

DATA ANALYSIS AND VISUALIZATION

HOMEWORK – 4

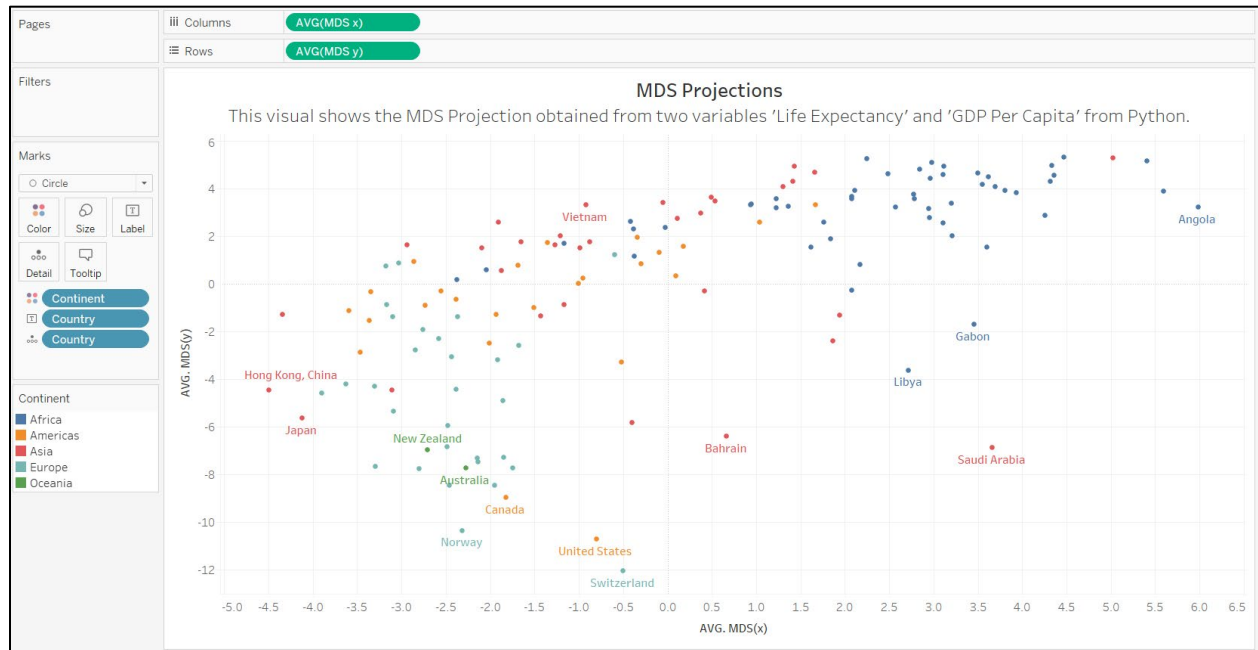
MDS Visual

Figure 1: MDS Projections on Life Expectancy and GDP Per Capita

- We combined two variables namely 'Life Expectancy' and 'GDP Per Capita' of the Gapminder Dataset, and obtained two MDS projections, i.e., MDS(x) and MDS(y).
- We have highlighted countries based on the continent's color.
- We have also highlighted (labelled) a few countries to do further analysis on it.
- Later on in this examination, we use graphics to comprehend the reasons for the proximity or distance between the chosen nations and the factors that could be responsible for it.

