EECE 5642 Final Project Report
Music Virtualization

Yifan Wang 002839146
Department of Electrical and Computer Engineering
Northeastern University
Boston, the United States of America
wang.yifan27@northeastern.edu

Shuyue Li 002697053
Department of Electrical and Computer Engineering
Northeastern University
Boston, the United States of America
li.shuyue@northeastern.edu

Qiwen Ding 002839771
Department of Electrical and Computer Engineering
Northeastern University
Boston, the United States of America
ding.qiw@northeastern.edu

***Abstract*—This project explores the integration of advanced technologies in live music performances through professional and scientific analysis, focusing on the enhancement of audience engagement using virtual effects and animations. Our objective was to innovate within the realm of live performance by converting audio data from .wav files into dynamic multi-track animations. Although initial attempts were not successful, these challenges provided valuable insights into the complex nature of audio-visual transformation. Further, we attempted to virtualize music data for live streaming platforms, aiming to create interactive and immersive experiences for users engaged in streaming top music tracks.**

***Index Terms*—Music Features Extract, K-means, AE animation, Billboard, Spotify**

I. Introduction

In the advent of digital media and data science, the representation of music through visual elements has emerged as a novel avenue for analysis and interpretation. Our project encapsulates this innovation through a multi-faceted approach to music visualization. Firstly, we have extracted and analyzed Traditional Music Features, utilizing computational methods to identify and quantify the elements that define various genres and styles. We then brought these auditory experiences to life with After Effects (AE), creating three distinct music virtualization animations that offer an immersive exploration of sound through sight.

Furthermore, we have delved into the realm of music industry analytics with our Billboard Top Music Data Virtualization. Here, we dissect and display trends and patterns within the music industry, offering insights into the popularity and commercial success of songs and artists. Complementing this, our Spotify Daily Streaming Data Virtualization provides a dynamic view of listening behaviors and preferences, showcasing the ebb and flow of streaming activity across a diverse global audience.

II. Dataset

A. GTZAN Dataset

The GTZAN dataset is the most-used public dataset for evaluation in machine listening research for music genre recognition (MGR). The files were collected in 2000-2001 from a variety of sources including personal CDs, radio, microphone recordings, in order to represent a variety of recording conditions http://marsyas.info/downloads/datasets.html

a) Dataset content:

- genres original - A collection of 10 genres with 100 audio files each, all having a length of 30 seconds (the famous GTZAN dataset, the MNIST of sounds)
- images original - A visual representation for each audio file. One way to classify data is through neural networks. Because NNs (like CNN, what we will be using today) usually take in some sort of image representation, the audio files were converted to Mel Spectrograms to make this possible.
- 2 CSV files - Containing features of the audio files. One file has for each song (30 seconds long) a mean and variance computed over multiple features that can be extracted from an audio file. The other file has the same structure, but the songs were split before into 3 seconds audio files (this way increasing 10 times the amount of data we fuel into our classification models). With data, more is always better. Acknowledgements

The url of the dataset: https://www.kaggle.com/datasets/andradaolteanu/gtzan-dataset-music-genre-classification

B. Billboard-the-hot-100-songs Dataset

The Billboard Hot 100 serves as the definitive record chart within the United States music industry, published on a weekly basis by Billboard magazine. Rankings on the chart are determined by a combination of sales, radio airplay, and online streaming activity within the United States. Each week, Billboard publishes "The Hot 100" chart, showcasing the top songs based on their performance in sales and airplay during that period. This dataset comprises all "The Hot 100" charts released since its establishment in 1958.

