Diary Entry

Ho Wei Ni

2023-10-25

Week 9

What is the topic that you have finalized?

Answer: My finalized topic is on music. I plan to present a data story on the popularity of songs in different regions of the world based on the song attributes (e.g. danceability / energy / key / mode / loudness / speechiness / acousticness / instrumentalness / liveness / valence / tempo), genres and artists etc.

What are the data sources that you have curated so far?

Answer: I will be using this Spotify Dataset found on Kaggle. Since the data set is rather large, I will be using Radiant to filter data extracted during 2022 only. I will also be removing some variables e.g. collab / source / previous rank / pivot, that I have deemed irrelevant to my data story.

Week 10

What is the question that you are going to answer?

Answer: What factors shape the popularity of songs in today's diverse music landscape?

Why is this an important question?

Answer: According Forbes, every society appears to have some form of music, a type of communication that is often overlooked, as part of their culture. From a business perspective, studying the factors behind song popularity aids artists and marketers in understanding audience preferences, which is crucial for creating content that resonates. From the CNM perspective, popular songs mirror cultural sentiments, offering insights into collective mood, social trends, and cultural shifts.

Which rows and columns of the dataset will be used to answer this question?

Answer: My Week 9 Diary Entry mentioned that I am filtering the data to show 2022 results only. However, I have since discovered that the file size is still too large so I will be further filtering the data to show **January 2022** results only.

(Note: The terms 'song' and 'track' are used interchangeably in the following paragraph.)

To begin, I want readers to understand the general music landscape of January 2022. This can be done by listing the tracks and artists listened to in January 2022 in terms of descending order of popularity.

The overall popularity of a track will be determined by calculating the total streams of each track, using the ***streams*** column, taking care to group by ***track_name***. Similarly, the popularity of an artist will be determined by calculating their total streams, taking care to group by ***artist_individual***. Additionally, I will be showing the general distribution of song attributes: ***danceability***, ***energy***, ***key***, ***mode***, ***loudness***, ***speechiness***, ***acousticness***, ***instrumentalness***, ***liveness***, ***valence***, ***tempo***, ***duration***. This indicates the type of songs currently being produced.

To understand the factors that shape the popularity of a song, I will investigate how each song attribute correlate with total streams and chart rankings. Total streams implies the long-term popularity of a song, identifying which attributes are more "timeless" whereas chart rankings implies the short-term popularity of a song that could be of a trend-specific nature. For the latter, ***peak_ranking***, ***previous_week*** and ***weeks_on_chart*** will be used. This analysis will be further refined using ***country*** and ***language*** to compare global and local trends in hopes of revealing similarities and differences.