

Segmentation Project: US Retail Store International Expansion

Key Business Decisions:

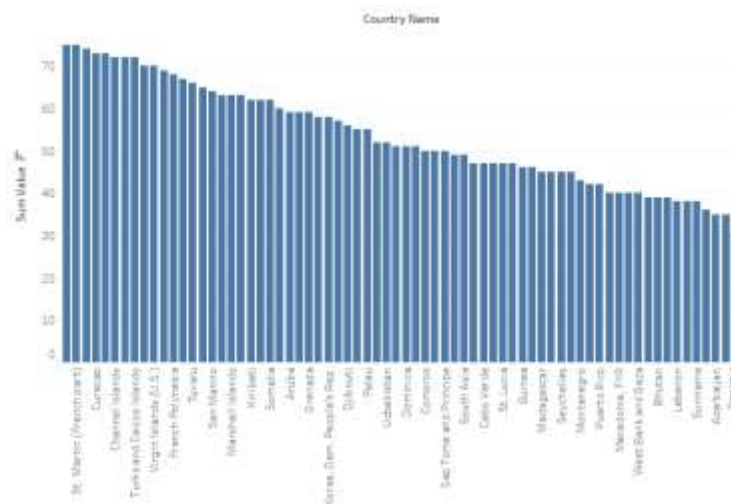
Determine which international countries are similar to the United States for which a retail store chain will expand into. The key comparison characteristics the international countries must share in common with the United States are: economic, demographic, education, and environment.

The specific types of data needed in order to complete the analysis will be data taken from the World Bank website. Sample data types with in the following segments will be used for analysis:

- Economic: Total general (local, regional and central) government expenditure on education (current, capital, and transfers), expressed as a percentage of GDP
- Demographic: Employment to population ratio is the proportion of a country's population that is employed
- Education: The percentage of population (age 25 and over) with a completed master's or equivalent degree (ISCED 7) degree as the highest level of educational attainment
- Environment: Population living in slums is the proportion of the urban population living in slum households

Explore & Clean the Data:

There were 215 countries with data provided from the World Bank. Removing countries with incomplete or missing data (> 25 null data points), total of 144 countries will be used for analysis. Below is chart highlighting the countries removed from eligibility along with the total number count of null values.



In order to represent the most accurate data, variables not relevant to the analysis are removed from the dataset. Below is a list of 9 variables that have been removed from the dataset:

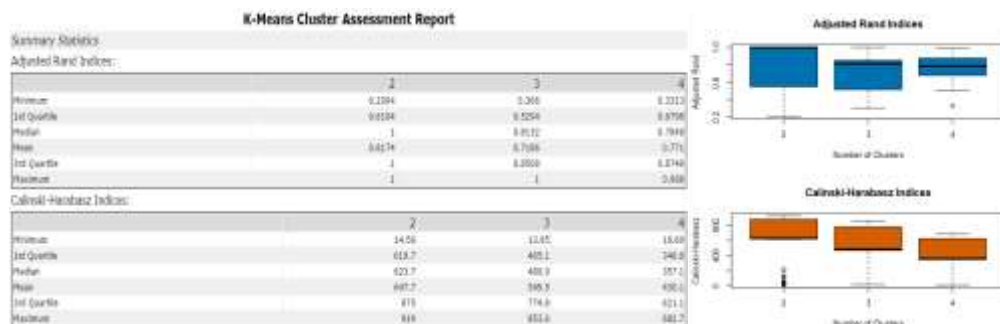
<i>Series Code</i>	<i>Category</i>	<i>Definition</i>
IT_NET_USER_P2	Background	Internet users are individuals who have used the Internet (from any location) in the last 12 months
SH_DYN_AIDS_ZS	Background	Prevalence of HIV refers to the percentage of people ages 15-49 who are infected with HIV.
SH_DYN_MORT	Background	Under-five mortality rate is the probability per 1,000 that a newborn baby will die before reaching age five.
SH_MED_PHYS_ZS	Health	Physicians include generalist and specialist medical practitioners.
SH_XPD_PCAP	Health	Total health expenditure is the sum of public and private health expenditures as a ratio of total population.
SN_ITK_DEFC_ZS	Health	Population below minimum level of dietary energy consumption shows the percentage of the population whose food intake is insufficient to meet dietary energy requirements continuously.
SP_POP_DPND	Health	Age dependency ratio is the ratio of dependents--people younger than 15 or older than 64--to the working-age population--those ages 15-64.
SG_VAW_BURN_ZS	Health	Percentage of women ages 15-49 who believe a husband/partner is justified in hitting or beating his wife/partner when she burns the food.
SH_TBS_PREV	Health	Prevalence of tuberculosis is the estimated number of TB cases (all forms) at a given point in time, expressed as the rate per 100,000 population.

