IS310 Final Submission Paper
Group Members: Jeffrey Fang, Wei-Ting Yang, Shoi Rathi, Charles Havyarimana, Yuktha Sureshkumar

Final Project: An Exploration of Billboard

The Billboard 100 is a music chart published by Billboard Magazine that seeks to rank the most popular songs in the United States across all genres. It serves as a barometer of mainstream music success by combining data from radio airplay, streaming figures, and sales (both physical and digital/streaming). In 2024, streaming services such as Spotify and Apple Music have become the predominant form of music consumption across the board. These services allow for on-demand access to music, usually through a monthly fee that users pay for the service. Together these two forces dominate the music industry in the present day, relying on each other to catalogue and shape the music industry. The focus of this project is to understand the metrics and function of Billboard and Spotify as a tool for popularity and access, but also their position as the primary arbiter of musical popularity. How have these tools shaped/affected music through shifting economies and changing times, and how can we challenge the social and economic hegemony that they have established. To explore this question we have decided to use a dataset that catalogues the Billboard Yearly Top 100 dataset through Spotify song entries.

Initially, we decided to approach this by searching for Billboard related datasets on Kaggle. In this process we found several that satisfied the requirements of having Spotify songs and metadata whilst being Billboard lists. In the first stages of our project we were split between using a dataset that had the Top 10000 songs from 1960-2023. We conducted basic data exploration using pandas to search for patterns and biases that might be obviously present within the dataset. As we were doing these data explorations we decided that we wanted to understand the relationship between economic conditions and how these might be reflected in the Billboard dataset, as an obvious strength of the dataset was the spans of time it covered. Furthermore we realized that this list of 10000 songs were for the weekly Billboard charts, and by using this we would have to be able to process an extremely dense amount of data. Thus for the next phase of this project we decided to switch to the 1946-2022 dataset, which although has a larger historical range, instead uses the Billboard Year End dataset, which means only 100 entries per year, making this more manageable and relevant. Furthermore, this dataset has considerably less missing data, especially in the Genre's section which was a large part of the missing data in the previous dataset.

In the next stage of the project, we seeked to update and augment the 1946-2022 dataset. **One of the first things we decided to do was adjust the number of features available. We saw that there were many features that weren't necessary for our analysis. We also went through and dropped rows that were missing data.** Our music dataset (music_sample.csv) includes essential music features, popularity, danceability, energy, valence, tempo, and loudness, which are critical for analyzing music trends and listener preferences. The main challenge in this initial phase is that the observations were too limited to continue music trend analysis as a small dataset can lead to biased insight and conclusion. To fix this issue, we found another dataset, a

curated collection of world music, from Kaggle as our reference to expand the dataset. This dataset is a curated collection of world music with 326 tracks from 66 diverse artists from 1958 to 2019. Each track is meticulously tagged with Spotify audio features, providing insights into tempo, key and energy. By organizing this dataset with our initial dataset, it captures more key attributes of popular music and provides a foundation for exploring trends, patterns, and relationships within a historical and societal context. The time frame of 1958 - 2019 encompasses critical moments in history such as the post-war era and economic recession/expansion. This allowed us to not only increase more observations but also improve complementary insights into musical attributes. In this phase, the challenge we encountered is integration. When merging music_sample_data and world_hits, we need to ensure the retention of shared columns and alignment of data types. To achieve this, we applied pd.concat() to combine the datasets based on shared columns. However, the process was far from straightforward. Some columns had different names - for instance, Track Name versus Track - which required renaming to ensure uniformity and seamless merging.

For the Mid-Semester phase of our project, we decided at the time that we wanted to further pursue our integration of historical data with the Spotify dataset. Thus we both attempted to clean the dataset of unnecessary data and also create a new dataset that was primarily oriented around historical events and recessions. The idea for this second dataset was to be able to cross reference times and dates of historical events to possible changes in the patterns of music listening habits. For the historical events dataset, web scraping played a crucial role in gathering accurate and relevant data. We found a website that had data of all crucial historical events that took place in America through the decades. Then using Python's BeautifulSoup library, we extracted timelines and descriptions of major historical American events from the 1960s to now. After setting up custom scraping scripts, we parsed through HTML structures to isolate event dates and key descriptions of the events. The goal was to build a strong dataset that could help us see possible connections between historical events and changes in music listening habits. After these historical events were scraped into our dataset, We decided to continue to add on to this dataset by manually compiling a list of major recessions by date and cause and include them into this dataset. The idea here stems from the notion of recession pop, or that during times of economic downturn, hedonistic and upbeat pop music becomes popular as a means of escape and leisure for people who may be struggling. Our intentions weren't to seek trends that correlated with this absolutely, but rather to see if change of any kind was perceivable during these periods of recession. The standards, sources and rationale are compiled in this document.

