For IRIS Dataset,

<u>K-Means (when init is ‘random’):</u>

The K-means (when init is ‘random’) cluster analysis are:

[0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
2 1 2 2 2 2 1 2 2 2 2 2 2 1 1 2 2 2 2 1 2 1 2 1 2 2 1 1 2 2 2 2 2 1 2 2 2 2 1 2 2 2 1 2 2 2 1 2 2 1]

Cluster Centres for k-means (when init is 'random') are:

[[5.006 3.428 1.462 0.246]

[5.9016129 2.7483871 4.39354839 1.43387097]

[6.85 3.07368421 5.74210526 2.07105263]]

Distortion for k-means (when init is ‘random’): 26.59

<u>K-Means (when init is ‘k-means++’)</u>:

K-Means Cluster analysis for k-means (when init is 'k-means++') are:

[1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
0 0 2 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 2 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
2 0 2 2 2 2 0 2 2 2 2 2 2 0 0 2 2 2 2 0 2 0 2 0 2 2 0 0 2 2 2 2 2 0 2 2 2 2 0 2 2 2 0 2 2 2 0 2 2 0]

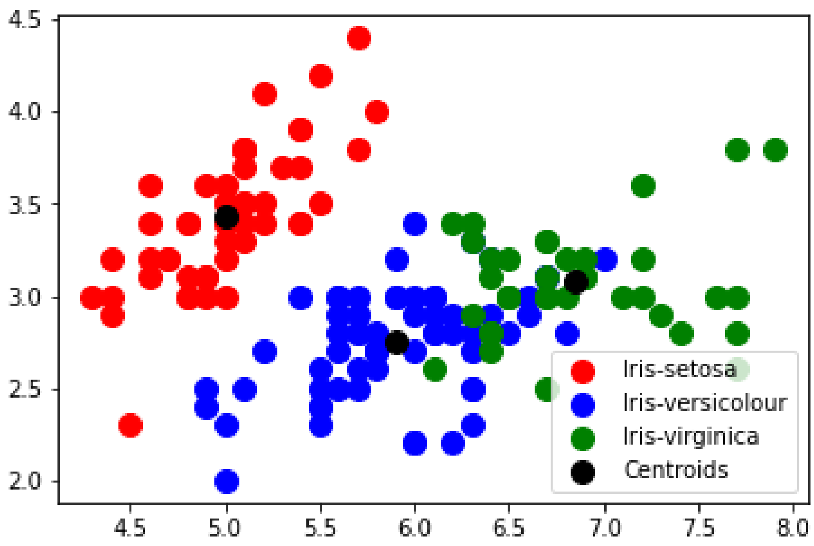
Cluster Centres for k-means (when init is 'k-means++') are:

[[5.9016129 2.7483871 4.39354839 1.43387097]

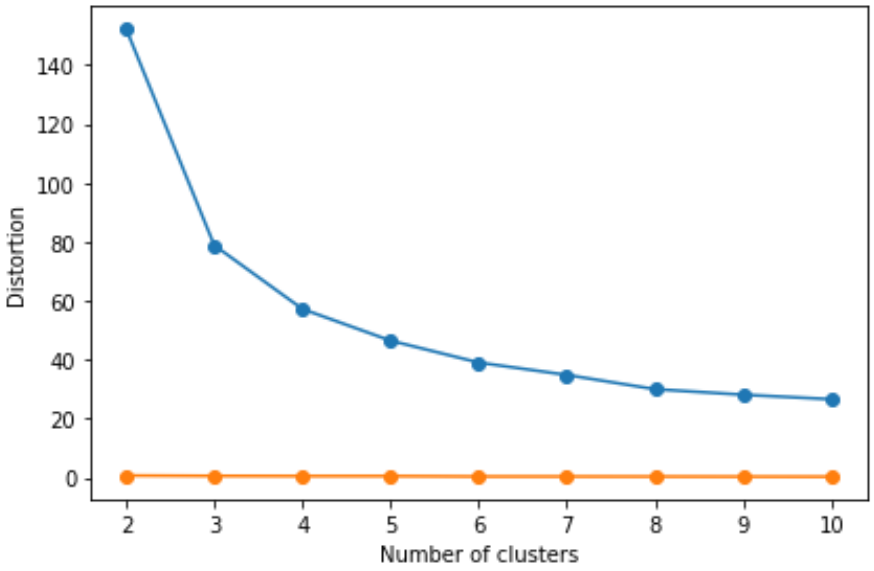
[5.006 3.428 1.462 0.246]

