1 a) Experiment with some different sizes of k and observe the range of the Sum of Squares Error (SSE) (see Appendix for more details on SSE). What k value would you pick to best cluster the countries into continents? Briefly justify why you select the k value.

b) Report the final parameters you set including the value of k selected to obtain your final clusters. Also, report the coordinates of the centroids, sum of squared differences to centroids and the number of iterations from the best clustering you have found. Explain in one sentence why you think this is your best clustering.

c) Name the continent each cluster represents in the table below. Describe each cluster according to the centroid values of each attribute. For each cluster, be sure to report the attribute centroid in terms of the original attribute values. Also, you can visualize the clusters n a scatter plot to help you describe and identify the continent represented by each cluster. You can also concatenate the cluster labels with longitude, latitude and country names to analyze the countries in each cluster.

|  |  |  |
| --- | --- | --- |
| Cluster | Centroid | Continent |
| 0 | 5.97218533 | 41.94617853 |
| 1 | -15.84210807 | -164.35100116 |
| 2 | 9.62003573 | -69.14907131 |
| 3 | 39.84178493 | 7.59405209 |
| 4 | 6.47298142 | 129.70481495 |

A picture containing screenshot, colorfulness

Description automatically generated

2 a) Use the AgglomerativeClustering algorithm. Experiment with different number of clusters (n\_clusters) and other parameters (e.g., affinity, linkage, etc.) to get the best clusters to represent continents. Use dendrogram and scatter plot to help you visualize the clusters. Report the best parameters, the number of clusters you have selected. Explain in one sentence why you think this is your best clustering.

b) Name the continent each cluster represent in the table below. To analyze each cluster, observe the data points in each cluster on the scatter plot or look at what country names are in the clusters.

|  |  |
| --- | --- |
| Cluster | Continent |
| 0 | Asia |
| 1 | Europe |
| 2 | South America |
| 3 | Africa |
| 4 | Oceanic |
| 5 | Oceanic |
| 6 | South America |
| 7 | Asia |
| 8 | Antarctica |

A picture containing screenshot, text, colorfulness

Description automatically generated

3 a) Based on the algorithm you have selected, provide a final mapping of the cluster number to the continent name in the following table.

|  |  |
| --- | --- |
| Cluster | Continent |
| 0 | Asia |
| 1 | Europe |
| 2 | South America |
| 3 | Africa |
| 4 | Oceanic |
| 5 | Oceanic |
| 6 | South America |
| 7 | Asia |
| 8 | Antarctica |