Creating a spotify playlist using clustering

Tejal Sungra, tejalsungra@gmail.com
Shavi, skambojkhera@gmail.com
Shravanti , shrav1708@gmail.com
René Wollny, wollny.rene@gmail.com

Task

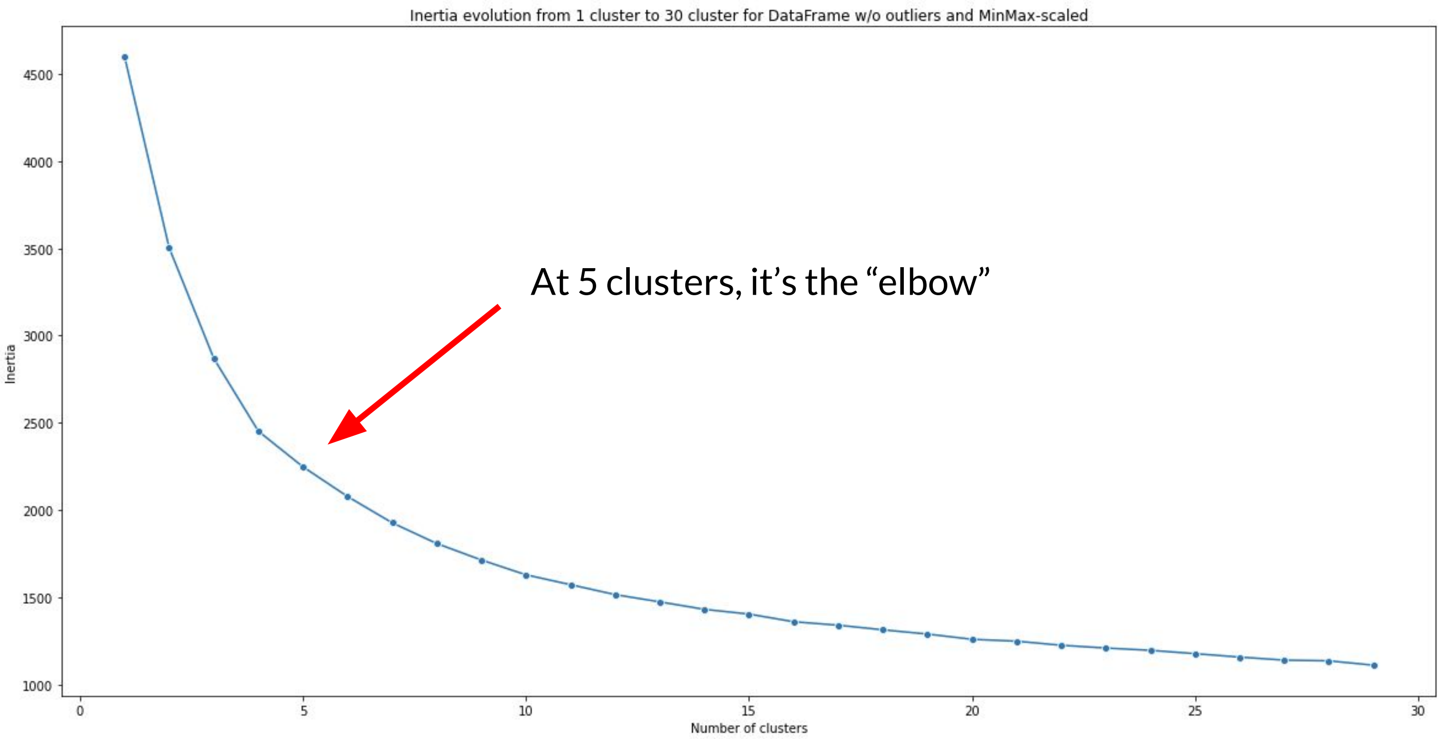
- Create playlist(s) out of 5000 songs using Kmean-clustering

