Synesthete: A deep learning engine that sees sound

Vivas Kaul
Sprint 3 Deliverable
09/05/2023

Table of Contents

Project Overview

- **Problem Statement:** Recommend similar songs to a user based on a single audio track as an input. The audio can come via a pre-saved audio file or as a recording in real time.

- **Proposed Solution:** Vectorize audio data and calculate cosine similarity. Then present the user with the **top ten** matches based on that similarity. Two models would be created to start with. One would use only pairwise cosine similarity and the other would use a CNN in conjunction with cosine similarity.

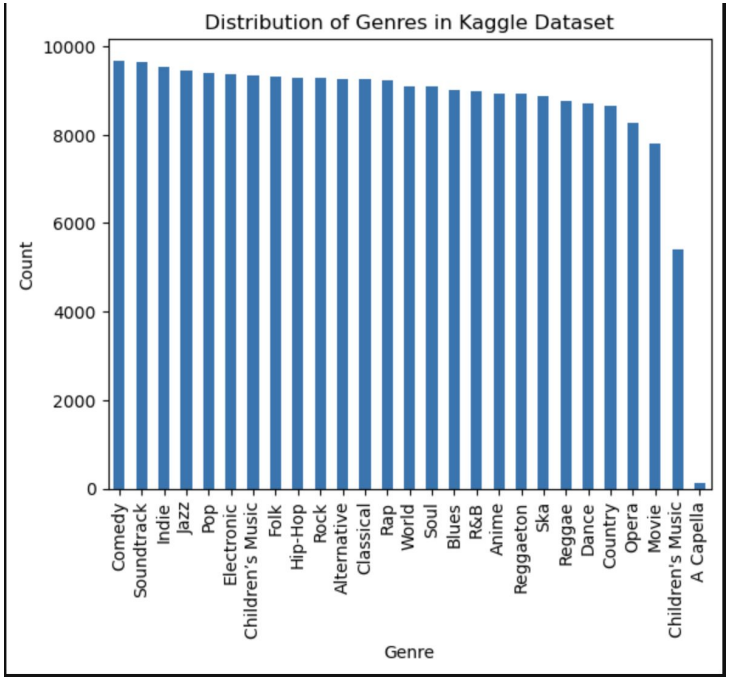
- **Potential Impact:** Provide the user tools to find similar artists via a standalone app or create value-add for existing app.

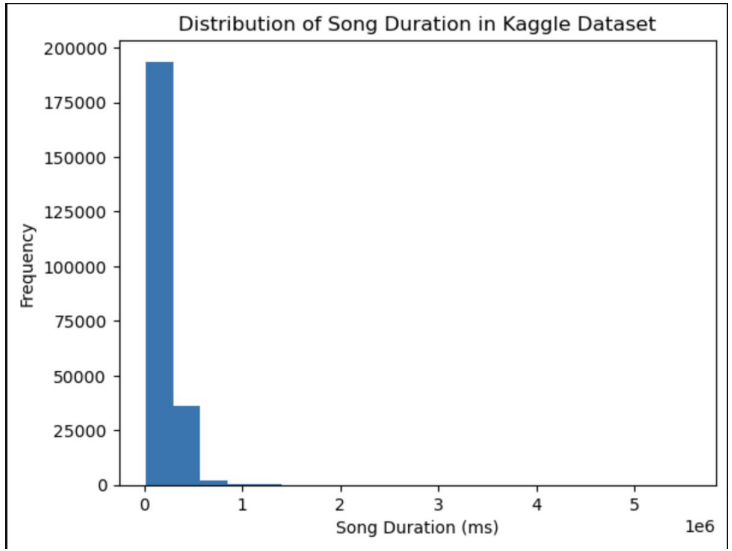
Data Overview and Preprocessing

- Kaggle Data
  - 232,725 rows each representing a song
  - No Nulls - very clean
- Vectorized MP3s
  - 11578
- Encoded MP3s using Librosa package
  - Mel Spectrum
  - Mel Frequency Cepstral Coefficients (MFCC)
  - Chroma Vector
  - Tonnetz Spectrum
- Set track_id to category in order to set hierarchy for ordering track_id when constructing a playlist

Findings from EDA

Kaggle Dataset





Findings from EDA (cont.)

