



TASK

Exploratory Data Analysis on the Taylor Swift Lyrics Data Set

Visit our website

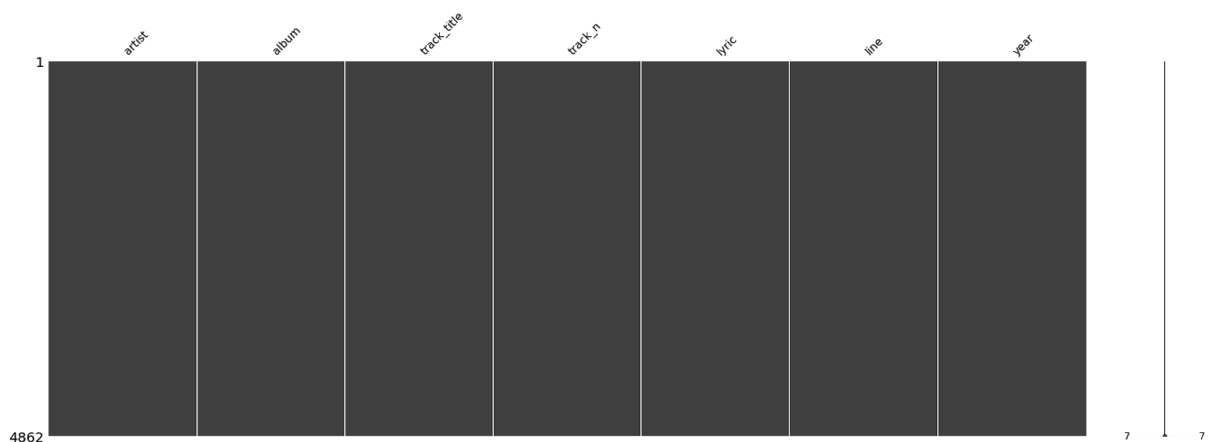
Introduction

The data set contains Taylor Swift's lyrics from her first 6 studio albums. It includes 94 songs, album names, track names, and year of release.

DATA CLEANING

I removed any duplicate rows, unnecessary or redundant columns, and I used the missingno library to create a visualization to check for any missing values in the data set. The lyrics per song were separated over several rows, which represented the different lines in the song; in preparation for analysing the lyrics, I merged all the song lyrics to be in one row each, making it easier to find the word count and unique word count. Part of this was removing all punctuation and making the letters lowercase.

MISSING DATA



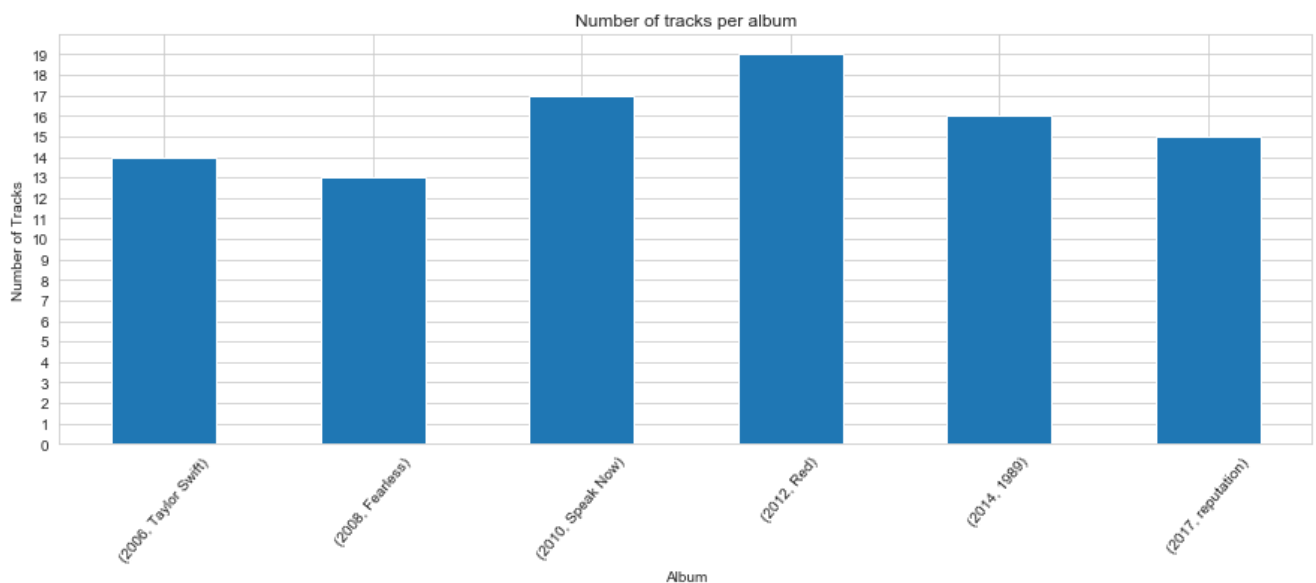
Fortunately, this data set did not have any missing data. It is based on media that is widely available, so it is understandable for all data values to be captured effectively.

