2/12/2018

MTECH KE-5107 (DMMM)

PCA/FACTOR/CLUSTER ANALYSIS REPORT

FMA MUSIC ANALYSIS

SUBMITTED TO

INSITITUTE OF SYSTEM SCIENCE
NATIONAL UNIVERSITY OF SINGAPORE

PREPARED BY

SAURABH SEMWAL
VIKNESHKUMAR BALAKRISHNAN
ANJALI SINHA
GOPALAKRISHNAN SAISUBRAMANIAM

MASTER OF TECHNOLOGY IN KNOWLEDGE ENGINEERING
BATCH KE-30(2018)

1. Abstract

Principal Component Analysis (PCA) is a popular statistics technique used primarily in the field of data mining for **dimensionality reduction**. It uses an **orthogonal transformation** to reduce a set of possibly correlated variables in a multivariate dataset into a set of linearly combination of the original variables, referred to as **Principal Components**. This report serves to showcase the usefulness of PCA by applying it on a real world dataset - **FMA (Free Music Archive)**. The rotated loading matrix from the technique (varimax rotated PCA) is used to provide domain relevant names to orthogonal variables. The following independent sub-tasks are also considered after generation of the principal components: i. Using a **clustering technique** with the reduced feature set to **profile** music tracks, ii. Performing **nominal regression analysis** using the reduced number of orthogonal variables and analyzing the prediction quality using standard metrics. All related experiments and analysis are conducted using the software - JMP Pro 13.

2. Dataset

a) Data Source

The dataset under consideration is a dump of the **Free Music Archive (FMA)**, an interactive library of high-quality, legal audio downloads [1] The dataset was created by a group of researchers at EPFL, Switzerland.[2]

b) Subset Selection

The full version of the dataset has ~13k observations and has information suitable for multi-label prediction, but for the purpose of this assignment a subset was constructed specifically for the simpler problem of single-label prediction by careful selection of higher quality tracks that have a single root genre and have meta-data completeness. This subset shall be henceforth referred to simply as the *dataset* in this document.

c) Data Description

The original dataset contains 781 variables and 13129 observations.

Index	Column Number	Variable Name	Type	Description
A	1	Track id	Continuous	Unique identifier of the song
B	2-13	Social & Audio features	Continuous	Scores of songs on the basis of various features such as valence, liveness, danceability, tempo, acousticness, energy, and instrumentalness. Scores of artists on basis of discovery, familiarity and hotness.
C	14-237	Temporal Echonest features[4]	Continuous	It consists calculations of statistical moments of various segments such as pitch, timbre, loudness, derived using Echonest API [5]

D	238-241	Album attributes	Various	Consists of album type, album listens, album tracks and album active days
E	242-244	Artist info	Various	Consists of artist id, artist location
F	245-246	Set info	Categorical	Column for dataset split and subset.
G	247-263	Track info	Continuous	Track details such as genre, duration, days active, number of time it was selected as favorite, number of listens.
H	264-515	Chroma calculations	Continuous	Chroma features represent audio by projecting the entire spectrum onto 12 bins representing the 12 distinct semitones of the musical octave. Different representations of chroma values using various techniques are derived using Librosa package[3]
I	516-655	Mfcc stats	Continuous	Features representing speech in compact form.[6]
J	656-662	Rmse stats	Continuous	The RMS level is proportional to the amount of energy over a period of time in the signal.
K	663-774	Tonnetz stats	Continuous	A planar representation of pitch relations showing harmonic relationships in European classical music.[7]
L	775-781	Zcr stats	Continuous	Rate at which the signal changes from positive to negative or back.

[Table 1: Data Description]

3. Preliminary Exploratory Analysis

i. Cleaning and Treatment of Missing Values

ii. Transformations

- track_num_days_active: The difference between the date of track creation and current date.
- artist_num_days_active: The difference between the start of artist career and current date.
-

iii. Final Predictors

Out of the original 781 variables, we identified and removed those that exhibit near-zero variance and whose p-values were insignificant (p-value > 0.05). We finally ended up with 40 input predictors, and the track_genre as the response variable.

