

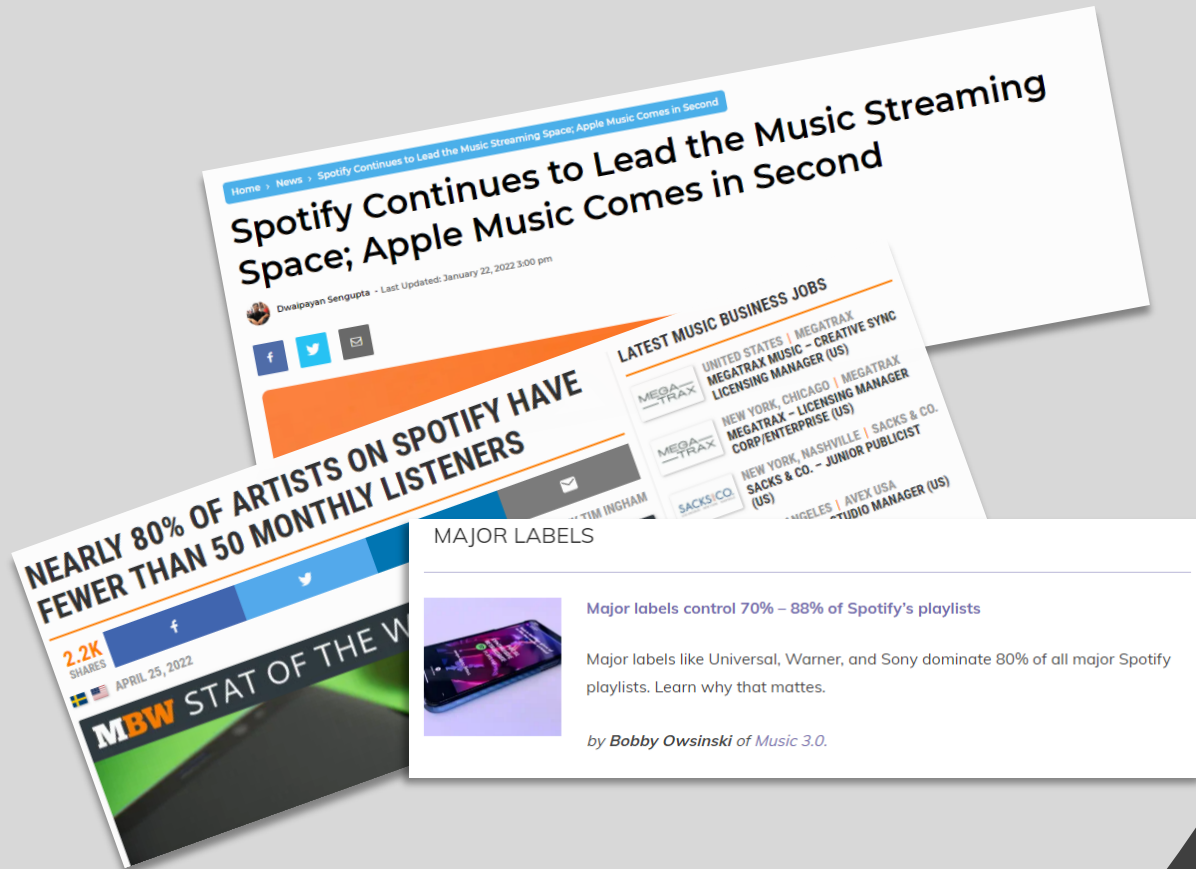


Mystery Box

Find the answers.

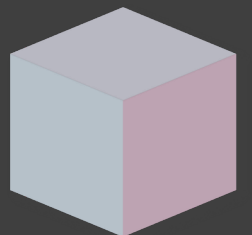
Mystery Box aims to use machine learning to take the guess work out of finding Spotify playlists and artist similar to you.

What answers are we trying to find?



How do I break into the Spotify Top Tier?