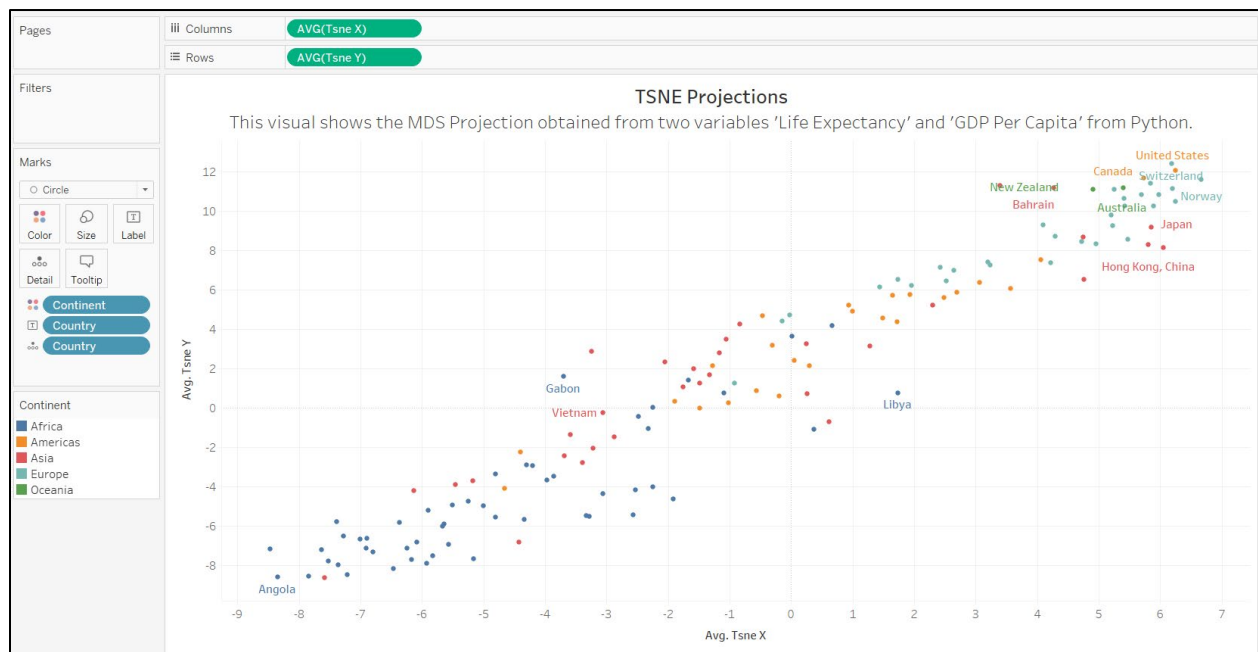
TSNE Visual

Figure 2: TSNE Projections on Life Expectancy and GDP Per Capita

- We combined two variables namely ‘Life Expectancy’ and ‘GDP Per Capita’ of the Gapminder Dataset, and obtained two TSNE projections, i.e., TSNE(x) and TSNE(y).
- We have highlighted countries based on the continent’s color.
- We have also highlighted (labelled) a few countries to do further analysis on it.
- Later on in this examination, we use graphics to comprehend the reasons for the proximity or distance between the chosen nations and the factors that could be responsible for it.

