Convolutional Genre Classifier

Overview

Just for fun I decided to create a neural network to classify music genres because I'm interested in the intersection of music and technology, and because I see it as a stepping stone to harder problems like building recommender systems. I trained my network on a variation of the GTZAN dataset which contained spectrogram images instead of audio clips. This allowed me to approach the task as a computer vision problem. I've had trouble generalizing beyond the dataset, I think because the spectrograms I've been producing from mp3's in my music library appear quite a bit different than the ones in the dataset.

Model Architecture

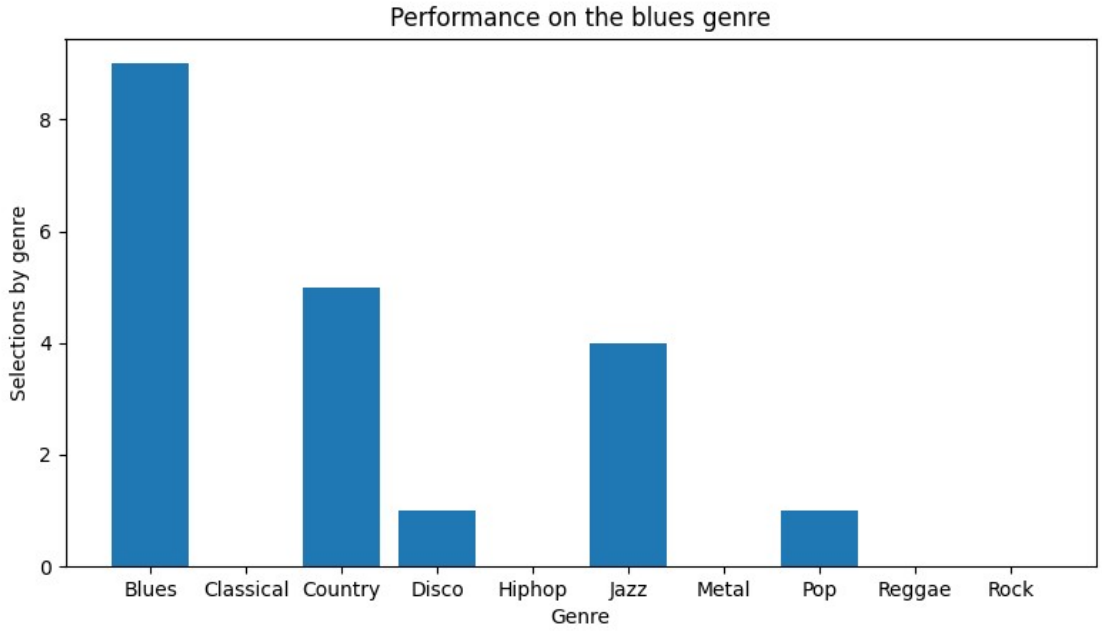
The model architecture is simple. It consists of a pair of convolutional layers, each followed by a max pooling layer, then one last convolutional layer and a global average pooling layer to flatten the data, which is then fed through a small dense network. There is dropout on the final convolutional layer as well as the intermediate dense layer, and the output has a softmax activation to allow for multi-class categorization. I used the Nadam optimizer with learning rate decay and sparse categorical crossentropy for my loss function.

