CHRISTIAN MUSIC - ANALYTICAL COMPARISON BETWEEN CTM AND CCM

DIGHUM 100: THEORY AND METHOD IN DIGITIAL HUMANITIES BY DR. ADAM ANDERSON | WILLIAM IGNATIUS WIJAYA | JUNE 27, 2021

Project Description

Christian Music has a very interesting history in it. From being rejected by different culture and ethnic groups, and now it is accepted by more people nowadays. As a person who loves to hear Christian music, I do not lean toward one type of Christian Music, whether it is traditional or contemporary, as both of the genres serve the same purpose, i.e., to praise and bring glory to the Father. In this project, I want to compare two of the most common Christian Music genres: Christian Traditional Music (CTM) and Christian Contemporary Music (CCM). I want to compare how do they deliver the song to the audience differently given that the creation of both genres have the same purpose.

Hypothesis

My hypothesis is each of the top rated songs are unique as each of the songs are trying to convey different messages with the same purpose. And, when listening to Christian Music, the emotions that people have are joy and positive. In term of the words used, CTM tends to use words that are "about" God and CCM tends to use words that are "to" God. Finally, in term of the music elements, CTM have music elements toward musical style of the past, while CCM have music elements toward pop music

Question for Exploratory Data Analysis

1. Are the top rated songs from each of the genres unique? Or are they similar to each other?
2. What kind of emotions do people have while listening to CTM and CCM?
3. How is the word choice between CTM and CCM differs?
4. How CTM and CCM differs in term of their audio components?

Audience

The audience of this project could be anyone, but this specifically address to individuals who love to listen to Christian Music. This project gives an insight to the audience about the most common genres in Christian Music and how they differs in term of their word usage, audio features, message, and emotions.

Are your questions with the methods you are proposing?

Q1. would be answerable because I can perform cosine similarity analysis and see from the cosine similarity plot whether the correlation between each songs are high or not.

Q2. might be answerable by performing emotion analysis using a phyton library NRCLex. From this analysis, I can categorize each songs into their respective emotions category.

Q3. would be answerable by using a phyton library WordCloud use WorldCloud which creates an image composed of prevalent words/bi-grams used in each genres. The size of each word/bi-gram indicates their frequency and impotance.

Q4. would be answerable by using Spotify API and spotipy in Phyton to obtain the audio features analysis by Spotify. The audio features will include terms like (danceability, energy, loudness, etc). From the "Reference of Tools and Methods" section, see the link for "Audio features" to understand the terms in depth.

Description of Dataset

For the data collection, I will be using Spotify top rated 50 songs people are listening now. I obtained such data by searching for a playlist for each genres containing top 50 songs people are listening now. I will type the title and artist of the song manually to the dataset. After that, I will use Genius API and Musixmatch API to obtain the lyrics for each song. For the audio components, I will be using Spotify API to pull out the audio features analysis by Spotify and then store it into my dataframe. To help bridge the gap between Spotify API and phyton, I will be using Spotipy, a simple library that allows the users to access the data from their account. Then, I can proceed using Pandas to transform my dataset into a dataframe. In the process of data collection, I'm referencing to David Panangian's notebook. Overall, The dataset has a total of 12 columns: genre, title, lyrics, danceability, energy, loudness, speechiness, acousticness, instrumentalness, liveness, valence, and tempo . All songs are in English.

After constructing the dataframe, I then proceed into cleaning my dataset by removing unecessary words from the lyric of the song.

Description of Tools and Methods

a) Text Scrapping, Data Cleaning, Data Manipulation

Scrapped data for the lyrics using Genius API. I use pandas to create the dataframe. Then, I am using NLTK library to do data cleaning. Data cleaning includes doing tokenization to the lyrics, then removing stopwords and digits, and do lemmatization.

b) Cosine Similarity Analysis

I will be using SKLEARN to use its cosine-similarity, Count Vectorizer, and TfidVectorizer functions. These function are used to compare the 15 top rated songs in each genres. Then, I am using SEABORN package to plot the heat map of the cosine similarity values of the top 15 songs in each genres.

c) Sentiment Analysis

I am using NRCLEX library to perform emotional sentiment analysis on my lyrics data. The library provides 10 emotions, but I will be extracting 6 emotions: joy, positive, trust, sadness, negative, and anger. Then, I will be using MATPLOTLIB library to create stacked bar graph of my emotional sentiment analysis.

