**Executive Summary**

**Authors: Ashish Khare, Tejas Magar, Prakhar Vashisht, Shubham Malpani, Devan Kumar, Vikram Rao**

**Introduction:**

The **ECHONEST SONG DATABASE** is a freely-available collection of audio features and metadata for a million contemporary popular music tracks.

**Objective:**

* On the basis of some of the variables such as tempo, loudness, familiarity, betas confidence and other facts we are going to decide the song hotness level for the future songs as per year.

**Purpose**:

* To encourage research on algorithms that scale to commercial sizes
* To provide a reference dataset for evaluating research
* As a shortcut alternative to creating a large dataset with APIs

Below is the description of each data fields: -

|  |  |  |
| --- | --- | --- |
| Field name | Type | Description |
| analysis sample rate | float | sample rate of the audio used |
| artist 7digitalid | int | ID from 7digital.com or -1 |
| artist familiarity | float | algorithmic estimation |
| artist hotttnesss | float | algorithmic estimation |
| artist id | string | Echo Nest ID |
| artist latitude | float | latitude |
| artist location | string | location name |
| artist longitude | float | longitude |
| artist mbid | string | ID from musicbrainz.org |
| artist mbtags | array string | tags from musicbrainz.org |
| artist mbtags count | array int | tag counts for musicbrainz tags |
| artist name | string | artist name |
| artist playmeid | int | ID from playme.com, or -1 |
| artist terms | array string | Echo Nest tags |
| artist terms freq | array float | Echo Nest tags freqs |
| artist terms weight | array float | Echo Nest tags weight |
| audio md5 | string | audio hash code |
| bars confidence | array float | confidence measure |
| bars start | array float | beginning of bars, usually on a beat |
| beats confidence | array float | confidence measure |
| beats start | array float | result of beat tracking |
| danceability | float | algorithmic estimation |
| duration | float | in seconds |
| end of fade in | float | seconds at the beginning of the song |
| energy | float | energy from listener point of view |
| key | int | key the song is in |
| key confidence | float | confidence measure |
| loudness | float | overall loudness in dB |
| mode | int | major or minor |
| mode confidence | float | confidence measure |
| release | string | album name |
| release 7digitalid | int | ID from 7digital.com or -1 |
| sections confidence | array float | confidence measure |
| sections start | array float | largest grouping in a song, e.g. verse |
| segments confidence | array float | confidence measure |
| segments loudness max | array float | max dB value |
| segments loudness max time | array float | time of max dB value, i.e. end of attack |
| segments loudness max start | array float | dB value at onset |
| segments pitches | 2D array float | chroma feature, one value per note |
| segments start | array float | musical events, ~ note onsets |
| segments timbre | 2D array float | texture features (MFCC+PCA-like) |
| similar artists | array string | Echo Nest artist IDs (sim. algo. unpublished) |
| song hotttnesss | float | algorithmic estimation |
| song id | string | Echo Nest song ID |
| start of fade out | float | time in sec |
| tatums confidence | array float | confidence measure |
| tatums start | array float | smallest rhythmic element |
| tempo | float | estimated tempo in BPM |
| time signature | int | estimate of number of beats per bar, e.g. 4 |
| time signature confidence | float | confidence measure |
| title | string | song title |
| track id | string | Echo Nest track ID |
| track 7digitalid | int | ID from 7digital.com or -1 |
| year | int | song release year from MusicBrainz or 0 |

**Personas**:

In order to understand the objectives better we created 3 personas from the music industry and interviewed them to interpret their challenges.

|  |  |
| --- | --- |
| Kriztofer Olan  Music Producer  Los Angeles, California | Age: 45  Education: BS in Music, UC Berkeley.  Challenges:   * Create a final master of a song with optimal likeness. * Able to advise arrangements & methods to improve song popularity * Compose songs which would break records * Trends in the songs that are liked most. |
| Samantha Simson  Vocalist/Pianist  Louisiana, Missouri | Age: 33  Education: BS in Music, concentration in vocal performance.  Challenges:   * Compose & Write songs that would be big hit * To understand what pitch must be used to increase the feel of the song |
| DJ Limbu  Professional Disk Jockey  Miami, Florida | Age: 28  Education: High school  Challenges:   * To understand what tempo must be used to increase the danceability of a song * Able to understand hotness of the song |

**Let us understand each objective from the point of view of the personas:** -

**Kriztofer Olan, Music Producer**

The dataset results will help Kriztofer to understand what kind out of songs are in demand in the market i.e. songs which are more likely to be successful. So that he could produce those kinds of songs with the respective artist and earn more fame in the future.