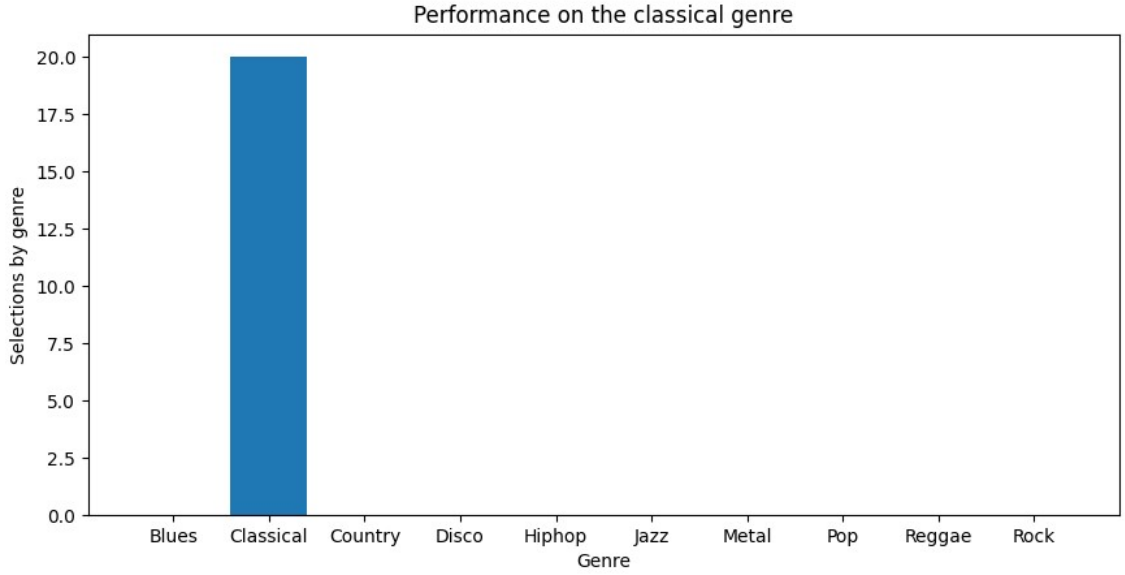
Training

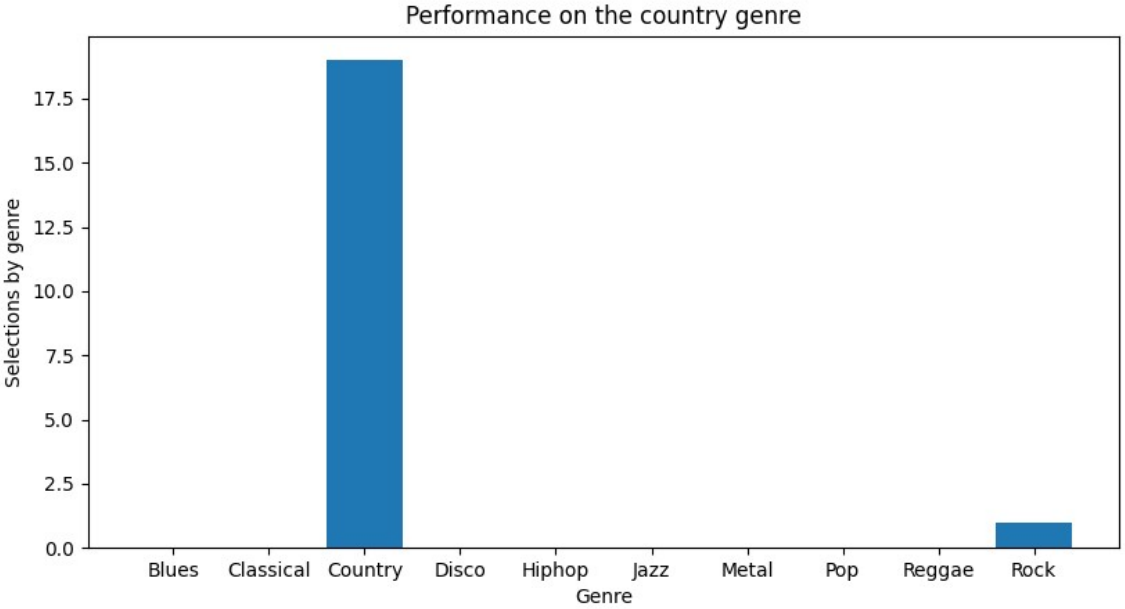
The model was trained for 150 epochs on a training set consisting of 800 spectrogram images, distributed evenly across all genres, with a validation split of 0.05. The genres were blues, classical, country, disco, hip-hop, jazz, metal, pop, reggae, and rock. The model was evaluated against a test set of 199 spectrogram images from the same distribution, 20 from each genre, except for jazz, which only had 19.