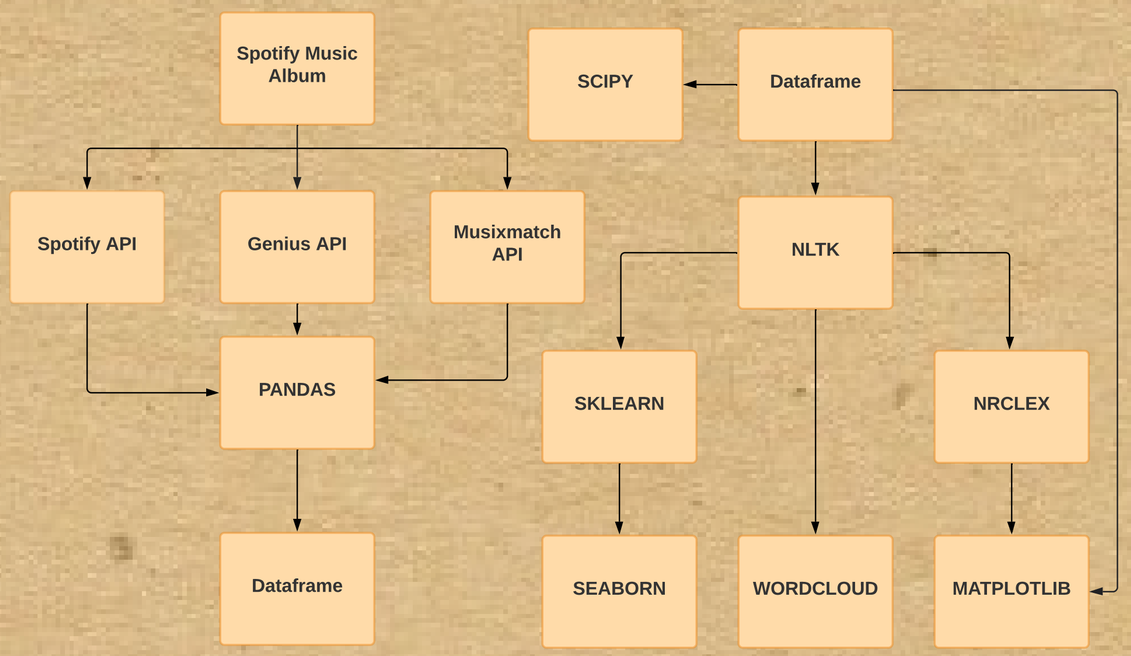
d) WordCloud

I am using Wordcloud to generate words that are more frequent in each genres.

e) Audio component analysis

I will be using MATPLOTLIB to plot superimposed histogram for each of the listed audio component. The superimposed histogram will contain two histograms describing each genres' audio component. In total, there will be 8 superimposed histograms. I am choosing superimposed histogram because it is easier to analyze and compare using this type of data visualization. Then, I will use .describe command to display the mean, median, and quartiles of the histogram. The superimposed histogram will be used to compare each audio features of CTM and CCM, and then make the conclusion whether the audio features of CTM is higher or lower compared to that of CCM. After that, I will to perform hypothesis test using unpaired z-test. This test is performed to verify whether the conclusion that I made from each of the histograms is correct or not. This test will be performed by defining a z-test function and using SCIPY package in phyton. .

