Spotify: What Makes a Top Song?

DigHum 100 Dr. Anderson | Gloria Yang | June 10, 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 songs released from 1959 to 2019. The data set breaks down each song by the following attributes: genre, year, BPM, danceability, energy, loudness, liveness, valence, length, acousticness, speechiness, and popularity. Definitions for attributes can be found <u>here</u>.

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	Top Genre	Year	Minute (BPM)	Energy	Danceability	Loudness (dB)	Liveness	Valence	Length (Duration)	Acousticness	Speechiness	Popularity
1	Sunrise	Norah Jones	adult standards	2004	157	30	53	-14	11	68	201	94	3	71
2	Black Night	Deep Purple	album rock	2000	135	79	50	-11	17	81	207	17	7	39
3	Clint Eastwood	Gorillaz	alternative hip hop	2001	168	69	66	-9	7	52	341	2	17	69
4	The Pretender	Foo Fighters	alternative metal	2007	173	96	43	-4	3	37	269	0	4	76
5	Waitin' On A Sunny Day	Bruce Springsteen	classic rock	2002	106	82	58	-5	10	87	256	1	3	59
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
1990	Heartbreak Hotel	Elvis Presley	adult standards	1958	94	21	70	-12	11	72	128	84	7	63
1991	Hound Dog	Elvis Presley	adult standards	1958	175	76	36	-8	76	95	136	73	6	69
1992	Johnny B. Goode	Chuck Berry	blues rock	1959	168	80	53	-9	31	97	162	74	7	74
1993	Take Five	The Dave Brubeck Quartet	bebop	1959	174	26	45	-13	7	60	324	54	4	65
1994	Blueberry Hill	Fats Domino	adult standards	1959	133	50	49	-10	16	83	148	74	3	56

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

Csv file from Kaggle

↓

Jupyter Notebook

↓

Data Cleaning with Regex

↓

Data Visualizations with Seaborn and Matplotlib

↓

Data Analysis

↓

Interpretation of Results

↓

Github

Methodology:

1) 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.

2) 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.

3) 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.

4) 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 if and which attributes are helpful in predicting a song's popularity.

SPOTIFY'S MOST PLAYED ALL-TIME

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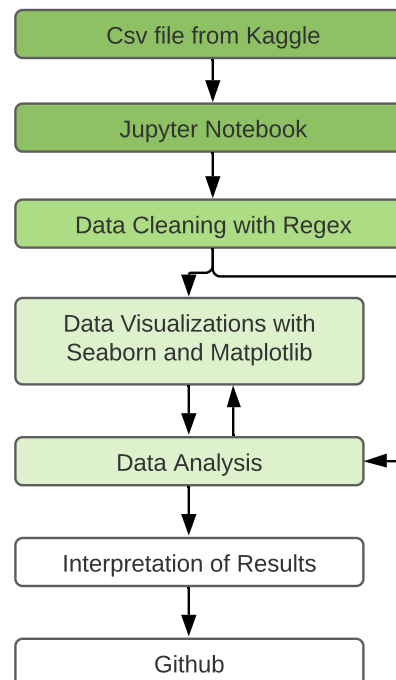
Correlation between Song Elements
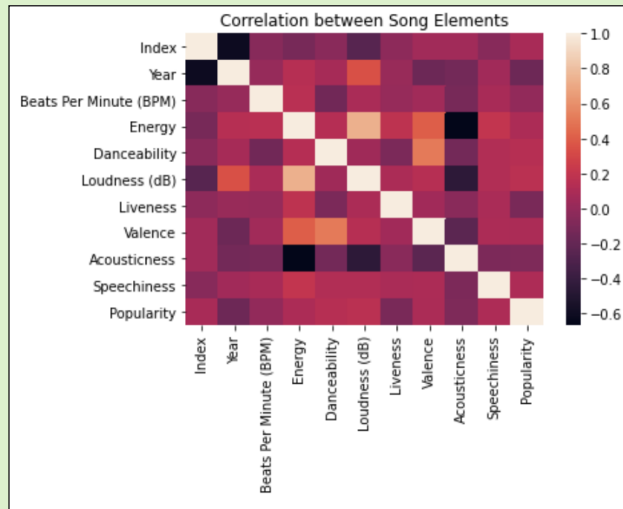
I started playing around with my dataset to get a beginning idea of existing correlations between song attributes. My heatmap indicated a strong correlation between loudness and energy. Further analysis on a scatter plot showed the correlation was positive; the more energetic a song, the more likely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

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

I then 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.