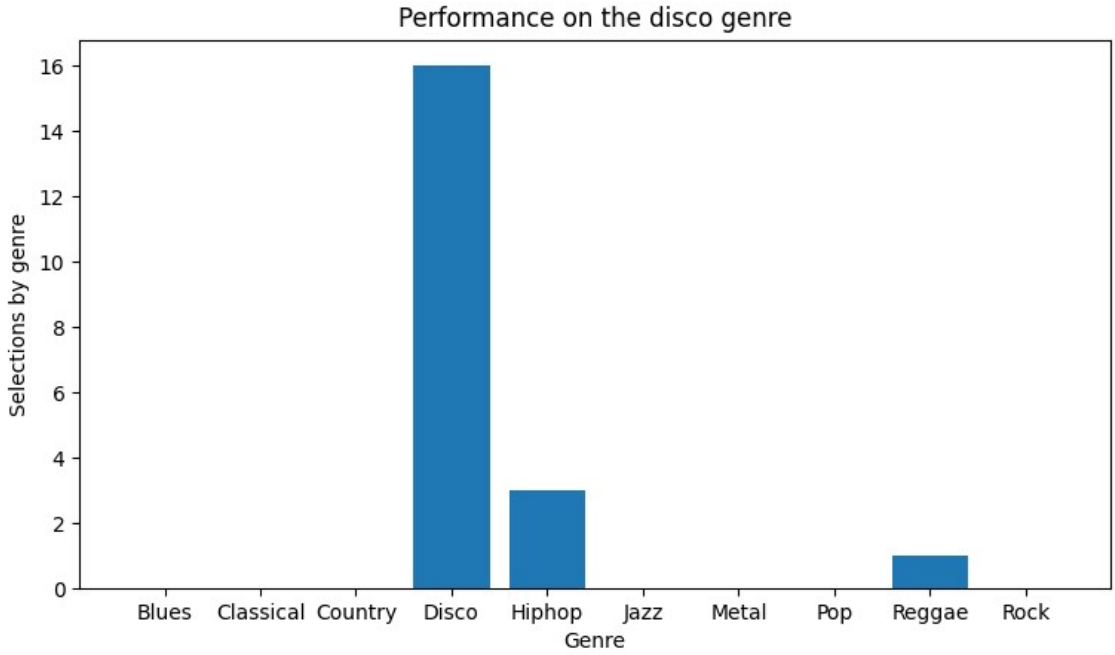
Performance

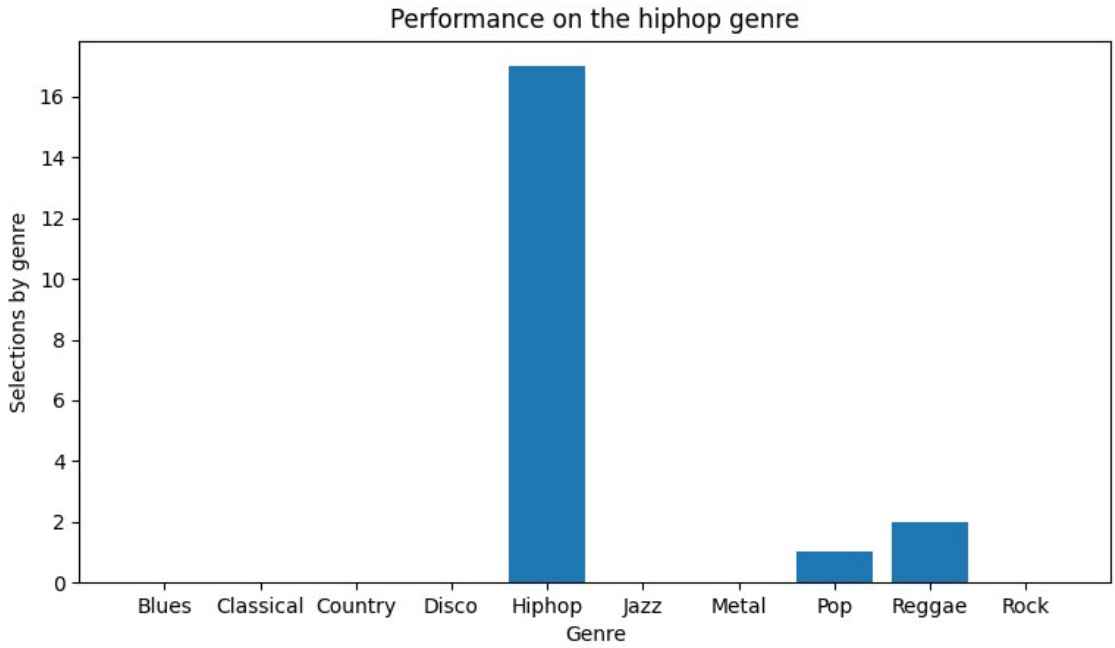
The model as described achieved an overall test set accuracy of 70.85%. It performed extremely well classifying the classical, country, hip-hop, disco, and pop genres, but poorly classifying the jazz, blues, and reggae genres. For the genres that the model could not classify fairly accurately, it seemed to be making