**PROJECT REPORT**

**TITLE: SPOTIFY DATA ANALYSIS REPORT**

**DATE: 7-11-2024**

**INTRODUCTION**

Spotify is one of the most popular music streaming platforms worldwide,with millions of active users.

The aim of this data analysis was to investigate what variables may influence a song popularity while developing a greater understanding of statistical procedures.

More specifically ,the following questions were to be addressed.

1. What variables are associated with popularity of song choice by spotify users?
2. Is one variable associated with popularity above others?

In this report, I analyzed a data-set containing information on spotify songs, including their features such as songs popularity, genre, dance-ability, energy, tempo.The primary goal of this analysis is to uncover trends, patterns and correlation in music data, as well as to gain insights to user preferences and track characteristics.

**DATA SOURCE**: The data was gotten from the spotify website using the Spotify API.

**DESCRIPTION OF DATA**

The original data set included over 2060 observations and 18 variables that describe the spotify Tracks Database .The key variables include song name, genre, artist, popularity, dance-ability, energy, key, loudness, acoustiness, instrumentalness, liveness, valence, tempo. Unfortunately,the data included some unwanted character, missing values, outliers, and un-transformed data. Data cleaning was done to remove the unwanted characters , missing values, outliers and un-transformed data.

**STEPS FOR DATA CLEANING**

* I copied the data into another sheet for cleaning so as not to tamper with the original data I was given,I searched for duplicates but no duplicate was found.
* I inserted columns were necessary to filter the data,I cleaned the artist and song name using proper,trim and replace functions.
* I filled the blank spaces in the explicit using countif() function.
* I removed in the unwanted character in the year using find and replace function.
* I looked for the average of the data set and filled it into the blank spaces for the dance-ability and energy.
* I deleted the rows of outliers for the key variable
* I used find and replace function for the valence and tempo
* I used the trim,find and replace function for the genre variable

The resulting data set was 2039 observations and 18 variables.

**DATA TRANSFORMATION**

The data was raw and it was difficult to interpret so I transformed the data by grouping some of the variables into levels.

* POPULARITY- I grouped the popularity into High popularity,medium popularity and low popularity using `=ifs()` the cells of popularity is `=>70`, ”high popularity”, cells `>=60` and `>70`, ”medium popularity”,cells `<60 `,”low popularity”
* DANCE-ABILITY- I grouped the dance-ability into High dance-ability ans low dance-ability using `=ifs()` function (ifs the cells of dance-ability is <0.5,”less dance-ability”,cells>=0.5,”high dance-ability”
* VALENCE- I grouped the valence into High valence ans low valence using ifs function =ifs () the cells of valence is <0.5,”less valence”,cells>=0.5,”low valence”.
* ENERGY MEASURE- I grouped the Energy measure into High intensity and low intensity using ifs function (ifs the cells of intensity is <0.5,”less dance-ability”,cells>=0.5,”high intensity”

**DATA REPORT**

1. In what year do we have popularity that are 80 and above: 2001-2019

I used pivot chart to visualize the years and the popularity

1. Which key has the highest popularity- Key 1

I used pivot chart to visualize the key as a percentage of popularity

1. Which mode has the highest popularity-mode 0 and 1 has the same number of popularity

I used pivot chart to visualize the mode that has the highest number of popularity

1. Which song has the highest popularity-Sweater weather

I used vlookup() formula to get the song with the highest popularity

1. Which Energy has the most popularity-0.718

I used pivot chart to visualize the energy of song with the highest popularity

1. How many hip hop genre do we have in the popularity- 127

I used pivot chart to visualize and also confirmed the answer using =countif formula

1. How many rock, pop genre do we have in the popularity- 4

I used formula `=countif()` to generate my answer

**DATA VISUALIZATION**

1. **COUNT OF GENRE BY YEAR AND POPULARITY LEVEL**

I used the excel pivot table to create a chart that shows genre by year and popularity level. From this chart, I observed that the highest popularity level was in 2018 and the lowest popularity was in 1998 and 2020 probably due to an outbreak of disease which caused lockdown.