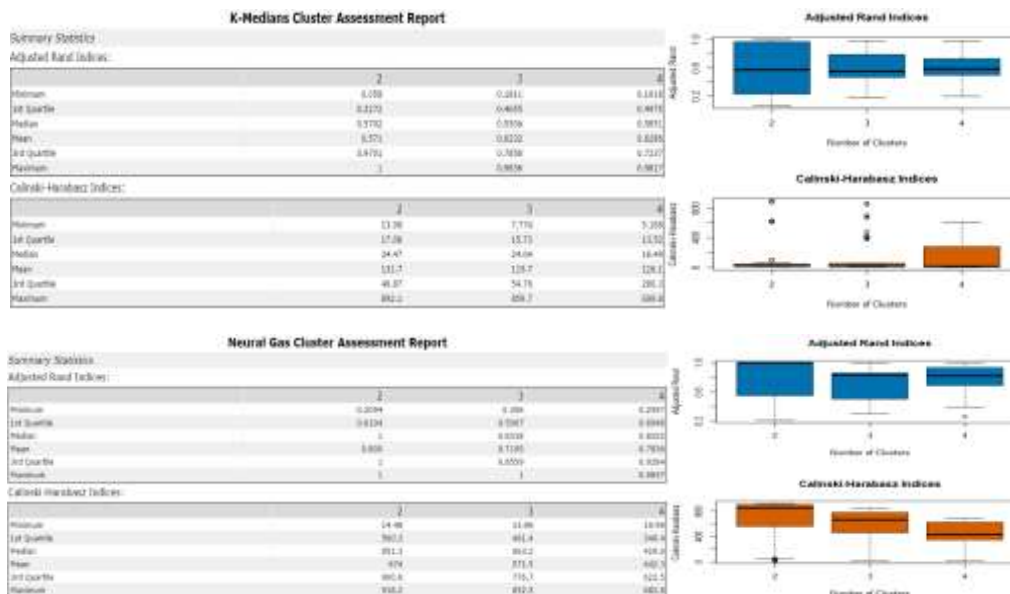
Determine Clusters & Methodology:

Building cluster models is considered an iterative process. The goal of running 3 different clustering methods (K-Means / K-Median / Neural Gas) is to compare the clustering models against each other to determine the best method. The best clustering method has the following characteristics:

- High median value (value closest to 1)
- Tight box & whisker plot (1st quartile & 3rd quartile are relatively close to each other in value)

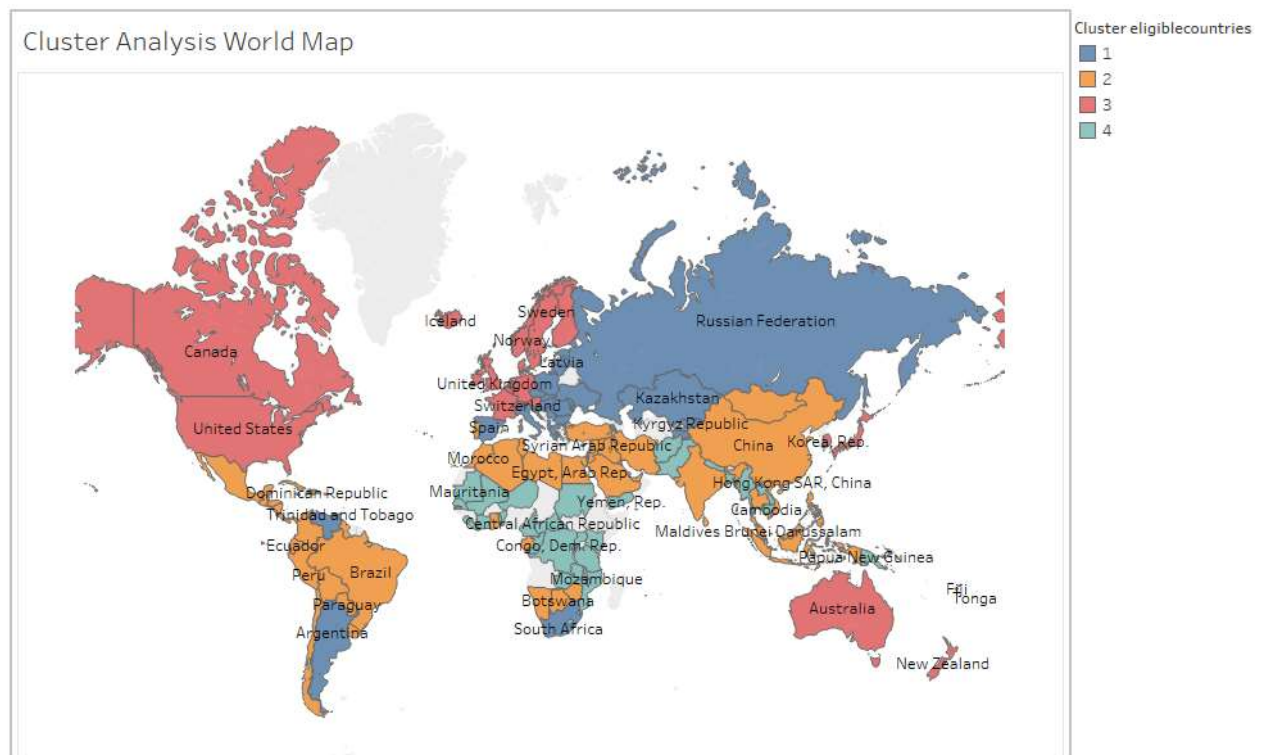
From the 3 different clustering methods, the Cluster 4 Neural Gas method best displayed the above listed characteristics.





Run the Data & Validate:

The clustering model has resulted in highlighting countries that are quite similar to the United States on the following 4 characteristics: economic, demographic, education, and environment. The below listed are countries that are clustered closest to the characteristics similar to the United States:



Country.Name
Australia
Austria
Barbados
Belgium
Canada
Denmark
Finland
France
Germany
Hong Kong SAR, China
Iceland
Ireland
Japan
Korea, Rep.
Luxembourg
Netherlands
New Zealand
Norway
Sweden
Switzerland
United Kingdom
United States

Recommendation:

It is recommended that the retail store chain consider the above listed countries to expand internationally. The countries listed have the characteristics closest to the United States based on economic, demographic, education, and environment traits.