Data Visualization - Process Book
Spotify Analytics (Spotiviz)

Rishitha Guttapalle Mohan (34735824, rguttapallem@umass.edu)
Sreya Talasila (34733781, stalasila@umass.edu)
Muhammad Yusuf Hassan (34870192, mdhassan@umass.edu)

Website: https://md-hassan.github.io/CS571-Project-Spotiviz/index.html
GitHub: https://github.com/md-hassan/CS571-Project-Spotiviz
Project Screen Cast: https://www.youtube.com/watch?v=wFi1tBVjQGc

Initial Project Proposal

Background and Motivation:

Music consumption has evolved significantly over time, influenced by cultural shifts, technological advancements, and platform-driven recommendations. Understanding these trends can provide valuable insights into how genre preferences, artist longevity, and album popularity change across different periods. Additionally, with the rise of social media platforms like TikTok, certain songs experience viral success, significantly impacting their streaming performance.

This project aims to explore how music tastes evolve, both at a societal level and on an individual scale. Which genres have grown or declined in popularity? Do certain albums maintain their relevance years after release? How long do artists typically stay in the mainstream before fading out or reinventing themselves?

Beyond genre and artist longevity, this project also aims to explore how mood influences listening habits. With streaming services providing audio feature analysis—such as energy, danceability, and valence—it is possible to map how people gravitate toward different moods in their music choices over time. Additionally, the rise of social media has introduced a new dynamic: songs can go viral overnight due to trends on TikTok, Instagram, or YouTube, leading to surges in streaming numbers. This raises the question of how external cultural forces shape music consumption patterns.

Podcasts have also emerged as a major component of digital audio streaming. Understanding podcast popularity, genre trends, and user engagement can provide insights into how people consume spoken-word content compared to music.

Finally, this project is not just about broad trends—it also aims to provide personal insights. Many listeners are curious about their own habits: How have their preferences shifted over time? Do they gravitate toward different moods in different seasons? Which artists have been constants in their playlists? By combining population-wide analysis with user-specific insights, this project will offer an engaging and interactive way to explore music data.

Possible Objectives and Insights we could gain :

Some questions that this project aims to answer are:

Genre Evolution

- How have people's genre preferences shifted over time?

Artist Longevity & Activity

- Which artists maintain relevance over extended periods?
- How frequently do they release new content?
- Have artists had to change their styles to keep up with current trends?
- Which artists saw the fastest rise in popularity? Did they sustain it or fade quickly?

Album Popularity

- Which albums remain popular long after release?
- How do album sales and streaming numbers compare over time?

Mood-Based Listening Trends

- How do different moods correlate with listening habits?
- Are certain genres or moods more popular in specific seasons? (e.g., do people listen to more melancholic music in winter?)

Cultural Impact of Songs

- Can we track songs that have gone viral due to external factors like TikTok?
- Do viral songs fade quickly, or do they lead to sustained popularity?

Podcast Consumption

- What are the most popular podcast genres?
- How does podcast streaming behavior compare to music streaming habits?
- Are longer or shorter podcast episodes more engaging?

Personalized Insights

- Provide users with a way to analyze their own listening history through similar metrics and questions as above.
- How often do people replay their favorite songs/albums vs. exploring new music?

The insights derived from our analysis could serve not only the music industry, but also everyday users who want to explore their own musical journey in a more structured and visual way.

Data Sources:

We plan to collect data from multiple sources for this project:

- **Spotify Charts:** this will be our main source of weekly and daily statistics of top songs, albums, and artists

- **Official Spotify API:** we plan to use this source to collect additional metadata on the individual songs. eg. song length, featured artists, genres, features like danceability, energy, instrumentalness, liveness, etc.

- **Kworb.net, Last.fm:** contains historical music trends

- **Stats.fm:** provides detailed personal listening statistics

- **Tokhits.com, Tokchart.com:** to extract TikTok music trends

Data Processing:

Since these data sources originate from different platforms, significant data cleaning and preprocessing may be required.

Data Aggregation: aggregate and merge artist, album, song trends from Spotify, Last.fm, and Kworb.net while handling discrepancies.

Genre Classification: categorizing music genres using Spotify's API.

Mood Estimation: Using Spotify's audio features (e.g., valence, energy) to categorize songs by mood.

Cultural Impact Indicators: Identifying spikes in song popularity from external sources (e.g., TikTok music trends).

Visualization Preparation: Formatting the processed data for different visualization types.

Visualization Design:

Prototype 1:
The user can search for songs, artists, and albums. The user can then click on any of the results to get more details on that entry. Specifically, we plan to show the following three visualizations:

Fig1: Album popularity over time:
- Displays the artist's total album streams for a specific year using a barchart.

- The user can select from the dropdown the specific year they want to visualize.
- This chart allows us to show three points of information: the artist's top albums, album streams and the year under consideration.

Fig2: Genre-based listening trend:
- Displays comparative popularity of music genres across time, highlighting the searched entity's (song/artist/album) genre.
- A line chart is apt here because it can show a continuous comparative trend

Fig3: Artist collaborations:
- Shows the selected artist and their collaborations and features in the form of a network.
- A network works well here since that is the nature of data we are working with: multiple artists, their albums, and the connections between them.

Prototype 2:
This page shows weekly analytics. The user selects a week they are interested to know the statistics of, and the following visualizations are displayed:

Fig4: Total streams of top artists weekly
- Shows the number of streams of top artists throughout the week.
- Line chart used as it makes it easy to visualize trends in the data - peaks, declines and patterns.

Fig5: Popular albums of the week
- Shows the number of streams for the most popular albums of the week. Bar chart used.
- The bar chart allows for an easy comparison of multiple albums and their changing popularity levels.

Fig6: Popular podcast genres of the week
- Shows the number of streams of popular podcast genres in the week on a donut chart.
- The donut chart provides a clear, visually appealing breakdown of the most popular music genres, allowing us to compare proportions easily.

Fig7: Popular music genres of the week
- Shows the number of streams of popular music genres in the week.
- Uses a bubble chart to show similar genres close to each other, bubble size represents the number of streams

Prototype 3:
This prototype shows the page layout of the analytics dashboard. On choosing Music Analytics, visualizations on genre popularity and album popularity are displayed. On choosing Podcast Analytics, visualizations on genre popularity along with top podcasts charts are displayed.

Visuals

Visual 1: Listening Trends Over Time
We used a line chart to effectively show how listening trends fluctuate over time, making it easy to identify peaks, declines, and seasonal patterns in music consumption.

Visual 2: Album Popularity Over Time
A horizontal bar chart was chosen to display album popularity trends over time, as it allows for an easy comparison of multiple albums and their changing popularity levels.

Visual 3: Most Popular Song Genre
The donut chart provides a clear, visually appealing breakdown of the most popular music genres, allowing us to compare proportions easily.

Visual 4: Music Network: Genre, Artists, and Albums
A network chart was selected to showcase the relationships between music genres, artists, and albums, revealing connections and collaborations effectively.

Visual 5: Trending Podcasts Over Time
We used a line chart to track how different podcasts have trended over time, making it easy to observe growth, popularity spikes, and seasonality in podcast listening.

Visual 6: Most Popular Podcast Genre (Based on Views and Average Rating)
A vertical bar chart was chosen to compare podcast genres based on views and average ratings, providing a clear visual representation of audience preferences.

Must-Have Features:

Our must have features include:

- Artists trends
- Song trends
- Album trends

Optional Features:

Our optional features include:

- Mood-based trends
- Podcast trends
- Impact of virality (Tiktok)
- Personalized insights

Project Schedule:
Mar 3 - Mar 10: Data scraping
Mar 10 - Mar 17: Data cleaning
Mar 17 - Mar 23: Music Analytics
Mar 23 - Mar 30: Podcast Analytics
Mar 30 - Apr 6: Key Performance Indicators
Apr 6 - Apr 13: Settings Page
Apr 13 - Apr 20: Finishing up
Apr 20 - end: Process book, Demo video

---x---

Clarifications to the initial proposal, based on Gradescope regrade request submitted on Mar 18

Links to data sources and APIs:

- Spotify Charts: https://charts.spotify.com/charts/overview/global
- Spotify API: https://developer.spotify.com/documentation/web-api
- Kworb.net: https://kworb.net/
- last.fm: https://www.last.fm/
- stats.fm: https://stats.fm/
- Tokhits.com: https://www.tokhits.com/
- Tokchart.com: https://tokchart.com/

Brainstorming:

Filter: We have not explicitly mentioned the filtered Ideas, but we have mentioned all the ideas in the brain storming page, and filtered what we were implementing in the further sheets.

Combine and Refine:

Discussion: All the prototype designs we have proposed are complementary to each other and are not alternative. Furthermore, we have mentioned advantages and disadvantages of each visualization, present in the design.

Must-Have Features:

Clarification on which component refers to which graph:

We plan to have parallel charts for all the three main categories of information: artists, albums and songs. Based on this, we plan to have the following charts (please refer to page 5 and 6 of the design sheets)

- Artist trends: Visual 1 (line chart), Visual 2 (bar chart), Visual 3 (donut chart), Visual 4 (network chart)
- Album trends: Visual 1 (line chart), Visual 2 (bar chart), Visual 3 (donut chart), Visual 4 (network chart)
- Song trends: Visual 1 (line chart), Visual 2 (bar chart), Visual 3 (donut chart), Visual 4 (network chart)

--x--

Process Book

Overview and Motivation:

Music consumption has evolved significantly over time, influenced by cultural shifts, technological advancements, and platform-driven recommendations. Understanding these trends can provide valuable insights into how genre preferences, artist longevity, and album popularity change across different periods. Additionally, with the rise of social media platforms like TikTok, certain songs experience viral success, significantly impacting their streaming performance.

This project aims to explore multiple themes related to music statistics. How long does a viral song or album usually on the top? Do older songs/albums maintain their relevance years after release or have a sudden resurgence? Which genres have grown or declined in popularity? How long do artists typically stay in the mainstream before fading out or reinventing themselves?

Additionally, the rise of social media has introduced a new dynamic: songs can go viral overnight due to trends on TikTok, Instagram, or YouTube, leading to surges in streaming numbers. This raises the question of how external cultural forces shape music consumption patterns.

Apart from a global analysis, we also plan to explore local trends in a few select cities—it would be interesting to see if certain artists or genres are more popular in some regions.

Related Work:

The inspiration for this project came from Spotify's annual Wrapped personalized summaries. However, we wanted to take it a step further and explore beyond just static yearly numbers and statistics. We wanted to create an interactive dashboard that:

1. Provides trends up to a **week's granularity**, not just yearly.
2. Allows comparisons between **artists, genres, or songs**.
3. Offers **heatmaps, timelines, and dynamic charts** for exploration.