- Nearly **80%** of artists on Spotify have less than **50** monthly listeners
- Major labels control **70% – 88%** of Spotify's playlists - hypebot.com
- Spotify holds a **31 percent market share** in the music streaming industry. Apple Music, follows in second place with only a 15 percent market share. - BaseBorn.com



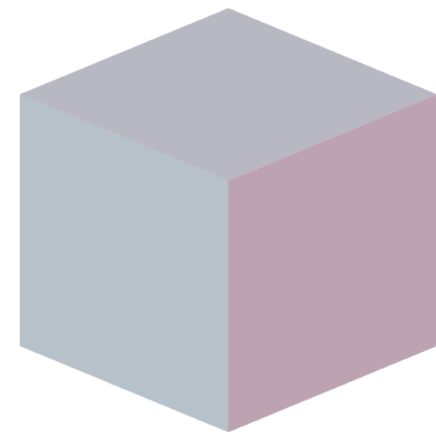
So what's in the **Box**?





Flask

web development,
one drop at a time



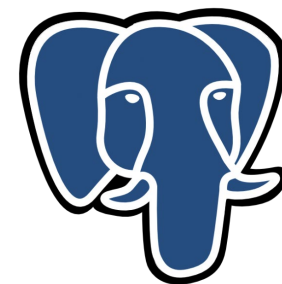
MYSTRY BOX

aws



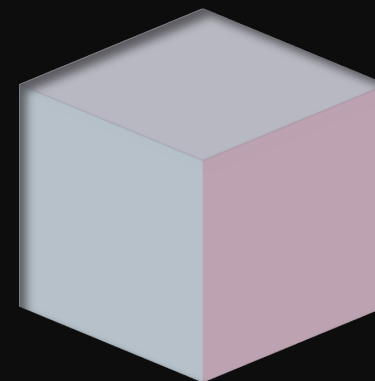
Mystery Box technologies

- **Easy to deploy**
- **Easy to customize**
- **Always learning**



PostgreSQL

How to Find the answers.



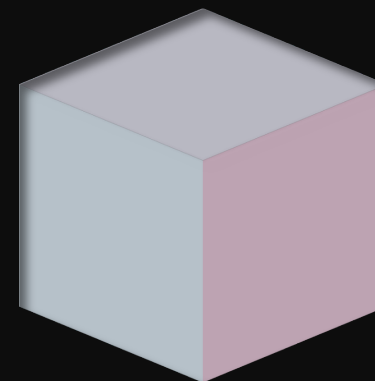
Danceability describes how suitable a track is for dancing based on a combination of musical elements including tempo, rhythm stability, beat strength, and overall regularity. A value of 0.0 is least danceable and 1.0 is most danceable.

Energy is a measure from 0.0 to 1.0 and represents a perceptual measure of intensity and activity. Typically, energetic tracks feel fast, loud, and noisy. For example, death metal has high energy, while a Bach prelude scores low on the scale. Perceptual features contributing to this attribute include dynamic range, perceived loudness, timbre, onset rate, and general entropy.

Valence is a measure from 0.0 to 1.0 describing the musical positiveness conveyed by a track. Tracks with high valence sound more positive (e.g. happy, cheerful, euphoric), while tracks with low valence sound more negative (e.g. sad, depressed, angry).

Speechiness detects the presence of spoken words in a track. The more exclusively speech-like the recording (e.g. talk show, audio book, poetry), the closer to 1.0 the attribute value. Values above 0.66 describe tracks that are probably made entirely of spoken words. Values between 0.33 and 0.66 describe tracks that may contain both music and speech, either in sections or layered, including such cases as rap music. Values below 0.33 most likely represent music and other non-speech-like tracks.