DATA STORIES AND VISUALIZATIONS

TRACKS PER ALBUM

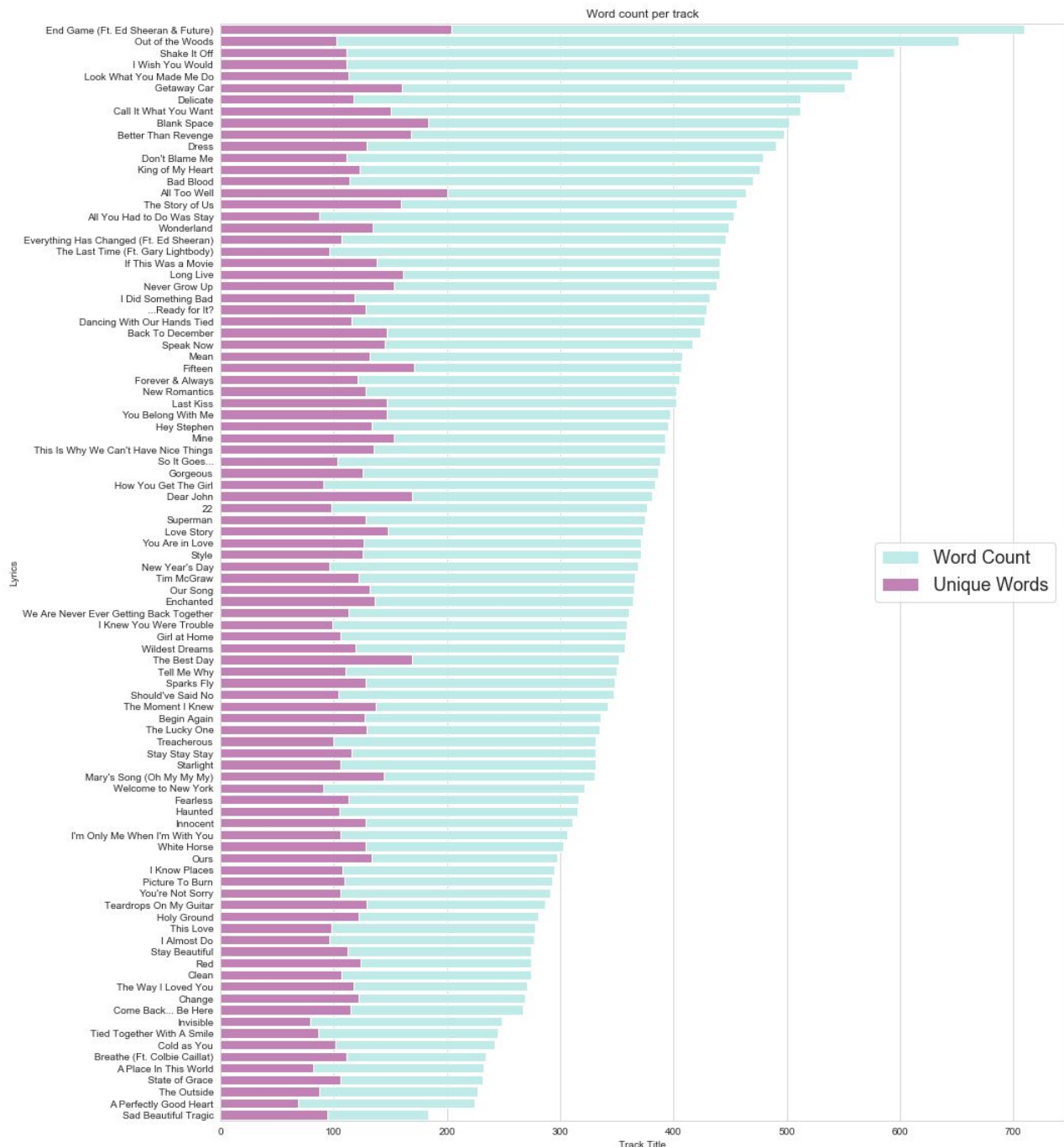
I started off the analysis by looking at how many tracks were on each album.

