

FINAL PROJECT

+ + + CLUSTERING TOP SPOTIFY SONGS

Based on Rock Genre

Dalitri Oktaviani Saptarini

- + **BACKGROUND**
- + **DATA PREPROCESSING**
- + **MODELING**
- + **ANALYSIS**
- + **RECOMMENDATION**

CONTENTS




BACKGROUND

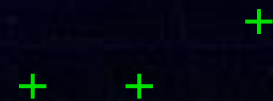


SPOTIFY

Spotify's platform revolutionized music listening forever when it launched in 2008. Spotify's move into podcasting brought innovation and a new generation of listeners to the medium, and in 2022 Spotify entered the next audio market primed for growth with the addition of audiobooks.



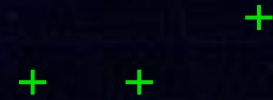
CASE STUDY



In today's era, there are many platforms to listen to music. Therefore, it is important to improve user experience and engagement by providing personalized song recommendations (in this project specifically in the rock genre).

By clustering songs based on user preferences and characteristics, platforms can offer customized playlists and recommendations, thereby increasing user satisfaction and retention, but what efforts should be made to increase user satisfaction and retention?

BUSINESS OBJECTIVE



Why is improving user experience and engagement important?

It is important to increasing user satisfaction and retention.

What efforts should be made?

By clustering songs based on user preferences and characteristics, platforms can offer customized playlists and recommendations.

CURRENT DATA

Data

Source: [Kaggle](#)

There are 105669 songs data with 25 features

Rock Genre

will find the best number of clusters using relevant features

Non Rock Genre

will not be analyzed further in this project

OBJECTIVES

+

+

Modeling

Clustering uses features that have an influence on the type of rock music, namely **Energy, Loudness, Tempo, Valence, Mode**.

Analysis

Analysis of the characteristics of each cluster resulting from segmentation.

Recommendation

Provide a playlist of rock sub-genre songs



DATA

PREPROCESSING



PREPROCESSING STEPS

1. Data Collecting
2. Data Cleansing
3. Data Scaling

DATASET INFORMATION

The dataset consist of:

105669 rows

where there are no duplicated values but have missing values in the database

25 features

- spotify_id
- name
- artists
- daily_rank
- daily_movement
- weekly_movement
- country
- snapshot_date
- popularity
- is_explicit
- duration_ms
- album_name
- album_release_date
- danceability
- energy
- key
- loudness
- mode
- speechiness
- acousticness
- instrumentalness
- liveness
- valence
- tempo
- time_signature

DATA CLEANSING

Fill the NaN in country column with unknown country

country
NaN
NaN
NaN
NaN
NaN



country
unknown country
unknown country
unknown country
unknown country
unknown country

DATA CLEANSING

Drop missing values

```
# check missing values per column
df.isna().sum()

spotify_id      0
name            21
artists         21
daily_rank      0
daily_movement  0
weekly_movement 0
country         0
snapshot_date   0
popularity      0
is_explicit     0
duration_ms     0
album_name      22
album_release_date 22
danceability     0
energy          0
key             0
loudness        0
mode            0
speechiness     0
acousticness    0
instrumentalness 0
liveness        0
valence         0
tempo           0
time_signature   0
dtype: int64
```