- Explore different data scaling models

Proceedings

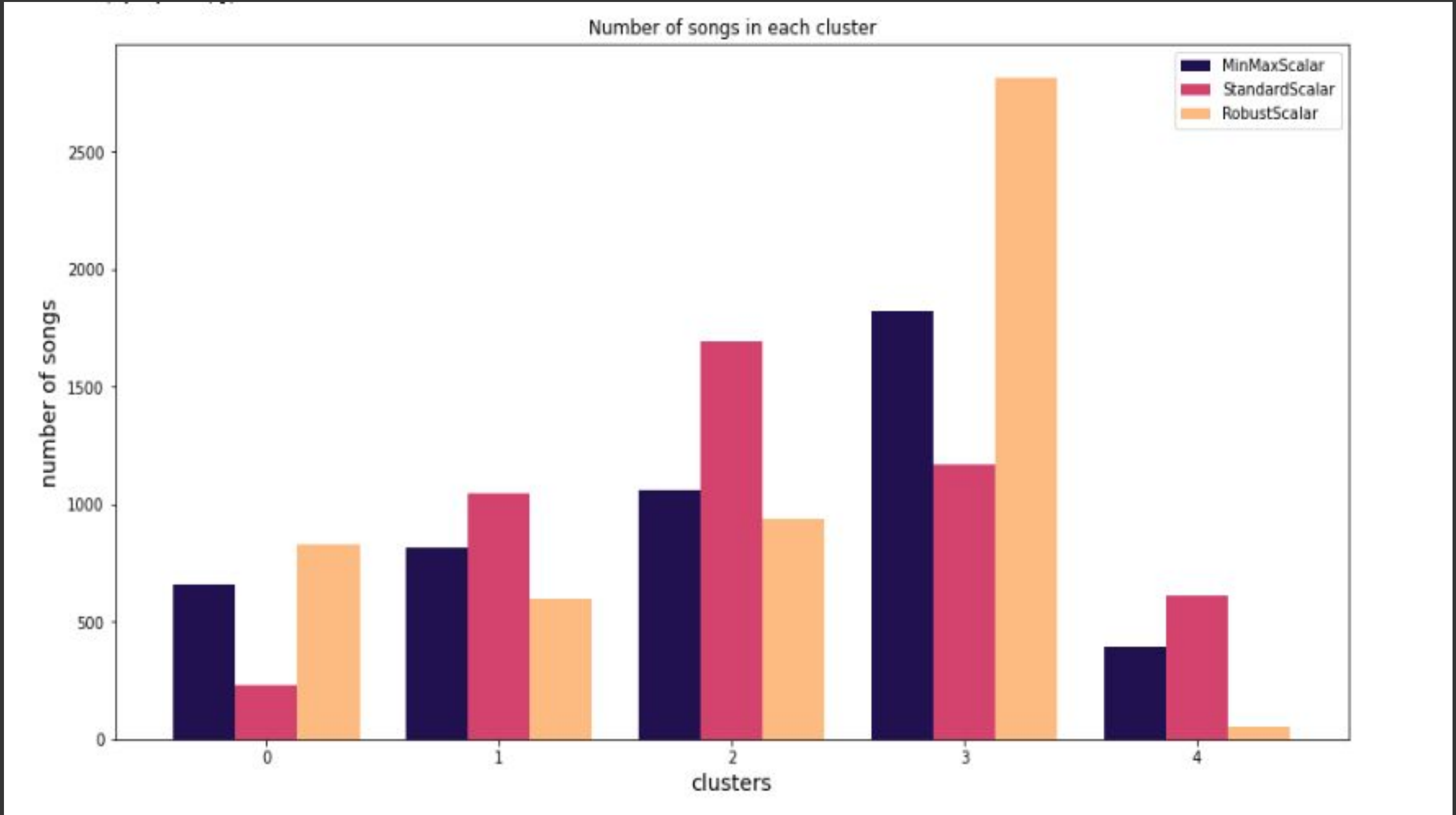
- Clean data set (drop unused columns, remove outliers, clean column names)
- Use different scaling methods (MinMax-, Standard-, Robust-Scaler)
- Use “Elbow-Method” to define amount of clusters (=playlists)
- Use Silhouette score to decide on best scaler
- Check different song-features for each cluster to name playlists

Selecting the number of clusters using the “Elbow-Method”



At 5 clusters, it’s the “elbow”