the same sort of mistakes that a human might make. In the example of blues, the model selected country 25% of the time. That's a reasonable mistake as country music and the blues are derivative of each other and often blend elements.

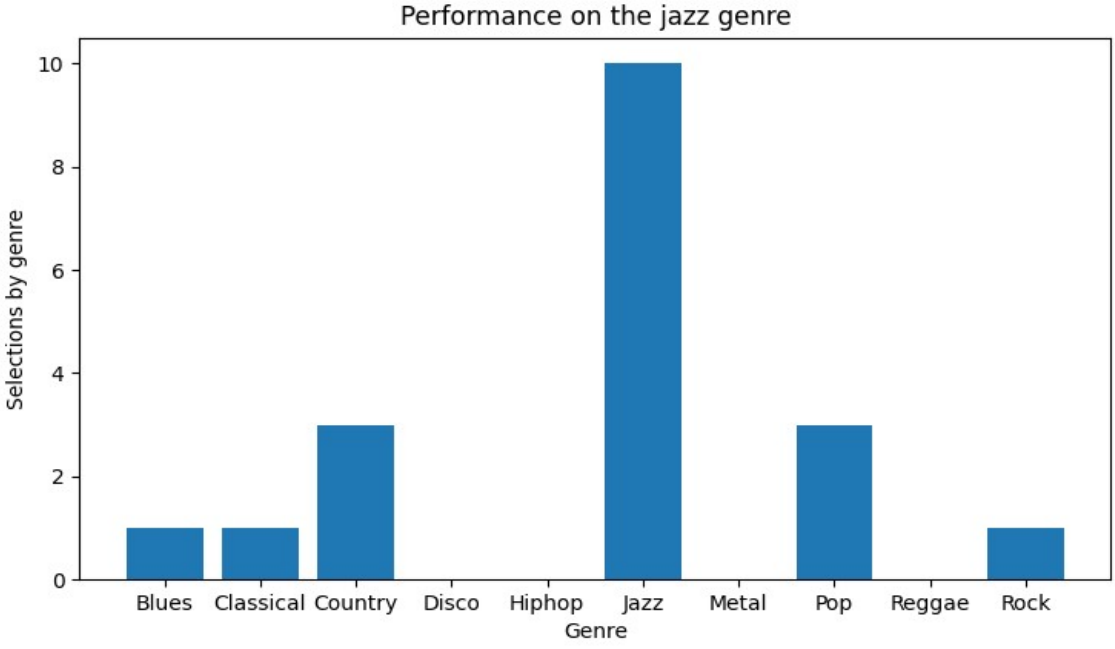

Performance on the blues genre
Selections by genre
0
2
4
6
8
Blues
Classical
Country
Disco
Hiphop
Jazz
Metal
Pop
Reggae
Rock
Genre

