Multi-faceted Analysis and Classification of Spotify Tracks

A project report by Anu Bazarragchaa

Topic: Multi-faceted Analysis and Classification of Spotify Tracks
You can visit the link to the Colab here: Spotify.ipynb (Includes interactive plots)
Dataset: https://huggingface.co/datasets/maharshipandya/spotify-tracks-dataset
Submitted: 2023/10/30
NOTE: In order to explore the dataset, it's also useful to read the documentation available with the current dataset.

Tasks I & II: Data Exploration and Feature Engineering
In the Data Exploration and Feature Engineering tasks, I began by importing and examining the dataset. I found that each genre contained exactly 1000 songs, which was a happy surprise, and indicated a well balanced dataset. In most cases, we expect a normal distribution, but here, I noticed an unexpected surplus of less popular songs and genres.

I started exploring the data quickly afterwards. I found that there was a concentration of popularity scores in the 20-30 range, which could provide insights into the composition of online streaming platforms' music libraries. Then I found the danceability of tracks through a histogram. It revealed that most songs had a *relatively high* danceability value, with a peak around <u>0.6</u>. I also examined the energy of the tracks, which exhibited a right-skewed distribution, indicating the prevalence of high-energy songs in the dataset. This could impact various musical attributes like tempo and time signature.I noticed a slight left-skewed distribution, suggesting that the dataset contained relatively more moody songs compared to happier ones. I also looked into the duration of tracks, showing a concentration of songs in the 3-4 minute range. After I computed statistical measures like the average, median, and mode to quantify this observation, the data showed that the majority of tracks in the dataset were non-explicit, likely to be radio-friendly. In the music industry, non-explicit songs or versions of popular songs are made for younger listeners, and this might be a factor.

I also explored the tempo distribution, identifying peaks at 100 and 140 beats per minute, with a smaller peak around 160-170 BPM. In my favorite genre, Techno music, the average tempo was around 112.64 BPM. Finally, I looked into time signatures, discovering that the 4/4 time signature was the most common, followed by 3/4, with occasional occurrences of 5/4 and 1/4 time signatures. Most scores are written in 4/4 and most modern songs are too. To improve data quality, I normalized artist names and focused on the primary artist when multiple artists were listed. The 'explicit' column was also converted into a binary format, streamlining our data for further analysis and modeling. This was done by the guidelines given in the task pamphlets.
Task III. Data Visualization
I used a specific genre to see how different it would look from the general data. I created a general aggregated interactive plot using plotly.js, and afterwards I visualized specific genres(e.g.: “Techno”), to see how different it would look. There were some differences in popularity, valence and loudness etc., but the tempo and time signature attributes were very similar all around.
I then visualized a correlation matrix between all the classes in order to find highly similar classes, and found that loudness was very close to energy. I then took out

“loudness” on the grounds that high correlation would be corr>0.5; and I did the visualization again. I used PCA to find two important patterns in the data based on the class features. The first pattern (PC1) is mainly linked to features like song length, danceability, energy, mood, and tempo. It shows a negative connection with danceability, energy, and mood. The second pattern (PC2) is associated with song length, energy, and tempo, and it has positive links with energy and tempo. These patterns help us simplify the data and understand key relationships among the class features. This is more helpful later on, as I build recommendation algorithm prototypes and I will need classes that matter more.

Task IV. Recommendation Algorithm

My goal here was to create a song recommendation system. In this step, I wanted to find songs that are similar to the infamous Jason Mraz's "I'm Yours" based on how they sound. First, I found the unique code for "I'm Yours" using a short script. Then, I used a method called cosine similarity to compare its sound to other songs. This method gave us a list of 10 songs that sounded most like "I'm Yours." It’s a very cookie-cutter song that scores high in valence, tempo and has an acousticness and a cheerful vibe. After creating the recommendation algorithm as specifically noted in the docket, the algorithm said the closest song to Jason Mraz's I'm Yours is this song: Old Crow Medicine Show - Wagon Wheel (give it a listen :D). After playing around with the features, I concluded that this is a promising result – the banjo and 4/4 time signature matches Jason Mraz's guitar strut and happy-go-lucky song well. It’s important to also note that cosine similarity also has its downfalls. One thing I found was that it increased in similarity when I used as many classes as possible, so it heavily depends on scaling audio-specific features.

Task V. Classification Based on Audio Features

In this task, my objective was to categorize songs into various genres based on their audio attributes. I adopted several machine learning algorithms, including Random Forest, Gradient Boosting, Feedforward Neural Network, and Multi-Layer Perceptron to accomplish this task.

Random Forest achieved an accuracy rate of **64.7%** when classifying the test data, with precision, recall, and F1-scores reported for each genre. Notably, it demonstrated strengths in accurately recognizing genres like acoustic and afrobeat genres. However, it faced challenges in classifying genres like alt-rock and alternative tracks, highlighting specific areas where its classification abilities excelled or fell short. My conclusion here would be that it’s very hard to categorize more experimental genres than genres with more structure, such as pop and rock. Meanwhile, Gradient Boosting had an accuracy rate of **66.7%**, and displayed an improved overall performance compared to Random Forest. It excelled in identifying genres like acoustic and afrobeat genres, showing its ability. However, like Random Forest, it encountered difficulties in accurately categorizing genres like alt-rock and alternative tracks.

