



SONGS RECOMMENDATION SYSTEM

P180 GROUP 1



GROUP MEMBERS :

Aji Thomas
Ananthalakshmi Saripalle
Chimala Rajesh
Jagadeesh Korukonda
Margam Navya
Sanjusha Suresh

Mentor : Ms. Neha Ramchandani

CONTENT

ABSTRACT

INTRODUCTION

1. Business Objective
2. Project Architecture
3. Challenges faced

DATA COLLECTION

Variable information

EDA

Interpretations

DATA VISUALIZATION

MODEL BUILDING

K-Means Clustering

MODEL DEPLOYMENT

Streamlit

RESULT

Abstract

RECOMMENDATION ENGINE

A recommendation system is an artificial intelligence or AI algorithm, usually associated with machine learning, that uses Big Data to suggest or recommend additional products to consumers. These can be based on various criteria, including past purchases, search history, demographic information, and other factors.



INTRODUCTION

BUSINESS OBJECTIVE

The aim of this project is to build a feature of recommendation system to support a music app. As the first phase we need to develop the proof of concept to make the client understand how effective the feature could be.

PROJECT ARCHITECTURE

Data Collection



EDA & Data Cleaning



Data Visualization



Model Building



Model Deployment

CHALLENGES FACED

LIMITED TARGET TIME

We overcame this challenge by delegating responsibilities equally to all the group members.

DATA COLLECTION

The data is from Spotify which was collected from the Kaggle.



VARIABLE INFORMATION

song_name: The names of the different songs.

song_popularity: The popularity of the song which ranges between 0 and 100.

song_duration_ms: The length of the track in milliseconds.

acousticness: The value that describes how acoustic a song is. Higher values indicates that the song is most likely to be an acoustic one. Ranges between 0-1.

danceability: The relative measurement of the track being danceable. Higher values indicate that the song is more danceable. Ranges between 0-1.

energy: The energy value of the track. Higher values indicates that the song is more energetic. Ranges between 0-1.

instrumentalness: The relative ratio of the track being instrumental. Higher values indicate that the song contains more instrumental sounds. Ranges between 0-1.

key: All keys on octave encoded as values ranging from 0 to 11, starting on C as 0, C# as 1, etc.

liveness: Detects the presence of an audience in the recording. Higher values represent an increased probability that the track was performed live. A value above 0.8 provides a strong likelihood that the track is live. Ranges between 0-1.

loudness: The overall loudness of a track in decibels (dB). Loudness values are averaged across the entire track and are useful for comparing the relative loudness of tracks. Loudness is the quality of a sound that is the primary psychological correlate of physical strength (amplitude). Values typically range between -60 and 0 dB.

audio_mode: Mode indicates the modality (major or minor) of a track, the type of scale from which its melodic content is derived. Major is represented by 1 and minor by 0.

speechiness: The relative length of the track containing any kind of human voice. Ranged between 0-1.

tempo: The tempo of the track in Beat Per Minute (BPM).

time_signature: It is a notational convention used in Western musical notation to specify how many beats (pulses) are contained in each measure (bar), and which note value is equivalent to a beat.

audio_valence: The positiveness of the track. Higher values indicates, that the track evokes positive emotions (like happiness, excitement) otherwise indicates, that it evokes negative emotions (like anger, fear). Ranges between 0-1.

