CS 6140 : Final Project Report
Clustering Songs of Spotify Data

Team Members : Kiranmayee Dobbali, Sai Vamshi Dobbali, Sachin Boban

Introduction: Problem and Motivation:

We want to analyse different songs by going beyond the conventional genres they fall into, we felt usual genres are not enough to correctly categorize the varied collection of songs on Spotify. These genre labels offer a rough measure that gauges similarities and differences between tracks. However, how these labels applied are subjective. One person might label a track 'pop' but be labelled as 'folk' by someone else.This way we will be able to create a better playlist of songs and categorise songs which are more similar into the same bucket. We could further explore into sub genres and create very accurate clusters for different Genres of songs.

This project aims to cross the boundaries imposed by genre classification and instead find similarities among music instead of relying on human-assigned labels. By applying clustering techniques on audio features, specific "genres" can be inferred that can describe music differently than conventional music genres. Descriptive subgenres can strengthen a music recommendation engine and give users more options for discovering otherwise unknown music.

The Data:

The dataset that we are working on is the Spotify Songs dataset collected from Kaggle (https://www.kaggle.com/zaheenhamidani/ultimate-spotify-tracks-db). It consists of around 2 Lakhs records and 18 parameters of different songs. It is very important to understand these features below to understand the clusters.

Definition of Audio Features:

Acousticness: A confidence measure from 0.0 to 1.0 of whether the track is acoustic. 1.0 represents high confidence the track is acoustic.
Danceability: Danceability describes how suitable a track is for dancing. A value of 0.0 is least danceable and 1.0 is most danceable.
Energy: Energy is a measure from 0.0 to 1.0 and represents a perceptual measure of intensity and activity. Typically, energetic tracks feel fast, loud, and noisy.
Instrumentalness: Predicts whether a track contains no vocals. Rap or spoken word tracks are clearly “vocal”. The closer the instrumentalness value is to 1.0, the greater likelihood the track contains no vocal content. Values above 0.5 are intended to represent instrumental tracks, but confidence is higher as the value approaches 1.0.
Liveness: Detects the presence of an audience in the recording. Higher liveness values represent an increased probability that the track was performed live.
Speechiness: Speechiness detects the presence of spoken words in a track. The more exclusively speech-like the recording (e.g. talk show, audio book, poetry), the closer to 1.0 the attribute value.
Tempo: The overall estimated tempo of a track in beats per minute (BPM). In musical terminology, tempo is the speed or pace of a given piece and derives directly from the average beat duration.
Valence: A measure from 0.0 to 1.0 describing the musical positiveness conveyed by a track. Tracks with high valence sound more positive (e.g. happy, cheerful, euphoric), while tracks with low valence sound more negative (e.g. sad, depressed, angry)

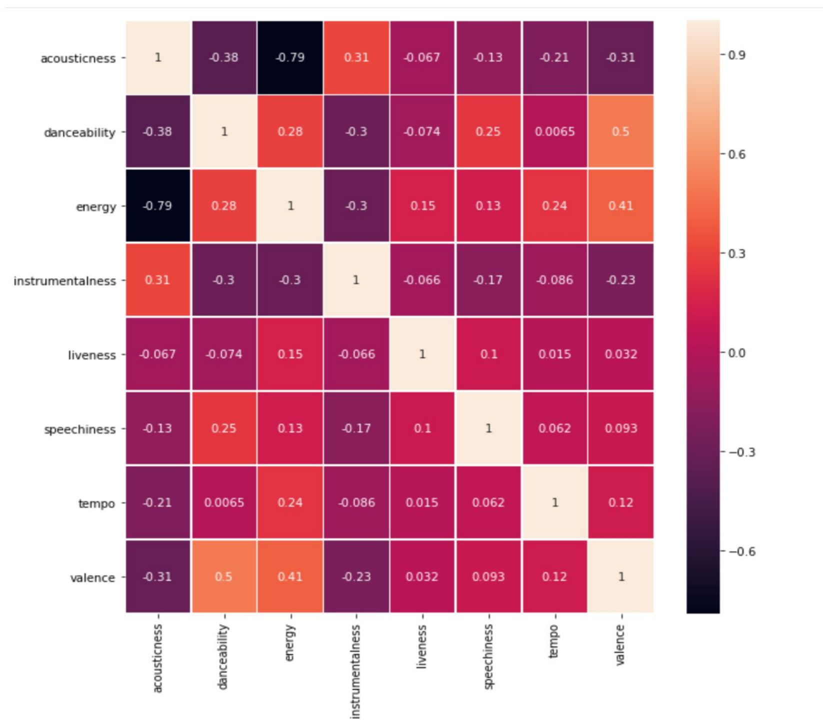
Data Pre-Processing:

This data set did not have much noise in it. However, one of the genres “Children’s Music” is represented in 2 different features with slight name differences. So combined 2 of these columns into a single column. Also as the value in tempo has a higher range, normalize it to the range [0,1] to match with other features.

