Search

Home

Library

Capstone:

Spotify Track Recommender

Kyle Ness

Problem

- Overabundance and aimlessness - search for new songs by genre? By artists?
- Waste of time, less entertainment, less discovery
- Can a recommender help?
  - How would it work?
  - Can it be accurate?

Table of Contents

01 Data

Data sourcing, cleaning, preprocessing.

▶ 3:15min

02 EDA

Looking for trends in music over time, and by genre

▶ 3:15min

03 Building

Fleshing out the recommender

▶ 3:15min

04 Demo

Live action look at recommender

▶ 3:15min

Search

Home

Library

Capstone

01

Data

Foundation for recommendation.

Sourcing

- An accurate recommender requires a lot of data
- Although Spotipy offers an easy to use wrapper, limited time led me to Kaggle
- Close to 600,000 tracks w/ audio features, metadata available for download

Features

- Meta: song name, artists, release date (yr), genre
- Audio: explicit, popularity, danceability, energy, key, loudness, mode, speechiness, acousticness, instrumentalness, liveness, valence, tempo, duration_ms, time signature

Cleaning

- Plenty of data -> dropped nulls and bad entries, fixed data types
- Converted strings to their literals in various columns
- Sorted by popularity (for search purposes), dropped duplicates
- Reduction of ~100,000 tracks to 490,000.

Preprocessing

- Scaled numeric features by z-score for more accurate distance calculations
- Working with cleaned tracks data and scaled tracks data going forward
- For plotting fun, scaled data was also reduced to 2 principal components with PCA