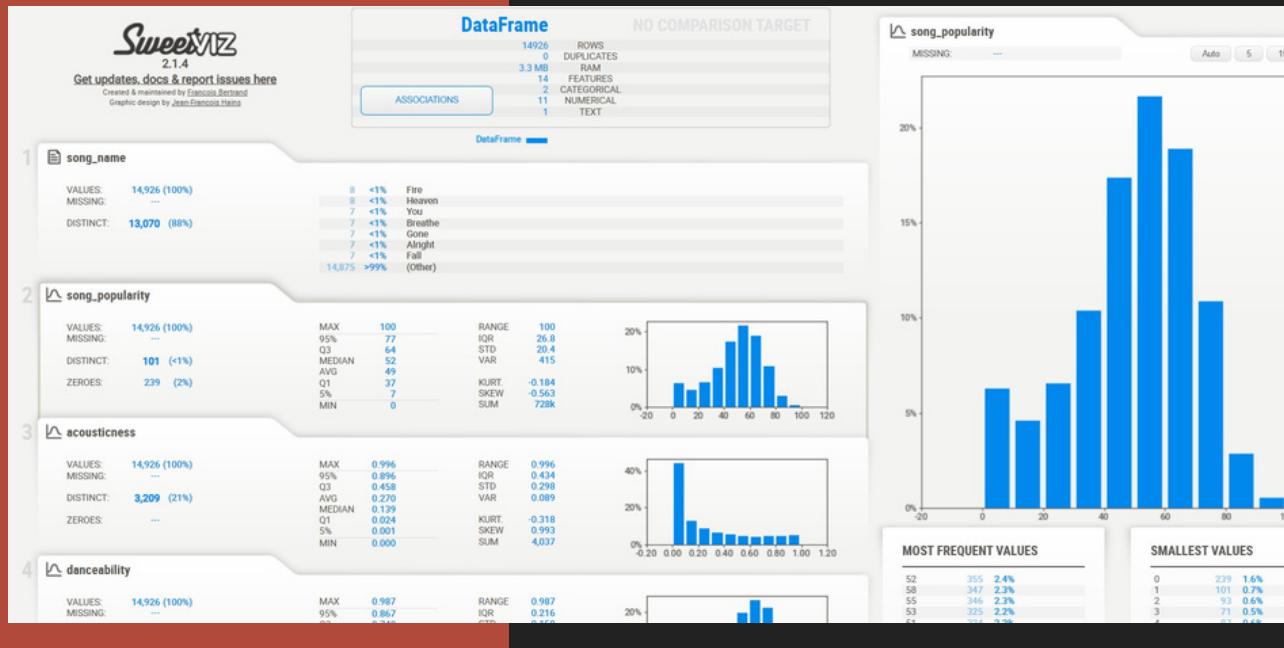


EDA



- The shape of the dataset is (18835, 15).
- There are no null values in the dataset.
- There are 3909 duplicate values which has to be dropped.
- After the data cleaning i.e. after dropping the duplicates and the insignificant column, the final shape of the dataset is (14926, 14).
- From the correlation, Energy and Loudness has the highest positive correlation of 0.76570. Energy and Acousticness has the highest negative correlation of -0.67923.
- Number of songs: 13070
- Number of users: 101
- Number of ratings: 14926
- Highest popularity rating 100 is given only 1 time. 52 popularity rating is given the max no. of 355 times.
- Maximum number of songs has the key value 0 and the least number of songs has the key value 3.
- Most of the songs have major modality of 1.
- Most of songs has time signature of 4 i.e. a song has four beats per measure. 4/4 time is also called "common time" because it is the most common time signature for pop, rock, folk music etc.

