International Segmentation of Countries Using Unsupervised Machine Learning Algorithms

Udara Rangana Jayathilaka
Department Of Global Business Management,
Gangseo University, Seoul, South Korea.

ORCID

0000-0002-1682-9812

Abstract

As markets grow more global, global market segmentation becomes increasingly important for creating, marketing, and selling items globally. Operating a business in many nations introduces unique obstacles. International markets are more diverse than domestic markets due to differences in historical, institutional, cultural, and political environments among nations. Clustering algorithms enable multinational companies to give better-personalized products and customer services to their foreign customers. By gaining a deeper understanding of the nation's characteristics, companies may identify the appropriate products or services to distribute, choose the best marketing channels for the target consumers, discover new and important insights, and launch new business strategies. This study used unsupervised machine learning algorithms such as Affinity Propagation, DBSCAN, and Hierarchical clustering to segment 150 countries. The segments of countries were established based on their brand awareness, gross national income, and trade liberalization, which are considered to be some of the most relevant qualities to employ when determining the macro segmentation of countries in the context of international business. This research emphasizes and recommends that various machine learning technologies be used to construct segmented and countrywide business strategies and marketing tactics in order to advance the global market expansion.

Keywords: *Hybrid work, Affinity Propagation, Clustering, Market segmentation, Multinational companies*

Introduction

Understanding international market and consumer base is crucial to the sustained expansion of many successful multinational corporations. Market segmentation is a prominent approach that organizations may use to better focus their goods to the correct clients.

Segmentation was formerly conducted on a national basis, before the advent of globalization. The conventional wisdom in international marketing was to choose nations for exporting or more sophisticated kinds of international commerce based on macro characteristics such as economic wealth, infrastructure, system of government, and society, trade, and then divide the chosen markets. As globalization has increased, there has been significantly more discussion regarding international segmentation, which is defined as segmenting a collection of prospective clients that crosses national lines.

The process of partitioning its overall worldwide market into one or more pieces (segments or sub-markets), with each tends to be homogenous in all major features, is referred to as global market segmentation [1]. International market segmentation can be described as the practice of dividing the global consumer market into subsets of buyers with comparable preferences and habits. Distinctions and resemblances among prospective purchasers in overseas markets may be found and classified using global market segmentation.

International segmentation is classified into three types: international, regional, and distinctive. When there is a collection of customers with shared demands that cuts across many nations, a global segmentation strategy is used; when there is a cluster of consumers with comparable wants and preferences in a single area or a few nations, a regional or a distinct segmentation strategy is called for; and when customer expectations are very country-specific, a unique segmentation strategy is employed [2].

International market segmentation techniques have emerged as a critical challenge in the development, marketing, and sale of goods across national boundaries. It enables businesses to target prospective clients at the worldwide level and achieve a suitable cross-border positioning. A fundamental difficulty for businesses is efficiently dealing with the complexity of diversity in customer demands and desires across borders, as well as targeting consumer groups in various nations. These segments represent geographic groups or groups of persons and are made up of prospective customers who are expected to respond similarly to marketing activities.

International segmentation enables businesses to determine their positioning approach. The following are some of the main advantages of segmenting the overseas markets mentioned in international business literature: it allows for the generalization of market research findings; samples can be obtained from many segments rather than each market; customers' requirements can be more clearly recognized; growth prospects can be discovered; and marketing messages can be better targeted [3], [4]. To get these advantages, sectors should be as homogeneous as possible.

In essence, if research from many markets show significantly comparable outcomes, they may be combined to establish a segment. As a consequence, insights from one market may be applicable to the remaining of the sector as well.

One of the most significant advantages of segmenting the market is that it may help businesses acquire a competitive edge over its rivals. Marketing segmentation helps to examine different consumer groups and target particular items to fulfill their needs [5].

A segmentation plan allows company's marketers to research the demands of each category and understand its rivals' present offers. This allows businesses to uncover market possibilities for increased client satisfaction. When the market segmentation strategy identifies disappointed clients, company may join this market by offering the items or services they need.

Leveraging customer segmentation models enables businesses to create goods that are properly tailored to the needs of their target audience. Among market segmentation's many advantages is its potential to provide light on consumers' rational expectations of the company's product or service. A market segmentation plan enables to assess what is performing well for the company and what is not, providing management with the information they need to make the necessary adjustments to better fulfill the demands of the customers[6].

