Aditya Patel

CS50025

Homework 5

5d)

Inertia is the sum of the differences between the actual data and the fitted values, or simply put, how well a dataset was clustered by K-means. A larger inertia value indicates a poor fit, while the closer

I selected a k-value of 8 based on the plot of the inertia value vs the number of clusters below. The inertia scores represent the squared standard error at each value of K. At 8 clusters, there is a ‘bend’ in the plot, suggesting that the change in slope will not reduce much further after K=8.

A blue line on a white background

Description automatically generated

5e)

The results of the clustering and selection operation appears as follows.

A screenshot of a computer

Description automatically generated

From the way that the data is clustered, it does appear that clusters were assigned by the artist’s genre with the most tracks – for example, Iron Maiden has 213 Rock songs, but has 227 Metal songs. The dominant genre is Metal, which it shares with Metallica.