MINI PROJECT REPORT

on

SPOTIFY MUSIC ANALYSIS

Submitted in partial fulfillment of the requirements

of the degree of

Bachelor of Engineering in

Information Technology

by

Shivani Pandeti – 120A3050

Shreya Idate - 120A3051

Shreyas Prabhakar - 120A3052

Sonakshi Jha – 120A3056

Under the Guidance of:

Prof. Savita Lohiya

Department of Information Technology
SIES graduate School of Technology
2022-2023

CERTIFICATE

This is to certify that the Data Science Lab Mini project entitled “Spotify Music Analysis” is a bonafide work of the following students, submitted to the University of Mumbai in partial fulfillment of the requirement for the award of the degree of Bachelor of Engineering in Information Technology.

Shivani Pandeti –120A3050

Shreya Idate – 120A3051

Shreyas Prabhakar – 120A3052

Sonakshi Jha – 120A3056

PROJECT REPORT APPROVAL

This Data Science Lab Mini project report entitled ***Spotify Music Analysis*** by following students is approved for the degree of ***Bachelor of Engineering in Information Technology***.

Shivani Pandeti –120A3050

Shreya Idate – 120A3051

Shreyas Prabhakar – 120A3052

Sonakshi Jha – 120A3056

Name of External Examiner: ----------

Signature:----------

Name of Internal Examiner: ----------

Signature:----------

Date:

Place:

DECLARATION

I declare that this written submission represents my ideas in my own words and where others ideas or words have been included, I have adequately cited and referenced the original sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. I understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

Shivani Pandeti –120A3050

Shreya Idate – 120A3051

Shreyas Prabhakar – 120A3052

Sonakshi Jha – 120A3056

Date:

ACKNOWLEDGEMENT

It gives us immense pleasure to thank Dr. Atul Kemker, Principal for extending his support to carry out this project. We also thank the Head of Department Dr. K. Lakshmi Sudha for her support in completing the project. We wish to express our deep sense of gratitude and thanks to our Internal Guide, Ms. Savita Lohiya for her guidance, help and useful suggestions, which helped in completing our project work in time.

Also we would like to thank the entire faculty of the Information Technology Department for their valuable ideas and timely assistance in this project. Last but not the least, we would like to thank our non-teaching staff members of our college for their support in facilitating timely completion of this project.

Project Team

Shivani Pandeti –120A3050

Shreya Idate – 120A3051

Shreyas Prabhakar – 120A3052

Sonakshi Jha – 120A3056

CONTENTS

SR NO	TOPIC	Page No.
Chapter 1	**Abstract**	7
Chapter 2	**Introduction**	8
Chapter 3	**Literature Survey**	9
Chapter 4	**Problem Definition**	10
Chapter 5	**Proposed Methodology**	11
Chapter 6	**Code And Result**	12-18
Chapter 7	**Business Intelligence Decision**	19
Chapter 8	**Application**	20
Chapter 9	**Conclusion**	21
References		22

ABSTRACT

The Spotify Music Analysis project is focused on providing valuable insights into user behavior and music preferences on the Spotify music streaming platform.

The project's goal is to discover trends in user behavior and determine critical aspects that impact music preferences. This is accomplished through the use of machine learning algorithms and data visualization tools. The resulting insights can guide businesses in making informed decisions related to marketing strategies, product development, content creation, and other areas.

Overall, the Spotify Music Analysis project aims to leverage the power of data and analytics to provide valuable insights that can help companies make smarter, data-driven decisions in the highly competitive music streaming industry. By understanding user behavior and preferences, companies can develop better products and services that meet the needs and preferences of their customers.

INTRODUCTION

This report presents the findings of the Spotify Music Analysis project using visualization tools and K-medoids clustering algorithm. The main objective of this project is to acquire a better understanding of the behavior of Spotify users and their music preferences, and to identify the primary factors that shape these preferences.

The report also discusses the results of the K-medoids clustering algorithm, including the identification of meaningful user groups based on their listening behavior and preferences. It highlights the potential applications of these insights, including personalized music recommendations, content creation, marketing strategies, and product development.

Overall, the report provides a valuable resource for businesses and individuals interested in understanding user behavior and preferences on the Spotify platform. It demonstrates the power of visualization tools and clustering algorithms in identifying patterns and trends in large datasets, and shows how these insights can be leveraged to make informed decisions that lead to increased user engagement and retention.

LITERATURE SYSTEM

Sr.No	Author	Title & Publications	Key findings
01	Mariangela Sciandra, Irene Carola Spera.	A Model Based Approach to Spotify Data Analysis	This paper proposes a beta generalized linear mixed model (GLMM) approach for analyzing Spotify data, allowing for the identification of key factors that influence user behavior and music preferences.
02	Van den Hoven, M. J., & Boucherie, R. J.	ANALYZING SPOTIFY DATA	This paper presents a performance analysis of a music streaming service, focusing on its queuing system and resource allocation strategies.
03	T.Velmurugan, Dr.T.Santhanam	Performance Analysis Of K-Means And K-Medoids Clustering Algorithms For A Randomly Generated Data Set	This paper presents a performance analysis of the K-Means and K-Medoids clustering algorithms for a randomly generated data set.