Kaggle Dataset - Duplication of track_id

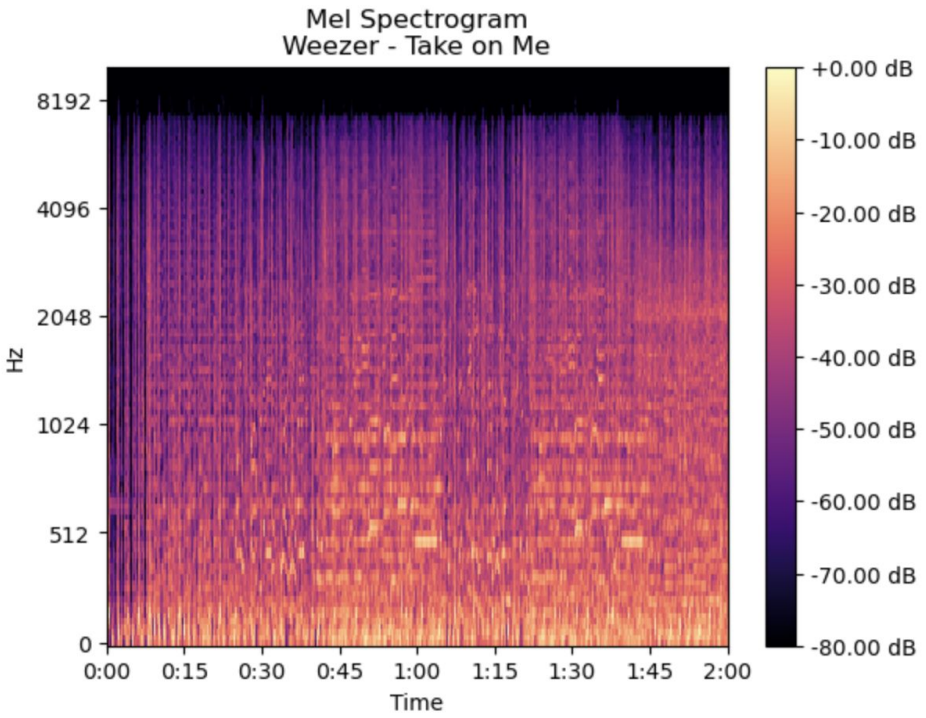
```
kaggle_df.track_id.value_counts(ascending=False).head() # some of the track_ids appear more than once
# This seemed odd considering that the track_id should be a unique identifier.
```

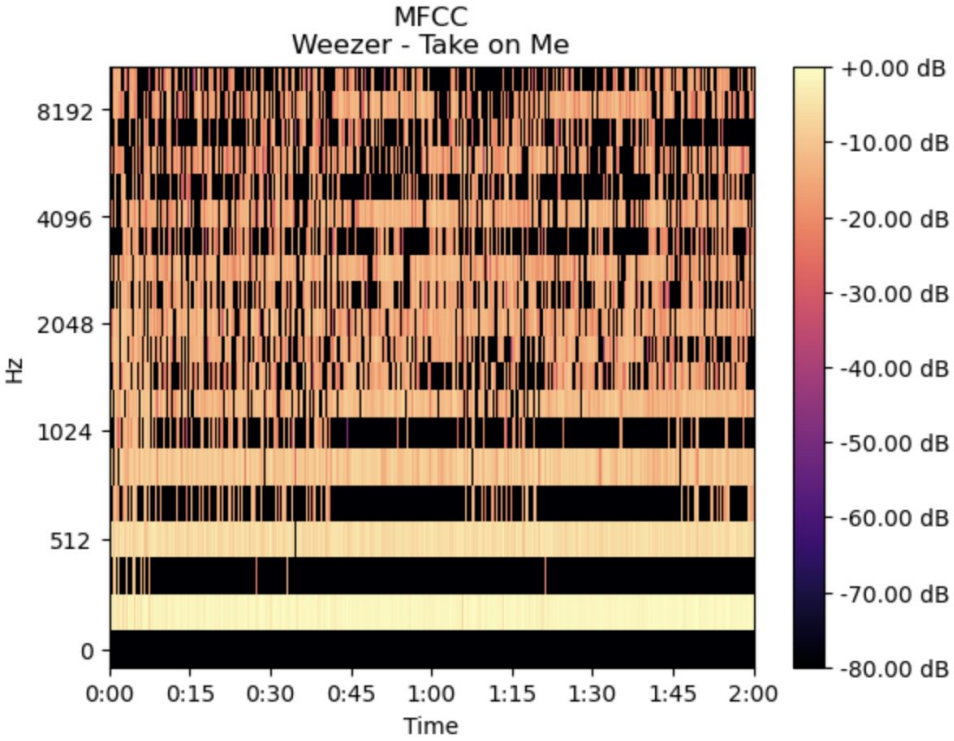
```
3R73Y7X53MIQZWnKloWq5i    8
0wY9rA9fJkuESyYm9uzVK5    8
6sVQNUvcVFTXvlk3ec0ngd    8
0UE0RhnRaEYsiYgXpyLoZc    8
6AIte2Iej1QKlaofpjCzW1    8
Name: track_id, dtype: int64
```

```
kaggle_df.loc[kaggle_df.track_id == "3R73Y7X53MIQZWnKloWq5i"]
# This shows that a single track id does correspond to a single song, but that the song can repeat within the data set
# based on the genre. All of the other columns are the same though.
```

