

# Automating Playlist Creation

**Can Unsupervised Machine  
Learning help us?**

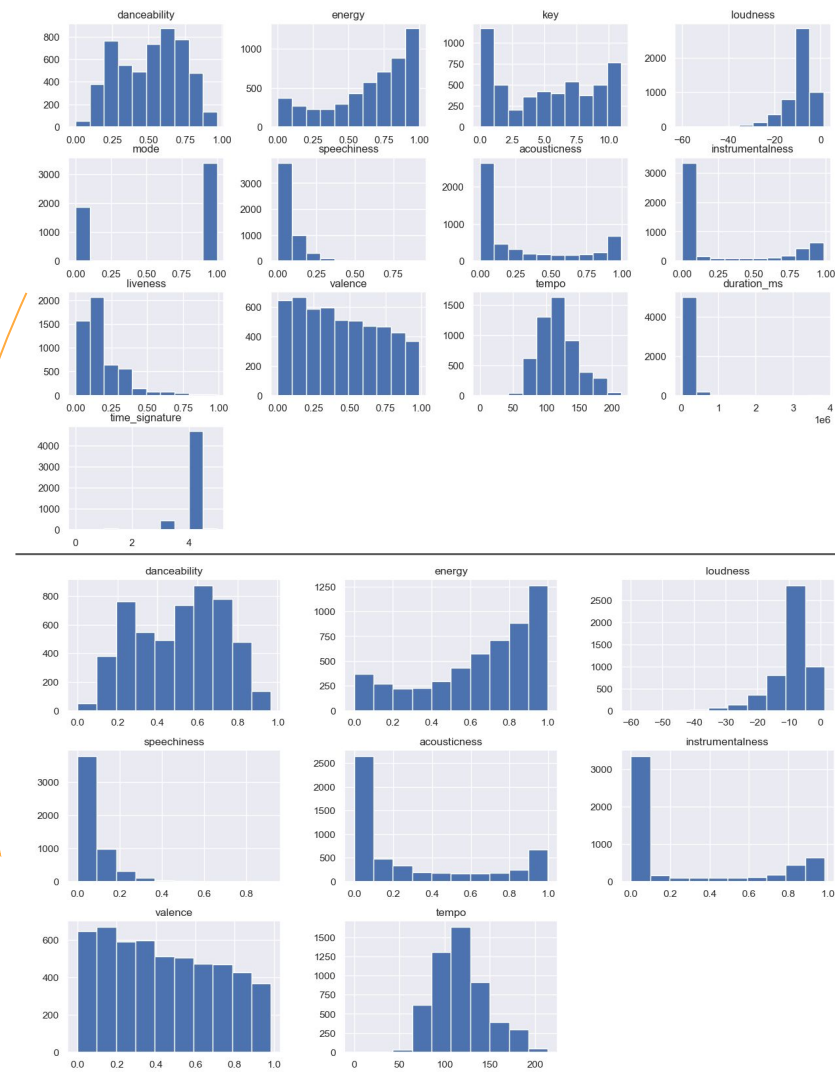
# Introduction & Challenge

**Challenge:** Moosic's current manual playlist process can't keep up with business growth.

**Our goal:** Prototype automated playlist creation using machine learning clustering based on Spotify audio features.

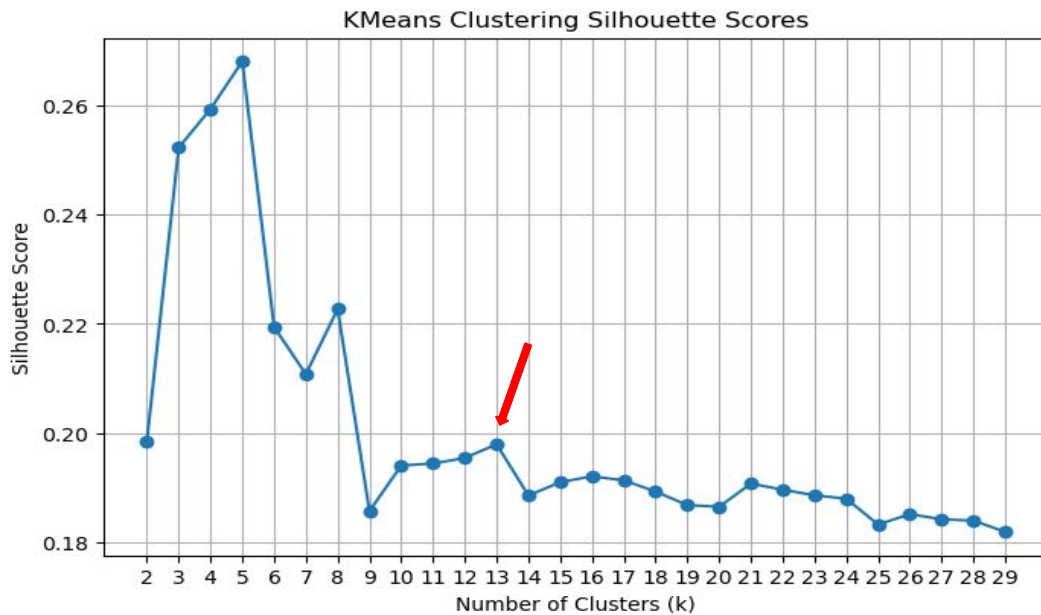
# Choosing Best Features.

- In order to achieve our business goal, we had to choose the more relatable features we needed. While some features worth experimenting, some of them were more general and skewed and some were unrelated to our goal.
- After carefully examining we decided to go with **8 out of 13**:  
Danceability, Energy, Speechiness, Acousticness, Instrumentalness, Liveness, Valence, Tempo.
- To prepare data, tested different scaling techniques and found **RobustScaler** worked best as it reduces outlier impact, preserving meaningful distances.