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.
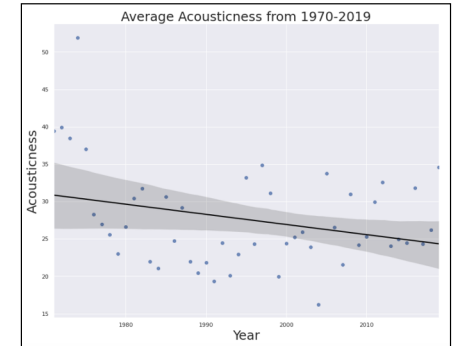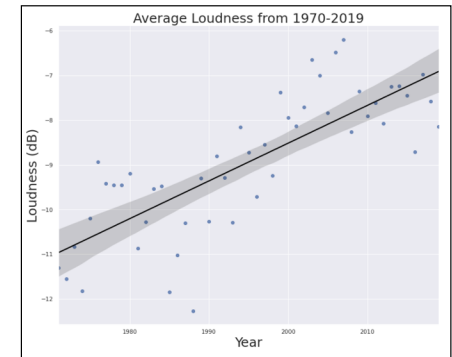
Results Cont:

Popularity of Genres in the Spotify Top 2000 from 1956-2019
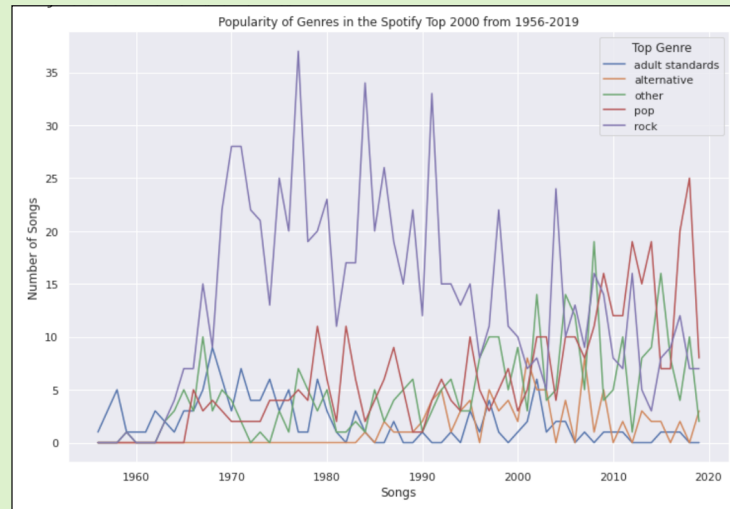
The previous graph shows the number of songs included in Spotify's Top 2000 over time in the following genres: adult standards, pop, rock, alternative, and other. While the number of of rock songs decline into the 21st century, genre diversification is shown through growth for other genres in the number of songs in Spotify's Top 2000 over time. Due to this genre diversification, I expected to see additional shifts in song attributes such as length and BPM.

Over time, we can see that the most popular genre shifted from rock to pop. When I examined other attributes for trends over time, I also noticed that attributes such as accousticness, valence, and duration tended to decrease while attributes like loudness and da tended to increase. According to The Atlantic, music, particularly pop music, has become intrinsically louder due to current trends in the industry; valence has also decreased due to widespread content for more negative and sadder content. Likewise, due to new technology, electronic sounds have become more prominent over time. Music has become more electronic and less acoustic, mirroring our modern society's integration with technology. With further research, I also learned that due to the emergence music streaming platforms like Spotify. Artists are paid by how many songs are streamed rather than physical sales, which leads to shorter songs and new song structures such as pop overture. Therefore, besides gender diversification, a song's popularity is influenced by emotional preference and technological and streaming modernization.
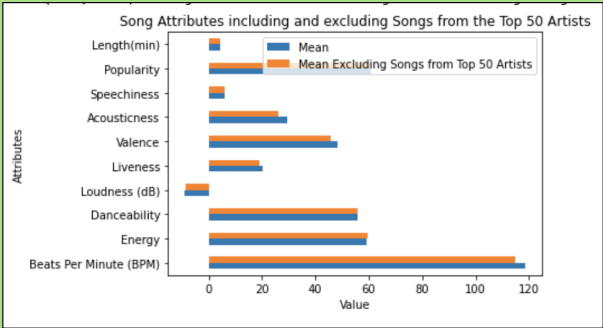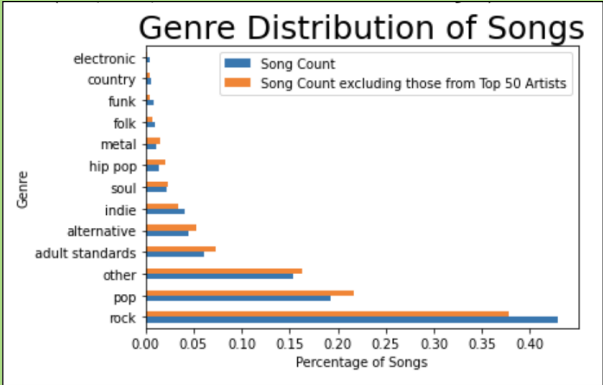
Graphs for song attribute trends over time can be viewed in the right column.