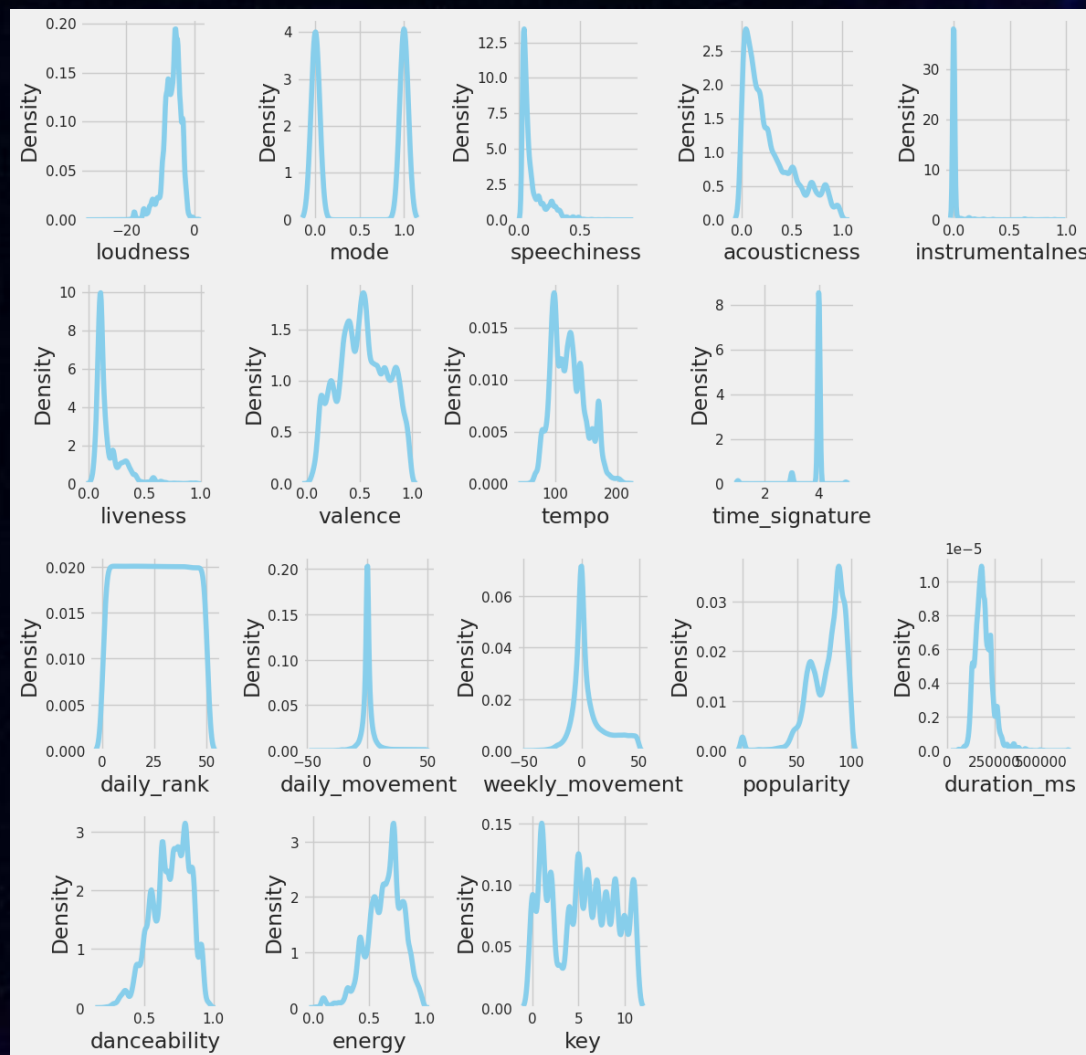


```
# check missing values per column

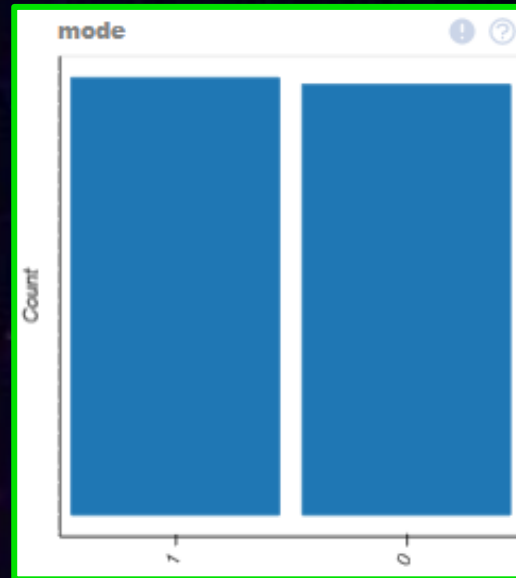
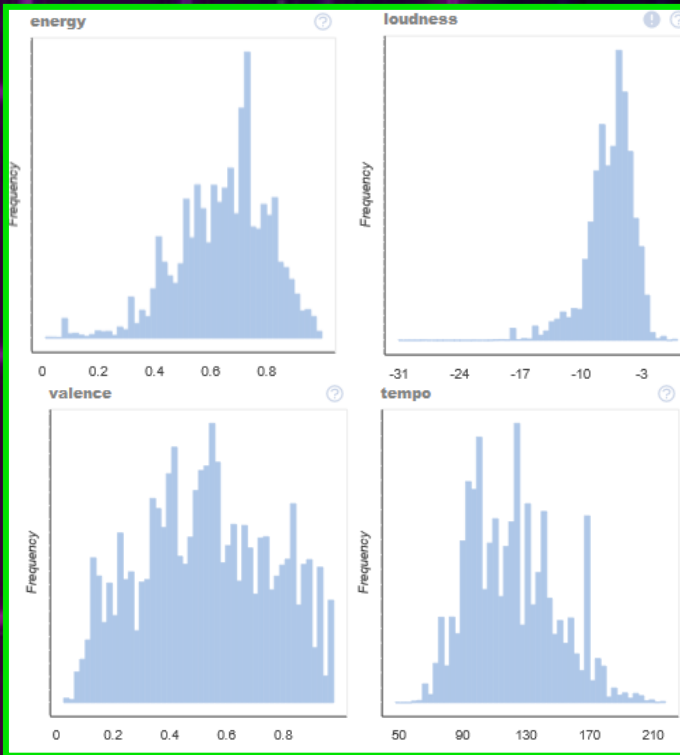
df.dropna(inplace=True)
df.isna().sum()

spotify_id      0
name            0
artists         0
daily_rank      0
daily_movement  0
weekly_movement 0
country         0
snapshot_date   0
popularity      0
is_explicit     0
duration_ms     0
album_name      0
album_release_date 0
danceability     0
energy          0
key             0
loudness        0
mode            0
speechiness     0
acousticness    0
instrumentalness 0
liveness        0
valence         0
tempo           0
time_signature   0
dtype: int64
```

