*The idea is quite original. Good presentation. The project has an important descriptive part. Good potential for BI. Not a lot of detail on the dataset is provided. Also, little information on the quality of the data is provided. There is a clear definition of the business problem.*

The data view has to summarise relevant information of the target dataset of the application.

The analytics view has provide relevant analytics constructed from the data.

1. Milestone II. Week 6 (Oct 26) Application

By this date each group has to define the deliverables of the dashboard. Each group has to submit a one page (max) report describing what the “data” and “analytics” views of the dashboard are going to do.

**Milestone II - Computing Lab and Data Warehousing and BI Project**

By: Niti Mishra, Miquel Torrens and Bálint Ván

Data view

For our project we are using data from the Million Song Dataset. Due to resource limitations we are currently working on a subset of 10,000 songs. This data is collected from different sources related to online music services, and has an extensive set of variables. In the data view of the dashboard we are going to display the characteristics of groups of variables which share similar features. Detailed information about the meaning of each variable will be available in dropdown lists. We will include only those variables which are used in each analysis. Here we present the description of the main groups and a few example variables:

**Musical summary variables**

They show technical characteristics of the songs calculated out of the sound files. They are exhaustive at song level and complete for all the songs. Some examples of the most relevant variables: duration, energy, key, mode, loudness, tempo, danceability, among many others.

**Metadata**

These are obtained from various sources. Many observations have missing values on some of the variables. These data include artist name, year of release, location of the artist, genre of the song, artist related terms (tags), related artists, or song and artist “hotness”, among others.

**User data**

Complete play counts on each song broken down by user for more than one million of anonymous users of The Echo Nest.

**Lyrics**

Lyrics are represented in a bag of words format. The 5,000 most important words were selected when creating the data set. We know how many times these words occurred in each song.

Analytics view

We will address different questions in the analytics part of the project and the dashboard will be partitioned in different tabs according to these questions, which will at the same time be related. We will use different subsets of the data to address each question. In the final version of the project we might not include all of these questions, depending on time and resource constraints.

1. **Genre. How different are genres in their musical variables or lyrics?**
2. **Time. How have songs evolved over decades** in their structure or lyrics?
3. **Lyrics. What are the characteristics of the lyrics of the most popular songs?** How are they different from the others? **Does the occurrence of certain words in a song imply its musical characteristics?**
4. **Popularity. What are the key features to explain the popularity of a song?**
5. **Origin. How different is the music produced in different places?**
6. **Recommender. We will try to find closest neighbors to songs or bands to suggest likely matches to users’ selections.**