**FINAL PROJECT - DATA VISUALISATION**

**SUBJECT – DATA ANALYTICS**

**COURSE INSTRUCTOR: KIRTI DHIR**



**SUBMITTED BY:**

**KRISHNA VIJAY RAMAYI VADAKEYIL**

**SAGARIKA SURESH KALYANKAR**

**NIKHILA ELANGOVAN**

**GROUP NUMBER:1**

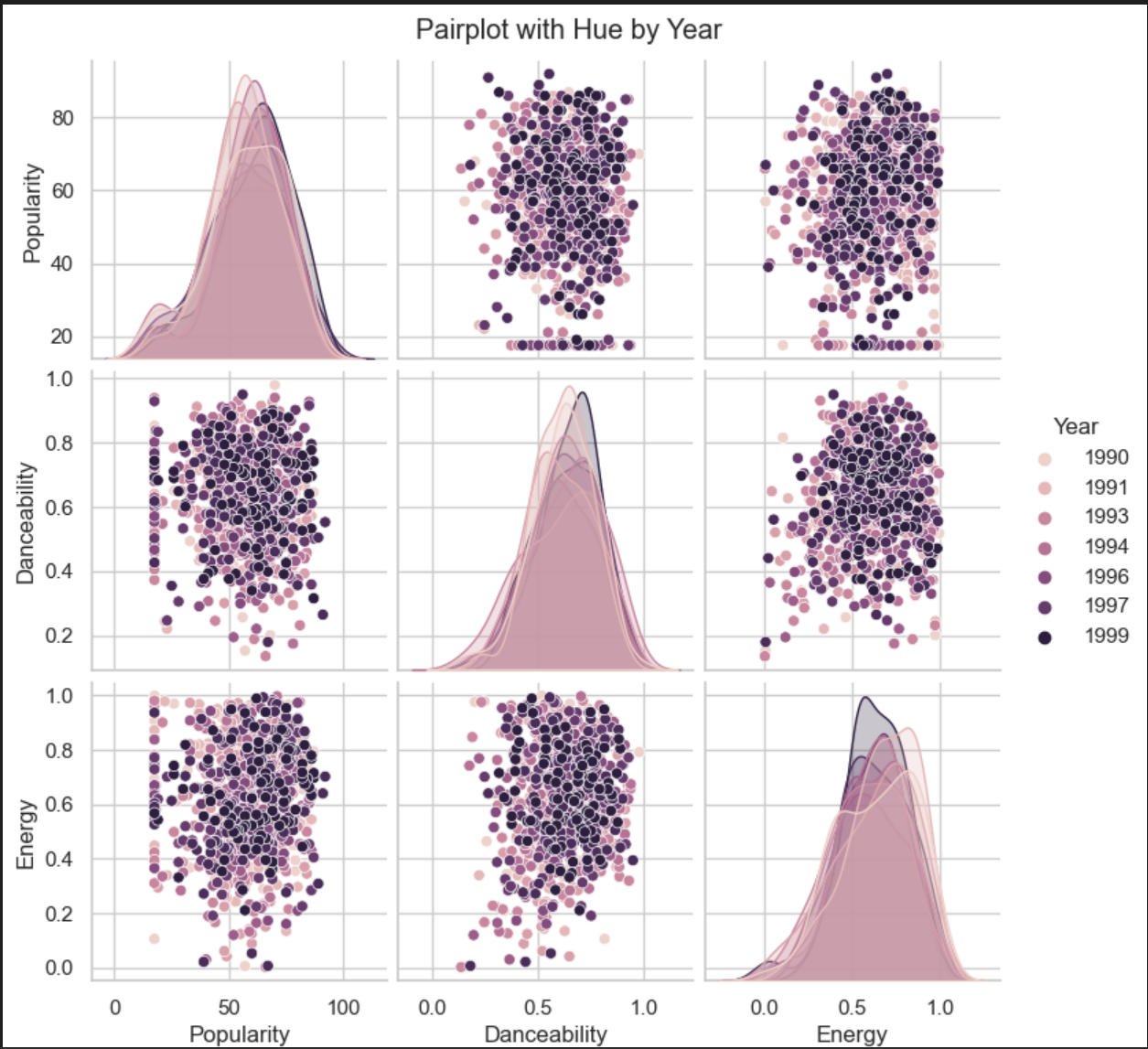
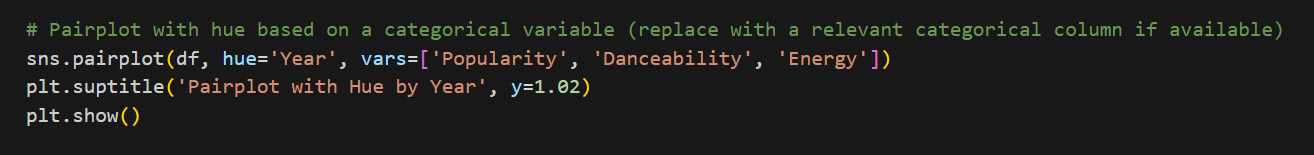
**SUBMITTED DATE - 11 August 2024**

