Unveiling Musical Triumphs

Table of Contents

List of Figures 4
1.0 Introduction 6
1.1 Overview of research area and focus of the study 6
1.2 Aims and Importance 6
1.3 Summarize of literatures 7
1.4 Hypothesis and research questions 7
Hypothesis 7
Research Questions 8
1.5 Approach to achieving aims 8
1.6 Key findings and contribution of work 8
Key findings 8
Contribution of work 11
2.0 Methodology 11
2.1 How was the data gathered 11
2.1.1 Summary on data cleaning and feature selection 12
Data cleaning 12
Feature selection 13
2.2 Research methods selection 14
2.2.1 Justification for method choice 15
2.2.2 How were these methods applied to analyse the research question or problem? 15
3.0 Results 16
3.1 Restate the study purpose 16
3.2 Key findings 17
3.2.1 Popularity Vs Duration (Spotify) 17
3.2.2 Feature Importance Vs Popularity 18
3.2.3 Explicit Vs Popularity 18
3.2.4 Song Clusters Vs Popularity 19
3.2.5 Top 10 Artists Vs Followers in Spotify 22
Ed Sheeran 23
Ariana Grande 24
Drake 25
Justin Bieber 26
Eminem 27
3.2.6 Popularity Vs Duration (YouTube) 28
3.2.7 Top 10 Artists Vs Followers in YouTube 29
Luis Fonsi 30
Wiz Khalifa 31
Katy Perry 32
3.3 Report of insignificant research findings 33
3.4 Summary of key study outcomes 33
4.0 Discussion 33
4.1 Key findings interpretation 33
4.2 Comparative analysis and critical assessment 34

4.2.1 Compare and contrast to previous study 34
4.2.2 Strengths and limitations of this study 35
Strengths 35
Limitations 36
4.2.3 Discussion of unexpected findings 36

5.0 Discussion* *36

5.1 Summary of hypothesis and purpose of this study 36
Hypotheses 37
Purpose of this study 37

5.2 Significance of this study 38

5.3 Unanswered questions and future enhancements 39

6.0 References* *39

Appendix* *40

Home 40

About 43

Popularity 43

Song clusters vs Popularity 45

Artist (Spotify) 47

Artists (YouTube) 48

Prediction 49

List of Figures

Figure 1.0	Acousticness Prediction (15 Years)
Figure 2.0	Energy Prediction (15 Years)
Figure 3.0	Danceability Prediction (15 Years)
Figure 4.0	Heatmap
Figure 5.0	Spotify Merged Dataset
Figure 6.0	YouTube Dataset
Figure 7.0	Popularity Vs Duration
Figure 8.0	Popularity Vs Duration with OLS Implemented
Figure 9.0	Feature Importance Vs Popularity
Figure 10.0	Explicit Vs Popularity
Figure 11.0	Song Clusters Vs Popularity
Figure 12.0	Cluster 8 Songs Duration (Mean)
Figure 13.0	KMeans Without Using Methods To Choose Number of K
Figure 14.0	Elbow Method To Select Number of K
Figure 15.0	KMeans With The Use of Method To Choose Number of K
Figure 16.0	3D Graph For Song Clusters
Figure 17.0	Top 10 Artists Vs Followers in Spotify
Figure 18.0	Average Values of Music Characteristics For Ed Sheeran
Figure 19.0	Ed Sheeran Song Characteristics Over Time
Figure 20.0	Average Values of Music Characteristics For Ariana Grande
Figure 21.0	Ariana Grande Song Characteristics Over Time
Figure 22.0	Average Values of Music Characteristics For Drake
Figure 23.0	Drake Song Characteristics Over Time
Figure 24.0	Average Values of Music Characteristics For Justin Bieber
Figure 25.0	Justin Bieber Song Characteristics Over Time
Figure 26.0	Average Values of Music Characteristics For Eminem
Figure 27.0	Eminem Song Characteristics Over Time
Figure 28.0	Popularity Vs Duration
Figure 29.0	Popularity Vs Duration with OLS Implemented
Figure 30.0	Top 10 Artists Vs Followers in YouTube
Figure 31.0	Average Values of Music Characteristics For Luis Fonsi

Figure 32.0	Luis Fonsi Song Characteristics Over Time
Figure 33.0	Average Values of Music Characteristics For Wiz Khalifa
Figure 34.0	Wiz Khalifa Song Characteristics Over Time
Figure 35.0	Average Values of Music Characteristics For Katy Perry
Figure 36.0	Katy Perry Song Characteristics Over Time

1.0 Introduction

1.1 Overview of research area and focus of the study

Over the past century, the landscape of popular music has undergone a profound metamorphosis, mirroring the swift evolution of culture and musical innovation. This research endeavours to delve into the dynamic development of musical genres and production techniques that have continually shaped our auditory experiences. Spanning from the exuberant Jazz Age of the 1920s to the contemporary triumphs of modern pop, the investigation aims to unravel the intricate interplay between cultural shifts and musical evolution (Welch et al., 2020).

The focal point of this study revolves around discerning the fundamental elements that contribute to the success and popularity of songs across different eras. The nuanced exploration of energy, tempo, rhythm, and lyrical content serves as a lens through which to comprehend the essence of musical triumphs over time. By undertaking this inquiry, we seek to illuminate the complex interrelationship between artistic expression and audience preference, thereby shedding light on the enduring allure of musical compositions spanning earlier epochs to the current musical landscape.

1.2 Aims and Importance

The primary objective of this research is to unravel the essential components that underpin the success and popularity of songs across diverse historical periods. By dissecting the subtleties of energy, tempo, rhythm, and lyrical content, we aim to decipher the underlying factors that have consistently resonated with audiences throughout the decades. This investigation aspires to contribute to a comprehensive understanding of the symbiotic relationship between artistic innovation and the ever-shifting tastes of listeners.

Furthermore, the study holds significance in providing insights into the enduring appeal of musical works from earlier eras in the context of the contemporary musical landscape. Unraveling the threads that connect the past to the present will enrich our appreciation for the cultural significance of music, elucidating its role as a vital and evolving component of our cultural fabric (Welch et al., 2020).

1.3 Summarize of literatures

The music landscape has undergone significant transformations over the last century, with shifts in consumer preferences and musical genres occurring across various eras. According to a study investigating inequality and unpredictability in cultural markets, the success and popularity of songs are not solely contingent on the intrinsic quality of the music but are also influenced by social factors (Salganik, 2006).

Research conducted by DataFace has revealed a cyclical pattern in the life cycles of musical genres, characterized by their ascent in popularity, reaching a peak, and eventually making way for other forms (Beckwith, 2016). This cycle is attributed to the interplay of several elements, including technological advancements, sociocultural shifts, and evolving consumer tastes. Notably, the emergence of contemporary genres like Electronic Dance Music (EDM) and Dubstep can be linked to advancements in electronic devices used in music creation. These findings underscore the significance of identifying key features crucial for the creation of potentially successful songs.

1.4 Hypothesis and research questions

Hypothesis

In general terms, hypothesis is a statement that is utilised to conduct an experiment on the connection of two or more variables or a suggested clarification about several observation phenomena. Thus, in this report the hypothesis that will be made are stated below:

- The musical features will have varying importance on influencing the popularity of the song, with some having a (larger/smaller) impact on the popularity than others.
- The combination of multiple and unique musical features, including tempo, acoustic, etc., will have a (higher/lower) probability of creating a popular song.
- The optimal range for a song’s duration (increases/decreases) the likelihood of a song gaining popularity compared to other songs which are not in this range.
- The song characteristics have (evolved/not evolved) over the past century for few of the famous artists, which refers to the change in consumer tastes as time goes by.

Research Questions

In general terms, research questions are specified questions where the research is conducted to deliver the answers to, thus normally positioned at the first step of the project. Consequently, research questions emphasize on research, dictate the methodology and hypothesis, and provide the route for all steps in analysis, inquiry, and report. Thus, in this report the research questions that will be made are stated below:

- Does the impact on a song's popularity vary across different musical features?
- Does the combination of these features such as acoustic, tempo, etc., in harmony increase the likelihood of producing a hit song?
- Does an optimal duration exist for a song, maximizing its potential for popularity?
- Does the characteristics of songs differ as time progresses from 1920 to 2020?

1.5 Approach to achieving aims

The aim of this research is to pinpoint the factors that wield the most significant influence on a song's popularity and to delineate the most effective combination of these attributes for crafting a song with optimal chances of success. This will be achieved through meticulous cleaning and analysis of a dataset encompassing a diverse array of artists and spanning multiple eras. Machine Learning algorithms such as K-Means to cluster songs based on popularity, Ordinary Least Square to detect outliers, and Long Short Term Memory (LSTM) for prediction purpose, will be employed to assess the importance of various elements such as duration, danceability, energy, explicitness, and acousticness. Additionally, we will employ data visualization techniques to depict and scrutinize patterns within the data and refer to appendix for the dashboard created. Beyond providing a comprehensive understanding of the dynamics driving musical success, our study will yield practical insights for artists and producers by identifying the ideal blend of these components. This research contributes to the knowledge base on factors influencing song success and serves as a valuable resource for future studies in music psychology as well as for artists aspiring to create chart-topping singles.

1.6 Key findings and contribution of work

Key findings

As depicted in Figures 1.0 to 3.0, the researchers anticipated the musical traits of Acousticness, Energy, and Danceability. These three components were chosen for prediction over the next 15 years based on the heatmap depicted in Figure 4.0, the most important feature based on popularity as illustrated in Figure 9.0, as well as insights from a study conducted by (Millecamp et al., 2018). The mentioned research indicated that a majority of participants expressed a preference for the energy attribute, followed by acousticness. The prediction algorithm employed for this analysis is LSTM.

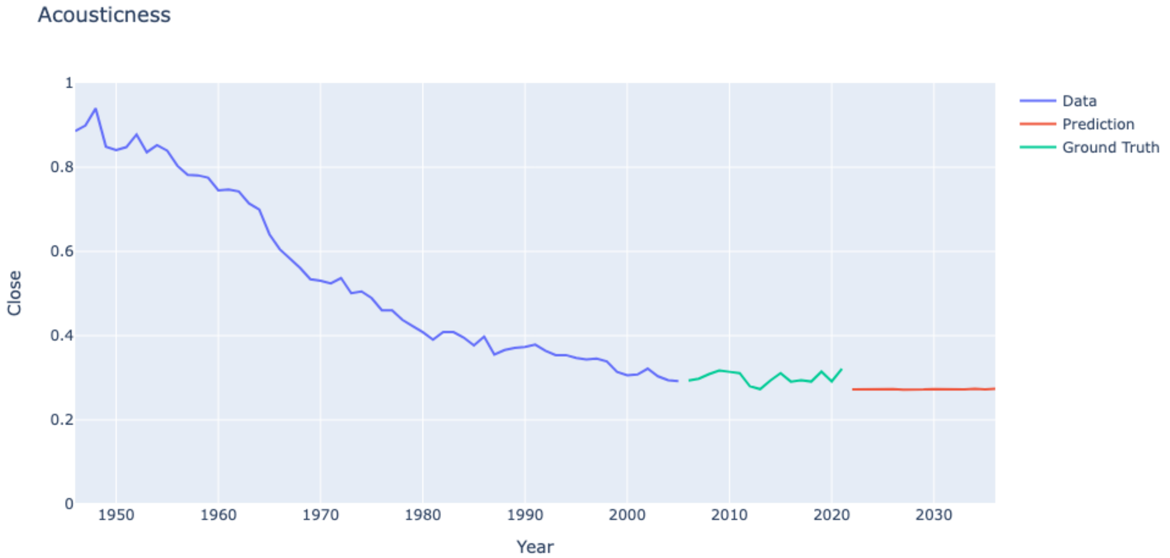


Figure 1.0 - Acousticness Prediction (15 Years)

Figure 1.0 illustrated a projected stable pattern for acousticness variable over the next 15 years.

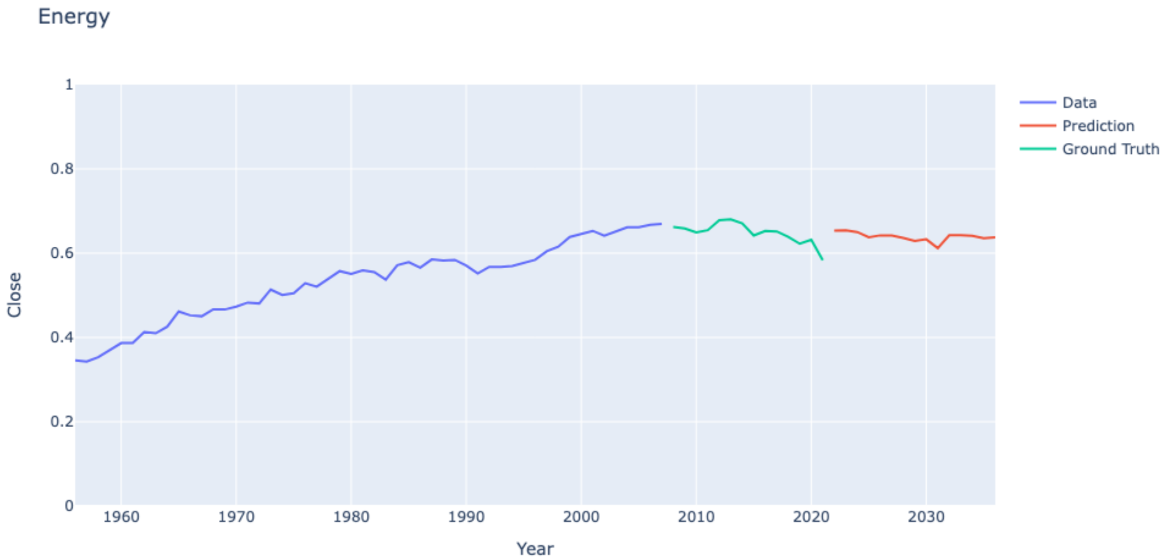


Figure 2.0 - Energy Prediction (15 Years)

Figure 2.0 illustrated a gradual decline in energy variable over the next 15 years.

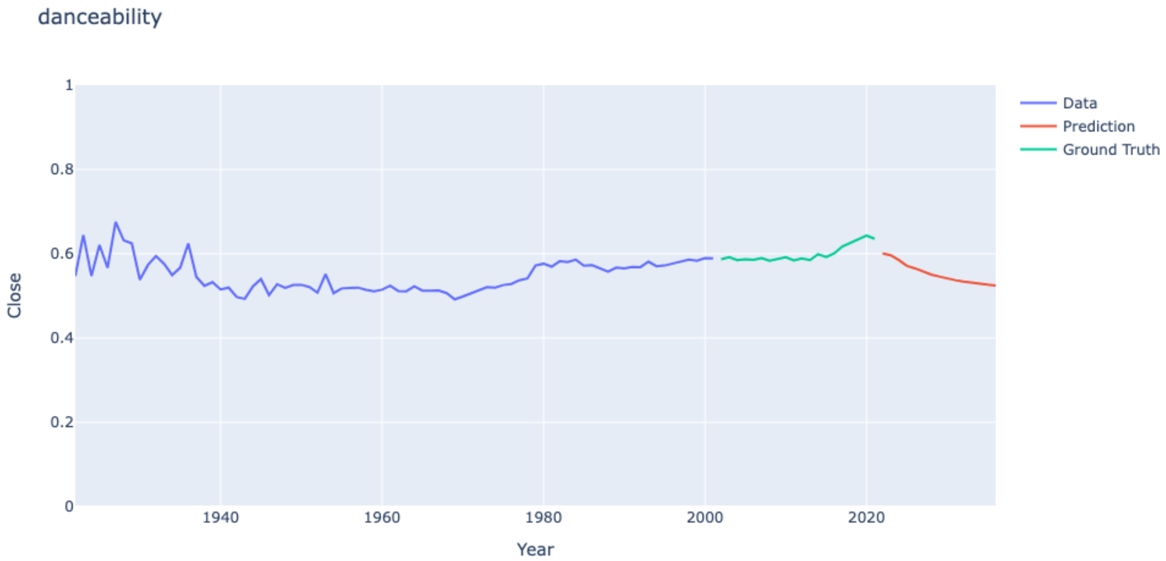


Figure 3.0 - Danceability Prediction (15 Years)

Figure 3.0 illustrated a decline in the danceability variable over the next 15 years.

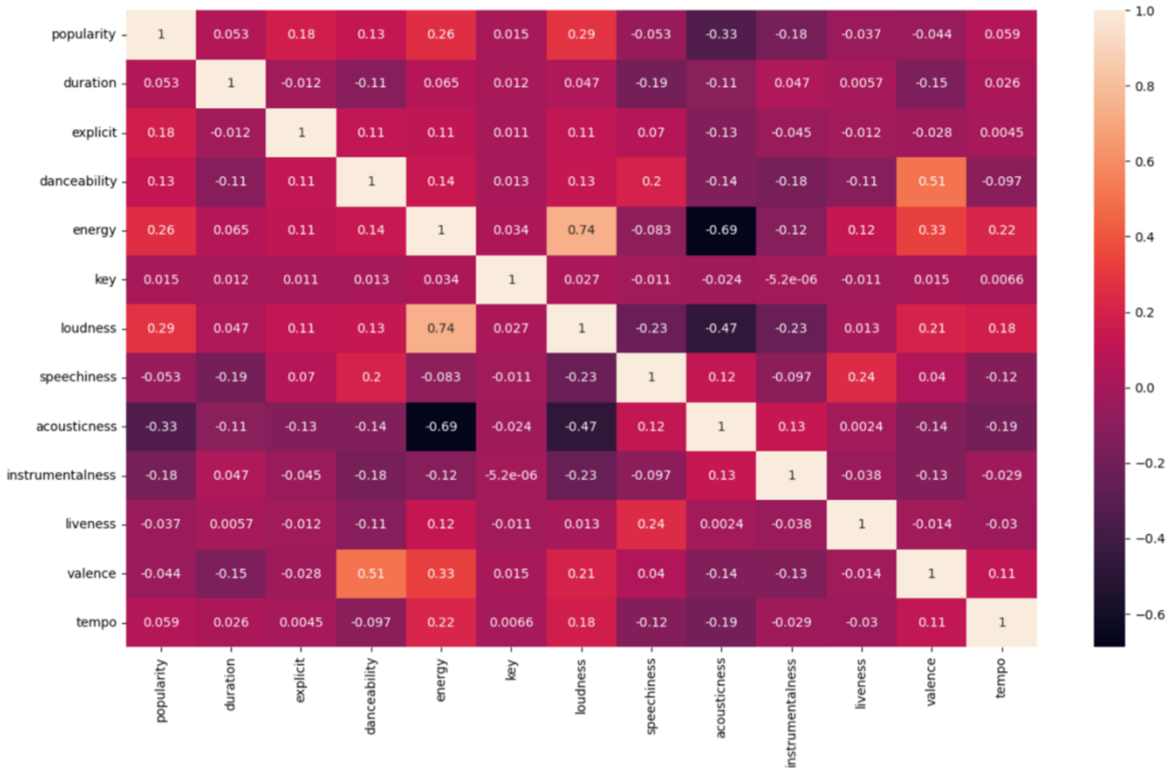


Figure 4.0 – Heatmap

Figure 4.0 illustrated the correlation between various features, revealing a noteworthy observation: loudness and energy exhibited the highest correlation, followed by danceability and valence. Consequently, for prediction purposes, energy was chosen over loudness, taking into account insights from a study conducted by (Millecamp et al., 2018) which highlighted that a majority of participants favored the energy attribute, followed by acousticness. Additionally, the preference for danceability over valence is grounded in the feature importance versus popularity chart presented in Figure 9.0. According to the chart, danceability exhibits greater popularity compared to valence. Hence, only three variables - Acousticness, Energy, and Danceability - have been selected for prediction over the next 15 years where this prediction can assist music creators in crafting songs that have the potential to garner greater popularity.

Contribution of work

As per the research carried out by (Salganik, 2006), the researcher noted substantial changes in the musical landscape over the past century, encompassing shifts in consumer preferences and the evolution of musical genres across different eras. Figures 1.0 to 2.0 reveal observable trends, indicating both a decline and an increase in acousticness and energy.