[6.85 3.07368421 5.74210526 2.07105263]]

Distortion for k-means (when init is ‘k-means++’): 26.09

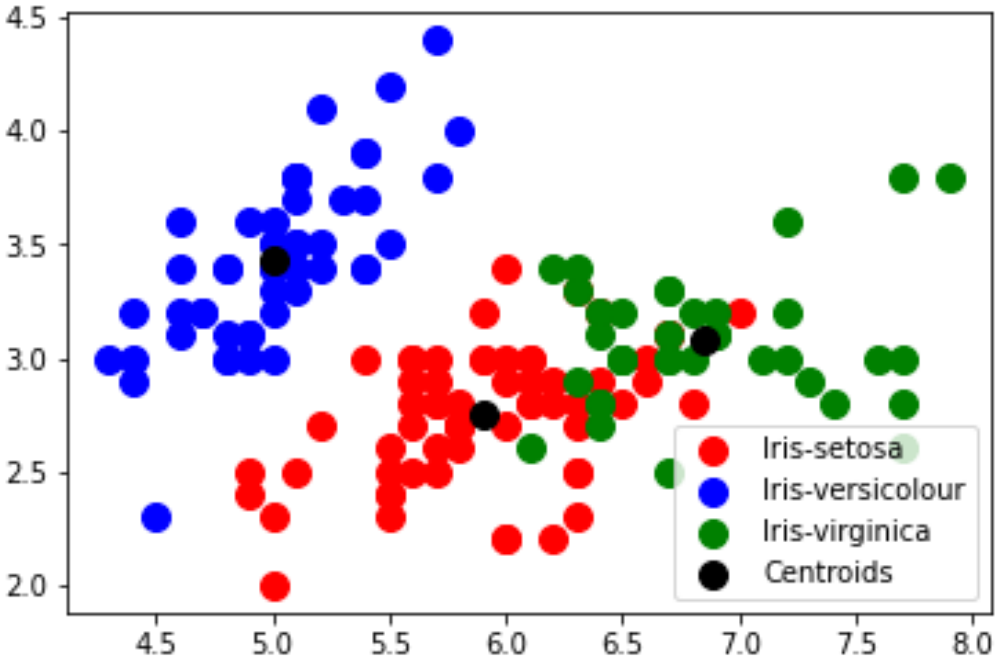


k-means plot (when init is 'random')