The FNN method outperformed the others with the highest accuracy of **68.8%**. It provided detailed precision, recall, and F1-scores for each genre. The FNN particularly excelled in recognizing afrobeat and ambient genres. This is probably due to the fact that afrobeat has a very distinctive set of drums, like drum n' bass, and ambient songs usually have similar valence scores. The MLP achieved an accuracy of **60.5%**, which was slightly lower than the other methods. It provided precision, recall, and F1-scores for each genre that were satisfactory, although it demonstrated a lower overall performance compared to the other methods. The MLP faced challenges in accurately recognizing alt-rock and alternative songs. The Feedforward Neural Network (FNN) emerged as the most accurate method, achieving an accuracy of **68.8%**. Gradient Boosting also performed well ,Random Forest coming in third and Multi-Layer Perceptron (MLP) at the end of the roster.

Task VI: Analysis of the 'popularity' variable

In this task I aimed to find out the relationship between audio features and track popularity using regression models. The question in sight: What makes a track popular?

In order to find what makes a track popular, we need to build an accurate regression model that predicts popularity. I first tried a method called linear regression. The mean absolute error was 18.34, and the mean squared error was 482.38. These numbers help us understand how well the model works. But the most important number is the R-squared score, and it was only 0.0225, which is pretty low. The model didn't do particularly well.

Then, I thought maybe a different method could work better, so I tried polynomial regression. It improved the R-squared score a bit to 0.0554, but it was still not great. After that, I attempted logistic regression, which gave an accuracy score of 0.1406. This score is also quite low. Then I decided to test a decision tree regression model. It did better than the others with a moderate R-squared score of 0.1303. While it was an improvement, it still does not model the variable appropriately.

The analysis revealed that none of them fit the bill for explaining popularity. The linear and polynomial regression models had low R-squared scores, indicating weak relationships, while logistic regression had a low accuracy score. Although the decision tree regression model performed somewhat better, it still produced a moderate R-squared score. Overall, the analysis suggests that audio features alone may not be enough to predict track popularity in a reliable manner, and that other factors or better models (or better programmers!) were needed for this task.

Task VII. Genre-based analysis

I examined the average values of key audio features for each genre in this genre-based analysis, providing insights into genres that stand out in terms of danceability, energy, and valence. Kids' music had the highest average danceability score among the genres studied, indicating its suitability for dancing. This was closely followed by chicago-house, reggaeton, latino, and reggae, all of which demonstrated similarly high levels of danceability, indicating their energetic and dance-friendly nature.

In terms of energy, death-metal emerged as the most energetic genre (unsurprisingly!), with high-intensity and activity in the music. Grindcore, metalcore, happy, and hardstyle were also among the most energetic, owing to their vibrant and dynamic musical styles. Salsa, forró, rockabilly, afrobeat, and ska are examples of genres that convey positive musical emotions and high valence scores. Because these genres are known for evoking feelings of happiness and positivity, they are popular choices for upbeat and cheerful music. The genre with the highest energy level was identified as death-metal, which is distinguished by its intense and powerful musical composition. Other energizing genres included grindcore, metalcore, happy, and hardstyle, all of which exhibited vibrant and lively musical qualities. The kids genre had the highest danceability, making it ideal for movement and dancing. Genres such as chicago-house, reggaeton, latino, and reggae were also among the most danceable, indicating their suitability for grooving and rhythm. The genres with the highest valence scores, indicating positive emotional content, are salsa, forró, rockabilly, afrobeat, and ska. These genres are known for their ability to evoke positive emotions in the listener.

Task VIII. Conclusion

During the course of the project I aimed to classify, analyze and eventually build a recommendation model and saw that in the initial stages, the genres contained each a 1000 songs in total, creating a great balance for further analysis. This structural distinction was combined with a concentration of popularity scores in the 20-30 range, providing insights into the composition of streaming music libraries. Key audio features such as danceability, energy, valence, duration, tempo, explicit content, and time signatures were examined to reveal the dataset's musical characteristics and demographic appeal.

I used data visualization methods to delve into genre-specific trends, which brought out distinctions in attributes like popularity, valence, and loudness. Tempo and time signature remained quite consistent across different genres. I also discovered that the Feedforward Neural Network (FNN) proved to be the most accurate classification model, achieving a *commendable* 68.8% accuracy, particularly excelling in identifying afrobeat and ambient genres. These results suggest possibilities for future research where we could focus on improving track popularity prediction models and as a consequence, helping Spotify and other platforms make better recommendation algorithms. These could hold the potential to help better music recommendation systems that are already in place, and help listeners discover new music.