2.0 Methodology

2.1 How was the data gathered

The datasets employed in this study were sourced from Kaggle, a widely utilized platform for datasets, data science, and machine learning. The initial dataset comprised more than 500 thousand rows of track data, encompassing features like duration, popularity, and song features. The second dataset encompassed data on over a million artists, featuring attributes such as popularity and the number of followers for each artist. Lastly, the third dataset provided statistics on tracks sourced from YouTube.

Subsequently, comprehensive data cleaning procedures were implemented across all datasets. In the initial dataset, we removed symbols such as "[]" and " ‘’ " from the 'artists' column. Additionally, null values and duplicates were identified and subsequently removed from the 'name' column. Four features - 'id', 'id_artists', 'mode', and 'time signature' - were deemed irrelevant to our research and were consequently dropped. It's noteworthy that this dataset exclusively contained track details. To enhance the scope of our analysis, we opted to merge it with the second dataset, which primarily centered on artist statistics.

Regarding the second dataset, we renamed the column name 'name' to 'artists' to ensure consistency during the merging process. Subsequently, we addressed null values in the 'followers' column by replacing them with the mean value of the column. Duplicate values in the 'artists' column were identified and removed. In parallel with the first dataset, two features, namely "id" and "genres" were excluded from the dataset.

Through the merging of the first and second datasets based on the 'artists' column, we obtained a more comprehensive dataset comprising 18 features. Following the merger, two popularity columns, namely 'popularity_X' and 'popularity_Y,' emerged. We opted to utilize the 'popularity_X' column as it aligns more closely with our research focus, consequently discarding the 'popularity_Y' column. Additionally, the 'duration_ms' column, representing song duration in milliseconds, was converted to minutes to enhance data presentation and visualization, ensuring consistency. The final result was a merged dataset encompassing 17 features.

Our third dataset focuses on the popularity of songs on YouTube, diverging from the first and second datasets that utilize track and artist data from Spotify. The purpose of incorporating this dataset is to juxtapose the musical preferences of the Spotify user base with those of the YouTube user base. Initially, we engaged in feature selection by excluding elements deemed irrelevant for this comparison, including "Unnamed: 0," "Url_spotify," "Album_type," "Uri," and several others. Subsequently, null values were identified in the "Views" column, and we replaced them with the column's mean. Simultaneously, we eliminated null values in the "Duration_ms" column. Much like the merged dataset, we converted the duration column into minutes and transformed the "Views" feature from the 'Object' data type to a numeric format.

2.1.1 Summary on data cleaning and feature selection

Data cleaning

- Symbols Removal: Eliminated symbols "[]" and " ‘’ " from the 'artists' column within the initial dataset.
- Null Values and Duplicates Management: Identified and removed both null values and duplicates in the 'name' column of the initial dataset and substituted null values in the 'followers' column of the second dataset with the mean value of the column and discarded duplicate values present in the 'artists' column of the second dataset.
- Feature Exclusion: Eliminated the 'id', 'id_artists', 'mode', and 'time signature' features from the initial dataset and eliminated 'id' and 'genres' features from the second dataset.
- Column Renaming: Changed the name of the 'name' column in the second dataset to 'artists' to maintain consistency during the merging process.

Feature selection

- Dataset Integration: Combined the initial and second datasets by aligning them based on the 'artists' column, resulting in a more inclusive dataset.
- Relevance-Based Feature Selection: Chose the 'popularity_X' column over 'popularity_Y' due to its closer alignment with the research objective, leading to the exclusion of the 'popularity_Y' column.
- Data Conversion: Transformed the 'duration_ms' column, indicating song duration in milliseconds, into minutes to enhance data visualization and analytical insights.
- Final Feature Count: Concluded with 17 pertinent features in the merged dataset following thorough data cleaning and feature selection procedures.

Following rigorous data cleaning, feature elimination, and selection procedures, the combined dataset resulted in 17 pertinent features. This extensive dataset played a pivotal role in uncovering the connections between musical attributes and their impact on song popularity. Additionally, it facilitated comparisons between the music preferences of Spotify and YouTube user bases, laying the groundwork for subsequent analyses. Figure 5.0 represents the merged dataset, while Figure 6.0 depicts the YouTube dataset after the rigorous data cleaning, feature elimination, and selection process.

	name	popularity	duration	explicit	artists	release_date	danceability	energy	key	loudness	speechiness	acousticness	instrumentalness
0	Carve	6	2.115050	0	Uli	1922-02-22	0.645	0.44500	0	-13.338	0.4510	0.674	0.744000
1	Lazy Boi	0	2.622217	0	Uli	1922-02-22	0.298	0.46000	1	-18.645	0.4530	0.521	0.856000
2	Sketch	0	1.450667	0	Uli	1922-02-22	0.634	0.00399	5	-29.973	0.0377	0.926	0.919000
3	L'enfer	0	0.666667	0	Uli	1922-02-22	0.657	0.32500	10	-14.319	0.2540	0.199	0.856000
4	Graphite	0	1.740000	0	Uli	1922-02-22	0.644	0.68400	7	-8.247	0.1990	0.144	0.802000
...	...	...	...	...	...	...	...	...	...	...	...	...	...
351773	The Cutest Puppy	67	1.375000	0	Laureen Conrad	2020-10-30	0.609	0.01720	8	-28.573	0.1180	0.996	0.973000
351774	同行 (新加坡電視劇《愛...沒有距離》主題曲)	43	3.420633	0	Boon Hui Lu	2020-03-03	0.743	0.67900	8	-3.952	0.0323	0.269	0.000000
351775	John Brown's Song	66	3.087500	0	Gregory Oberle	2020-03-20	0.562	0.03310	1	-25.551	0.1030	0.996	0.961000
351776	云与海	50	4.304450	0	阿YueYue	2020-09-26	0.560	0.51800	0	-7.471	0.0292	0.785	0.000000
351777	blind	72	2.554883	0	ROLE MODEL	2020-10-21	0.765	0.66300	0	-5.223	0.0652	0.141	0.000297

351778 rows × 17 columns

Figure 5.0 - Spotify Merged Dataset

	Artist	Track	Duration_ms	Views
0	Gorillaz	Feel Good Inc.	222640.0	693555221.0
1	Gorillaz	Rhinestone Eyes	200173.0	72011645.0
2	Gorillaz	New Gold (feat. Tame Impala and Bootie Brown)	215150.0	8435055.0
3	Gorillaz	On Melancholy Hill	233867.0	211754952.0
4	Gorillaz	Clint Eastwood	340920.0	618480958.0
...	...	...	...	...
20713	SICK LEGEND	JUST DANCE HARDSTYLE	94667.0	71678.0
20714	SICK LEGEND	SET FIRE TO THE RAIN HARDSTYLE	150857.0	164741.0
20715	SICK LEGEND	OUTSIDE HARDSTYLE SPED UP	136842.0	35646.0
20716	SICK LEGEND	ONLY GIRL HARDSTYLE	108387.0	6533.0
20717	SICK LEGEND	MISS YOU HARDSTYLE	181500.0	158697.0

20716 rows × 4 columns

Figure 6.0 - YouTube Dataset

2.2 Research methods selection

Research methods comprise a wide range of organised techniques and strategies used to collect information or proof, with the ultimate objective of promoting more in-depth understanding and enlightening analysis of a given topic. These approaches function as helpful frameworks that direct the gathering and analysis of data with the ultimate goal of revealing new information or developing a more thorough comprehension of the studied field (Tech, 2023). There are two types of research methods:

Quantitative research methods which utilizes statistical and computational techniques, this analytical approach quantifies, measures, and examines different facets of the data to enable researchers to draw insightful conclusions. In order to test hypotheses, find correlations, and extrapolate results to a broader population, quantitative data analysis is frequently employed. This approach allows for a more thorough comprehension of the underlying patterns and behaviours present in the dataset (Lakshman, 2000).

1. Quantitative research methods which include:
    a. Surveys
    b. Questionnaires
    c. Tests
    d. Databases
    e. Experiments

Qualitative research methods involve the investigation of non-numerical data with the goal of revealing underlying themes and meanings. With its rich, descriptive insights into a particular research subject, it provides a deeper understanding of subjective interpretations, complex phenomena and human interaction.

2. Qualitative research methods:
    a. Interviews
    b. Observations
    c. Focus Group

Acquiring the data from Kaggle constitutes a form of Secondary Data Analysis (SDA), which involves using information derived from previously gathered primary data sources such as books, surveys, journals, and more. Given that our data predominantly involves statistical information, it falls within the realm of Quantitative Research methods (Wickham, 2019).

2.2.1 Justification for method choice

SDA was used in this project due to the various advantages it provides. Although it is not directly under Quantitative or Qualitative research, it provides benefits such as; Reducing the time taken in collecting data, Convenience of utilizing existing data, Cheaper in comparison to primary data collection and more. Due to these factors, we decided that SDA would be the more convenient and time-efficient option in comparison to other research methods.

2.2.2 How were these methods applied to analyse the research question or problem?

The development of research questions and hypotheses guided the experimental design and offered a framework for assessing the influence of musical characteristics on popularity, laying the groundwork for the analysis in this study. These guiding concepts played a crucial role in determining which model parameters and visualization techniques worked best.

Additionally, the data collection strategy described in section 2.1 was crucial in creating the dataset needed for this project. One of the most important steps in creating an enriched dataset that was specifically designed to fulfil the study's objectives was combining carefully cleaned datasets that had been feature-selected via a rigorous process. This step was vital in order to discover important features that will correlate most with our research questions and avoid data inaccuracies and consistency issues.

Section 2.2 provides further information on the use of Secondary Data Analysis (SDA) as the primary research method in the field of quantitative research. SDA was a wise strategic decision because it aligned with the goals of the study. It was used as a potent tool to extract knowledge from pre-existing datasets and explore the complex relationships between musical features and popularity, rather than as a stand-alone research method. Utilizing this method we were able to be time and cost efficient, and achieve the best possible visualization of our hypotheses.

In summary, the methodologies outlined in the hypothesis and research questions section served as the foundation for designing experiments aimed at visualizing the impact of individual musical features on song popularity. The utilization of predictive algorithms allowed for a dynamic exploration of changing trends within the dataset. Visualization techniques played a pivotal role in capturing the nuanced relationships between musical features and popularity dynamics. Through graphical representations and data visualizations, the evolving trends were brought to life, offering a rich and intuitive understanding of the intricate connections between different musical elements and the changing landscape of song popularity. Concurrently, the implementation of data preparation techniques, including thorough data cleaning, judicious feature selection, and strategic dataset merging, proved indispensable in sculpting a dataset conducive to robust predictive modeling. This iterative process facilitated a meaningful exploration of the evolving musical landscape and its correlation with song popularity.

3.0 Results

3.1 Restate the study purpose

This research aims to explore the dynamic evolution of musical genres and production techniques over the past century, focusing on the fundamental elements contributing to the success of songs across different eras. The study delves into the interplay between cultural shifts and musical evolution, emphasizing the nuanced exploration of energy, tempo, rhythm, and lyrical content. It also aims to decipher the factors resonating with audiences, highlighting the symbiotic relationship between artistic expression and audience preference. The significance of the research lies in understanding the enduring appeal of music from earlier eras in the contemporary landscape. Building on existing literature, the study considers social factors and identifies key features crucial for creating successful songs. Hypotheses and research questions are formulated to test assertions about the impact of musical features on popularity, the influence of feature combinations, the existence of an optimal song duration, and potential evolution of song characteristics over the past century. The research adopts a comprehensive approach, utilizing a diverse dataset and employing Machine Learning algorithms and data visualization techniques. Ultimately, the study aims to contribute valuable insights for artists and producers, offering practical guidance in navigating the ever-changing music industry.

3.2 Key findings

In this section, most visualizations focus on Popularity which is our dependent variable. As highlighted in section 1.1, the study aims to identify key elements influencing the success of songs over time. Hence, popularity is being a crucial factor in gauging their success across different eras.

3.2.1 Popularity Vs Duration (Spotify)

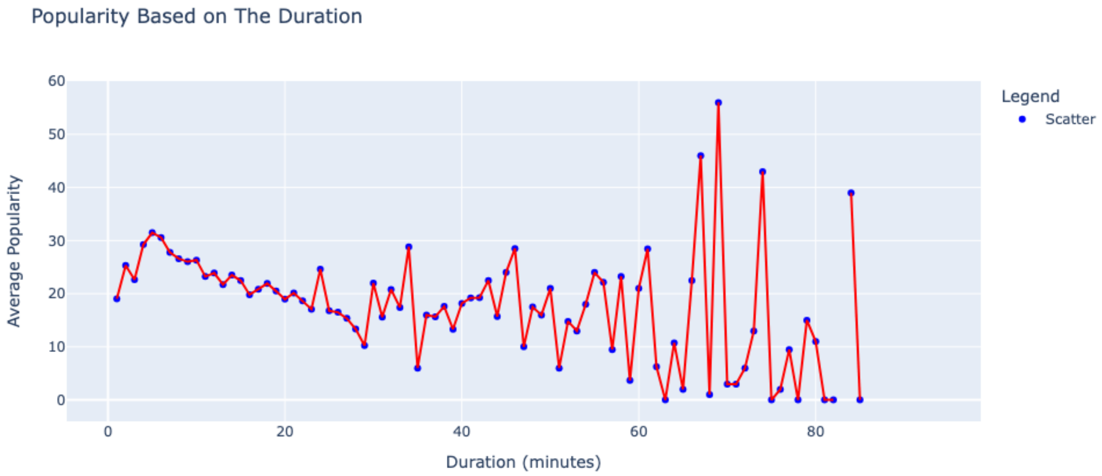


Figure 7.0 - Popularity Vs Duration

In relation to Figure 7.0, identifying outliers poses a challenge for researchers. Consequently, a best-fit line was created using Ordinary Least Squares (OLS) regression to address this difficulty as illustrated in Figure 8.0.

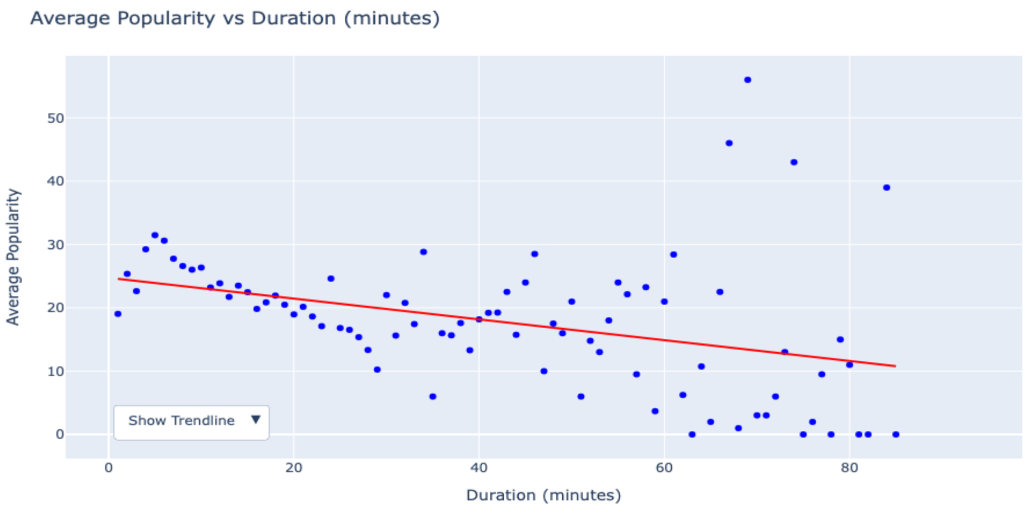


Figure 8.0 - Popularity Vs Duration with OLS Implemented

3.2.2 Feature Importance Vs Popularity

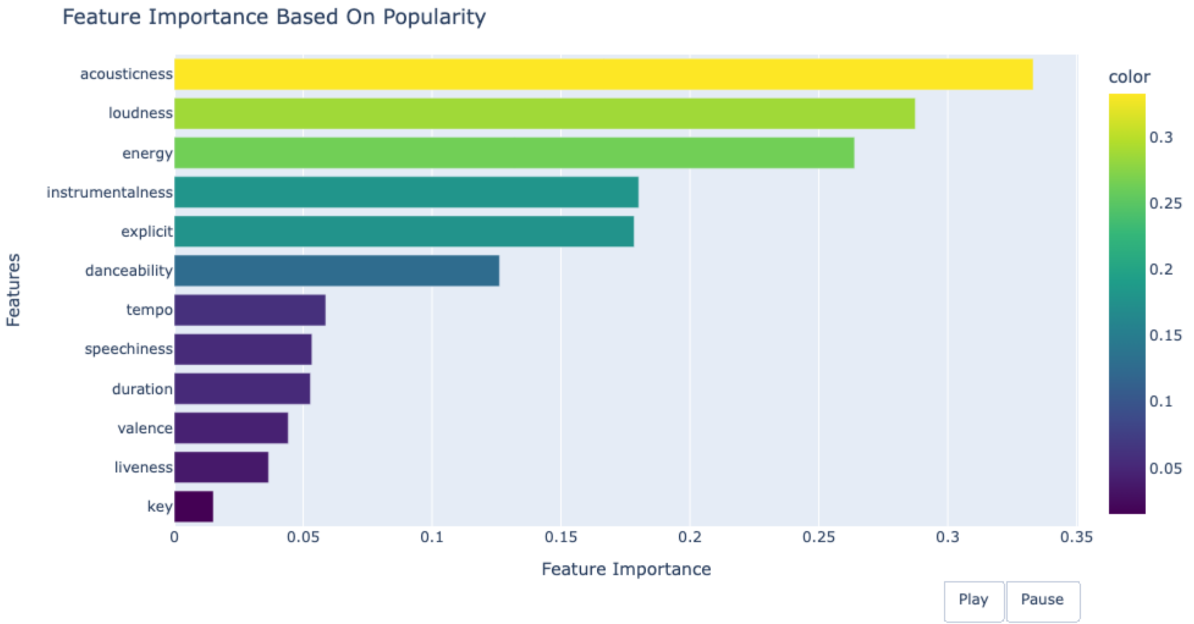


Figure 9.0 - Feature Importance Vs Popularity

Figure 9.0 depicted the significance of individual features according to their popularity, arranged in descending order from the most popular feature, acousticness, followed by loudness, energy, and so forth.

3.2.3 Explicit Vs Popularity

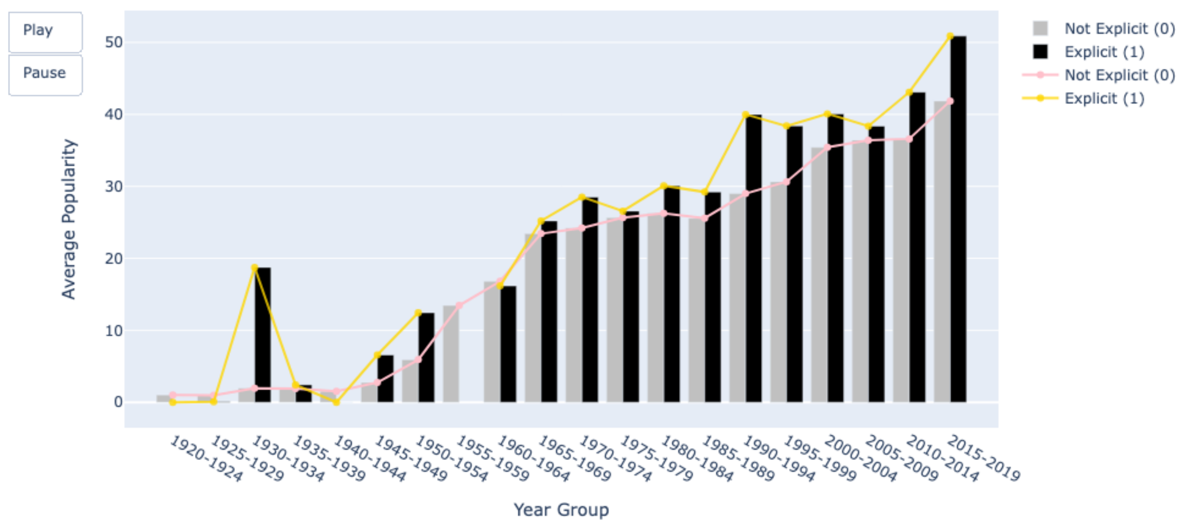


Figure 10.0 - Explicit Vs Popularity

Figure 10.0 depicted the explicitness versus popularity chart, revealing a noticeable upward trend for both explicit and non-explicit content from 1920 to 2019. Furthermore, researchers observed that the distinction between the yellow line representing explicit content and the pink line representing non-explicit content is not substantial, indicating that explicitness does not significantly impact the popularity of a song.

