

Anthony Kang

Professor Solomon

CS 277

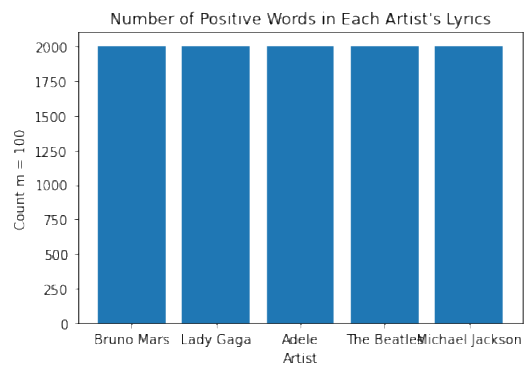
2023/3/12

Final Visualization

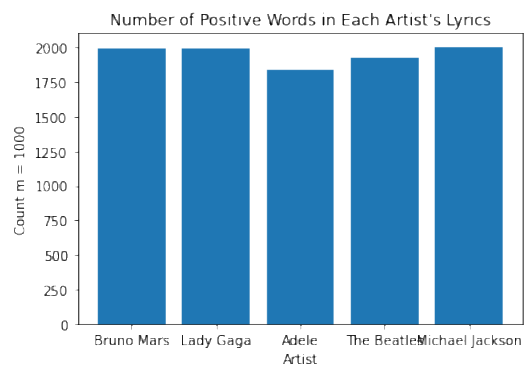
Bloom Filter finding positive words in lyrics