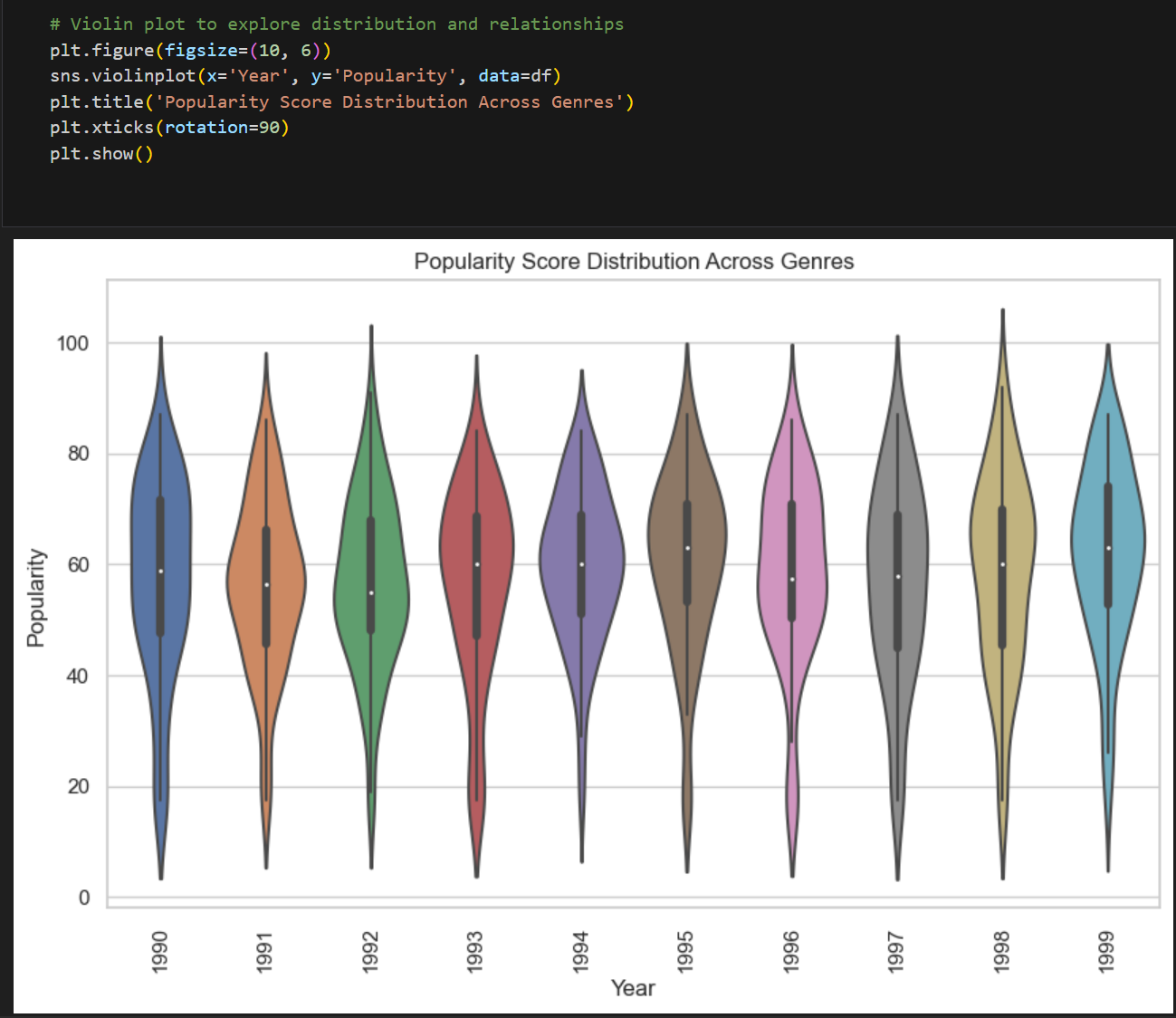
**INTRODUCTION**

The music industry provided the dataset we've selected, which focuses on several aspects of songs, including danceability, energy, and loudness. This dataset contains 980 classic 1990s hits from 536 artists, enriched with Spotify audio features, offering a detailed view of the musical attributes that shaped the decade. This industry's noteworthy influence on entertainment and culture makes it especially fascinating. We selected this subject because we love music and are interested in learning how various musical components might affect a song's level of popularity. This dataset is interesting and enlightening as it provides an opportunity to blend our personal interest with data research.

**PROBLEM STATEMENT**

Our aim is to identify the elements—rhythm, energy, and loudness—that contribute to a song's popularity. Our goal is to identify trends that can be used to forecast which songs will be popular. This will provide producers and musicians with insightful information that will help them make music that connects with listeners. The analysis seeks to enhance decision-making during the composition of music.