3.2.4 Song Clusters Vs Popularity

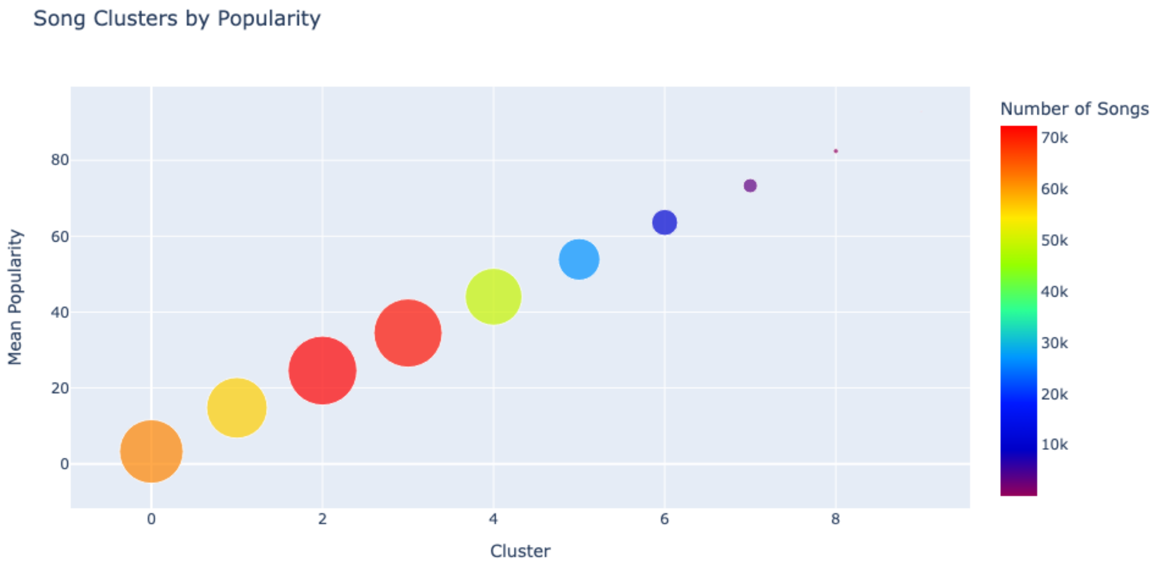


Figure 11.0 - Song Clusters Vs Popularity

Figure 11.0 has depicted eight distinct song clusters determined by popularity. These clusters are computed using the mean of 'popularity' and the count of 'name' (representing the number of songs). Notably, cluster 8 stands out with the highest mean popularity, reaching 82.41. Consequently, the researchers have chosen to calculate the mean duration for all songs in that cluster to determine the optimal duration for creating a song with the potential for high popularity as illustrated in Figure 12.0.

```
average_duration_cluster_8 = cluster_8_songs['duration'].mean()
print(f"Average Duration for Cluster 8 Songs: {average_duration_cluster_8:.2f} minutes")
```

```
Average Duration for Cluster 8 Songs: 3.50 minutes
```

Figure 12.0 - Cluster 8 Songs Duration (Mean)

Based on this statistic, it can be concluded that songs lasting 3.50 minutes tend to garner the highest popularity. Consequently, music creators may opt to produce songs of this duration to attract a larger audience, as it aligns with the preference of the majority. Moreover, as additional evidence supporting the idea that songs lasting approximately 3.50 minutes enjoy greater popularity, the researcher employed KMeans to categorize songs into clusters based on their popularity, as depicted in Figure 13.0.

```
Cluster 1 Center - Duration: 4.09 minutes, Popularity: 11.54
Cluster 2 Center - Duration: 3.40 minutes, Popularity: 49.28
Cluster 3 Center - Duration: 9.02 minutes, Popularity: 28.41
Cluster 4 Center - Duration: 32.79 minutes, Popularity: 16.87
Cluster 5 Center - Duration: 3.67 minutes, Popularity: 64.39
Cluster 6 Center - Duration: 2.28 minutes, Popularity: 15.88
Cluster 7 Center - Duration: 61.70 minutes, Popularity: 15.65
Cluster 8 Center - Duration: 1.72 minutes, Popularity: 32.97
Cluster 9 Center - Duration: 5.38 minutes, Popularity: 45.59
Cluster 10 Center - Duration: 17.11 minutes, Popularity: 20.44
Cluster 11 Center - Duration: 5.09 minutes, Popularity: 27.28
Cluster 12 Center - Duration: 6.30 minutes, Popularity: 9.12
Cluster 13 Center - Duration: 3.71 minutes, Popularity: 37.11
Cluster 14 Center - Duration: 2.80 minutes, Popularity: 2.57
Cluster 15 Center - Duration: 3.47 minutes, Popularity: 24.58
```

Figure 13.0 - KMeans Without Using Methods To Choose Number of K

Figure 13.0 represents the clusters generated by KMeans without employing any method to determine the appropriate number of K. In order to obtain more precise results, the researchers opted to utilize the elbow method to select the number of K, as demonstrated in Figures 14.0 and 15.0.

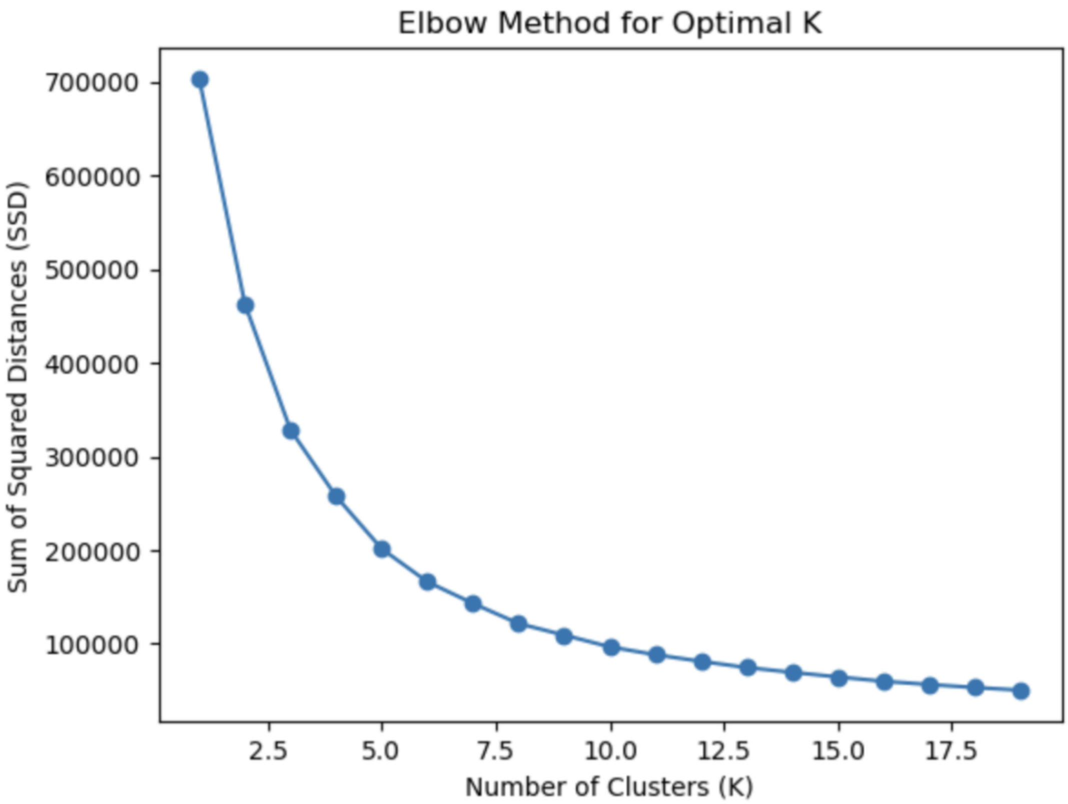


Figure 14.0 - Elbow Method To Select Number of K

As shown in Figure 14.0, the ideal choice for the number of clusters (K) was determined to be 8.

```
Cluster 1 Center - Duration: 4.16 minutes, Popularity: 38.52
Cluster 2 Center - Duration: 2.33 minutes, Popularity: 28.41
Cluster 3 Center - Duration: 52.82 minutes, Popularity: 17.04
Cluster 4 Center - Duration: 3.68 minutes, Popularity: 56.73
Cluster 5 Center - Duration: 7.79 minutes, Popularity: 25.45
Cluster 6 Center - Duration: 2.89 minutes, Popularity: 5.61
Cluster 7 Center - Duration: 19.94 minutes, Popularity: 19.63
Cluster 8 Center - Duration: 4.36 minutes, Popularity: 18.95
```

Figure 15.0 - KMeans With The Use of Method To Choose Number of K

As depicted in Figure 15.0, it is evident that Cluster 4 exhibited the highest popularity with a mean of 56.73, and its duration was approximately 3.68 minutes. This finding substantiates the earlier assumptions made in Figure 12.0, confirming that songs crafted around the 3.50 minutes mark indeed achieve the highest popularity. Moreover, a three-dimensional graph created with KMeans, depicted in Figure 16.0, was created to offer various insights, including observations on cluster patterns, the degree of separation between clusters, and more. It's evident that cluster 4 (maroon colour) stands out with the most tightly grouped points, signifying that the songs in this cluster exhibit more uniform characteristics compared to other clusters. In contrast, the other clusters show more separated points, suggesting a greater diversity within those clusters.

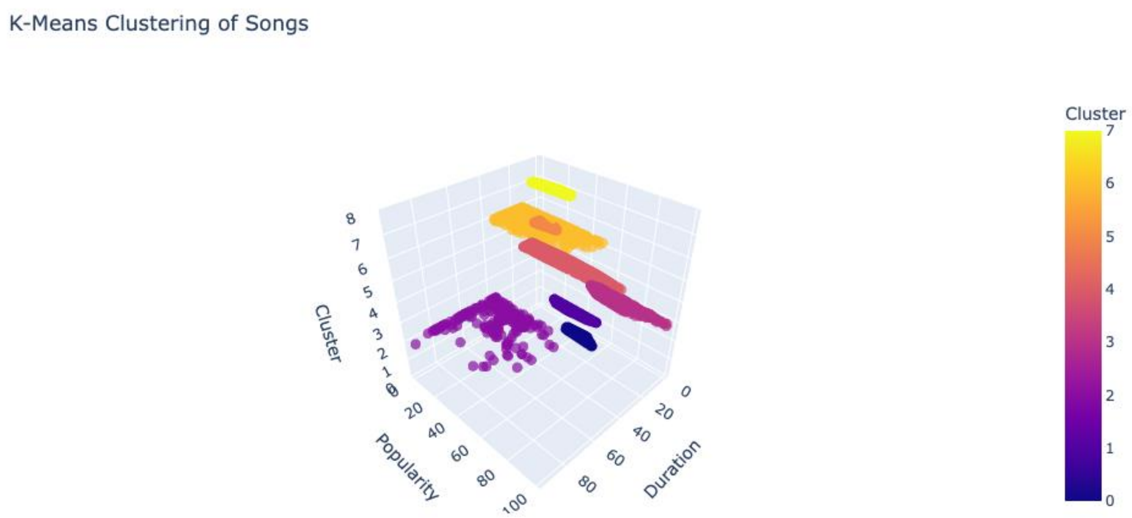


Figure 16.0 - 3D Graph For Song Clusters

3.2.5 Top 10 Artists Vs Followers in Spotify

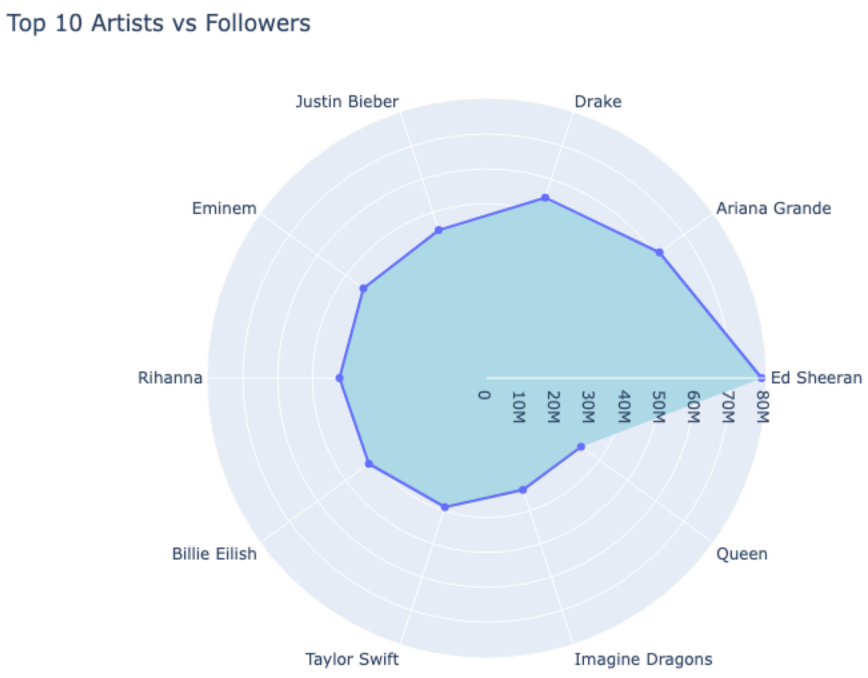


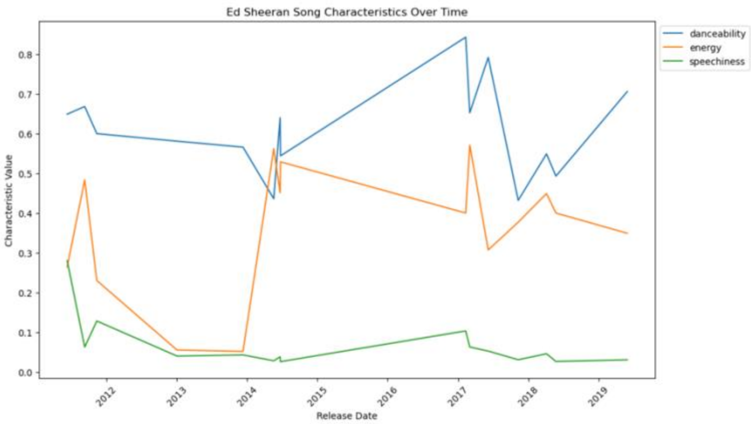
Figure 17.0 - Top 10 Artists Vs Followers in Spotify

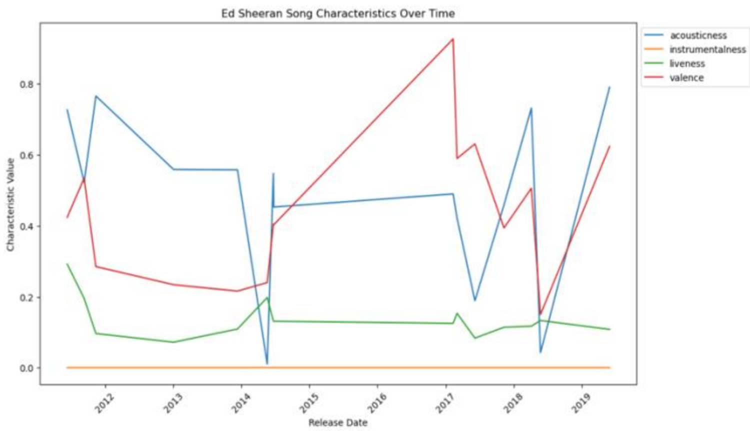
Figure 17.0 displays the Top 10 Artists ranked by followers, with Ed Sheeran leading the list, followed by Ariana Grande, Drake, and others. Additionally, the researchers have presented the average values for various music characteristics, encompassing danceability, energy, speechiness, acousticness, instrumentalness, liveliness, valence, tempo, key, and loudness as illustrated in Figure 18.0 to Figure 27.0 for the Top 5 artists with the most followers in Spotify. This statistical information is intended to assist music creators in tailoring their songs to personal preferences. For example, if a music creator seeks to produce music akin to Ed Sheeran's, known for genres like pop, folk-pop, or soft rock, these statistics can guide the creation of similar music.

Ed Sheeran

```
Average danceability: 0.6101242424242425
Average energy: 0.36527404040404043
Average speechiness: 0.06625641414141414
Average acousticness: 0.48460666666666674
Average instrumentalness: 2.0326424242424236e-05
Average liveness: 0.1372319191919192
Average valence: 0.4375181818181818
Average tempo: 113.24173232323233
Average key: 6.842424242424243
Average loudness: -9.398309595959596
```

Figure 18.0 - Average Values of Music Characteristics For Ed Sheeran





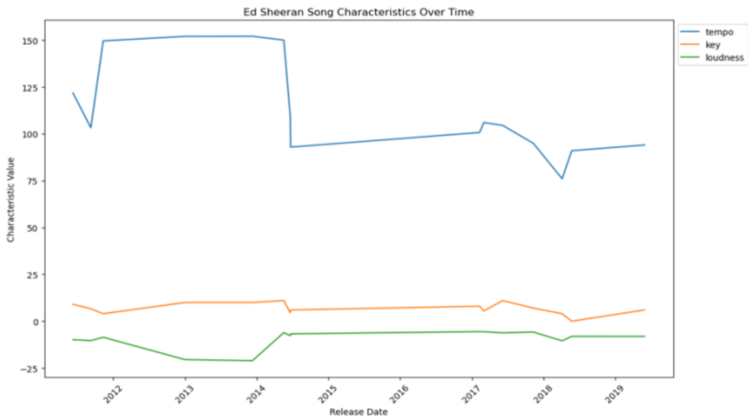


Figure 19.0 - Ed Sheeran Song Characteristics Over Time

Ariana Grande

```
Average danceability: 0.6364916666666667
Average energy: 0.6222722222222222
Average speechiness: 0.09249222222222221
Average acousticness: 0.20252305555555555
Average instrumentalness: 0.0006108144166666666
Average liveness: 0.2058311111111111
Average valence: 0.4704083333333333
Average tempo: 122.05839166666667
Average key: 5.977777777777778
Average loudness: -5.845752777777777
```

