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.

Challenges & Errors Faced

Embed Shiny into Quarto

Prior to Professor uploading the instructions on how to embed Shiny into Quarto, I was facing difficulties making my Shiny app interactive in the qmd file. Initially, I simply pasted the code in my qmd file, but the output was an image of my Shiny app that is non-interactive.

Thanks to this discussion, I found out that Quarto is a static website and will only run the html files generated, and hence will not run the R code. After trying various methods, I eventually found this tutorial that allows me to embed multiple Shiny apps into my qmd file directly.

I have yet to try the Shinyapps.io method provided by Professor, but have noticed that I will be limited to 5 apps only. While I understand that I can combine my apps such that it appears as a carousel, I am unsure if doing so will disrupt the flow of my data story. I will decide on the method to use after finalising the structure and flow of my story.

Changing the hovertext in my Plotly graph

When playing around with the data, I was trying to display a horizontal bar chart showing the top 10 tracks in January based on their total streams. However, the default hovertext is based on the x and y variable names, making it look messy as there were some redundant information

I came across the tooltip method suggested by cpsievert in this discussion, where I use the text aesthetic specify what information I want to be displayed, before supplying the tooltip text as a character vector, then the tooltip argument in ggplotly(). This will definitely be helpful in making my interactive visualisations neat and hence effective.

Plot distribution of selected attributes on the same histogram based on checkbox selection

Since there are so many song attributes, I wanted to plot them in the same histogram as I thought it looked nice, added to the interactive portion and makes for easier comparison. However, I had no idea how to go about linking the input selection to the data shown.

After lots of research, I found this discussion where the user was attempting to do something similar. Taking inspiration from this line `filter(value %in% as.vector(input$picker)`, I filtered my data set based on the inputs selected and it seems to be working as of now.

However, I am now facing difficulties making the histogram show proportion instead of frequency of count. Showing proportion would make for a fairer comparison as the frequency range for each attribute vary greatly. I will be exploring the solutions to this challenge next week.