Average Loudness from 1970-2019

Average Acousticness from 1970-2019

Average Song Duration from 1970-2019

Average Valence from 1970-2019

Conclusion:

If we were to predict the popularity of a song based on its attributes without the context of time, we would be able to make an better approximation based on whether it had rock or pop as a genre. The genre would have also influenced other attributes that we could use to gauge the song's popularity such as whether the BPM was within 100-140 or whether it had a length around 4.15 minutes.

However, we've learned through our analysis that there has been an increase in genre diversification into the 21st century. The presence of genre diversification has in turn led to more variation for typical ranges in song attributions, particularly danceability, length, loudness, and energy. There's also been a shift of the most popular musical genre from rock to pop from the mid 20ths century to today. Therefore, if we were to predict the popularity of a song released in the modern day, the song would likely be more popular if it were a pop song as opposed to rock.

Although there is a possibility that we there are reccuring artists that may have skewed genre and attribute, we compared the dataset when including and excluding these artist and found minimal difference in genre distribution and attribute means for different genres. We can therefore conclude that our dataset when including songs from the top 50 artists was representative of all 2000 songs in the dataset.
Different genres have different means and typical ranges. As genres of popular songs have become more diversified, this has led to a shift in the mean of attributes. For example, we learned that danceability and loudness has increased over time while song duration has decreased over time.

Further Questions:

- How do the genre distributions and attributes change with different demographics such as gender and age?
- How representative are popular songs for trends in the music industry?
- Are there any specific artists or genres that are continuing to lead trends in song attributes?
- How have song attributes changed in the past year as a result of COVID 19?
  ◦ I am particularly interested in investigating attributes like valence to see if artists' songs are mirroring listener's feelings of loneliness and sadness

Hyperlinks:

- **Hyperlink to Jupyter:** https://colab.research.google.com/drive/1DsZ4s_TyKT_TuBdc2z3GnCYQGoZT_HXB?usp=sharing
- **Hyperlink to CSV:** https://drive.google.com/file/d/16BOmSQKc0WRAMaaHU9TBXTGERMbiz2uI/view?usp=sharing
- **Hyperlink to Github:** https://github.com/gloriiayang/DIGHUM100
- **Hyperlink to Dataset Song Attribute Definitions:** https://docs.google.com/document/d/126tfKNlIyKkwNw7I9ragkKSapQ4q9GpVwKqjPxnNJJI/edit

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.

Hinkes-Jones, Llewellyn. "The Real Reason Music's Gotten So Loud."*The Atlantic*, Atlantic Media Company, 25 Nov. 2013,
    www.theatlantic.com/entertainment/archive/2013/11/the-real-reason-musics-gotten-so-loud/281707/.

Iqbal, Mansoor. "Spotify Revenue and Usage Statistics (2021)."*Business of Apps*, Business of Apps, 2 Apr. 2021, www.businessofapps.com/data/spotify-statistics/.

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

"Shareholder-Letter-Q4-2019."*Financials*, Spotify AB, 2 Feb. 2020, investors.spotify.com/financials/default.aspx.

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

Stuart. "Pop Music Is Louder, Less Acoustic and More Energetic than in the 1950s."*The Guardian*, Guardian News and Media, 25 Nov. 2013,
    www.theguardian.com/technology/2013/nov/25/pop-music-louder-less-acoustic.