Genre-based Playlist Generation using Supervised and Unsupervised Learning

Sonali Johari, Pranav Prajapati
Instructor: Amir H. Gandomi

STEVENS
INSTITUTE of TECHNOLOGY
THE INNOVATION UNIVERSITY®
1870

Business Intelligence & Analytics

Motivation

We often come across recommendations based on our recent activity on the streaming platform. Our favorite artists release music albums generally once every 2 years. Therefore, such recommendation systems are suboptimal for the best music experience. Our project aims to primarily provide listeners with playlist recommendations based on their favorite artists

Data

Scrapped 10,000 songs with more than 25 features using Spotify Web API.

Key Methods

Web Scraping, PCA, KNN Classification, K-medoids and K-means Clustering

Data and Feature Extraction

• Assigned binary classes to songs. 1 for the favorite artist and 0 for the songs to be classified.
• However to avoid overfitting some songs were assigned class 1 by matching some average feature scores of the favorite artists.

Dimension Reduction and Pre-processing in Supervised Learning
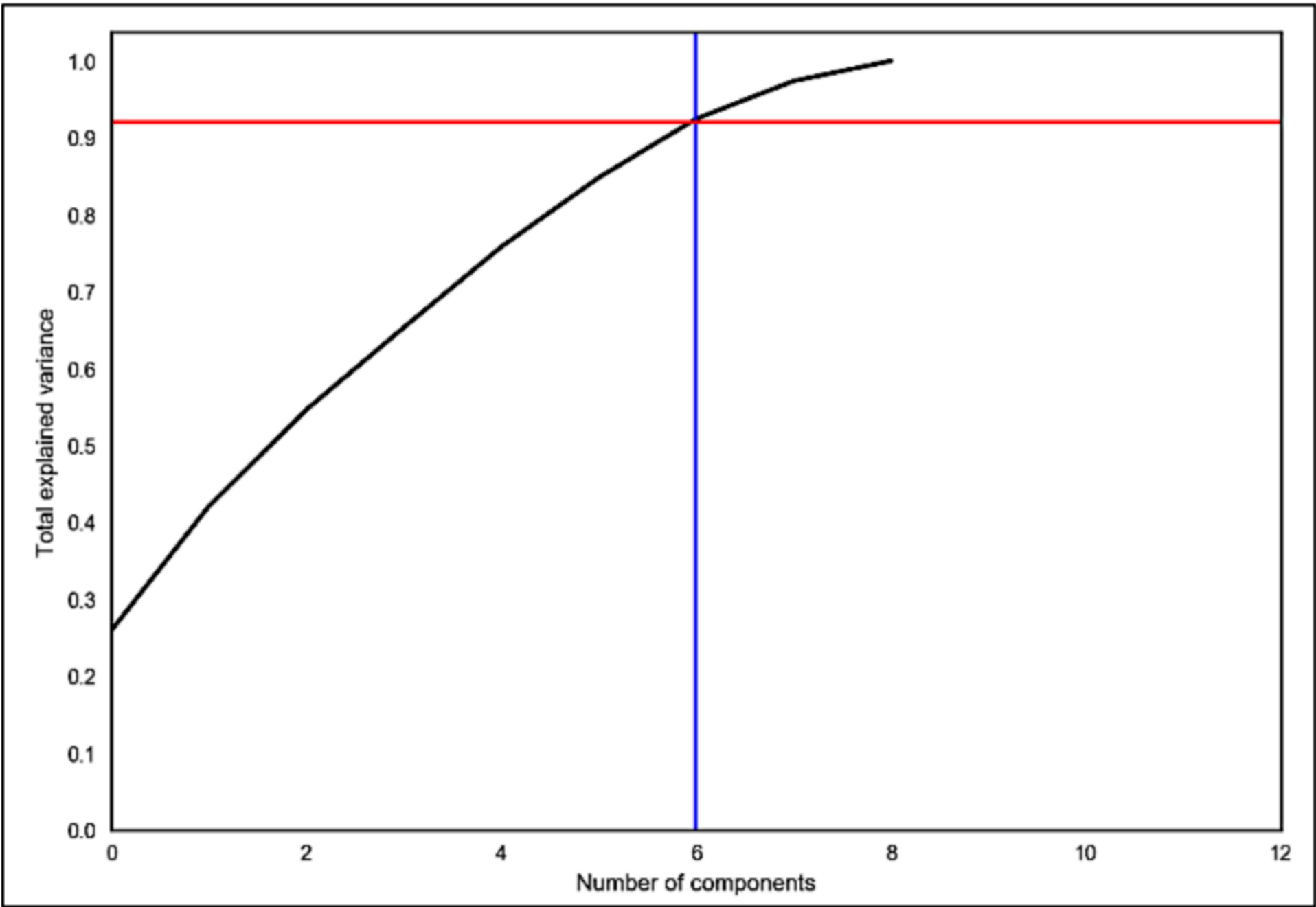
• Dropped unimportant columns from the like "album name", reducing the number of total features to 8.
• Performed Dimension reduction using PCA.
• Plotting the PCA generates the adjacent graph that shows a 92% variance for 6 components.