K-Means Clustering Visual

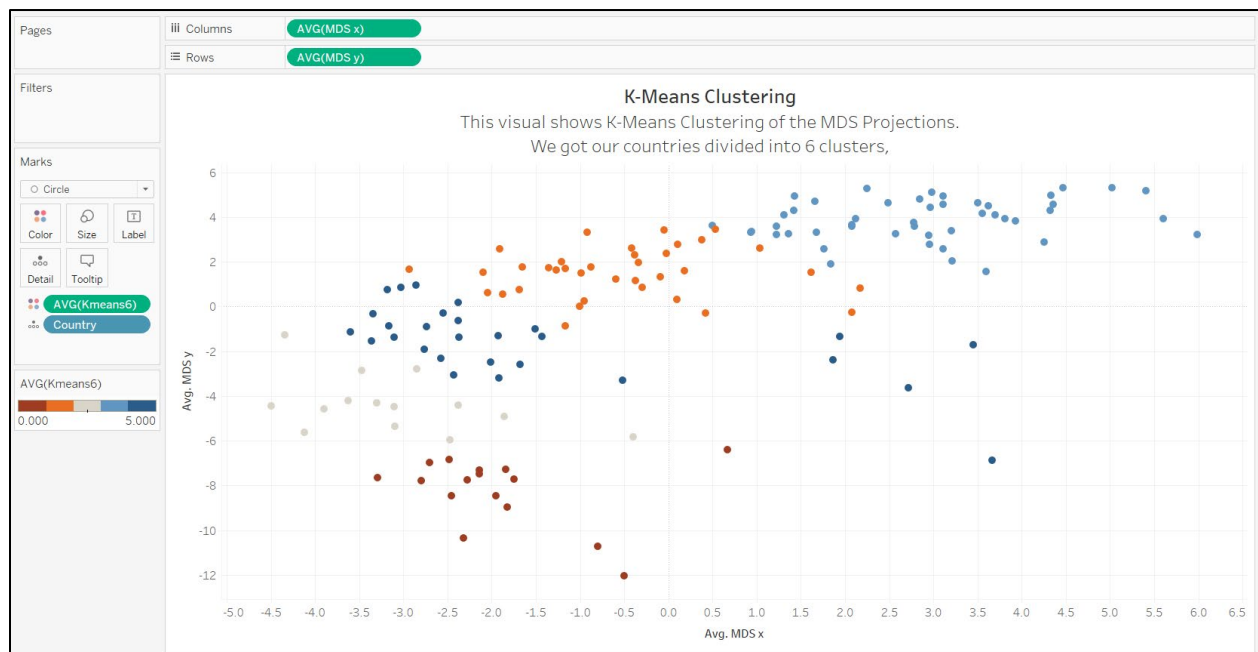


Figure 3: K-Means Clustering on MDS Projections

- We can see that countries have been divided into 6 different clusters using K-Means.
- The color scale has been used to visualize different clusters.
- This visualization uses MDS(x) and MDS(y) on x-axis and y-axis respectively, with K-Means clusters as different color.
- To find the insights about clusters, we plotted another visual which is plotted below.

K-Means Clustering (Supportive Visual)

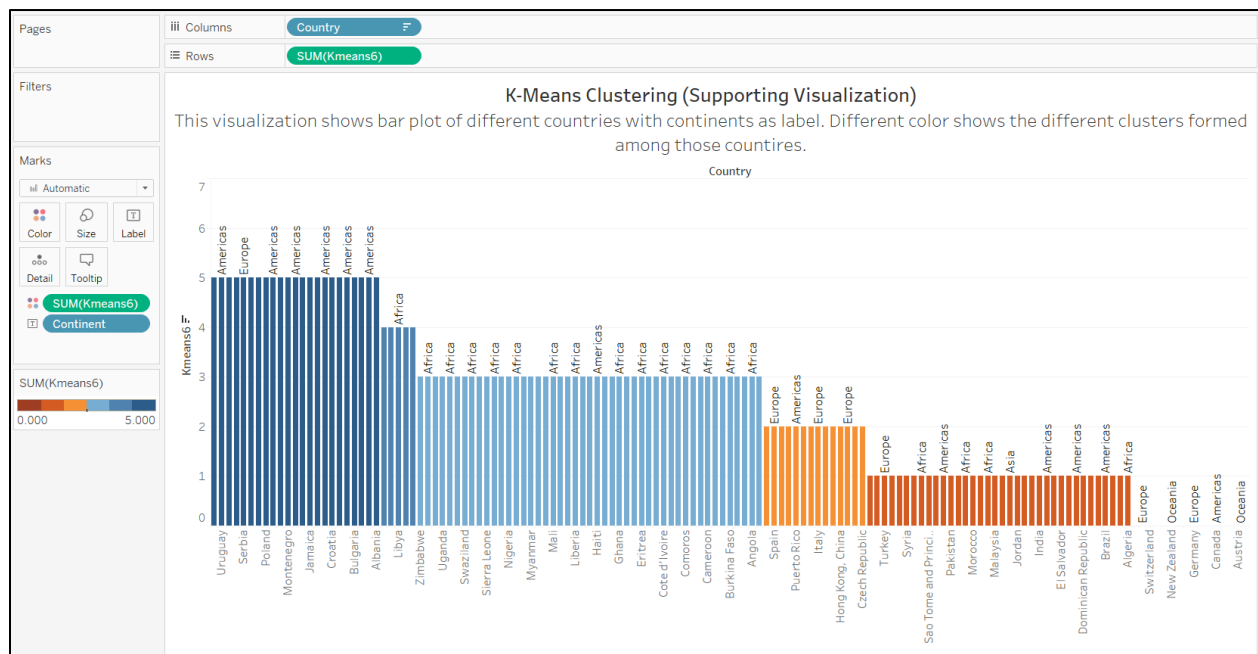


Figure 4: K-Means Clustering (Supportive Visualization)