In developing **SpotiViz**, a tool for visualizing weekly Spotify chart data, we were also inspired by academic work that highlights the power of large-scale music analytics. One particularly relevant paper is *"Close reading big data: The Echo Nest and the production of (rotten) music metadata"* [1] by Maria Eriksson (2016). The paper explores how The Echo Nest—acquired by Spotify—analyzed vast amounts of online music-related text to power recommendation systems, and discusses both the potential and pitfalls of using big data in cultural applications.

While SpotiViz does not use personal user data or recommendation algorithms, the paper underscores the broader value of **music metadata and chart analytics** for uncovering **cultural trends and shifts in listening behavior**. This aligns directly with SpotiViz's goal of providing **insightful visualizations of public Spotify chart data** over time, helping users explore how artists and songs rise and fall in popularity week by week.

There are many tools for data analysis, but few are designed for **interactive exploration of music trends** over time. We wanted SpotiViz to fill that gap by giving users a way to:

1. Explore Spotify charts week-by-week.
2. Track artist popularity,
3. Discover patterns in listening behavior.

Questions:

At the outset of our project, we aimed to explore a broad set of questions surrounding Spotify listening trends, ranging from genre evolution to mood-based patterns and podcast consumption. Our initial research questions were ambitious and designed to provide comprehensive insight into music and audio consumption:

Original Research Questions:

- **Genre Evolution:**
  How have people's genre preferences shifted over time?

- **Artist Longevity & Activity:**
  Which artists maintain relevance over extended periods?
  How frequently do they release new content?
  Have artists had to change their styles to keep up with current trends?
  Which artists saw the fastest rise in popularity? Did they sustain it or fade quickly?

- **Album Popularity:**
  Which albums remain popular long after release?
  How do album sales and streaming numbers compare over time?

- **Mood-Based Listening Trends:**
  How do different moods correlate with listening habits?
  Are certain genres or moods more popular in specific seasons? (e.g., melancholic music in winter?)

- **Cultural Impact of Songs:**
  Can we track songs that have gone viral due to external factors like TikTok?

Do viral songs fade quickly, or do they lead to sustained popularity?

- **Podcast Consumption:**
  What are the most popular podcast genres?
  How does podcast streaming behavior compare to music streaming habits?
  Are longer or shorter podcast episodes more engaging?

- **Personalized Insights:**
  How often do users replay favorite songs/albums versus exploring new music?

After examining the structure and limitations of the Spotify dataset we were working with (which consisted of weekly top chart data with artist names, song rankings, and week-by-week performance), we recognized that many of the original questions—especially those involving mood labels, podcast consumption, or user-level personalization—would require metadata or APIs we did not have access to.

We discussed these limitations as a team and decided to **reorient our project goals** to focus on the following refined, **data-driven** questions:

Final Questions Addressed:

1. **Popularity Over Time:**
   Which artists/songs/albums/genres consistently maintained high chart rankings over time?
   → Visualized through **line charts and heatmaps of artist rank trends**.

2. **Rapid Rise and Fall Patterns:**
   Which entities saw sudden spikes in popularity, and how sustainable were those trends?
   → Identified through **rank volatility and temporal trajectories**.

3. **Chart Dominance and Longevity:**
   Which entities remained active and visible in the top charts across multiple weeks?
   → Explored using **interactive artist activity visualizations**.

4. **Genre Trend Evolution:** How has the streaming popularity of selected music genres changed over time globally/locally ?
   → Explored using Genre popularity over time globally and number of songs in the Top 100 of a genre in the local level.

5. **Genre Representation in Top Charts:**
   What proportion of the top globally streamed tracks belongs to each music genre?
   → Explored using Genre Share

6. **City-Level Genre Uniqueness:**
   What are unique tracks of a genre across various cities?
   → Explored using number of songs in Top 100 of a genre in the local level and Genre Share

These questions formed the core of our final visualizations and allowed us to still answer meaningful and insightful aspects of musical popularity—even if on a narrower scale than originally envisioned. In this final version after our milestone, we have included many other questions and other visualizations that answer these questions.

Data:

Our initial plan was to scrape different statistics of popular songs, albums and artists across a span of a few years. The data would include the stream counts, the leaderboard ranks, and streak information.

Although Spotify releases some statistics of popular songs and artists at the end of each year, it does not provide any numbers or statistics on a more granular scale. Since our project requires data with more frequent intervals, we chose to resort to scraping data. From our research, the only publicly available source that satisfied all our needs was charts.spotify.com, a website that collects weekly Spotify data of the top 200 songs, artists and albums.

Challenges Encountered with the Scope of the Data:

After starting with the scraping process on charts.spotify, we were met with a few challenges which influenced the scope of data that we could collect, and hence influenced the set of questions that our project would answer.

- We realized that charts.spotify only provides information on the top 200 items in a week, and does not provide any information outside of that. This means that if a song is present in the chart, we can only track it and get its statistics for the duration it remained within the top 200.
- Additionally, charts.spotify provides streaming counts and rankings only for song data. For albums or artists, it only provides their weekly ranks with no other numerical indicators.
- Initially, we had plans to make visualizations using audio feature analysis (such as energy, danceability, valence, etc.) provided by Spotify. However, we realized later that Spotify has deprecated the public availability of these statistics due to misuse by some parties in training Machine Learning models.

Due to these unforeseen challenges in gathering the data, the scope of information available to us has changed compared to our initial proposal. This has naturally prompted us to slightly change our motivations and the questions we would like to answer.

Process of Scraping the Data:

We used the Selenium (https://selenium-python.readthedocs.io/) package in Python for data scraping. For each of the categories of data i.e. songs, artists, albums, we start from the main page, (https://charts.spotify.com/charts/view/regional-global-weekly/latest) and collect the metadata such as the *rank*, *song name*, *artists name(s), number of streams* (if available), *streak*, *release date*, etc. Then we keep moving to the previous week, collecting the data as far back in time as the site allows us to go. For instance, we were able to collect song data ranging from Dec 2016 to the present day, but artist-specific charts were available starting only Oct 2021.

Since we were scraping the website with a high frequency, the website would occasionally ask to fill in a captcha, which would halt the scraping process. To get around this, we switched to Undetected Chromedriver (https://github.com/ultrafunkamsterdam/undetected-chromedriver), which is an optimized version of Selenium that is less susceptible to being caught. If the scraper was caught still, we would randomize the *user_agent* argument and resume the process.

One problem we faced here was that in the case of multiple artists for a song. The artist field would be collected as a single string with the artists separated by a comma. We could get around this by simply splitting the string by comma and treating each part as a separate artist. However, if the artist's name itself contained a comma, this method would not work (for instance the artist “Tyler, The Creator” could wrongly be considered as two artists “Tyler” and ”The Creator”). We were stuck with this issue for some time, until we discovered the solution, as discussed ahead.

```
{
    "week": "2022-07-14",
    "rank": 16,
    "song": "Master of Puppets (Remastered)",
    "song_id": "2MuWTIM3b0YEAskbeeFE1i",
    "artist": {
        "Metallica": "2ye2Wgw4gimLv2eAKyk1NB"
    },
    "peak": "17",
    "previous": "22",
    "streak": "2",
    "release_date": "Mar 3, 1986",
    "first_entry": "Jul 7, 2022",
    "first_entry_pos": "22",
    "total_weeks": "2",
    "producers": [
        "James Hetfield",
        "Lars Ulrich",
        "Cliff Burton",
        "Kirk Hammett",
        "Flemming Rasmussen"
    ],
    "songwriters": [
        "Cliff Burton",
        "James Hetfield",
        "Kirk Hammett",
        "Lars Ulrich"
    ],
    "source": "Blackened Recordings"
},
```


Sample data collected for the song “Master of Puppets” by Metallica

Scraping the Genre Data:

Once we had the song metadata from charts.spotify, our plan was to extract genre data and audio features (like energy, danceability, valence, etc.) from the Official Spotify API.

The problem we faced here was that in order to search a song’s metadata on the Spotify API, we needed to have its Spotify ID. However, since we had scraped the data from an external site, we did not have the Spotify IDs. So, we thought of searching for the IDs by using the ‘search’ endpoint of the Spotify API. Since multiple songs can have the same name, we needed additional information to arrive at the correct song entry. We could improve the search by including the album name, but we did not have those at hand since charts.spotify did not specify album name in a song’s metadata (we could try collecting those externally, but it would take time and had no guarantee of working). Alternatively, we could search using the release date as an additional filter, but there was still ambiguity here in the case of remastered songs (which date is the release date, original or the remaster date?). We faced a similar issue with getting the artist’s Spotify ID since we could not separate the artist names due to the comma ambiguity issue mentioned above.

After spending some time, we finally got around this issue of Spotify IDs when we took a closer look at charts.spotify’s HTML code. We realized that the Spotify IDs were embedded into the

URLs linking either to the song's preview picture or the artist's profile page. This discovery allowed us to gather IDs for all the songs, albums and artists. It also allowed us to overcome the artist comma separation issue since now we could simply scrape the IDs instead of the comma-separated string.

Now that we had the Spotify IDs for songs, we were in a position to scrape the genre and audio feature data from the Official Spotify API. Unfortunately, we then got to know that Spotify had deprecated the public availability of these statistics due to misuse by some parties in training Machine Learning models.

We thought of using the Apple Music API instead, but that would have cost us a $100 subscription to the Apple Developer Program. Finally, we decided to scrape the genre data from the genre tags provided by last.fm and rateyourmusic.com. The scraping process for these websites was very similar to scraping charts.spotify so we will not go in-depth for that in this report.

Cleaning the Data:

The ranking data about songs, artists and albums and related metadata collected from charts.spotify was very clean and did not require any additional steps.

However, the genre data collected from last.fm was quite noisy as the tags are crowd-sourced and anyone can contribute to it. One of the most common tags associated with a song were the year and the decade in which it was released. We used a Regular Expression in Python to filter out such strings. In particular, we used the following regex:

```
re.match(r'^\d+$', s) or re.match(r'.*\d\d\d\d.*', s) or re.match(r'.*\d\d.*', s)
# strings like '2020' or 'from the 1990s' or '10s'
```

Another popular genre tag was the artist name itself, which we also filtered by matching the from the lists of artist names associated with that song. We then filtered out any tags that had a frequency of less than 5 in the entire dataset, except those tags that contained certain keywords. Few of the keywords we used include 'metal', 'rock', 'classical', 'jazz', 'pop', 'rap', 'hip hop' among many others.

