published in several groups on social media composed of expats of both cities, and academic groups of both cities. To complete the questionnaire, the participants had to have a job (paid position of regular employment) where they listen to music of their choice and reside in Copenhagen or Lisbon. The questionnaire makes these same questions to exclude anyone who attempted to participate without conforming to all the requirements. The survey population completing the survey comprises 58 participants, 28 residents of Lisbon (48%) and 30 of Copenhagen (52%). 31 are male (53%), 26 are female (45%), and 1 does not identify with either of the two genders (2%). 36 participants are portuguese (62%), 4 are danish (7%), and 18 have other nationalities (31%). Of all the participants, 6 have double nationality (10%). The next chapter follows an analysis of how working against not working (independent variable) affects the playback methods the participants choose (dependent variable). Additionally, a comparison of the average results between participants from Copenhagen and Lisbon attempts to answer the question of how the activity of working is related to passive listening behavior and if significant differences can be observed among the population of the cities. The imposition of histograms is used to observe the data.

Chapter 3

Results

3.1 Population sample

The population sample composed of 30 individuals residing in Copenhagen and 28 residing in Lisbon, is composed of 18 individuals between 16 and 25 years old, 31 between 26 and 35 years old, 7 between 36 and 45 years old, 1 between 46 and 55 years old, and 1 between 56-65 years old. In terms of years residing in one of

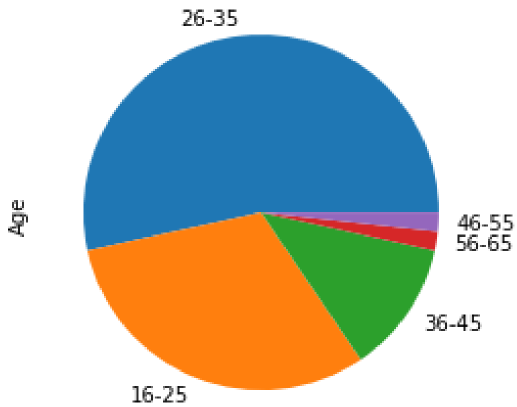


Figure 3.1: Age of the population sample.

the two cities, 19 individuals are residing in Copenhagen for less than 5 years, 5 residing in between 5 and 10 years, 1 between 21 and 25 years, 2 between 25 and 30 years, and 1 for more than 30 years. In Lisbon, 5 individuals are residing for less than 5 years, 4 residing in between 5 and 10 years, 4 residing in between 11 and 15 years, 2 between 16 and 20 years, 3 between 21 and 25 years, 3 between 25 and 30 years, and 8 above 30 years.

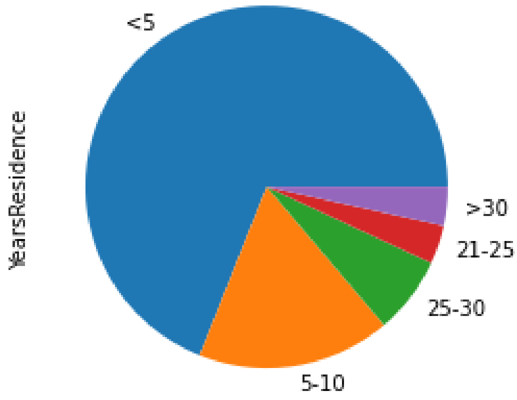


Figure 3.2: Population sample number of years residing in Copenhagen.

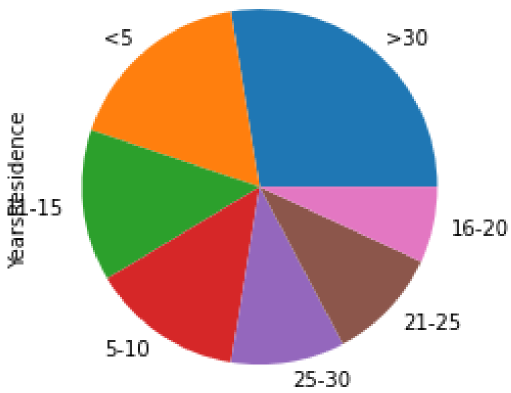


Figure 3.3: Population sample number of years residing in Lisbon.

3.2 Music streaming apps and media players

On a quick note, the results provide, for both cities, clear use of music streaming apps much higher than music media player apps. Wherefrom all participants, 28 in Copenhagen and 23 in Lisbon use more music streaming apps, 1 in Copenhagen and 4 in Lisbon use more music media players, and equally 1 for both cities use both means equally.

3.3 Playback method and activity context

At section A.2 in the appendix, a group of histograms present the differences and similarities in playback method for passive listening depending on activity context among all participants, Copenhagen and Lisbon residents, when working and not, and when working and not working among the two groups of participants.

3.3.1 General population comparison: working and not working

A general comparison follows, including all participants, without focus on their city of residence (Copenhagen and Lisbon), between each playback method individually when working and not.

Song after song passive listening when working plotted against not working shows an overlap that doesn't present a significant difference from being on or off job work activity. In both contexts, there's a tendency for listening more to a song after song than not. A concern to the term song after song arises. The intention is to mean selection, by the user, of each track getting played (like placing songs on queue). It appears unlikely to be the case when working due to the focus required and sense of responsibility. Address to figure A.1 in the appendix for histogram data.

The most significant difference observed for passive listening to a single artist when working and not is that it is more the case when not working, although it still happens while working and with less frequency. Address to figure A.2 in the appendix for histogram data.

No significant difference in listening to album or genre is detected. It seems likely preferred for both working and not working activity. Address to figure A.3 in the appendix for histogram data.

Listening playback method playlist or folder, plotted for working against not working, shows a higher frequency to always listening when working. Additionally, the plot shows a high frequency for always listening (7 in the Likert scale) in both contexts combined. Address to figure A.4 in the appendix for histogram data.

Shuffle on collection shows a very identical plot for working against not working, with peaks on both extremes of the Likert scale. Address to figure A.5 in the appendix for histogram data.

Online recommendation shows a low incidence in the higher end of the Likert scale, with an overall spread all along with its range. Address to figure A.6 in the appendix for histogram data.

Radio has the highest selection on one of the Likert scale values in both working and not working contexts, with never (1) selected for radio in both contexts and few choices of the items from the middle (4) to the higher end (7) of the scale. The researcher assumes radio is a less popular method of music listening, whatever the activity context is (unless driving [24]). Address to figure A.7 in the appendix for histogram data.

3.3.2 Working comparison: Copenhagen and Lisbon

A comparison between the two populations (residents of Copenhagen and Lisbon) follows, between each playback method individually when working.

In the Likert scale, song after song when working shows a higher selection for residents in Copenhagen than in Lisbon. Still, the general tendency falls from the middle to the higher end of the scale. Again, a concern to the term song after song being misunderstood by the participants is considered. Address to figure A.8 in the appendix for histogram data.

Passive listening to an artist when working shows a higher preference in the residents of Lisbon than those of Copenhagen, with both having a similar selection on the middle range of the Likert scale. Address to figure A.9 in the appendix for histogram data.

Likewise, in terms of passive listening to album or genre when working, the plot shows Lisbon residents to have a higher preference for this playback method rather than Copenhagen residents. Address to figure A.10 in the appendix for histogram data.

The playlist or folder passive listening when working plot shows a clear higher (end of the Likert scale) preference of the residents of Copenhagen in comparison to those of Lisbon that find themselves more present at the lower end of the Likert scale. Address to figure A.11 in the appendix for histogram data.

The plot for shuffle on collection, when working, shows a higher preference for this playback method by the residents of Lisbon. The opposite is the case for the residents of Copenhagen. Address to figure A.12 in the appendix for histogram data.

The plot for online recommendation listening when working shows residents of

Copenhagen more at the lower and higher end of the Likert scale. In comparison to those of Lisbon, found in the middle of it. In general, almost no differences are visible between the two populations. Address to figure A.13 in the appendix for histogram data.

As for radio passive listening when working, a slightly higher preference in the plot is shown for the population of Lisbon. Although, in general, the tendency of the result is for the lower end of the Likert scale. Address to figure A.14 in the appendix for histogram data.

3.3.3 Not working comparison: Copenhagen and Lisbon

A comparison between the two populations (residents of Copenhagen and Lisbon) follows, between each playback method individually, when not working.

When not working, the song after song plot shows a similar distribution in the Likert scale for the residents of both cities, with a higher tendency towards the higher end of the scale. Again, a concern to the term song after song getting misunderstood by the participants is under consideration. Address to figure A.15 in the appendix for histogram data.

The plot for passive listening to an artist when not working shows a similar distribution in the Likert scale among the residents of both cities. The same plot shows a higher tendency from the residents of Lisbon to the higher end of the scale. Address to figure A.16 in the appendix for histogram data.

Passive listening to an album and genre, when not working, has the respective plot showing a tendency of the participants of Lisbon to the central values of the Linker scale. The participants of Copenhagen tend to the edges of it. In general, a higher tendency to the higher values of the scale shows. Address to figure A.17 in the appendix for histogram data.

The plot shows that while the residents of Lisbon, when not working, do passive listen of a playlist or folder with a tendency on the lower values of the Likert, the ones of Copenhagen present the opposite behavior (7 on the scale gets a significant amount of selection per participants residing in Copenhagen). Address to figure A.18 in the appendix for histogram data.

The plot for residents of Copenhagen against Lisbon, on their preference for passive listening of music to shuffle on collection when not working, shows a higher tendency for residents of Lisbon on the lower end of the Likert scale. The opposite can be observed for the residents of Copenhagen. Address to figure A.19 in the appendix for histogram data.

When not working, online recommendation has the residents of Lisbon hitting higher on the lower end of the Likert scale. The Copenhagen population places its answers on the edges of the scale. Address to figure A.20 in the appendix for histogram data.

For residents of both cities, radio, when not working, has most selections on the Likert scale to the lower end of it, with most selection on the last value (1) of the scale. Address to figure A.21 in the appendix for histogram data.

3.3.4 Copenhagen and Lisbon comparison: working and not working

A final comparison follows, between the activity of when working and not working, and between the population of each city, for each playback method individually.

For passive listening of a song after song in Copenhagen, working has a higher incidence over the edges of the Likert scale, and not working has a higher incidence at the center. In Lisbon, working and not working tend to have higher values on the Likert scale. Like the previous cases, a concern to the term song after song getting misunderstood by the participants is under consideration. Address to figure A.22 in the appendix for histogram data.

For passive listening of an artist in Copenhagen, working and not working have a higher incidence on the higher values of the Likert scale. Not working reaches the higher edge of the range (6 and 7) and working places below that (4 and 5). Lisbon, similarly, tends to the higher values, both working and not working, with less incidence (both) on the smaller values in comparison to Copenhagen. Address to figure A.23 in the appendix for histogram data.

For passive listening of an album or genre in Copenhagen, working and not working both tend to the higher values of the Likert scale. Working shows a solid selection at the middle value (4) of the scale. Lisbon tends to the middle of the scale, both working and not working. Address to figure A.24 in the appendix for histogram data.

In Copenhagen, for passive listening of a playlist or folder, there's a high incidence on the highest value of the Likert scale (7) for both working and not working, and in general more the higher values than the smaller ones of the scale. Lisbon shows a similar frequency throughout the Likert scale for both working and not working. Address to figure A.25 in the appendix for histogram data.

In Copenhagen, for passive listening of shuffle on collection, there is a common higher incidence on the higher values of the Likert scale for working and not working. Lisbon shows a distribution of frequency on both the edge values of the Likert scale. Address to figure A.26 in the appendix for histogram data.

In Copenhagen, for passive listening of online recommendation, the plot shows a similar frequency through the whole range of the Likert scale for working. Not working, the frequency tends to both edges of the scale. In Lisbon, working presents a higher incidence on the smallest values on the edge of the Likert scale, while working is placed at the central ones. Address to figure A.27 in the appendix for histogram data.