- We can see that the clusters are formed based on the continents, that is the countries that are nearer to each other belong to one cluster.
- For example, we can see that for cluster number 5, most of the countries are from the continent 'Americas'.
- Similarly, cluster number 3 has most of the countries from the 'African' continent.
- Therefore, we can conclude from this visual that the Life Expectancy and GDP Per Capita of the neighboring countries are quite similar, and most of them belong to the same continent as well

S-Cluster Visual

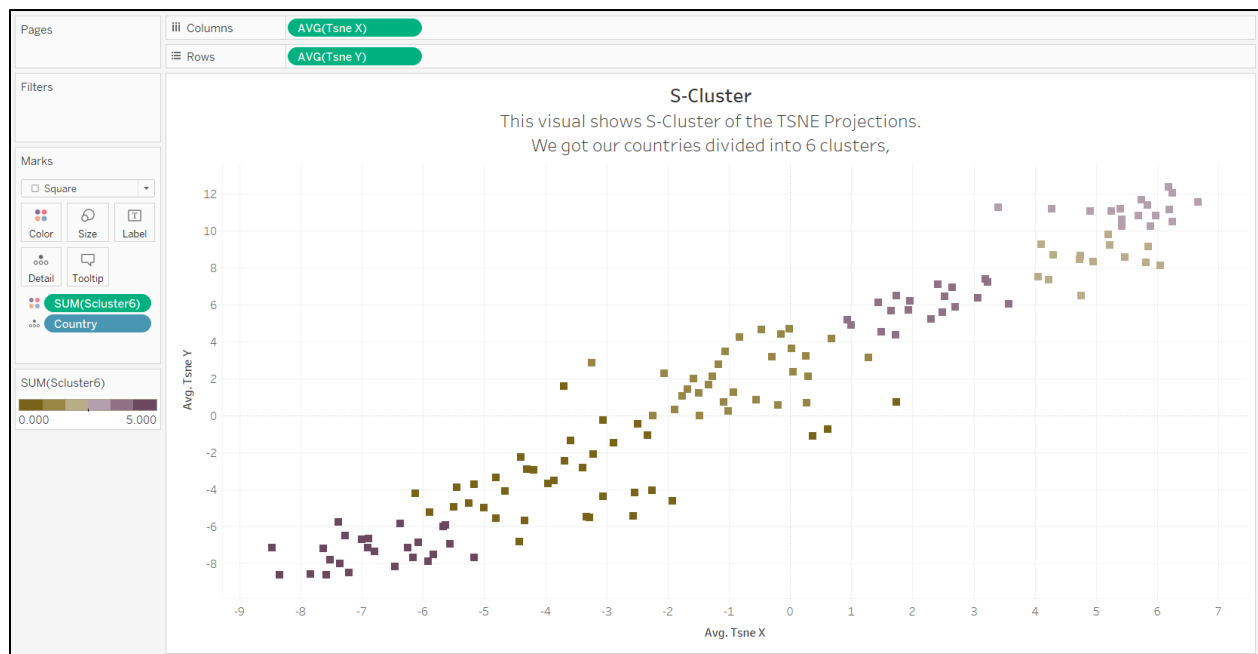


Figure 5: S-Cluster on TSNE Projections

- We can see that countries have been divided into 6 different clusters using S-Cluster.
- The color scale has been used to visualize different clusters.
- This visualization uses TSNE(x) and TSNE(y) on x-axis and y-axis respectively, with S-Clusters in color.
- To find the insights about clusters, we plotted another visual which is plotted below.