Doing the above operations did clean up the data to some extent but there were still over a thousand unique genres remaining in the dataset. This is because a large number of genre tags were very hybrid in nature like 'goth-pop', 'hindi-rock', 'folk metal', etc. In order to classify the tags in a more reliable and intelligent manner, we decided to use Large Language Models (LLMs). Since an LLM can reference a large volume of information from its training knowledge, and can also reason quite well, it is the perfect choice to do human-like "intelligent guessing" at a large scale. We chose to use the Qwen2.5-7B-Instruct model, which is a state of the art LLMs that excels on multiple tasks including reasoning, question answering, math, etc. We ran the model on the Unity Server at UMass, on an NVIDIA L40S GPU with 48 GB of VRAM. We asked the model to classify the genres into 15 main categories, also asking it to make an intelligent guess in case it knew about the artist, but the tags did not contain any useful information.

We provide our prompt below:

"

```
I will provide you with json data containing a list of artist genre
tags from a crowdsourced website. The data is quite noisy and often
contains irrelevant or incorrect tags. Please classify the artist
into one or more of the below provided main genre types.

Here are some rules to follow:
1. Ignore any irrelevant tags like "2019", "my favorite artist",
"album of the year", "male/female artist", artist names, dates,
awards, or subjective opinions.
2. Classify based on real genre information.
3. In case the tags are not helpful/absent but you know about the
artist, you may give the genre list followed by "(guess)"
4. If no valid genre tags are found, return ["N/A"].
5. Only use the main genres from the provided list. Do not invent new
genres.

Main genres to classify into:
1.   Pop
2.   Rock
3.   Hip-Hop / Rap
4.   R&B / Soul
5.   Electronic / Dance
6.   Metal
7.   Country
8.   Reggae
9.   Jazz
10.  Blues
11.  Folk
12.  Classical
13.  Latin (includes mexican, salsa, bachata, etc.)
14.  International (Afrobeats, K-Pop, J-Pop, Brazilian, etc.)
15.  Soundtrack (movie scores, game soundtracks, etc.)

Examples:
1. Input genres: ["rap", "male and female vocals", "olivia",
"brazilian"]
Output: ["Hip-Hop / Rap", "International"]
2. Input genres: ["my top artists", "rock", "alt-metal"]
Output: ["Rock", "Metal"]
```

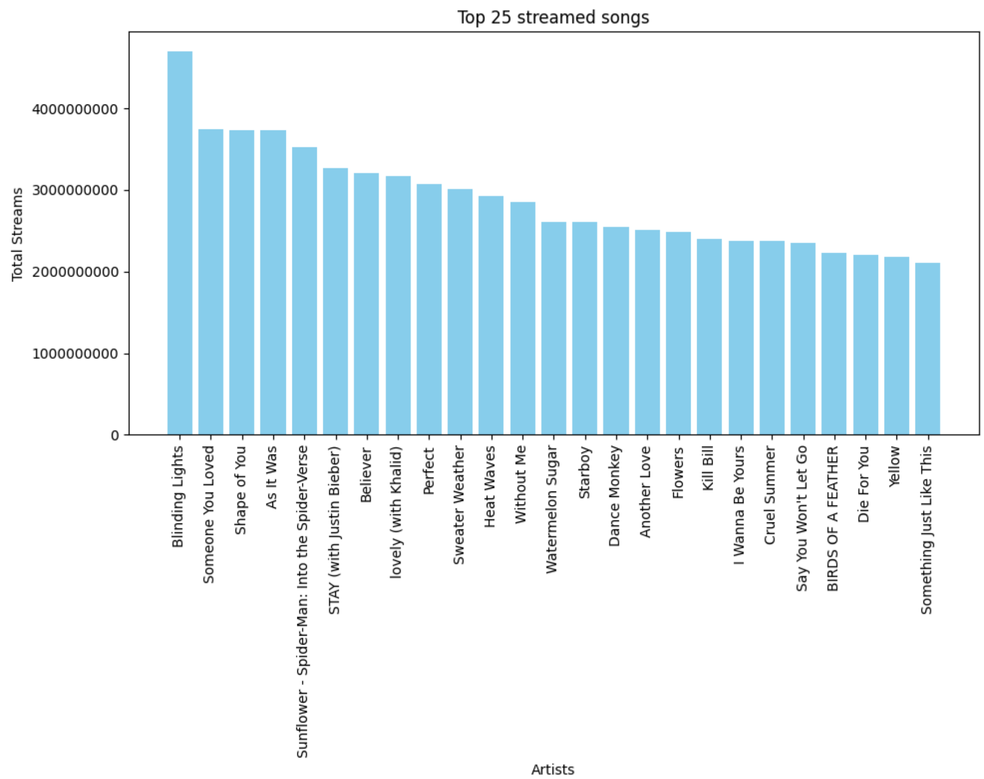
```
3. Input genres: ["5 stars"]
Output: ["N/A"]
4. Input genres: ["mexico", "corridos", "corridos tumbados", "latin",
"sierreno"]
Output: ["Latin"]
5. Input genres: ["brazilian", "samba", "bossa nova"]
Output: ["International"]
6. Input genres: ["instrumental", "film score"]
Output: ["Soundtrack"]
7. Input genres: ["1990s"], but artist is "Metallica"
Output: ["Metal"] (guess)

**Format your output exactly as a comma-separated list with no extra
explanation. Only output the list.**

Now classify the following:
”
```

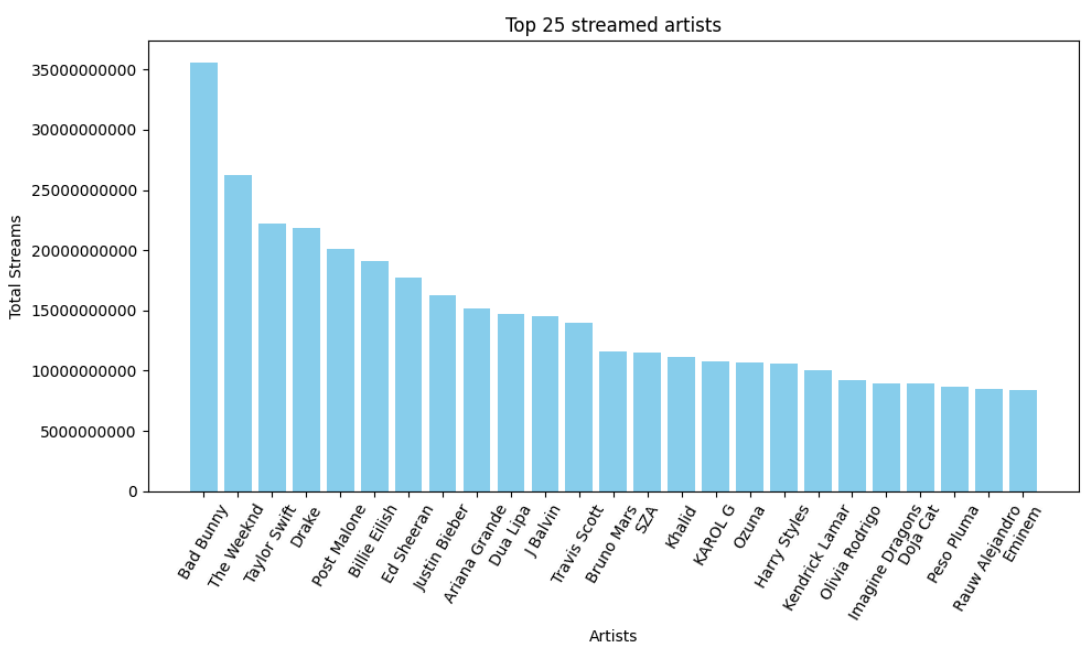
The outputs provided by the LLM were usually clean and required minimal post processing from our side.

Exploratory Data Analysis:

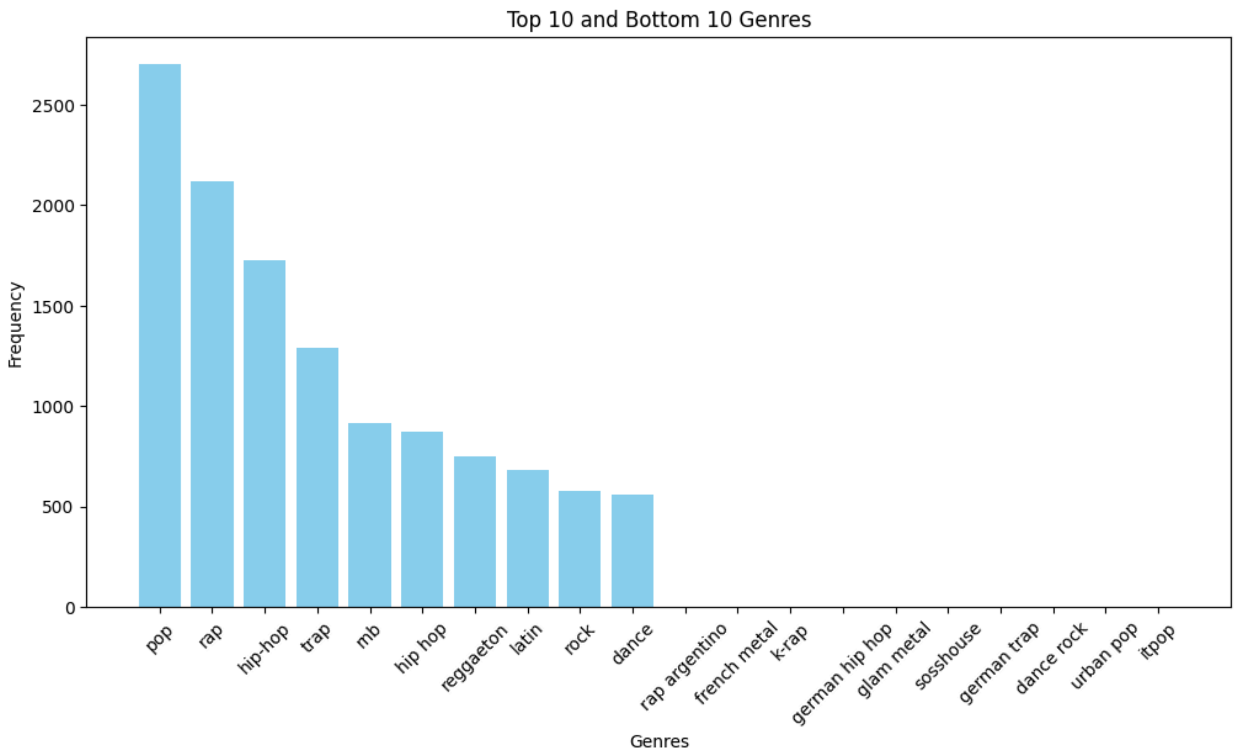
As an initial exploration we looked at basic statistics of the data we collected. From our current song data which ranges from Dec 2016 to Mar 2025, we have collected 5981 distinct songs. Surprisingly, these songs span only 2290 distinct artists. This short analysis reveals that most artists had multiple songs on the leaderboard, 2.6 on average.