DATA DISTRIBUTION



UNIVARIATE ANALYSIS



Based on the graph:

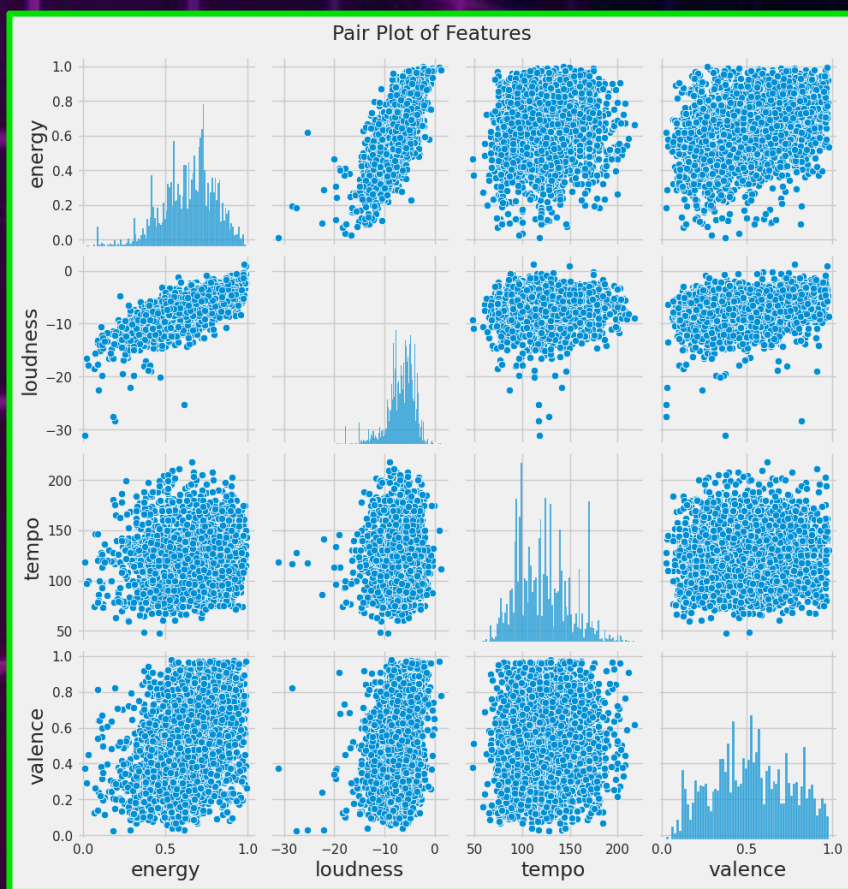
- The songs in this dataset mostly have quite high energy and loudness.
- Songs in this dataset have an average valence value of 0.529590.
- Songs in this dataset have an average tempo value of 121.686182.
- Songs in this dataset have the most major modes.

MULTIVARIATE ANALYSIS

+

+

+



Based on the graph, it can be seen that there is a relationship between some of the variables.

