

ST344 Report

Group 7

15/11/2019

Contents

1	Introduction	1
2	The Problem	1
3	Popularity prediction	2
3.1	Exploratory Data Analysis	2
3.2	Prediction	3
3.3	Modelling	5
3.4	Conclusions	6
4	Track recommendations	7
4.1	Testing	7
4.2	Conclusion	8
	References	8

1 Introduction

The advancement of the internet and digital technology has had a wide range of implications in most industries. Within the music industry, the turn of the 21st century came with a difficult set of challenges. With little regulation and the initial rise of file-sharing services such as Napster and LimeWire, many predicted that the music industry was heading towards disastrous losses in sales. At first, sales continued to plummet partially due to piracy (Halliday 2010). However, due to rapid technological change, it created opportunities for companies to have potentially different ways of distributing, marketing and promoting creative content. Also, one can conclude that the advancement of technology has indirectly increased the popularity of concerts creating alternative business models and opportunities (Wang 2016).

Fast forward to today, more than half of record companies income comes from streaming services alone. In 2018 the big record labels (Sony, Warner and Universal) made a combined \$6.93 billion, \$1.6 billion more than the companies made in 2017. As the influence of streaming platforms like Spotify and Apple Music continues to grow, the domination of physical music sales has been decreasing drastically. All three major record companies saw multimillion-dollar declines in sales of physical formats, such as CDs, in 2018, with Sony losing as much as a quarter-billion dollars (Wang 2019).

Many streaming services, such as Spotify and Deezer, have APIs (Application Program Interface) that allow the use of data collection by streaming services. Within Spotify's API, developers can use calculated audio features to learn about the quantifiable variables (Audio Features) of tracks such as Danceability, Energy, Key, Loudness, Speechiness, Acousticness, Instrumentalness, Liveness, Valence and Tempo (Spotify 2019a) (Spotify 2019b). With all of this available, record companies are now able to attempt to use the collected data to predict what type of music will be popular. This will give the companies an idea on which artists to invest to make a profit.

2 The Problem

In this report, we were tasked with providing some insight on the type of music that will be popular within the near future using the audio features to conclude our prediction. We will also try to create a model that will recommend new music to consumers based on their tastes. This is to help the record company to figure out the

type of artists they should invest in and in return, maximise their profit. The data set we have been provided with, **‘edited_spotify.xlsx’**, contains 35 of the best-selling albums from each decade, in the UK, from the 1960s to 2010s, along with some less successful albums. The data was obtained from Spotify on 25th September 2019. We will be testing the relationship between the audio features of the albums against its popularity to predict the type of artists and albums that would become popular.

We have learnt that the albums were picked based on what the team leader perceived to be popular at the time adding a bias to the data set he created. This will therefore lead our results to be biased. The team leader has given us a test set which contains 10 albums, 2 from each decade, to test the accuracy of our prediction popularity models. He also requires us to give recommendations on the tracks from the test set only. This can be difficult given the small sample size available in the test set and may not present the actual population (all albums released from the 1960s to 2010s).

3 Popularity prediction

Initially for this task it was thought that genre would be a useful variable to consider in the model, however, after some research into how Spotify assigns genres, the opposite conclusion was reached. The main issue is the genre of an artist (the data we have) is not a good indicator of the actual genre a track might be. This is highlighted in an article talking to Glenn McDonald - Spotify’s ‘Data Alchemist’(McDonald 2013). In this article from 2013 it shows that searching for the top ten artists in Rock gives Rihanna in first place, followed by Justin Timberlake, Bruno Mars, Taylor Swift, Demi Lovato and Lil Wayne. Although people may have different opinions on what counts as Rock music, these artists would not appear on most people’s list. To further this point, searching for Pop artists resulted in the same top 9 with only 10th place changing. However, searching for the top 10 songs in the Rock genre gives a probably more agreed upon list of Rock including songs such as Queen “We Are The Champions”, Cheap Trick “Surrender” and Led Zeppelin “Black Dog”. Our data set only includes artist genre and due to this reasoning, we didn’t investigate genre any further than the frequency of each genre in our data set. For a more complex model to predict popularity it would have been useful to have access to the genres allocated to each individual track.

