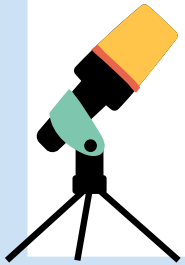
Spotify Song Recommendations

Daniel Wang, Yan Wang, Kate Yee, Xinyang Zhou

Table of contents



Business Objective

How can Spotify leverage the commonalities among songs to optimize song recommendations to users'?

Data Mining Methodology

Dimensionality Reduction

- tSNE for visualization
- PCA

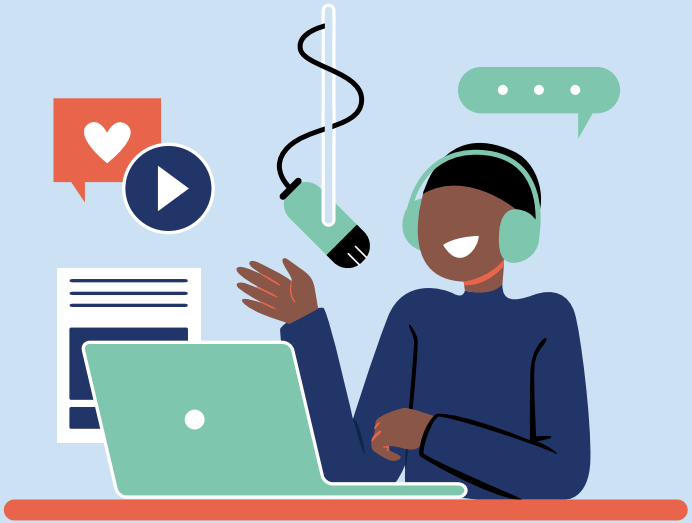
Clustering

- K-Prototype
- K-means
- DBScan

Recommendation Systems

Recommendation Systems
- Filter by song similarity
- Sort by popularity

Data Overview and EDA



Dataset Overview

- Most Streamed Spotify Songs in 2023
  - 954 rows, 24 columns

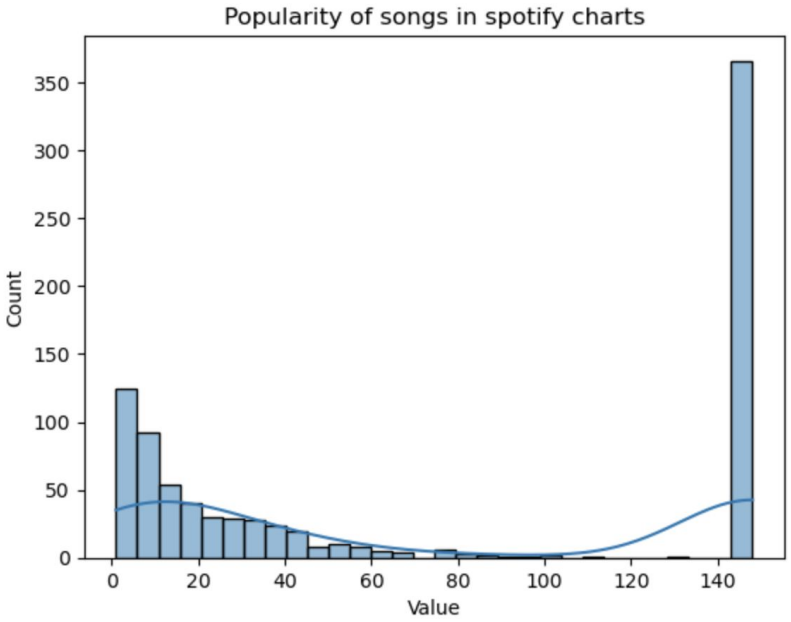
- Columns Kept
  - Categorical: track_name, artist(s)_name, key, mode
  - Numerical: in_spotify_playlists, in_spotify_charts, bpm, danceability_%, valence_%, energy_%, acousticness_%, instrumentalness_%, liveness_%, speechiness_%, released_year

Feature Explanations

- **track_name**: *Name of the song*
- **artist(s)_name**: *Name of the artist(s) of the song*
- **in_spotify_playlists**: *Number of Spotify playlists the song is included in*
- **in_spotify_charts**: *Presence and rank of the song on Spotify charts*
- **bpm**: *Beats per minute, a measure of song tempo*
- **key**: *Key of the song*
- **mode**: *Mode of the song (major or minor)*
- **danceability_%**: *Percentage indicating how suitable the song is for dancing*
- **valence_%**: *Positivity of the song's musical content*
- **energy_%**: *Perceived energy level of the song*
- **acousticness_%**: *Amount of acoustic sound in the song*
- **instrumentalness_%**: *Amount of instrumental content in the song*
- **liveness_%**: *Presence of live performance elements*
- **speechiness_%**: *Amount of spoken words in the song*

Data Preprocessing

- Removed nulls
- Handled data types
- Removed redundant columns
  - Chose in_spotify_charts as the preferred popularity metric
  - 1 is more popular
  - 148 is not appearing in charts

