**Question 1 (Answer)**

**Problem Statement:**

HELP International is an international humanitarian NGO that is committed to fighting poverty and providing the people of backward countries with basic amenities and relief during the time of disasters and natural calamities, In this case study, identifying the direst countries based on categorize of the countries using some socio-economic and health factors that determine the overall development of the country with the help of unsupervised machine learning technique using K-means and Principle Component Analysis

**Solution Methodology:**

The following steps has to follow to complete the task

Step 1: Importing Data Set and Understanding data as per the Business point of view

Step 2: Data Cleaning, missing value treatment and outliers treatment

Step 3: Exploratory Data Analysis (EDA)

Step 4: Applying Principle Component Analysis (PCA) for dimension reduction and also avoid multico-linearity

Step 5: Applying K-Means Clustering to make countries as a clusters based on socio-economic and health factors, etc. (features) to make funding decisions.

**Question 2 (Answer)**

**Shortcomings of Principal Component Analysis**

**1. Independent variables become less interpretable:** After implementing PCA on the dataset, original features will turn into Principal Components. Principal Components are the linear combination of your original features. Principal Components are not as readable and interpretable as original features.

**2. Data standardization is must before PCA:** You must standardize your data before implementing PCA, otherwise PCA will not be able to find the optimal Principal Components.

**3. Information Loss:** Although Principal Components try to cover maximum variance among the features in a dataset, if we don't select the number of Principal Components with care, it may miss some information as compared to the original list of features.

**Question 2 (Answer)**

**Difference between K - Means and Hierarchical clustering**

- Hierarchical clustering can't handle big data well but K - Means clustering can. This is because the time complexity of K Means is linear i.e. O(n) while that of hierarchical clustering is quadratic i.e. $O(n^2)$.
- In K Means clustering, since we start with random choice of clusters, the results produced by running the algorithm multiple times might differ. While results are reproducible in Hierarchical clustering.
- K Means is found to work well when the shape of the clusters is hyper spherical (like circle in 2D, sphere in 3D).
- K Means clustering requires prior knowledge of K i.e. no. of clusters you want to divide your data into. But, you can stop at whatever number of clusters you find appropriate in hierarchical clustering by interpreting the dendrogram