Figure 20.0 - Average Values of Music Characteristics For Ariana Grande

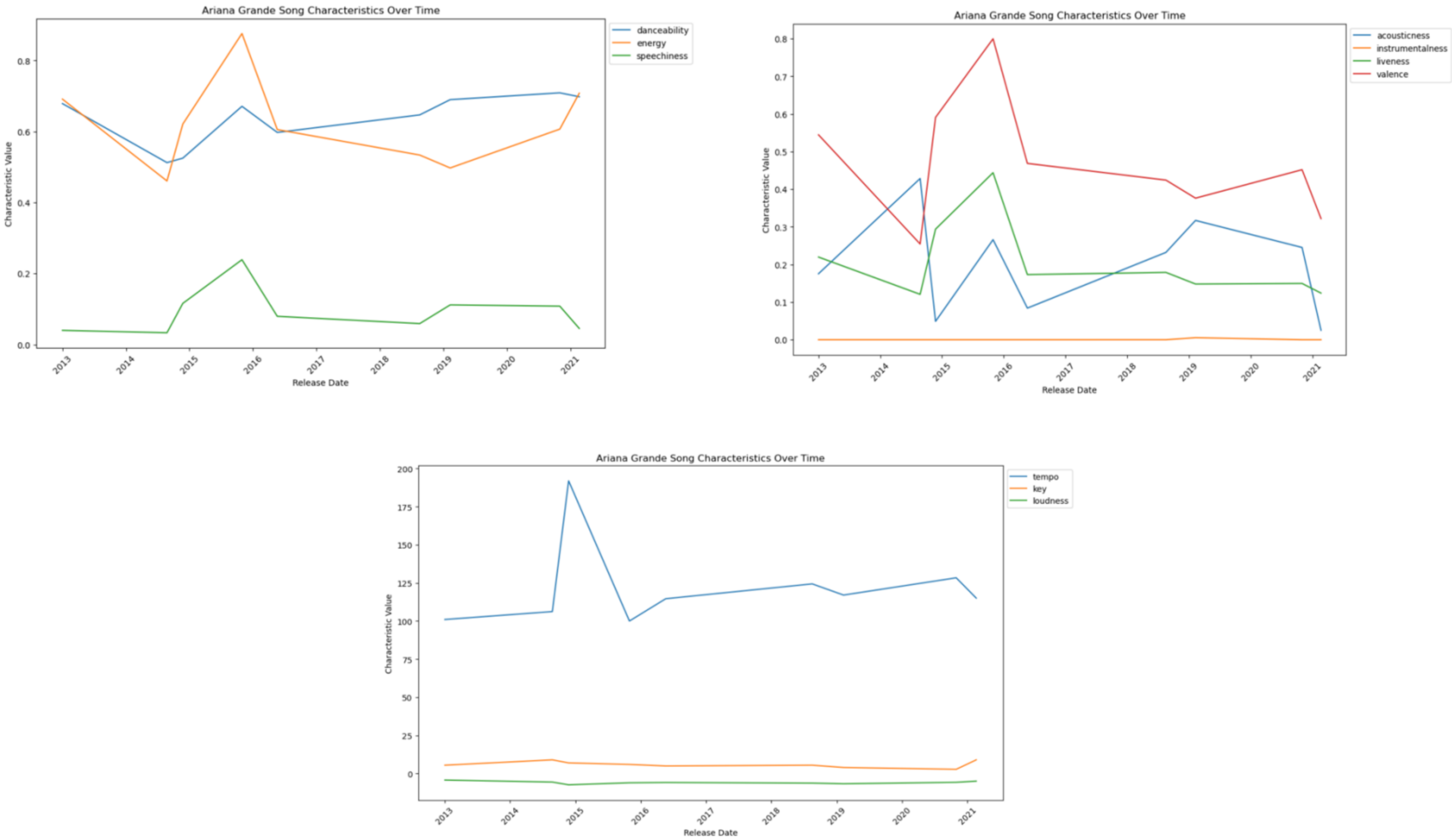


Figure 21.0 - Ariana Grande Song Characteristics Over Time

Drake

```
Average danceability: 0.6811075156985872
Average energy: 0.5558854461015175
Average speechiness: 0.21959447475143906
Average acousticness: 0.305928280873888
Average instrumentalness: 0.0017430243502093144
Average liveness: 0.16371287545787547
Average valence: 0.43257849947671373
Average tempo: 116.07188688513868
Average key: 5.096487441130298
Average loudness: -7.684185171376242
```

Figure 22.0 - Average Values of Music Characteristics For Drake

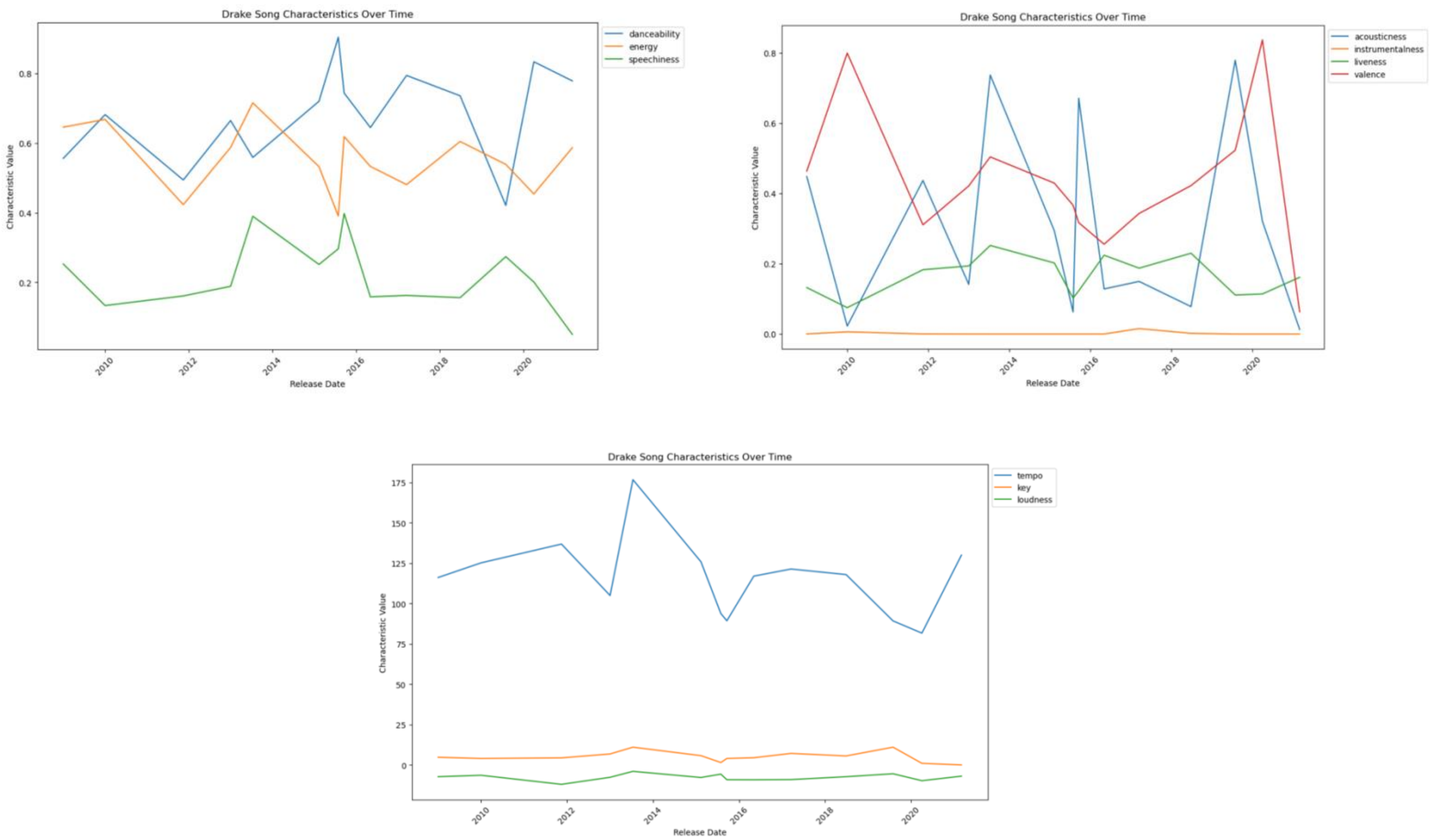


Figure 23.0 - Drake Song Characteristics Over Time

Justin Bieber

```
Average danceability: 0.6278299319727891
Average energy: 0.5220773809523809
Average speechiness: 0.06391787414965985
Average acousticness: 0.36654545068027217
Average instrumentalness: 0.00029537210884353741
Average liveness: 0.14748212585034012
Average valence: 0.5232750000000002
Average tempo: 117.15049302721084
Average key: 5.611734693877551
Average loudness: -7.377285204081631
```

Figure 24.0 - Average Values of Music Characteristics For Justin Bieber

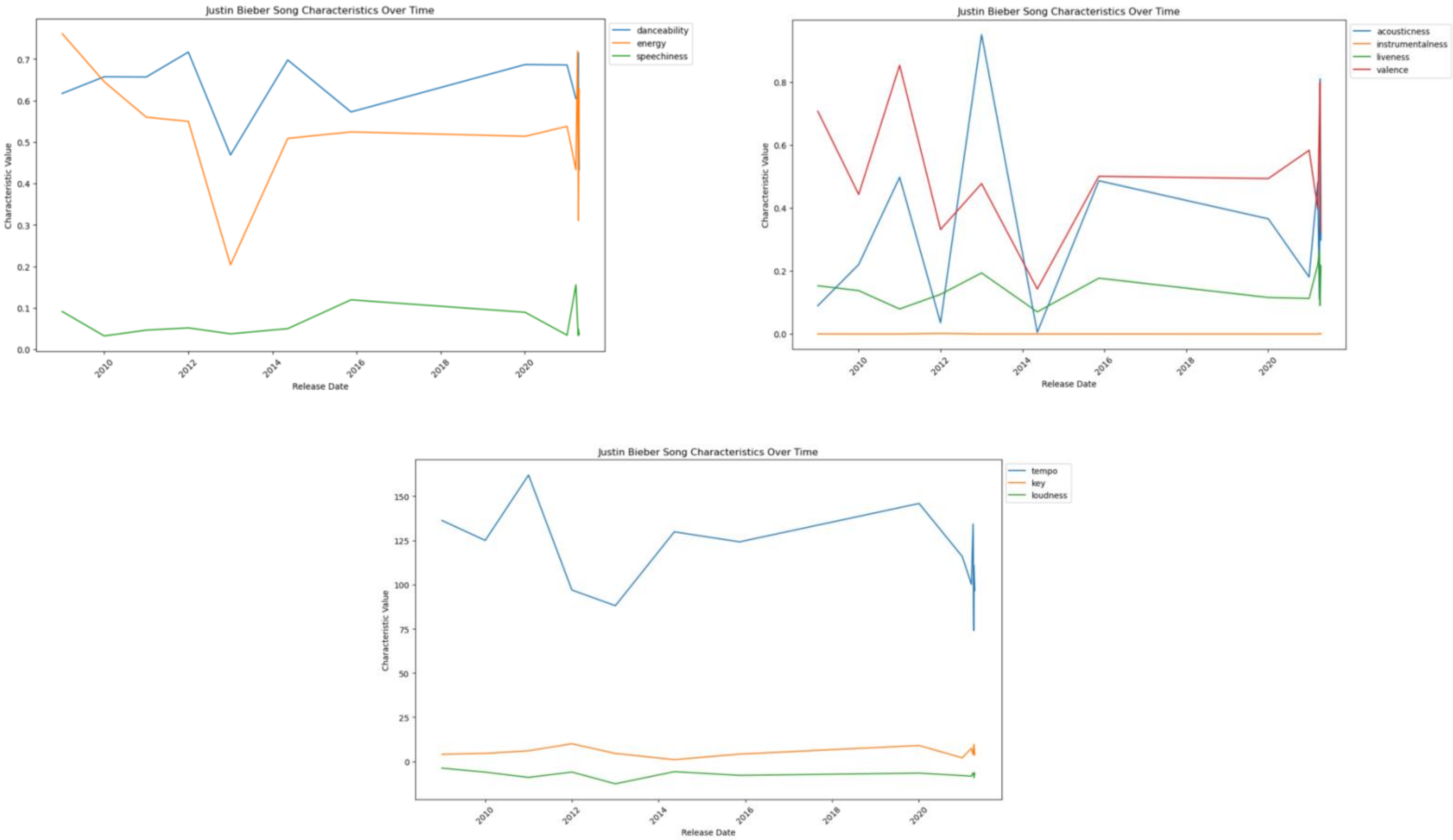


Figure 25.0 - Justin Bieber Song Characteristics Over Time

Eminem

```
Average danceability: 0.7280574074074074
Average energy: 0.7541617283950618
Average speechiness: 0.26814000000000004
Average acousticness: 0.11001537037037036
Average instrumentalness: 0.019169781728395506
Average liveness: 0.24624324074074078
Average valence: 0.4612006172839506
Average tempo: 110.97063842592593
Average key: 5.778858024691358
Average loudness: -4.900637191358026
```

Figure 26.0 - Average Values of Music Characteristics For Eminem

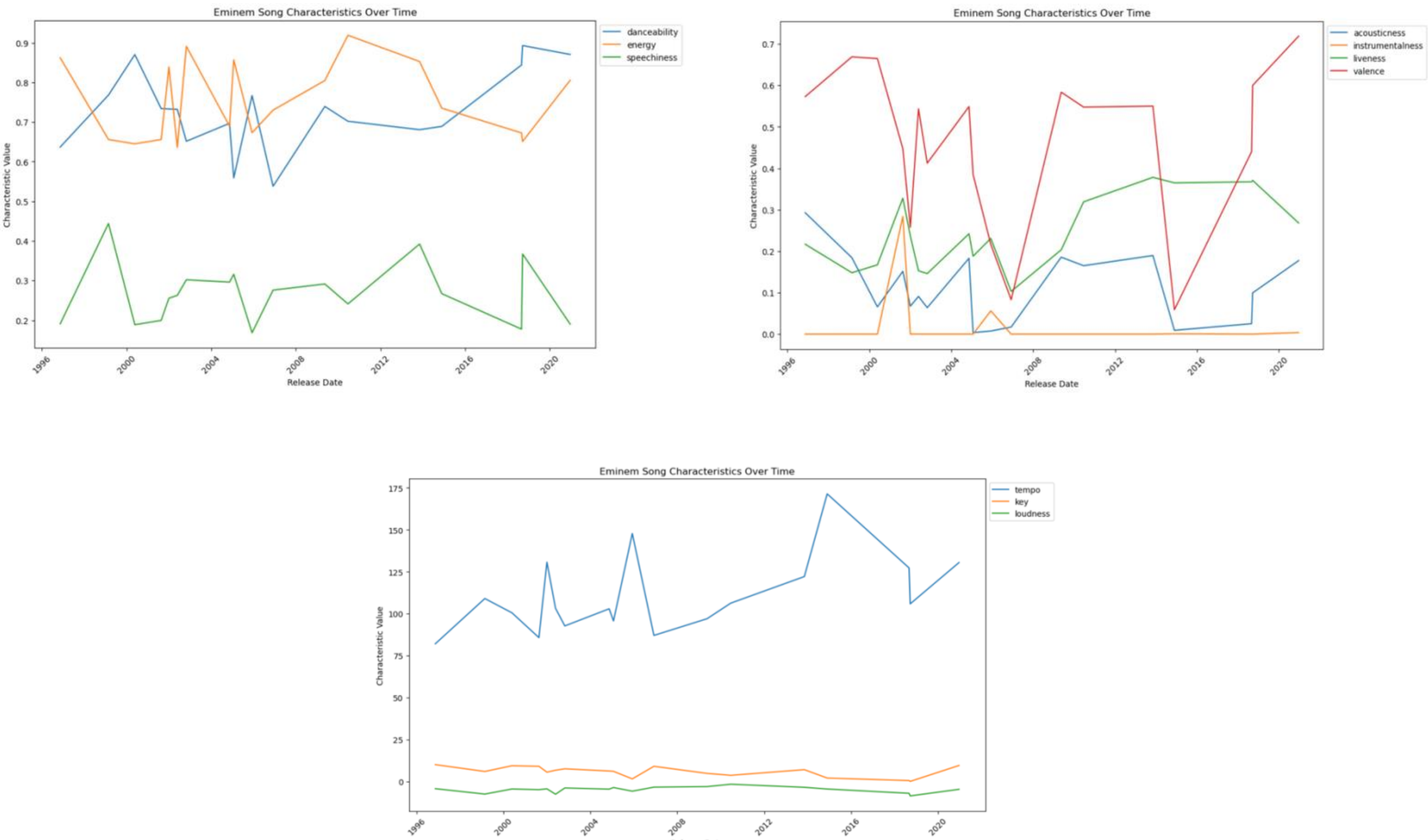


Figure 27.0 - Eminem Song Characteristics Over Time

3.2.6 Popularity Vs Duration (YouTube)

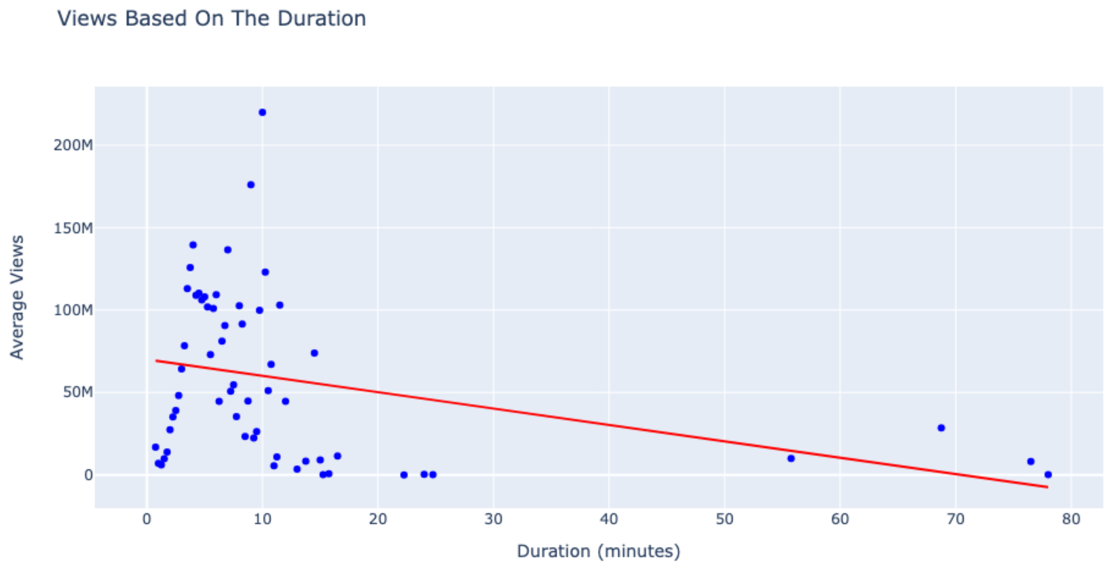


Figure 28.0 - Popularity Vs Duration

Figure 28.0 shows the average views (popularity) plotted against duration, with OLS used to identify outliers. Additionally, in Figure 29.0, after excluding outliers, it's evident that the peak average views were 139 million, corresponding to a duration of 4 minutes. This supports our findings in Figures 12.0 and 15.0, indicating that durations around 3.50 to 4 minutes tend to have the highest popularity. In summary, the optimal duration for maximum popularity is likely to fall between 3.50 and 4 minutes.