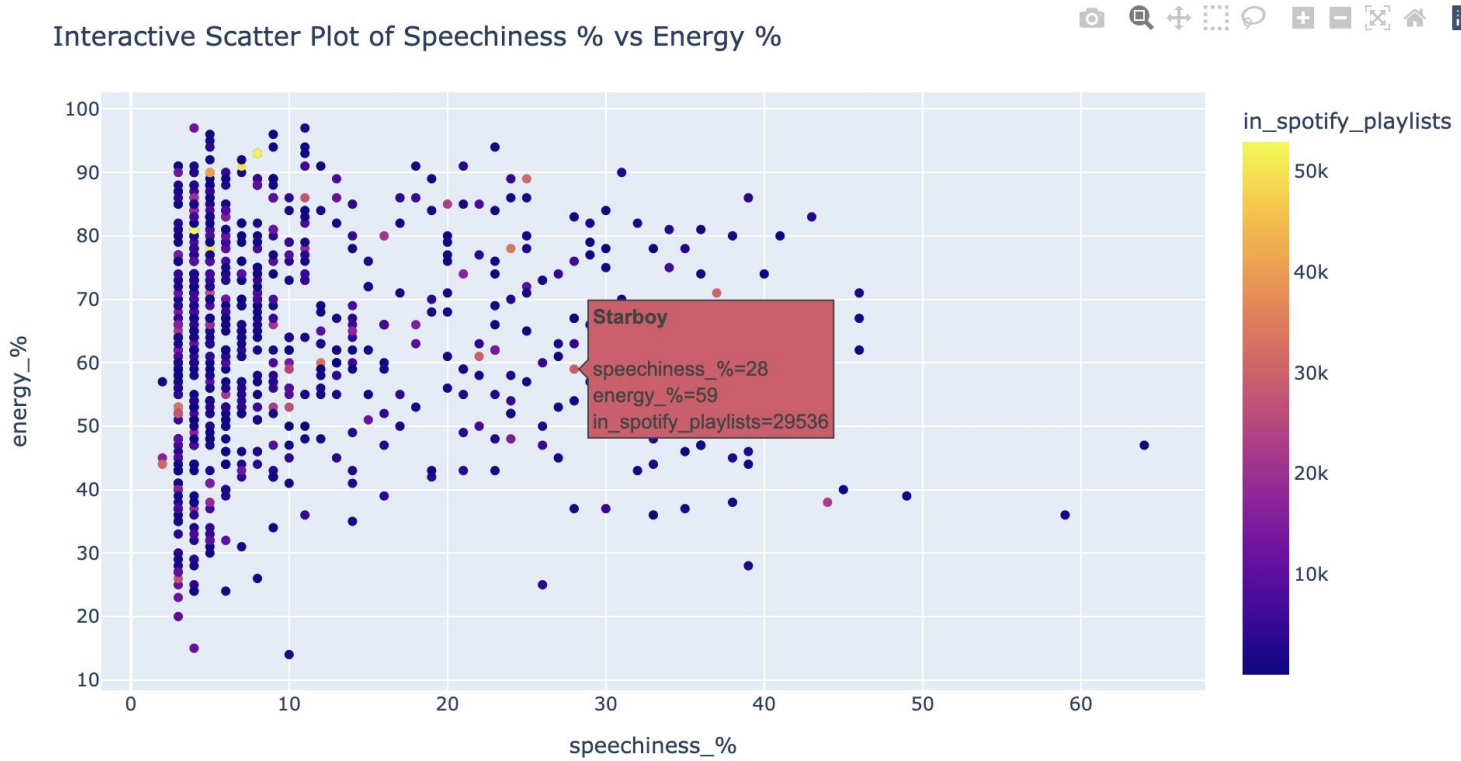


Top Artists on Spotify in 2023

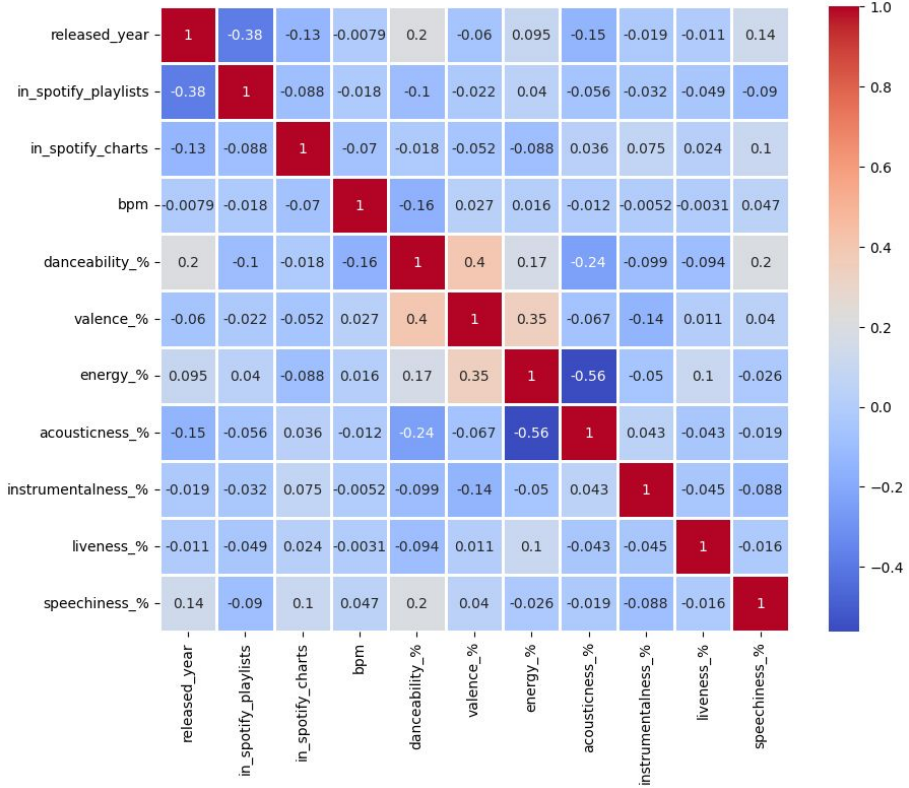
- The rank is based on the number of tracks, from highest to lowest
- Do you see your favorite on the list🎤?

