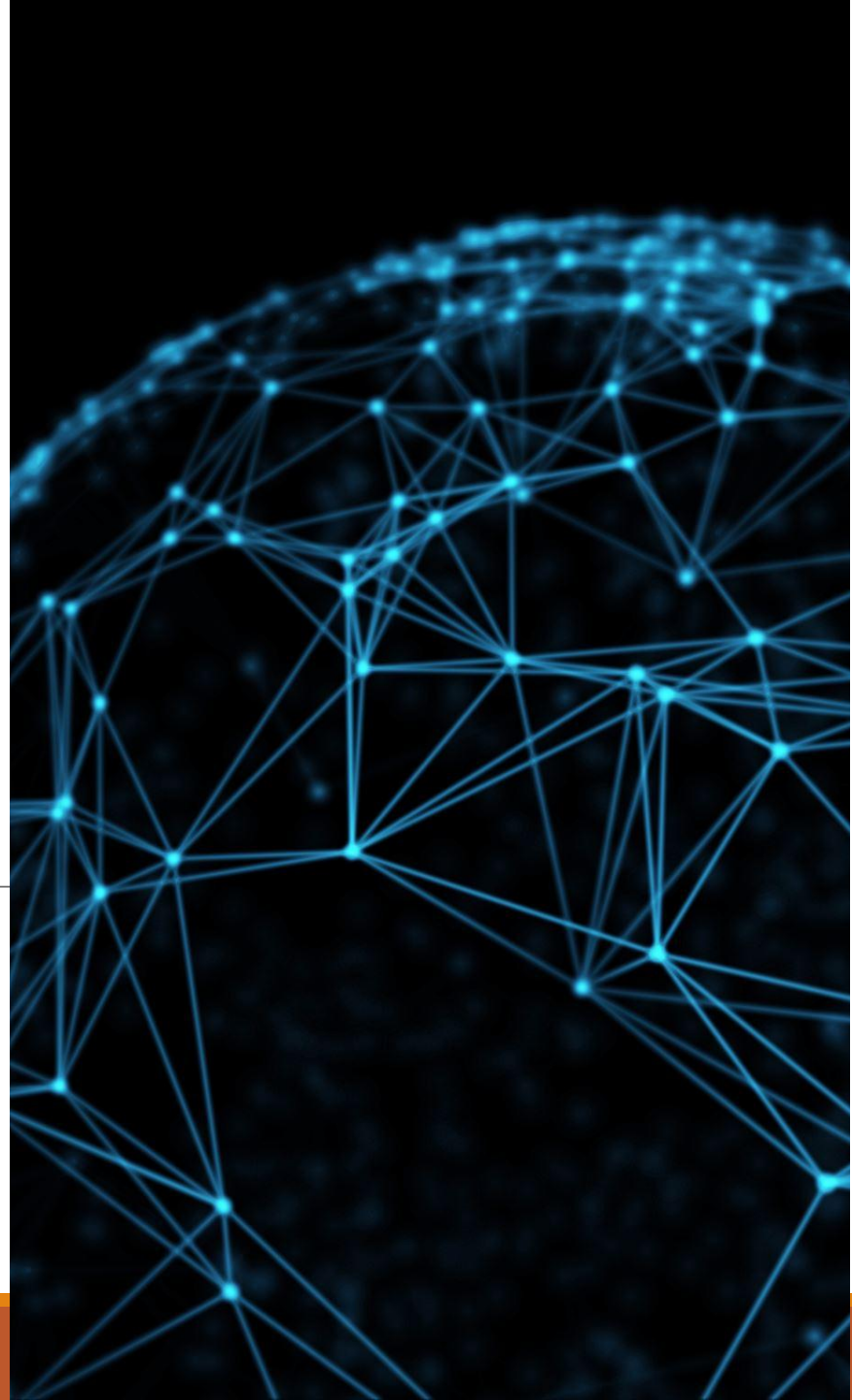
SPOTIFY SONG PLAYLIST CREATION



Team

Participant Name	Email Address	Role
Deepika Chintala	deepikachintala@my.unt.edu	Powerpoint Presentation and Executive Summary
Veepura Chari Dharmavarapu	VeepuraChariDharmavarapu@my.unt.edu	Coding
Kavya Parvathapuram	kavyaparvathapuram@my.unt.edu	Coding

Presentation Overview

- Introduction
- How ML Is Used in Music Industry
- Problem To Solve
- Dataset Explanation
- Data Preparation
- Exploratory Data Analysis
- Algorithms Used
- Algorithm Comparison
- Predictions
- How Does This Help The Company?