It was decided to predict track characteristics such as speechiness, acousticness or tempo as they describe well what type of music is going to be popular.

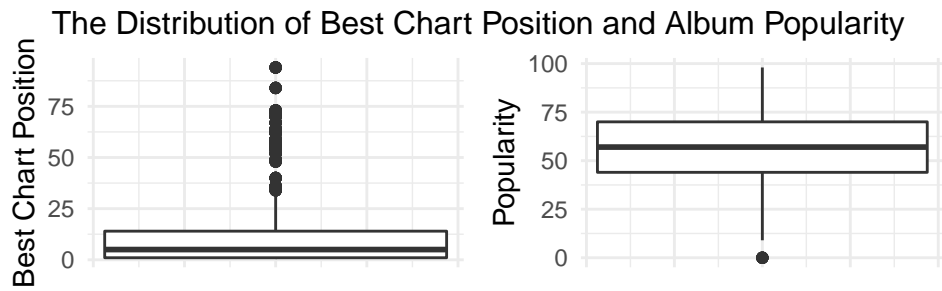
It needs to be defined what we understand by popularity. All the factors describing popularity from the data set (e.g. `AlbumBestChartPosition` or `AlbumPopularity`) are related to a whole album instead of particular tracks, even though it is often the case that an album is popular because of a specific hit single. Therefore, a major limitation of the analysis is the impossibility to evaluate how the characteristics of a certain track affect its popularity.

From now on, it will be assumed that the features of every track on the album influence its popularity.

3.1 Exploratory Data Analysis

We will focus on `AlbumBestChartPosition` and `AlbumPopularity` to define popularity. If the album sells well and is streamed a high number of times, it is high on the charts, hence generating income for the producers. If it is popular by Spotify’s standards (`AlbumPopularity`), it generates sufficient income from streams. `AlbumWeeksOnChart` or `AlbumWeeksNumberOne` will not be considered, assuming that if an album reaches a certain position on the charts, it has already brought additional income for the record label.

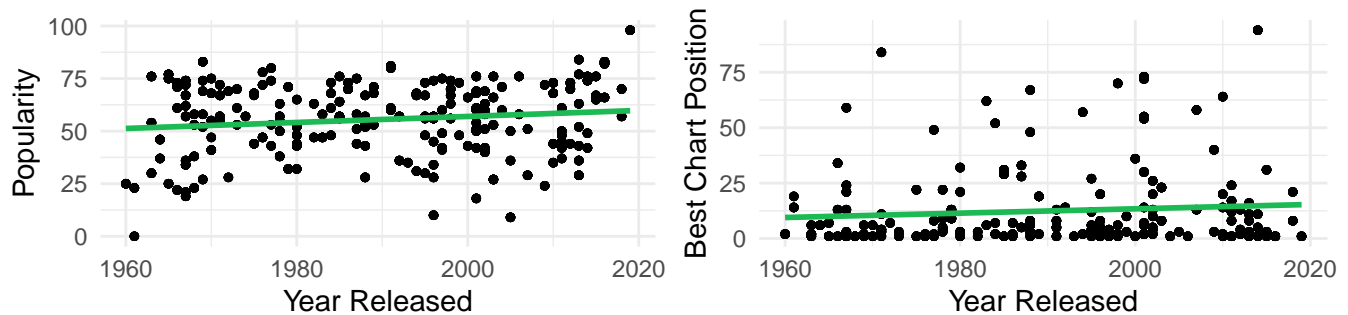
Let us consider the distribution of `AlbumBestChartPosition` and `AlbumPopularity`. We will divide the data set into popular and unpopular tracks based on the distribution, assuming that an artist generates income for their record company over a certain popularity threshold.



22.7 % of tracks reached the number one position and 22 % have popularity over 70. Reaching number one is the intuitive sign of success accomplished by only a fraction of tracks from the data set. We will accept the popularity of 70 to serve as a similar benchmark for streaming, given that it was reached by a similar fraction of tracks. Therefore, popularity will be defined as reaching number one on the charts or being above 70 in `AlbumPopularity`.