Workflow

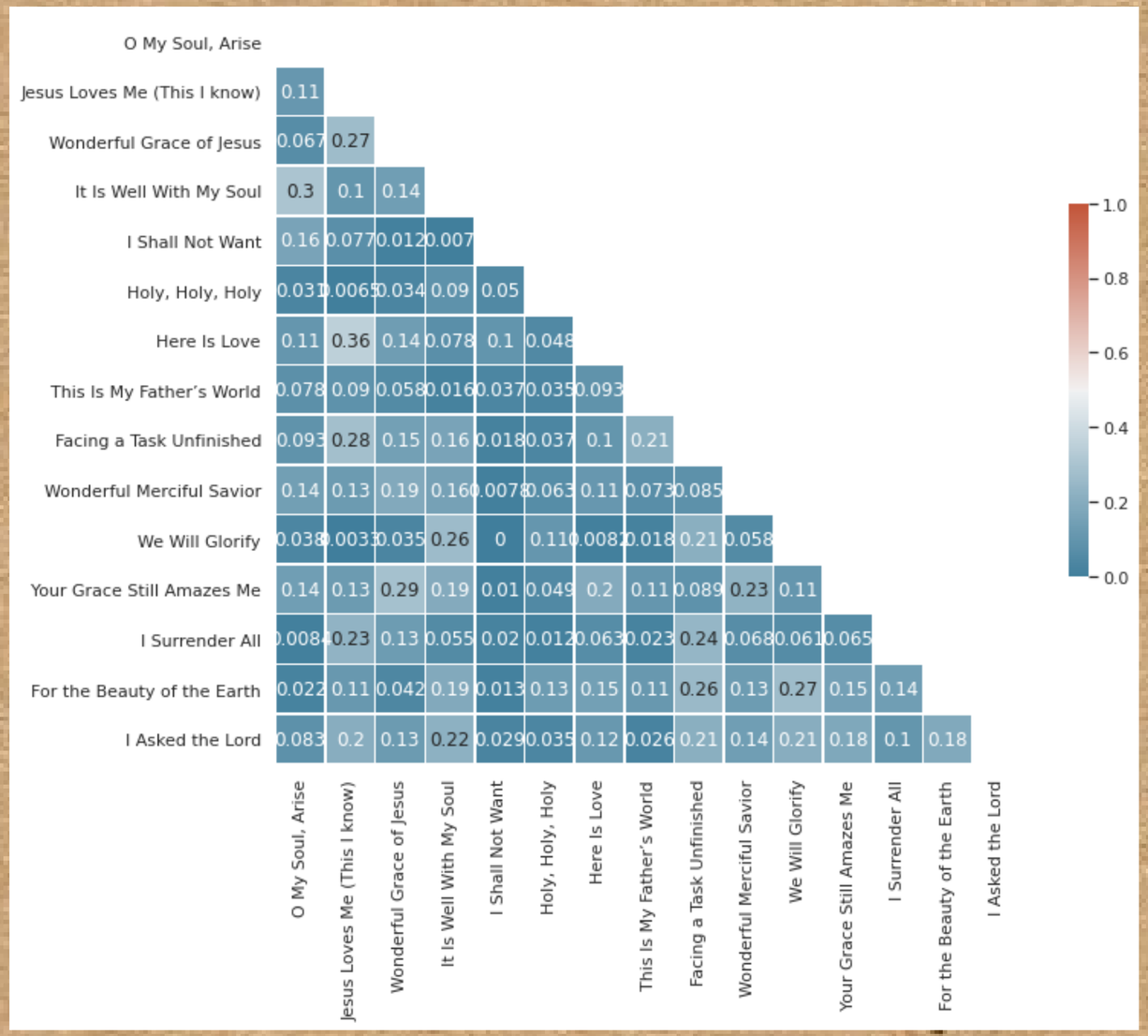


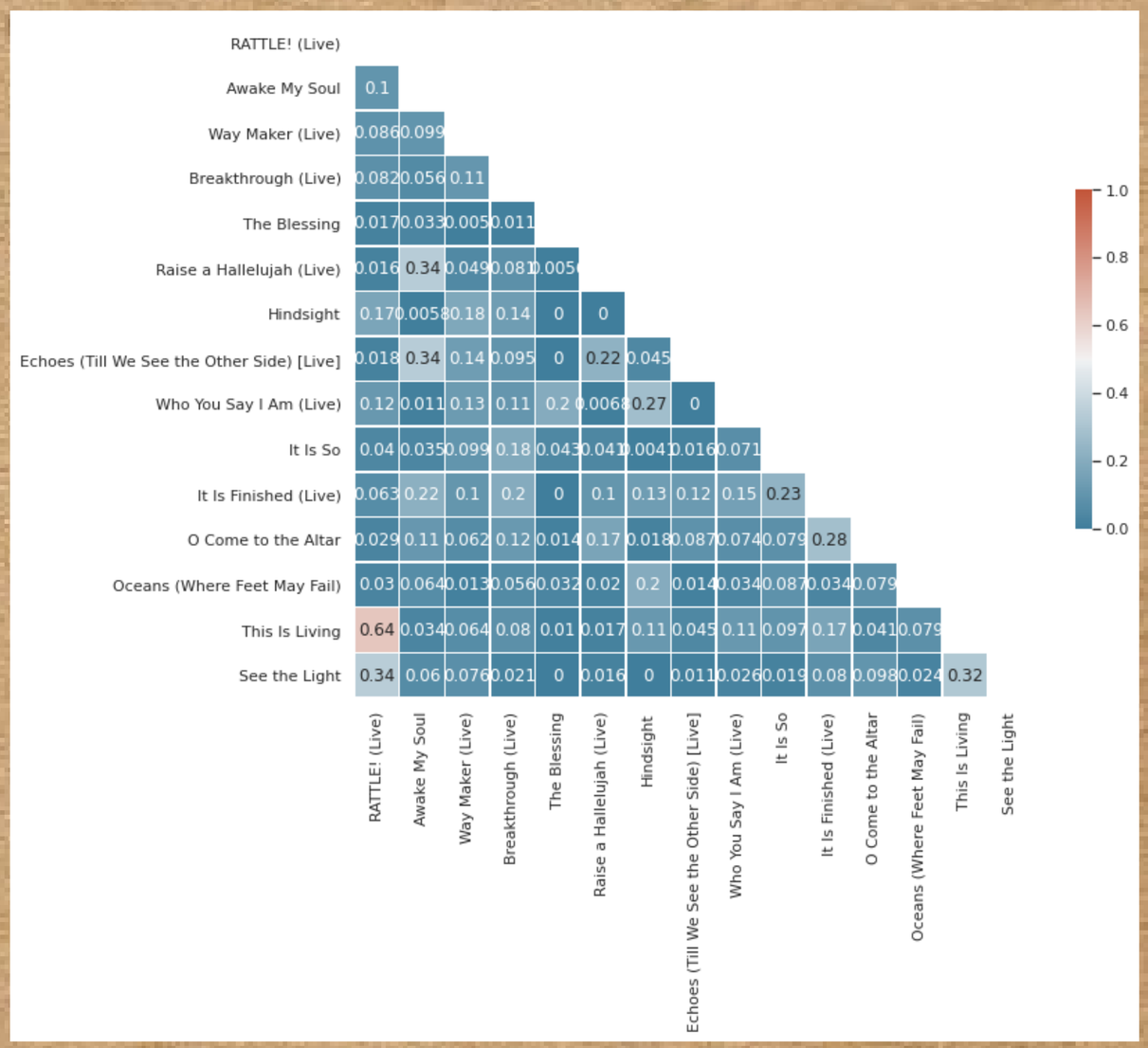
Reference of Tools and Methods

- CTM Spotify Playlist ID- 7pOcjSKD8Is86adSn6ST8C
- CCM Spotify Playlist ID- 39XGsc1HB7SVLFljc7aBpq
- Spotify Web API- https://developer.spotify.com/documentation/web-api/
- Spotipy- https://spotipy.readthedocs.io/en/2.12.0/
- Spotipy continued- https://medium.com/@maxtingle/getting-started-with-spotifys-api-spotipy-197c3dc6353b
- Lyric Genius- https://docs.genius.com/
- Musixmatch API- https://developer.musixmatch.com/documentation
- Musixmatch- https://pypi.org/project/musixmatch/
- Hudson Brendon's Github- https://github.com/hudsonbrendon/python-musixmatch
- John Miller's Github- https://github.com/MageJohn/genius-lyric-finder
- David Panangian's notebook- https://colab.research.google.com/drive/1VLrT0-OZbj998J-Z_rusYxXDsYCm6n8C#scrollTo=tJIbvpsJsmKk
- Kezia Liman's notebook- https://colab.research.google.com/drive/1hbNIVNFb4FCa46vLCq4D86N98QdjOz0R#scrollTo=Qm5qRS5DMMpd
- Sherren Kasmir's notebook-https://colab.research.google.com/drive/1Q32ufWwmFs-6kBDv4Pg2iynMeekHWA0P#scrollTo=KRDQf9_lUX4Y
- Eric Dratvas' notebook-https://colab.research.google.com/drive/1EWKNMvH10Ch6YWFTZJluyF4WxGQGdmz-
- NRCLex github - https://github.com/metalcorebear/NRCLex
- Stacked bar graph- https://stackoverflow.com/questions/43530494/a-third-stacked-bar-in-matplotlib
- Superimposed Histogram- https://datavizpyr.com/overlapping-histograms-with-matplotlib-in-python/
- Audio features- https://medium.com/@boplantinga/what-do-spotifys-audio-features-tell-us-about-this-year-s-eurovision-song-contest-66ad188e112a
- z distribution- https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.norm.html
- z test- http://courses.ieor.berkeley.edu/ieor165/lecture_notes/ieor165_lec18.pdf
- I am also taking reference from Chapter 7 of the Big Data & Social Sciences Textbook on text analysis. This particular chapter helped me in performing text analysis of the song lyrics..

- Link to notebook: https://colab.research.google.com/drive/1Ud11ow7BxNaXwDpHFlISSj_-5szpTWgW?usp=sharing
- Link to final presentation slide: https://docs.google.com/presentation/d/1CV86j4syKdfymUSHSsvmhTY6Wiod6vgi4G1QeaBFaCM/edit?usp=sharing