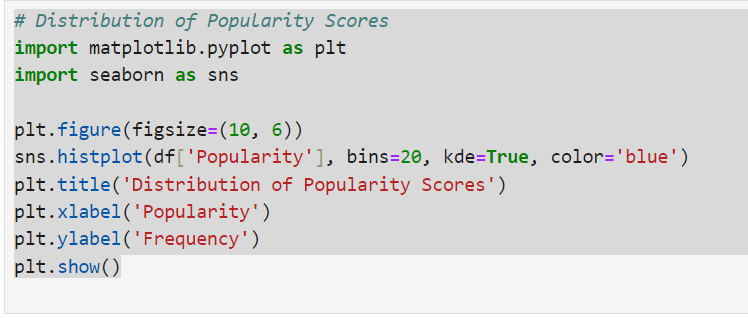
**EXPLORATORY DATA ANALYSIS**  
  
**1.**  
  
**INSIGHTS -**The graph shows a year-based pairplot that illustrates the correlations and dispersions of three musical attributes—Popularity, Danceability, and Energy—across a range of years from 1990 to 1999. Off-diagonal scatter plots show correlations between various features, broken down by year, whereas diagonal plots show the distribution of each feature across time. This graphic aids in the analysis of the changes in song qualities and musical tastes over the course of the ten years, offering insights into patterns and changes in the dynamics of the music business at that time.

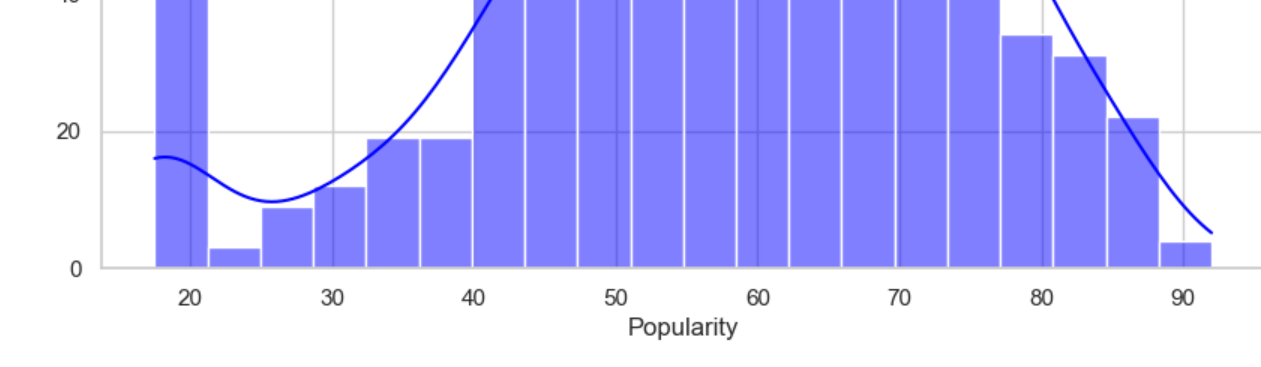
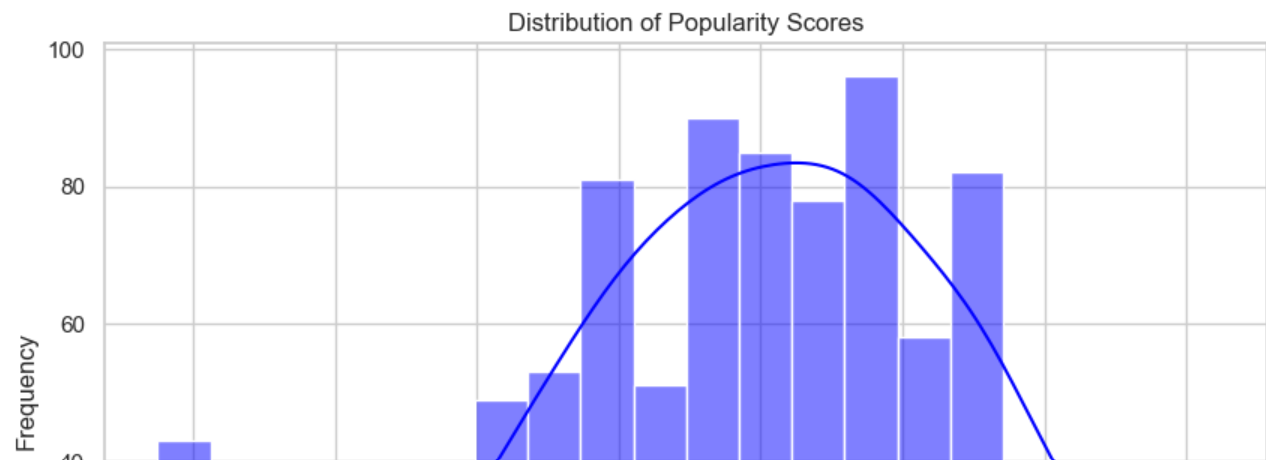
**2.**  