In Copenhagen, for passive listening of radio, working and not working tend to the smaller values of the Likert scale with incidence on the lowest value (1). Lisbon, comparatively, has its results more spread through the whole Likert scale, similarity, between working and not working, although sharing the incidence towards the lowest values of the scale. Address to figure A.28 in the appendix for histogram data.

3.4 Final observations

Repeated measures mixed ANOVA [45] with "Working" and "Not working" as the within-subject factor and "City" (Copenhagen and Lisbon) as the between-subject factor. The tests of within-subject effect present no significant (above 0.05) F values. The "Sig. value" shows a significant value, 0.007, for listening to an artist as a playback method. Finally, according to the partial eta squared results, for all playback methods, the test of within-subjects showed significant results (below 0.05). The tests of between-subjects effects show no significant F value, all significant "Sig. values" (all 0), and no significant partial eta squared values (all above 0.05). For difference withing-subject of working and not working, listening to a single artist has a significant Sig. value, and therefore the only significant variance when working and not working among the playback methods in focus. More deductions may be inferred through the histograms addressing the comparisons in the previous section (consult. A.2). Finally, the plots of estimated marginal means for each playback method studied follow .

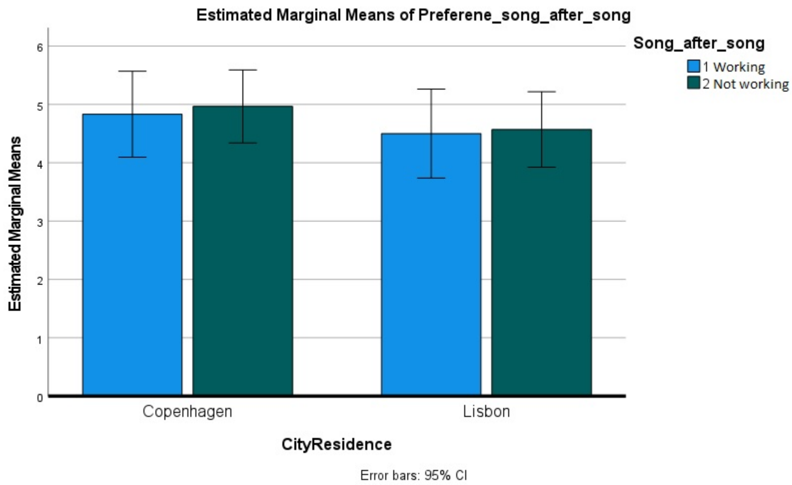


Figure 3.4: Repeated measures mixed ANOVA with "Working" and "Not working" as the within-subject factor and "City" as the between-subject factor for passive listening to song after song.

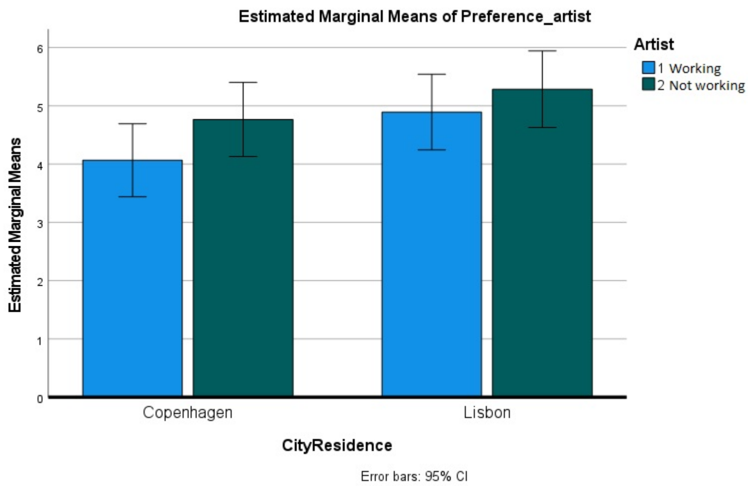


Figure 3.5: Repeated measures mixed ANOVA with "Working" and "Not working" as the within-subject factor and "City" as the between-subject factor for passive listening to an artist.

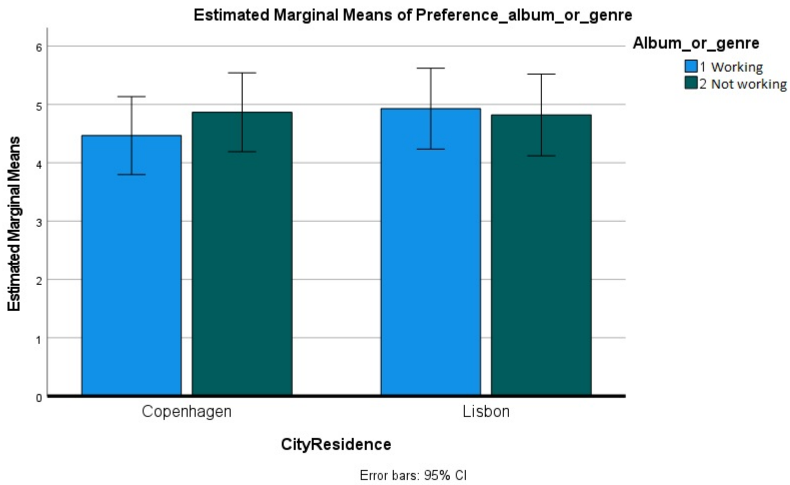


Figure 3.6: Repeated measures mixed ANOVA with "Working" and "Not working" as the within-subject factor and "City" as the between-subject factor for passive listening to an album or genre.

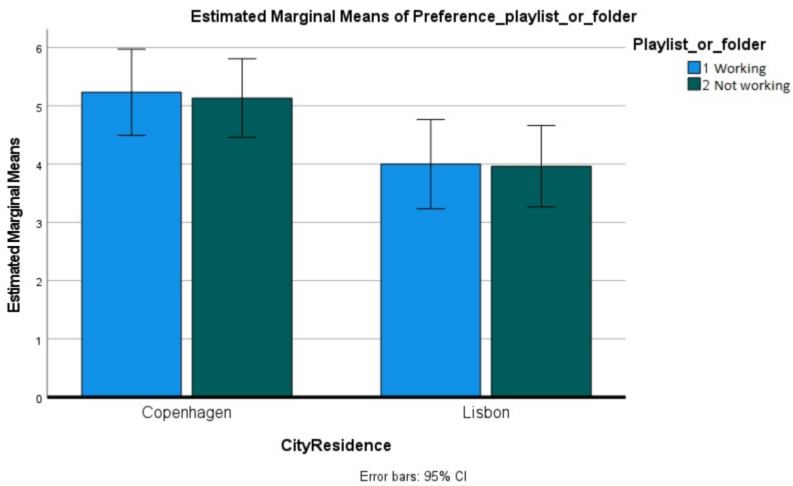


Figure 3.7: Repeated measures mixed ANOVA with "Working" and "Not working" as the within-subject factor and "City" as the between-subject factor for passive listening to a playlist of folder.

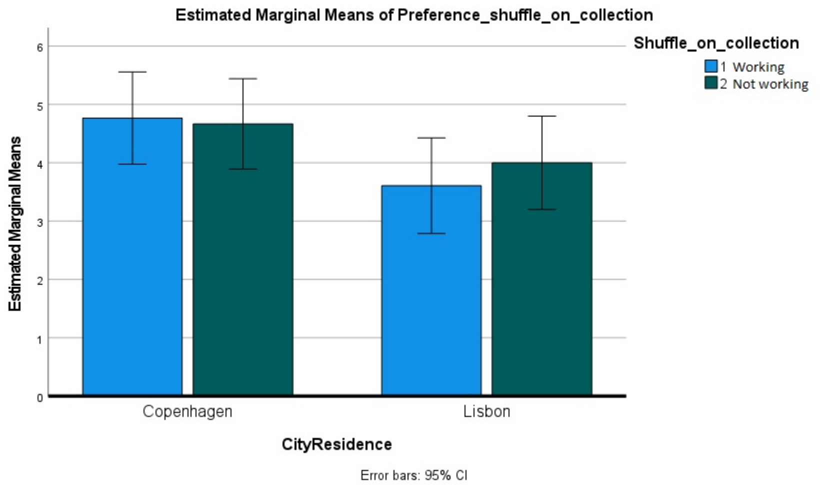


Figure 3.8: Repeated measures mixed ANOVA with "Working" and "Not working" as the within-subject factor and "City" as the between-subject factor for passive listening to shuffle on collection.

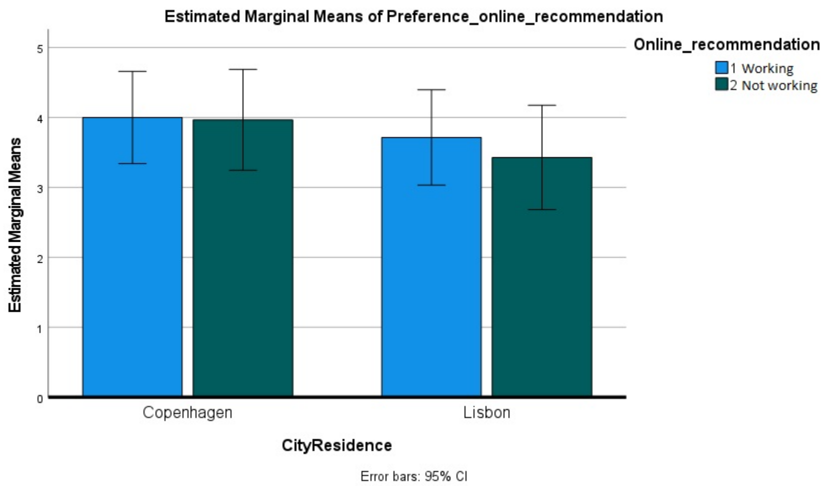


Figure 3.9: Repeated measures mixed ANOVA with "Working" and "Not working" as the within-subject factor and "City" as the between-subject factor for passive listening to online recommendation.

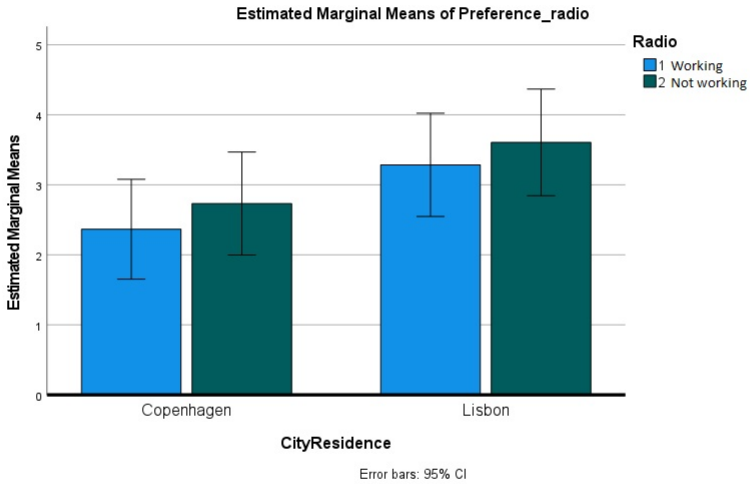


Figure 3.10: Repeated measures mixed ANOVA with "Working" and "Not working" as the within-subject factor and "City" as the between-subject factor for passive listening to radio.

Chapter 4

Research contribution

Murciego et al. in their article "Context-Aware Recommender Systems in the Music Domain: A Systematic Literature Review" (2021) [53] present a review of research and development of context-aware recommendation systems in the music domain. The results of the review are divided per contextual factors (more commonly employed), devices and technologies (for extracting contextual information), recommendation algorithms (context information exploited), the recommendation process and employment of algorithms), and evaluation metrics (to validate the effectiveness of the recommendation system).

4.1 Contextual factors

The context in the music domain divides into two groups, environment-related and user-related [61]. Environment-related refers to the context of the physical environment and devices used by the user. User-related refers to the current activity, knowledge background, and social context of the user.