Exploratory Data Analysis

Relation between Audio features:

Given the definitions of these features, it's worthwhile to examine how each of them relate to one another. The heatmap indicates indicates 2 significant cross-correlations:

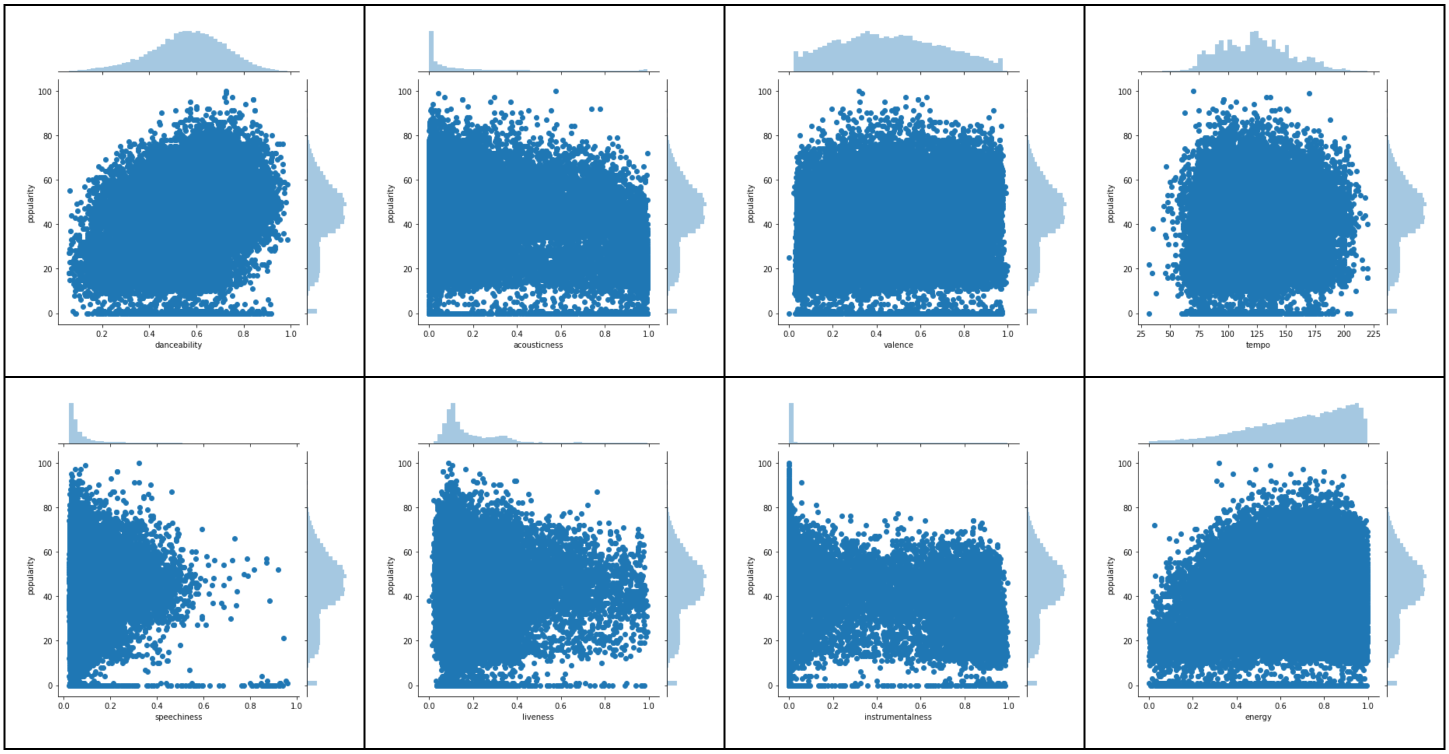


- Acousticness vs. Energy (-0.48). The higher the acousticness, the lower the energy.
- Danceability vs. Valence (0.43). The more danceable a track is, the higher valence/happier it tends to be.