Performance

• Used GridSearchCV, which performs an exhaustive search over specified parameter values for an estimator.
• All the possible combinations of parameter values are evaluated and the best combination is retained.
• We can see the best score is obtained for KNN classification at 80%, with Decision Trees ranking 2nd at 74%.

Algorithm	GridSearchCV Score
KNN	0.8032786885245902
Random Forest	0.7180327868852459
Decision Trees	0.740983606557377

Classification Results

• The results of precision and recall can be interpreted from the adjacent image.
• There is a high precision in our model to return only actual instances of unsimilar songs, and a high recall which identifies relevant instances correctly for similar songs.

precision: [0.81818182 0.36363636]
recall: [0.26865672 0.875   ]

name	Unnamed: 0	popularity	danceability	acousticness	energy	instrumentalness	liveness	loudness	speechiness	tempo	class	new_class
Sunflower - Spider-Man: Into the Spider-Verse	0	92	0.757	0.5410	0.501	0.000000	0.0718	-5.593	0.0466	89.931	0	0
Lucid Dreams	3	96	0.511	0.3490	0.566	0.000000	0.3400	-7.230	0.2000	83.903	0	0
Shallow	8	91	0.572	0.3710	0.385	0.000000	0.2310	-6.362	0.0308	95.799	0	0
when the party's over	9	89	0.498	0.9790	0.104	0.000079	0.0895	-14.080	0.0621	124.001	0	0
Mo Bamba	15	91	0.729	0.1940	0.625	0.009860	0.2480	-5.266	0.0315	146.034	0	0
Falling Down	25	96	0.669	0.0175	0.574	0.002940	0.1460	-6.442	0.0286	120.013	0	0
Noticed	27	80	0.702	0.4560	0.571	0.000000	0.1160	-7.540	0.0536	80.012	0	0
changes	41	91	0.669	0.8830	0.308	0.000000	0.0984	-10.068	0.0290	64.934	1	0
SLOW DANCING IN THE DARK	47	66	0.515	0.5440	0.479	0.005980	0.1910	-7.458	0.0261	88.964	0	0
Trip	51	84	0.477	0.2250	0.610	0.000000	0.1070	-5.628	0.1440	79.882	1	0
I Fall Apart	52	88	0.556	0.0689	0.538	0.000000	0.1960	-5.408	0.0382	143.950	1	0
Beautiful (feat. Camila Cabello)	59	93	0.638	0.3460	0.717	0.000000	0.1050	-4.722	0.0337	100.027	0	0