Taylor Swift	29
The Weeknd	17
SZA	17
Bad Bunny	16
Harry Styles	13
Kendrick Lamar	12
Ed Sheeran	9
Morgan Wallen	9
BTS	8
Feid	8
Olivia Rodrigo	7
Drake, 21 Savage	7

Interactive Scatterplot

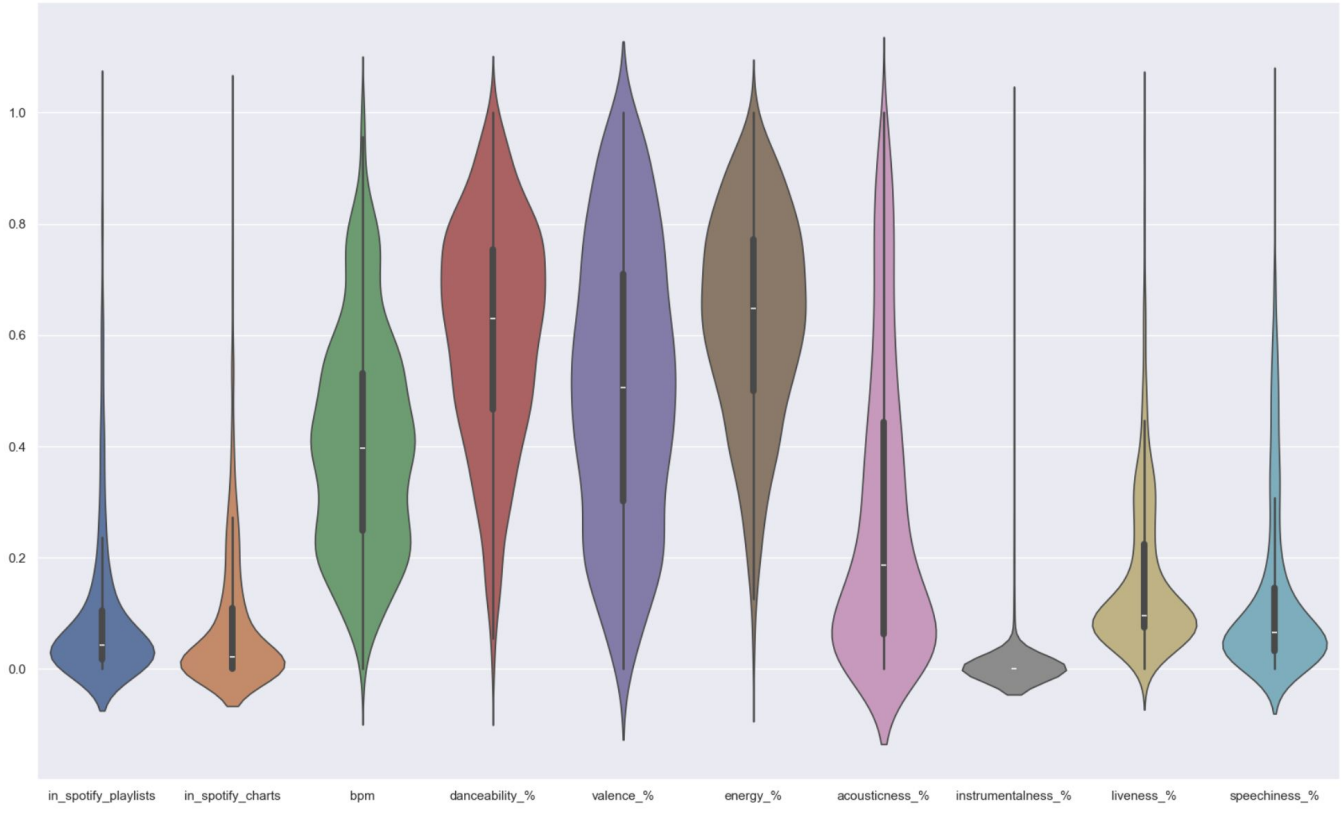


Correlation Matrix

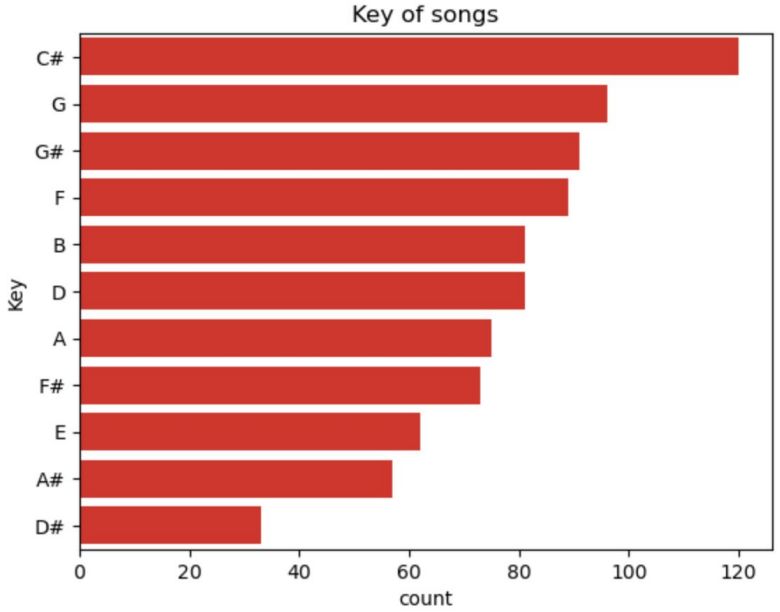


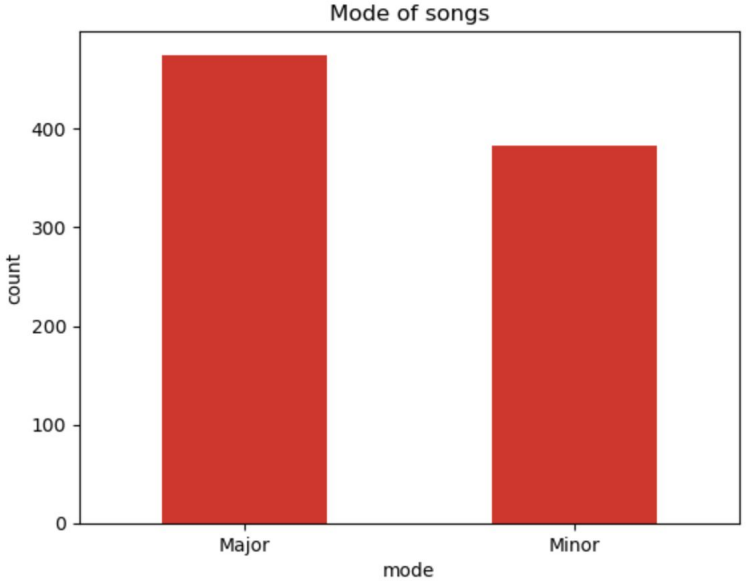
- Some correlation between numerical columns, but not an overwhelming amount

