Predicting Country Chart Success

October 17, 2019

By Kaylin Bugos

Overview

- Country radio: most popular radio format in the U.S.
- More than 2,100 stations across the country.
- This project seeks to predict how successful a single will be on country radio, using audio analysis of the song as the predictive variables.

"If you aren't on country radio, you don't exist"

- *Sony Nashville CEO for The Tennessean*

Data Sources

Billboard Charts

The target variable for this project is peak performance on the Billboard Country Airplay Chart, accessed using the Billboard Python library.

Spotify

The predictive variables for this project are the audio features data available through the Spotify API, accessed using the Spotipy Library.

Manual Additions

Gender and solo artists/groups were manually coded, allowing this information to serve as predictive variables as well.

Accessing the Data

Billboard Country Airplay Chart

- 5 years of chart data
- Data imported:
  - Song title
  - Artist name
  - Rank
  - Chart date
- Removed duplicate songs, keeping only the peak chart position.

Preparing the Data for Matching

- Billboard Charts no longer include Spotify links, making data matching more difficult
- Kept only primary artists for matching purposes
- Not all songs available on Spotify

Spotify Matching

- **Matched chart data to Spotify's unique identifier, called the Track ID, by querying their API**
- **Next, used the Track ID to bring in all of the audio features from Spotify:**

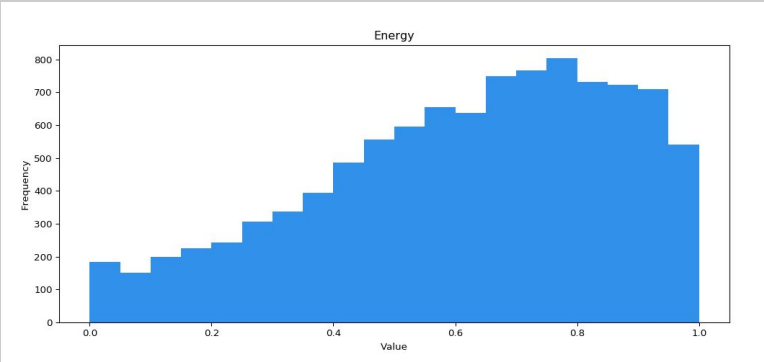
Danceability	*Energy*	*Key*
Loudness	*Mode*	*Speechiness*
Acousticness	*Instrumentalness*	*Liveness*
Valence	*Tempo*	*Type*

Exploratory Data Analysis

Spotify Audio Features

- **Spotify API documentation provides definitions of all of the audio features available.**
- **Example:**

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. The distribution of values for this feature looks like this:



Manually Coded Data

- Due to the relatively small number (224) of unique artists in the dataset, two key variables could be coded manually using information that is easily knowable with a simple Google search, but not available in a comprehensive dataset.

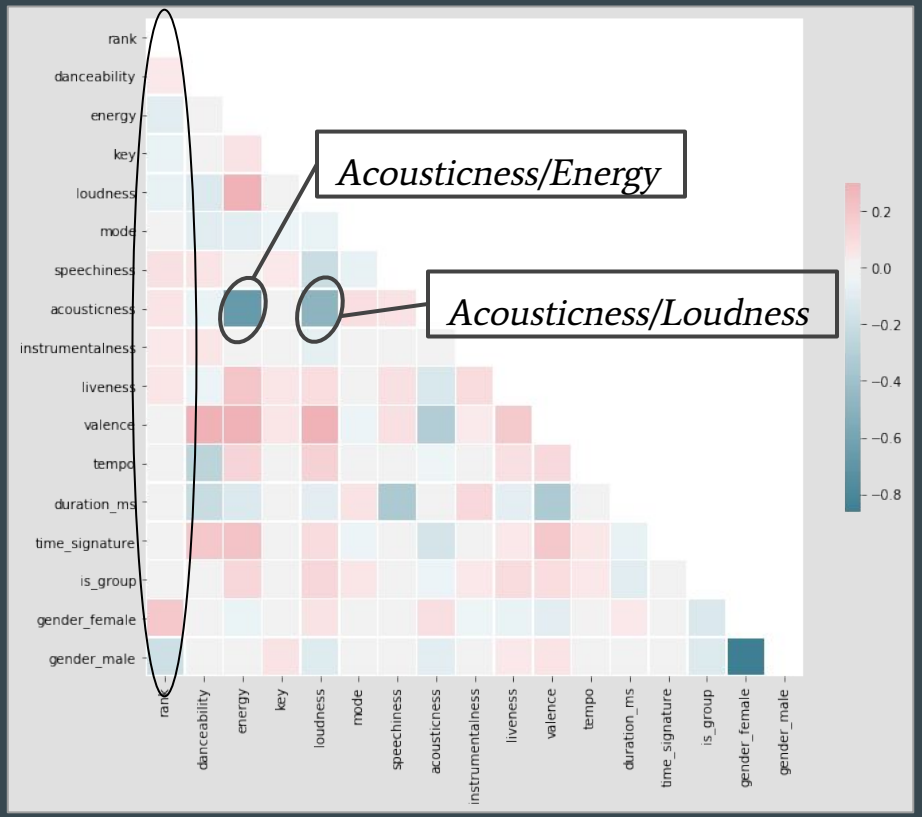
Gender:

1. Male
2. Female
3. Mixed (groups only)

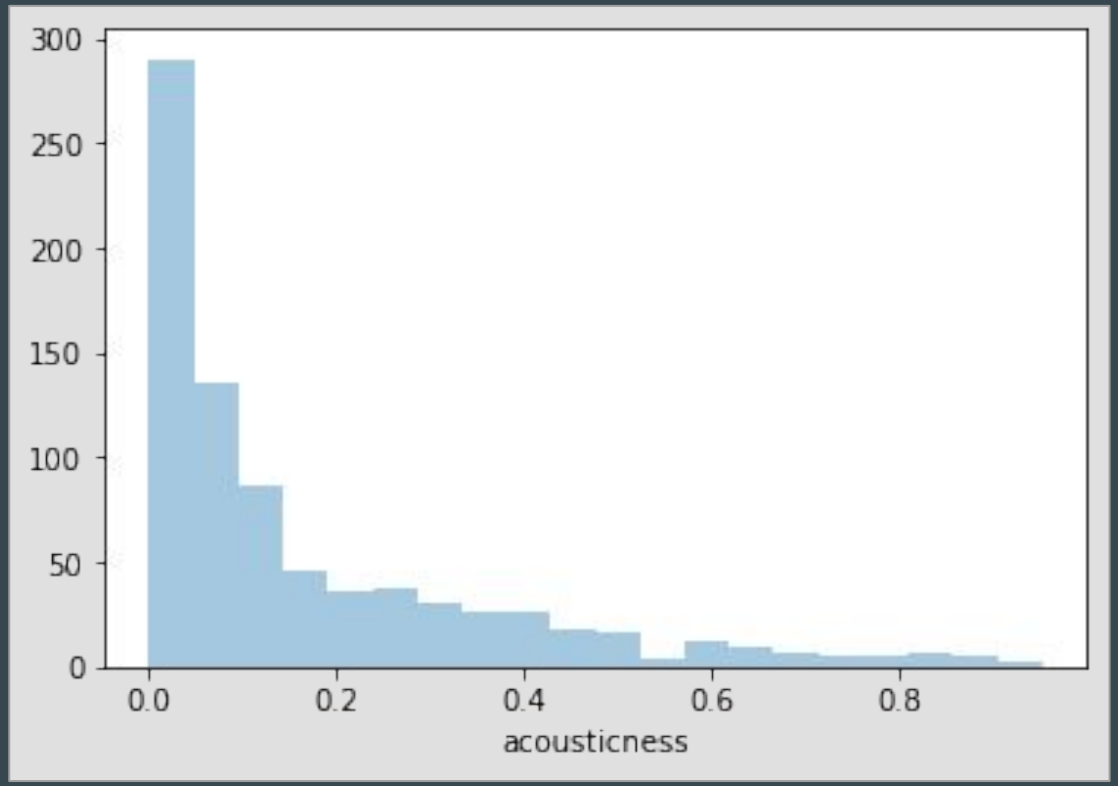
Is Group

1. Solo artist
2. Two or more people billed as a single artist