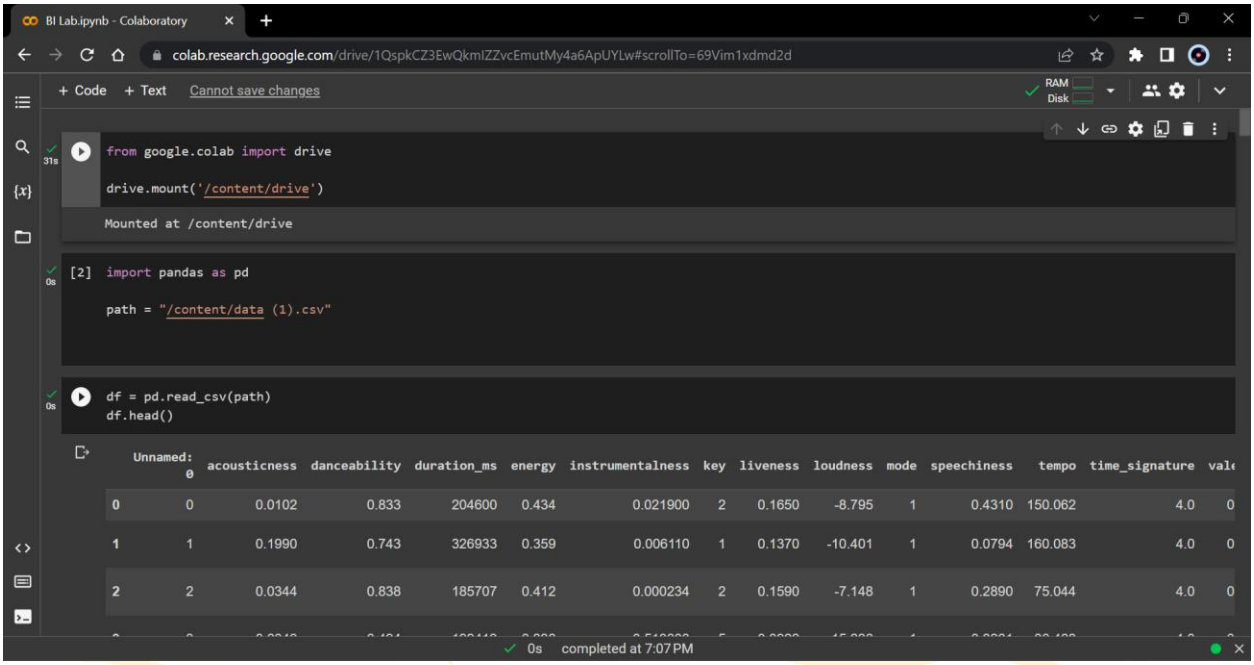
PROBLEM DEFINITION

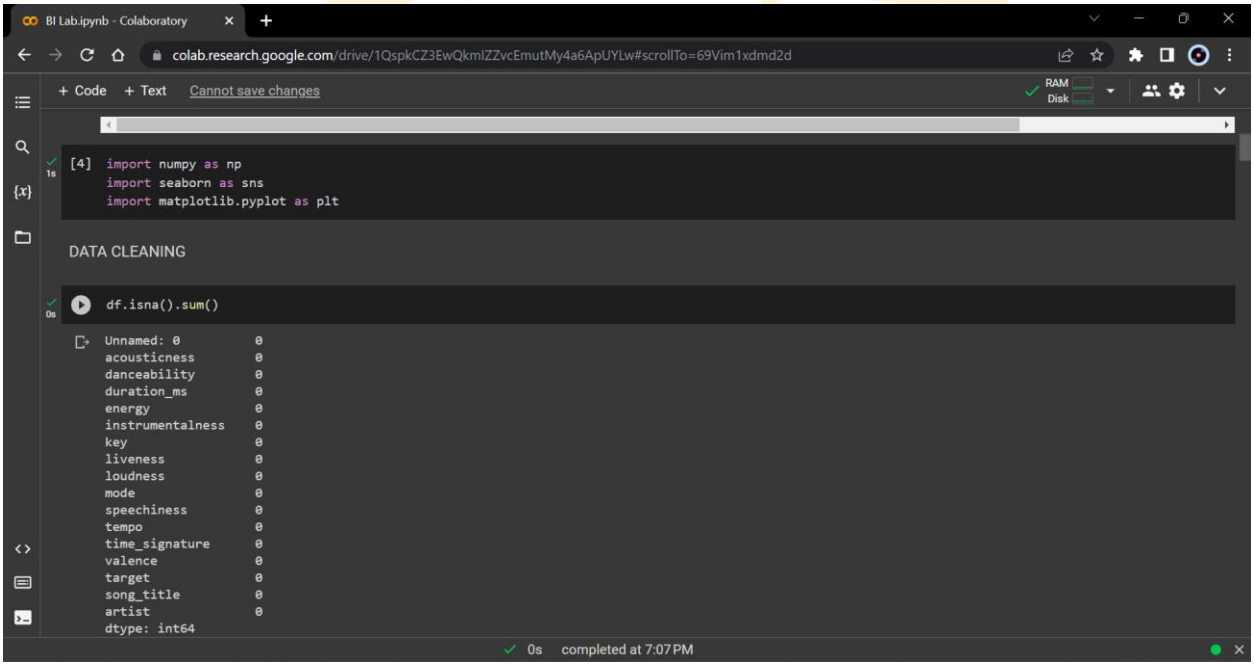
The problem definition for the Spotify Music Analysis project using visualization tools and K-medoids clustering algorithm is to identify patterns in user behavior and music preferences on the Spotify platform, and to determine the key factors that influence these preferences. The goal is to gain insights that can inform business decisions related to marketing strategies, content creation, and product development, among other areas.