4.1.1 Environment-related

In terms of physical context, environment-related context covers location, weather, time, and others such as traffic activity or ambient light. In terms of interactive media, the use of mobile devices, desktop devices, vehicles, wearables, virtual assistants, and session context are considered.

4.1.2 User-related

In terms of social context, through the access to social media data of the user, user-related context covers people met by the user at a given time, people the user connects or interacts with, and people's role and relationships of trust in the user's environment. In terms of modal context (user's state of mind, goals, mood, experience, and cognitive capabilities), the user's emotional state, skills, and culture are considered. Finally, user activity refers to tasks performing specific physical activities, and the physiological state (measured through biometric variables).

4.1.3 Further considerations

The knowledge the recommendation system has of the context can be classified as fully or partially observable, and unobservable. Respectively, means that the acquisition of contextual data is either complete, incomplete or nonexistent. Considering changes in the user's context over time, the context can be considered as static (steady music listening behavior over whole context activity) or dynamic (accentuated changes on music listening behavior over whole context activity). Finally, in terms of the recommendation system's acquisition of the user's context information, the acquisition of the data can be considered explicit, explicit, or inferred. Inferred means that the acquired information is a prediction based on the interaction of the user with the system.

4.2 Devices and technologies

The sensor can be classified into four categories: physical or hardware sensors, virtual or software sensors, social sensors, and human sensors [38].

4.2.1 Physical sensors

Physical sensors are the ones that provide a raw measurement of the environment. Such sensors are the GPS, accelerometer and gyroscope, WiFi and Bluetooth, camera, microphone, and biosensors. GPS provides the geographical location of the user. Accelerometer and gyroscope can detect physical activity through the capture of movement. WiFi and Bluetooth give the user's presence and position to devices nearby. The camera can give the emotional state of the user through emotion recognition through facial analysis as well as the geographic location in some cases. The microphone provides information on the user environment and captures spoken conversation. Last, biosensors refer to EEG and heart rate monitors to capture the state of the user.

4.2.2 Virtual sensors

Virtual sensors refer to the combination of measurements from different sensors. Its use can provide information on external devices and their APIs, such as geolocation acquisition through an IP. More precise measures of a contextual factor are possible through information fusion techniques.

4.2.3 Social sensors

Social sensors extract data from the social media content of the user. Facial expression recognition can provide information on the user's emotions by accessing their photos. Sentiment analysis of text published recently on the user's social media is considered.

4.2.4 Human sensors

Human sensors regard explicit information that the user provides such as a textual description of a playlist or a mood description of the activity being performed.

4.3 Recommendation algorithms

Context-aware recommendation systems have information added to them through different paradigms. 2D methods, contextual prefiltering, contextual post filtering, and contextual modeling are the recommendation processes in use [75].

4.3.1 2D methods

2D methods work in a two-dimensional space, most commonly through the following function: $User \times Item \rightarrow Rating$. With contextual attributes taken into account, the following paradigms (contextual prefiltering, postfiltering, and modeling) apply.

4.3.2 Contextual prefiltering

Contextual prefiltering uses contextual attributes to filter the data before other recommendation systems algorithms. Therefore, any recommendation algorithm can be applied independently of contextual prefiltering, although, consideration of the amount of data available after prefiltering must be taken so that there is enough information available to generate relevant recommendations. To counter such a situation, generalization techniques can be applied to acquire contexts less specific and have contextual group attributes in hierarchies or use dimensionality reduction models. This approach includes item splitting, user splitting, and user-item splitting, which, if the ratings are very different, splitting the profile of an item or user will create new entities of items or users that are linked to a given context.

4.3.3 Contextual postfiltering

Contextual postfiltering applies contextual attributes filtering to the recommendations obtained after the usual methods of recommendation are applied. The recommendations acquired within this paradigm are contextualized according to filtering or selection (discards recommendations irrelevant to the context), and ranking adjustment (ranking is altered due to the context). Additionally, these techniques can be classified into heuristic (pursuing the common characteristics of an item for a specific user in a given context), and model-based (predictive model considering a given context).

4.3.4 Contextual modeling

Contextual modeling adds contextual information into the recommendation algorithm (not before or after), creating the multidimensional space: $User \times Item \times Context \rightarrow Rating$. The contextual information can be multidimensional itself and therefore incorporating such information into the recommendation model, using the contextual dimensions as predictors of the rating from the user to the item. Model-based contextual modeling is an approach where the contextual dimension is directly added to the recommendation space and makes possible the employment of machine learning techniques such as classification and regression, and support vector machines. Additionally, collaborative filtering based on matrix factorization is possible with tensor factorization, factorization machine, and context-aware matrix factorization approaches. Heuristics in contextual modeling employ an extension of k-nearest neighbor techniques. Contextual modeling can be the combination of any of the other paradigms with this one.

4.4 Evaluation metrics

In addition to the first section on the introduction chapter of this research, some further considerations should be had when performing either user studies, online or offline evaluations. This concerns coverage, novelty, serendipity, diversity, sequence-aware evaluation measures. Coverage concerns the proportion of items on which the system generates recommendations, and therefore, the importance of coverage deals with situations where a music catalog won't get recommended because of a cold start or popularity bias. Novelty concerns the items that get recommended to the user and that he wasn't aware of before (such as recommendation of the new artist). Serendipity is the attempt to measure recommendations considered relevant and surprising, although the literature hasn't provided a consensus on how to measure this element. Sequence-aware evaluation measures consider the transition between songs during a listening session, this concerns mostly music genre and aims to predict a satisfactory suggestion by analyzing the order the user is listening to songs.

4.5 Datasets

Among the most used datasets in music recommendation systems, there are the MillionMusicalTweets (location and timestamp), LastFM1K (timestamp), and Spotify Playlists Dataset (playlist name). In terms of contextual information, the InCarMusic dataset contains multiple contextual factors, which are modal, weather, driving style, road type, traffic conditions, social, timestamp, and tags. The InCarMusic dataset is one of the most used in context-aware recommendation systems, as well as datasets from Last.fm (social tags and timestamps) which has become a reference.

4.6 Current challenges

Apart from the cold start problem, popularity bias, and dimensionality, a couple of other concerns challenge the state of the art of context-aware recommendation systems (CARS). In regards to the evaluation of CARS, an increase of datasets with more contextual information is likely to lead to improvements in the subject. The recommendation of less popular items, as popular items are generally more recommended, can boost the enhancement of novelty and diversity in CARS. Increase of emotion-aware methods (using NLP on the text and artificial vision on photos from recent social information), consideration of the situation in intention (inferring the intention the user is listening to music for and the situation the user's in), and voice-driven interaction (voice assistant that engages in dialogue with the user) are suggested open lines of research by Murciego et al. (2021).

4.7 Research project

4.7.1 Research considerations

From the topics covered above, the present research considers some of them for the future development of a context-aware recommendation system that considers the playback methods as the target of suggestion according to the activity context. In terms of contextual factors, the research considers location/time (environment-related), activity (user-related), and explicit data. This means, location and time or activity provided explicitly by the user giving their activity context. In terms of devices and technologies, these are not considered as explicit data (text or photos) and are to be provided by the user explicitly without recurring to implicit sensor data (due to ethical motives [56] that are not covered in this research). For the recommendation algorithms, contextual prefiltering is of interest as it shall select the playback method, upon explicit information of the user's activity at a given moment, before any other type of recommendation operation takes place. User studies are of most interest as an evaluation metric, which could be useful for the creation of new datasets that have further consideration in contextual elements. Furthermore, a sequence-aware recommendation for the automatic change of the playback methods is considered. The challenge of situation and intention-awareness is an important consideration for future work and is extended in the next chapter.

4.7.2 State of the art

Keeping close to the aspects considered in the previous section, state of art approaches to context-aware recommendation systems explore the importance of the emotion a specific place can arouse and has been studied to match music to a specific place [43], giving special attention to the specifics of the user's venue and the atmosphere that surrounds [14][15]. Time-aware music recommendation development has paid attention to overlooked aspects such as users' listening habits over time, far fewer ratings than listening records in music providing systems [49], and has improved the accuracy of music recommendation. In other research [72], again, the use of user listening habits data has been shown to improve the recommendation accuracy. A relevant tool to access this temporal information from music stream apps is ListenBrainz [51]. More research considers that tracks are short, listened to multiple times, typically consumed in sessions with other tracks, and relevance is highly dependent on context [31]. Approximate nearest-neighbor search algorithms are popular for tackling such challenges, such as t-SNE [54]. As most of the research focuses on the use of implicit data, extracted from social media, for example, much more than explicit (25 works on explicit and 81 in implicit covered in the CARS review research [53]). Zhou et al. (2020) [82] propose a conversational model, relying on explicit data from the user, in a "user ask, system respond" dialog fashion. Additionally, the user is permitted to express their music requirements via text. From the acquired data to the recommendations, the bandit-based algorithm [44] is applied. Beside the range of research considerations, in terms of implicit data, promising results have been attained with the use of click and cursor movement to access the emotional state of the user and use this information for music recommendations [41]. Using prefiltering has presented good results when applying semantic similarity of tags [17], particle swarm optimization, and unsupervised learning [25]. Finally, a recent user-based study (2020) has been conducted to look into the influence of context in music preference [66]. The result proposed a new framework for music context-aware recommendation system divided into three steps, combination of ratings, clustering-based predictive model generation, and generation.

4.7.3 Present research novelty

Extending the research of Kamalzadeh et al. (2011) [53], the questionnaire considered preferred methods of music playback for different activities (passive listening) and attached a cultural aspect by comparing the results between residents of two different cities instead of the ones of a single city, as conducted in the mentioned research. Additionally, although there's a higher download number of music media player apps in Portugal than Denmark [68], Spotify is the most downloaded app in both countries. While there's a higher number of participants residing in Lisbon who use media player apps more (and equally use to online stream) in comparison to Denmark, music streaming app is without contesting wider used. Should be considered that the data from SimilarWeb on most downloaded music apps refers to a whole country population, where the present research focuses on the residents of two cities, as this data represents additional populations to the ones in focus in this research. As for the user study performed in this research, there seems to be no significance in considering the autonomous playback method as a core attribute of context in a context-aware recommendation system. This research is therefore relevant as it provides novel information on the playback methods aspect and invites researchers to add additional contextual consideration for a new user study that may provide significant insight and apply it to the development of more accurate music recommendations.

Chapter 5

Conclusion and future work

Although the present research can statistically prove a significant variance between music listening preferences among the residents of two different cities, it is concluded that no significant information on how an autonomous recommendation system, that considers activity to select a music playback method, can bring forth the improvement of context-aware music recommendation systems. This is due to the lack of statically proven significance among the preference for playback methods among the two different activity contexts studied. Nevertheless, the research is driven by a recent increase of interest in context-aware recommendation systems in academic literature and attempts to motivate the readers and researchers interested in the topic to pursue similar lines of research.

From a music listening behavior discipline angle, the present research presented a novel approach by considering two populations and their work culture to analyze differences in terms of music playback method of preference, a parameter of music listening that hasn't been given yet attention in previous literature concerning context-aware recommendation system for music. Significance has been found between the passive music listening between the residents of Copenhagen and Lisbon. Except for one case (artist playback method preference), no significance is found in the different results for working and not working. Therefore, future research should get new insights on both populations by considering further social-economic factors concerning music genre preferences [65] as this paper shows how people living in Portugal and Denmark, with concern to the cultural and socio-economic background of the listener, place in distinct country archetypes. Due to the qualitative complexity of emotion concepts and terms [19], an important parameter overlooked in the present research is affective states. In future work, affective states [81] shall be applied with the use of NLP techniques for sentiment analysis (using VADER [16]) of explicit text data provided by the user in a fashion that goes along with the works of Polignano et al. (2021) [60], where an emotion-aware computational model based on affective user profiles is proposed, an affective coherence score between an item and the user profile is defined. In this way, preferences depend on user emotional state varies from user to user. Likely important, as considered in the CARS review research (2021) [53], situation and intention-aware, are to be considered too, for accurately inferring the activity that the user is performing more precisely, identifying the intention with which the user listens to music and in which situations.