CHART 1:COUNT OF GENRE BY YEAR AND POPULARITY LEVEL.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Count of genre** | **POPULARITY LEVEL** |  |  |  |
| **year** | **HIGH POPULARITY** | **MEDIUM POPULARITY** | **LOW POPULARITY** | **Grand Total** |
| 1998 | - | - | 1 | 1 |
| 1999 | 10 | 17 | 11 | 38 |
| 2000 | 15 | 20 | 41 | 76 |
| 2001 | 17 | 50 | 45 | 112 |
| 2002 | 26 | 28 | 34 | 88 |
| 2003 | 15 | 37 | 43 | 95 |
| 2004 | 18 | 30 | 48 | 96 |
| 2005 | 32 | 47 | 26 | 105 |
| 2006 | 25 | 30 | 42 | 97 |
| 2007 | 27 | 33 | 36 | 96 |
| 2008 | 35 | 37 | 26 | 98 |
| 2009 | 28 | 30 | 28 | 86 |
| 2010 | 41 | 41 | 28 | 110 |
| 2011 | 39 | 33 | 27 | 99 |
| 2012 | 55 | 36 | 31 | 122 |
| 2013 | 34 | 25 | 34 | 93 |
| 2014 | 48 | 31 | 28 | 107 |
| 2015 | 54 | 15 | 34 | 103 |
| 2016 | 42 | 24 | 34 | 100 |
| 2017 | 53 | 24 | 41 | 118 |
| 2018 | 78 | 19 | 10 | 107 |
| 2019 | 51 | 18 | 20 | 89 |
| 2020 |  |  | 3 | 3 |

TABLE 1:COUNT OF GENRE BY YEAR AND POPULARITY LEVEL

2.**COUNT OF SONG VS POPULARITY LEVEL**: I used the pivot table to create a chart comparing the count of songs and their popularity level and I observed that 743 song have high popularity level,625 songs has medium popularity level and 671 songs has low popularity level.

CHART 2: COUNT OF SONG BY POPULARITY LEVEL

1. **PERCENTAGE OF GENRE BY YEAR AND POPULARITY LEVEL**

I Used the pivot table to create a chart comparing the percentage of genre by year and popularity level and I observed that the highest popularity was in 2018 which had 72.90% of high popularity of genre,17.76% of medium popularity of genre and 9.35% of low popularity of genre.the lowest popularity of genre was in 1998 and 2020.

CHART 3: PERCENTAGE OF GENRE BY YEAR AND POPULARITY LEVEL

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Count of genre** | **POPULARITY LEVEL** |  |  |  |
| **year** | **HIGH POPULARITY** | **MEDIUM POPULARITY** | **LOW POPULARITY** | **Grand Total** |
| 1998 | 0.00% | 0.00% | 100.00% | 100.00% |
| 1999 | 26.32% | 44.74% | 28.95% | 100.00% |
| 2000 | 19.74% | 26.32% | 53.95% | 100.00% |
| 2001 | 15.18% | 44.64% | 40.18% | 100.00% |
| 2002 | 29.55% | 31.82% | 38.64% | 100.00% |
| 2003 | 15.79% | 38.95% | 45.26% | 100.00% |
| 2004 | 18.75% | 31.25% | 50.00% | 100.00% |
| 2005 | 30.48% | 44.76% | 24.76% | 100.00% |
| 2006 | 25.77% | 30.93% | 43.30% | 100.00% |
| 2007 | 28.13% | 34.38% | 37.50% | 100.00% |
| 2008 | 35.71% | 37.76% | 26.53% | 100.00% |
| 2009 | 32.56% | 34.88% | 32.56% | 100.00% |
| 2010 | 37.27% | 37.27% | 25.45% | 100.00% |
| 2011 | 39.39% | 33.33% | 27.27% | 100.00% |
| 2012 | 45.08% | 29.51% | 25.41% | 100.00% |
| 2013 | 36.56% | 26.88% | 36.56% | 100.00% |
| 2014 | 44.86% | 28.97% | 26.17% | 100.00% |
| 2015 | 52.43% | 14.56% | 33.01% | 100.00% |
| 2016 | 42.00% | 24.00% | 34.00% | 100.00% |
| 2017 | 44.92% | 20.34% | 34.75% | 100.00% |
| 2018 | 72.90% | 17.76% | 9.35% | 100.00% |
| 2019 | 57.30% | 20.22% | 22.47% | 100.00% |
| 2020 | 0.00% | 0.00% | 100.00% | 100.00% |

