

Visualizing the Evolution of Commercial Music

Team members: Sarthak Motwani, Milan Anand Raj, Pradumna Awasthi, Ujjwal Kumar, Abhinav Kumar Singh, Amanjit Singh, Satya Bikash Nayak, Kunal Nayak
Member emails: {sarthakm20, manandraj20, pradumna20, ujjwalk20, abhinavk20, amanjit20, satyabn20, kunalmayak20}@iitk.ac.in

IIT Kanpur

1 Introduction:

Music is like a time capsule, preserving memories, emotions, and experiences within its melodies and rhythms. It's more than just a form of entertainment; it's a reflection of our culture, society, and values. Over the years, the way we listen to music, the genres we love, and the artists we idolize have changed dramatically. From the days of vinyl records to the era of digital streaming, the musical landscape has undergone a fascinating evolution. The music industry also has a huge economic impact across the globe. It creates jobs for musicians, producers, and others involved in making and selling music. Music also makes a lot of money through things like concerts, streaming, and merchandise. It brings in tourists and helps local businesses when there are music festivals and events. Plus, it drives innovation in technology, like better sound systems and streaming services, which helps more than just the music industry. We have designed an interactive visual analytics platform that allows the users to understand the evolution of commercial music, in terms of its consumption, distribution and impact, across the globe over the years. By delving into the evolution of the musical landscape, we aim to unearth its cultural significance and economic impact. Understanding these dynamics is essential not only for acknowledging the impact of the music industry but also for anticipating future trends and innovations in the industry.

2 Tasks

We have performed the following tasks:

1. General Information about Global Music

- Visualize country-wise popular music genres globally.
- Highlight top artists from different countries, spanning both past and present.
- Showcase famous songs and albums that have left a significant mark on the global music industry.

2. Comparison between Music of the Past (1970s - 2000s) and Present (2000s - Present)

- Trace the evolution of commercial music sources, such as CDs, vinyl records, and digital platforms, in terms of units sold and revenue generated.
- Analyze major artists from the past, focusing on their commercial success, critical acclaim, and global impact.
- Identify best-selling albums and songs of the past, exploring their geographical origins and cultural significance.
- Examine popular music genres of the past across different regions.
- Repeat the analysis conducted for past music (last three points), but now focusing on the contemporary era.

3. Case Study: Cultural and Economic Impact of Two Top Artists

- For the case study of a top artist in the era of 1970s to 2000s, we choose **Michael Jackson**. We visualize Michael Jackson's top selling albums, units sold, and revenue generated over time.
- Analyze the success of Michael Jackson's tours/concerts in terms of attendance and revenue generated.
- For the case study of a contemporary top artist, we choose **Taylor Swift**. We explore the cultural and economic impact of Taylor Swift, similar to the case study on Michael Jackson.

4. Analysis of Streaming Music: Spotify's Case Study

- Investigate the rise of streaming services and their contribution to the modern music landscape.
- Analyze top songs, artists, and albums on Spotify, highlighting their geographical distribution.
- Explore characteristics of top songs on Spotify, such as danceability, energy, liveness, and instrumentality.

For the various plots made corresponding to each task, we first performed **Data Pre-processing** and then analysed this data to plot the best-suited figure based on the available data.

3 Proposed Solution

We have developed a standalone interactive visual analytics platform, in which we utilized **Plotly** and **Dash** libraries in Python for creating interactive data visualizations and embedding them into a web-based dashboard. Plotly is a powerful graphing library that enables the creation of various types of charts, while Dash allows the creation of interactive web applications using Python. Here are four key reasons for choosing Plotly Dash for our application:

1. **Interactive Visualizations:** Plotly Dash enables the creation of highly interactive visualizations, enhancing data exploration and insights.
2. **Web-Based Dashboards:** Dash allows the development of web-based dashboards, making data organized in the dashboard form accessible from a web browser.

3. **Pythonic Syntax:** With its Pythonic syntax, Dash is accessible to Python users, making it quite easy to create web applications.
4. **Integration with Plotly:** Dash seamlessly integrates with Plotly, providing a wide range of customizable charts and graphs for data visualization.

3.1 Libraries Used

1. **NumPy:** NumPy is a fundamental package for scientific computing in Python, providing support for large, multi-dimensional arrays and matrices, along with a collection of mathematical functions to operate on these arrays.
2. **Pandas:** Pandas is a powerful data manipulation and analysis library for Python, offering data structures like DataFrame and Series that simplify handling structured data and performing operations like filtering, grouping, and aggregation.
3. **Plotly:** Plotly is a Python graphing library that produces interactive, publication-quality visualizations. It offers a high-level interface to create a wide range of charts, including line plots, scatter plots, bar charts, and more.
4. **Geopandas:** Geopandas extends Pandas to handle geospatial data. It provides support for geometric operations and data manipulation with geospatial data types, allowing for the analysis and visualization of geographic datasets.
5. **Dash:** Dash is a Python framework for building web applications. It enables the creation of interactive, web-based dashboards using Python, allowing for easy integration with data analysis and visualization libraries like Plotly.
6. **Wordcloud:** Wordcloud is a library for generating word clouds from text data. It visualizes the frequency of words in a text by varying the size and color of each word based on its frequency.
7. **Matplotlib:** Matplotlib is a comprehensive plotting library for Python. It provides a wide variety of static, interactive, and animated visualizations, allowing users to create publication-quality plots with ease.
8. **Scikit-learn (sklearn):** Scikit-learn is a machine learning library for Python. It provides simple and efficient tools for data mining and data analysis, including clustering, dimensionality reduction, and more.

4 Results:

4.1 Homepage

The standalone visual analytics platform starts with the Homepage. The homepage contains a drop-down menu for selecting one of the four visualizations:

1. **Global Music Overview:** Explore country-wise music genres, top artists, and iconic songs that have shaped the global music landscape.

2. **Music: Past vs Present:** Compare the evolution of commercial music sources, major artists, best-selling albums, and popular genres between the past and present eras.
3. **Top Artists Case Studies:** Analyze the cultural and economic impact of top artists from different eras, focusing on Michael Jackson and Taylor Swift.
4. **Spotify Analysis:** Investigate the rise of streaming services, analyze top songs and artists on Spotify, and explore characteristics of popular songs.



Figure 1: Homepage

4.2 Visualization: Global Music Overview

4.2.1 Genre Popularity by Country

A choropleth map is a type of thematic map where areas are shaded or patterned in proportion to the value of a variable being represented.

This plot illustrates the distribution of music genres across the world. Each country is shaded according to the percentage of artists belonging to the chosen genre relative to all artists in that country. The dropdown menu allows users to select a genre of interest, providing insights into the genre's popularity in different regions.

This visualization helps identify the geographical spread of music genres, highlighting cultural preferences and trends worldwide.

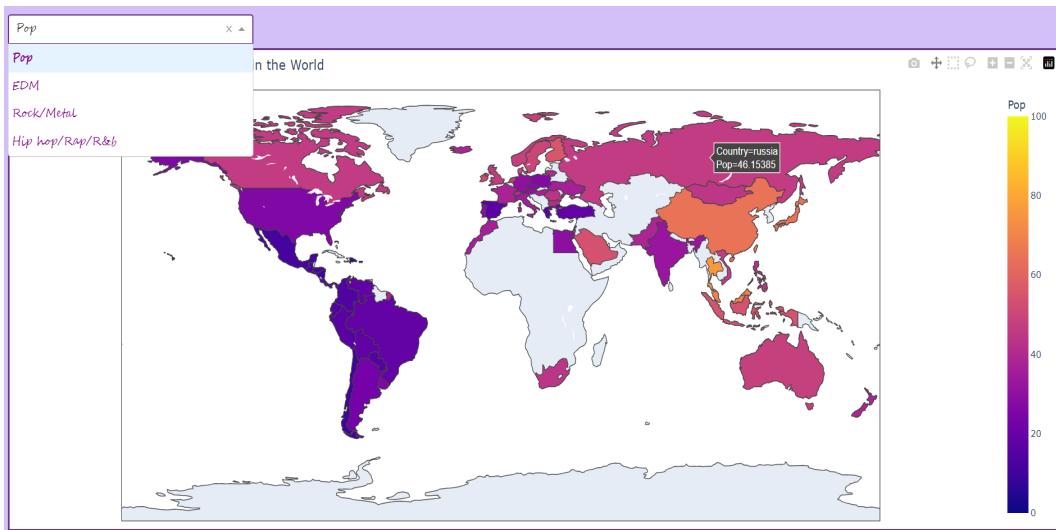


Figure 2: Genre Popularity by Country

4.2.2 Decade-wise Top Albums by Genre

This animated bar chart depicts the top albums of different genres over the years, grouped by decade. As the animation progresses decade-wise, the chart updates to display the top albums of each genre for that particular decade.

This visualization provides a dynamic representation of the most influential albums in various music genres over time.