Bibliography

[1] Femi Adeyemi. *NTS*. https://www.nts.live/. 2011.

[2] Gediminas Adomavicius et al. "Context-Aware Recommender Systems". In: *AI Magazine* 32 (Sept. 2011), pp. 67–80.

[3] AudienceNet. *The overall music landscape*. https://musicbiz.org/wp-content/uploads/2018/09/AM_US_2018_V5.pdf. 2018.

[4] Diverse group of authors and crews. *Radio Quântica*. https://www.radioquantica.com/. 2015.

[5] Mary Bazire and Patrick Brézillon. "Understanding Context Before Using It". In: vol. 3554. July 2005, pp. 29–40. ISBN: 978-3-540-26924-3. DOI: 10.1007/11508373_3.

[6] Joeran Beel, Stefan Langer, and Marcel Genzmehr. "Sponsored vs. Organic (Research Paper) Recommendations and the Impact of Labeling". In: Sept. 2013, pp. 395–399. ISBN: 978-3-642-40500-6. DOI: 10.1007/978-3-642-40501-3_44.

[7] Joeran Beel et al. "Persistence in Recommender Systems: Giving the Same Recommendations to the Same Users Multiple Times". In: vol. 8092. Sept. 2013. ISBN: 978-3-642-40500-6. DOI: 10.1007/978-3-642-40501-3_43.

[8] Joeran Beel et al. "The Impact of Demographics (Age and Gender) and Other User-Characteristics on Evaluating Recommender Systems". In: vol. 8092. Sept. 2013, pp. 396–400. ISBN: 978-3-642-40500-6. DOI: 10.1007/978-3-642-40501-3_45.

[9] Joeran Beel et al. "Towards reproducibility in recommender-systems research". In: *User Modeling and User-Adapted Interaction* 26 (Mar. 2016), pp. 69–101. DOI: 10.1007/s11257-016-9174-x.

[10] Steven Bird, Ewan Klein, and Edward Loper. *Natural Language Processing with Python*. 1st. O'Reilly Media, Inc., 2009. ISBN: 0596516495.

[11] Barry Brown, Erik Geelhoed, and Abigail Sellen. "The Use of Conventional and New Music Media: Implications for Future Technologies". In: (May 2001).

[12] Robin Burke and Robin. "Hybrid Web Recommender Systems". In: vol. 4321. Jan. 2007. ISBN: 978-3-540-72078-2. DOI: 10.1007/978-3-540-72079-9_12.

[13] Mario Cataldi et al. "Good Location, Terrible Food: Detecting Feature Sentiment in User-Generated Reviews". In: *Social Network Analysis and Mining* 3 (Dec. 2013). DOI: 10.1007/s13278-013-0119-7.

[14] Zhiyong Cheng and Jialie Shen. "On Effective Location-Aware Music Recommendation". In: *ACM Trans. Inf. Syst.* 34.2 (2016). ISSN: 1046-8188. DOI: 10.1145/2846092. URL: https://doi.org/10.1145/2846092.

[15] Zhiyong Cheng and Jialie Shen. "VenueMusic: A Venue-Aware Music Recommender System". In: *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*. New York, NY, USA: Association for Computing Machinery, 2015, 1029–1030. ISBN: 9781450336215. DOI: 10.1145/2766462.2767869. URL: https://doi.org/10.1145/2766462.2767869.

[16] cjhutto. *vaderSentiment*. https://github.com/cjhutto/vaderSentiment. May 2013.

[17] Víctor Codina, Francesco Ricci, and Luigi Ceccaroni. "Exploiting the Semantic Similarity of Contextual Situations for Pre-Filtering Recommendation". In: vol. 7899. June 2013. ISBN: 978-3-642-38843-9. DOI: 10.1007/978-3-642-38844-6_14.

[18] *Covariance (MathWorld)*. https://mathworld.wolfram.com/Covariance.html.

[19] Roddy Cowie, Naomi Sussman, and Aaron Ben-Ze'ev. "Emotion: Concepts and Definitions". In: Oct. 2011, pp. 9–30. ISBN: 978-3-642-15183-5. DOI: 10.1007/978-3-642-15184-2_2.

[20] *Cristinia's Radio 90.4*. http://christianiaradio.com/. 2018.

[21] Sofia Dahl and Prithvi Kantan. *Design and Analysis of Experiments*. https://github.com/SMC-AAU-CPH/Design_and_Analysis_of_Experiments.

[22] Mukund Deshpande and George Karypis. "Item-Based Top-N Recommendation Algorithms". In: *ACM Trans. Inf. Syst.* 22.1 (Jan. 2004), 143–177. ISSN: 1046-8188. DOI: 10.1145/963770.963776. URL: https://doi.org/10.1145/963770.963776.

[23] Mayukh Dewan. "Understanding Ethnography: An 'Exotic' Ethnographer's Perspective". In: Feb. 2018, pp. 185–203. ISBN: 978-981-10-7490-5. DOI: 10.1007/978-981-10-7491-2_10.

[24] Nicola Dibben and Victoria Williamson. "An exploratory survey of in-vehicle music listening". In: *Psychology of Music* 35 (Oct. 2007), p. 571. DOI: 10.1177/0305735607079725.

[25] V.s Dixit and Parul Jain. "Recommendations with context aware framework using particle swarm optimization and unsupervised learning". In: *Journal of Intelligent Fuzzy Systems* 36 (Feb. 2019), pp. 1–12. DOI: 10.3233/JIFS-179001.

[26] Xiaoyue Feng et al. "The Deep Learning–Based Recommender System "Pubmender" for Choosing a Biomedical Publication Venue: Development and Validation Study". In: *J Med Internet Res* 21.5 (2019), e12957. ISSN: 1438-8871. DOI: 10.2196/12957. URL: http://www.jmir.org/2019/5/e12957/.

[27] Andy Field and Graham Hole. *How to Design and Report Experiments*. Jan. 2003.

[28] *Folkehuset Absalon*. https://absaloncph.dk/. 2015.

[29] *Foreigners' Services Provider*. https://sefstat.sef.pt/.

[30] Joan Grusec and Paul Hastings. *Handbook of Socialization*. Jan. 2015. ISBN: 978-1-4625-1834-0.

[31] Casper Hansen et al. "Contextual and Sequential User Embeddings for Large-Scale Music Recommendation". In: *Fourteenth ACM Conference on Recommender Systems*. RecSys '20. Virtual Event, Brazil, 2020, 53–62. ISBN: 9781450375832.

[32] Khalid Haruna et al. "Context-Aware Recommender System: A Review of Recent Developmental Process and Future Research Direction". In: *Applied Sciences* 7 (Dec. 2017), p. 1211. DOI: 10.3390/app7121211.

[33] Christian Helmers, Pramila Krishnan, and Manasa Patnam. "Attention and saliency on the internet: Evidence from an online recommendation system". In: *Journal of Economic Behavior Organization* 161 (2019), pp. 216–242. ISSN: 0167-2681. DOI: https://doi.org/10.1016/j.jebo.2019.04.010. URL: https://www.sciencedirect.com/science/article/pii/S0167268119301131.

[34] Jong-yi Hong, Euiho Suh, and Sung-Jin Kim. "Context-aware systems: A literature review and classification. Expert Systems with Applications, 36(4), 8509-8522". In: *Expert Systems with Applications* 36 (May 2009), pp. 8509–8522. DOI: 10.1016/j.eswa.2008.10.071.

[35] João Horta. *Rádio Paranóia*. https://radioparanoia1000.com/. 2021.

[36] Minqing Hu and Bing Liu. "Mining and Summarizing Customer Reviews". In: *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD '04. Seattle, WA, USA: Association for Computing Machinery, 2004, 168–177. ISBN: 1581138881. DOI: 10.1145/1014052.1014073. URL: https://doi.org/10.1145/1014052.1014073.

[37] Minqing Hu and Bing Liu. "Mining Opinion Features in Customer Reviews". In: July 2004.

[38] Sergio Ilarri et al. "A Review of the Role of Sensors in Mobile Context-Aware Recommendation Systems". In: *International Journal of Distributed Sensor Networks* 2015 (Nov. 2015), pp. 1–30. DOI: 10.1155/2015/489264.

[39] F. O. Isinkaye, Y. Folajimi, and B. Ojokoh. "Recommendation systems: Principles, methods and evaluation". In: *Egyptian Informatics Journal* 16 (2015), pp. 261–273.

[40] Niklas Jakob et al. "Beyond the stars: Exploiting free-text user reviews to improve the accuracy of movie recommendations". In: *International Conference on Information and Knowledge Management, Proceedings* (Jan. 2009). DOI: 10.1145/1651461.1651473.

[41] Saba Jazi, Marjan Kaedi, and Afsaneh Fatemi. "An emotion-aware music recommender system: bridging the user's interaction and music recommendation". In: *Multimedia Tools and Applications* 80 (Apr. 2021). DOI: 10.1007/s11042-020-10386-7.

[42] Mohsen Kamalzadeh, Dominikus Baur, and Torsten Möller. "A survey on music listening and management behaviours". In: *Proceedings of the 13th International Society for Music Information Retrieval Conference (ISMIR 2012), Porto, Portugal* (Jan. 2012).

[43] Marius Kaminskas and Francesco Ricci. "Emotion-Based Matching of Music to Places". In: July 2016, pp. 287–310. ISBN: 978-3-319-31411-2. DOI: 10.1007/978-3-319-31413-6_14.

[44] Michael N. Katehakis and Arthur F. Veinott. "The Multi-Armed Bandit Problem: Decomposition and Computation". In: *Math. Oper. Res.* 12 (1987), pp. 262–268.

[45] B.M. King. "Analysis of Variance". In: Dec. 2010, pp. 32–36. ISBN: 9780080448947. DOI: 10.1016/B978-0-08-044894-7.01306-3.

[46] Joseph Konstan and John Riedl. "Recommender systems: From algorithms to user experience". In: *User Modeling and User-Adapted Interaction* 22 (Apr. 2012), pp. 101–123. DOI: 10.1007/s11257-011-9112-x.

[47] Saurabh Kulkarni and Sunil Rodd. "Context Aware Recommendation Systems: A review of the state of the art techniques". In: *Computer Science Review* 37 (Aug. 2020), p. 100255. DOI: 10.1016/j.cosrev.2020.100255.

[48] *Last.fm*. https://www.last.fm/.

[49] Mengxian Li, Wenjun Jiang, and Kenli Li. "When and What Music Will You Listen To? Fine-Grained Time-Aware Music Recommendation". In: *2017 IEEE International Symposium on Parallel and Distributed Processing with Applications and 2017 IEEE International Conference on Ubiquitous Computing and Commu-*

nications (ISPA/IUCC). 2017, pp. 1091–1098. DOI: 10.1109/ISPA/IUCC.2017.00165.

[50] Blerina Lika, Kostas Kolomvatsos, and Stathes Hadjiefthymiades. “Facing the cold start problem in recommender systems”. In: *Expert Systems with Applications: An International Journal* 41 (Mar. 2014), pp. 2065–2073. DOI: 10.1016/j.eswa.2013.09.005.

[51] *ListenBrainz*. https://listenbrainz.org/.

[52] B. Liu, Minqing Hu, and J. Cheng. “Opinion observer: analyzing and comparing opinions on the Web”. In: *WWW '05*. 2005.

[53] Álvaro Lozano Murciego et al. “Context-Aware Recommender Systems in the Music Domain: A Systematic Literature Review”. In: *Electronics* 10.13 (2021). ISSN: 2079-9292. DOI: 10.3390/electronics10131555. URL: https://www.mdpi.com/2079-9292/10/13/1555.

[54] Laurens van der Maaten and Geoffrey Hinton. “Viualizing data using t-SNE”. In: *Journal of Machine Learning Research* 9 (Nov. 2008), pp. 2579–2605.