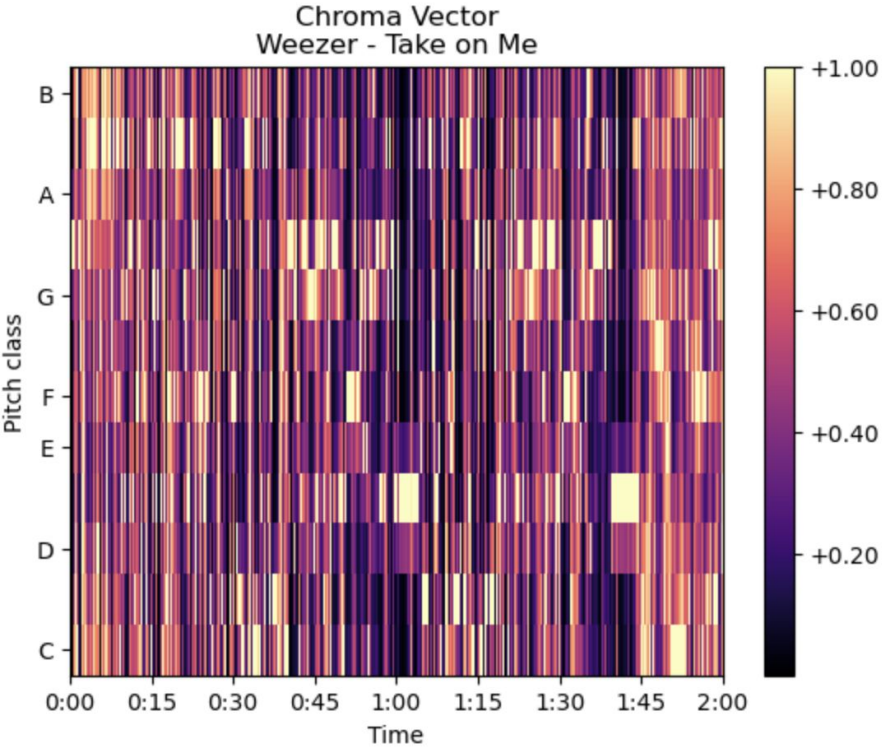
	genre	artist_name	track_name	track_id	popularity	acousticness	danceability	duration_ms	energy	instrumentalness	key	livene
5506	Alternative	Toro y Moi	Monte Carlo (feat. WET)	3R73Y7X53MIQZWnKloWq5i	50	0.08	0.65	125133	0.38	0.00	C#	0.
15615	Dance	Toro y Moi	Monte Carlo (feat. WET)	3R73Y7X53MIQZWnKloWq5i	61	0.08	0.65	125133	0.38	0.00	C#	0.
41367	Folk	Toro y Moi	Monte Carlo (feat. WET)	3R73Y7X53MIQZWnKloWq5i	61	0.08	0.65	125133	0.38	0.00	C#	0.
55106	R&B	Toro y Moi	Monte Carlo (feat. WET)	3R73Y7X53MIQZWnKloWq5i	61	0.08	0.65	125133	0.38	0.00	C#	0.
77769	Children's Music	Toro y Moi	Monte Carlo (feat. WET)	3R73Y7X53MIQZWnKloWq5i	61	0.08	0.65	125133	0.38	0.00	C#	0.
94721	Indie	Toro y Moi	Monte Carlo (feat. WET)	3R73Y7X53MIQZWnKloWq5i	61	0.08	0.65	125133	0.38	0.00	C#	0.
149585	Pop	Toro y Moi	Monte Carlo (feat. WET)	3R73Y7X53MIQZWnKloWq5i	61	0.08	0.65	125133	0.38	0.00	C#	0.
225238	Rock	Toro y Moi	Monte Carlo (feat. WET)	3R73Y7X53MIQZWnKloWq5i	61	0.08	0.65	125133	0.38	0.00	C#	0.