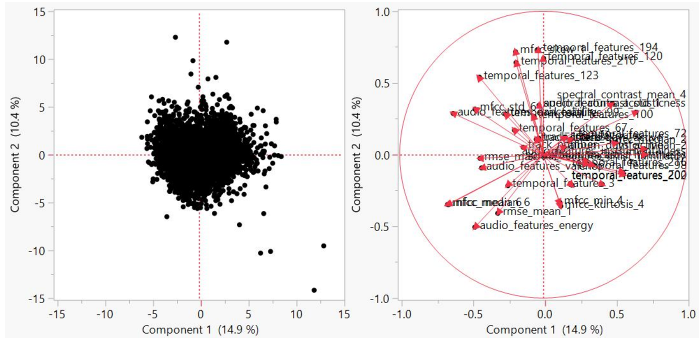
4. Dimensionality Reduction

i. Generation of Principal Components

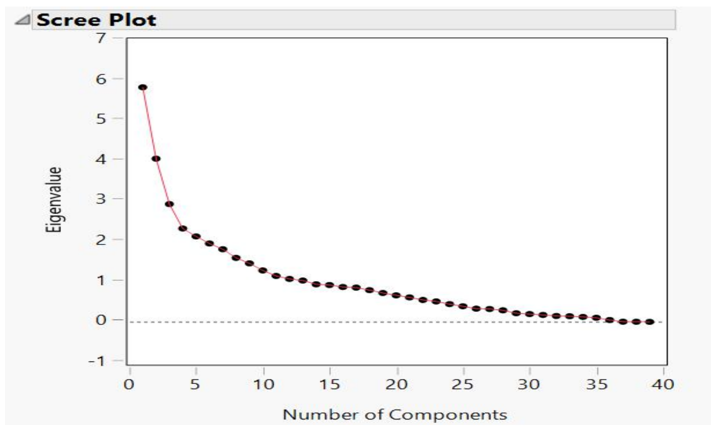
In JMP, we used Principal Component Analysis (under Multivariate option) and generate the orthogonal components. The below figures represent the variance explained by each component (Fig 4-1) and the orthogonal vectors representation. (Fig 4-2)

Number	Eigenvalue	Percent	Cum Percent
1	5.8224	14.929	14.929
2	4.0517	10.389	25.318
3	2.9262	7.503	32.821
4	2.3207	5.950	38.772
5	2.1245	5.447	44.219
6	1.9494	4.998	49.217
7	1.8065	4.632	53.849
8	1.5924	4.083	57.933
9	1.4566	3.735	61.667
10	1.2805	3.283	64.951
11	1.1440	2.933	67.884
12	1.0703	2.744	70.629
13	1.0305	2.642	73.271
14	0.9363	2.401	75.672
15	0.9194	2.357	78.029
16	0.8700	2.231	80.260
17	0.8553	2.193	82.453
18	0.7923	2.032	84.484
19	0.7214	1.850	86.334
20	0.6632	1.700	88.035

[Figure 4-1: Principal Components Eigen Value]



[Figure 4-2: Orthogonal Components]



[Figure 4-3: Scree Plot]

As per Kaiser's criterion, the eigen values over 1 are considered stable. For eigen values >=1, 73.271% of the data is retained (13 components) but in order to slightly improve the retention, we included the 14th component also. (75.672%)

ii. Naming the Orthogonal Variables from the Rotated Loading Matrix

The loading matrix is rotated for the sake of interpretation of extracted components in PCA.