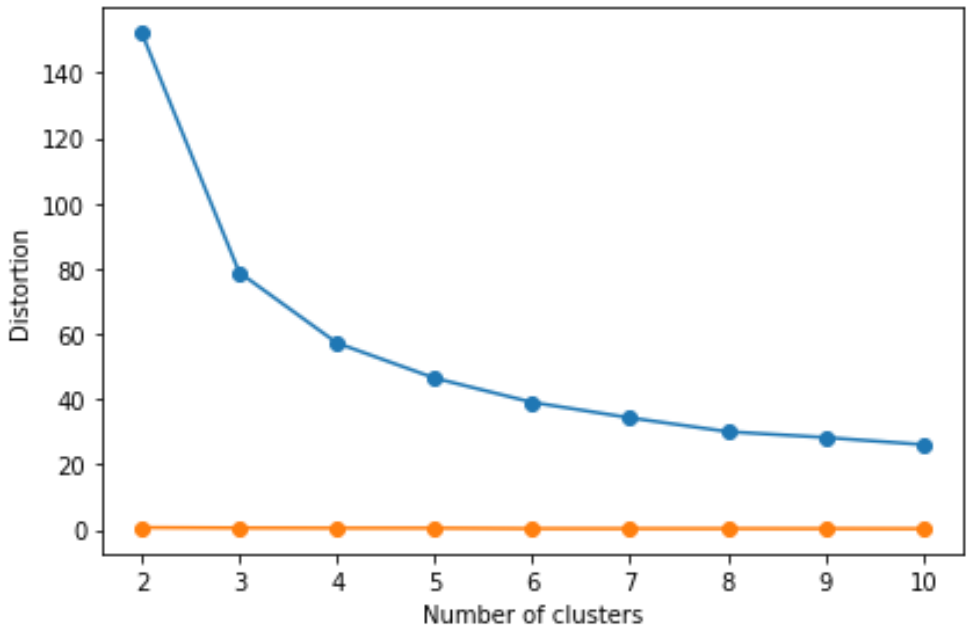


Elbow approach and the silhouette (when init is 'random')

The Blue colour line is the Elbow approach for the k-means (when init is ‘random’) as we can see from the plot the K value for the elbow method is unsure what to take so I compared the elbow with that of the silhouette score (orange colour line) so that the K value can be fixated on what to take.

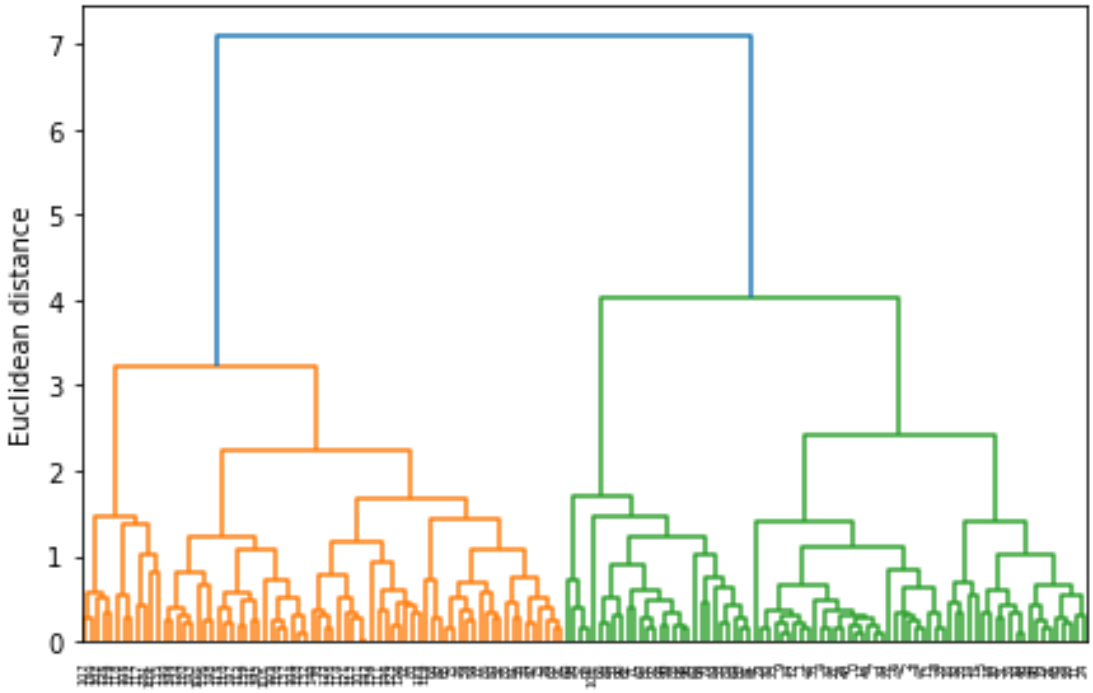


k-means plot (when init is 'k-means++')

