Python for Data Analytics

Final Project

Matthew De Nero 5/22/2022

The Beatles
"Dance Hits" by Album

For my final project I chose the Beatles because they're my favorite band and I thought it would be interesting to analyze every song by danceability and put the top dance songs into a new album called "The Beatles Dance Hits."

I found my data in Kaggle under user Chad Wambles called The Beatles Spotify Song Data. This data shows every song broken down by album, year, vocals, danceability, energy, speechless, acousticness, liveliness, and valence. I chose to work with song name, album, year, danceability, energy and vocals. Using this data my main goal was to create a new album that takes the top dance songs from every album and put them into one new dance album.

There are many ways to analyze this data and others calculated the number of songs by each of the albums. I also found some interesting data showing the relationship between energy and loudness (speechless) by song. But the most interesting I found was a regplot showing dependence between danceability and liveliness. I thought this was very interesting data and wondered what the top dance songs from each album would look like in their own album.

When writing my code I first started with something simple showing the release date of each album by year. I then analyzed every album by song and danceability and took the top dance song from each album and put that song into a new data frame that would become the Beatles new "Dance Hits" album. These songs ranged from a dance ability score of .6 (Yellow Submarine) to a .88 (for you blue). I then thought it would be interesting to see vocals and who sang the most songs either solo or together. John and Paul individually almost split the data in half with John singing solo slightly more at 37% with Paul singing solo 29%. It's interesting to see all four members only sang less than 1% of their songs together. Finally I wanted to see the danceability of every song and break them up into a pie chart showing light, light medium, medium, and high danceability scores. Less then 2 percent of all their songs have a light danceability score and almost half of their songs fell in the medium category with, the high category taking up almost 32% of their songs! No wonder they were so groovy!

I tired to combine all the data into a stem plot showing every song by danceability and the corresponding album by the color of the stem but I was unable to

demonstrate this in a legend. I kept it in the data anyways because it showed a nice bell curve that equated to my pie chart showing danceability by light, light medium, medium, and high. I hope to work with Bri later on to get this to work.

In the end, I was able to demonstrate the best dance songs from every album and plot them in their own data frame. This was a very fun and engaging project but I know I need to keep practicing and perfect my graphs especially the stem plot. I will certainly be creating a new Spotify playlist with these songs and take them to the next dance party!

Google Colab Link:
https://colab.research.google.com/drive/19FHedDPiGfBoY4lxo8PdmV3jAGtD-4ii#scrollTo=JjRoLAQWFJK_

References:

Data Source:
https://www.kaggle.com/datasets/chadwambles/allbeatlesspotifysongdata2009remaster/code

Data Analytics by other users:
https://www.kaggle.com/code/mpwolke/hello-goodbye-beatles-spotify