Search

Home

Library

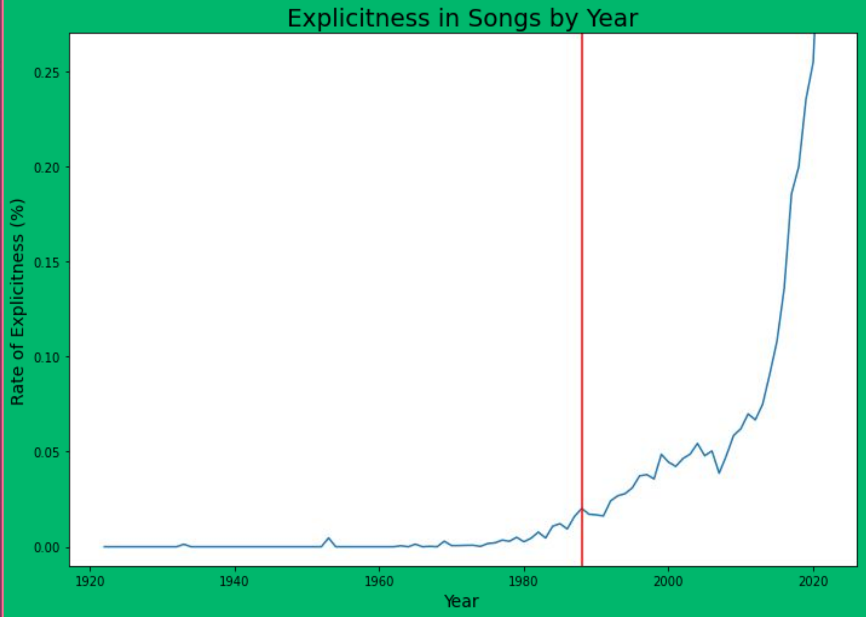
Capstone

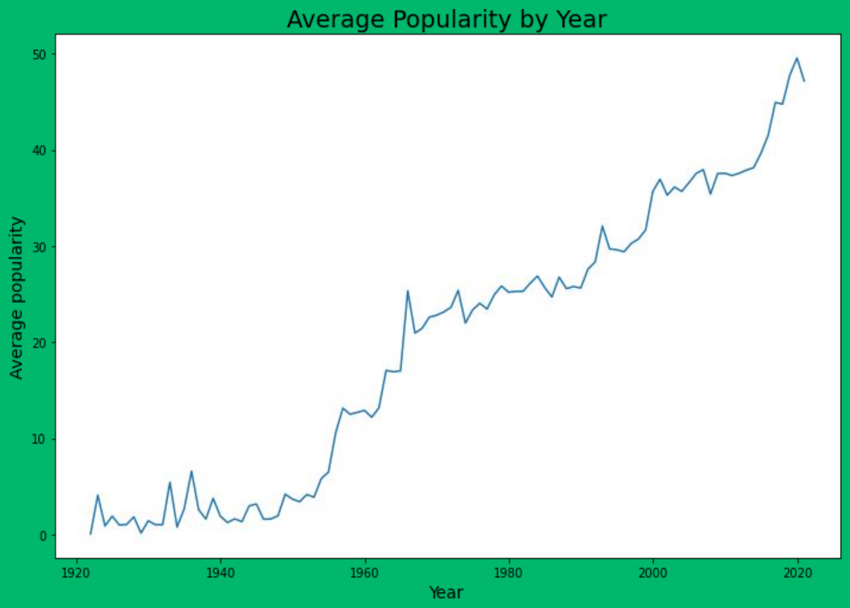