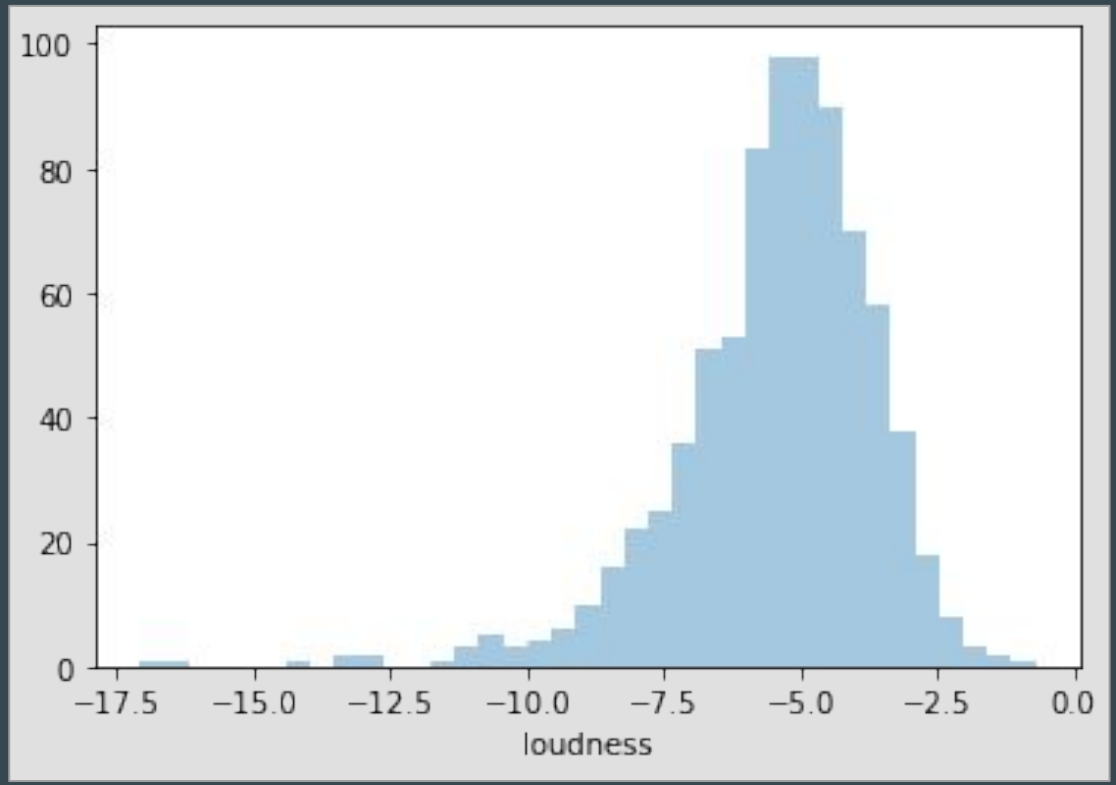
Audio Feature Correlations



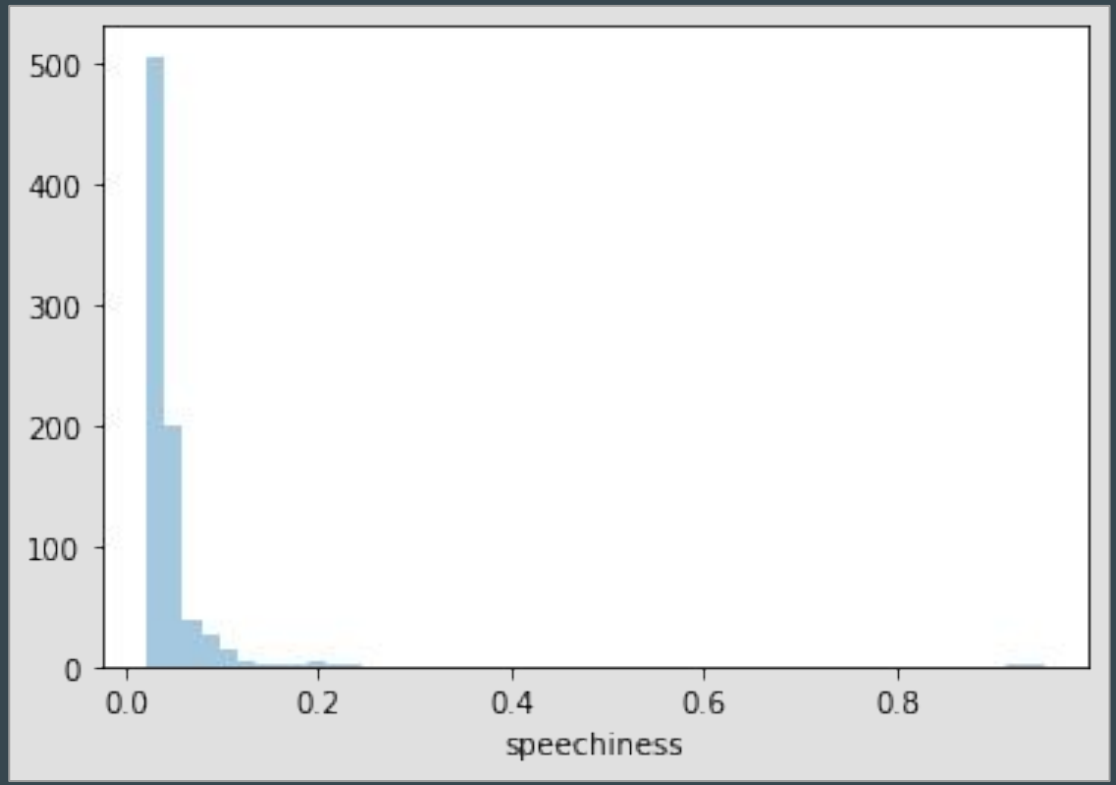
Audio Feature Distributions: Acousticness