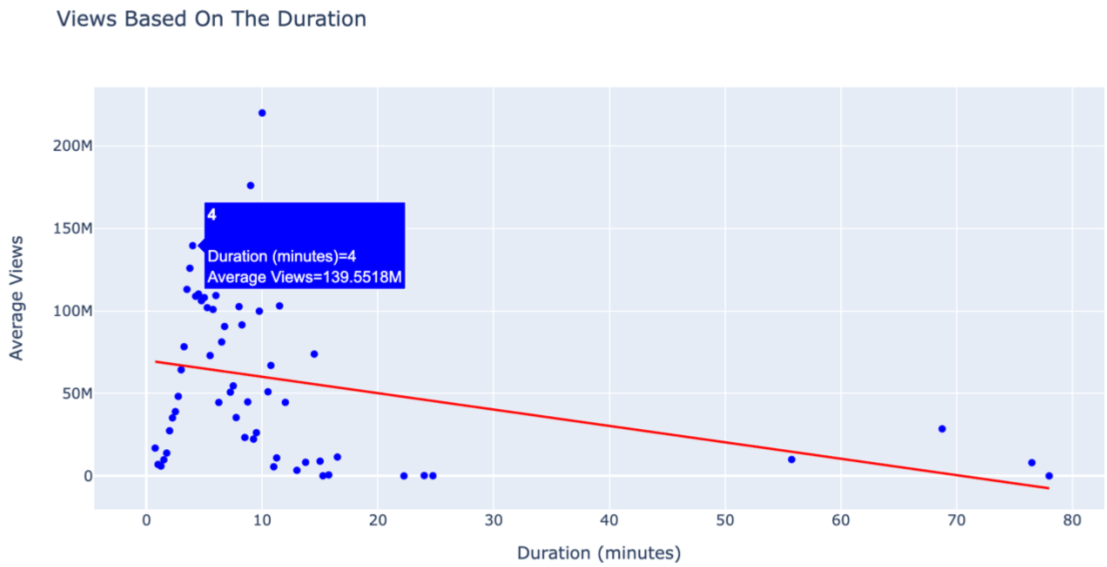


Figure 29.0 - Popularity Vs Duration with OLS Implemented

3.2.7 Top 10 Artists Vs Followers in YouTube

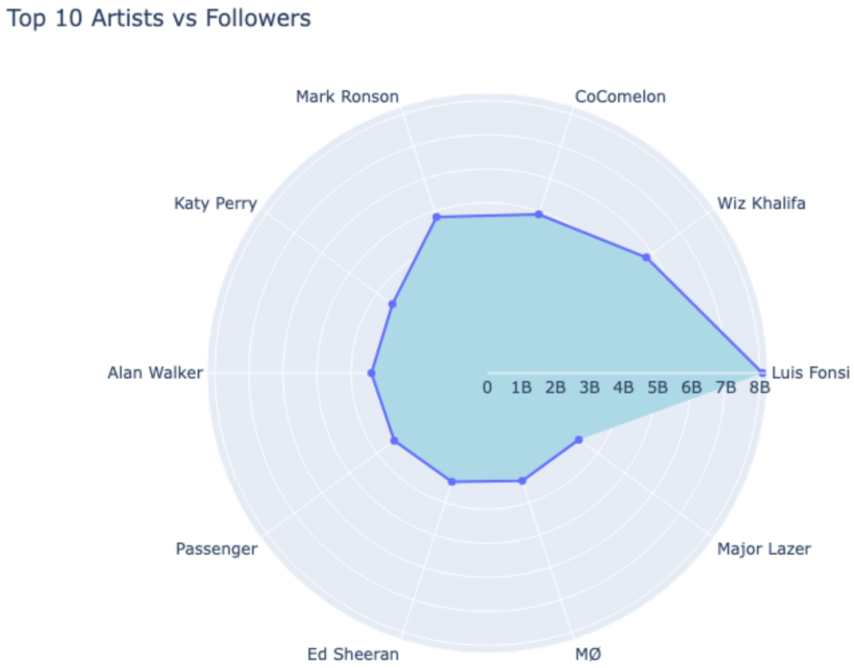


Figure 30.0 - Top 10 Artists Vs Followers in YouTube

Figure 30.0 displays the Top 10 Artists ranked by followers, with Luis Fonsi leading the list, followed by Wiz Khalifa, CoComelon, and others. Additionally, the researchers have presented the average values for various music characteristics, encompassing danceability, energy, speechiness, acousticness, instrumentalness, liveliness, valence, tempo, key, and loudness as illustrated in Figure 31.0 to Figure 36.0 for the Top 5 artists with the most followers in YouTube. However, only three artists - Luis Fonsi, Wiz Khalifa, and Katy Perry - will be showcased, excluding Mark Ronson and CoComelon due to constraints within the dataset. This statistical information is intended to assist music creators in tailoring their songs to personal preferences. For example, if a music creator seeks to produce music akin to Luis Fonsi's, known for genres like pop, these statistics can guide the creation of similar music. Moreover, the rationale behind presenting the statistics for Spotify, as shown in Figures 17.0 to 27.0, is to guide music creators when publishing their songs on the platform. By following the statistics of the specific platform they choose for song release, creators can increase their chances of gaining popularity. For example, if a creator intends to release songs on Spotify, they may find it more advantageous to emulate the statistics of Ed Sheeran, who has greater popularity on Spotify compared to Luis Fonsi, especially considering the differences in popularity between the two platforms.

Luis Fonsi

```
Average danceability: 0.5087875
Average energy: 0.6087083333333333
Average speechiness: 0.038872083333333335
Average acousticness: 0.31153333333333333
Average instrumentalness: 0.0
Average liveness: 0.15039166666666667
Average valence: 0.39712916666666664
Average tempo: 144.6866875
Average key: 4.275
Average loudness: -6.447245833333334
```

Figure 31.0 - Average Values of Music Characteristics For Luis Fonsi

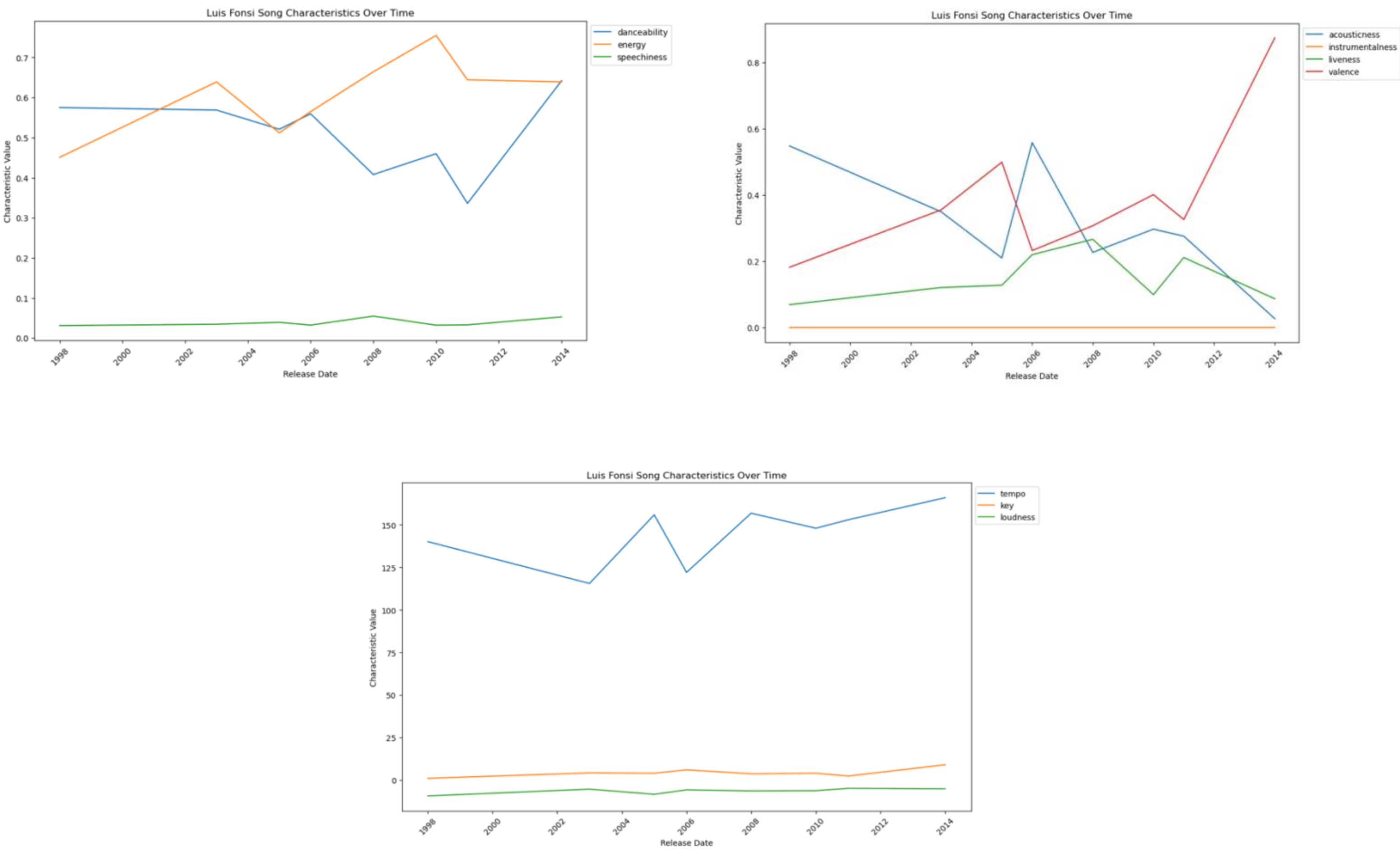


Figure 32.0 - Luis Fonsi Song Characteristics Over Time

Wiz Khalifa

```
Average danceability: 0.7196250000000001
Average energy: 0.7100833333333333
Average speechiness: 0.08350416666666666
Average acousticness: 0.07173208333333334
Average instrumentalness: 4.144999999999999e-06
Average liveness: 0.2378916666666667
Average valence: 0.5443333333333333
Average tempo: 151.45041666666668
Average key: 7.166666666666667
Average loudness: -4.916416666666667
```

Figure 33.0 - Average Values of Music Characteristics For Wiz Khalifa

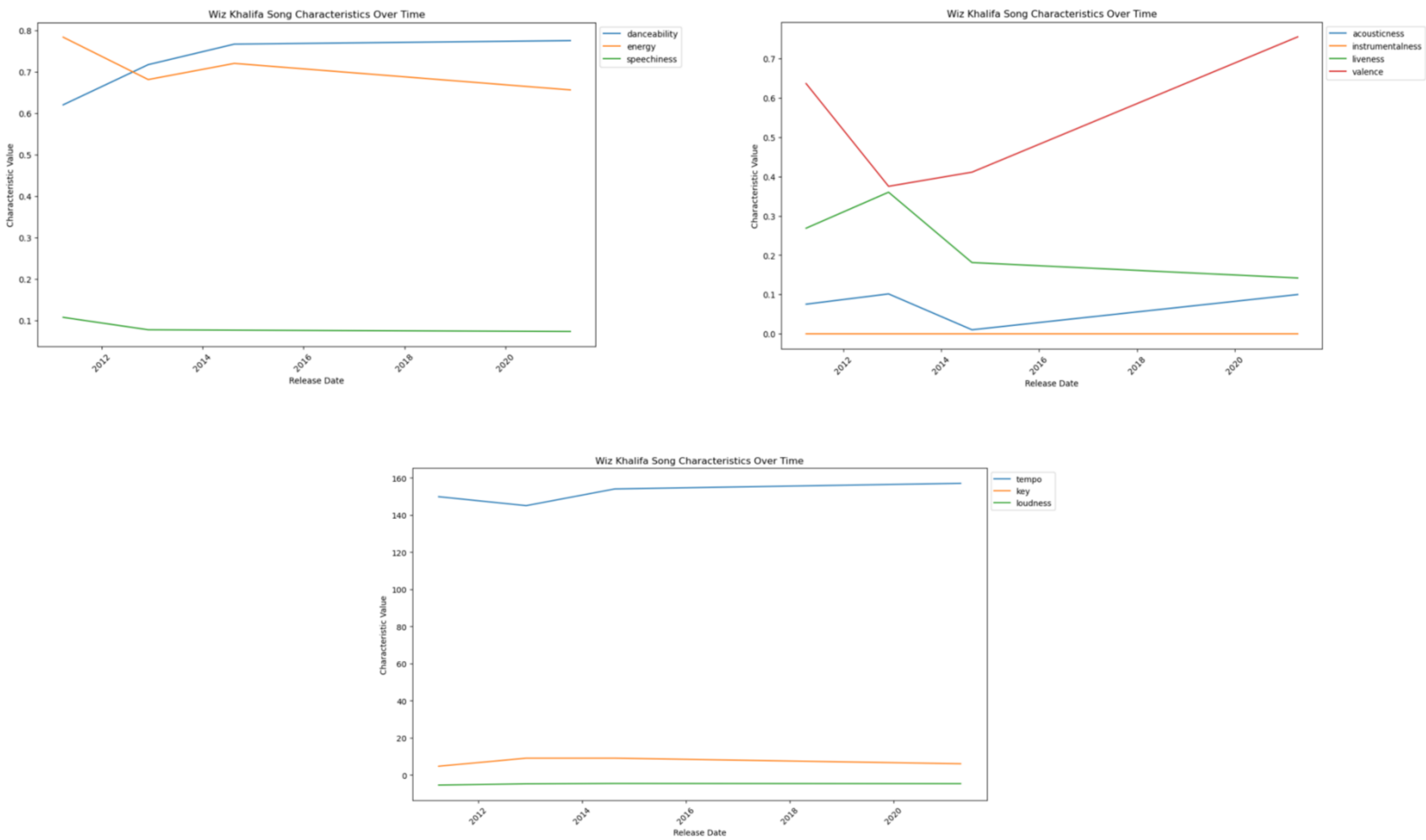


Figure 34.0 - Wiz Khalifa Song Characteristics Over Time

Katy Perry

```
Average danceability: 0.5881574074074074
Average energy: 0.7087222222222223
Average speechiness: 0.04400555555555555
Average acousticness: 0.10804958796296296
Average instrumentalness: 1.7148148148148148e-06
Average liveness: 0.19758240740740737
Average valence: 0.4718888888888897
Average tempo: 129.50292592592595
Average key: 5.694444444444445
Average loudness: -5.202851851851851
```

Figure 35.0 - Average Values of Music Characteristics For Katy Perry

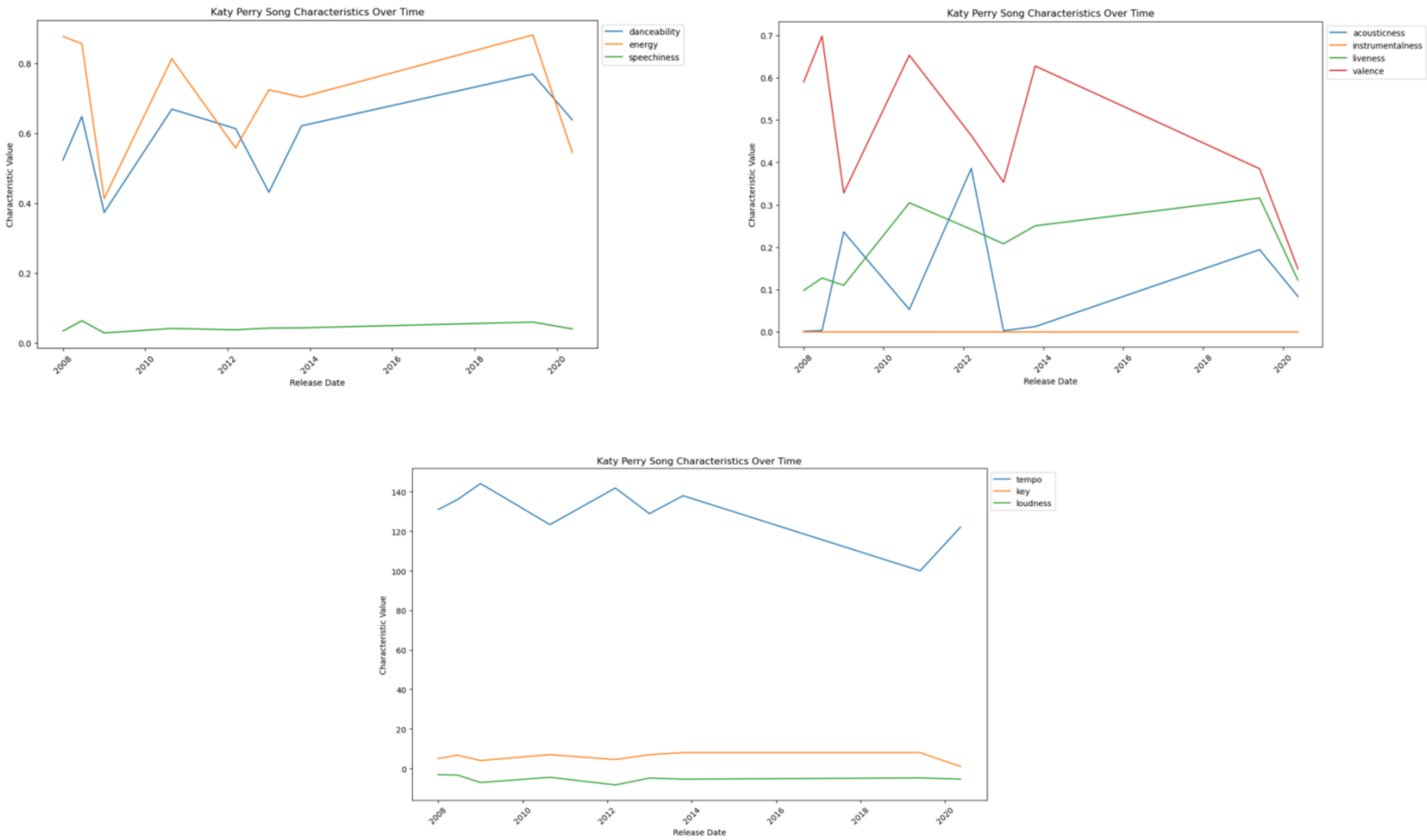


Figure 36.0 - Katy Perry Song Characteristics Over Time

3.3 Report of insignificant research findings

In the course of this study, a comprehensive and targeted research approach was employed to address specific objectives. Through rigorous analysis and careful consideration, it is noteworthy that no findings deemed insignificant were encountered. The research focused on obtaining meaningful and relevant insights, contributing to the absence of any outcomes classified as insignificant within the scope of this investigation.

3.4 Summary of key study outcomes

In this study, the objective is to investigate the dynamic evolution of musical genres and production techniques over the last century, with a particular emphasis on the essential components that contribute to the success of songs across various time periods. Hence, the researchers have identified specific elements that play a significant role in enhancing the success of songs. These elements are outlined below:

1. Songs exhibiting elevated acousticness, loudness, and energy, as illustrated in Figure 9.0, where these three attributes emerge as the most crucial features influencing popularity.
2. Songs falling within the duration range of 3.50 minutes to 4.0 minutes, as depicted in Figures 12.0, 15.0, and 29.0.
3. Examining musical traits derived from the top artists on Spotify and YouTube, as depicted in Figures 18.0 to 27.0 for Spotify and Figures 31.0 to 36.0 for YouTube.

4.0 Discussion

4.1 Key findings interpretation

The key findings of this research underscore the intricate relationship between musical characteristics and the success of songs across different eras. The emphasis on acousticness, loudness, and energy, as highlighted in Figure 9.0, points to these attributes as pivotal factors influencing the popularity of songs. This suggests that a higher degree of acousticness, loudness, and energy contributes significantly to the overall appeal of a song.

Furthermore, the identified optimal song duration, falling within the range of 3.50 minutes to 4.0 minutes, as depicted in Figures 12.0, 15.0, and 29.0, provides valuable insights for artists and producers. This finding suggests that songs within this duration bracket are more likely to resonate with audiences, emphasizing the importance of conciseness in musical compositions.

The exploration of musical traits from top artists on Spotify and YouTube, as illustrated in Figures 18.0 to 27.0 for Spotify and Figures 31.0 to 36.0 for YouTube, reveals patterns and characteristics that contribute to the success of these artists. This insight provides a nuanced understanding of the elements that resonate with contemporary audiences on these platforms.

In summary, these key findings contribute to a comprehensive understanding of the dynamic evolution of musical genres and production techniques. The identified musical elements, song duration preferences, and traits of successful artists offer valuable guidance for those navigating the ever-changing landscape of the music industry. The research not only sheds light on the factors contributing to the enduring appeal of music but also provides practical implications for artists and producers aiming to create impactful and resonant musical content.

4.2 Comparative analysis and critical assessment

4.2.1 Compare and contrast to previous study

In comparing the findings of the current research with those of the previous study mentioned in section 1.3, here are few of the comparisons and contrasts that could be drawn.

1. Social Factors and Song Success
    a. Previous study: Acknowledges the influence of social factors on the success and popularity of songs (Salganik, 2006).
    b. Current research: Corroborates this notion by emphasizing the intricate relationship between musical and the overall appeal and popularity of songs.
2. Cyclical Patterns in Musical Genres
    a. Previous study: Identifies cyclical patterns in the life cycles of musical genres, influenced by technological advancements, sociocultural shifts, and evolving consumer tastes (Beckwith, 2016).
    b. Current research: While not explicitly addressing cyclical patterns, the current study delves into specific musical elements and optimal song duration, offering a nuanced understanding of the features contributing to song success.
3. Platform-Specific Insights
    a. Previous study: Does not specifically explore traits from top artists on specific platforms.

b. Current research: Investigates musical traits from top artists on Spotify and YouTube, offering platform-specific insights into the elements that contribute to the success of these artists.

