

# Machine Learning Project: Clustering Analysis

## 1. Initial Data Inspection and Data Cleaning:

- Begin by examining your dataset. Look for:
  - Duplicates: Check for identical rows.
  - Missing values: Identify any null or NaN entries.
  - Irrelevant entries: Remove any data points that don't contribute to your analysis.
  - Outliers: Detect extreme values that might skew results.
  - Erroneous entries: Correct any data inconsistencies.
- Clean the data by addressing the issues you've identified.

## 2. Data Refinement:

- Based on your initial inspection, refine the data further. This may involve:
  - Dropping unnecessary columns.
  - Imputing missing values.
  - Standardizing or normalizing features.

## 3. Exploratory Data Analysis (EDA) and Feature Engineering:

- Visualize the data:
  - Create suitable visualizations (e.g., bar plots, scatter plots) to identify patterns.
  - Specifically, focus on identifying two albums with the most popular songs.
- Feature engineering:
  - Create new features if needed (e.g., aggregating song features).
  - Explore relationships between features.

## 4. Song Popularity Analysis:

- Investigate the relationship between song popularity and various factors:
  - Features like danceability, energy, valence, etc.
  - How this correlation has evolved over time (if your data includes a temporal aspect).
- Use visualizations to illustrate your findings.

## 5. Dimensionality Reduction Techniques:

- Explain the significance of dimensionality reduction:
  - High-dimensional data can be challenging to work with.
  - Techniques like Principal Component Analysis (PCA) or t-SNE reduce dimensions while preserving important information.
  - We used PCA to reduce the dimensionality

## 6. Cluster Analysis:

- Determine the right number of clusters:
  - Using method like the elbow method we determined the appropriate number of clusters
- Choose appropriate clustering algorithms:
  - K-means algorithm was used for clustering
  - Cluster count is taken as 20
- Define each cluster based on features:
  - Interpret the characteristics of each cluster (e.g., high-energy songs, acoustic songs).