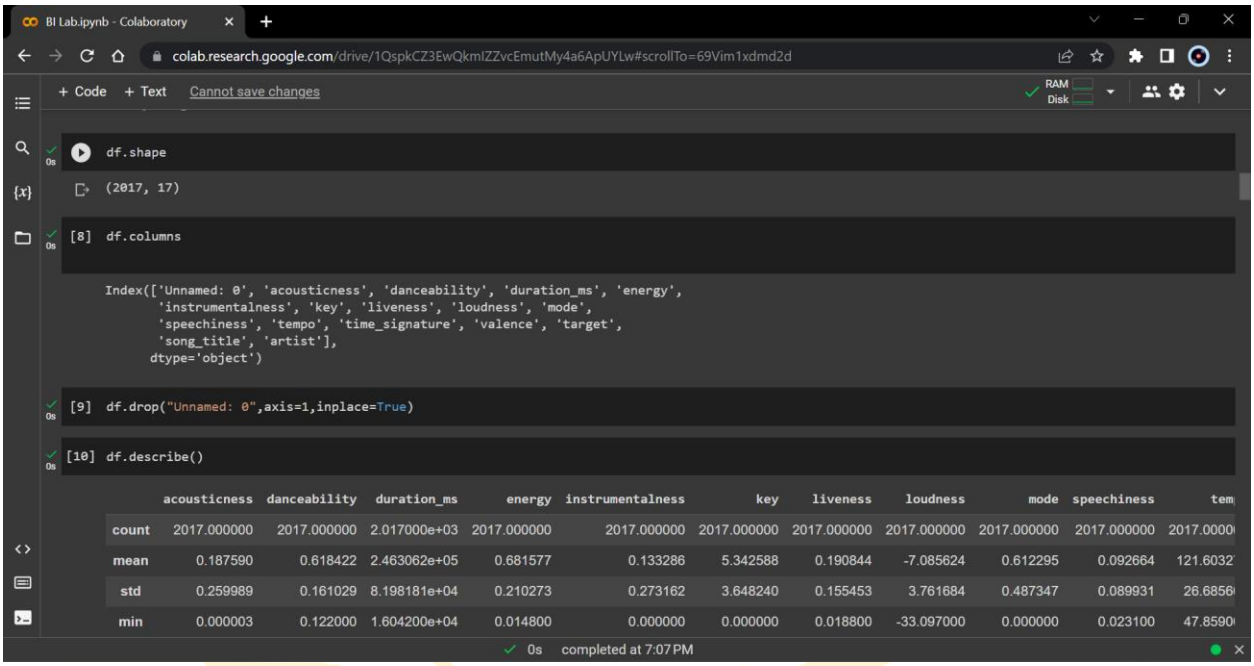
PROPOSED METHODOLOGY

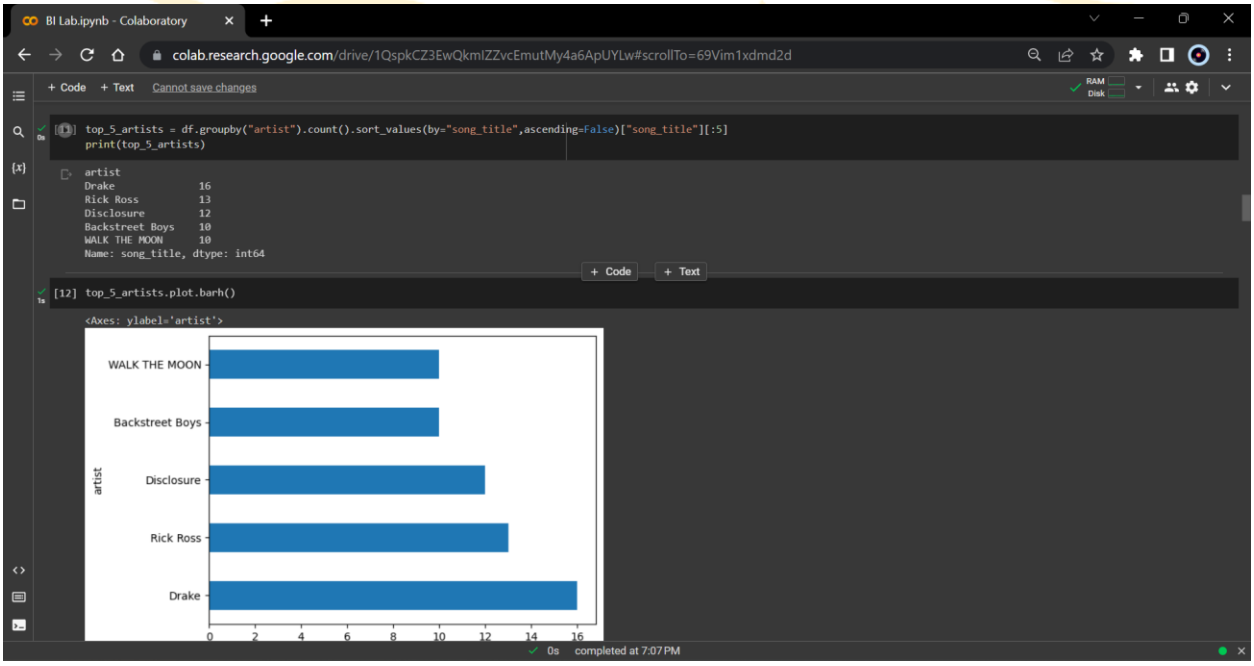
The proposed methodology for Spotify Music Analysis using K-Medoids involves several steps. First, data is collected on user behavior and music preferences from the Spotify platform. The data is then preprocessed to remove duplicates, missing values, and outliers. Relevant features are extracted from the preprocessed data, which are used to cluster users based on their music preferences using the K-Medoids clustering algorithm. The identified clusters are then analyzed to understand the characteristics of each group and identify key factors that influence their music preferences. Finally, the performance of the K-Medoids algorithm is evaluated on the Spotify data. This methodology allows for gaining insights into user behavior and preferences, which can be useful for informing business decisions in the music industry.