In summary, while the previous study provides a foundation by recognizing the influence of social factors and identifying cyclical patterns in musical genres, the current research builds upon this foundation. It specifically examines musical elements, optimal song duration, and platform-specific insights, offering a more detailed and contemporary understanding of the factors contributing to the success of songs across different eras. Together, these studies provide valuable insights for navigating the complex and ever-changing landscape of the music industry.

4.2.2 Strengths and limitations of this study

The following list includes this study's strengths and limitations:

Strengths

- Comprehensive investigation
    - The study undertakes a comprehensive examination of the dynamic evolution of musical genres and production techniques over the past century, providing a thorough analysis that spans from the Jazz Age of the 1920s to contemporary pop.
- Incorporation of multiple data sources
    - By examining musical traits from top artists on both Spotify and YouTube, the study incorporates a diverse range of data sources, enhancing the richness and relevance of the findings.
- Practical implications
    - The research aims to offer practical guidance for artists and producers, aligning the findings with real-world applications in the ever-changing landscape of the music industry.

Limitations

- Potential bias in data sources
    - The reliance on data from top artists on Spotify and YouTube may introduce a bias toward mainstream or popular content, potentially neglecting certain niche genres or artists that contribute to the overall musical landscape.
- Generalization challenges
    - The findings, while valuable, may not be universally applicable to all genres or cultural contexts. Musical preferences and success factors can vary significantly, and the study's scope might not capture the full diversity of the music landscape.

4.2.3 Discussion of unexpected findings

Due to the fact that this study was primarily focused on confirming or validating specific hypotheses and expanding upon the existing body of knowledge in the realm of popular music dynamics, it appears that there were no particularly unexpected findings in this research. Similarly, in alignment with the comprehensive and targeted research approach employed, and after rigorous analysis and careful consideration, no outcomes classified as insignificant were encountered. The study maintained a focused objective of investigating the dynamic evolution of musical genres and production techniques, particularly emphasizing the essential components contributing to the success of songs across various time periods.

5.0 Discussion

5.1 Summary of hypothesis and purpose of this study

The overarching purpose of this study is to unravel the complex interplay between cultural shifts and musical evolution over the past century, with a focus on discerning the fundamental elements contributing to the success and popularity of songs across different eras. The investigation delves into the dynamic development of musical genres and production techniques, aiming to shed light on the enduring allure of music from the Jazz Age of the 1920s to contemporary pop. To achieve this, the study formulates specific hypotheses and research questions, guiding the exploration of key features influencing song success.

Hypotheses

- Variability in musical feature impact
    - This study posits that the impact on a song's popularity varies across different musical features. Some features are expected to have a more substantial influence than others.
- Combination of features and song popularity
    - The hypothesis suggests that a unique combination of multiple musical features, including tempo, acousticness, etc., increases the probability of creating a popular song. The synergy of these features is anticipated to play a crucial role in song success.
- Optimal song duration
    - The study hypothesizes the existence of an optimal song duration, with songs falling within the range of 3.50 to 4.0 minutes having a higher likelihood of gaining popularity compared to songs outside this range.
- Evolution of song characteristics
    - The hypothesis proposes that the characteristics of songs have evolved over the past century for famous artists, reflecting changes in consumer tastes and cultural preferences.

Purpose of this study

- Comprehensive investigation
    - The research conducts a comprehensive examination, spanning from the exuberant Jazz Age to modern pop, providing an in-depth analysis of the dynamic evolution of musical genres and production techniques.
- Incorporation of diverse data sources
    - By examining musical traits from top artists on both Spotify and YouTube, the study incorporates diverse data sources, enriching the analysis and ensuring relevance to contemporary music platforms.
- Practical implications
    - The study aims to offer practical insights for artists and producers, aligning the findings with real-world applications in the ever-changing landscape of the music industry.

5.2 Significance of this study

This study holds significant implications and practical applications for various stakeholders within the music industry, academia, and the broader community. The insights derived from this research contribute to the understanding of musical dynamics and offer valuable assistance in several key areas:

1. Artistic guidance for musicians and producers
    a. Optimal Song Composition: The identification of crucial musical features such as acousticness, loudness, and energy, along with the optimal song duration, provides practical guidance for artists and producers. This knowledge empowers them to craft songs that align more closely with audience preferences, increasing the likelihood of success.
2. Forecasting trends for future innovation
    a. Predictive Modeling for Future Trends: The study employs machine learning algorithms to project trends in acousticness, energy, and danceability over the next 15 years. This forward-looking approach assists industry professionals in anticipating future musical trends and adapting their strategies accordingly.
3. Cross-platform understanding for streaming services
    a. Platform-Specific Insights: Streaming platforms such as Spotify and YouTube can benefit from platform-specific insights. Understanding the traits of successful artists on each platform allows these services to optimize recommendations, enhance user experiences, and tailor content to the preferences of their user base.

In summary, the significance of this study extends beyond academic inquiry, offering practical implications for musicians, industry professionals, educators, and streaming services. By bridging the gap between artistic expression and audience preference, this research enriches the music landscape and contributes to the ongoing dialogue surrounding the evolving nature of musical success.

5.3 Unanswered questions and future enhancements

According to the research questions in section 1.4, there aren't any unanswered questions. While the current research provides valuable insights into the dynamic evolution of musical genres and the factors influencing song success, there are several avenues for future enhancements and refinements.

1. Qualitative analysis of lyrical content
    a. Inclusion of Lyrical Analysis: While the current research focuses on quantitative musical features, future studies could incorporate a qualitative analysis of lyrical content. Exploring how lyrics contribute to the success of songs could provide a more holistic view of audience preferences.
2. User surveys and feedback
    a. User Feedback Integration: To complement quantitative analyses, incorporating user surveys and feedback could capture subjective preferences and perceptions. Understanding the emotional and personal connections that listeners form with songs adds depth to the analysis.

6.0 References

Millecamp, M., Htun, N. N., Jin, Y., & Verbert, K. (2018). Controlling Spotify recommendations. *Proceedings of the 26th Conference on User Modeling, Adaptation and Personalization*. https://doi.org/10.1145/3209219.3209223

Beckwith, J. (2016, September 7). *The evolution of music genre popularity*. The DataFace. https://thedataface.com/2016/09/culture/genre-lifecycles

Lakshman, M. (2000). Quantitative vs qualitative research methods - springer. https://link.springer.com/content/pdf/10.1007/bf02820690.pdf

Salganik, M. J. (2006, February 10). Reports experimental study of inequality and ... - Princeton University. https://www.princeton.edu/~mjs3/salganik_dodds_watts06_full.pdf

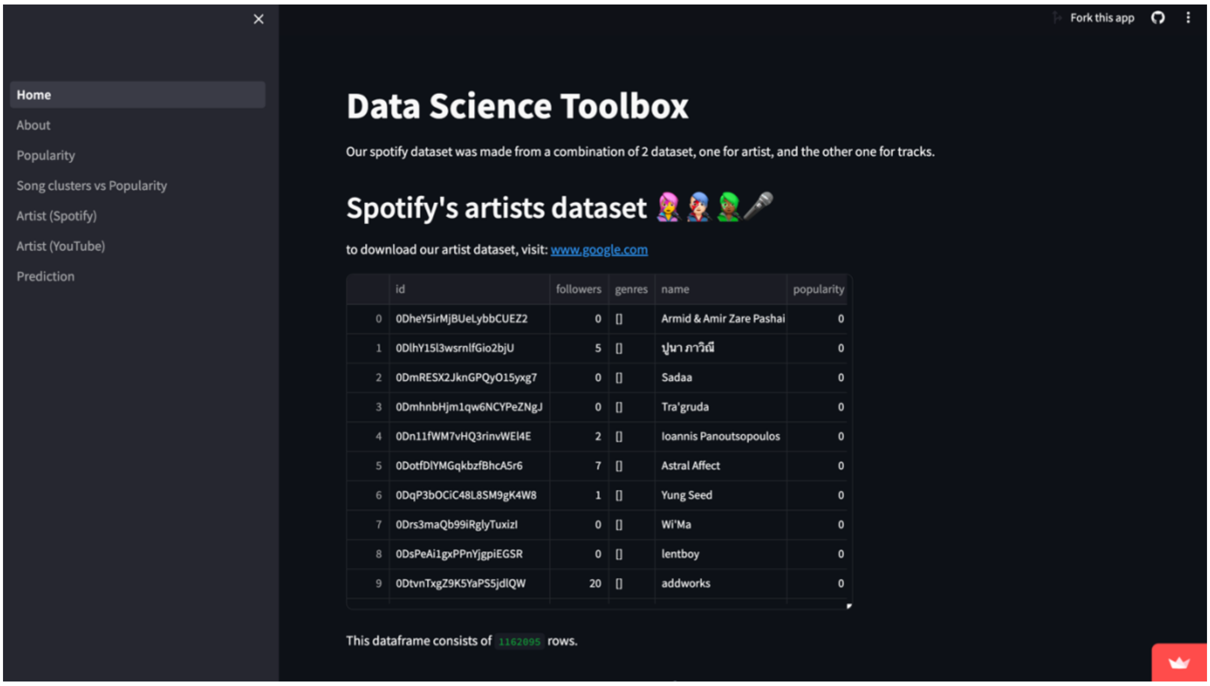
Tech, V. (2023). *Research guides: Research methods guide: Research design & method*. Research Design & Method - Research Methods Guide - Research Guides at Virginia Tech. https://guides.lib.vt.edu/researchmethods/design-method

Welch, G. F., Biasutti, M., MacRitchie, J., McPherson, G. E., & Himonides, E. (2020, May 13). *Editorial: The impact of music on human development and well-being*. Frontiers. https://www.frontiersin.org/articles/10.3389/fpsyg.2020.01246/full

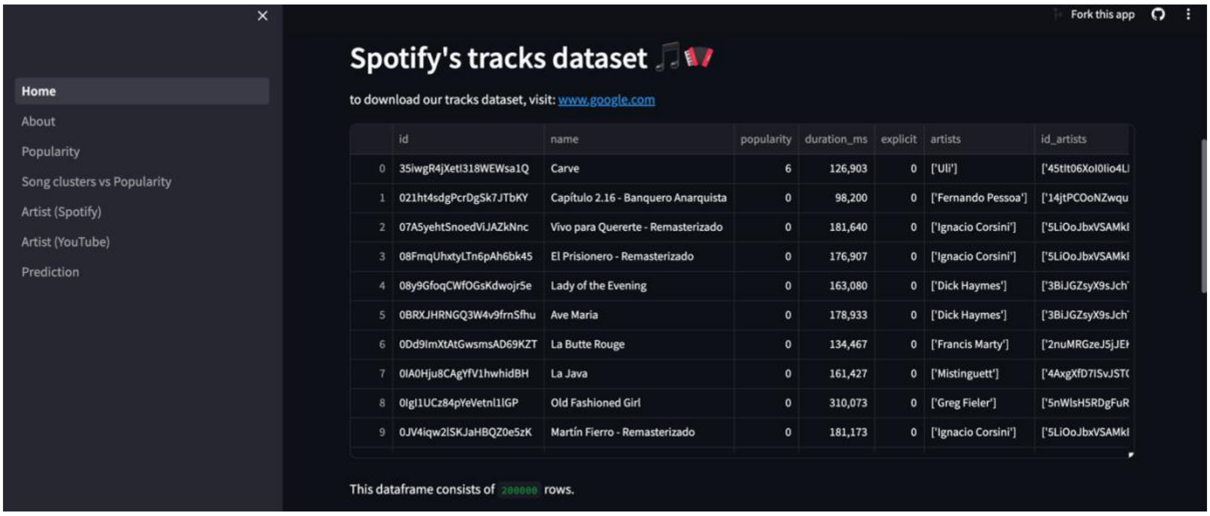
Wickham, R. J. (2019). *Secondary Analysis Research*. Journal of the advanced practitioner in oncology. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7520737/#:~:text=Most%20SDAs%20use%20quantitative%20data,%2C%20a%20medical%20record%20review).

Appendix

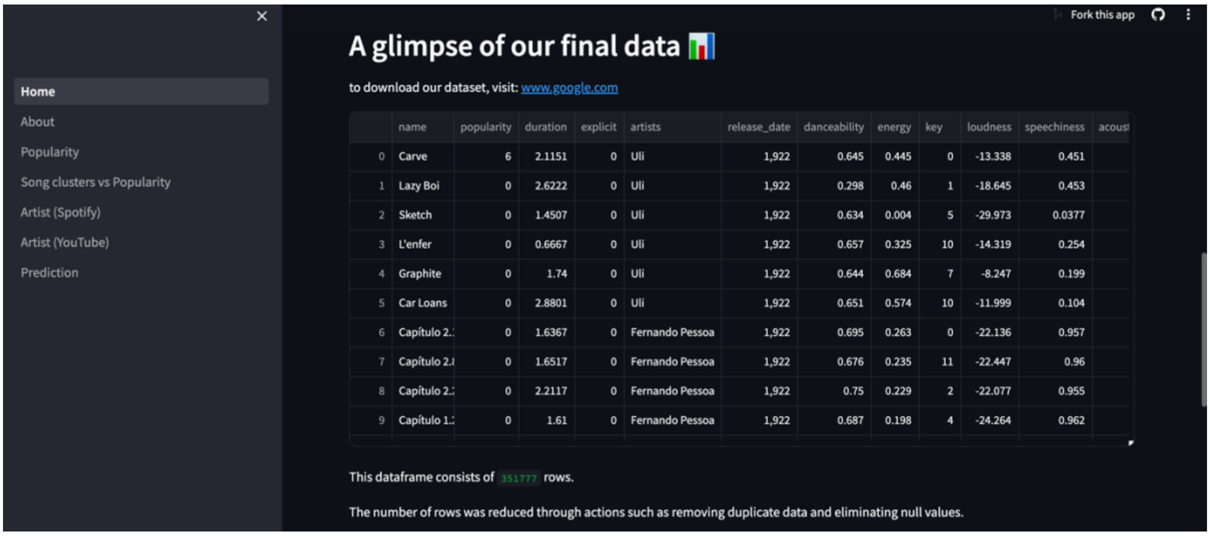
Home



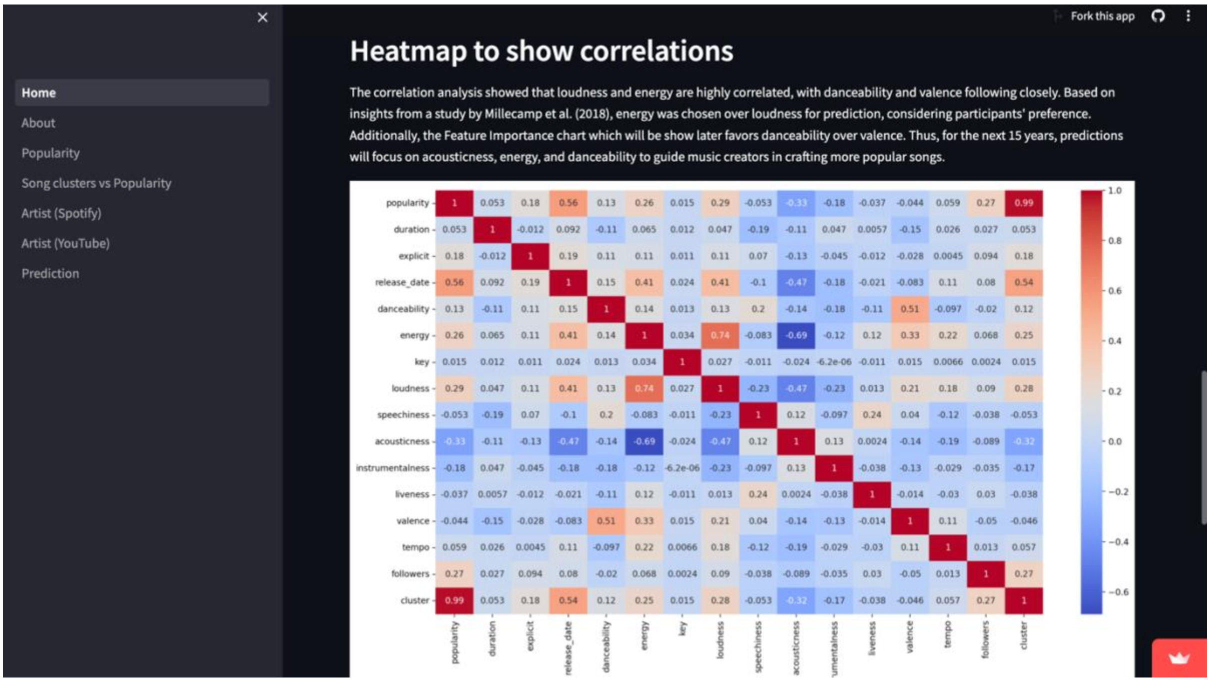
The artists dataset in Spotify is accompanied by information about the data frame's row count, which is located within the Home.



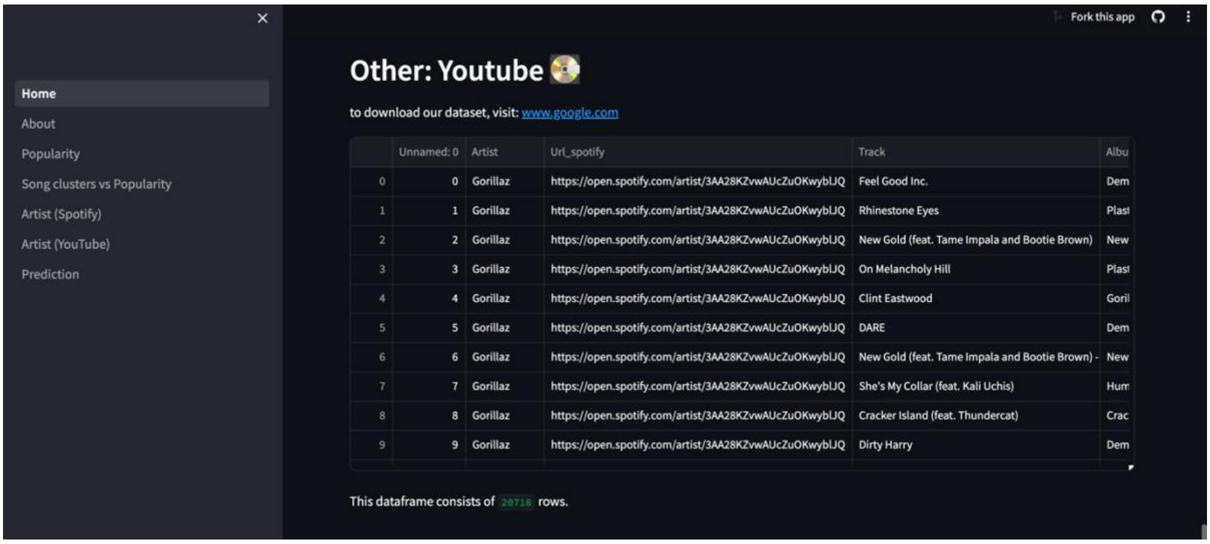
The tracks dataset in Spotify is accompanied by information about the data frame's row count, which is located within the Home.



The final dataset by combining Spotify's artists and tracks dataset is accompanied by information about the data frame's row count, which is located within the Home.

