Project Proposal NLP Unsupervised

Abstract:

A revolution in music production, distribution and consumption made millions of songs available to virtually anyone on the planet, through the Internet. To allow users to retrieve the desired content from this nearly infinite pool of possibilities, algorithms for automatic music indexing and recommendation are a must.

Question/need:

I will use Natural Language Processing to create a content-based recommendation system.

Data Description:

The Million Song Dataset, a collaborative project between The Echo Nest and LabROSA, is a freely available collection of audio features and meta data for a million contemporary popular music tracks.

Tools:

I will use python because it contains libraries for Machine Learning also for exploratory analysis and data cleaning will use (pandas, Matplotlib, NumPy ,etc.).