The following Features were used for clustering:

Acousticness
Danceability
Energy
Instrumentalness
Liveness
Loudness
Speechiness
Tempo
Valence

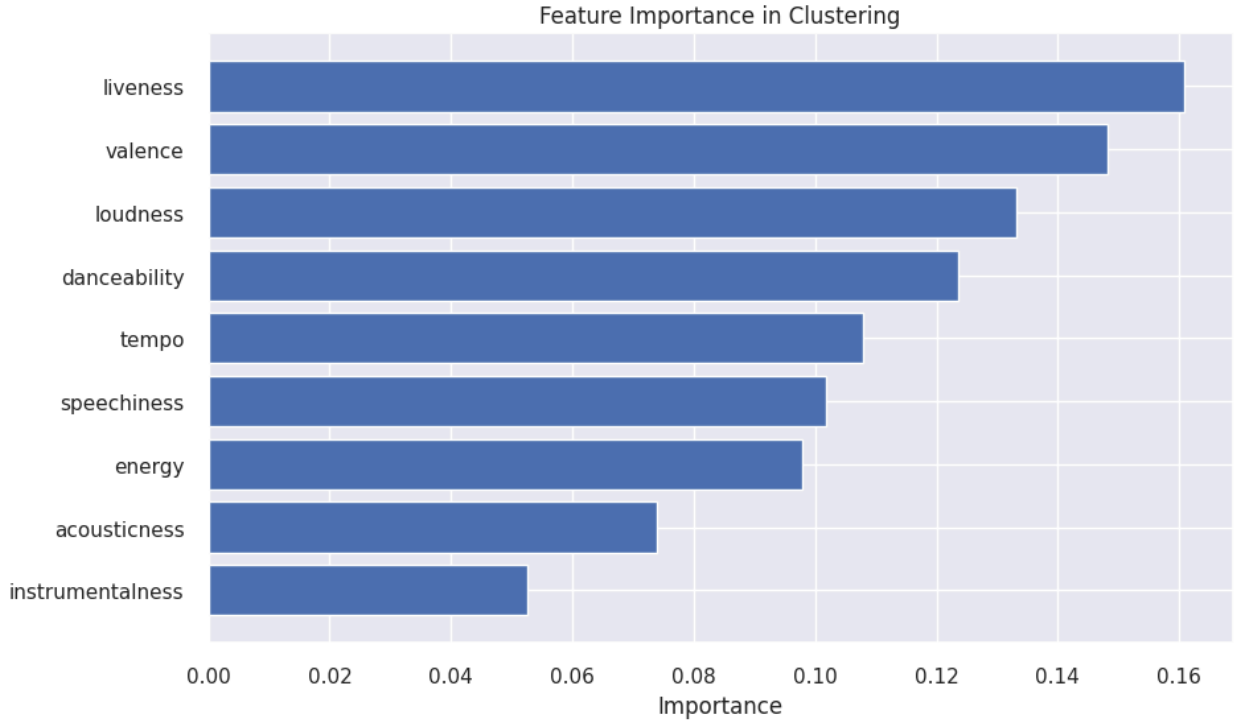
View the following link for a data dictionary [Web API Reference | Spotify for Developers]

The data was from Spotify. For the hierarchical and DBSCAN clustering, we pulled a playlist called Top 50 global songs using the Spotify API. For the kmeans algorithm, a dataset with 500 Spotify songs was used.