After submitting this for feedback to professor Leblanc, we recognized that the direction of our project was still too broad, and that a lot of the historical dataset that we had created was not obviously relevant to the explorations we were trying to conduct. Trying to conduct effective analysis for every recession and its related changes in the dataset was also beyond the scope of this project, thus we decided to continue to narrow the scope. An important step in us effectively narrowing the scope was thinking and arriving at computational methodologies we wanted to pursue. These methods included time series modelling, K-means clustering, data visualization

with seaborn and matplotlib, sentiment analysis, Spotify Algorithm simulation, and topic modeling. **Initially, we used a time series model to do some exploratory analysis on the data. We looked at different pairings of features over time and found some insights to help guide our next steps. We saw trends and variability showing us that the model was less confident pre-1983. This is likely due to missing metadata. This suggests focusing on post-1980s data and performing additional cleaning. Using time series helped us understand the data structure and guide adjustments before proceeding.** Exploratory Data Analysis (EDA) and k-Means clustering was conducted to allow scholars to see that pop music likely served as a source of entertainment and escapism during the economic downturn. Economic downturns may create opportunities for specific music genres or characteristics to be created. Scholars can base on this trend to forecast music trends during future economic shifts or global events. The music features such as tempo, loudness, or valence are essential in the dataset for music trend analysis over significant historical cultural periods. Topic modeling was then used to attempt to observe any trends in lyric content within the different groups of years surrounding the great recession. The data columns that were important to this computational method are primarily the Lyrics column and the Hot100 Ranking Year column.

The first step that we took to verify our data was to manually consult the Hot 100 Year End chart to ensure the integrity of the data. What we noticed was that the song titles were in line with the lists of Top 100 songs, however there were songs scattered throughout the dataset that were the Karaoke versions of the songs in the actual Top 100. This indicates that this dataset was most likely created with some sort of scraping or retrieval technique, as manual data entry would not have warranted these effects. This likely affected a few of our computational methods, specifically the topic modeling, as lyrics are not retrievable if either the song name and artist names are incorrect or not available. Luckily the Karaoke versions of songs weren't anywhere close to a majority of the dataset, however this is not ideal. More work should have been done to verify the integrity of the data earlier, to ensure that the cultural objects being represented through data are accurate. Another way that we verified the integrity of the dataset was using the requests library to ping the links of the songs under the "Spotify Link" column. The integrity of the Spotify links were good, and the few that timed out where checked manually, and all of those links directed to a song that loaded and played as expected.

Ultimately, our final dataset consists of the original Billboard 1946-2022 data, but filtered for post-1960's data. This is because when verifying the integrity of the dataset, many years before 1960 had less than 100 songs, indicating some sort of missing data. Although we didn't use all of the columns included in the dataset for our computational methods, we chose to leave them in the dataset as they could easily offer some avenue of statistical and critical significance for others using the dataset. The lyrics column was added in the dataset, and the lyrics were retrieved from the lyrics.ovh API. These song lyrics add a lot of detailed and specific information about the music represented in the billboard dataset, and flesh out the entries of the dataset from being simply abstract representations of culture to entries containing direct references and information from the culture that it is attempting to represent. Unfortunately due

to the limitations of the lyrics.ovh dataset, about ⅓ of the dataset has the lyrics column as "Lyrics not found". Lyrics.ovh was the only API that we were able to find that could handle the amount of requests that it took to retrieve the entirety of the dataset, however many of the older songs did not have lyrics available in the API database. We decided to leave the songs without lyrics anyways so as to not break the integrity of the Hot 100 structure.

