About the dataset:

The name of the dataset is ‘Spotify Track DB’ it is taken from the website Kaggle the link for the dataset is <https://www.kaggle.com/datasets/zaheenhamidani/ultimate-spotify-tracks-db>

The dataset is used to study and analyze the features of the music data available from Spotify API. In total there are 26 genres for the total 232,725 tracks available.

In depth detail about the dataset:

There are total about 18 columns which include the following features:

1. trackid
2. acousticness
3. danceability
4. duration\_ms
5. Energy
6. instrumentalness
7. key
8. liveness
9. loudness
10. mode
11. tempo
12. time\_signature
13. valence
14. artist name
15. genre
16. track name
17. type
18. time signature

Visualization 1:

Chart, bar chart

Description automatically generated

About the visualization:

The above bar graph is a graph of Popularity of the songs based on various genres. A filter has been added for the user to play around with that could help isolate certain genres or be used as per the need of the visualization. The genres have been color coded using Marks and filters have been applied on the dimension ‘Genres’.

Inference from the visualization:

The graph tells us about the popularity of various tracks based on their genres. Here, we can see that the top 3 genres are Pop, Rap and Rock whereas the 3 least genres are Movie, Children’s Music and A Capella. This graph can be used to track popularity of various genres can filters can be used to segregate data and analyze the popularities of different tracks separately.

Visualization 2:

Chart, scatter chart

Description automatically generated

About the visualization:

The above scatter plot is a comparison of duration of various keys in genres (also the mode of the keys has been isolated using the different color codes). A total of 12 keys have been plot for each genre against the duration of each in the category.

Chart, scatter chart

Description automatically generated

Inference from the visualization:

The scatter plot helps us inferring that the genres that have a high acoustic value have a lower duration of keys as the instrumentalness for such songs are also low. Similarly, for the genres like World track have a higher instrumentalness value and hence the duration is much more scattered it varies from high to low whereas for the genres like A Capella where the value of acousticness/speechiness is more have a less variance in the duration of keys as not many keys are used.

Visualization 3:

Chart, bar chart

Description automatically generated

About the visualization:

The above bar plot is a genres vs speechiness and genres vs instrumentalness. The reason for plotting this graph was to see if the speechiness and instrumentalness are inversely correlated with each other as one is related to vocality of the singer and other the beats or the sound of different instruments.

Inference from the visualization:

As per our initial hypothesis speechiness and instrumentalness of tracks are inversely correlated with each other. This is true for all the genres given in the dataset which is clearly visible from the graph above. As the values for the speechiness increases the value of instrumentalness decreases and vice versa.

Visualization 4:

Chart, bar chart

Description automatically generated

About the visualization:

The above bar plot was drawn to understand the relationship between the measures popularity and acoustic of tracks. Also, the acousticness and the popularity are further categorized using color code depending on mode major and minor.

Inference from the visualization:

The graph depicts that the popularity and acousticness of each track in most of the categories of the genres in the dataset are inversely related. But this does not hold true for all the tracks but does for most hence, the relationship can be described as moderately inverse between the 2 features.

Visualization 5:

Table

Description automatically generated

Chart, bar chart

Description automatically generated

About the visualization:

The two visualizations above have been done to understand the relationship and correlation between energy and loudness of the track. Because while initial analysis using the first plot using highlight table a general direct correlation was evident. Hence, bar plots were used to confirm the hypothesis which turned out to be true. Also, filter has been used to genres measure to give the user a greater control of how the graph could be visualized.

Inference from the visualization:

The highlighted table and the bar plot help us prove that there is direct correlation between the 2 features. As the value of energy increase the value of loudness also increases and as the value of energy is low the value of loudness would also be low. Color coding has been done to depict different types of genres.

Visualization 6:

Timeline

Description automatically generated

About the visualization:

The above horizontal bars have been plotted to analysis the popularity of modes in various keys across the tracks in the dataset. The modes have been color codes where orange is the color for minor codes and blue for the major codes.

Inference from the visualization:

We can infer from the above graph that for all the keys except E, B and A# the major key is more popular that the minor key for the tracks in different genres. But only for the above listed keys. The thickness of the major keys also shows us that most used major key is C and least used major key is D#. The most popular minor key B and least popular is D#.

Summary:

All the visualization above gives us a general understanding about how the dimensions and measures are related to each other (strong/ moderate/ weak correlation and whether the relations inversely correlated or directly correlated).

The measures genre and artist name have been organized in hierarchy. Plus, filters have been used in different visualizations and marks have been used in various sheets for better understanding of the plots.