*There were 2005 positive words

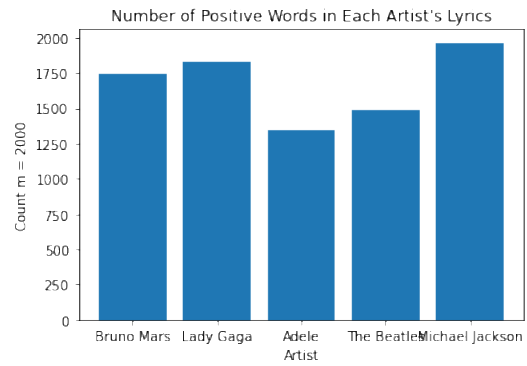
When $m = 100$



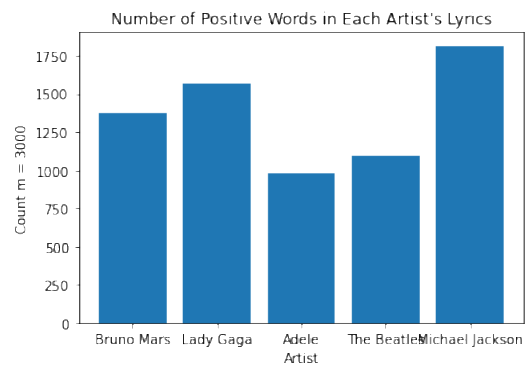
When $m = 1000$



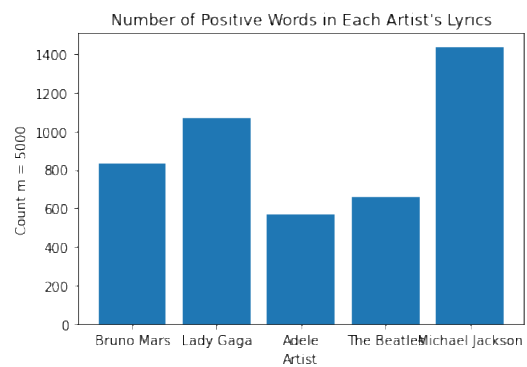
When $m = 2000$



When $m = 3000$



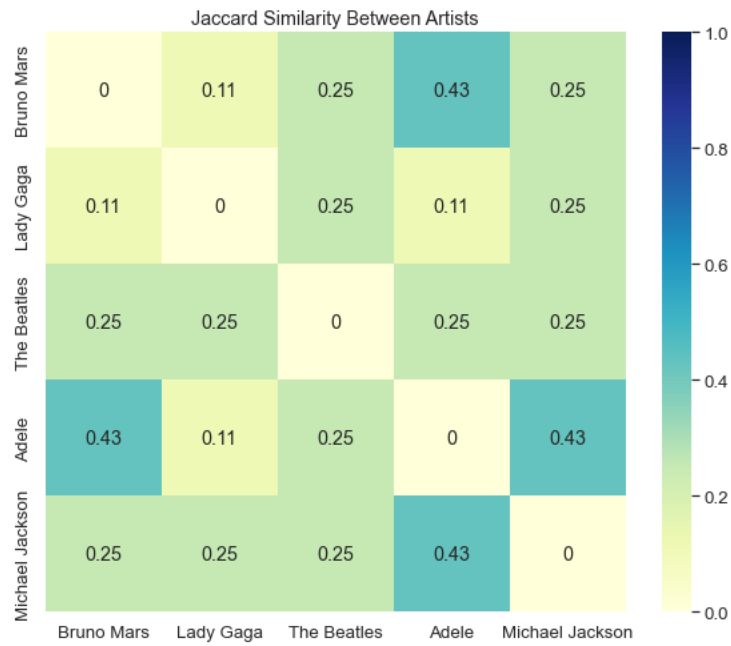
When $m = 5000$



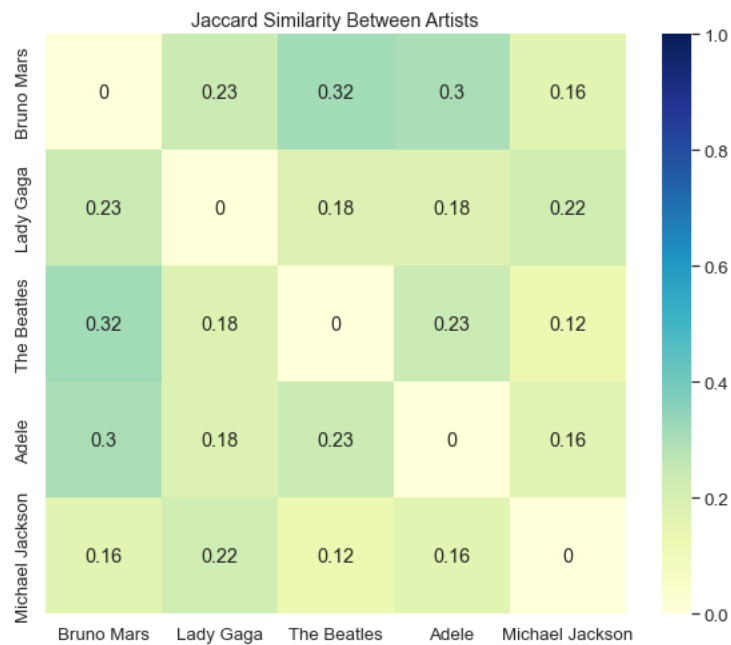
After 5000, the bar chart almost looked the same with no changes.

