Final Project

Jesigga

2023-12-07

Stat 184 Final Project

Taylor Swift

A R script that analyzes the dataset of Taylor Swift songs pulled from Spotify's API.

Dataset: https://raw.githubusercontent.com/rfordatascience/tidytuesday/master/data/2023/2023-10-17/taylor_all_songs.csv.

Introduction

Taylor Swift is a superstar who managed to completely cross over from country to the mainstream, becoming an enduring pop culture icon and conquering the world in the process. Taylor moved from her country roots, revealing that she was perhaps the savviest singer/songwriter of her generation, one who could make it personal and still fun and enjoyable.

In this project, I will analyze an imported dataset from Spotfiy's API on Swift's discography. From the raw data, I have decided to extract information regarding featured artists and ratings of the artist's songs on danceability and energy.

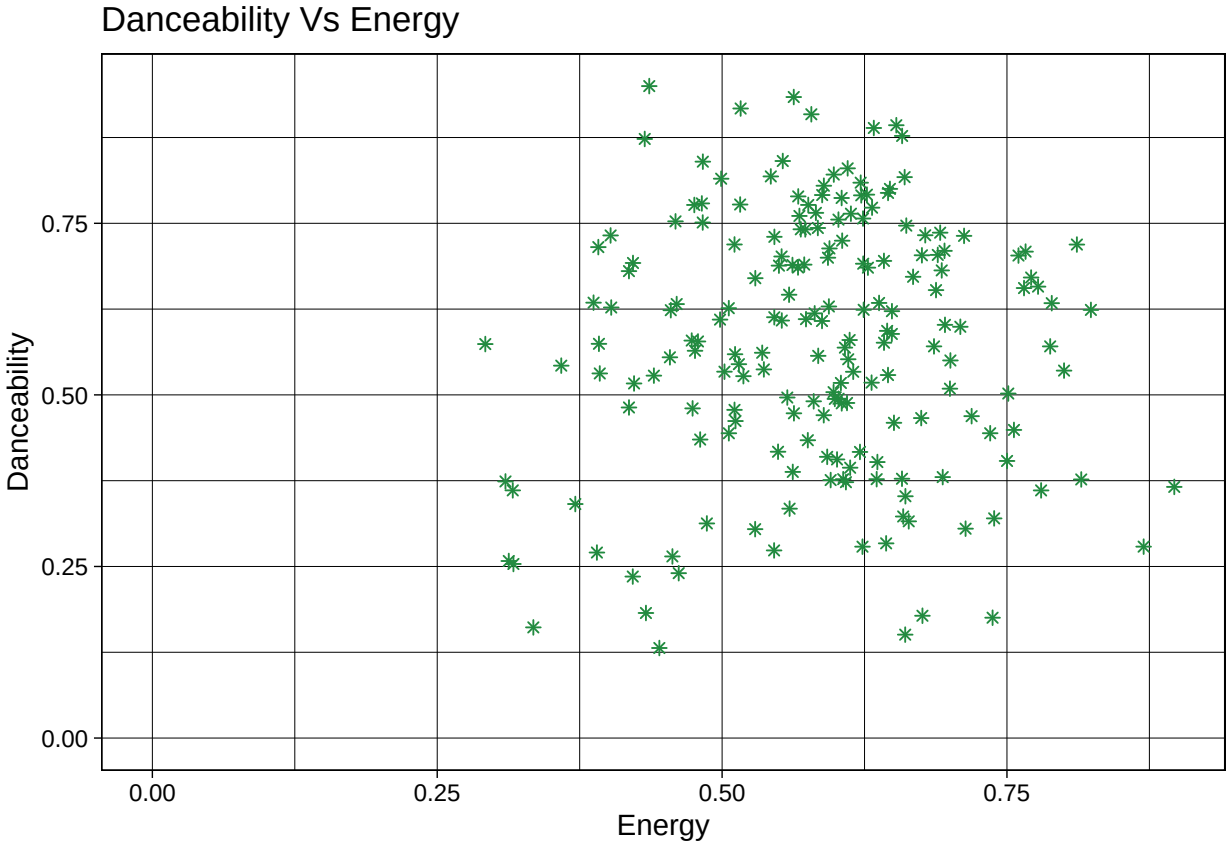
This topic holds significant importance in current times and society, as Swift is a major artist within the music industry. This data also holds benefits for those wanting to explore how music streaming services, such as Spotify, input billions of songs and run many complex analyses of songs.

I have tried to calculate how and why Taylor Swift has become so popular with the general public by analyzing Spotify's data set on danceability and

energetic-ness. I have created visual graphs to support my hypotheses and to highlight the trends and statistics of her musical catalog.