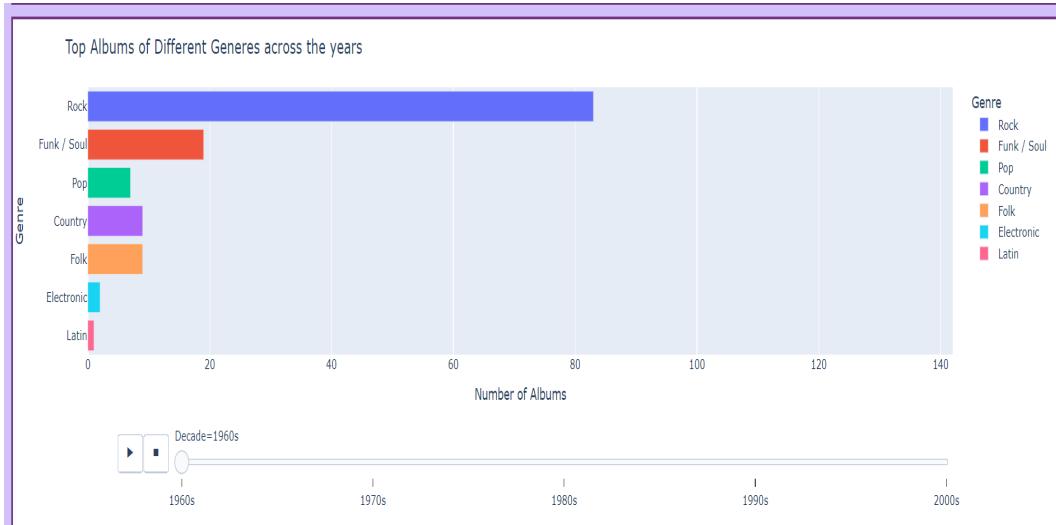


Figure 3: Decade-wise Top Albums by Genre

4.2.3 Grammy Winners Global Distribution

These Scattergeo plots display the geographical distribution of Grammy Album of the Year winners around the world and in the USA specifically because most of these winners are from USA.

It highlights the diversity of talent worldwide and offers a glimpse into the cultural influence of music from various parts of the globe.



Figure 4: Grammy Album of the Year Winners Global Distribution



Figure 5: Grammy Album of the Year Winners from the US

4.2.4 Top 10 Highest Grossing Years for an Artist

This stacked bar chart visualizes the top 10 highest grossing years for an artist based on data from Billboard 2022. The bars are stacked to represent different sources of income for the artist, including touring, streaming, sales, and publishing.

This visualization allows us to identify the most financially successful years for artists and understand the sources contributing to their earnings.

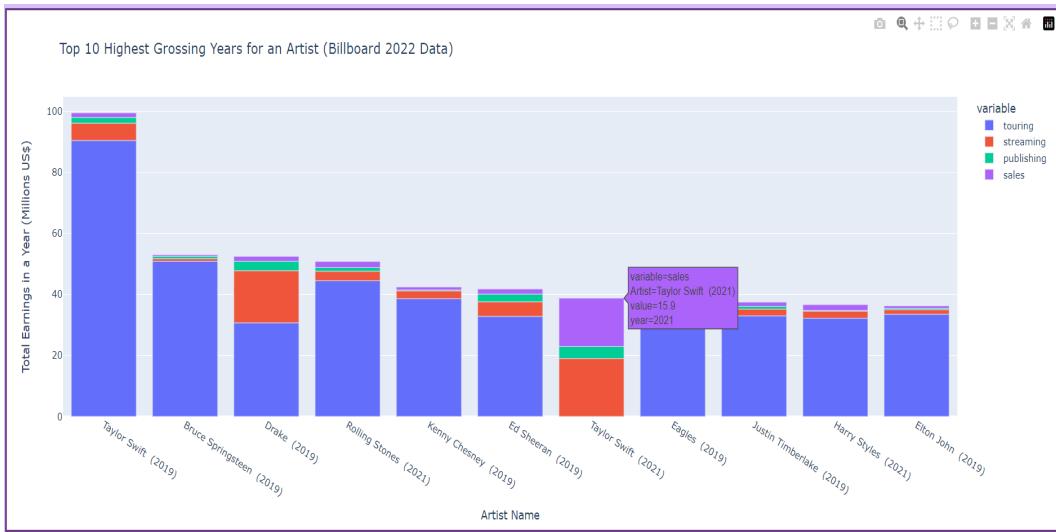


Figure 6: Top 10 Highest Grossing Years for an Artist

4.3 Visualization: Music: Past vs Present

4.3.1 Inflation Adjusted USA Music Revenues by Physical Format (2019)

This line chart illustrates the inflation-adjusted music revenues in the USA over time for different physical formats, including CDs, vinyl records, cassettes, etc.

This visualization provides insights into the historical trends of music consumption in the USA across various physical formats, revealing shifts in consumer preferences and the impact of technological advancements on the music industry.

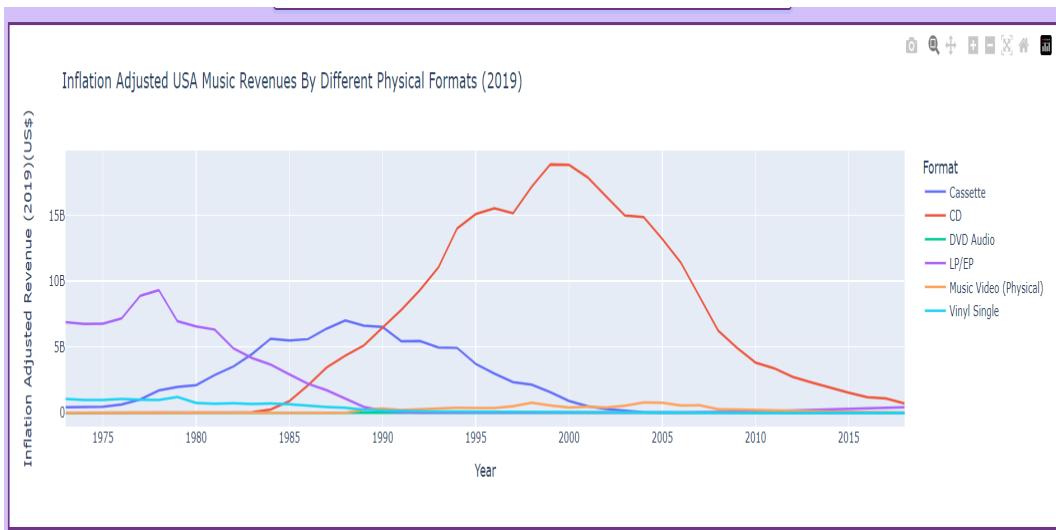


Figure 7: Inflation Adjusted USA Music Revenues by Physical Format (2019)

4.3.2 Inflation Adjusted USA Music Revenues by Digital Format (2019)

This line chart illustrates the inflation-adjusted music revenues in the USA in 2019 for different digital formats, including download albums, download singles, and on-demand streaming, etc.

This visualization highlights the revenue distribution among various digital music formats

in 2019. It provides insights into consumer preferences for accessing music digitally and shows how different formats contribute to the overall music revenue.

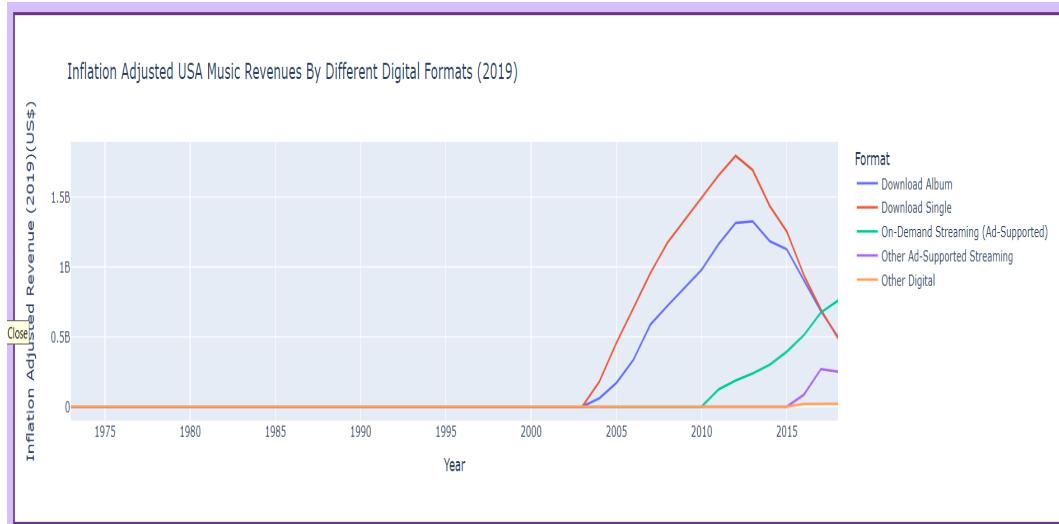


Figure 8: Inflation Adjusted USA Music Revenues by Digital Format (2019)

4.3.3 Spotify Monthly Active and Premium Users Over Time

This line chart illustrates the number of Spotify monthly active users and premium users over the years.