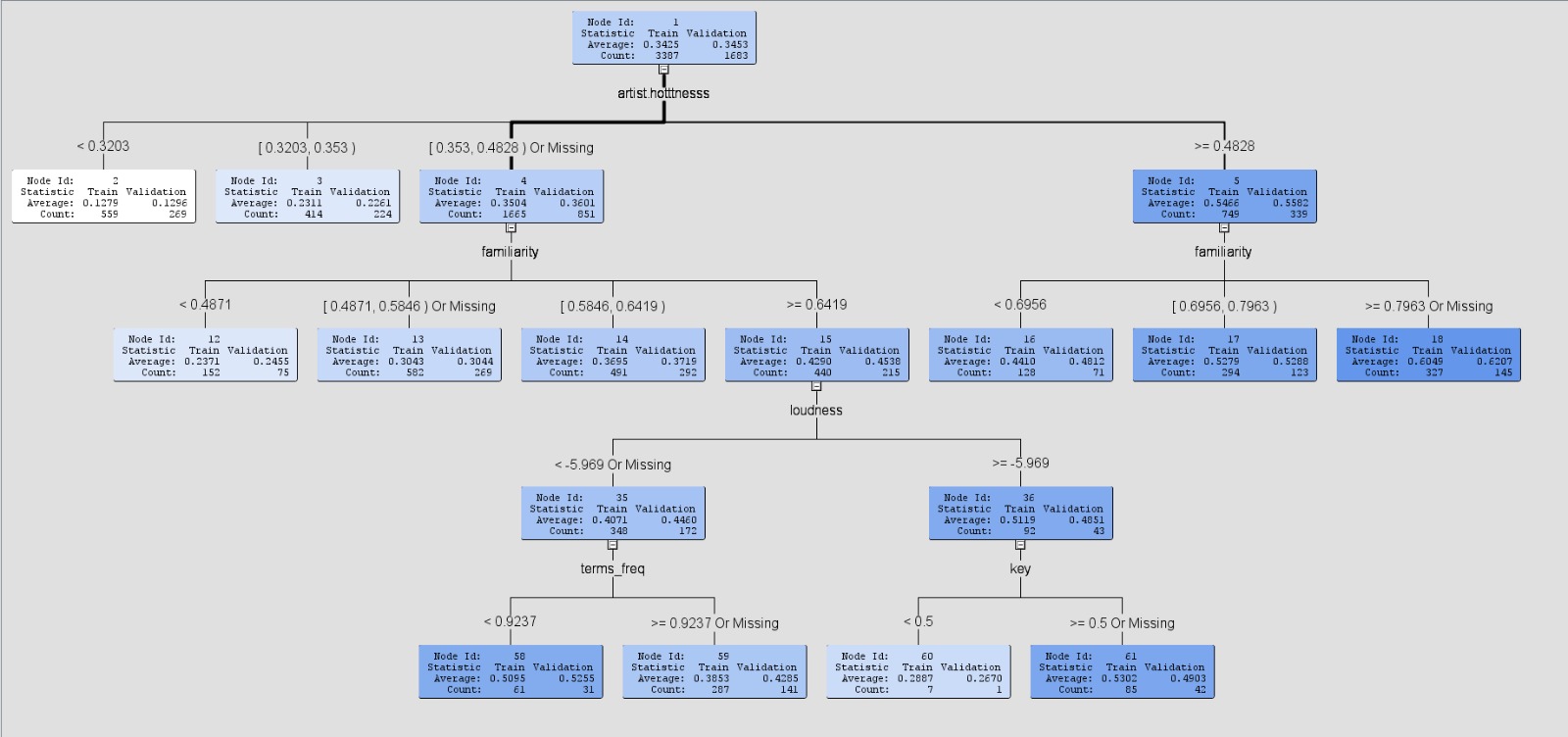
**Samantha Simson, Vocalist/Pianist**

The dataset results will help Samantha to compose songs which could attract music companies and also to make songs of particular beats and rhythm.

**DJ Limbu, Professional Disk Jockey**

The dataset results will help DJ Limbu to make and select songs which are popular among his listening audience.

We used decision tree to find the factors affecting the hotness of a song



**Findings from the dataset: -**

The hotness of a song is dependent mainly on the following factors: -

* Familiarity of the artist
* Artist’s hotness
* Loudness of the song

So, there are higher chances for a song to be successful is these factors are implemented correctly.

**How the personas benefit from the findings**: -

**Kriztofer Olan, Music Producer**

As a music producer with the help of the data results Kriztofer should take artists which are well known to the audience and have a good fan base. Also Kriztofer should focus on songs which are of higher frequency rather than mellow songs. If these factors are handled by Kriztofer as a music director there will be a more likelihood of the song to be a hit in the market.

**Samantha Simson, Vocalist/Pianist**

As a vocalist based on the dataset results Samantha should sing songs which have many beats and tempo. This would lead to increase her fame and make her to come into notice in the eyes of the audience. Eventually, it would lead to more music produces like Kriztofer signing upcoming and famous singers like her and producing more songs in the future.

**DJ Limbu, Professional Disk Jockey**

Being a DJ, the findings from the dataset would help DJ Limbu to understand which songs he could change the beats of to make in more popular among the club/parties visiting audience.

We as a group believe that the analysis of this dataset would provide more insight into prediction of song hotness not just for the above-mentioned personas but shed more light to answering more research questions in the music industry.