When the emphasis is fixed and target market are clearly determined by the marketing and management team, the competition begins. The management team will bring fresh unique concepts to boost their brand effectively in order to differentiate products as the target market becomes known to the firm. Different promotions and discounts can assist the marketing team attract more customers, and getting to know your customers better will help develop brand loyalty.

Customer retention is critical for businesses, and market segmentation aids in this process. When a company understands its customers, it can better respond to their requirements, and the customer 's interaction with the market brand may assist them in connecting with the product/service. Consumers often favor the products or services they have previously used, particularly in the case of airlines and healthcare facilities.

Market segmentation leads or steers a company's business activities to be customer and market-oriented within a particular market segment, allowing enterprises to serve their customers more effectively and increase customer happiness, which is the primary purpose of the business [7].

Time, money, and other resources may be saved without sacrificing effectiveness in a marketing campaign if proper market research and segmentation are conducted. Furthermore, since consumers are classified based on their requirements, similarities, preferences, statuses, and so on, marketers may more easily address their customers. Thus, market segmentation allows the campaign management operation to run smoothly and effectively while being cost-effective.

One of the most significant advantages of market segmentation seems to be that it allows businesses to choose the optimal moment to undertake certain marketing initiatives. By dividing target audience into subgroups, a business can determine whether it is the right moment to launch a marketing campaign, or if it should wait. Customers are more inclined to react and believe that a company genuinely understand their demands when the company identifies the optimal timing to launch particular marketing campaigns [8].

Segmentation, Global and Macro-segmentation:

Market clustering is the process of splitting a targeted customer into clusters or subgroups depending on various factors. Market segmentation lets businesses identify its prospective clients based on their habits, interests, demography, interests, as well as other aspects, rather than developing a single marketing strategy for all of the consumers.

Even yet, totally standardized marketing tactics cannot serve the whole diversified population. Many businesses know that groups of customers from various nations sometimes have more similarities than consumers from the same country [9]. As a result, they opt to service parts that cross national boundaries.

Organizations may segment overseas markets by employing one or more factors. They may divide the world into areas based on geography, such as Central Europe, the Arab World, Latin America and the Caribbean, or the Western Pacific. Geographic segmentation suggests that neighboring countries will share many features and habits. This is usually the case, although there are several exceptions. Although the USA and Canada have many similarities, they are both economically and culturally distinct from neighboring Mexico.

Economic considerations may also be used to separate global marketplaces. Nations, for instance, might be classified according to their degree of economic advancement and population prosperity [10]. The economic structure of a nation determines its population's product requirements and, as a result, the marketing possibilities it provides.

Countries may be classified according to political and legal variables such as form of governance and social stability, openness to international enterprises, monetary rules, and bureaucratic depth and complexity. Such variables are critical in determining which nations to approach and how to enter them. Cultural variables, such as similar language, faith, values, and behaviors, may also be utilized to classify markets [11], [12].

Segmenting international markets based on the aforementioned variables indicates that segments will be made up of clusters of nations. Many businesses, however, choose a distinct strategy known as inter-market segmentation. Using this method, businesses may create groups of customers that have comparable wants and purchasing habits while being in different nations [13]. Luxury cars, for example, targets the world's rich groups irrespective of their nation of origin.

Customer factors like as age, gender, socioeconomic status, attitude, customer loyalty, product consumption, and sentiments about a certain brand are frequently considered as segmentation basis in domestic markets [14]. In international marketplaces, however, a second feature, namely country features, must be addressed. International markets can thus be split in two steps.

First, depending on national market characteristics, a macro segment formed of individual or groups of nations may be determined. The market may then be further subdivided depending on client characterization within each macro-segment.

