DS203 E7
ASSIGNMENT

DS-PROJECT

DEV ARORA (23B1271)
FORAM TRIVEDI (23B1269)
MANYA SHAH (23B3969)
SHUBH TANGRI(23B3931)

Executive Overview

Problem Analysis, Model Formulation, and Resource Exploration

Feature Selection: Identifying Key Attributes, Eliminating Correlated Features

Implementing Clustering Techniques and Developing a Verification Framework

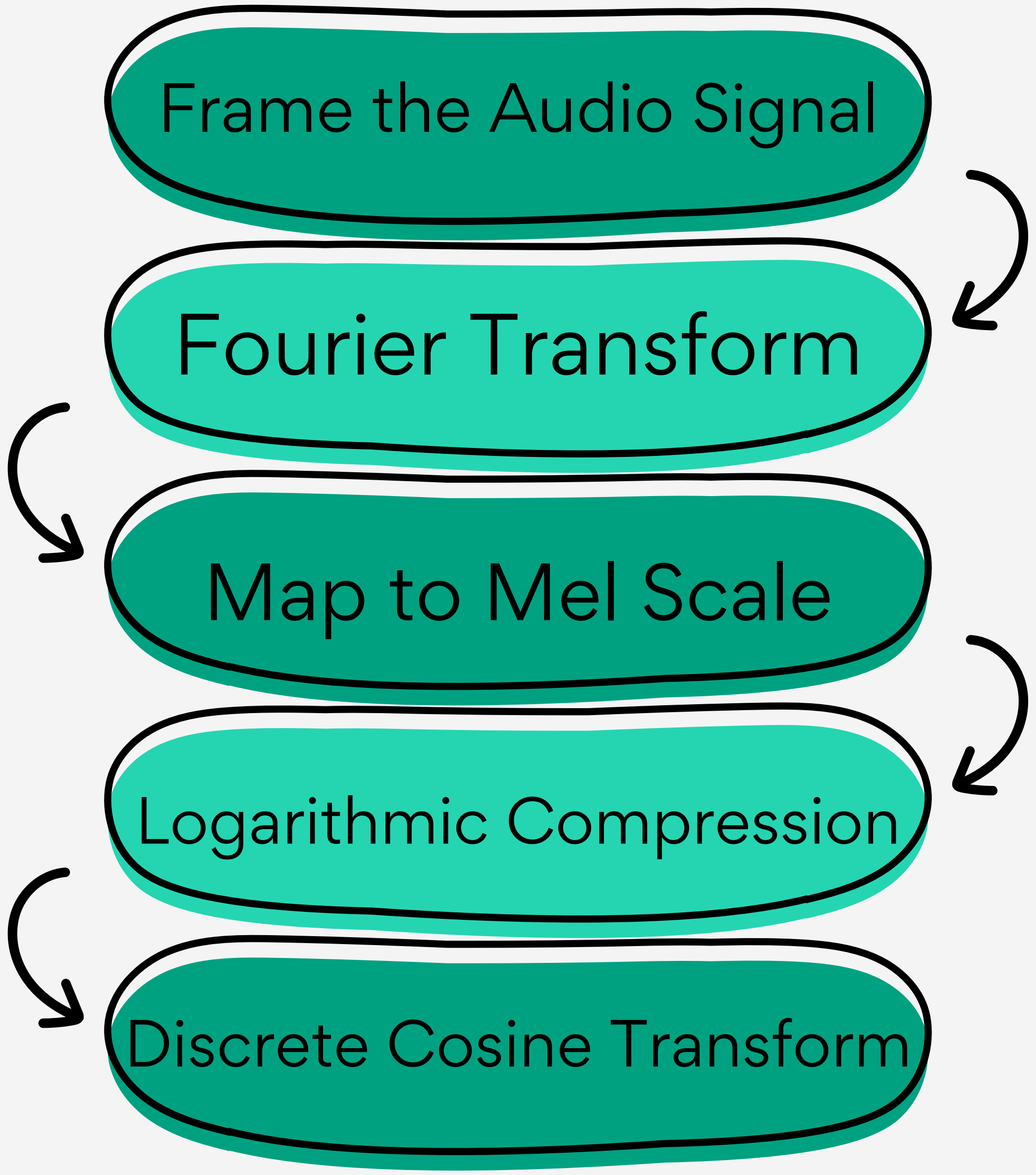
Results, Rationale, Explanations, and Key Challenges

MFCC

What Is it?

MFCCs (Mel-frequency cepstral coefficients) capture essential sound features, like frequency and timbre, by mapping audio to the Mel scale, which reflects our pitch perception. By emphasizing perceptually important frequencies and applying transformations, MFCCs compress sounds into a manageable form, making them ideal for music and speech analysis.

How is it done?



Insights on the Problem

What is the Data Complexity?

Each audio file contains 20 MFCC coefficients sampled at 44100 Hz, resulting in approximately 20 rows by 10,320 columns per file for a 2-minute song, creating a high-dimensional feature space.

01

02

What Are the Classification Goals?

Classify files into predefined categories, such as National Anthem, traditional genres, and specific artists, and detect solo artists like Asha Bhosale, Kishor Kumar, and Michael Jackson.

Insights on the Problem

What Are the Technical Challenges?

High-dimensionality of MFCC features requires dimensionality reduction for efficiency, while feature selection is essential to eliminate irrelevant features and focus on attributes critical for classification.