To further explore the emotional content of the songs, lyric sentiment analysis was conducted using the VADER sentiment analysis tool from the NLTK library. This method computes a compound sentiment score for each song's lyrics, ranging from -1 (most negative) to +1 (most positive). By analyzing the sentiment of song lyrics across decades, we were able to identify shifts in emotional tone over time, which reflect broader cultural and societal trends. For example, periods of economic downturn often coincided with songs that had more somber or reflective sentiment scores, while periods of economic stability and growth showed an increase in more positive, upbeat scores. This sentiment analysis provides another layer of insight into how music serves as both a mirror and a means of emotional expression during significant historical and cultural moments.

Labor Division

Jeff:

The work that I did included conducting topic modelling using BERTTOPIC. I did this by stratifying the years close to the great recession (2007-2009) into a period of 2000-2004 being well before the recession. The years 2005-2006 being directly before the recession, 2007-2009 being during the recession and 2010-2013 being after the recession. I then conducted topic modelling on all the stratified groups and created visualizations for all of these years. During this process I used an API to retrieve lyrics and augment the final dataset, as well as creating the documentation for the dataset. I also verified the integrity of the dataset, and cleaned the dataset for the final submission. While working on the experimentation update, I processed the lyric data by using Matplotlib word clouds, which stood in as a similar function to the topic modelling before I had figured out how to do that. During the mid semester update, I compiled a list of recession dates by cross referencing sources and adding them to a dataset that we had made. For the proprietary dataset creation aspect I reviewed and completed the documenting the data portion of the write up.

Weiting:

My work included data merging, exploratory analysis, and how our dataset can be applied in future research. I first integrated music_sample.csv and world_hits datasets, which allowed us to gather data that spans important cultural and historic time periods. This combined dataset became crucial for our analysis for music trends and how they relate to the states of society. For Exploratory Data Analysis (EDA), I used visualizations to find the underlying patterns in the merged dataset. Using graphs such as scatter plots, histograms, and line graphs, I identified correlations between musical attributes (e.g., tempo, valence, and energy) and their popularity

during declines in economy. This showed that music at the time served as a form of escapism, especially during depression, like the 2008 economic crisis. This finding aligns with existing studies, like Pettijohn et al. (2012), which suggest that musical characteristics change with socio-economic conditions. As for k-means, I used k-means clustering to group songs with similar features, which allowed me to find music trends relating to different time periods. Clusters with high valence and tempo often correlate to upbeat, danceable music that were popular during stable economic periods, while clusters with lower energy were more popular during times of societal difficulty, showing a preference for reflective and mellow music during such periods. To extend the dataset's utility for further research, I also mentioned the database's potential for predictive modeling and cross-genre comparison, where future scholars can analyze the evolution of attributes like tempo and energy across genres and explore how listener preferences have shifted over decades. Lastly, I documented the guidelines for dataset usage, ensuring that future researchers can effectively utilize our work. These guidelines include detailed descriptions of the dataset's attributes, potential analytical techniques, and recommendations for avoiding misinterpretation due to the dataset's inherent biases (e.g., its focus on popular songs).

Shoi:
The work I did included creating the historical dataset and also creating a simulation of the spotify algorithm to gain insights on what goes into measuring the factors that seem subjective. For creating the historical dataset I used a website with information for significant American historical events by decades, and scraped each of the decades' information then combined the dataset into one. Then I helped in documenting the dataset. Jeff and I worked on combining the dataset and editing it so that it fit our context in the way that fit best for our project. For example we handpicked some data points that were not adding relevant information and we removed them. The other aspect of the majority of my work consisted of creating the algorithm which simulated the algorithm of spotify and helped us gain some validity and explanation on the information that our dataset had. We were able to understand the processes that go into extracting factors from music such as energy, danceability, etc. This helped us explore the dataset further and understand the computational methods it uses.

Charles:
For the mid semester update, the work I did was the initial data cleaning. This included column adjustment, and removal of missing data rows. Furthermore I served as the project manager for the last weeks of the project. I delegated work, set goals for the project and adjusted deadlines to work with our team. I ran a time series model for our initial computational analysis as well as working with everyone on the final dataset essay.