02

EDA

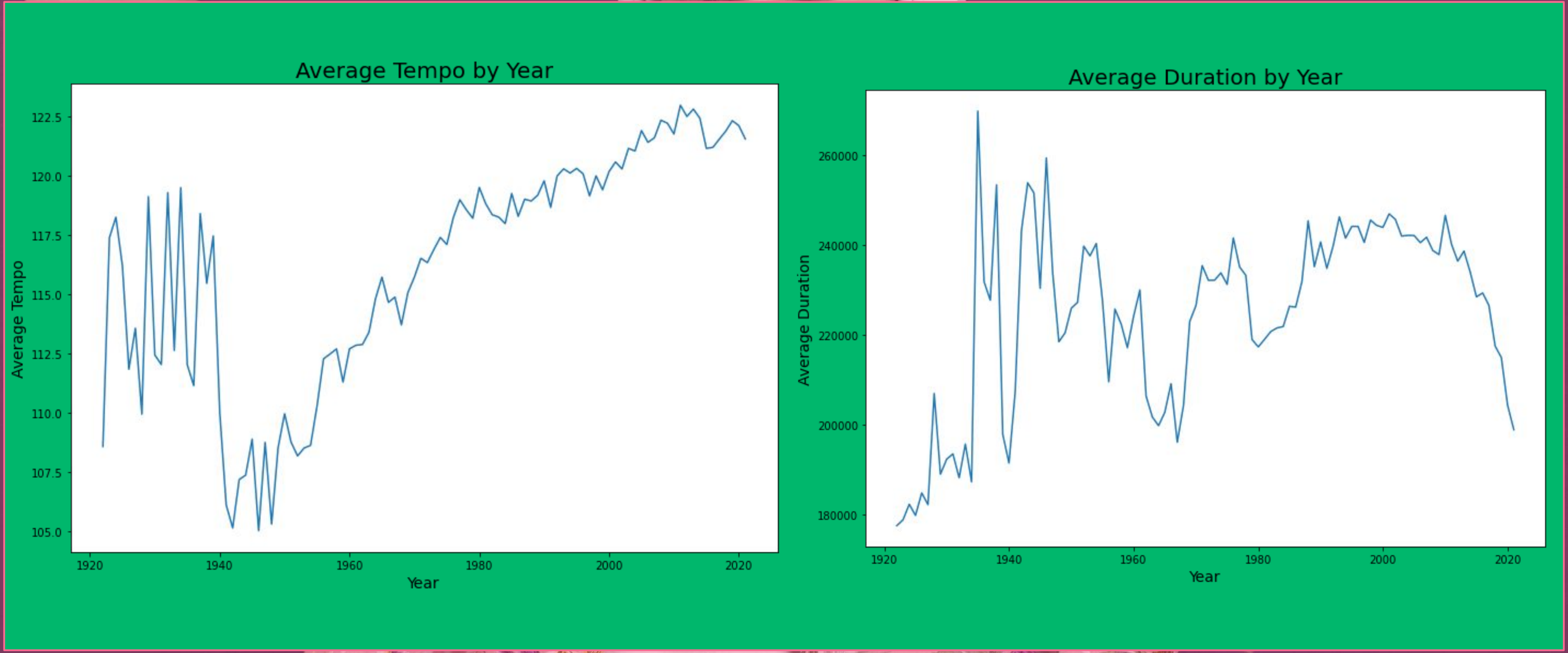
Exploring music the data scientist way.