--RESULT AND ANALYSIS--

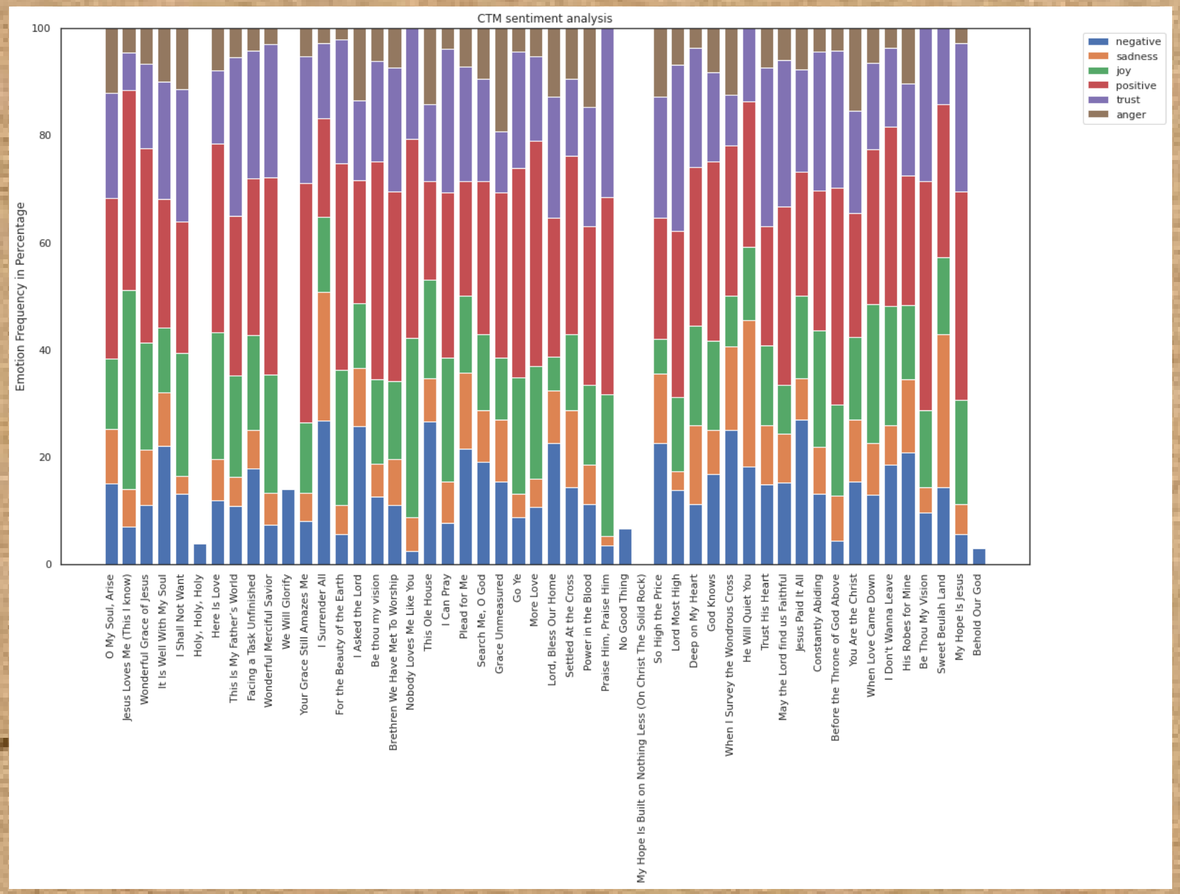


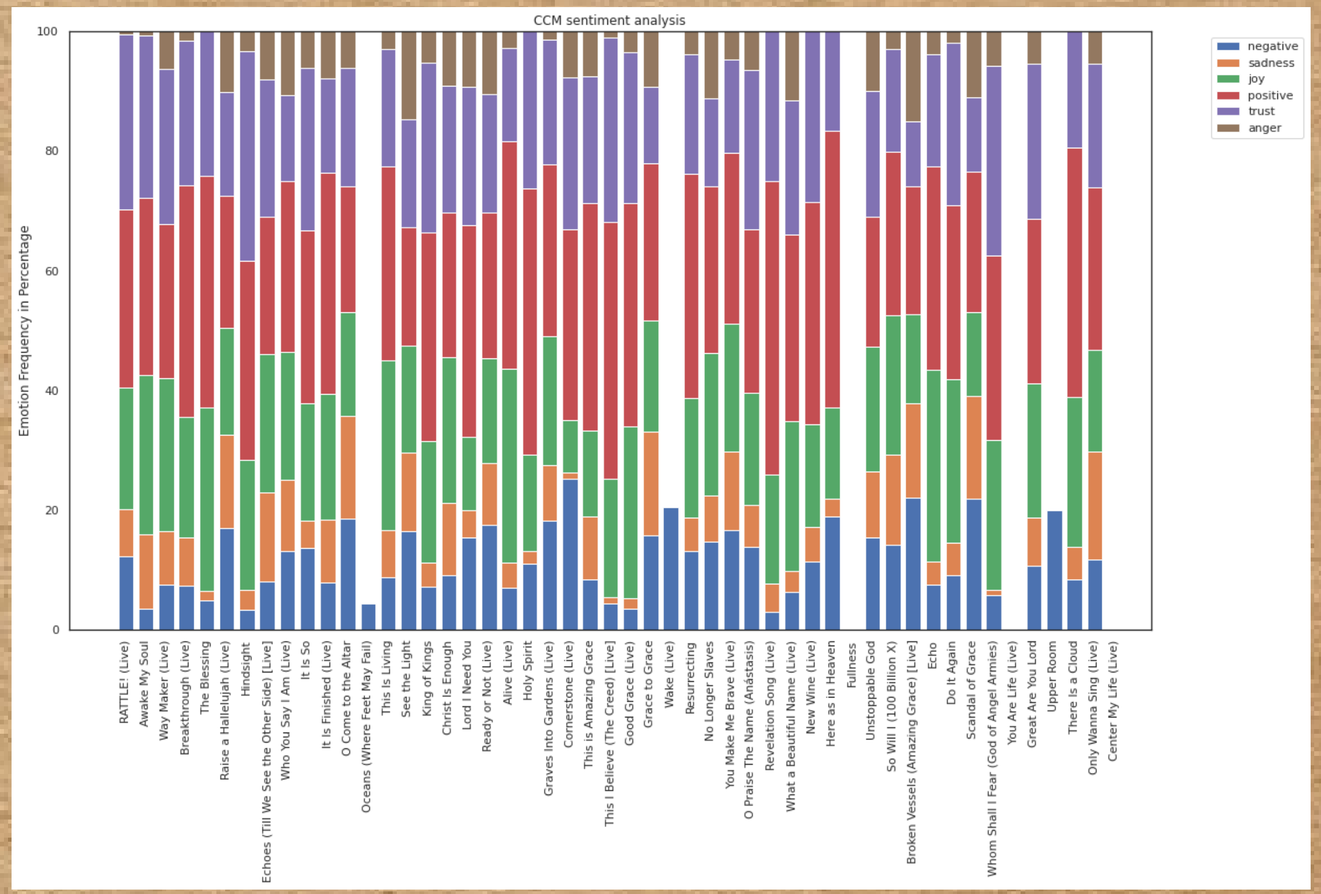


Cosine Similarity Analysis

The left part of the graph is for CTM and the right part is for CCM.The x-axis of the figures are top 15 songs from each genres going from left to right. The y-axis of the figures are top 15 songs from each genres but going from up to bottom. As I hypothesized there will be no or little correlation within each top rated songs because each songs deliver different messages given that purpose of the each songs are same. There is only one correlation from CCM that is quite high, that is between "This is Living" and "RATTLE!"; however, this alone still suggest that top rated songs are unique according to the cosine similarity analysis.

--RESULT AND ANALYSIS--





Sentiment Analysis

For the sentiment analysis, I want to compare the emotions as the following: positive with negative, joy with sadness, and trust with anger. From both of the plots above, we can see that positive emotions exceeds negative emotions, joy exceeds sadness, and trust