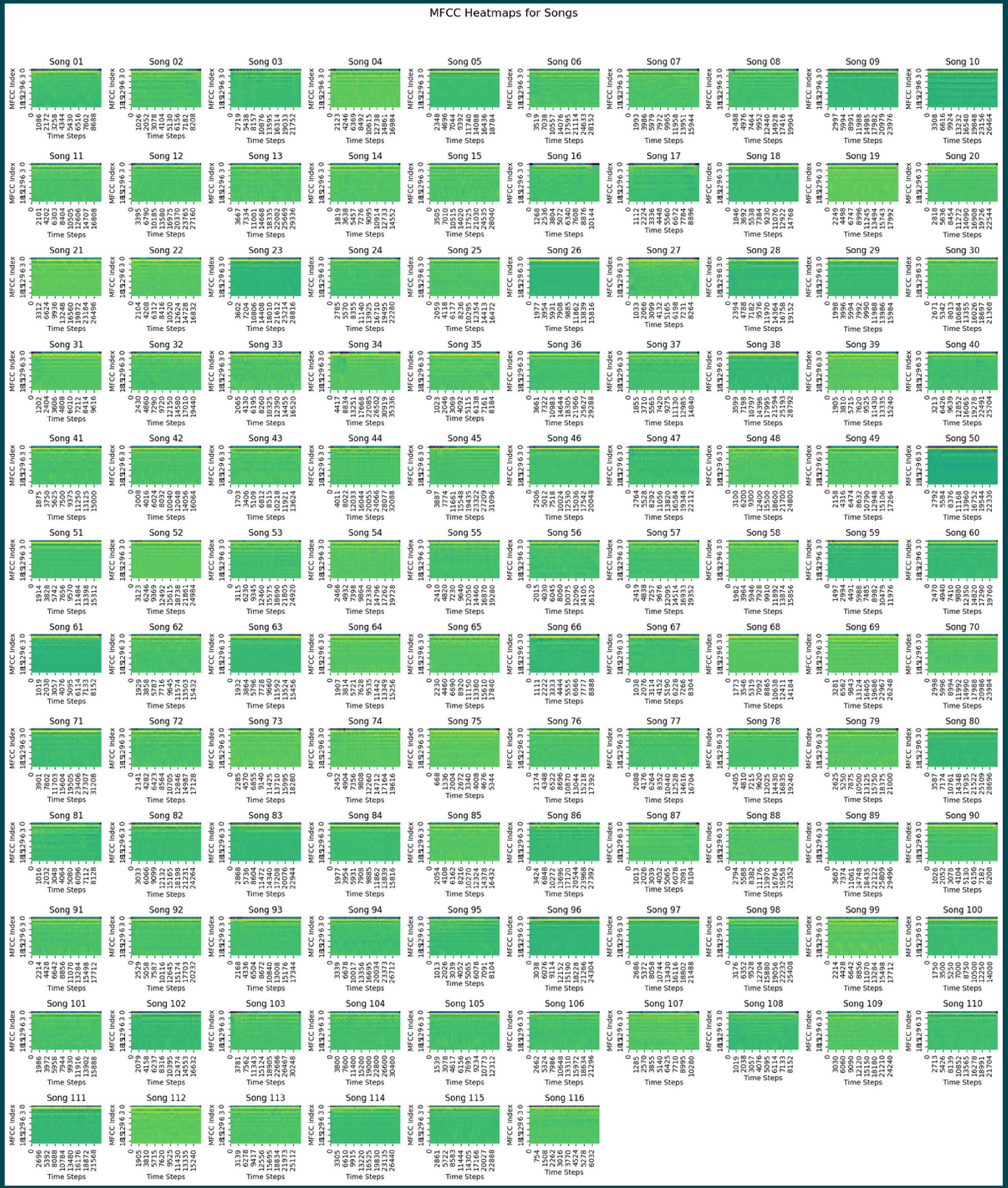
03

04

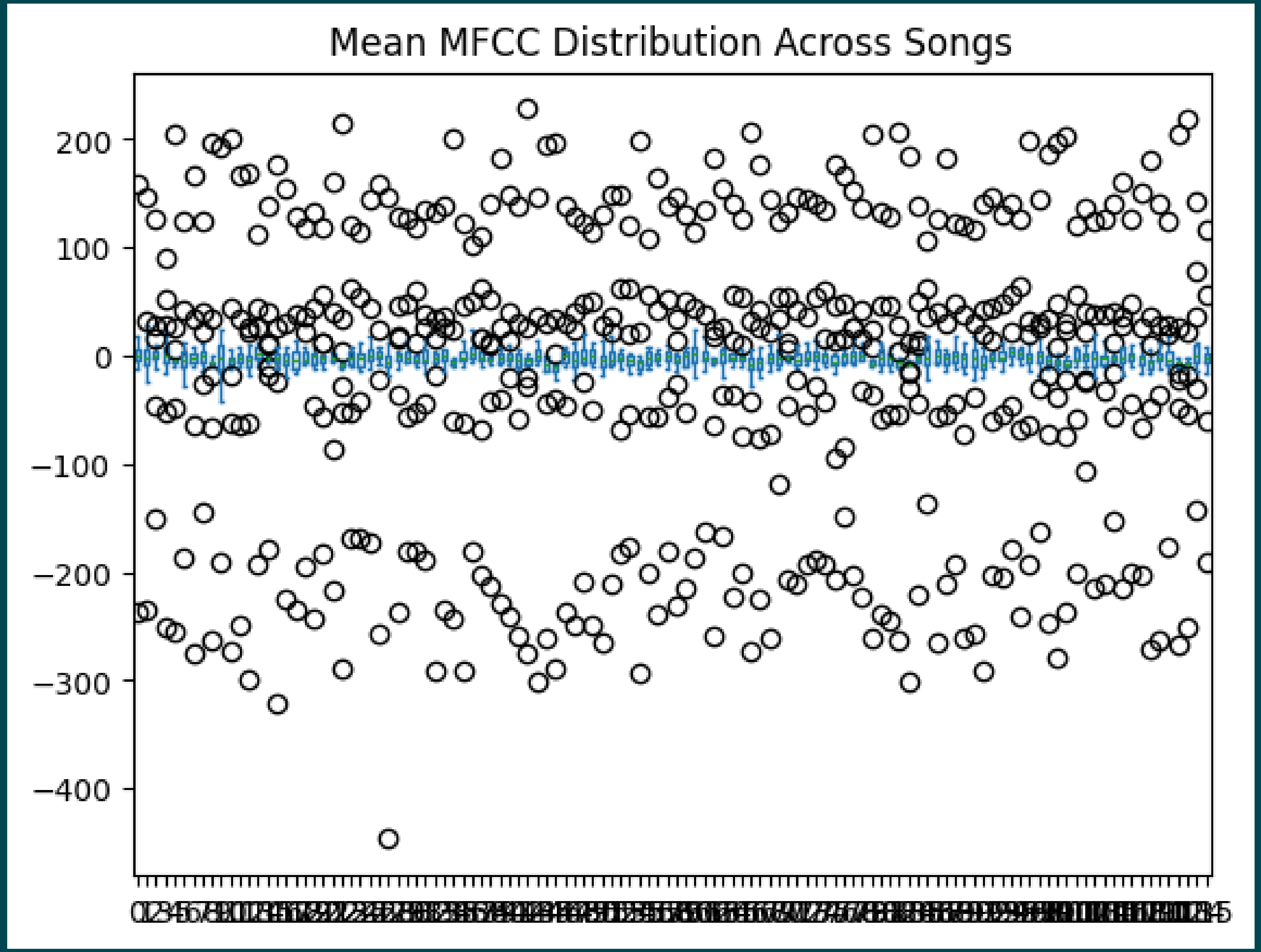
What Is the Proposed Approach?

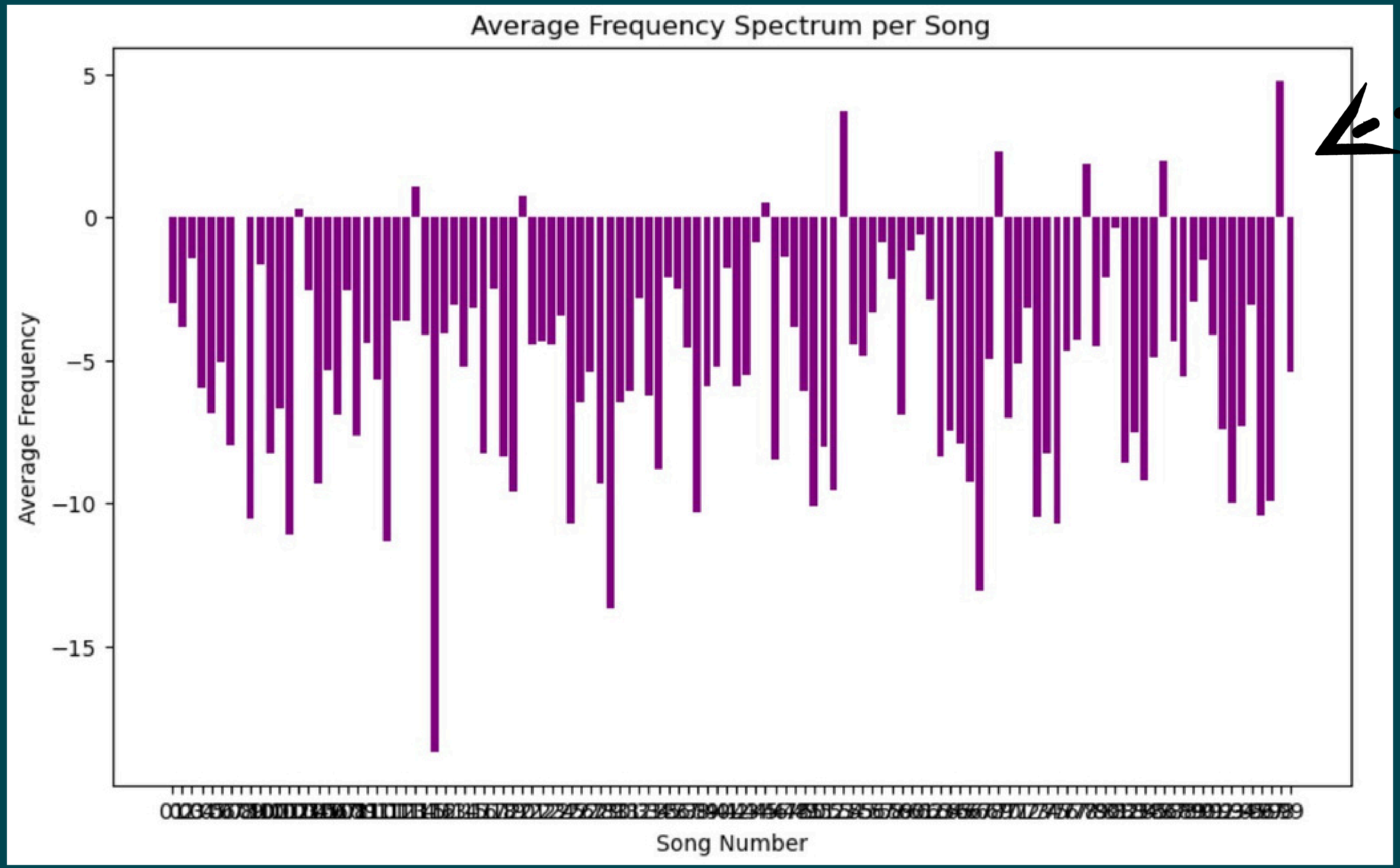
Apply unsupervised clustering to group files by MFCC pattern similarities into potential categories and verify classifications by cross-referencing with known artist and genre attributes.

MFCC Heat Maps – Visualizing Patterns and Structure in Frequency Components



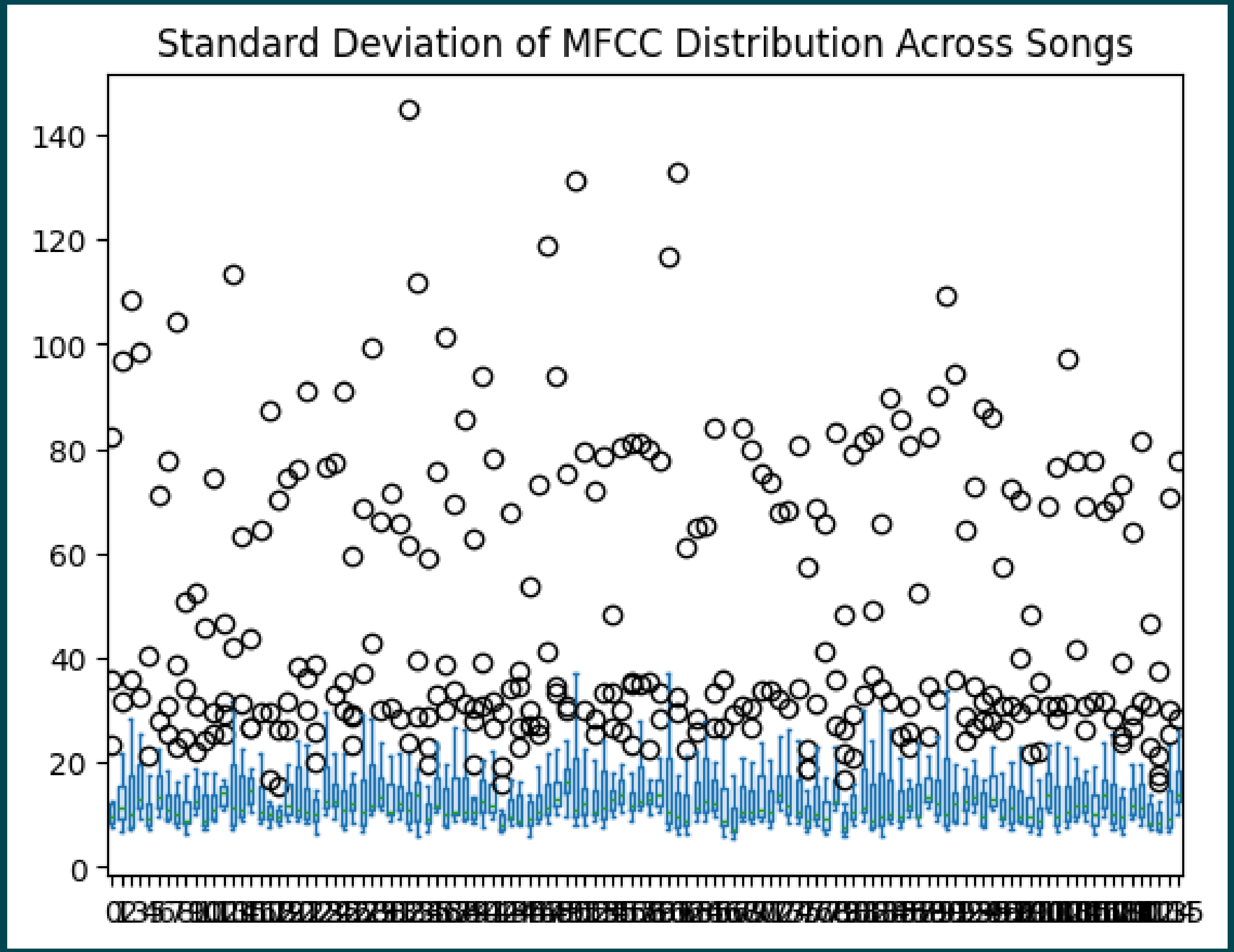
Mean of MFCC Files – Highlighting Average Frequency Characteristics

