Spotify

MUSIC FOR EVERYONE

EXPLORING MUSIC GENRE PREFERENCES ON SPOTIFY

-Group 4

Agenda

- Introducing Dataset
- Goal
- Exploratory Data Analysis
- Pre-processing the data
- Clustering
- Mixture Model
- Neural Network algorithm
- Conclusion
- Limitations and Future Direction

Have you ever wondered how music streaming platforms like Spotify create personalized playlists for millions of users?

Introducing Dataset

- Spotify Dataset from Kaggle
- 114000 observations with 20 variables
- 14 numeric variables, 1 binary variable and 5 character variables
- Consists of characteristics of the songs released on spotify like energy, danceability, loudness, acousticness, liveness, speechiness etc.,
- Contains NA’s in “Year” column
- Scaled our dataset

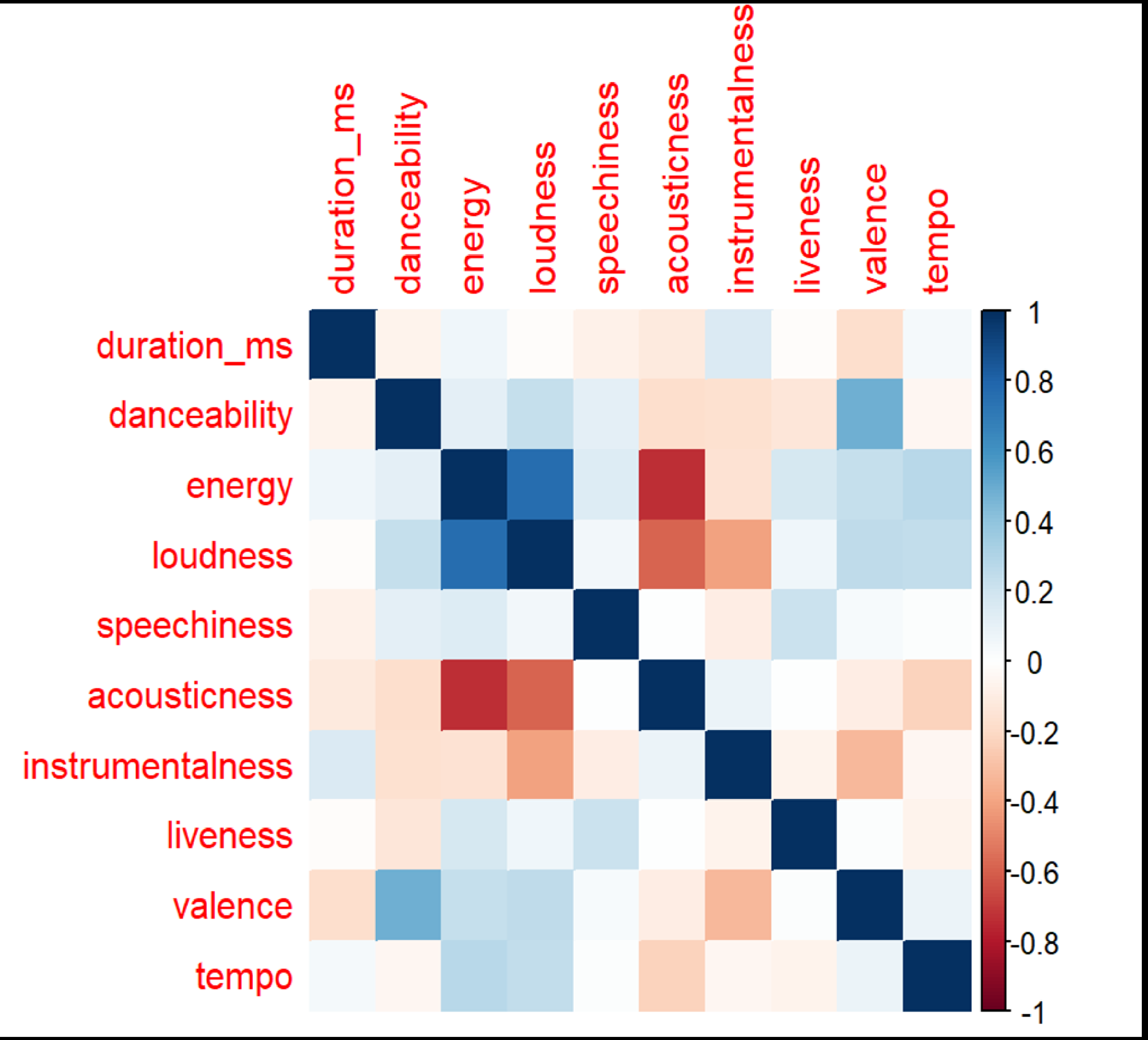
Goal of our Analysis

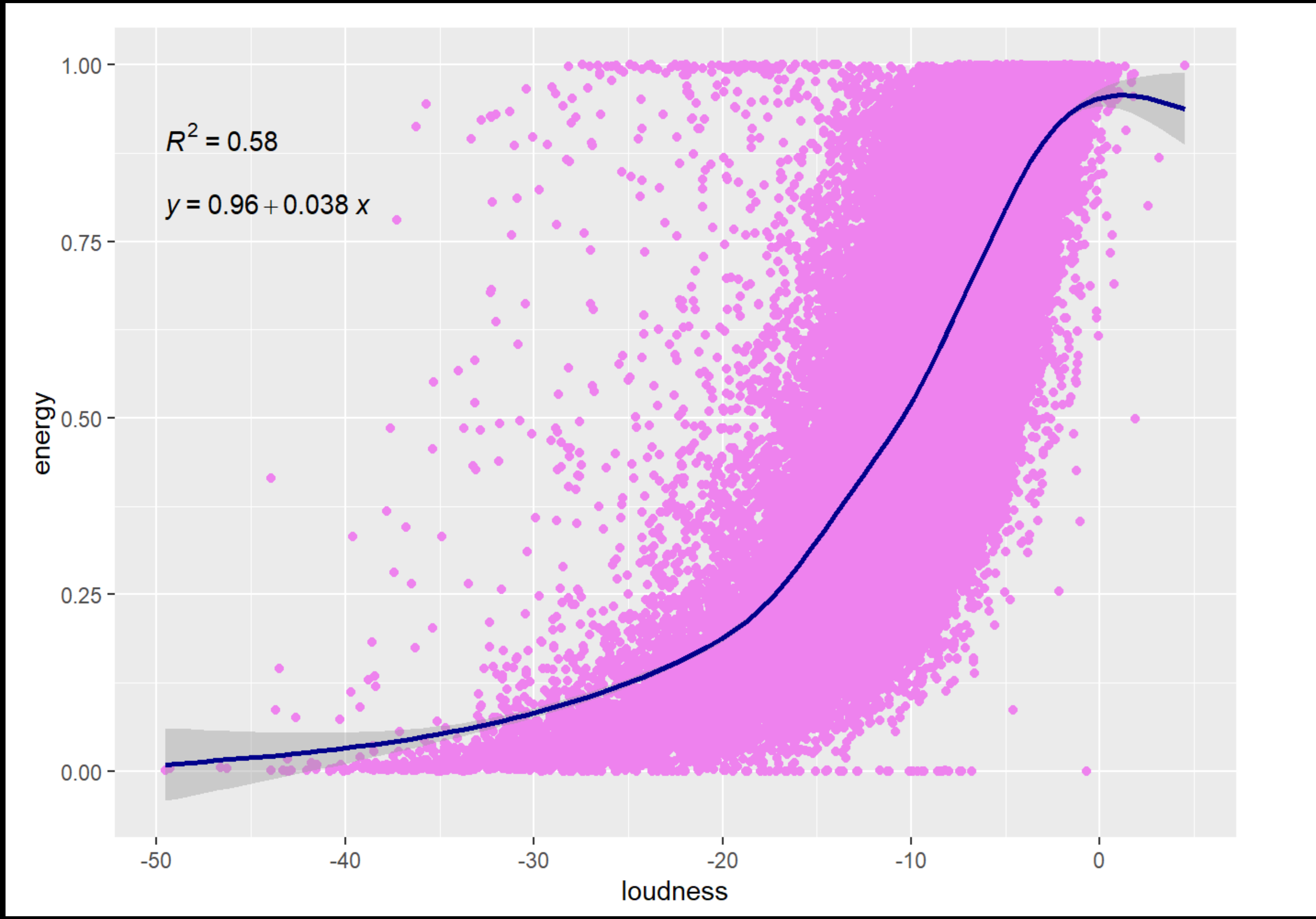
Grouping similar songs of Spotify together based on their audio features and identify the characteristics of the clusters, which can be used to understand user preferences.

Applications in real world:

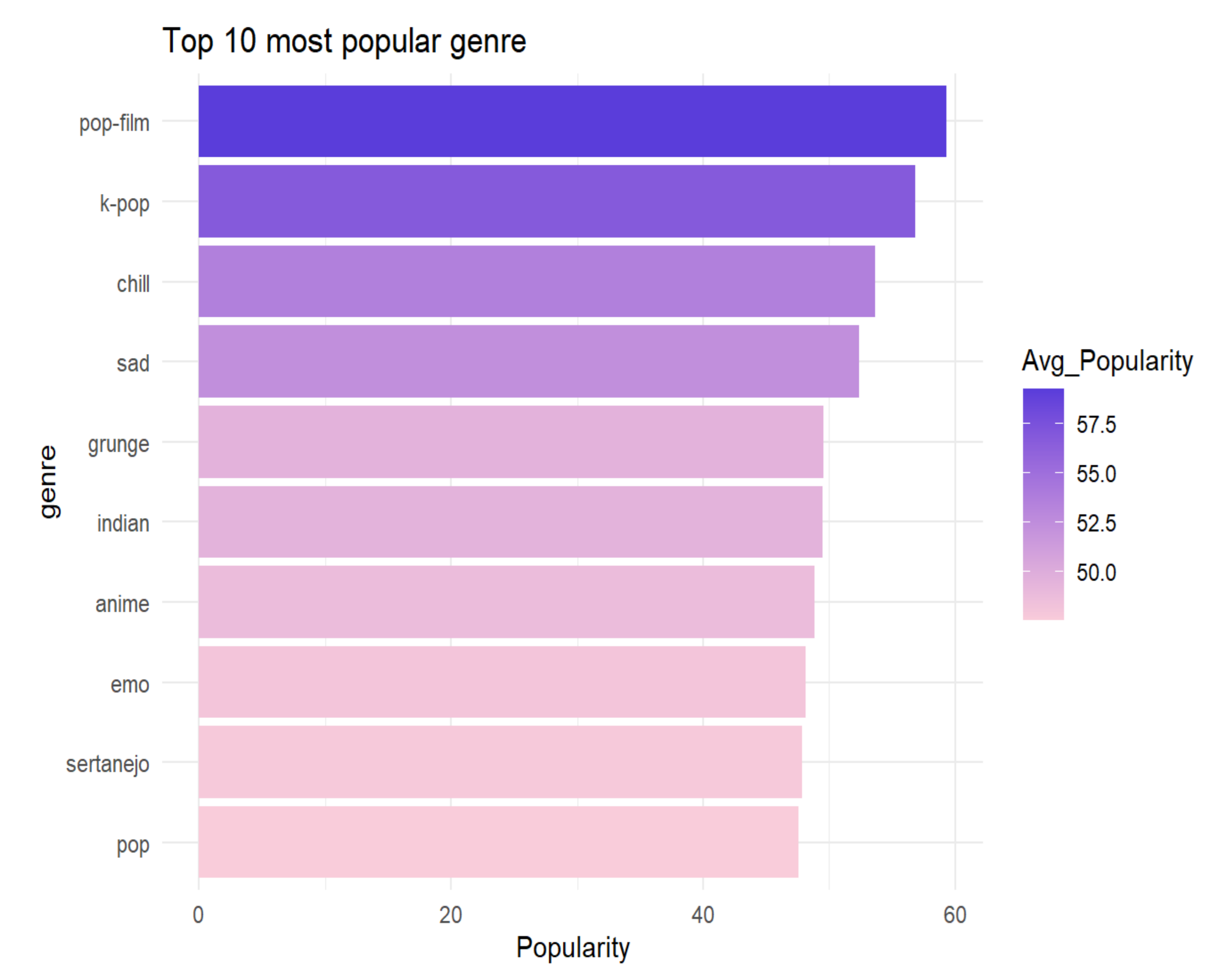
- Create playlists with similar music
- Personalized music recommendations to users
- Gain insights on popular songs and their characteristics

