Diary Entry Week 11

Chua Jieh Yih Audra

2023-11-03

Week 9

Finalised topic (Changed)

I originally wanted to analyse the ratings for Animal Crossing by critics and users, and its sales in relation to that. However, upon receiving feedback on my dataset I have decided to change my topic.

My new topic is: what factors affect the number of streams a song gets on Spotify?

Data sources I have curated so far

Here is my new data source: https://www.kaggle.com/datasets/salvatorerastelli/spotify-and-youtube

Week 10

What is the question that I am going to answer?

What factors affect the number of streams a song gets on Spotify?

Why is this an important question?

According to the Bulgarian Comparative Education Society (BCES), 2021, music is present everywhere around us, and acts as a medium which shapes our environment, self-perception, and interpersonal interactions, thus leaving a large impact on society. Source: https://files.eric.ed.gov/fulltext/ED614071.pdf

According to Business of Apps statistics for 2023, Spotify is the most popular music streaming platform in the world, and it has over 350 million users and 150 million subscribers. Source: https://www.businessofapps.com/data/music-streaming-market/

Finally, according to Statista, in 2022, total revenue generated by the recorded music industry was 26.2 billion U.S. dollars, of which the majority of revenue came from streaming services, showing the economic significance they have. Source: https://www.statista.com/chart/4713/global-recorded-music-industry-revenues/

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

Rows: all Columns: Stream, Danceability, Energy, Speechiness, Instrumentalness, Liveness, Valence, Tempo, Duration_ms, Views, Likes, Title, Description, Licensed, official_video

Upon opening my dataset in Excel, it gave me a message letting me know that some numbers had gotten stored as text. Excel helped me to automate the process of fixing these and converting them back to numbers.

Week 11

Visualisations I will use in my project

I will use ggplot line graphs, whereby no. of Spotify streams would be the x axis, and my other variables would be the y axis.

How do I plan to make it interactive

I would like to let the users be able to view each specific data by hovering over the area they want to know the information of. This will allow them to view the peaks in data to learn about what they want to know. I would also like to create a plot that allows users to select more than one variable to view at once (So the x axis, no. of Spotify streams, would remain the same, but the y axis would be determined by the user) Most likely it seems I'll have to use a reactive() function and replace the variable name in ggplot's 'aes' section for y with a variable that changes depending on user input.

What concepts incorporated in your project were taught in the course and which ones were self-learnt?

Table of Concepts Used

Topic/ Concept	Weeks
Variables (numeric and categorical)	3
Manipulating data.	4
Functions (most likely for input)	5
Visualization with ggplot2	7
Exploratory data analyses	9
If/ Else	
Plotly	
Reactive() Expression/ Function	
Text analysis (str_detect() probably)	