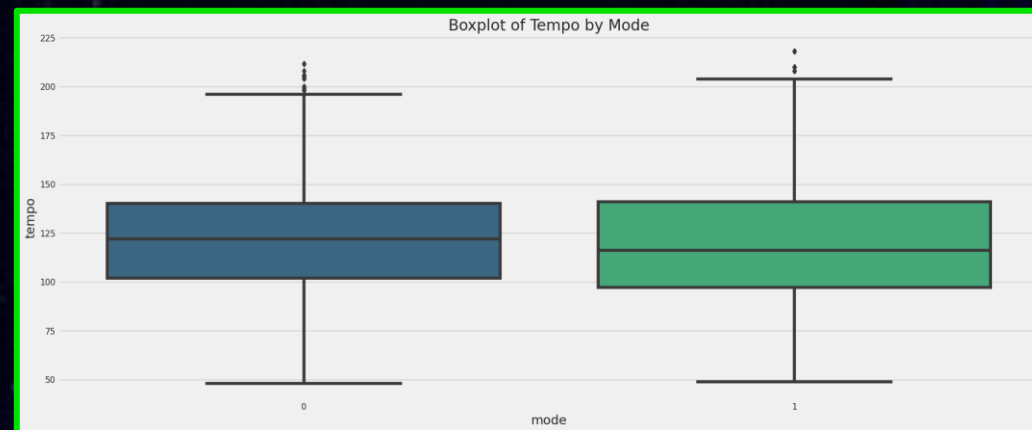
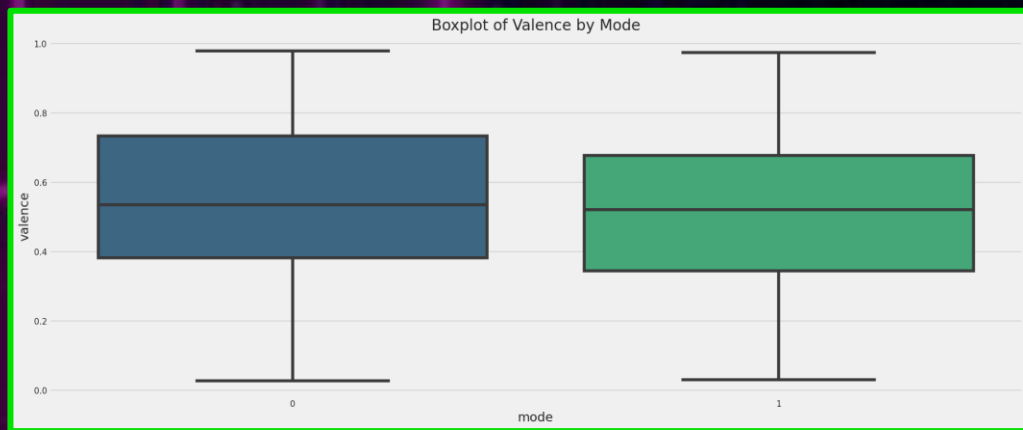
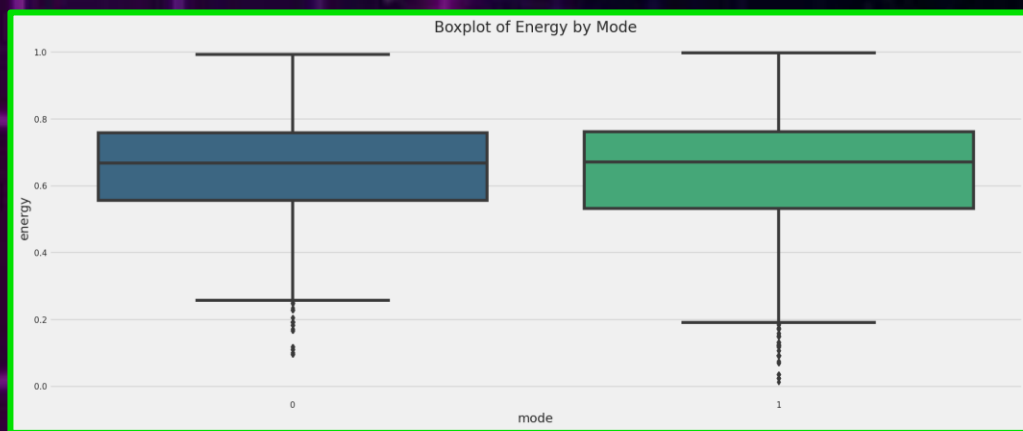
For example, **energy** and **loudness** are **positively correlated**, which means that songs with high energy tend to have high loudness as well. **Tempo** and **valence** are also **positively correlated**, which means that songs with a fast tempo tend to have high valence as well.