— Data scaling models

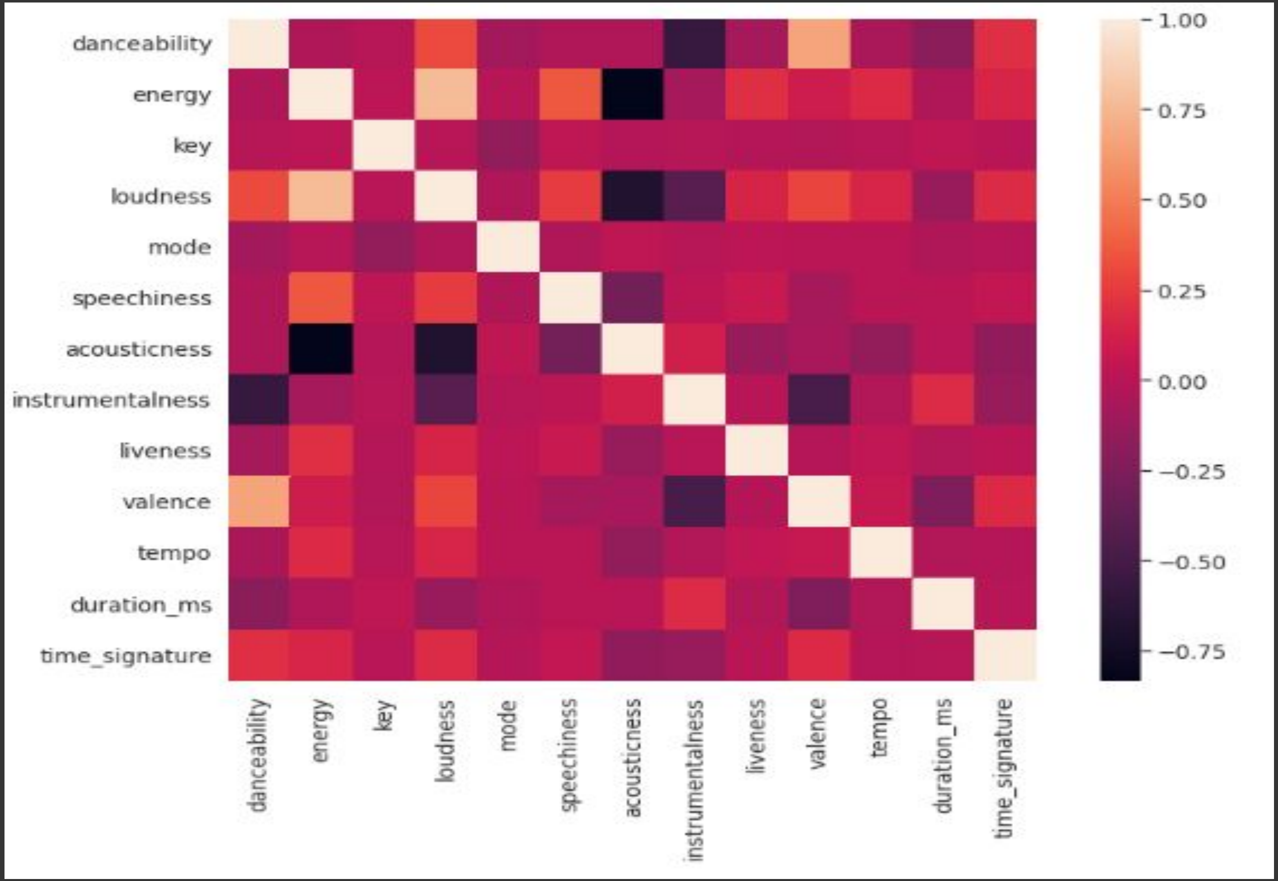


1. MinMaxScaler
Silhouette score : 0.253

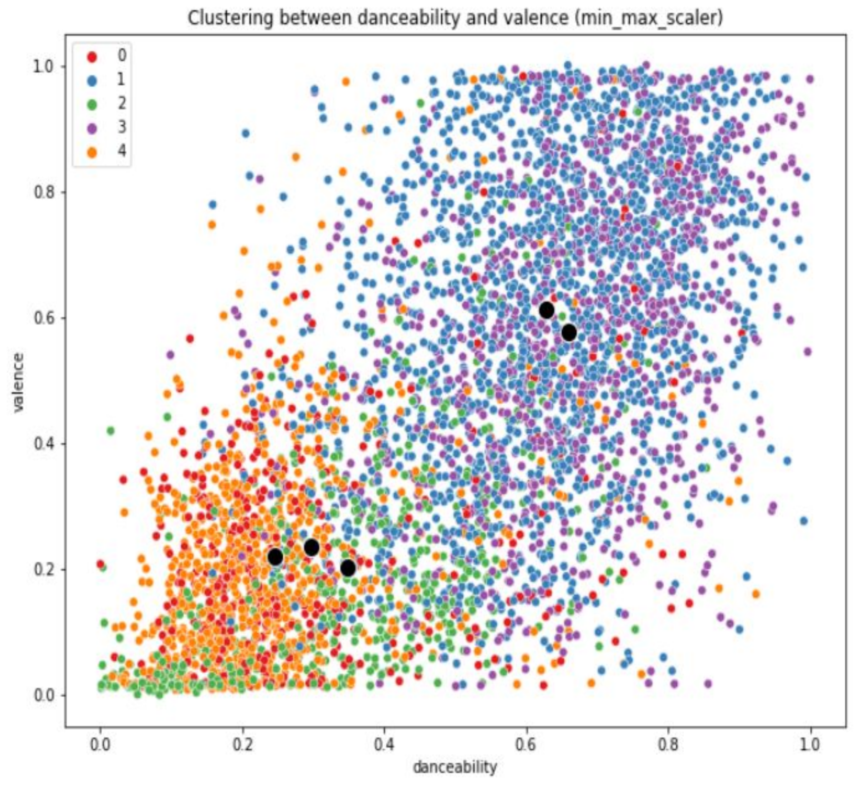
2. StandardScaler