Clustering of songs

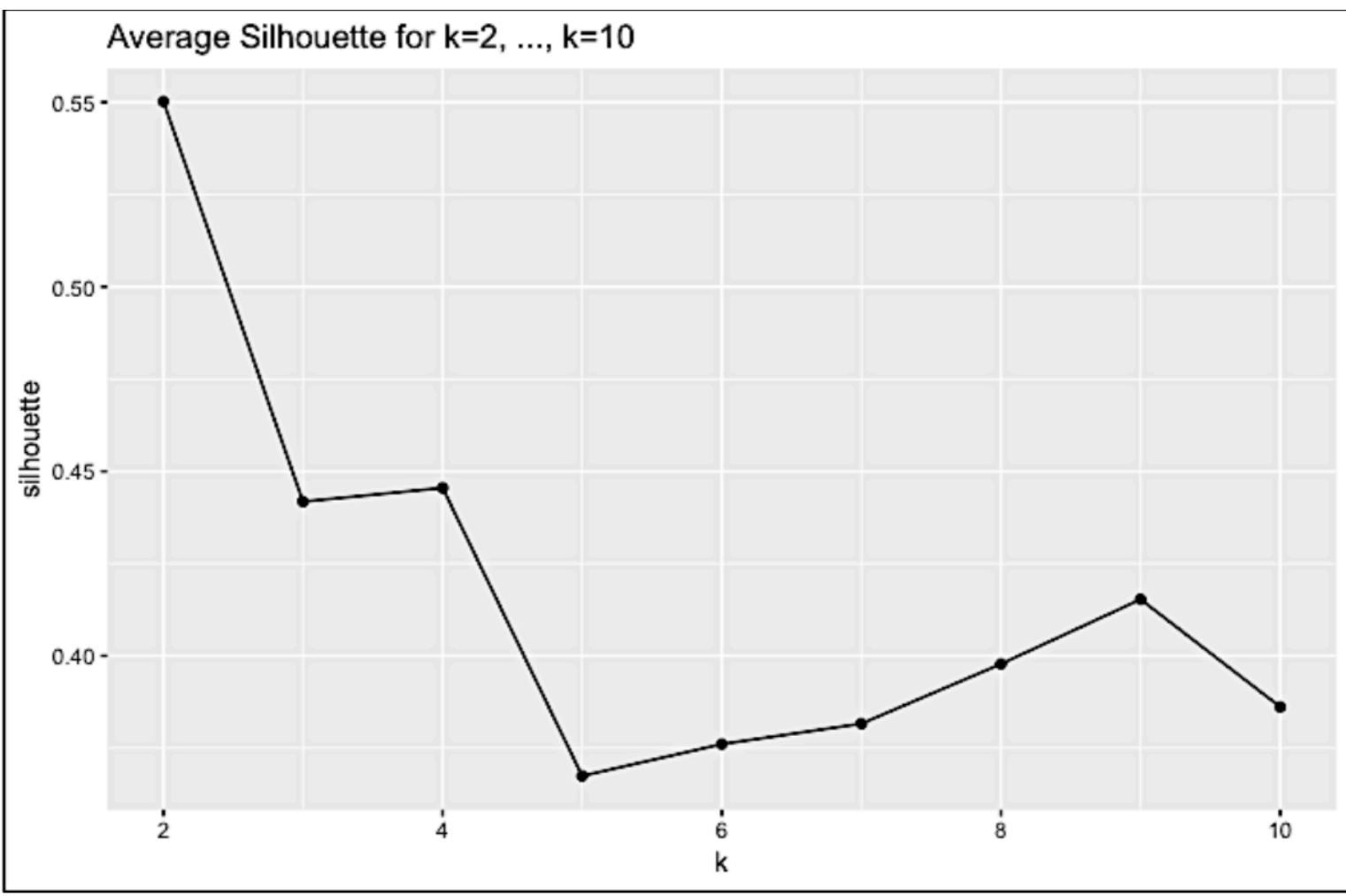### K- Means

Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4

• Summarized the song clusters by generating a radar chart