It has to be noted that unpopular albums are not represented well in the data set. It is not true that 22.7% of randomly selected albums reach the number one spot. The selection bias in favour of popular albums will affect the predictions by not providing a sufficient sample of unpopular albums.

We may also ask if the average value of `AlbumPopularity` or `AlbumBestChartPosition` change over the decades.

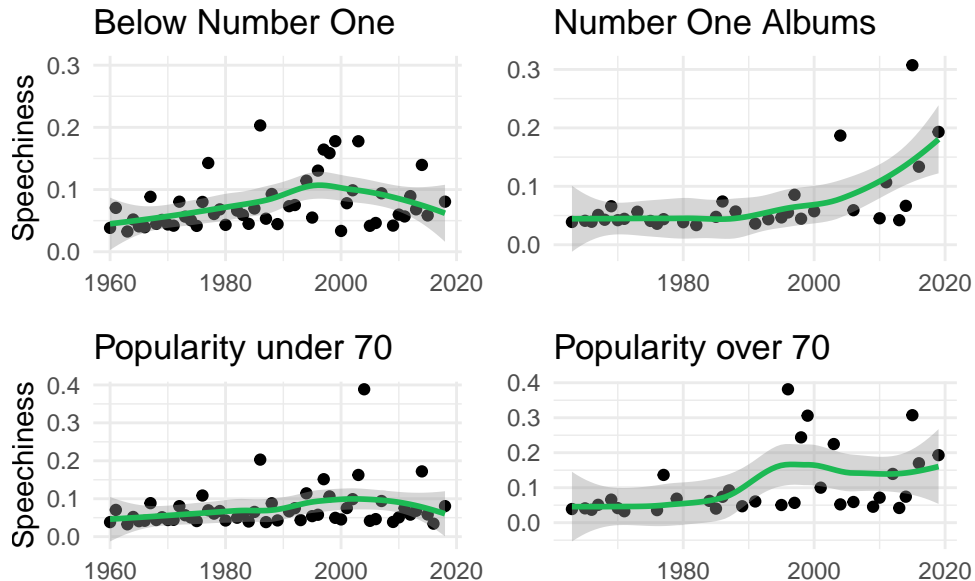


There is a minor increase over time, but not to the extent where this would have a significant impact on the analysis.

3.2 Prediction

3.2.1 Speechiness

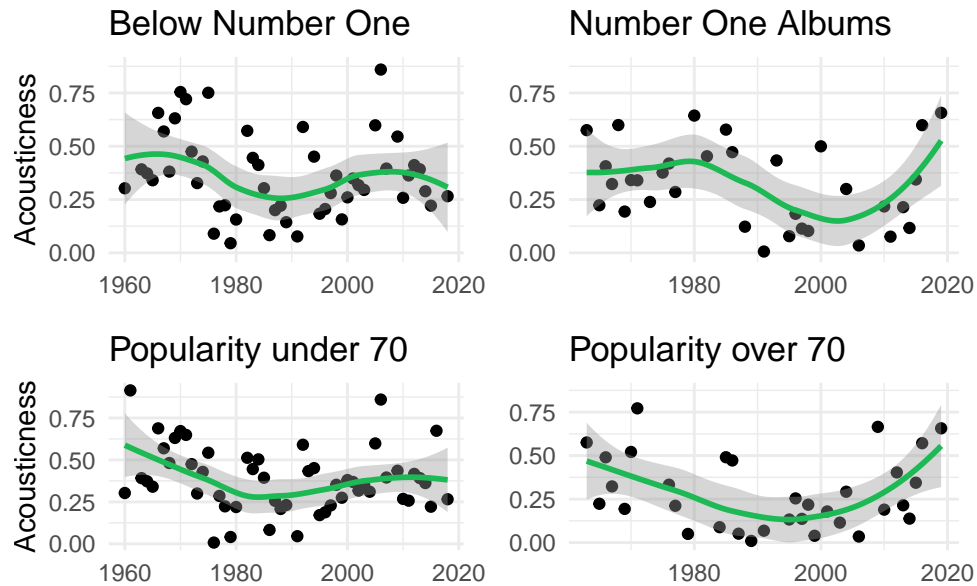
The gradual increase in the popularity of rap music is acknowledged by music experts; let us check if the speechiness of the popular tracks reflects that.