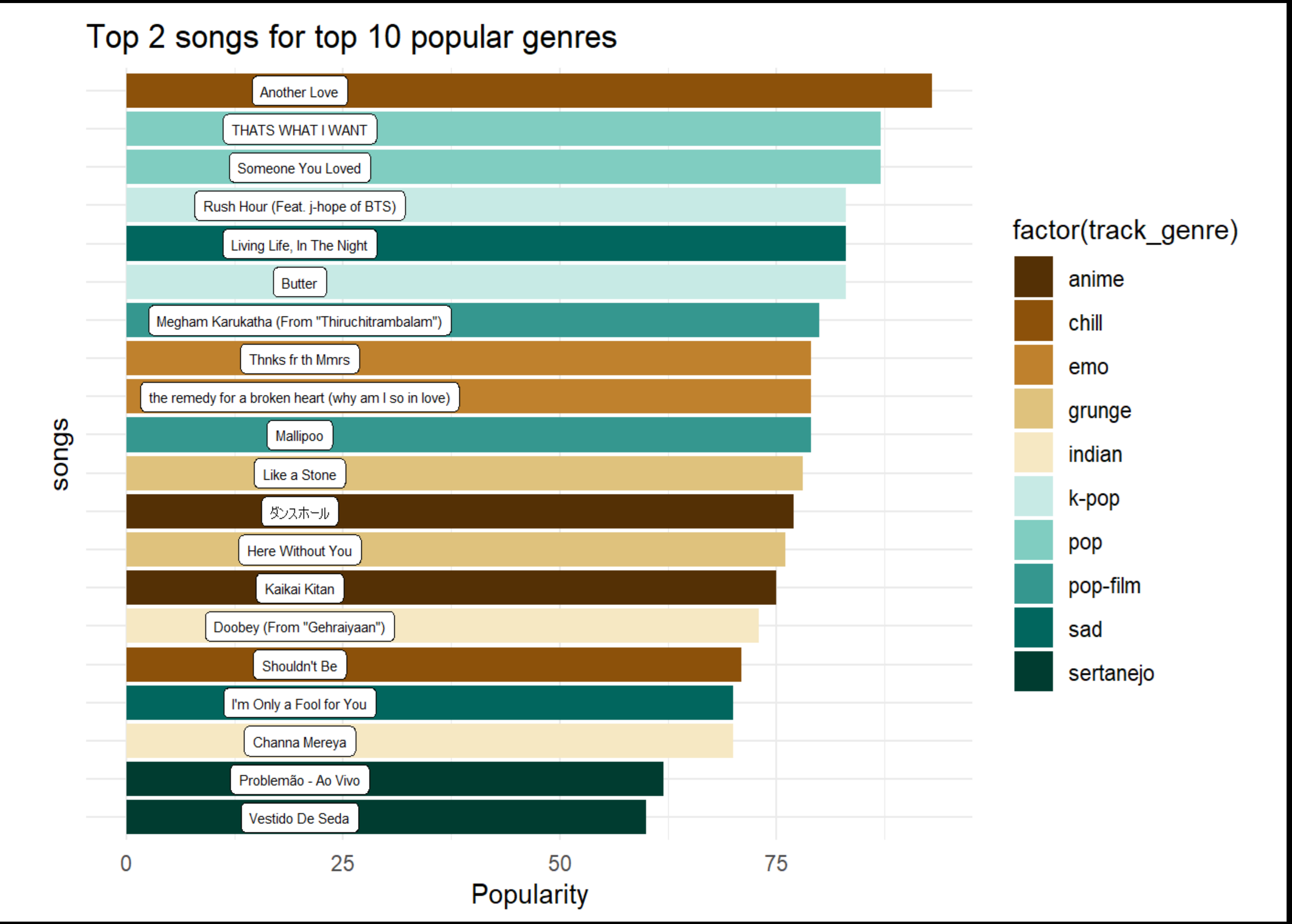
Exploratory Data Analysis



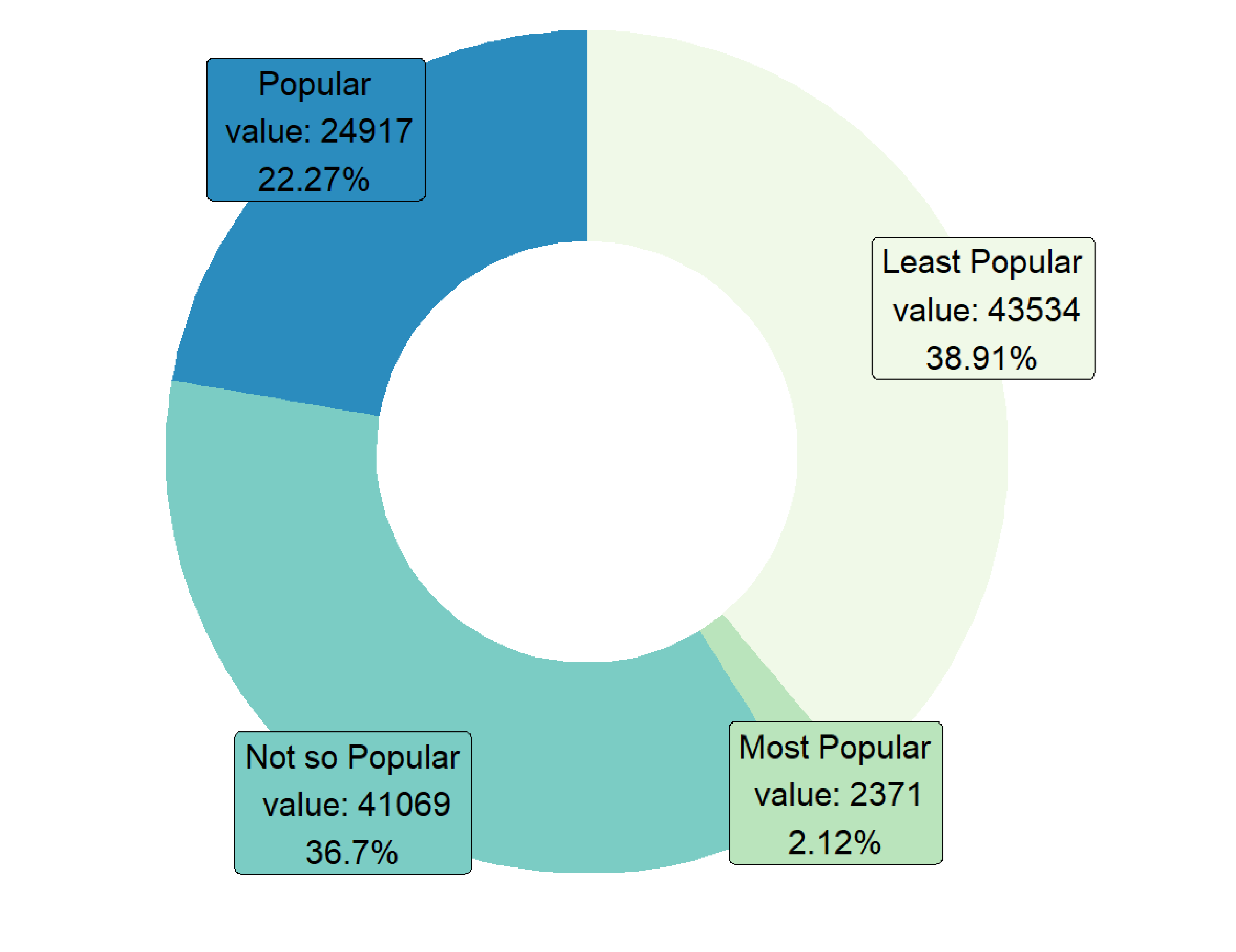


Exploratory Data Analysis





Exploratory Data Analysis



Feed-Forward Neural Network

Song Elements:
- duration_ms
- time_signature
- key
- mode

Predicting Song Characteristics:
- Speechness
- acousticness
- instrumentalness
- liveness
- valence
- danceability
- energy
- loudness
- tempo

Feed-Forward Neural Network

- Sampled: training 70% & Testing 30% data
- Input & hidden layer - ReLU activation function
- Output layer - Linear activation function

```r
# Define MLP model architecture
model=keras_model_sequential()
model %>%
  layer_dense(units=64, activation='relu', input_shape=ncol(x_train)) %>%
  layer_dropout(rate = 0.4) %>%
  layer_dense(units = 32, activation = 'relu') %>%
  layer_dropout(rate = 0.4) %>%
  layer_dense(units =8, activation='linear')
```

```
Model: "sequential"
________________________________________________________________________________
 Layer (type)                        Output Shape                    Param #