Music Over Time


Explicitness in Songs by Year

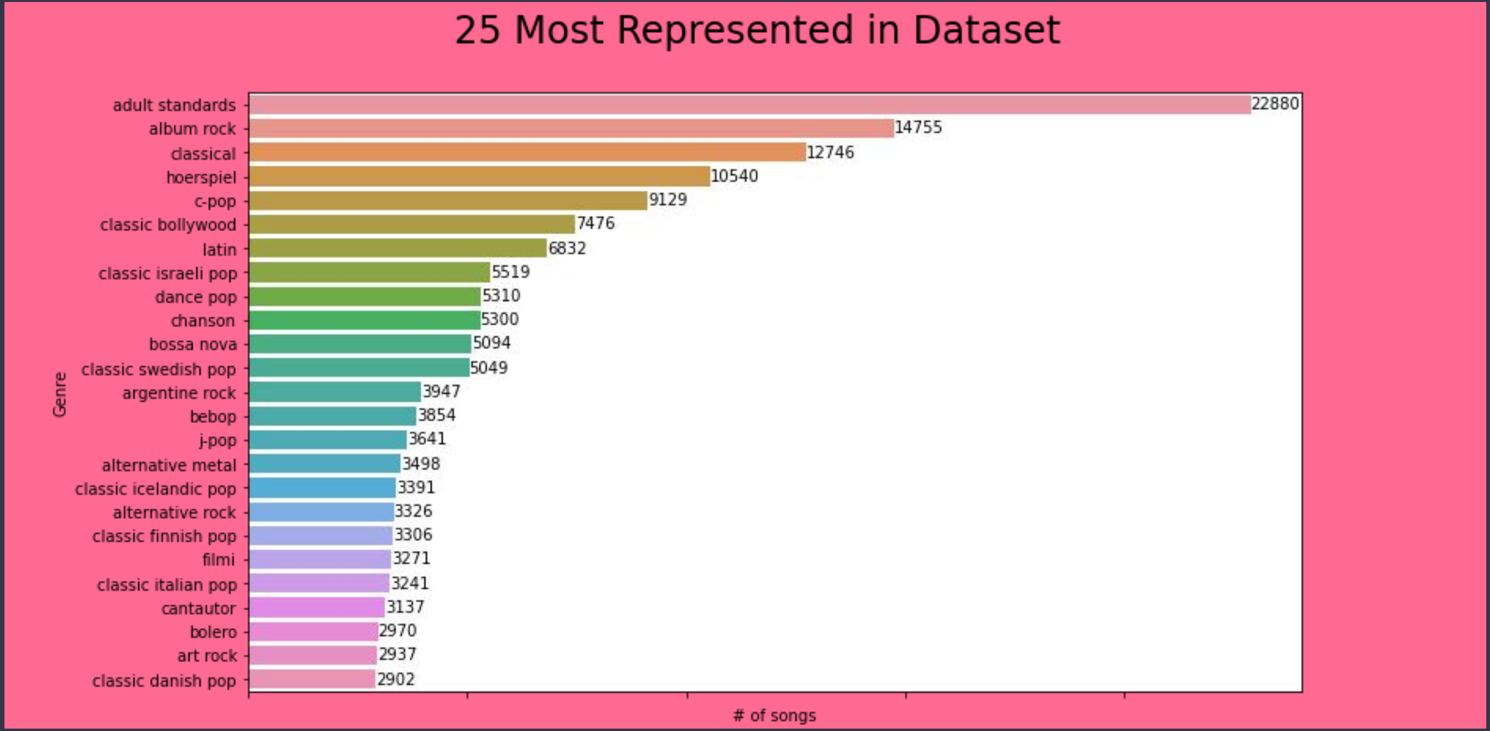

Average Popularity by Year

Music Over Time


Average Tempo by Year

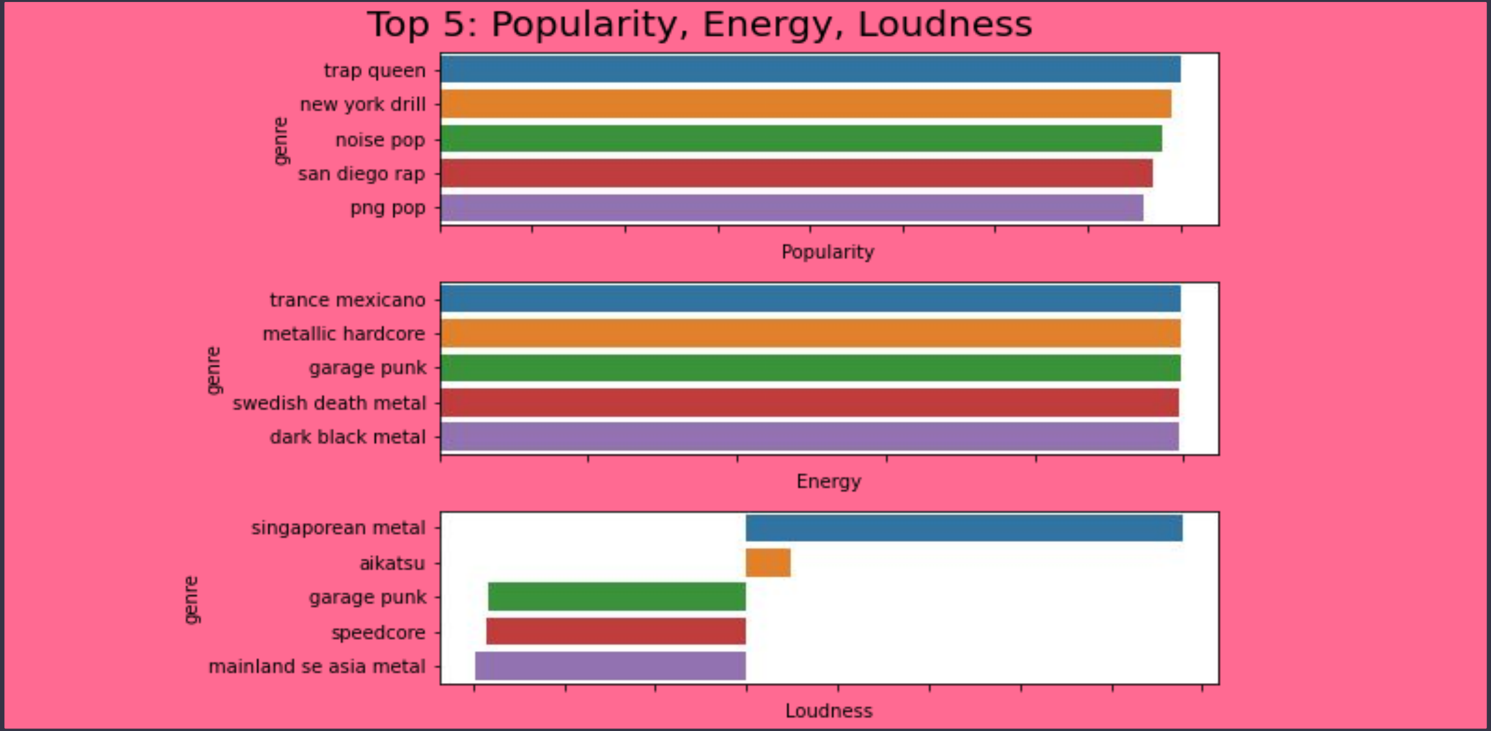
Average Duration by Year

By Genre



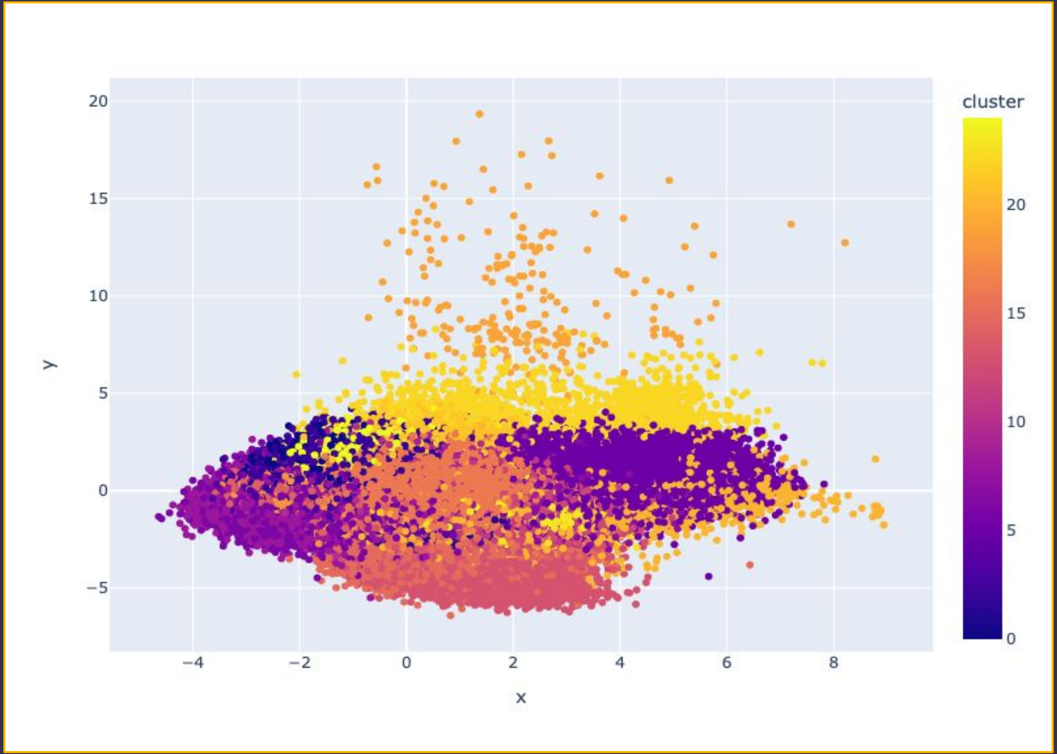
25 Most Represented in Dataset

By Genre



Top 5: Popularity, Energy, Loudness
genre
trap queen
new york drill
noise pop
san diego rap
png pop
Popularity
genre
trance mexicano
metallic hardcore
garage punk
swedish death metal
dark black metal
Energy
genre
singaporean metal
aikatsu
garage punk
speedcore
mainland se asia metal
Loudness