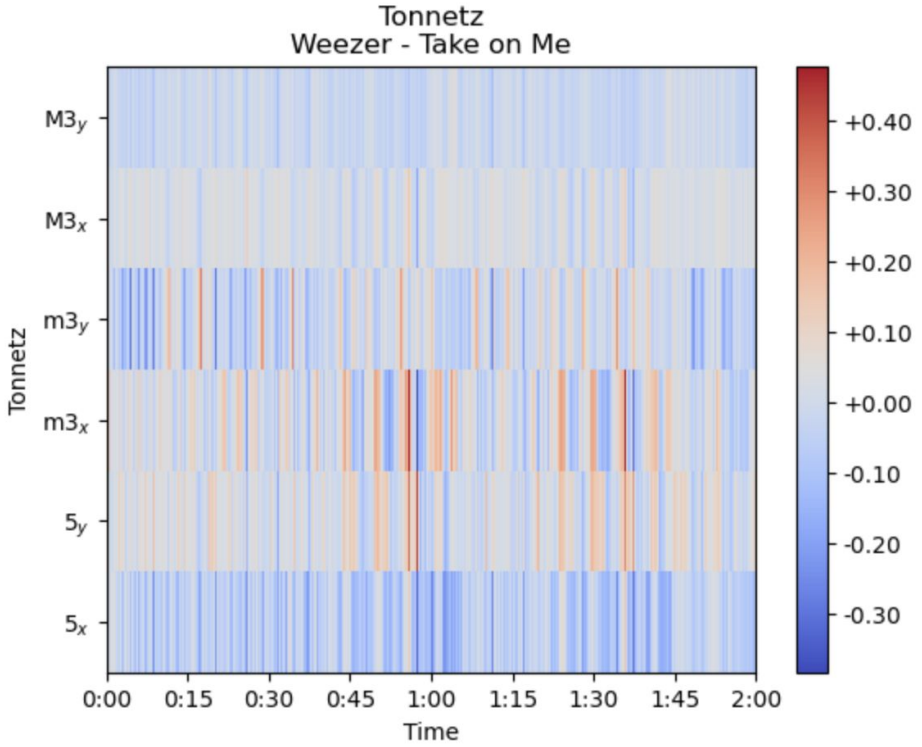
Findings from EDA (cont.)