================================================================================
 dense_2 (Dense)                     (None, 64)                      384
 dropout_1 (Dropout)                 (None, 64)                      0
 dense_1 (Dense)                     (None, 32)                      2080
 dropout (Dropout)                   (None, 32)                      0
 dense (Dense)                       (None, 8)                       264
================================================================================
Total params: 2,728
Trainable params: 2,728
Non-trainable params: 0
```

Model & RESULT

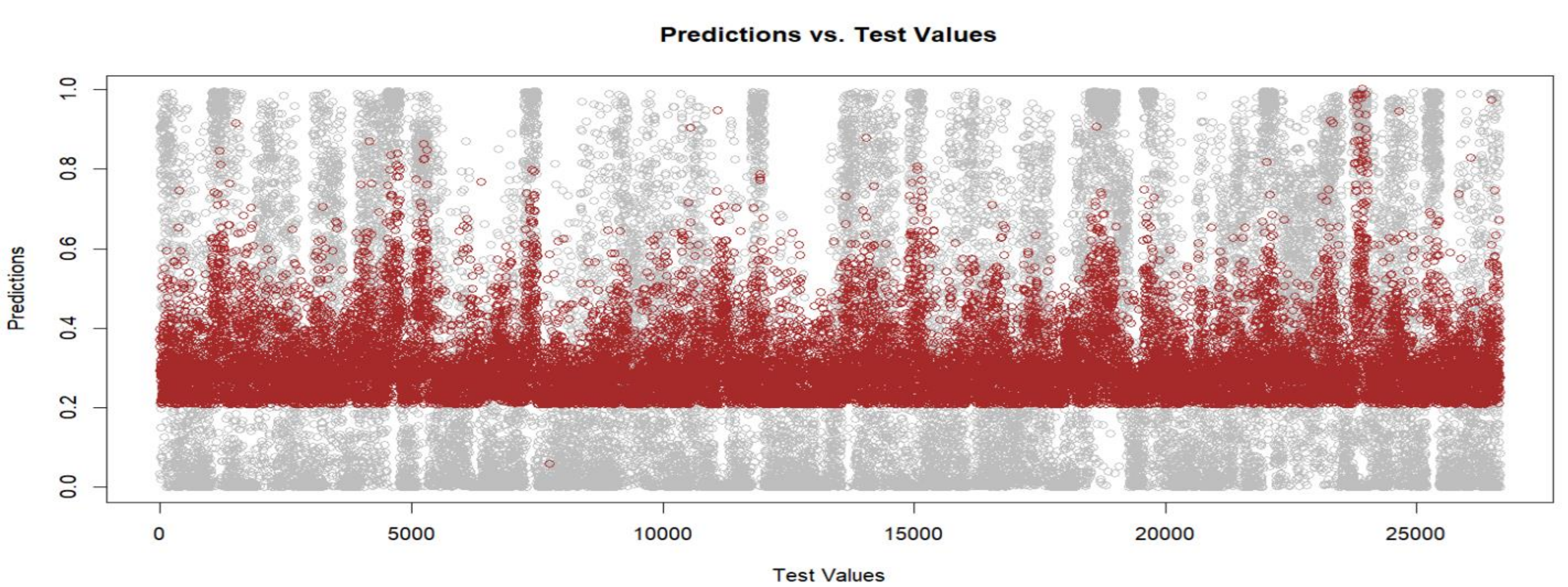
- Loss Function - MSE
- Optimizer - rmsprop
- metric - mae

```
# Compile model
model %>%
  compile(loss = 'mse',
          optimizer = 'rmsprop',
          metrics=c('accuracy', 'mae')
          )

# Train model
history<-model %>%
  fit(x_train, y_train, epochs = 50, batch_size = 64,verbose=1,validation_split = 0.2)
```

