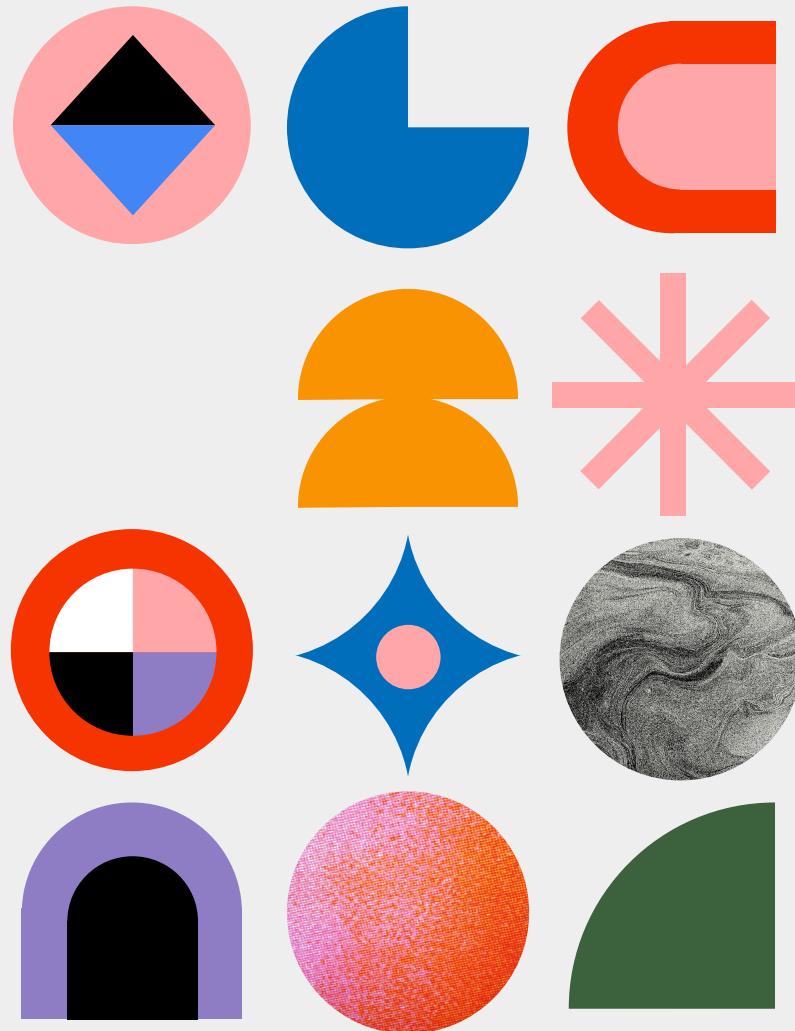


# Data-Driven DJ: Crafting Spotify Playlists with Python & ML





# Our new task as data scientist:

We trained a model to create the perfect playlists, making it data-driven and scalable.



# 1. Clean the dataset

---

## Step 1

- Stripped the column names

## Step 2

- Combined name and artist and used it as index
- Removed duplicates

## Step 3

- Deleted columns, that won't give us valuable information

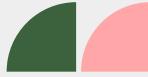


## 2. Scaling

We tested various scalers and found that the MinMax scaler performs better than the others for our needs.

For accurate analysis, we applied MinMaxScaler to columns that weren't already within the 0 to 1 range.





