Gisma University
of Applied Sciences

Assessment Submission Form

Student Number (If this is group work, please include the student numbers of all group participants)	GH1023448
Assessment Title	Individual Final Project
Module Code	B106R
Module Title	B106R Data Visualisation (SS0324 Retake)
Module Tutor	Mehran Monavari
Date Submitted	11.07.2024

Declaration of Authorship

I declare that all material in this assessment is my own work except where there is clear acknowledgement and appropriate reference to the work of others.

I fully understand that the unacknowledged inclusion of another person's writings or ideas or works in this work may be considered plagiarism and that, should a formal investigation process confirms the allegation, I would be subject to the penalties associated with plagiarism, as per GISMA Business School, University of Applied Sciences' regulations for academic misconduct.

Signed *Stefan-Bogdan Zbyrko* Date 11.07.2024

Github Link:
https://github.com/LvivGuyOrSmth/Data-Visualisation.git

About Dataset

This dataset presents a comprehensive compilation of the most streamed songs on Spotify in 2023. It provides extensive insights into each track's attributes, popularity, and presence on various music platforms, offering a valuable resource for music analysts, enthusiasts, and industry professionals. The dataset includes information such as track name, artist, release date, ISRC, streaming statistics, and presence on platforms like YouTube, TikTok, and more.

Research Question

What musical characteristics had the greatest impact on the streaming popularity of songs released on Spotify in 2024?

About Study

The study appeals to a range of stakeholders in the music industry, including artists, record labels, producers, music promoters, and streaming platforms such as playlist curators. It aims to provide insights into music consumption patterns and cultural influences, benefiting those involved in music creation and promotion. The analysis not only addresses commercial interests but also enhances comprehension of current music trends and listener preferences.

Goal:

My objectives are to pinpoint critical factors influencing streaming success and to gain a deeper understanding of Spotify listener preferences, as I aspire to pursue a career there in the future.

Outline

- Data Extraction: Gathering and compiling relevant data from the dataset.
- Data Cleaning: Removing duplicates, handling missing values, and filtering out irrelevant data to ensure dataset integrity.
- General Information: Providing an overview and summary of key aspects within the dataset.
- Correlation Heatmap: Visualizing correlations between variables to understand relationships and patterns.

- Scatter Plot with Key as Hue: Creating a graphical representation to explore the influence of musical keys on streaming success.
- Exploring Feature Impact on Streaming Success: Investigating how various musical features affect the popularity of songs on Spotify.
- Feature Importance Analysis with Random Forests: Assessing the importance of different features using a random forest model.
- Conclusion: Summarizing findings, implications, and potential avenues for future research based on the analysis.

Conclusion

In the analysis, we implemented the use of scatter plots, rankings of top songs and artists, and tried to compare different features which I found to be most crucial with the hope of answering the question. In truth, the data was hard to analyze and make much sense of. To the best of my understanding, valence, danceability, liveness, and energy are the most influential features according to the Random Forest model which was used towards the end. On the other hand, instrumentals seemed to have the lowest impact on streaming popularity. I would conclude that it is too difficult to determine which characteristics have a greater impact on song popularity with the given data, considering popularity divides into many branches, one of which is also the complexity of human's musical preferences.

Further Considerations:

Even being this detailed and well-documented, this dataset does not account for other factors such as marketing, the artist's popularity at the time, trends and so on. It is safe to assume that these factors also play a crucial role in influencing streaming numbers. Also, understanding how factors like danceability, energy and valence are measured is puzzling, therefore limiting the understanding of the data at hand.

Giriyewithana, N. (2023). *Top Spotify Songs 2023*. Kaggle. Available at: https://www.kaggle.com/datasets/nelgiriyewithana/top-spotify-songs-2023 (Accessed: 11 July 2024).