Surviving The Hot 100 🎶

An Analysis of Spotify Song Attributes in Relation to The Billboard Hot 100

Ricky Chau, Jonathan Goshu, Tanadol (Tiger) Lamlertprasertkul, Yunfei (Cynthia) Xing

Background

With the current trends in music constantly changing, our team is curious to see how these measurable traits have changed overtime for the songs making it on the Billboard Hot 100 list and how they have affected the success of any given song after its release. Presently, influences from digital platforms, such as TikTok and other social media sites, have shifted the paradigm for mainstream, popular music and with the result of our research project, we will be able to analyze which direction these shifts are taking the music industry.

Hypothesis

We formulated the following hypotheses to analyze:

Hypothesis One: There is a significant correlation between a song's Danceability score and whether they are on the Billboard for more than 10 weeks.

Hypothesis Two: The average Valence score for songs with 10 weeks or more on the Billboard is significantly different than the overall average Valence score for songs that have ever been on the Billboard.

Hypothesis Three: The average Energy score for songs on the 1970s Billboard is not significantly different from the average Energy score for songs on the 2010s Billboard.

Data

Our Data was collected from the following sources for relevant songs from 1970 to 2022:

We collected weekly Billboard Hot 100 songs from the first week of January, 1970 to the second week of April, 2022. These contained information about song name, artist name, weeks on chart and date of first appearance on Billboard Hot 100. This is about 200,000 data points.

For every song that we collected from the previous step, we collected song features using Spotify API. Songs that are not found on Spotify are eliminated. These contained information about danceability, energy, key, loudness, mode, speechiness, acousticness, instrumentalness, liveness, valence, tempo, type, id, uri, track_href, analysis_url, duration_ms, and time_signature.

Methodology

The following statistical tests were conducted to test our hypotheses:

Hypothesis One -> We used a **chi-square independence test** to check if danceability is related to a song being on the Billboard chart for an extended period of time (10 weeks).

Hypothesis Two -> We used a **two-sample t-test** to check if the valence (positivity) of songs that are on the Billboard for a long time are statistically different from the songs that are only on the Billboard for a relatively short amount of time.

Hypothesis Three -> We used a **two-sample t-test** to check if the mean of energy scores in the songs from the 1970s decade was significantly different compared to songs from the 2010s.

Machine Learning Component -> We used a **feedforward neural network with 3 dense layers** to predict the number of weeks that a song can stay on the chart.

Results

Hypothesis One

Test Statistic	P-Value
51522.998	1.0

The large t-statistic and p-value suggest that our the test is statistically insignificant and thus fails to reject our null hypothesis for hypothesis one. This leads us to determine that **there is a no statistically significant correlation between a song's danceability and the number of weeks it remains on the Billboard Hot 100.**

Hypothesis Two

Tests for **2010s, 2020s** and **overall** had a **statistically significant p-value**, which means that we will reject the null hypothesis. This suggests that in those three time periods, **there is a statistically significant difference in valence score for songs that stayed longer on the Billboard.**

Decade	Test Statistic	P-Value
1970s	0.35	0.72
1980s	-0.50	0.62
1990s	-0.42	0.68
2000s	-0.27	0.78
2010s	3.41	0.00066
2020s	3.64	0.00028
All	-4.099	$-4.17E^{-5}$

Hypothesis Three

This test was statistically significant and lead us to reject the null hypothesis. This suggests that **there is significant difference between the energy of Billboard Hot 100 songs in the 1970s versus the 2010s.** This is also supported by our energy bar chart plot.

Test Statistic	P-Value
-17.86	$5.48E^{-70}$

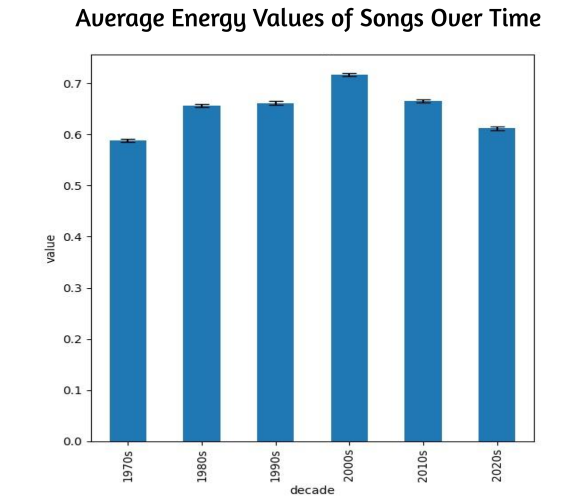

Figure 2. Bar chart displaying average energy levels for songs on the Billboard Hot 100 over the decades, based on songs from the 1970's to 2000's