This visualization provides insights into the growth of Spotify's user base over time, distinguishing between free and premium users. It can indicate trends in digital music consumption habits.



Figure 9: Spotify Monthly Active and Premium Users Over Time

4.3.4 Highest Earning Musician Yearly Earnings (1987 - 2021)

This line chart displays the yearly earnings of the highest-earning musician from 1987 to 2021.

It highlights trends in the music industry's economic landscape over time.

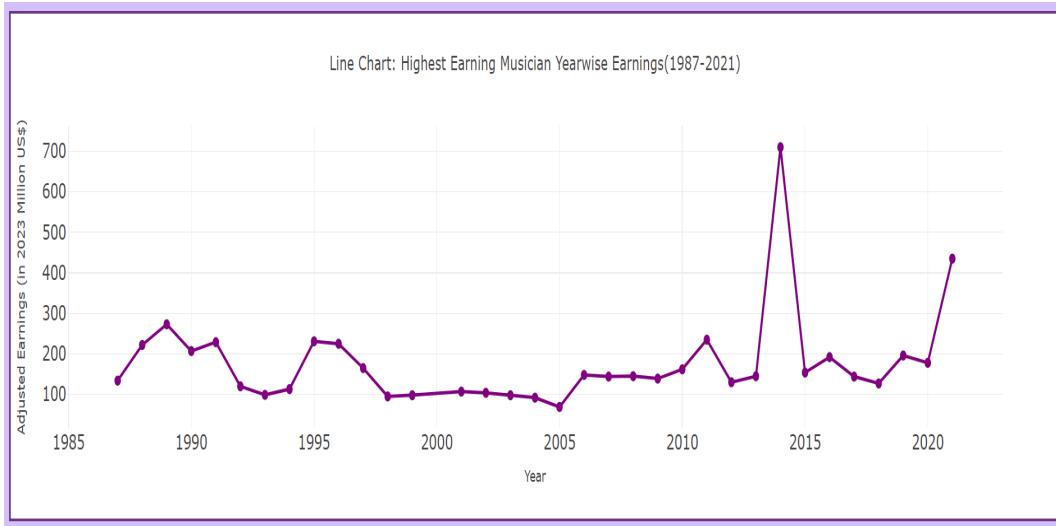


Figure 10: Highest Earning Musician Yearly Earnings (1987 - 2021)

4.3.5 Highest Earning Musician Year wise Earnings Distribution (1987-2021)

These two box plots visualize the distribution of yearly earnings for the highest-earning musician from 1987 to 2021. The box represents the interquartile range (IQR) of earnings, with the median value indicated by the line inside the box. Since there are two values of earnings which are very high as compared to the highest earnings for remaining years, the second box plot is drawn after excluding these values.

This visualization provides insights into the variability and distribution of earnings for the highest-earning musician over the specified period.

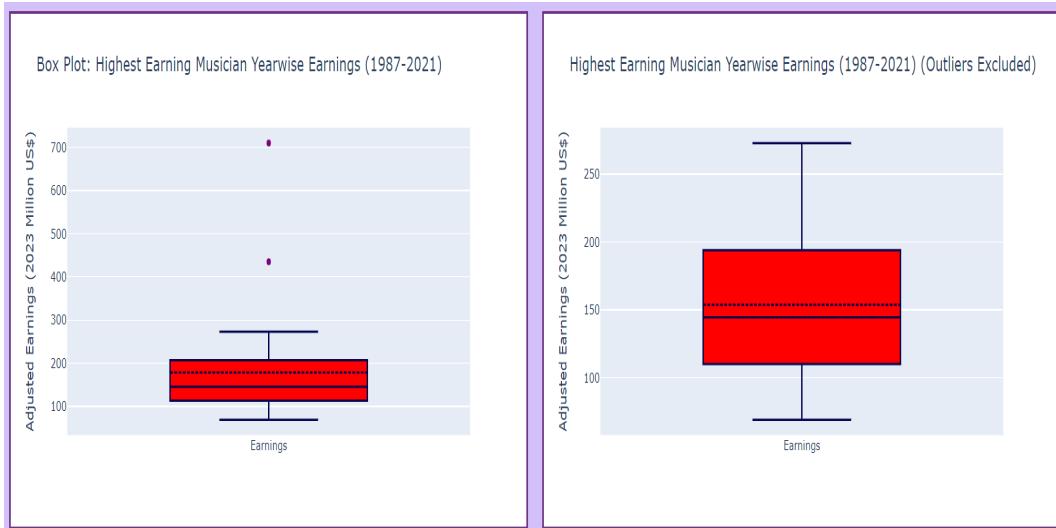


Figure 11: Highest Earning Musician Year wise Earnings Distribution (1987-2021) (Outliers Excluded)

4.3.6 Proportion of Top Albums by Genre

These two pie charts show the proportion of different genres among the top albums. The left chart represents albums from 1987 to 2000, while the right chart represents albums from 2001 to 2020.

These visualizations offer insights into the changing landscape of music genres among top albums over the years.

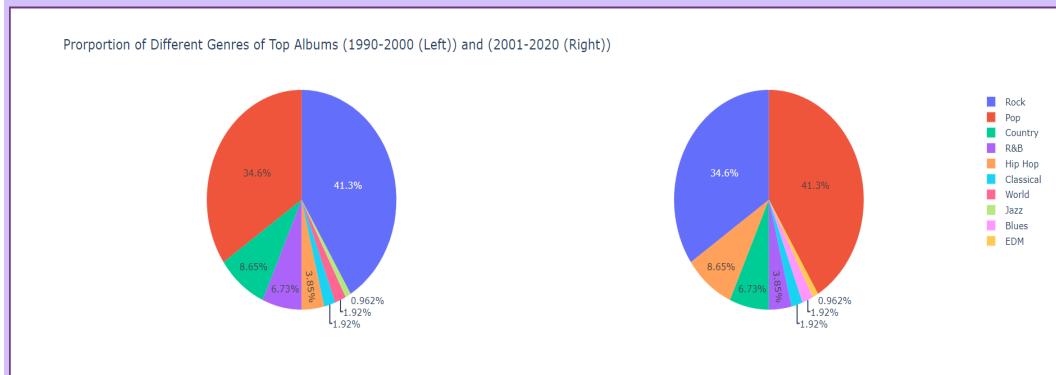


Figure 12: Proportion of Top Albums by Genre

4.3.7 Spotify Genre Word Clouds (1980s vs. 1990s-2010s)

These two word clouds display the frequency of music genres in Spotify data, with one representing the 1980s and the other representing the combined years from the 1990s to the 2010s.

By comparing the word clouds, we can observe changes in genre popularity and trends in music consumption across the decades.

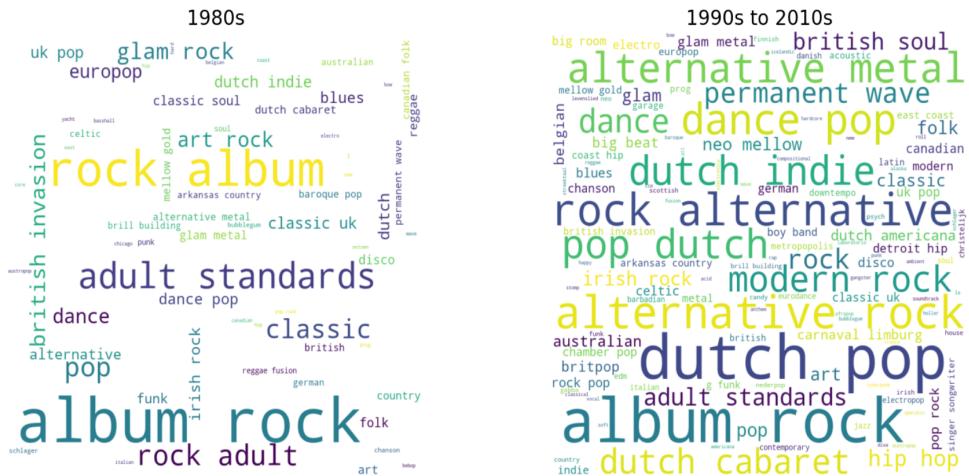


Figure 13: Spotify Genre Word Clouds (1980s vs. 1990s-2010s)

4.3.8 Top Selling Artists and Albums Treemaps

A treemap is a type of hierarchical visualization that displays hierarchical data using nested rectangles.

These two treemaps visualize the top-selling artists and their best-selling albums. The hierarchy starts with the artist, followed by the album, and upon hovering, the sales for that album and its genre are displayed. The first treemap covers albums from the 1990 to 2000, while the second treemap covers albums from 2001 to 2020.

