Know-IT(Pune)

"Music Recommendation System"

PG-DBDA May 2021

Submitted By:

Group No:01

Rutuja Arsule (210543083003)

Gaurav Langar (210543083008)

Priya Mali (210543083011)

Pooja Chaskar (210543083021)

Submitted in partial fulfillment for the award of

Post Graduate Diploma in Big Data Analytics (PG-DBDA)

From Know-IT(Pune).

Guided by:

Mr. Amey Manjrekar.

CERTIFICATE

This is to certify that the Project Entitled

Music Recommendation System

Submitted by:

Rutuja Arsule (210543083003)

Gaurav Langar (210543083008)

Priya Mali (210543083011)

Pooja Chaskar (210543083021)

is a Bonafede student of this institute and the work has been carried out by him/her under the supervision of Mr. Amey Manjrekar and it is approved for the partial fulfilment of the requirement of Post Graduate Diploma in Big Data Analytics.

ACKNOWLEDGEMENT

This project **"Music recommendation System"** was a great learning experience for us and we are submitting this work to CDAC Know-IT (Pune).

We all are very glad to mention the name of **Mr. Amey Manjrekar** and **Mr. Sanjay Sane** for his valuable guidance to work on this project. His guidance and support helped us to overcome various obstacles and intricacies during the course of project work.

We are highly grateful to **Mr. Vaibhav Inamdar** Manager (Know-IT), C-DAC, for his guidance and support whenever necessary while doing this course Post Graduate Diploma in Big Data
Analytics (PG-DBDA) through C-DAC ACTS, Pune.

Our most heartfelt thanks goes to **Mrs. Bakul Joshi** (Course Coordinator,PG-DBDA) who gave all the required support and kind coordination to provide all the necessities like required hardware, internet facility and extra Lab hours to complete the project and throughout the course up to the last day here in C-DAC Know-IT, Pune.

From:

Rutuja Arsule (210543083003)

Gaurav Langar (210543083008)

Priya Mali (210543083011)

Pooja Chaskar (210543083021)

INDEX
TABLE OF CONTENTS

1.Abstract 1
2.Introduction 2
3.Purpose 3
4.Functional Requirements 4
5.System Requirements.....6
5.1. Hardware Requirements 6
5.2.Hardware Requirements 6
6.System Architecture…….8
7.Methodology..... 9
8.Machine Learning Algorithm..... 10
9.Data Visualization..... 11
10.Conclusion……16
10.Future Scope……..17
11.References …….18

ABSTRACT

Music is one of the common languages that connects us all together and it has always been a major part of all our lives. For that first we downloaded data using python, performed ETL operations using spark and dump that data in MongoDB for recommendation system and visualization using PowerBI.

We used Neighborhood Collaborative Filtering using the similarity metrics method to recommend more accurate songs to the user.

INTRODUCTION

In simple words, a recommendation system is any system that automatically suggests content for website readers and users. These systems were evolved as intelligent algorithms, which can generate results in the form of recommendations to users.
[1] Theyrequire a large dataset and a fast-computing system that can perform analytics on same within fraction of seconds.
[2] A variety of techniques have been proposed till today for performing recommendations. Recommendation system is defined as the computer program that helps the user determine goods and content by predict the users rating of eachitem and presentation them the substance that they would rate highly. There are different ways to approach a recommendation system. They are basically divided into Content-basedand Collaborative Filtering systems. In our project we are using Collaborative filtering.

1.1 Collaborative Filtering

The collaborative filtering is based on customer's behavior's and past likes. It is used to make automatic predictions, filtering collecting preferences, interest of a user and taste information of many users. Collaborative filtering can be performed in two ways mainly, model-based and memory-based.

Purpose

The objective of the problem is to recommend Songs. Clearly, we have to create a recommendation system here. Here we used Neighborhood Collaborative Filtering using the similarity metrics method.

Data Collection:

To create a Spotify recommendation system, we will be using a dataset that has been collected from Spotify. The dataset contains over 175,000 songs with over 19 features grouped by artist, year and genre. we will begin the task of building a music recommendation system with machine learning by importing the necessary Python libraries and dataset.

FUNCTIONAL REQUIREMENTS

Python:

- Python is a general-purpose interpreted, interactive, object-oriented, andhigh-level programming language.
- Python supports multiple programming pattern, including object-oriented, imperative, and functional or procedural programming styles.
- Python has a large and broad library and provides rich set of module andfunctions for rapid application development.