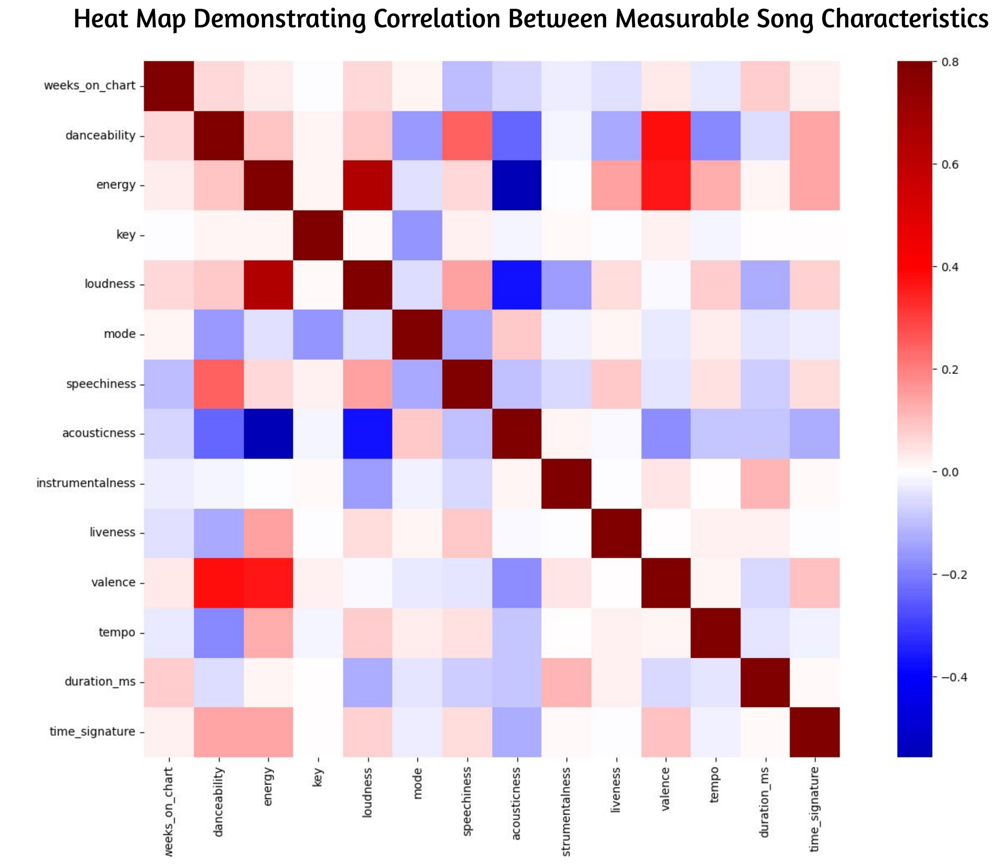

Figure 1. Heat map displaying correlation between spotify song attributes for songs on the Billboard Hot 100, based on songs from the 1970's to 2022.

Machine Learning Component

Our model achieves **30.53% accuracy** on the testing dataset compared to **21.76% accuracy** achieved by a dummy model which always predicts the range with the highest likelihood.

