

## **Data Appendix**

### **Original Data**

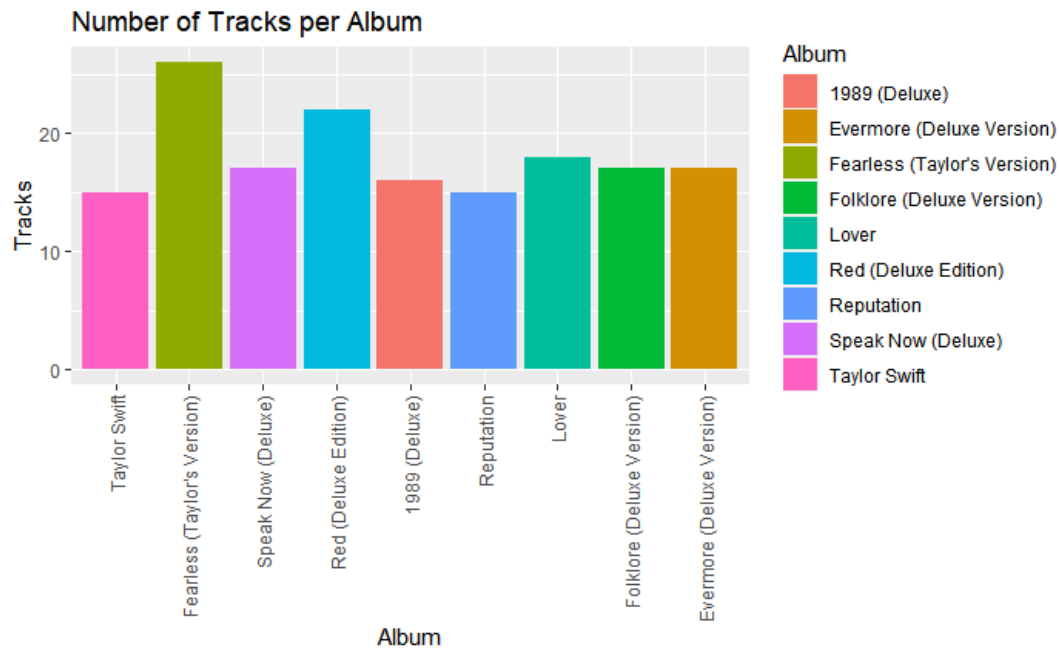
Data Dictionary for all initial data in the DATA folder. This includes the following files:

- 01-taylor\_swift.csv
- 02-fearless\_taylors\_version.csv
- 03-speak\_now\_deluxe\_package.csv
- 04-red\_deluxe\_edition.csv
- 05-1989\_deluxe.csv
- 06-reputation.csv
- 07-lover.csv
- 08-folklore\_deluxe\_version.csv
- 09-evermore\_deluxe\_version.csv

| <b>Column</b> | <b>Description</b>                            | <b>Potential Responses</b>                                     |
|---------------|---|--|
| albumn_name   | What the album that the song is from is named | Various; text  |
| track_title   | What the song name is                         | Various; text  |
| track_number  | The song number within the album              | 1-18, the first track on each album will always start with one |
| lyric         | The actual song lyric line                    | Various; text  |
| line          | The line number within the song               | 1-67; the first line on each track will always start with one  |