• Gives an understanding of which cluster belongs to which genre.
• One cluster is not defined correctly
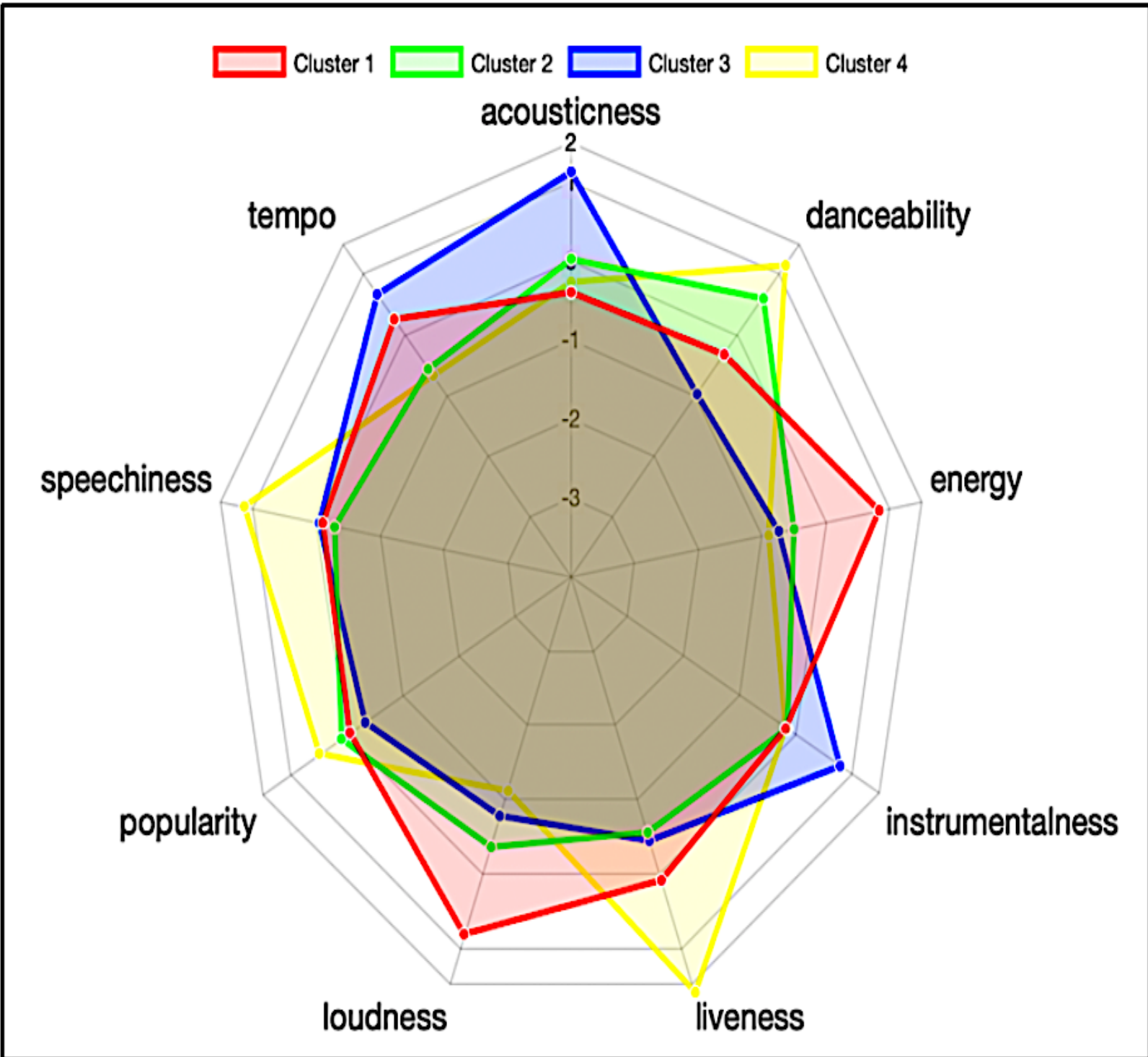
K- Medoids

