# Natural Language Processing

Song Lyrics

# The Artists

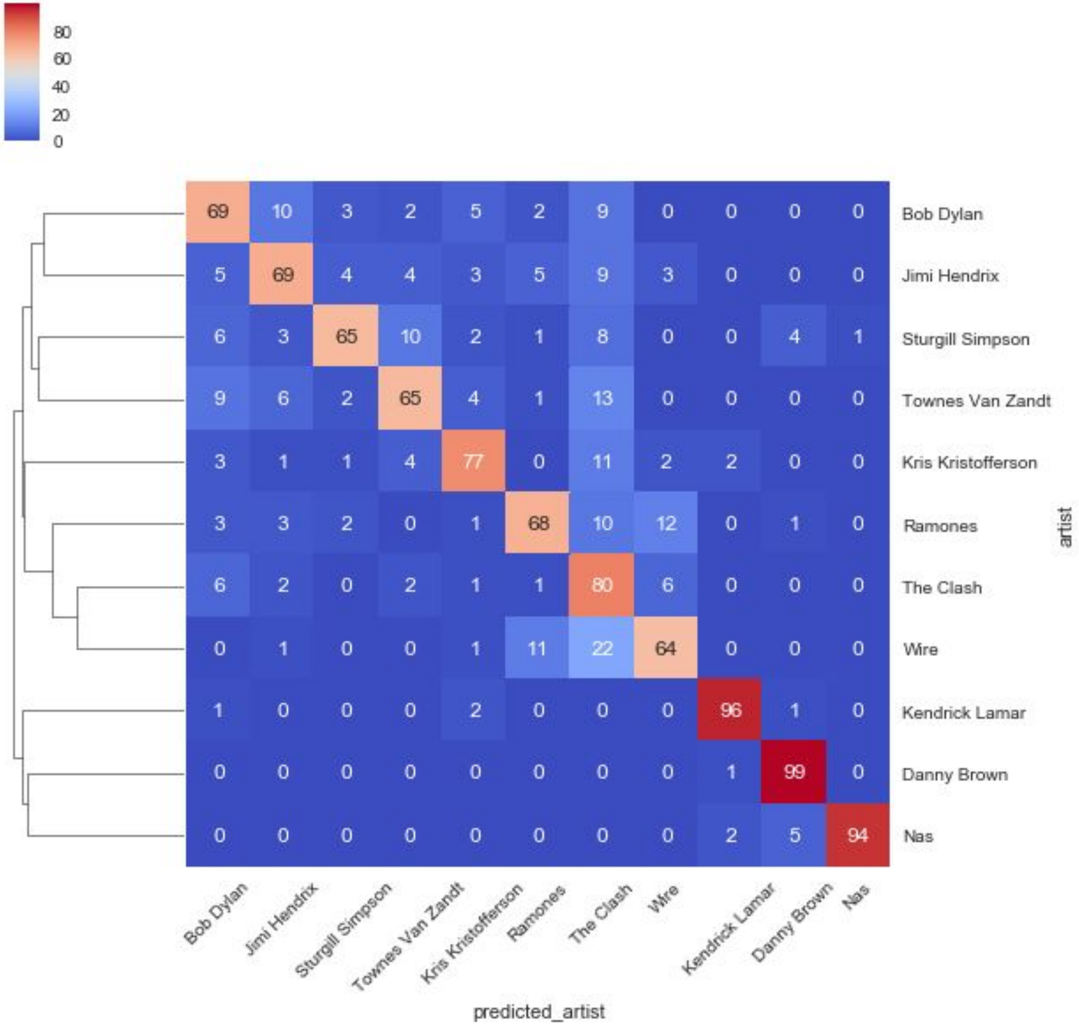
- Bob Dylan
- Jimi Hendrix
- Sturgill Simpson
- Townes Van Zandt
- Kris Kristofferson
- Ramones
- The Clash
- Wire
- Kendrick Lamar
- Danny Brown
- Nas

# Data

- All data was taken from the Genius Lyrics API using R
- Regex processing to get rid of data quirks
- Combined into one line per song
- Expanded contractions
- Identified covers and created dictionary to filter

# Machine Learning Approach

- Created a pipeline using:
  - TfidfTransformer from gensim
  - CountVectorizer and SGDClassifier from sklearn
- Ran 100 random iterations of train_test_split and averaged results
- Got an average F-score of .79



| artist \ predicted_artist | Bob Dylan | Jimi Hendrix | Sturgill Simpson | Townes Van Zandt | Kris Kristofferson | Ramones | The Clash | Wire | Kendrick Lamar | Danny Brown | Nas |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Bob Dylan | 69 | 10 | 3 | 2 | 5 | 2 | 9 | 0 | 0 | 0 | 0 |
| Jimi Hendrix | 5 | 69 | 4 | 4 | 3 | 5 | 9 | 3 | 0 | 0 | 0 |
| Sturgill Simpson | 6 | 3 | 65 | 10 | 2 | 1 | 8 | 0 | 0 | 4 | 1 |
| Townes Van Zandt | 9 | 6 | 2 | 65 | 4 | 1 | 13 | 0 | 0 | 0 | 0 |
| Kris Kristofferson | 3 | 1 | 1 | 4 | 77 | 0 | 11 | 2 | 2 | 0 | 0 |
| Ramones | 3 | 3 | 2 | 0 | 1 | 68 | 10 | 12 | 0 | 1 | 0 |
| The Clash | 6 | 2 | 0 | 2 | 1 | 1 | 80 | 6 | 0 | 0 | 0 |
| Wire | 0 | 1 | 0 | 0 | 1 | 11 | 22 | 64 | 0 | 0 | 0 |
| Kendrick Lamar | 1 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 96 | 1 | 0 |
| Danny Brown | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 99 | 0 |
| Nas | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 5 | 94 |

# Results

- Clear clustering around Country, Rap, Punk, and Rock (whatever that means)
- Rappers are the most distinctive, likely due to the volume of words available for their genre
- Lots of artists were mistaken for The Clash
    - Wide range of topics for their songs, wide range of emotions
- Danny Brown is the most unique artist
    - The only mix-up features a song with Kendrick Lamar and thus got picked up
- Interesting cross-genre confusion

# Next Steps

- Dig in to each artist’s word profile
- Answer questions like:
  - What words make an artist distinctive?
  - Why are artists confused for each other?
  - Why is that one random song predicting consistently for the wrong artist?