Analysis of different factors that affect Popularity of songs:

1) How accousticness of songs is affecting its popularity. we found more acoustic songs are less popular.
2) How Loudness of a song Affects its popularity. We realised popular songs are not loud.

3) How Speechiness of a song affects its popularity. We realised songs with almost no words that are more instrumentalness are popular.
4) How Danceability of a song affects its popularity. We realised songs with higher danceability have higher popularity.



Boxplots of Audio Features

We have plotted multiple box plots to understand predominant features in each of the genre’s and came to these conclusions.

- **Acousticness**: Metal typically scores low even though the music is heavily reliant on stringed instruments (guitars and bass). Folk's distribution is centered towards high acousticness scores (not surprising) but has plenty of outliers that score low. Rock's distribution spans through a large chunk of the available range.
- **Danceability**: For the most part, each genre centers around medium values for danceability.
- **Energy**: Folk's distribution contrasts how it scores in acousticness, visibly clustering at lower values than other genres. This is partly
- **Instrumentalness:** The distribution for Metal and Electronic are nearly similar, scoring high in instrumentalness with plenty of outliers at low values.
- **Speechiness:** To nobody's surprise, A genre that is mostly centered around vocals (Hip-Hop) vastly outscores all genres in speechiness.

- **Tempo:** Each genre seems to share roughly the same distribution here, with some outliers with high tempos. Those Rock outliers may represent Punk songs.
- **Valence:** Hip-Hop takes the honor for happiest genre of this sample. Metal on the other hand is the saddest / angriest genre of the sample.

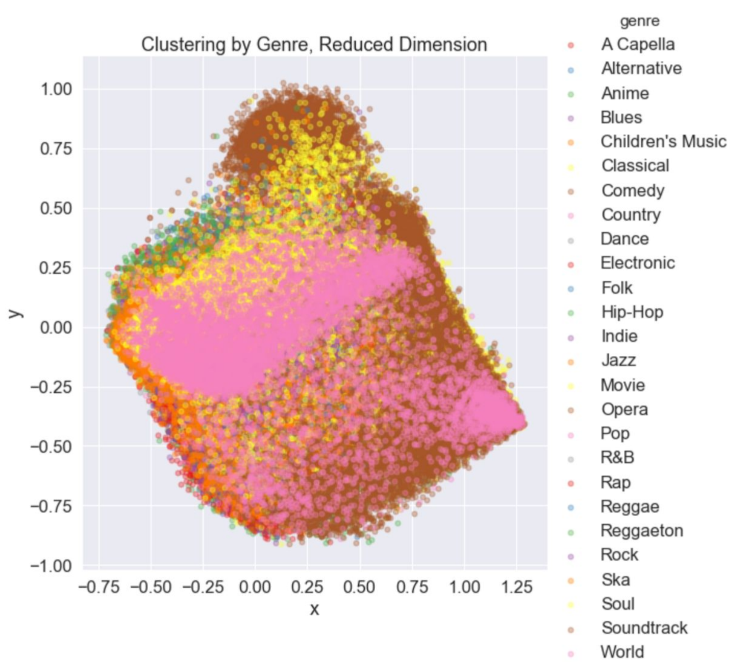
What We Did:

1. We want to see how similar are songs in a particular cluster, and judge if these Genres are really categorising well or how good they are doing it.
   We have computed Mean values of all the attributes of the songs per Genre as shown below.
   Then in order to visualise these clusters we applied Principal Component Analysis on the clusters to reduce its dimensions from 9 to 2.

	genre	popularity	acousticness	danceability	duration_ms	energy	instrumentalness	liveness	loudness	speechiness	tempo	valence	method
0	A Capella	9.302521	0.829941	0.412252	204467.697479	0.250313	0.007491	0.136924	-13.660387	0.042414	0.381792	0.328724	Mean
1	Alternative	50.213430	0.162313	0.541898	233241.364245	0.711519	0.061303	0.196985	-6.540803	0.088783	0.433624	0.449590	Mean
2	Anime	24.258729	0.286843	0.472090	229937.067927	0.665356	0.280592	0.192391	-7.917802	0.065102	0.452891	0.441682	Mean
3	Blues	34.742879	0.327840	0.528232	251931.826000	0.606171	0.095175	0.233125	-9.053807	0.061809	0.427051	0.579425	Mean
4	Children's Music	36.202426	0.320112	0.598829	199663.490919	0.593204	0.087013	0.183986	-8.399591	0.097763	0.429407	0.532251	Mean