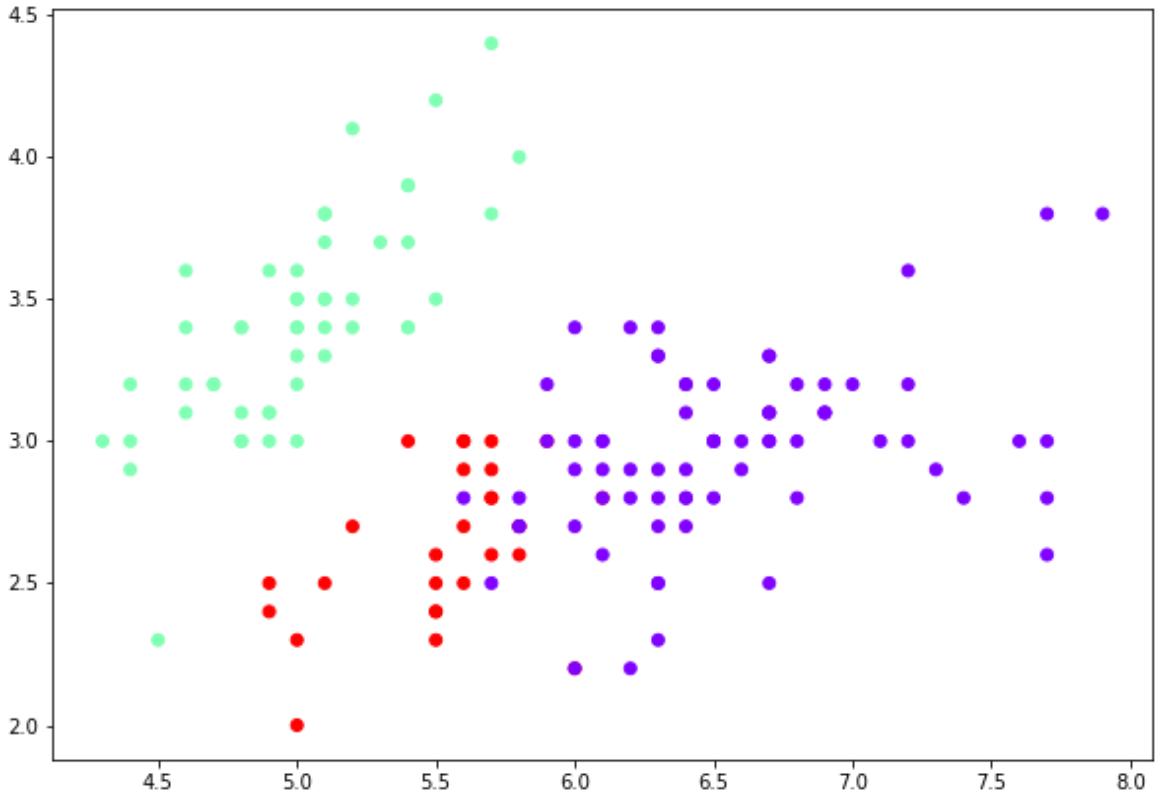


Elbow approach and the silhouette (when init is 'k-means++')

The Blue colour line is the Elbow approach for the k-means (when init is ‘k-means++’) as we can see from the plot the K value for the elbow method is unsure what to take so I compared the elbow with that of the silhouette score (orange colour line) so that the K value can be fixated on what to take.



Hierarchical Clustering (using SciPy library) Dendrograms



Hierarchical clustering (using Scikit library) Dendrograms

Algorithms	Sum of Squared Errors	Silhouette Scores	Time Taken
K-means (when init is 'random')	78.851	0.5528	0.2076
Elbow Method (when init is 'random')	--	0.3044	0.1713
K-means (when init is 'k-means++')	78.851	0.5528	0.0255
Elbow Method (when init is 'k-means++')	--	0.3312	0.3763
Hierarchical approach (using SciPy library)	--	Cluster1 (when max_d is taken) = 0.4998 Cluster2 (when k is taken) = 0.5159	0.2689
Hierarchical approach (using scikit library)	--	0.5135	0.0025

For MNIST Dataset,

I took a subset of test - size ‘0.90’.

K-Means (when init is ‘random’):

The K-means (when init is ‘random’) cluster analysis are:

[2 2 1 ... 2 0 1]

Cluster Centres for k-means (when init is 'random') are:

[[0. 0. 0. ... 0. 0. 0.]

[0. 0. 0. ... 0. 0. 0.]