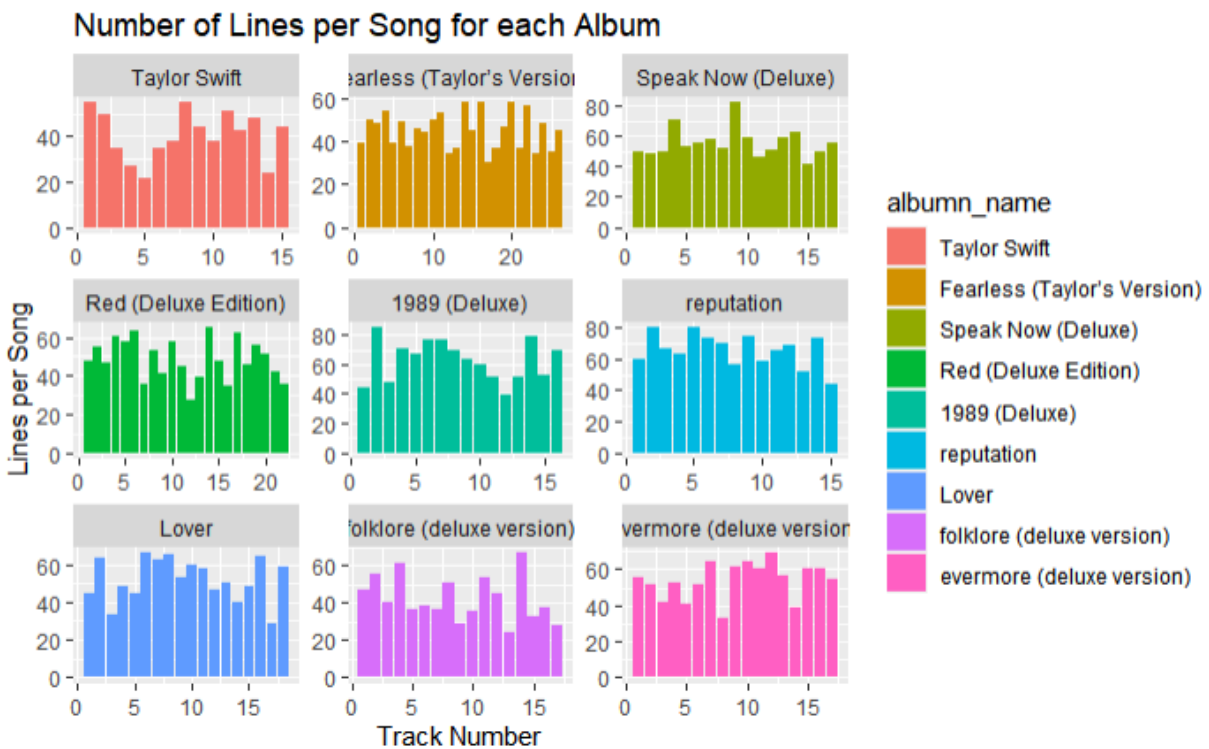
**Figure 1.**

A histogram illustrating the number of tracks per album.



**Figure 2.**

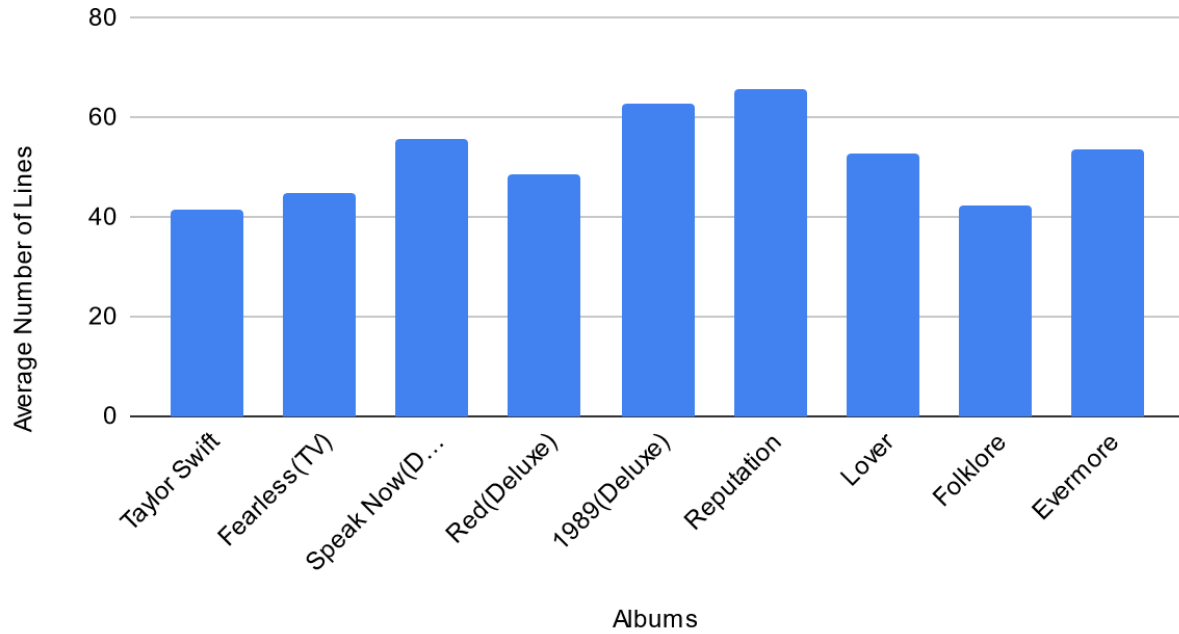
A histogram for each album illustrating the number of lines for each track.



