

Part 1 : Dataset Description

Problem Statement:

The dataset used for this project is sourced from Kaggle and contains various attributes related to songs and their characteristics on the Spotify platform. The primary objective for this dataset is to build a recommendation engine that can analyze and understand the factors contributing to the popularity and user engagement of songs on Spotify. This recommendation engine aims to provide personalized song recommendations to users based on their preferences, ultimately enhancing the user experience on the platform.

Source of the Dataset:

The dataset was obtained from Kaggle, a well-known platform for sharing and discovering datasets. Kaggle sourced the data from Spotify, a popular music streaming service. Using data from Spotify ensures that the dataset contains real-world information about songs available on the platform, making it valuable for building a music recommendation system.

Brief Description of the Dataset:

The dataset obtained from Kaggle consists of information about songs available on Spotify. Each row in the dataset represents a unique song. The dataset comprises 14 columns, each representing different attributes or variables related to these songs.

Description of the Attributes/Variables/Columns of the Dataset:

1. **artists:** This column contains the names of the artists associated with the songs. It is of data type object.
2. **acousticness:** This column represents the acousticness of the songs, measured as a float value. Acousticness refers to the extent to which a song is acoustic in nature, with values closer to 1 indicating a more acoustic sound.
3. **danceability:** This column represents the danceability of the songs, measured as a float value. Danceability indicates how suitable a song is for dancing, with higher values indicating songs that are easier to dance to.
4. **energy:** This column represents the energy level of the songs, measured as a float value. Energy indicates the intensity and activity of a song, with higher values indicating more energetic songs.
5. **id:** This column contains unique identifiers for the songs. It is of data type object.
6. **instrumentalness:** This column represents the instrumentalness of the songs, measured as a float value. Instrumentalness indicates the extent to which a song contains no vocals, with values closer to 1 indicating purely instrumental songs.

7. **key:** This column represents the key of the songs, measured as an integer. The key is a musical attribute that determines the overall pitch and tonality of a song.
8. **liveness:** This column represents the liveness of the songs, measured as a float value. Liveness indicates whether a song was recorded live, with higher values indicating a live recording.
9. **loudness:** This column represents the loudness of the songs, measured as a float value. Loudness refers to the overall volume of a song, with negative values indicating quieter songs.
10. **mode:** This column represents the mode of the songs, measured as a float value. Mode indicates whether a song is in a major or minor key.
11. **name:** This column contains the names of the songs. It is of data type object.
12. **speechiness:** This column represents the speechiness of the songs, measured as a float value. Speechiness indicates the presence of spoken words in a song, with higher values indicating more speech-like content.
13. **tempo:** This column represents the tempo (beats per minute) of the songs, measured as a float value.
14. **valence:** This column represents the valence of the songs, measured as a float value. Valence indicates the positivity or happiness of a song, with higher values indicating more positive or happy songs.