The url of the dataset: https://www.kaggle.com/datasets/dhruvildave/billboard-the-hot-100-songs

C. *Spotify Daily Streaming Dataset*

This dataset encompasses a comprehensive collection of global music streaming data from Spotify, spanning from the year 2017 to 2019, it meticulously records the daily top 200 chart positions of various tracks across different countries, capturing essential details such as the date, the Spotify track URL, chart position, track name, artist, number of streams, and the country of streaming. The primary purpose of this dataset is to facilitate an in-depth analysis of music streaming trends over time, enabling researchers and analysts to gauge the popularity of tracks and the global reach of artists. By examining these elements, the dataset serves as a valuable resource for understanding how different genres and artists perform in various markets worldwide, thus offering insights into the dynamic landscape of the music industry during this period.

The url of the dataset: https://mkt.tableau.com/Public/Datasets/Spotify_Daily_Streaming.csv.zip

III. Data Virtualization Tasks

A. *Music Feature Extract and Classification*

In this part of project, we finished the tasks of music feature extraction and classification, as well as the data analysis procedures. The work primarily involves utilizing Python libraries such as Numpy, librosa, sklearn, and matplotlib to analyze and visualize music data. Specific methods include applying a series of techniques to extract waveform features from each song, using dimensionality reduction techniques Principal Component Analysis (PCA) and unsupervised learning for data classification. In addition, to gain deeper insights into musical characteristics, pre-emphasis filtering was applied, and a total of 1045 features were extracted for each song. These included time-domain feature extraction (3 features), frequency-domain feature extraction (3 features), Mel Frequency Cepstral Coefficients (MFCCs, 13 features), Delta features of MFCC (26 features), and Fast Fourier Transform (FFT) features (1000 features). These detailed feature sets allow us to use visualization methods to present the effects of classification and clustering, providing a comprehensive perspective to understand the attributes of music across various dimensions.

1) Feature Extraction:

a) Time-domain Feature Extraction: The figure 1,2,3 below show the Time-domain extracted Features, we only randomly choose one song to virtualize these features.

- Root Mean Square (RMS) Energy measures the average power of the signal. In music analysis, it can be used to estimate the loudness of music recordings.
- Zero Crossing Rate (ZCR) refers to the rate of sign changes in the signal. It's an indicator of time-domain features of an audio signal and is commonly used to distinguish between voiced and unvoiced parts in music and speech.
- Autocorrelation measures the similarity of the signal with itself at different time delays. It can be used to detect repetitive patterns, such as the presence of beats in music signals. Peak values indicate duplicated musical segments, most of the dupilicated segments are drum beats, the peak is more likely the repeat of the melody.

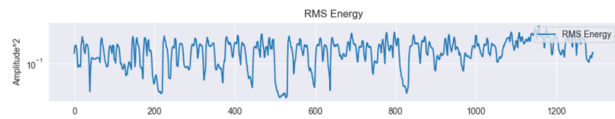

Fig. 1. Root Mean Square (RMS) Energy

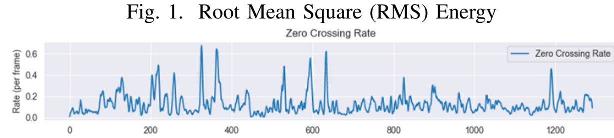

Fig. 2. Zero Crossing Rate (ZCR)

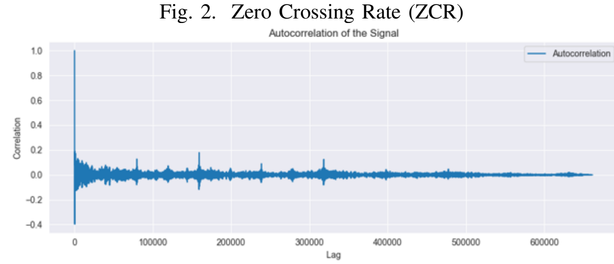

Fig. 3. Autocorrelation

b) Frequency-domain Feature: The figure 4,5,6 below show the Frequency-domain extracted Features

- Spectral Centroid reflects the "brightness" of a sound, calculated as a weighted average frequency. There will be more Sounds with higher brightness feel more "sharp.
- Spectral Rolloff Point is the frequency below which a specified percentage (usually 85%) of the total spectral energy lies. It's used to distinguish between the harmonic content and noise content of sounds.
- Spectral Contrast considers the difference between peaks and valleys in the signal across different frequency bands. This helps in identifying different timbres and instruments in music.

c) Mel Frequency Cepstral Coefficients (MFCCs): Mel Frequency Cepstral Coefficients (MFCCs) are a set of features that describe the overall shape of an audio signal. They are obtained by performing a cepstral analysis on the spectrum on a mel scale, very suitable for characterizing human voices and musical sounds.