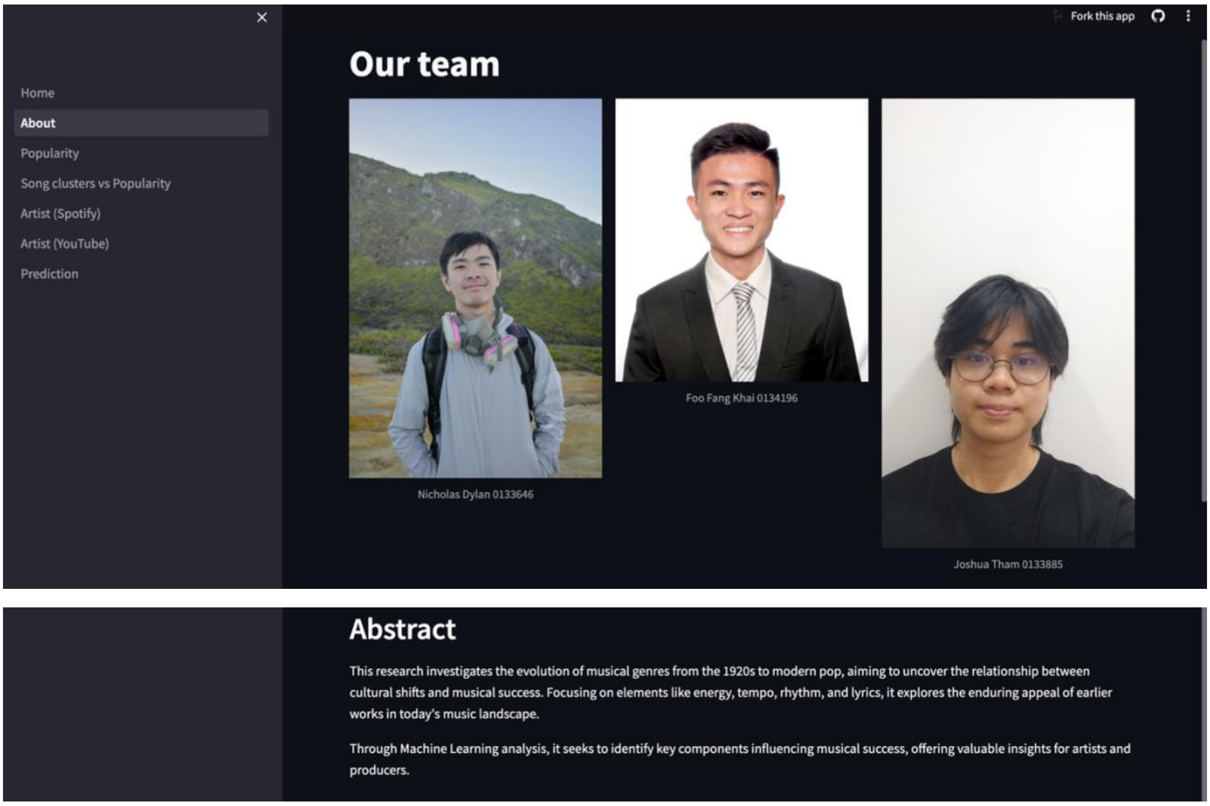


A heatmap is utilized to illustrate the relationships among various variables, serving as a visualization tool to identify the relevant features for prediction. Additionally, the rationale behind the selection of specific variables as prediction values is elucidated.



The YouTube dataset is accompanied by information about the data frame's row count, which is located within the Home.

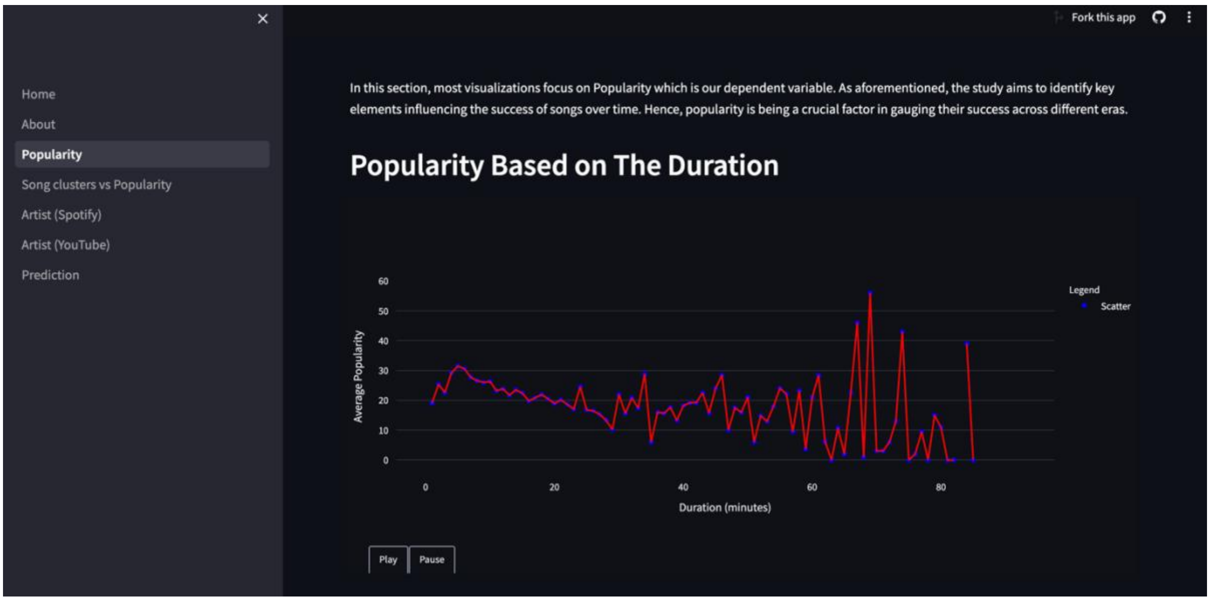
About

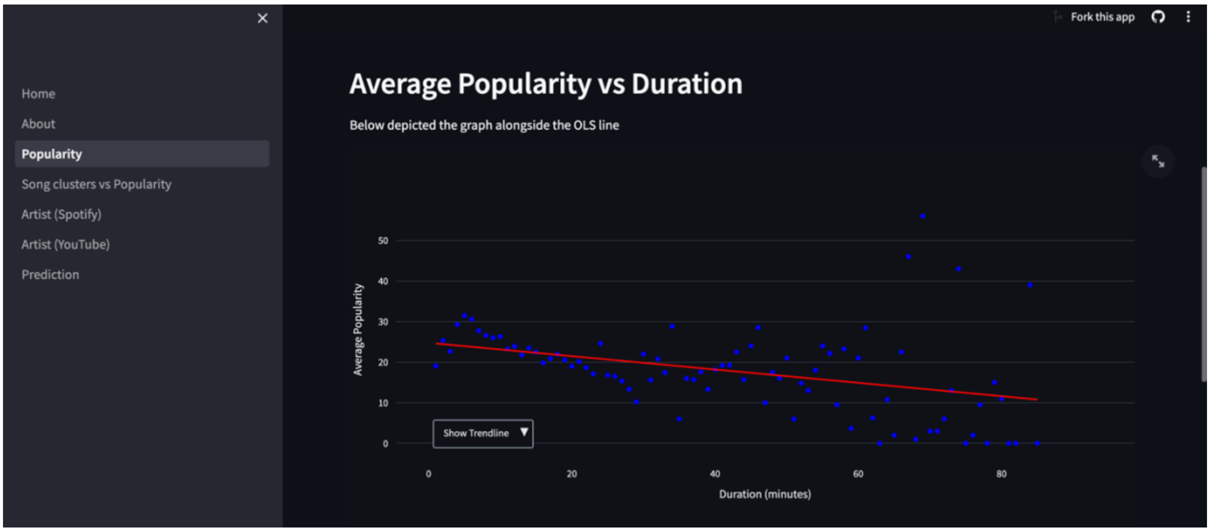


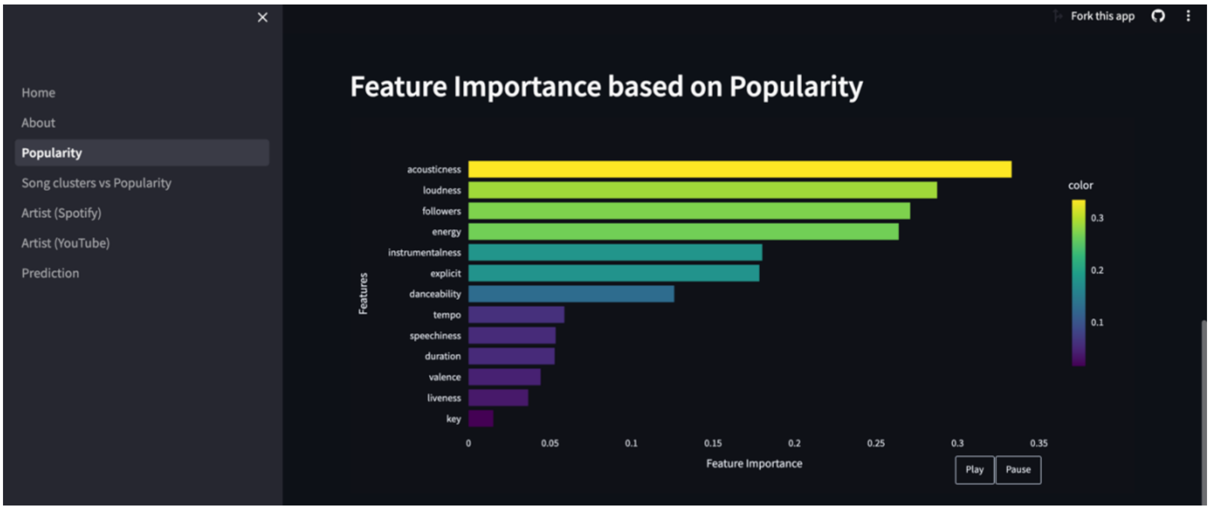
Our group comprises three individuals: Foo Fang Khai, Nicholas Dylan, and Joshua Tham. An abstract will be presented along with an overview explaining the purpose of this dashboard.

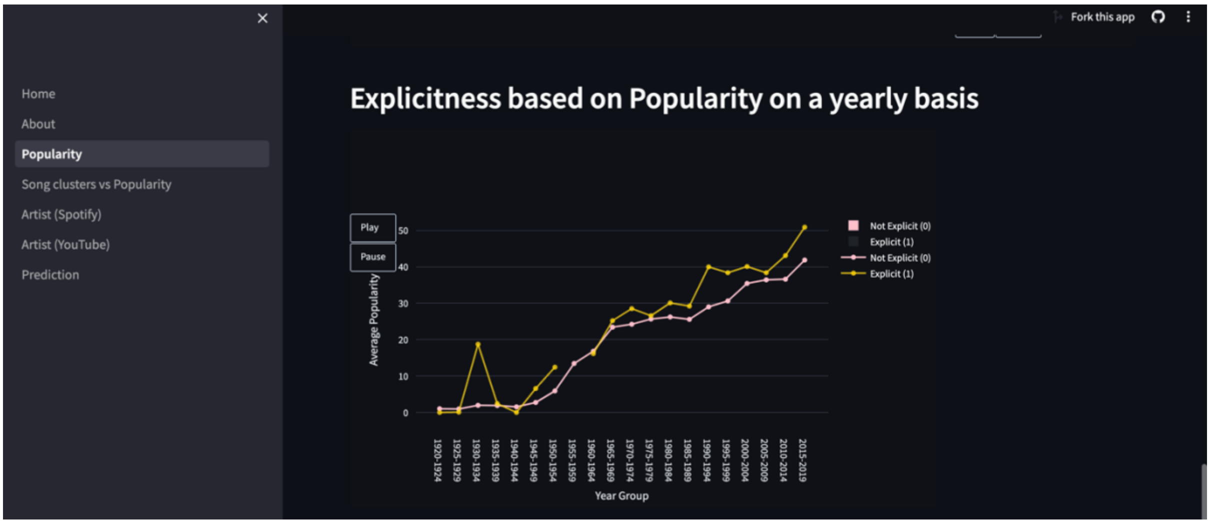
Popularity

This section will present various visual representations, encompassing Popularity Based on Duration, Average Popularity vs Duration, Feature Importance based on Popularity, and Explicitness based on Popularity on a yearly basis, with Popularity as the dependent variable.



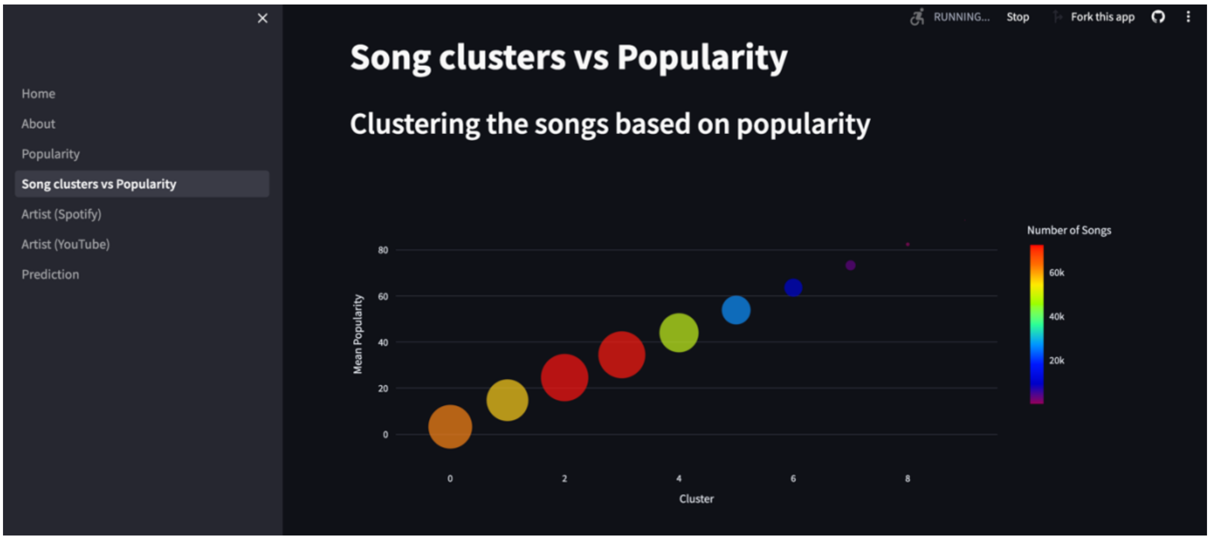

Fork this app
Home
About
Popularity
Song clusters vs Popularity
Artist (Spotify)
Artist (YouTube)
Prediction
Average Popularity vs Duration
Below depicted the graph alongside the OLS line
Average Popularity
Show Trendline
Duration (minutes)

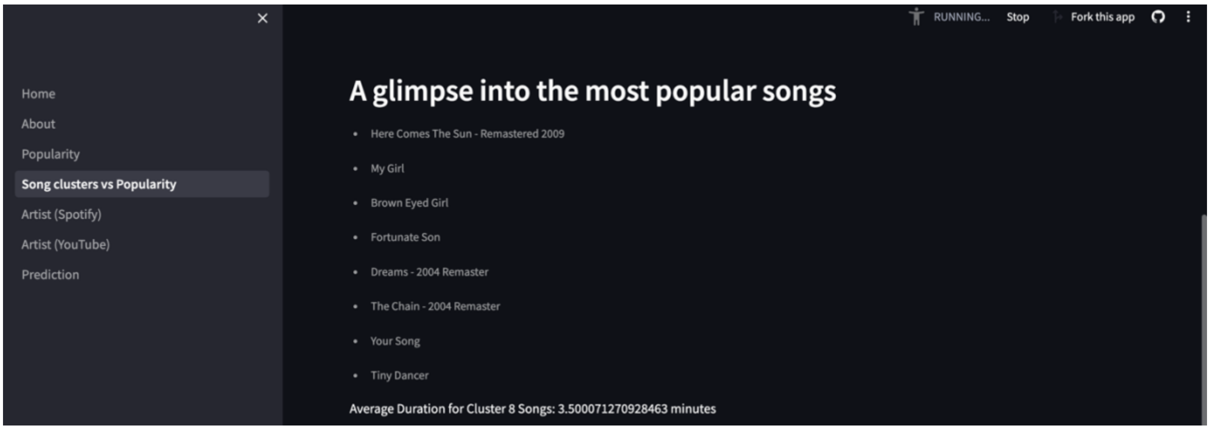

Fork this app
Home
About
Popularity
Song clusters vs Popularity
Artist (Spotify)
Artist (YouTube)
Prediction
Feature Importance based on Popularity
acousticness
loudness
followers
energy
instrumentalness
explicit
danceability
tempo
speechiness
duration
valence
liveness
key
Features
color
Feature Importance
Play
Pause

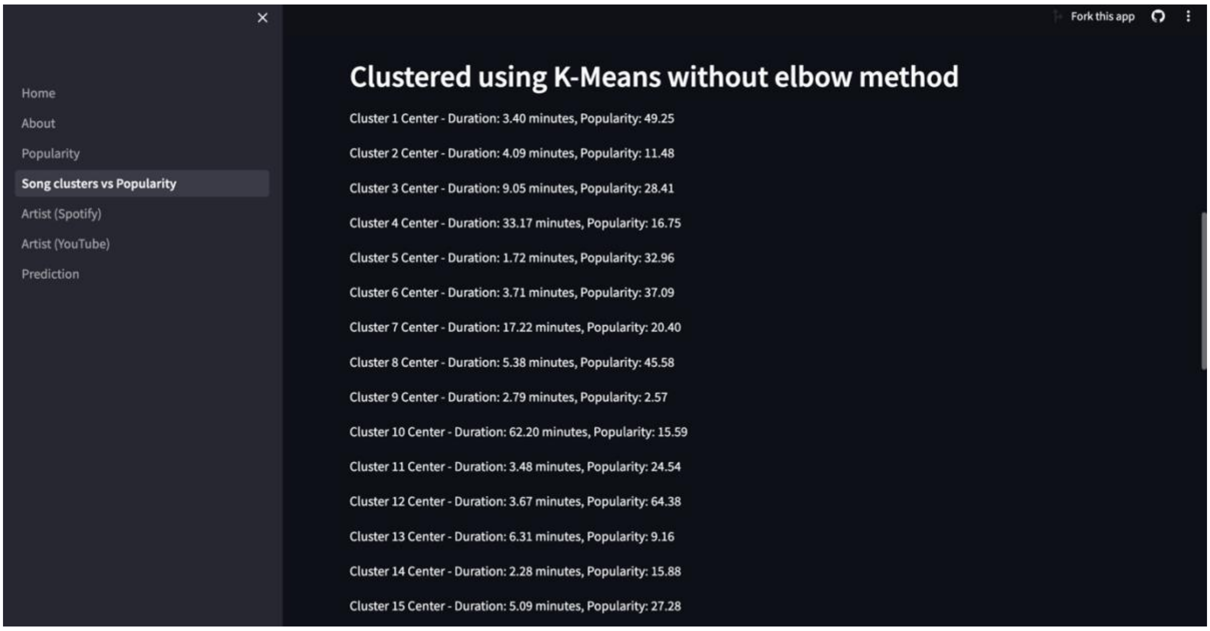

Fork this app
Home
About
Popularity
Song clusters vs Popularity
Artist (Spotify)
Artist (YouTube)
Prediction
Explicitness based on Popularity on a yearly basis
Play
Pause
Average Popularity
Not Explicit (0)
Explicit (1)
Not Explicit (0)
Explicit (1)
Year Group