Silhouette score : 0.15

3. Robust Scaler
Silhouette score : 0.20

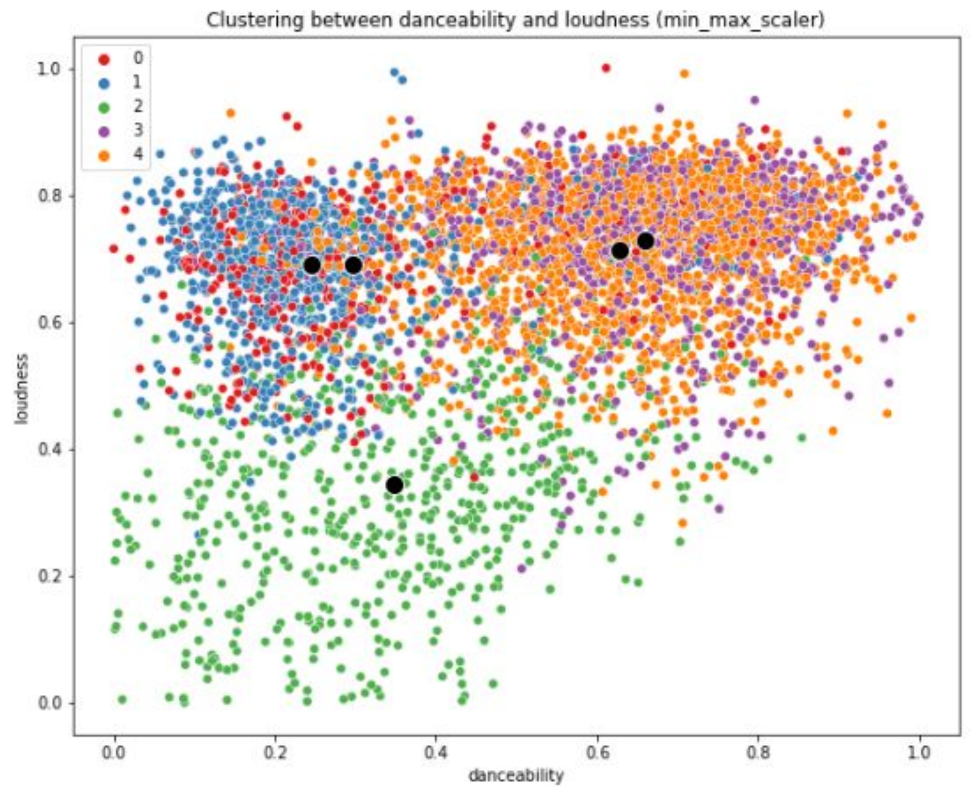
Correlation between different song-features