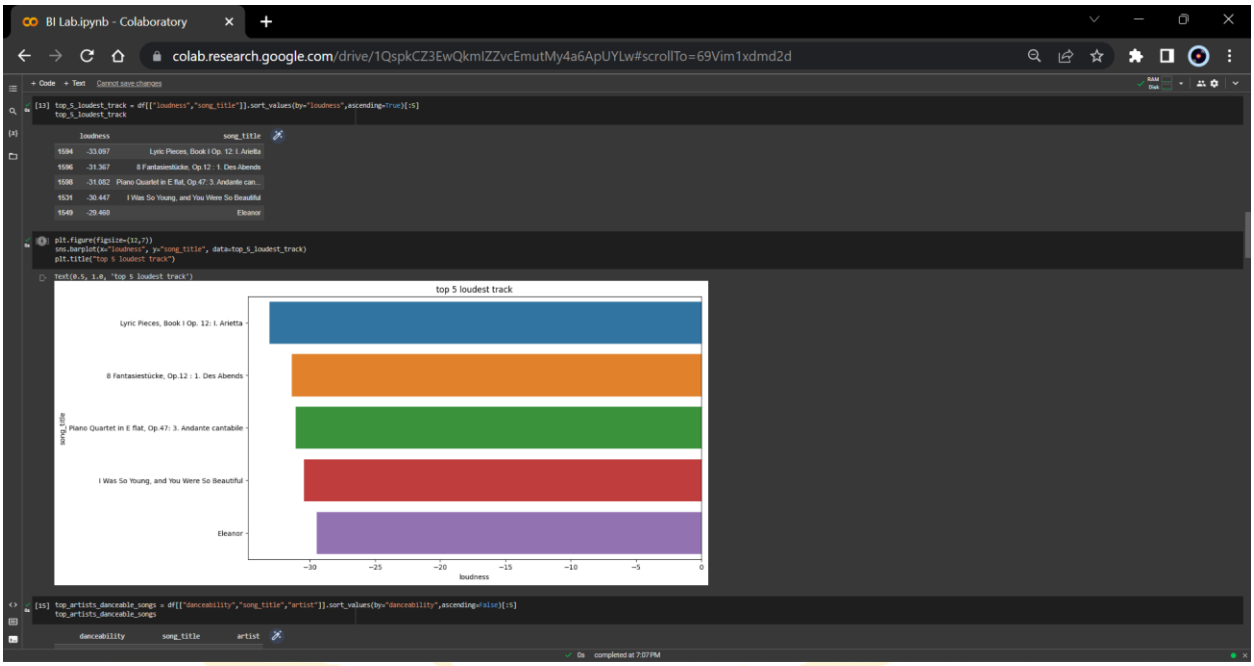
CODE AND RESULT

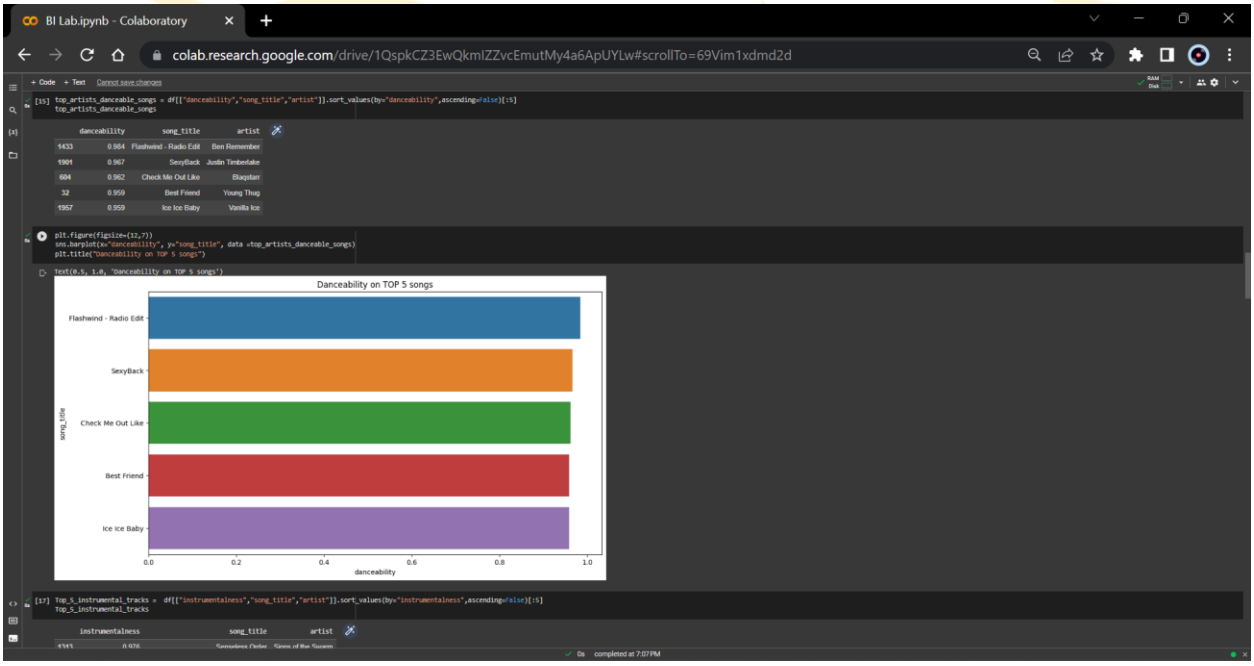


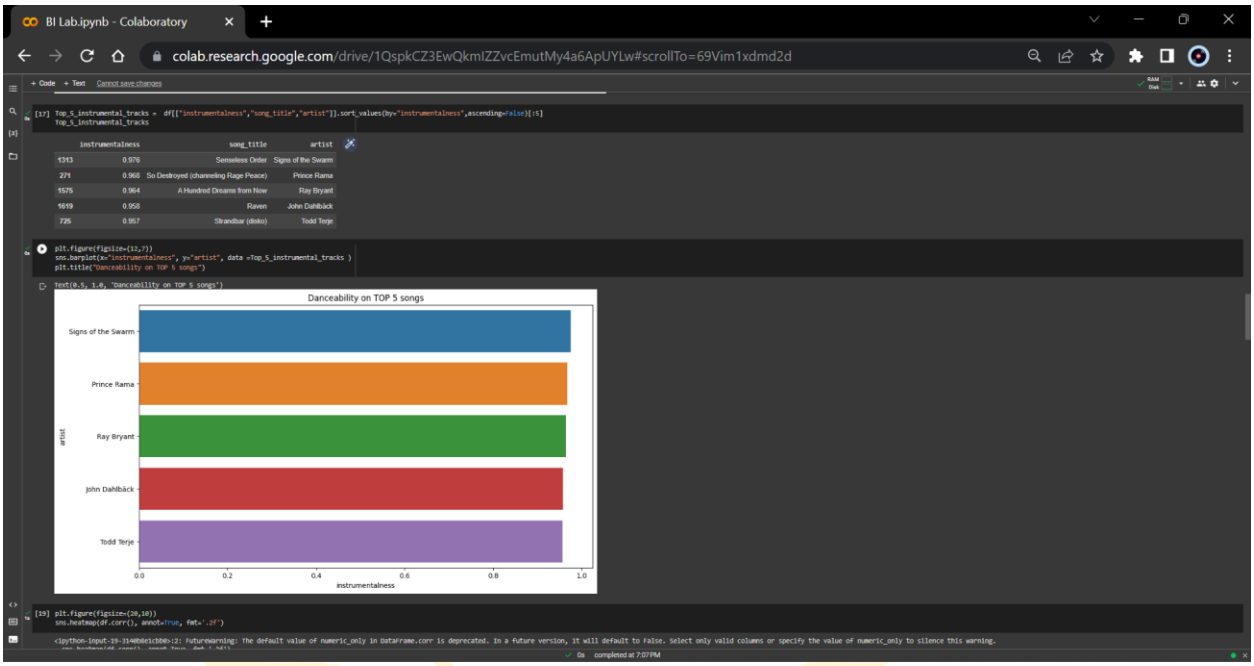



BI Lab.ipynb - Colaboratory
colab.research.google.com/drive/1QspkCZ3EwQkmIZZvcEmutMy4a6ApUYLw#scrollTo=69Vim1xdmd2d
+ Code + Text Cannot save changes
RAM
Disk
df.shape
(2017, 17)
[8] df.columns
Index(['Unnamed: 0', 'acousticness', 'danceability', 'duration_ms', 'energy',
'instrumentalness', 'key', 'liveness', 'loudness', 'mode',
'speechiness', 'tempo', 'time_signature', 'valence', 'target',
'song_title', 'artist'],
dtype='object')
[9] df.drop("Unnamed: 0",axis=1,inplace=True)
[10] df.describe()
acousticness danceability duration_ms energy instrumentalness key liveness loudness mode speechiness tem
count 2017.000000 2017.000000 2.017000e+03 2017.000000 2017.000000 2017.000000 2017.000000 2017.000000 2017.000000 2017.000000 2017.0000
mean 0.187590 0.618422 2.463062e+05 0.681577 0.133286 5.342588 0.190844 -7.085624 0.612295 0.092664 121.6032
std 0.259989 0.161029 8.198181e+04 0.210273 0.273162 3.648240 0.155453 3.761684 0.487347 0.089931 26.6856
min 0.000003 0.122000 1.604200e+04 0.014800 0.000000 0.000000 0.018800 -33.097000 0.000000 0.023100 47.8590
0s completed at 7:07 PM