The human ear doesn't perceive units like Hz in a linear relationship. For instance, when we're accustomed to a tone of 1000Hz, increasing the frequency to 2000Hz would only

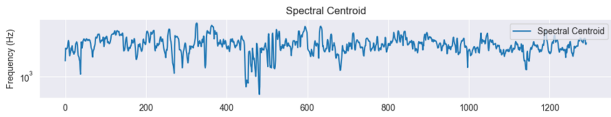

Fig. 4. Spectral Centroid

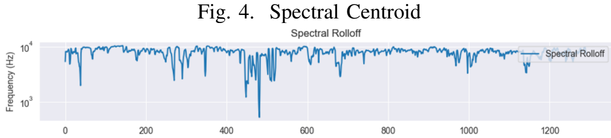

Fig. 5. Spectral Rolloff Point

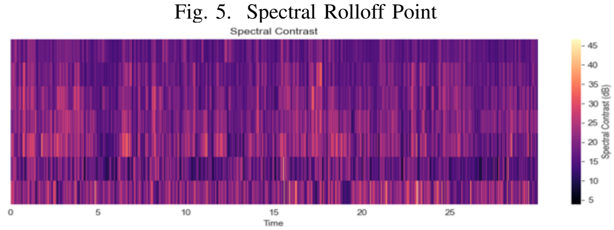

Fig. 6. Spectral Contrast

be perceived as a slight change by our ears, not doubling. However, in the Mel scale, if two speech segments differ by a factor of two in Mel frequency, the perceived pitch by the human ear would also differ by about the same factor. The mapping from Hz to Mel have a logarithmic relationship, when frequencies are low, Mel changes rapidly with Hz; when frequencies are high, the increase in Mel is slow, and the slope of the curve is small. This suggests that the human ear is more sensitive to low-frequency tones, while it becomes less sensitive at higher frequencies.

According to the basic rules of Mel Frequency Cepstral Coefficients [1], we extracted 26 different frequencies among the whole frequency space.

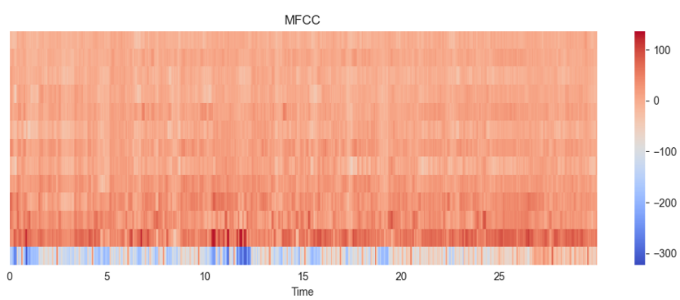

Fig. 7. Mel Frequency Cepstral Coefficients

d) FFT Features: We used Fast Fourier Transform (FFT) to transform the signal from time domain to frequency domain. The extracted features describe the frequency components of the signal. In our project, only the first 1000 coefficients are taken, reflecting the most significant frequency components of the audio signal.

2) K-means: K-means is a widely used clustering algorithm that partitions data points into K clusters, where K is the number of clusters specified by the user. The objective of this algorithm is to minimize the distance between data points within clusters and the cluster center, thereby making the clusters as tight as possible.

a) The selection of the number of clusters k: The figures 8 below show the results of the Silhouette Coefficient and the Elbow Method, which are used to determine the optimal number of clusters, k.