[0. 0. 0. ... 0. 0. 0.]]

Distortion for k-means (when init is ‘random’): 17839831688.70

K-Means (when init is ‘k-means++’):

K-Means Cluster analysis for k-means (when init is 'k-means++') are:

[0 0 1 ... 0 2 1]

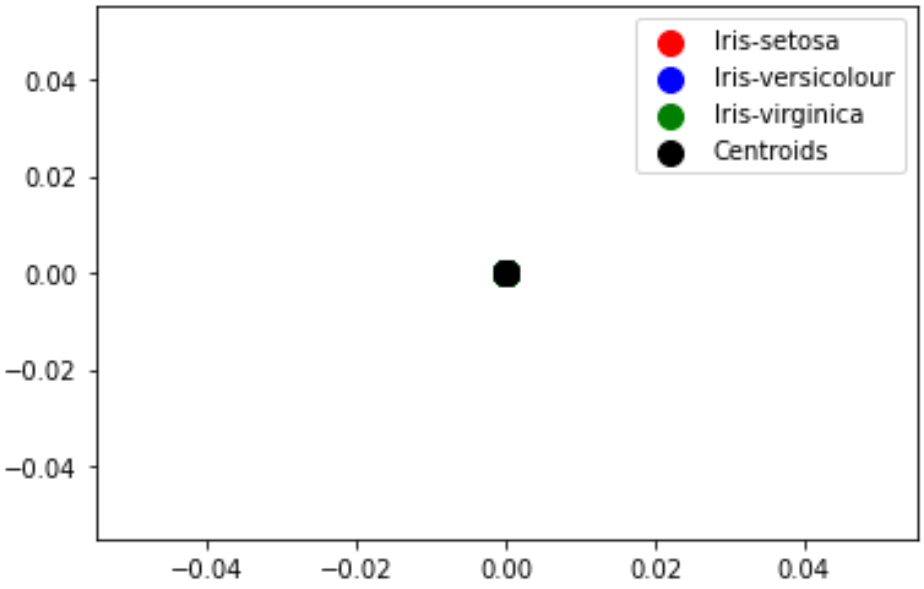
Cluster Centres for k-means (when init is 'k-means++') are:

[[0. 0. 0. ... 0. 0. 0.]

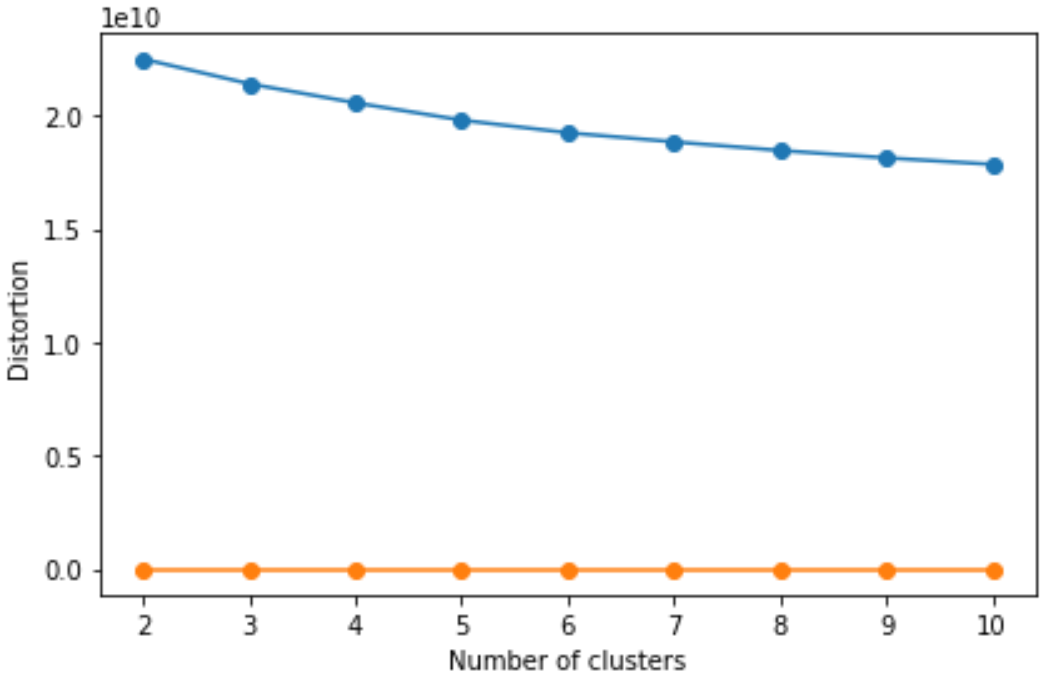
[0. 0. 0. ... 0. 0. 0.]