S-Cluster (Supportive Visual)

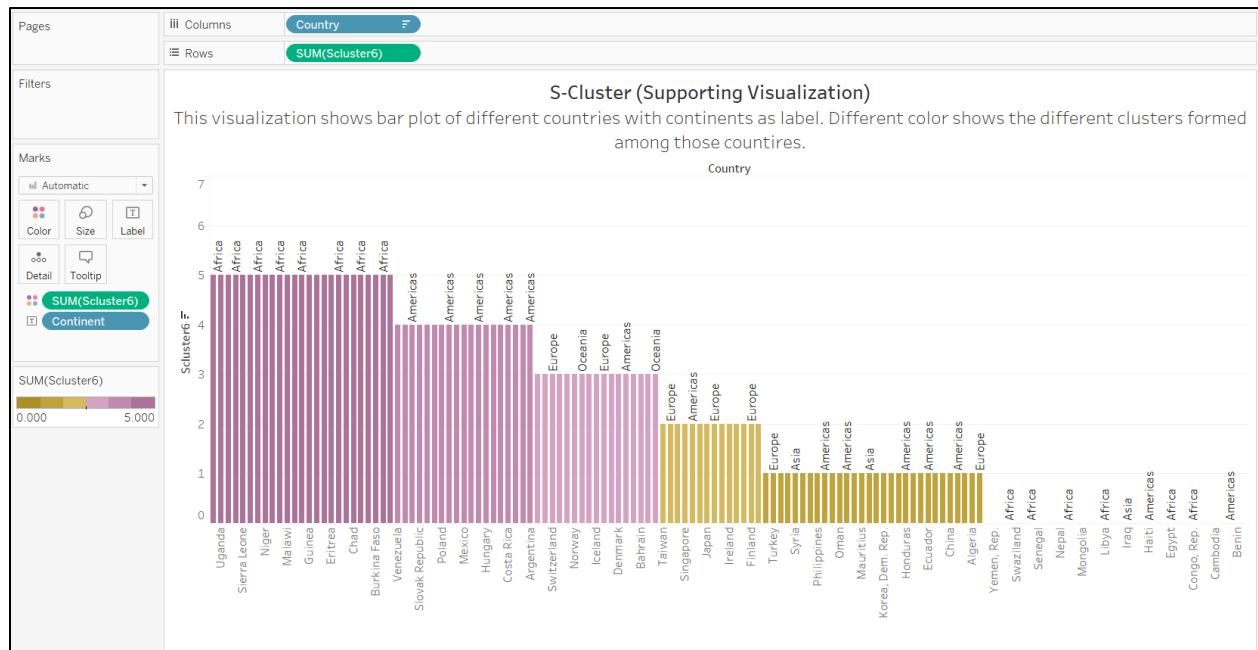


Figure 6: S-Cluster (Supportive Visualization)

- We can see that the clusters are formed based on the continents, that is the countries that are nearer to each other belong to one cluster.
- For example, we can see that for cluster number 5, most of the countries are from the continent 'Africa.
- Similarly, cluster number 4 has most of the countries from the 'American' continent.
- Therefore, we can conclude from this visual that the Life Expectancy and GDP Per Capita of the neighboring countries are quite similar, and most of them belong to the same continent as well.