They allow viewers to explore the sales distribution among different artists and albums

and understand the contribution of each album to an artist's overall sales. Additionally, the treemaps offer insights into genre preferences among top-selling albums across the decades.

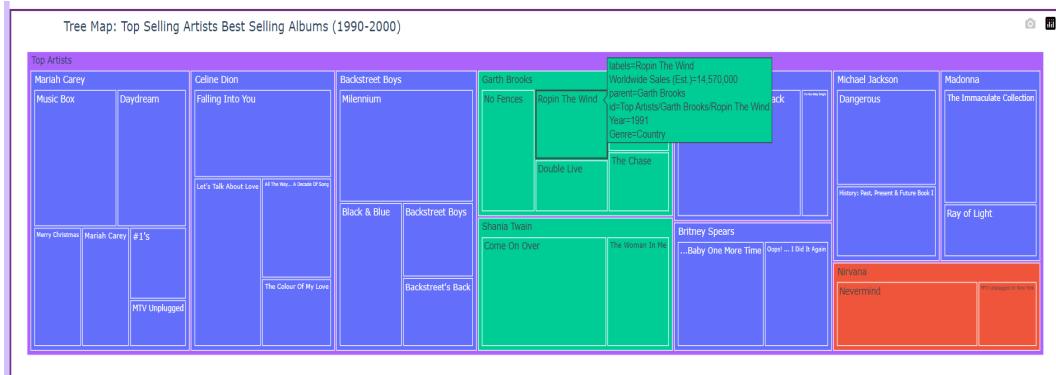


Figure 14: Top Selling Artists and Albums Treemaps (1990-2000)



Figure 15: Top Selling Artists and Albums Treemaps (2001-2020)

4.3.9 Average Song Duration Over Time (1999 - 2019) (Spotify Data)

This line chart illustrates the average duration of songs released over the years from Spotify data, spanning the years 1999 to 2019.

This visualization provides insights into the trend of song duration in the music industry over two decades. This chart reveals that trends in shorter songs are more prevalent.



Figure 16: Average Song Duration Over Time (1999 - 2019) (Spotify Data)

4.3.10 Average Album Duration of Top Albums Over Time (1990 - 2020)

This scatter plot with a trend line visualizes the average duration of top albums yearwise from 1990 to 2020.

By observing the trend line, we can identify that album durations have decreased on an average over time.

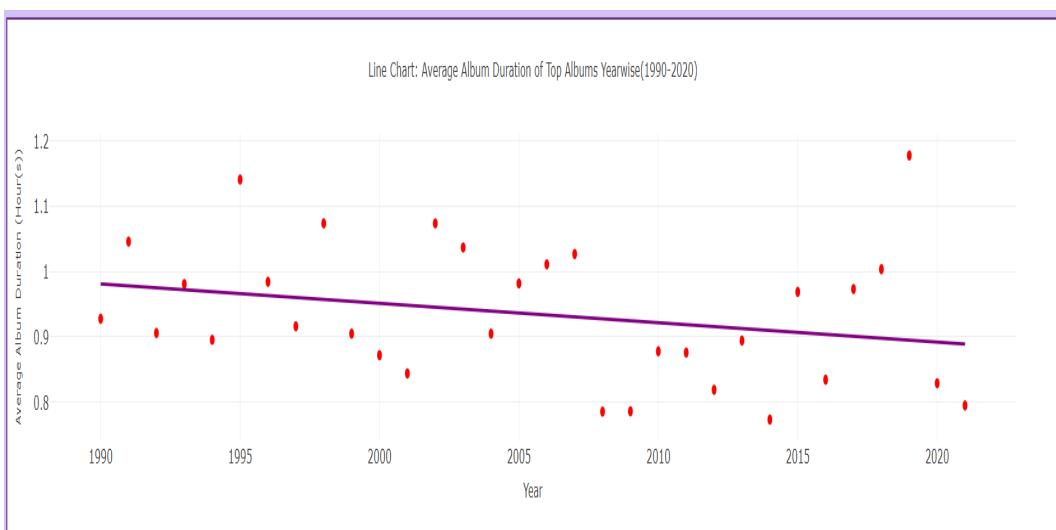


Figure 17: Average Album Duration of Top Albums Over Time (1990 - 2020)

4.3.11 Album Duration vs. Worldwide Album Sales Histogram (1990 - 2020)

This histogram plot visualizes the distribution of album duration on the x-axis and the total number of worldwide album units sold for each duration bin.

This visualization allows us to understand the relationship between album duration and album sales. By observing the distribution of album sales across different duration bins, we can identify patterns in consumer preferences that albums having duration of between 40 minutes to 1 hour have the largest sales.

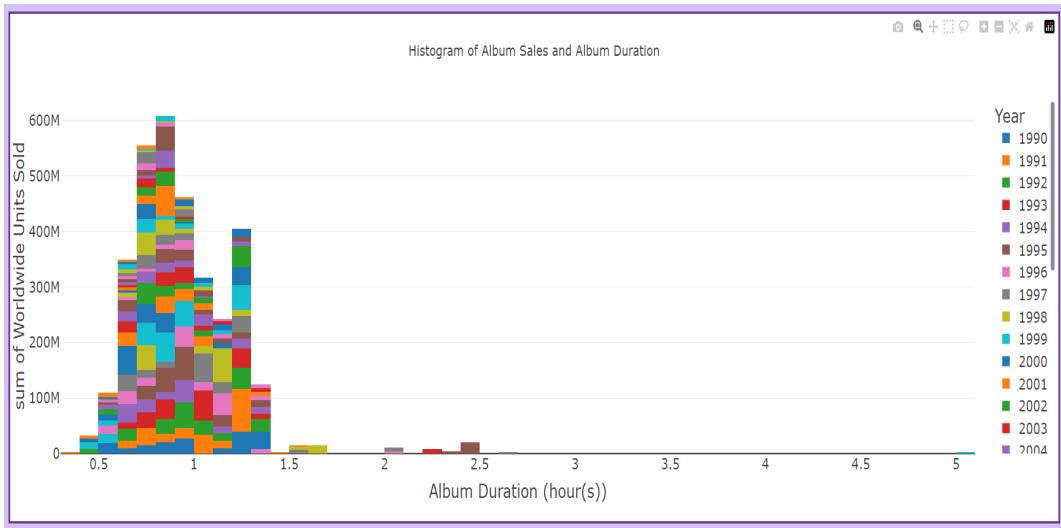


Figure 18: Album Duration vs. Worldwide Album Sales Histogram (1990 - 2020)

4.4 Top Artists Case Studies

4.4.1 Countrywise Album Sales of Taylor Swift and Michael Jackson

These choropleth plots compare the countrywise album sales of Taylor Swift (contemporary artist) and Michael Jackson (top artist from the past). Each plot represents album sales in different countries, with colors indicating the level of sales intensity.

These plots highlight the geographical regions where each artist's music has made the most significant impact.