Output:
Metric of interest: mae
mae=0.578

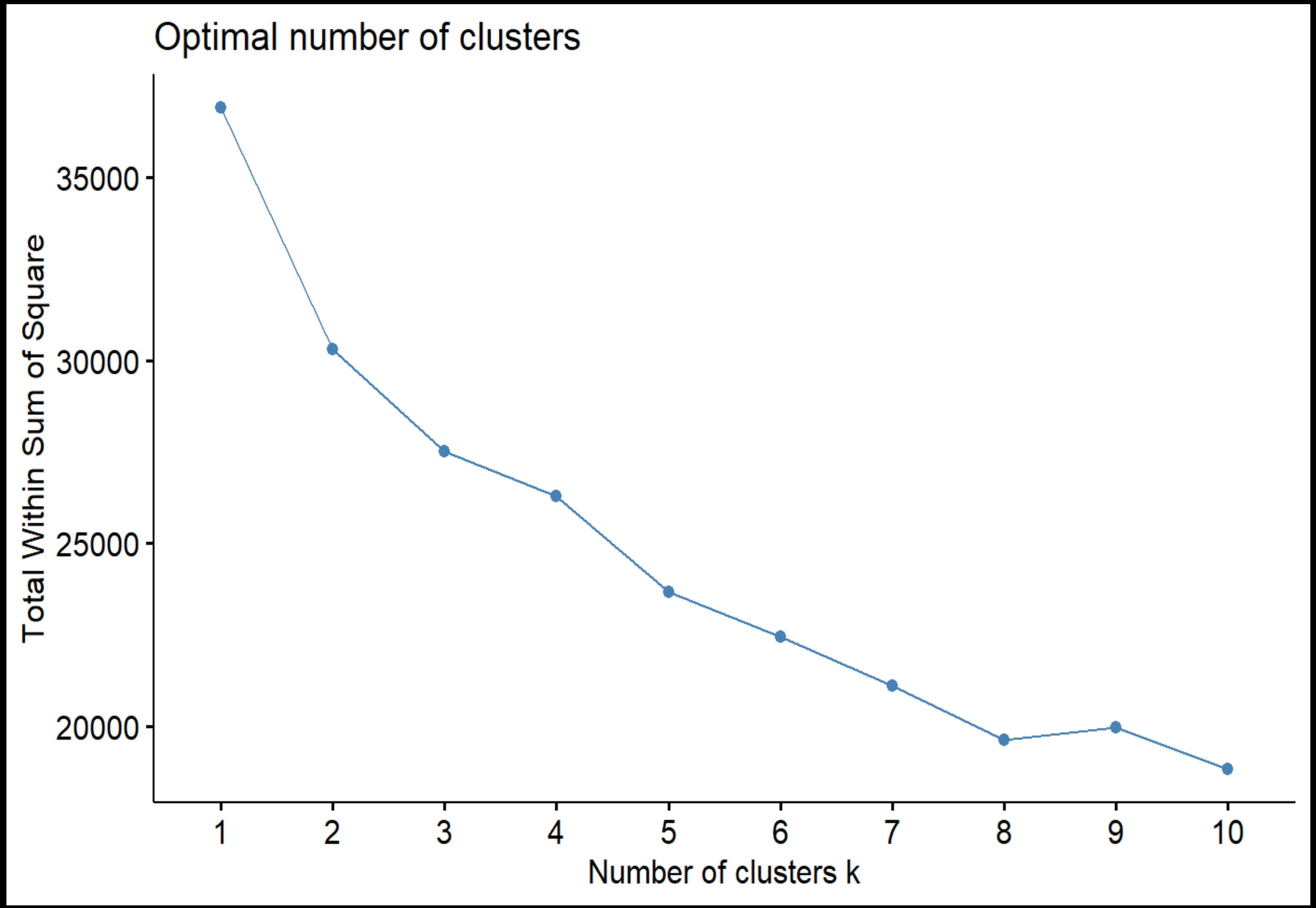
```
      mae
0.5784451
```



- Acousticness of Prediction Vs Test

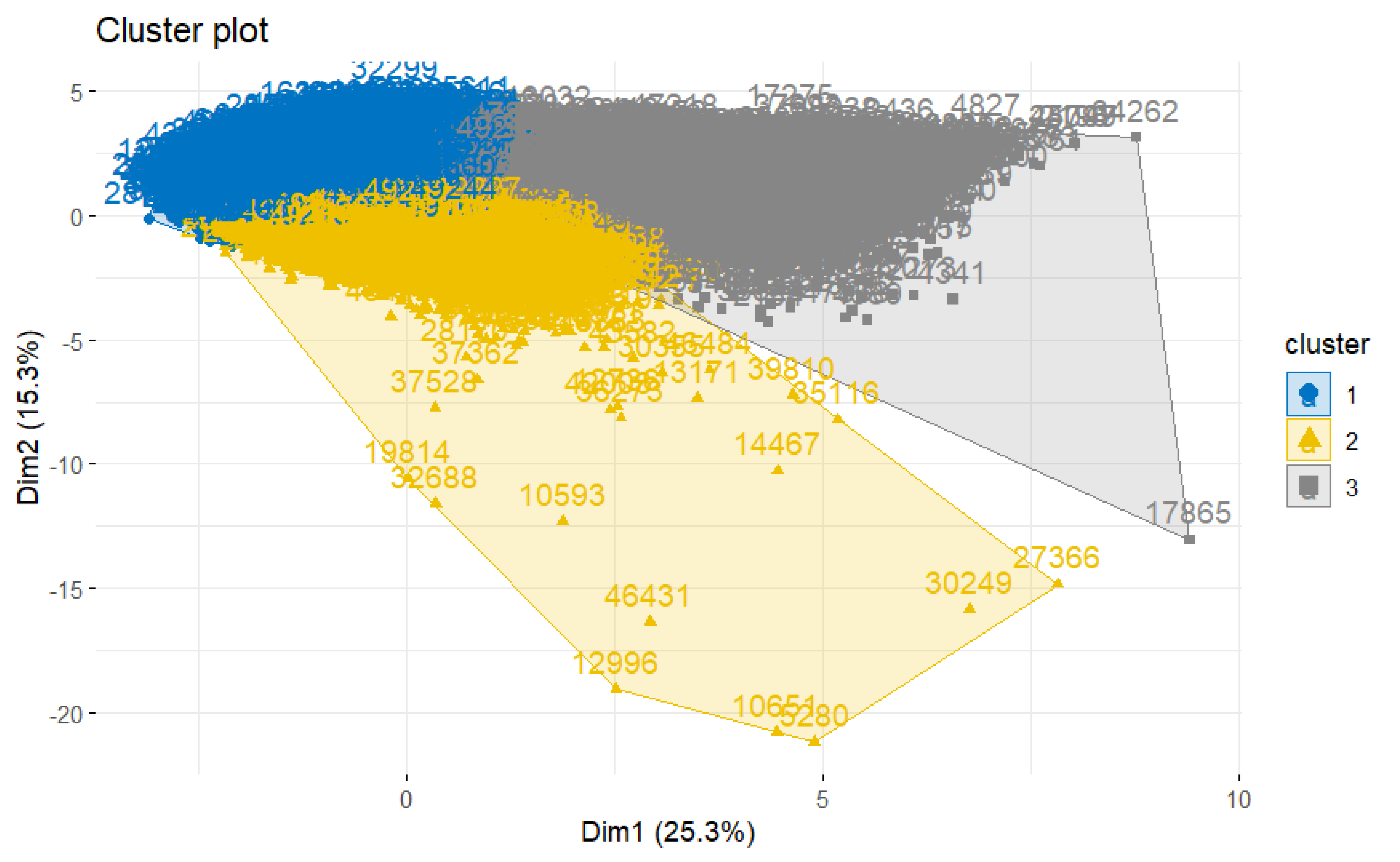
K-means Clustering Method

- K means(Euclidean distance, centroids)
- Optimal number of clusters(Elbow method)
- #clusters selected, k=3



K-means Clustering

Clusters	Cluster 1	Cluster 2	Cluster 3
Characteristics	Danceability, Loudness	Liveness, Loudness, High Tempo	Instrumentalness, Acousticness