It is clearly visible that the speechiness factor is rising for the popular tracks.

3.2.2 Acousticness

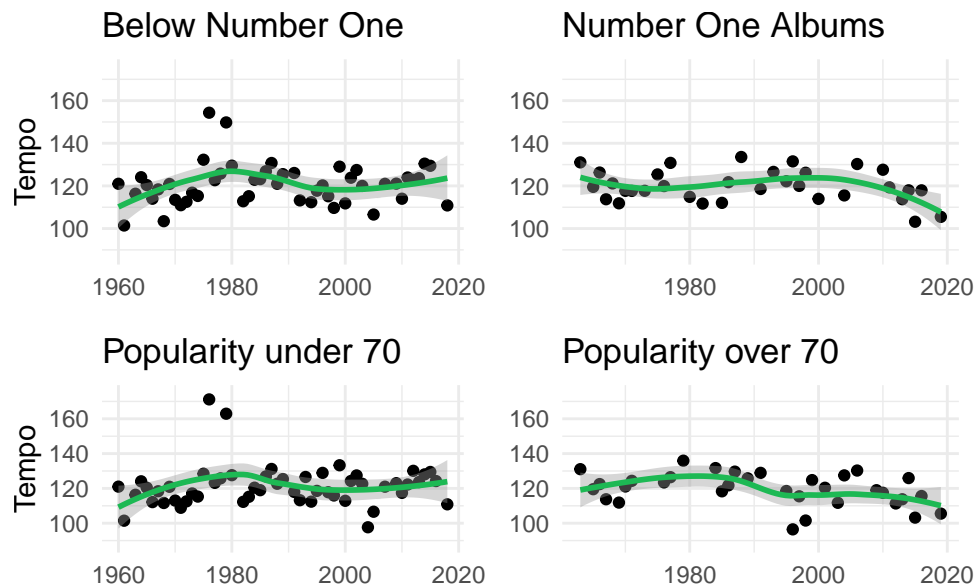
The popularity of artists such as Ed Sheeran, John Mayer or Jason Mraz may be proof of the comeback of acoustic music, a trend highlighted in the recent article by LastMinuteMusicians (LastMinuteMusicians 2018).



The data supports this hypothesis with a consistent upward trend in the popular tracks.

3.2.3 Tempo

The decrease in the average tempo of hit songs was noted in this article in the Rolling Stone (Leight 2017).



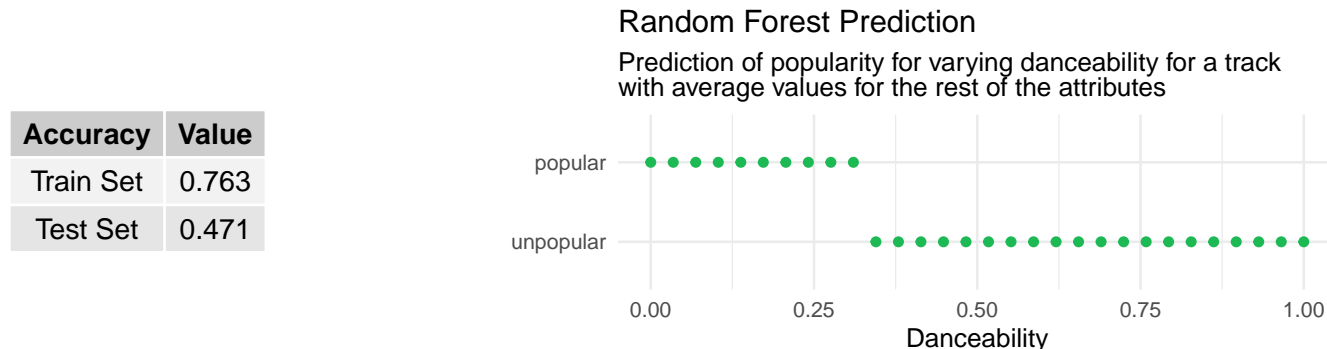
The data reflects this tendency with a downwards trend in more popular songs.

3.3 Modelling

It may be a natural question to ask if it is possible to predict if a song will be popular based on its characteristics. Four different models will be investigated.