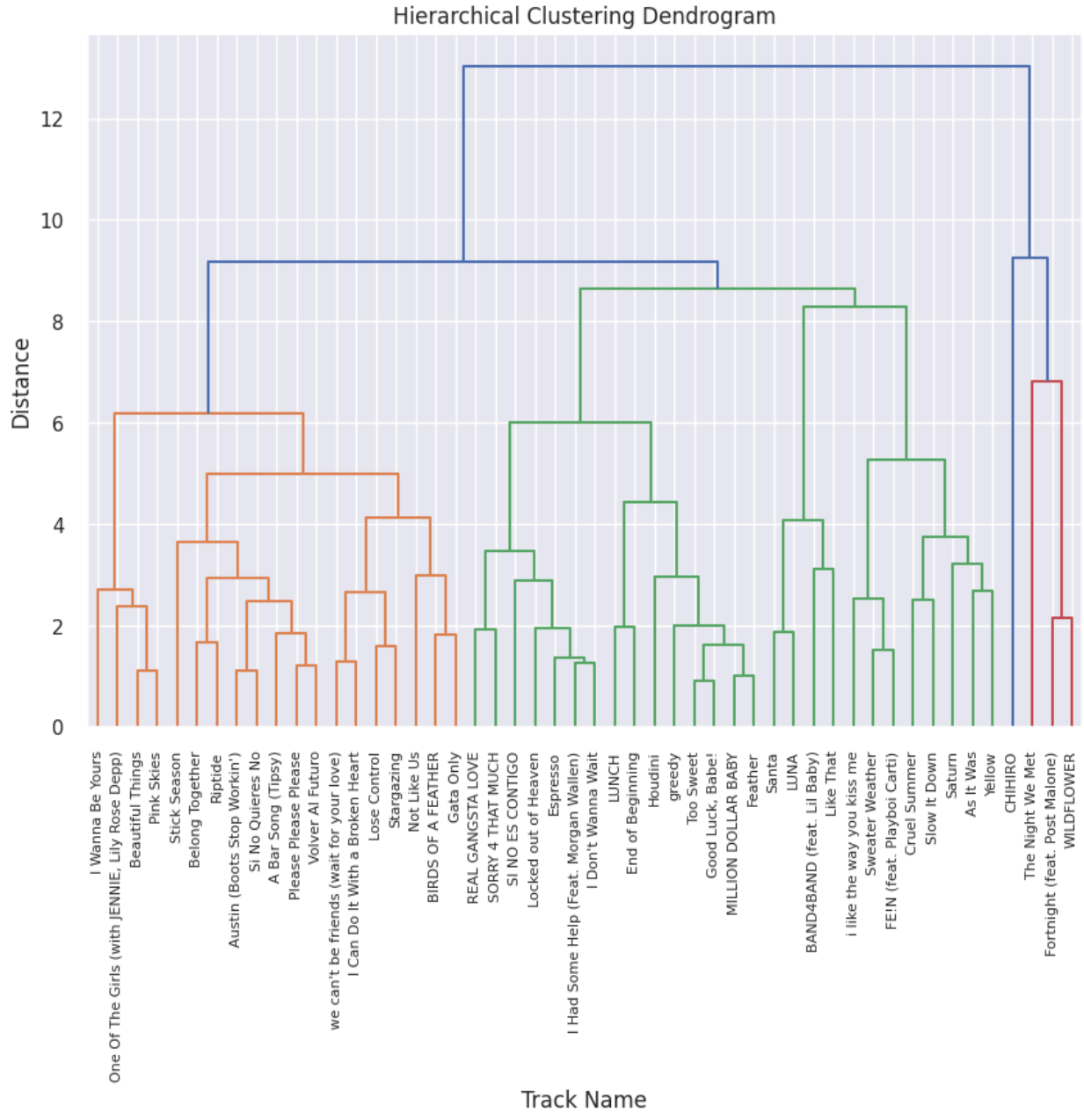
Hierarchical Clustering

Clustering using all features

Features' Importance in Clustering



Clusters' Interpretation using all features



There are 3 main clusters.

1. Orange Cluster (Left): This cluster contains songs like "I Wanna Be Yours", "One Of The Girls", "Beautiful Things", "Stick Season", and others. These songs seem to have a more indie, folk, or

alternative pop vibe. Many of these tracks are by artists known for their introspective or emotional lyrics.