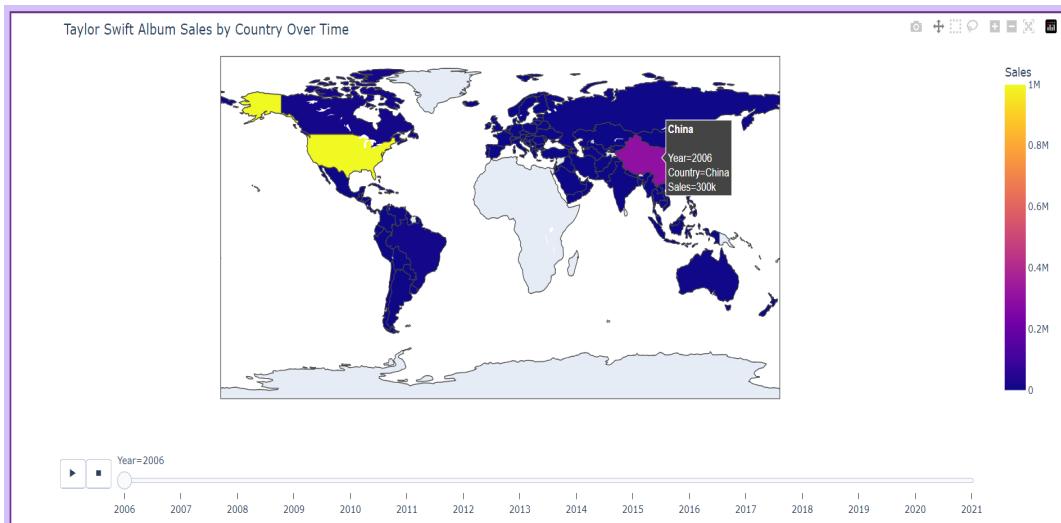


Figure 19: Countrywise Album Sales of Taylor Swift over the Years

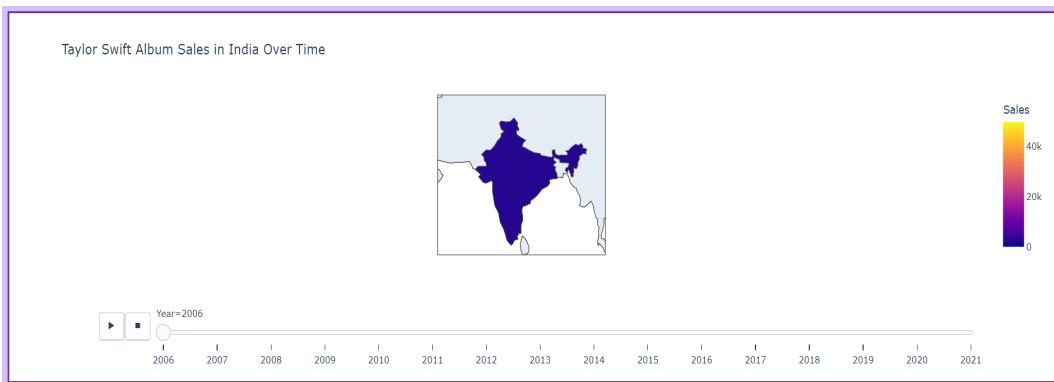


Figure 20: Sales of Taylor Swift Albums in India over the years

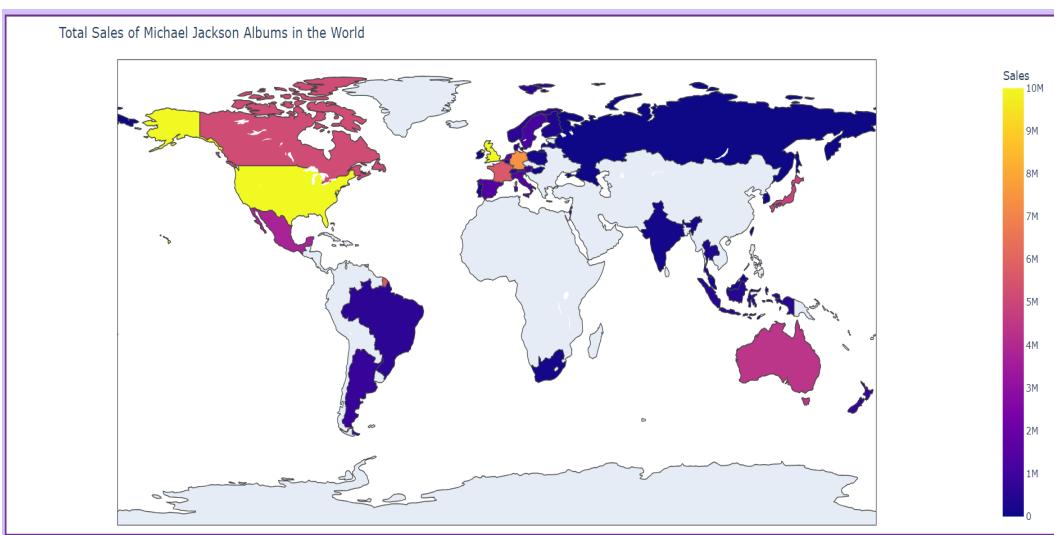


Figure 21: Countrywise Total Album Sales of Michael Jackson

4.4.2 Grouped Bar Chart: Total Album Sales Comparison

This grouped bar plot compares the total album sales of Taylor Swift and Michael Jackson in each country. For each country, there are two bars: one representing the total album sales of Taylor Swift, and the other representing the total album sales of Michael Jackson. By examining the height of the bars, viewers can quickly understand which artist has sold more albums in each country and identify any geographical patterns in their popularity. This comparison helps in understanding the global impact and popularity of both artists across different regions.

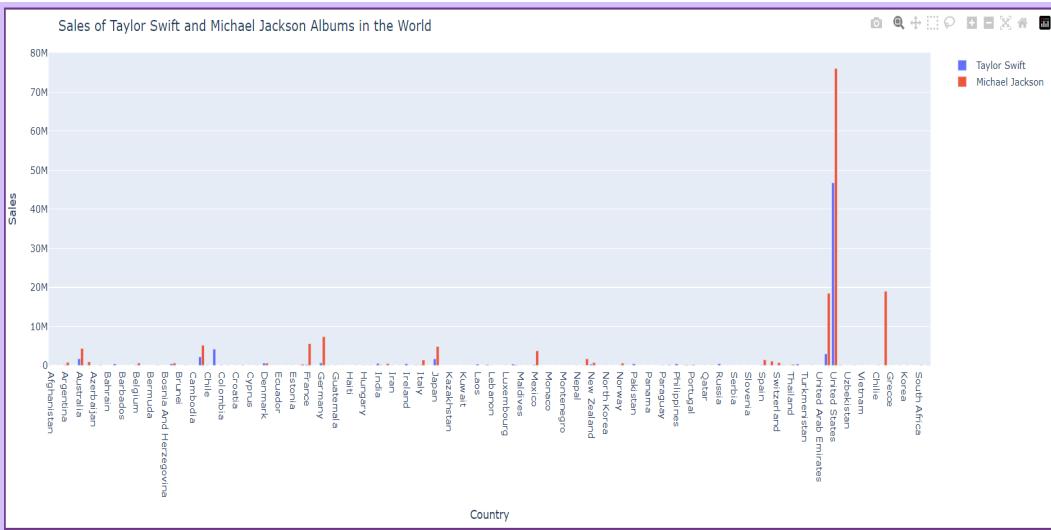


Figure 22: Grouped Bar Chart: Total Album Sales Comparison

4.4.3 Global Comparison of Album Sales

This choropleth plot displays a global comparison of album sales between Michael Jackson and Taylor Swift. Countries marked with red indicate those where Michael Jackson has sold more albums, while countries marked with blue are where Taylor Swift has sold more albums up to the present.

This visualization offers a clear overview of the global dominance of Michael Jackson and Taylor Swift in different countries in terms of album sales.

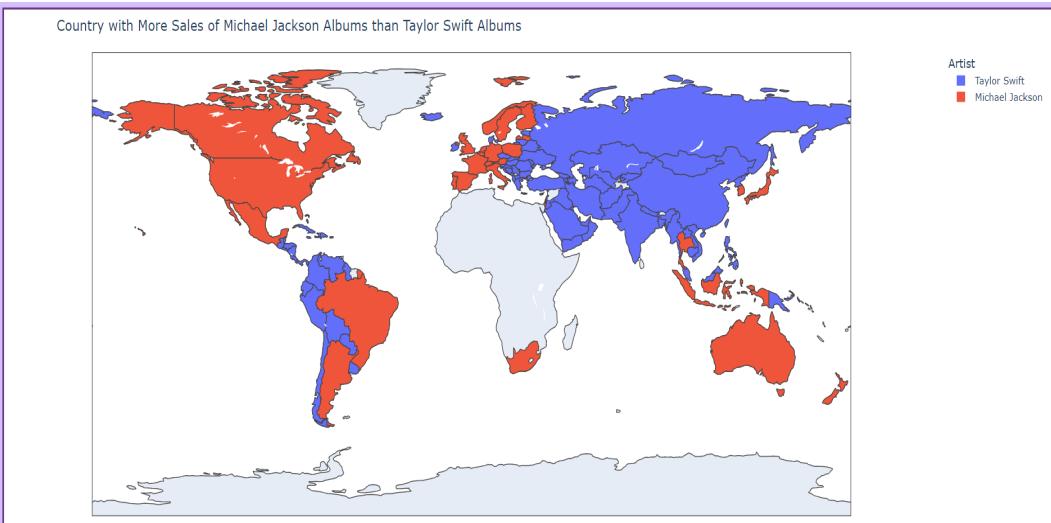


Figure 23: Global Comparison of Album Sales

4.4.4 Taylor Swift's Concert Revenue by Country

This choropleth visualization showcases Taylor Swift's concert revenue by country on a globe. Each country is color-coded based on the revenue generated from Taylor Swift's concerts.

This visualization provides valuable insights into the geographical distribution of Taylor Swift's concert revenue. It allows event organizers and promoters to identify regions

where Taylor Swift's concerts are particularly lucrative, aiding in strategic planning for future tours.