To get an idea of numbers and their distribution, we plotted the frequency of the top songs, albums and artists. As an example, below are the distributions of the 25 most streamed songs, and the 25 most streamed artists in our dataset.





With respect to genre data, our dataset contained a total of 37,519 genre tags (non-distinct), out of which 2,699 were either numbers or were the artist names themselves. After this initial round of cleaning, we were left with a total of 34,820 tags, comprising 2,961 distinct tags. We realized that many of the tags were still quite noisy and decided to remove any tags with frequency less than 5, keeping the ones that contained any keywords from a list as described in the previous section. This led us to have 1,012 distinct tags. Below, we plot the top 10 and bottom 10 genres from our dataset.



How these insights informed our design:

From our analysis into the chart data, we realized that since there is such a large number of songs, artists and albums, it is not possible to view all of the ranked entities at once. Hence, we decided to give the user an option to limit the top N entities they want to visualize and also select a time duration of their choice.

From our analysis into the genre data, we realized that even after cleaning, there is a very large number of genres available. However, most of the genres can be categorized into one or more archetypes such as metal, rock, classical, jazz, pop, rap, etc. Hence, as described in the previous section, we used an LLM to classify noisy genre data in 15 main categories, significantly simplifying the categorization.

Furthermore, since last.fm and rateyourmusic.com provide genre data not only for songs, but also for artists and albums, we created visualizations for genre data in the context of songs, artists and albums as well.

Design Evolution:

From Proposal:

For Album Visualizations,

We had thought of showcasing a horizontal bar chart showing album streaming count with a single metric per album.

This developed into our Album Rank Movement visualization, which transformed by a bit from the basic horizontal bar chart into a horizontal diverging bar chart while incorporating an essential aspect of comparative analysis over a specified time frame. Instead of presenting absolute figures, we opted to display rank changes, which more effectively emphasized progress or decline.

From Milestone:

Since the milestone, we have successfully implemented genre analytics as part of our Spotify data exploration. This includes collecting and processing genre-related data and visualizing it through various charts, which we have showcased below.

For Final submission:

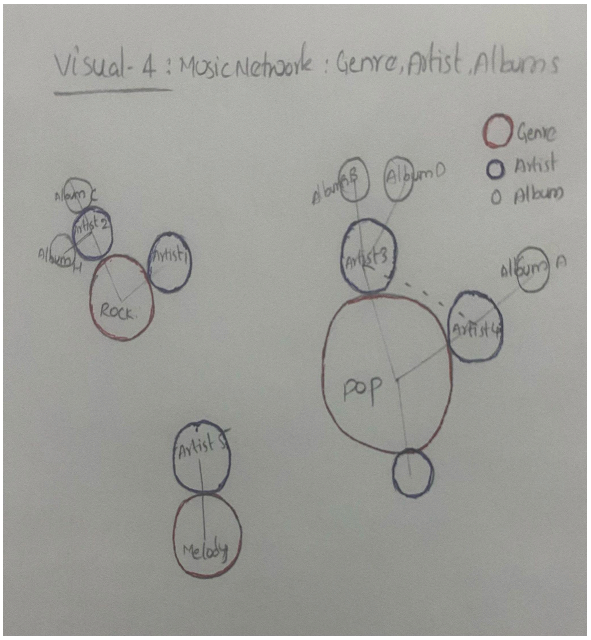
We have effectively showcased visualizations for albums, artists, and genres, progressing from the artist level to the album and genre levels, along with key indicators for all sections. Our designs have also been developed to allow filtering of specific information based on user requirements for each data group that the visualization presents. Additionally, we have provided visualizations that enable deeper insights at various stages, spanning from a global to a local level as needed. Our design has undergone some changes and adjustments from the initial proposal to the final product, and these modifications, along with minor deviations, will be detailed clearly in the following sections.

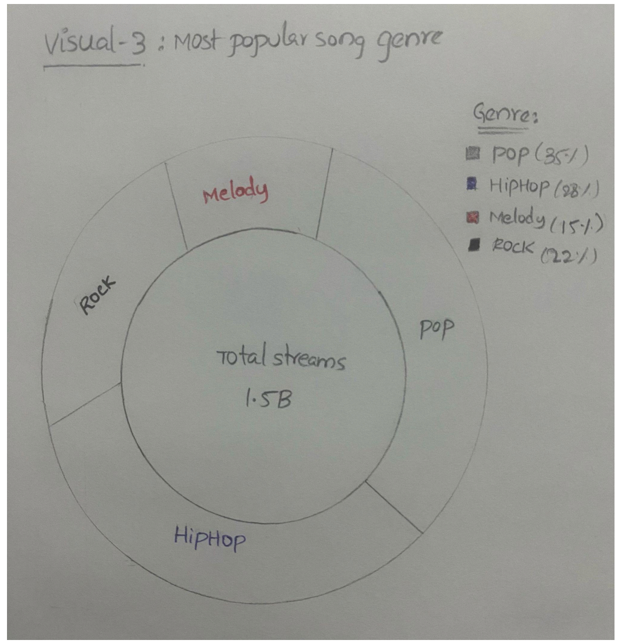
Deviations from proposal:

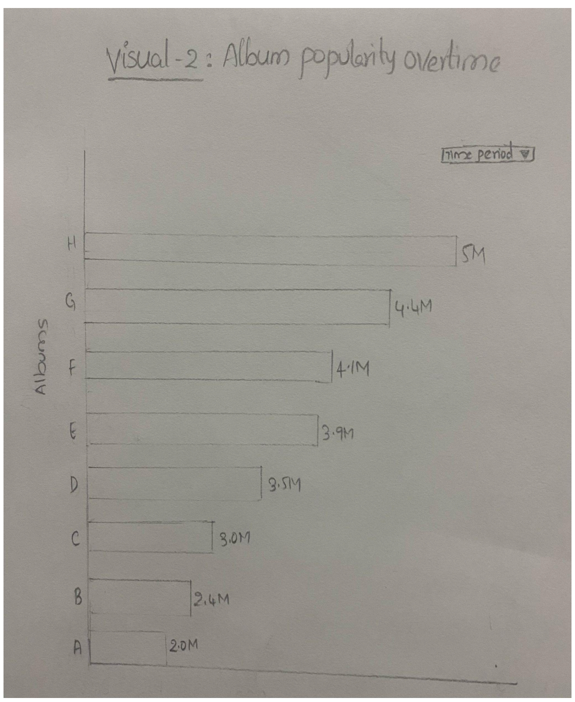
- We introduced some additional visualization types beyond the conventional ones like line and bar charts that we mentioned earlier in the proposal, aiming to enhance the communication of data insights. This choice was made to facilitate better understanding with more commonly recognized chart formats.

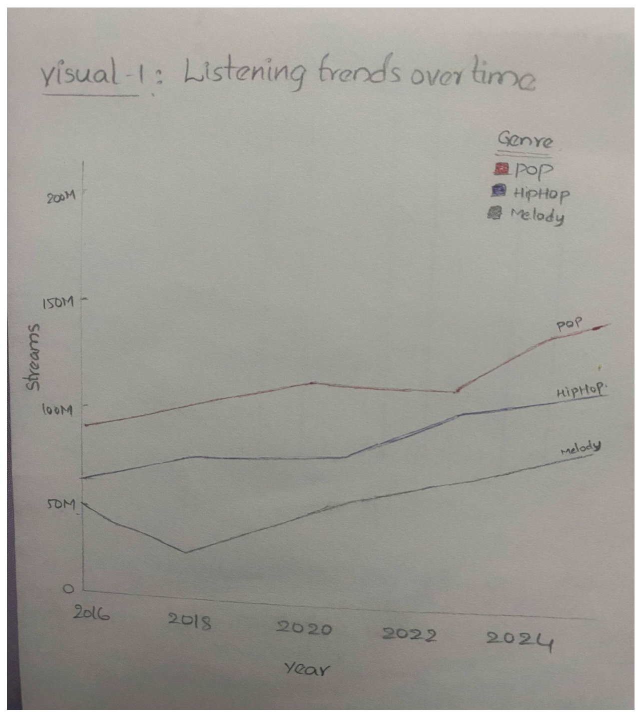
- From the data we gathered from Spotify, which was restricted to a few anticipated charts, we adjusted some questions to align better with the information we collected. However, these adjustments still centered around our primary features: albums, artists, and songs, without straying from the original concept of showcasing the proposed features.

- We enhanced the interactive features beyond the original proposal, incorporating filtering options, time range selection, and tooltip details for all visualizations. This enhancement was driven by initial feedback indicating that static visualizations did not permit adequate exploration of the extensive dataset.









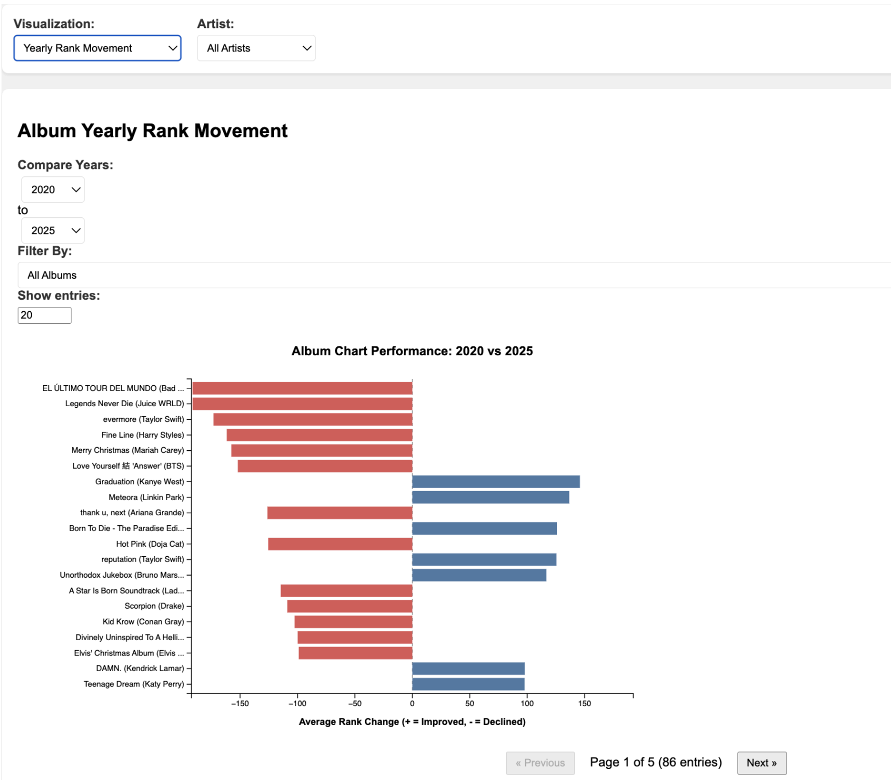
Implementation: Describe the intent and functionality of the interactive visualizations you implemented. Provide clear and well-referenced images showing the key design and interaction elements.

