Ehsan Kahrizi and kishore ragul_ Project 3_ Group 1

Introduction

Our objective for this project is to analyze various aspects of music listening trends. We collected our data from the Spotify API using the Spotify library. In our study, we tried to explore various aspects of music listening, like song and artist popularity, and the influence of audio features (tempo, energy, danceability) on popularity of singers. Also, we analyzed the mentioned influence on the changes in listening patterns over time. Further, we tried to find out the correlations between these audio features and popularity metrics. This can provide a valuable insights for industry professionals and companies in the music market. Our analyses include many different analyses, including visualizations of the relationship between audio features and song popularity, comparisons across genres, and seasonal listening trends. To address these questions, we used different techniques such as plotting, correlation matrices, and statistical tests like the t-test. Our key findings show that attributes like energy and danceability significantly can impact on song popularity, with variations observed across genres and over time periods. The insights from this analysis are valuable for music companies to better understand audience behavior and track trends in the industry.

The detailed results and visualizations are avaliable in our GitHub repository and Slides.

Dataset

In this study, we used data from Spotify to find trends about music trends and other analyses. We accessed the web content dataset through an API[1](Spotify API) and extracted relevant data and columns for each analysis. Spotify's data covers different aspects such as artists' names, albums, and audio features. These aspects are: top tracks and audio features for a selection of artists, including Billie Eilish, The Weeknd, Taylor Swift, and Bruno Mars. Our dataset have some information about popularity of a track, release date, and name of the artists. Also, additional features that we retrieved with the API were danceability, energy, tempo, valence, and the loudness. Further, we used our dataset to do some temporal analysis like trends to listen music in different times like morning and evening.