Silhouette Plot

Average Silhouette for k=2, ..., k=10

• Evaluated the silhouette coefficient, a measure of how similar a data observation is to those on its own clusters compared with the observations in other clusters.
• +1 indicates good cohesion, -1 indicates object is in the incorrect cluster.
• Good cohesion can be seen for k=2 and k=3.
• Since we wanted to obtain playlists of different genres, k=3 was chosen.


Silhouette plot of the cluster
n = 77

3 clusters C_j
j : n_j | $ave_{i \in C_j} s_i$

1 : 24 | 0.46

2 : 24 | 0.55

3 : 29 | 0.34

Silhouette width s_i

Average silhouette width : 0.44

• The adjacent figure is of the silhouette plot of the clusters.
• The numbers to the right represent the average silhouette of each cluster.
• Cluster 2 is the best one with an average silhouette of 0.55.
• The negative silhouette points represent the songs that are probably assigned to the wrong cluster.
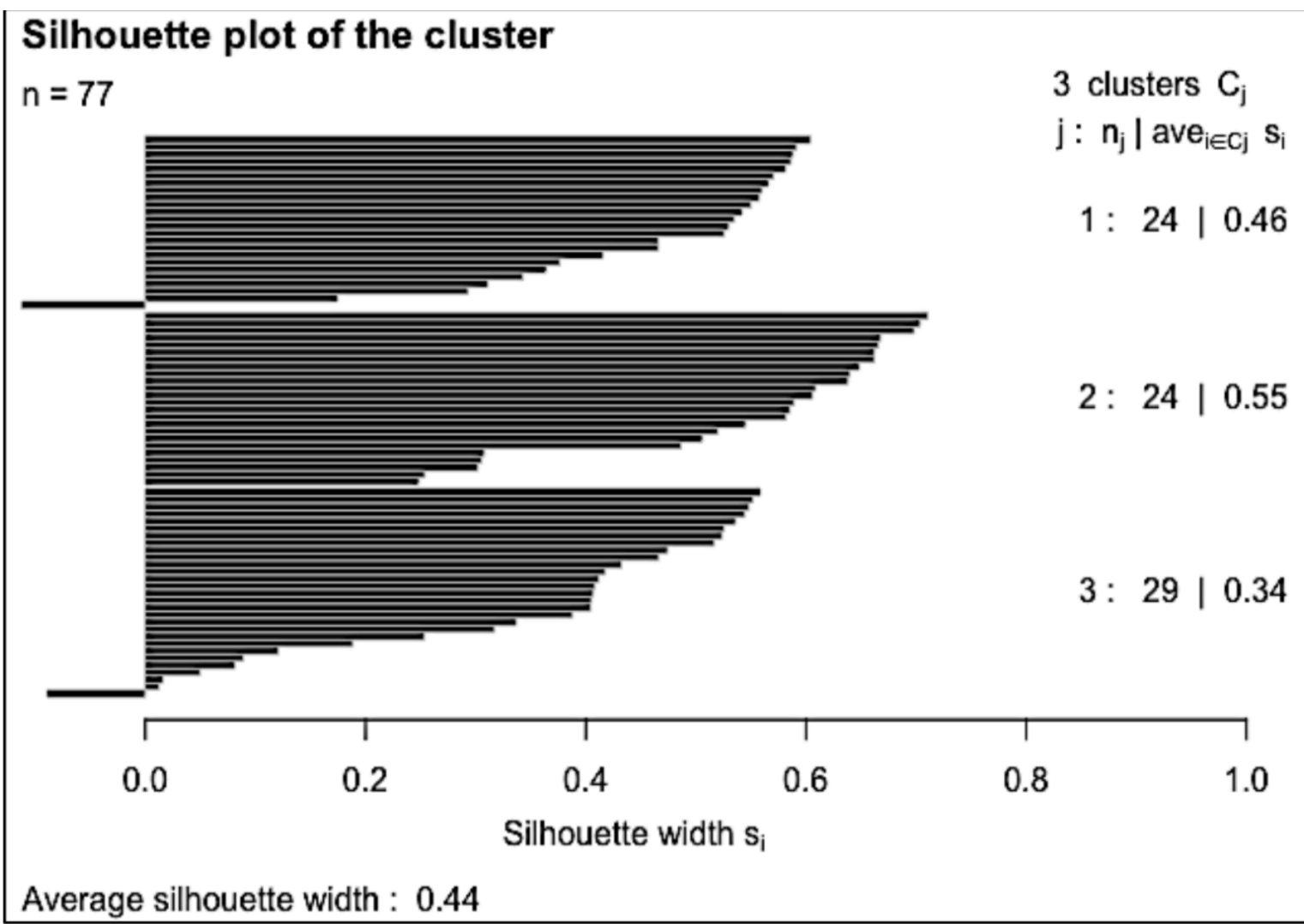
Pre-processing in Unsupervised Learning