The number of tracks per album had a variation from 13 to 19 songs, not showing any patterns in particular. This information became more interesting once we started looking at the number of words used over the years of the albums being released, but we will come back to this.



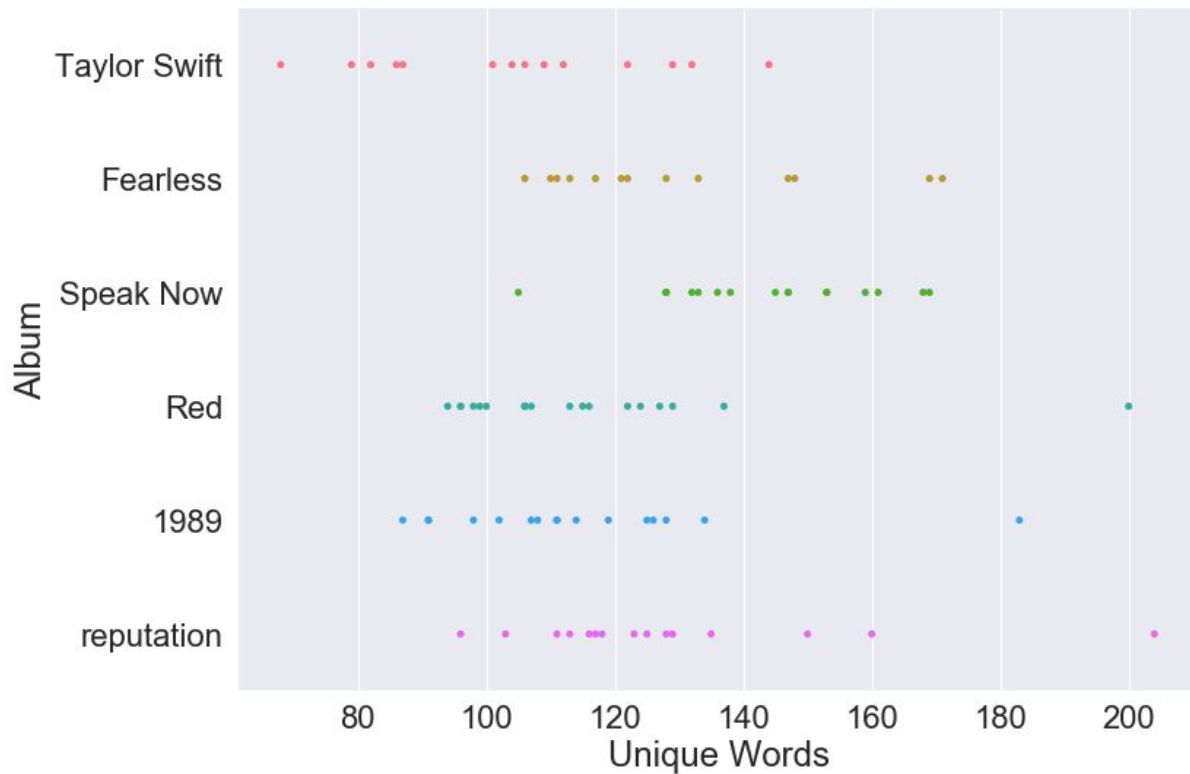
WORD COUNT PER TRACK

The next analysis I ran was looking at the number of lyrics per song, as well as the number of unique words used. This revealed an interesting truth about pop music, because even though there are tracks with hundreds of words, the number of unique words seldom go over 200, and they usually stay around 100-150 unique words per song. The truth is that pop lyrics are repetitive, this is not necessarily a bad thing, because pop music remains one of the most popular genres in the music world.