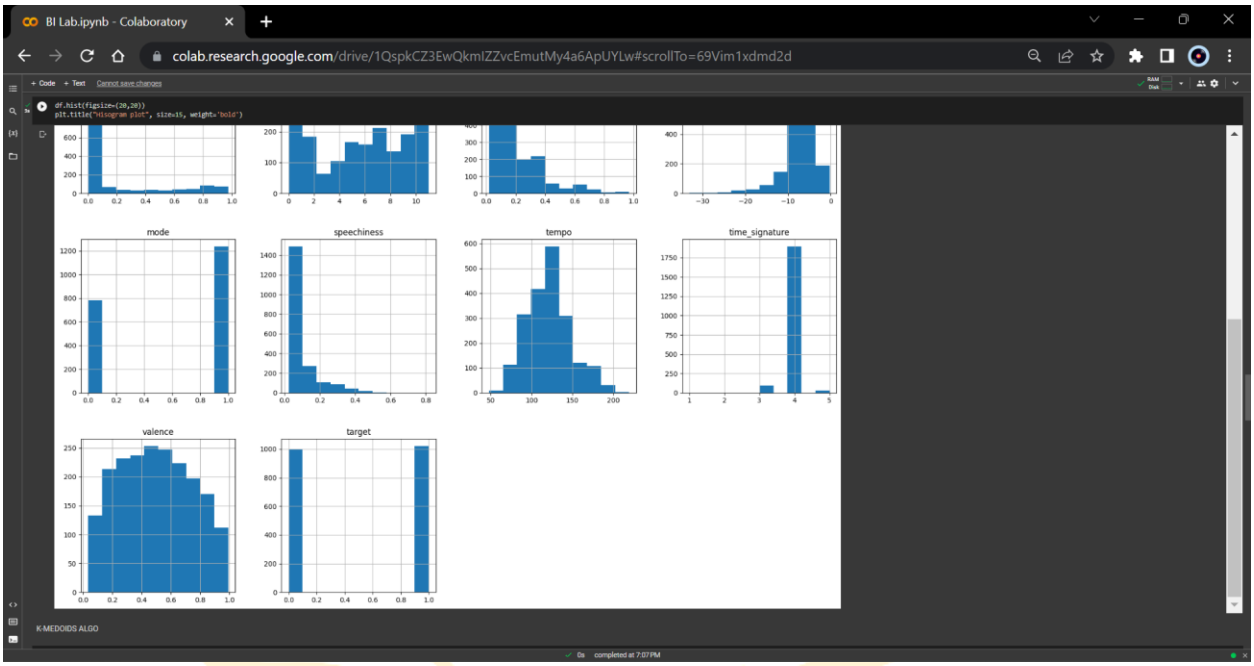

BI Lab.ipynb - Colaboratory
colab.research.google.com/drive/1QspkCZ3EwQkmIZZvcEmutMy4a6ApUYLw#scrollTo=69Vim1xdmd2d
+ Code + Text Cannot save changes
top_5_artists = df.groupby("artist").count().sort_values(by="song_title",ascending=False)["song_title"][:5]
print(top_5_artists)
artist
Drake 16
Rick Ross 13
Disclosure 12
Backstreet Boys 10
WALK THE MOON 10
Name: song_title, dtype: int64
+ Code + Text
[12] top_5_artists.plot.barh()
<Axes: ylabel='artist'>
WALK THE MOON
Backstreet Boys
Disclosure
Rick Ross
Drake
artist
0 2 4 6 8 10 12 14 16
0s completed at 7:07 PM

