Report

Oluwamayowa Adesoye

1. Which insights did you gain from your EDA? Were any columns highly correlated? If so, name them.

   I hypothesize that newer releases have more violent content, and this was proven true. Another interesting discovery was that rock was a genre with lyrics that had something to do with violence, which was shocking, and next was reggae. In addition, a majority of the grass had a right, skewed distribution, indicating that the majority of the data is located on the left side of the graph, and the mean or average is greater than the median. When the mean is greater than the median, there are some extremely high values outliers in the data set that are influence in the mean to be higher than the median these outliers are typically much larger than most of the other values in the data set, in this instance, I chose to keep the outliers because they help with the analysis of this project

2. How did you determine which columns to drop or keep? If your EDA informed this process, explain which insights you used to determine which columns were not needed.

   I knew what columns to keep because they were interesting correlations said about them on the box plot. They told more of a story and allowed for discoveries to be made.

3. What was the optimal number of clusters in your cluster model? Explain how you determined this value.

   There were 5 optimal numbers of clusters in my cluster model and this was discovered using the elbow plot.

4. Take a look at the respective songs that fell into your clusters. Describe these clusters in human terms to the best of your ability using the columns in your dataset (for example high-gospel songs, low-gospel songs, etc). Feel free to listen to these songs as well to get a sense of what nuance your algorithm picked up on.

   There are 5 different clusters.

- Yellow cluster: between ages 0-0.37 has a high amount of violence in it.
- Green cluster: between ages 0.37-1 has a high amount of violence in it and it encompasses a wider age range.
- Dark Purple cluster: has low amount of violence and is between the age of 0-0.36.
- Greenish-blue cluster: between age 0.36-0.63 has a low amount of violence.
- Light purple: between 0.63- 1 has a wider age range and lower amount of violence.

5. Take a look at the clusters that your algorithm assigned to your test samples. Based on these clusters, which songs would you recommend to this user?

Report

Oluwamayowa Adesoye

- For 0 and 9 I would recommend songs from the artist Paul Anka.
- For 1 and 6 I would recommend songs from godsmack
- For 2 and 5 I would recommend songs from Jerry Lee Lewis.