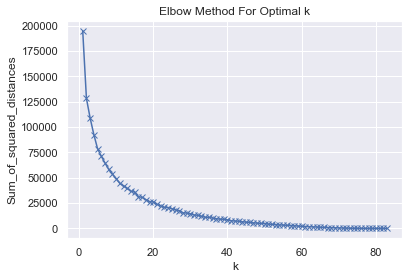
In this project we are going to cluster the countries based on the five dimensions below:

1. ***Power Distance* (PDI)**: This dimension explains the extent to which members who are less powerful in a society accept and also expect that the distribution of power takes place unequally.
2. ***Individualism vs. Collectivism*** (IDV) (Universalism vs. Exclusionism): This dimension describes preferential treatment of “in-group” members in collectivist societies vs. equal treatment for all (including strangers) in more individualistic societies.
3. ***Masculinity vs. Femininity* (MAS)**: Masculinity implies a society’s preference for assertiveness, competition, achievement and material rewards. On the contrary, femininity represents a preference for modesty, cooperation, quality of life and caring for the weak.
4. ***Uncertainty Avoidance* (UAI)**: It is a dimension that describes the extent to which people in society are not at ease with ambiguity and changes.
5. ***Indulgence vs. Restraint* (IVR)**: Indulgent societies tend to allow relatively free gratification of natural human desires related to enjoying life and having fun, whereas restraint societies are more likely to believe that such gratification needs to be curbed and regulated by strict norms.

In order to conduct the clustering, I used the data from [*countries5.csv*](https://drive.google.com/file/d/1CIQt7xP6qpDPpb0z_bU0VFy2BUMoC_M7/view?usp=sharing) collected by Prof. Barsky. Some data for IVR is missing and I filled the missing data by using global average (see the code in the other pynotebook).

The first clustering method I used is the K-means clustering method and the major problem is how to find the best value of K. Based on the method provided on the assignment document, I applied the method to compare the cluster quality based on [this example](https://blog.cambridgespark.com/how-to-determine-the-optimal-number-of-clusters-for-k-means-clustering-14f27070048f). Thus, I obtained the graph below:



From the graph, I determined the turning point as K = 15 or K = 10. Based on my personal knowledge and for simplicity, I set K = 10 and get the following results:

Cluster 1: China, Hong Kong, India, Jamaica, Singapore, Vietnam

Cluster 2: Kuwait, Lebanon, Libya, Malaysia, Morocco, Philippines, Saudi Arabia, Slovak Rep, United Arab Emirates

Cluster 3: Australia, Austria, Canada, Ireland, New Zealand, South Africa, Switzerland, United Kingdom, United States

Cluster 4: Argentina, Belgium, Czech Republic, France, Germany, Hungary, Israel, Italy, Japan, Malta, Poland, Spain

Cluster 5: Estonia, Latvia, Lithuania

Cluster 6: Brazil, Chile, Costa Rica, Greece, Peru, Portugal, Slovenia, South Korea, Suriname, Turkey, Uruguay