UNIQUE WORDS PER ALBUM

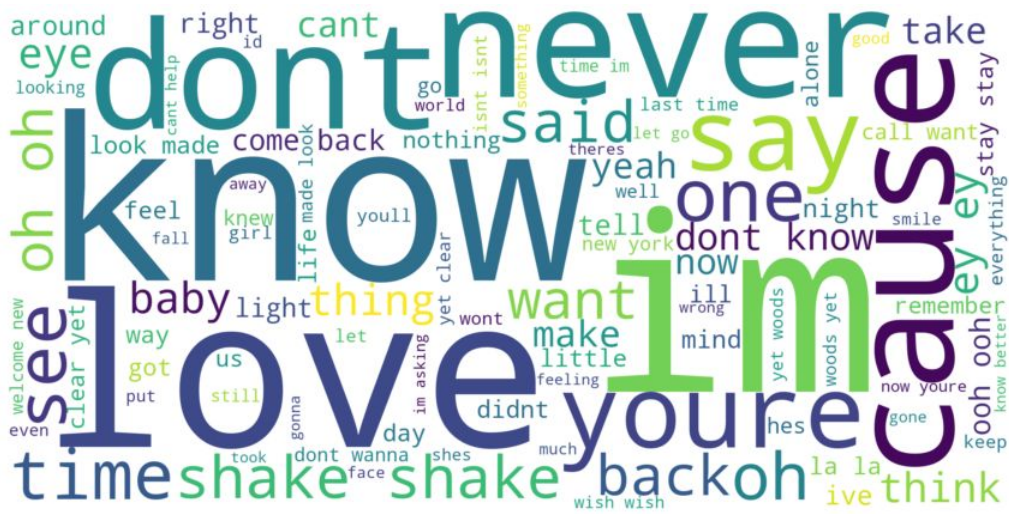
The previous visualization was a lot to take in, so I took a different perspective and made a visualization showing the unique words per album, just to give a sense of which albums had more diversity in the lyrics.



As we can see, this category plot is much more comfortable to look at, although it does not reveal the same trend as the previous plot. Nevertheless, it does reveal which albums had the songs with the most unique lyrics, with '*reputation*' coming in first, and '*Red*' with a very close second.

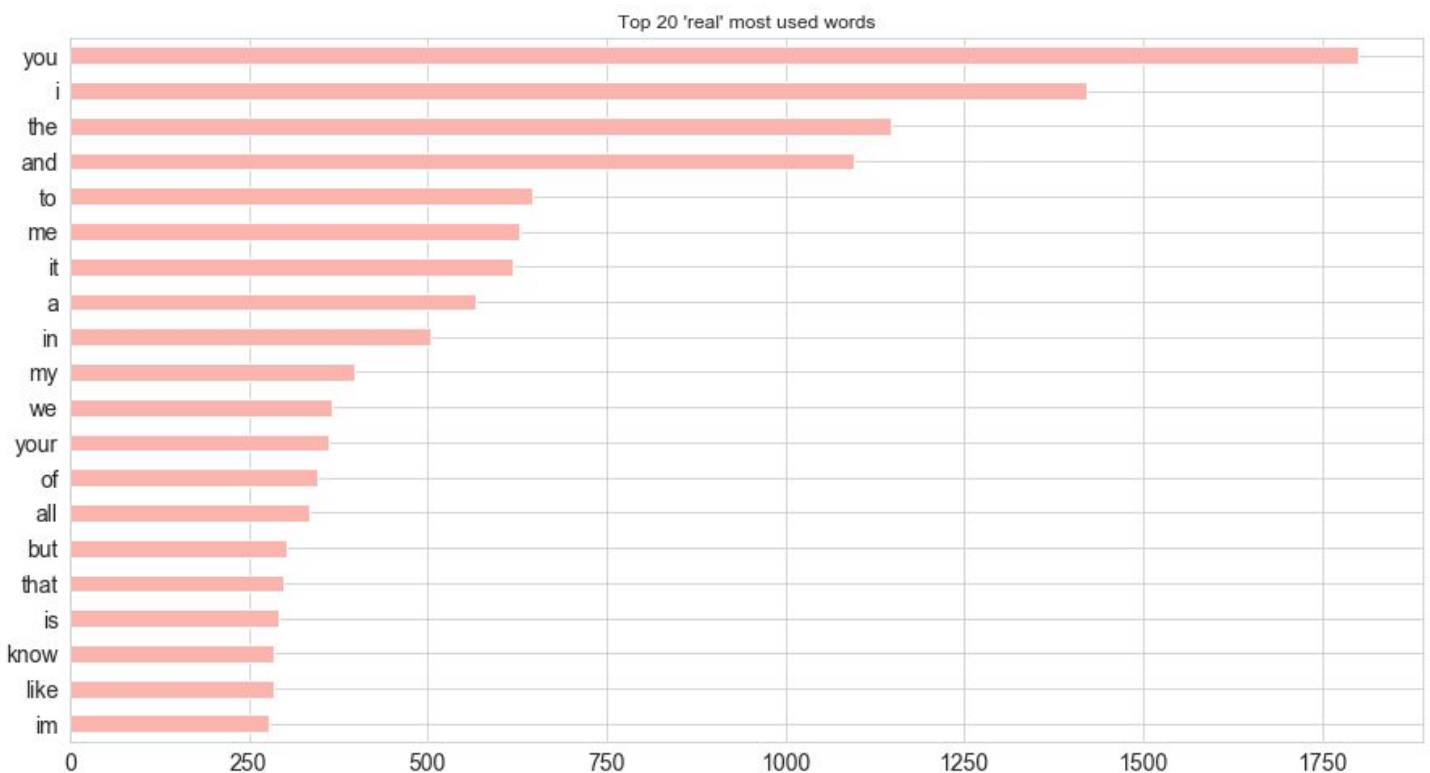
MOST POPULAR WORDS USED

I then made a word cloud to show which words Taylor Swift uses the most.



