

+Masters Data in Business

Assignment 1

Spotify - World Music Tracks & Characteristics

Group 2 : Gabrielle Shamu, Omone Akha, Georgia Sapsani

Task 1: Real-World Problem

Identifying a real-world problem you are looking to solve with Data Analysis

A current and real-world problem lies with music apps in modern-day society. These apps have an overwhelming amount of choice regarding artists, genres, popularity, etc. This overflow of music and recommendations makes it harder for users to discover new music they may like. Due to features such as “Top 100”, and “Top Picks” (which showcase current and popular music), users end up only listening to a fraction of available songs, relying on well-known tracks. This may limit their enjoyability/ overall experience on such music apps due to the lack of tailoring to user preferences. There is also an issue with algorithms within music apps. Current recommendation algorithms recommend music based on user metrics (popular genres and artists) rather than refined preferences such as key, danceability, tempo, instrumentality, liveliness etc.

Justify why it is important to solve

Having a database rich in information about musical aspects such as danceability, energy, key, speechiness, acousticness, tempo, valence, etc could assist in igniting a more nuanced algorithm/recommendation engine that will tailor the music to user preferences.

Who benefits from the solution?

This sort of information will provide a more personalised experience for users, which will aid in bringing more users to the music app and spending more time on the app (due to the tailored experience). This allows users to discover new music that they wouldn't have normally come across. This, in turn, benefits the user, as well as stakeholders, as it will increase the user population as well as produce user loyalty for existing users, leading to an increase in revenue for the music app. Having such an intricate database can also aid in promoting new and upcoming artists as well as lesser-known artists whose songs match a user's nuanced preferences, boosting engagement for both the artists and the music app itself, which again brings more revenue, benefiting stakeholders.

Task 2: Database - WorldHits

Source: [300 World Music Tracks \(with Spotify Data\) \(kaggle.com\)](#)

The source above includes variables important for music tracks such as Duration, Time_Signature, Danceability, Energy, Key, Loudness, Mode, Speechiness, Acousticness, Instrumentalness, Liveness, Valence, Tempo, and Popularity. These variables give us an insight into how specific characteristics in a music track make it more or less appealing to Spotify's audience.

Task 3: Exploratory Data Analysis (EDA)

Correlations

The following *correlations* show the relationship between the variables of our database. Although we didn't identify any strong significant correlations, there is a positive association between loudness and danceability which shows that the louder a song, the more suitable it is for dancing.

Also, there is a negative relationship between acousticness and danceability, which shows that songs that solely or primarily use instruments, are less suitable for dancing. There is also a negative relationship between instrumentalness and danceability. This shows that music without any vocals is less ideal for dancing. These two relationships show that the most popular music possibly has vocals and instruments playing simultaneously.

The negative association between liveness and danceability shows that the more the presence of an audience in the recording, the less suitable it is for dancing. It possibly has to do with the clarity of the song as any additional noise in the track makes it less attractive for people to dance.

Moving to popularity, the correlations are not strong but songs that last long (duration), have increased liveness and more spoken words (speechiness) are less popular. On the other hand, songs that are more suitable for dancing and are higher in valence, tend to be more popular. Valence in Spotify¹ describes the musical positiveness conveyed by a track. For example, tracks with high valence sound more positive (happy, cheerful), while tracks with low valence sound more negative (sad, depressed, angry).

	Correlations
Duration vs Time_Signature	0

Duration vs Danceability	-0.04
Duration vs Energy	-0.01
Loudness vs Danceability	0.24
Acousticness vs Danceability	-0.23
Instrumentalness vs Danceability	-0.21
Liveness vs Danceability	-0.16
Tempo vs Danceability	0.01
Popularity vs Duration	-0.13
Popularity vs Danceability	0.16
Popularity vs Liveness	-0.14
Popularity vs Tempo	-0.02
Popularity vs Valence	0.13
Popularity vs Instrumentalness	-0.09
Popularity vs Acousticness	0.03
Popularity vs Speechiness	-0.13
Popularity vs Mode	-0.05
Popularity vs Energy	-0.01
Popularity vs Key	0.06

¹<https://community.spotify.com/t5/Spotify-for-Developers/Valence-as-a-measure-of-happiness/td-p/4385221>

Visualisations

A clear challenge when analysing this dataset is the need for more diversity within the data. As shown in Chart 1 below, most of the tracks were released between 2000 and 2009. Very few tracks were released between 2010-2019 and 1950-1959. This skews other results when analysing this data, as shown in Chart 2.