Cluster Visualization



Search

Home

Library

Capstone

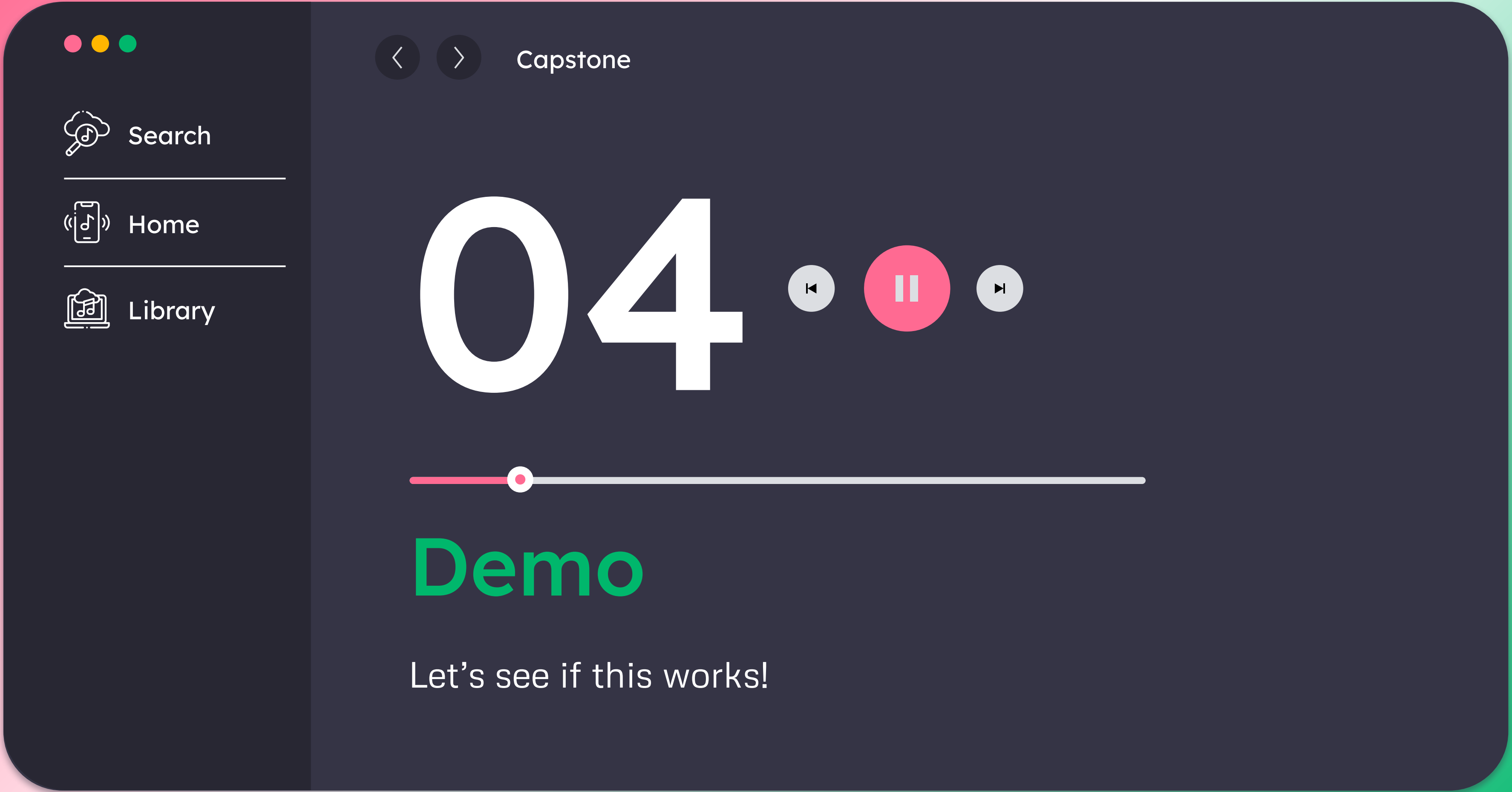
03

Building

How it’s made.

Thought Process

- Given a list of tracks:
    - Search for each in local scaled data. If not available, pull and scale
    - With list of scaled vectors, calculate mean
    - Calculate cosine distances from mean vector to all tracks
    - Sort and grab 10 closest to the mean
- The 10 adjacent tracks should be similar in composition, and thus probably enjoyable



04

Demo

Let’s see if this works!

Conclusion

- Recommendations can be quite accurate, but may also miss.
- Potential shortcomings:
  - Data (amount, representation)
  - Using mean vector, cosine distances
  - Only able to recommend from local
- Very fun to work with and explore, and personalization can be a lucrative avenue.

Search

Home

Library

Capstone

Thanks!

Do you have any questions?

CREDITS: This presentation template was created by **Slidesgo**, and includes icons by **Flaticon** and infographics & images by **Freepik**

Please keep this slide for attribution