**Project Overview**

You work for a retail store chain in the United States of America. The company is thinking of expanding to other countries and want to figure out which countries are similar economically, demographically, education, and environment to the United States of America. You are tasked with finding out which countries fit this criteria.

**How Do I Complete this Project?**

This project uses skills learned throughout the "Segmenting and Clustering ” course. To complete this project:

* Go through the course
* Apply the skills learned in the course to solve the business problem given in the project details section.
* Use our guidelines and rubric to help build your project.
* When you're ready, submit it to us for review using the submission template found in the supporting materials section.

**Skills Required**

In order to complete this project, you must be able to:

* Cleanup, format, and blend a wide range of data sources
* Segment out data and run K-means/K-median/Neural Gas clustering to find clusters in a dataset

**Project Details**

You work for a retail store chain that only has a presence of stores in the United States. The company is thinking of expanding to other countries, but would like to start this process with countries that are similar economically and demographically to the United States.

Your manager has asked you to segment the countries of the world based on various economic, demographic, education, and environment data. From this you should be able to provide a list of countries that are similar to the United States. This will be the “short list” for further consideration by management.

You have the following information to work with:

A set of various data taken from the [**World Bank web site**](http://databank.worldbank.org/data/databases.aspx?qterm=&pagenumber=1). A clean version of this data is included with the supporting materials.

**Steps to Success**

**Step 1. Key Decisions**

Determine the key business questions and specific types of data you need to look at in order to complete this analysis.

**Step 2. Explore and Cleanup the Data**

Clean up the data in the provided CSV file. The data contains 215 countries with 77 variables.

Not all of the countries have complete data and not all of the variables may be applicable for the analysis. Based on the request from your manager, you will need to:

* Remove countries with a high number of missing data points (>25). HINT: There should be 144 countries left.
* Remove variables that are not relevant for the analysis (any variable not under the economic, education, demographic, or environment categories). HINT: You should remove 9 variables.

**Step 3. Determine Clusters and Methodology**

Your manager would like to see four clusters in the results. You also need to determine which clustering method provides the optimal clusters.

**Step 4. Run the Data and Validate**

After you run the model with the data you’ve decided to use, append the Cluster IDs back to the data in order to visualize the data.

Feel free to use any visualization software you want, such as Tableau.

**Step 5. Recommendation**

Provide your recommendations and justify why your countries should considered to be explored first for international expansion.

### Review

Use the [**project rubric**](https://review.udacity.com/#!/rubrics/424/view) to review your project. If you are happy with your submission, then you're ready to submit your project. If you see room for improvement, keep working to improve your project.

### Submission Template

Use the submission template at the bottom of this section to submit your project. After filling it out, save it as a PDF and submit the PDF in the next section.

#### Data

country\_data.csv - This file contains country information

variables\_catalog.csv - This file contains descriptions for the data fields in country\_data.csv

#### Supporting Materials

**[Country Data](https://d17h27t6h515a5.cloudfront.net/topher/2016/October/57f40d44_country-data/country-data.csv" \t "_blank)**

**[P7 Submission Template](https://d17h27t6h515a5.cloudfront.net/topher/2016/December/58498118_p7-submission-template/p7-submission-template.docx" \t "_blank)**

**[Variables Catalog](https://d17h27t6h515a5.cloudfront.net/topher/2016/December/5849817d_variables-catalog/variables-catalog.csv" \t "_blank)**