Introduction

Spotify is one of the most popular music streaming services offering over 50 million songs and 700,000 podcasts. About 40,000 new songs are added to Spotify every day!

How ML is used in Music Industry?

- Personalized Recommendations
- Predictive Analytics
- Genre Identification
- User Behavior Analysis

Problem To Solve

So how can we create personalized playlists? Why are they so popular these days?

The aim of our project is to create playlists based on various factors catered to the song.

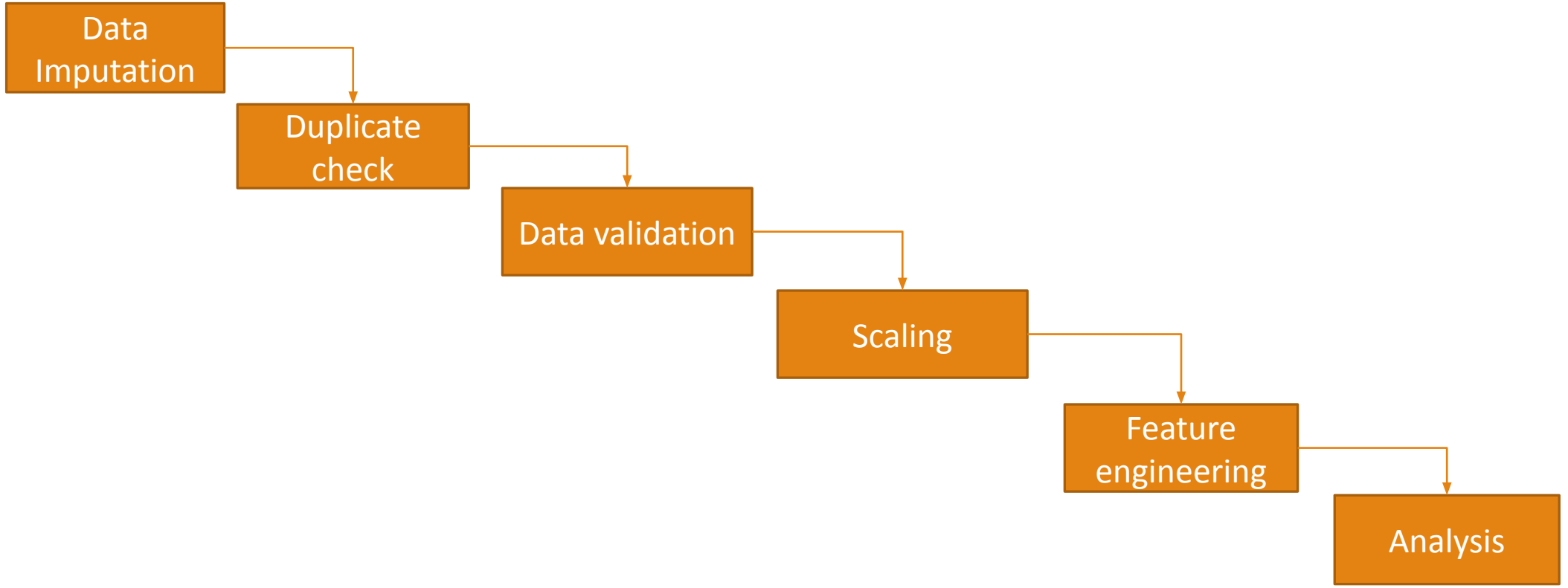
Dataset Explanation

track_name	Name of the song
artist_name	Artist for the track
genre	Which genre the song belongs to
duration_ms	Song duration in milliseconds
acousticness	Confidence measure of 0.0-1.0 whether a song is acoustic.
danceability	Measure of 0.0-1.0 on how suitable the song is for dancing.
energy	Represents measure of intensity and activity.
instrumentalness	Predicts whether a track contains vocals or not.
key	Tells which key the track is in.
liveness	Detects if audience is present in the song, i.e. whether the song was performed live.

