HW4

Python part:

1. import and check the data. (We are using median-age-vs-children-per-woman.csv)

There are 222 unique countries, and the time is from1950-2020.

2. change columns names and select required columns

3.unstack the table and merge (matched there are 222 rows).

4. applying algorithms.

Chart, scatter chart

Description automatically generatedTableau part:

MDS shows no clear relationship between the continent and the country. But we do find that as the x axis being positive and y axis being negative, there is a clearer view of clustering over continent and most of them tread to be developed countries. The left side view is more disorganized since there are huge differences between developing countries. And we may also say that there is a trend of less birth giving and bigger median age.

Chart, scatter chart

Description automatically generated

When using the Tsne algorithm, there is a clearer view of cluster over continents. We can see that the less developed African countries and on the bottom left, Asia, SA, and part of NA’s island countries are mixed together in the middle part. Some unique countries like China, Japan, South Korean, most part of EU and NA are on the top right. It ‘s better than MDS in general.

Chart, histogram

Description automatically generatedGraphical user interface

Description automatically generated with medium confidence

We can see that in the plot above even though the two plots are given by different algorithms, they all indicate the same results which is their will be smaller number of birth and bigger median age over time. We should also notice that the plot bellow’s y axis is starting with 1.

Chart, scatter chart

Description automatically generatedChart, scatter chart

Description automatically generatedAs we can see in the plots above, we all set the number of clusters to 6, and k means algorithm is doing a great job splitting the data into generally equal size but on the other hands, the spectral clustering algorithm are failed only detected 2 of the Cluster NO.1(Orange) which is no doing great. This is because the spectral clustering algorithm is good at detecting circle shape of clusters showed on the class not the strip shape. But some countries located in different but end up in the same pattern shows that these countries have showed the same sociology pattern due to the similar development of their counties.

Chart, scatter chart

Description automatically generatedChart, scatter chart

Description automatically generatedSame result above, but more clear view of the clusters of bottom left and top right. So the Tsne and K mean may be the best match to plot these theme.