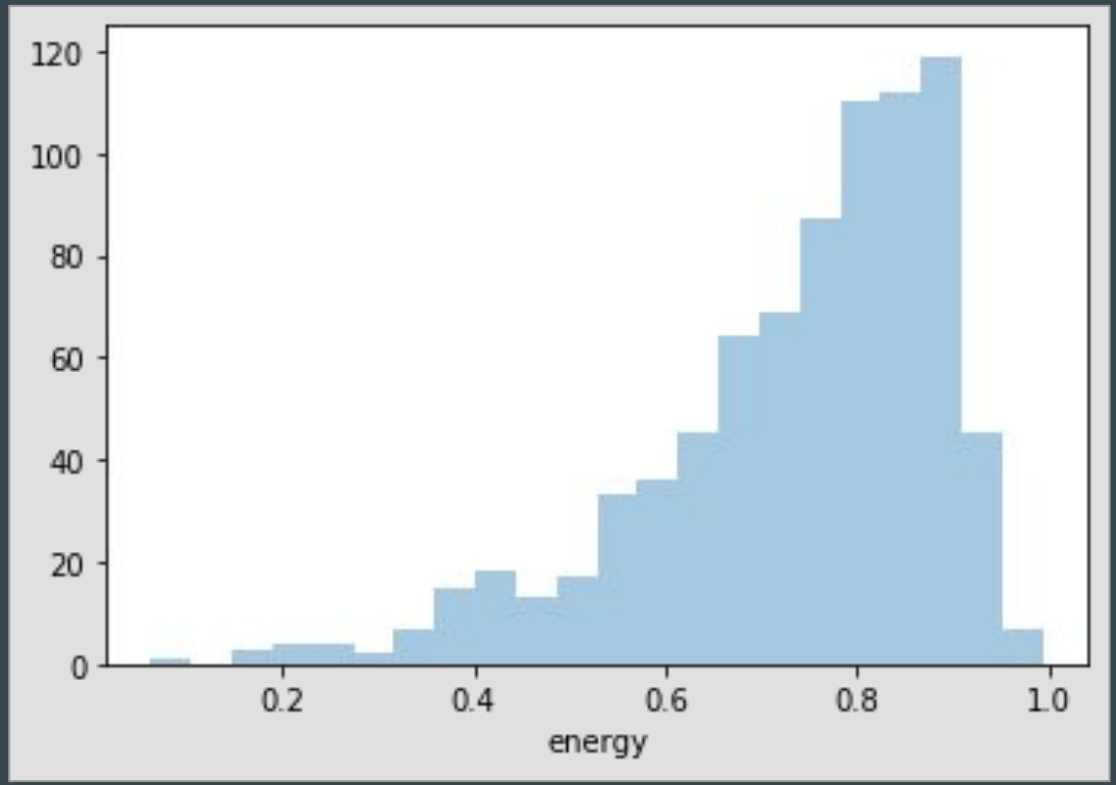
Audio Feature Distributions: Loudness



Audio Feature Distributions: Speechiness

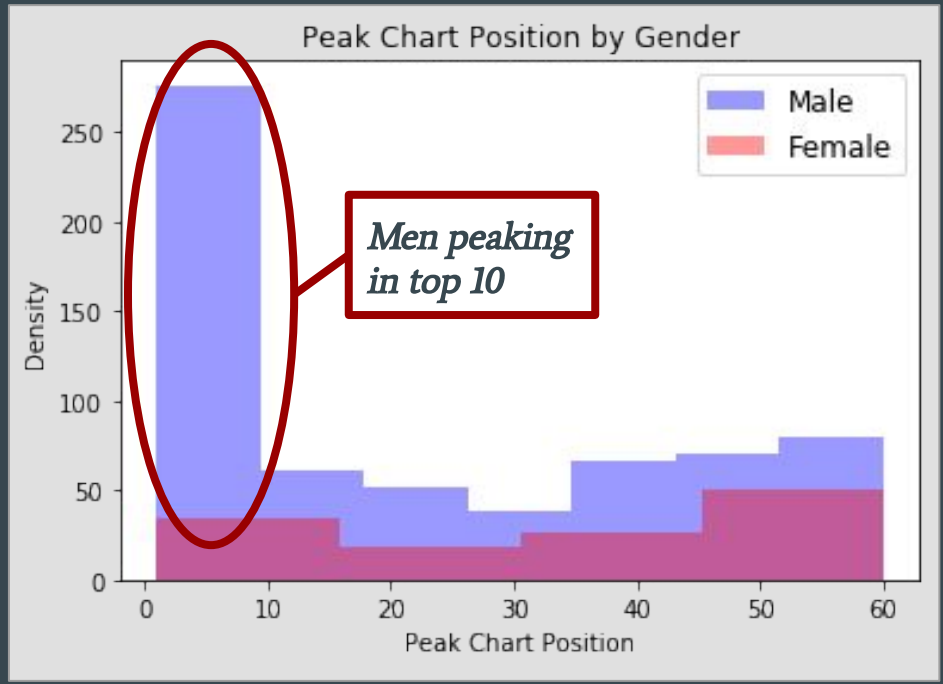


Audio Feature Distributions: Energy



Gender Analysis

- Men are heavily overrepresented
- Approximately 16% of songs are by female artists.
- Men are much more likely to peak in the top 10 than women are.



Gender Analysis

Male		Female
29%	Percent that peak at #1	9%
53	Number of unique artists to hit #1	7

Group Analysis

Solo		Group
25%	Percent that peak at #1	23%
162	Number of unique artists to hit #1	37

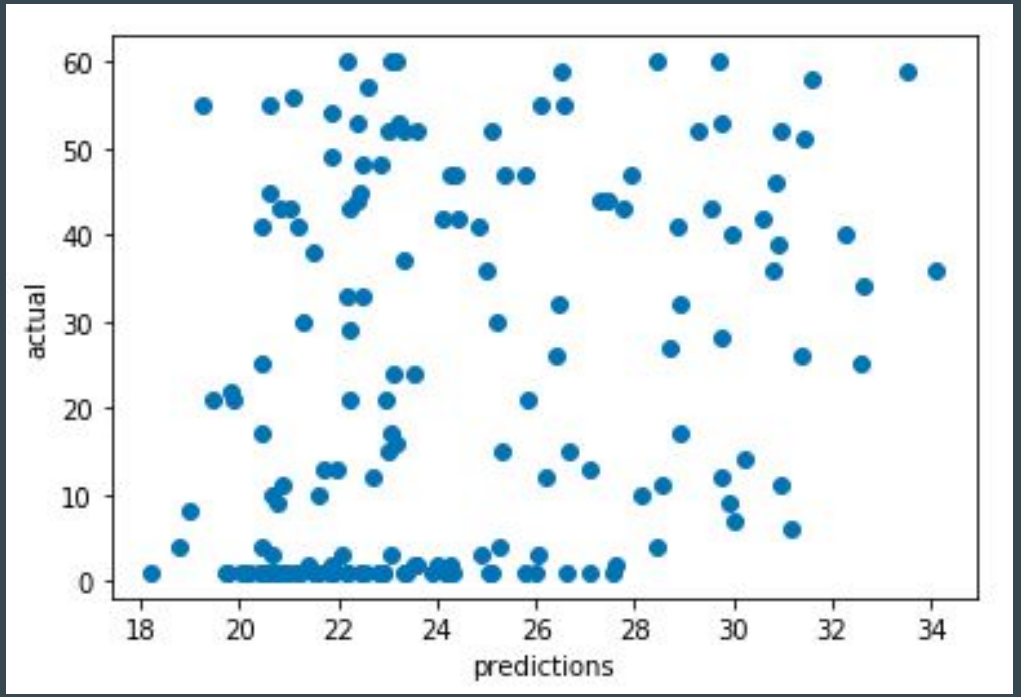
Gender + Group Analysis

Male Groups		Female Groups
29%	Percent that peak at #1	9%
33	Number of unique artists to hit #1	1

Modeling

Random Forest

- Grid Search
  - Max Depth: 3
  - Max Features: 3
  - Min Sample Leaf: 4
  - N_Estimators: 200



Predicting a Song's Success

Predicted Peak Chart Position

30

Predicting a Song's Success

Predicted Peak Chart Position

25

Predicting a Song's Success

Predicted Peak Chart Position

21

Actual Peak Chart Position

1

Future Additions

Metadata	• Songwriters • Producers
Artist History	• How successful has the artist been previously? • How many times have they hit the top 10?
Label Data	• Which labels are most likely to succeed? • Are labels just trading the number 1 spot?
Chart Duration	• How long was the song on the chart? • How long did the song stay at its peak?