Connected Scatter Plot

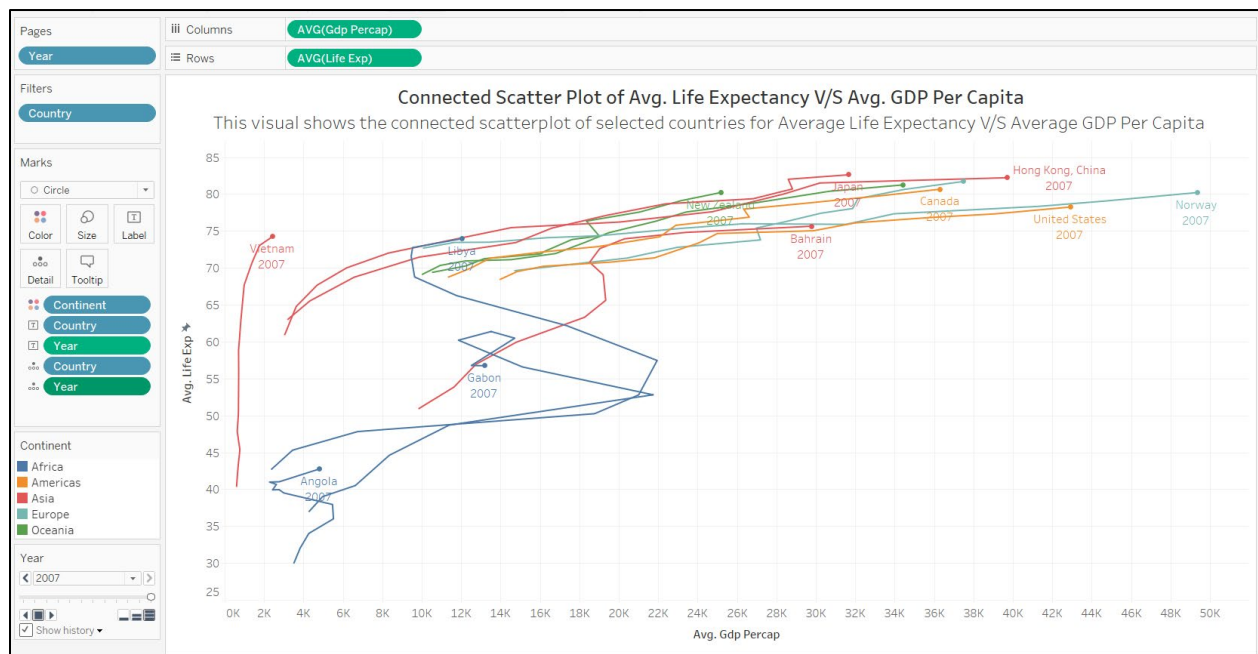


Figure 7: Connected Scatterplot of Average Life Expectancy V/S Average GDP Per Capita of selected countries

- This visual shows the connected scatterplot of average life expectancy v/s average GDP per capita of some selected countries.
- We have color coded the countries based on the continents.
- The scatterplot shows the journey from 1952 to 2007.
- We can infer from this scatterplot that the countries from K-Means clustering and S-Clusters which were together, are also nearer in this scatterplot.
- Lastly, as the average GDP per capita increases, average life expectancy also tends to increase for most of the plotted countries.