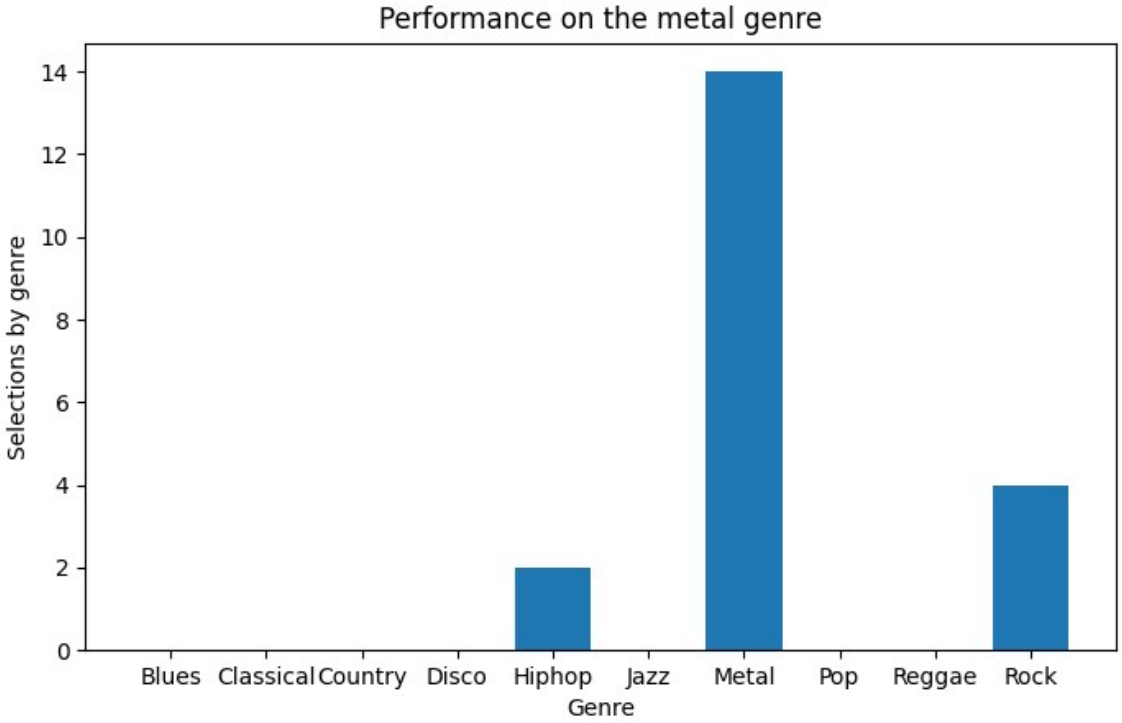

Performance on the classical genre
Selections by genre
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0
Blues
Classical
Country
Disco
Hiphop
Jazz
Metal
Pop
Reggae
Rock
Genre