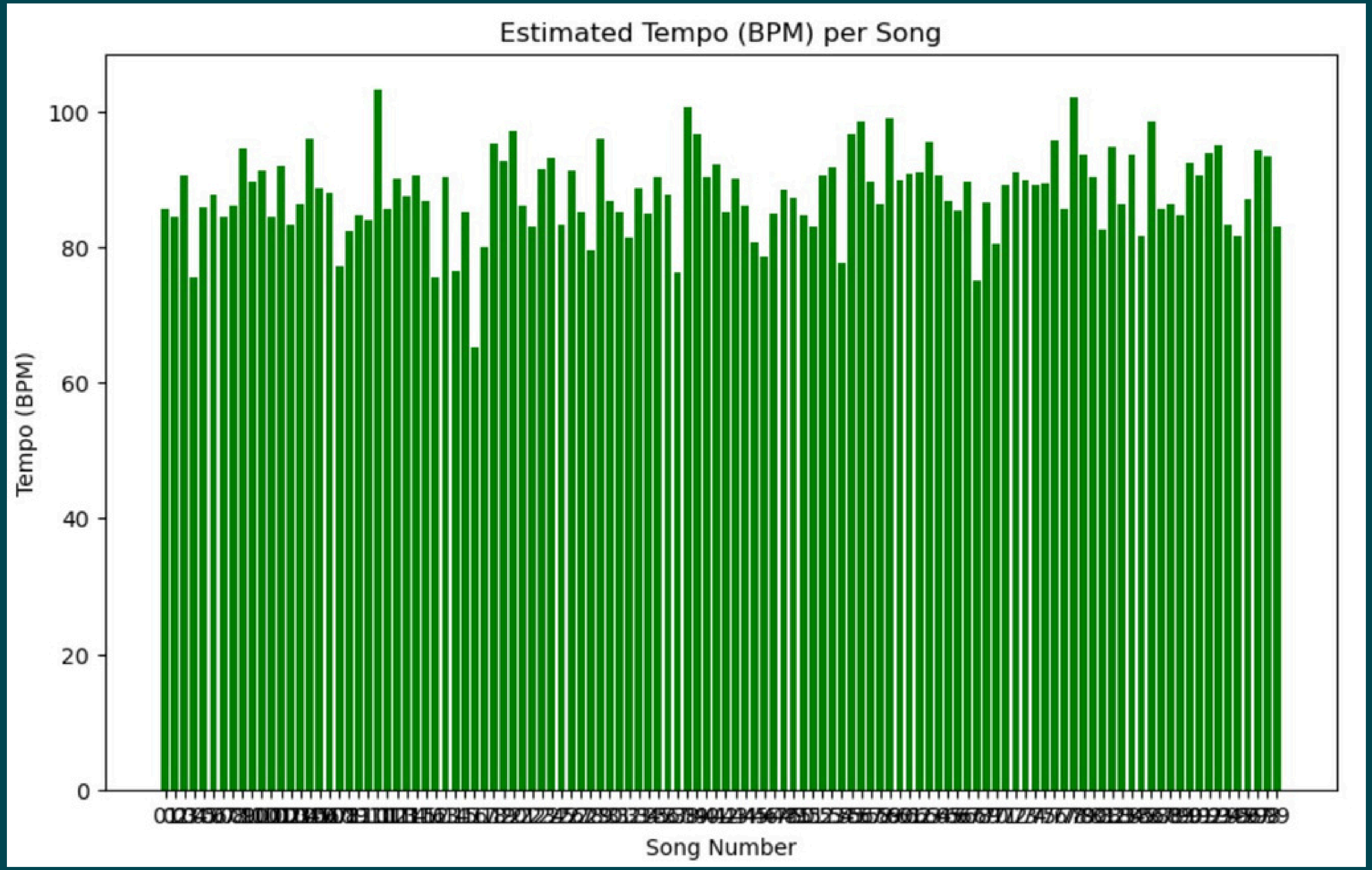




Average Frequency Spectrum – Highlighting spread in frequency of each file

Std Dev of MFCC Coefficients – Exploring Variability Across Frequency Bands





Tempo – Spread in tempo of each file

Finding the Right features

- Basic features failed to produce well-defined clusters..

Column Count	Minimum	Maximum	Mean	Median	Skew
Kurtosis	Std. Deviation	Variance	Delta Kurtosis	Delta Std. Deviation	Delta Variance
Delta (Mean, Median, Variance, Min, Max, Std deviation)	Delta Kurtosis	Delta Delta (Mean, Median, Variance, Min, Max, Std deviation)			

Why Initial Features Fell Short:

Limited Musical Representation: Basic MFCCs, delta, and delta-delta captured spectral and temporal dynamics but missed rhythmic and harmonic nuances, leading to overlapping clusters.

Lack of Genre Differentiation: These features struggled to distinguish complex genres like Lavani and Bhaav Geet, which share similar melodic or structural patterns.

Initial Clusters

Finding the Right features

Roll-off Frequency

Determines the frequency below which most spectral energy is concentrated, helping differentiate genres by emphasizing timbral qualities unique to each cluster.

Maps harmonic characteristics, useful for distinguishing Western pop (e.g., Michael Jackson) from traditional or genre-specific songs (e.g., Bhav Geet or Lavani).

Tonnetz

Spectral Contrast

Highlights amplitude variations across frequency bands, aiding in separating genres and vocal styles based on harmonic and timbral differences.

Analyzes rhythmic pace and beat intensity, useful for identifying clusters with fast-paced styles, such as Lavani.

Tempo and Beat Strength

Spectral Flatness

Indicates how tonal versus noise-like a song is, helping distinguish melodic songs from those with more percussive elements.

Finding the Right features

Spectral Bandwidth

Measures the range of frequencies in a song, distinguishing between denser, complex compositions and simpler ones across the clusters.

Measures how often the signal crosses the zero line

Zero-Crossing Rate (ZCR)

Global MFCC Features

Captures key statistical characteristics of each song's frequency content, aiding in distinguishing broad genre or artist-based variations across the six clusters.

Identifies rhythmic patterns and periodicity within the song, useful for differentiating rhythm-heavy genres like Lavani from less rhythmic styles like Bhav Geet.

Autocorrelation

Energy Band Features

Analyzes energy distribution across frequency bands, helping differentiate clusters by emphasizing bass-heavy songs (e.g., Lavani), midrange vocals (e.g., Bhav Geet), and high-frequency elements (e.g., Michael Jackson's songs).

All Features that we tried

Column Count	Minimum	Maximum	Mean	Median	Skew
Kurtosis	Std. Deviation	Variance	Delta Kurtosis	Delta Std. Deviation	Delta Variance
Delta (Mean, Median, Variance, Min, Max, Std deviation)	Delta Kurtosis	Global MFCC mean	Global Mean	Global Variance	Delta Skew
Global Kurtosis	Global Minimum	Global Maximum	Low MFCC autocorrelation	Mid MFCC autocorrelation	Low Band Energy
Mid Band Energy	High Band Energy	tonnetz (Mean, Median, Variance, Min, Max, Std deviation)	Tempo	Beat Strength (Mean, Median, Variance, Min, Max, Std deviation)	Spectral rolloff (Mean, Median, Variance, Min, Max, Std deviation)
Spectral flatness (Mean, Median, Variance, Min, Max, Std deviation)	Spectral Bandwidth (Mean, Median, Variance, Min, Max, Std deviation)	Delta Delta (Mean, Median, Variance, Min, Max, Std deviation)			

Testing Method

Data Organization and Preprocessing:
Worked in collaboration to organize songs into 6 subfolders by artist, converted each song to MFCC, and stored the outputs as CSV files.

Initial Clustering and Feature Extraction:
Applied clustering using standard feature extraction and filtering to create initial clusters.

Final Feature Mapping and Artist Identification:
Extracted final features from test songs and mapped them to existing clusters, identifying the artist for each cluster.

Solutions And Achievements

- **Stage1: Dimensionality Reduction**

Initially, 527 columns were added, resulting in infinite VIF (Variance Inflation Factor) values for all columns.

To address this, we reduced the number of columns, as infinite VIF leads to poor clustering performance with PCA.

```
Original Normalized Features (first few rows):
        song_id  num_columns  mfcc_1_min  mfcc_1_max  mfcc_1_mean  \
0   70-MFCC.csv        26979  -609.03960  -73.390305  -264.318414
1  104-MFCC.csv        34193  -507.75558   -2.346409  -179.180796
2   67-MFCC.csv         9341  -594.88116 -142.483080  -301.255447
3  110-MFCC.csv        24413  -604.53700  -47.797054  -216.665971
4   92-MFCC.csv        22757  -560.72473  -70.374140  -203.180603

   mfcc_1_median  mfcc_1_std  mfcc_1_variance  mfcc_1_skew  mfcc_1_kurtosis  \
0     -255.84961   80.625633      6500.492697    -1.430020         4.126565
1     -177.26149   63.096498      3981.168110    -1.961053         9.269314
2     -295.13766   65.813001      4331.351080    -0.958775         2.539263
3     -194.19955   91.176897      8313.226535    -1.871377         4.741894
4     -188.89359   69.878202      4882.963167    -2.759736        10.405252

   ...  spectral_flatness_mean  spectral_flatness_std  spectral_flatness_max  \
0  ...                0.045851               0.031891               0.177387
1  ...                0.050083               0.033506               0.172262
2  ...                0.034109               0.029011               0.152825
3  ...                0.065001               0.045277               0.189844
4  ...                0.045737               0.036078               0.152734

   spectral_flatness_min  spectral_flatness_variance  spectral_bandwidth_mean  \
0               0.000386                    0.001017              5795.370642
1               0.000755                    0.001123              5806.697154
2               0.000069                    0.000842              5467.679150
3               0.000137                    0.002050              5307.862927
4               0.000099                    0.001302              5790.067660

   spectral_bandwidth_std  spectral_bandwidth_max  spectral_bandwidth_min  \
0              967.501763             7841.826864                     0.0
1              923.283589             8201.822261                     0.0
2              743.031812             7287.021671                     0.0
3              874.706774             7669.273766                     0.0
4              853.405334             7527.037644                     0.0

   spectral_bandwidth_variance
0                  936059.6606
1                  852452.5865
2                  552096.2734
3                  765111.9396
4                  728300.6644

[5 rows x 527 columns]
```

Solutions And Achievements

- **Variance Threshold Method**

```
Filtered DataFrame with variance threshold applied:
       song_id  num_columns  mfcc_1_min  mfcc_1_max  mfcc_1_mean  \
0   70-MFCC.csv      26979.0  -609.03960  -73.390305  -264.318414
1  104-MFCC.csv      34193.0  -507.75558   -2.346409  -179.180796
2   67-MFCC.csv       9341.0  -594.88116 -142.483080  -301.255447
3  110-MFCC.csv      24413.0  -604.53700  -47.797054  -216.665971
4   92-MFCC.csv      22757.0  -560.72473  -70.374140  -203.180603

   mfcc_1_median  mfcc_1_std  mfcc_1_variance  mfcc_1_skew  mfcc_1_kurtosis  \
0     -255.84961   80.625633      6500.492697    -1.430020         4.126565
1     -177.26149   63.096498      3981.168110    -1.961053         9.269314
2     -295.13766   65.813001      4331.351080    -0.958775         2.539263
3     -194.19955   91.176897      8313.226535    -1.871377         4.741894
4     -188.89359   69.878202      4882.963167    -2.759736        10.405252

   ...  spectral_rolloff_mean  spectral_rolloff_std  spectral_rolloff_max  \
0  ...           10322.055460           3290.806809           18568.42105
1  ...           10622.618820           3339.281314           19728.94737
2  ...            8832.846421           2429.348784           16247.36842
3  ...            8248.326819           3645.354205           18568.42105
4  ...           10149.926340           3034.337074           18568.42105

   spectral_rolloff_min  spectral_rolloff_variance  spectral_bandwidth_mean  \
0                   0.0               1.082941e+07              5795.370642
1                   0.0               1.115080e+07              5806.697154
2                   0.0               5.901736e+06              5467.679150
3                   0.0               1.328861e+07              5307.862927
4                   0.0               9.207201e+06              5790.067660

   spectral_bandwidth_std  spectral_bandwidth_max  spectral_bandwidth_min  \
0              967.501763             7841.826864                     0.0
1              923.283589             8201.822261                     0.0
2              743.031812             7287.021671                     0.0
3              874.706774             7669.273766                     0.0
4              853.405334             7527.037644                     0.0

   spectral_bandwidth_variance
0                  936059.6606
1                  852452.5865
2                  552096.2734
3                  765111.9396
4                  728300.6644

[5 rows x 350 columns]
```

Applied a threshold-based method to remove highly correlated features, reducing the columns to 350, though further optimization is needed.

