**1 Import reduced dimensions data and required libraries**

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 1600 entries, 0 to 1599

Data columns (total 9 columns):

# Column Non-Null Count Dtype

--- ------ -------------- -----

0 PC\_1 1600 non-null float64

1 PC\_2 1600 non-null float64

2 PC\_3 1600 non-null float64

3 PC\_4 1600 non-null float64

4 PC\_5 1600 non-null float64

5 PC\_6 1600 non-null float64

6 PC\_7 1600 non-null float64

7 PC\_8 1600 non-null float64

8 PC\_9 1600 non-null float64

dtypes: float64(9)

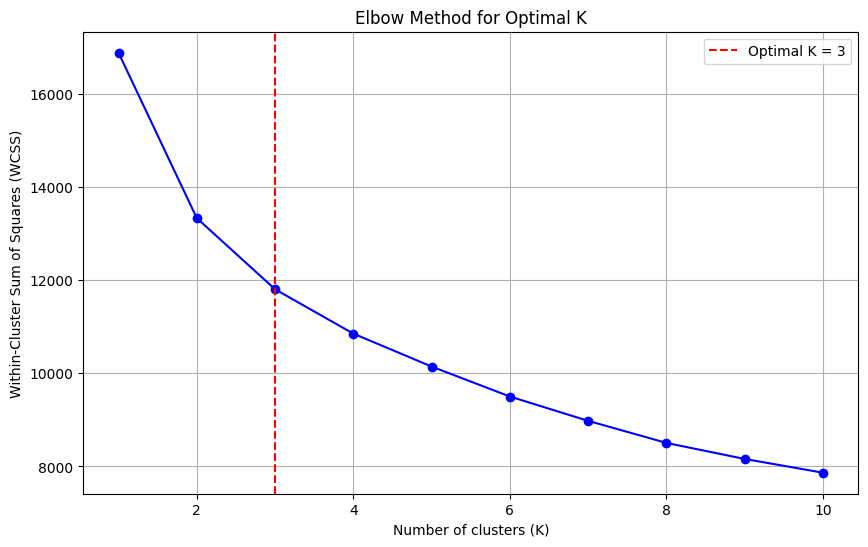
memory usage: 112.6 KB

None

# 2. Determine optimal number of clusters

Optimal number of clusters: 3

WCSS at elbow point: 11800.524424263393



# 3. Apply clustering methods: K-Means, Hierarchical Clustering, GMM.

\*\*\*\*\*\*KMeans clustering \*\*\*\*\*\*\*\*\*\*\*\*\*\*

Best number of clusters by silhouette (2..10) = 2 score= 0.19538401053708315

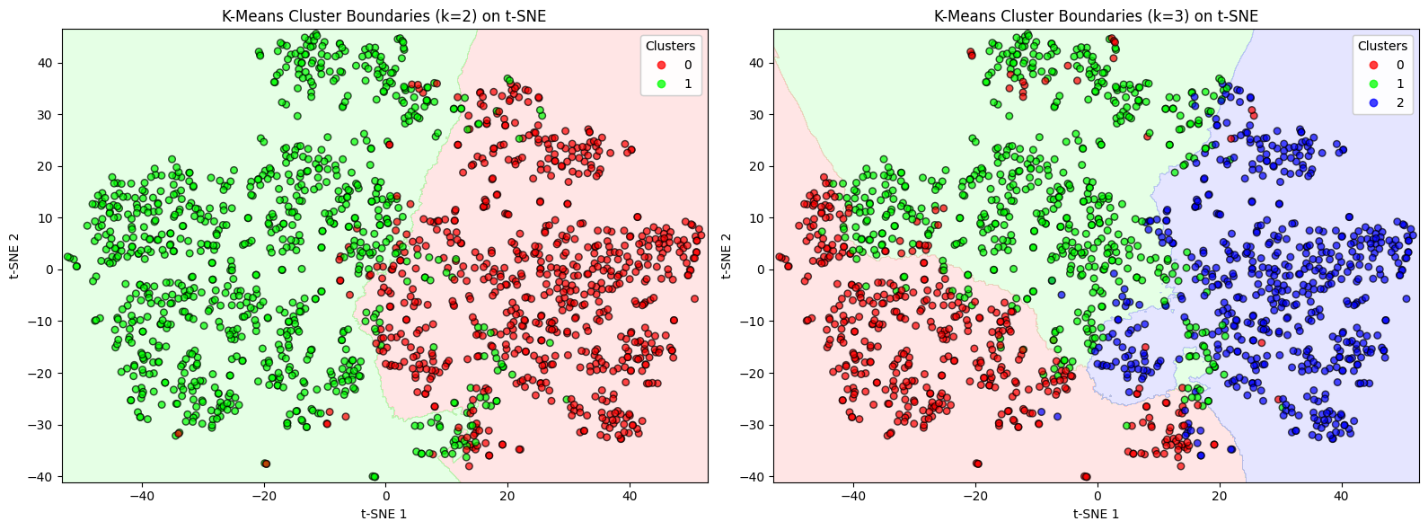
\*\*\*\*\*\*Agglomerative clustering \*\*\*\*\*\*\*\*\*\*\*\*\*\*