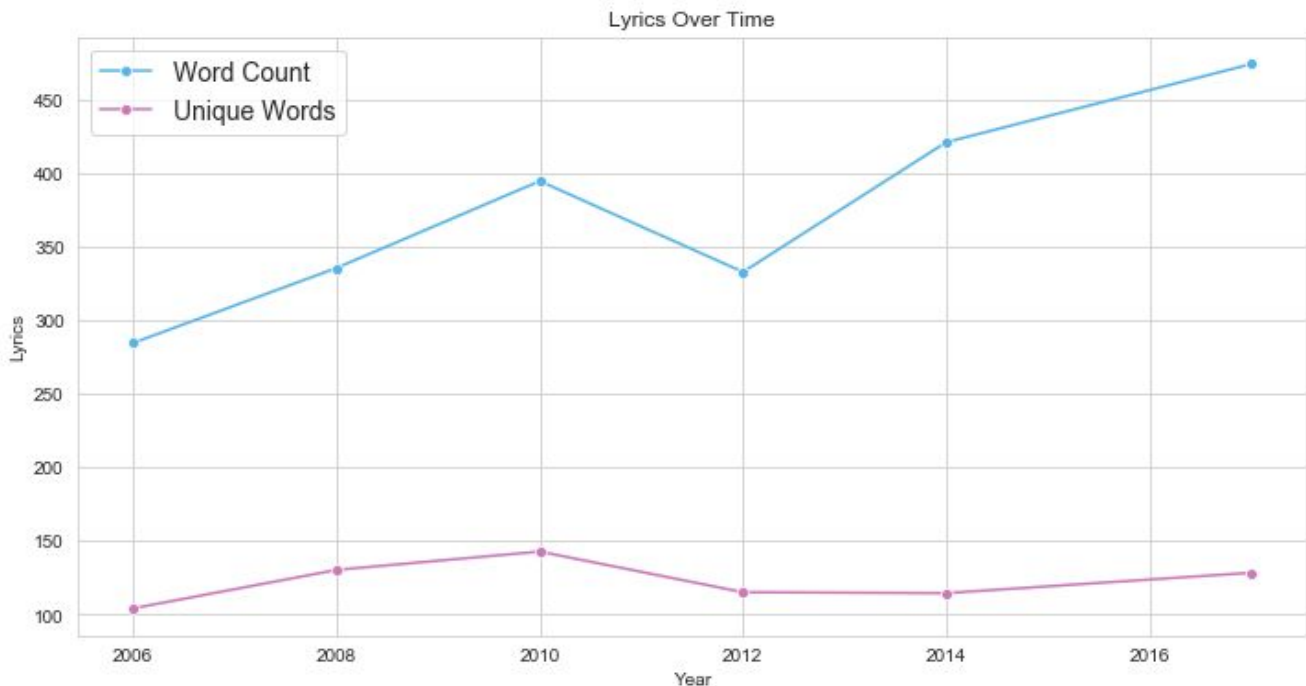
It is worth mentioning that this word cloud has a list of stopwords, common words that it filters out. This is helpful for when you are trying to get a sense of what the songs are about, instead of purely looking at which word appeared the most.

And, speaking of words that appeared the most, that's what I did next, in a more literal sense. The results were unsurprising, as it is mostly common english words that don't really give you a sense of the music.



LYRICS OVER TIME

The last visualization I made was to see the number of lyrics, unique and otherwise, of the years of the albums' releases.



I thought this quite an interesting one, especially noting the difference in growth between the total word count and the unique words used in the songs. When Taylor Swift began her music career her style was more inline with country music, and as she progressed she became more in-tune with pop music. Generally, pop music is known for having catchy lyrics which are repeated throughout the songs, which is apparent here in this graph, where even though the word count goes up, the unique words are more or less the same.

Another interesting point is that even though the '*Red*' album came out in 2012, containing a song that had the second-most unique lyrics of all the songs, and the album that had the most tracks (19 tracks), there is a clear dip in total lyrics used for that album.

THIS REPORT WAS WRITTEN BY : CALVIN LIANG