3.3.1 Random Forest

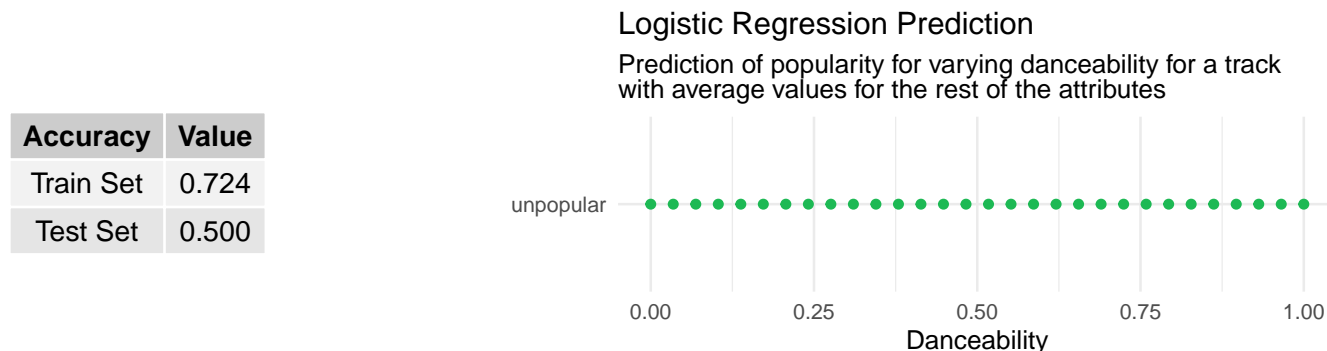
A random forest classifier uses a multitude of decision trees to come up with a final prediction. It will be considered first due to its flexibility and the lack of underlying assumptions. It does not require independence of the variables allowing for each song to be considered separately. It needs to be mentioned that the limitation of this approach is the unavailability of popularity data for specific songs.



Analysis: The model has a lower accuracy than a random guess on test data with a high accuracy on the train part. This may be a result of over fitting, a typical drawback of decision trees models.

3.3.2 Logistic Regression

Logistic regression is used to predict binary classification (here popular or unpopular). It assigns a probability of an observation belonging to a certain category. One of the assumptions is independence of observations, so it requires us to consider average data for albums instead of separate songs. A major limitation of this approach is the inability to distinguish between varied albums (e.g. where a value ranges between 0 and 1, averaging at 0.5) and more consistent albums (e.g. where all tracks are close to the average of 0.5). What is more, while it is vaguely reasonable to ask an artist to write a song with acousticness of 0.5, it may be impossible to demand a whole album with such an average value. Keeping that in mind, let us consider the accuracy of the model.

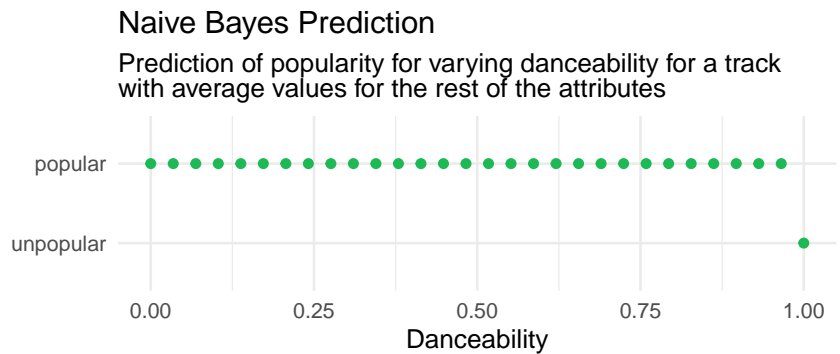


Analysis: The accuracy of the model on the test data is equal to that of a random guess with high accuracy on train data.

3.3.3 Naive Bayes

Naive Bayes Classifier is a probabilistic classifier requiring strong independence based on Bayes' theorem. Because of the assumption of independence, we will use albums' averages instead of songs once more.