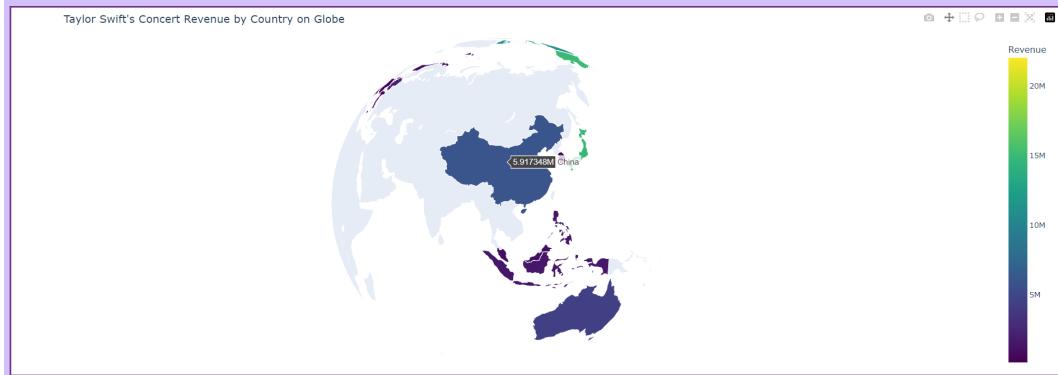


Figure 24: Taylor Swift's Concert Revenue by Country

4.5 Concert Attendance of Taylor Swift and Michael Jackson

These two choropleth visualizations depict the concert attendance of Taylor Swift and Michael Jackson by country. Users can select a specific concert from a dropdown menu, and the choropleth dynamically updates to display the attendance for that concert in each country.

These visualizations offer insights into the popularity and global reach of Taylor Swift and Michael Jackson's concerts.

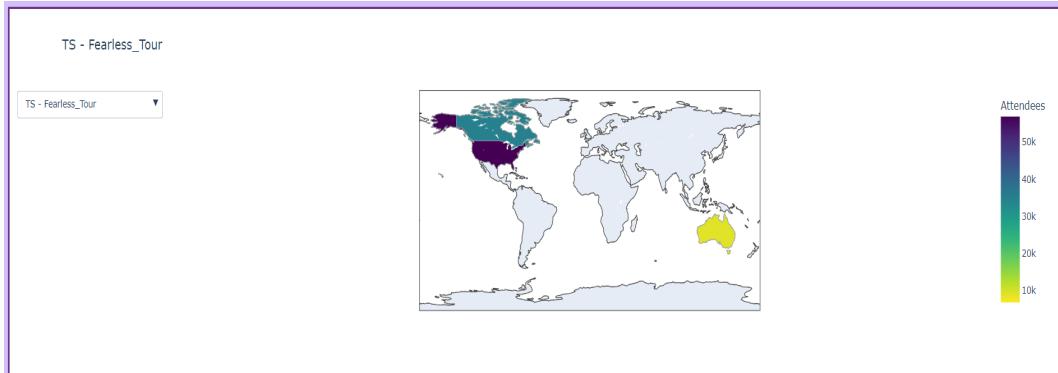


Figure 25: Taylor Swift's Concert Attendance of Different Tours

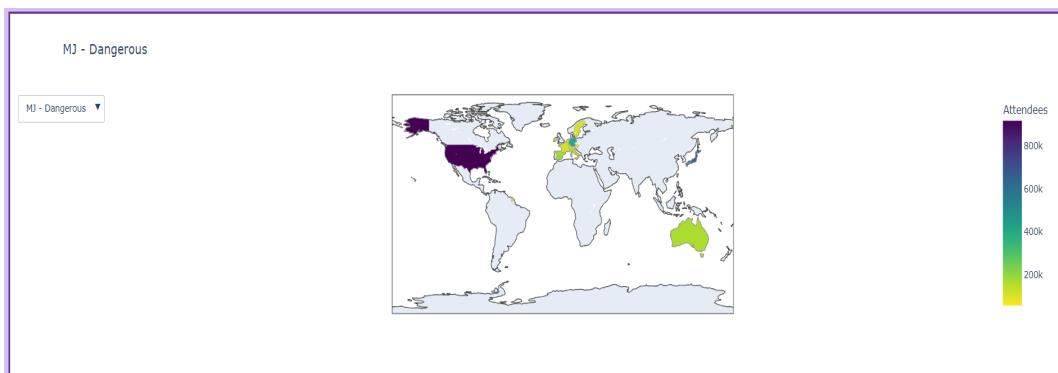


Figure 26: Michael Jackson's Concert Attendance of Different Tours

4.6 Spotify Analysis

4.6.1 Spotify Monthly Active and Premium Users Over Time

This grouped bar chart illustrates the trend of Spotify monthly active users and premium users over the years.

This visualization provides a comparison of the growth of Spotify's user base and premium subscribers over time. This comparison helps in assessing Spotify's market penetration in retaining users.



Figure 27: Spotify Monthly Active and Premium Users Over Time

4.6.2 Spotify Users by Region Over Time

This area chart displays the number of Spotify users by region over the years. Each region (Europe, North America, Latin America, and Rest of World) is represented by a colored area, and the combined areas represent the total number of Spotify users.

This visualization offers insights into the growth of Spotify's user base across different regions over time.

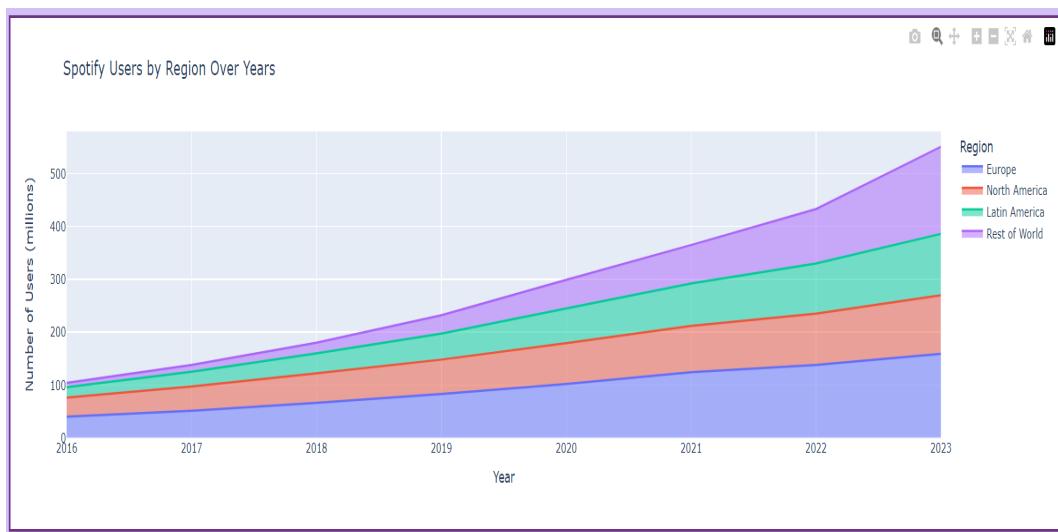


Figure 28: Spotify Users by Region Over Time

4.6.3 Genre-wise Popularity Distribution

This histogram distribution plot illustrates the distribution of popularity for tracks belonging to a selected genre. Users can use a dropdown menu to select the genre of interest, and the plot dynamically updates to show the distribution of popularity scores for tracks in Spotify data within that genre.

This information is valuable for music enthusiasts, artists, and industry professionals to understand the popularity trends within specific music genres.

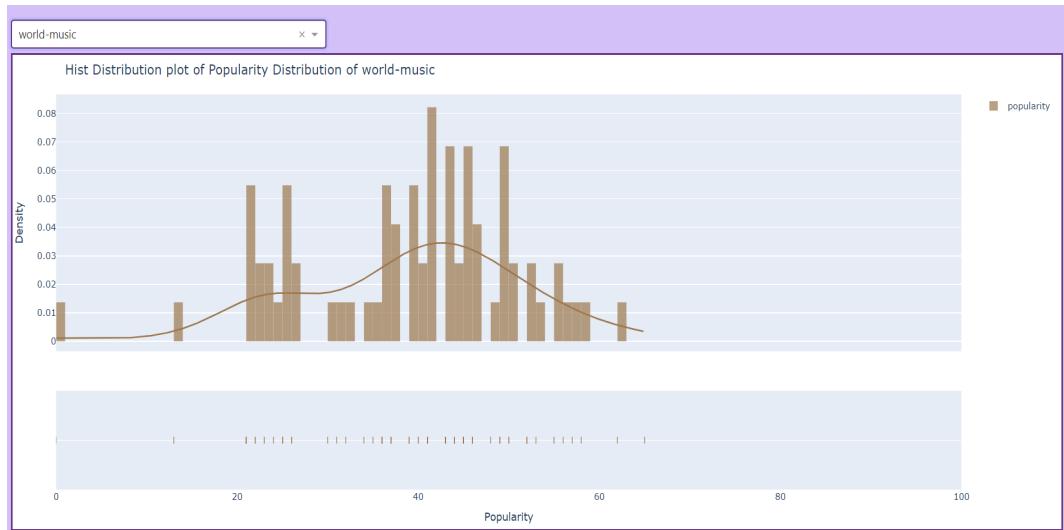


Figure 29: Genre-wise Popularity Distribution

4.6.4 Genre-wise Distribution of Song Characteristics

This histogram distribution plot illustrates the distribution of various characteristics of songs within a selected genre. Users can use a dropdown menu to choose the genre they are interested in, and the plot dynamically updates to display the distribution of characteristics such as valence, energy, liveness, instrumentality, etc.