[55] Glenn McDonald. *Every Noise at Once*. https://everynoise.com/. 2017.

[56] Silvia Milano, Mariarosaria Taddeo, and Luciano Floridi. “Recommender systems and their ethical challenges”. In: *AI SOCIETY* 35 (Dec. 2020). DOI: 10.1007/s00146-020-00950-y.

[57] Miquel Montaner, Beatriz López, and Josep Rosa. “Developing trust in recommender agents”. In: Jan. 2002, pp. 304–305. DOI: 10.1145/544741.544811.

[58] Lori Morris. “The Cambridge Grammar of the English Language (review)”. In: *The Canadian Journal of Linguistics / La revue canadienne de linguistique* 48 (Jan. 2003), pp. 90–92. DOI: 10.1353/cjl.2004.0007.

[59] *Pandora*. https://www.pandora.com/.

[60] Marco Polignano et al. “Towards Emotion-aware Recommender Systems: an Affective Coherence Model based on Emotion-driven Behaviors”. In: *Expert Systems with Applications* 170 (Dec. 2020), p. 114382. DOI: 10.1016/j.eswa.2020.114382.

[61] Francesco Ricci, Lior Rokach, and Bracha Shapira. *Recommender Systems Handbook*. 2nd. Springer Publishing Company, Incorporated, 2015. ISBN: 1489976361.

[62] John A Rice. *Mathematical statistics and data analysis*. China machine press, 2003.

[63] Ellen Riloff. “An empirical study of automated dictionary construction for information extraction in three domains”. In: *Artificial Intelligence* 85.1 (1996), pp. 101–134. ISSN: 0004-3702. DOI: https://doi.org/10.1016/0004-3702(95)00123-9. URL: https://www.sciencedirect.com/science/article/pii/0004370295001239.

[64] Ellen Riloff and Michael Thelen. "Bootstrapping method for learning semantic lexicons using extraction pattern contexts". In: (July 2002). DOI: 10.3115/1118693.1118721.

[65] Markus Schedl et al. "Listener Modeling and Context-aware Music Recommendation Based on Country Archetypes". In: *CoRR* abs/2009.09935 (2020). arXiv: 2009.09935. URL: https://arxiv.org/abs/2009.09935.

[66] Imen Ben Sessi and Sadok Ben Yahia. "How does context influence music preferences: a user-based study of the effects of contextual information on users' preferred music". In: (Nov. 2020). DOI: 10.1007/s00530-020-00717-x.

[67] Guy Shani and Asela Gunawardana. "Evaluating Recommendation Systems". In: *Recommender Systems Handbook*. Ed. by Francesco Ricci et al. Boston, MA: Springer US, 2011, pp. 257–297. ISBN: 978-0-387-85820-3. DOI: 10.1007/978-0-387-85820-3_8. URL: https://doi.org/10.1007/978-0-387-85820-3_8.

[68] *SimilarWeb*. https://www.similarweb.com/.

[69] Reid Smith. "Knowledge-Based Systems Concepts, Techniques, Examples". In: Jan. 1985.

[70] *Spotify*. https://www.spotify.com/.

[71] *Statistics Denmark*. https://www.dst.dk/en.

[72] Diego Sánchez-Moreno, Yong Zheng, and María N. Moreno-García. "Time-Aware Music Recommender Systems: Modeling the Evolution of Implicit User Preferences and User Listening Habits in A Collaborative Filtering Approach". In: *Applied Sciences* 10.15 (2020), p. 5324. ISSN: 2076-3417. DOI: 10.3390/app10155324. URL: http://dx.doi.org/10.3390/app10155324.

[73] William Thompson and L.-L Balkwill. "Cross-Cultural Similarities and Differences". In: *Handbook of Music and Emotion: Theory, Research, Applications* (July 1993), pp. 755–788. DOI: 10.1093/acprof:oso/9780199230143.003.0027.

[74] Cheng-Hung Tsai et al. "Personal Recommendation Engine of User Behavior Pattern and Analysis on Social Networks". In: *2015 International Conference on Computational Science and Computational Intelligence (CSCI)*. 2015, pp. 404–409. DOI: 10.1109/CSCI.2015.46.

[75] Katrien Verbert et al. "Context-aware Recommender Systems". In: 16.16 (Aug. 28, 2010). |http://www.jucs.org/jucs$_{1}6_{1}6/context_{a}ware_{r}ecommender_{s}ystems$|, pp. 2175–2178.

[76] Hanna Wallach et al. "Evaluation methods for topic models". In: vol. 382. Jan. 2009, p. 139. DOI: 10.1145/1553374.1553515.

[77] Hao Wang, Zonghu Wang, and Weishi Zhang. "Quantitative analysis of Matthew effect and sparsity problem of recommender systems". In: Apr. 2018, pp. 78–82. DOI: 10.1109/ICCCBDA.2018.8386490.

[78] Ann Werner. "Organizing music, organizing gender: algorithmic culture and Spotify recommendations". In: *Popular Communication* 18.1 (2020), pp. 78–90. DOI: 10.1080/15405702.2020.1715980. eprint: https://doi.org/10.1080/15405702.2020.1715980. URL: https://doi.org/10.1080/15405702.2020.1715980.

[79] *"What is Ethnography?"*. https://anthropology.princeton.edu/undergraduate/what-ethnography.

[80] J. Wiebe and Ellen Riloff. "Finding Mutual Benefit between Subjectivity Analysis and Information Extraction". In: *IEEE Transactions on Affective Computing* 2.4 (2011), pp. 175–191. DOI: 10.1109/T-AFFC.2011.19.

[81] yi-hsuan Yang and Jen-Yu Liu. "Quantitative Study of Music Listening Behavior in a Social and Affective Context". In: *Multimedia, IEEE Transactions on* 15 (Oct. 2013), pp. 1304–1315. DOI: 10.1109/TMM.2013.2265078.

[82] Chunyi Zhou et al. "Conversational Music Recommendation based on Bandits". In: *2020 IEEE International Conference on Knowledge Graph (ICKG)*. 2020, pp. 41–48. DOI: 10.1109/ICBK50248.2020.00016.

[83] Cai-Nicolas Ziegler et al. "Improving recommendation lists through topic diversification". In: Jan. 2005. DOI: 10.1145/1060745.1060754.

Appendix A

Appendix A name

A.1 Questionnaire

Questionnaire on music listening behavior when and when not at work.
In order to take this test you must be residing in either the city of Copenhagen or Lisbon, be employed and working. Please answer to all questions the most accurate way possible.

In which city do you have a residence right now?

- Copenhagen
- Lisbon

In total how many year have you lived in this city?

- <5
- 5-10
- 11-15
- 16-20
- 21-25
- 25-30
- >30

How old are you at the present moment?

- <15

- 16-25
- 26-35
- 36-45
- 46-55
- 56-65
- >65

Select your nationality.

If you have double nationality, select your other nationality. Otherwise, choose the option "Doesn't apply".

Select the gender you identify with.

- Female
- Male
- Other

Select the type of platforms you use to listen to music. Check all that apply.

- Music media player (audio files required)
- Music streaming app (online audio stream)
- Both equally
- None of the above

From 1 to 7, where 1 is never and 7 is always, select how often you listen to music in a determined playback method when working.

- Song after song
- Artist
- Album or genre
- Playlist or folder
- Shuffle on collection
- Online recommendation

- Radio

From 1 to 7, where 1 is never and 7 is always, select how often you listen to music in a determined playback method when not working.

- Song after song
- Artist
- Album or genre
- Playlist or folder
- Shuffle on collection
- Online recommendation
- Radio

The questionnaire is over. Thank you for completing it.

A.2 Histogram data & Copenhagen and Lisbon populations

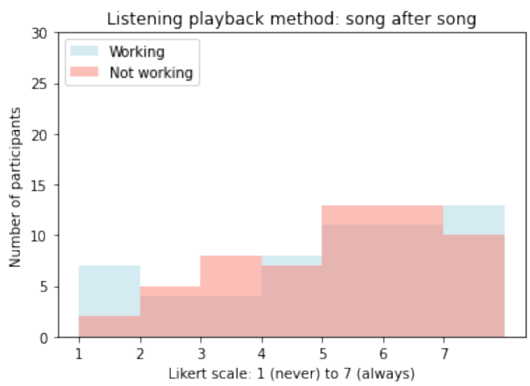


Figure A.1: Comparison of working and not working passive music listening preference to song after song among all participants.

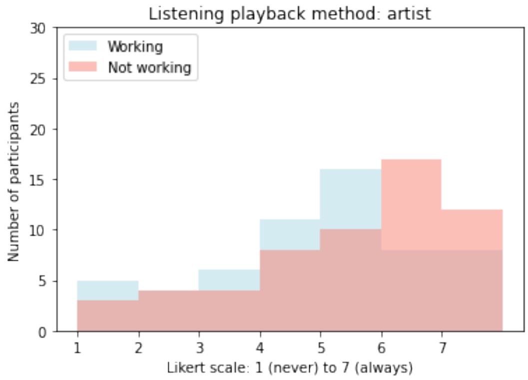


Figure A.2: Comparison of working and not working passive music listening preference to artist among all participants.

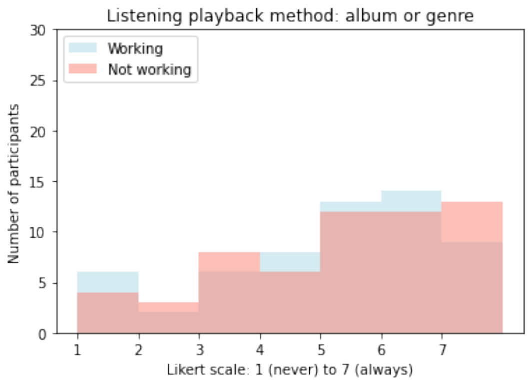


Figure A.3: Comparison of working and not working passive music listening preference to album or genre among all participants.

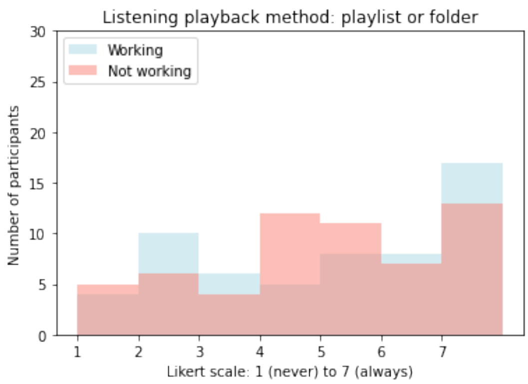


Figure A.4: Comparison of working and not working passive music listening preference to playlist or folder among all participants.

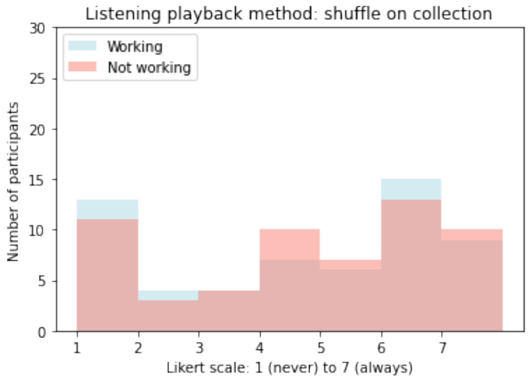


Figure A.5: Comparison of working and not working passive music listening preference to shuffle on collection among all participants.

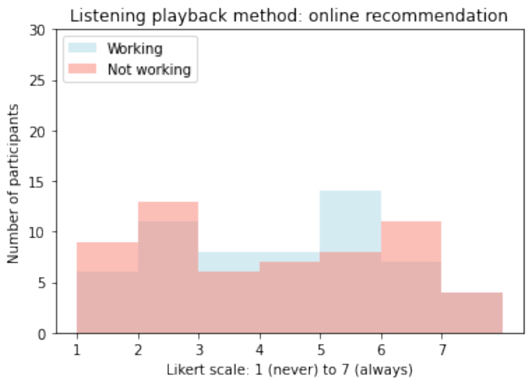


Figure A.6: Comparison of working and not working passive music listening preference to online recommendation among all participants.