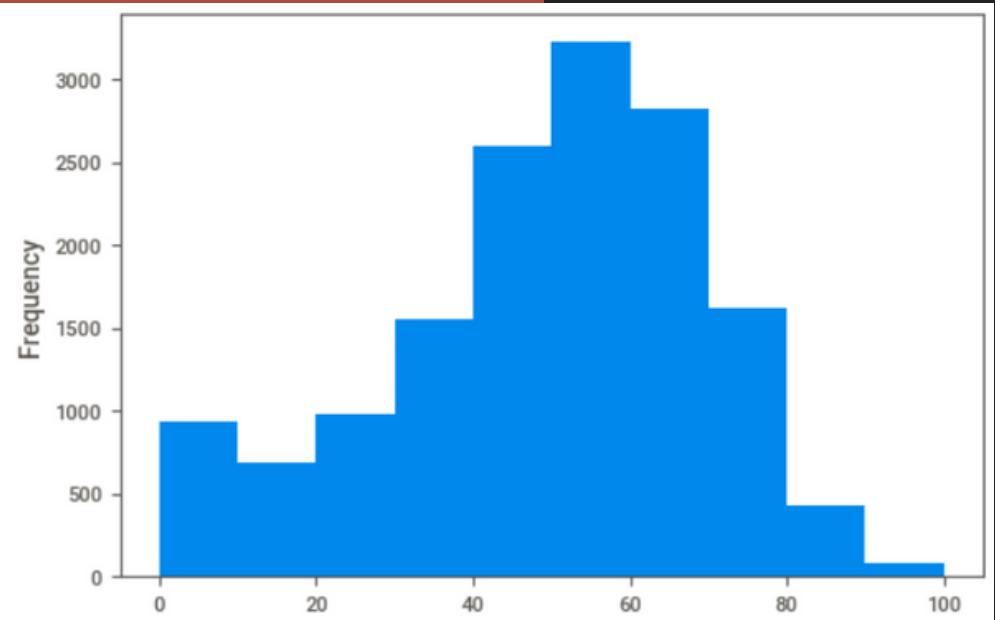
Data Visualization



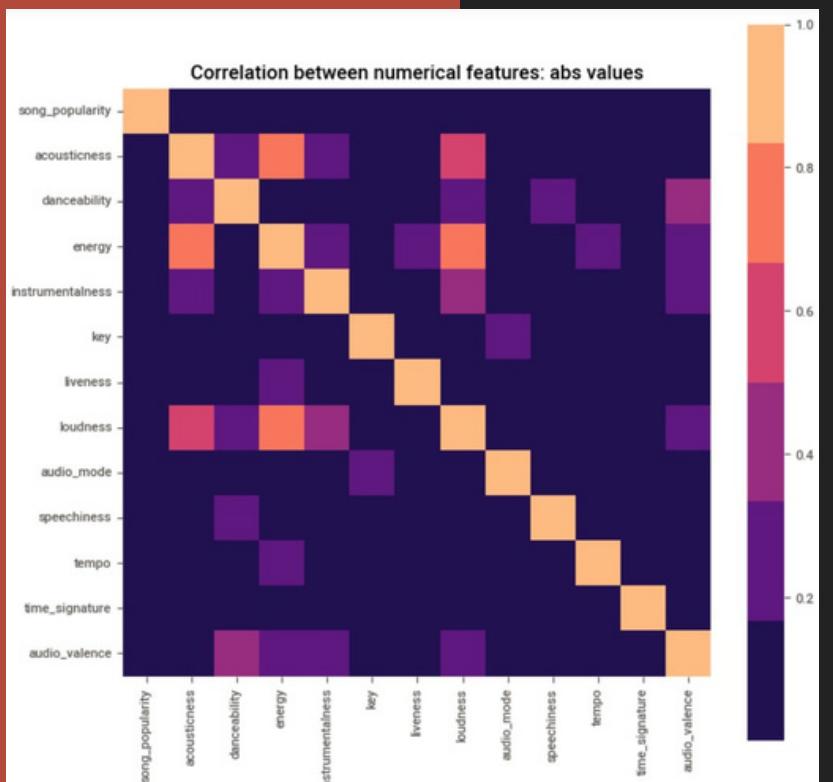
SWEETVIZ



PLOTY

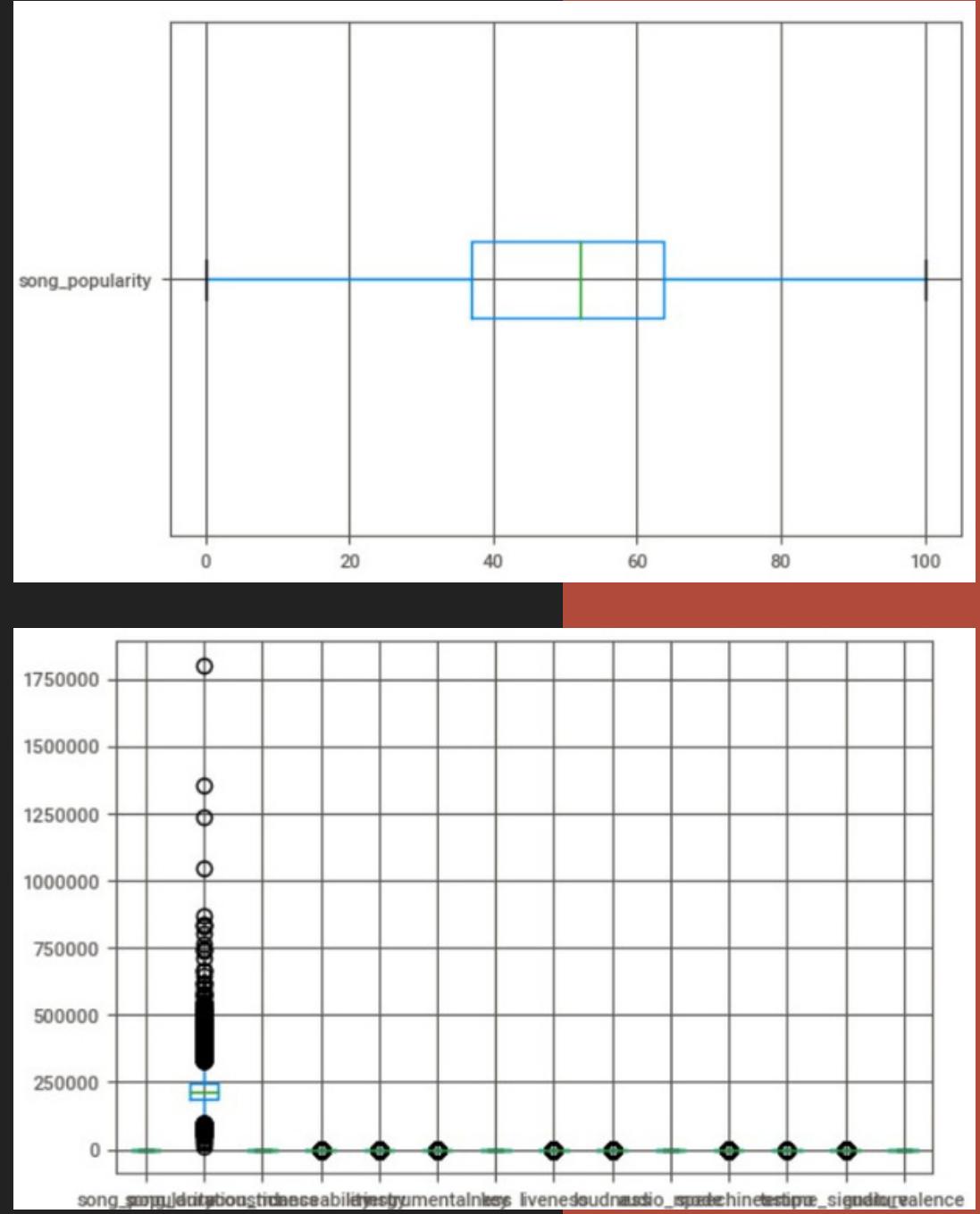


HISTOGRAM



HEAT MAP

BOXPLOTS



Model Building

K-MEANS CLUSTERING

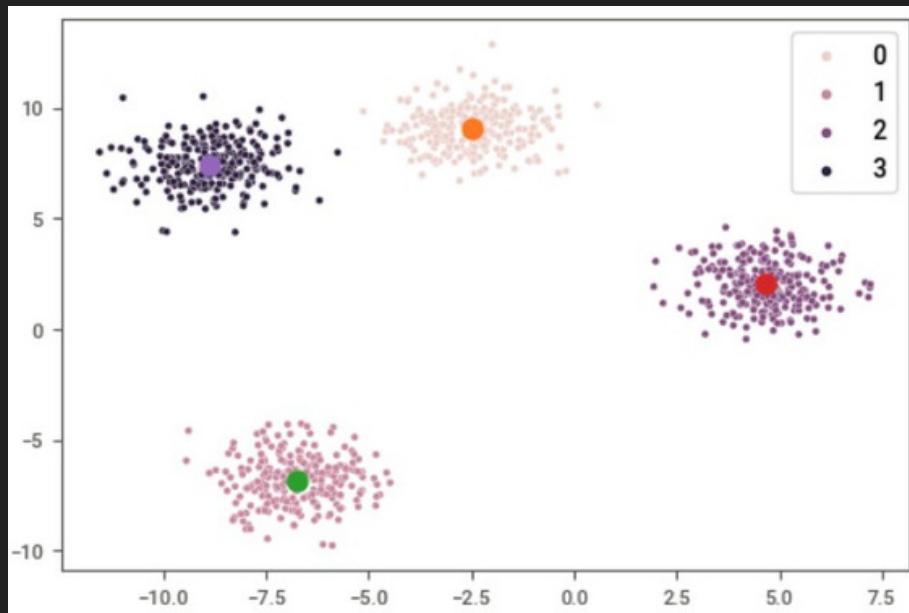