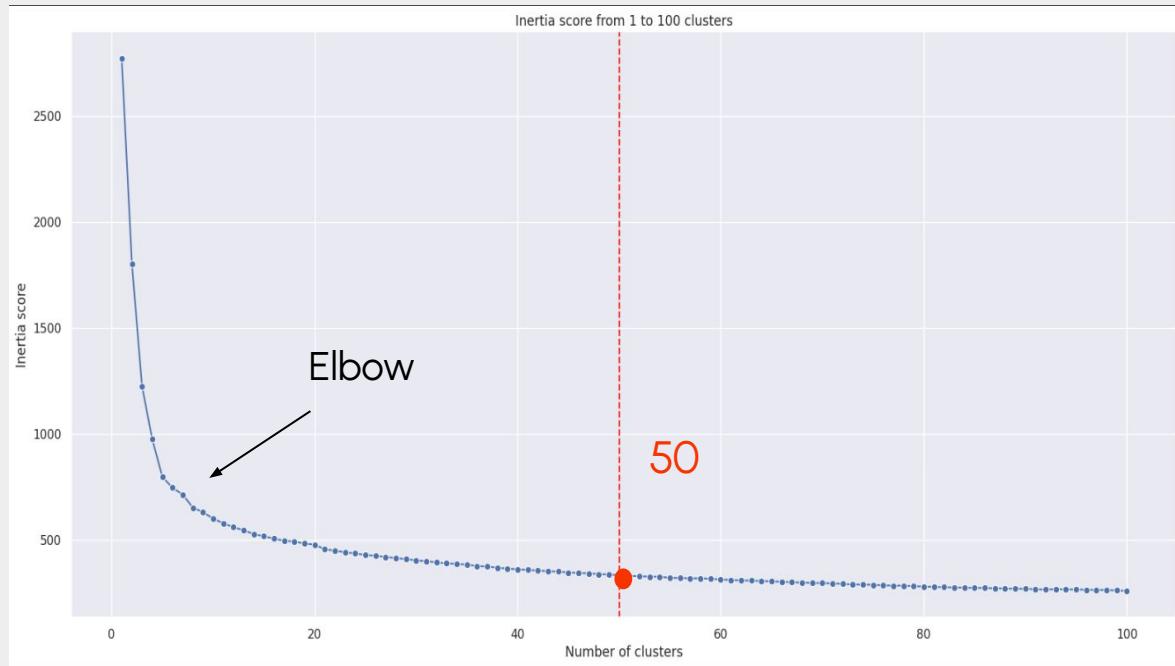
### 3. Finding the correct K numbers

We analysed inertia for data set

Elbow suggested an optimal range between 2 and 10

According to business needs from 20 to 100

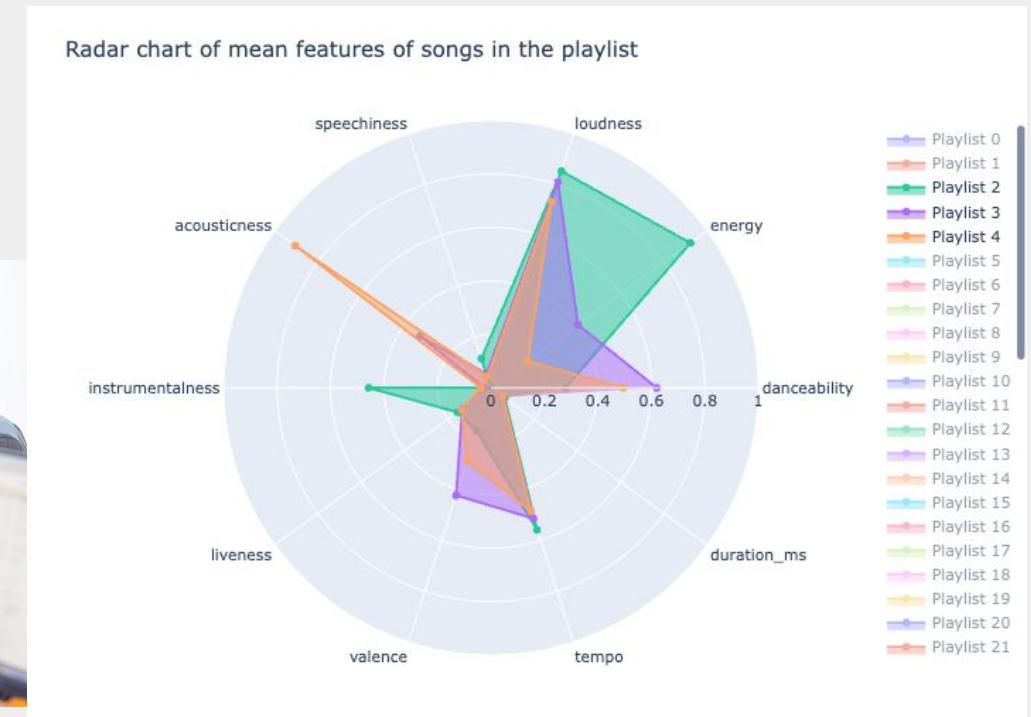
We decided  
to take  
**50**

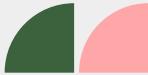




# 4. Radar chart

With the help of a radar chart we spotted out meaningful differences between certain playlists.





# 5. Results: Our built playlists

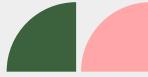


We created a bridge between our Colab notebook and Spotify, instantly generating playlists with songs from the clusters.

The songs grouped into clusters are highly accurate, each fitting well within a specific style, leading to great results!

The screenshot shows the Spotify mobile application interface. On the left, the 'Your Library' screen displays a list of playlists, all created by 'Kee.anyaoha'. The playlists are named 'my\_playlist\_cluster\_49', 'my\_playlist\_cluster\_48', 'my\_playlist\_cluster\_47', 'my\_playlist\_cluster\_46', 'my\_playlist\_cluster\_45', and 'my\_playlist\_cluster\_44'. Each entry includes the name, type (Playlist), creator, and the date added (34 minutes ago or 33 minutes ago). On the right, a specific playlist titled 'this is a test list' by 'Kee.anyaoha' is selected. The playlist summary shows 10 songs and a total duration of 35 min 26 sec. The tracklist on the right lists nine songs, each with a play button, title, and duration:

#	Title	Duration
1	Who Knew	3:28
2	Bringt ihn heim - Single...	3:48
3	Now She Walks with th...	2:22
4	Für immer	3:25
5	Crack A Bottle E	4:57
6	Heart Attack	3:30
7	Frauen regier'n die Welt	2:58
8	Applaus, Applaus - Rad...	3:07
9	Arthur's Theme (Best T...	3:54



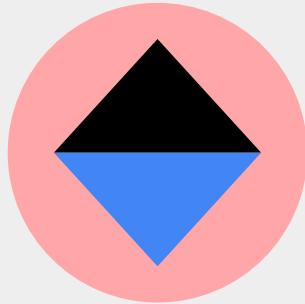
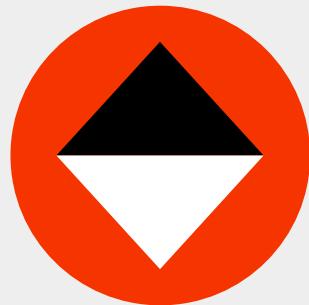
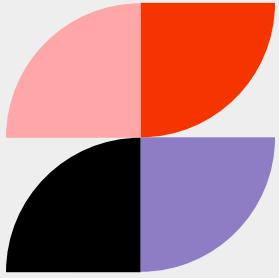
## 6. Our takeaways:

K-means is a good starting point to initially build clusters but there's always room for improvement, and it requires an additional touch to refine the results further.

F.ex.

- Using PCA to reduce the count of features used
- Choosing an alternative clustering method
- Providing more features to have more various song characteristics, like:
  - Genre, artist popularity, rhythm patterns lyrics sentiments





# THANK YOU!