MULTIVARIATE ANALYSIS

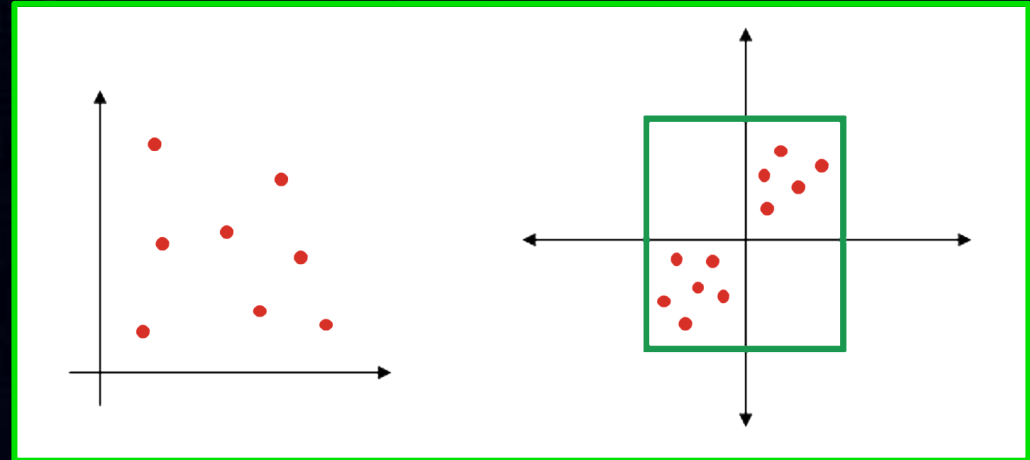
+

+

+



DATA SCALING

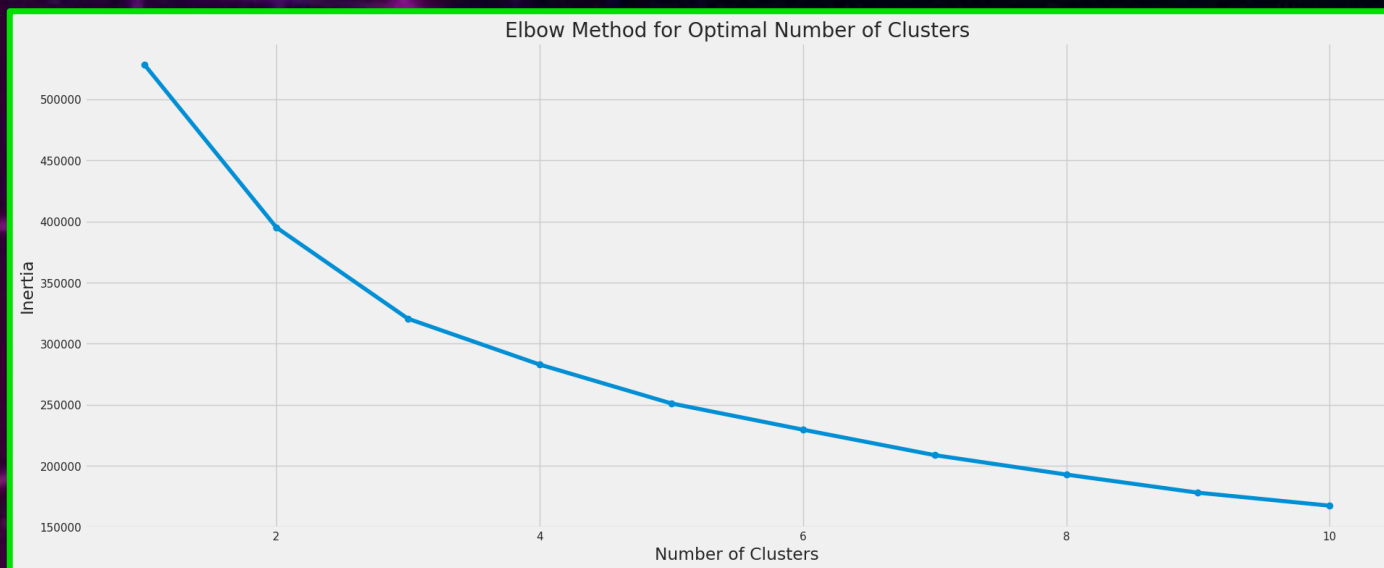


StandardScaler to normalize the data so that the data used does not have large deviations.