- For the milestone, we have highlighted two of the three essential features outlined in the proposal, specifically the visuals related to Artist and Album.

- We need to focus on genre data. Therefore, for the final presentation, we will continue to work on genre data along with the visuals associated with it.

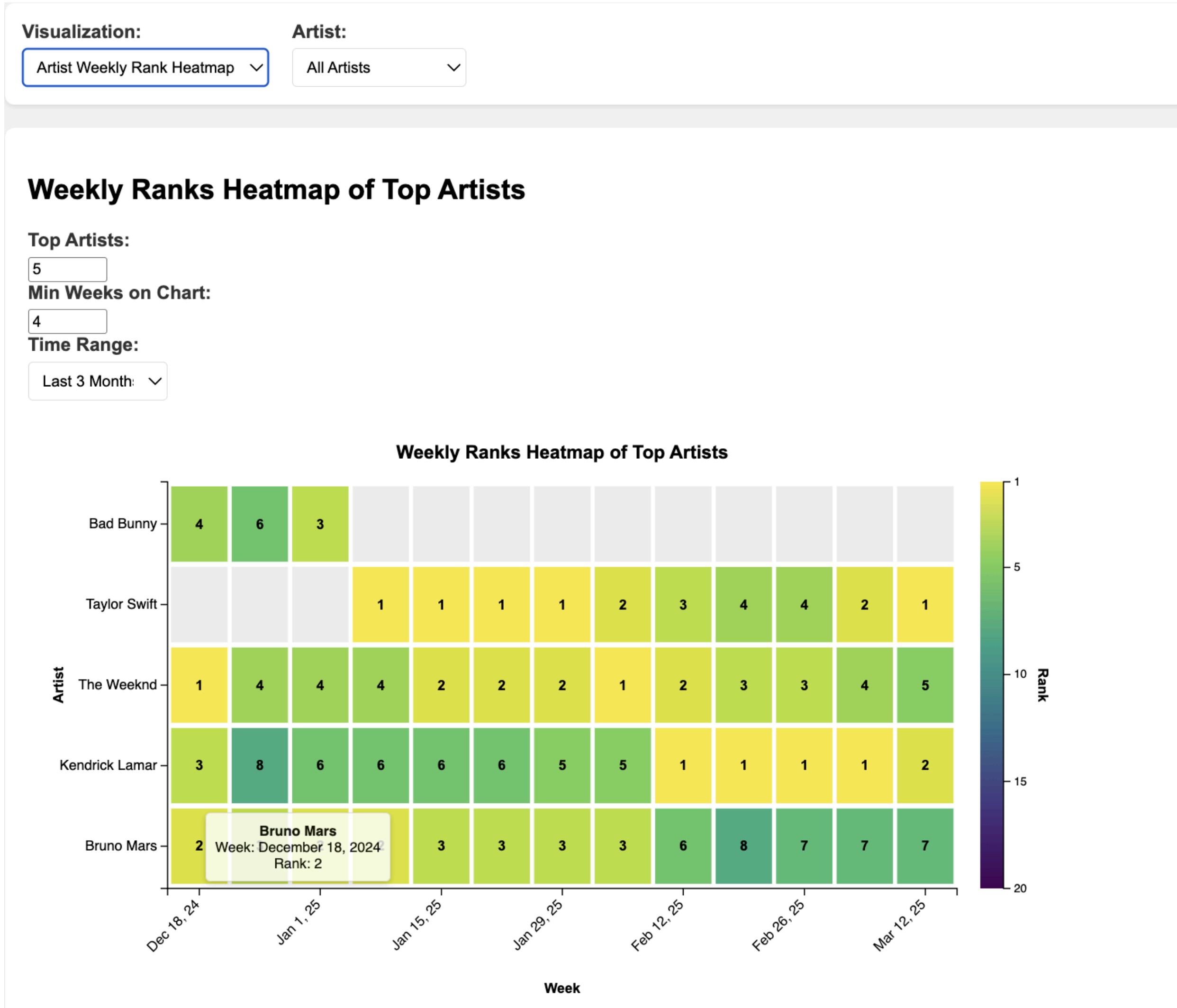
Visuals Implementations:

Visual-1:



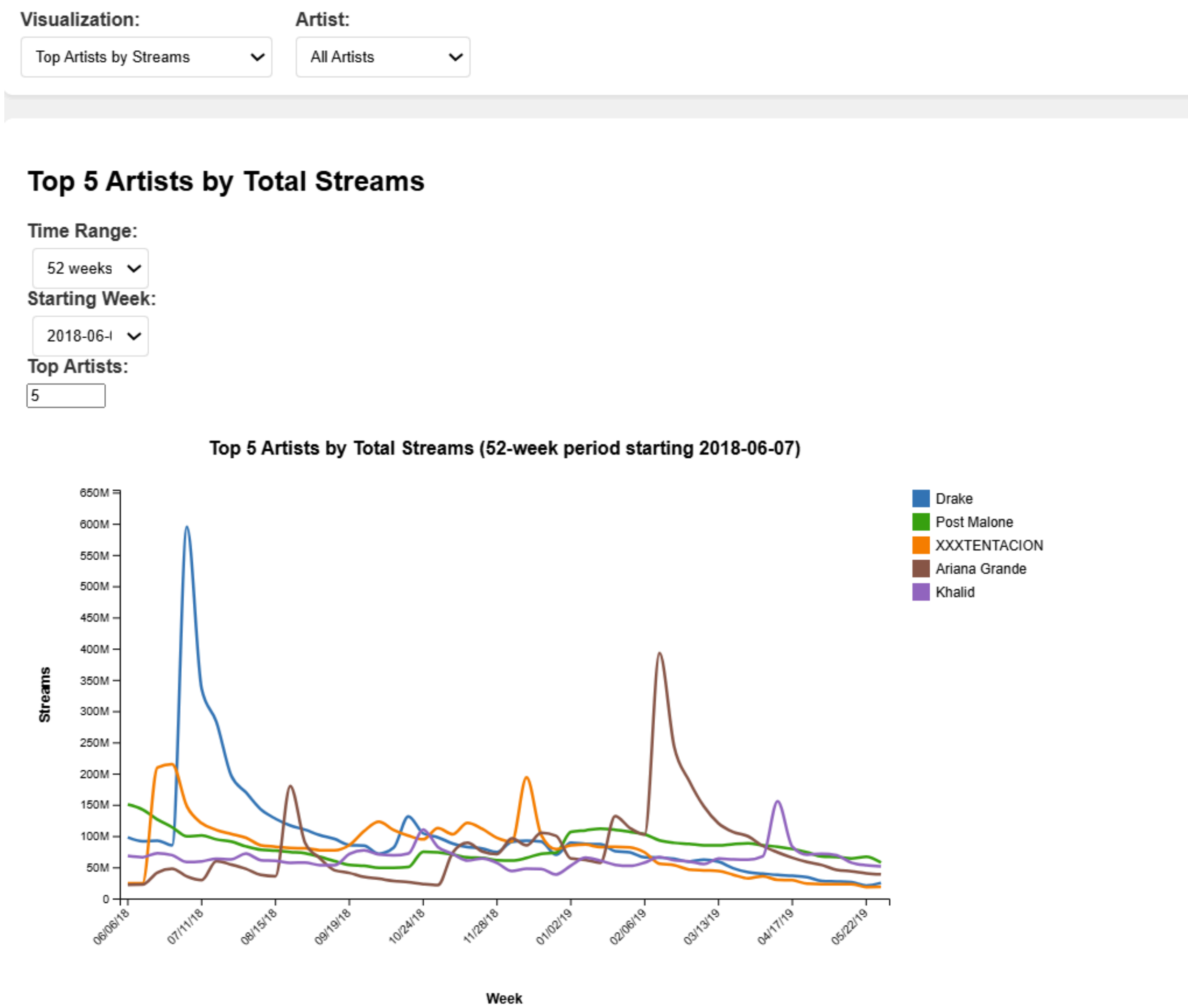
- The Intent of this visualization is to compare streaming performance of albums over years.
- Visualization: A horizontal diverging bar chart displaying rank shifts of improvements or declines across from one particular year selected to other and it is built with D3.js having positive/negative axis to show directional changes.
- Interactivity: Users are given a choice to choose the years they want to compare, filter by albums, and modify the number of entries displayed. Also it has hovering which displays the average change over the years along with their ranks which shows a decline or improvement.

Visual-2:



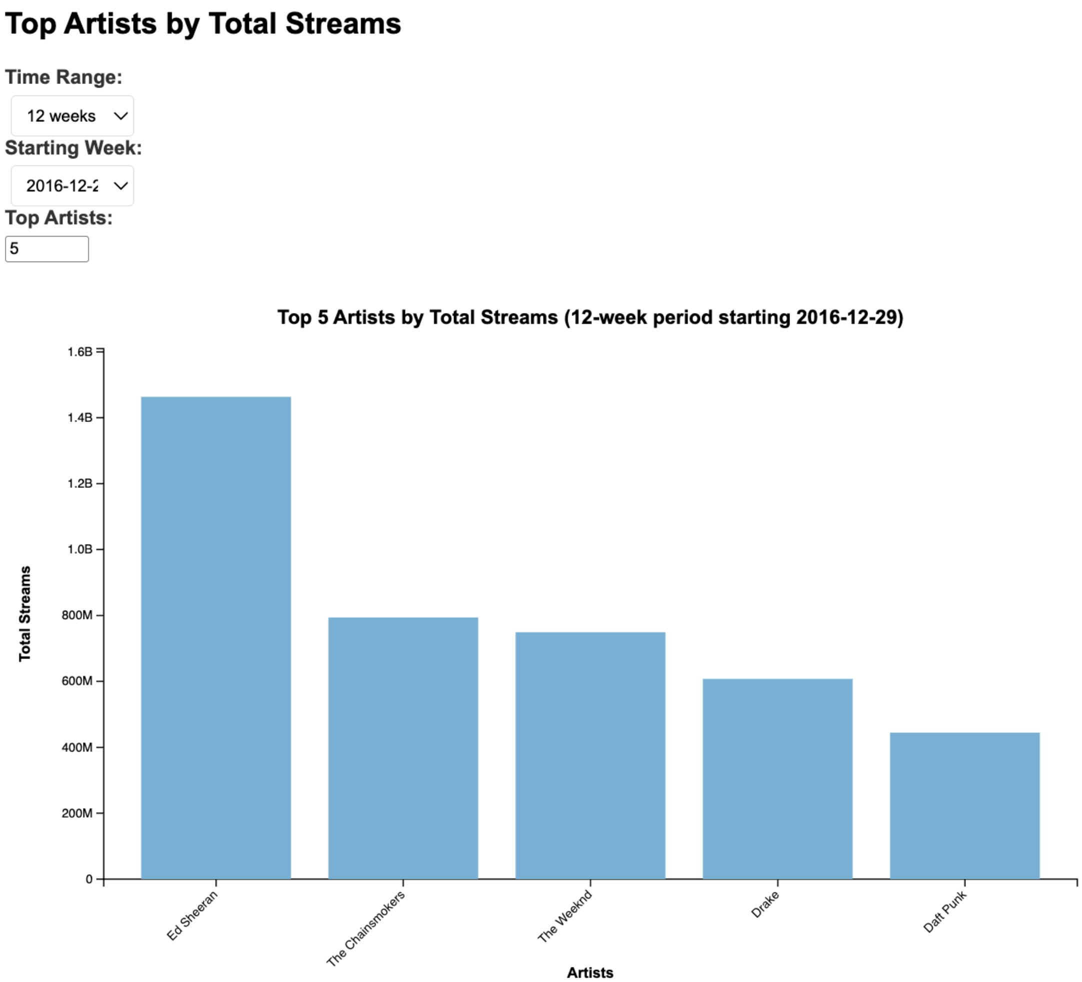
- The Intent of this visualization is to demonstrate visualization of consistency and trends in artist ranking charts.
- Visualization: A heat map displaying weekly ranking positions, with the color intensity reflecting the rank and it is built with D3.js to create responsive and customizable charts.
- Interactivity: Users have the ability to filter by the minimum number of weeks displayed on the chart, choose a time period, and top artists of the week. We will try to implement hovering over the cells to view precise rank details for the final one.