Rotated Factor Loading

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7	Factor 8	Factor 9	Factor 10	Factor 11	Factor 12	Factor 13	Factor 14
temporal_features_120	0.968601	0.016322	-0.023928	-0.001228	0.035894	0.040776	0.027832	0.008562	-0.033912	-0.044970	-0.134279	-0.039160	-0.012117	-0.035832
temporal_features_194	0.913638	0.004985	-0.044923	-0.038609	0.056660	0.192074	0.071231	0.000246	-0.037821	-0.039641	-0.130734	-0.021376	0.023906	0.011805
temporal_features_210	0.895891	0.016088	0.001575	-0.145773	-0.004934	0.046013	-0.002406	-0.012723	0.035894	-0.129013	0.132930	0.025044	0.014881	-0.038674
temporal_features_123	0.279733	0.008108	-0.018998	-0.311868	0.053224	0.489827	0.062527	-0.097406	0.130547	-0.198767	0.390868	0.009613	0.165870	0.152806
mfcc_skew_1	0.258222	-0.378806	-0.092631	-0.044072	0.237402	0.657476	0.034889	-0.058807	-0.033668	-0.165137	-0.063585	-0.020213	0.121049	0.078970
spectral_contrast_std_1	0.176514	-0.083333	-0.240907	-0.276578	0.196560	-0.059733	-0.002679	0.085332	0.146544	-0.472168	-0.053268	0.091518	-0.079865	-0.052604
temporal_features_216	0.096719	0.848119	-0.108457	0.062855	0.156640	-0.145880	-0.005186	0.064121	-0.058039	0.075758	-0.230342	0.109075	0.030487	-0.026374
rmse_max_1	0.083067	-0.060857	0.050141	-0.122475	-0.110801	0.121553	0.042335	-0.023075	0.857962	-0.064951	0.055435	-0.006310	-0.042932	-0.046597
temporal_features_67	0.079279	-0.227862	-0.008483	-0.032145	-0.065980	0.168106	-0.043542	0.052615	0.232425	-0.042871	-0.076301	-0.342162	0.288834	-0.132494
temporal_features_99	0.063884	-0.094903	-0.164778	-0.104707	-0.181660	0.161773	0.026379	-0.023400	-0.024726	-0.482661	0.520172	0.461721	-0.063501	-0.016408
audio_features_acousticness	0.061150	0.040450	-0.474049	0.398631	0.367205	-0.103800	-0.007063	-0.034504	-0.179073	0.049777	0.012481	-0.132781	0.357674	0.085306
track_interest	0.055223	-0.011562	0.003312	-0.008780	-0.034671	-0.002303	0.949385	-0.029816	-0.015680	0.049286	0.028745	-0.021207	0.013329	-0.034637
mfcc_std_6	0.055031	-0.086048	0.222531	-0.045406	-0.135969	0.384303	-0.023711	-0.151374	0.022499	-0.558690	0.075474	-0.394448	0.028796	0.002538
track_listens	0.039132	-0.014407	0.005958	-0.022228	-0.027213	0.019221	0.945058	-0.014041	0.024639	0.043312	0.030163	0.002837	-0.008628	-0.045219
temporal_features_72	0.019970	0.046918	-0.082727	0.264275	0.601231	-0.192872	0.002528	0.055730	-0.106361	0.127674	-0.003637	-0.037981	-0.049697	-0.034825
temporal_features_200	0.019454	0.957695	-0.081902	0.023069	-0.001098	-0.114556	-0.006843	0.013372	-0.009341	0.032962	-0.115542	-0.043548	0.038835	0.001326
spectral_contrast_mean_4	0.018022	0.007049	-0.238450	0.194361	0.700794	-0.020335	0.057782	0.005741	-0.139232	-0.162693	0.007545	-0.068202	-0.029105	-0.000172
track_days_active	0.014001	0.072743	-0.043139	0.093857	-0.000245	-0.048945	-0.013351	-0.048914	-0.120406	-0.048589	0.097785	0.135965	0.807774	-0.008260
temporal_features_100	0.009883	0.002255	-0.468606	-0.369779	0.097045	0.032420	0.015119	0.006726	-0.094135	-0.114442	0.117680	-0.470873	-0.300583	0.075834
temporal_features_3	0.005775	0.082907	-0.012660	0.053194	-0.714711	-0.092191	0.026640	-0.075834	-0.024318	-0.077470	0.097032	-0.108936	0.037610	0.076008
social_features_artist_familiarity	-0.000238	0.032430	-0.062702	0.096519	0.028980	-0.085417	0.008914	0.930574	-0.064284	0.046267	-0.004284	0.026001	0.068661	0.004490
social_features_artist_hotttnesss	-0.003439	0.035099	-0.056609	0.046083	0.063424	-0.058753	0.060521	0.936831	0.005798	0.007286	-0.060235	-0.011692	-0.070251	-0.046121
audio_features_instrumentalness	-0.008874	0.016104	-0.110189	0.167462	0.015639	-0.028666	0.035707	0.029322	0.070209	0.026317	-0.097689	0.801345	0.165755	-0.080307
artist_favorites	-0.010955	0.022409	-0.062717	0.074202	0.144080	0.022911	0.483191	0.260477	0.122835	-0.203865	-0.093247	0.118418	-0.044459	0.097584
audio_features_danceability	-0.013464	-0.197945	0.101250	-0.299342	-0.044975	0.512856	0.042910	-0.062553	0.126498	-0.100117	0.554265	-0.010606	-0.013424	0.102951
album_days_active	-0.023471	-0.003145	-0.040814	-0.110649	-0.072347	-0.004297	-0.007900	0.157061	-0.287050	-0.073299	-0.115412	0.100622	0.416714	-0.430345
track_number	-0.026992	-0.044661	0.018133	-0.145779	-0.145178	0.083457	-0.018315	-0.090276	-0.131904	-0.095377	-0.162488	0.094307	0.005481	0.744493
mfcc_median_6	-0.027815	-0.099330	0.903526	-0.230679	-0.134338	-0.037444	-0.016237	-0.074693	0.092031	0.028710	0.066634	-0.097269	-0.047102	0.001690
temporal_features_98	-0.030729	0.149490	-0.312371	0.427798	-0.069298	-0.549931	-0.040562	0.068752	-0.224846	-0.336131	0.008858	0.081596	-0.127366	-0.069125
mfcc_mean_6	-0.032371	-0.098943	0.904211	-0.221812	-0.121660	-0.025439	-0.017332	-0.076615	0.088260	0.056823	0.078498	-0.091289	-0.039450	0.006614
temporal_features_202	-0.041862	0.909030	-0.020680	-0.092259	-0.083525	0.014817	-0.002005	0.008646	0.020876	0.011729	0.040828	0.009163	0.012807	-0.029140
album_tracks	-0.043592	-0.011772	-0.057917	0.114401	0.054801	-0.051890	-0.014010	0.117247	0.023580	0.114001	0.303996	-0.154846	-0.014291	0.674798
mfcc_min_4	-0.098533	0.040968	0.031781	-0.071133	0.086952	-0.038308	-0.014622	0.031191	-0.019082	0.830641	-0.016921	0.042487	-0.098174	0.002626
mfcc_median_2	-0.099921	-0.033243	-0.250258	0.843840	0.229944	-0.005762	0.016009	0.089792	-0.036885	0.011120	-0.113208	0.135337	0.047043	0.011414