MODELING

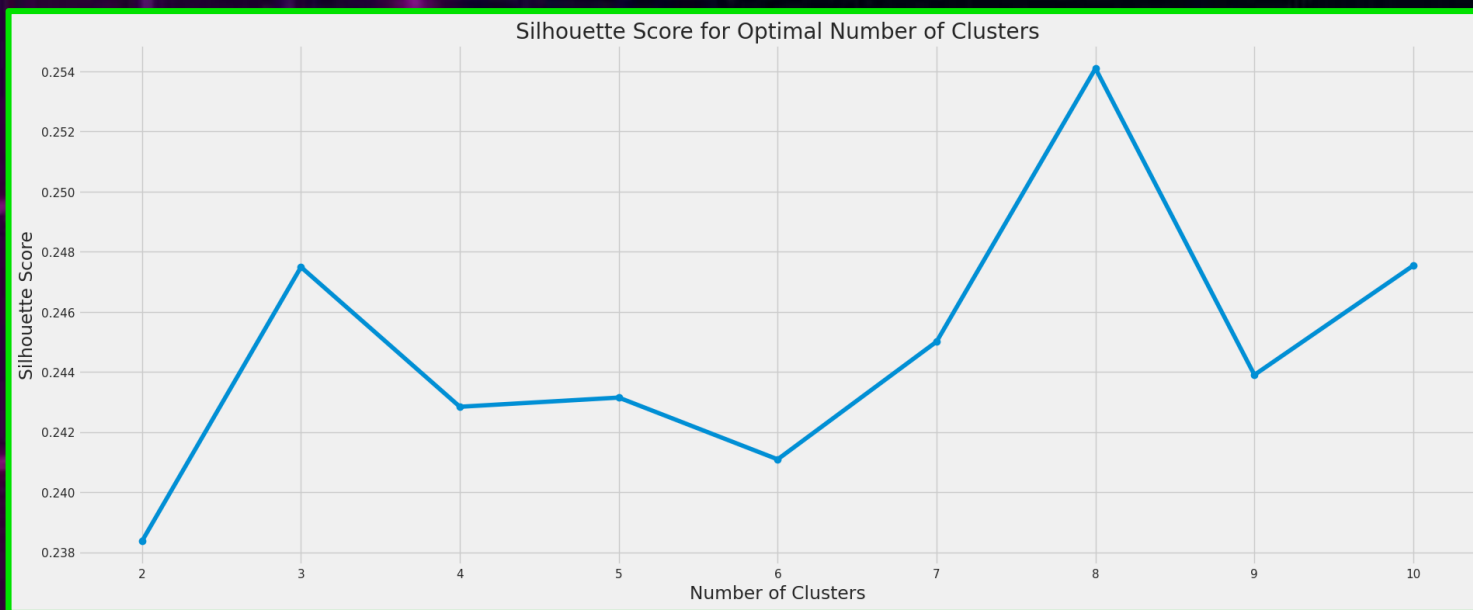
INERTIA



Before doing clustering, it would be better to determine the best and right number of clusters first.

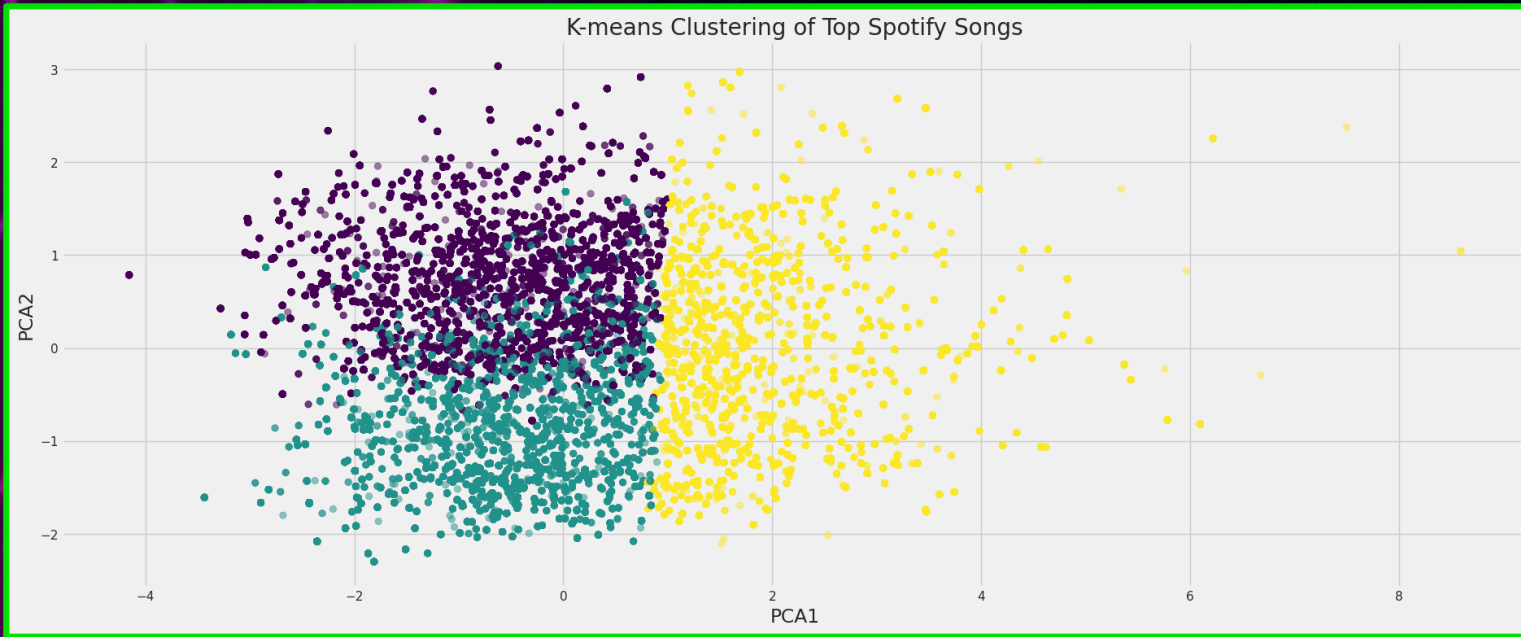
According to the graphic (Elbow method), the angle change starts to occur at point 3, then the correct K value for K-Means Clustering is $K = 3$.

