

Introduction

Our overall goal was to complete a rigorous analysis of trends in the music industry during the pandemic - more specifically, mid-March 2020 to the end of June 2020. As the pandemic began and social restrictions tightened, people all over the world naturally became lonelier than usual and generally started to experience more negative emotions. We were curious if this had any impact on what music people listened to, and if, perhaps, the popular songs during the pandemic were “sadder” than before. We were able to measure this in several ways, outlined in the rest of our poster.

Our Questions

- How did overall listening habits change?
- How did the sentiment of popular songs’ lyrics change?
- Did the prevalence of COVID (e.g. higher case numbers) contribute to any shifts in the music people listened to?

To answer these questions, we put our many skills to use that we have learned from previous courses (especially Intro to Data Science, taught by Dr. Mattei).

There were also a vast amount of evaluating metrics we were able to analyze, which we found through a lot of research and time spent on many different websites.

Below is a QR code which will take the scanner to our Dash application. This is where we visualized the most important results from our analysis.



QR Code to <https://pandemic-music-trends.herokuapp.com>.

Approach

- Split the timeline into two parts
 - 1/1/20 - 3/10/20 (pre-pandemic)
 - 3/11/20 - 6/30/20 (during the height of the pandemic, when lockdowns and restrictions were most prevalent)
- Used the Python libraries *pandas* and *plotly*
 - To store our massive datasets into easy-to-work-with dataframes
 - To neatly visualize our data
 - Dash runs on *plotly*, so this was the obvious choice out of the visualization libraries
- Analyzed lyrics for trending songs
 - Obtained lyrics using Genius and stored them in a new database
 - Performed a sentiment analysis on lyrics using *spacy* and *nltk*
- Created multiple “graph generators”
 - Easily allowed us to pick metrics we were interested in comparing/plotting over time
- Built a Dash application (QR code above)
 - Visualization of our analysis that is easy to consume

Results

Once we completed our analysis, we had a massive amount of statistical information detailing how the most popular songs throughout our chosen six-month period had changed. We were able to measure these changes quantitatively. While all of our most important findings are on our Dash application, we offer two of our data visualizations below.

Figure 1 shows “danceability” over time in our chosen time periods. According to Spotify’s API, danceability “describes how suitable a track is for dancing based on a combination of musical elements including tempo, rhythm stability, beat strength, and overall regularity.” We standardized the danceability ratings across the dataset. We believed danceability would decrease once the heights of the pandemic began, and while it didn’t decrease *too* much, the ratings did drop.

Figure 2 shows the “sentiment” of lyrics over the same time constraint as our first figure. Sentiment is an evaluating metric we created based on *nltk*’s “VADER” scores. VADER (Valence Aware Dictionary and sEntiment Reasoner), created by C.J. Hutto and Eric Gilbert, “is a lexicon and rule-based sentiment analysis tool that is specifically attuned to sentiments expressed in social media, and works well on texts from other domains.” We also standardized the sentiment scores and hypothesized that they would decrease (become more negative) as time went on, which they did.