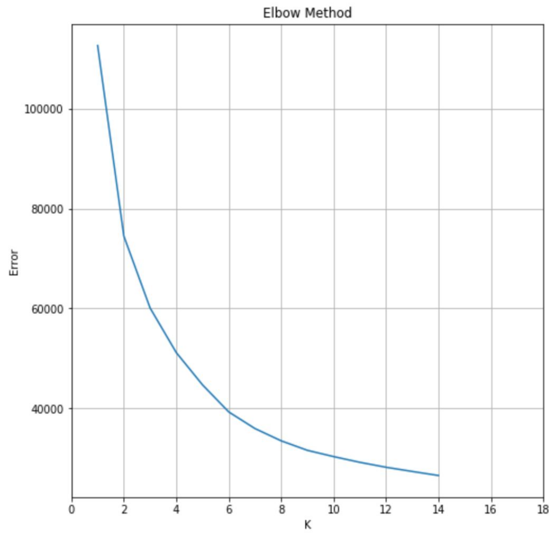
2. **Principal Component Analysis (PCA)** is a dimensionality reduction technique
that can be used to reduce the attribute space such that the existing set of variables is reduced to a smaller set with the most important information about the dataset. The first principal component accounts for as much of the variability in the data as possible. In the similar way, each succeeding principal component accounts for as much of the remaining variability too.

In the Reduced dimension, conventional Genres are not properly clustered, so when a user asks for similar

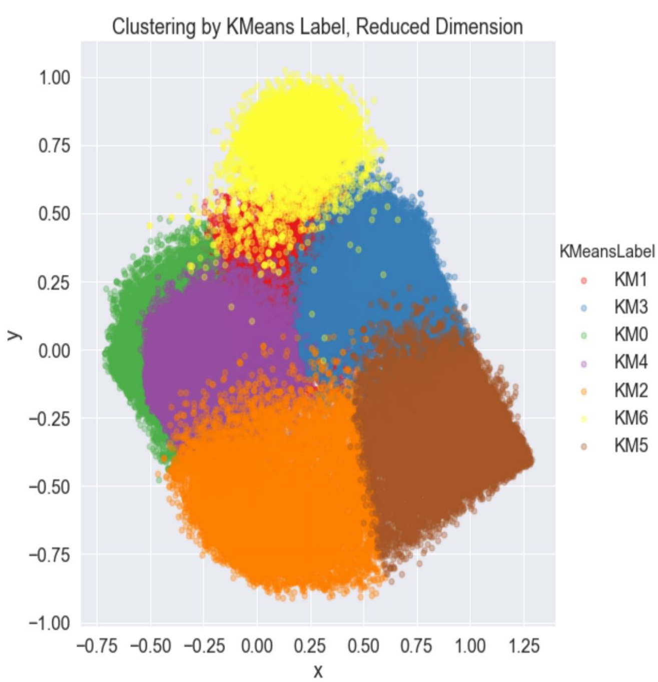


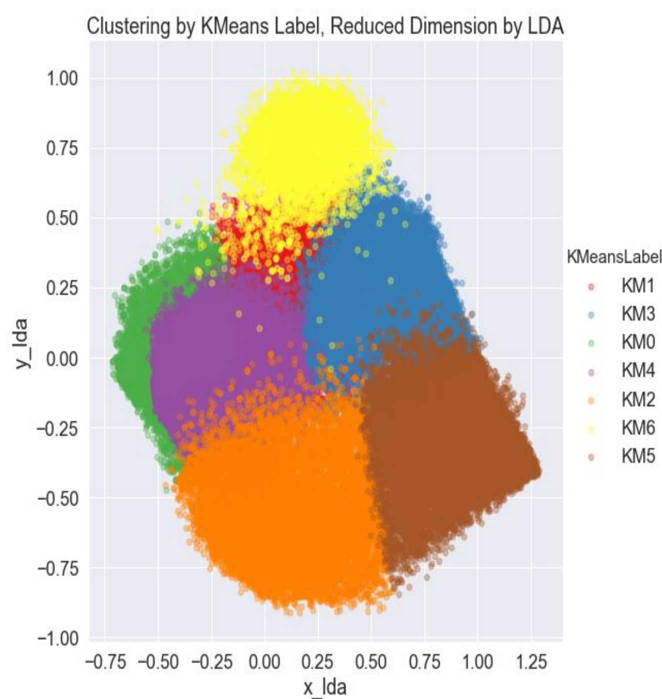
songs we might most likely end up giving incorrect suggestions by selecting songs from the same Genre.