Factor 1: This component is highly loaded with 3 of the temporal features (120,194,210). **Name:** Temporal features

Factor 2: This component represents three temporal features (200,202,216) whose eigen vectors lie along one direction and the vector of the mfcc component(mfcc_skew_1) is in the opposite direction. **Name:** High frequency temporal features

Factor 3: This component is loaded with speech recognition components whose eigen vectors lie along one direction (mfcc_median_6 & mfcc_mean_6) and vector of acousticsness component which is a measure of usage natural musical sounds. Lower the value of the acousticsness, it means usage of electronic instruments rather than natural sounds. **Name:** Measure of Low frequency instrumental sounds

Factor 4: This component is highly dependent low frequency speech components (mfcc_mean_2 & mfcc_mean_median 2) and also the acousticsness which is basically natural sounds (like human voice) **Name:** Measure of Human voice

Factor 5: This component is highly dependent on spectral contrast mean 4 which explains the frequency band of the sound and the acousticsness variable. It inversely varies with the energy variable. **Name:** Measure of Low frequency natural sounds

Factor 6: It is highly loaded with two low frequency speech components of mfcc(mfcc_skew_1 & mfcc_std_6), 1 temporal feature(123) & also highly correlated with danceability of the audio feature. **Name:** Track danceability

Factor 7: This component is highly loaded with track interest, track listens and artist favorites. On the whole, this explains the popularity of a track. **Name:** Track Popularity

Factor 8: This consists of artist hotness and familiarity. **Name:** Artist Popularity

Factor 9: This component highly varied by the values of rmse_max_1 and rmse_mean_1 which is a function of energy component of the audio. **Name:** Track energy

Factor 10: This component is inversely dependent on spectral_contrast_1, temporal feature_99 and mfcc_std_6 which makes it a measure of high frequency range and also directly varied by mfcc min 4. **Name:** Measure of high frequency sounds

Factor 11: This component mostly influenced by n the valence, danceability and the temporal feature_99. Valence explains the emotions of the audio in the scale of sad to happy. **Name:** track emotion

