**Data Analysis of Music Hits from the 90s**

Lewis King

ID:000051992

4/16/25

Objective: To analyze and clean the 1990s Hit Classics DataSet. While also finding trends and patterns throughout the dataset and visualizing them.

The 1990s were a significant decade for music, and many popular songs were released during this time. Analyzing this data set will reveal which songs from the 1990s were the most popular, identify the artists who dominated the decade, and ultimately highlight the songs that people enjoyed the most. This will be explored as the data set is cleaned and analyzed.

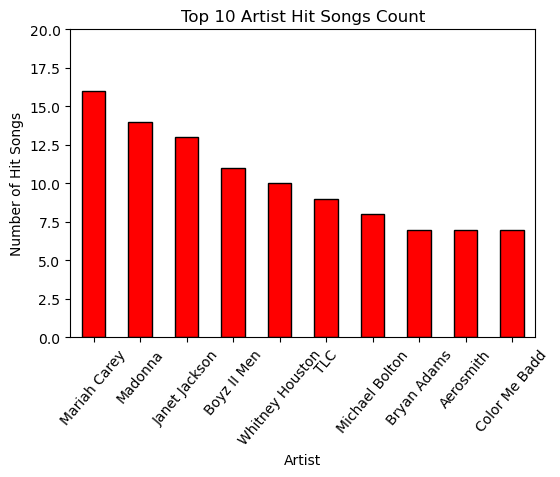
DataSet Origin: [1990s Classic Hits Dataset](https://www.kaggle.com/datasets/thebumpkin/1990s-classic-hits-with-spotify-data)

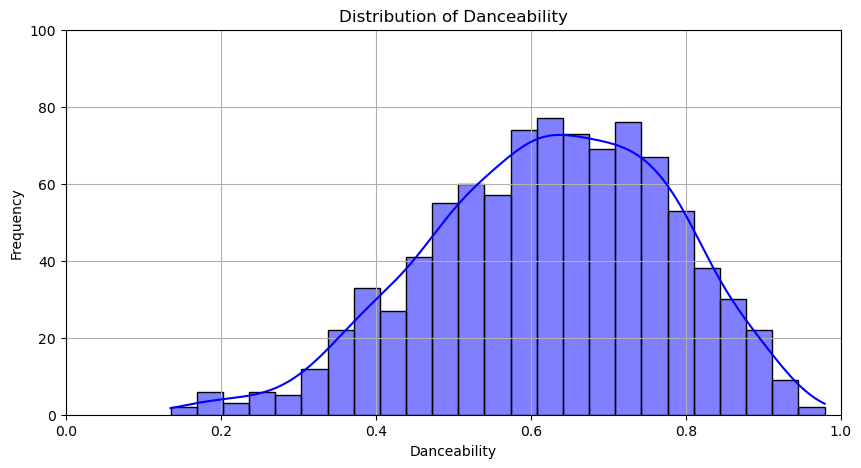
To start, the 1990s Classic Hits Dataset has a shape of 919 rows and 17 columns. The dataset includes both categorical and numerical data types. The categorical columns are 'Track,' 'Artist,' and 'Duration,' while the numerical columns are 'Time Signature,' 'Danceability,' 'Energy,' 'Key,' 'Loudness,' 'Mode,' 'Speechiness,' 'Acousticness,' 'Instrumentalness,' 'Liveness,' 'Valence,' 'Tempo,' 'Popularity,' and 'Year.'

After copying my dataset into a new variable, I used a command to find the sum of null values and discovered that there are no null values. However, if there had been missing values, I would have addressed them differently based on the data type. For categorical data, I would have replaced missing values with "Unknown," and for numerical data, I would have calculated the average of the column to fill in the missing values.