**Figure 3.**

A histogram displaying the average number of lines seen across songs in each album.

### Average Number of Lines in Each Album



### **Analyzed Data**

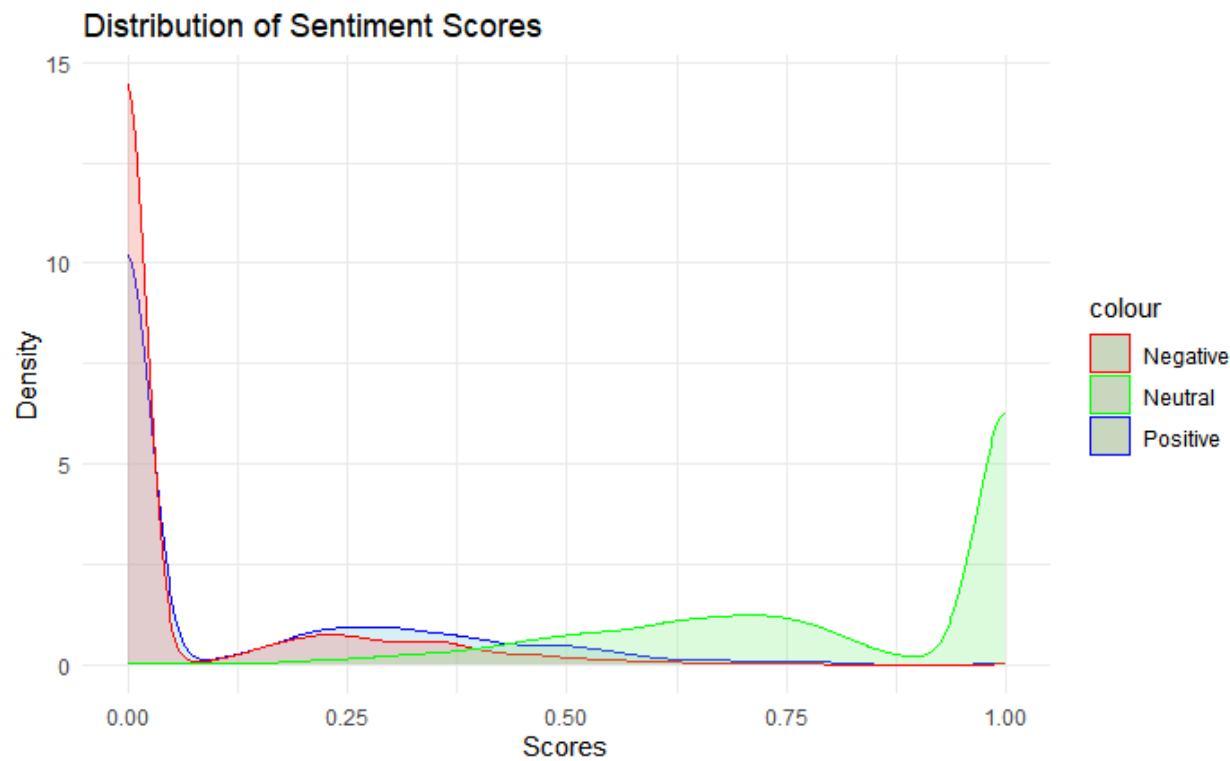
The analyzed dataset is a dataset consisting of all initial data combined into one dataset. It also includes positive, negative, neutral, and overall sentiment scores for each line for each song. To create this dataset, we ran a script called taylor swift.py that puts all created data into a csv called taylor\_swift.csv.