Figure on the left is a plot of the Silhouette Coefficient, The Silhouette Coefficient is a measure that ranges between -1 and 1, where higher values indicate that points within a cluster are closer to each other than to points in other clusters. Ideally, we look for the point where the silhouette score is highest to select (First time), This graph shows some fluctuations without a clear peak, but we might look for where the silhouette scores start to stabilize, usually indicating that increasing the number of clusters does not significantly improve the quality of clustering.

Figure on the right displays the Elbow Method, which is the variation of the sum of squared distances within clusters, k. The ideal k is often at the point where the decline in inertia starts to slow down, which is visually represented as an "elbow" on the graph. This chart shows a rapid decrease in inertia as k increases, but there is a deceleration in the rate of decrease around k of about 10, suggesting that $k = 10$ might be a good choice for the number of clusters.

Based on these two figures, we chose k where there is a relatively high silhouette score and where inertia begins to plateau in the elbow graph. However, given that the silhouette coefficient graph does not show a distinct peak, and the elbow graph presents an apparent elbow around $k = 10$, which may be a relatively good choice.

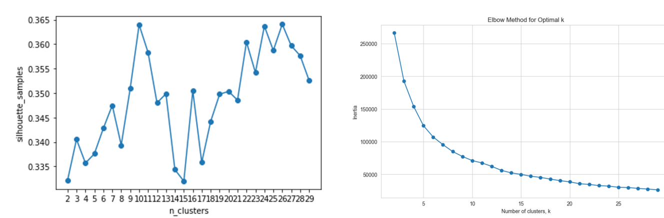

Fig. 8. Silhouette Coefficient (left) and Elbow Method (right)

b) The virtualization result of K-means: To show the result of classification, we successfully convert the 1045 dimensions vector of features into 3 dimensions and 2 dimensions by using PCA. We also create a heatmap show in figure 10 to represent the comparison between clusters identified by a k-means algorithm and the actual genres of songs.

- Cluster 1 is predominantly classical, showing the algorithm's effectiveness in grouping these features.
- Disco and pop are concentrated in Cluster 7, suggesting similar features or less differentiation by the algorithm.
- Rock and metal are distributed across clusters, indicating diversity within these genres or feature overlap with other genres.

- Clusters such as 2, 3, 5, and 8 are less genre-specific, perhaps representing unique or uncommon musical features.
- Hip hop is well-defined within Cluster 0, and jazz and reggae show moderate associations in Clusters 4 and 9, respectively.

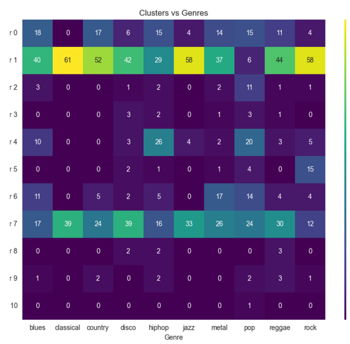

Fig. 9. Heatmap

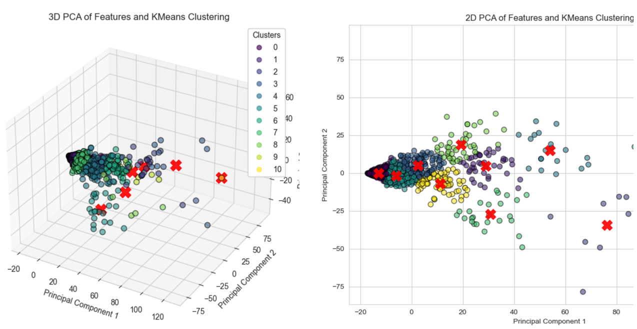

Fig. 10. Result of K-means in 3D (left) and 2D (right)

Unfortunately the result isn't quite well, although the coefficient shows that when k=10 we will have the best result, but the songs in each cluster are quite different from the original, 70% of the songs are squeezed into two groups, some songs are too far away from the central point. There are three main reasons for why we got such classification result:

- There are too many features, we don't know which one can represent the music very well.
- A song can belong to the different class, such as Spotify, each song has their own tags, we can't say that it must belong to only one cluster, so in the k-means algorithm they are likely hard to distinguish.
- The K-means algorithm is sensitive to the selection of initial cluster centers. If the initial cluster centers are poorly chosen, it may lead the algorithm to converge to a local optimum, thereby affecting the quality of clustering. For eaample the FFT, due to the time limit we don't have time to test every features, we only pick about 1000 points. And we can't ganrantee that this 1000 point can represent the songs best.

B. AE Music Visualization

In the practical realization of our music visualization project, a blend of advanced digital tools was employed to seamlessly translate auditory data into visual formats. Predominantly, After Effects (AE) and Flourish stood out as the cornerstone technologies facilitating this creative conversion. After Effects was instrumental in the meticulous crafting of video content, offering robust features for audio analysis, layer and graphic design, particle effects parameterization, camera control, and the setting of particle motion trajectories. Meanwhile, Flourish provided the necessary capabilities to produce intricate and informative charts that further enriched our visual outputs. This section of the report will delve into the detailed application of these tools, illuminating how they collectively contributed to the aesthetic and functional success of our visualization endeavors.

For our music visualization project, we selected primarily electronic and instrumental music, as these genres exhibit a pronounced rhythmic quality that enhances visual representation. The distinct beats and tempo of electronic music, in particular, lend themselves superbly to dynamic visualization techniques, enabling a vivid portrayal of sound as it plays.

In terms of visual elements, we emphasized the use of gradient colors to create a fluid and evolving aesthetic that mirrors the ebb and flow of the music. Lines were utilized to depict irregular beats and suggest movement, simulating running tracks that animate in response to the audio. Additionally, particle effects were employed to add depth and texture, creating an immersive visual field. Perspective changes were strategically integrated to shift viewer focus and provide a three-dimensional feel to the visual experience, making the music not only audible but also visually palpable and engaging.

1) Music analysis: Analyzing the intricacies of sound is pivotal in crafting a responsive visual landscape, and the graphics presented here delineate the core aspects of our audio analysis phase. Through the utilization of the FreqReact script, we dynamically mapped frequencies within a selected range, as illustrated by the prominent blue peaks that respond to the music's tempo and intensity. The bounding box in the Figure 11 is our chosen "drum beat", When the Amplitude in the rectangle is more than 50%, we control script to emit particles and the particle effect will trigger whenever the audio exceeds this threshold.

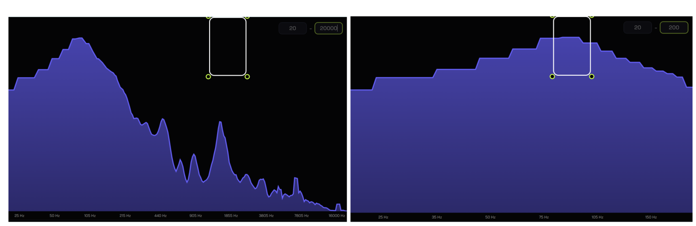

Fig. 11. FreqReact script

To further dissect the auditory elements, we applied the 'Sound Keys' effect within After Effects, which allowed us to extract and quantify specific audio features. This is demonstrated by the granular breakdown into 'High amplitude treble' and 'Low amplitude treble' waveforms, offering a visual dichotomy between intense and subtle sound fragments. The image below showcases the fluctuating amplitude of treble frequencies, effectively translated into a tangible visual format, laying the foundation for our subsequent visualization layers.