TABLE 3:PERCENTAGE OF GENRE BY YEAR AND POPULARITY LEVEL

4.**COUNT OF SONG BY DANCEABILITY AND POPULARITY LEVEL**

I used the pivot table to create a pivot chart comparing the percentage of song by danceability and popularity level and I observed that the danceability level (describes how suitable a track is for dancing based on a combination of musical elements including tempo, rhythm stability, beat strength, and overall regularity)of the songs greatly affects the popularity level of song I.e the popularity level was low when the danceability level was low.

CHART 4: COUNT OF POPLUARITY LEVEL BY DANCEABILITY LEVEL

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Count of POPULARITY LEVEL** | **POPULARITY LEVEL** |  |  |  |
| **DANCEABILITY LEVEL** | **HIGH POPULARITY** | **MEDIUM POPULARITY** | **LOW POPULARITY** | **Grand Total** |
| HIGH DANCEABILITY | 655 | 549 | 596 | 1800 |
| LESS DANCEABILITY | 88 | 76 | 75 | 239 |

TABLE 4:COUNT OF SONG BY DANCEABILITY AND POPULARITY LEVEL

1. **PERCENTAGE OF THE POPULARITY LEVEL BY EXPLICIT**

I used the pivot table to create a chart comparing the percentage of valence level by explicit and popularity level and I observed that whether there is presence of explicit ( lyrics or content of a song or a music video contain one or more of the criteria which could be considered offensive or unsuitable for children) or not in a song doesn’t affect their popularity level.

CHART 5:PERCENTAGE OF POPULARITY LEVEL BY EXPLICIT

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Count of POPULARITY LEVEL** | **POPULARITY LEVEL** |  |  |  |
| **explicit** | **HIGH POPULARITY** | **LOW POPULARITY** | **MEDIUM POPULARITY** | **Grand Total** |
| FALSE | 34.70% | 33.97% | 31.33% | 100.00% |
| TRUE | 41.44% | 29.85% | 28.71% | 100.00% |

TABLE 5 : TABLE OF POPULARITY LEVEL BY EXPLICIT

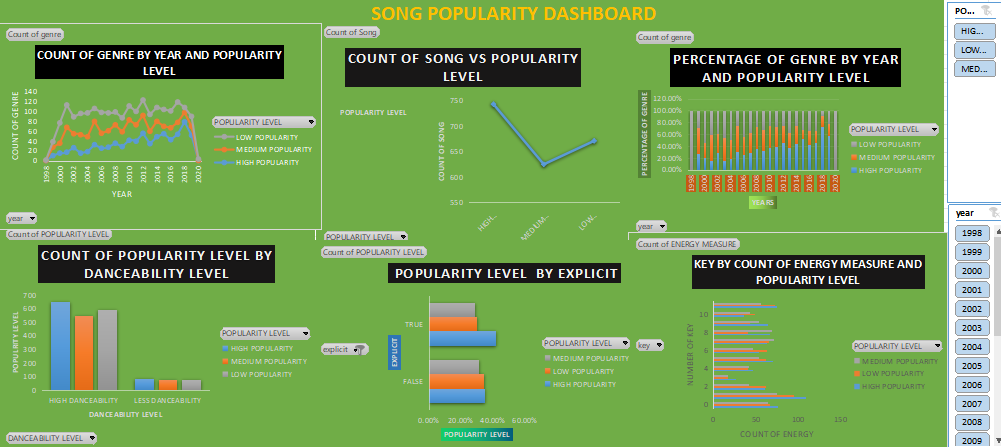
6.**THE KEY OF SONG BY COUNT OF ENERGY MEASURE AND POPULARITY LEVEL**

I used the pivot table to create a pivot chart to compare the key of song by the count of energy measure and popularity level.I observed that key 1(KEY 1 = C♯/D♭)has the highest popularity level and the highest count of energy measures.