# **INSIGHTS -** Using distinct colors for each year between 1990 and 1999, the image shows a violin plot that shows how popularity scores distributed across that time. Combining elements such as kernel density and box plots, this graph shows data points' density at various popularity levels and the median popularity. The width of the violin indicates the density of records; bigger sections of the violin reflect a higher concentration of songs at that popularity score. Each violin shape represents a year. Plotting offers a clear visual comparison of how popularity has changed over the course of the decade, displaying themes like variations in song popularity and changes in the central tendency. It is especially helpful to utilize this kind of visualization to spot changes, providing information about which years in the music industry had more varied or concentrated popularity scores among tracks over time. 3. **INSIGHTS** - A boxplot shows the distribution of popularity scores for tracks in a dataset is seen in the image. The interquartile range (IQR), represented by the box's width, the range of popularity excluding outliers, and the median popularity score, are all clearly shown by the boxplot. Notably, the plot implies a narrow IQR, meaning most tracks have popularity values close to one another. The whiskers sticking out of the box indicate the presence of outliers, which are tracks with popularity scores that are higher or lower than most of the data. The central tendency, dispersion, and possible outliers in the popularity scores may be rapidly understood with the use of this boxplot, which helps with judgements relating to data processing or more in-depth investigation of the elements affecting popularity. 4. **INSIGHTS** - Using shades of red and blue to represent positive and negative correlations, respectively, the graphic displays a correlation heatmap from a collection of musical features. These results indicate that Danceability and Energy have a significant positive connection, indicating that danceable music tend to be more energetic. Notably, there are negative correlations that show that louder songs are typically less acoustic, such as the one between loudness and acousticness. Additionally, the heatmap indicates that Popularity is driven by a complicated combination of variables, as it does not closely correspond directly with most other features. **DATA VISUALIZATION**

1. Analysing distribution of Popularity Scores

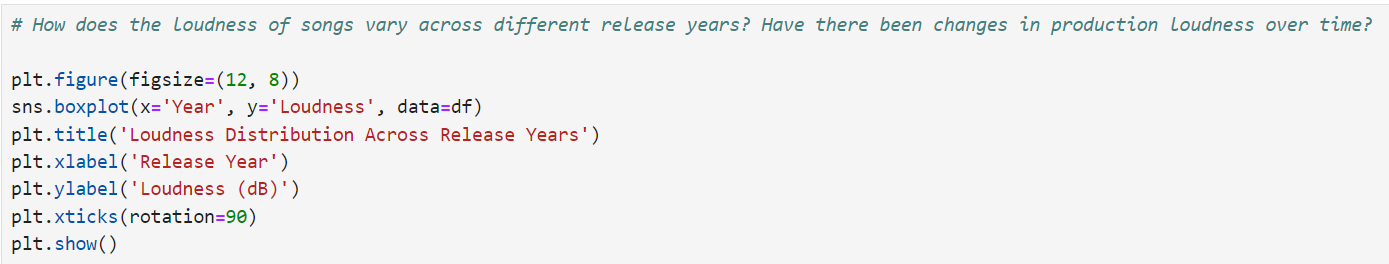
We chose a histogram to show the distribution of popularity ratings since it clearly shows how popularity is dispersed across all songs. This illustration is significant because it contributes to our understanding of average popularity levels, which is necessary for identifying patterns that predict a song's success and is closely related to our investigation into the variables that affect song popularity.

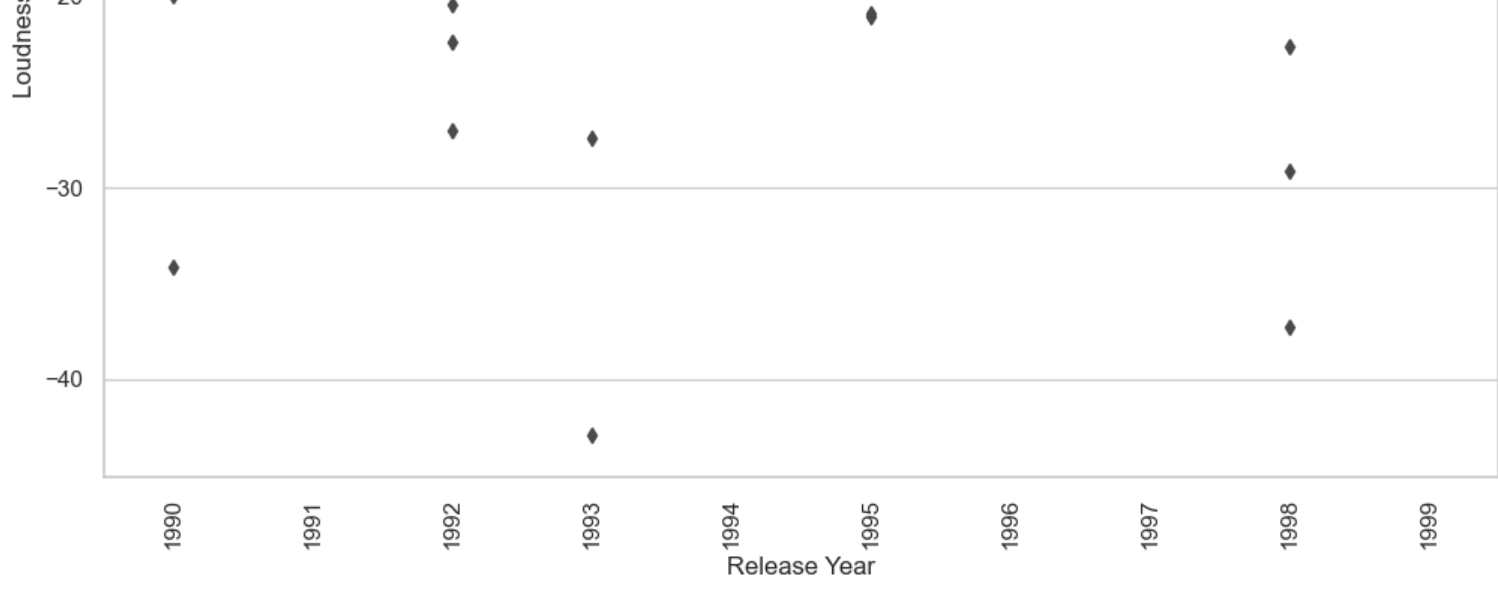




Most songs have average popularity ratings, especially in the 50–70 range where frequency is strongest. The smooth curve shows that fewer songs are very popular at either end, suggesting that these scores are the most average. By examining this frequency distribution, we may investigate what factors affect these overall popularity levels and the causes of the variations in the success of songs.

