**PROJECT 3 – Group 2 – Most Streamed Spotify Songs 2023**

**Topic**

*Project proposal:* Reviewing the most famous songs of 2023 as listed on Spotify. Using a CSV from <https://www.kaggle.com/datasets/nelgiriyewithana/top-spotify-songs-2023?resource=download> we are going to examine the data and evaluate the top 10-20 streams of 2023, then drill down to the released year(s) of the top streamed 2023 songs, after that we will consider if the song is in Spotify playlists and/or apple playlists and the corresponding artists. Once we had the dataset we had to scrub the data, and found a few errors in the data, we had to ensure that the data set was in the right format for CSV and SQL use.

**Dataset**

*Project proposal and submission*: Most Streamed Spotify Songs 2023, <https://www.kaggle.com/datasets/nelgiriyewithana/top-spotify-songs-2023?resource=download>, from link: This dataset contains a comprehensive list of the most famous songs of 2023 as listed on Spotify. The dataset offers a wealth of features beyond what is typically available in similar datasets. It provides insights into each song's attributes, popularity, and presence on various music platforms. The dataset includes information such as **track name, artist(s) name, release date, Spotify playlists and charts, streaming statistics, Apple Music presence, Deezer presence, Shazam charts, and various audio features**. A screenshot of the metadata: A screenshot of a computer

Description automatically generated

*Project submission only:* We used plotly express for our added/new package. This gave us the functionality to have hovers with our dataset information. The website we used was https://plotly.com/python/plotly-express/

**Project Inspiration & Initial Design**

*Project proposal and submission*: Three or four screenshots of relevant, “inspiring” visualizations that show your creative ideas (visualization track only)

A screenshot of a graph

Description automatically generated

Inspiration for “Top 10-25 streamed songs” with hover of track name, artist, and # of streams.

A screenshot of a computer

Description automatically generated A screenshot of a graph

Description automatically generated

Inspiration for release year vs number of streams

A graph of a bar chart

Description automatically generated with medium confidence

Inspiration for comparison of Spotify vs apple playlists

*Project proposal:* A webpage that will look similar or use this as a starting point and evolve from there. This was changed to a PowerPoint with interactive Jupyter Notebook charts, instead of a webpage.

A collage of images of charts and graphs

Description automatically generated

**GitHub Link**

*Project proposal only*: <https://github.com/Nalfrey/Project3.git>

**Ethical Considerations**

*Project submission only:* For ethical considerations we could argue that even though Spotify has Spotify Wrap that monitors the number of times you play a song, or the length in which you use their app, this was a summary of overall streams, and did not contain any individual or private data for the users.

**Repository Structure**

*Project submission only:* Our overall project is housed in a Jupyter Notebook (JN) and runs as a normal JN runs using a SQL database. The first 7 cells should be executed as a normal Jupyter Notebook. In our repository you will find a Project proposal which we used in our ideation phase of this project, a readme, an initial CSV that was used for the data, a jupyter notebook for the execution of analysis, the html for our hover design, and a total steams with the hover design.

For the Jupyter Notebook:

For the first visualization we showed the top 10 songs that were streamed for 2023, you must import plotly express, reset the indices, and then create the bar chart with the hover functionality.

For the second visualization we looked at the streams by song released year, we grouped the data by the year the song was released(released\_year) and then by the number of streams(streams). We made sure that the graph was more easily readable by showing only songs that were released after the year 2000.

For our last visualization we compared the dataset information on the number of times a song was found in a Spotify playlist and the number of times a song was found in an Apple playlist. We created an aggregation of the data columns “in\_spotify\_playlists” and “in\_apple\_playlists”. Those were then compared by platform and playlist count. These were presented in a pie chart.

Lastly for user interaction we created a user feedback box that allows the user to input their favorite artist to see if they had songs that were part of the most streamed songs of 2023. We created a function to get the list of songs by the artist, we also considered that if the artists name was in collaboration with another artist, it would pull as well. We then printed an if, else statement that if the artist name was not found, JN would print that there were no songs found for the artist. If songs were found then it would iterate through the rows and print all songs that that artist had in the most streamed songs of 2023, the corresponding number of streams, and what date the song was released. In the next cell we created the functionality to execute this dataset.