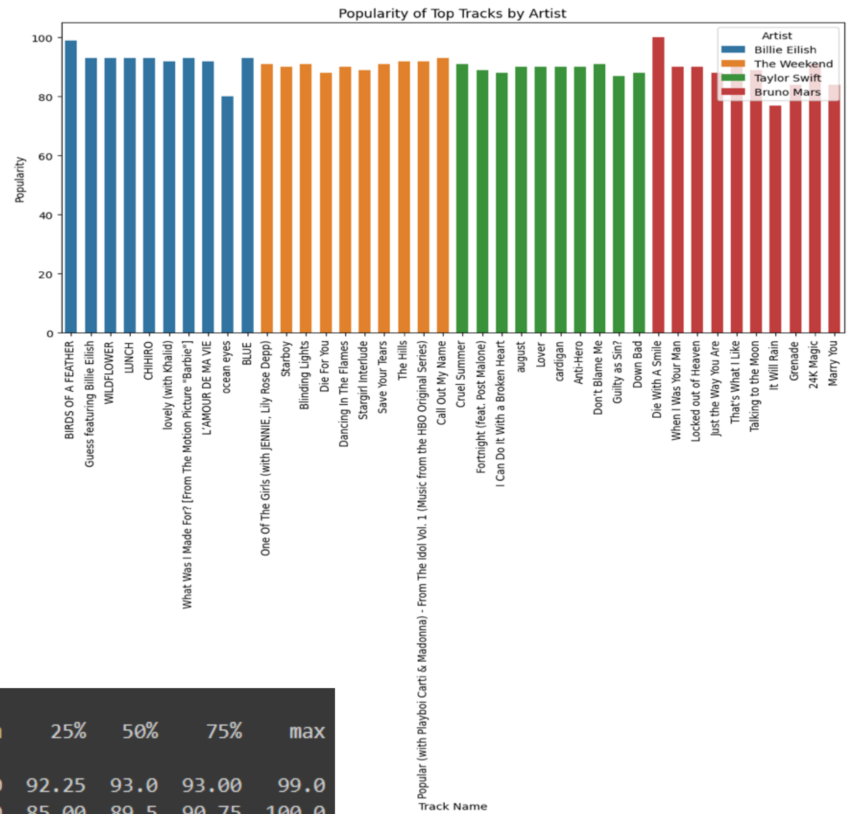
Analysis technique

In this study we used different types of analysis techniques; including correlation (scatter) plots, t-tests, kernel density estimate (KDE) plots, bar charts, and descriptive statistics such as mean and standard deviation. The correlation plots were used to show the relationship between some audio features (energy, danceability, tempo) and track popularity; to present how specific features affect of popularity scores. The T-test were applied to find the differences in listening times between morning and evening or between seasons were statistically significant. This offering insights about temporal and seasonal variations. The KDE plots presented the distribution some of audio features across tracks. This plot helped us to understand the feature distributions. Lastly, the bar charts represented a well comparison of track popularity for different artists, which highlighting signicficant performances. These different methods provided a comprehensive analytical framework to address the our study questions and providing meaningful interpretations of the dataset.

[1] Application programming interface

Ehsan Kahrizi and kishore ragul_ Project 3_ Group 1

Result

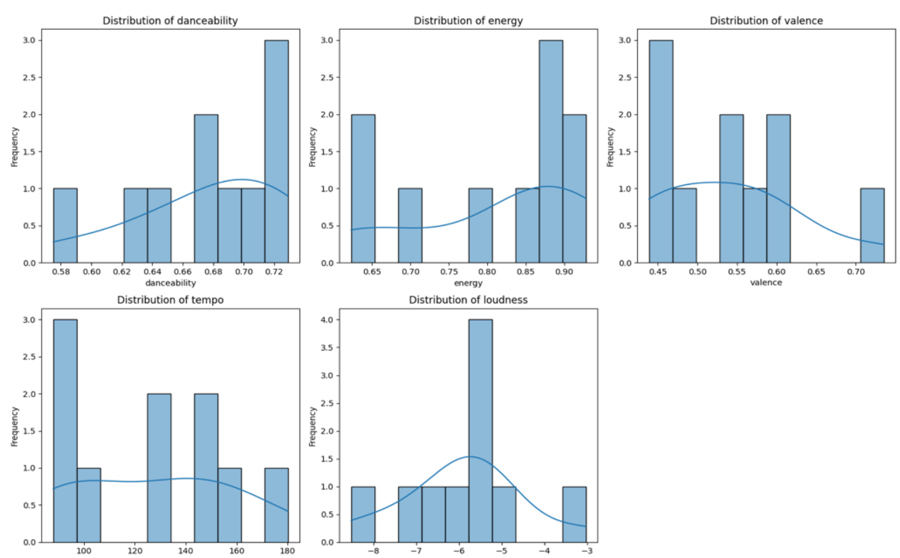


In our first analysis we looked at 4 famous singers. This bar chart shows the comparison between the popularity of the top tracks of four artists: Billie Eilish, The Weeknd, Taylor Swift, and Bruno Mars. The x-axis is the track names, and the y-axis is the popularity scores. As we can see, the Billie Eilish (Birds of a feather) and Bruno Mars's top tracks (Die with a smile) have higher popularity on average compared to other artists.