Violin Distribution Plots



Categorical Variables Plots



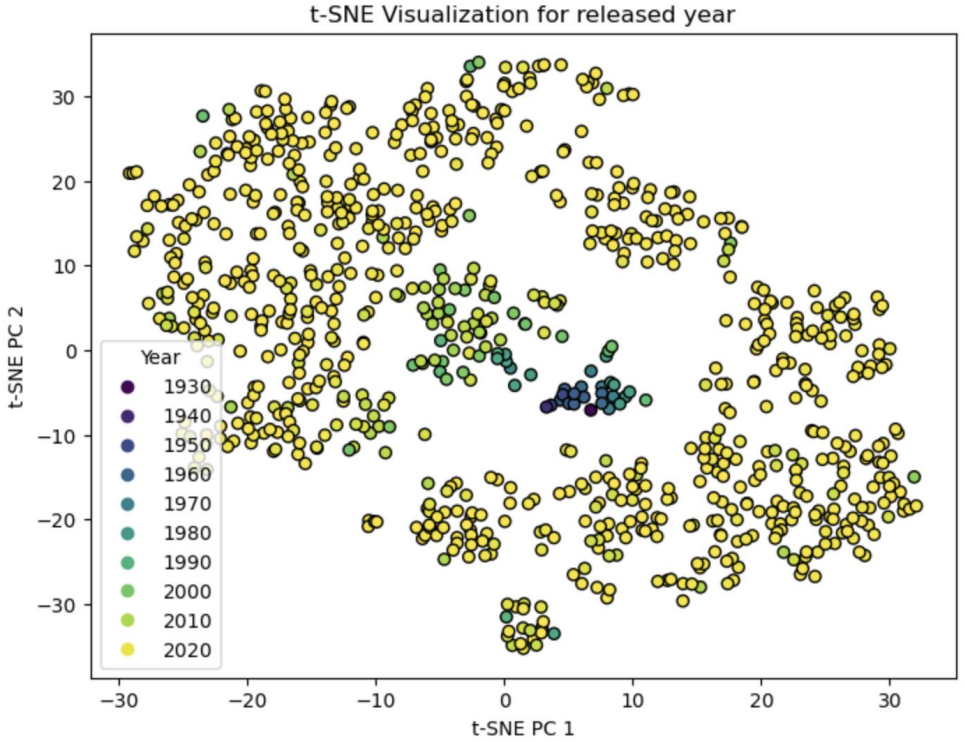


- C# is the most common key
- There are more songs in major than minor

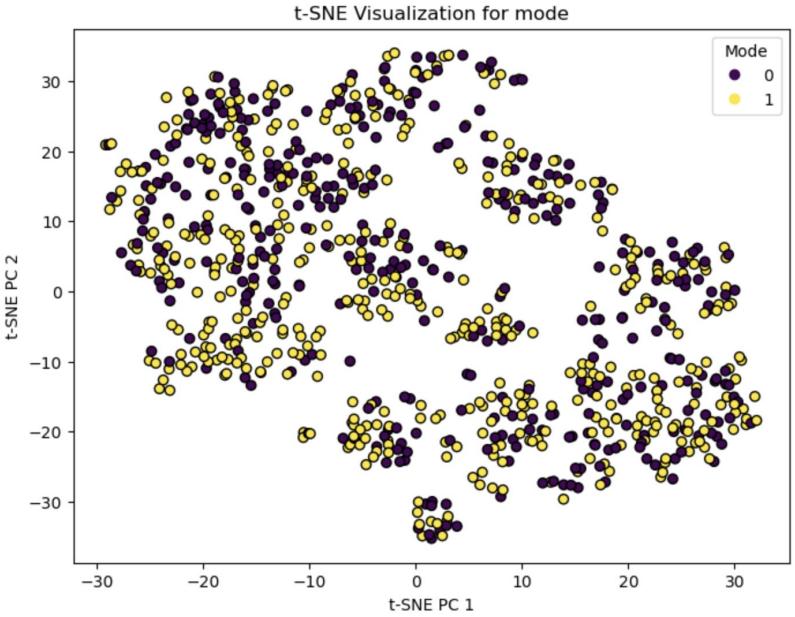
Dimensionality Reduction

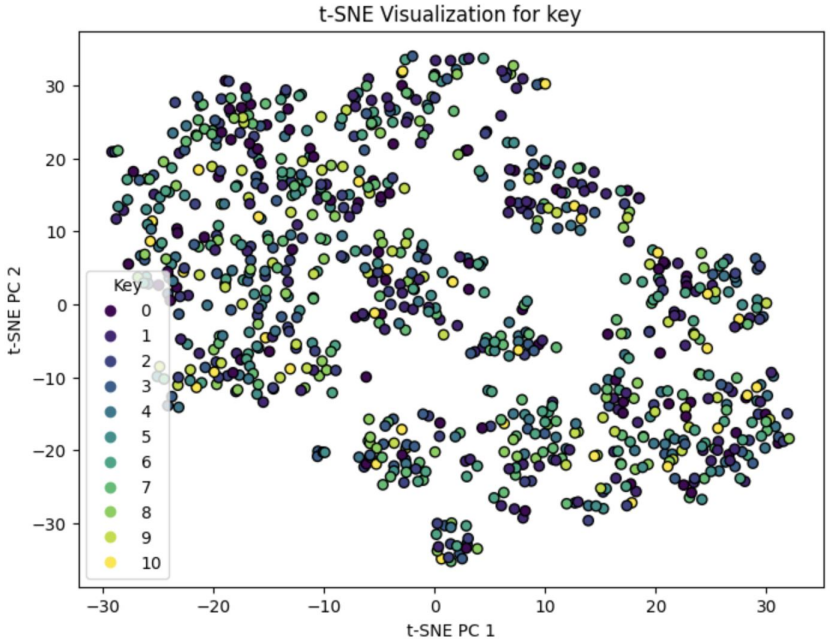