• Although K-medoid was successful in generating clusters, there were a few outliers yet to be removed
• A 3-D representation using T-SNE on the main cluster gives a better understanding of the outliers in our clusters

3D Representation of the Main Cluster using t-SNE
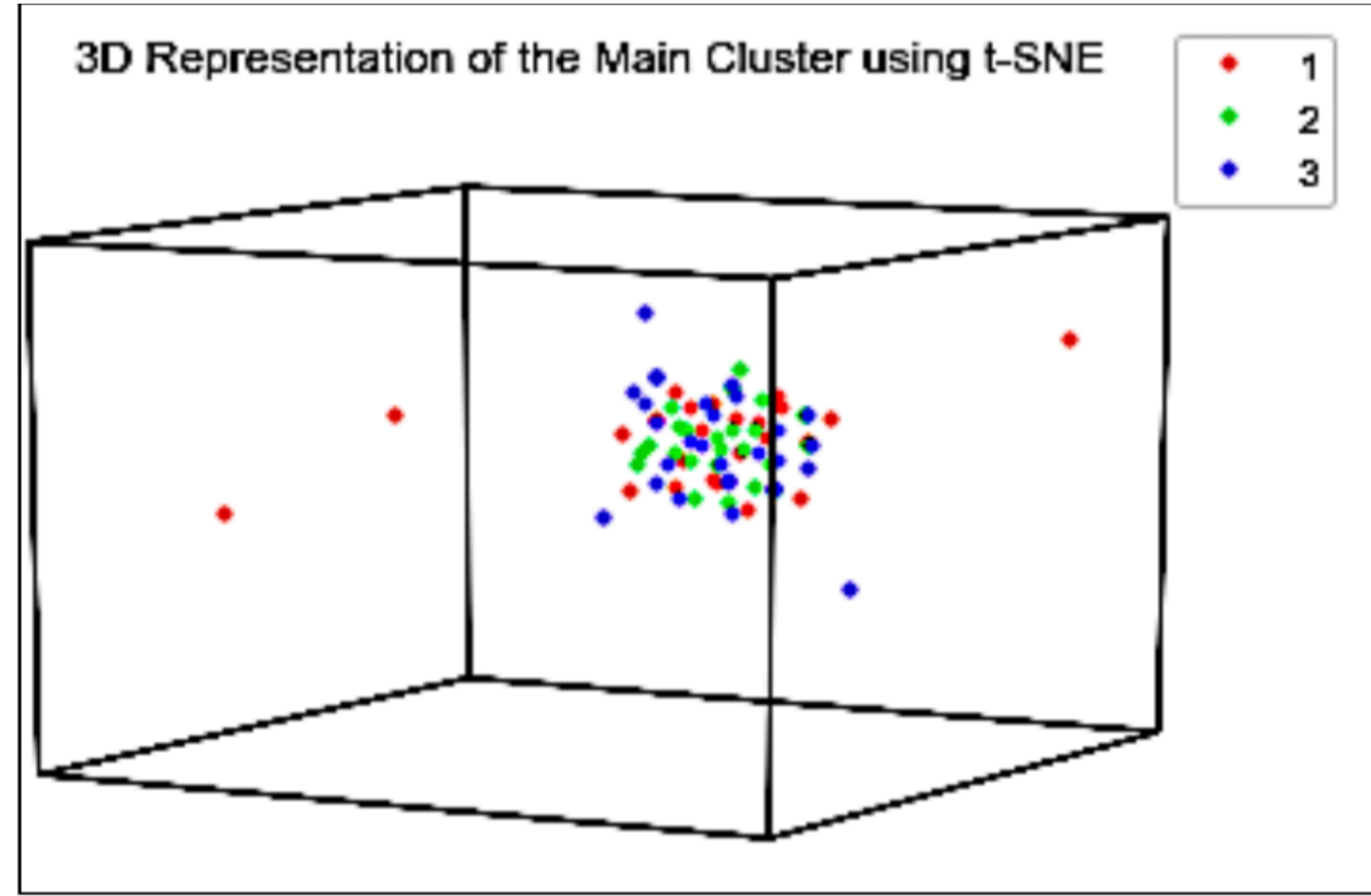
Results