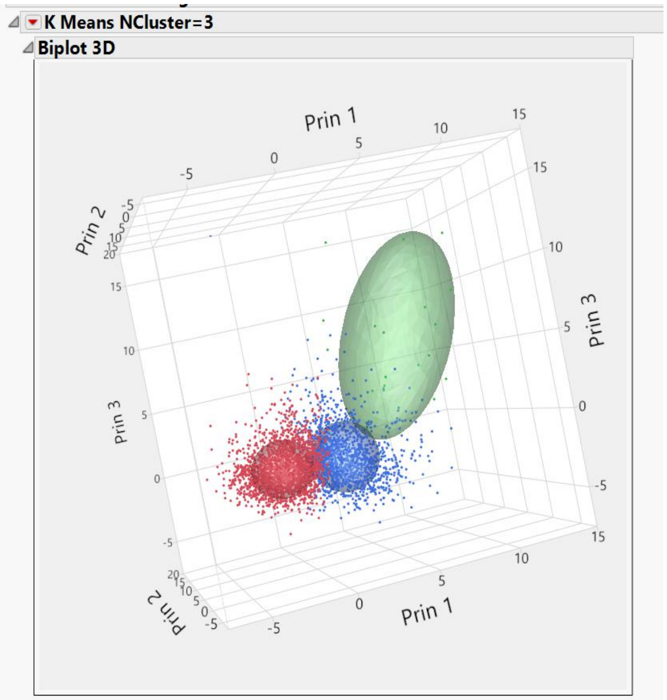
Factor 12: This is directly proportionate to instrumentalness and temporal feature_99 and inversely proportional to low frequency mfcc component (mfcc_std6) and the temporal feature_100. **Name:** Instrumental usage

Factor 13: It is heavily loaded with variables track days active and album days active. **Name:** track lifespan

Factor 14: It is heavily loaded with variables like album tracks, track number and negatively correlated to album days active. **Name:** Album Size

5. Clustering

i. Profile / Description



[Figure-5-1 Clusters]

Cluster 1: The mean of the variables like audio_danceability, audio_energy, track interest and track listens are very high in this cluster. **Profile:** <u>Upbeats</u>

Cluster 2: The mean of the audio_acousticness is high which means tracks which doesnot use electronic musical instruments come under this group. Also, the mean of audio_danceability, energy and valence are low. Low valence signifies the songs revolve around sadness which will have less energy and danceability. **Profile:** <u>Underrated Blues</u>

Cluster 3: This cluster has higher means for artist_familiarity, artist_hotness and artist_favorites. The tracks under this cluster are just listened and liked because of the popularity of the artist. **Profile:** <u>Celebrity Hypes</u>

6. Regression Analysis

i. Methodology

Since our problem statement is classifying four types of genre, we have opted to use multinomial logistic regression (nominal) for the same. In JMP, this is achieved by selecting the *Fit Model* option. The Principal Components and artist_location features are added as input predictors (X variables) and the track_genre is selected as the response (Y variable). The personality (method) is Nominal Logistic with degree=2.

ii. Performance Metrics

Various performance metrics observed using the JMP software are shown below.

▪ EFFECT SUMMARY

The effect summary table gives us the **LogWorth** and **p-values** corresponding to the fields. The LogWorth of all the Sources is > 2, therefore all sources play a significant role in determining the genre. Based on LogWorth, Prin1 and artist_location seem to be the prime sources.

Nominal Logistic Fit for track_genre

Effect Summary

Source	LogWorth	PValue
Prin1	339.254	0.00000
artist_location	170.035	0.00000
Prin11	71.585	0.00000
Prin6	57.652	0.00000
Prin2	56.077	0.00000
Prin5	40.738	0.00000
Prin12	31.458	0.00000
Prin14	29.721	0.00000
Prin13	26.479	0.00000
Prin4	25.081	0.00000
Prin3	13.324	0.00000
Prin7	12.161	0.00000
Prin8	10.475	0.00000
Prin9	5.811	0.00000
Prin10	5.108	0.00001

Remove Add Edit FDR

[Figure 6-1: Effect Summary of Nominal Logistic fit]

▪ WHOLE MODEL TEST

In the Whole Model Test report, the chi-square statistic (5811.857) has a small p-value (<0.0001), which indicates that the overall model is significant. The RSquare (U) tends to be small for logistic models. [10]

Whole Model Test

Model	-LogLikelihood	DF	ChiSquare	Prob>ChiSq
Difference	2905.9283	156	5811.857	<.0001*
Full	1879.9663			
Reduced	4785.8946			

RSquare (U)	0.6072
AICc	4093.38
BIC	5055.31
Observations (or Sum Wgts)	3453

[Figure 6-2: Whole Model Test]

▪ CONFUSION MATRIX

The confusion matrix for the genre prediction task using Nominal Regression:

Confusion Matrix

Training

Actual	Predicted Count			
track_genre	Electronic	Folk	Hip-Hop	Rock
Electronic	616	53	80	87
Folk	33	702	37	90
Hip-Hop	66	28	757	43
Rock	65	79	37	680

[Figure 6-3: Confusion Matrix for multiclass genre prediction]

Accuracy:

- (TP+TN)/(TP+TN+FP+FN) = 2755/3453 = **0.797**
  We have close to 80% accuracy from this model.