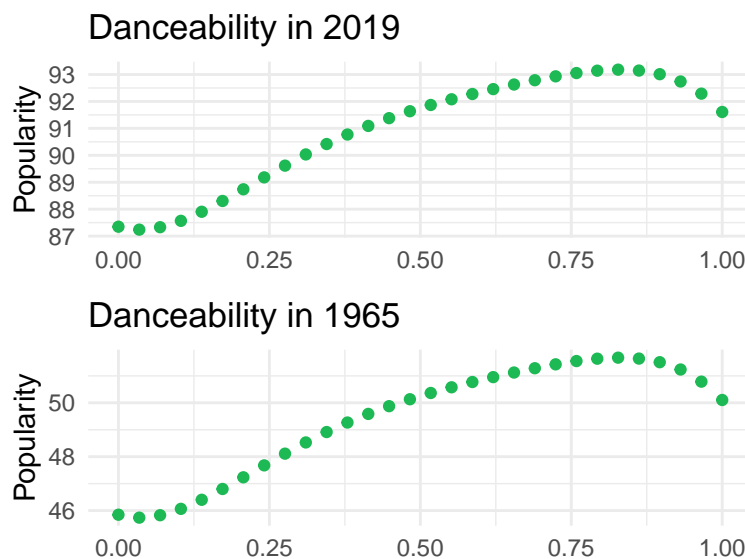
Accuracy	Value
Train Set	0.673
Test Set	0.500



Analysis: The accuracy on the test set is equal to that of a random guess with a higher one for the train set. Moreover, the graph shows the unrealistic nature of the models. Rationally, we know that albums with extreme values of those attributes are unlikely to be successful (e.g. here a whole album with a danceability of 0), but the models do not seem to notice that based on the data. This is a sign of a data limitation with the lack of representation of unusual albums with extreme values.

3.3.4 Polynomial Regression

Typical linear regression is not considered in this report for the reason that it is not sensible to try to model 60 years of trends in music with a straight line. A polynomial of a sufficient degree would be more suitable for describing the ever changing fashion. Nevertheless, it is incapable of noticing different tendencies in data based on the year a song is released.



3.4 Conclusions

Four different modelling approaches were analysed leading to a conclusion that it may be an impossible task in the present day. There were limitations stemming from the qualities of the data set, but it may not be the case that better data would result in a useful model. After all, making such a model would be incredibly lucrative for music companies and most definitely was attempted numerous times on much larger data sets by larger teams with better resources without any success. After all, two songs could have exactly the same Spotify characteristics with one being a hit and the other a failure. The current technology cannot quantify all the features that make a work of art. Creating a relatively successful prediction model for hit songs would mean getting really close to creating art with AI, a task still not reachable with the current state of technology.

4 Track recommendations

Before beginning to tackle this task, we decided to conduct a little background research into how successful platforms conduct these recommendations. The obvious choice was the origin of our data set, Spotify, which is well renowned for its high-quality recommendation algorithms. After looking into Spotify's approach, it was clear that they tackled the problem in three ways: -

1. Collaborative Filtering models, which analyse both your behaviour and others' behaviours. It picks up the trends present (for example diversity and self esteem music).
2. Natural Language Processing (NLP) models, which analyse text to identify what type of context the listener enjoys.
3. Audio models, which analyse the raw audio tracks themselves and fundamentally understand the similarities between songs in their listening history and recommend tracks based on this. (Ciocca 2017)

Unfortunately, the first two approaches were inaccessible to us for this project as we didn't have the necessary data available. However, the third method of using audio models seemed like a great starting point as it would allow us to make use of the numerical data in columns 17 through 28 of the edited Spotify database provided.

Although, as a group we were content with using the audio characteristics of the tracks to make recommendations, we felt that it may not be sufficient with regards to utilising as much data as possible. To be more specific, we also wanted to incorporate the genre as a factor as we felt that it could play an important role in people's taste in music.

Initially we considered that the genre of a track may be sufficiently represented within the audio data itself however we decided that we didn't see enough evidence for this in the data. This is likely due to how Spotify assigns genres to albums which, as mentioned above, is not the most scientific approach. Therefore, we decided that we would split our recommendation process into two separate approaches, which were the following: -

1. Recommendation based on similarities in the audio data alone.
2. Recommendation based on similarities in genre.

4.0.1 Approach 1

For approach 1 our method was simple but effective. In order to find songs that had similar audio data we used an R function that takes the listener's track as the input and finds the 'k' nearest neighbours to that track in the edited Spotify database. It finds the nearest neighbours by calculating the Euclidean distance between the audio data values of the input track and the audio data values of the edited Spotify tracks and then returns the tracks that correspond to the 'k' smallest distances. Which, in theory at least, are the 'k' most similar tracks audibly (see technical appendix for R code).