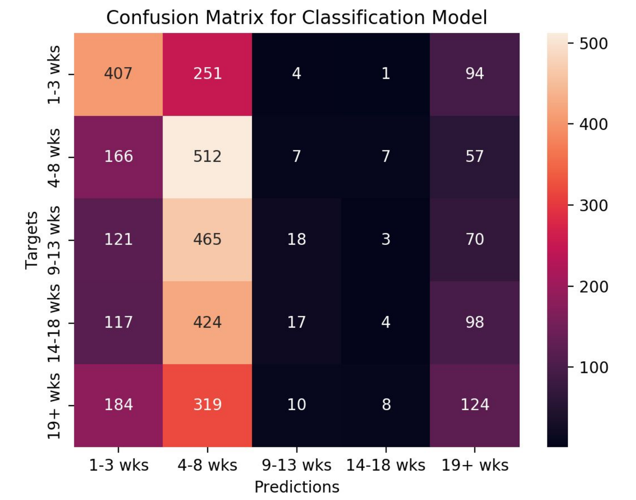

Figure 3. Confusion matrix of the classification model trained on Spotify API song features

Music Over The Years

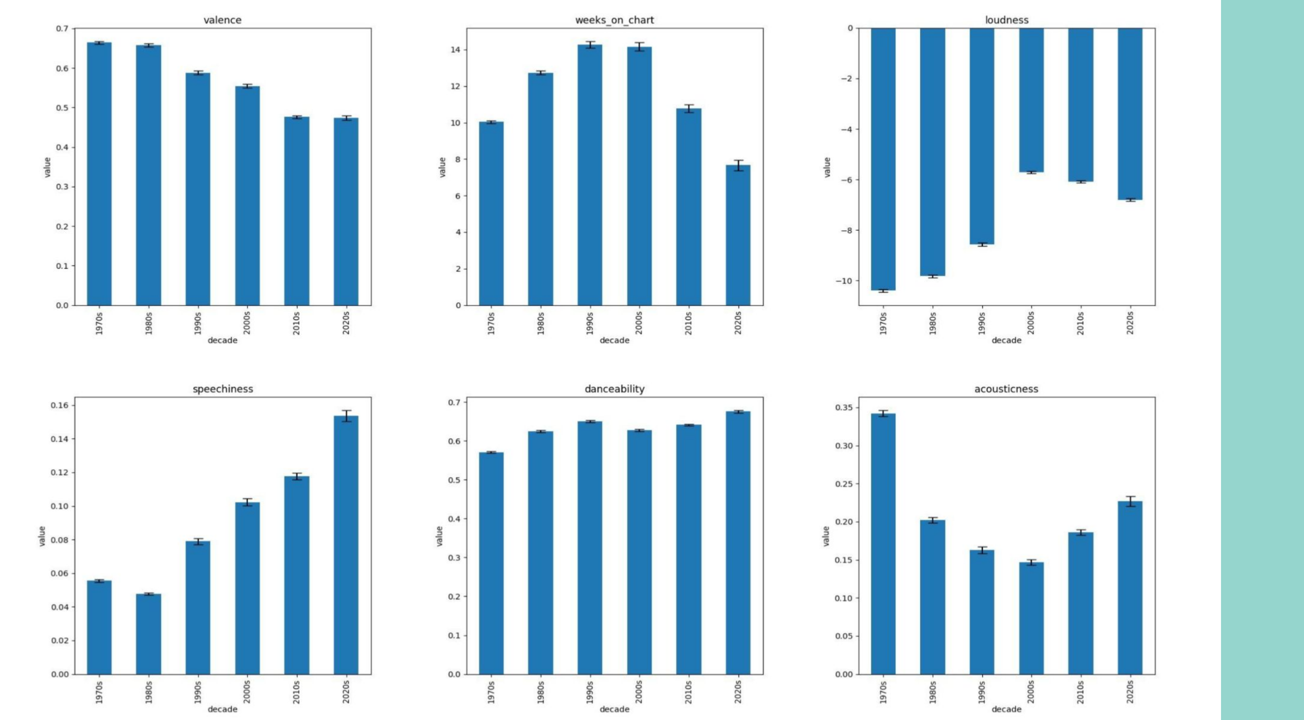

Figure 4. Collection of bar charts representing changes in the average values of various Spotify attributes and average number of weeks on chart for songs on the Billboard Hot 100, based on songs from the 1970's to 2022.

Conclusion

Based on the statistical tests we conducted, we found that although **danceability** may not have a correlation with the number of weeks a song is on the charts, that the magnitude and important of characteristics like **energy** and **valence** can vary by decade, capturing the changing landscape of the music industry.