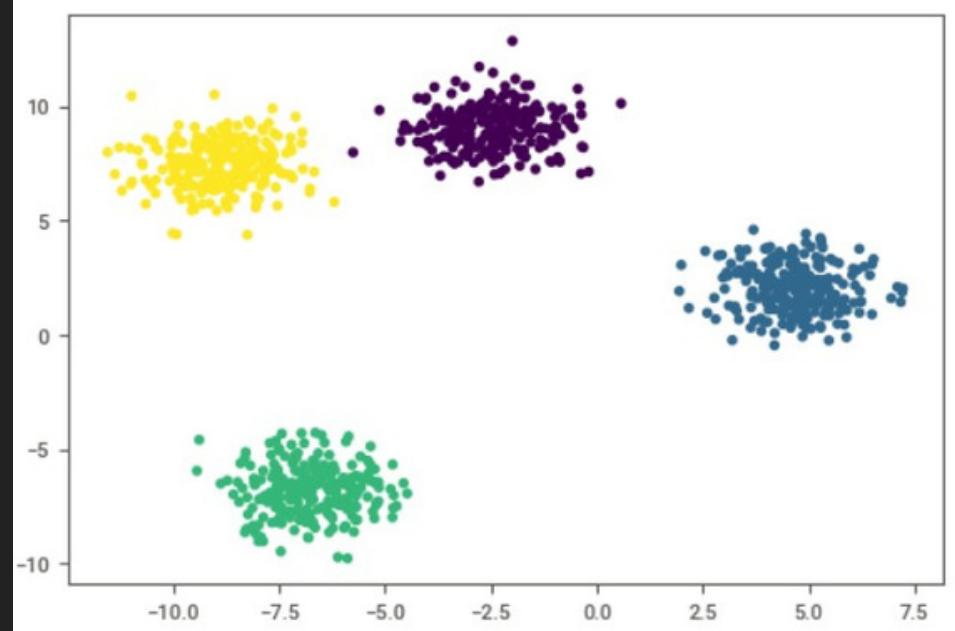
K-Means Clustering is an unsupervised learning algorithm that is used to solve the clustering problems in machine learning or data science.

It allows us to cluster the data into different groups and a convenient way to discover the categories of groups in the unlabeled dataset on its own without the need for any training. It is a centroid-based algorithm, where each cluster is associated with a centroid. The main aim of this algorithm is to minimize the sum of distances between the data point and their corresponding clusters.