tSNE for Visualization

- 2 Components
- KL Divergence: 1.02486
- Number of iterations: 999
- Effective for released year



tSNE for Visualization





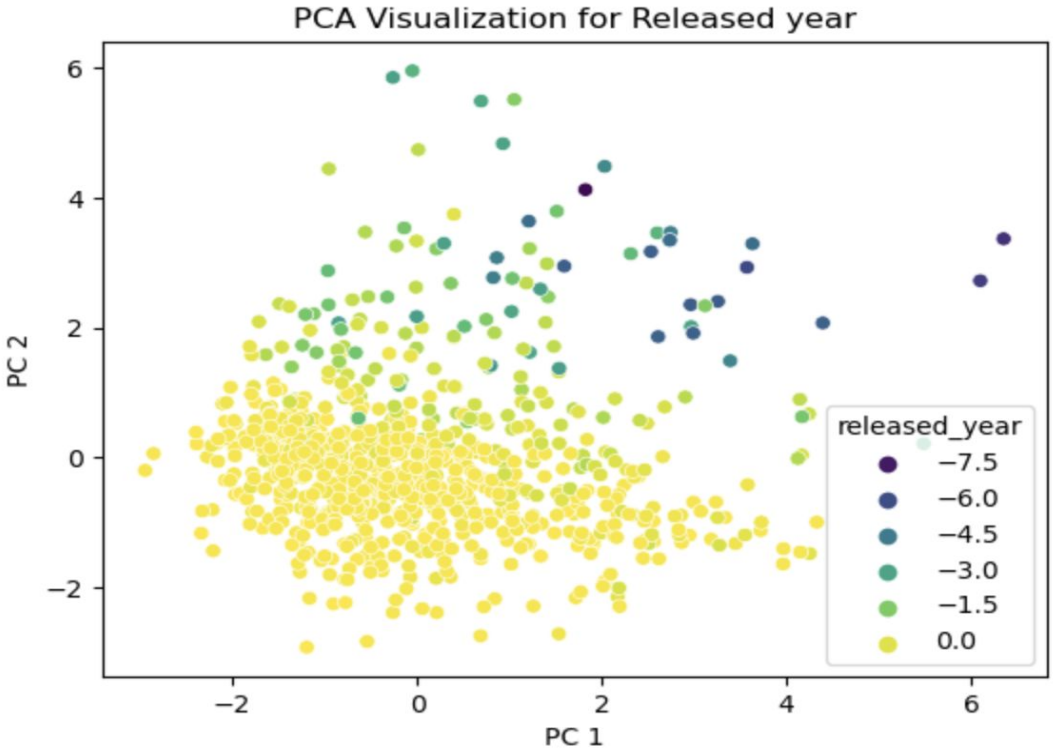
- Not as effective for other categorical variables