[0. 0. 0. ... 0. 0. 0.]]

Distortion for k-means (when init is ‘k-means++’): 17838960304.95

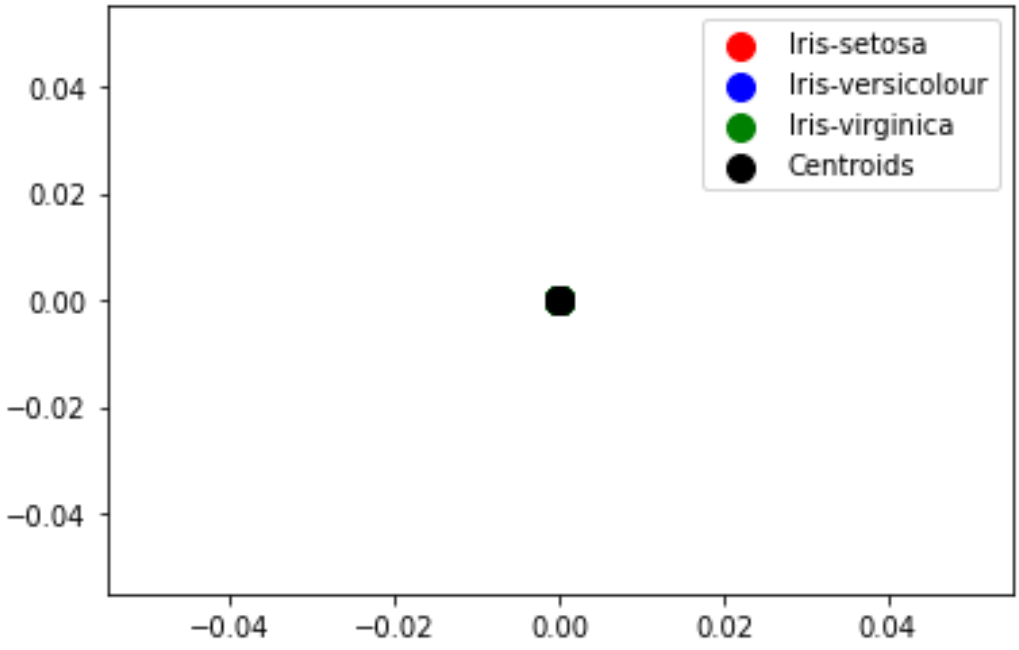


k-means plot (when init is 'random')