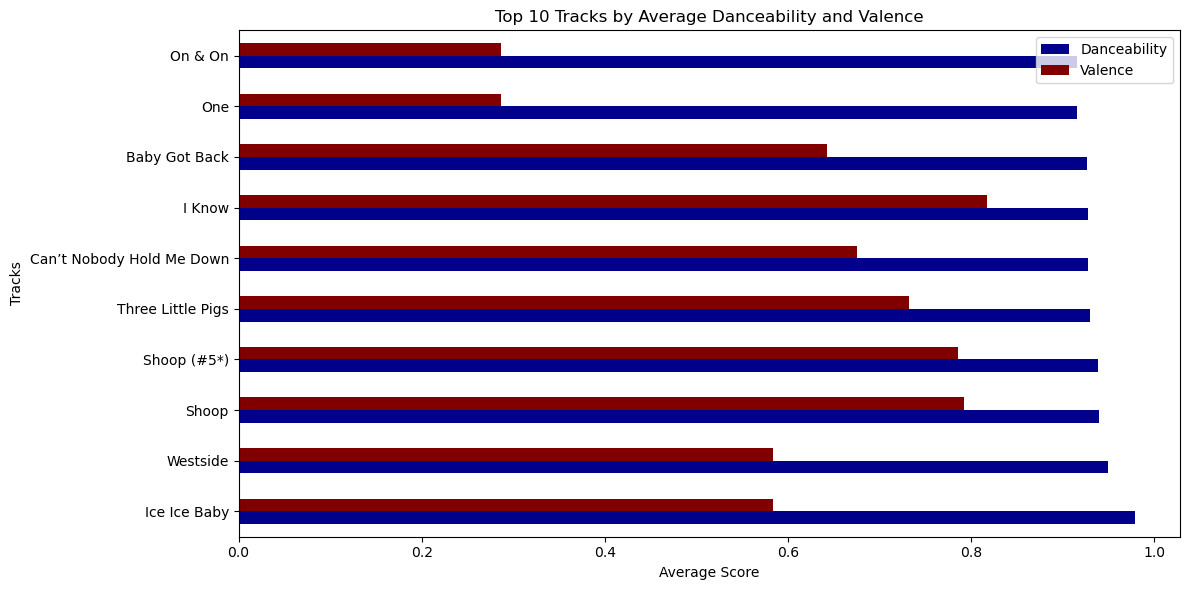
Next, I checked the dataset for duplicates and found that there are none. If duplicates were present, I would have used the `.drop\_duplicates()` function to remove them. With the dataset thoroughly checked and cleaned, I then moved on to identifying trends and patterns in the data for visualization.

Data Visualization:



This bar plot illustrates the top 10 artists with the highest number of hit songs in the dataset. Each bar represents the count of hit songs made by each artist. One observation I've made is that there is a significant drop in the number of hits from the top three artists to the others in the top 10. This indicates a steep popularity curve among these artists, suggesting that only a few dominated the charts in the 1990s and defined the decade. The most prominent artists of that era were Mariah Carey, Madonna, and Janet Jackson. This trend may imply that these three artists produced and released more music during the 90s than others, which is why they have a larger number of hits.

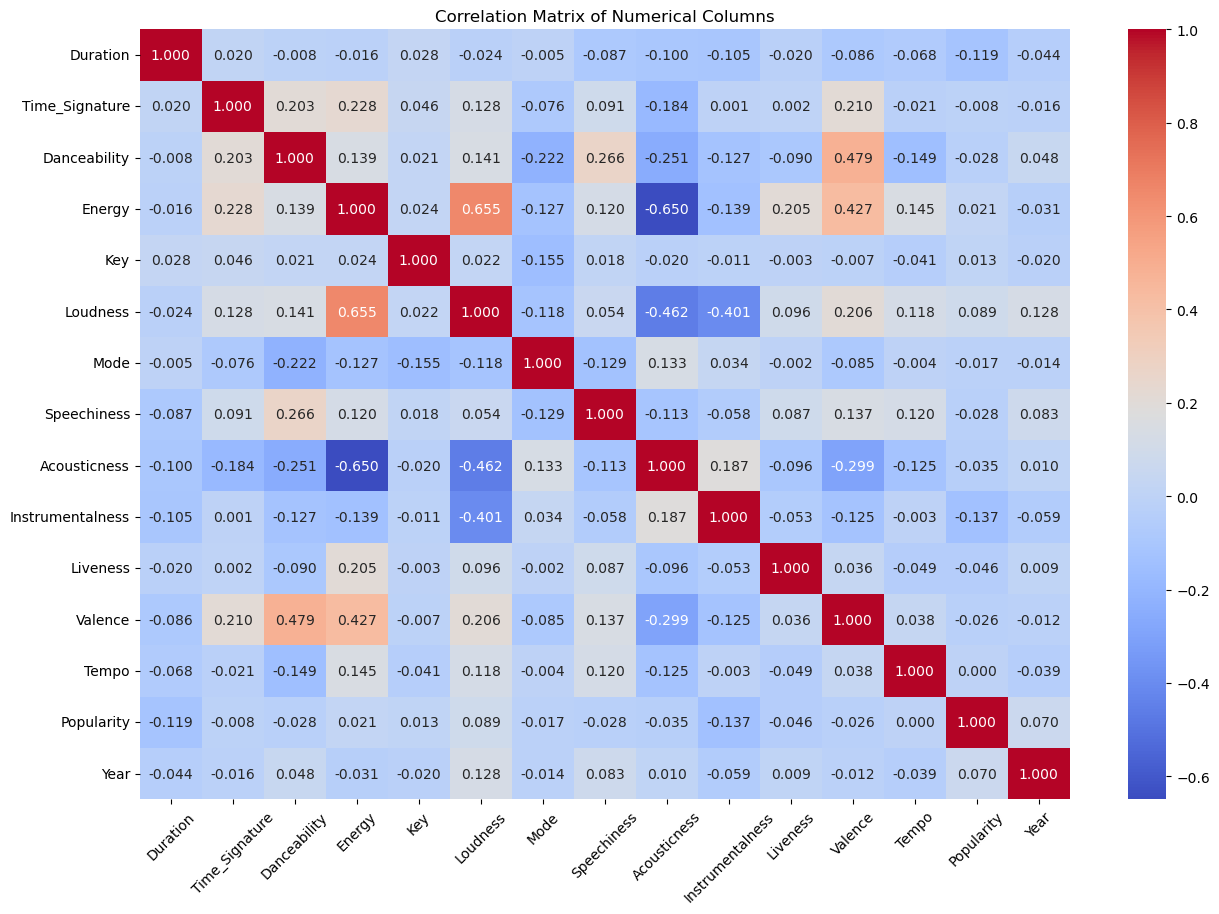
The histogram above illustrates the distribution of danceability scores, highlighting which scores are the most common. Upon analyzing the histogram, we observe that the highest bar and density peak occur between 0.6 and 0.8. This indicates that the majority of songs in this dataset from the 1990s are very danceable and enjoyable.

