

Project Proposal

Song Lyrics Analyzer

Valentine Sénégas

CART 451 - Networked Media Studio
Presented to Sabine Rosenberg
October 4th 2022

Project Description

Music is probably the most essential artistic means of self-expression because it is often used in other artistic disciplines. Today, we can hardly imagine a new movie coming out without sound or music. The same goes with video games, advertisement, or any other experience from the entertainment industry. Music has a very big influence on people's lives. What we listen to can define our social circles, help shape our opinions, open our eyes on issues that we don't know about or in the contrary, relate to an artist for faces the same hardships as us. Music has a great power on the mind: it can help us be more focused, relax, even motivated.

This project aims to analyze song's lyrics from more than 40 major artists from the 20th and the 21st century. The main goal is to compare word similarities or differences between all the artists. A graphic representation will show the relationships by creating a set of words for each artist, which will be linked to others. Clusters that are closer to each other are more similar than ones that are far apart. Then, the data will be compared between artists to understand how they are similar.

Research Questions

Do artists who are in the same musical category use the same kind of vocabulary? When put apart and not in sentences, do these words mean something on their own? We can think of the song Hotel California, which is about drug use, but does not clearly mention the word drug or any drug whatsoever.

With this project, I also want to know more about the **diversity of vocabulary according to each artist**. Do they use a variety of words or are the same words always repeated?

What kind of words are used? Does this reflect the level of depth of the songs? Do the words used reflect a specific social background or belonging to a group?

Data

The data that I will use was collected by Paul Mooney on free lyric hosting websites. The database has no copyright and can be found on Keggle (<https://www.kaggle.com/datasets/paultimothymooney/poetry>). This data must be accurate in the sense that these are the words sung by each artist. It is easily verifiable to the source as all of them sing in English. One potential bias could be the songs that were selected. Also, Paul Mooney selected pop artists, but they may not reflect all the music created by pop artists. It is uncertain whether we should make conclusions and assumptions on all of pop music based only on this small sample.

Inspiration from in-class readings

Hito Steyerl

Hito Steyerl's text highlights the importance of putting data in context. She starts the text by showing an abstract image revealed by Edward Snowden. Taken out of context, this image does mean anything: "Snowden's image of noise could stand in for a more general human inability to perceive technical signals unless they are processed and translated accordingly". In a similar way, my project will extract words from songs, exactly as they were written, and try to make analogies and conclusions from these individual words. My project will not consider groups of words or expressions. Although the data will be accurate strictly considering the words, the meaning behind the songs is not reflected in the end result.