Apache Spark:

- Streaming Data - Apache Spark's key use is its ability to process streaming data.
- Traditional ETL (Extract, Transform, Load) tools used.
- Interactive Analysis - Among Spark's, most notable features are its capability for interactive analytics.

MongoDB:

- MongoDB's defining features is its schema-less or non-relational data structure.
- Traditional ETL (Extract, Transform, Load) tools used.
- MongoDB is the opportunity for horizontal scaling through sharding.

Machine Learning:

- Machine Learning is the field of study that gives computers thecapability to learn without being explicitly programmed.
- Machine learning combines data with statistical tools to predict anoutput.

PowerBI:

- Microsoft Power BI provides such tools that will let you visualize key data points accurately from various sources in a single dashboard.
- Power BI takes input from various sources and provides that data in various contexts
- Power BI helps businesses to check any kind of details to ensure that the business is running smoothly.

System Requirements

Hardware Requirements:

- Platform – windows 10
- RAM – 8 GB of RAM,
- A network connection for data recovering over network.

Software Requirements:

- Python 3
- Apache Spark
- MongoDB
- PowerBI
- Flask

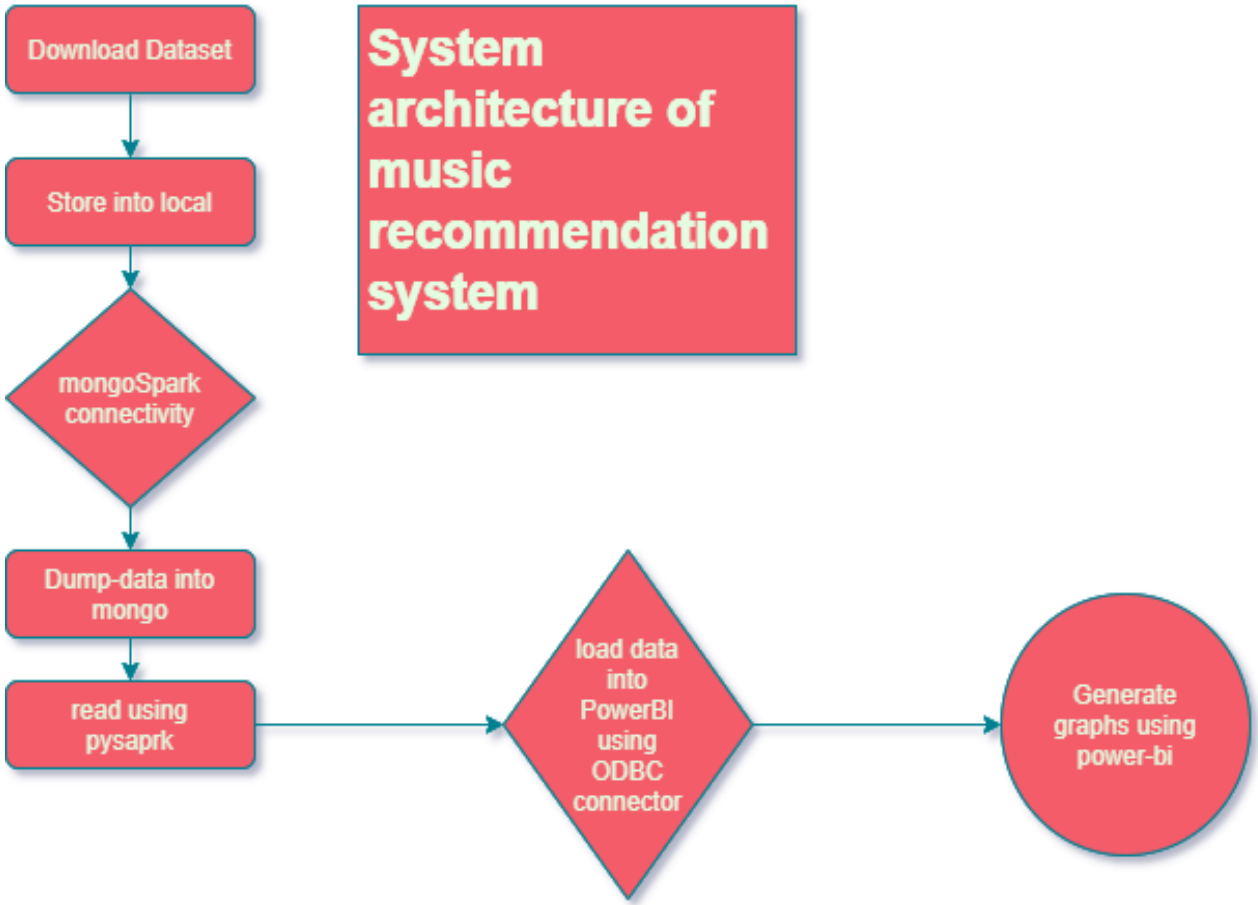
Data Engineering and Analysis:

Data engineering is the aspect of data science that focuses on practical applications of data collection and analysis. For all the work that data scientists do to answer questions using large sets of information, there have to be mechanisms for collecting and validating that information.

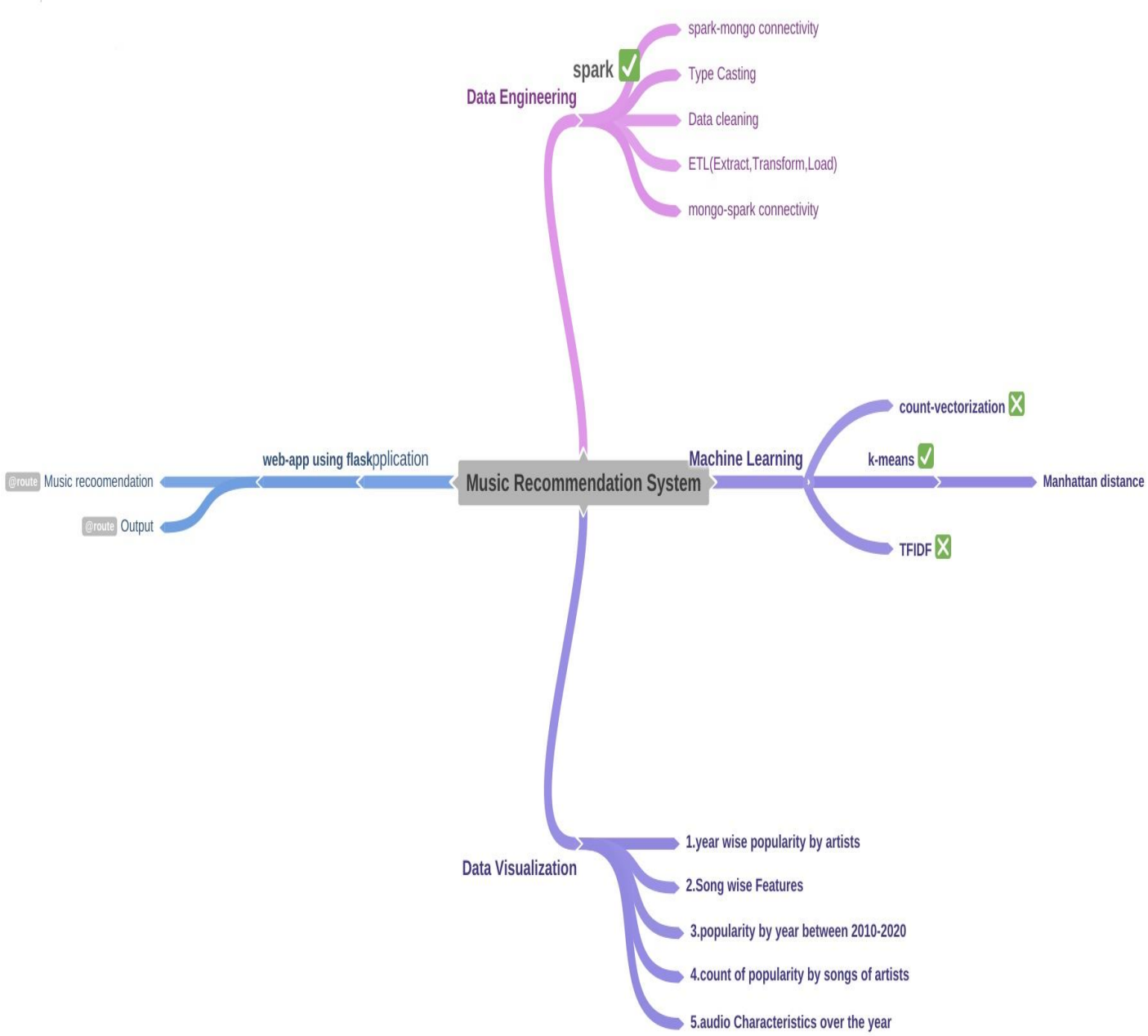
Data Cleaning:

Data cleansing or data cleaning is the process of detecting and correcting (or removing) corrupt or inaccurate records from a record set, table, or database and refers to identifying incomplete, incorrect, inaccurate or irrelevant parts of the data and then replacing, modifying, or deleting the dirty or coarse data. Data cleansing may be performed interactively with data wrangling tools, or as batch processing through scripting. After cleansing, a data set should be consistent with other similar data sets in the system. The inconsistencies detected or removed may have been originally caused by user entry errors, by corruption in transmission or storage, or by different data dictionary definitions of similar entities in different stores. Data cleaning differs from data validation in that validation almost invariably means data is rejected from the system at entry and is performed at the time of entry, rather than on batches of data.

System Architecture:



Methodology:



Machine Learning Algorithms:

1. K-means:

K Means is a clustering algorithm which divides observations into k clusters. Since we can dictate the amount of clusters, it can be easily used in classification where we divide data into clusters which can be equal to or more than the number of classes.

2. Song Recommendation System:

Song Recommendation System
Building a recommendation system where it recommends similar songs for any given Song.
Here I have used Neighborhood Collaborative Filtering using the similarity metrics method. Calculated Manhattan Distance using all numerical features available in the dataset and find the neighbor songs which have relatively less distance.

3. Manhattan distance

The Manhattan distance between two points $\mathbf{x} = (x_1, x_2, ..., x_n)$ and $\mathbf{y} = (y_1, y_2, ..., y_n)$ in n-dimensional space is the sum of the distances in each dimension.

$$d(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{n} | x_i - y_i |.$$

Song wise Recommendation :

```
#Get recommendations 'Red Roses (feat. Landon Cube)' song
recommender.get_recommendations('Red Roses (feat. Landon Cube)', 5)
```

```
100%|
```

	artists	name	year
71204	Post Malone, Halsey, Future	Die For Me (feat. Future & Halsey)	2019
124891	Dirty Heads	Believe	2012
25213	Vince Staples, 6LACK, Mereba	"Yo Love - From ""Queen & Slim: The Soundtrack"""	2019
127781	Cuco, Jean Carter	Bossa No Sé (feat. Jean Carter)	2019
91402	Baby Bash, Z-Ro, Berner, Baby E	Light Up Light Up	2014