Jaccard Similarity between each artists

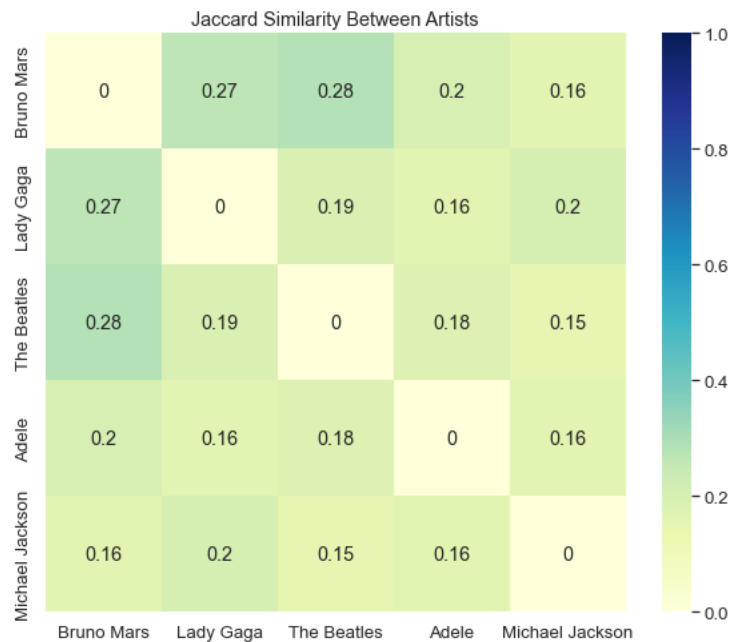
Minhash with k = 5



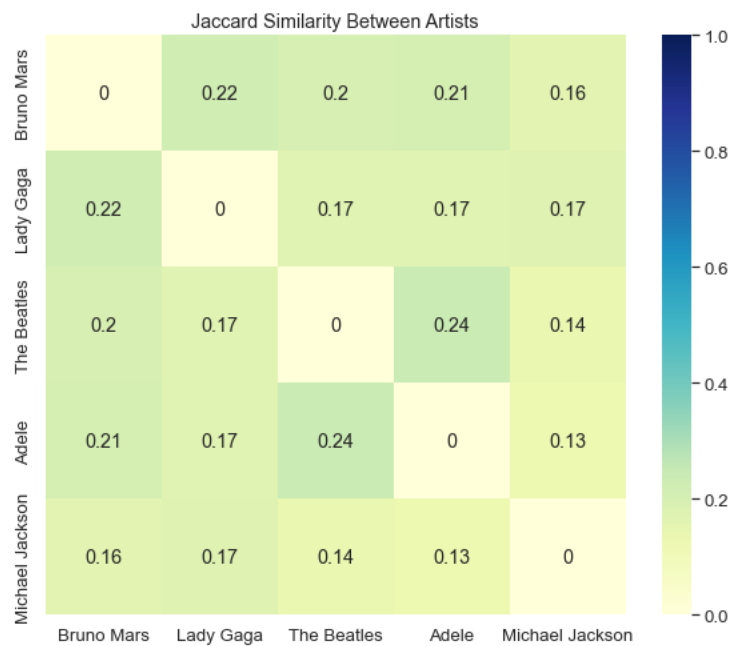
Minhash with k = 50



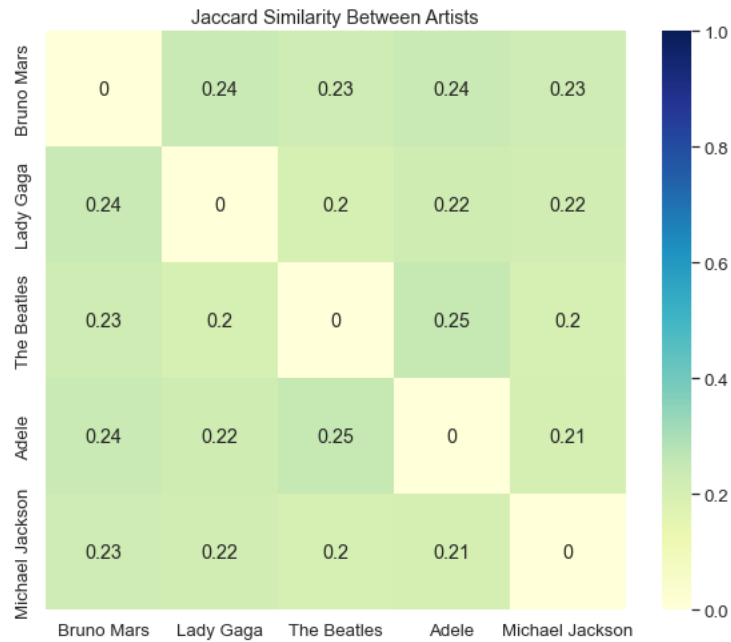
Minhash with k = 100



Minhash with k = 1000

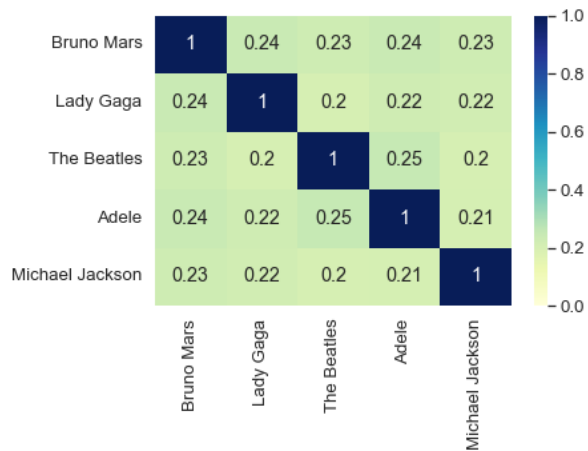


Minhash with k = 100000



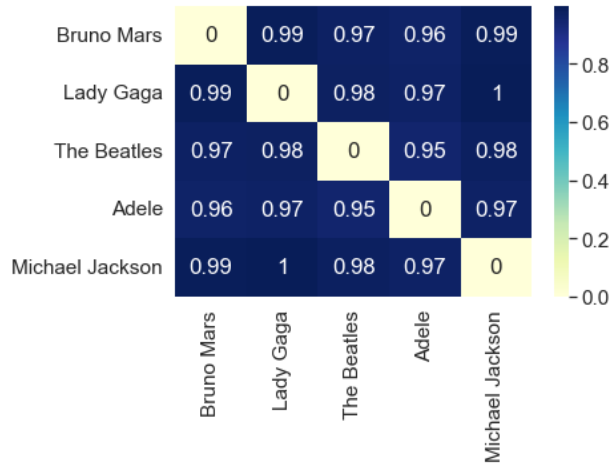