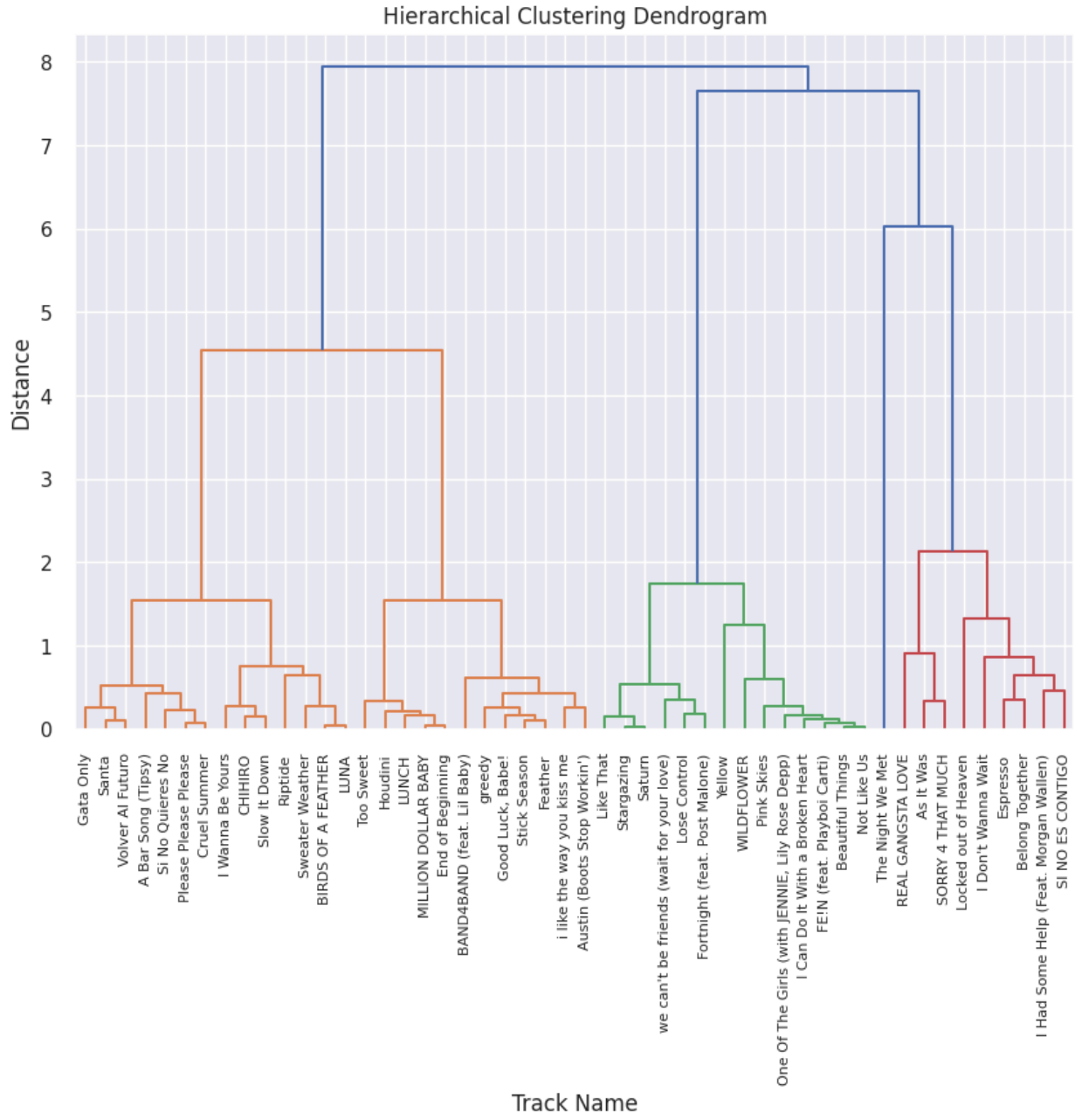
2. Green Cluster (Middle): This larger cluster includes songs like "Greedy", "Paint The Town Red", "Lovin On Me", "Feather", and "Water". This group appears to represent more mainstream pop and R&B tracks. These songs are likely more upbeat, danceable, or chart-oriented.
3. Red Cluster (Right): This small cluster only contains two songs: "Fortnight (feat. Post Malone)" and "VAMPIRE". These might be outliers or represent a distinct style, possibly more experimental or genre-blending tracks.

Interpretation:

- The orange cluster might represent more alternative or indie-leaning tracks, possibly with more introspective or emotional themes.
- The green cluster likely represents popular, mainstream hits that are more radio-friendly and danceable.
- The red cluster, being small and separate, might indicate songs that are unique in style or don't fit neatly into the other categories.

The blue lines at the top of the dendrogram show how these clusters relate to each other. The orange and green clusters are more closely related to each other than to the red cluster, suggesting that while they represent different styles, they still share some common characteristics that set them apart from the songs in the red cluster.

Clusters' Interpretation using Top 2 features (Liveness and Valence)



Analyzing this dendrogram, we can identify 4 main clusters.

1. Orange Cluster (Left):
This large cluster includes songs like "Gata Only", "Snooze", "A Bad Song (Tipsy)", "PERRO NEGRO", and others. This group seems to represent a mix of genres, possibly including Latin pop, R&B, and some alternative tracks. It's the largest and most diverse cluster.

2. Green Cluster (Middle-Left):
This smaller cluster contains songs like "Lovin On Me", "Lose Control", "MELTDOWN", and "Paint The Town Red". These tracks appear to be more mainstream pop and hip-hop hits, likely upbeat and danceable.