2. How does the loudness of songs vary across different release years? Have there been changes in production loudness over time?

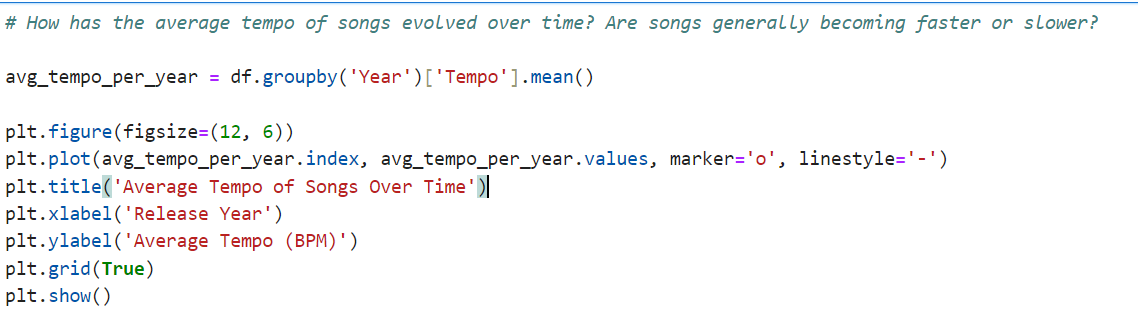
We used a box plot, which makes trends, medians, and outliers easier to identify, to show the annual variations in loudness. This aids in our understanding of whether music loudness affects song popularity, which is directly related to our goal of figuring out what makes songs successful.

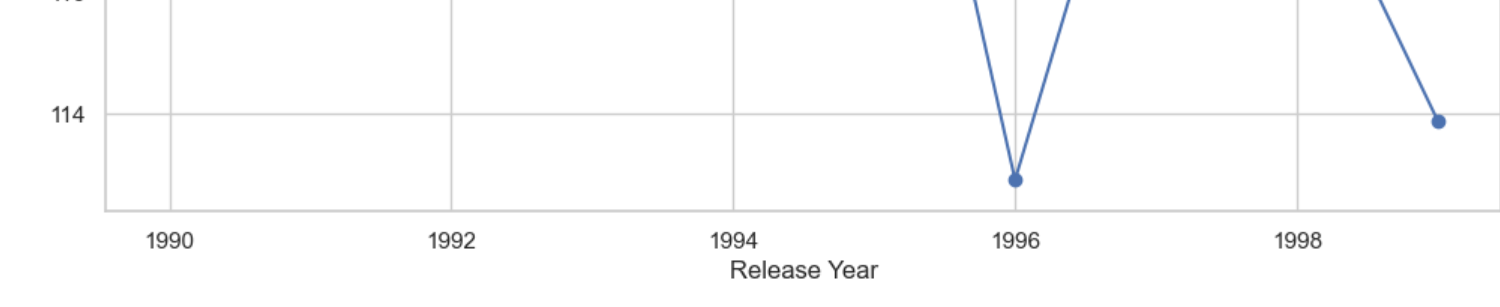
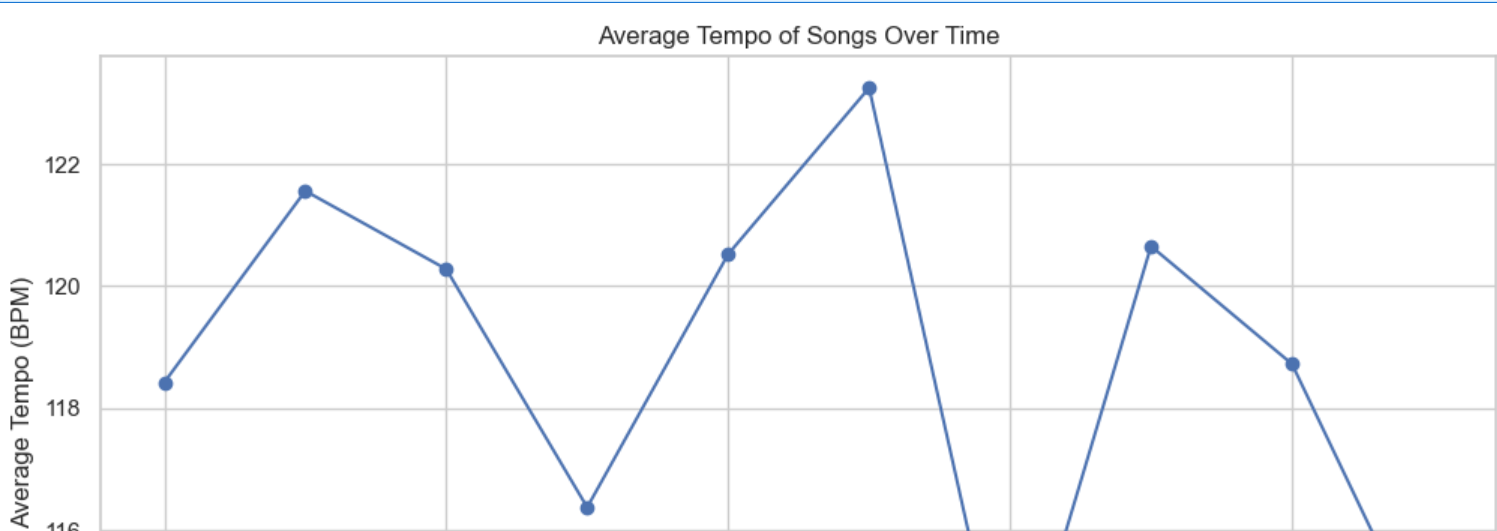