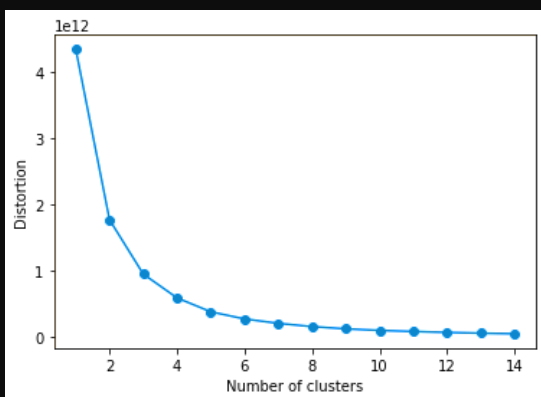
How to Find the answers.



MYSTRY BOX

	one_playlist	two_playlist
0	0.279503	0.378882

- A way around this accuracy issue was to provide the user with two playlist options



- Used **Kmeans** to create an extra feature based off clustering. This was meant to simulate 'genres' or 'moods' at an extremely rudimentary level.

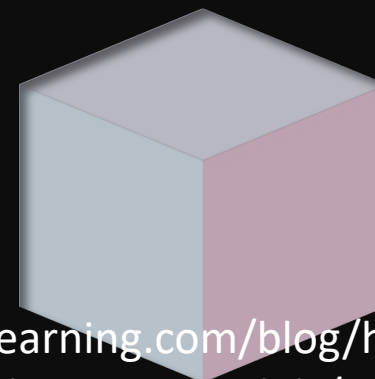
- Used the **Elbow Method** to find the best number of clusters to used

	test_type	lr	knn	cart	randf	grad	n_classes
0	wo_stacking	0.638519	0.632593	0.638148	0.729630	0.700000	2
1	w_stacking	0.638519	0.632593	0.644444	0.743333	0.713704	2
0	wo_stacking	0.403333	0.325833	0.516667	0.564167	0.561667	10
1	w_stacking	0.403333	0.325833	0.519167	0.562500	0.555833	10
0	wo_stacking	0.219813	0.112844	0.157685	0.229442	0.203666	24
1	w_stacking	0.219813	0.112844	0.163245	0.230846	0.206418	24

- Testing stacking vs no stacking. Here you see results using **cross_val_score**

- The larger the number of targets to predict, the lower the accuracy score

How Spotify does it



<https://www.mygreatlearning.com/blog/how-spotify-uses-machine-learning-models/>

- **1. Collaborative Filtering**

- On Spotify, the collaborative filtering algorithm compares multiple user-created playlists that have the songs that users have listened to. The algorithm then combs those playlists to look at other songs that appear in the playlists and recommends those songs.

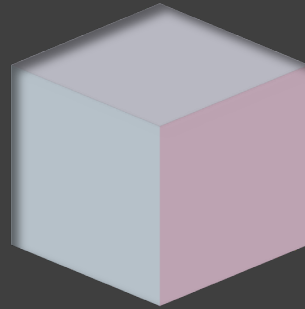
2. Natural Language Processing

- Spotify's NLP constantly trawls the web to find articles, blog posts, or any other text about music, to come up with a profile for each song.

3. Convolutional Neural Networks

- Each song is converted into a raw audio file as a waveform. These waveforms are processed by the CNN and is assigned key parameters such as beats per minute, loudness, major/minor key and so on. Spotify then tries to match similar songs that have the same parameters as the songs their listeners like listening to.

How to
Find the
answers.



MYSTRY BOX

Mystry Box database

29000+ songs

400+ playlists

1400+ song features

 **Spotify** for Developers



aws



PostgreSQL



Flask
web development,
one drop at a time

Mystery Box

Future Updates.

- **Increase accuracy by training on a larger dataset**
- **Provide more insight into the song metrics**
- **Try and use clustering to create mood tags such as 'chill', 'energetic', etc..**

Mystery Box

Some good reads.

- **How Spotify Does It**
- **Predicting my mood using my spotify data**
- **Hyperparameter tuning RandomForest**
- **Stacking Classifiers**