In addition, our model was more successful than a dummy model in predicting the length of time a song was popular for. However, with only **~30% accuracy**, we would in the future like to account for more song characteristics such as the artist of the song and the lyrical content.

Challenges

One challenge we encountered during the data collection process was that the search engine provided by Spotify API was not intelligent enough to search for a specific version of a song or songs from old decades. To tackle this, we decided to drop songs that don't appear on search result or use an alternative version, assuming that they share similar features.

Another challenge was the uneven distribution of the dataset. Initially we build a regression machine learning model which learned to predict the exact number of weeks that a song stayed on the chart. However, because our data is skewed toward lower number of weeks, the model failed to learn and relied on the greedy algorithm which predicted the average value to minimize the loss function. To alleviate this, we turned to a classification model and divided the dataset into 5 groups of equal number of songs.

Background

With the current trends in music constantly changing, our team is curious to see how these measurable traits have changed overtime for the songs making it on the Billboard Hot 100 list and how they have affected the success of any given song after its release. Presently, influences from digital platforms, such as Tiktok and other social media sites, have shifted the paradigm for mainstream, popular music and with the result of our research project, we will be able to analyze which direction these shifts are taking the music industry.

Hypothesis

We formulated the following hypotheses to analyze:

Hypothesis One: There is a significant correlation between a song's Danceability score and whether they are on the Billboard for more than 10 weeks.

Hypothesis Two: The average Valence score for songs with 10 weeks or more on the Billboard is significantly different than the overall average Valence score for songs that have ever been on the Billboard.

Hypothesis Three: The average Energy score for songs on the 1970s Billboard is not significantly different from the average Energy score for songs on the 2010s Billboard.

Data

Our Data was collected from the following sources for relevant songs from 1970 to 2022:

We collected weekly Billboard Hot 100 songs from the first week of January, 1970 to the second week of April, 2022. These contained information about song name, artist name, weeks on chart and date of first appearance on Billboard Hot 100. This is about 200,000 data points.

For every song that we collected from the previous step, we collected song features using Spotify API. Songs that are not found on Spotify are eliminated. These contained information about danceability, energy, key, loudness, mode, speechiness, acousticness, instrumentalness, liveness, valence, tempo, type, id, uri, track_href, analysis_url, duration_ms, and time_signature.