Yuktha: My work focused on applying sentiment analysis to explore the emotional tone of song titles over time and its connection to cultural and societal trends. I utilized the VADER sentiment

analysis tool from the NLTK library to compute compound sentiment scores for each title, generating a range from -1 (most negative) to +1 (most positive). These scores were added as a new column, 'Sentiment,' to the dataset, enabling further analysis of emotional trends across different time periods. I also explored practical applications for this analysis, such as its use in psychological research to understand how music reflects or influences collective moods during significant cultural moments. Additionally, I highlighted the potential for marketing research, where sentiment trends could inform strategies to align music and advertising with public sentiment. This analysis not only revealed emotional patterns in song titles but also provided insights into the broader interplay between music, societal shifts, and consumer behavior.

Dataset with Documentation
See Folder: Final_Submission/Dataset

Guidelines for Data Use
  In general, with our dataset applied and experimentation done above, scholars can use our dataset and experiments to see how popularity correlates with important music features like tempo, valence, and energy. Relevant studies include examining how musical attributes and listening habits have changed over historical time. In addition, the dataset can be used to explore what genres rise and fall along with popularity during broader societal and cultural dynamics with different data analysis and machine learning techniques.
  According to "And the Beat Goes On: Popular Billboard Song Beats Per Minute and Key Signatures Vary with Social and Economic Conditions", it talks about how characteristics of Billboard songs, such as Tempo, beats per minute (BPMs) and key signature, changed with the socio-economic conditions of the United States between 1955 and 2008. The research showed that during times of socio-economic difficulty, slower tempos are more prevalent in songs, which reflects a mood of introspection. Similarly, during periods of economic and social stability, faster tempos and standard key signatures dominate, signaling positivity and energy (Pettijohn et al., 2012). In this article, the authors did not explicitly show any code from their studies, however, they did use statistics strategies such as correlation analysis and time series modeling to strengthen their argument. These methods were applied to analyze the relationships between musical features (e.g., BPM and key signatures) and socioeconomic indicators (e.g., General Hard Times Measure).
  Based on the summary of the journal article above, we conducted dataset exploration with Exploratory Data Analysis (EDA) and k-Means clustering with important music features. Doing these two experiments may allow scholars and our group to testify if pop music with higher tempo likely served as a source of entertainment and escapism during the economic downturn as we assumed that during the economic recession (2007- 2009), people are more likely to seek higher tempo or energetic songs. In our EDA, we observed a clear increase in the number of pop songs during the recession. This could mean that pop music became more

prominent or was produced more frequently during this challenging economic time. Pop music's upbeat energy may have provided comfort for the public in between 2007 to 2009. Then, we continued for deeper analysis with k-Means clustering. We identified two clusters: songs with slightly lower popularity (average 86), moderate energy (80), loudness of around -4 decibels, and a tempo averaging 125 beats per minute and songs with higher popularity (average 95), higher energy levels (90), louder volumes at -2 decibels, and a faster tempo of 140 bpm. There is a slight difference between the clusters, but they suggest that more energetic, louder, and faster songs were generally more popular during economically challenging times. This aligns with the idea that upbeat and dynamic music resonates well with listeners when people may seek uplifting or engaging tunes when they face financial difficulties.

A short observation we got here is that economic downturns may create opportunities for specific music genres or characteristics to be created. Scholars can base on this trend to forecast music trends during future economic shifts or global events. The recession likely pushed people to seek affordable forms of entertainment, and music, especially pop songs. Overall, these two experiments can be a good start point and inspiration for researchers specializing in music trends and cultural research. One potential extension of this analysis is cross-genre comparison. Researchers could examine trends across genres like rock, hip-hop, or electronic music. This would involve analyzing how attributes such as tempo, energy, and popularity differ across genres and how these characteristics have evolved within each genre over time. Such insights could reveal the shifting preferences of listeners and the cultural significance of various genres during different periods.

Additionally, the sentiment analysis we conducted on the song titles could offer many opportunities for research. There are various ways for scholars to utilize this data to explore the interplay between music, culture, and societal trends. Sentiment Analysis is a key application of Natural Language Processing, and has become a very important tool for understanding the emotional tone of text data. In this context, we applied these techniques to explore the emotional content of song titles over time, which offers perspectives on how the music industry has evolved in terms of emotional expression and marketing strategies. Our methodology centered on utilizing the VADER (Valence Aware Dictionary and sEntiment Reasoner) sentiment analysis tool, which is part of the Natural Language Toolkit (NLTK) library. Our function computes a compound sentiment score for each song title, ranging from -1 (most negative) to +1 (most positive). The resulting sentiment scores were then appended as a new column, 'Sentiment', to our dataframe.