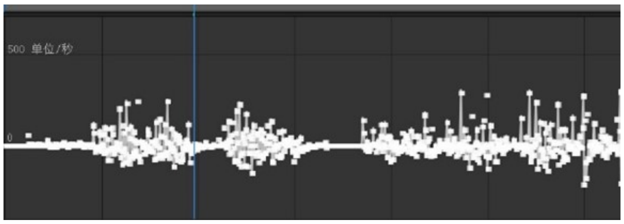

Fig. 12. High amplitude treble

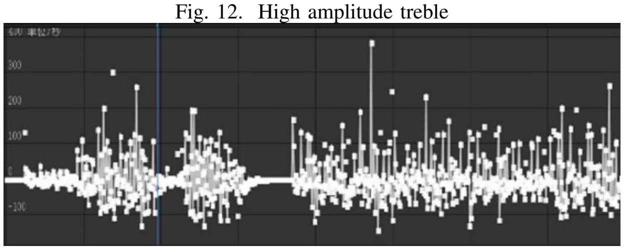

Fig. 13. Low amplitude treble

C. Graphic Design

The Graphic Design section of our music visualization project illustrates the creative synergy between music and visual elements. The 'Line dancing' feature depicts animated lines that groove to the rhythm, their movements synchronized with the musical beats, producing a digital dance of colors. 'Particle bloom' showcases a burst of particles that scatter and twirl, embodying the music's crescendos and diminuendos with visual splendor. In 'Snowflakes and tetrahedrons', geometric shapes intertwine with delicate snowflakes, creating a complex, yet harmonious, visual representation of the music's texture. Lastly, 'Emitter' represents a swarm of particles aligned to form a stream that flows in harmony with the melody, like a visual echo of the sound waves. Each design element has been carefully crafted to visually echo the auditory experience, ensuring that the visual journey is as compelling as the auditory one.

D. Dynamic Leaderboard Video

In this part of project, Our goal is to delve into this rich historical archive to uncover and visualize the transformation of music preferences over the decades, offering a unique lens through which to view the metamorphosis of popular culture.

1) Data Processing: The integrity and structure of data are critical for insightful analysis, particularly when examining historical trends within Billboard's "The Hot 100" charts. Our approach commenced with meticulous data cleansing, a phase where we eradicated duplicate records, rectified entry and formatting inconsistencies, and purged any data irrelevant

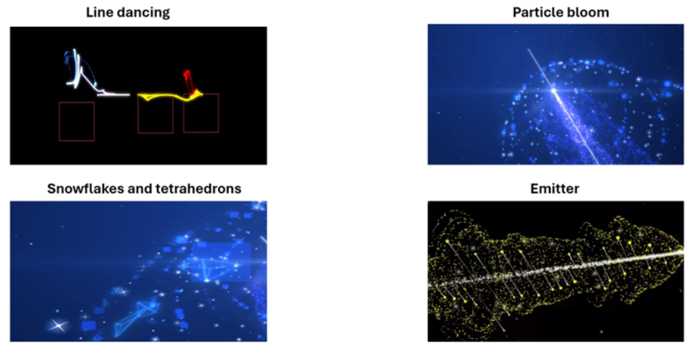

Fig. 14. AE Virtualization Result

to our analytical goals. This step was crucial in establishing a reliable foundation for our exploration. Subsequently, we progressed to data transformation. This process entailed reformatting and converting data into a schema that aligned with the requirements of our analytical methods. By tailoring the dataset in this way, we ensured compatibility with the tools and techniques we intended to deploy later in the analysis. Finally, data reconstruction allowed us to rearrange the dataset's architecture to better serve our investigation's objectives. This involved creating pivot tables for multidimensional data examination and transforming the dataset from a wide format to a long format to enhance the clarity and accessibility of the information. Through these stages of data preparation, we set the stage for a comprehensive and nuanced analysis of the ebb and flow of musical trends over time. Figure 6 shows the complete process of processing our dataset.