With occasional outliers and a little amount of fluctuation, song volume has largely remained steady over time. loudness may not be the sole significant signal of a song's potential popularity, as evidenced by the lack of a discernible pattern showing a rise in loudness. This constancy in loudness enables us to focus on other elements of the music that can be more important to a song's success.

3. How has the average tempo of songs evolved over time? Are songs generally becoming faster or slower?

We used a line plot to track tempo variations over time to make patterns easier to see. This is crucial to our analysis since it allows us to ascertain whether tempo influences the level of popularity of a song.

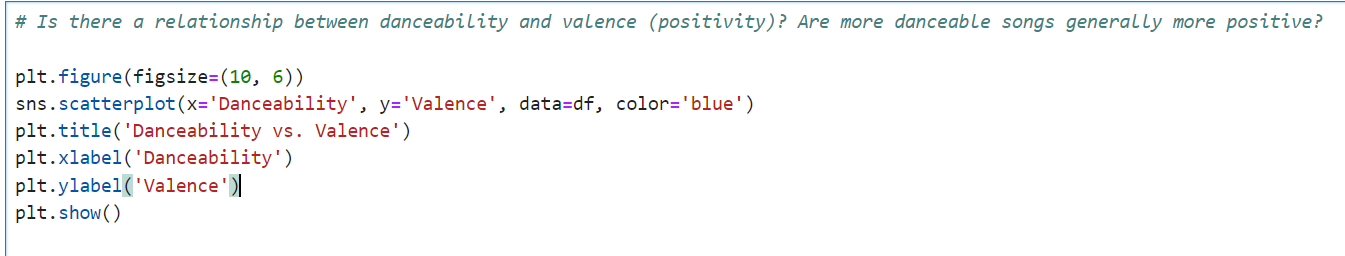


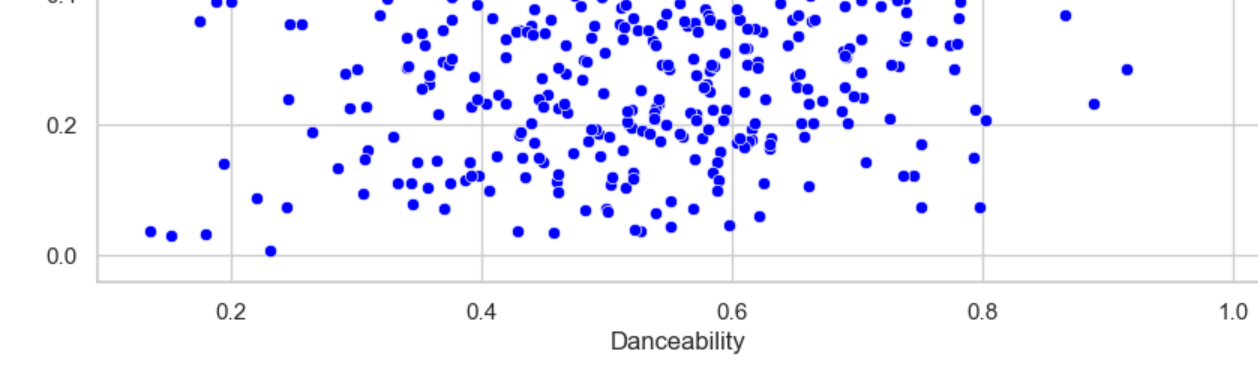
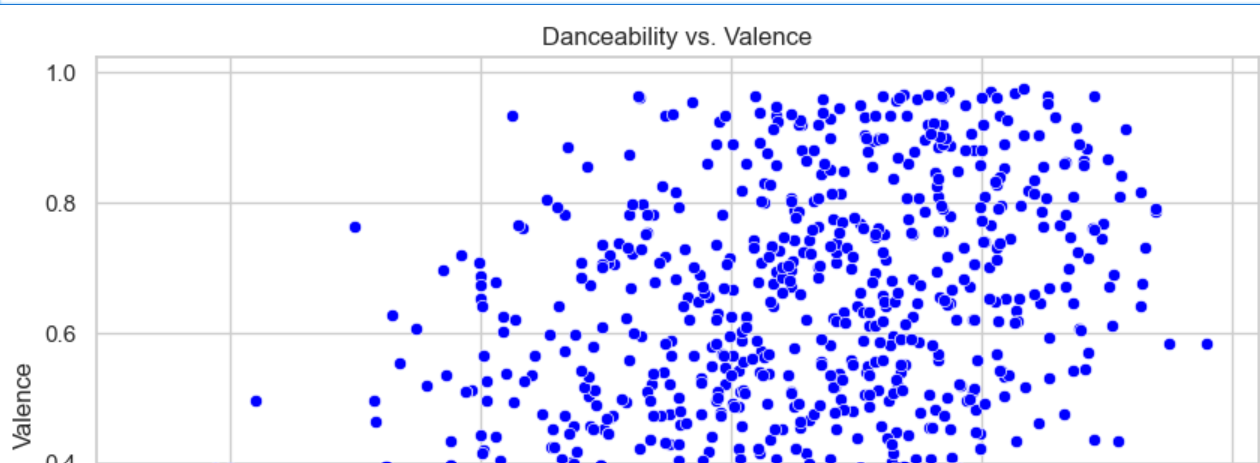