Understanding these distributions helps in analyzing the musical traits that define each genre and can aid in music production, recommendation systems, and genre classification.

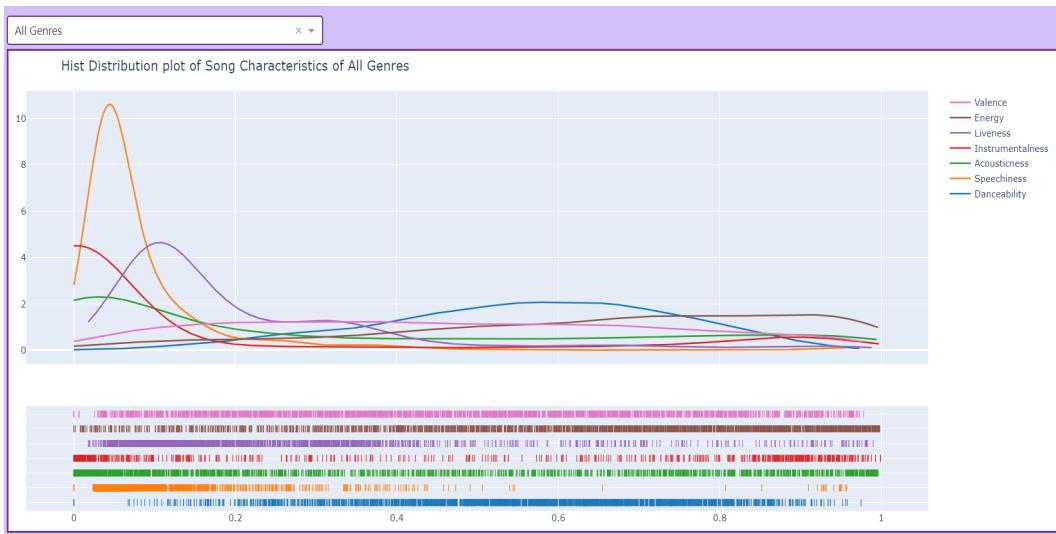


Figure 30: Genre-wise Distribution of Song Characteristics

4.6.5 Correlation Heatmap of Song Characteristics

This correlation heatmap illustrates the correlation between various song characteristics such as instrumentalness, energy, valence, acousticness, etc.

This visualization helps in understanding the relationships between different musical traits. By analyzing the correlation heatmap, users can identify which characteristics tend to co-occur in songs. Understanding these correlations is valuable for music analysis, production, and recommendation systems.



Figure 31: Correlation Heatmap of Song Characteristics

4.6.6 Popularity of Top Artists

This bubble plot displays the popularity of the top 10 artists along with their names. This plot helps in understanding the relative popularity of different artists in Spotify.

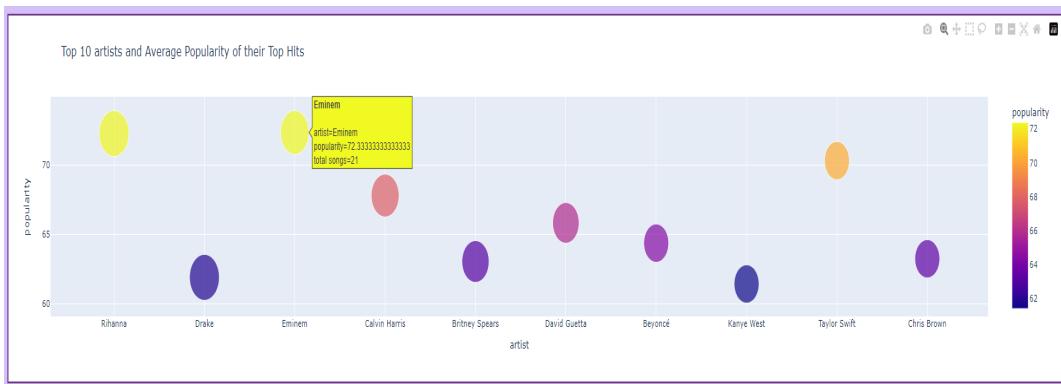


Figure 32: Popularity of Top Artists

4.6.7 Top 15 Artists with Most Top Songs in Spotify (2023)

This bar chart displays the top 15 artists with the most top songs in Spotify for the year 2023.

This visualization highlights the artists who have consistently produced hit songs and maintained a strong presence on Spotify in 2023.

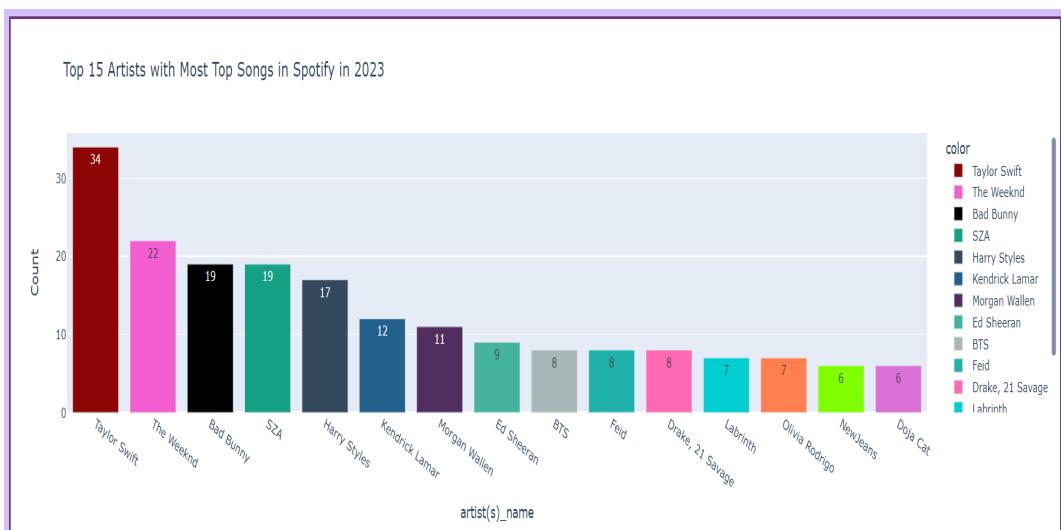


Figure 33: Top 15 Artists with Most Top Songs in Spotify (2023)

4.6.8 Percentage of Artists from Different Countries in Spotify

This choropleth world map illustrates the percentage of artists from different countries in Spotify. Each country is shaded according to the percentage of artists it contributes to Spotify's artist pool.

This visualization provides insights into the global representation of artists on Spotify.

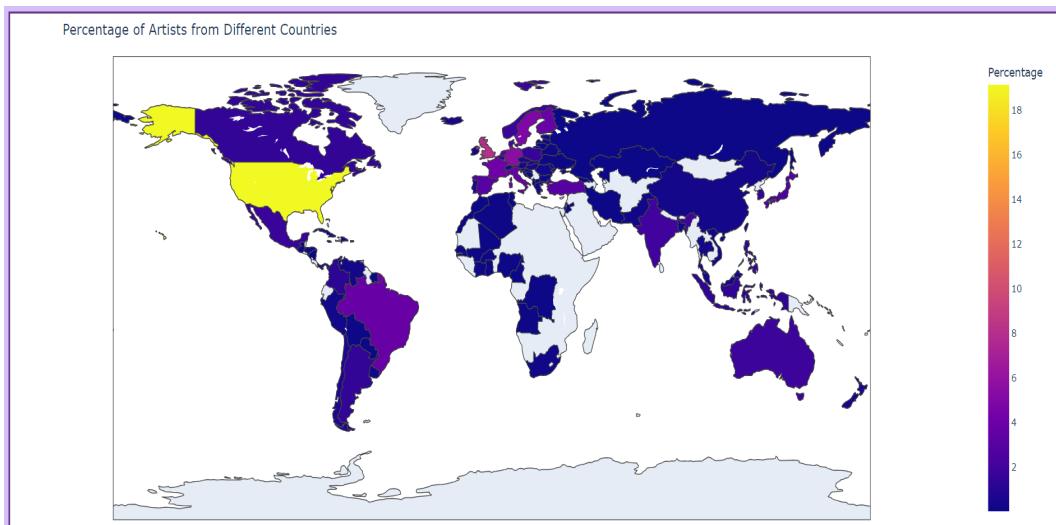


Figure 34: Percentage of Artists from Different Countries in Spotify

4.6.9 Genre Popularity and Top Artists in Selected Country

This visualization consists of two components. Firstly, there is a pie chart showing the popularity of different genres in a selected country. Users can choose a country from a dropdown menu, and the pie chart dynamically updates to display the distribution of genre popularity in that country. Secondly, there is a table displaying the top 5 popular artists from Spotify in the selected country.