Music Over The Years

Capturing the Trend in Mean Values for Musical Characteristics Over time

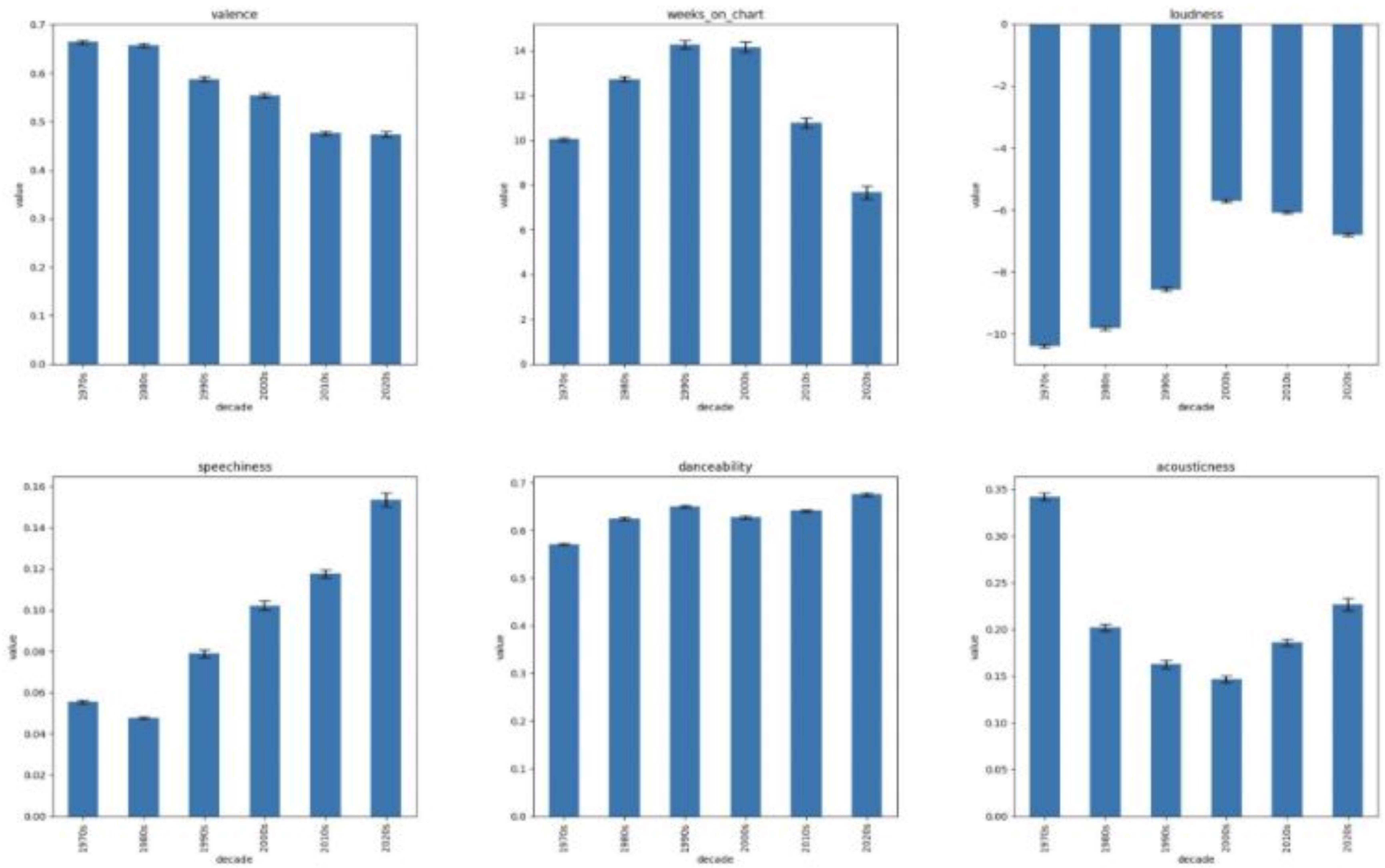

Figure 4. Collection of bar charts representing changes in the average values of various Spotify attributes and average number of weeks on chart for songs on the Billboard Hot 100, based on songs from the 1970's to 2022.

Methodology

The following statistical tests were conducted to test our hypotheses:

Hypothesis One -> We used a **chi-square independence test** to check if danceability is related to a song being on the Billboard chart for an extended period of time (10 weeks).

Hypothesis Two -> We used a **two-sample t-test** to check if the valence (positivity) of songs that are on the Billboard for a long time are statistically different from the songs that are only on the Billboard for a relatively short amount of time.

Hypothesis Three -> We used a **two-sample t-test** to check if the mean of energy scores in the songs from the 1980s decade was significantly different compared to songs from the 2010s.

Machine Learning Component -> We used a **feedforward neural network with 3 dense layers** to predict the number of weeks that a song can stay on the chart.

Results

Hypothesis One

Test Statistic	P-Value
51522.998	1.0

The large t-statistic and p-value suggest that our the test is statistically insignificant and thus fails to reject our null hypothesis for hypothesis one. This leads us to determine that **there is a no statistically significant correlation between a song's danceability and the number of weeks it remains on the Billboard Hot 100.**

Hypothesis Two

Tests for **2010s, 2020s** and **overall** had a **statistically significant p-value**, which means that we will reject the null hypothesis. This suggests that in those three time periods, **there is a statistically significant difference in valence score for songs that stayed longer on the Billboard.**

Decade	Test Statistic	P-Value
1970s	0.35	0.72
1980s	-0.50	0.62
1990s	-0.42	0.68
2000s	-0.27	0.78
2010s	3.41	0.00066
2020s	3.64	0.00028
All	-4.099	$-4.17E^{-5}$

Hypothesis Three

This test was statistically significant and lead us to reject the null hypothesis. This suggests that **there is significant difference between the energy of Billboard Hot 100 songs in the 1970s versus the 2010s.** This is also supported by our energy bar chart plot.