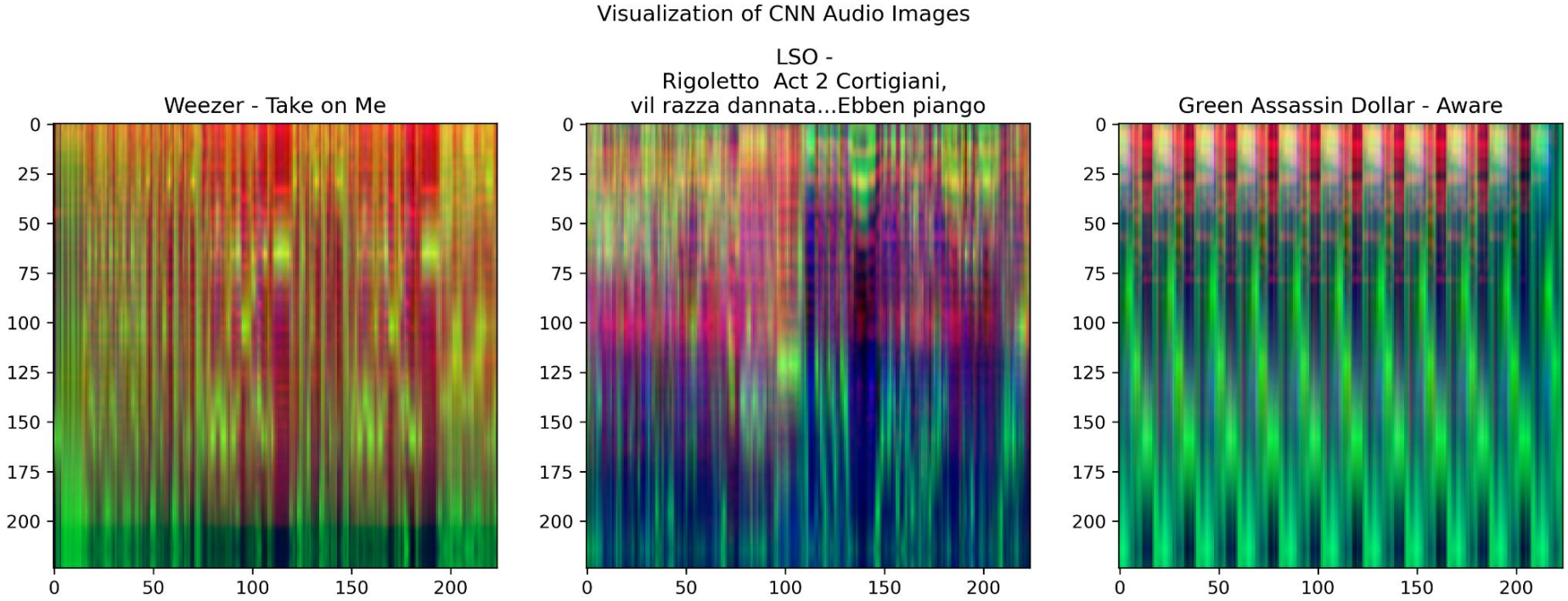


Findings from EDA (cont.)