3. Red Cluster (Right):
This cluster includes songs like "Fortnight (feat. Post Malone)", "Vampire", "Is It Over Now?", and "NOW AND THEN". These might represent more alternative or indie pop tracks, possibly with more emotional or introspective themes.

4. Blue Cluster (Far Right):
This small cluster contains only a few songs like "I Had Some Help (Feat. Morfar)", "CONTIGO", and "Beneath Malen". These could be outliers or represent a distinct style, possibly more experimental or genre-blending tracks.

Interpretation:
- The orange cluster seems to represent a broad mix of popular music styles, suggesting these songs share some common characteristics despite genre differences.
- The green cluster likely represents chart-topping, radio-friendly pop and hip-hop tracks.
- The red cluster might indicate more alternative or indie-leaning pop songs, possibly with deeper lyrical content.
- The blue cluster, being small and separate, might indicate songs that are unique in style or don't fit neatly into the other categories.

The height at which clusters merge in the dendrogram indicates how similar or different they are. The orange and green clusters merge at a lower height, suggesting they're more similar to each other than to the red and blue clusters.

The clustering provides insight into how these songs relate to each other in terms of their musical characteristics, which might not always align with traditional genre classifications.

Clusters' Musical Features Statistics using Top 2 features (Liveness and Valence)

	cluster	index	liveness	valence
0	1	count	26	26
1	1	mean	-0.523157	0.555334
2	1	std	0.235778	0.71002
3	1	min	-1.06821	-0.627916
4	1	25%	-0.635904	-0.00811083
5	1	50%	-0.499863	0.441351
6	1	75%	-0.388511	1.11862
7	1	max	0.0523605	1.61734

+----+-----------+---------+------------+------------+

+----+-----------+---------+------------+-----------+
| | cluster | index | liveness | valence |
|----+-----------+---------+------------+-----------|
0	3	count	9	9
1	3	mean	1.14516	0.520024
2	3	std	0.605487	0.422442
3	3	min	0.394981	0.0124125
4	3	25%	0.667062	0.209437
5	3	50%	0.999605	0.455717
6	3	75%	1.443	0.738939
7	3	max	2.29955	1.30949
+----+-----------+---------+------------+-----------+

+----+-----------+---------+------------+-----------+
| | cluster | index | liveness | valence |
|----+-----------+---------+------------+-----------|
0	2	count	14	14
1	2	mean	-0.118878	-1.23336
2	2	std	0.388321	0.246679
3	2	min	-0.723574	-1.74439
4	2	25%	-0.360799	-1.37804
5	2	50%	-0.08368	-1.24567
6	2	75%	0.0221292	-1.06096
7	2	max	0.888757	-0.878301
+----+-----------+---------+------------+-----------+

+----+-----------+---------+------------+-----------+
| | cluster | index | liveness | valence |
|----+-----------+---------+------------+-----------|
0	4	count	1	1
1	4	mean	4.95989	-1.85193
2	4	std	nan	nan
3	4	min	4.95989	-1.85193
4	4	25%	4.95989	-1.85193
5	4	50%	4.95989	-1.85193
6	4	75%	4.95989	-1.85193
7	4	max	4.95989	-1.85193
+----+-----------+---------+------------+-----------+

Cluster 1 (26 songs):
- Low liveness (-0.523 mean) suggests studio-polished tracks
- Moderate to positive valence (0.555 mean) indicates generally upbeat mood

- This cluster likely represents mainstream, radio-friendly pop songs that are produced for mass appeal. These tracks could be great for licensing in commercials or as background music in retail spaces.

Cluster 2 (14 songs):
- Slightly negative liveness (-0.119 mean) indicates a mix of live and studio elements
- Very low valence (-1.233 mean) suggests somber or melancholic tracks
- This cluster might represent emotional or introspective songs, potentially appealing to listeners seeking depth or catharsis. These could be valuable for soundtrack licensing in dramatic films or TV shows.

Cluster 3 (9 songs):
- High liveness (1.145 mean) indicates a more raw, live-sounding production
- Moderate valence (0.520 mean) suggests a balanced emotional tone
- This cluster likely represents more authentic or "indie" sounding tracks. These could appeal to listeners valuing perceived authenticity and could be great for branding campaigns targeting younger, trend-conscious demographics.

Cluster 4 (1 song):
- Extremely high liveness (4.960) and very low valence (-1.852)
- This outlier track is unique, possibly experimental or avant-garde. It could represent a niche market or be used to create standout moments in marketing campaigns.