Test Statistic	P-Value
-17.86	$5.48E^{-70}$

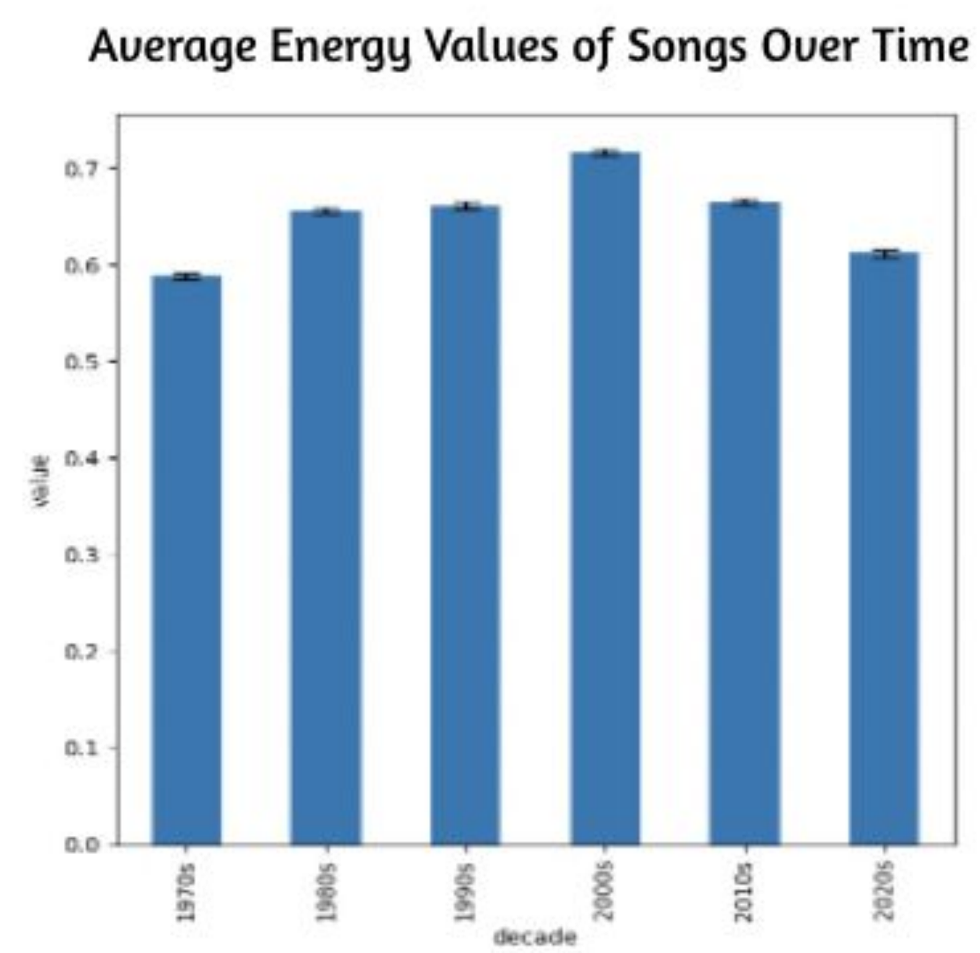

Figure 2. Bar chart displaying average energy levels for songs on the Billboard Hot 100 over the decades, based on songs from the 1970's to 2000's