PCA

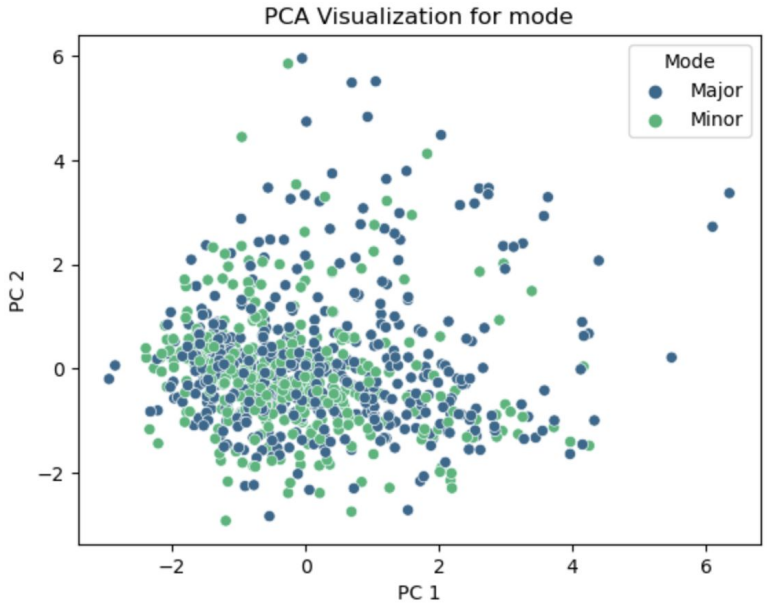
	Component 1	Component 2
released_year	-0.260289	-0.523572
in_spotify_playlists	0.107906	0.599054
in_spotify_charts	0.108565	-0.137120
bpm	0.034223	0.059526
danceability_%	-0.463341	-0.190716
valence_%	-0.407372	0.143267
energy_%	-0.508831	0.315087
acousticness_%	0.463723	-0.218230
instrumentalness_%	0.172079	-0.039959
liveness_%	-0.038403	0.088803
speechiness_%	-0.148249	-0.362459

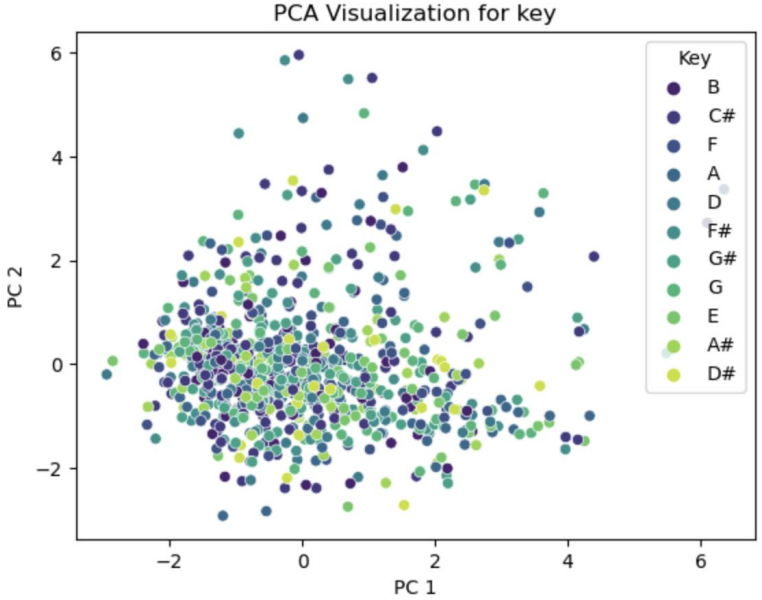
- Chose 2 Components
- Loadings Interpretation:
  - PC1: Mostly composed of 4 song quality attributes
    - Lower value: darker more acoustic music
    - Higher value: happier poppier music
  - PC2: Mostly composed of release year and playlist frequency
    - Lower value: More recent songs
    - Higher value: Classic hits