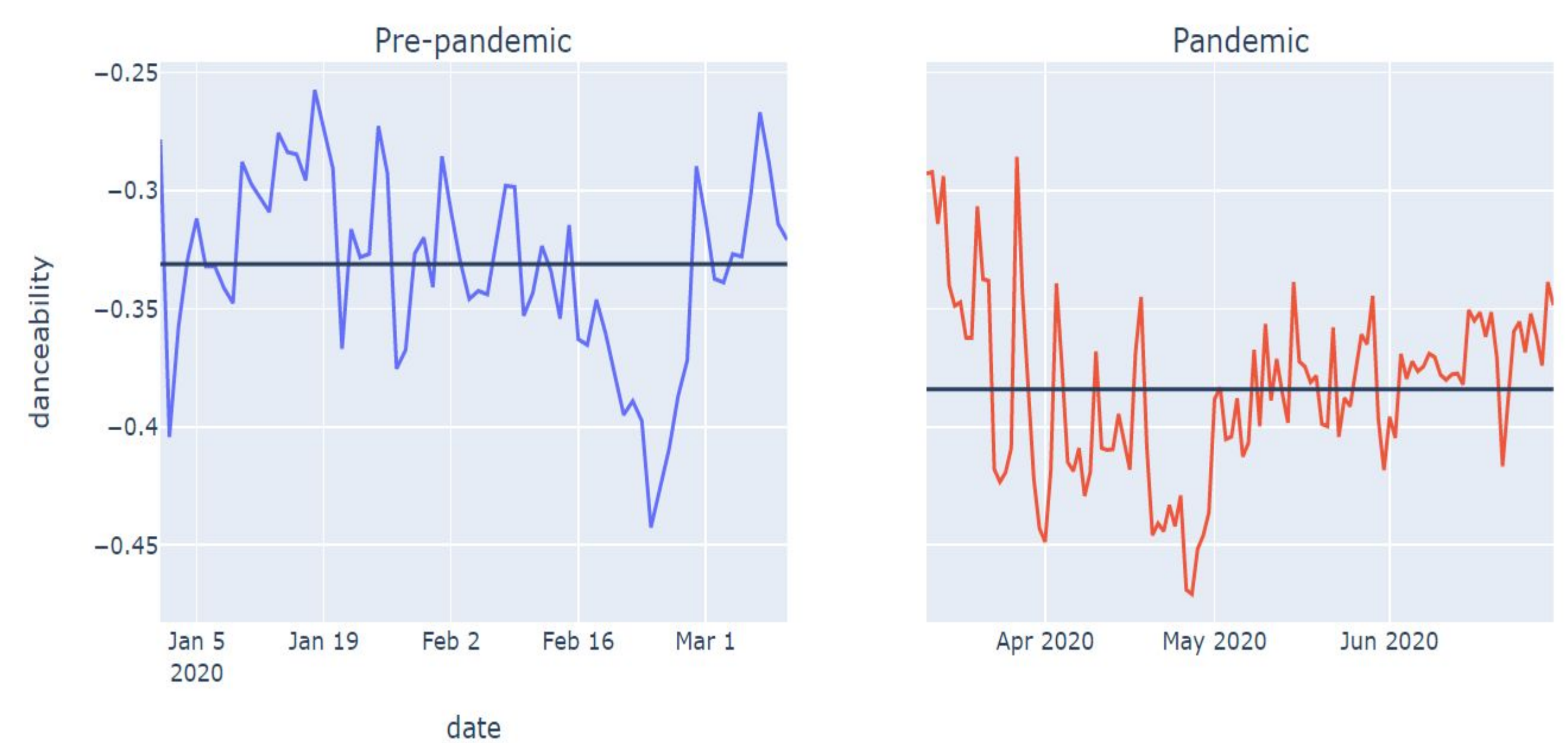


Fig. 1 Danceability Over Time - split into pre-pandemic and during the pandemic.

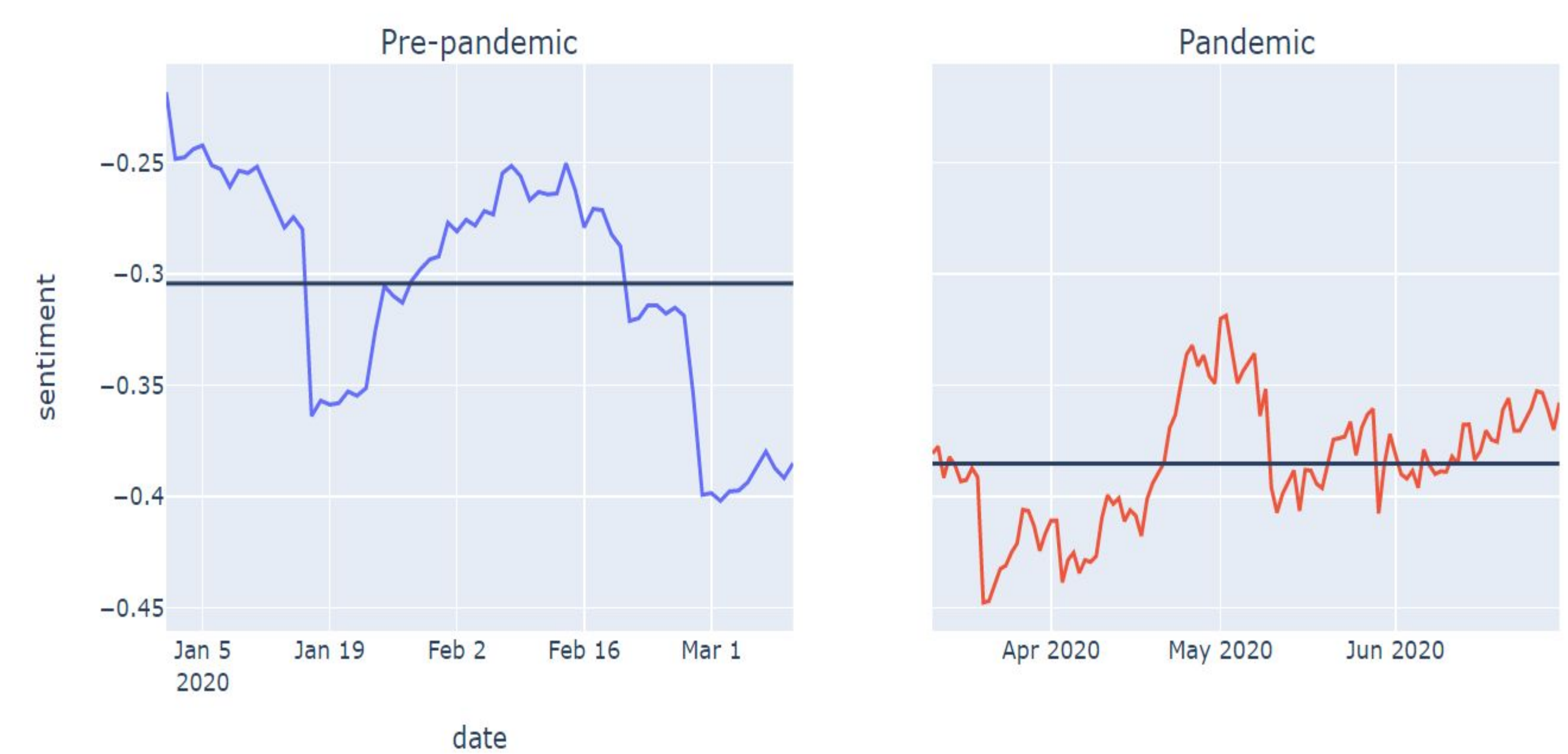


Fig. 2 Sentiment Over Time - split into pre-pandemic and during the pandemic.

Future Considerations

With more time and resources, researchers can take this analysis much further. We can consider how music changed in the long term due to the COVID-19 pandemic. It would be fascinating to see not only how lyrics or statistical measurements have changed, but also how the music industry changed as a whole.

References

- Pepepython, 2021, “Spotify HUGE database - daily charts over 3 years”, Kaggle.com, <https://www.kaggle.com/datasets/pepepython/spotify-huge-database-daily-charts-over-3-years>
- Lyrics provided by Genius, see website.
- Google trends provided by Google.
- *Spacy* and *nltk* packages used for sentiment analysis.