Visual-3:



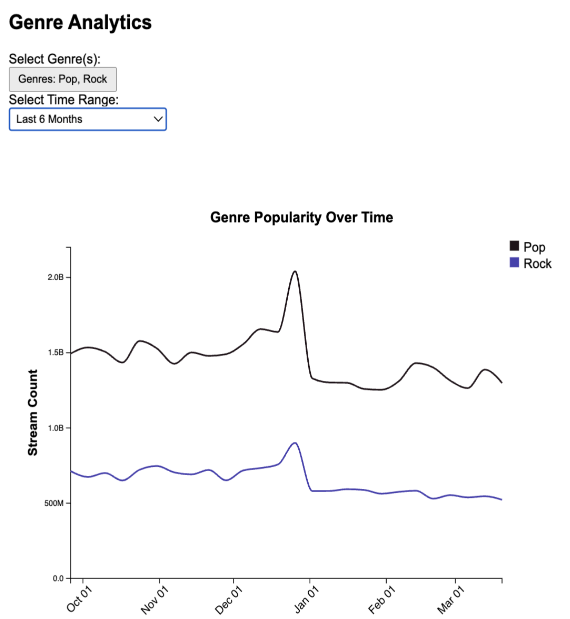
- The Intent of this visualization is to compare streaming performance of top artists over time.
- Visualization: The Line chart portrays stream counts for top n number of artists over a chosen time period and it is built with D3.js to create responsive, dynamic line charts.
- Interactivity: Users are given an option to select time range, starting week, and number of top artists to display.

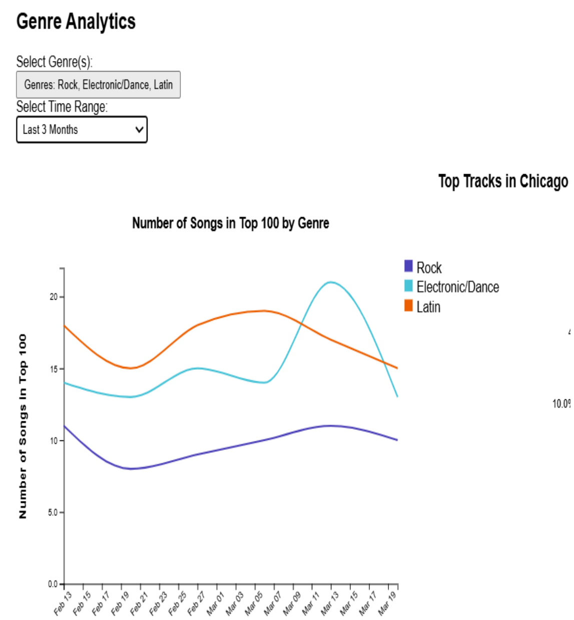
Visual-4:



- The intent of this visualization is to compare streaming performance of top music artists over a specific time period, allowing users to see which artists were most popular on streaming platforms.
- Visualization: The bar chart portrays the total streams for the top 5 artists during a 12-week period starting December 29, 2016, with Ed Sheeran clearly dominating with approximately 1.5 billion streams.
- Interactivity: Users are given an option to adjust the time range, select a different starting week, and change the number of top artists displayed, allowing for customized analysis of streaming data.

Visual-5:

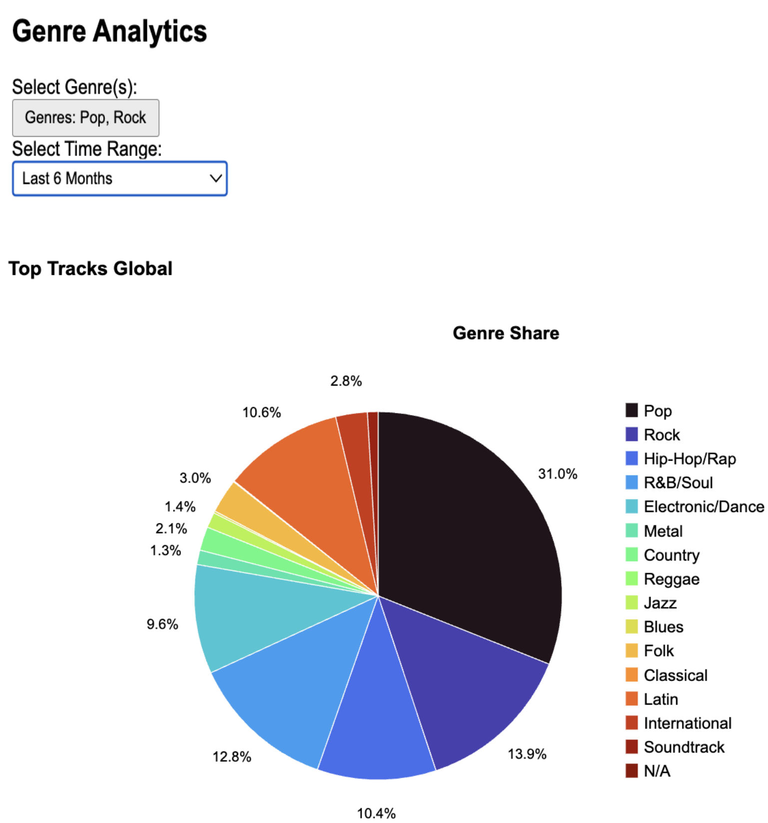




- The intent of this visualization is to compare the streaming popularity of selected music genres (in this case, Pop and Rock) over a defined time period, helping users understand genre trends globally.
- Visualization: The line chart displays the total stream counts of Pop and Rock genres over the past six months (from early October to early March).
- Interactivity: Users are given the choice to select genres using a multi-select dropdown (e.g., Pop, Rock). Adjust the time range (e.g., "Last 6 Months") using a dropdown menu to dynamically update the chart and analyze different periods.

Note: The first visualization represents Global Genre Analytics, using *Stream Count* as the primary metric to analyze and compare the popularity of selected music genres over time.The second visualization focuses on genre presence by visualizing the *Number of Songs in the Top 100*, offering insight into how frequently different genres appear among the most streamed tracks.

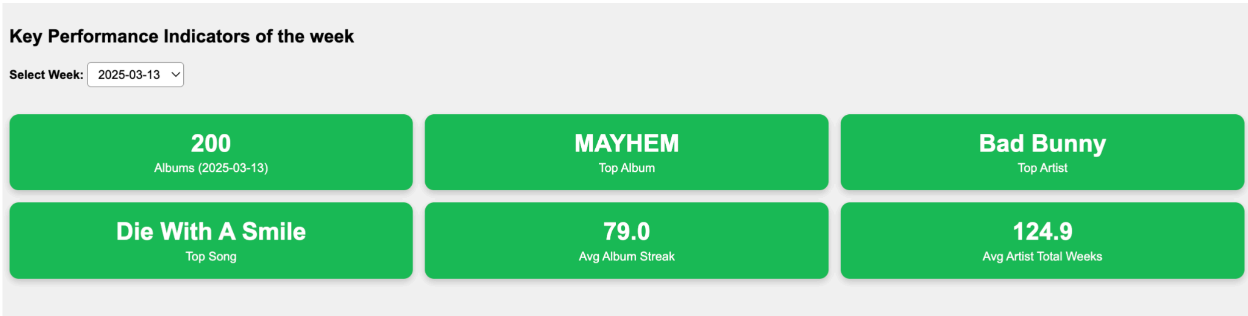
Visual-6:



- This visualization aims to illustrate the global distribution of music genres within the Top 100 tracks, highlighting the relative representation of each genre.
- Visualization: The pie chart presents the Genre Share across the global Top 100 tracks. Each slice represents the proportion of songs in that genre. The diversity of genres shows a strong dominance of mainstream categories, with niche genres having minimal presence.
- Interaction:.Adjust the time range (e.g., "Last 6 Months") using a dropdown menu to dynamically update the chart and analyze different periods.

Note: The same chart is used to visualize genre data both locally and globally.

Visual-7:



- This dashboard highlights essential weekly Key Performance Metrics in the global music landscape, helping users quickly identify top-performing albums, songs, and artists as well as broader engagement trends.
- Visualization: The visual showcases critical weekly highlights including: The total number of albums featured that week. The top-performing album, artist, and song. The average streak of albums remaining on the charts. The average total weeks artists have appeared on the charts.
- Interactivity: A dropdown menu allows users to select different weeks, dynamically updating the KPIs to support historical comparisons and trend analysis over time.