From the first playlist 3 songs - Happier, Taki Taki, Eastside were the outliers and God is a woman, Woman Like me were the outliers from cluster 3. However, a close look at cluster 2 gave us one song – no tears left to cry as an outlier although Cluster 2 was perfect.

HIPHOP	RNB	TRAP
name	name	name
Drip Too Hard (Lil Baby & Gunna)	Without Me	Nonstop
Taki Taki (with Selena Gomez, Ozuna & Cardi B)	Moonlight	Better Now
Happier	FEFE (feat. Nicki Minaj & Murda Beatz)	SICKO MODE
I Love It (& Lil Pump)	Never Recover (Lil Baby & Gunna, Drake)	Fine China
In My Feelings	Money	Wake Up in the Sky
ZEZE (feat. Travis Scott & Offset)	Yes Indeed	No Stylist
MIA (feat. Drake)	Jocelyn Flores	Leave Me Alone
Arms Around You (feat. Maluma & Swae Lee)	Fuck Love (feat. Trippie Redd)	BIG BANK (feat. 2 Chainz, Big Sean
Better	Mona Lisa (feat. Kendrick Lamar)	STOOPID
SAD!	Psycho (feat. Ty Dolla $ign)	All Girls Are The Same
Taste (feat. Offset)	Hard Work Pays Off	rockstar (feat. 21 Savage)
Eastside (with Halsey & Khalid)	Last Memory	Lean Wit Me
God's Plan	REEL IT IN	Lucky You (feat. Joyner Lucas)
Suncity (feat. Empress Of)	Youngblood	WORKIN ME
Topanga	Look Alive (feat. Drake)	Woman Like Me (feat. Nicki Minaj)
Kiss and Make Up	Everybody Dies In Their Nightmares	STARGAZING
Uproar	YOSEMITE	Saturday Nights
High Hopes	Look At Me!	Wasted (feat. Lil Uzi Vert)
Ric Flair Drip (& Metro Boomin)	Congratulations	XO TOUR Llif3
Te BotÃ© - Remix	Sweet but Psycho	God is a woman
Plug Walk	no tears left to cry	Love Lies (with Normani)
Kamikaze	Get Dripped (feat. Playboi Carti)	Hope
Killshot	goosebumps	Close To Me (feat. Swae Lee)
FeRRIs WhEEL (feat. Trippie Redd)		