Solutions And Achievements

- **Correlation Threshold Method**

Calculated the correlation matrix for numeric features, identified pairs with correlations above 0.80, and dropped the highly correlated features, reducing the feature set to 196 columns, which was saved to a new CSV file.

```
Reduced DataFrame:
       song_id  num_columns  mfcc_1_min  mfcc_1_max  mfcc_1_mean  mfcc_1_std  \
0   70-MFCC.csv      26979.0  -609.03960  -73.390305  -264.318414   80.625633
1  104-MFCC.csv      34193.0  -507.75558   -2.346409  -179.180796   63.096498
2   67-MFCC.csv       9341.0  -594.88116 -142.483080  -301.255447   65.813001
3  110-MFCC.csv      24413.0  -604.53700  -47.797054  -216.665971   91.176897
4   92-MFCC.csv      22757.0  -560.72473  -70.374140  -203.180603   69.878202

   mfcc_1_skew  mfcc_2_min  mfcc_2_max  mfcc_2_std  ...  global_mfcc_skewness  \
0    -1.430020         0.0   223.88953   30.733178  ...             -2.773363
1    -1.961053         0.0   206.41900   31.398634  ...             -1.373236
2    -0.958775         0.0   247.00610   34.255390  ...             -1.812997
3    -1.871377         0.0   257.14000   36.254801  ...             -1.758957
4    -2.759736         0.0   214.04141   28.291418  ...             -1.509140

   mid_band_energy  high_band_energy       tempo  beat_strength_max  \
0       244.971819        143.627300   97.508844          347.84494
1       279.374255        106.293856   77.133862          302.58400
2       253.754365        117.637790   93.963068          364.80390
3       117.785441        104.729146  107.666016          372.52054
4       236.872841        120.616427  117.453835          322.01288

   beat_strength_min  spectral_rolloff_mean  spectral_rolloff_std  \
0           6.702176           10322.055460           3290.806809
1           4.986100           10622.618820           3339.281314
2           5.842791            8832.846421           2429.348784
3           4.508741            8248.326819           3645.354205
4           5.896157           10149.926340           3034.337074

   spectral_rolloff_min  spectral_bandwidth_min
0                   0.0                     0.0
1                   0.0                     0.0
2                   0.0                     0.0
3                   0.0                     0.0
4                   0.0                     0.0

[5 rows x 198 columns]
```

Solutions And Achievements

- **Normalization**

```
VIF values for each feature:
                   Feature  VIF
0              num_columns  inf
1               mfcc_1_min  inf
2               mfcc_1_max  inf
3              mfcc_1_mean  inf
4               mfcc_1_std  inf
..                     ...  ...
192      beat_strength_min  inf
193  spectral_rolloff_mean  inf
194   spectral_rolloff_std  inf
195   spectral_rolloff_min  inf
196 spectral_bandwidth_min  inf

[197 rows x 2 columns]
```

Normalized the data for efficient VIF calculation, but the VIF values remained infinite, indicating the need for further optimization.

Solutions And Achievements

- **Singular Value Decomposition (SVD):**

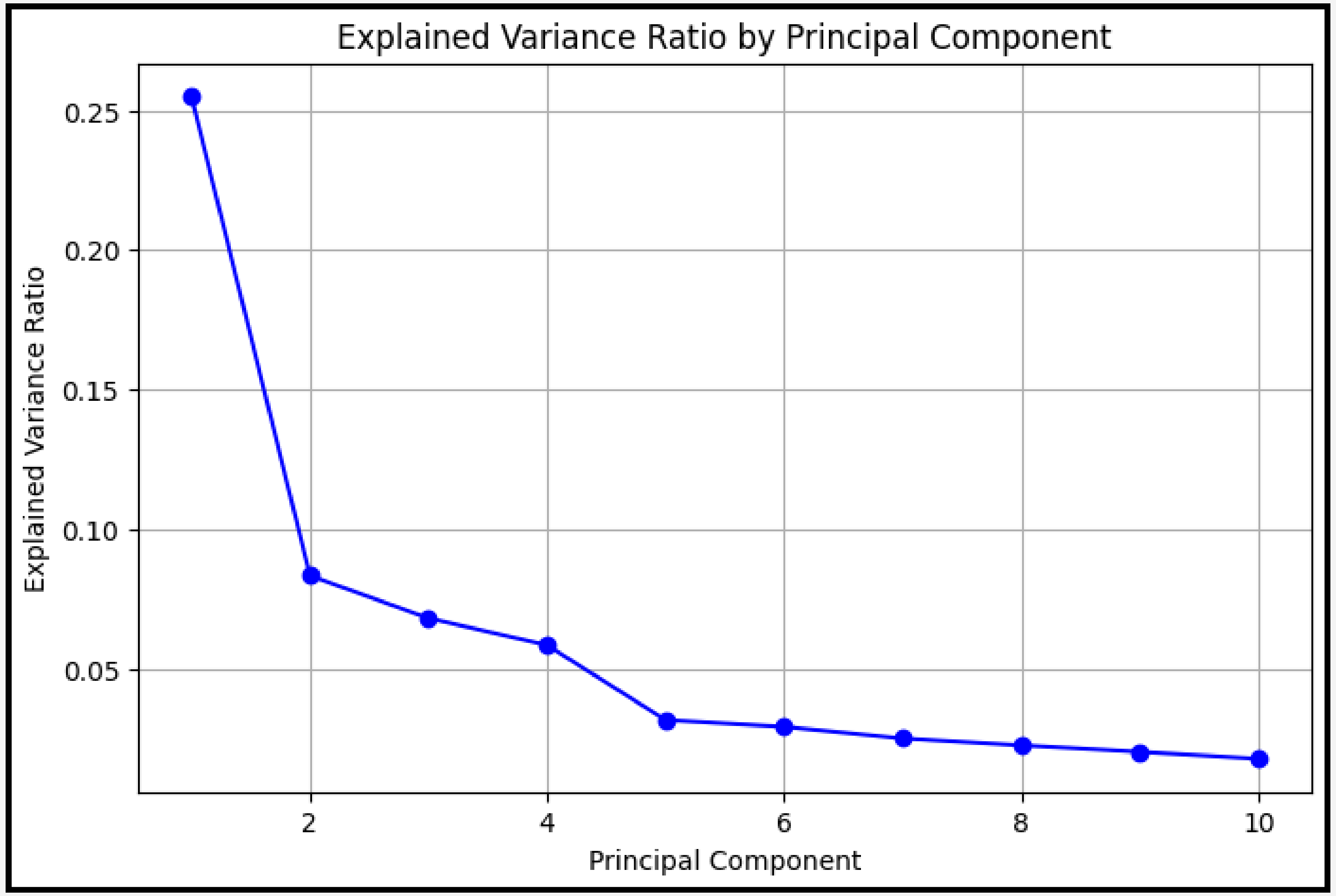
To address collinearity, I applied SVD, reducing the feature set to 116 columns, and achieved a VIF of 1, indicating minimal multicollinearity.