One way lyric sentiment analysis could be used in real-life research is in psychological studies. Psychologists could investigate how the emotional content of popular music titles reflects or influences collective mood and mental health trends. "Music Use for Mood Regulation: Self-Awareness and Conscious Listening Choices in Young People With Tendencies to Depression" explores how young people with depression use music to manage their moods and their awareness of its impact on their mental health. The findings from this study on music listening habits of young people with depression can be directly connected to our lyric sentiment

analysis project. The study highlights how individuals use music to regulate their moods, often with varying levels of awareness about its impact. This aligns with our analysis of song title sentiments over time, as it provides context for the emotional trends we've observed. For instance, the study's observation that participants often chose music mirroring their current mood to cope with negative feelings could explain periods of predominantly negative sentiment in song titles. Similarly, the finding that some participants used music with different moods to shift their emotional state might correspond to fluctuations in title sentiment over time. The study's emphasis on the importance of lyrical content in influencing mood further validates our focus on analyzing song titles, as they often encapsulate the song's emotional theme. Moreover, the temporal aspect of our sentiment analysis could potentially reflect the changing awareness and music listening strategies described in the study, offering insights into broader cultural shifts in emotional expression through music. By incorporating these insights, our sentiment analysis can provide a more nuanced interpretation of the trends we've observed, considering how they might reflect not just cultural moods but also individual listening habits and their potential impact on mental health.

Additionally, marketing researchers can indeed leverage sentiment analysis of song titles to gain valuable insights into consumer behavior and advertising effectiveness. This approach builds upon Nielsen's 2015 study on the emotive power of music in advertising, extending it to explore how the emotional content of song titles correlates with commercial success. Nielsen's research demonstrated that advertisements with music performed better across key metrics like creativity, empathy, and emotive power compared to those without music. The sentiment analysis of song titles offers valuable insights for marketers to enhance their strategies and connect with consumers more effectively. By analyzing the emotional tone of popular song titles, marketers can gauge broader public moods and predict consumer responses to specific marketing messages. This understanding allows for the creation of more impactful advertisements by selecting music that resonates with the target audience's emotional state.

An important part of our Spotify dataset is the subjective categories such as tempo, danceability, energy, loudness, and instrumentalness. We wanted to understand how Spotify decides those categories and makes those classifications. The reason we wanted to figure that out is so that we understand the context around it and are able to gain some algorithmic explainability behind the processes that led to making our data. To figure this out, I created a code that would simulate the Spotify algorithm. There are a few aspects that go into the algorithm. The main aspect includes audio aspect extraction which is done in my code through Librosa. This is a python library with which audio aspects of an Mp3 audio file can be extracted. With those extracted features I was able to either directly use a function to get one of the factors, or I had to create an equation which put multiple factors together to get the final factor. For example, one of the first features we analyzed was **tempo**, which measures the speed of a song in beats per minute (BPM). Using Librosa's beat.beat_track function we estimated the tempo by detecting beat positions within the audio signal. Another key feature was **danceability**, which reflects how suitable a song is for dancing. Calculating danceability involved synthesizing

several attributes such as tempo, beat strength, spectral contrast, and spectral flatness. Tempo provides a baseline for rhythm, while beat strength measures the prominence of beats. Spectral contrast captures the difference in amplitude between peaks and valleys in the frequency spectrum which is often associated with dynamic and rhythmic music. Spectral flatness quantifies how noise-like a sound is. Lower flatness generally corresponds to more danceable tracks. By combining and normalizing these features on a scale of 0 to 1 we were able to derive a danceability score. We also calculated **loudness**, which represents the perceived volume of the audio. We used a perceptual weighting function that simulates human sensitivity to different frequencies. First, the audio signal was converted to the frequency domain using the Short-Time Fourier Transform (STFT). This was followed by computing a power spectrogram to represent the energy of various frequencies. Finally, perceptual weighting was applied which gave us a loudness value. Determining the **key** of a song required the use of the chroma feature which gave the simple output of the key directly. To measure **instrumentalness**, we used harmonic-percussive source separation (HPSS), which separates the audio signal into harmonic (melodic) and percussive (rhythmic) components. The energy of each component was calculated and the ratio of harmonic energy to total energy was used to get the instrumentalness. The last feature we examined was **energy** which is a measure of a song's intensity and dynamics. Using the Root Mean Square (RMS) function we calculated the magnitude of variations in the audio signal over time to get the energy. Fundamentally, these formulas are similar to the factors that spotify considers. The significant difference is that spotify uses a more machine learning approach to take on a large number of songs and have more accurate results. Overall, this code helped us understand how aspects of our datasets were made. This is useful for others who want to use our dataset because it adds validity and trust.