The average tempo of songs has increased or decreased over time, reflecting changes in musical tempo. These tempo trends help us understand how shifting musical preferences may affect a song's level of popularity. Tempo analysis is crucial when trying to figure out what it is about songs that aged nicely.

4. Is there a relationship between danceability and valence (positivity)? Are more danceable songs generally more positive?

We used a scatter plot to quickly identify any relationships between danceability and valence. This helps identify whether the popularity of a song is influenced by its mood, which is crucial for understanding what makes songs successful.



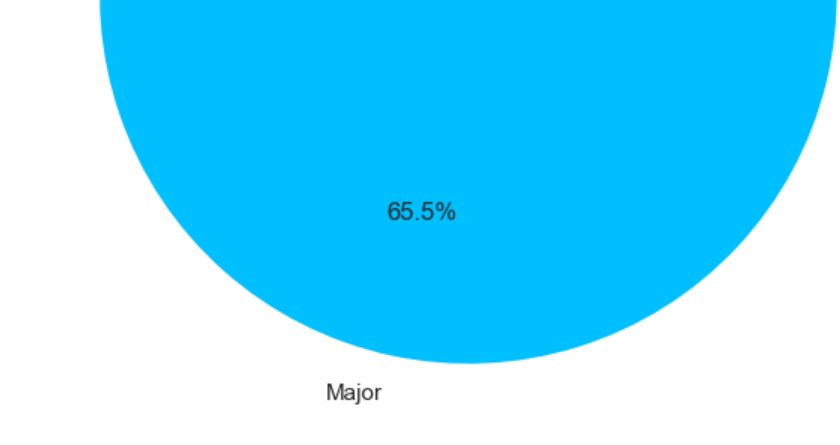
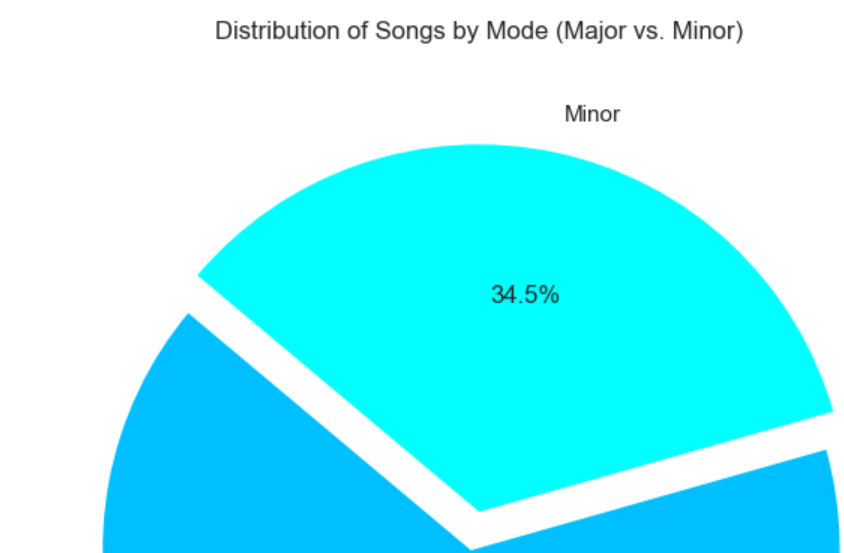


Most songs that are danceable have positive valence (danceability 0.4–0.8). This implies that positive tunes are probably more danceable, which could account for their increased popularity. We understand how a song's popularity is influenced by its mood by recognizing this relationship.