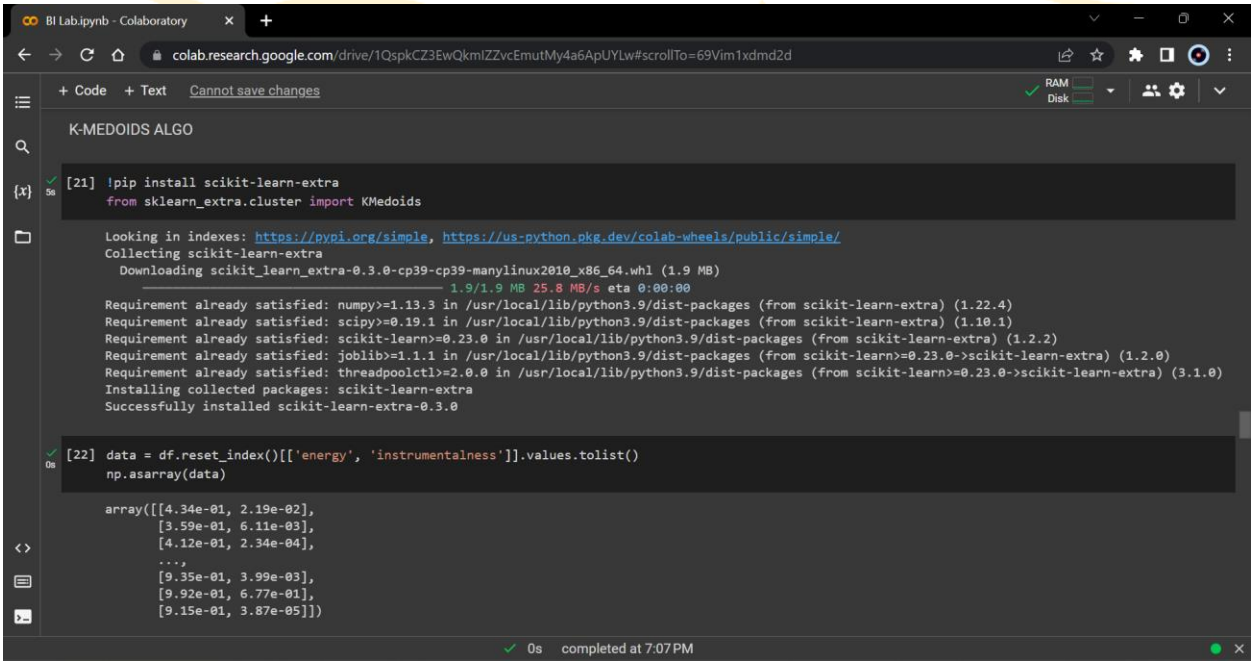

BI Lab.ipynb - Colaboratory
colab.research.google.com/drive/1QspkCZ3EwQkmIZZvcEmutMy4a6ApUYLw#scrollTo=69Vim1xdmd2d
top_5_loudest_track = df[["loudness","song_title"]].sort_values(by="loudness",ascending=True)[:5]
top_5_loudest_track
loudness song_title
1594 -33.097 Lyric Pieces, Book I Op. 12: I. Arietta
1596 -31.367 8 Fantasiestücke, Op.12 : 1. Des Abends
1598 -31.082 Piano Quartet in E flat, Op.47: 3. Andante can...
1631 -30.447 I Was So Young, and You Were So Beautiful
1549 -29.469 Eleanor
plt.figure(figsize=(12,7))
sns.barplot(x="loudness", y="song_title", data=top_5_loudest_track)
plt.title("top 5 loudest track")
Text(0.5, 1.0, 'top 5 loudest track')
top 5 loudest track
Lyric Pieces, Book I Op. 12: I. Arietta
8 Fantasiestücke, Op.12 : 1. Des Abends
Piano Quartet in E flat, Op.47: 3. Andante cantabile
I Was So Young, and You Were So Beautiful
Eleanor
song_title
−30
−25
−20
−15
−10
−5
0
loudness
top_artists_danceable_songs = df[["danceability","song_title","artist"]].sort_values(by="danceability",ascending=False)[:5]
top_artists_danceable_songs
danceability song_title artist
completed at 7:07 PM

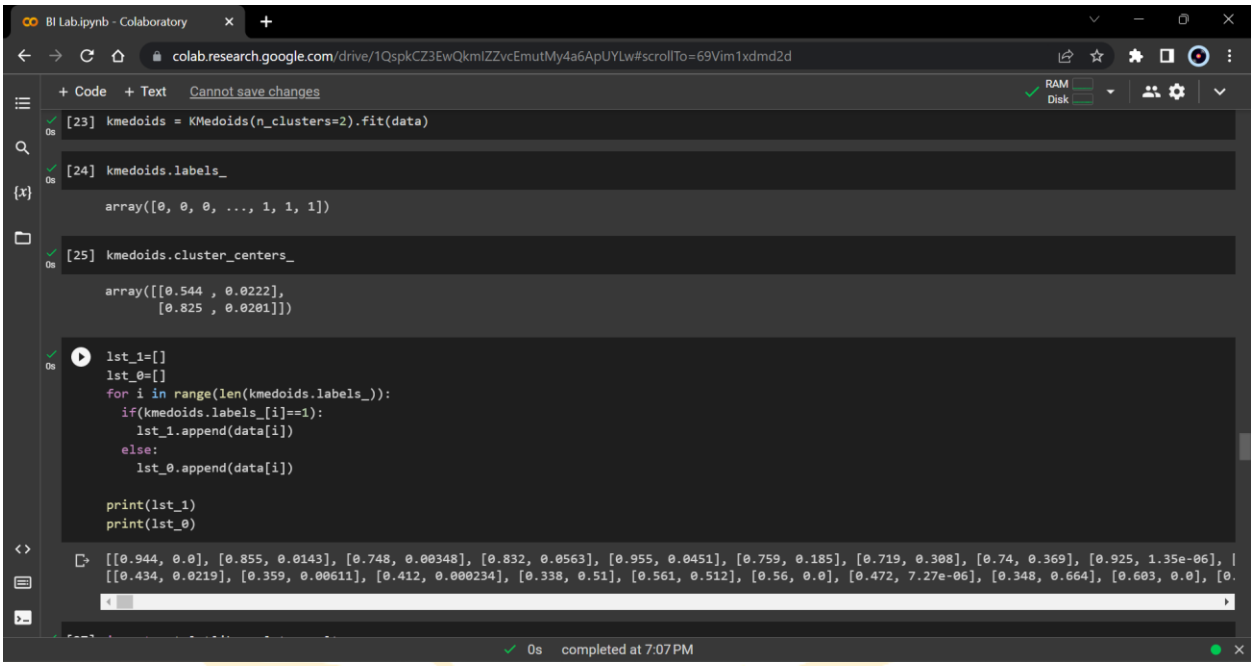

BI Lab.ipynb - Colaboratory
colab.research.google.com/drive/1QspkCZ3EwQkmIZZvcEmutMy4a6ApUYLw#scrollTo=69Vim1xdmd2d
top_artists_danceable_songs = df[["danceability","song_title","artist"]].sort_values(by="danceability",ascending=False)[:5]
top_artists_danceable_songs
danceability song_title artist
1433 0.984 Flashwind - Radio Edit Ben Remember
1901 0.967 SexyBack Justin Timberlake
604 0.962 Check Me Out Like Blagstarr
32 0.959 Best Friend Young Thug
1957 0.959 Ice Ice Baby Vanilla Ice
plt.figure(figsize=(12,7))
sns.barplot(x="danceability", y="song_title", data =top_artists_danceable_songs)
plt.title("Danceability on TOP 5 songs")
Text(0.5, 1.0, 'Danceability on TOP 5 songs')
Danceability on TOP 5 songs
Flashwind - Radio Edit
SexyBack
Check Me Out Like
Best Friend
Ice Ice Baby
song_title
0.0
0.2
0.4
0.6
0.8
1.0
danceability
Top_5_instrumental_tracks = df[["instrumentalness","song_title","artist"]].sort_values(by="instrumentalness",ascending=False)[:5]
Top_5_instrumental_tracks
instrumentalness song_title artist
completed at 7:07 PM