exceeds anger. This suggests that people experience more toward positive, joy, and trust when listening to both CTM and CCM music, which is in accordance to my hypothesis.

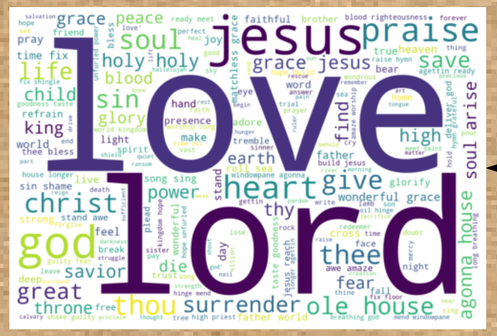
--RESULT AND ANALYSIS--

Wordcloud Analysis

We can see from the WorldCloud that CTM songs tends to use words from the past. We can frequent use of words from the past, such as, "thou", "lord", "thee", "holy", etc. And, CTM songs tends to use words that described the attribute and activities God: "thou", "great", "thee", "holy", and so on.

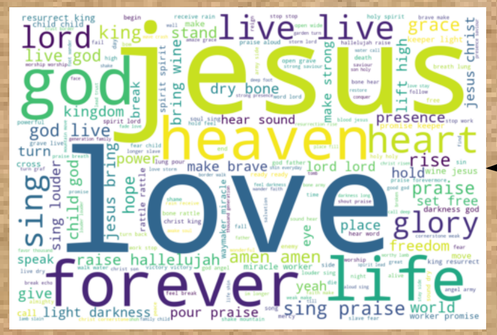
We can see from the WorldCloud that CCM songs tends to use words from the present. We can see that there is less use of words from the past and more frequent use of words that people use today, such as, "forever", "life", "live", "glory", "jesus" instead of "lord", etc. CCM focus less on the word of the song and more on the sense of union in the hearing and singing together. This can be seen by the words like: "glory", "sing", "jesus" (more frequent compared to CTM), "hallelujah", and so on.