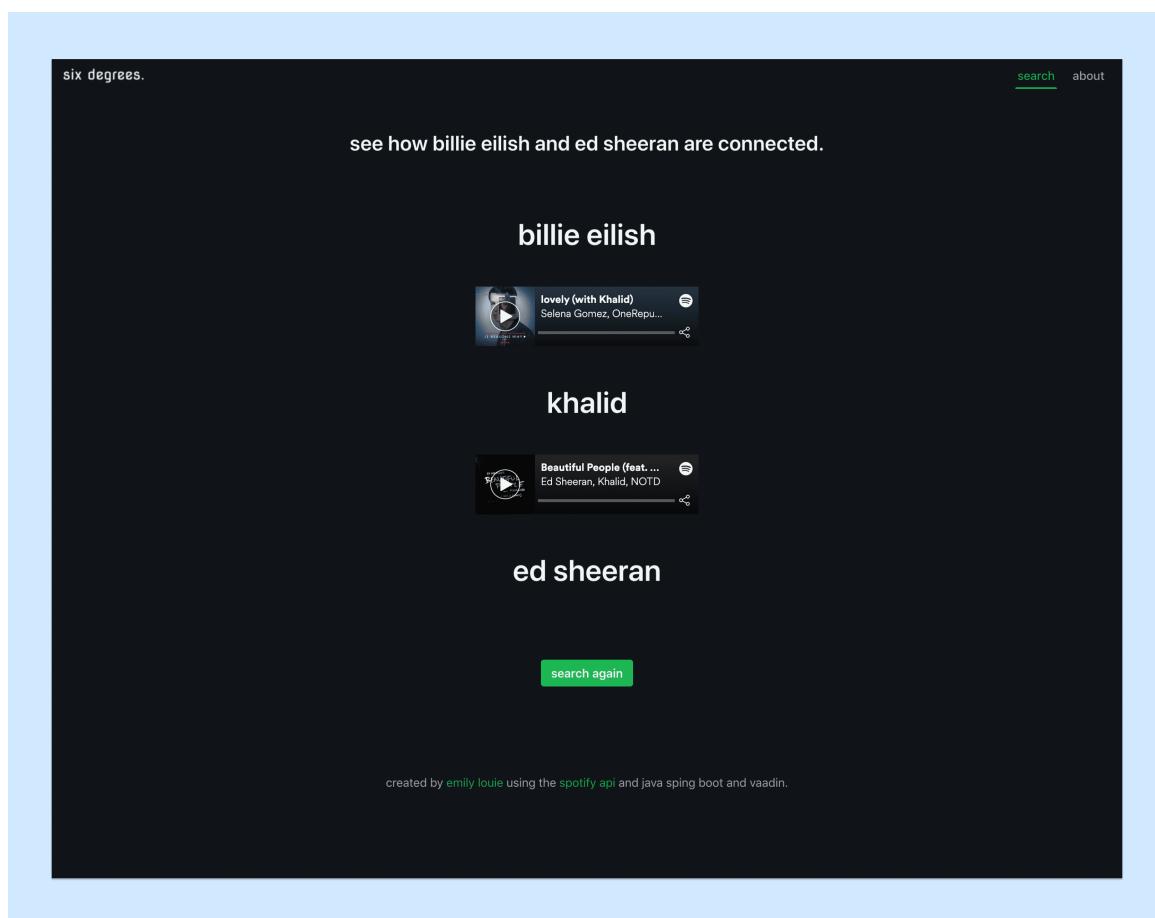
Mimi Onuoha

At first, I thought that my dataset could not be biased because the lyrics of the songs can easily be verified. Then, I realized that the artists and songs in this dataset are only a small fraction of all pop music. My project could demonstrate that pop artists use a lot of swear words, that there is no meaning behind the songs and portray a negative image of the genre. The algorithm will not be able to make sense of all the words together and analyze them individually, which could distort the intended meaning behind the songs.

Other Inspiration

Six Degrees of Spotify

A very interesting project made using a lot of data is Six Degrees of Spotify by Emily Louie. In 2020, she created a website that relates the theory of six degrees of separation to music. This theory is that every person in the world is connected to anyone in less than 6 connections. Using the Spotify API, she was able to determine how two artists were connected to each other by looking at their collaborations on songs.



Other Inspiration

Music-map.com

Music-map.com is a website illustrating the connections between different music artists. The user inputs the name of an artist, and a representation appears, showing similar music artists. The user can then click on each artist to see artists similar to this one. I plan on using a similar system to visualize information.



Visual Storyboard

The Song Lyrics Analyzer

Are you curious about the words
used in most pop songs?

Show me!

Data based on influential music artists of the 20th and 21st century.

[See the dataset](#)

Home page of the website.

The link *See the dataset* takes the viewer to the original dataset for transparency purposes.



This is the graphic representation of all the words sung by the analyzed artists. The ones that are closer sing similar words.

Visual Storyboard

Kanye West

The most used word of Kanye West is

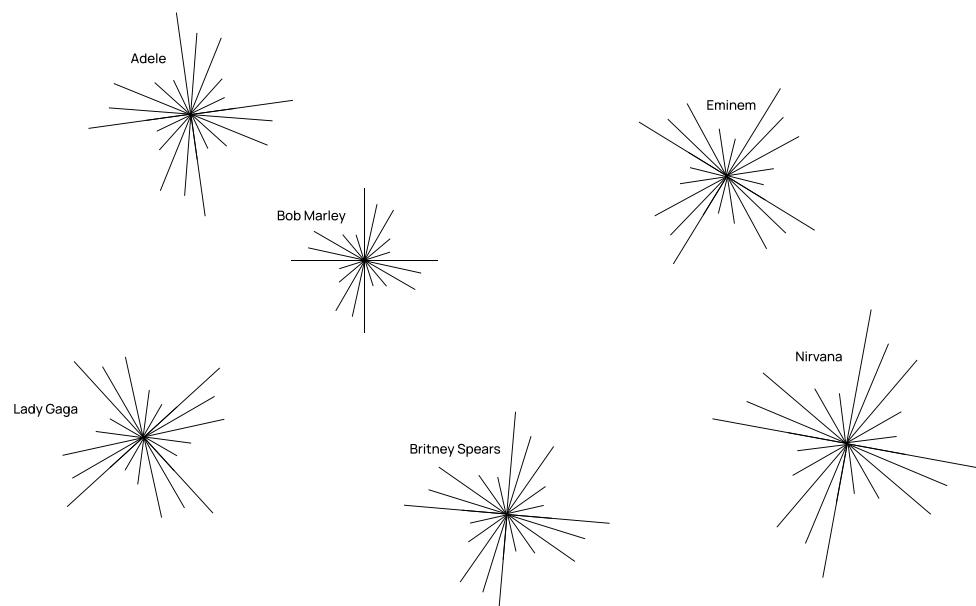
nigga

Found 128 times in our database.

This artist uses words similar to

Drake Lil Wayne

When clicking on an artist, there is a description page with more precise information.



Idea of visual representation where each cluster represents an artist, each line represents a word, and the length of the line represents the number of times the word was found for the artist. Each word could be color coded.