JAMMIN' WITH SPOTIFY

Christophe Vivensang - Data Engineer
Gia Liu - Data Modeler
Joe Brueneman - Data Design Specialist
Julia Min - Data Visualizer
Minh Le - Data Research Specialist

SPOTIFY MUSIC RECOMMENDATION

- Spotify has 574 million monthly active listeners as of 2024
- 4 Billion Playlists
- Over 100 million tracks

DATA

1 MILLION PLAYLISTS

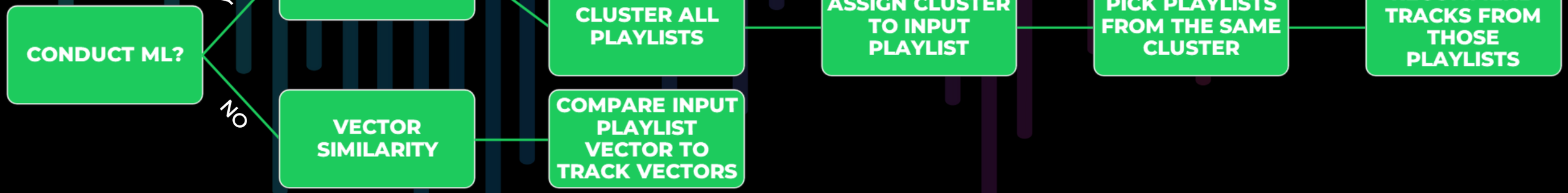
66.3 MILLION TRACKS

2.2 MILLION UNIQUE TRACKS

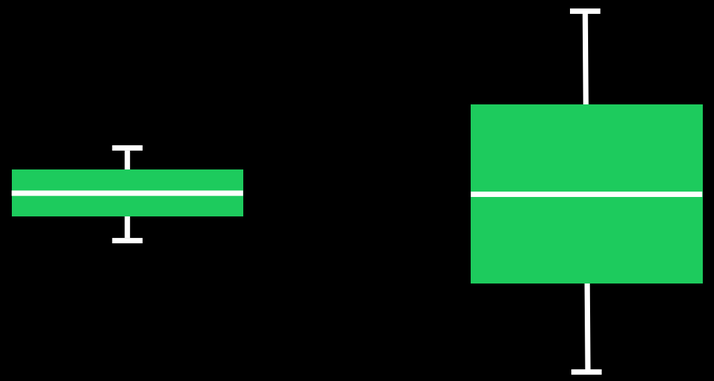
AUDIO FEATURES


```
{
  "acousticness": 0.00242,
  "danceability": 0.585,
  "duration_ms": 237040,
  "energy": 0.842,
  "instrumentalness": 0.00686,
  "key": 9,
  "liveness": 0.0866,
  "loudness": -5.883,
  "mode": 0,
  "speechiness": 0.0556,
  "tempo": 118.211,
  "time_signature": 4,
  "valence": 0.428
}
```

IDEAS


CONDUCT ML?
YES
NO
CLUSTER
VECTOR SIMILARITY
CLUSTER ALL TRACKS
CLUSTER ALL PLAYLISTS
COMPARE INPUT PLAYLIST VECTOR TO TRACK VECTORS
INPUT USER PLAYLIST
ASSIGN CLUSTER TO INPUT PLAYLISTS
ASSIGN CLUSTER TO INPUT PLAYLIST
RECOMMEND TRACKS FROM SAME CLUSTER OF TRACKS
PICK PLAYLISTS FROM THE SAME CLUSTER
RECOMMEND TRACKS FROM THOSE PLAYLISTS

MEAN VS. HOMOGENEITY



$$\sigma_x \quad < \quad \sigma_x$$

PLAYLIST FINGERPRINT

Track 1
"acousticness": 0.00242,
"danceability": 0.585,
"duration_ms": 237040,
"energy": 0.842,
"instrumentalness": 0.00686,
"key": 9,
"liveness": 0.0866,
"loudness": -5.883,
"mode": 0,
"speechiness": 0.0556,
"tempo": 118.211,
"time_signature": 4,
"valence": 0.428

Track 2
"acousticness": 0.0295,
"danceability": 0.581,
"duration_ms": 204160,
"energy": 0.963,
"instrumentalness": 0,
"key": 11,
"liveness": 0.139,
"loudness": -4.087,
"mode": 1,
"speechiness": 0.0981,
"tempo": 129.192,
"time_signature": 4,
"valence": 0.788

...