Data Dictionary for **taylor\_swift.csv**.

| Column       | Description   | Potential Responses  |
|--------------|---|--|
| albumn_name  | What the album that the song is from is named                                     | Various; text  |
| track_title  | What the song name is   | Various; text  |
| track_number | The song number within the album  | 1-18, the first track on each album will always start with one |
| lyric        | The actual song lyric line  | Various; text  |
| line         | The line number within the song   | 1-67; the first line on each track will always start with one  |
| negative     | Numeric designation of how negative the VADER software determined the line to be. | 0 to 1 with detail to the thousandth decimal point*            |
| neutral      | Numeric designation of how neutral the VADER software determined the line to be.  | 0 to 1 with detail to the thousandth decimal point*            |
| positive     | Numeric designation of how positive the VADER software determined the line to be. | 0 to 1 with detail to the thousandth decimal point*            |
| overall      | Overall designation of the lyric line.  | Negative, Neutral, or Positive                                 |

\* Negative, Neutral, or Positive should add up to 1 for each line. 1 is considered to be strongly indicative while 0 is considered very weakly indicative. For example, a negative rating of 1 means that line is very negative.

**Figure 4.**  
A density curve illustrating the distribution of positive, negative, and neutral sentiment scores across all lines of each song in the nine albums we are analyzing.



**Figure 5.**  
A table showing summary statistics for the final dataset.

| albumn_name                            | track_title         | track_number   | lyric               | line             | negative           | neutral           | positive          | overall             |
|--|---------------------|----------------|---------------------|------------------|--------------------|-------------------|-------------------|---------------------|
| Fearless<br>(Taylor's<br>Version):1170 | Length:8358         | Min. : 1.000   | Length:8358         | Min. :<br>1.00   | Min.<br>:0.00000   | Min.<br>:0.0000   | Min.<br>:0.0000   | Length:8358         |
| Red (Deluxe<br>Edition) :1072          | Class<br>:character | 1st Qu.: 5.000 | Class<br>:character | 1st<br>Qu.:13.00 | 1st<br>Qu.:0.00000 | 1st<br>Qu.:0.6490 | 1st<br>Qu.:0.0000 | Class<br>:character |
| 1989 (Deluxe)<br>:1006                 | Mode<br>:character  | Median : 9.000 | Mode<br>:character  | Median<br>:26.00 | Median<br>:0.00000 | Median<br>:1.0000 | Median<br>:0.0000 | Mode<br>:character  |
| reputation : 988                       | NA                  | Mean : 9.627   | NA                  | Mean<br>:27.78   | Mean<br>:0.06828   | Mean<br>:0.8146   | Mean<br>:0.1171   | NA                  |
| Speak Now<br>(Deluxe) : 945            | NA                  | 3rd Qu.:14.000 | NA                  | 3rd<br>Qu.:40.00 | 3rd<br>Qu.:0.00000 | 3rd<br>Qu.:1.0000 | 3rd<br>Qu.:0.2380 | NA                  |
| Lover : 934                            | NA                  | Max. :26.000   | NA                  | Max.<br>:85.00   | Max.<br>:1.00000   | Max.<br>:1.0000   | Max.<br>:1.0000   | NA                  |
| (Other) :2243                          | NA                  | NA             | NA                  | NA               | NA                 | NA                | NA                | NA                  |

## References

- [1] “Welcome to Vadersentiment’s documentation!” Welcome to VaderSentiment’s documentation! - VaderSentiment 3.3.1 documentation, 2021. [Online]. Available: <https://vadersentiment.readthedocs.io/en/latest/> [Accessed Feb. 1, 2024].
- [2] “Taylor Swift Song Lyrics (All Albums)”, Kaggle, 2022. [Online]. Available: <https://www.kaggle.com/datasets/thespacefreak/taylor-swift-song-lyrics-all-albums?resource=download> [Accessed Feb. 8 2024].