Project Details

You work for a retail store chain in the United States of America. The company is thinking of expanding to other countries and wants to figure out which countries are similar economically, demographically, in education and environment to the United States of America.
Your manager has asked you to segment the countries of the world based on various economic, demographic, education and environmental data. From this, you should be able to provide a list of countries that are similar to the United States.

What decisions need to be made

The retail store needs to determine the countries that are similar to the United States of America, based on demographic, education, economic and environmental characteristics.

What data is needed to inform these decisions

Economic	Education	Environmental	Demographic
• GDP per Capita • Employment Rate • Total Labour Force	• Literacy Rate • Total no. of schools (primary, secondary, tertiary)	• Crime Rates • Access to Electricity • Population density	• Age/Gender • Marital Status • Life Expectancy ratio

Data Wrangling

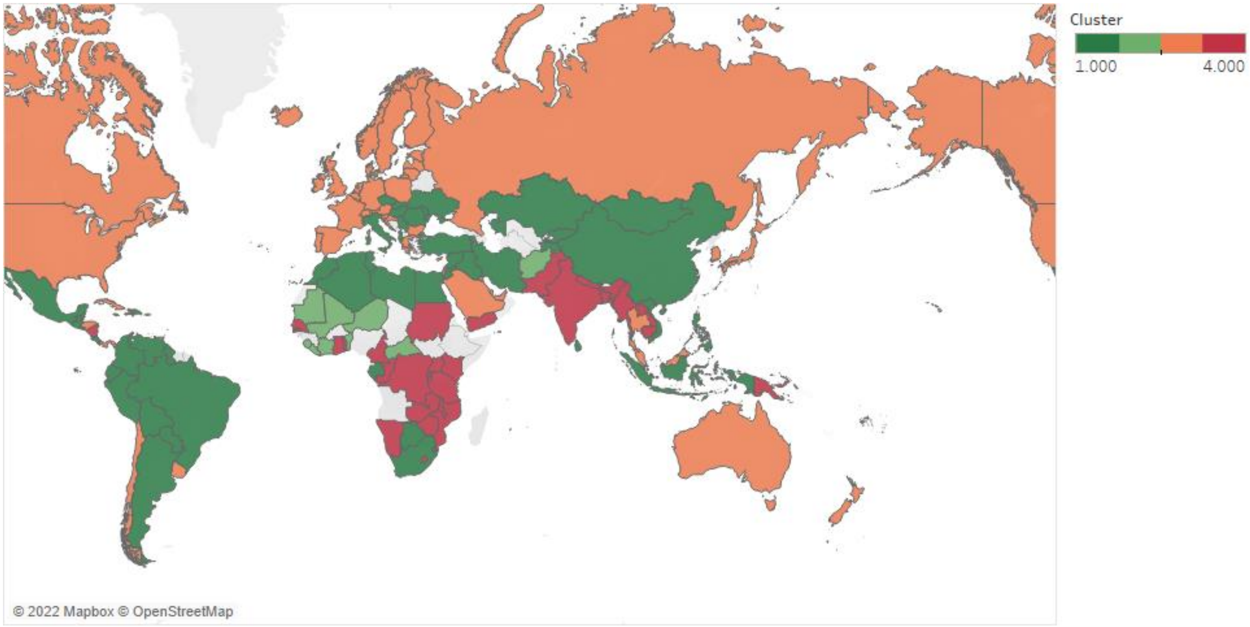
The dataset provided was gathered from World Bank and it contained 76 key indicators for 215 countries. Countries containing more than 25 missing indicators were removed from the dataset. In addition, variables not related to Education, Economy or Environment were excluded. After cleansing the data, there were 67 indicators and 144 countries left. The variables were wrangled and variables with a lot of missing values/countries were removed to prevent bias in the data.

Education literacy and years in education were decomposed using PCA to check for dimensionality in the data.

Analysis using K Means Clustering

The optimal number of clusters were tested using the Elbow method for KMeans clustering and the Calinksi Harabasz and the optimal number of clusters selected was 4. The countries and their respective clusters can be found in the table below.

Countries that fall under cluster 3 are more similar to the United States in terms of Education, Environmental factors and Demographics.



Countries similar to United States

United Arab Emirates	Australia	Austria	Belgium	Bulgaria
Bahrain	Brunei Darussalam	Canada	Switzerland	Chile
Cuba	Cyprus	Germany	Denmark	Spain
Estonia	Finland	France	United Kingdom	Greece
Hong Kong SAR, China	Honduras	Croatia	Ireland	Iceland

Israel	Japan	Korea	Kuwait	Lithuania
Luxembourg	Latavia	Macao SQR, China	Malta	Mauritius
Malaysia	Netherlands	Norway	New Zealand	Panama
Poland	Portugal	Qatar	Russian Federation	Saudi Arabia
Singapore	Slovenia	Sweden	Thailand	Trinidad
Uruguay				