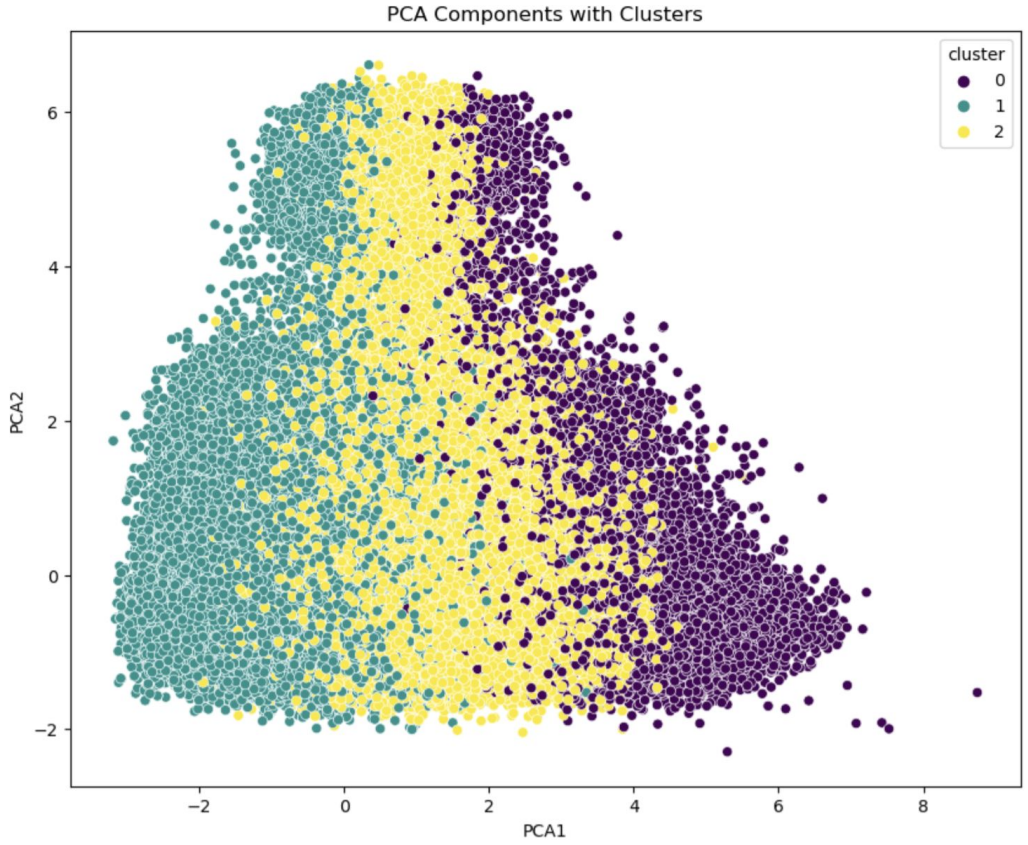
Dataset Explanation

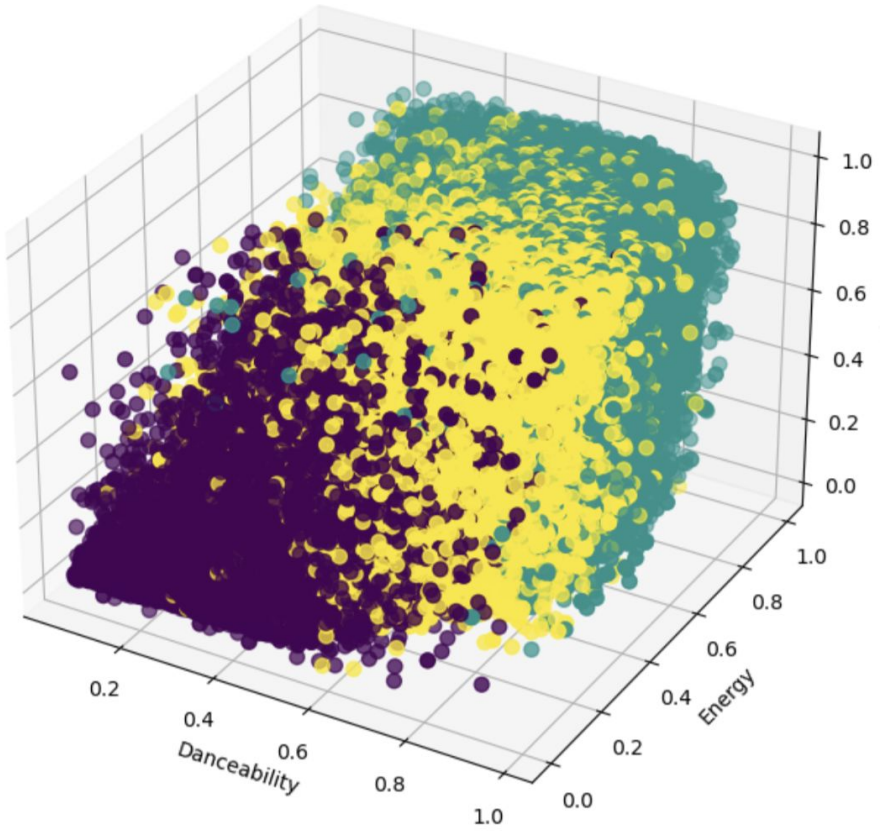
mode	Indicates if a track is major or minor.
speechiness	Detects the presence of spoken words in a song.
tempo	Tempo in beats per minute (BPM)
time_signature	Overall time signature of a track.
valence	Describes the musical positiveness.
loudness	Loudness of a track in decibels (dB)
popularity	If a song is popular or not.
track_id	Unique id which identifies a track.

