

# Project Report Paper

Prepared by:

Adrian Dominic P. Danao

Gabriel K. De Jesus

Carlo D. Mendoza

Leo Isaiah P. Yap III

Submitted on: December 8, 2017

**CS 129.1 B Contemporary Database**

### A. Approved Big Data Problem

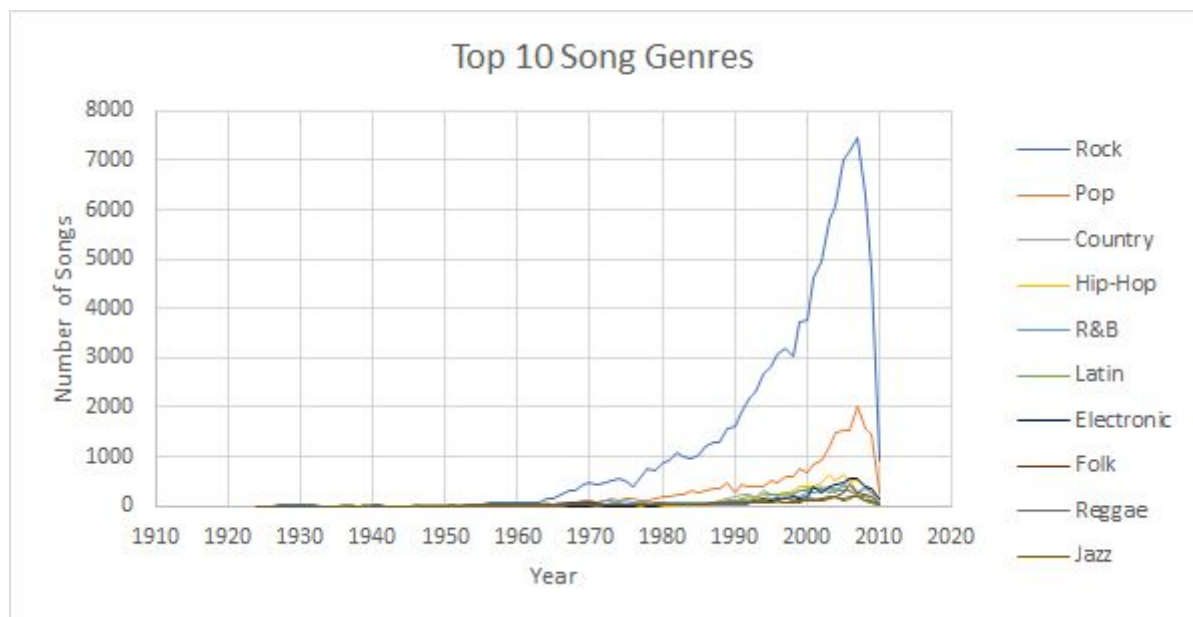
Given a collection of songs how do we determine how many songs belong to each genre and what is the most frequently used word in a given genre per time period?

### B. Description of Source Dataset

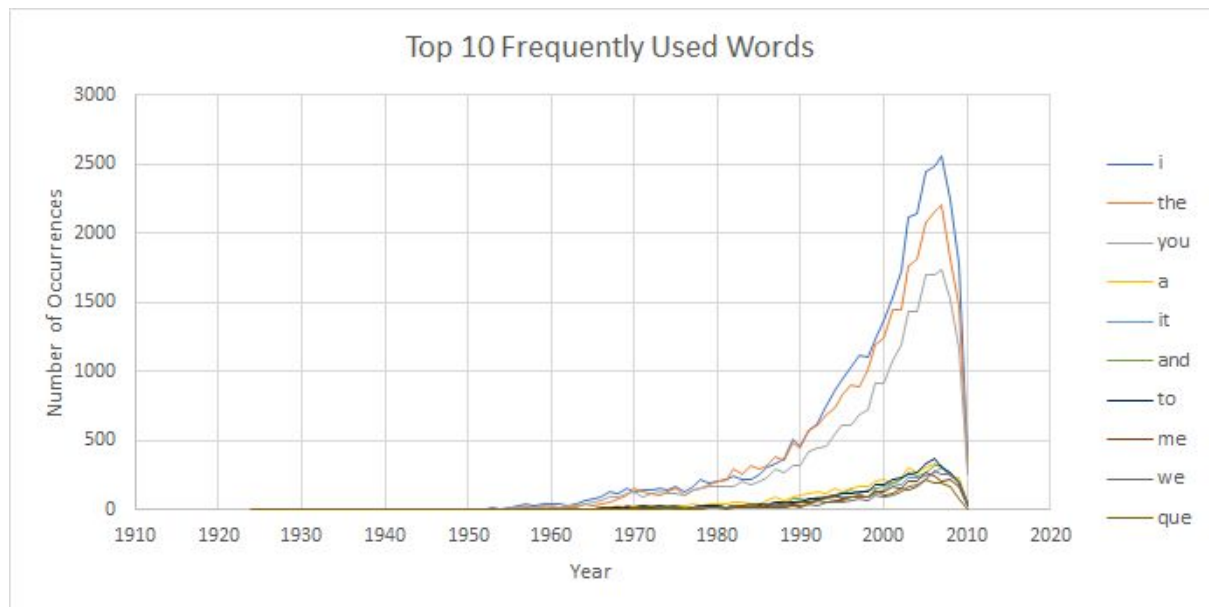
The group used the Million Song Dataset as the source of data. It is an open source collection of audio features and metadata for a million contemporary music tracks. It also has complementary datasets which provide other important features of each track such as the genre and the most used word in each track.

The group used only a subset of 189, 481 songs for the dataset. The year of release<sup>[1]</sup>, genre<sup>[2]</sup>, and most used word<sup>[3]</sup> of each track were retrieved from separate sources so they needed to be joined in one collection.

### C. Visualization of Output



**Figure 1.** A line chart showing the top 10 song genres per year from 1924 to 2010.



**Figure 2.** A line chart showing the top 10 frequently used words per year from 1924 to 2010.

#### D. Discussion of Output

The group only got the top 10 genres and top 10 frequently used words per year for better visualization of the output.

The group noticed a huge spike in Rock songs towards the 21st century (as shown in Figure 1). This may have been the result of other song genres having less valid songs during the database parsing, or that more Rock songs were indexed by the source database.

The visualization of the top 10 frequently used words, and top 10 song genres, both spike up towards the end of the graph. Because of the abundance of Rock songs, words like *I*, *the*, and *you*, also rose in number (as shown in Figure 2). Had a different genre of song been the most popular, such as Jazz, then the top 10 frequently used words may have also been different.

#### E. External Links

[1] [https://labrosa.ee.columbia.edu/millionsong/sites/default/files/Additional Files/tracks\\_per\\_year.txt](https://labrosa.ee.columbia.edu/millionsong/sites/default/files/Additional%20Files/tracks_per_year.txt)

[2] [http://www.tagtraum.com/msd\\_genre\\_datasets.html](http://www.tagtraum.com/msd_genre_datasets.html)

[3] <https://labrosa.ee.columbia.edu/millionsong/musixmatch>