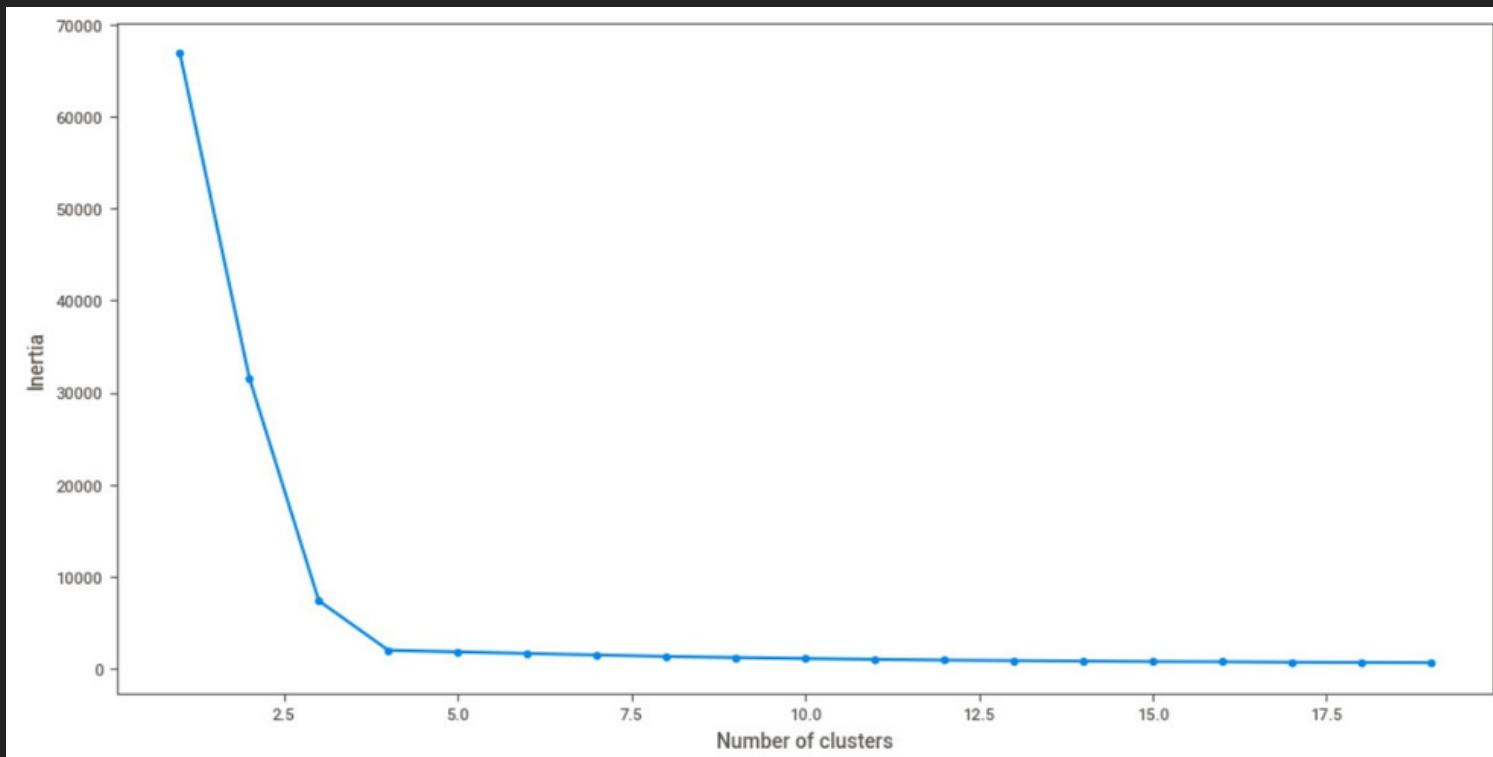
Using K-Means we plotted the graph, which shows four main clusters of users, where each clusters are grouped based on user similarities.