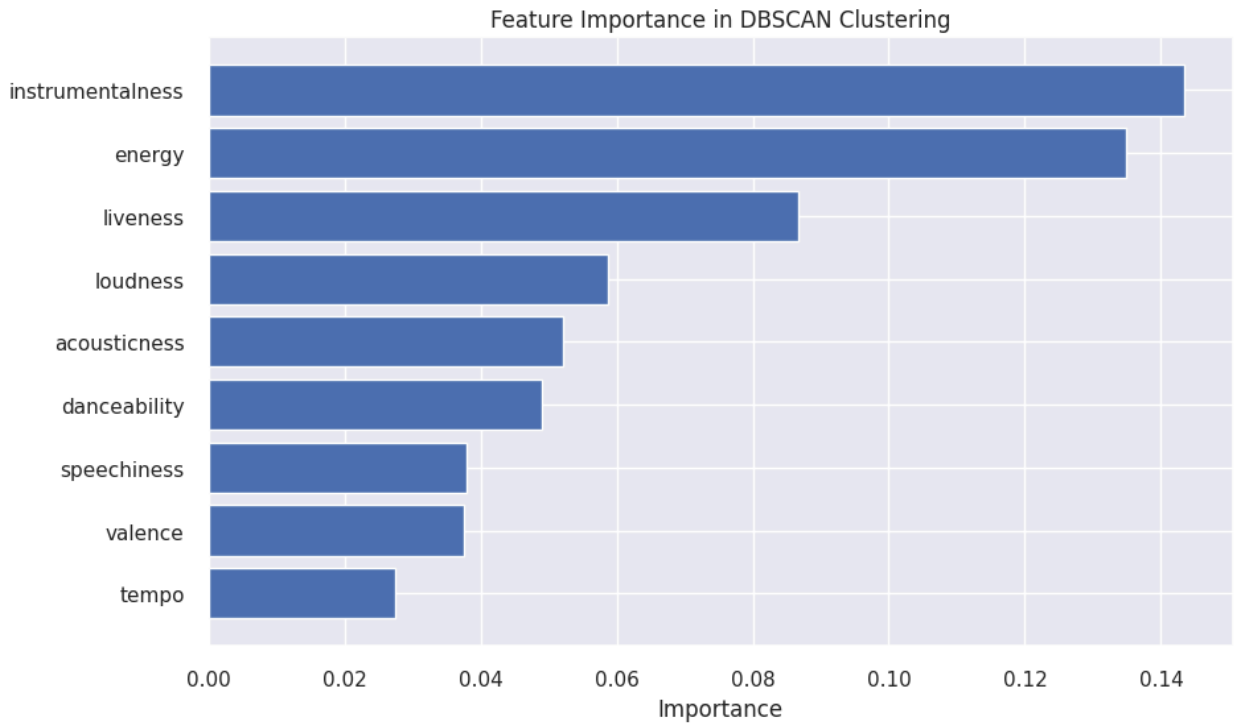
Business Implications:
1. Diverse Portfolio: The clusters show a range of musical styles, allowing for a balanced catalog that can cater to various listener moods and preferences.

2. Targeted Marketing: Each cluster can be marketed differently. Cluster 1 for mass appeal, Cluster 2 for emotional impact, Cluster 3 for authenticity-seeking listeners, and Cluster 4 for niche markets.

3. Playlist and Radio Programming: Understanding these clusters can help in creating more effective playlists or radio stations, potentially increasing listener engagement and retention.

4. Licensing Opportunities: Different clusters may be more suitable for various licensing scenarios (e.g., Cluster 1 for commercials, Cluster 2 for dramatic scenes, Cluster 3 for indie film soundtracks).

DBSCAN
The DBSCAN clustering is still a work in progress and needs more tuning to get interpretable results (see code for details) but most of them are
Features' Importance in Clustering



Kmeans - Overall:
Loudness and energy have positive correlations.
No other significant insights

Rhythm/ Energy:
High danceability, with high energy, and low tempo
High Danceability, lower energy and middle tempo
Low energy, low danceability and low tempo

Mood:
Liveness is one of the main characteristics that distinct cluster 0 from others
Cluster 2: high acoustic
Cluster 3: High valence
Cluster 1: High energy and low in other aspect

Speechiness:
Cluster 3: high instrumentalness
Cluster 0: high speechiness
Cluster 1 high loudness and low instrumentalness and speechiness
Cluster 2: Low in instrumentalism and middle for other aspects

PCA: no significant findings