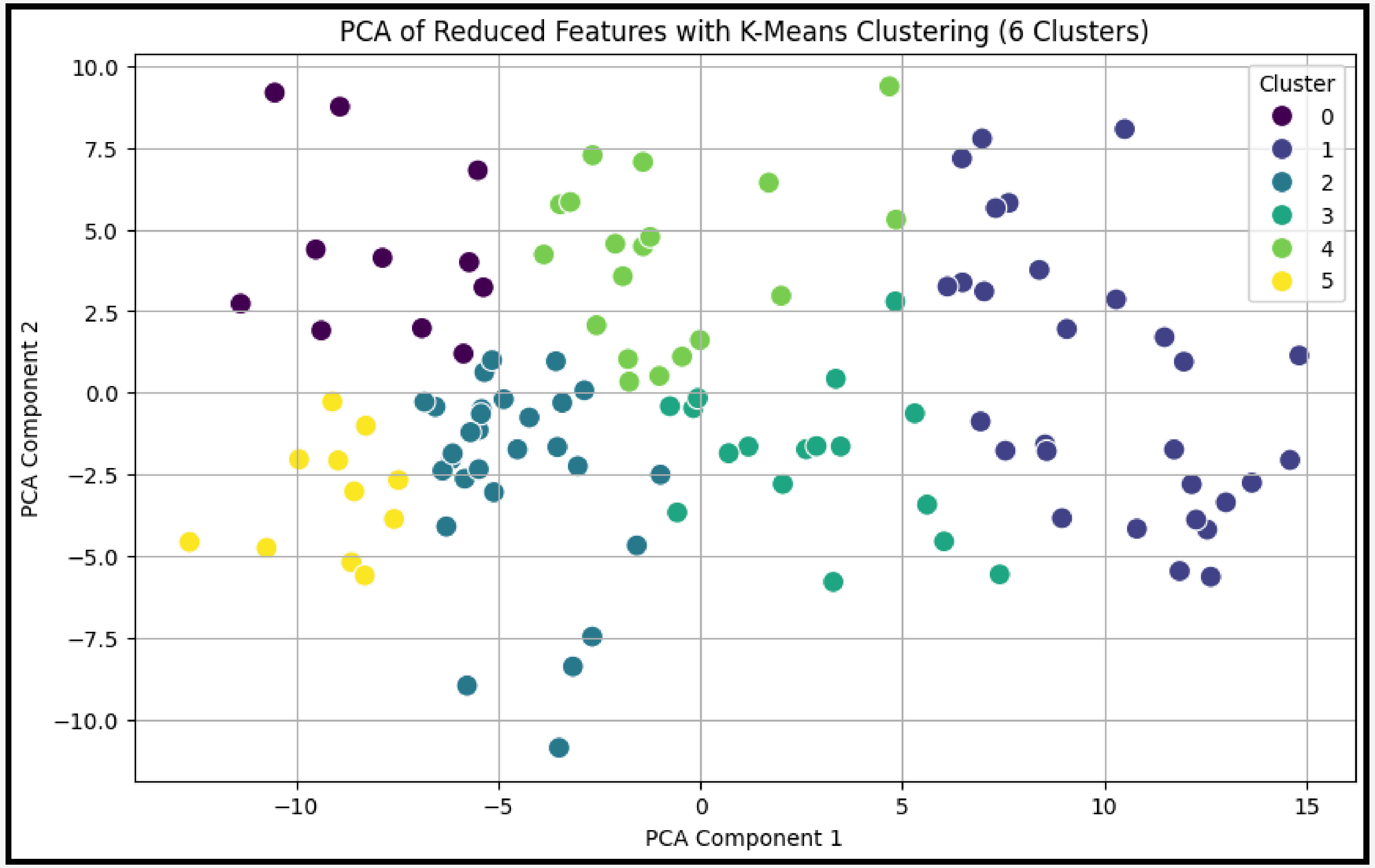
```
Number of features: 197
Shape of reduced features: (116, 116)
      SVD_1     SVD_2     SVD_3     SVD_4     SVD_5     SVD_6     SVD_7  \
0  -5.411210 -0.503476 -1.859148 -0.994620 -1.853217  1.530600  2.616344
1  -4.863628 -0.193596 -3.045104  2.034780  1.487988 -3.076766 -0.471716
2  11.972534  0.949673  3.189433 -4.380771 -4.829559 -1.878644  0.524347
3  -1.568310 -4.664805  3.649952 -7.470024  8.774862  1.445513 -0.886039
4   3.356655  0.429853 -2.098160  0.412222 -1.569736 -0.894637  0.265695

      SVD_8     SVD_9    SVD_10  ...   SVD_107   SVD_108   SVD_109   SVD_110  \
0 -1.883947 -0.979717  1.552497  ... -0.202672  0.516280  0.018467  0.098265
1  1.838459  0.191701 -1.194897  ... -0.356009  0.494130 -0.134710  0.097928
2  0.080470 -2.803012  0.122771  ... -0.088574  0.011964 -0.402476  0.140782
3 -0.512473  3.103010  3.234703  ... -0.056845  0.042861 -0.066343 -0.284212
4 -0.414574 -2.396303  0.346076  ... -0.111754  0.050189  0.029843 -0.217695

    SVD_111   SVD_112   SVD_113   SVD_114   SVD_115       SVD_116
0  0.062995 -0.263466 -0.122438 -0.017656 -0.006657  2.857861e-15
1 -0.235056 -0.040714  0.172008 -0.074854 -0.021319  2.136216e-15
2 -0.007058 -0.503905 -0.036682 -0.059142 -0.049003  2.115399e-15
3 -0.232688 -0.010820  0.104399 -0.101826  0.007336  2.441527e-15
4  0.024452  0.111550 -0.061510 -0.250262  0.006015  2.517855e-15

[5 rows x 116 columns]
```

Solutions And Achievements

- **Stage 2: Clustering**

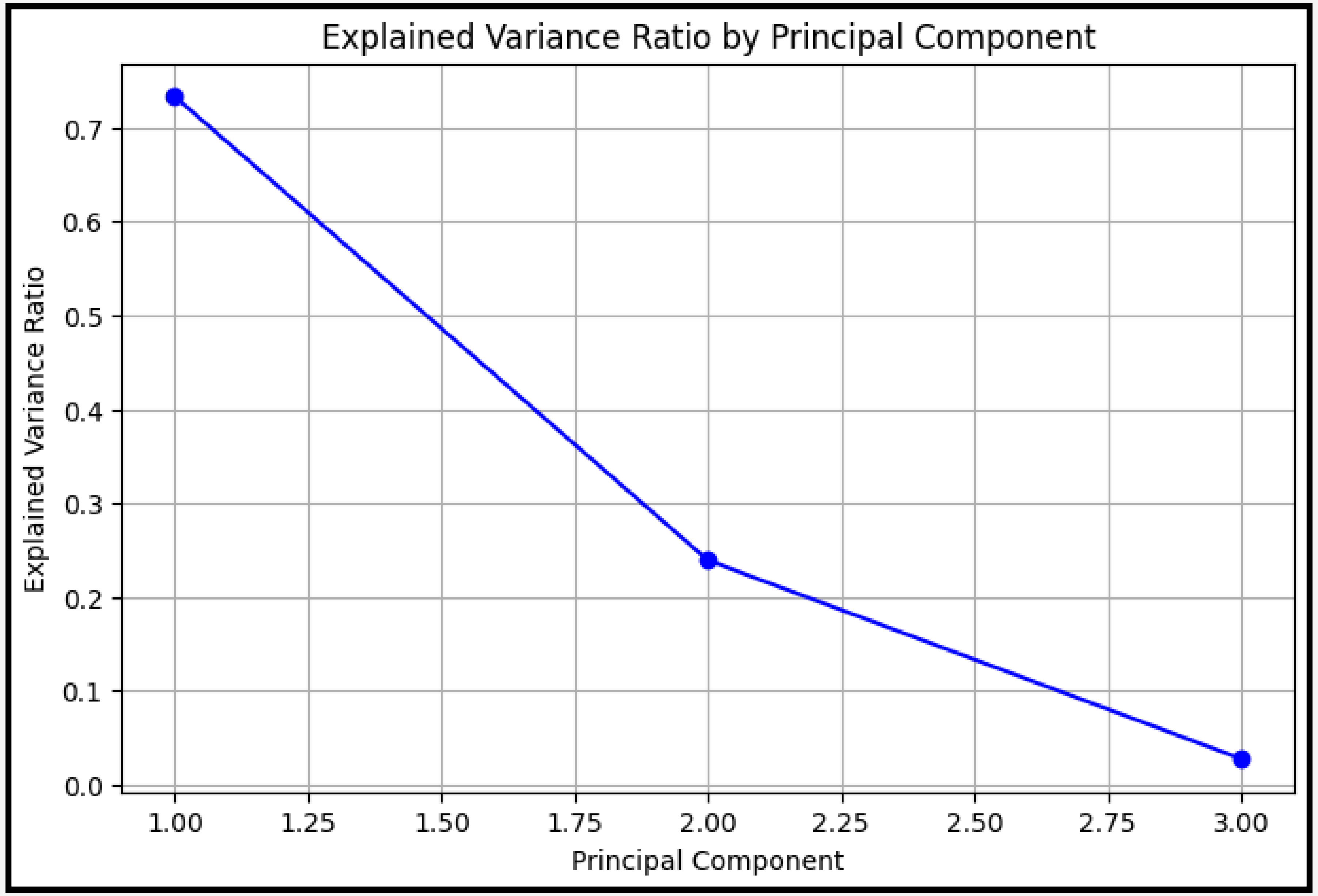
With the features optimized and VIF reduced to 1, I was ready to apply PCA, ensuring better clustering by minimizing correlation among the columns.