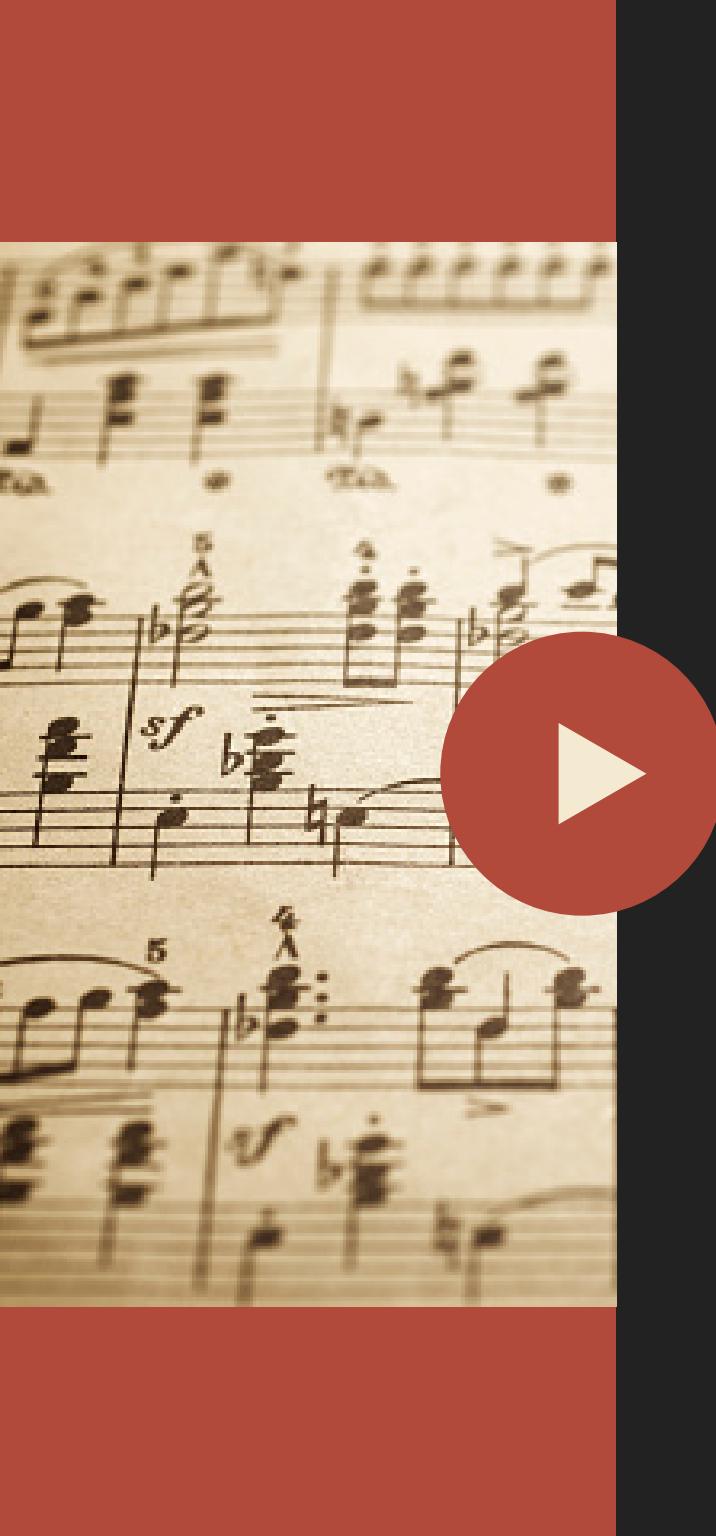
Then we fitted multiple k-means algorithms and stored the values in an empty list.



And by plotting the centroids for each clusters, we get the graph as shown here.

Finally by converting the results into a data frame and plotting them, we get the below elbow plot. Which also shows the decrease in the deviation from the value four.





MODEL DEPLOYMENT

- Created a feature matrix using the song name and popularity features.
- And used the CountVectorizer to convert the song name feature into a numerical feature matrix.
- Then, combined the song name feature matrix with the popularity feature matrix.
- Calculated the cosine similarity between all songs.
- Included a function to recommend songs based on a given song name.
- Saved the model using pickle and created a deployment file in .py using Streamlit.
- Finally, running the necessary code in anaconda prompt we get the Model Deployment.

RESULT



K-MEANS CLUSTERING

K-Means is one of the most widely used unsupervised clustering methods. Using clustering can address several known issues in recommendation systems, including increasing the diversity, consistency and reliability of recommendations.

ABOUT

Made with streamlit



recommendations.

ABOUT

Made with streamlit



P180 Group 1:

Aji Thomas , AnanthaLakshmi Saripalle

Chimala Rajesh , Jagadeesh Korukonda

Margam Navya , Sanjusha Suresh

Under the guidance of

Neha Ramchandani

Himavanth Ila

SONGS RECOMMENDATION SYSTEM

Similar songs based on popularity



| | song_name | song_popularity | song_duration_ms | acousticness | danceability | energy |
|-----|-------------|-----------------|------------------|--------------|--------------|--------|
| 173 | Sex on Fire | 81 | 203346 | 0.0017 | 0.5420 | 0.9050 |

Please type the song name for which you want the recommendation

rely

Recommended songs are

| | song_name |
|-------|-------------------|
| 1006 | Selling The Drama |
| 17280 | The Locomotion |
| 15025 | The Hearse |
| 2852 | The Cutter |
| 14029 | Becoming the Bull |
| 4650 | The Middle |
| 5244 | The Grand Tour |
| 4009 | The Great Escape |
| 4414 | The Ceremony |
| 16835 | The Promise |



A dark, moody photograph of a person playing a guitar. The guitar's neck and strings are visible, and a hand is shown strumming. The background is dark with blurred, colorful bokeh lights from what appears to be a Christmas tree.

THANK YOU

Created by: Sanjusha Suresh