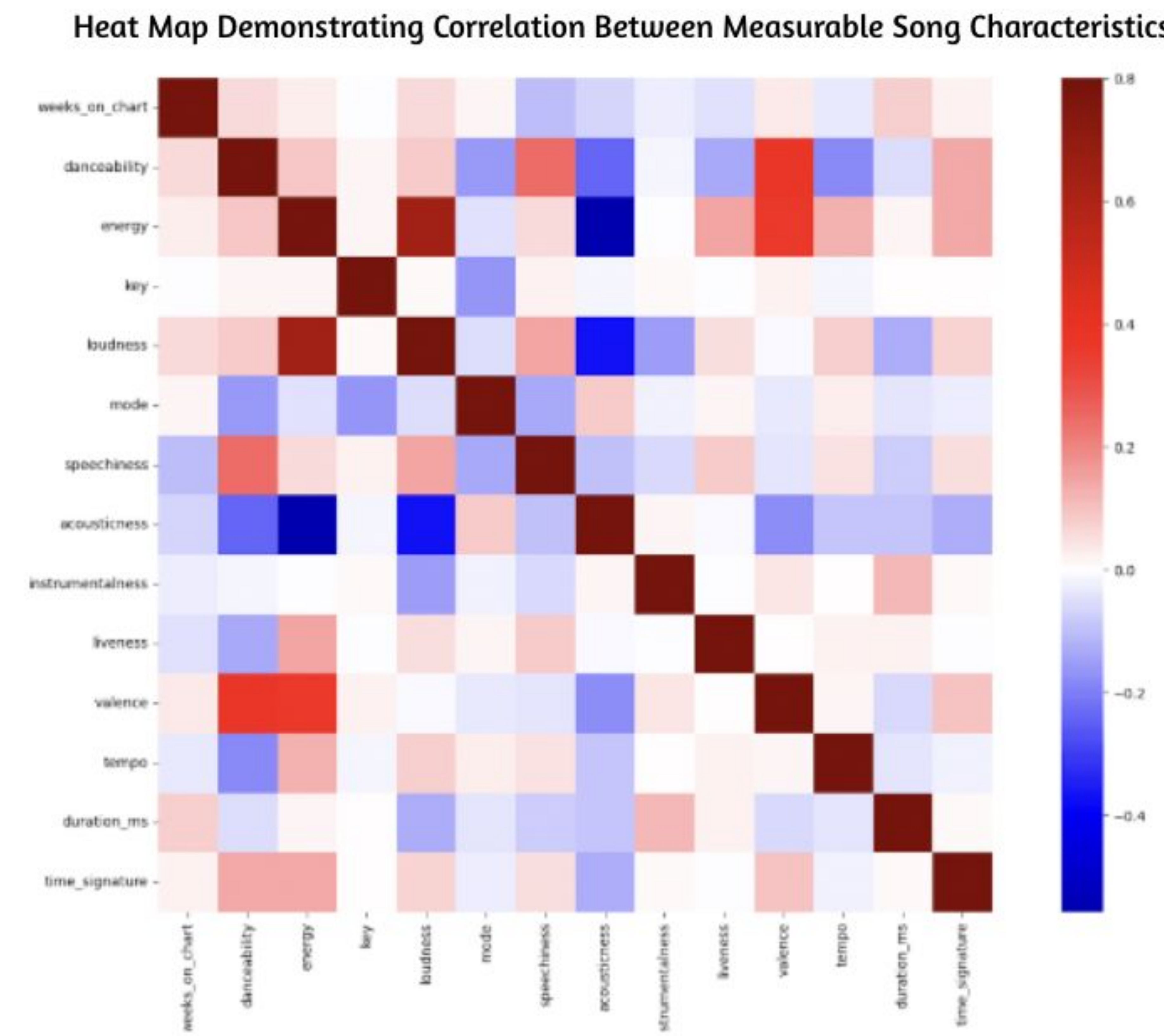

Figure 1. Heat map displaying correlation between spotify song attributes for songs on the Billboard Hot 100, based on songs from the 1970's to 2022.

Machine Learning Component

Our model achieves **30.53% accuracy** on the testing dataset compared to **21.76% accuracy** achieved by a dummy model which always predicts the range with the highest likelihood.

