

Spotify: What Makes a Top Song?

DigHum 100 Dr. Anderson | Gloria Yang | May 27, 2021



Description: I've noticed that features of the most popular songs have changed a lot over the decades. While there are influences and consistencies in popular song attributes from the 20th century, the 21st century brought up new genres and styles of music. Modernization has led to different approaches and greater diversity of sounds such as louder sounds, faster tempos, and speechiness. In this project, I will be analyzing the popular songs on Spotify over time and looking for common attributes and trends.

Dataset: "Spotify: All Time Top 2000 Mega Dataset" is a dataset taken from Kaggle which contains audio statistics and attributes for Spotify's top songs. The dataset is contained within a csv file and includes data on 1995 songs released from 1959 to 2019. The data set breaks down each song by the following attributes: genre, year, BPM, energy, danceability, loudness, liveness, valence, length, acousticness, speechiness, and popularity.

I will be able to cluster and manipulate the dataset's columns to make our visualizations more clear during exploration. I will be using the dataset in order to explore the attributes of Spotify's most popular songs and the changes in these attributes over time.

Index	Title	Artist	Pop. Genre	Year	Minute	Energy	Danceability	Loudness (dB)	Liveness	Valence	Speechiness	Acousticness	Popularity
1	Smells Like Teen Spirit	Nirvana	grunge	1992	3:43	0.66	0.51	-1.7	0.16	0.05	0.04	0.04	71
2	Billie Jean	Michael Jackson	pop	1983	7:07	0.69	0.69	-1.4	0.26	0.99	0.01	0.01	69
3	Stayin' Alive	Bee Gees	disco	1977	6:54	0.59	0.59	-1.2	0.16	0.59	0.01	0.01	68
4	The Pretender	Fox Fighters	alternative	2007	3:29	0.66	0.51	-1.7	0.16	0.05	0.04	0.04	71
5	Water On A Burning Day	Bruce Springsteen	classic rock	2002	3:06	0.66	0.51	-1.7	0.16	0.05	0.04	0.04	71
...
1990	How Deep Is Your Love	Elvis Presley	adult standards	1958	3:04	0.66	0.51	-1.7	0.16	0.05	0.04	0.04	71
1991	Hotel California	Elvis Presley	adult standards	1958	3:36	0.66	0.51	-1.7	0.16	0.05	0.04	0.04	71
1992	Johnny B. Goode	Chuck Berry	blues rock	1959	3:00	0.66	0.51	-1.7	0.16	0.05	0.04	0.04	71
1993	Take Five	The Dave Brubeck Quartet	bebop	1959	5:14	0.66	0.51	-1.7	0.16	0.05	0.04	0.04	71
1994	Blueberry Hill	Fats Domino	adult standards	1959	3:03	0.66	0.51	-1.7	0.16	0.05	0.04	0.04	71

Hyperlink to Jupyter: https://colab.research.google.com/drive/1DsZ4s_TyKT_TuBdc2z3GnCYQGoZT_HXB?usp=sharing
Hyperlink to CSV: <https://drive.google.com/file/d/16BOMsQKc0WRAMaaHU9TBXTGERMbiz2ul/view?usp=sharing>

Questions for EDA:

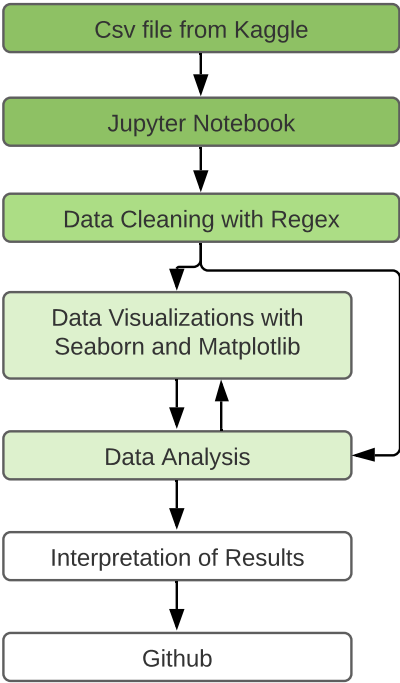
- Can we predict a song's popularity based on attributes such as danceability, valence, loudness, and length?
- Are there any trends in the popularity of different music genres?
- How do song attributes differ according to different genres?
- How have the attributes of songs on Spotify's Top 2000 changed over time?

Tools:

- Matplotlib and Seaborn:** tool to create data visualizations such as heatmaps, line graphs, and scatter plots
- Pandas:** to manipulate, clean, and store data within dataframes
- Jupyter Notebooks:** tool for data cleaning, data manipulation, and developing visualizations
- Regex:** implemented to clean genre data from subgenres
- Github:** tool for creating a portfolio repository that includes the original dataset, Jupyter notebook, and storyboard

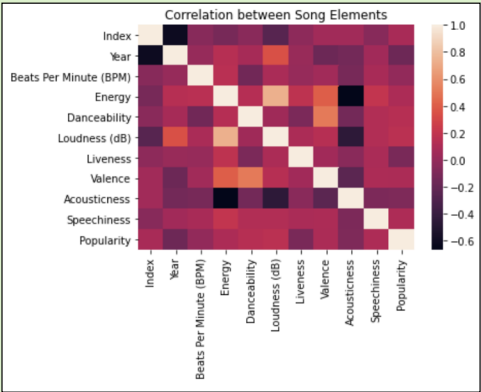
Methodology:

- I will be converting the "Spotify: All Time Top 2000 Mega Dataset" csv file into a dataset. Tools such as Matplotlib and Regex will also be imported.
- Using regex, songs will be categorized from 10 main genres; songs that have other genres will be categorized as "Other". Data will be further cleaned and formatted.
- I will be exploring correlation and typical ranges between different song attributes. A heatmap will be derived from the dataset to see if there are any correlations between any two song attributes.
- Using Pandas and data visualization, I will explore possible shifts in song attributes and genre distribution. If there are shifts or existing attribute characteristics among popular songs, I will try to see such attributes are helpful in predicting a song's popularity.