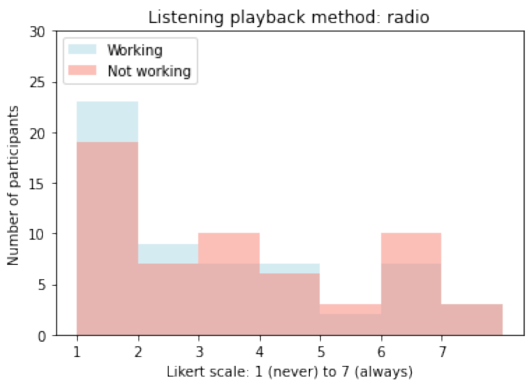


Figure A.7: Comparison of working and not working passive music listening preference to radio among all participants.

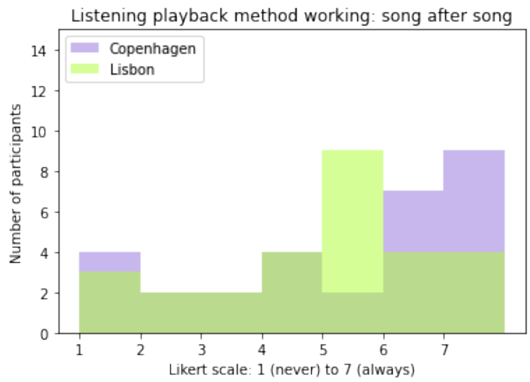


Figure A.8: Comparison of Copenhagen and Lisbon residents passive music listening preference to song after song when working.

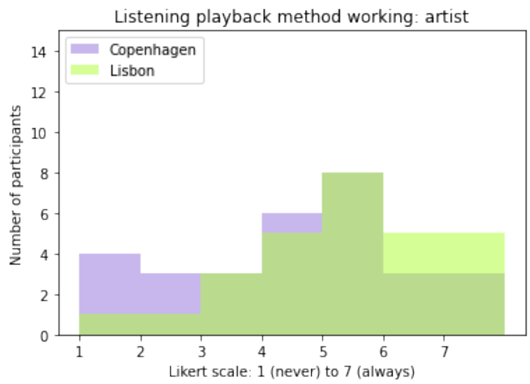


Figure A.9: Comparison of Copenhagen and Lisbon residents passive music listening preference to artist when working.

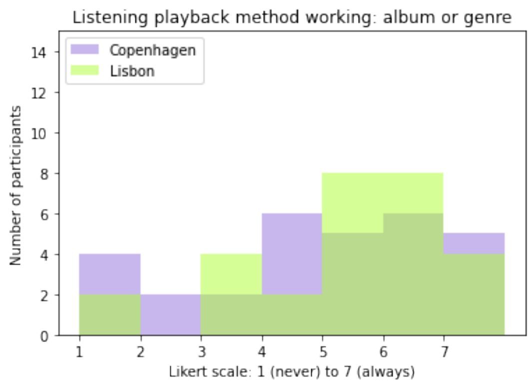


Figure A.10: Comparison of Copenhagen and Lisbon residents passive music listening preference to album or genre when working.

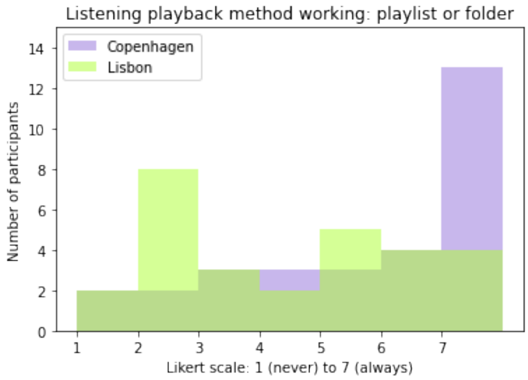


Figure A.11: Comparison of Copenhagen and Lisbon residents passive music listening preference to playlist or folder when working.

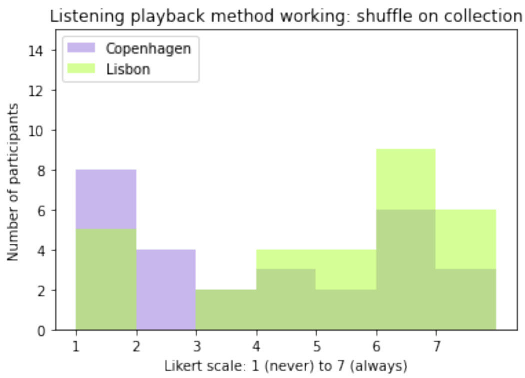


Figure A.12: Comparison of Copenhagen and Lisbon residents passive music listening preference to shuffle on collection when working.

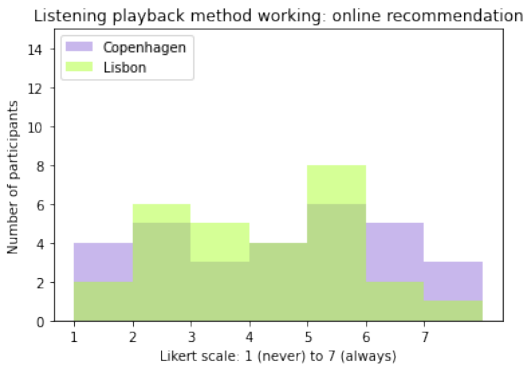


Figure A.13: Comparison of Copenhagen and Lisbon residents passive music listening preference to online recommendation when working.

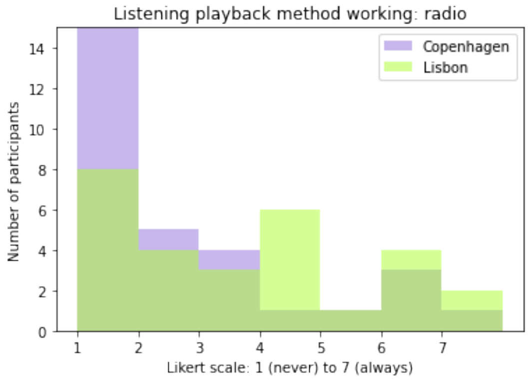


Figure A.14: Comparison of Copenhagen and Lisbon residents passive music listening preference to radio when working.

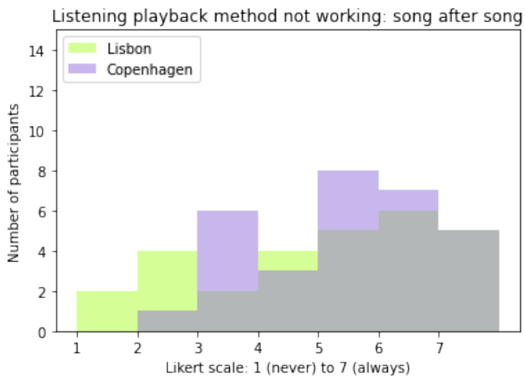


Figure A.15: Comparison of Copenhagen and Lisbon residents passive music listening preference to song after song when not working.

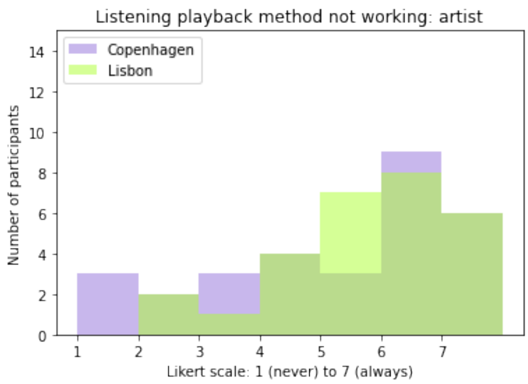


Figure A.16: Comparison of Copenhagen and Lisbon residents passive music listening preference to artist when not working.

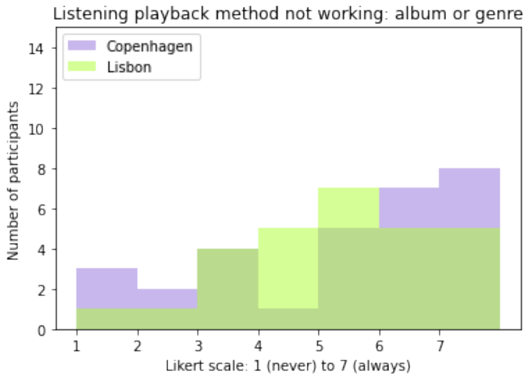


Figure A.17: Comparison of Copenhagen and Lisbon residents passive music listening preference to album or genre when not working.

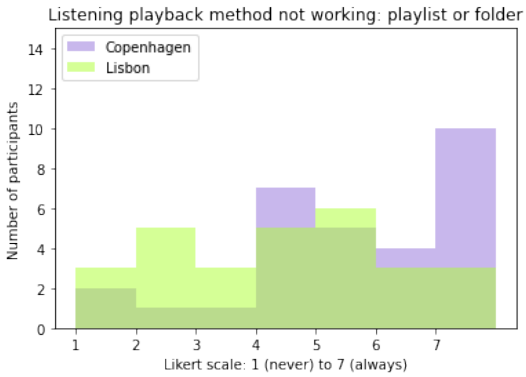


Figure A.18: Comparison of Copenhagen and Lisbon residents passive music listening preference to playlist or folder when not working.

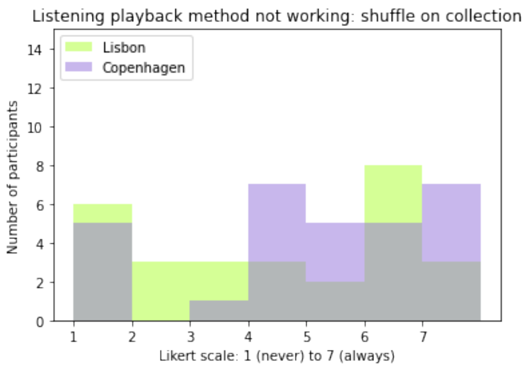


Figure A.19: Comparison of Copenhagen and Lisbon residents passive music listening preference to shuffle on collection when not working.

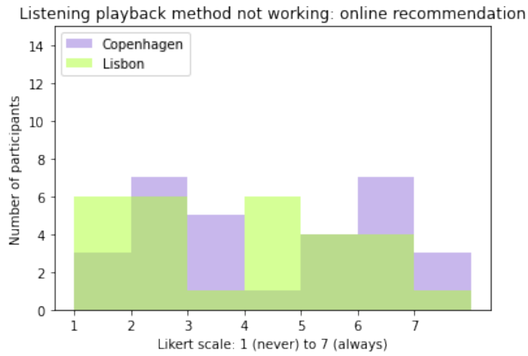


Figure A.20: Comparison of Copenhagen and Lisbon residents passive music listening preference to online recommendation when not working.

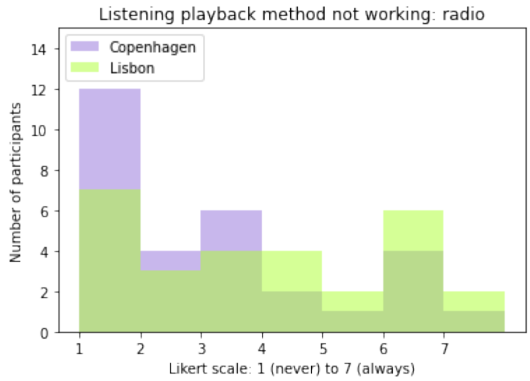
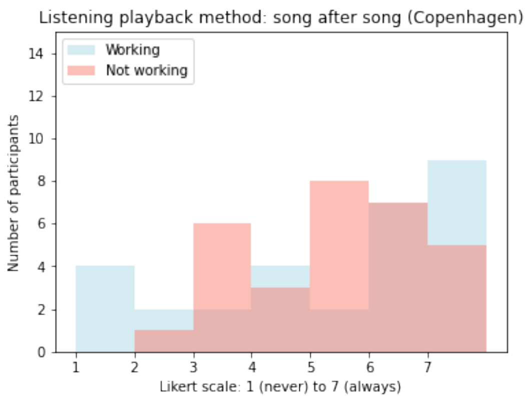


Figure A.21: Comparison of Copenhagen and Lisbon residents passive music listening preference to radio when not working.