Limitations and Considerations:

One of the glaring weaknesses of our dataset is simply the size of it. A general rule for most machine learning applications is to have 10 times the number of features. This means if our dataset has fewer rows (observations) than 10 times the number of features (columns), it poses a risk of overfitting when building machine learning models. Models trained on such datasets may memorize rather than generalize, leading to poor performance on unseen data. For our project, we intentionally narrowed our scope to make the project more manageable. However, with more experience with data science and/or time, working on a larger dataset would be better.

Bias Toward Popular Songs:

The dataset includes only popular songs, this limits its generalizability. Our analyses and models based on this data won't fully be representative of the entire music landscape. This also affects the diversity of the analysis and model as it can omit culturally or regionally significant music that does not achieve chart status. There are songs that didn't gain commercial success until much later that wouldn't be represented in the original year it was released.

Another thing to consider is Billboard's role within the dataset. As an authoritative and prominent figure in the music industry, they decide what charts are based on their chosen metrics. While these metrics provide a quantitative basis, certain subjective elements and methodological choices can influence the rankings, potentially affecting analyses that rely on this data. As times have changed, Billboard has adjusted its weightings for the metrics it uses to chart a song. Only recently has Billboard allowed album cuts to chart. Album cuts are songs that weren't released as commercial singles and therefore were not eligible to be ranked on the Hot 100 until 1998. This means before 1998, songs that were immensely popular but not released as commercial singles were excluded from the Hot 100, potentially skewing historical analyses and underrepresenting artists whose work resonated widely without traditional single releases. Another instance of this is in 2018 when Billboard adjusted its methodology to give greater weight to paid subscription streams over free streams. Such changes can impact chart positions, especially for genres or artists that perform better on specific platforms. In future uses, users should consider this and split up the analysis focusing on different eras of music consumption.

Furthermore, the obvious effect of Billboard's rules and regulations on the performance and popularity of music raise an important question regarding the feedback loop created through Billboard and music popularity. Is it that Billboard is simply a cataloguer of music's popularity, or has it become a more influential and controlling force? Not only has Billboard become a ubiquitous institution, but the very idea of charting on the Billboard Top 100 has become a sort of benchmark of success. At what point does Billboard no longer just categorize and describe the popularity of music trends, and begin to prescribe an aesthetic standard as to which to adhere to, contributing to the possible homogeneity of music and the further commodification of music? This can be seen in the recent emphasis on streaming in Billboard's metrics for popularity, which has caused a shift in album formats to include a higher quantity of songs that are often shorter, thus generating more streams and counting as more sales. Furthermore, the music industry and record labels consider Billboard as a legitimizing benchmark for artists, which often means that there is pressure from these institutions to consider Billboard and its preferences during the creative process. These factors combined with the fact that the actual Billboard magazine often still focuses on pop music instead of leveraging it's platform to give exposure to underground or lesser known artists reinforce the control and influence Billboard has.

Considerations

As this dataset is publicly available on Kaggle, there are no privacy concerns for future users. I would recommend they have a well-defined goal before they begin, as this was something we struggled with. This would then inform whether this dataset is right for their use case. Considering the size and bias, it might not be suitable for use cases requiring comprehensive analysis of the entire music industry. It would not be ideal for studying niche music trends, regional music preferences, or the evolution of underground genres that are not represented in the Billboard charts. That said, this dataset works well for analyzing trends and patterns among popular songs. It's perfect for studying characteristics of chart-topping music,

looking at how certain genres have dominated over time, or predicting which song features contribute to commercial success.