SILHOUETTE SCORE



According to the graph (silhouette), the highest points occur at points 3 and 8, so I choose the right K value for K-Means Clustering is $K = 3$.

K-MEANS CLUSTERING



From the results of this clustering and visualized with a scatterplot as shown below.

This diagram shows the distribution of data which is divided into clusters according to the K-Means Clustering algorithm.



ANALYSIS

MEDIAN PER CLUSTER

	energy	loudness	tempo	valence	mode
cluster					
0	0.706	-5.678	123.010	0.5690	0.0
1	0.720	-5.501	120.001	0.5745	1.0
2	0.441	-9.319	111.982	0.3240	1.0





RECOMMENDATION

Provide a playlist of rock sub-genre songs

ROCK GENRE CLUSTERS

+

+

+

Cluster 1

With high energy, loudness, tempo, and valence, as well as minor modes, this music most likely falls into the energetic and aggressive rock genre. One subgenre of rock that can include these features is **Hard Rock**.

Hard rock : Loud, energetic, and often fast-paced sound. Minor modes can also give the music a darker and more intense feel. Some famous hard rock bands such as AC/DC, Guns N' Roses, or Led Zeppelin often have the characteristics you mentioned.

ROCK GENRE CLUSTERS

+

+

+

Cluster 2

With very high energy, loudness, tempo, and valence, as well as major modes, this music may fit into some rock subgenres that have these elements. A subgenre that might fit these characteristics is **Glam Rock**.

Glam Rock: Tends to have glamorous, eccentric and stylish elements. Some glam rock bands like David Bowie or T. Rex can display high energy and an upbeat atmosphere.

ROCK GENRE CLUSTERS



Cluster 3

With its energy, loudness, tempo, very low valence, and major modes, these characteristics are more in line with some types of rock music that have a melancholic or introspective feel. One subgenre of rock that can include these features is **Slowcore**.

Slowcore: Slow tempo, low dynamics, and often introspective atmosphere. Some slowcore bands such as Low or Codeine tend to display quiet and melancholic musical features.

“

WHERE WORDS FAIL,
MUSIC SPEAKS.

HANS CHRISTIAN ANDERSEN

”

THANK YOU!

HAVE ANY QUESTION?



CONTACT

- + **dalitrios@gmail.com**
- + **+62 821 1686 5885**
- + **www.linkedin.com/in/dalitrios**