Also, both CTM and CCM emphasize more on "love", and "God" or "Jesus". This means that most of the songs from both of these genres delivers a message: to love Jesus or love God.



CTM

	danceability	energy	loudness	speechiness	acousticness	instrumentalness	liveness	valence	tempo	duration_ms
count	50.000000	50.000000	50.000000	50.000000	50.000000	50.000000	50.000000	50.000000	50.000000	50.000000
mean	0.363540	0.274214	-10.856200	0.035668	0.835260	0.031358	0.163876	0.293384	107.820400	220977.400000
std	0.141915	0.115110	3.405425	0.011865	0.176403	0.145521	0.135741	0.137152	28.933259	55292.813693
min	0.102000	0.011300	-25.914000	0.024700	0.121000	0.000000	0.075900	0.072200	69.099000	111865.000000
25%	0.294250	0.200250	-12.425000	0.030125	0.786500	0.000000	0.097800	0.192250	82.769750	181119.750000
50%	0.321000	0.267500	-10.365500	0.032350	0.895500	0.000000	0.132000	0.272000	99.073000	222000.000000
75%	0.440000	0.327750	-9.228250	0.034775	0.934750	0.000012	0.163000	0.356750	127.841500	254863.500000
max	0.732000	0.587000	-4.367000	0.095200	0.993000	0.948000	0.866000	0.666000	188.698000	339613.000000



CCM

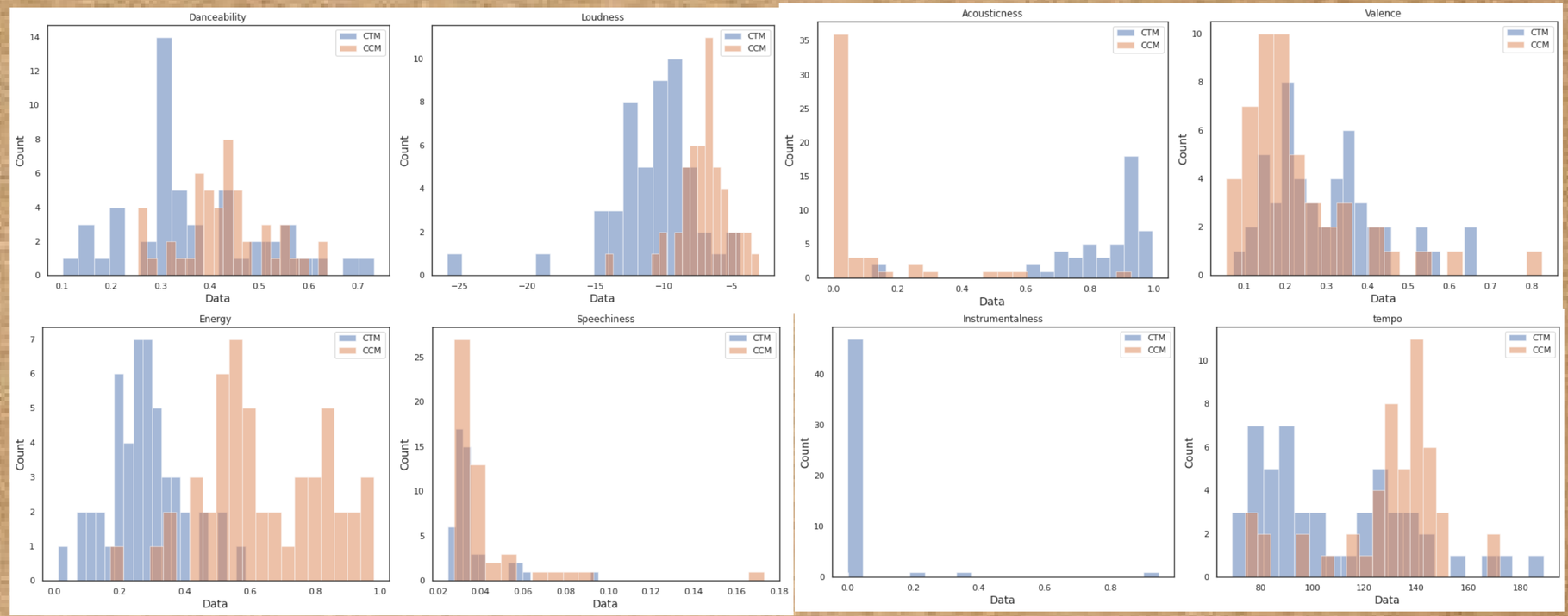
	danceability	energy	loudness	speechiness	acousticness	instrumentalness	liveness	valence	tempo	duration_ms
count	50.000000	50.000000	50.000000	50.000000	50.000000	50.000000	50.00000	50.000000	50.000000	50.000000
mean	0.428520	0.640980	-7.007480	0.041288	0.088166	0.000106	0.42586	0.228360	129.246260	349943.800000
std	0.092032	0.187391	1.899354	0.023160	0.182540	0.000473	0.31353	0.146255	22.358305	100707.909401
min	0.254000	0.172000	-14.300000	0.027800	0.000192	0.000000	0.07140	0.056000	73.998000	195480.000000
25%	0.381000	0.521750	-7.815250	0.031325	0.004215	0.000000	0.10950	0.138250	126.222000	265786.500000
50%	0.429500	0.598500	-6.795000	0.033500	0.011000	0.000000	0.33250	0.184500	135.898000	344486.500000
75%	0.465000	0.813250	-6.066500	0.039175	0.061500	0.000004	0.70825	0.277500	142.049500	407300.250000
max	0.638000	0.979000	-3.037000	0.173000	0.928000	0.002680	0.96400	0.825000	171.827000	568787.000000