Data Preparation



Exploratory Data Analysis





Algorithms Used

- K-Means
- DBSCAN
- K Nearest Neighbors

Algorithm Comparison

Model	Clustering Approach	Cluster Shape and Size	Handling Outliers
K Means	Partition data into k clusters	Assumes spherical clusters of equal size	Sensitive to Outliers
DBSCAN	Groups data based on density	Can identify clusters of arbitrary shapes	Robust to Noise. Identifies noise points as not belonging to any cluster
K Nearest Neighbors	Considers majority class among neighbors	Does not create clusters	Sensitive to Noise.

Predictions

Silhouette Scores - DBSCAN - 0.66

KMeans - 0.587

```
I'm a Little Teapot (3t61HuZKUHvGkMNE2OgeLG)
Alouette (4BOgMQcgtLIFOlYAnlyr4n)
Mary Had a Little Lamb (4bUsXMWaeostMYmX0BUBoE)
Twinkle Twinkle Little Star (5yd7jAXUv6WBIcxIkbx7w9)
Playtime : Here We Go Round the Mulberry Bush / Pat a Cake / One, Two, Buckle My Shoe / Itsy Bitsy Spider / This Little Piggy (1VVDj8rUL7R6J(
Little Lucas (57x2ZL9CLBeP17gXeXcZng)
The Alphabet Song (58TyIObSY8oU3WjuglwwVr)
Incy Wincy Spider (5QJPzvMzMCmYVE9bBHYWxR)
Pirouette, cacahuète (7t8fnv1MB20YltARxIREVq)
The Wheels on the Bus (10aXwOUqXyX6RNqKqSFWpf)
Rock a Bye Baby (5AWi9QmC4B8h6Cgzd45RJK)
By the Chinny Chin Chin (4WQAeGcKVUJHk3C1UgRN1A)
One Elephant (4X4twP069xb8Bhua6BKJaD)
Goosey Goosey Gander (5SUkK5jlXw2v0ImMaACvKP)
London Bridge (0Aeag2nF2wHTGXjYg7WChQ)
Just Plain Silly : Jack and Jill / Three Blind Mice / Rub-A-Dub / Humpty Dumpty / What Are Little Boys Made of / I Saw a Ship a Sailing / Mos
Chloe (06opIyj8JWWeQv44xvJ47R)
My Father Gave Me a Hubby (36VJEBMjNvQqVktjRNZrqH)
You Are My Sunshine (5KRyT9R1fwqRc3Vm5wRYmS)
The Baker Has a Lot of Coins (7Fg8IEdtyMtk7RZIURHyEQ)
Polly Wooddle Doodle (2xqqpuSsXxIc8KpH3evNMU)
The Ants Go Marching In (4EVrOAM6MADmije4yyXM4Z)
I Had A Little Rooster (7rwVK6xpznLkKLXzlKXS83)
Round & Round The Garden (52jvLmqOyzBYjmzs89duGl)
```

How Does This Help The Company?

- Users spend more time with the app.
- Allows users to share playlists to others.
- Increases new user registrations.

References

- T. S. Bohra, V. Kumar and S. Ganesan, "Segmenting music library for generation of playlist using machine learning," 2015 IEEE International Conference on Electro/Information Technology (EIT), Dekalb, IL, USA, 2015, pp. 421-425, doi: 10.1109/EIT.2015.7293429.
- A. Ferraro et al., "Melon Playlist Dataset: A Public Dataset for Audio-Based Playlist Generation and Music Tagging," ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Toronto, ON, Canada, 2021, pp. 536-540, doi: 10.1109/ICASSP39728.2021.9413552.
- https://www.kaggle.com/code/dimasikson/similar-songs-clustering-k-means-cosine-simil/input
- https://towardsdatascience.com/generating-spotify-playlists-with-unsupervised-learning-abac60182022

Thank You

Any Questions?