



Python Case Study **Clustering the** **Countries** **WEEK 7**



Disclaimer: Country development was defined based on the socio-economic and health factors.

Dataset description: The dataset contains 167 rows and 10 columns

country	Name of the country
child_mort	Death of children under 5 years of age per 1000 live births
exports	Exports of goods and services per capita. Given as %age of the GDP per capita
health	Total health spending per capita. Given as %age of GDP per capita
imports	Imports of goods and services per capita. Given as %age of the GDP per capita
Income	Net income per person
Inflation	The measurement of the annual growth rate of the Total GDP
life_expec	The average number of years a new born child would live if the current mortality patterns are to remain the same
total_fer	The number of children that would be born to each woman if the current age-fertility rates remain the same.
gdpp	The GDP per capita. Calculated as the Total GDP divided by the total population.

Task: You will use the Country-data dataset. Your Job as a Data scientist is to categorise the countries using some socio-economic and health factors that determine under-developing, developing and developed countries. You are asked to use PCA, K-Means and Hierarchical methods.