Bubble Chart

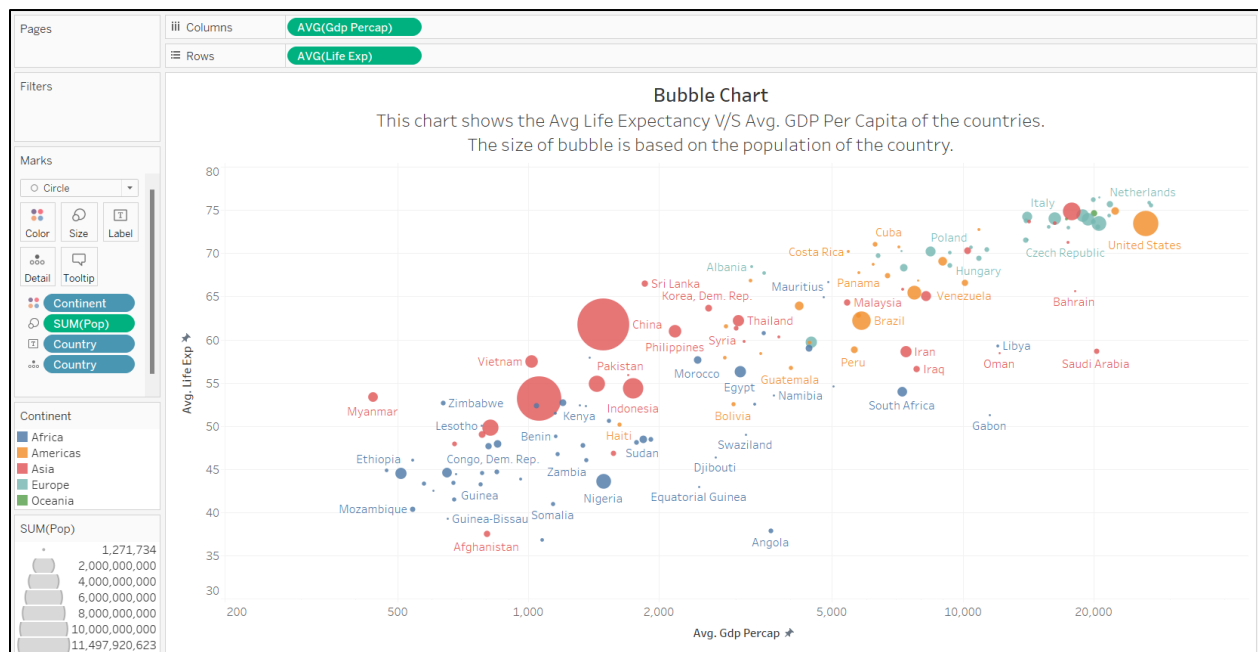


Figure 8: Bubble Chart of Average Life Expectancy V/S Average GDP Per Capita

- This visual shows the bubble chart of average life expectancy v/s average GDP per capita.
- We have color coded the countries based on the continents.
- The size of the bubble is based/denotes the population of that specific country.
- We can infer from this bubble chart that the countries which were near in the MDS and TSNE visual are very identically closely located in this chart as well.

World Map

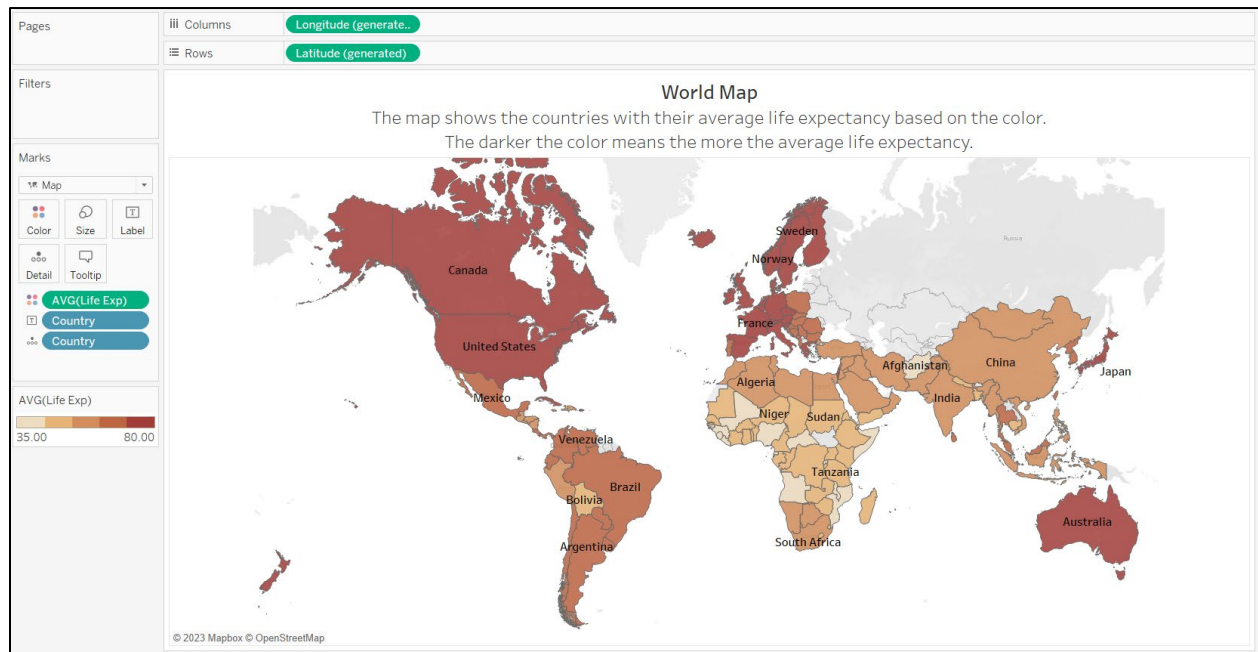


Figure 9: World Map showing average Life Expectancy of countries

- This visual shows the world map of the countries showing average life expectancy.
- We have color coded countries based on life expectancy.
- The darker the color, the higher their life expectancy.
- Similarly, the lighter the color, the lower their life expectancy.
- A few of the countries are highlighted which were quite interesting to watch out for.
- By looking at the color shades, we can clearly see that why K-means and S-Cluster were having countries in the same clusters.
- We can see that all the African countries have mostly similar life expectancy, and therefore they belonged to the same clusters in our earlier plots.
- Similarly, we can see that for Asian, and American continents as well.