Results:

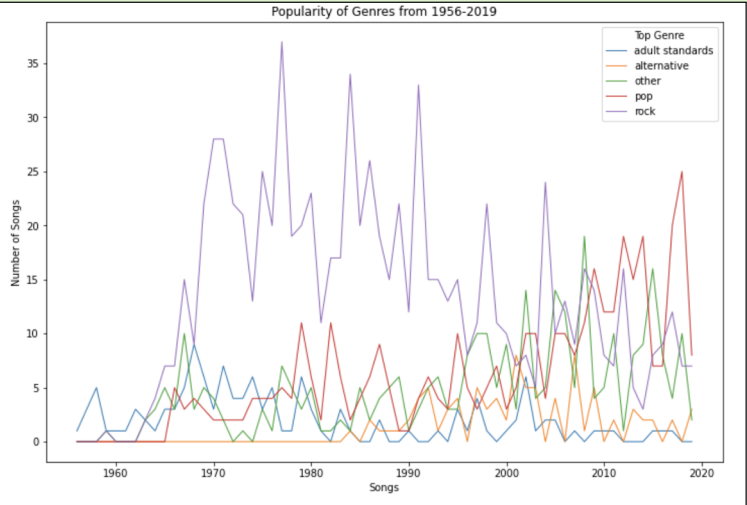
I started playing around with my dataset to get a beginning idea of existing correlations that I can explore regarding song attributes. My heatmap indicated a strong correlation between loudness and energy. Further analysis on a scatter plot showed that the correlation was positive; the more energetic a song, the more likely to the song is to feel loud.



I was also interested in seeing whether popular songs had ideal BPMs and song lengths. With histograms, I was able to see that the mean for song duration was approximately 4.15 minutes. The majority of songs also had a BPM with the range of 100-140.

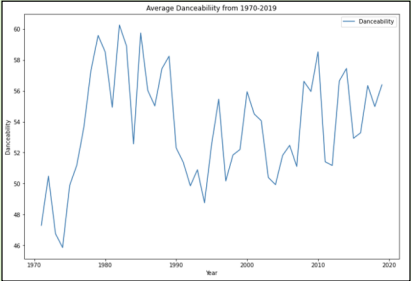
However, in terms of popularity prediction, attributes in the dataset are taken from top songs over a broad period of time and might not be representative of songs in the present day. Modernization has led to different approaches and greater diversity of sounds such as louder sounds, faster tempos, and speechiness. We must therefore also examine trends in changes to attributes over time.

I used regex to categorize subgenres into the following genres: adult standards, alternative, country, electronic, folk, funk, hip pop, indie, metal, pop, rock, soul, and other. Across all songs in the dataset, the most popular genres were different types of rock and pop.

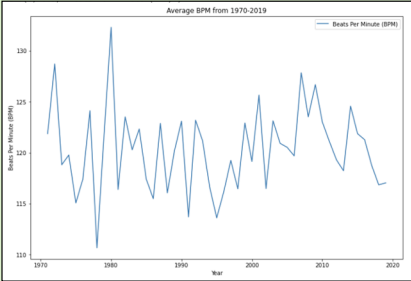
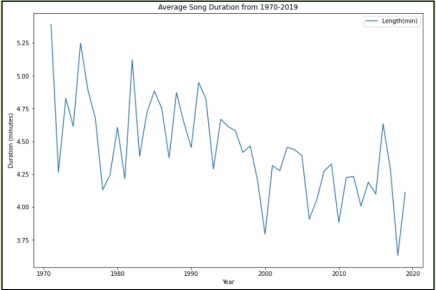


Over time, we can see that the most popular genre shifted from rock to pop. Songs within the dataset became more diverse in terms of genre. Due to genre diversification, I expected to see additional shifts in song attributes such as length and BPM.

Results Cont:



We can see that average danceability increased from 1970-2019.



MY FREE TRIAL NO LONGER
ALLOWS ME ADD FURTHER TEXT
OR IMAGES

Works Cited

Biss, Madars. "Rhythm Tips for Identifying Music Genres by Ear." Musical U, Easy Ear Training Ltd., 14 Feb. 2017, www.musical-u.com/learn/rhythm-tips-for-identifying-music-genres-by-ear/.

Esty, Thomas. "Trends Over Time - Music Popularity Data Analysis." Google Sites, sites.google.com/site/musicpopularitydataanalysis/trends-over-time.

Fadelli, Ingrid. "Using Spotify Data to Predict What Songs Will Be Hits." Tech Xplore - Technology and Engineering News, Tech Xplore, 9 Sept. 2019, techxplore.com/news/2019-09-spotify-songs.html.

Przybyla, Matt. "Predicting Spotify Song Popularity." Medium, Towards Data Science, 3 Feb. 2021, towardsdatascience.com/predicting-spotify-song-popularity-49d000f254c7.

Singh, Sumat. "Spotify - All Time Top 2000s Mega Dataset." Kaggle, 4 Feb. 2020, www.kaggle.com/iamsumat/spotify-top-2000s-mega-dataset.