DATA VISUALIZATION AND REPRESENTATION

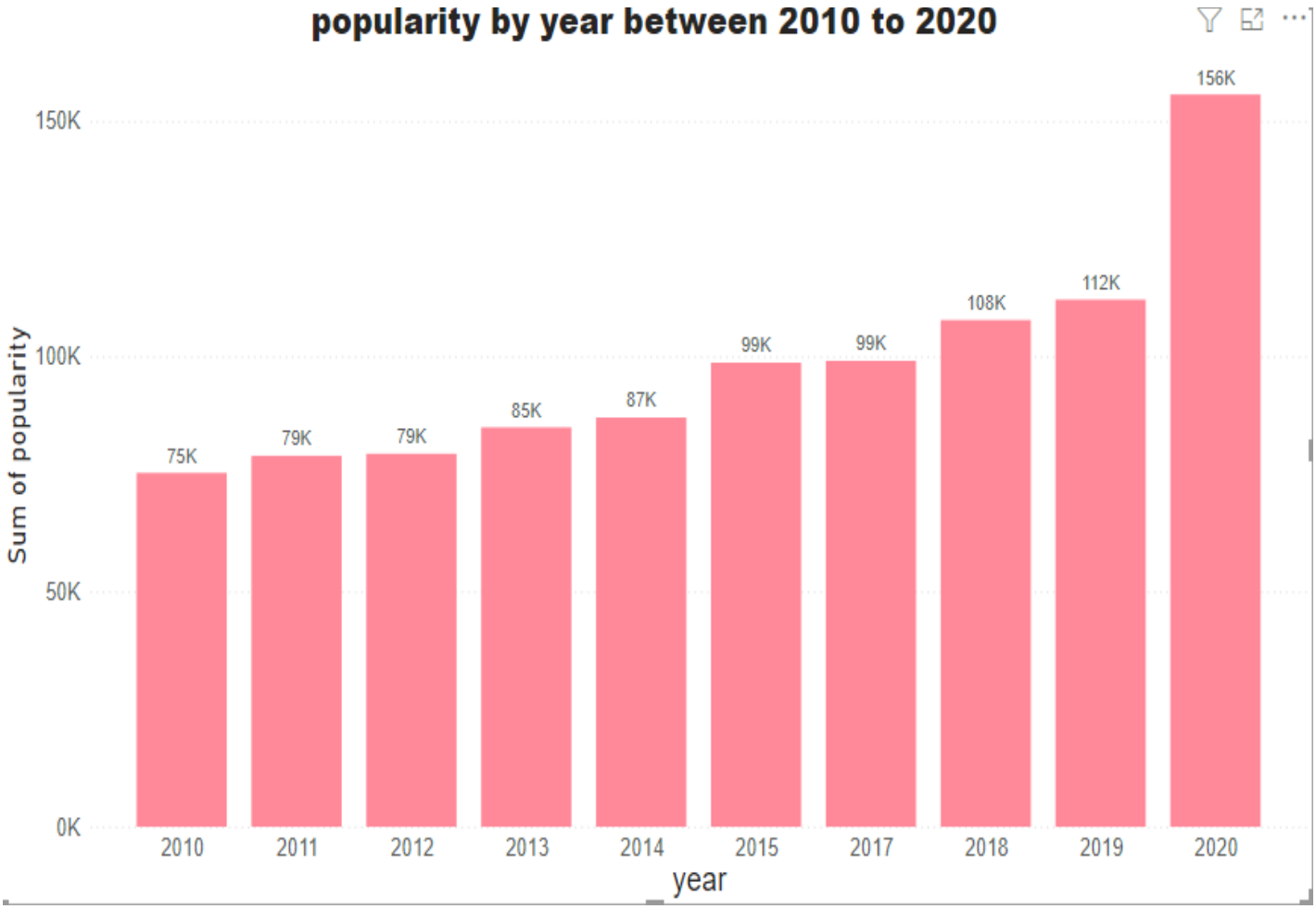


Fig 1. Popularity by Year

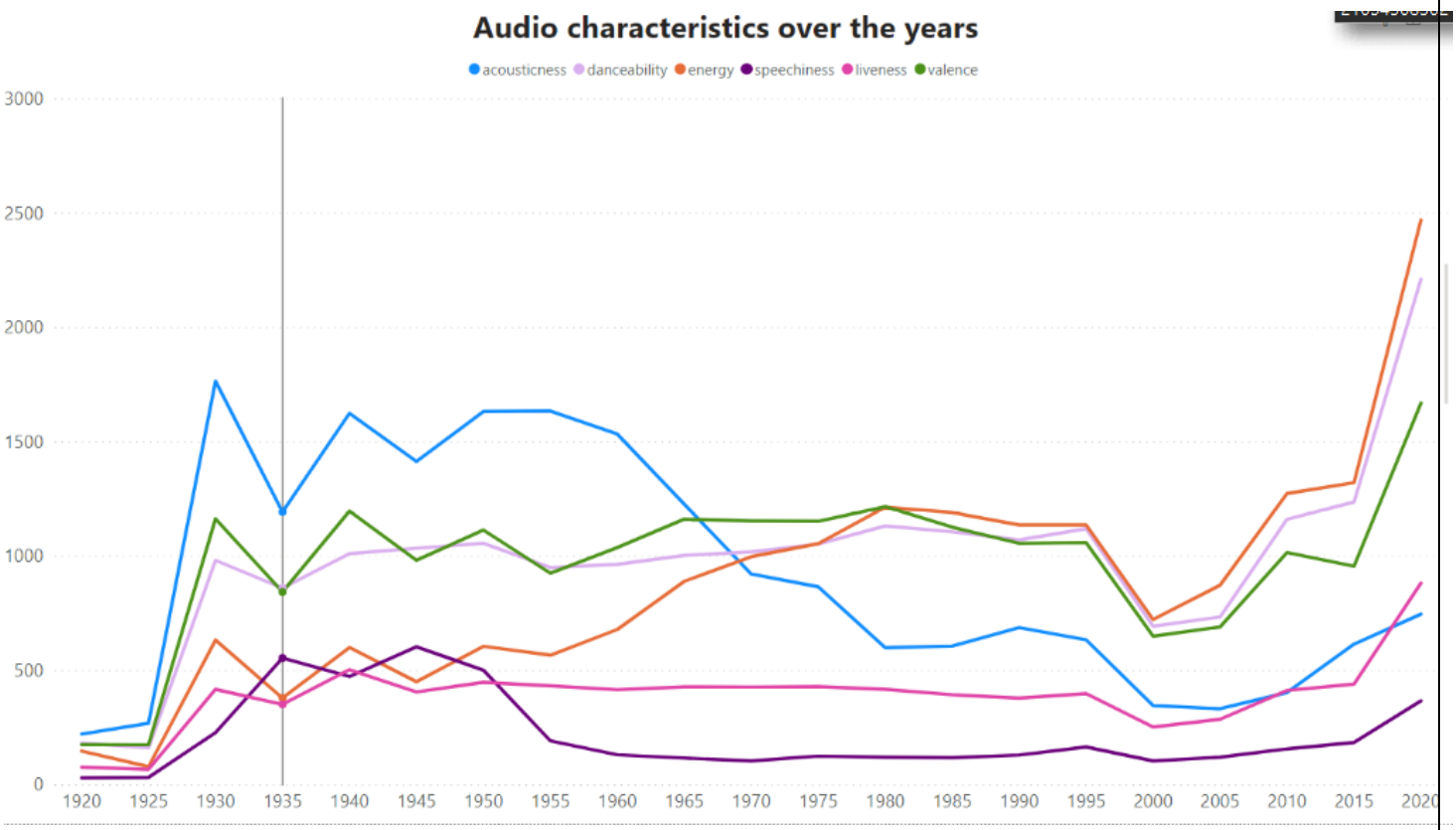


Fig 2. Audio Characteristics

Year-wise Popularity of Artists

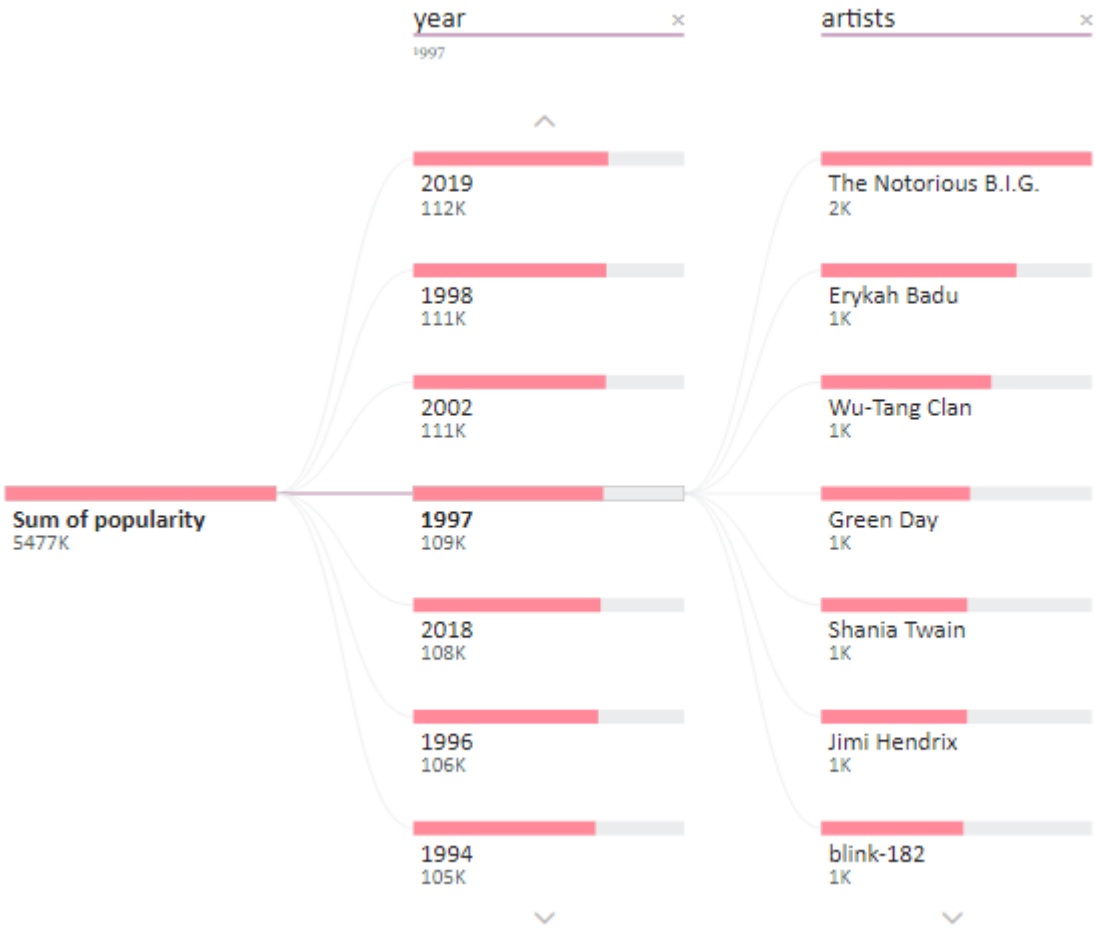


Fig 3. Popularity of Artists

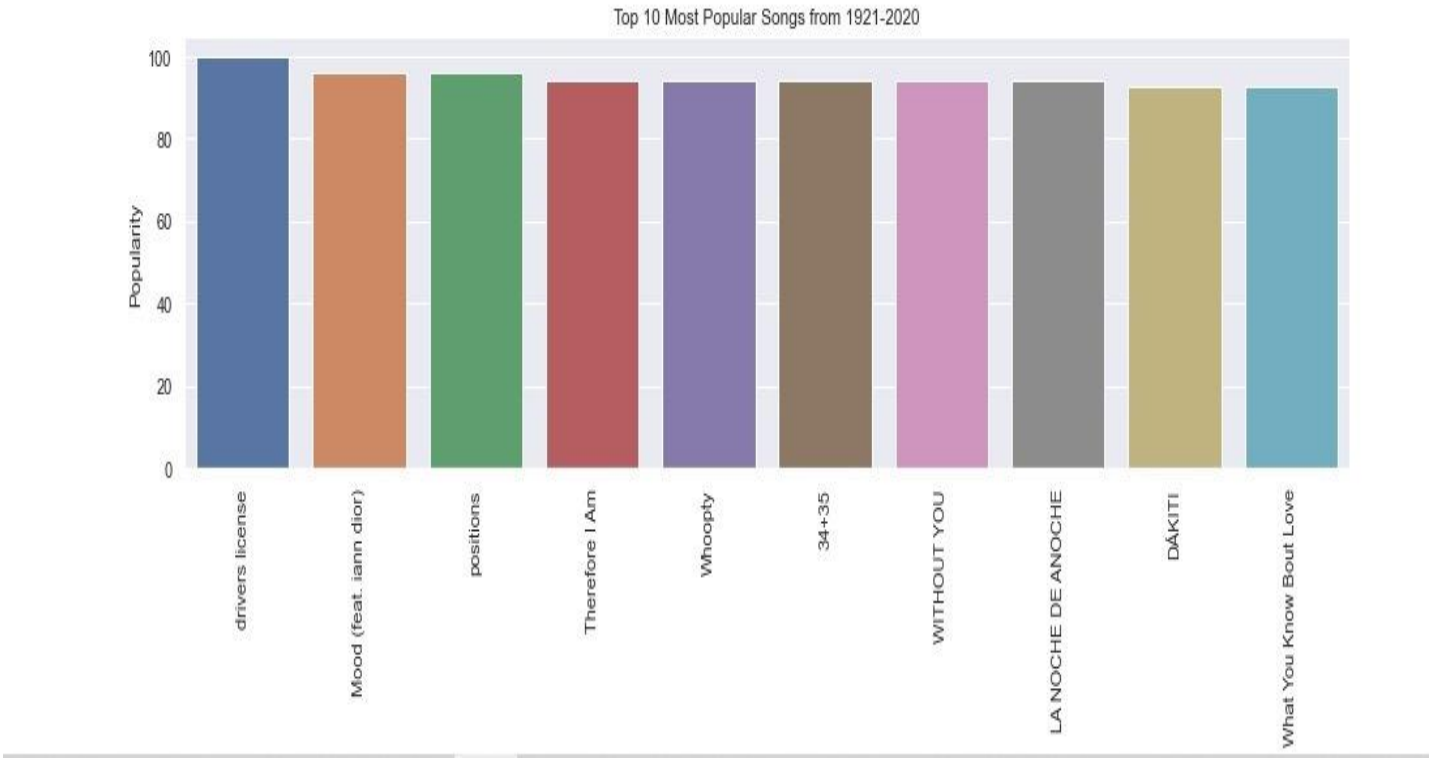


Fig 4. Top 10 Popular songs

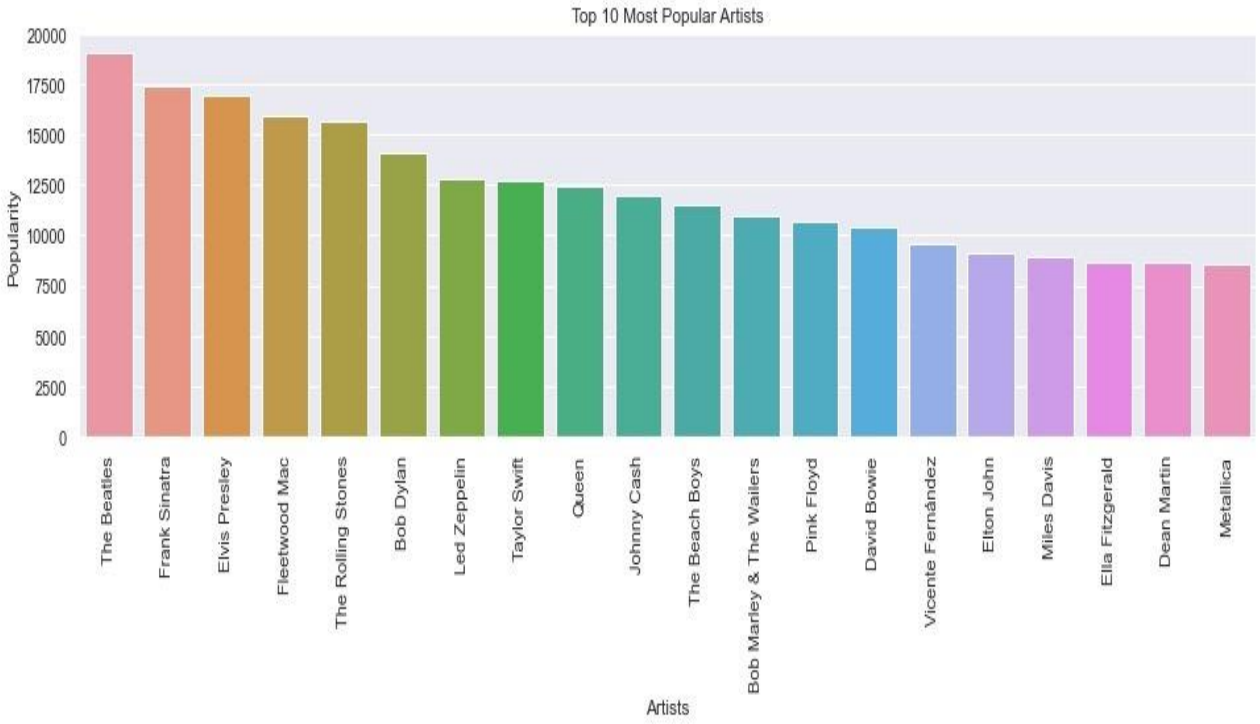


Fig 4. Top 10 Popular Artists

CONCLUSIONS

