Lab 05 – MATH 240 – Computational Statistics

Brendan Mariano
Colgate
Mathematics
`bmariano@colgate.edu`

Abstract

In this lab, we used the `tidyverse` to group each artist and summarize their data for each song measuring feature. This ultimately allowed us to compare each artists similarity to Allentown for select features and determine who was the biggest contributor.

Keywords: Tidyverse; Summarizing; ggplot

1 Introduction

The overarching goal for this lab was to determine whether The Front Bottoms, Manchester Orchestra or All Get Out contributed most to their joint song Allentown. We had data from three different places: Essentia (Bogdanov et al., 2013)– a program that analyzes the sound of a track, additional Essentia models (Alonso-Jiménez et al., 2020), and the LIWC, Linguistic Inquiry and Word Count (Boyd et al., 2022), which provides features about the lyrics of a track. Our job was to determine two things which features were relevant and to then analyze them in order to reach a conclusion about who contributed the most to Allentown. The methods sections explains how we determined which features were relevant and how we presented them in an effective way. On the other hand, we convey the trends shown for each artist in the results section and analyze our findings to reach an overall conclusion in the discussion section.

1.1 Intro Subsection

2 Methods

Overall, the problem that we needed to solve was to determine who contributed most to the song. We had a collection of features from 10s of songs from each of the artists, however, we needed to summarize each of their data. This led us to using a summarize function, which created a table grouping each artist with all of their songs and analyzed their min, max, lower fence and upper fence for each quantitative feature. From here, we now had a table for comparison among artists for each feature. In order to translate this towards our overall goal, we determined whether Allentown's value for each feature was an outlier, out of the range or within the range. This allowed us to determine which features mattered and ultimately create plots displaying them. When choosing a feature, I ensured that at least one (often two) of the artists was not within range. This was because, if Allentown was within range for every artist, it wouldn't demonstrate any difference in individual contribution, which wouldn't help us to figure out who contributed more. There were 40 different features which fit our criteria, and it was at first difficult to choose. In order to determine which features were relevant, I researched what each of the features represented– allowing me to select features that were dissimilar. Lastly, I made sure to have four of the features be lyrical so that I was also analyzing the words rather than solely the sound of the music itself. To ultimately answer our question, we created plots to be used for analysis using ggplot2 (Wickham, 2016). I found that box plots demonstrated my data effectively because they are easy to understand and I could insert a line indicating the allentown value.

2.1 Methods Subsection

3 Results

For the features related to the lyrical portion of the song– Cognition, Perception, power and Authentic, the results were split mainly split between Manchester Orchestra and the Front Bottoms. Manchester Orchestra was within range for all four of the lyrical features while The Front Bottoms was in range for two and All Get Out was in range for one. This suggests that both contributed to the lyrical portion of Allentown, but Manchester Orchestra contributed the most.

For the features related to the production of the song, the Manchester Orchestra continued to have the most features within its range with comparison to the other bands. For instance, the Allentown value for `erbbands skewness`, which is otherwise known as the skewness of the error bands feature, is above the fence of All Get Out and the Front Bottoms, but within range of Manchester Orchestra. This suggests that Allentown has a right skewed distribution of error bands, meaning that they have more variation in higher frequencies which can be characteristic of songs by the Manchester Orchestra. Out of the eight different production related features, Allentown was within range for the Manchester Orchestra every single time while it was only within range for All Get Out twice and zero times for The Front Bottoms.

3.1 Results Subsection

4 Discussion

In total, between the lyrical features and the production related features, Allentown was within the range of Manchester Orchestra 12 times while the The Front Bottoms was

within range twice and All Get Out was within the range three times. This suggests that Manchester Orchestra contributed the most across all of portions of the song; however, since The Front Bottoms were within range for half of the lyrical features and none of the production related ones, it is likely that they contributed to the lyrical portion of the song. As for All Get Out, with two indicators within range out of eight for the production and one for lyrical out of four, it is likely that they contributed small amounts to different portions of the song. Since Allentown was within range for a diverse array of features(lyrical and production related) for the Manchester Orchestra, we conclude that they were the greatest contributors to the song.

Bibliography:

References

Alonso-Jiménez, P., Bogdanov, D., Pons, J., and Serra, X. (2020). Tensorflow audio models in essentia. In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 266–270. IEEE.

Bogdanov, D., Wack, N., Gómez, E., Gulati, S., Herrera, P., Mayor, O., Roma, G., Salamon, J., Zapata, J. R., Serra, X., et al. (2013). Essentia: An audio analysis library for music information retrieval. In *ISMIR*, volume 13, pages 493–498.

Boyd, R. L., Ashokkumar, A., Seraj, S., and Pennebaker, J. W. (2022). The development and psychometric properties of liwc-22. *Austin, TX: University of Texas at Austin*, 10:1–47.

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.

5 Appendix