

# Sick Beats

Team rim-dj, Logins: dfang8, rdrew, jchu25, makel

## Hypothesis

Over the past two years, COVID-19 has affected every facet of our lives. We sought out to investigate whether the prevalence of COVID-19 (measured by the number of new cases per day) has an effect on the songs that people listen to (quantified by the song attributes of popular songs). Specifically, we wanted to test the following 3 hypotheses: popular songs during COVID-19 peaks will have higher loudness (1), valence (2), and energy (3) attributes than popular songs during COVID-19 troughs. Additionally, using machine learning, we wanted to test the following hypothesis: popular songs during COVID-19 peaks will have more negative lyric sentiment than popular songs during COVID-19 troughs.

## Data

We visually examined national covid trends according to The World in Data to identify five peaks and five troughs in new cases. Using these dates, we conducted web scraping on the Billboard 100 to extract the title and authors of the songs on the list per respective date. As the premier music-rating organization in the US, Billboard publishes the top 100 songs every week according to record sales, radio plays, and streams in the US. Using the Spotify API, we collected data on the loudness, energy, and valence of these top songs for use in our statistical hypothesis testing. Using a combination of the Genius API and web scraping, we extracted lyrics for use in our sentiment analysis component. To find the sentiment of these lyrics, we turned to the Flair NLP library, which has a reputation for slow runtime but accurate predictions. Our lyrics are skewed toward songs by individual artists/bands, rather than collaborations or features.

## Findings

**Claim #1:** Loudness is a feature that measures the average loudness of a track in decibels. Popular songs during COVID-19 peaks tend to be louder than popular songs during COVID-19 troughs.

**Support for Claim #1:** We created two lists as inputs to a two-tailed independent t-test with different variance (no homoscedasticity assumption), with one list containing the loudness values of songs during COVID-19 peaks and one list containing the loudness values of songs during COVID-19 troughs. We found a statistically significant difference between the samples with a p-value of 0.00021. It would be highly unlikely to find a result as extreme as ours if the null hypothesis were true, so we are able to reject the null hypothesis at the 0.05 significance level.

**Claim #2:** Energy is a feature that measures the perceptual intensity and activity of a track (energetic tracks feel fast, loud, noisy, etc.). Popular songs during COVID-19 peaks tend to be more energetic than popular songs during COVID-19 troughs.

**Support for Claim #2:** We created two lists as inputs to a two-tailed independent t-test with different variance (no homoscedasticity assumption), with one list containing the energy values of songs during COVID-19 peaks and one list containing the energy values of songs during COVID-19 troughs. We found a statistically significant difference between the samples with a p-value of 0.03622. It would be highly unlikely to find a result as extreme as ours if the null hypothesis were true, so we are able to reject the null hypothesis at the 0.05 significance level. This result is consistent with Claim #1, given that loudness and energy are intuitively related.

**Claim #3:** Valence is a feature that measures the musical positiveness conveyed by a track (high valence tracks sound more happy, cheerful, euphoric, etc.). There is no evidence to suggest that there is a difference in valence for popular songs during COVID-19 peaks and popular songs during COVID-19 troughs.

**Support for Claim #3:** We created two lists as inputs to a two-tailed independent t-test with different variance (no homoscedasticity assumption), with one list containing the valence values of songs during COVID-19 peaks and one list containing the valence values of songs during COVID-19 troughs. We found a p-value of 0.82926, which is insufficient to reject the null hypothesis that there is no difference in valence between the two samples, so we conclude that valence of popular songs is similar between COVID-19 peaks and COVID-19 troughs.

**Claim #4:** There is no difference between the sentiment of lyrics of popular songs at COVID-19 peaks and popular songs during COVID-19 troughs.

**Support for Claim #4:** We scraped lyrics from Genius.com and then used Flair's NLP library to conduct sentiment analysis on the lyrics. Flair is known to trade off runtime for high accuracy. We then ran two-tailed independent t-tests with different variance on sentiment, but we failed to arrive at a statistically significant result. To explore additional analyses, we also compared the number of total words and unique words in the lyrics of top songs, but we did not uncover any statistically significant difference in these metrics either.