5.Analyse distribution of songs by Modes

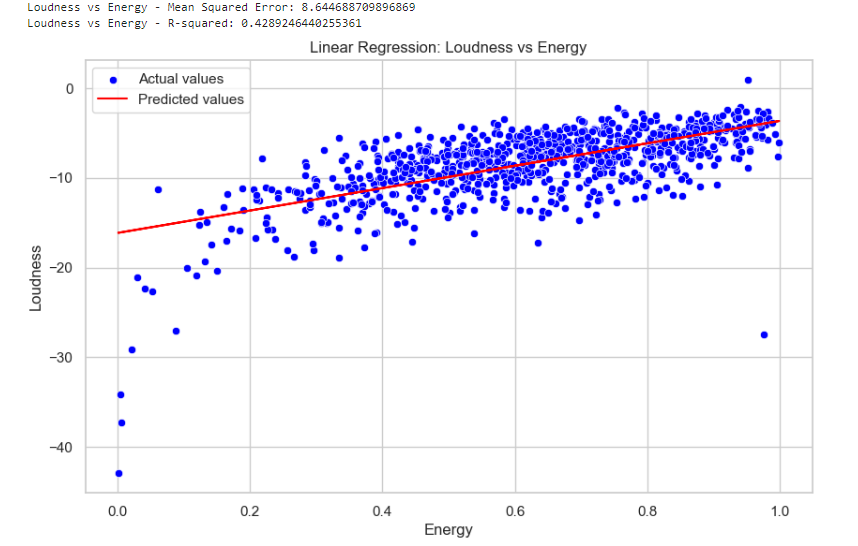
We made a pie chart to show the proportion of songs in major and minor keys. This is important for our analysis since it makes clear how a song's key choice may affect its popularity.





The pie chart shows that while most songs are in major keys (65.5%), 34.5% are in minor keys. This suggests that major keys are utilized more often, which may affect a song's appeal and level of popularity. Understanding this distribution allows us to look at how the musical key affects a song's success.

**PREDICTIVE ANALYSIS**



**Energy vs. Loudness**

* The scatter figure illustrates the correlation between Loudness (y-axis) and Energy (x-axis) by fitting a line of linear regression to the data.
* The red line indicates the values predicted by the linear regression model, while the blue dots indicate the actual data points.

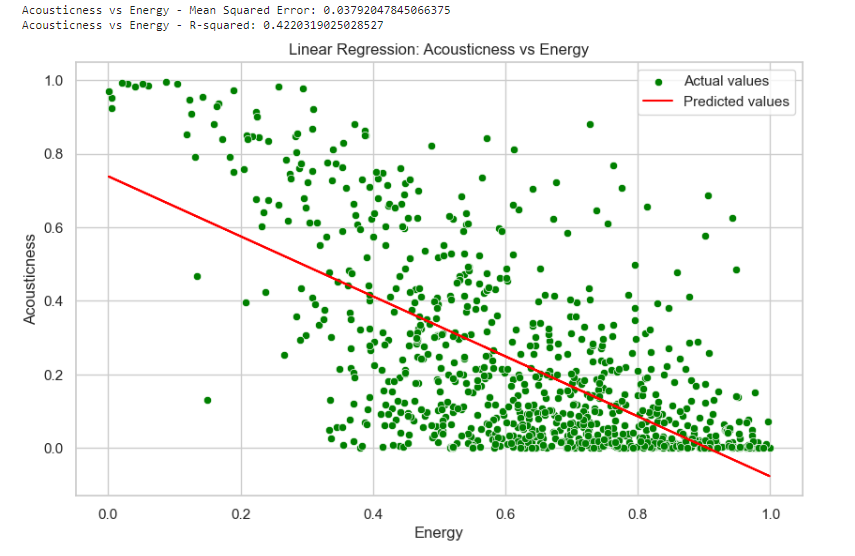
**Interpretation**

* **Positive Slope:** The regression line's positive slope suggests that loudness and energy have a positive link. This implies that a song's tendency to get louder goes hand in hand with its energy.
* **R-squared Value:** The energy variable accounts for roughly 42.89% of the variance in loudness, with an R-squared value of 0.4289. This points to a moderate relationship—while energy has a significant influence on loudness, loudness is also influenced by other variables.
* **Mean squared error:** MSE is 8.644. This shows the average squared variance between the model-predicted Loudness values and the observed ones.

Although the model is useful, it is not flawless, as indicated by the number, which represents the error margin in the model's predictions.

**Summary**

The positive association suggests that louder music tends to be greater in energy. The moderate R-squared number indicates that additional factors are likely at work in addition to Energy, which is important in determining Loudness.



Energy vs. Acousticity

The scatter plot depicts the correlation between Acousticness (y-axis) and Energy (x-axis) using a linear regression line. The analysis is done as follows.

The red line indicates the values predicted by the linear regression model, while the green dots indicate the actual data points.

Interpretation:

It can be observed that there is a negative association between Acousticness and Energy based on the regression line's negative slope. According to this, a song's acoustic qualities—such as its organic, raw sound—tend to deteriorate as its intensity rises.

Approximately 42.20% of the variance in Acousticness can be explained by the Energy variable, according to the R-squared value of 0.4220. Comparable to the plot of loudness vs. energy, this connection is similarly modest.

• Mean Squared Error (MSE): With an MSE of 0.0379, which is low, the model's average forecast accuracy is pretty close to the real Acousticness values.

In conclusion, it appears that songs with more energy tend to be less acoustic due to the negative association, which might be explained by the fact that higher energy songs typically feature more processed or electronic sounds. Although there are additional elements that affect acoustic characteristics, the R-squared value indicates that energy plays a substantial role in defining acousticness.