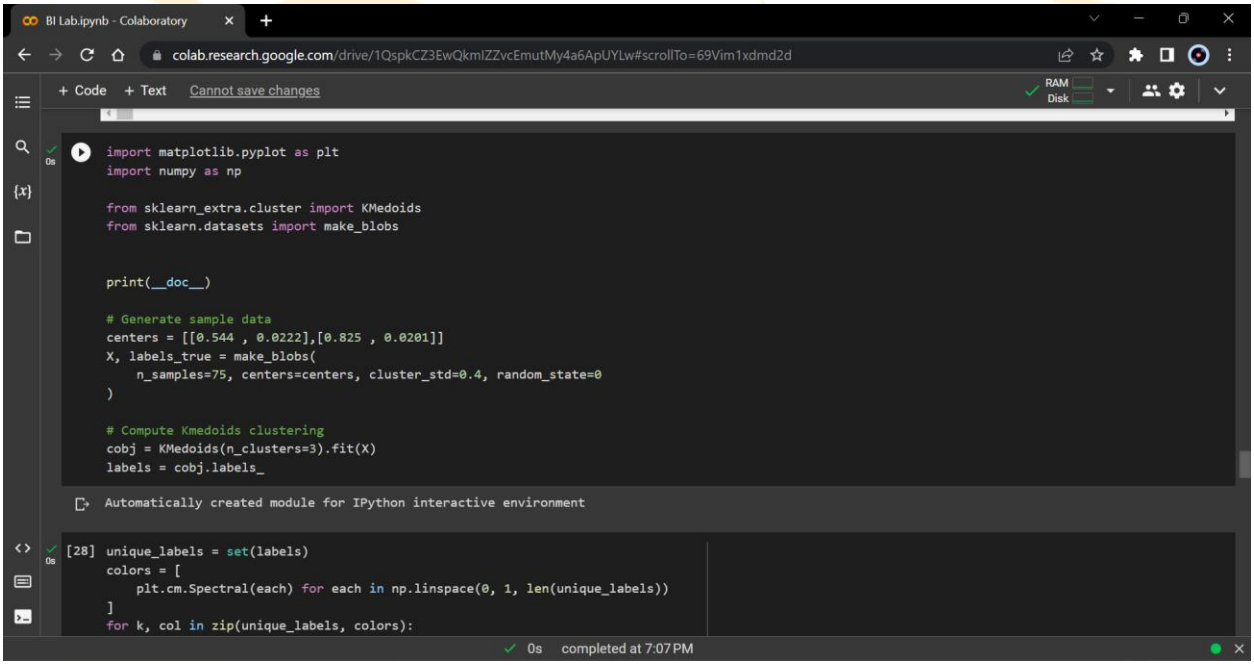

BI Lab.ipynb - Colaboratory
colab.research.google.com/drive/1QspkCZ3EwQkmIZZvcEmutMy4a6ApUYLw#scrollTo=69Vim1xdmd2d
Danceability on TOP 5 songs
Signs of the Swarm
Prince Rama
Ray Bryant
John Dahlbäck
Todd Terje
artist
instrumentalness

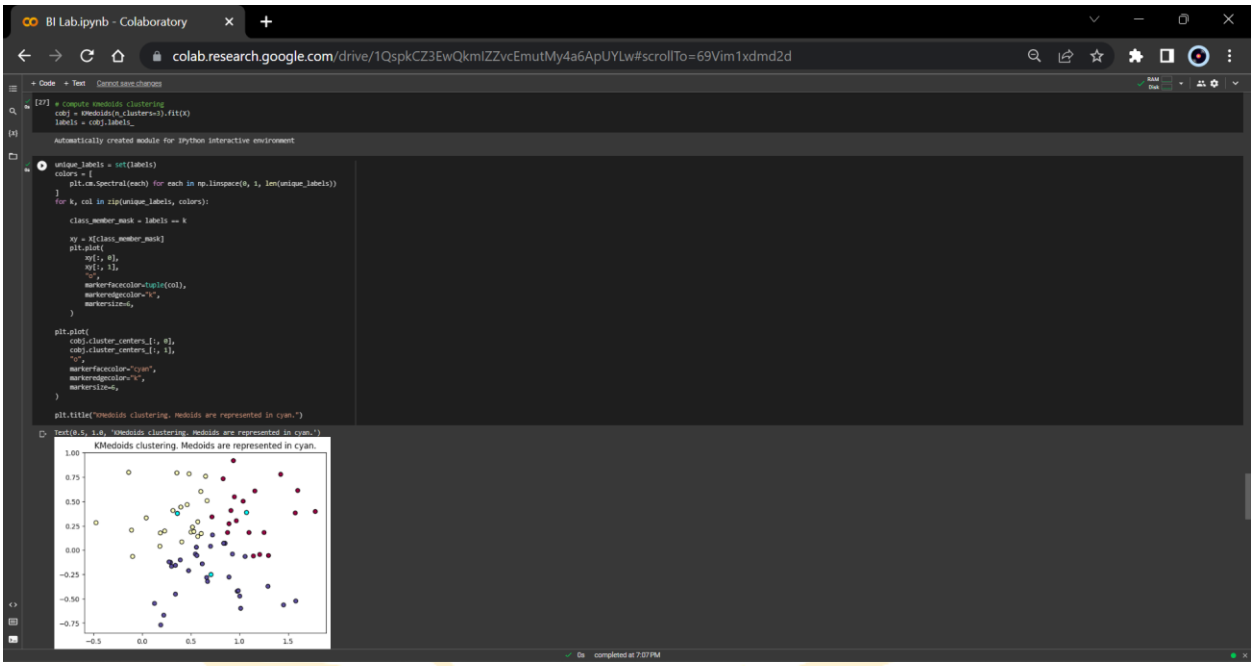

BI Lab.ipynb - Colaboratory
colab.research.google.com/drive/1QspkCZ3EwQkmIZZvcEmutMy4a6ApUYLw#scrollTo=69Vim1xdmd2d
acousticness
danceability
duration_ms
energy
instrumentalness
key
liveness
loudness
mode
speechiness
tempo
time_signature
valence
target


BI Lab.ipynb - Colaboratory
colab.research.google.com/drive/1QspkCZ3EwQkmIZZvcEmutMy4a6ApUYLw#scrollTo=69Vim1xdmd2d
mode
speechiness
tempo
time_signature
valence
target
K-MEDOIDS ALGO


BI Lab.ipynb - Colaboratory
colab.research.google.com/drive/1QspkCZ3EwQkmIZZvcEmutMy4a6ApUYLw#scrollTo=69Vim1xdmd2d
+ Code + Text Cannot save changes
RAM
Disk
K-MEDOIDS ALGO
[21] !pip install scikit-learn-extra
from sklearn_extra.cluster import KMedoids
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/public/simple/
Collecting scikit-learn-extra
Downloading scikit_learn_extra-0.3.0-cp39-cp39-manylinux2010_x86_64.whl (1.9 MB)
1.9/1.9 MB 25.8 MB/s eta 0:00:00
Requirement already satisfied: numpy>=1.13.3 in /usr/local/lib/python3.9/dist-packages (from scikit-learn-extra) (1.22.4)
Requirement already satisfied: scipy>=0.19.1 in /usr/local/lib/python3.9/dist-packages (from scikit-learn-extra) (1.10.1)
Requirement already satisfied: scikit-learn>=0.23.0 in /usr/local/lib/python3.9/dist-packages (from scikit-learn-extra) (1.2.2)
Requirement already satisfied: joblib>=1.1.1 in /usr/local/lib/python3.9/dist-packages (from scikit-learn>=0.23.0->scikit-learn-extra) (1.2.0)
Requirement already satisfied: threadpoolctl>=2.0.0 in /usr/local/lib/python3.9/dist-packages (from scikit-learn>=0.23.0->scikit-learn-extra) (3.1.0)
Installing collected packages: scikit-learn-extra
Successfully installed scikit-learn-extra-0.3.0
[22] data = df.reset_index()[['energy', 'instrumentalness']].values.tolist()
np.asarray(data)
array([[4.34e-01, 2.19e-02],
[3.59e-01, 6.11e-03],
[4.12e-01, 2.34e-04],
...,
[9.35e-01, 3.99e-03],
[9.92e-01, 6.77e-01],
[9.15e-01, 3.87e-05]])
0s completed at 7:07 PM