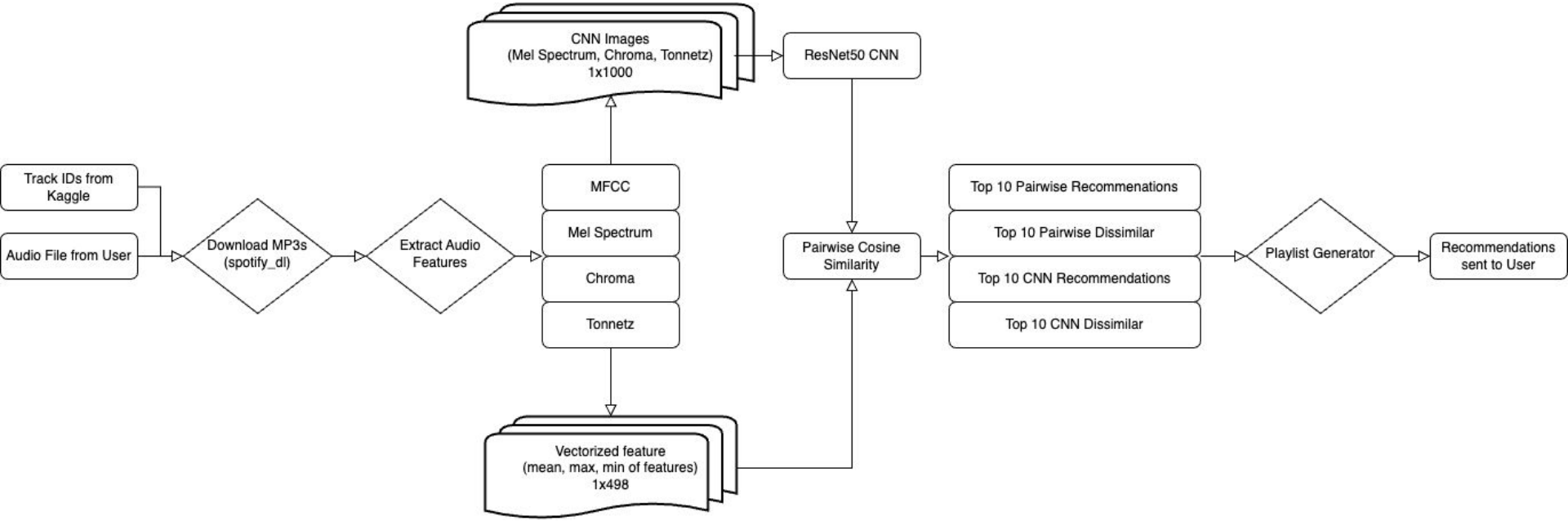




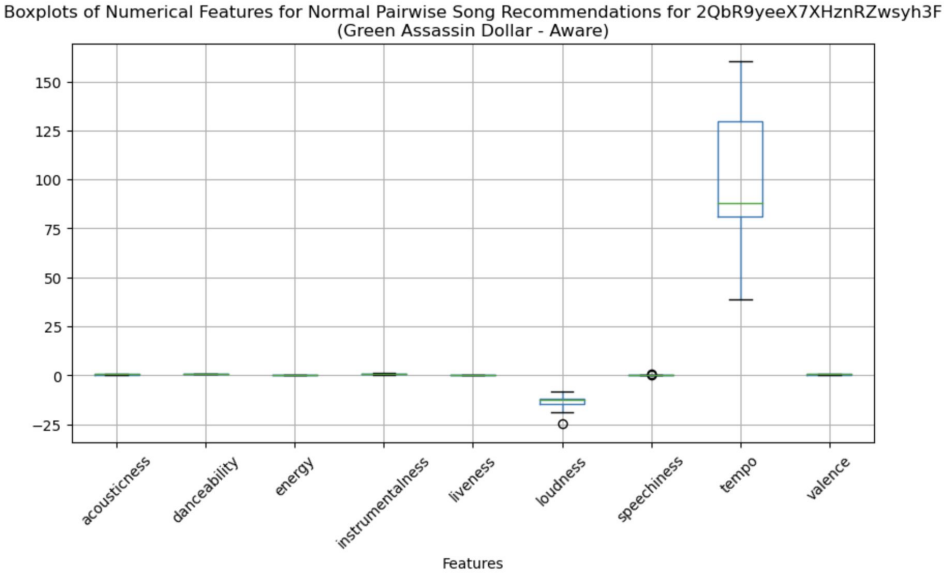
Findings from EDA (cont.)

