Lab 5 – MATH 240 – Computational Statistics

Sankalp Ojha
Colgate University
Mathematics
`sojha@colgate.edu`

Abstract

This study aims to answer of who contributed most to the making of the song "Allentown". Was it The Front Bottoms, All Get Out, or Manchester Orchestra? Each band's style was anaylzed and compared with the characteristics of Allentown. The study concluded that Manchester Orchestra had the most contribution followed by All Get Out as the second largest contributor. The Front Bottoms were the band which contribiuted the least to the making of Allentown.

1 Introduction

Through this lab, we will be exploring who contributed more to the song "Allentown". Was it The Front Bottoms, All Get Out, or Manchester Orchestra? To answer the question at hand, we will analyze all the songs released, excluding all joint albums, live albums, and single releases contained in a full album or an Extended Play (EP), by all the bands before Allentown. Through the analysis, we will aim to determine the style of each band which will allow us to compare Allentown's style with each band. The comparison will allow us to determine which band's style matches the most with Allentown, leading to the conclusion of which band contributed in a greater quantity.

2 Methods

2.1 Making A batfile.txt

To accomplish this lab, we will first need to create a `batfile.txt` file using the `R` package `stringr` (Wickham, 2023) that we can execute in the command line to produce a `.json` file. Using the `R` package `jsonlite` (Ooms, 2014) we will gather all relevant information needed, such as tempo in beats per minute, musical key or musical mode, to analyze the similarities between the each band's style and Allentown's style.

To create the `batfile.txt` file, I downloaded all the given songs which were assorted by artist with a sub folder for each album. Each album folder contains all the songs in that given album. We retrieved the artist name, album name, and each track's name from the file path to create a file name for each song's `.json` file. Each song's file path and `.json` file name were pasted into an execuatble `batfile.txt` file which will be used to create the necessary `.json` file for each song.

2.2 Gathering All Relevant Data And Merging `.csv` Files

We used Essentia Extractor (Bogdanov et al., 2013) to obtain all the `.json` files for the songs. Once we had all 181 `.json` files one for each song, it was time to sift through all the data and gather all relevant information needed for our analysis. We also were provided with Essentia models (Alonso-Jiménez et al., 2020) which came in the form of a `.csv` file. The grand outcome of this portion of the lab is to sift through the data and create a new data frame with only the relevant information we will need for the final portion of the lab which will be to analyze the data. Cleaning up the data first requires combining similar variables into one average variable. For example, for the party measurement, we took an average of the `eff party` and `nn party` variables to get a mean average party variable. We used this combining process for all the measurements we wanted to analyze at the end. The final step of compiling all the relevant information was creating a new data frame which only contained the necessary information such as artist, album, track, and all the new variables we created. We also used LIWC (Boyd et al., 2022) to extract information about the frequency of keywords and emotions in the songs. We used the `R` function `merge()` to merge the `streaming music extractor.csv`, the key information.`csv` we made, and the LIWC.`csv` file. The new merged data frame will allow us to extract various pieces of information which we will be able to use for making graphs and plots for analyzing which band contributed more to the making of Allentown.

2.3 Making A Descion

Once, we had our master data frame, I created a `R` function which would compare each bands characteristics with Allentown's characteristics. The function finds each band's median, interquartile range (IQR), maximum, and minimum value for a given characteristic. The function finally determines whether the given characteristic is "Within Range", Allentown's level falls within the IQR, "Outlying", Allentown's level falls outside of the IQR but within the max and min, "Out Of Range", Allentown's level falls outside of the max and min values. I furthered summed up the number of occurrences of the three categories each band had. I also made a points which award each band one point for being "Within Range" and half a point for being "Outlying". This way I could quantify the impact of falling into one of the three categories.

3 Results

In the appendix, there is a figure which displays four plots. Each plot shows the number of occurrences each band had for a specific label and also displays the points table. From my observations, Manchester Orchestra had the most contribution to the making of Allentown. All Get Out had the second most contribution followed by The Front Bottoms, who had the least amount of contribution.

4 Discussion

As can be seen in the plots, Manchester Orchestra had the most amount of 'Within Range", which is a clear indicator that they had a hefty contribution. Manchester Orchestra also had the least amount of "Out Of Range", which supports the conclusion that they had the most contribution. With the same logic, All Get Out came second both of those categories as well, which implies they had the second most amount of contribution. The points plot is also consient with the infrence of Manchester Ochestra having the most contribition and All Get Out being the second largest contributor.

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.

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.