4.0.2 Approach 2

The second approach was to consider the genres and recommend songs based on those genres. We advanced by first scanning the genres of the chosen album and then picking one of the genres in that register. We then further scanned the rest of the data file to identify other albums with the same genre and output a sample of those albums.

We decided that an appropriate number of tracks to recommend would be 10, 5 from each approach to increase the diversity of music. We decided on 10 as we felt that it was a good balance between the more tracks that we recommend the higher the likelihood of suggesting a track that the listener would enjoy, and recommending so many tracks that the similarities between the suggested tracks and the input tracks are seemingly non-existent.

4.1 Testing

When reviewing our function we found it hard to tell whether it was performing well or not, as there is no sufficient way to quantify similarity between different tracks or 'how good a recommendation was'. We therefore had to take a less scientific and more subjective approach which involved us using our function on tracks from the test set and using our experience and general knowledge of the music industry to 'sanity check' whether the recommendations

given seemed sensible. After doing this we were, for the most part satisfied with the results. The only issue was that the recommendations were quite categorical, especially for recommendations using approach 1. Meaning that the recommendations for a track were usually by the same or similar artist at a similar place in time. This may not necessarily be a bad thing as it is reasonable to assume that if a person likes a song of a certain type from 1970s say, then they may also enjoy a song of a similar type from the 1970s. However, we felt that a slightly more diverse set of recommendations would be preferred as some may prefer listening to a great variety of different music whilst our function recommends songs that are similar to the song tested.

4.2 Conclusion

As stated above, for the most part we were content with the performance of the recommendation function that we constructed, however, there were a few limitations. The first of which is the one mentioned in the testing section above. We felt that this was likely due to the way the data was constructed, being just a ‘snapshot’ from different decades it was not sufficiently representative of the whole music industry. A more comprehensive database would likely go a long way to elevate this problem. Another factor that may have contributed to this is down to the way that genre was assigned as we felt that a lot of the genres were very specific which resulted in them being quite niche. This meant that there may only be very few albums with certain genres, which would drastically reduce the sub-set of songs from which the genre approach could recommend. Likely resulting in it recommending tracks from the same or similar artists. Another limitation was due to the fact that genre was assigned to the albums and not the tracks themselves. If instead we had the track genres then it is likely that the genre approach would make more specific recommendations. At first we thought this may not be too much of a limitation as tracks on the same album were likely to be similar anyway, however, in practice this wasn’t always the case so we concluded that it would be more suitable to have individual track genres.

Looking forward, there is obviously a reason why the big platforms such as Spotify use an amalgamation of multiple approaches to make recommendations. This is likely because of the diverse and undefined nature of music taste. With this in mind we believe that the best next step to improve our current recommendation process would be to include another model approach. The one we had in mind would be a collaborative filtering method (defined above) as we think this single approach would add the biggest improvement based on how our process currently operates. However, implementing such a process would require significantly more (and more complicated) data such as the listening history of many individuals including the user that the recommendations are for. We appreciated that collecting such data and implementing this system would be very resource intensive but all things considered we believe it to be worth while.

References

- Ciocca, Sophia. 2017. “How Does Spotify Know You so Well?”
- Halliday, Josh. 2010. “LimeWire Shut down by Federal Court.”
- LastMinuteMusicians. 2018. “What’s Popular? - Acoustic Bands.”
- Leight, Elias. 2017. “Producers, Songwriters on How Pop Songs Got so Slow.”
- McDonald, Glenn. 2013. “How We Understand Music Genres.”
- Spotify. 2019a. “Audio Features & Analysis.”
- . 2019b. “Get Audio Features for a Track.”
- Wang, Amy X. 2016. “Recorded Music Is in Trouble, but Live Concerts Are Making More Money Than Ever.”
- . 2019.