Additionally, the histogram is right-skewed, suggesting that most songs from the 1990s were designed to be upbeat and dance-inducing. This trend aligns with the popularity of pop and R&B music during that decade. Overall, the histogram reflects that the 1990s were a time for fun, and the music of that era often aimed to be highly danceable, which likely allowed many songs to become hits.

The horizontal bar plot displayed above illustrates the top 10 tracks with the highest average Danceability and Valence. Key observations from the plot include the following:

1. Tracks with high Danceability usually also have high Valence, suggesting that upbeat, danceable songs often sound happier.

2. Some tracks in the top 10 show slightly lower Valence but are still highly danceable, carrying a more neutral or even melancholic tone.

Additionally, the bar plot highlights the fact that the 1990s were a period when music was crafted to be danceable, regardless of the song's mood, whether it was slow and somewhat sad or fast and energetic.

The heatmap above illustrates the correlation between the numerical columns in the dataset. It reveals how different variables influence one another, which can help identify trends and patterns within the data. Notable observations from the heatmap include:

1. Valence shows a strong positive correlation with both Danceability and Energy, indicating that more energetic or danceable songs tend to feel happier or more positive.

2. Loudness has a high positive correlation with Energy, suggesting that louder songs generally have more energy.

3. Acousticness exhibits a significant negative correlation with both Energy and Valence, implying that more acoustic songs tend to sound melancholic, resulting in lower energy levels.

This heatmap provides valuable insights into the relationships within the dataset and highlights how a single aspect of a song can impact various components, ultimately influencing the overall tone of the song.

In conclusion, the analysis of the 1990s Classic Hits Dataset provided meaningful insights into the music trends of the decade. The dataset comprised 919 rows and 17 columns, featuring both categorical and numerical data. It had no missing or duplicate values. Had there been any present, I would have used the appropriate data cleaning techniques, such as filling missing categorical values with "Unknown" and replacing missing numerical values with the average of their respective columns, while dropping duplicate values.

Through data visualization, key patterns emerged. A bar plot of the top 10 artists showed that Mariah Carey, Madonna, and Janet Jackson dominated the charts, while also indicating a sharp popularity curve. This suggests that a small number of artists had a significant impact on the music scene during the 90s. A histogram displaying danceability scores revealed that most songs fell within the range of 0.6 to 0.8, reflecting the decade's emphasis on upbeat, danceable music, which was consistent with the rise of pop and R&B genres at the time. Further analysis, using bar plots and heatmaps, demonstrated that songs with high danceability often correlated with higher valence, indicating a positive and energetic tone. However, some tracks managed to balance high danceability with more neutral or melancholic moods, showing the diversity in musical expression.

The heatmap revealed strong positive correlations between features such as Loudness and Energy, along with a negative relationship between Acousticness and both Valence and Energy. This indicates that acoustic songs tended to be softer and more somber. Overall, this dataset emphasizes how music in the 1990s was designed to be lively and engaging, with key characteristics like danceability and energy playing a central role in a song's popularity. The trends observed help illustrate the vibrant and expressive nature of 90s music.

**References/Acknowledgments:**

***1990s Classic Hits (with Spotify Data)*. (2024, July 30). Kaggle. https://www.kaggle.com/datasets/thebumpkin/1990s-classic-hits-with-spotify-data**