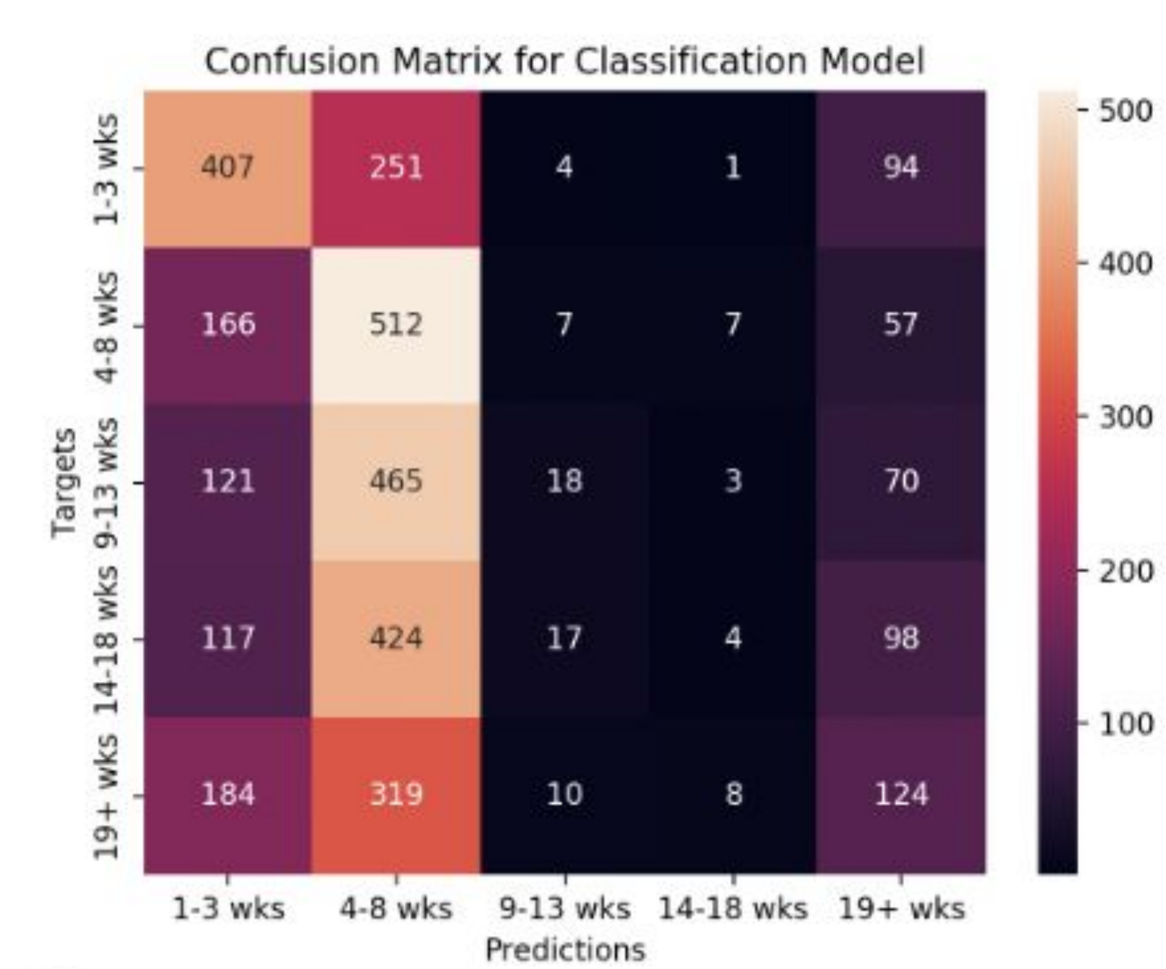

Figure 3. Confusion matrix of the classification model trained on Spotify API song features

Conclusion

Based on the statistical tests we conducted, we found that although **danceability** may not have a correlation with the number of weeks a song is on the charts, that the magnitude and important of characteristics like **energy** and **valence** can vary by decade, capturing the changing landscape of the music industry.

In addition, our model was more successful than a dummy model in predicting the length of time a song was popular for. However, with only **~30% accuracy**, we would in the future like to account for more song characteristics such as the artist of the song and the lyrical content.

Challenges

One challenge we encountered during the data collection process was that the search engine provided by Spotify API was not intelligent enough to search for a specific version of a song or songs from old decades. To tackle this, we decided to drop songs that don't appear on search result or use an alternative version, assuming that they share similar features.

Another challenge was the uneven distribution of the dataset. Initially we build a regression machine learning model which learned to predict the exact number of weeks that a song stayed on the chart. However, because our data is skewed toward lower number of weeks, the model failed to learn and relied on the greedy algorithm which predicted the average value to minimize the loss function. To alleviate this, we turned to a classification model and divided the dataset into 5 groups of equal number of songs.