Then we did K Means Clustering on this n-dimension dataset, using Euclidean distance between the songs as a parameter for similarity/distance between them.
Inorder to decide on the number of clusters we can take, we have applied elbow technique.



Based on the above graph we took k = 7 that is we are clustering the data into 7 Genres.
Inorder to Visualize the clusters, we have applied PCA on the above clusters.

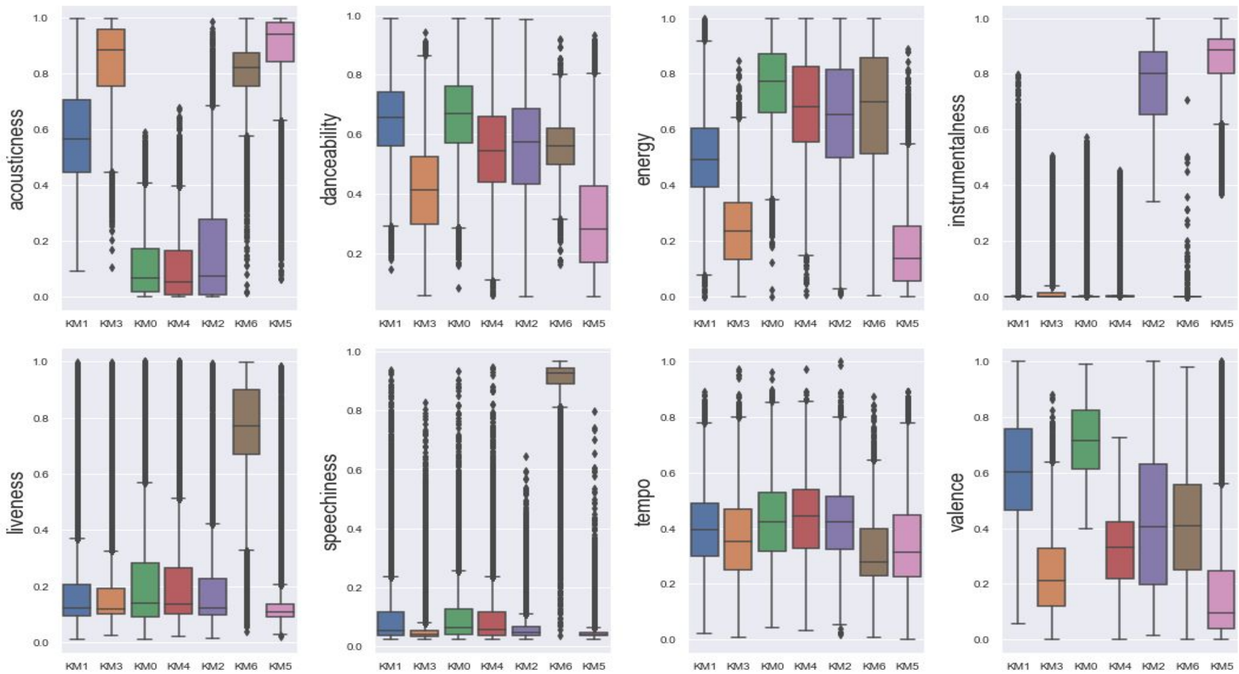




Above are the new clusters we obtained. As you can see these clusters are much better and give us better Genres. Later we tried to do LDA projection to visualize the clusters, but they are more or less similar to PCA projection.

Evaluation of Genre Clusters by KMeans:

The below box plots shows how each of these features(‘acousticness', 'danceability', 'energy', 'instrumentalness', 'liveness', 'speechiness', 'tempo', 'valence’) are in each cluster that our algorithm generated.



Cluster 6 Analysis - Fun and Lively Music

Comedy : 0.916878 , Movie : 0.035335, Children's Music : 0.015389, Hip-Hop : 0.004354

In cluster 5 we can observe that about 91% of the songs in the database comes under the comedy genre. From observing the features, speechiness, liveness, acoustics and energy are very high. Instrumentalness, tempo, and valence are very low for this cluster.As we can say that comedy genre usually involves high amount of usage of words, also has high energy, this cluster having high values for both the features makes sense. So we can name this cluster as a "Fun and Lively" group.