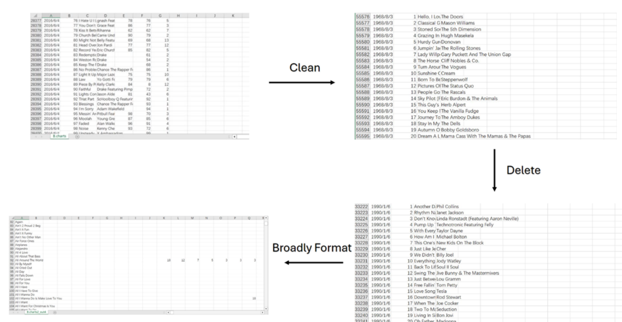

Fig. 15. Data processing

The Figure 16 showcases a colorful bubble chart, representing Billboard Hot 100 songs for the week of November 6, 2021. Each bubble corresponds to a song, with size likely indicating a metric such as popularity or weeks on the chart, and varying colors perhaps denoting different genres or categories. This visual offers an at-a-glance understanding of the chart's composition and the relative standing of each track within it.

E. Spotify Trends Analysis

Using the Spotify Daily Streaming Dataset, we can perform a multidimensional analysis to reveal trends in the music

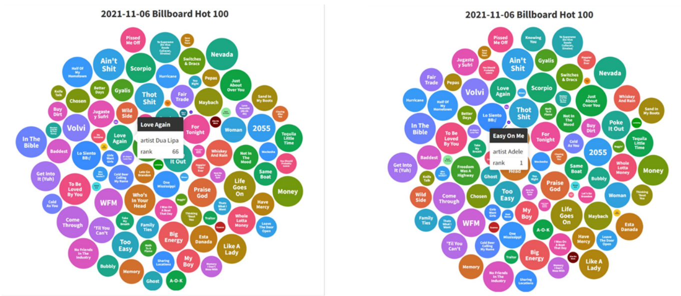

Fig. 16. Bubble chart

Fig. 18. Geographic Variations Analysis

streaming domain. This dataset, covering the years 2017 to 2019, allows us to observe shifts in musical preferences, the impact of seasonal trends on listening habits, and the rise of new artists and genres in different markets.

1) Temporal Trends Analysis: Firstly, temporal trend analysis focuses on how music consumption changes over time. By aggregating data monthly or quarterly, we can identify patterns such as a surge in streaming at the end of the year (possibly related to holiday seasons) or summer peaks (potentially linked to holidays and music festivals). Additionally, analyzing specific holidays like Christmas or Chinese New Year can reveal the music consumption characteristics of specific cultures and regions. This analysis also highlights the longevity of certain tracks or albums, showing how long they stay within the top charts, which is a crucial indicator of their popularity and endurance.

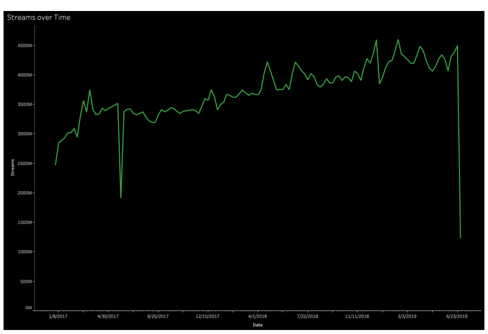

Fig. 17. Temporal Trends Analysis

2) Geographic Variations Analysis: Another significant area of analysis is the geographical differences in music streaming. With the dataset broken down by country, we can compare popular genres and artists across different regions. For example, reggae music might perform stronger in Latin American countries, while K-pop may have a significant impact in East and Southeast Asia. In geographical analysis, we can also explore differences in music preferences between urban and rural areas and how economic development levels influence music trends. This part of the analysis can also reveal how global hits transcend cultural and linguistic barriers to achieve worldwide popularity.

3) Artists and Genres Insights: Moreover, the dataset supports in-depth analysis of artists and genres. By tracking the streaming numbers for each artist or genre, we can identify rising stars in the music industry and trends in popular or declining genres. This analysis can help identify artists who consistently appear on the charts, indicating they have a strong and loyal fan base, or those who achieve one-hit-wonder status. The detailed analysis of artists and genres can further be segmented into impact analysis of emerging artists and market endurance of established artists.

