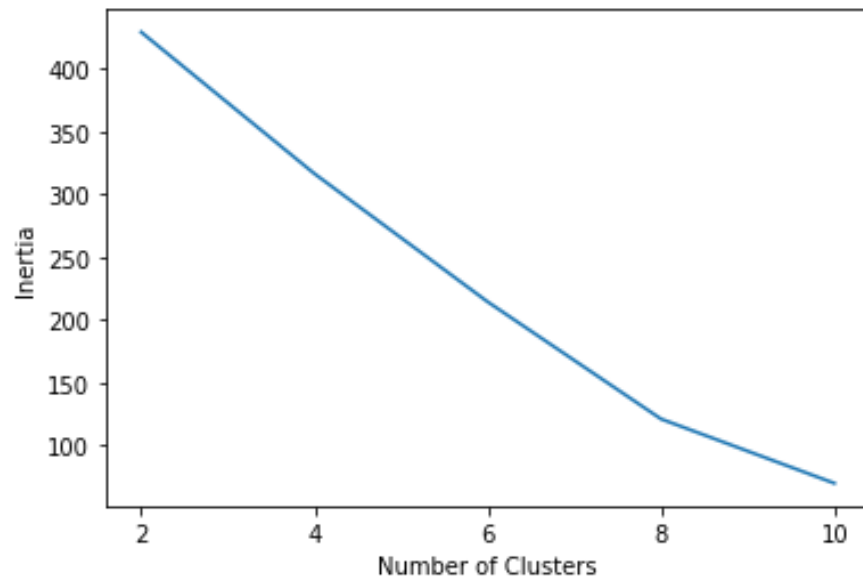


Aditya Patel
CS50025
Homework 5

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.



5e)

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

```
● aditya@Adityas-MacBook-Pro hw5 % /usr/local/bin/python3 /Users/aditya/Library/CloudStorage/OneDrive-purdue.edu/DEII/HW/hw5/hw5.py
```

	Artist	group_num	centroid_dist	ArtistId	Album	Track	Playlist	Rock	Latin	Metal	Alternative & Punk	TV Shows	Jazz	Blues
0	The Office	0	1.030241	156	3	52	1	0	0	0	0	72	0	0
1	Lost	0	1.030241	149	4	91	1	0	0	0	0	96	0	0
2	Metallica	1	2.533658	50	10	112	3	0	0	296	0	0	0	0
3	Iron Maiden	1	2.533658	90	21	150	3	213	0	227	0	0	0	18
4	Deep Purple	2	0.666262	58	11	87	2	226	0	0	0	0	0	0
5	Led Zeppelin	2	0.770153	22	14	91	2	252	0	0	0	0	0	0
6	U2	2	1.058766	150	10	124	2	287	0	0	0	0	0	0
7	Tim Maia	3	0.293653	145	2	30	2	0	67	0	0	0	0	0
8	Cássia Eller	3	0.384323	77	2	30	2	0	62	0	0	0	0	0
9	Milton Nascimento	3	0.600851	42	2	26	2	0	55	0	0	0	0	0
10	Skank	4	0.530865	130	2	23	2	57	0	0	0	0	0	0
11	Jamiroquai	4	0.558845	92	3	32	2	30	0	0	0	0	0	0
12	The Cult	4	0.572670	139	2	30	2	64	0	0	0	0	0	0
13	Spyro Gyra	5	0.885924	53	2	21	2	0	0	0	0	0	54	0
14	Miles Davis	5	0.885924	68	3	36	2	0	0	0	0	0	75	0
15	Green Day	6	0.421434	54	2	34	2	0	0	0	83	0	0	0
16	Smashing Pumpkins	6	0.442987	131	2	34	2	0	0	0	85	0	0	0
17	The Tea Party	6	0.705562	143	2	22	2	0	0	0	76	0	0	0
18	Eric Clapton	7	0.000000	81	2	47	3	0	49	0	0	0	0	96

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.