Cluster 5 analysis - Gloomy Acoustic music

Soundtrack:0.313404, Classical: 0.265783, Opera: 0.091940, Anime :0.067304,World: 0.064784

So, in the KM5 cluster, we can observe that 57% of the songs are in the soundtrack and Classical genre category. And we can observe that the values of instrumentalness and acousticness are very high and energy,danceability, speechiness and liveness are significantly low. So the cluster having high values of acousticness and instrumentalness makes sense. This cluster can be named "**Gloomy Acoustic**" .

Cluster 4 Analysis - Upbeat Music

Children's Music : 0.076797, Rap : 0.073560,Alternative : 0.073155,Hip-Hop : 0.068784,Pop : 0.062034

This cluster has high values of energy, danceability. Mid level tempo.
Also very low values of instrumentalness, acousticness, liveness and speechiness.

Cluster 3 Analysis - Happy Acoustic

Opera : 0.209852, Movie : 0.100126, Folk : 0.097856, Classical : 0.078306

This cluster has a high value of acousticness. Mid Level danceability and tempo. Also this cluster contains 20% of the songs from Opera which makes sense because opera usually involves acoustic music and doesn't have much instrumentalness or speechiness. Also, Folk classical Movie opera contributes to almost 50% of the songs in this category. So this group can be named **"Happy acoustic"** as opposed to cluster 5 which has high acoustic but very low danceability and valence.

Cluster 2 Analysis - Energetic Instrumentals

Electronic: 0.233340, Jazz: 0.161606, Soundtrack: 0.096665, Anime: 0.087630
This cluster contains songs of "Electronic and Jazz" of about 40% from all the songs. By observing the above boxplots, this cluster has high values of instrumentalness and energy. It also has significantly low values from acousticness, liveness and speechness.

So we can name this group as "**Energetic Instrumentals**"

Cluster 1 Analysis -Positive soul,dancing Jazz Blues

Children's Music:0.110416,Soul:0.076567,Blues:0.072667,Movie:0.070936,Jazz:0.065526,Folk:0.061562

This cluster contains valence,danceability and acousticness and mid level energy and tempo. It has low levels of instrumentalness, liveness and speechiness.

As we can see that 11% of the songs in this cluster has children's music, having high values of valence(positivity) makes sense. Soul, blues and Jazz are kind of overlapped in general music in the world. I will just name this category - **"Positive soul,dancing Jazz Blues"**

Cluster 0 Analysis - Upbeat positive music

Reggaeton:0.097368, Reggae : 0.089483, Ska : 0.084809,Dance : 0.060487

This cluster has high values of energy, danceability and valence. Mid level tempo. Also very low values of instrumentalness, acousticness, liveness and speechiness.

Now by observing the genre percentages in this group, each of the Reggaeton, Reggae, Ska,Dance, Children's Music, Hip Hop contribute to almost 45% of the songs. If we see Reggaeton, reggae and Hip Hop they are all child genre categories of the same parent genres. So our cluster having major contributions from these genres makes sense.
If we observe both cluster 0 and cluster 4, both of them have the same kind of features except for valence. Cluster 0 has high valence whereas cluster 0 has low valence. Valence means positivity. So we can categorize cluster 4 as **Upbeat music** and cluster 0 as **Upbeat positive music**.

Comparing Clustering Algorithms For Mini Sample Dataset:

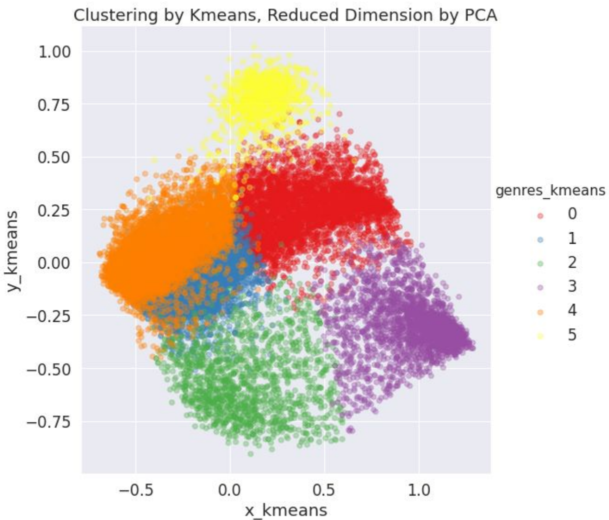
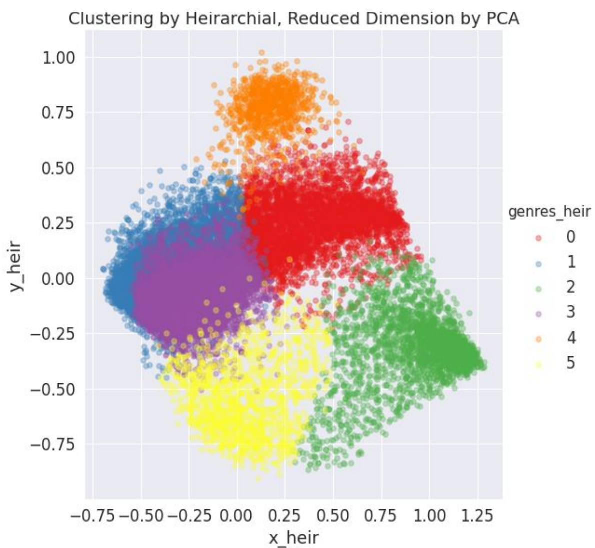
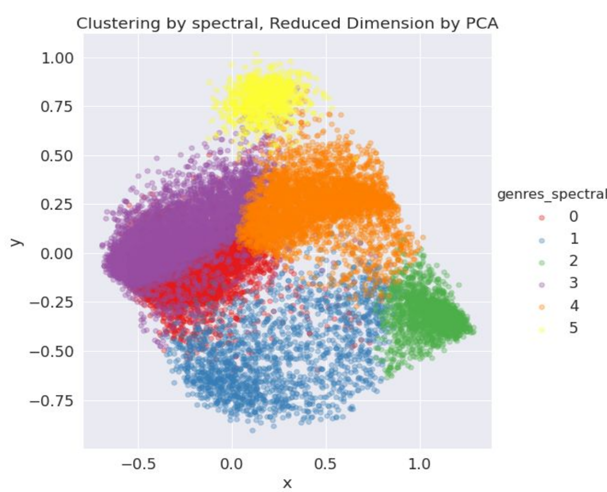
We have performed different Clustering Algorithms like

1) K Means Clustering
2) Hierarchical Clustering
3) Spectral Clustering

To perform these clustering algorithms and compare, we have reduced the original dataset which is really huge and has about 200k songs in it. It was not possible to perform Agglomerative and Spectral clustering on the original dataset and so we took a random sample dataset of about 23k songs.

We also performed Elbow technique on KMeans to get the best number of clusters in this reduced dataset.

By comparing all the 3 clustering algorithms, we can see that both K Means and Hierarchical generated the same kind of clusters.
Also by applying clustering evaluation metrics such as Adjusted_rand_score,Adjusted_mutual_info_score, v_measure_score and silhouette_score, we can see that k means and agglomerative works better than spectral. However, the overall scores are very low which means the clustering model performed poorly.

	**adjusted_rand_score:** 0.0545270436 **Adjusted_mutual_info_score**: 0.20120700 **v_measure_score**: 0.2021039751 **silhouette_score**: 0.26071096
	**adjusted_rand_score:** 0.0545270436 **Adjusted_mutual_info_score**: 0.20120700 **v_measure_score**: 0.2021039751 **silhouette_score**: 0.26071096
	**adjusted_rand_score:** 0.04822040 **Adjusted_mutual_info_score**: 0.19679008 **v_measure_score**: 0.19769781 **Silhouette_score**: 0.2459457

Conclusion:

Due to the variable nature of music, a lot of overlapping music genres which are slightly different from each other, it is difficult to evaluate performance of our clusters. In order to evaluate the clusters, it is required to have a domain knowledge of the genres and interpret the results which may vary on how we define genres. However, by observing the features and little research on original music genres, we were able to differentiate newly formed clusters and assign new genres names to them.

Some of the interesting observations that we found in these clusters are that many of the original music genres have a lot of similarities. For example, we found that cluster 1 which is “Positive soul,dancing Jazz Blues” has songs from Soul, Jazz and Blues music genres and they all are the most similar music genres.

We were able to apply different clustering algorithms on our sample data which is around 23000 songs (10% of the original data set) and observe the differences. We found that the clusters formed here are very much similar to that of our original dataset.

Individual Contributions:

- Kiranmayee and Saivamshi, we both have equally contributed to the project.
- Sachin participated in the initial project proposal submission.