To conclude about our project, we have made analysis of different research papers and algorithm implemented in it about recommendation systems. In our project we have improvised and modified the recommendation systems. This music Recommendation System has considered many parameters like Popularity, songs name, artists name, released year etc. We successfully implemented and found thesimilar books using Manhattan distance. Also, the recommendation system was implemented using Spotify dataset and popular dataset and k-means algorithm to display most similar Songs.

FUTURE SCOPE

- For the future studies, researchers can implement other exciting machine learning Methods and there outcomes.

- For future applications, an emotional detector system that will recommend the songs by recognizing our facial emotion can be developed.

- In future researchers can also build movie recommendation system along with music recommendation on a single platform like Amazon Prime.

REFERENCES

[1] https://thecleverprogrammer.com/2021/03/03/spotify-recommendation-system-with-machine-learning/

[2] https://ahsieh53632.github.io/music-attributes-and-popularity/

[3] https://jovian.ai/nilanjan-nath/spotify-music-analysis

[4] https://www.ijert.org/music-recommendation-system-using-content-and-collaborative-filtering-methods

[5] https://www.analyticsvidhya.com/blog/2021/06/spotify-recommendation-system-using-pyspark-

[6] https://medium.com/swlh/spotify-song-prediction-and-recommendation-system-b3bbc71398ad