BI Lab.ipynb - Colaboratory
colab.research.google.com/drive/1QspkCZ3EwQkmIZZvcEmutMy4a6ApUYLw#scrollTo=69Vim1xdmd2d
+ Code + Text Cannot save changes
RAM
Disk
[23] kmedoids = KMedoids(n_clusters=2).fit(data)
[24] kmedoids.labels_
array([0, 0, 0, ..., 1, 1, 1])
[25] kmedoids.cluster_centers_
array([[0.544 , 0.0222],
[0.825 , 0.0201]])
lst_1=[]
lst_0=[]
for i in range(len(kmedoids.labels_)):
if(kmedoids.labels_[i]==1):
lst_1.append(data[i])
else:
lst_0.append(data[i])
print(lst_1)
print(lst_0)
[[0.944, 0.0], [0.855, 0.0143], [0.748, 0.00348], [0.832, 0.0563], [0.955, 0.0451], [0.759, 0.185], [0.719, 0.308], [0.74, 0.369], [0.925, 1.35e-06],
[[0.434, 0.0219], [0.359, 0.00611], [0.412, 0.000234], [0.338, 0.51], [0.561, 0.512], [0.56, 0.0], [0.472, 7.27e-06], [0.348, 0.664], [0.603, 0.0], [0.
0s completed at 7:07 PM


BI Lab.ipynb - Colaboratory
colab.research.google.com/drive/1QspkCZ3EwQkmIZZvcEmutMy4a6ApUYLw#scrollTo=69Vim1xdmd2d
+ Code + Text Cannot save changes
RAM
Disk
import matplotlib.pyplot as plt
import numpy as np
from sklearn_extra.cluster import KMedoids
from sklearn.datasets import make_blobs
print(__doc__)
# Generate sample data
centers = [[0.544 , 0.0222],[0.825 , 0.0201]]
X, labels_true = make_blobs(
n_samples=75, centers=centers, cluster_std=0.4, random_state=0
)
# Compute Kmedoids clustering
cobj = KMedoids(n_clusters=3).fit(X)
labels = cobj.labels_
Automatically created module for IPython interactive environment
[28] unique_labels = set(labels)
colors = [
plt.cm.Spectral(each) for each in np.linspace(0, 1, len(unique_labels))
]
for k, col in zip(unique_labels, colors):
0s completed at 7:07 PM


BI Lab.ipynb - Colaboratory
colab.research.google.com/drive/1QspkCZ3EwQkmIZZvcEmutMy4a6ApUYLw#scrollTo=69Vim1xdmd2d
+ Code + Text Cannot save changes
[27] # compute kmedoids clustering
cobj = KMedoids(n_clusters=3).fit(X)
labels = cobj.labels_
Automatically created module for IPython interactive environment
unique_labels = set(labels)
colors = [
plt.cm.Spectral(each) for each in np.linspace(0, 1, len(unique_labels))
]
for k, col in zip(unique_labels, colors):
class_member_mask = labels == k
xy = X[class_member_mask]
plt.plot(
xy[:, 0],
xy[:, 1],
"o",
markerfacecolor=tuple(col),
markeredgecolor="k",
markersize=6,
)
plt.plot(
cobj.cluster_centers_[:, 0],
cobj.cluster_centers_[:, 1],
"o",
markerfacecolor="cyan",
markeredgecolor="k",
markersize=6,
)
plt.title("KMedoids clustering. Medoids are represented in cyan.")
Text(0.5, 1.0, 'KMedoids clustering. Medoids are represented in cyan.')
KMedoids clustering. Medoids are represented in cyan.
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−0.5
0.0
0.5
1.0
1.5
0s completed at 7:07 PM

BUSINESS INTELLIGENCE DECISION

- The project analyzes large amounts of data from the Spotify music streaming platform using K-melodies and data visualization tools.
- The goal of the project is to identify patterns in user behavior and determine the key factors that influence music preference.
- By leveraging the power of data and analytics, the project can provide valuable insights that help companies make informed decisions in the highly competitive music streaming industry.
- The project can help companies develop more effective marketing campaigns by targeting specific user groups with relevant music recommendations.

APPLICATIONS

The Spotify Music Analysis project has several potential applications that can benefit individuals and businesses alike. Here are some of them:

- Personalized music recommendations: By analyzing a user's listening history and preferences, the project can suggest highly personalized music recommendations to users, making it easier for them to discover new music that aligns with their taste.
- Industry analysis: The project can provide insights into the music streaming industry as a whole, such as identifying the most popular music genres and the factors that influence user behavior and preferences, enabling businesses to make informed decisions that help them stay ahead of the competition.
- Product development: The insights gained from the project can inform product development strategies, such as identifying the most popular music genres and features among users, enabling businesses to create products that cater to their users' needs and preferences.

CONCLUSION

In conclusion, the Spotify Music Analysis project that utilizes K-medoids clustering algorithm provides valuable insights into user behavior and music preferences on the Spotify music streaming platform.

The K-medoids clustering algorithm is particularly useful in this context as it enables the identification of user groups based on their listening behavior, making it easier to create targeted marketing campaigns and personalize music recommendations. By utilizing this algorithm, the Spotify Music Analysis project can provide a more personalized user experience, leading to increased user engagement and retention.

Overall, the Spotify Music Analysis project that utilizes K-medoids clustering algorithm is a valuable tool in the music streaming industry that can help businesses make informed decisions and stay ahead of the competition.

REFERENCES

[1]Raza, A., Baek, J., & Kwon, Y. (2020). A Model-Based Approach to Spotify Data Analysis: A Beta GLMM. In 2020 IEEE International Conference on Big Data (Big Data) (pp. 3265-3272).

[2] Van den Hoven, M. J., & Boucherie, R. J. (2011). Performance Analysis of a Music Streaming Service. In 2011 IEEE International Symposium on Performance Analysis of Systems and Software (ISPASS) (pp. 45-54)

[3] Jain, P. K., & Singh, U. (2012). Performance Analysis of K-Means and K-Medoids Clustering Algorithms for a Randomly Generated Data Set. In 2012 IEEE 2nd International Conference on Computational Intelligence and Computing Research (pp. 1-6).