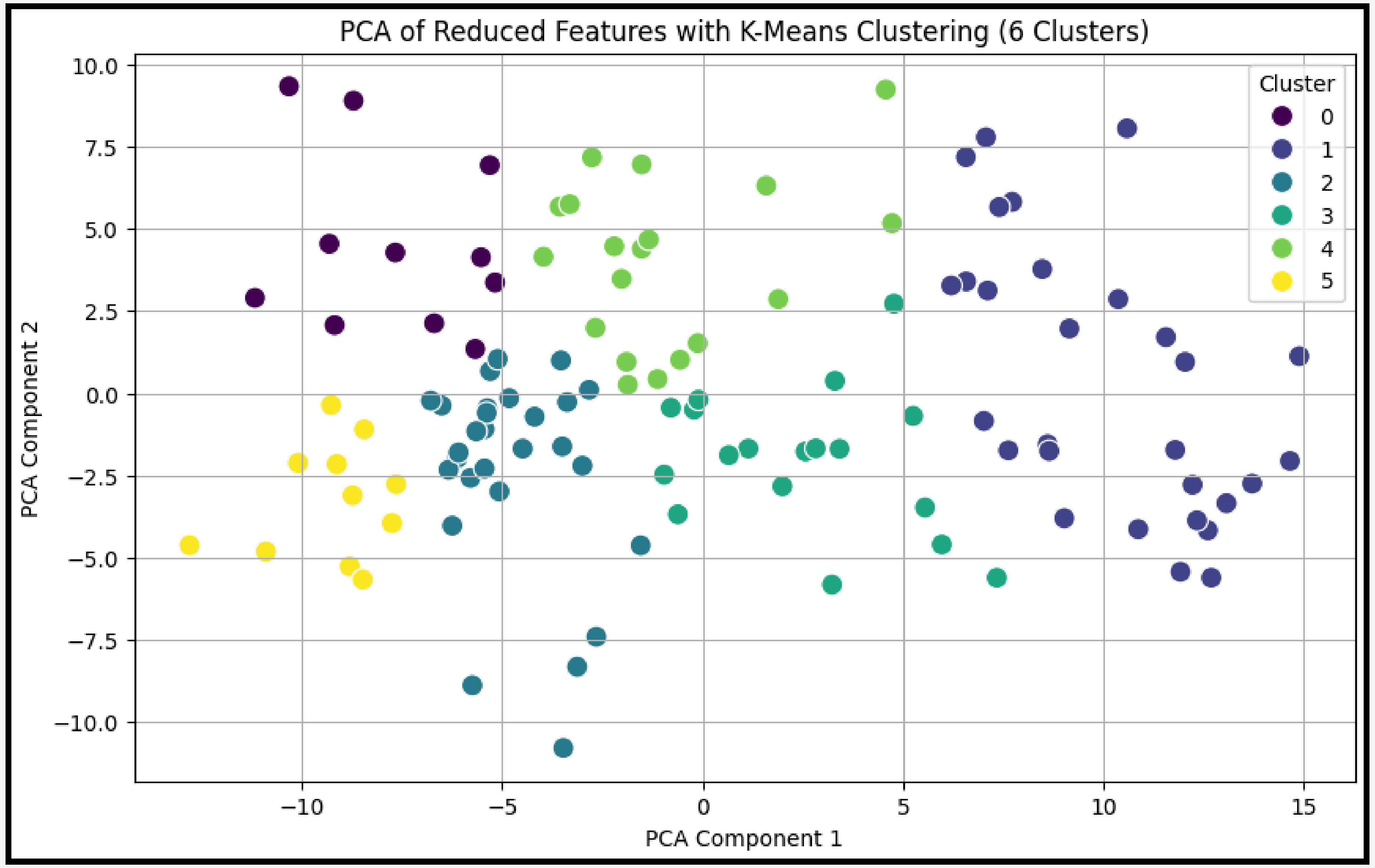




Solutions And Achievements

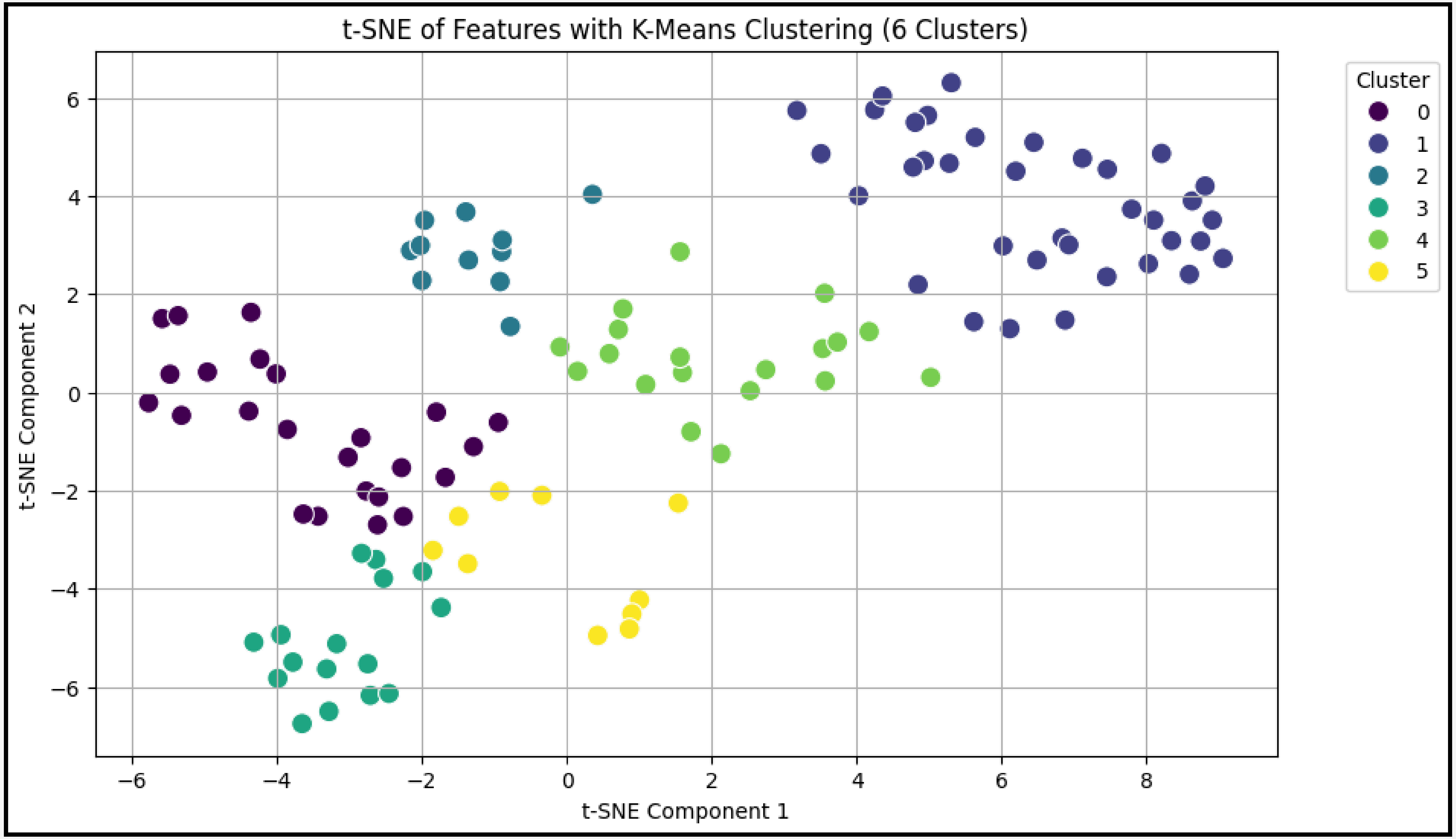
Despite the optimization, the clustering results were not satisfactory, so we applied PCA once again to improve the clustering.



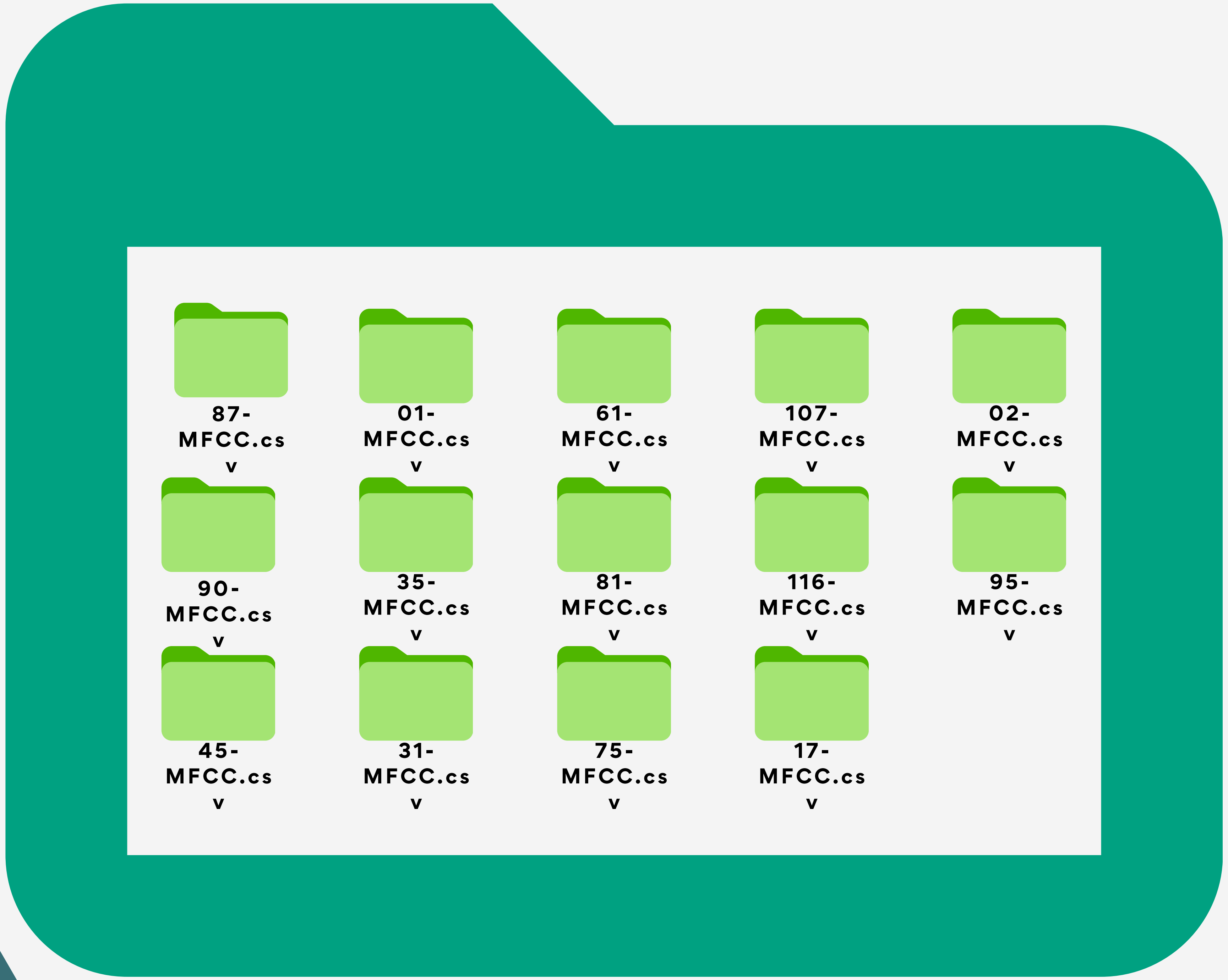


Solutions And Achievements

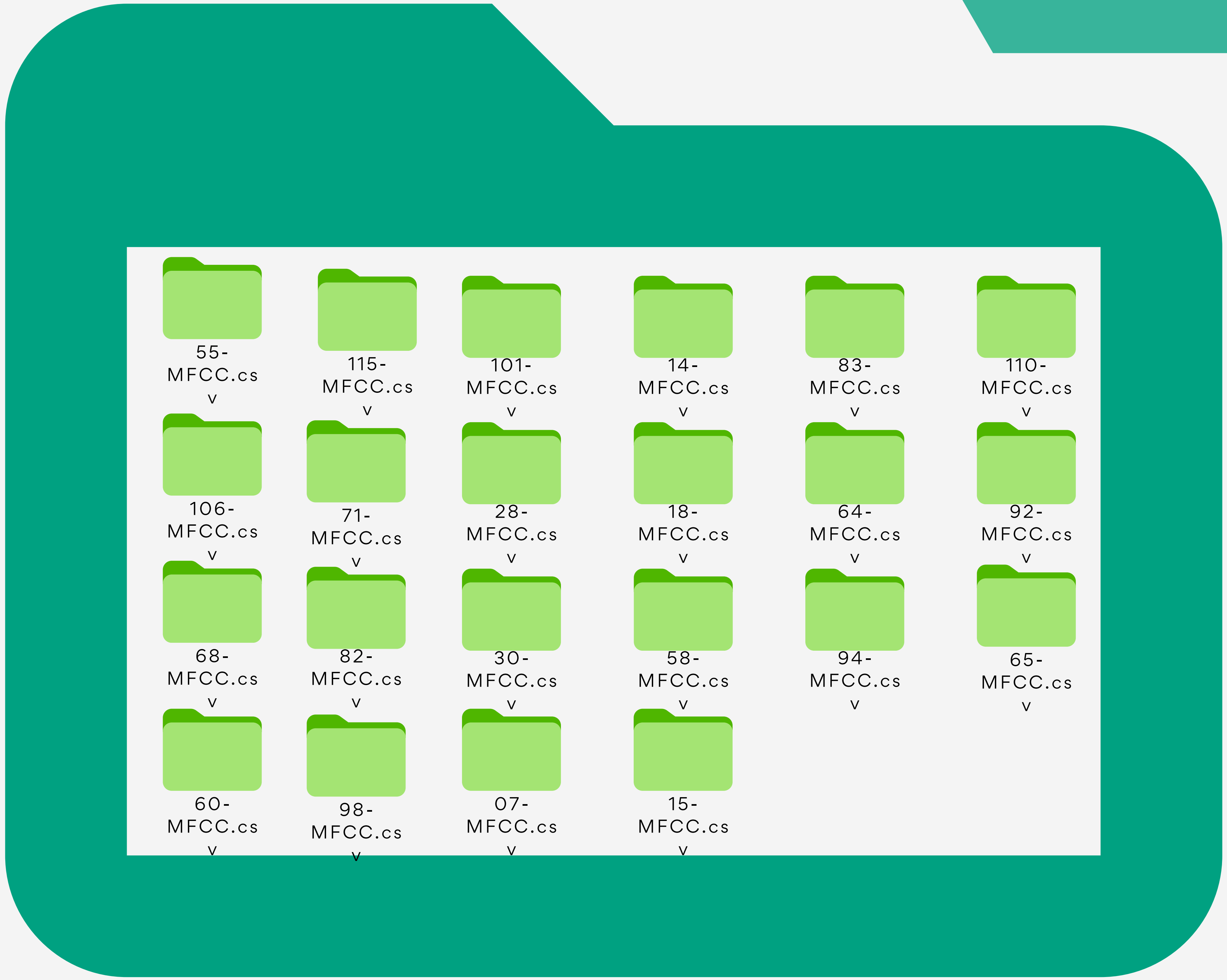
Finally, we applied t-SNE, which provided the best clusters, resulting in the desired outcome and highly satisfactory results for our project.