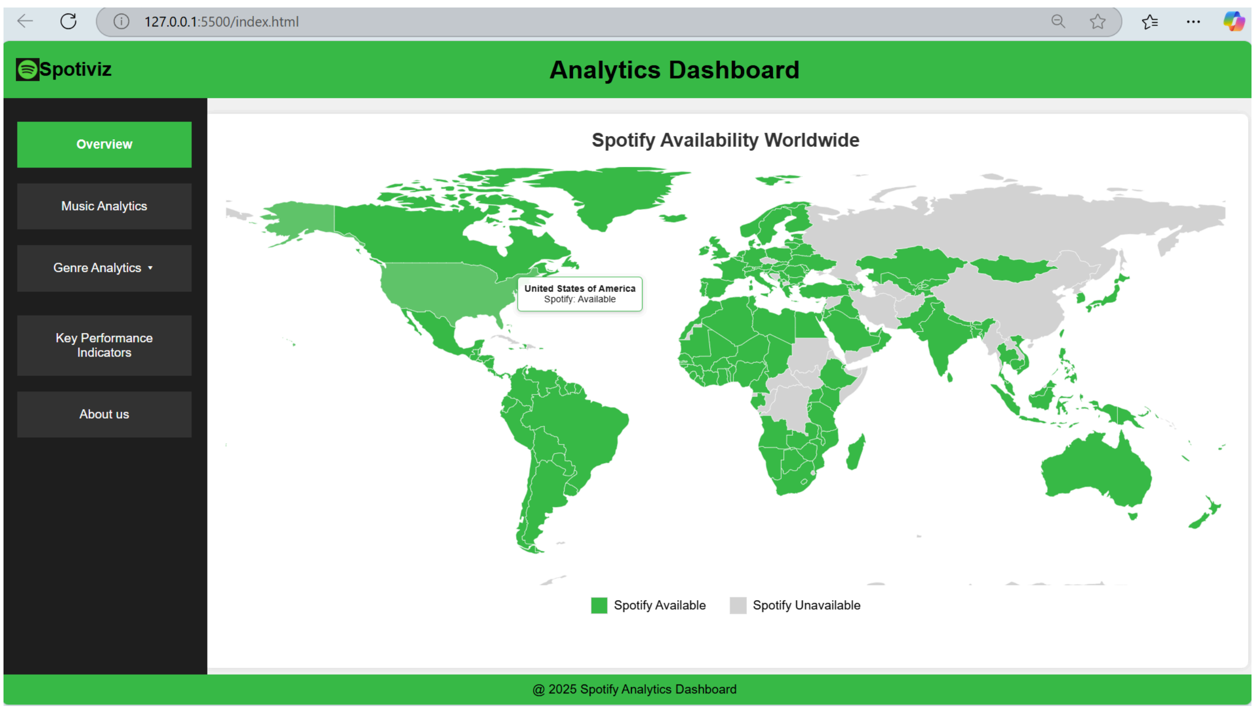
Website implementation:

Languages and technologies used for implementation:

- HTML
- CSS
- Javascript
- d3.js for visualization

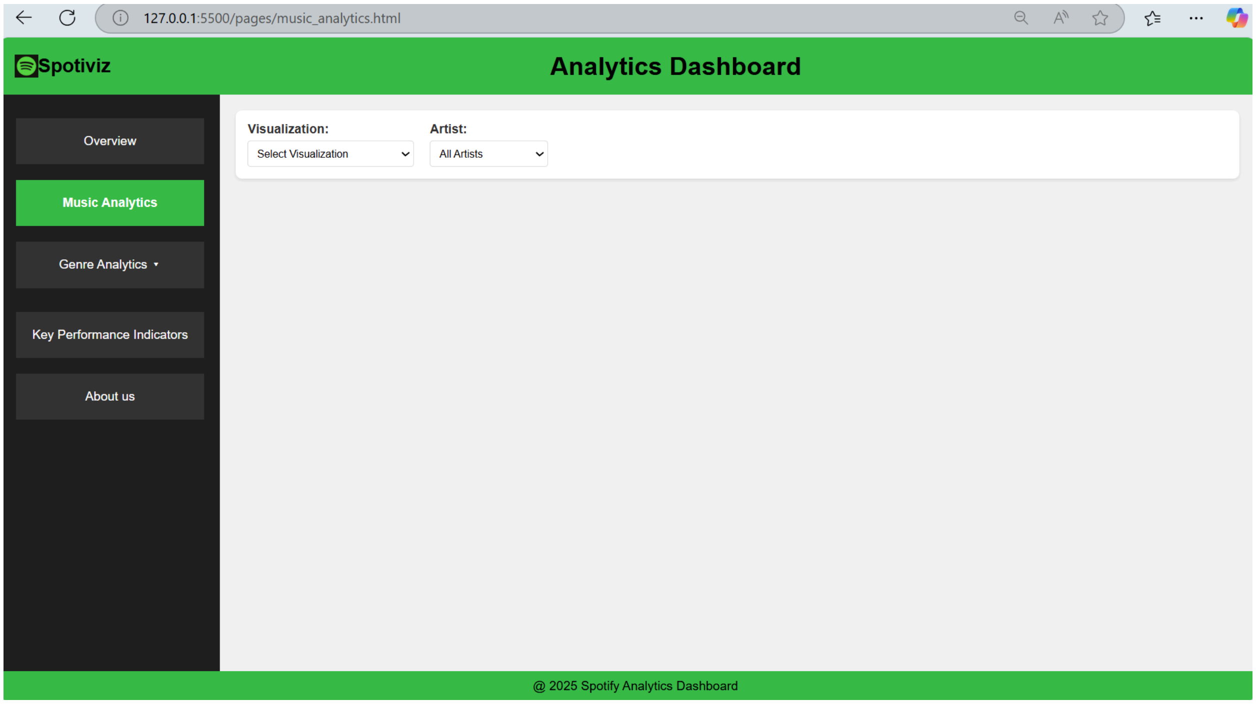
These are the webpages of Final Website implementation:

Page-1: Overview Page



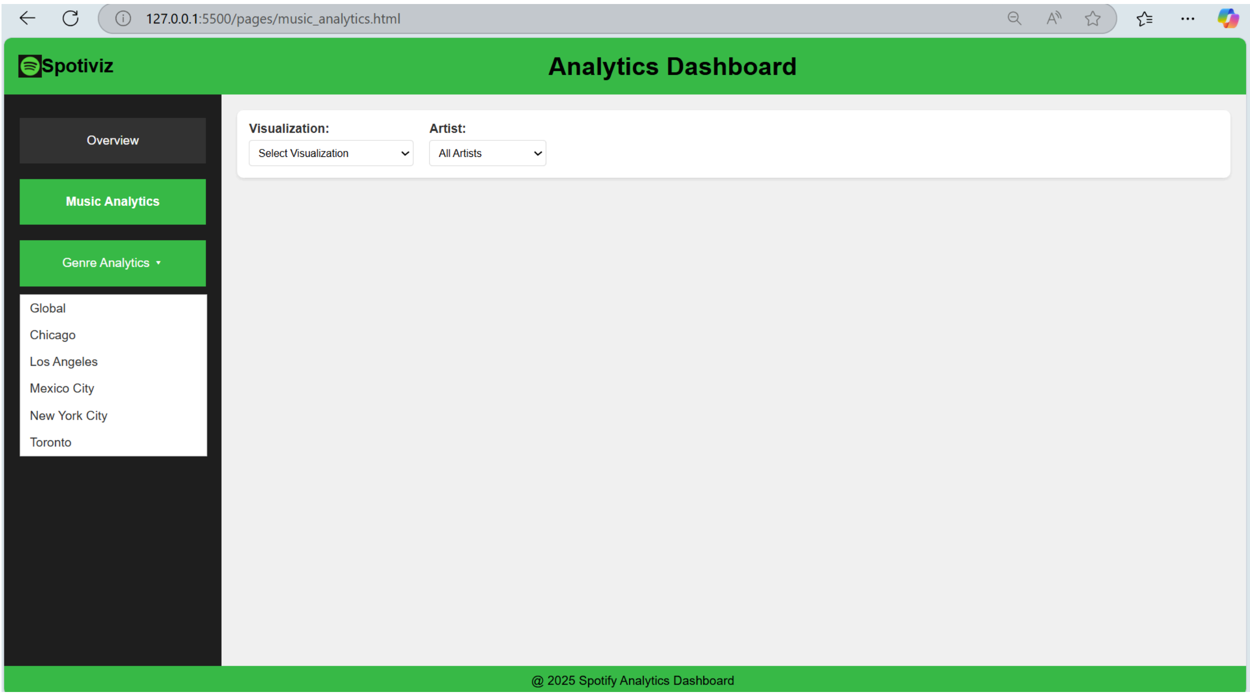
- This is the “overview” page which is the homepage of Spotiviz website which shows the Global spotify locations plotted on map.
- This is an interactive map where the countries having spotify are displayed in green and countries without spotify are displayed in grey. When it is hovered over each country, the country name along with its availability will be displayed.
- We have implemented this map visualization using d3 js.

Page-2: Music Analytics Page



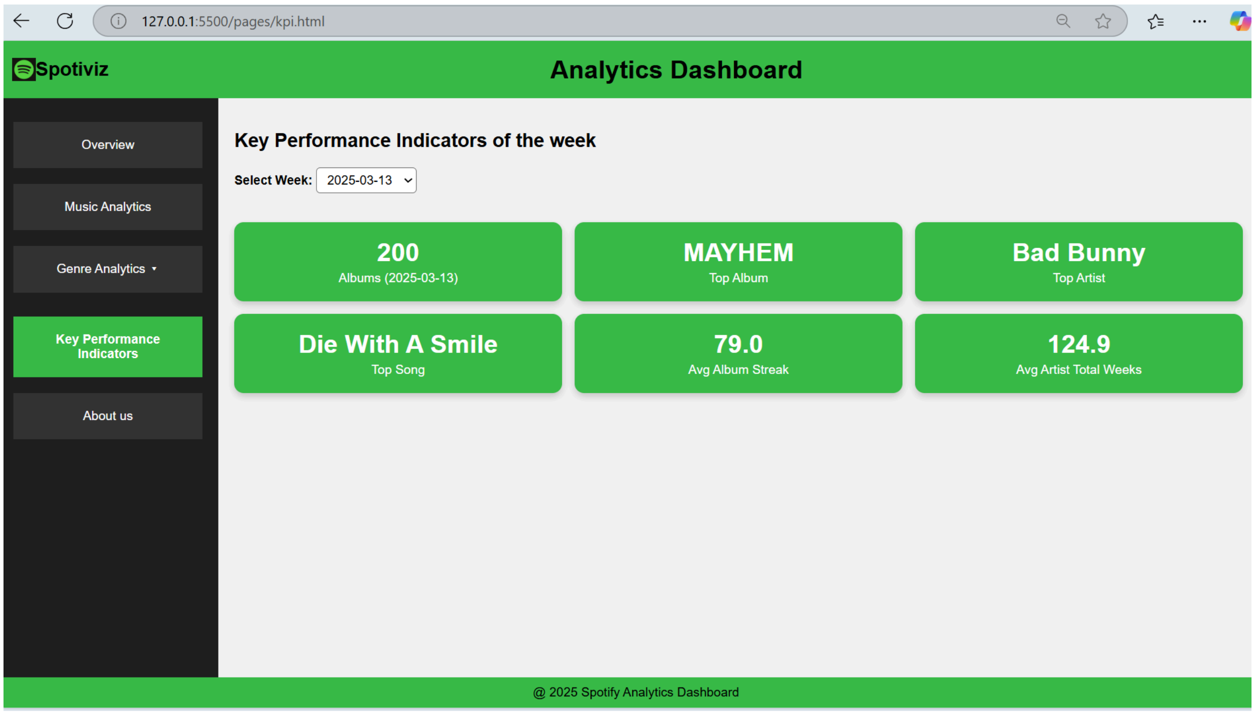
- This is the “Music Analytics” web page of the website which shows different trends and visualizations of the music data. Based on the visualization type selected using the filters, the respective visualization will be displayed.