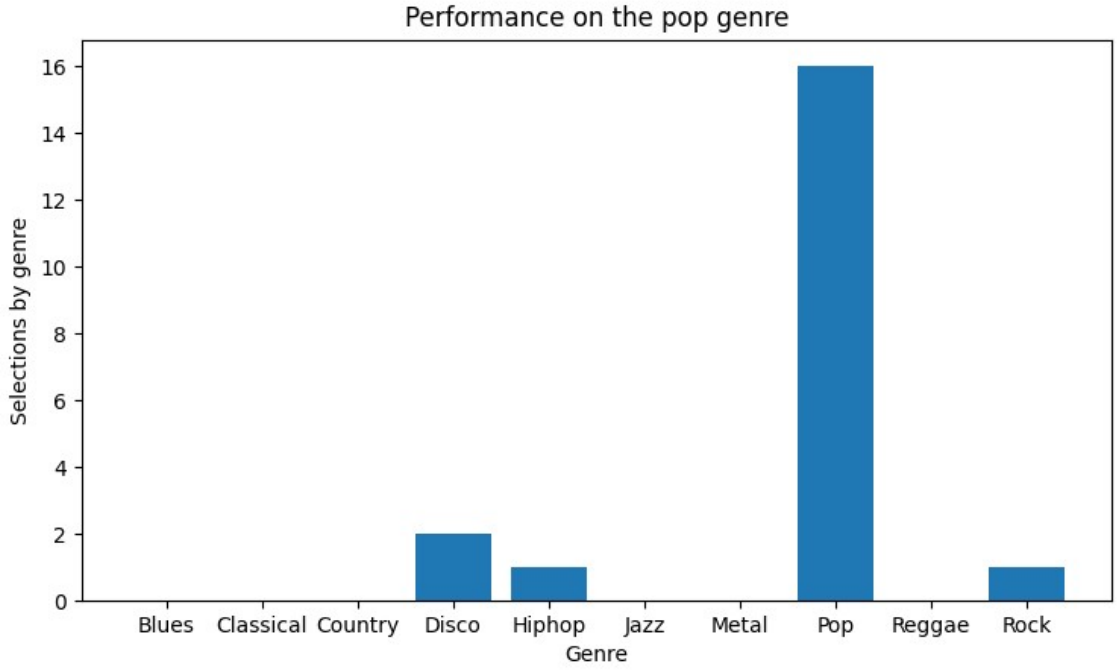

Performance on the country genre
Selections by genre
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
Blues
Classical
Country
Disco
Hiphop
Jazz
Metal
Pop
Reggae
Rock
Genre

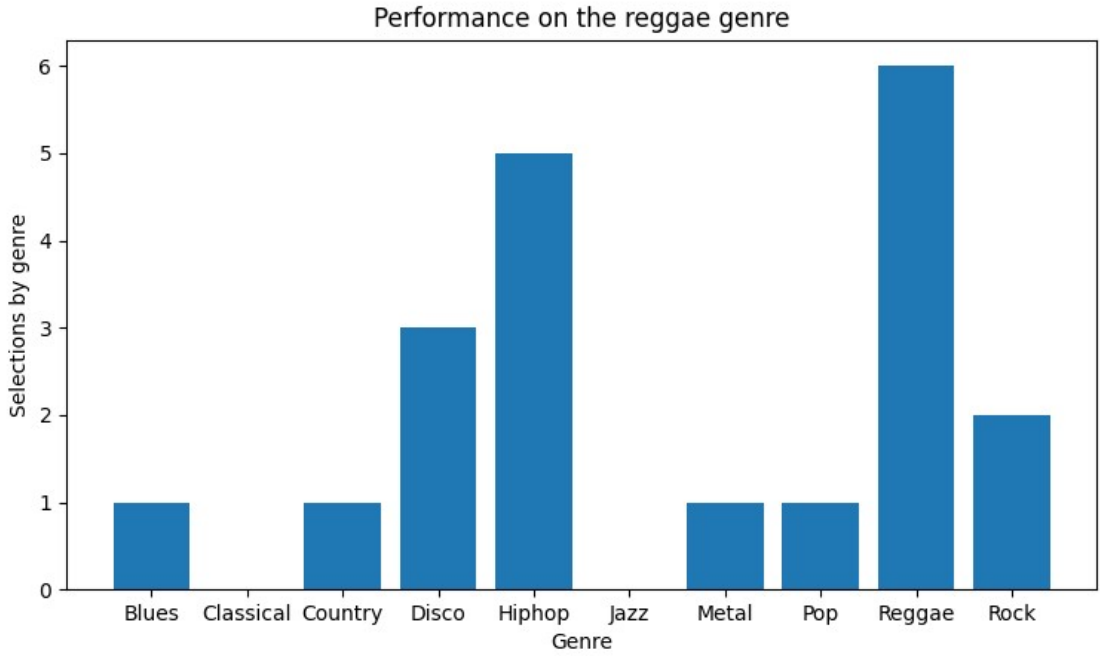

Performance on the disco genre
Selections by genre
0
2
4
6
8
10
12
14
16
Blues
Classical
Country
Disco
Hiphop
Jazz
Metal
Pop
Reggae
Rock
Genre