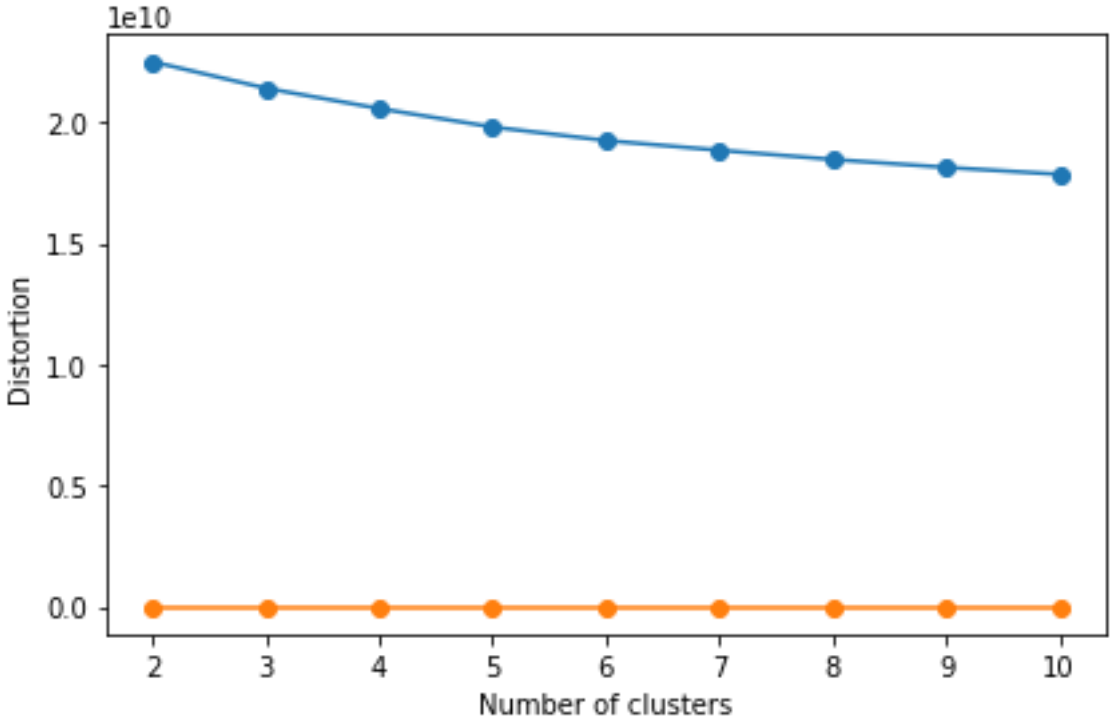


Elbow approach and the silhouette (when init is 'random')

The Blue colour line is the Elbow approach for the k-means (when init is ‘random’) as we can see from the plot the K value for the elbow method is unsure what to take so I compared the elbow with that of the silhouette score (orange colour line) so that the K value can be fixated on what to take.

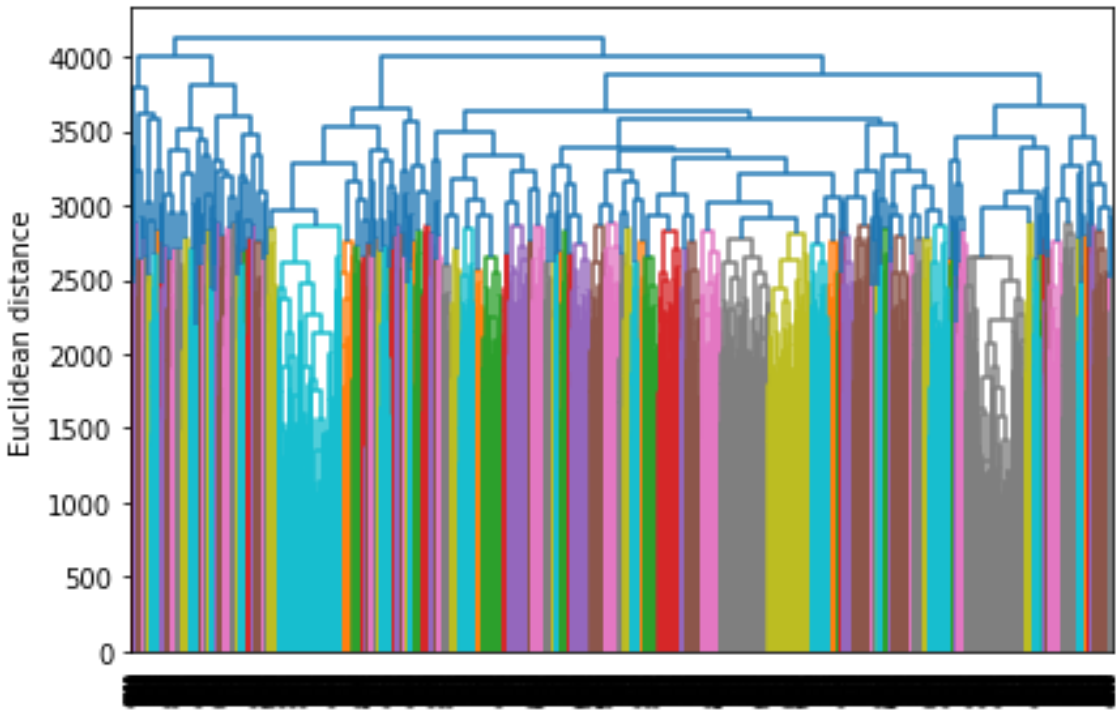


k-means plot (when init is 'k-means++')