# Playlists & Clusters

- Explored cluster numbers 2–29; best around **13 clusters** balancing distinctiveness and usability.
- Dimensionality reduction (**PCA**) used for visualization but didn't help with clustering accuracy.

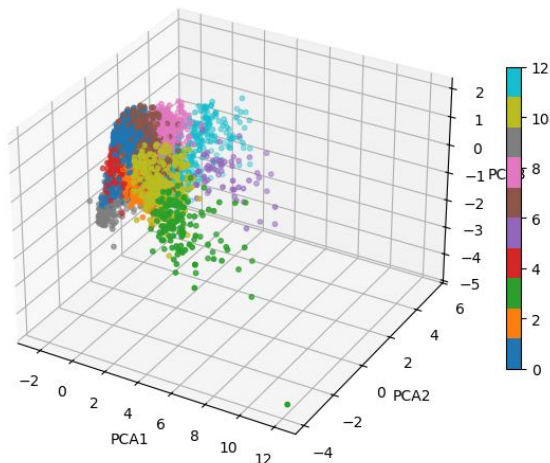


# Our Approach & Data

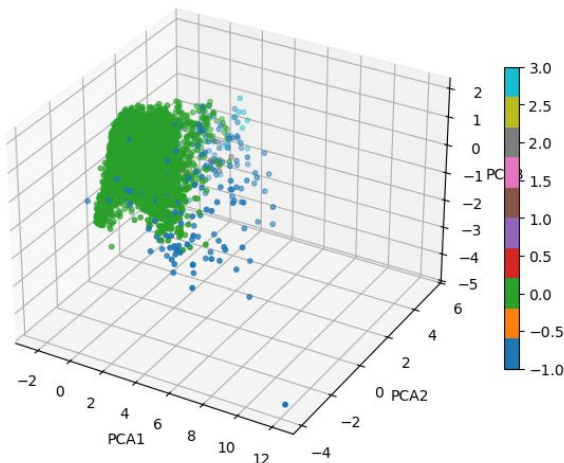
- Used unsupervised learning (**KMeans**, **DBSCAN**, **Agglomerative**) to group songs.
- Clusters represent meaningful song groups by mood/genre for playlist creation.
- As shown on 3D visualization which we created by reducing dimensions using **PCA**.

**KMeans** algorithm gives us the best Clusters (Playlists).

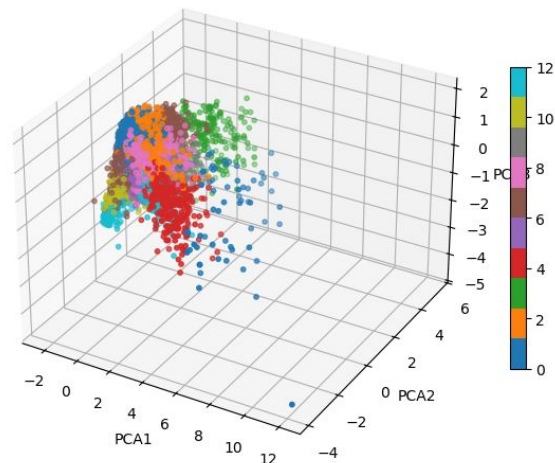
KMeans Clusters

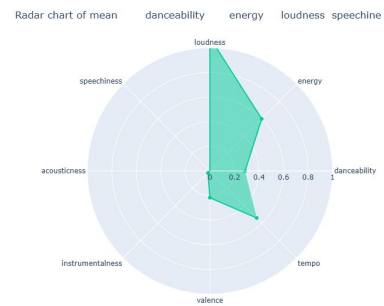
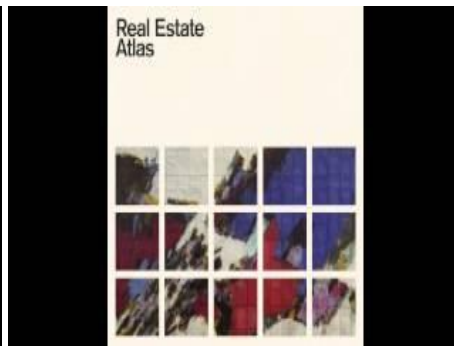


DBSCAN Clusters

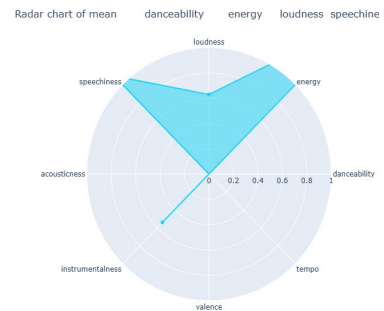


Agglomerative Clusters





## A sample of good playlist



## A sample of not well made playlist

# Prototype Results & Insights

- Our clustering prototype reveals some natural groupings, but the clusters are not sharply defined.
- Audio features like tempo, energy, and danceability help separate songs, but group overlap remains significant.
- Playlist clusters show only moderate cohesion, reflecting the complexity of musical similarity.
- Audio features alone do not fully capture nuances such as lyrics, cultural context, or listener preferences.
- While promising, the prototype currently falls short of perfectly replicating expert-curated playlists.

# Recommendation

*A Hybrid* solution with **80%** getting help from *machine learning* algorithms and **20%** *human* judgement.



**Thank you for your attention.**



**Feel free to ask questions!**



***Azumi, Victoria, Sahand***