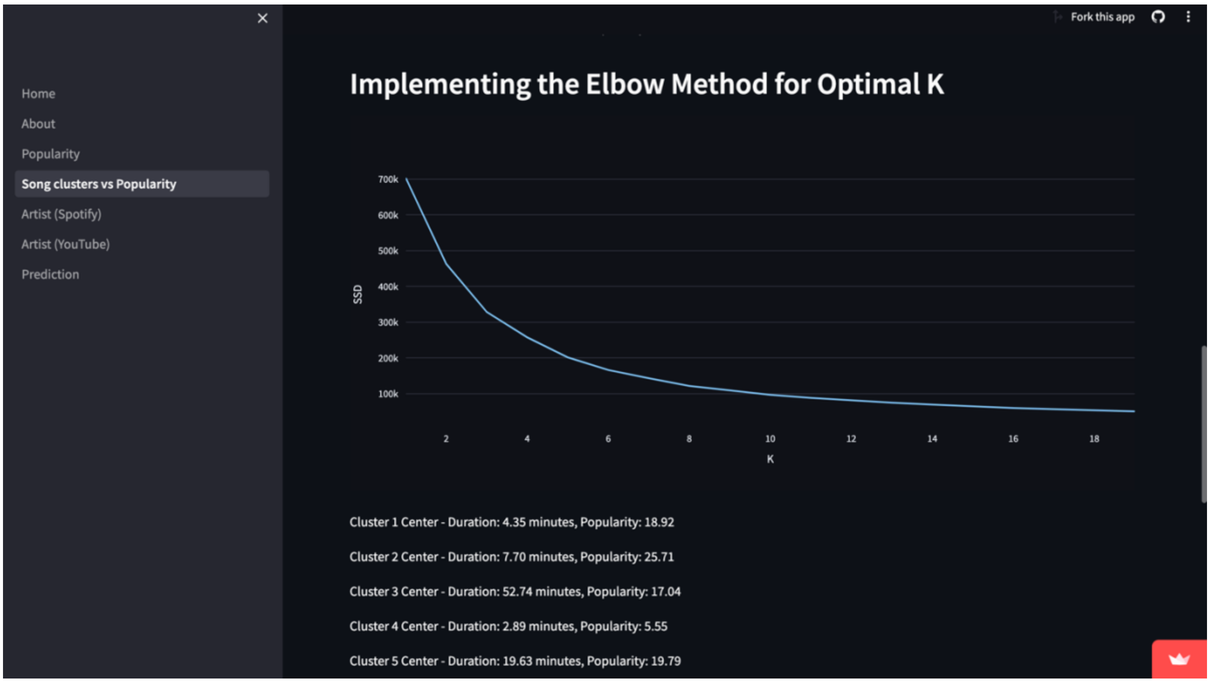
Song clusters vs Popularity

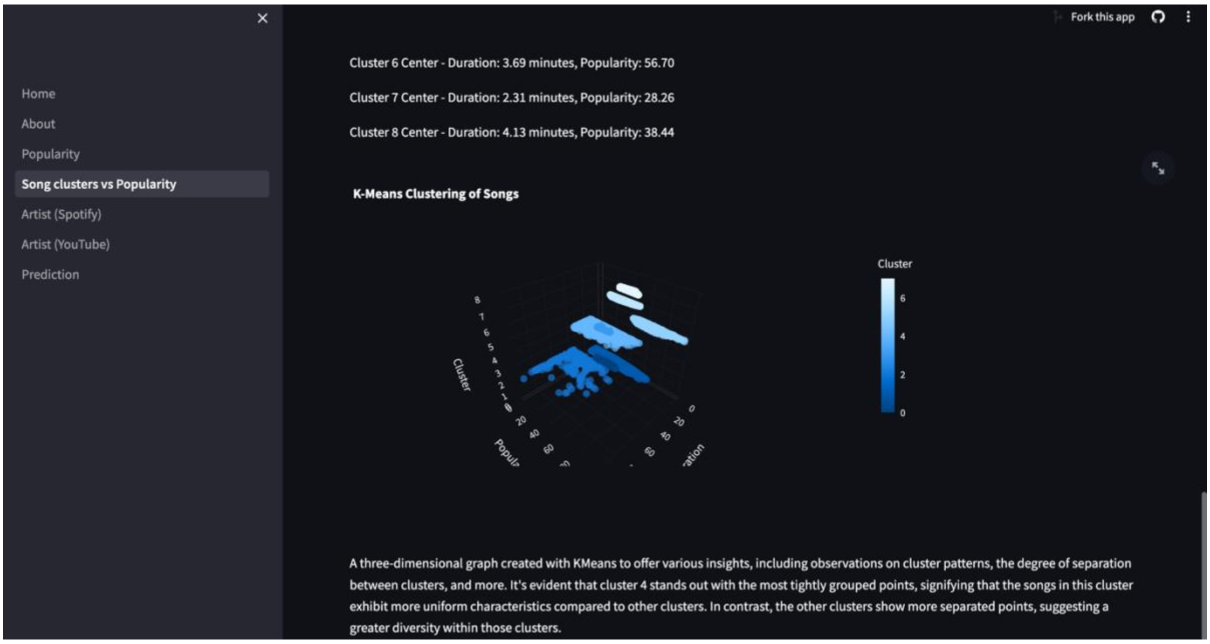
In this segment, information will be presented regarding the ideal duration for a song that has the highest likelihood of success.





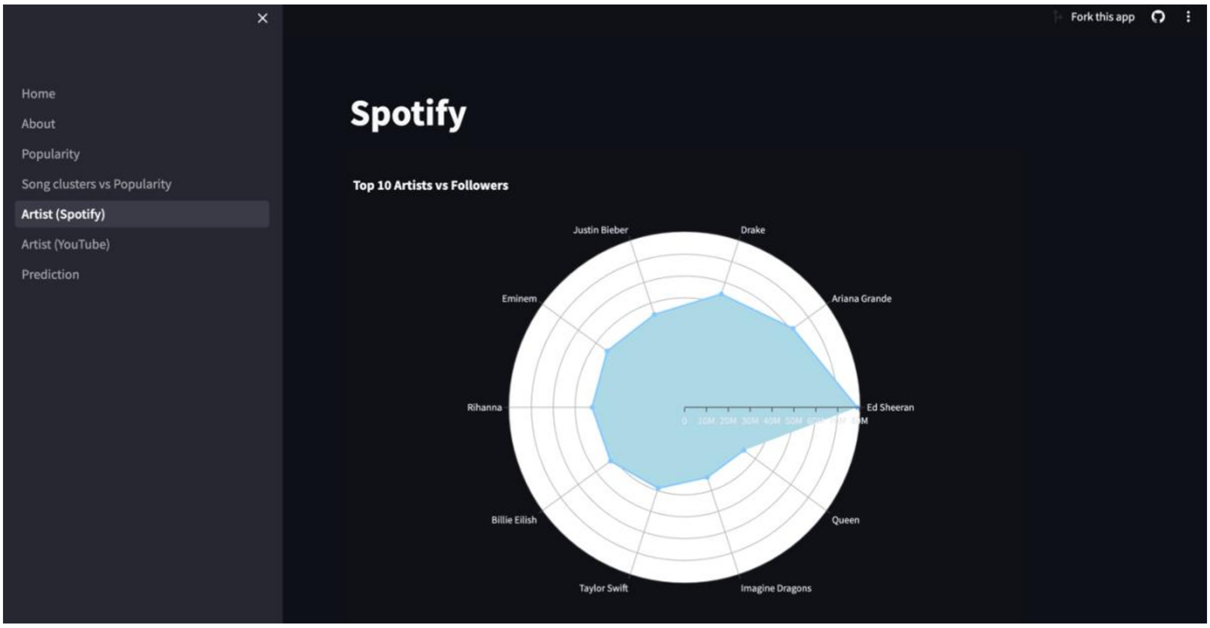



Fork this app
Home
About
Popularity
Song clusters vs Popularity
Artist (Spotify)
Artist (YouTube)
Prediction
Implementing the Elbow Method for Optimal K
SSD
700k
600k
500k
400k
300k
200k
100k
2
4
6
8
10
12
14
16
18
K
Cluster 1 Center - Duration: 4.35 minutes, Popularity: 18.92
Cluster 2 Center - Duration: 7.70 minutes, Popularity: 25.71
Cluster 3 Center - Duration: 52.74 minutes, Popularity: 17.04
Cluster 4 Center - Duration: 2.89 minutes, Popularity: 5.55
Cluster 5 Center - Duration: 19.63 minutes, Popularity: 19.79


Fork this app
Home
About
Popularity
Song clusters vs Popularity
Artist (Spotify)
Artist (YouTube)
Prediction
Cluster 6 Center - Duration: 3.69 minutes, Popularity: 56.70
Cluster 7 Center - Duration: 2.31 minutes, Popularity: 28.26
Cluster 8 Center - Duration: 4.13 minutes, Popularity: 38.44
K-Means Clustering of Songs
Cluster
6
4
2
0
A three-dimensional graph created with KMeans to offer various insights, including observations on cluster patterns, the degree of separation between clusters, and more. It's evident that cluster 4 stands out with the most tightly grouped points, signifying that the songs in this cluster exhibit more uniform characteristics compared to other clusters. In contrast, the other clusters show more separated points, suggesting a greater diversity within those clusters.

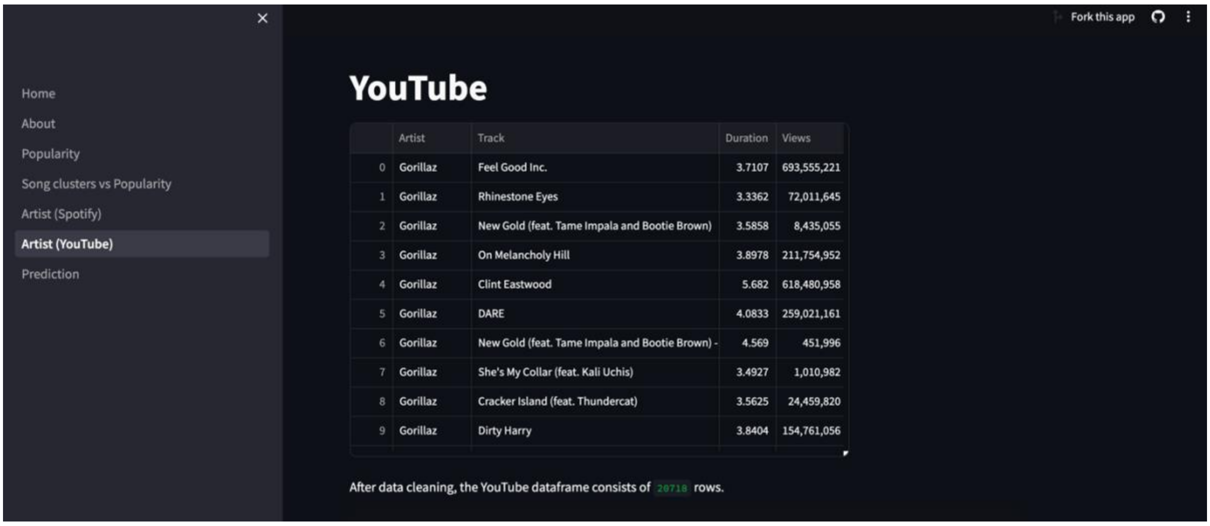
Artist (Spotify)

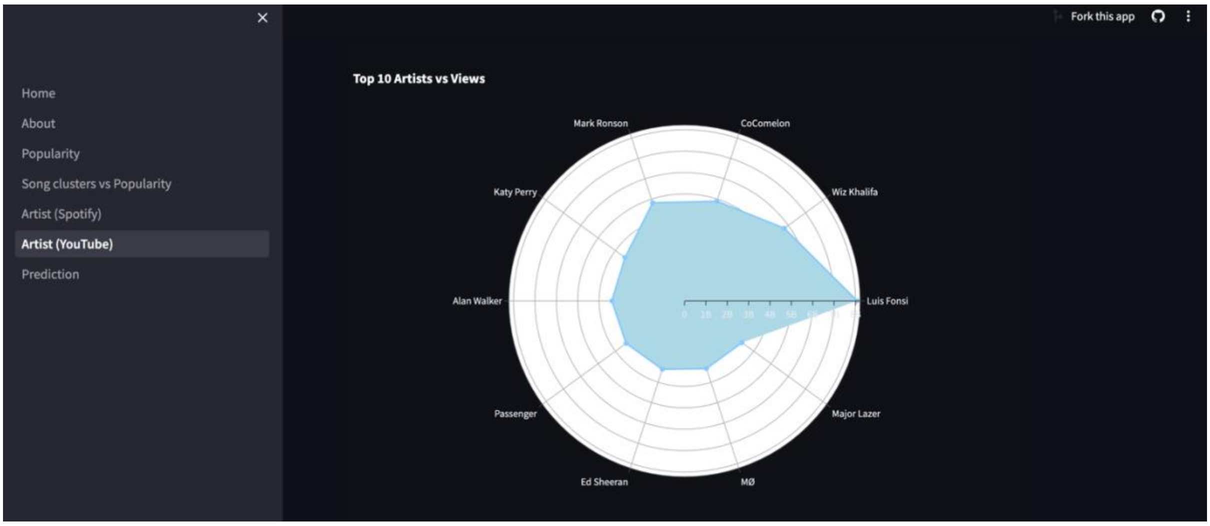
In this section, the statistical data aims to support music creators in customizing their songs to individual preferences by furnishing average values for diverse musical attributes, including danceability, energy, speechiness, acousticness, instrumentalness, liveliness, valence, tempo, key, and loudness.

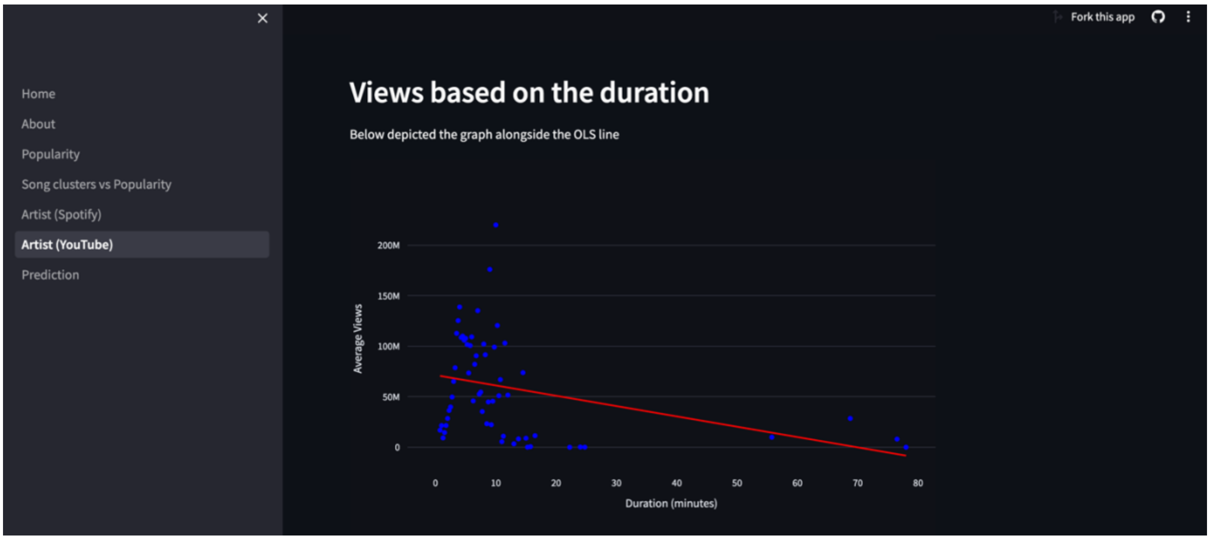


Artists (YouTube)

In this section, the statistical data aims to support music creators in customizing their songs to individual preferences by furnishing average values for diverse musical attributes, including danceability, energy, speechiness, acousticness, instrumentalness, liveliness, valence, tempo, key, and loudness.

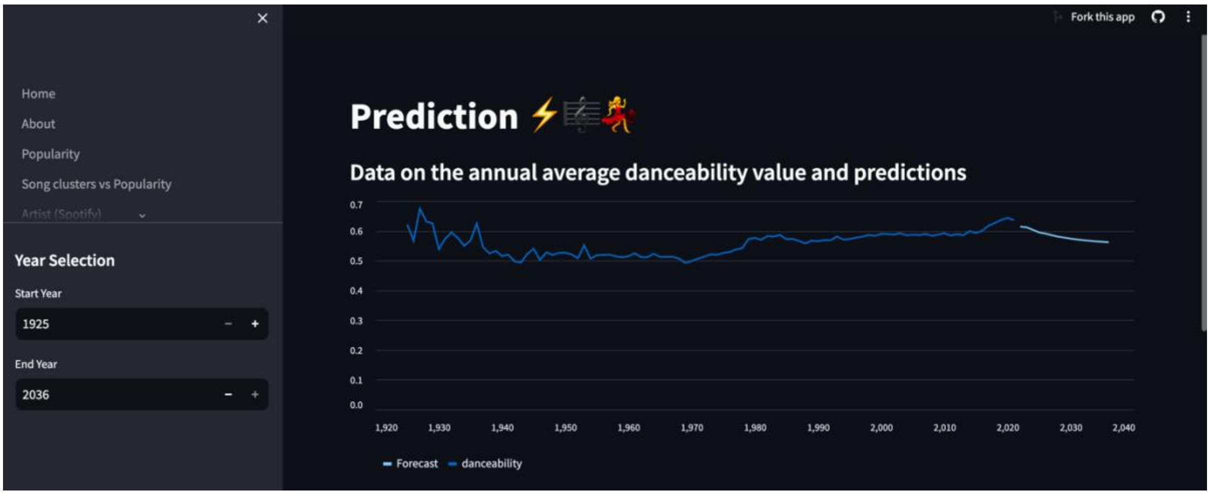


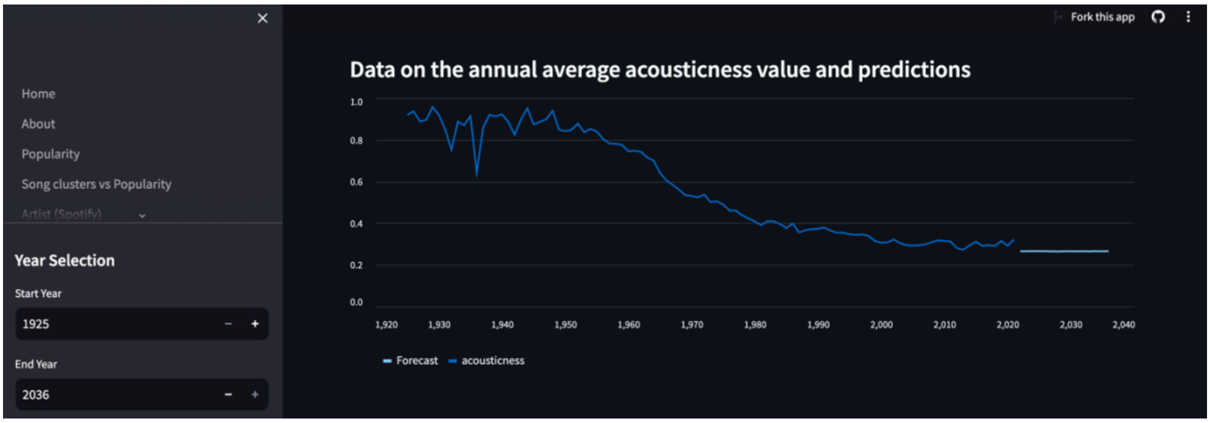


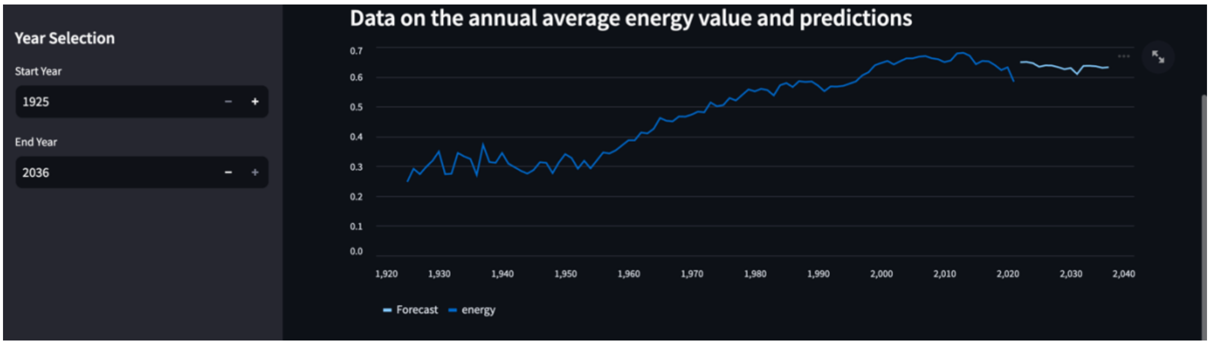


Prediction

In this segment, forecasts for danceability, acousticness, and energy will be presented to offer insights into the forthcoming trends of these three pivotal musical attributes.




Fork this app
Home
About
Popularity
Song clusters vs Popularity
Year Selection
Start Year
1925
End Year
2036
Data on the annual average acousticness value and predictions
Forecast
acousticness


Year Selection
Start Year
1925
End Year
2036
Data on the annual average energy value and predictions
Forecast
energy