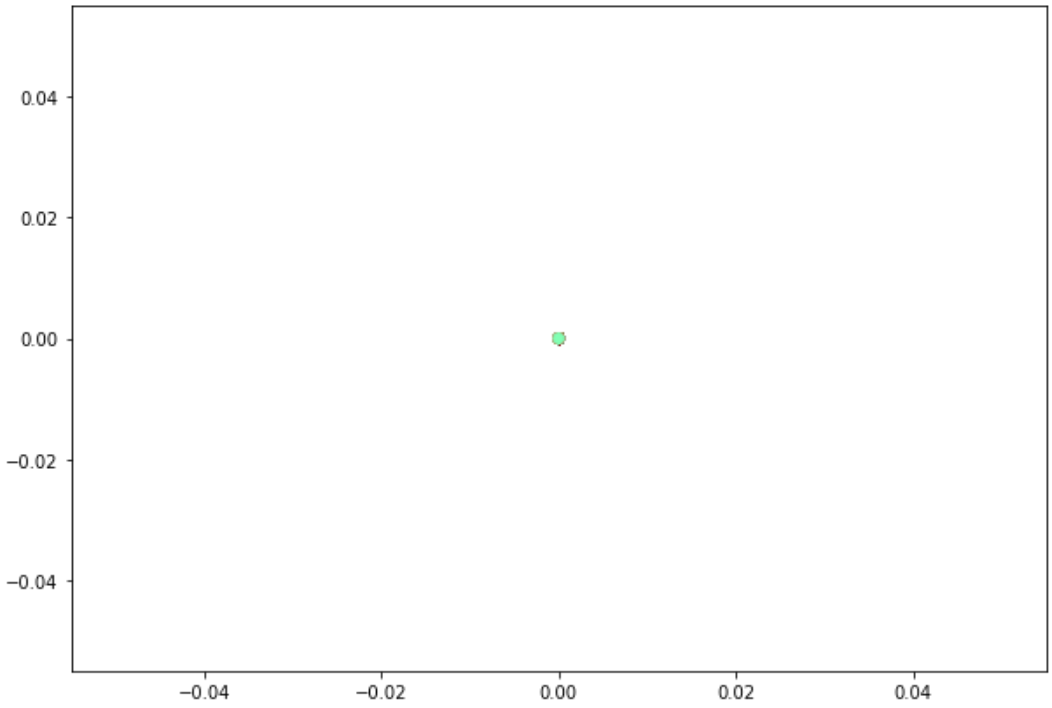


Elbow approach and the silhouette (when init is 'k-means++')

The Blue colour line is the Elbow approach for the k-means (when init is ‘k-means++’) as we can see from the plot the K value for the elbow method is unsure what to take so I compared the elbow with that of the silhouette score (orange colour line) so that the K value can be fixated on what to take.



Hierarchical Clustering (using SciPy library) Dendrograms



Hierarchical clustering (using Scikit library) Dendrograms

Algorithms	Sum of Squared Errors	Silhouette Scores	Time Taken
K-means (when init is 'random')	**21392925381.585**	**0.0538**	**2.9420**
Elbow Method (when init is 'random')	**--**	**0.0572**	**33.8220**
K-means (when init is 'k-means++')	**21392906205.685**	**0.0539**	**3.9344**
Elbow Method (when init is 'k-means++')	**--**	**0.0577**	**40.6300**
Hierarchical approach (using SciPy library)	**--**	**--**	**102.7345**
Hierarchical approach (using scikit library)	**--**	**0.0399**	**22.7918**