Ooms, J. (2014). The jsonlite package: A practical and consistent mapping between json data and r objects. *arXiv:1403.2805 [stat.CO]*.

Wickham, H. (2023). *stringr: Simple, Consistent Wrappers for Common String Operations*. R package version 1.5.1.
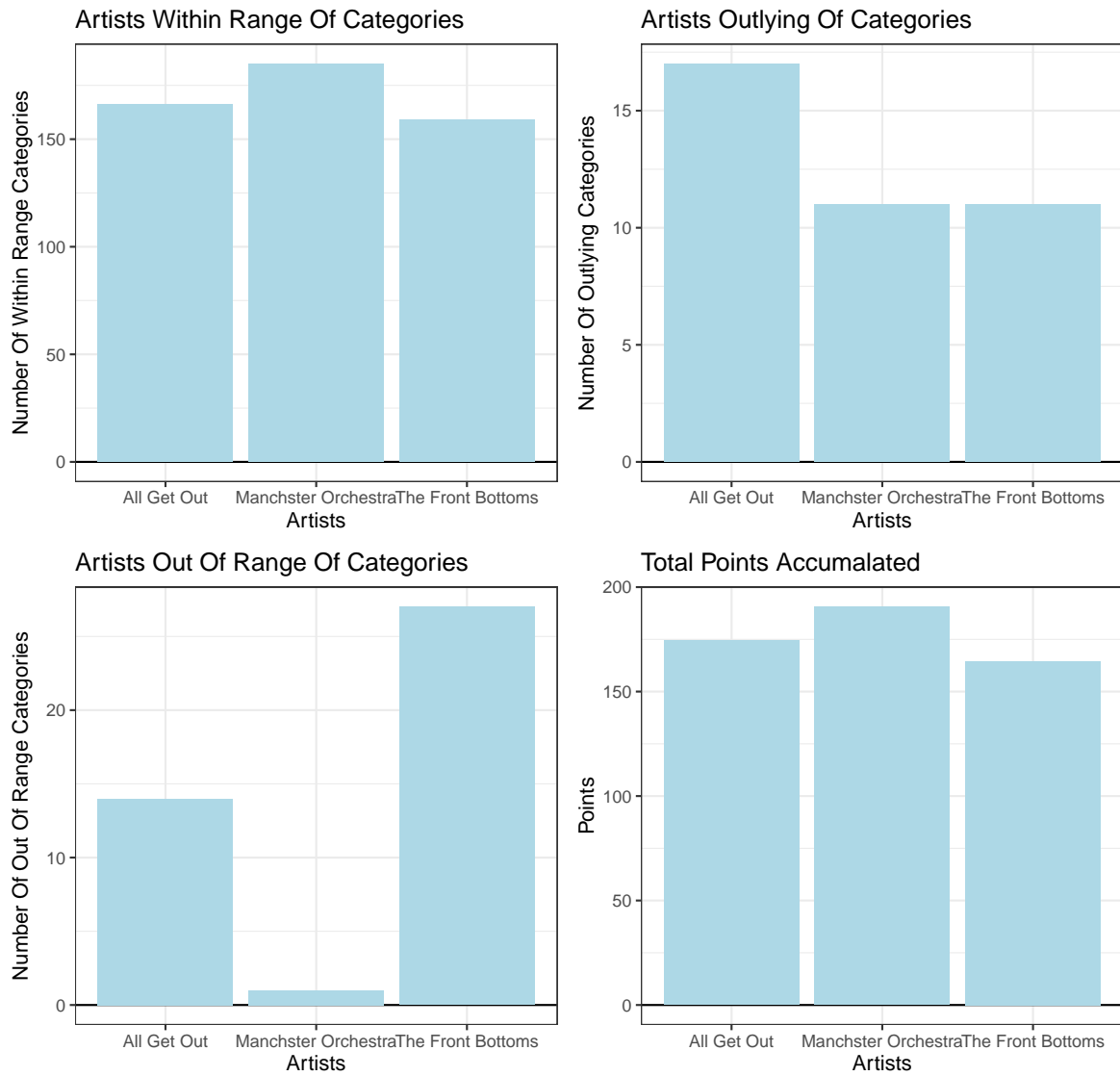
5 Appendix

Figure 1: Plots For Each Category

	artists	count.in.range	count.outlying	count.out.range	points
1	All Get Out	166.00	17.00	14.00	174.50
2	Manchester Orchestra	185.00	11.00	1.00	190.50
3	The Front Bottoms	159.00	11.00	27.00	164.50