PCA - Released Year



PCA - Mode and Key





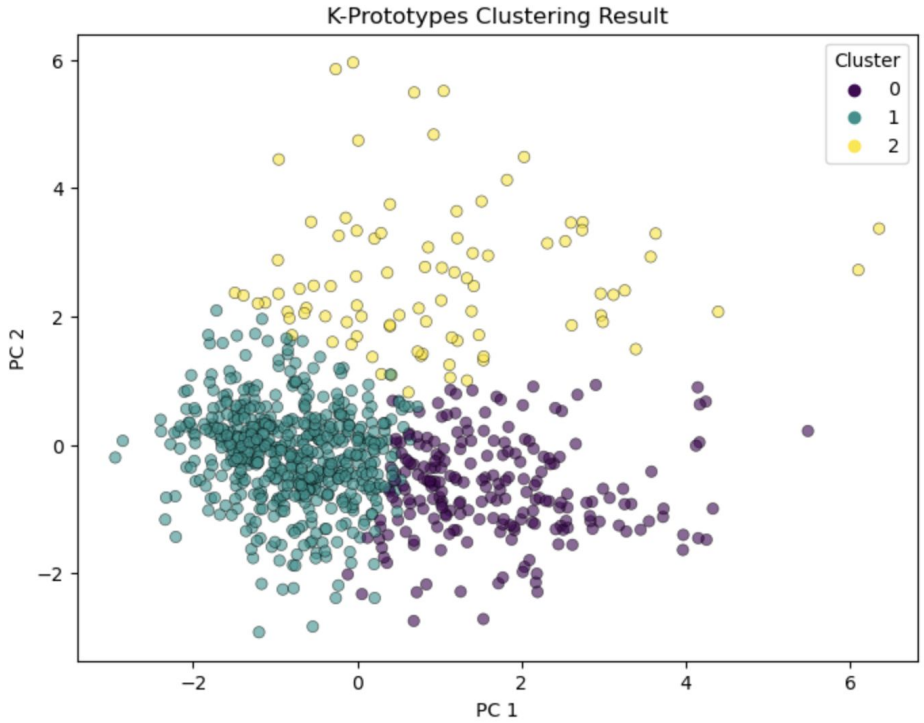
- Mode and key are spread across the PC mapping

Clustering

K-Prototype

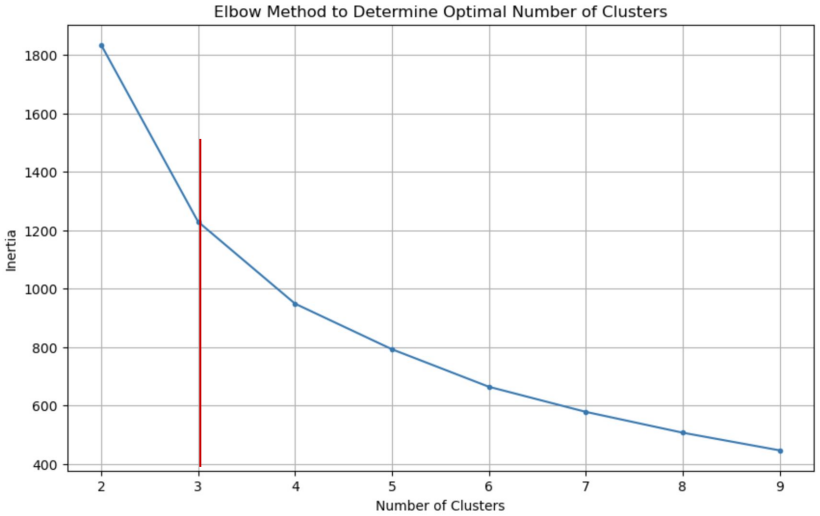
- Inputs: PC1, PC2, Mode, Key
- Best number of clusters: 3
- Best Silhouette Score: 0.4469

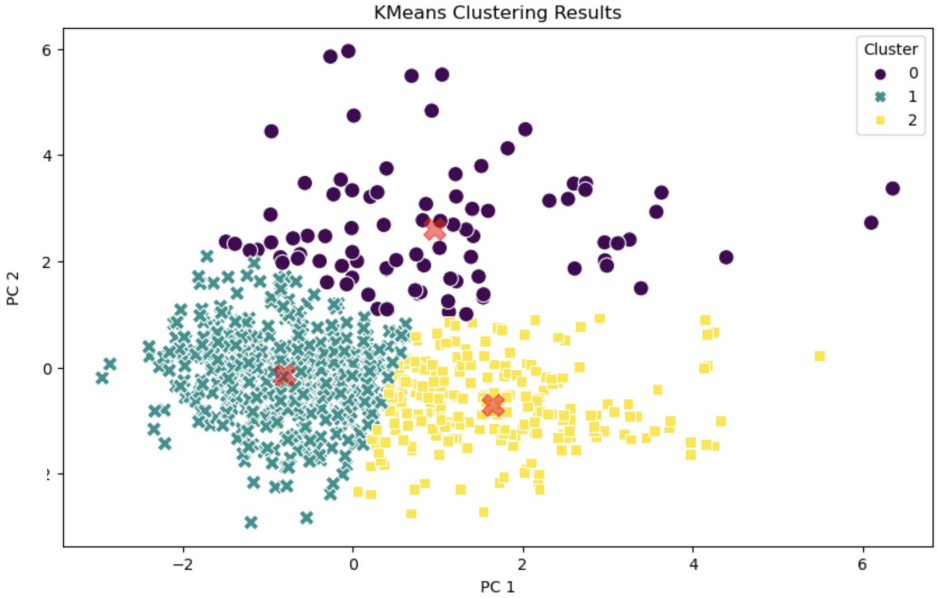
- Cluster interpretation
  - 0: Moody Modern Tracks
  - 1: Pop Bops
  - 2: Classic Hits



