The Trends and Basis of Music Streaming Services.

Jonathan K. Farias (AB), Eivy Cedeno Torres (AF), Liliana Alvarado Garcia(AB)

Computer Science 163 Intermediate Data Programming

**Summary of research questions & method**

1. *What music streaming services have had the highest number of active users in the last 3 years?*
   1. Description: We are interested in investigating which streaming services have been the most popular in recent years. To do this we will need to compare the total users for each service per year and the average user per month for the last 3 years. When thinking about this problem we might ask, who is on the top of the streaming services? Who is showing the most growth? How big is the difference between these services? What affects these outcomes?
   2. Method : To get the active users for a streaming service we will count up the total number of subscribers or accounts and sort them by year in each streaming service. Doing this will allow us to create a graph and look at the trend over time, putting them in the same graph will help to analyze and observe clear comparisons. Since we are comparing the services, there will be data included with the graph such as peak users for that year, average users, and growth from previous years.

For this process we will be using the column “numberOfSubscribers” and add them up where the date is sorted and filtered by month. After we have a new dataframe with the active subs/users per month corresponding to date, we will be using ggplot or the .plot() function to graph a line for each of the music services. The x axis will be the subs/users and the y axis will be time in months for 3 years. We will be making each line a different color for each music service. We will also be adding a key and a title for visuals and ease of understanding.

1. *In general of songs being released by artists, how has the general duration of songs changed for every year?*
   1. Description: Overtime songs have changed in the way they present themselves to listeners in different genres. All songs made are different from each other, but one thing they somewhat share in similarity is the idea of how long a song plays for (duration). So, how has the duration of music changed over time? What might we infer from these changes?
   2. Method: Using visualizations such as different plots, we will be able to visually see the change in song durations for every year. We would be comparing the average song duration for every year (Y-axis) to its corresponding year (X-axis). We would need to find the average duration for every year by grouping the years, adding the durations for each song in each year and dividing by how many songs were in each year. This would result in the average duration for each year, then we would use this information to plot (plotly or seaborn) in a bar chart the average duration for each year between 2010 and 2019.
2. *How has the Covid-19 outbreak affected the time usage of music streaming services between years before and after the outbreak?*
   1. Description: We will investigate data about the average time usage of music streaming services such as Spotify, Pandora and Apple Music to make a comparison during years before and during the Covid-19 outbreak. This will allow us to infer the motive of usage trends before and after quarantine.
   2. Method: We will have to start by adding the datasets to our program. Then we will find the average time usage column of each streaming service by filtering the time usage columns to calculate its average per year. Then, we must use visualizations to demonstrate the change between the years of usage for each dataset. With that being said, the (y-axis) will be the time of usage and the (x-axis) would be the years 2017- 2021. After making a visualization for each, we will make a visualization comparing the three datasets to answer our research question. We will color code for each different music streaming service to be able to identify the service being compared. With that, there will be a legend connected to the color scheme being used to identify the streaming services.

**Motivation**

The reasoning behind our interest in the music trends is caused by our curiosity of the increased popularity of music streaming services during the Covid-19 pandemic. As students who have been affected by the Covid-19 quarantine, we have explored and experienced the usage of various music streaming platforms. When exploring, we noticed that the usage of music streaming services are a very convenient method to ease the stress from the pandemic crisis to concentrate on studying by streaming unlimited music. We also started thinking about how music has changed over time, and has evolved based on its listeners and the music industry. We hope to analyze these changes within the music industry through the many popular streaming applications that we use everyday, such as Spotify, Apple music, or Youtube music, being the leading music streaming services. After analyzing the trends and basis of music streaming, we will be able to get a better understanding of why we are so attracted to this new trend and will help the world understand the conscious choices of trending music and reflect on our actions from before and after the pandemic.

**Datasets**

**Spotify Data**

The following link takes one to the Spotify dataset that we will be using for our proposal and project. The dataset is called “ Spotify Top 100 Songs of 2010-2019” and it shows data from Spotify from the years 2010 and 2019. The dataset contains 34 different columns ranging from integers, decimals and strings. It also contains 945 rows. The dataset ranges from song names that were in the top 100 for each of the years, duration of songs, genres, year songs were added, artists and much more features for each particular song. The dataset was taken from the Kaggle website and has access to a huge repository of community published data.

<https://www.kaggle.com/datasets/muhmores/spotify-top-100-songs-of-20152019>

(folder to more spotify datasets <https://drive.google.com/drive/folders/1L2H9-8-jSDk1jR3gPMC3NeRMwkSU7qKV?usp=sharing>)

**Apple Music Data**

The second dataset we will be working with is the apple music dataset. Unfortunately we do not currently have access to that dataset for another week since our request stated that. Using this dataset we will be able to answer the research questions such as the active users and duration of the songs. We will be looking at the date column to check for users over the recent year. The main purpose of this dataset is to focus on apple music specifically for which we cant access with the other datasets.

#### **Youtube Music Data**

Our final dataset will be a youtube music dataset. The dataset consists of 13 columns and 157 rows containing data on specific artists. The columns contain data such as subscribers, duration, and dates of songs which we will be using in answering the research question. This dataset is from the Kaggle website where we found some of the other datasets.

[Youtube Music Data | Kaggle](https://www.kaggle.com/datasets/yashtiwari1906/youtube-music-data)

**Challenge Goals**

The first challenge we are planning to meet is being able to use **multiple datasets**. We think our project will fulfill this goal by utilizing multiple datasets to compare the trends between music streaming services. This approach will give us a better insight of what we are intending to research. Another challenge we intend to incorporate in our project is to use a **new library** in a significant way to help with our analysis. We will fulfill this goal by utilizing a library such as plotly to incorporate different graphs and skills in our project.

**Work Plan**

|  | Name | Description | Due by | Estimate time |
| --- | --- | --- | --- | --- |
| Task 1 | Download Python and familiarize with program | Each member will need to download python individually, add the datasets to their program and play around with data. | May 15th | 1-2 hours |
| Task 2 | Code for visualizations | Answer the research questions by computing code and creating visuals as a group. | May 22nd | 5-10 hours |
| Task 3 | Edit/revise code and challenge goals | Look through code thoroughly and fix mistakes and style formatting. Make sure code meets all points and challenge goals are met. Clean code and check with TA with any uncertainties. | May 27 | 3-6 hours |
| Task 4 | Complete Code/Report | As a group work to finish the codes to answer the research questions to also finalize the report for the project. By trying to complete thus 2 days before its due date we will have time to seek help or feedback on our codes and report. | May 31st | 4 hours |
| Task 5 | Complete Poster | Come together as a group to create a poster to showcase the project. | June 6th | 3-4 hours |
| Tak 6 | Peer Feedback | Complete and turn in the Peer Feedback form individually. | June 10 | 15-30 min. |

**Workflow Description**

Each of our members have different skills when it comes to coding, so if any challenges arise within any of the parts, we know we can count on each other for support in overcoming coding challenges or help each other seek for help. Each of us in the group will focus on our specific research question that we came up with to try to solve our question and afterwards we will check each other's codes to be sure the code is efficient and accurate to what the research questions are asking for. We as a group plan to meet days before each due date to make sure everyone is doing well and not falling behind. We also plan to get together just whenever we can in order to be productive and hold each other accountable for this project.