BUILDING A BETTER PLAYLIST

Kelly O'Neill
Data Scientist / Chief Entertainment Officer

Goldman and Associates

PROBLEM STATEMENT

- Create better alternatives to Spotify's premade playlists
- Mounting frustration over suboptimal playlists
- Classify songs of a given genre into four moods
- Generate playlists in Spotify based on mood

01

AQUIRE/CLEAN DATA

02

DETERMINE BEST MODEL

03

INTERPRET RESULTS

04

MAKE SONG RECOMMENDATIONS

01

DATA ACQUISITION

- Artists used in MTV's 10,000 top artists
- Genre was provided
- Easier to subset data based on the types of songs I wanted to recommend
- Alternative/Indie etc.
- Scrape Spotify designated song features
- Started with all genres, narrowed to just alternative for this iteration

GENRES USED

ALTERNATIVE

INDIE ROCK

ALT ROCK

FOLK/ROCK

WHY THESE?

1. THEY'RE MY PERSONAL FAVORITE GENRES
2. TREATING ALL GENRES THE SAME WAY DOESN'T MAKE SENSE

02.A

DATA CLEANING/EDA

- Was able to bypass values that didn't exist in Spotify with try/except statements
- Some nulls, dropped them (mostly artists that did not have their discographies in Spotify)
- Only had to drop ~100 songs

02.B

DATA CLEANING/EDA

- Visualize feature distributions
- Drop outliers
- Ultimately focus on valence, danceability, energy, tempo - other features added noise

03
MODELING

- DECIDE ON MODEL
- FEATURE SELECTION
- FIT MODEL
- MAKE INTERPRETATIONS

03
MODEL SELECTION

KERAS

- Could use existing classifications
- Easier to interpret
- Dataset was too small ~700 rows
- Any errors would be exponentiated

KMEANS

- Specify number of clusters (moods)
- Interpretation more difficult
- Ultimately more accurate

FEATURE SELECTION

VALENCE

ENERGY

DANCEABILITY

TEMPO (BPM)

POPULARITY

LOUDNESS (DB)

LIVENESS

ACOUSTICNESS

SPEECHINESS

INSTRUMENTAL
NESS

MODEL RESULTS

24.3%

Silhouette Score using Four Clusters

67,100

Number of songs classified

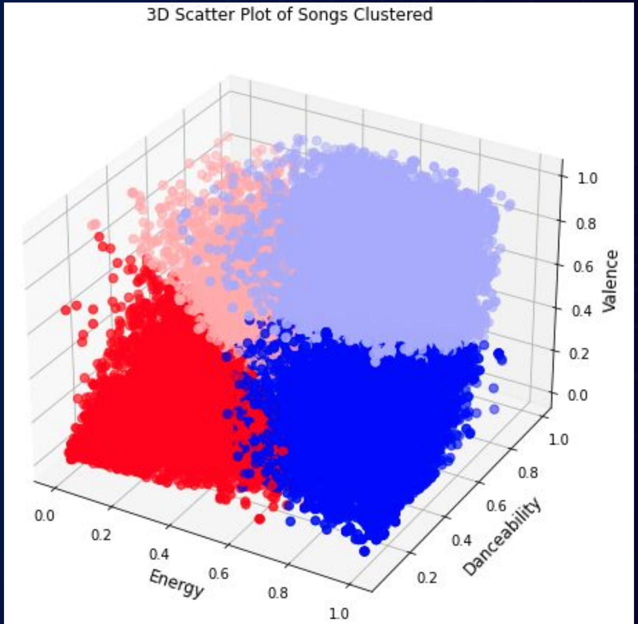
WHAT DOES THAT MEAN?

SCORE

- Score falls between -1 and 1
- Closer to 1: nicely separated
- 0: overlap
- -1: incorrect

CLUSTERS

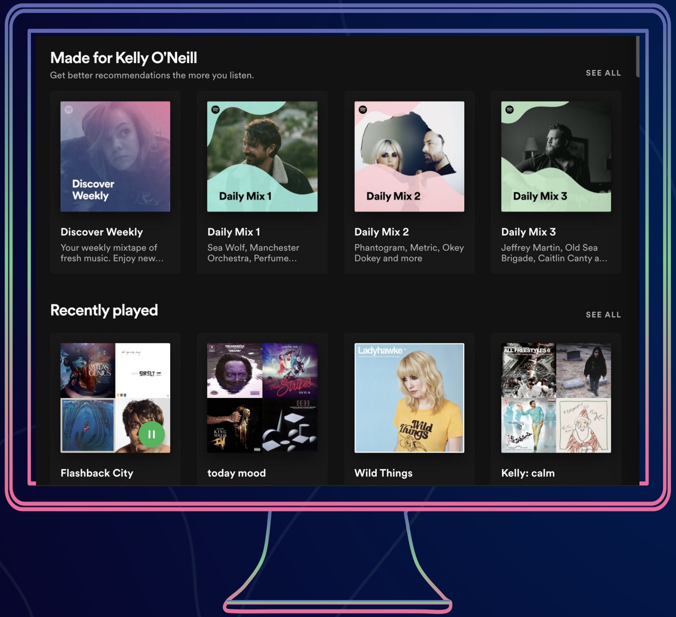
- Four clusters to classify songs into four moods
- Happy, sad, energetic, calm



MODEL INTERPRETATION

- Score was satisfactory
- Labels provided by model were 0, 1, 2, 3
- Separated songs in to four separate dataframes
- Analyzed median values of top four features
- Determined moods:
    - 0 = sad
    - 1 = calm
    - 2 = energetic
    - 3 = happy

PLAYLIST CREATION



- Python function takes user input for name, mood, and playlist title
- Generates song ids
- Creates empty playlist in user’s account and fills with song ids
- Currently only available on my account due to app permissions

NEXT STEPS

- Add additional clusters to provide more accurate and granular mood determinations AND
- Use separate model for each genre category (Rock, rap, pop, etc)
- Use additional genres
- Flask app for playlist maker
- Ability to create playlists for other users

SOURCES

MTV artist dataset
Spotify documentation
Spotipy documentation
Clustering algorithm research
Silhouette score analysis for num_clusters
3D Scatterplot
Additional reading on KMeans

THANK YOU!

Any questions?