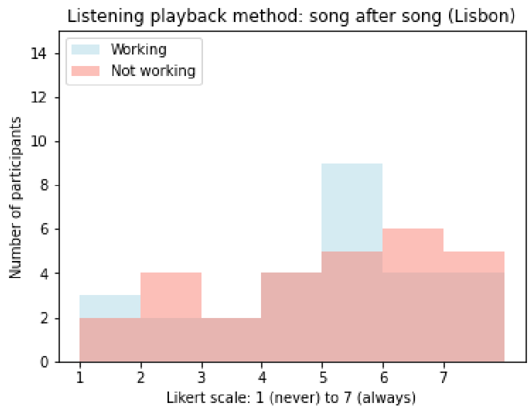
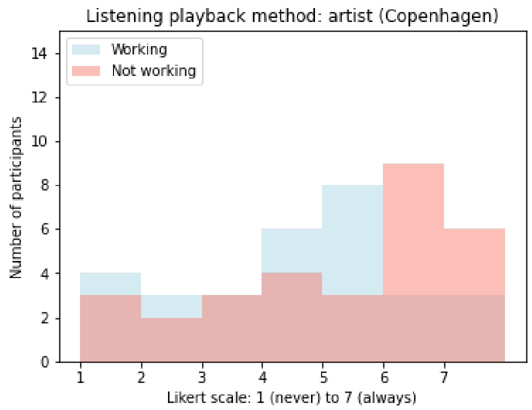


Figure A.22: Comparison of working and not working passive music listening preference to song after song between Copenhagen and Lisbon residents.



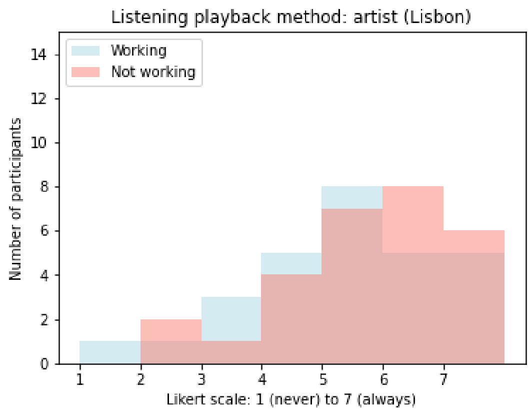
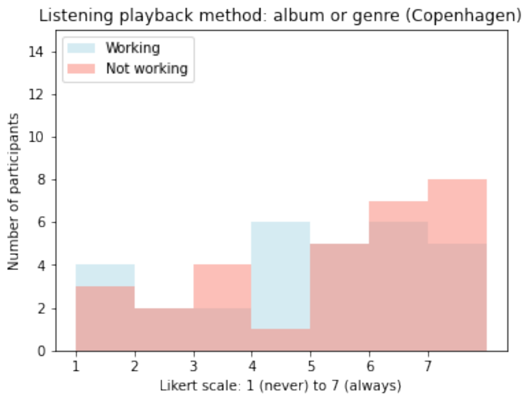


Figure A.23: Comparison of working and not working passive music listening preference to artist between Copenhagen and Lisbon residents.



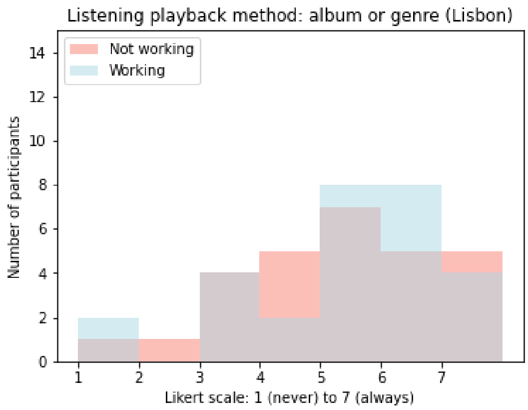
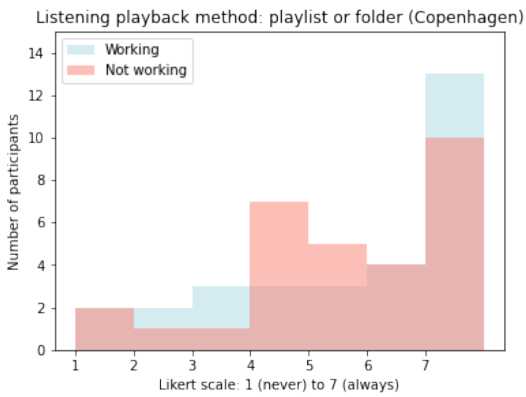


Figure A.24: Comparison of working and not working passive music listening preference to album or genre between Copenhagen and Lisbon residents.



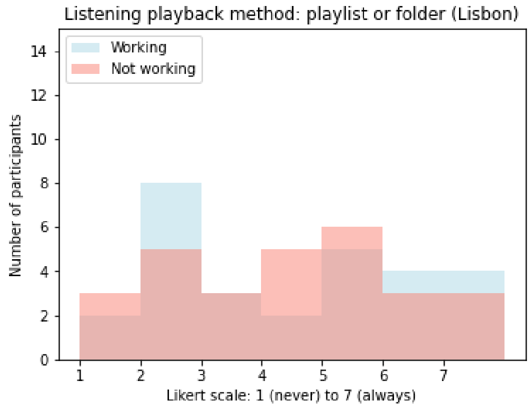
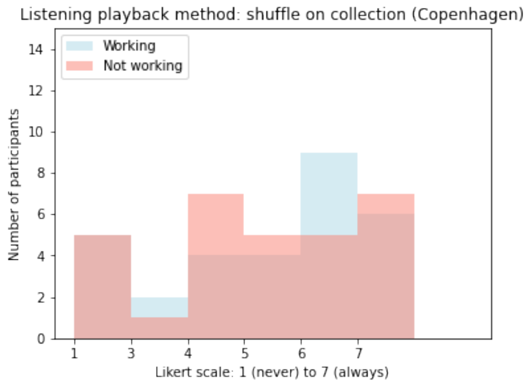


Figure A.25: Comparison of working and not working passive music listening preference to playlist or folder between Copenhagen and Lisbon residents.



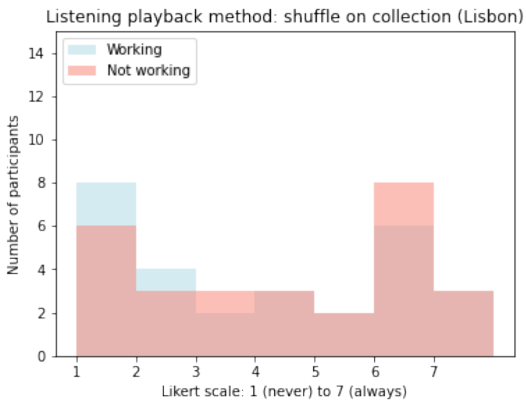
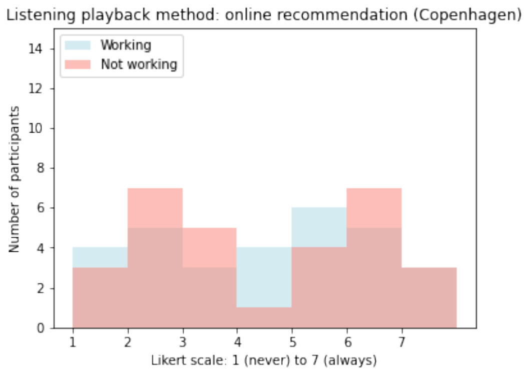


Figure A.26: Comparison of working and not working passive music listening preference to shuffle on collection between Copenhagen and Lisbon residents.



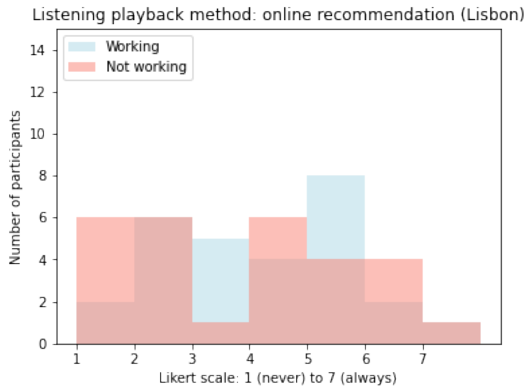
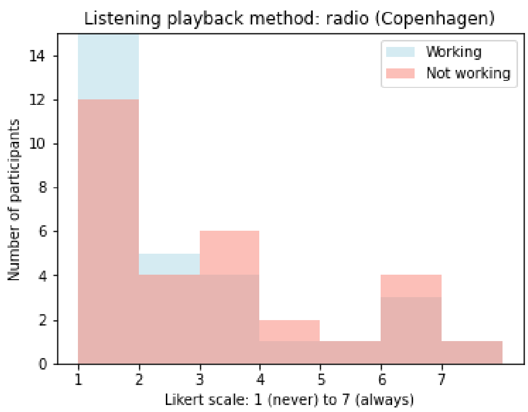


Figure A.27: Comparison of working and not working passive music listening preference to online recommendation between Copenhagen and Lisbon residents.



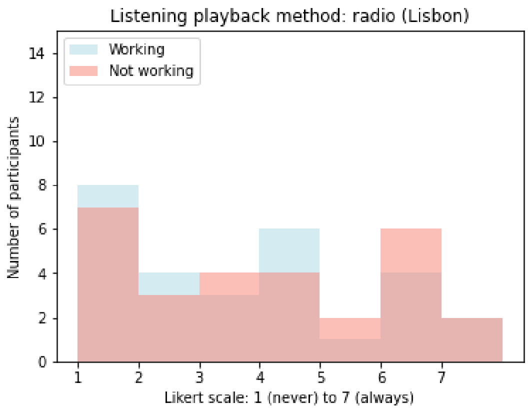


Figure A.28: Comparison of working and not working passive music listening preference to radio between Copenhagen and Lisbon residents.

