FindMe FM

Final Project: Group 3

Dorna Abdi
Manuel Elizadi
Christian Koehlinger
Cailyn Miller
Mike Prue
Joshua Rhodes

Project Outline

Topic: Can machine learning predict what songs a user will enjoy based on the audio features of a song they like?

Reason for topic: Interest in the use of Spotify API and other available spotify datasources.

Data Source: Kaggle dataset - Spotify Dataset 1922-2021 ~600k tracks

- Contains info on the audio features of each song (danceability, acousticness, tempo, etc,)
- Dataset is created using the Spotify API

Data Structure: Tracks

Primary:

- ID

Numerical

- acousticness (ranges from 0 to 1)
- danceability (ranges from 0 to 1)
- energy (ranges from 0 to 1)
- duration_ms (ranges from 0 to 1)
- instrumentalness (ranges from 0 to 1)
- valence (ranges from 0 to 1)
- popularity (ranges from 0 to 1)
- tempo (ranges from 0 to 1)
- liveness (ranges from 0 to 1)
- loudness (ranges from 0 to 1)
- speechiness (ranges from 0 to 1)

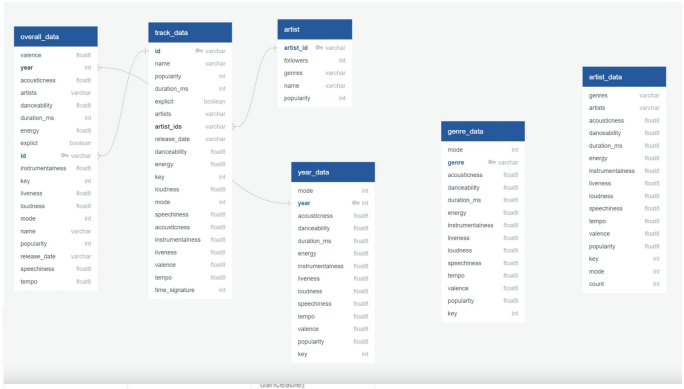
Why Song Prediction?

Music is relevant to our everyday lives and as a group we are all interested in the subject.

We chose Spotify because, after YouTube, Spotify is the most popular music streaming service and a trailblazer in connecting users to their listening data.

Spotify's API includes quantifiable data on specific audio features which we were eager to explore!

Descriptions of the data exploration phase of the project



id id of track	name name of track	popularity popularity of track				
586672 unique values	446475 unique values	0 100	114030 unique values	19700 unique values	0 0.99	0 1
35iwgR4jXetI318WEWsa1Q	Carve	6	['Uli']	1922-02-22	0.645	0.445
021ht4sdgPcrDgSk7JTbKY	Capitulo 2.16 - Banquero Anarquista	0	['Fernando Pessoa']	1922-06-01	0.695	0.263
07A5yehtSnoedViJAZkNnc	Vivo para Quererte - Remasterizado	0	['Ignacio Corsini']	1922-03-21	0.434	0.177
08FmqUhxtyLTn6pAh6bk45	El Prisionero - Remasterizado	0	['Ignacio Corsini']	1922-03-21	0.321	0.0946
08y9GfoqCWfOGsKdwojr5e	Lady of the Evening	0	['Dick Haymes']	1922	0.402	0.158
0BRXJHRNGQ3W4v9frnSfhu	Ave Maria	0	['Dick Haymes']	1922	0.227	0.261

Created a mapping between the different data sources

Explored the datatypes

FIND ME

Technologies, languages, tools, and algorithms

Heroku: Cloud platform being used to store all of data and website.

pgAdmin: Platform being used to store the database.

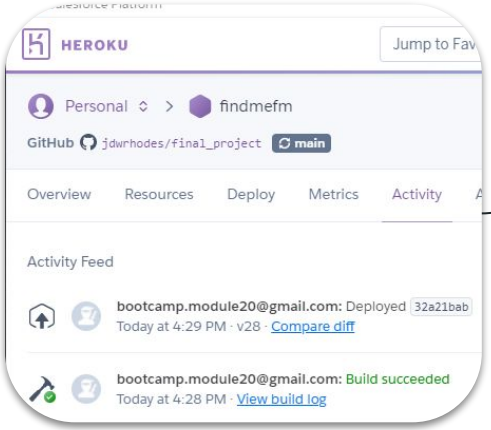
Flask: Module being used to create the website.