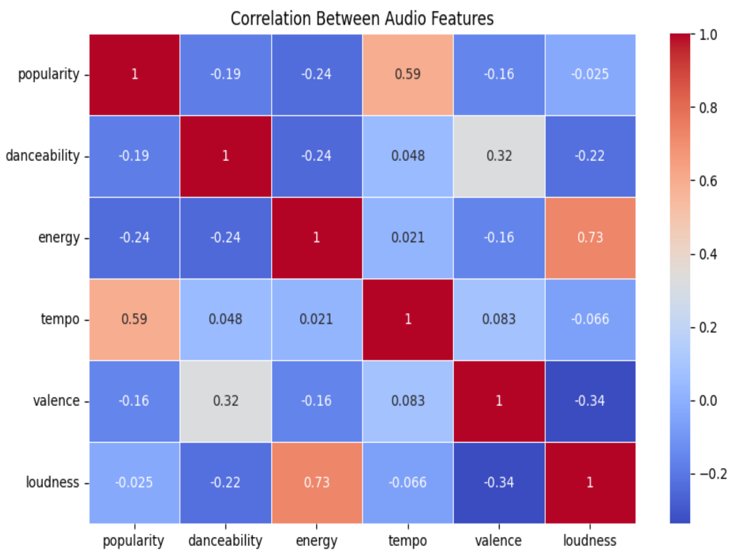
Then, we looked at the statistics, which are shown in the following figure, and we see that Billie Eilish's mean popularity is more than others:

```
Basic Statistics for Track Popularity:
               count  mean       std   min    25%   50%    75%    max
artist
Billie Eilish   10.0  92.1  4.701064  80.0  92.25  93.0  93.00   99.0
Bruno Mars      10.0  88.4  5.985167  77.0  85.00  89.5  90.75  100.0
Taylor Swift    10.0  89.4  1.349897  87.0  88.25  90.0  90.00   91.0
The Weekend     10.0  90.7  1.494434  88.0  90.00  91.0  91.75   93.0
```

In this plot, we took a look at the distribution of the five audio features (‘danceability’, ‘energy’, ‘valence’, ‘tempo’, ‘loudness’) tracks. The Y-axis of Each histogram shows the frequency of specific values for each feature, and the X-axis is the range of values for a specific audio feature (for danceability, energy, and valence, it is between 0~1 and for tempo is the number of bits/min, and for the loudness is dB). Also, the overlaid kernel density estimate (KDE) curves provide a representation of their probability distributions. The danceability and energy features show relatively uniform distributions, which indicate that the tracks have varying levels of these attributes. Valence, which measures the positivity of a song, indicates low valence tracks as the highest. Tempo, measured in beats/minute, is more

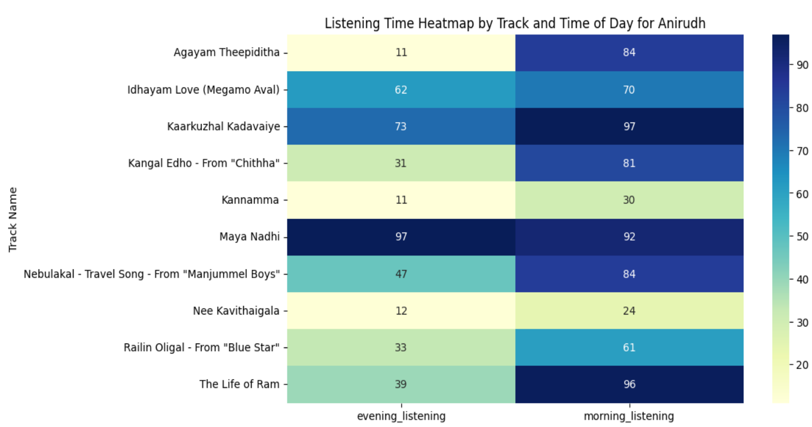


evenly spread across different ranges. Also, it shows a similarity between the distributions of Valence and Tempo distribution, indicating that we have the highest frequency in low-range beats/min and low-range Valence. Lastly, the loudness (dB) clusters around a specific range, reflecting similar sound intensities for the tracks, and it has a good distribution. Overall, these distributions provide insights into the diversity of the audio features present in the dataset for all artists together. It indicated that the tracks possess various characteristics, which can influence their popularity and listening patterns in the market.