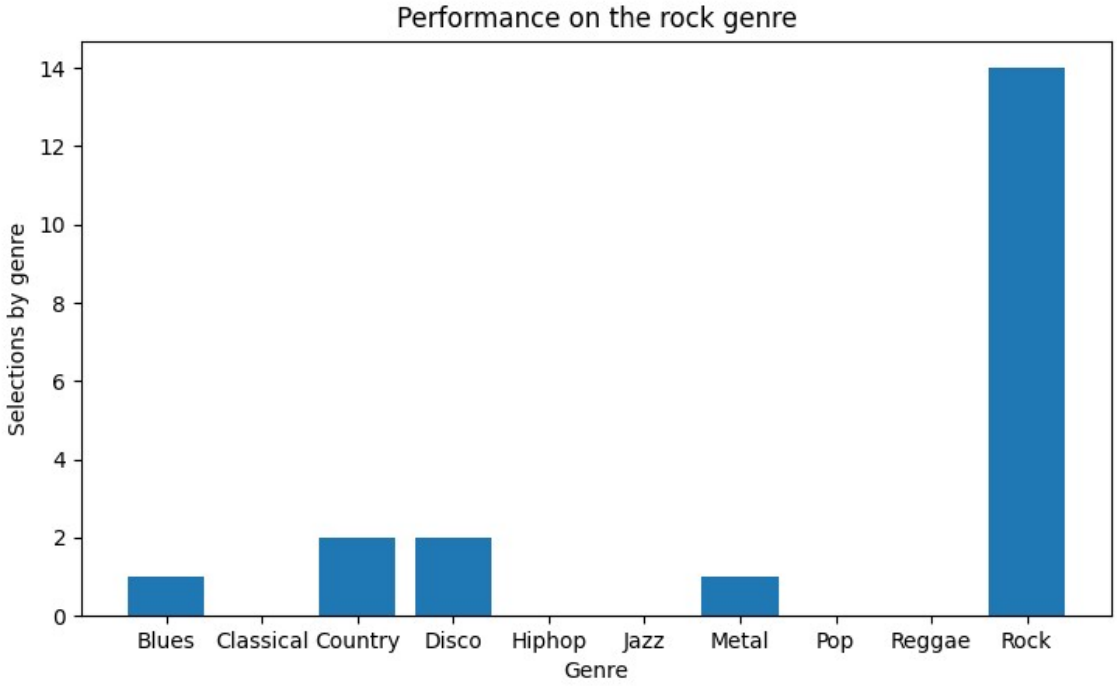

Performance on the hiphop genre
Selections by genre
0
2
4
6
8
10
12
14
16
Blues
Classical
Country
Disco
Hiphop
Jazz
Metal
Pop
Reggae
Rock
Genre


Performance on the jazz genre
Selections by genre
0
2
4
6
8
10
Blues
Classical
Country
Disco
Hiphop
Jazz
Metal
Pop
Reggae
Rock
Genre


Performance on the metal genre
Selections by genre
0
2
4
6
8
10
12
14
Blues
Classical
Country
Disco
Hiphop
Jazz
Metal
Pop
Reggae
Rock
Genre


Performance on the pop genre
Selections by genre
0
2
4
6
8
10
12
14
16
Blues
Classical
Country
Disco
Hiphop
Jazz
Metal
Pop
Reggae
Rock
Genre


Performance on the reggae genre
Selections by genre
0
1
2
3
4
5
6
Blues
Classical
Country
Disco
Hiphop
Jazz
Metal
Pop
Reggae
Rock
Genre


Performance on the rock genre
Selections by genre
0
2
4
6
8
10
12
14
Blues
Classical
Country
Disco
Hiphop
Jazz
Metal
Pop
Reggae
Rock
Genre