K-Means

- Inputs: PC1, PC2
- Best number of clusters: 3
- Maximum Silhouette Score: 0.4495
- Similar cluster interpretation

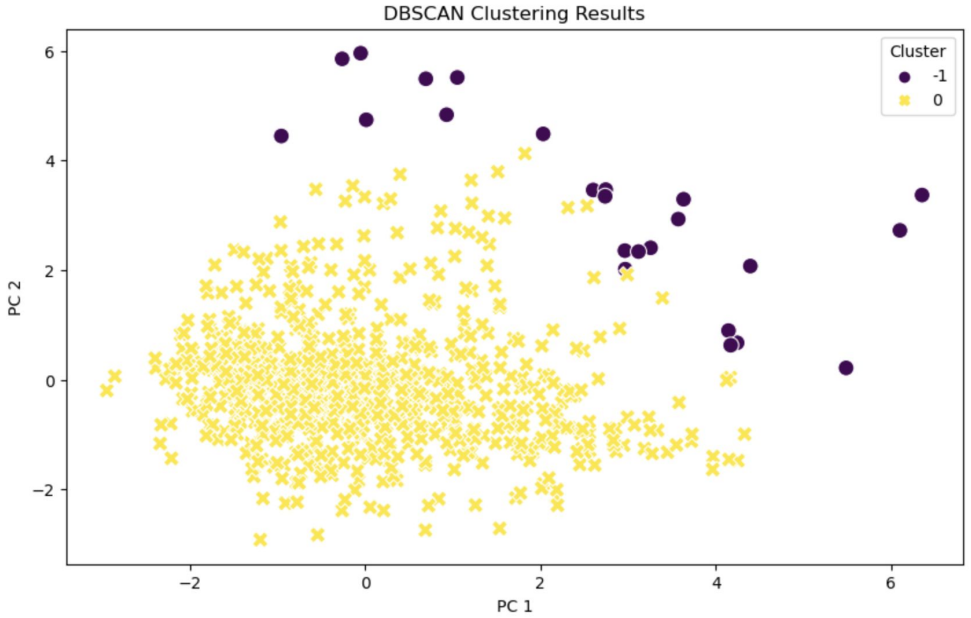




DBSCAN

- Inputs: PC1, PC2
- Best eps: 1
- Best min_samples: 10,
- Best silhouette score: 0.5719

- Cluster interpretation
  - 0: Modern Bops
  - -1: Classic Hits



Best Choice

- K-Prototypes
  - Pros: numerical and categorical data, interpretable
  - Cons: parameter choice complexity, computationally expensive than K-Means
- K-Means
  - Pros: simple, efficient, scalable, interpretable
  - Cons: sensitive to initial centroids, spherical cluster assumption
- DBscan
  - Pros: highest silhouette score, automatic cluster number, flexible shape
  - Cons: sensitive to eps and minPts, difficult to interpret

- Final Decision: **K-Prototypes**
  - Interpretable for Spotify insights
  - Contains numerical and categorical data

Recommendation System

Clustering Recommendation Strategy

Given a song a user likes follow these steps to return recommended songs:

1. Find what cluster that original song belongs to from our previous K-Prototype clusters
2. Filter the songs in that cluster (not including input song) by the same key and mode
3. Sort the remaining songs by popularity
4. Return as many of the top songs in the sorted list as requested
5. If the sorted list is shorter than the number of requested recommendations, return the remaining overall most popular songs from the cluster.

User Experience

1. Enter a track name
2. Enter numbers of recommendations

Enter “quit” to exit the system

```
==================================================
Welcome to Spotify Recommender System.
==================================================


******************** Main Page ********************
Enter "quit" at any time to exit the system.
Please enter a track name.
LALA
Please enter numbers of recommendations you want.
4

Start recommending ... Please wait...

Based on your favorite song style and popularity:
Recommendation 730:
Track Name: TUS Lï¿½ï¿½GR
Artist Name: Sech, Mora
Spotify Chart Rank: -1.122594647370691


Recommendation 699:
Track Name: Dos Mil 16
Artist Name: Bad Bunny
Spotify Chart Rank: -1.1071967254893416


Recommendation 388:
Track Name: KICK BACK
Artist Name: Kenshi Yonezu
Spotify Chart Rank: -1.091798803607992


Recommendation 563:
Track Name: Bar
Artist Name: Tini, L-Gante
Spotify Chart Rank: -1.091798803607992



******************** Main Page ********************
Enter "quit" at any time to exit the system.
Please enter a track name.
quit
Exiting Spotify Recommender System. Thank you for using!
```

Spotify Insights

- **Clustering System**
  - Can be used for song categorization
    - 0: Moody Modern Tracks
    - 1: Pop Bops
    - 2: Classic Hits

- **Recommendation System**
  - Filter by song similarity
  - Sort by popularity

Thanks!