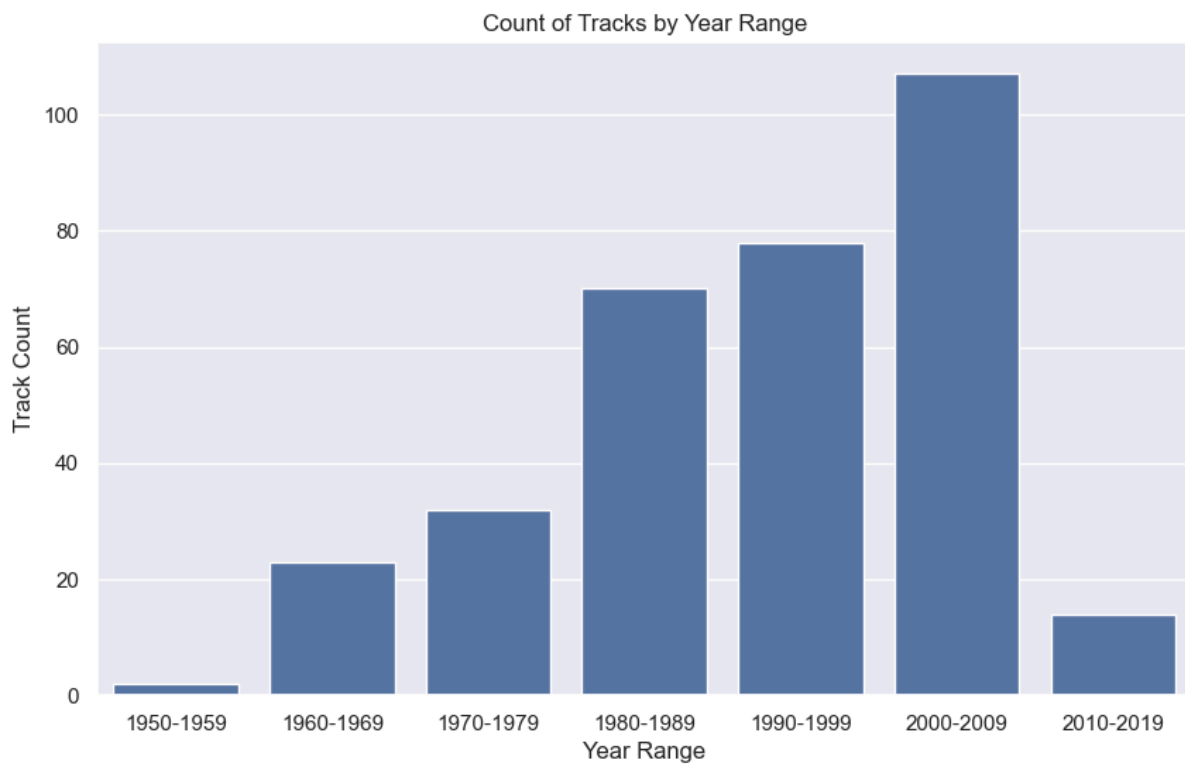


Chart 1

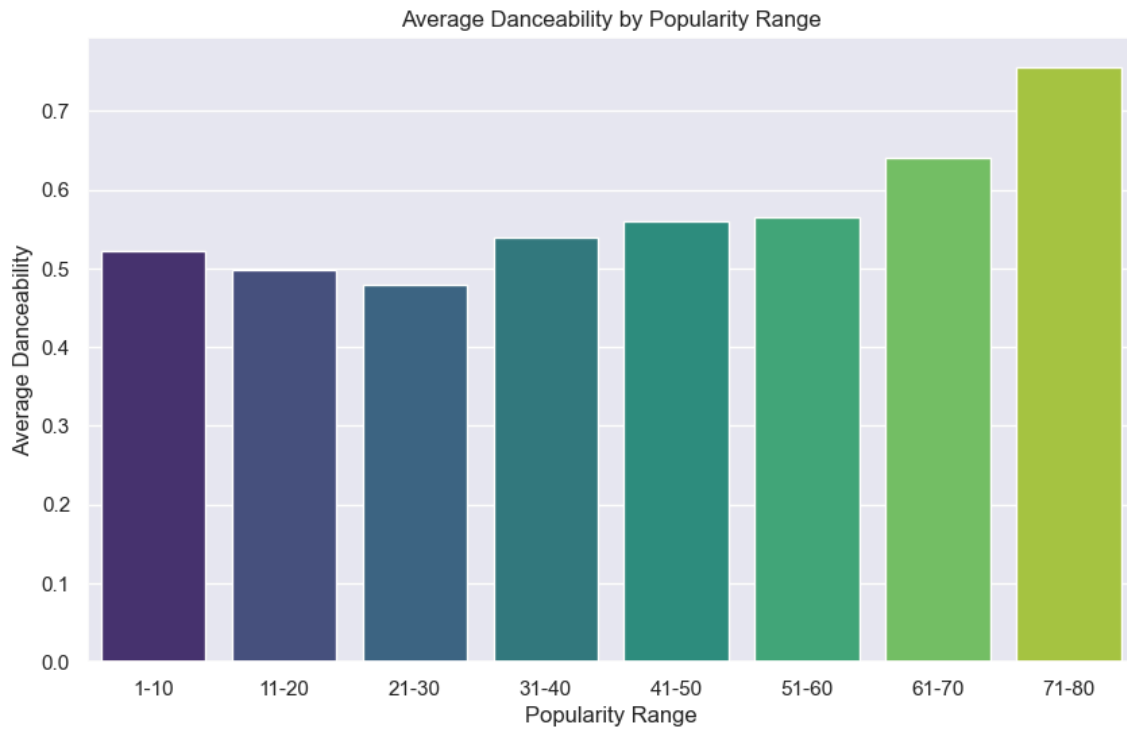


Chart 2

In bar chart 2, we can see that songs with popularity 71-80, are more suitable for dancing on average.

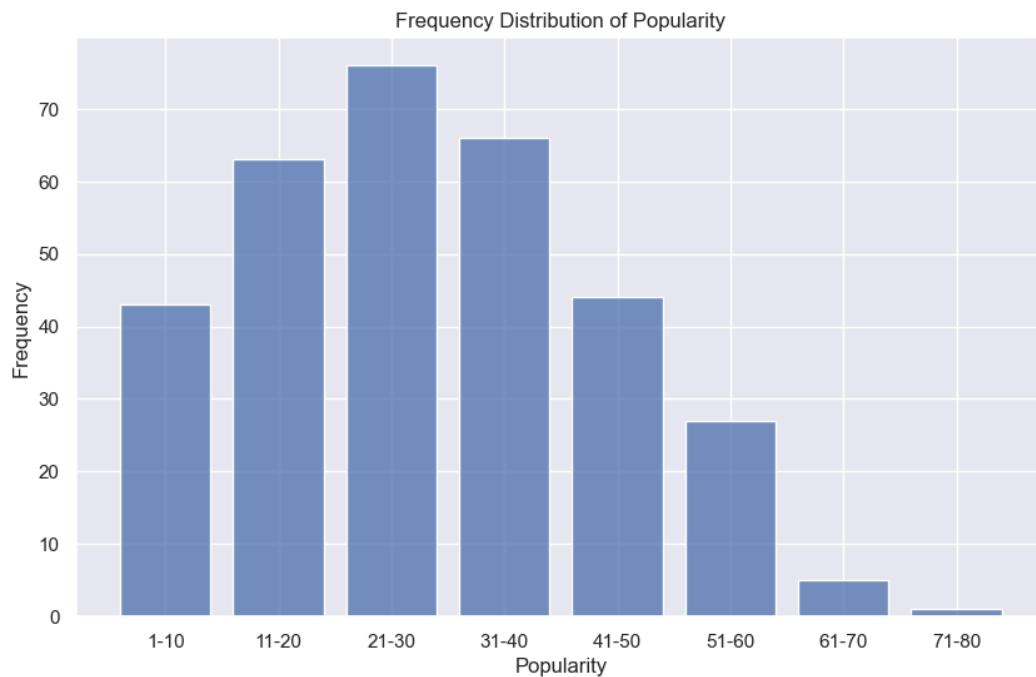


Chart 3

Most of the songs reached popularity between 21 and 30, as shown in Chart 3. Manu Chao is the artist with the highest popularity, with the track Bongo Bong from the album Clandestino in 1998.

Task 4: Data Cleaning Process

During data cleaning: no duplicates, no missing data, outliers and no white spaces were found. The names of the columns were modified for “Duration”, “Time_Signature”, “Danceability”, “Energy”, “Key”, and “Loudness” as shown in the ‘WorldHits_DataCleaning’ Excel file so that the context for these variables is clearer and easier to understand.

Also, we froze the first row that includes the names of the columns so that we can identify the data more easily while scrolling up or down.