Trend in Danceability and Energy

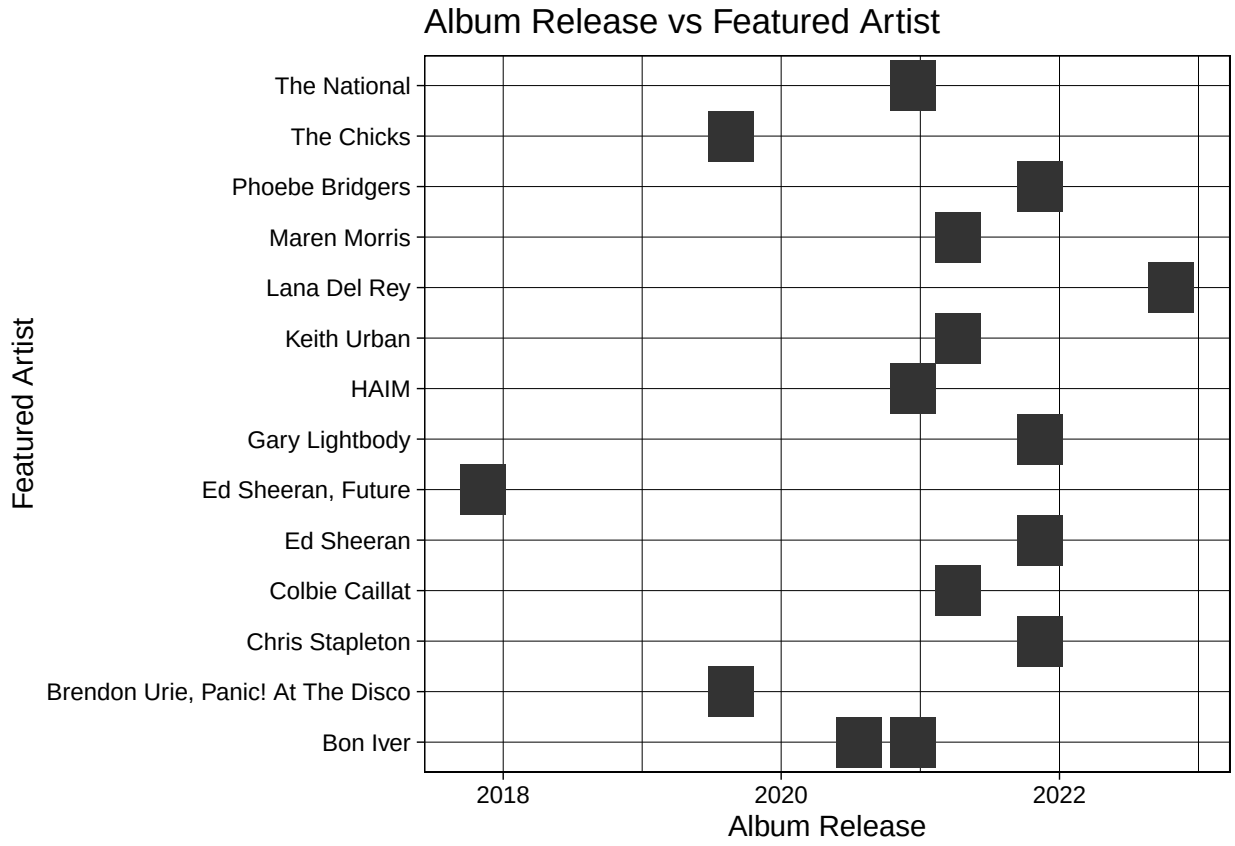
I was curious as to how someone can make songs that are fun, energetic, and easy to dance to while still connecting with people. I was curious to see if having songs that are more energetic and easy to dance to would correlate with each other. For example, would a song that is very high in energy be considered a more danceable song. This leads to my first visualization.



I wanted to find the correlation between danceability, and energy represented by a graph. From this graph, I was able to see that energy and danceability are related but not linearly.

Trend in Featured Artist over Time

Another question I found myself pondering was would artists being featured have anything to do with her popularity? For example would she start with having multiple featured artists on her albums and that's how she gained popularity or did she become famous and then start featuring other artists? This leads me to my next visualization.



I was curious to see if There was any trend in people being featured in her albums as her career progressed and as you can see there are more featured artists later in her career.

Final Thoughts

Overall this has been a lot of fun playing with Taylor's song data and I found some interesting information from this. I found how although it's not exactly linear her danceability versus energy graph shows that more energetic songs would be considered more danceable according to Spotify' data, which I thought was really interesting. This is very similar to what I thought, but I wasn't expecting the data to be so condensed in the higher sections of both danceability and energy.

I thought it was fun to see how in the very beginning of her albums being released, there were not any artists being featured on her albums. She didn't really start having anyone featured on her albums until about 2018, which is when the featured artists starts, but her first albums were released much earlier in her career when she was originally more of a country singer, which I thought was really fascinating to see.

Overall, I feel like my predictions versus what the data showed were quite similar and can see how she became very popular then featured artists on her albums. You can also see that the more energetic a song is the more danceable it would be according to Spotify.

Code Apendix

```
### Load Packages
library(ggplot2)
library(esquisse)
library(dplyr)
library(tidyverse)

## Import Data Set For Taylor Songs
taylor_album_songs <- readr::read_csv('https://raw.githubusercontent.com/

##Clean Data To specific Columbus
##Subset of the dataframe to select attributes that will be used.
certan_col = subset(taylor_album_songs, select = c(album_name, album_rele
Taylor_clean_data <- certan_col %>%
  filter(!is.na(album_name)) %>%
  filter(!is.na(artist)) %>%
  arrange(desc(album_release))

## Data visualization for dancability Vs energy
ggplot(Taylor_clean_data) +
  aes(x = danceability, y = energy) +
  geom_jitter(shape = "asterisk",  # changing shape
    size = 1.0,
    colour = "#228B40")+ # changing color
   expand_limits(x = 0, y = 0) + # starting at zero
  labs(
    x = "Energy",
    y = "Danceability", # labels
    title = "Danceability Vs Energy"
  ) +
  theme_linedraw()


## second data visualization Featured artists over time
```

```r
Taylor_clean_data %>%
  filter(!is.na(featuring)) %>%  # removes N/A songs
  ggplot() +
  aes(x = album_release, y = featuring) +
  geom_tile() +
  labs(
    x = "Album Release",
    y = "Featured Artist", # labels
    title = "Album Release vs Featured Artist"
  ) +
  theme_linedraw()
```