CHART 6:KEY BY COUNT OF ENERGY MEASURE AND POPULARITY LEVEL

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Count of ENERGY MEASURE** | **POPULARITY LEVEL** |  |  |  |
| **key** | **HIGH POPULARITY** | **LOW POPULARITY** | **MEDIUM POPULARITY** | **Grand Total** |
| 0 | 76 | 66 | 64 | 206 |
| 1 | 109 | 94 | 74 | 277 |
| 2 | 60 | 61 | 41 | 162 |
| 3 | 26 | 17 | 17 | 60 |
| 4 | 46 | 40 | 41 | 127 |
| 5 | 70 | 62 | 53 | 185 |
| 6 | 53 | 63 | 46 | 162 |
| 7 | 63 | 65 | 71 | 199 |
| 8 | 66 | 40 | 68 | 174 |
| 9 | 64 | 43 | 53 | 160 |
| 10 | 36 | 48 | 42 | 126 |
| 11 | 74 | 72 | 55 | 201 |
| TOTAL | 743 | 671 | 625 | 2039 |

TABLE 6:KEY BY COUNT OF ENERGY MEASURE AND POPULARITY LEVELS



SPORTIFY DATASET DASHBOARD

**SUMMARY and EXCERPTS**

The season a song is released greatly affects its popularity because based on this analysis the genre released in 2020 and 1998 had low percentage of popularity.Based on this data,most song from 1998-2020 has high popularity,but there was also large numbers of songs with low popularity and few songs with medium popularity.Also the danceability level (describes how suitable a track is for dancing based on a combination of musical elements including tempo, rhythm stability, beat strength, and overall regularity)of the songs greatly affects the popularity level of song I.e high danceablity level has high popularity and low danceability level has low popularity level.The presence of explicit( ( lyrics or content of a song or a music video contain one or more of the criteria which could be considered offensive or unsuitable for children) or not in a song doesn’t affect their popularity level.

The key (Integers map to pitches using standard Pitch Class notation )of songs affects its popularity.Songs that uses key 1(KEY 1 = C♯/D♭)has the highest popularity level and the highest count of energy measures.

**CONCLUSION**

Based on the analysis of music popularity trends from 1998 to 2020, several key insights emerge regarding factors that influence the success of songs.

1.Count Of Genre By Year And Population Level: The timing of a song's release appears to have a significant impact on its popularity. Specifically, the years 1998 and 2020 saw a lower number of popular songs, which could be attributed to external factors like the global disease outbreaks during these periods. This suggests that external circumstances, such as pandemics, may disrupt the normal patterns of music consumption and popularity, possibly due to altered social behaviors or economic conditions.

2.Count Of Song Vs Popularity Level: Between 1998 and 2020, the majority of songs experienced high popularity, but there was also a notable number of songs with low popularity and very few songs with medium popularity. This suggests that, in general, the music industry has seen a polarization of success, where either songs become highly popular or fail to gain traction.

3.Count Of Song Vs Danceability Level:Interestingly, the danceability of a song, which refers to its tempo, rhythm stability, and overall regularity, greatly influence its popularity level. Songs with high danceability, have high popularity and songs with less danceability level has low popularity. This indicates that danceability of a song appeal to most audiences, it is a determining factor for a song's overall success.

4.The presence of explicit content in a song or music video does not appear to have a significant effect on its popularity. Songs with both explicit and non-explicit content have achieved similar levels of success, suggesting that listeners are not particularly influenced by this criterion when choosing music.

5. One of the most notable factors influencing a song's popularity is its key. Songs that use **Key 1 (C♯/D♭)** have the highest popularity levels and the greatest number of energy measures, suggesting that certain musical keys may resonate more with listeners, contributing to their overall appeal. This insight could guide music producers and artists in crafting songs that align with key signatures proven to attract larger audiences.

**RECOMMENDATIONS**

Given these findings, it is clear that while many factors influence a song's popularity, the key of the song and the timing of its release play crucial roles. Artists and producers may want to focus on releasing songs in certain keys (such as Key 1) to maximize potential success, and consider the broader societal context when planning release schedules.Song that are more danceable should be produced. Additionally, while explicit content don't seem to affect popularity, paying attention to musical trends and external factors like global events could help navigate the fluctuating landscape of music popularity.