Best number of clusters by silhouette (2..10) = 2 score= 0.15490859063471274

\*\*\*\*\*\*GaussianMixture clustering \*\*\*\*\*\*\*\*\*\*\*\*\*\*

Best number of clusters by silhouette (2..10) = 2 score= 0.17981446259092018

# 4. Visualization of the clusters



# 5. Data split based on the clustering result

\*\*\*\*\*\*\*\*\*\*\*\*\*First group:\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

number\_row name album release\_date \

0 0 Concert Intro Music - Live Licked Live In NYC 2022-06-10

1 1 Street Fighting Man - Live Licked Live In NYC 2022-06-10

2 2 Start Me Up - Live Licked Live In NYC 2022-06-10

3 3 If You Can't Rock Me - Live Licked Live In NYC 2022-06-10

4 4 Don’t Stop - Live Licked Live In NYC 2022-06-10

track\_number id uri \

0 1 2IEkywLJ4ykbhi1yRQvmsT spotify:track:2IEkywLJ4ykbhi1yRQvmsT

1 2 6GVgVJBKkGJoRfarYRvGTU spotify:track:6GVgVJBKkGJoRfarYRvGTU

2 3 1Lu761pZ0dBTGpzxaQoZNW spotify:track:1Lu761pZ0dBTGpzxaQoZNW

3 4 1agTQzOTUnGNggyckEqiDH spotify:track:1agTQzOTUnGNggyckEqiDH

4 5 7piGJR8YndQBQWVXv6KtQw spotify:track:7piGJR8YndQBQWVXv6KtQw

acousticness danceability energy instrumentalness liveness loudness \

0 0.0824 0.463 0.993 0.996000 0.932 -12.913

1 0.4370 0.326 0.965 0.233000 0.961 -4.803

2 0.4160 0.386 0.969 0.400000 0.956 -4.936

3 0.5670 0.369 0.985 0.000107 0.895 -5.535

4 0.4000 0.303 0.969 0.055900 0.966 -5.098

speechiness tempo valence popularity duration\_ms

0 0.1100 118.001 0.0302 33 48640

1 0.0759 131.455 0.3180 34 253173

2 0.1150 130.066 0.3130 34 263160

3 0.1930 132.994 0.1470 32 305880

4 0.0930 130.533 0.2060 32 305106

\*\*\*\*\*\*\*\*\*\*\*\*\*Second group:\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

number\_row name \

23 23 Honky Tonk Women - Live At The El Mocambo 1977

27 27 Fool To Cry - Live At The El Mocambo 1977

30 30 Crackin’ Up - Live At The El Mocambo 1977

32 32 Around And Around - Live At The El Mocambo 1977

33 33 Tumbling Dice - Live At The El Mocambo 1977

album release\_date track\_number id \

23 Live At The El Mocambo 2022-05-13 1 7BPKw4y2CX3waC8IQnQDXB

27 Live At The El Mocambo 2022-05-13 5 4ORw1k0PzLvF4F8BybMLma

30 Live At The El Mocambo 2022-05-13 8 1TZos9kygOjAaHw8398LZi

32 Live At The El Mocambo 2022-05-13 10 3obSrugIHUiznbg0NkEBJg

33 Live At The El Mocambo 2022-05-13 11 3MnDqaLZto9mE5GKbLaphc

uri acousticness danceability energy \

23 spotify:track:7BPKw4y2CX3waC8IQnQDXB 0.1620 0.430 0.901

27 spotify:track:4ORw1k0PzLvF4F8BybMLma 0.4340 0.331 0.692

30 spotify:track:1TZos9kygOjAaHw8398LZi 0.3080 0.594 0.801

32 spotify:track:3obSrugIHUiznbg0NkEBJg 0.0645 0.429 0.869

33 spotify:track:3MnDqaLZto9mE5GKbLaphc 0.1330 0.450 0.842

instrumentalness liveness loudness speechiness tempo valence \

23 0.01100 0.694 -6.215 0.0450 112.077 0.890

27 0.01430 0.889 -7.105 0.0450 76.210 0.512

30 0.10300 0.699 -7.448 0.0277 105.572 0.925

32 0.00543 0.848 -7.467 0.0420 97.223 0.790

33 0.03580 0.758 -7.254 0.0343 107.825 0.768

popularity duration\_ms

23 32 216546

27 29 297386

30 27 254506

32 27 233200

33 32 296226