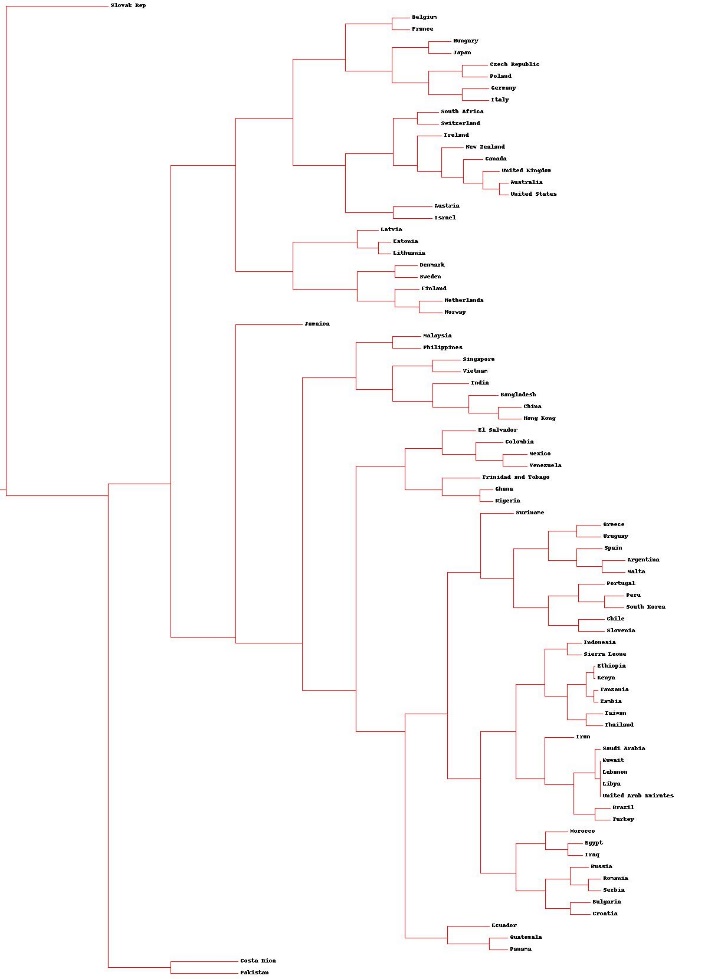
Cluster 7: Bulgaria, Croatia, Egypt, Guatemala, Iraq, Panama, Romania, Russia, Serbia

Cluster 8: Bangladesh, Ecuador, Ethiopia, Indonesia, Iran, Kenya, Pakistan, Sierra Leone, Taiwan, Tanzania, Thailand, Zambia

Cluster 9: Denmark, Finland, Netherlands, Norway, Sweden

Cluster 10: Colombia, El Salvador, Ghana, Mexico, Nigeria, Trinidad and Tobago, Venezuela

The method we developed during class was also used for hierarchical clustering, and the graph is generated as below:



By using Sklearn library I was able to make 10 clusters using hierarchical clustering, but since the are slightly worse than K-means clustering in my perspective, I decided not to include them in this report.

The description for all 10 clusters found using K-meaning clustering is:

Cluster 1: favor social order, lower individualism, neutral gender orientation, creative, restraint society

Cluster 2: strong favoring social order, favor collaboration, neutral gender orientation, moderate creativity, slightly restricted society

Cluster 3: higher individual power, favor personal rights, more aggresive, tend to be creative, happiness first

Cluster 4: lean to individual power, slightly favor individual rights, neutral gender, favor certainty, work slightly hard

Cluster 5: favor individualism, personal rights, highly modest, slightly favor creativity, work hard

Cluster 6: slightly favor social order, high individual power, tender, favor certainty, moderate work

Cluster 7: strong social order, low individual power, soft and modest, favor certainty, work hard

Cluster 8: slightly favor social order, low individual power, modest, slightly favor creativity, word hard

Cluster 9: high personal rights, high individual power, modest, favor creativity, happiness prior

Cluster 10: slightly favor social order, high individual power, neutral gender orientation, favor certainty, happiness is priority

The visualization of the clusters is within the submission and the word clouds are included as well.

**Discussion:**

Some interesting notes I found during the clustering:

1. Estonia, Latvia, and Lithuania are always clustered together, probably because their proximity.

2. Egypt is similar to Russia in that they are favoring strong social power and low individual rights. This is also similar to Iraq (considering when the data is collected)

3. To my surprise, Taiwan is not in the same cluster as China, as Hennig (2005) had shown, and is more similar to the countries in west Asia, east Africa and south Asia

4. Another interesting point is that South Korea is unlike any other Asia countries, probably because of the declining economy? I am uncertain about that.

References:

Hennig-Thurau, Thorsten & Gwinner, Kevin & Gremler, Dwayne & Paul, Michael. (2005). Managing Service Relationships in a Global Economy: Exploring the Impact of National Culture on the Relevance of Customer Relational Benefits for Gaining Loyal Customers. Advances in International Marketing. 15. 11-31. 10.1016/S1474-7979(04)15002-3.