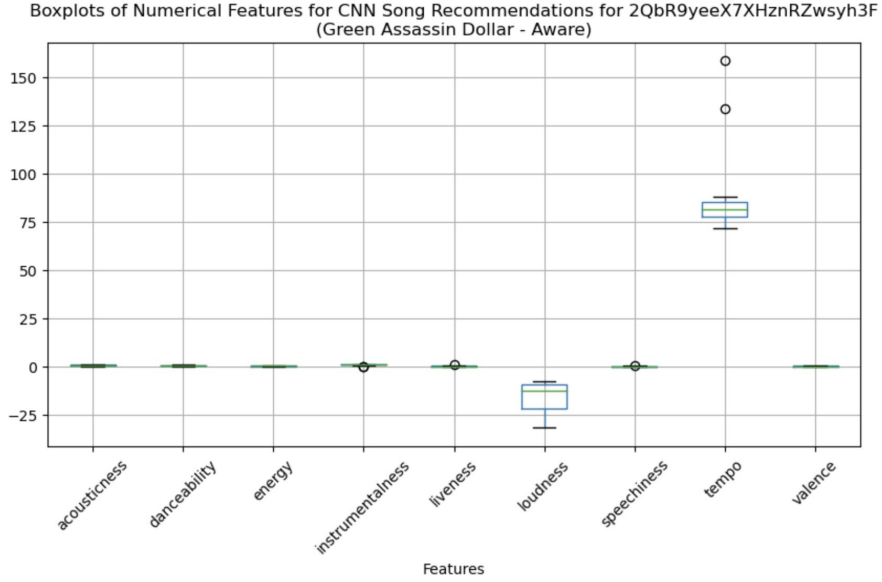


Modeling Procedure

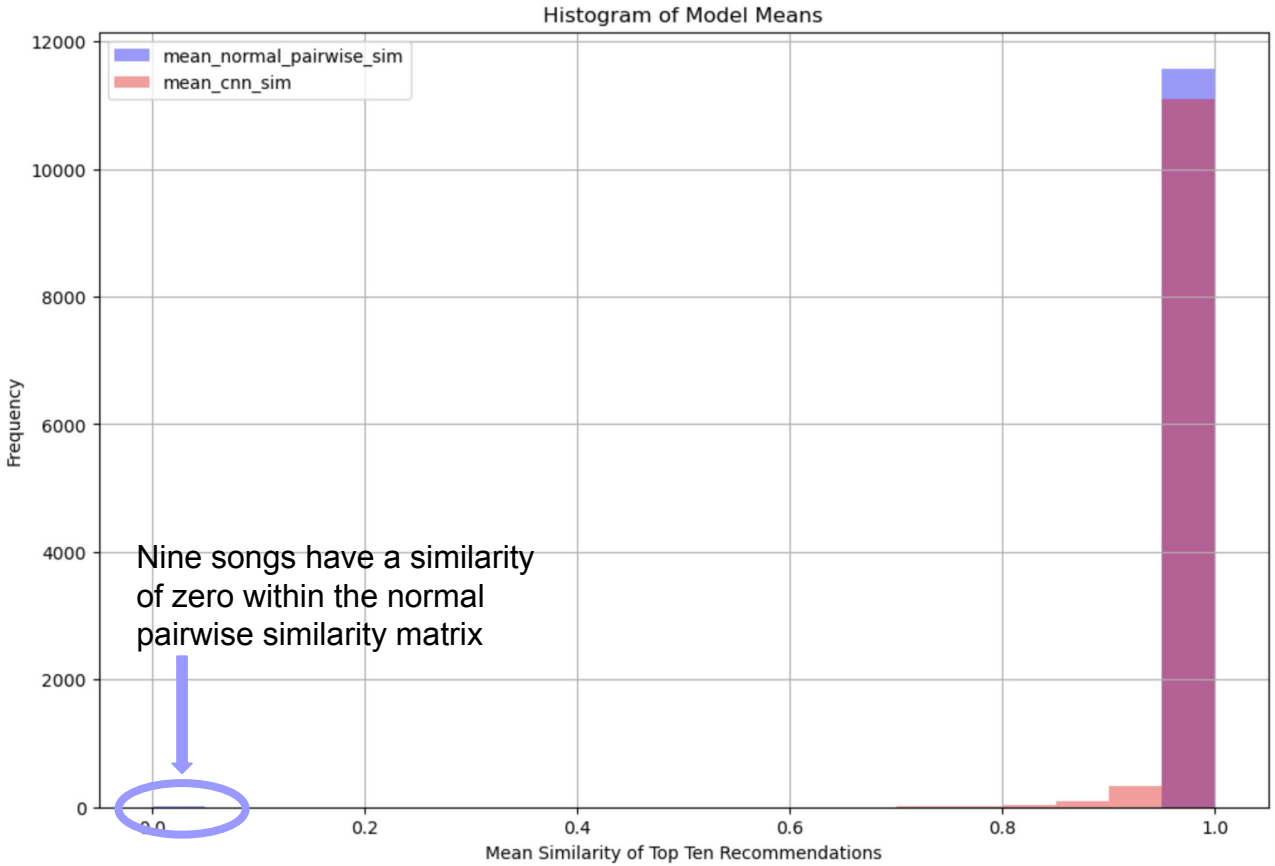


Model Comparison and Interpretation

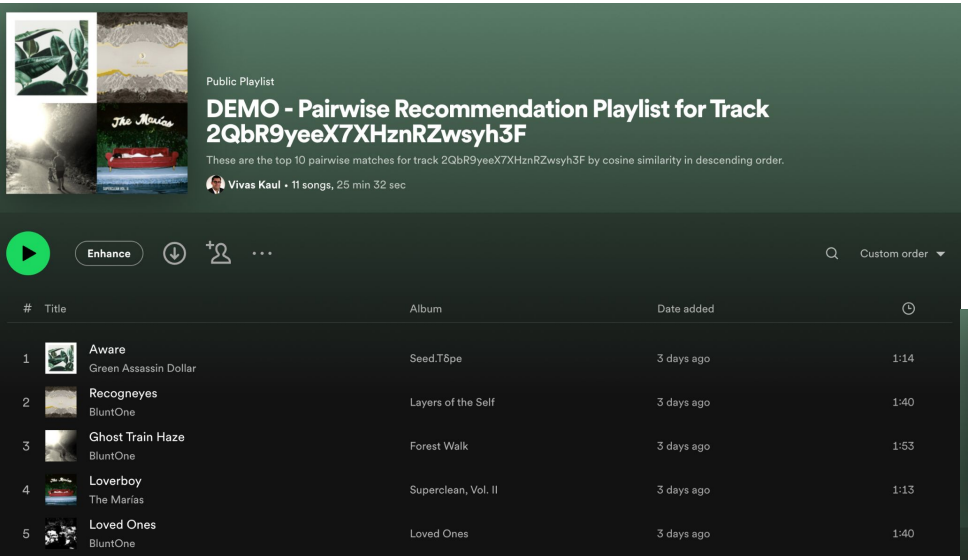


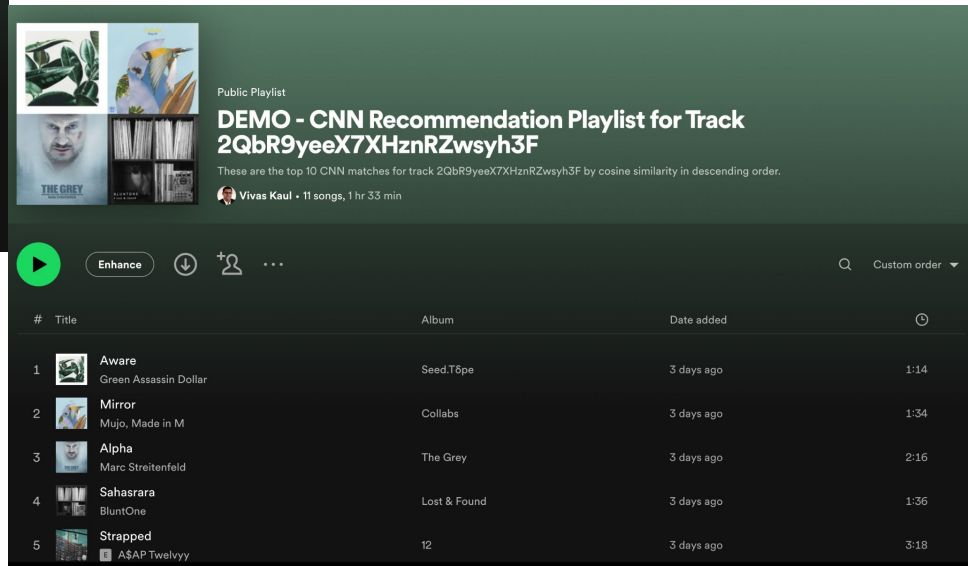


Model Comparison and Interpretation (cont.)



Model Comparison and Interpretation (cont.)





App Design

- Streamlit/Flask App
  - Return playlists (similar or dissimilar, normal or CNN)
    - This could be done via a post request or as a dashboard in streamlit using a search
  - Song/Artist search
    - Same as above, this could be setup to use a built-in string function or regex
    - User could search for a song within the database and return a set of tracks
    - Artist search let the user select a track before results are returned
- Backend
  - Similarity data
  - Sorted similarity data
  - Kaggle data
    - Subset to training tracks

Future Improvements

- Determine Evaluation Metric for Model
  - Spotify recommendation engine via API
  - Source song-recommendation playlist co-occurrence (Market Basket Analysis)
- Testing Different Models
  - Clustering (DBSCAN, K-Means)
  - Alterations to CNN
    - Layers and Neurons
  - Reinforcement Learning
  - Conversion to supervised learning problem via user feedback
- Productionize Models
  - Streamlit/Flask App
  - Containerize and mount in Cloud
    - AWS - EC2 and ECR

Questions?