	Country	Total	Men	Women
1	Pakistan	8699	4509	4190
2	Turkey	7470	3834	3636
3	Iraq	6960	3781	3179
4	Germany	6473	3113	3360
5	Poland	6346	2810	3536
6	Somalia	5397	2828	2569
7	Sweden	5326	2288	3038
8	Morocco	5287	2676	2611
9	United Kingdom	5196	3321	1875
10	Lebanon	5068	2657	2411
11	Norway	4695	1787	2908
12	Italy	4434	2625	1809
13	Iran	4123	2259	1864
14	China	4098	1723	2375
15	India	4098	2343	1755
16	USA	3738	1857	1881
17	Romania	3320	1636	1684
18	France	3229	1738	1491
19	Yugoslavia	3139	1546	1593
20	Spain	2988	1532	1456
21	Philippines	2505	724	1781
22	Nepal	2463	1326	1137
23	Republic of North Macedonia	2343	1226	1117
24	Iceland	2211	979	1232
25	Bosnia and Herzegovina	2193	1102	1091
26	Thailand	2015	487	1528
27	Argentina	1958	1074	884
28	Lithuania	1897	780	1117
29	Bulgaria	1870	975	895
30	Syria	1820	964	856
31	Afghanistan	1801	967	834
32	Russia	1651	605	1046
33	Greece	1628	925	703
34	Vietnam	1559	734	825
35	Brazil	1482	619	863
36	Finland	1367	485	882
37	Netherlands	1361	788	573
38	Portugal	1336	788	548
39	Hungary	1233	568	665
40	Jordan	1196	651	545
41	Bangladesh	1124	639	485
42	Australia	1053	588	465
43	Ukraine	978	350	628
44	Egypt	961	609	352
45	Latvia	924	386	538
46	Canada	870	445	425
47	Chile	812	434	378
48	Ghana	809	485	324
49	Slovakia	803	378	425
50	Croatia	679	343	336
51	Algeria	641	378	263
52	Ethiopia	625	307	318
53	Ireland	623	385	238
54	Sri Lanka	577	289	288
55	South Korea	574	209	365
56	Japan	569	198	371
57	Switzerland	562	265	297
58	Mexico	560	294	266
59	Tunisia	557	325	232
60	Nigeria	503	316	187
61	Israel	497	315	182
62	Uganda	495	250	245
63	Belgium	484	256	228
64	Austria	466	228	238
65	Czech Republic	456	201	255
66	Estonia	442	125	317
67	Colombia	434	218	216
68	Gambia, The	412	228	184
69	South Africa	383	221	162
70	Kenya	361	149	212
71	Peru	345	155	190
72	Soviet Union	344	142	202
73	Serbia	334	135	199
74	Kuwait	324	193	131
75	Venezuela	271	130	141
76	Tanzania	256	134	122
77	Indonesia	247	109	138
78	New Zealand	241	144	97
79	Moldova	238	116	122
80	Albania	236	132	104
81	Eritrea	233	129	104
82	Cuba	230	128	102
83	Ivory Coast	219	109	110
84	Sudan	214	129	85
85	Kosovo	212	103	109
86	Armenia	208	88	120
87	Cameroon	199	112	87
88	Singapore	195	82	113
89	Taiwan	173	61	112
90	Belarus	167	54	113
91	Malaysia	165	68	97
92	Slovenia	152	65	87
93	Ecuador	149	72	77
94	Sierra Leone	147	83	64
95	Montenegro	134	63	71
96	Czechoslovakia	132	54	78
97	Middle East not stated	132	85	47
98	Congo, Democratic Republic	124	68	56

99	Uruguay	124	62	62
100	Burundi	122	56	66
101	Saudi Arabia	122	66	56
102	Yugoslavia, Federal Republic	121	56	65
103	Azerbaijan	108	53	55
104	Stateless	101	59	42
105	Yemen	100	42	58
106	Guinea-Bissau	98	70	28
107	Kazakhstan	90	36	54
108	Georgia	86	31	55
109	Libya	86	53	33
110	Rwanda	80	29	51
111	Guinea	78	42	36
112	United Arab Emirates	77	44	33
113	Luxembourg	76	36	40
114	Senegal	76	37	39
115	Myanmar	71	50	21
116	Zimbabwe	71	40	31
117	Guatemala	69	38	31
118	Bolivia	66	27	39
119	Dominican Republic	63	32	31
120	Zambia	62	28	34
121	Cambodia	60	28	32
122	Mozambique	60	28	32
123	Togo	60	37	23
124	Nicaragua	59	26	33
125	Cyprus	54	23	31
126	Costa Rica	51	28	23
127	Honduras	51	20	31
128	Jamaica	50	30	20
129	Uzbekistan	47	23	24
130	Angola	46	21	25
131	Serbia and Montenegro	45	23	22
132	Not stated	44	25	19
133	Djibouti	40	20	20
134	Mauritius	40	22	18
135	Trinidad and Tobago	40	18	22
136	Liberia	39	22	17
137	Bahrain	38	14	24
138	Panama	37	19	18
139	Congo, Republic	34	19	15
140	North Korea	31	10	21
141	Malta	28	18	10
142	El Salvador	26	9	17
143	Guyana	24	7	17
144	Mali	23	12	11
145	Mongolia	20	4	16
146	Paraguay	20	8	12
147	Benin	19	12	7
148	Cape Verde	16	7	9
149	Kyrgyzstan	16	5	11
150	Madagascar	16	7	9
151	Africa not stated	14	6	8
152	Bhutan	14	8	6
153	Burkina Faso	14	10	4
154	Haiti	14	9	5
155	Namibia	13	9	4
156	Mauritania	12	6	6
157	Barbados	11	7	4
158	Eswantini	11	5	6
159	Laos	11	3	8
160	Malawi	11	8	3
161	Botswana	10	5	5
162	Comoros	10	4	6
163	South Sudan	10	3	7
164	Tajikistan	10	3	7
165	Lesotho	9	6	3
166	Suriname	9	6	3
167	Asia not stated	8	5	3
168	Dominica	7	2	5
169	Oman	7	5	2
170	Belize	6	4	2
171	Central African Republic	6	3	3
172	Qatar	6	3	3
173	Brunei	5	2	3
174	Equatorial Guinea	5	2	3
175	Fiji	5	4	1
176	Liechtenstein	5	2	3
177	Monaco	5	3	2
178	Seychelles	5	2	3
179	Bahamas	4	3	1
180	Grenada	4	2	2
181	Niger	4	3	1
182	Europe not stated	3	2	1
183	Pacific Islands	3	1	2
184	Saint Lucia	3	2	1
185	San Marino	3	1	2
186	Tonga	3	2	1
187	Andorra	2	2	0
188	Antigua and Barbuda	2	0	2
189	Chad	2	1	1
190	Sao Tome and Principe	2	0	2
191	Turkmenistan	2	1	1
192	Gabon	1	1	0
193	Kiribati	1	1	0
194	Papua New Guinea	1	1	0
195	Samoa	1	1	0
196	West Indies	1	0	1

Figure A.29: Non-danish population in Copenhagen by nationality and gender

	Country	Total	Men	Women
1	Brazil	60469	26293	34176
2	Cape Verde	23364	10622	12742
3	China	14662	7171	7491
4	Guinea Bissau	13779	7159	6620
5	Angola	13512	6090	7422
6	Italy	11684	6455	5229
7	Romania	10613	5547	5066
8	France	10189	5373	4816
9	Nepal	10080	6113	3967
10	Ukraine	8911	3865	5046
11	India	8715	6100	2615
12	Spain	7650	3963	3687
13	UK	6171	3605	2566
14	Sao Tome and Principe	6094	2597	3497
15	Bangladesh	6076	4753	1323
16	Germany	6004	3331	2673
17	Netherlands	3468	1940	1528
18	Pakistan	3461	2435	1026
19	Sweden	2159	1200	959
20	USA	1909	976	933
21	Russia	1854	640	1214
22	Moldavia	1845	806	1039
23	Bulgaria	1765	849	916
24	Mozambique	1704	727	977
25	Belgium	1665	961	704
26	Poland	1275	412	863
27	Guinea	1103	739	364
28	Turkey	950	467	483
29	Senegal	874	603	271
30	South Africa	792	378	414
31	Venezuela	786	312	474
32	Austria	764	386	378
33	Switzerland	734	395	339
34	Ireland	715	443	272
35	Philippines	686	214	472
36	Denmark	653	391	262
37	Colombia	529	224	305
38	Syria	510	285	225
39	Finland	491	214	277
40	Iran	449	253	196
41	Hungary	436	171	265
42	Norway	433	239	194
43	Thailand	432	233	199
44	Lebanon	429	244	185
45	Greece	407	189	218
46	Nigeria	394	255	139
47	Jordan	378	209	169
48	Vietnam	371	187	184
49	Morocco	368	191	177
50	Canada	344	181	163
51	Egypt	311	190	121
52	Argentina	283	132	151
53	Lithuania	276	78	198
54	Mexico	256	97	159
55	Iraq	253	131	122
56	Cuba	252	106	146
57	Japan	234	93	141
58	Czech Republic	195	65	130
59	Latvia	190	53	137
60	Australia	184	98	86
61	Algeria	179	103	76
62	Georgia	177	99	78
63	Croatia	174	79	95
64	Belarus	165	52	113
65	Saudi Arabia	138	73	65
66	Kazakhstan	136	44	92
67	Gambia	136	107	29
68	South Korea	134	56	78
69	Tunisia	134	78	56
70	Congo (Democratic Republic)	131	74	57
71	Estonia	131	38	93
72	Slovakia	130	44	86
73	Chile	123	55	68
74	Libya	123	65	58
75	Uzbekistan	123	61	62
76	Peru	119	38	81
77	Israel	100	53	47
78	Eritrea	98	65	33
79	Serbia	98	45	53
80	Luxembourg	97	51	46
81	Slovenia	91	38	53
82	Ecuador	87	43	44
83	Cameroon	86	44	42
84	Kenya	76	32	44
85	Costa do Marfim	70	48	22
86	Ghana	65	48	17
87	Uruguay	64	32	32
88	Sri Lanka	62	41	21
89	Indonesia	61	24	37
90	Cyprus	60	35	25
91	Saint Kitts and Nevis	59	32	27
92	Malaysia	53	18	35
93	Sudan	51	35	16
94	Palestine	49	27	22
95	Paraguay	49	16	33
96	East Timor	49	24	25

97	Albania	44	18	26
98	Armenia	44	21	23
99	Congo	44	23	21
100	Afghanistan	43	25	18
101	Azerbaijan	43	21	22
102	kuwait	42	25	17
103	Sierra Leone	42	21	21
104	Bolivia	40	11	29
105	Iceland	36	20	16
106	Panama	36	14	22
107	Zimbabwe	36	18	18
108	El Salvador	33	18	15
109	Ethiopia	33	15	18
110	Macedonia	33	15	18
111	New Zealand	33	15	18
112	Somalia	33	17	16
113	Dominican Republic	32	12	20
114	United Arab Emirates	30	18	12
115	Togo	30	16	14
116	Costa Rica	29	11	18
117	Malta	29	15	14
118	mali	26	24	2
119	Tanzania	23	16	7
120	Equatorial Guinea	22	13	9
121	Bosnia and Herzegovina	21	11	10
122	Mauritius (Islands)	21	13	8
123	Taiwan	21	9	12
124	Yemen	19	13	6
125	Uganda	19	7	12
126	Guatemala	18	10	8
127	Singapore	18	4	14
128	stateless person	17	12	5
129	Honduras	15	6	9
130	Hong Kong	15	11	4
131	Gabon	14	6	8
132	Kyrgyzstan	14	3	11
133	Kosovo	13	6	7
134	Mauritania	13	12	1
135	Rwanda	13	6	7
136	Nicaragua	10	3	7
137	Burkina Faso	9	7	2
138	Mongolia	9	4	5
139	Namibia	9	3	6
140	Benin	8	7	1
141	Dominica	8	6	2
142	Liberia	8	7	1
143	Montenegro	8	3	5
144	Bahrain	7	4	3
145	Jamaica	7	4	3
146	Liechtenstein	7	4	3
147	Malawi	7	3	4
148	Belize	6	3	3
149	Unknown	6	3	3
150	Djibouti	6	4	2
151	Madagascar	6	0	6
152	Southern Sudan	6	2	4
153	Turkmenistan	6	4	2
154	Guyana	5	4	1
155	Maldives	5	0	5
156	Oman	5	3	2
157	UK (British Subject)	5	1	4
158	Trinidad and Tobago	5	3	2
159	Cambodia	4	1	3
160	Andorra	3	3	0
161	Burundi	3	1	2
162	Fiji (Islands)	3	2	1
163	Qatar	3	1	2
164	Botswana	2	1	1
165	Lesotho	2	0	2
166	Papua New Guinea	2	1	1
167	Central African Republic	2	1	1
168	Swaziland	2	0	2
169	Vanuatu	2	2	0
170	Antigua and Barbuda	1	1	0
171	Bahamas	1	0	1
172	Barbados	1	1	0
173	Grenade	1	1	0
174	Haiti	1	1	0
175	Marshall (Islands)	1	1	0
176	Burma (Myanmar)	1	1	0
177	Nauru	1	1	0
178	Niger	1	1	0
179	Suriname	1	0	1
180	Tajikistan	1	0	1
181	Zambia	1	0	1

Figure A.30: Non-portuguese population in Lisbon by nationality and gender