Results



National Anthem



Lavni

Results

33-MFCC.csv
23-MFCC.csv
32-MFCC.csv
99-MFCC.csv
77-MFCC.csv
105-MFCC.csv
39-MFCC.csv
38-MFCC.csv
40-MFCC.csv
102-MFCC.csv

Asha Bhosle

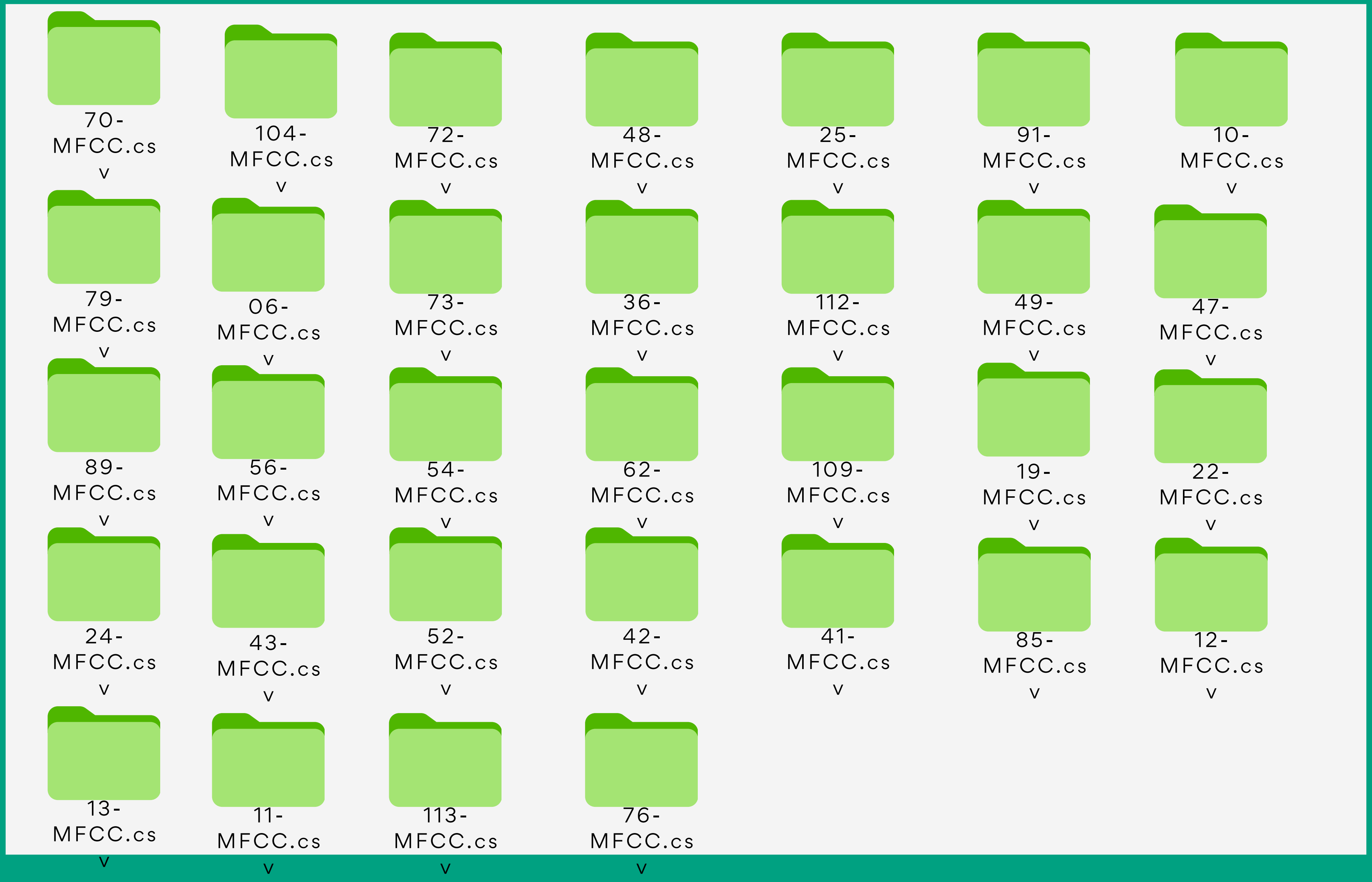
86-MFCC.csv
44-MFCC.csv
03-MFCC.csv
53-MFCC.csv
74-MFCC.csv
78-MFCC.csv
21-MFCC.csv
34-MFCC.csv
57-MFCC.csv
103-MFCC.csv
114-MFCC.csv
37-MFCC.csv
08-MFCC.csv
04-MFCC.csv
69-MFCC.csv
88-MFCC.csv
20-MFCC.csv

Michael Jackson

Results

Kishore Kumar

Results



Bhav Geet

Solutions And Achievements

- **Clustering Results**

```
Silhouette Score for existing 6 K-Means clusters: 0.43565982580184937
Davies-Bouldin Index for existing 6 K-Means clusters: 0.7336657962084835

Cluster Counts:
Cluster 0: 24 points
Cluster 1: 36 points
Cluster 2: 11 points
Cluster 3: 16 points
Cluster 4: 19 points
Cluster 5: 10 points
```

The Silhouette Score and DB Index were satisfactory based on our test data. Michael Jackson and National Anthem clusters were perfect, matching every test, while Asha Bhosle was highly scattered. The remaining clusters showed decent results, with about a 70% success rate.

Initial Observation:

- Michael Jackson and National Anthem clusters showed high accuracy and clear genre separation.
- Lavani, Kishore Kumar, Asha Bhosle, and Bhaav Geet clusters had significant overlap and misclassification.

What We Did:

- Removed Michael Jackson and National Anthem clusters to improve clustering precision.
- Focused on refining the algorithm for overlapping clusters, minimizing bias from distinct groups.

The Outcome:

- Re-clustering improved genre distinction and classification accuracy.
- Songs were grouped into more meaningful and precise clusters.

Results

Asha Bhosle

110-MFCC.csv
33-MFCC.csv
23-MFCC.csv
79-MFCC.csv
12-MFCC.csv
32-MFCC.csv
77-MFCC.csv
105-MFCC.csv
56-MFCC.csv
38-MFCC.csv
40-MFCC.csv
24-MFCC.csv
115-MFCC.csv
15-MFCC.csv

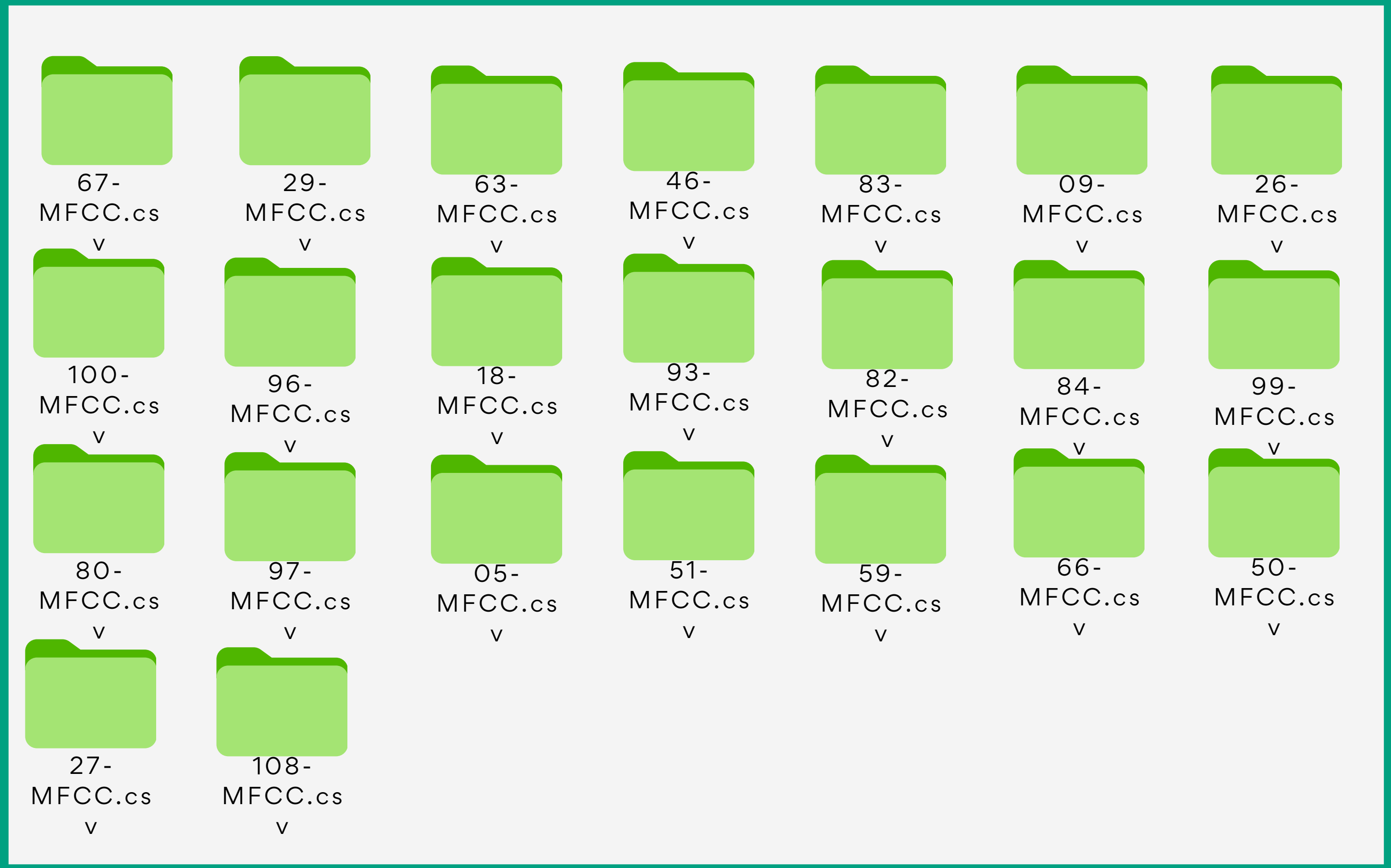
Lavani

92-MFCC.csv
65-MFCC.csv
94-MFCC.csv
64-MFCC.csv
14-MFCC.csv
58-MFCC.csv
101-MFCC.csv
28-MFCC.csv
30-MFCC.csv
71-MFCC.csv
07-MFCC.csv
15-MFCC.csv
60-MFCC.csv
68-MFCC.csv
106-MFCC.csv
55-MFCC.csv
98-MFCC.csv