The pie chart allows users to understand the genre preferences of listeners in a specific country, helping to identify the dominant music genres. Meanwhile, the table provides additional context by showcasing the top artists contributing to the popularity of those genres.

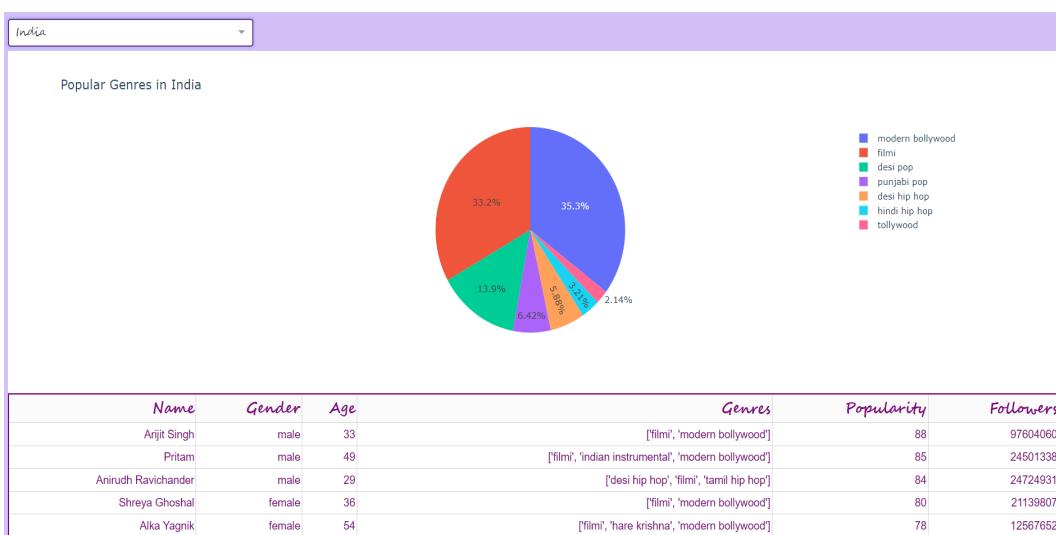


Figure 35: Genre Popularity and Top Artists in Selected Country

4.6.10 A small Recommendation System

We've implemented a recommendation system using k-means clustering to organize our music library into distinct clusters based on attributes like loudness, energy, and dance-

ability. Each cluster highlights the top 5 tracks, showcasing diverse playlists tailored to user preferences. Visual aids like radar maps illustrate the unique features of each cluster, enhancing user understanding and engagement by comparing the musical attributes visually. Additionally, genre information for each track deepens user insight into the musical styles present, promoting new music discovery that aligns with individual tastes. This approach exemplifies how modern music streaming apps leverage machine learning techniques to deliver personalized, dynamic recommendations, improving user experience and interaction in the digital music landscape.

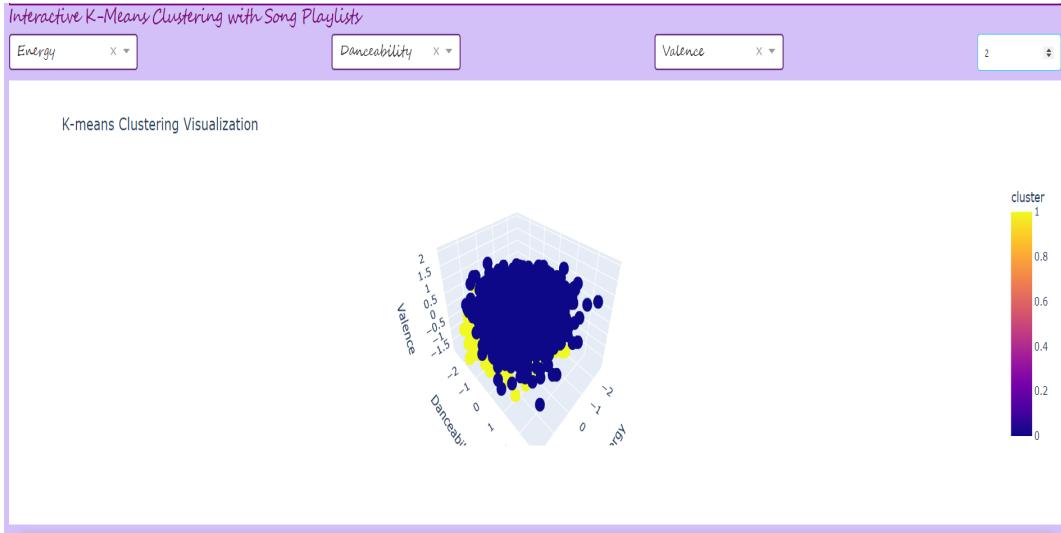


Figure 36: K-Means Clustering Visualization



Figure 37: Playlists Corresponding to Clusters

5 Conclusion:

In conclusion, our visual analytics system provides a comprehensive exploration of the evolution, trends, and preferences in the global music landscape. Through a diverse

range of visualizations, we have delved into various aspects of music consumption, artist popularity, and genre preferences, utilizing Spotify data spanning several decades.

We began by analyzing general trends in global music and then examined the evolution of music genres and top artists, showcasing the changing dynamics from the 1980s to the present. Our system highlighted the global impact of iconic artists like Michael Jackson and Taylor Swift, while also uncovering emerging trends in contemporary music. We also observed that digital forms of music consumption like streaming are more popular today than physical forms like CDs, vinyls.

Furthermore, we examined the economic and cultural impact of music through detailed analyses of album sales, revenue, streams of top-selling artists. By visualizing the distribution of album sales across different countries and regions, we gained insights into the global popularity of various artists and genres.

Our system also provided interactive visualizations for exploring song characteristics and genre preferences. Users could analyze the distribution of song characteristics within genres, understand correlations between different traits, and explore the popularity of specific genres in different countries.

Additionally, our system facilitated comparative analyses between artists, genres, and countries, allowing users to understand trends, patterns, and relationships in the music industry. Whether it's comparing the popularity of top artists, exploring genre preferences by region, or understanding the characteristics of hit songs, our system offers a versatile toolkit for music enthusiasts, industry professionals, and researchers.

In conclusion, our visual analytics system provides valuable insights into the multi-faceted world of music, empowering users to explore, understand, and appreciate the rich tapestry of musical diversity and creativity.

6 Link to source code:

- Click here to see our github repository: CS661 *Evolution of Music*
- Install all the requirements and run **code.ipynb** to interact with the application .

7 Data Sources

We have chosen the following datasets from **kaggle**:

1. *Top Music Genres in different countries*
2. *Rolling Stone's List of Top 500 Music Albums of all Time*
3. *Top Music Artists according to Grammys from 1960 to 2018*
4. *Music Industry Sales (Vinyls to Streaming) from 1973 to 2019*
5. *US Recorded Music Revenue of different formats from 1973 to 2019*
6. *Best Selling Albums globally from 1990 to 2021*
7. *Michael Jackson's Album Sales (from bestsellingalbums.org)*
8. Michael Jackson's Concert Attendance

9. *Taylor Swift's Album Sales*
10. *Taylor Swift's Concert Tours' Attendance and Revenue generated*
11. *Spotify's Top Songs Comprehensive Dataset (1960 - Present)*
12. *Most Streamed Songs of Spotify in 2023*
13. *Spotify - All Time Top 2000s Mega Dataset*

8 Individual Contributions

Member	Contribution
Sarthak Motwani	Fig. 2 (with dropdown), Fig. 4, Fig. 5, Fig. 6 Fig. 7, Fig. 8, Fig. 10, Fig. 11, Fig. 12 Fig. 14, Fig. 15, Fig. 17, Fig. 18, Fig. 29 Integrated Plots with Plotly-Dash, made Frontend, made report
Milan Anand Raj	Fig. 3, Fig. 4, Fig. 5, Fig. 13, Fig.17 Fig. 24 ,Fig. 30 (with drop-down), Integrated Plots with Plotly-Dash
Ujjwal Kumar	Fig. 16, Fig. 29, Fig.31, Fig. 30 Fig. 32, Fig. 6
Satya Bikash Nayak	Fig. 9, Fig. 11, Fig. 27, Fig. 28, Fig. 31 Fig. 33, Fig. 34, Fig. 35 (with table and dropdown)
Kunal Nayak	Fig. 27, Fig. 36, Fig. 37 (k-means clustering and a small recommendation system)
Abhinav Kumar Singh	Fig. 19, Fig. 20, Fig. 21, Fig. 22 Fig. 29, Fig. 23,
Pradumna Awasthi	Fig. 13, Fig. 22, Fig. 25 (with dropdown) , Fig. 26 (with dropdown)
Amanjit Singh	Fig. 6, Fig. 14, Fig. 15, Fig. 13, Fig. 19

Note: Fig. i means the student implemented the plot corresponding to figure i (as shown in the report). Also, more than one students together implemented many plots, so that is written in the contribution of that student.