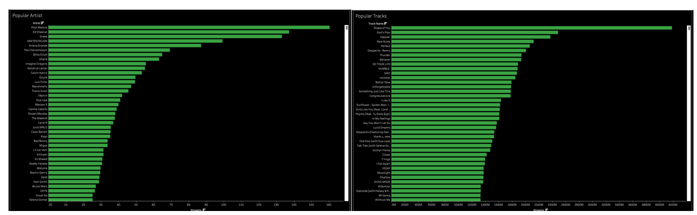

Fig. 19. Popular Tracks and Popular Artist

4) Total UI: To effectively communicate our analytical findings, we used an interactive dashboard to allow users to explore data based on their specific interests, such as focusing on a particular artist or time period. In addition to traditional charts, timeline animations can be used to showcase changes in the popularity of artists or music genres over time. User-interactive elements such as sliders and filters allow users to customize views based on different time periods, regions, artists, or genres, providing a more personalized data exploration experience.

The website of our virtualization: https://lishuyue0104.github.io/1/

IV. Teamwork

- Yifan Wang: K-means and music feature extract, Deploy and test MT3 model, AE animation
- Qiwen Ding: AE animation, Top music virtualization
- Shuyue Li: AE animation, Spotify Music Trends Analysis

Except for the ”Top Music Virtualization” and ”Spotify Music Trends Analysis” sections, we utilized third-party libraries to process our data. All other sections were implemented with our own visualization techniques. The codes

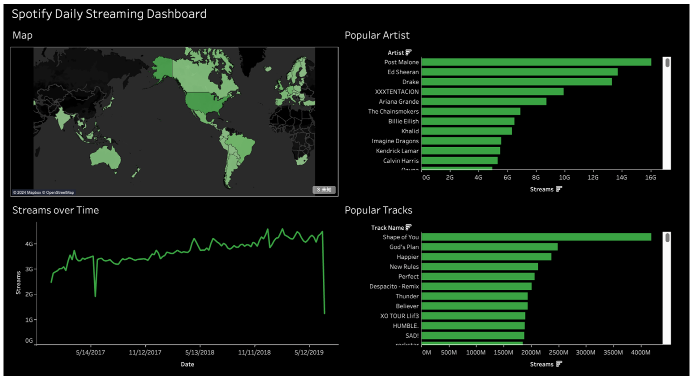

Fig. 20. Total UI

of our work is located at : https://github.com/AKAGIwyf/EECE-5642-Data-Virtualization

V. Conclusion

In conclusion, for the K-means and music feature extract part, exploring additional traditional features suitable for unsupervised learning could enhance our understanding of music representation. Comparing these traditional features with those extracted from deep learning models offers valuable insights into the effectiveness of each approach. However, for the AE animation,

For the AE animation, Top music virtualization part, using well-known software like After Effects and detailed data analysis, we've turned the sounds of music trends into clear visual diagrams. We've paid close attention to every note and beat in the songs, carefully choosing visual elements and camera movements to match. This work tracks the changes in what music people have enjoyed over the years. The project demonstrates our technical skill and adds to our knowledge of how music's popularity changes with time.

For the Spotify Music Trends Analysis part, the Spotify Daily Streaming Dataset not only provides a snapshot of the global music scene from 2017 to 2019 but also offers a rich resource for understanding the dynamic nature of music consumption. The insights derived from this dataset can inform stakeholders in the music industry about potential markets, successful promotional strategies, and the overall direction in which music trends are heading.

References

[1] Hugo B. Lima, Carlos G. R. Dos Santos, and Bianchi S. Meiguins. A survey of music visualization techniques. *ACM Comput. Surv.*, 54(7), jul 2021.