Track 224
"acousticness": 0.0291,
"danceability": 0.641,
"duration_ms": 258343,
"energy": 0.922,
"instrumentalness": 0,
"key": 2,
"liveness": 0.0862,
"loudness": -4.457,
"mode": 1,
"speechiness": 0.0786,
"tempo": 146.078,
"time_signature": 4,
"valence": 0.847

$$\overrightarrow{Track} \quad (x_1, x_2, x_3, \dots)$$

$$\overrightarrow{Playlist} \quad (\bar{x}_1, \sigma_{x1}, \bar{x}_2, \sigma_{x2}, \bar{x}_3, \sigma_{x3}, \dots)$$

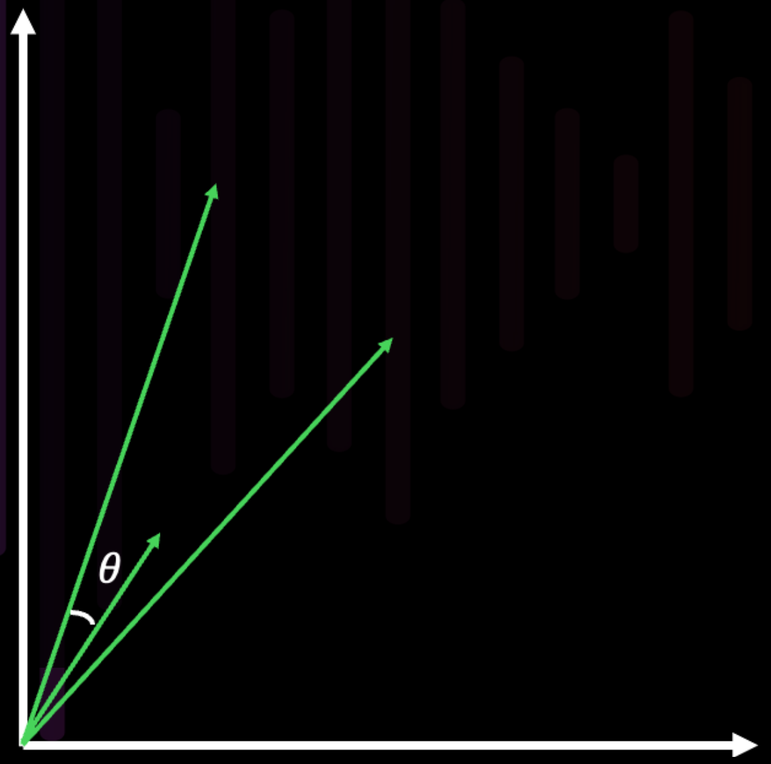
COSINE-SIMILARITY

TRACK VECTORS

- audio features
- standardize

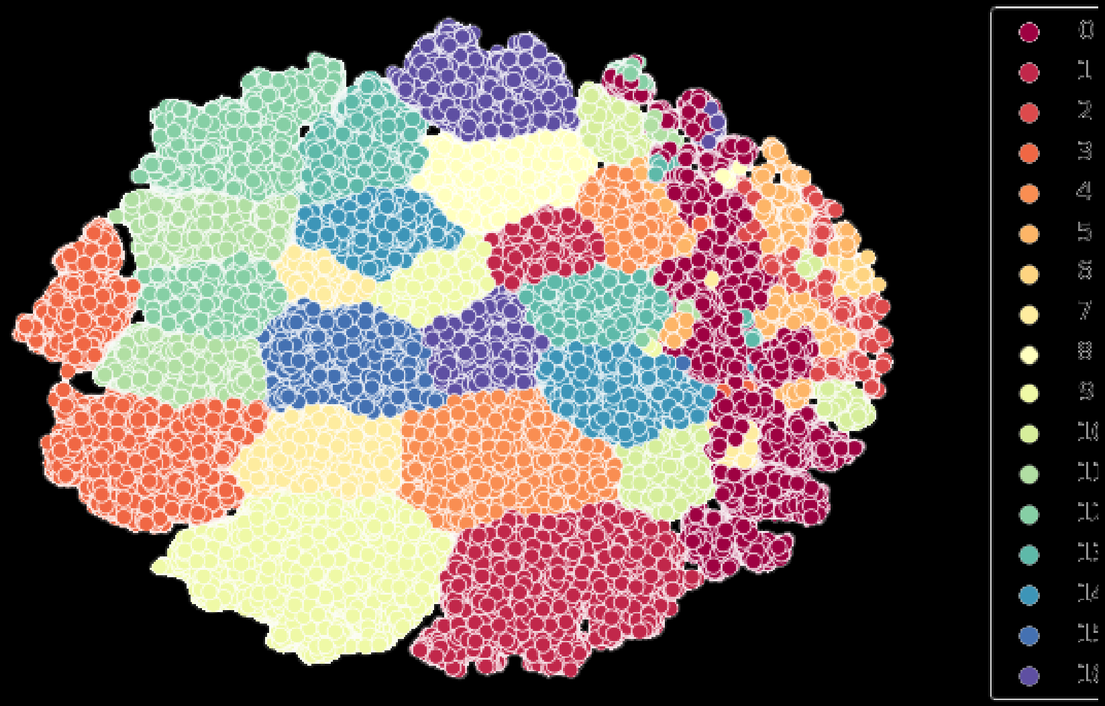
PLAYLIST VECTOR

- mean audio features
- standardize



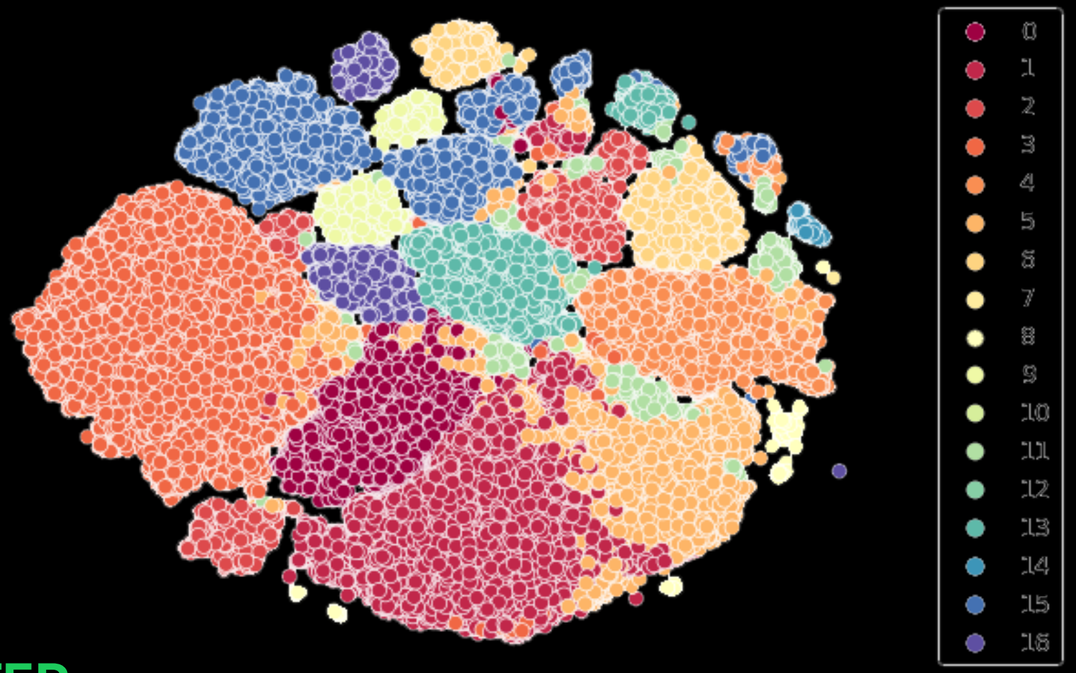
MODEL 1
TRACKS CLUSTERING