Results

Bhav Geet

Kishore Kumar



Solutions And Achievements

- ## Clustering Results2.0

```
Silhouette Score for existing 6 K-Means clusters: 0.4419495761394501
Davies-Bouldin Index for existing 6 K-Means clusters: 0.8154454065866006

Cluster Counts:
Cluster 0: 23 points
Cluster 1: 23 points
Cluster 2: 20 points
Cluster 3: 18 points
```

Asha Bhosle still overlapped significantly with other clusters, particularly with Bhaav Geet and Lavani. This overlap occurred because of the high similarity between Asha Bhosle and Bhaav Geet, and between Asha Bhosle and Lavani, as many songs in these genres were sung by Asha Bhosle, resulting in a large number of MFCC files for these clusters.

Evaluation-Driven Project Outcomes

How many of the Problems were correctly solved?

1. Analyze MFCC files to organize the 115 files into groups broadly corresponding to those listed above
2. Identify at least 3 files containing the National Anthem
3. Identify at least 3 files (each) containing solo songs by Asha Bhosale, Kishor Kumar, and Michael Jackson

1) Problem 1 has been addressed, and the solutions are presented in the slides.

2) These are surely National Anthem: 75, 116, 02
Many other have been written in the slides as well

3) Solo Songs by Asha Bhosle: 110,15,79
Solo Songs by Kishore Kumar: 09,05,97
Solo Songs by Michael Jackson:3, 114, 08

Where does our Innovation/Creativity Come in?

- **Creative Feature Engineering:** Applied energy band features, roll-off frequency, and tonal centroid analysis to effectively distinguish between different song genres and artists**.**
- **Advanced Techniques for Dimensionality Reduction:** Utilized PCA and SVD to reduce the feature set, minimizing collinearity and improving clustering efficiency.
- **Improved Clustering Visualization:** Used t-SNE for effective visualization and interpretation of clusters, enhancing the clarity and accuracy of the final results.

Relevance and Justifications

- **Relevance to Genre Differentiation:** Features like energy band features, roll-off frequency, and tonal centroid analysis were specifically chosen to capture the unique musical characteristics that differentiate genres and artists, enhancing the relevance of the features to the problem at hand.
- **Handling Complex Genre Overlap:** By focusing on both spectral and temporal features, including rhythm and harmonic content, the feature engineering process was tailored to address challenges like genre overlap, particularly between similar genres sung by the same artist (e.g., Asha Bhosle in Bhaav Geet and Lavani).
- **Dimensionality Reduction for Precision**: Techniques like PCA and SVD were applied to reduce the feature set while maintaining the most relevant information, optimizing the model's ability to capture critical genre characteristics and improving clustering outcomes.

Timeline

- **Oct 21** — Initial discussions regarding the project, building understanding of concepts like MFCC, and speech recognition
- **Oct 27** — Pitching ideas, extracting features out of the MFCC coefficients and beginning cluster formation
- **Nov 1** — Analyzing errors in clusters, discussing appropriate changes to make
- **Nov 5** — Finalizing the clusters, generating required plots, and started working on report
- **Nov 7** — concluding with the report and collecting all the notebooks/code into a zip file

Major Challenges And Overcoming them

CHALLENGE 1

Infinite VIF and Poor PCA Performance:
Infinite VIF values led to ineffective PCA; dimensionality reduction was applied to address this.

CHALLENGE 2

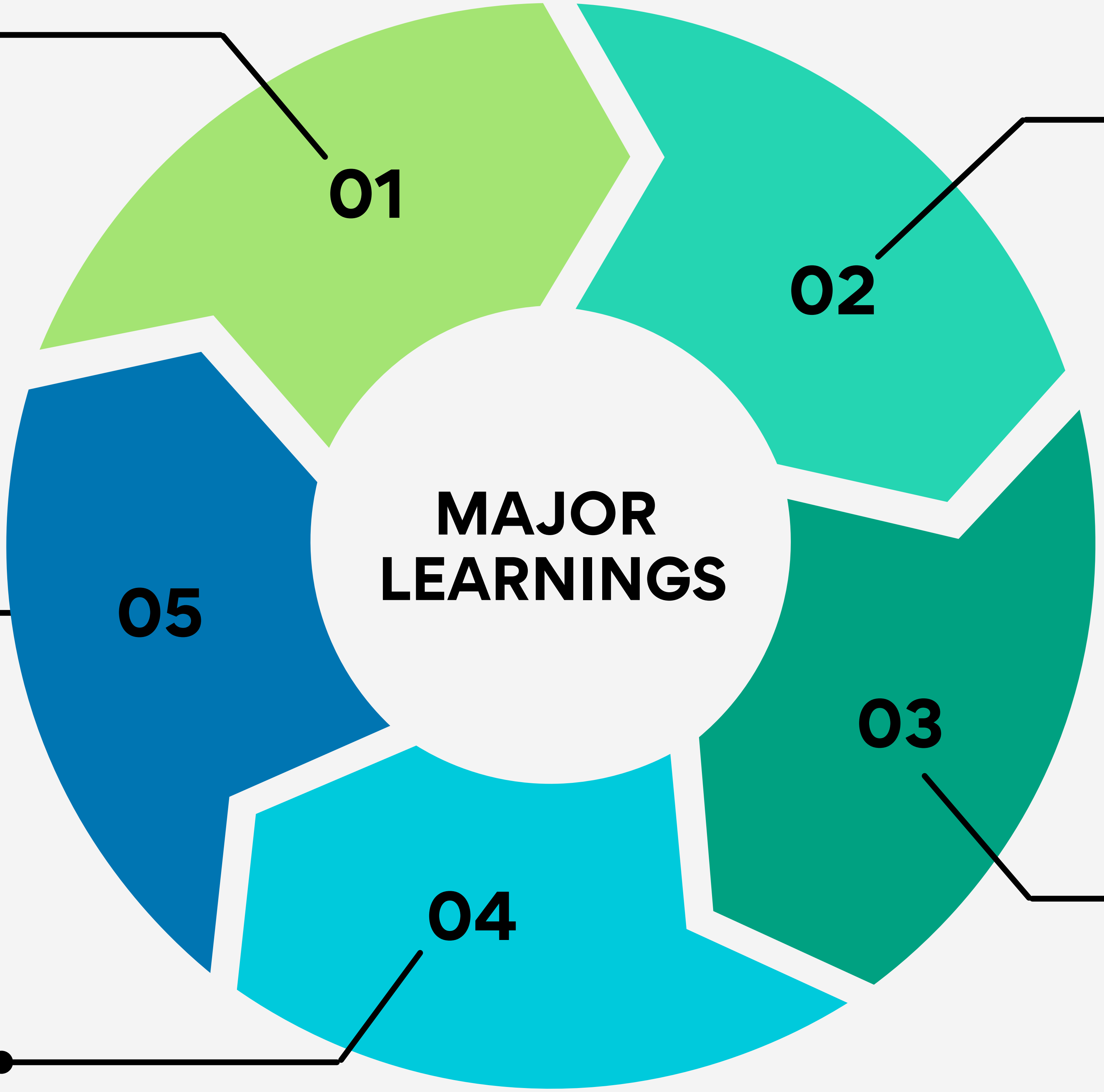
Cluster Confusion for Michael Jackson:
Michael Jackson's songs were highly mixed, leading to difficulty in forming a distinct cluster

CHALLENGE 3

Asha Bhosle's Genre Overlap:
Asha Bhosle's songs overlap with multiple genres due to her collaborations, causing difficulty in isolating her cluster.

CHALLENGE 4

Limited MFCC Data and Testing Methods:
There was a lack of available MFCC data online, and no established testing methods to validate the clustering results.



01 Learned to extract and utilize MFCCs to represent sound characteristics for music analysis.

02 Applied machine learning to categorize songs by singer, genre, and instrumentation.

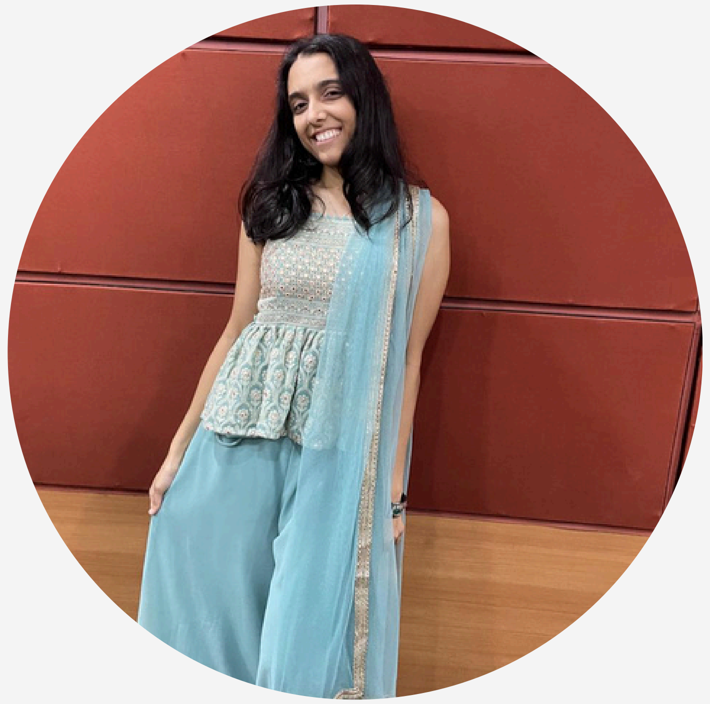
03 Managed large MFCC datasets, refining skills in dimensionality reduction.

04 Identified audio patterns to differentiate genres and artists based on frequency and timbre.

05 Practiced teamwork, dividing tasks and streamlining complex problem-solving processes.

Team Profile

Dev Arora
Electrical(23B1271)



Foram Trivedi
Electrical(23B1269)

Manya Shah
Electrical(23B3969)

Shubh Tangri
Electrical(23B3931)

THANK YOU