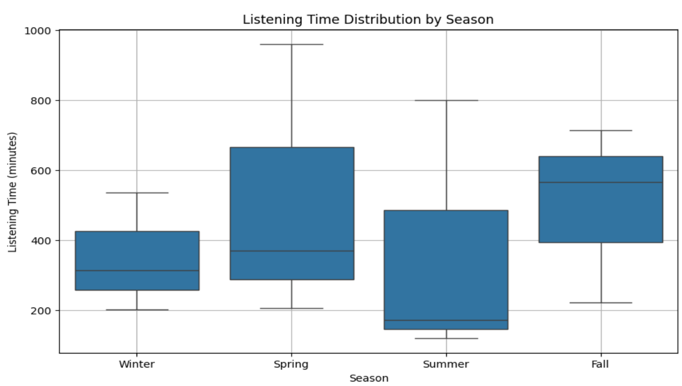
The heatmap shows the correlation matrix between the audio features that we have in our study, including popularity, danceability, energy, tempo, valence, and loudness. Each cell represents the correlation coefficient between two features that values ranging from -1 to 1. A positive value indicates a direct relationship (1), while a negative value indicates an inverse relationship (-1). The strongest positive correlation that we have seen is between energy and loudness, which is around 0.73, which indicates that higher energy tracks are often louder. The second top positive correlation is a moderate positive correlation between tempo and popularity, which is around 0.59. This indicates higher tempo tracks tend to be more popular. Further, weak correlations are observed between popularity and danceability, which is around -0.19, and also valence and energy, which is -0.16. This heatmap confusion matrix can help artists and market managers to know the relationships between different types of music and see which type may be more successful and popular.

The next step was to understand the difference between listening to songs at different times of the day (morning and evening). Darker shades indicate higher listening times. As we can see in this heatmap most of the tracks in the morning are more listened. Notably, "Maya Nadhi" has the highest listening time in both the morning and evening.



The last analysis was about the listening times during different seasons (Winter, Spring, Summer, and Fall). This box plot shows that the Spring has the highest median listening time, whereas the Winter has the lowest. The range of listening times in Spring and Fall is wider in comparison with Winter and Summer. Also, the t-test results in the table indicate that there are no statistically significant differences in listening times between Winter and Summer, as the p-value is around 0.958. Also, for Spring and Fall there are no statistically significant differences because the p-value is 0.968.

In conclusion, our results showed that the Billie Eilish and Bruno Mars's tracks generally have higher popularity scores than others. Also, we found that the Weeknd's tracks have the most consistent popularity scores (low standard deviation), while Bruno Mars has the widest range of popularity that indicates varied reception among his tracks. For analyzing the future of the track, we found that High energy and loudness are positively correlated, which indicates a strong relationship between these features in popular tracks. The listening time heatmap shows that most tracks are more popular in the morning. Also, some tracks like "Maya Nadhi" have consistently high listening times during both morning and evening. The seasonal box plot shows that listening times vary across seasons, with Spring having the highest median listening time. But, the t-test results tell us that these variations are not statistically significant, so we can say that seasonal changes do not have a major effect on listening generally.




```
T-test results between Winter and Summer:
T-statistic: -0.055659908336819976
P-value: 0.9582819899498194
T-test results between Spring and Fall:
T-statistic: 0.04296031064463953
P-value: 0.9677921496186013
```

Technical

Our work generally had some important steps. Data Preparation used the Spotify API to collect data on top tracks and audio features for artists like Billie Eilish, Taylor Swift, Bruno Mars, and The Weeknd. This dataset had information like popularity, release dates, and audio features (danceability, energy, valence). We added the listening times (morning and evening) to our analysis to find the temporal listening patterns. For data cleaning, we used the ‘pandas’ to remove the missing values and ensure consistency. Also, the release dates were converted to date-time format, and additional synthetic data was generated to evaluate seasonal listening behaviors. For the Analysis section, we performed many correlation plots to identify relationships between features such as energy and loudness. Also, we performed the T-tests to assess differences in listening times between seasons and times of day. Eventually, for the visualizations, we used histograms, KDE plots, and heat maps to show feature distributions and correlations. Lastly, we displayed the seasonal listening patterns in a box plot.