Sensitivity (True Positive Rate):

- Sensitivity(Electronic) = (TP)/(TP+FN) = 616/836 = 0.74
- Sensitivity(Folk) = (TP)/(TP+FN) = 702/862= 0.81
- Sensitivity(Hip-Hop) = (TP)/(TP+FN) = 757/894 = 0.85
- Sensitivity(Rock) = (TP)/(TP+FN) = 680/861 = 0.79

Specificity (True Negative Rate):

- Specificity(Electronic) = (TN)/(FP+TN) = 2453/2617 = 0.94
- Specificity(Folk) = (TN)/(FP+TN) = 2431/2591 = 0.94
- Specificity(Hip-Hop) = (TN)/(FP+TN) = 2405/2559 = 0.94
- Specificity(Rock) = (TN)/(FP+TN) = 2372/2592 = 0.92

Eg:- For calculating the Specificity of Electronic genre,

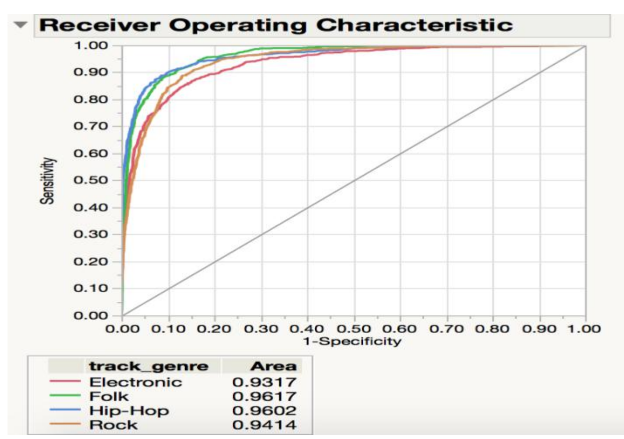
TN = 702+37+90+28+757+43+79+37+680 = 2453

FP = 33+66+65=164

Specificity = (TN)/(FP+TN) = 2453/(164 + 2453) = 0.94

- ROC

The area under the curve(AUC) is > 0.9 for all genres and this indicates that a value closer to 1 is likely to have better overall classifier performance.



[Figure 6-4: ROC curve]

- MEASURE OF THE MODEL

The value of Generalized RSquare (0.8685) is closer to 1, indicating a better model fit.
RMSE(0.4122) indicates a decent fit.
Misclassification Rate (0.2021) indicates a better model fit. [10]

Fit Details

Measure	Training	Definition
Entropy RSquare	0.6072	1-Loglike(model)/Loglike(0)
Generalized RSquare	0.8685	(1-(L(0)/L(model))^(2/n))/(1-L(0)^(2/n))
Mean -Log p	0.5444	∑ -Log(ρ[j])/n
RMSE	0.4122	√ ∑(y[j]-ρ[j])²/n
Mean Abs Dev	0.2939	∑ |y[j]-ρ[j]|/n
Misclassification Rate	0.2021	∑ (ρ[j]≠ρMax)/n
N	3453	n

[Figure 6-5: Fit details]

7. Limitations in the Findings

The limitations

8. Conclusion & Future Scope

We conclude

9. References

1. FMA: A Dataset for Music Analysis https://arxiv.org/pdf/1612.01840.pdf
2. FMA Dataset Github Page https://github.com/mdeff/fma
3. Librosa, a python package for music and audio analysis https://librosa.github.io/librosa/
4. Echonest https://developer.spotify.com/web-api/get-several-audio-features/
5. Temporal Echonest Features http://www.ifs.tuwien.ac.at/~schindler/pubs/AMR2012.pdf
6. MFCC http://musicweb.ucsd.edu/~sdubnov/CATbox/Reader/logan00mel.pdf
7. Detecting harmonic change in music audio http://www.ofai.at/~martin.gasser/papers/oefai-tr-2006-13.pdf
8. Effect of MFCC filters in distinguishing low vs high frequency sound http://www.speech.cs.cmu.edu/15-492/slides/03_mfcc.pdf
9. Instruments and Frequencies https://www.headphonezone.in/blogs/audiophile-guide/mid-range-high-range-frequencies
10. Whole Model Test https://www.jmp.com/support/help/13-2/Whole_Model_Test.shtml