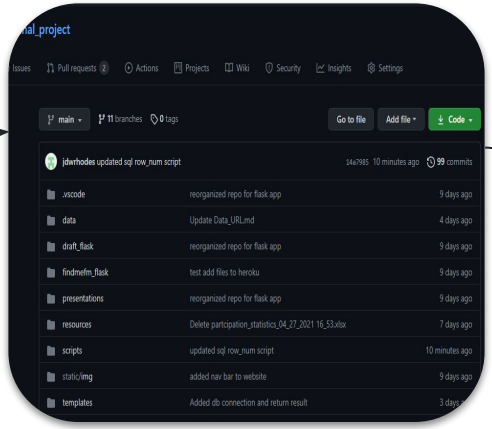
Tableau: Platform being used to visualize the data analysis.

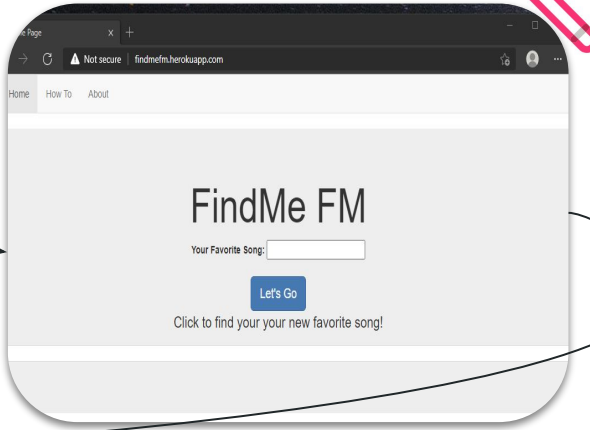
SQL: Language being used to query and restructure the database.

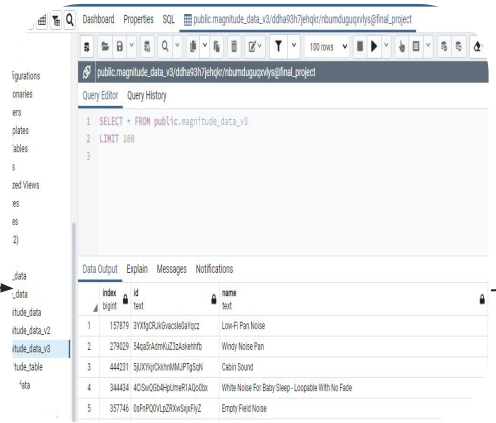
Python: Language being used to clean the data and run Machine Learning algorithms.

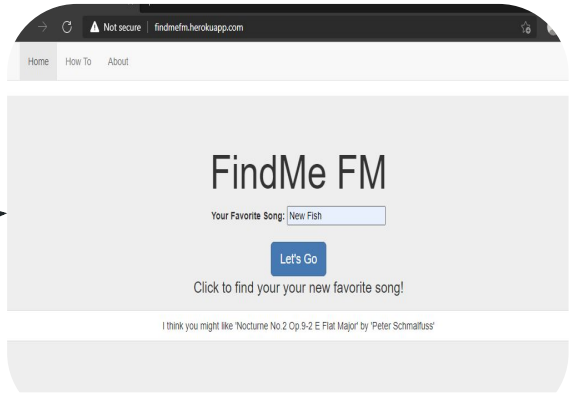
So, how does this all work?











Website Demo

Analysis Results

PCA Analysis

Z-Scores

Magnitude Table

Adjusting Coefficients

Control Group

Description of the analysis phase of the project

acousticness	danceability	energy	instrumentalness	liveness	loudness	popularity	speechiness	tempo
1.597267	-1.402608	-0.434803	0.527065	0.681875	-2.560544	0.401476	-0.384211	0.481201
-0.026886	-0.362256	0.666427	-0.519699	0.799534	-1.552361	1.193404	-0.071579	0.280969
3.212922	-1.633991	0.005412	1.118260	0.249359	-1.904048	-0.143154	1.033555	0.752712
-1.105168	-1.309363	-0.297432	0.131040	0.100094	-0.986246	-0.980184	-1.432502	0.749119
-1.037801	-0.886208	1.048483	1.097711	0.860369	-0.453503	0.875575	-1.344493	1.213830

We ran a PCA analysis to understand which audio elements had the greatest variability.

Magnitude/Z-Scores

Recommendations for Future Analysis

- Add sliders on the user interface for users to customize their recommendations
- Recommend artists/albums based on audio features
- “Fuzzy Search”
- Add a “Did you like this recommendation” button for future refining

What We Would Have Done Differently

Use the actual Spotify API

Started testing features sooner

Nail down the magnitude formula and refine it

Analysis and testing of data source and the validity of each column (valence, danceability)

Use Regex to limit characters from different languages

Results

At first, the recommendation results did not really line-up

Then, we refined it by normalizing by z-scoring and continued to test the results

Once we adjusted certain magnitudes, the song recommendations came out correct!

Check it out for yourself!!

findmefm.herokuapp.com/