Audio Component Analysis

From the histogram, statistic table, and z-test, these are the conclusions I made:

- Danceability: CTM is lower than CCM
- Energy: CTM is lower than CCM
- Loudness: CTM is lower than CCM
- Speechiness: CTM is lower than CCM
- Acousticness: CTM is higher than CCM
- Instrumentalness: CTM is higher than CCM
- Liveness: CTM is lower than CCM
- Valence: CTM is higher than CCM
- Tempo: CTM is lower than CCM

We can see that CTM songs tends to use more acoustic instrument and have audio features more toward the musical elements of the past. On the other hand, CCM songs tends to be more energetic, loud, more beat, and more words in it, which matches more toward pop music



--FINAL EVALUATION OF PROJECT--

Overall, all of my questions are answerable using the tools and visualization that I described previously.

What are some obstacles that I face?

There is an obstacle that I face when using these tools to answer my questions. The obstacle I face is when I am collecting the data. To pull out the lyrics from either Genius API or Musixmatch API, I have to type the title of the song and artist of the song manually. And, I have to type the track id manually when using the Spotify API to pull out the audio features. This process takes huge amount of time and concentration. Other than the data collection process, all other processes are fine.

What I will continue for future work?

I would really love to continue working on this project. First and foremost, I would like to expand my dataset. 50 songs for each genres are maybe too few. I am considering to have 100 songs for each genres, but the data collection takes too much time. Perhaps, I will find another ways to do the data collection without doing much manual works. Other than that, I would like to explore more subgenres. CTM and CCM have each subgenres, but I am unable to explore those due to time constraint. If I am given more time, I would like to explore and compare those subgenres. Finaly, it would be more interesting if I do a cross genre-analysis on the cosine similarity analysis rather than within genre; however, it seems that cosine similarity analysis does not really work in this case. Hence, I would like explore the method to do a cross genre-analysis using the method other than cosine similarity analysis.

Ending

I would like to thank Dr. Adam Anderson for teaching us this course and all his efforts in grading our assignment. I would also thank the Ash Tan for providing us the tools and resources related to the technical parts of the coding. Finally, I would like to also thank to my classmates who helped me into shaping my project and providing me feedbacks during the planning phase of the project..

REFERENCES

1. Anderson, Elder Karl. “What Is the Difference between Traditional and Contemporary Gospel?”*Gainesville Sun*, 2 Feb. 2006, eu.gainesville.com/article/LK/20060202/news/604165237/GS.

2. Ekspears. “Three Differences Between Traditional and Contemporary Worship.”*Healthychurchnetwork*, 25 July 2012, healthychurchnetwork.wordpress.com/2012/07/25/three-differences-between-traditional-and-contemporary-worship.

3. Halim, Michael Adrian. “Music During Hard Times.”*bcourses*, 2 July 2020, https://bcourses.berkeley.edu/courses/1505337/files/folder/2020_Student_Work/Posters/FinalVersions?preview=80594716.

4. Panangian, David Julio. "Lyrical Transformation of Kanye West."*bcourses*, 5 July 2020, https://bcourses.berkeley.edu/courses/1505337/files/folder/2020_Student_Work/Posters/FinalVersions?preview=80594730

5. Wikipedia contributors. "Christian Music." Wikipedia, 28 May 2021. en.wikipedia.org/wiki/Christian_music.

6. “What Is the Difference Between Traditional and Contemporary Worship?”*Protestant Worship Diversity in Naperville*, www.worshipdiversityinnaperville.hopedla.org/contemporary-vs-traditional-worship. Accessed 16 June 2021.