Correlation between danceability and other features

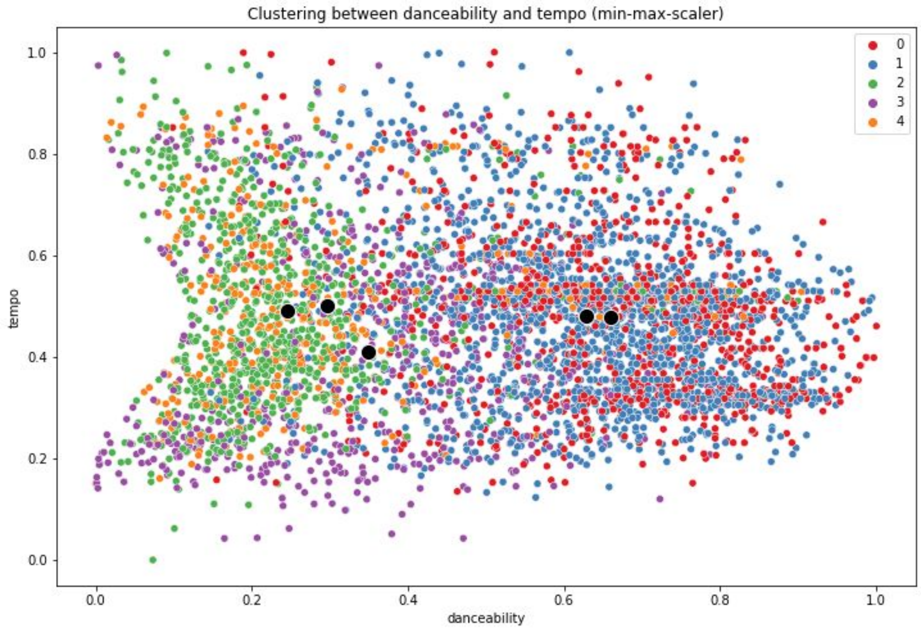


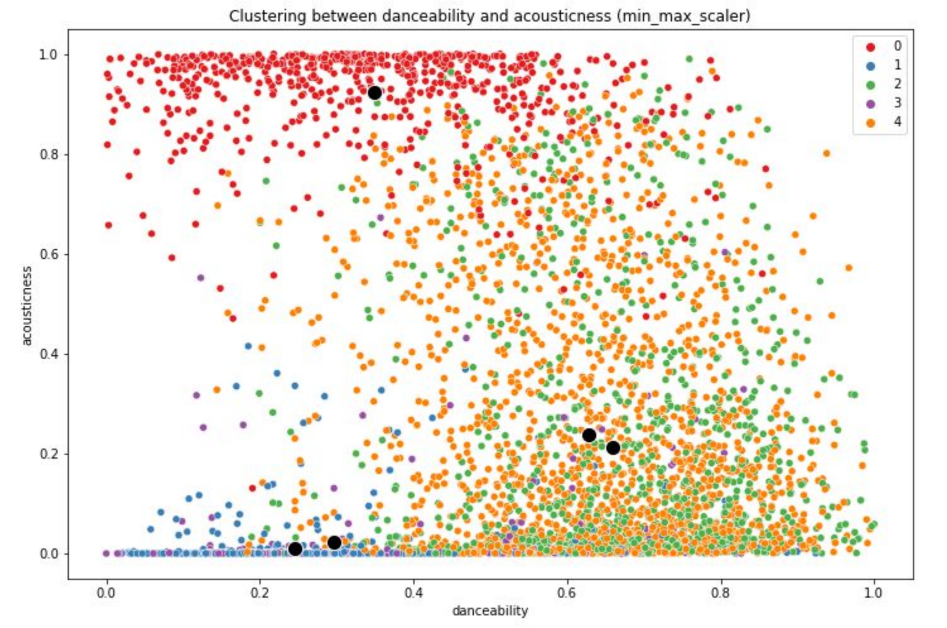
Clustering between danceability and valence (min_max_scaler)



Clustering between danceability and loudness (min_max_scaler)

Correlation between danceability and other features


Clustering between danceability and tempo (min-max-scaler)


Clustering between danceability and acousticness (min_max_scaler)

Cluster 4

- Loud
- Instrumental
- Acoustic
- Orchestra
- Symphonies

Name of playlist: “Best of Instrumental Music”

Pro’s of K-Means

- Relatively simple to implement.
- Scales to large data sets.
- Can warm-start the positions of centroids.
- Easily adapts to new examples.

Con’s of K-Means

- Choosing the k values manually is a tough job.
- Being dependent on initial values.
- It is sensitive to the outliers.
- As the number of dimensions increases its scalability decreases.
- Overlapping the clusters.

Conclusion

- Spotify’s audio features is a good way to describe a song in numbers, but not close to human understanding of the song
- K-Means is not the best way to create playlists

Recommendation

- Add more audio features: like pitch, melody, frequency, ….
- Explore different methods to create playlist