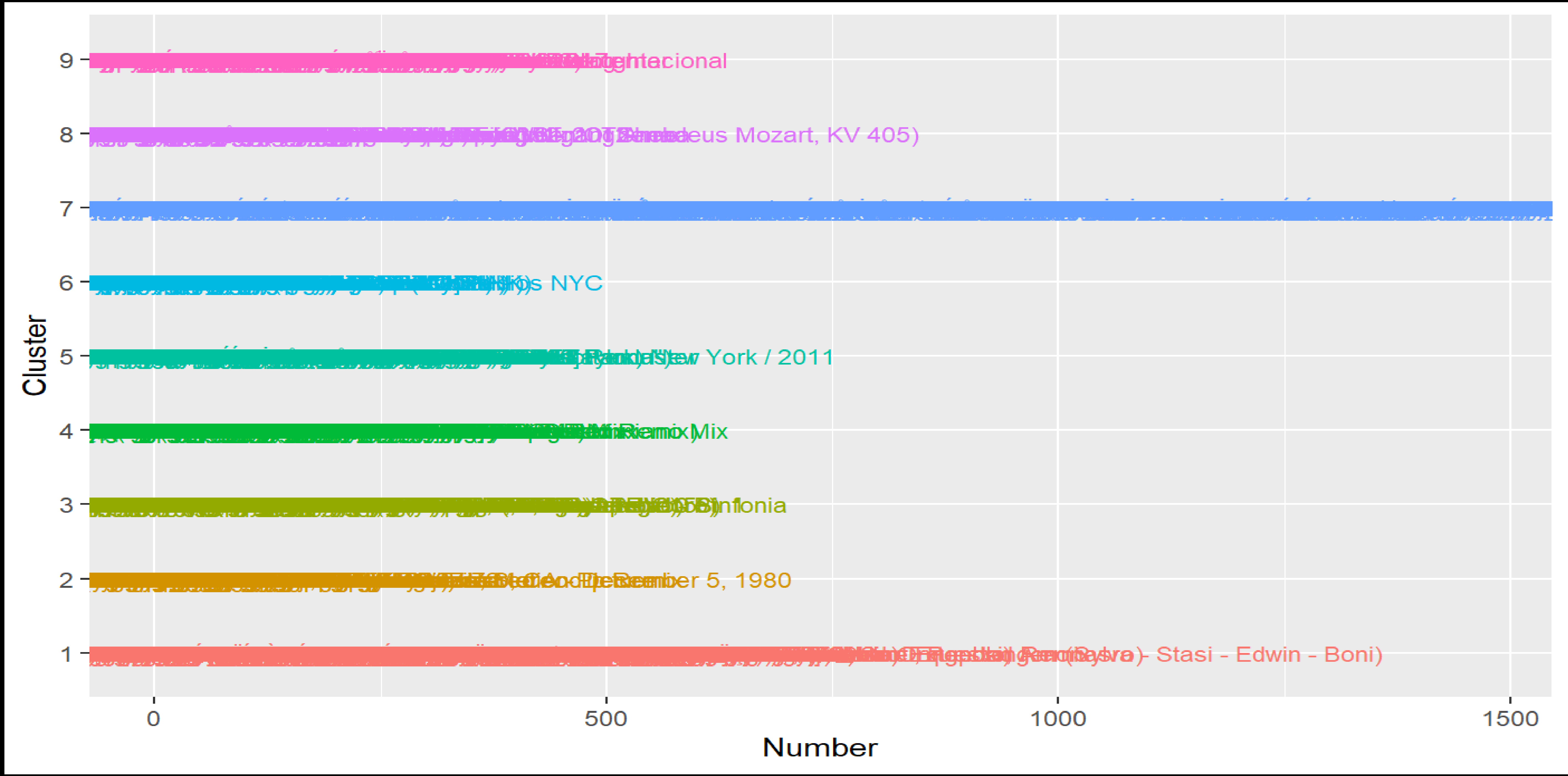
	artists	track_name	cluster
1	Justin Bieber	Mistletoe	1
2	Omar S;L'Renee	S.E.X - C.G.P (Conant Gardens Posse) Remix	1
3	STU48	夏の"好き"はご用心	1
4	Keenan Te	Halfway There	1
5	BBS Paranoicos	La Rabia	2
6	Stabil	Kovala - Orijinal Film Müziği	1
7	Tale Of Us;Mind Against	Astral	2
8	Weston Estate	Silence	2
9	Bob Marley & The Wailers	Jamming	1
10	Rex Williams	You Are My Heart	1
11	MC Arraia;DJ Guih Da ZO	Tu Só Foi Mais Uma	1
12	Maneva	Vem Ver (Ao Vivo)	1
13	Vanna	Toxic Pretender	2
14	Alka Yagnik;Hariharan	Bahon Ke Darmiyan - Khamoshi - The Musical / Soundtrack ...	2
15	Melorman	Love in the 90's	3
16	Front Line Assembly	Killing Ground	

Mixture Models

- Gaussian Mixture Models
- Used Mclust function
- Mclust has chosen the number of optimal clusters as 9 using BIC criteria

Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8	Cluster 9
0.206305	0.035373	0.0461497	0.076210	0.083821	0.033110	0.392357	0.038849	0.080.087821
Valence, Loudness	Acousticness, Instrumentalness, Liveness	Acousticness, Instrumentalness	Danceability	Loudness, Speechiness	Danceability, Valence	Danceability, energy, loudness	Loudness, Speechiness	Instrumentalness, Loudness

Mixture Models



Conclusion

CLUSTERS	Characteristics	Songs
Cluster 1	Danceability, Loudness	Attention
Cluster 2	Liveness, Loudness, High Tempo	Castle of Glass
Cluster 3	Instrumentalness, Acousticness	Kun Faya Kun

Limitations and Future Direction

- It’s difficult to interpret the clusters through mixture models.
- Can build music recommendation system for personalized playlists to users based on the clusters of the songs

THAK YOU

Questions?