- CLUSTER THE TRACKS
- ASSIGN A CLUSTER TO INPUT PLAYLIST
- GET CLOSEST SONGS FROM THE SAME CLUSTER



CLUSTER = 17

MODEL 2
PLAYLISTS CLUSTERING

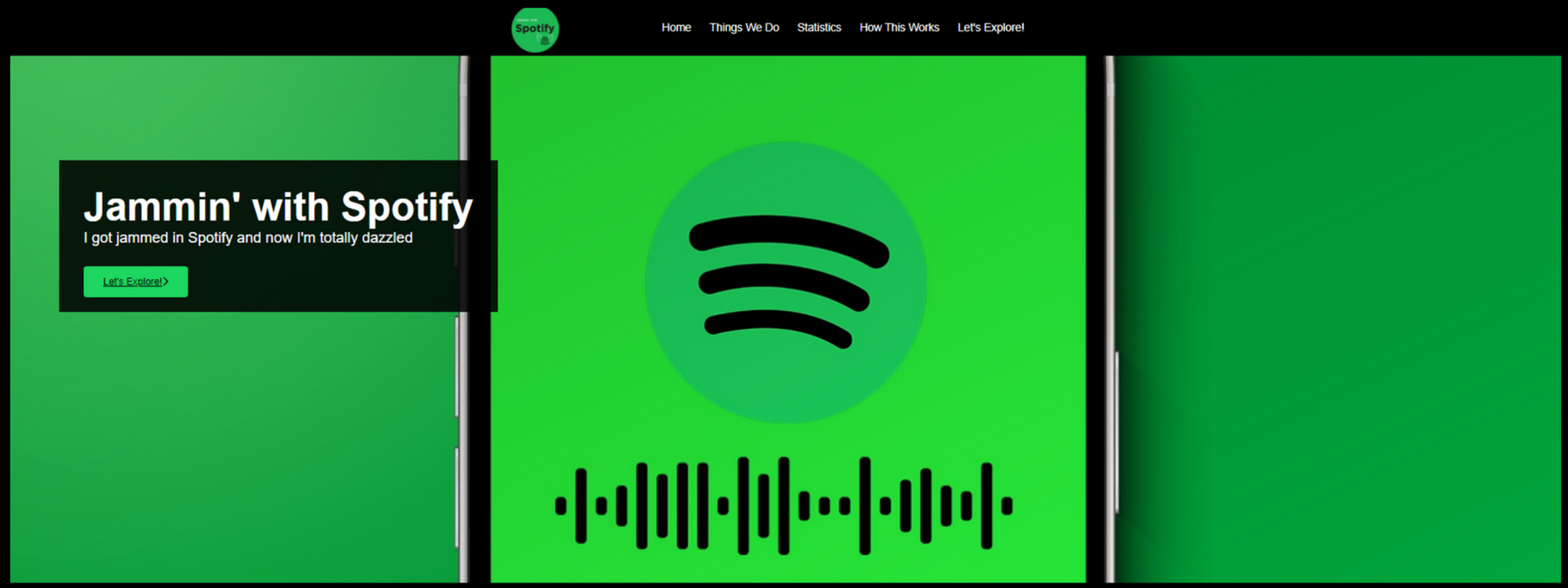


- CLUSTER THE PLAYLISTS
- ASSIGN A CLUSTER TO INPUT PLAYLIST
- GET CLOSEST PLAYLISTS FROM THE SAME CLUSTER
- RECOMMEND SONGS FROM THOSE PLAYLISTS

CLUSTER = 17

WEBPAGE DEMO

LET'S GO



This is a website that helps you find your next favorite song.

Are you curious about how Spotify recommends your next song?

What kind of features do they study to know you better?

How do they predict your taste in songs based on those features?

Here we show you how a song recommendation system could answer all these questions.

SUMMARY

We believe that playlist clustering is the most effective method

Track clustering does slightly more than vectorization

NEXT STEPS

- EXPLORE MORE COLLABORATIVE FILTERING METHOD
- FIELD TEST TO GATHER USER DATA
- IMPROVE MODEL DATA

RESOURCES

https://www.aicrowd.com/challenges/spotify-million-playlist-dataset-challenge
https://open.spotify.com/?
https://www.demandsage.com/spotify-stats/#:~:text=Monthly%20Spotify%20Stats,second%20with%20a%2019%25%20share.
https://medium.com/@onersarpnalcin/standardscaler-vs-minmaxscaler-vs-robustscaler-which-one-to-use-for-your-next-ml-project-ae5b44f571b9
https://bootcamp.uxdesign.cc/utilizing-spotifys-friend-feature-7e53ffe9d003