Page-3: Genre Analytics Page



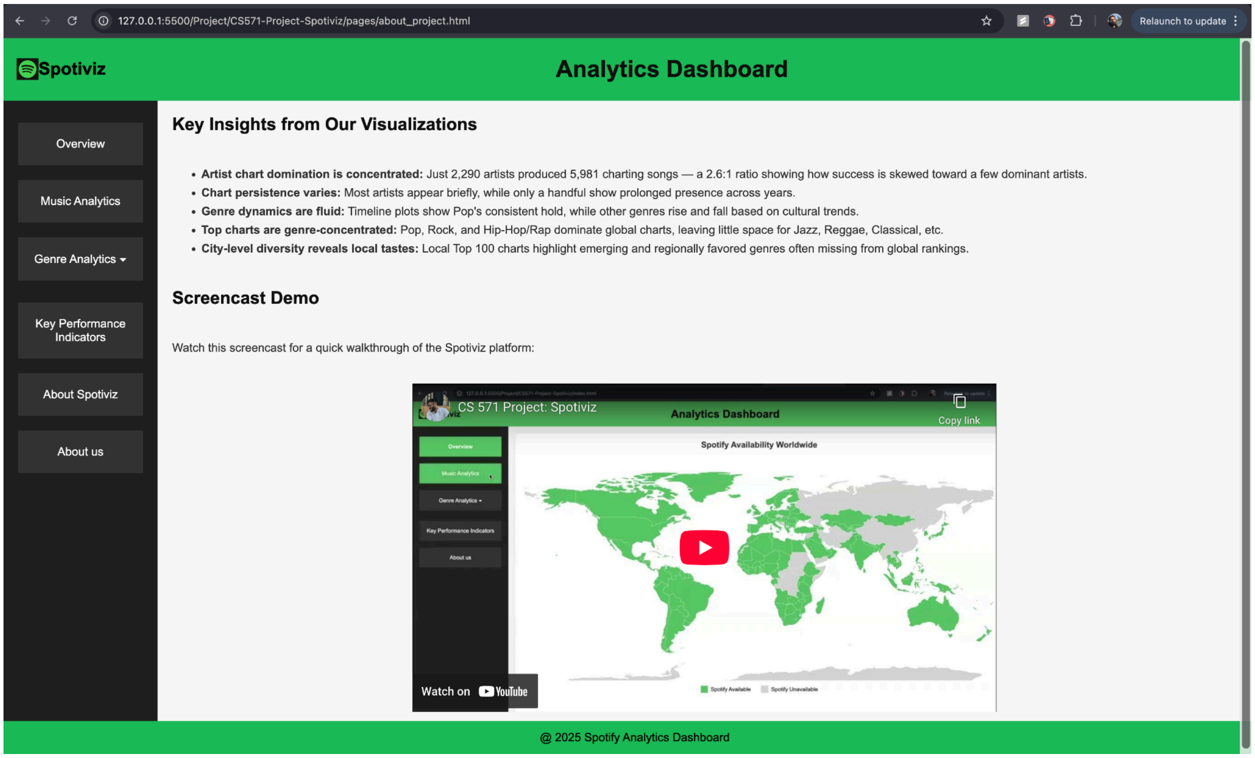
- This is the “Genre Analytics” page and there are filters given while moving on to this page itself to select based on the requirement of the user at local or global level.

Page-4: Key Performance Indicators Page



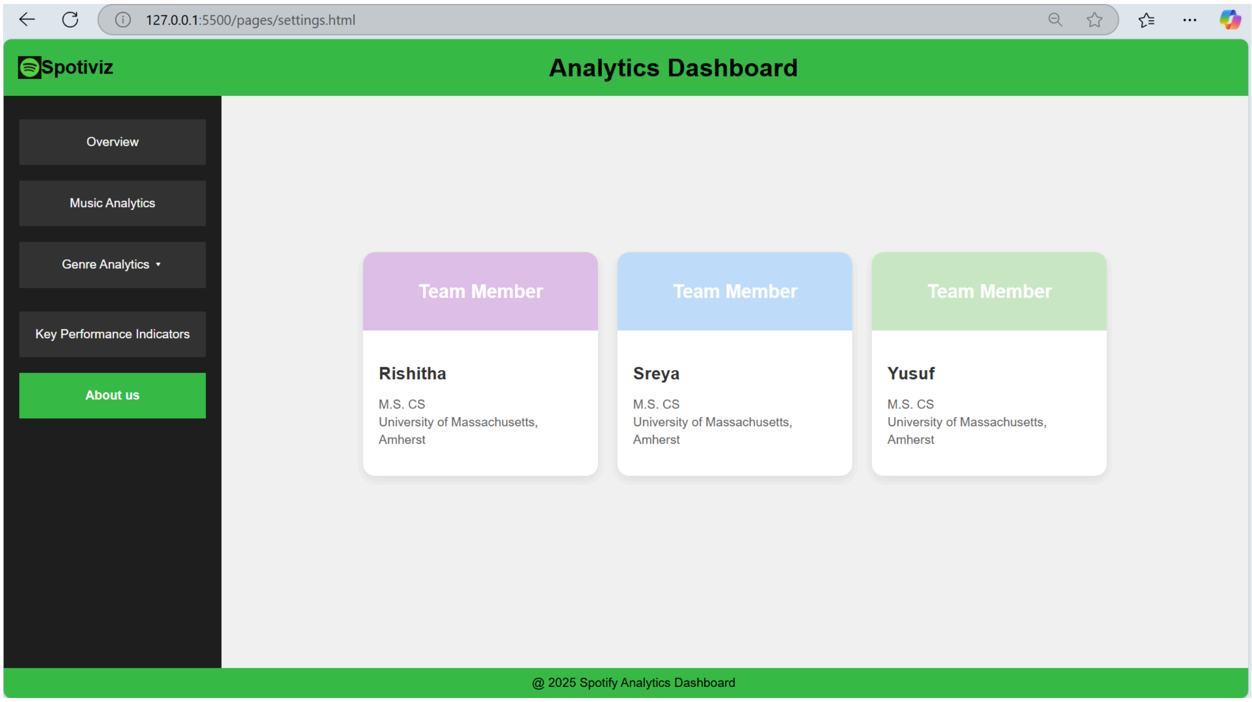
- This is the "KPI (Key performance Indicators)" page where different KPIs of artist, album and genre will be indicated on this page in distinct forms. These give a high level clarity about the entire data and major metrics from data.

Page-5: About Spotiviz



- The "About Spotiviz" page provides an overview of our motivation, key findings, embedded visualizations, screencast, and data links.

Page-6: About us



- This is the "About Us" page that we replaced the proposed settings page with. This page provides information about all project members, and each team member's card is interactive, featuring a hover effect when the cursor is moved over it.

Evaluation:

What we learned from our visualizations:

- **Artist chart domination is concentrated:** The data showed surprising concentration among top artists, with just 2,290 distinct artists accounting for 5,981 charting songs. This 2.6:1 ratio highlights how chart success is unevenly distributed, with established artists dominating charts with multiple songs simultaneously.
- **Chart persistence varies dramatically:** While examining artist longevity through our timeline visualizations, we discovered that few artists maintain consistent chart presence over extended periods. Again, it is usually a select few artists who consistently dominate the rankings.
- **Genre trends reveal dynamic shifts in popularity:**
  Our analysis of genre popularity over time—globally and locally—shows notable fluctuations in streaming patterns. Using timeline visualizations, we observed that while some genres like Pop maintain stable dominance, others experience periodic surges. The number of songs from each genre appearing in local Top 100 charts further emphasizes how regional preferences contribute to evolving genre success.
- **Top charts remain genre-concentrated:**
  The Genre Share distribution shows a clear dominance by a few genres, with Pop, Rock, and Hip-Hop/Rap collectively accounting for a majority of the top global tracks. This indicates that despite increasing diversity in music production, the charts remain concentrated, with limited representation from niche genres such as Jazz, Reggae, and Classical.
- **City-level genre diversity highlights local listening behavior:**
  By analyzing the number of unique tracks by genre across different cities, we uncovered meaningful variations in regional musical taste. While global charts reflect mainstream dominance, local Top 100 rankings surface emerging trends and genre-specific uniqueness not always visible at the global level. These insights were derived by combining local genre share data with track-level chart appearances.

Effectiveness of our visualizations:

Our visualizations succeeded in several aspects:

- The homepage presents Spotify locations on an interactive and visually engaging world map for users in a clear and communicative way.
- Interactive time range selection proved highly effective for exploring different time periods and identifying seasonal patterns.
- Rank movement visualizations clearly communicated directional changes in popularity.

- Filtering capabilities allowed for focused analysis of specific artists, songs or albums.
- Genre data provides an option to select at global or local level based on user selection.
- KPI shows all the major data of artist, album and genre for any week chosen by the user uniformly.
- Added hovering effects to most of our visualizations.

Limitations:

In the initial stages, our project aimed to explore a broad set of questions surrounding Spotify listening trends, ranging from genre evolution to mood-based patterns and podcast consumption. However, due to unforeseen limitations relating to the data we could collect, we had to realign our research questions and tone them down a bit.

For instance, while we were able to extract information for only the top 200 ranked entities for each week, we believe that the overall message and trends are still visible from our analyses.

Although we had initially thought of using mood including mood analysis of songs, we are currently restricted by data. We have, however, found an alternative to it by analysis genre data and hopefully extracting some insights from that.

Future Work:

In the future there is a scope to improve by creating a more engaging website that presents any insights in a more informative manner, if there is an opportunity to gather additional data, we have the possibility to improve it by incorporating more detailed information. This could include data reflecting various moods of individuals and unique local statistics, which can be applied across all charts, not solely limited to genres. Moreover, we could integrate other content forms beyond music, such as podcasts, with enhanced interactive features that would be visually appealing to users.

References:

[1] The Echo Nest paper, Maria Eriksson (2016)