After 100 thousands, it didn't really change much so I was not able to observe meaningful differences after it.

Jaccard Similarity with raw data

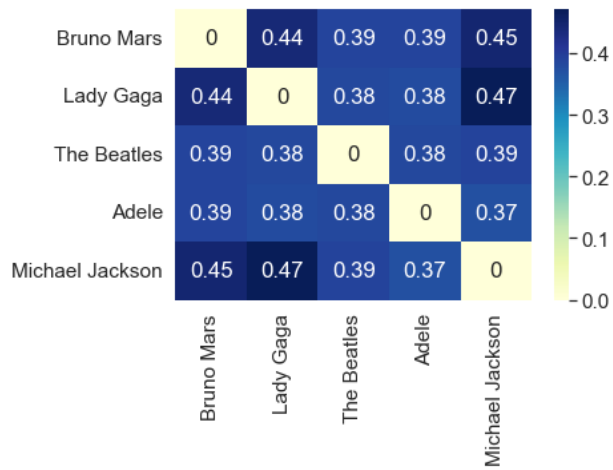


Jaccard Similarity with bloom filter

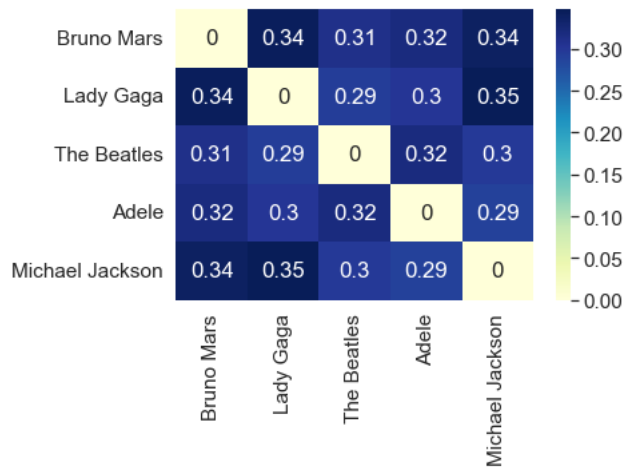
M = 5000, k = 10



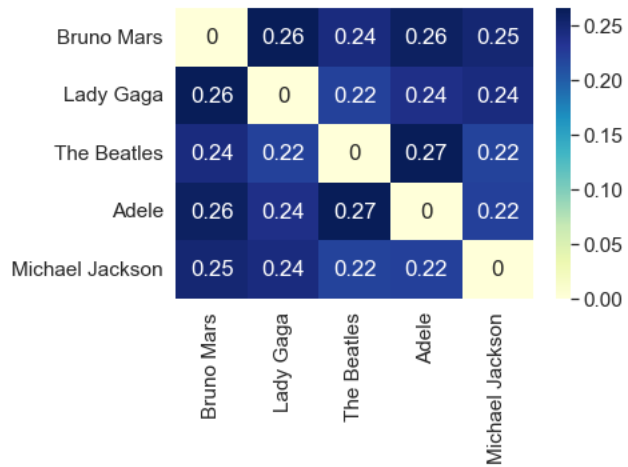
M = 50000, k = 10



M = 100000, k=10



M = 500000, k=10



After 500 thousands, it didn't really have meaningful changes.