Macro-segment approach, also known as country-based segment approach, discovers groups of nations that have common buying patterns. Macro-segmentation groups nations into market segments based on geographic, demographic, and socioeconomic criteria such as geography, Gdp per capita, size of population, or household size, and then determines one or several segments to design marketing techniques for each of the chosen segments. This method allows a business to consolidate its activities and save money on manufacturing, sales, transportation, and customer support [15]. However, macro-segmentation ignores consumer distinctions inside each nation and across country markets that are grouped together, as well as the presence of segments that extend beyond the limits of a certain geographic area. As a result, the corporation may be operating at a loss by passing up possibilities to meet the needs of customer groups throughout different geographical segments. Macro-segmentation leads to inaccurate national stereotyping, which neglects within-country variability. Ignoring similarities in requirements across national lines leads in governments missing economies of scale advantages that may be obtained by fulfilling the needs of a larger population across national (macro-segment) borders [16].

While there are several segmentation methodologies, two stand out: micro and macro clustering. The distinction between the two segmentation procedures is the amount of precision needed by each. Micro segmentation concentrates on customer particular data such as favorite items, experience with the brand, and date since last purchase, while macro segmentation concentrates on high level consumer data such as region, language, or origin.

Because of the variances in these methodologies, the segments produced by macro and micro classification procedures will vary in size and variation. Macro segmentation produces bigger groups with a greater degree of variability, making marketing personalization tougher to perform. Micro segmentation, on the other side, will result in tiny segments with few, homogeneous clients.

Although marketers often begin with macro segmentation techniques, as they grow and become more skilled, they shift to micro clustering as their primary approach. By grouping nations together based on their desire for comparable goods, "macro-segmentation" or "country-based segmentation" reveals patterns in global consumption [17], [18]. Using factors like geography, GDP per capita, population size, and family size, macro-segmentation divides the global market into distinct chunks that can be more easily targeted with advertising. This method allows a business to consolidate its activities and save money on manufacturing, sales, logistics, and support.

Micro-segmentation also categorizes customers into groups according to the similarities in their preferences, beliefs, and attitudes, as well as their lifestyle choices. This method makes use of psychographic and/or behavioristic segmentation factors. One of the key distinctions between referring to a company as "international" or "global" is based on the breadth and basis of segmentation. An international company is one that focuses on a few national marketplaces. A global company is one that creates segmentation plans based on responsive portions of the market, regardless of where in the world those segments may be located, and considers the whole world to be a significant part of its overall market.

Methodology

Clustering, also known as clustering algorithms, is a procedure that organizes objects based on their shared similarities. This method identifies similar structures within an unlabeled data set. It does this by dividing the available heterogeneous datasets into subgroups so that each cluster is more comparable to each other than the whole set. In other words, these clusters are collections of samples that are comparable to one another but unique from samples identified in other clusters.

Hierarchical clustering

The initial phase of hierarchical clustering is to regard each observation as its own independent cluster. Following that, it repeats the following two stages: (1) Identify the two distinct clusters that are most closely connected to one another, and then (2) merge the clusters with the highest degree of similarity [19].

This iterative procedure will proceed until all clusters have been merged into one. When the approach is performed correctly, the distances are stored in a structure called the proximity matrix, designated by the symbol M. This matrix indicates, at the location indicated by the symbols "i,j", the distance between the ith and jth clusters. $M_{ij} = M(X_i, X_j)$.

Affinity Propagation

Affinity Propagation generates clusters by transmitting messages between sample pairs until convergence. After that, a dataset is characterized by employing a limited set of exemplars, which are samples that have been determined to be the most typical of other types of data. The messages that are sent between pairs provide a representation of the appropriateness of one sample to serve as the example of the other, which is then modified in accordance with the values provided by other pairings. This update takes place in cycles until it reaches convergence, at which time the ultimate exemplars are selected, and the ultimate clustering is determined as a result. In a manner that is somewhat more formal, the obligation of a sample *k* to serve as the example of sample *j* is stated by the following [20]:

$$r(j,k) \leftarrow s(j,k) - max[a(j,k') + s(j,k') \forall k' \neq k]$$

Where *s(j,k)* denotes the degree to which samples j and k are comparable to one another. Because of the following factors, it is possible for sample k to serve as an example for sample j:

$$a(j,k) \leftarrow min[0, r(k,k) + \sum_{j' \; s.t. \; j' \notin \{j,k\}} r(j',k)]$$

All of the values for r and 'a' are initialized to zero, and then the computation for every iterates until convergence is reached. The factor lambda is added to the iteration process in order to prevent numerical oscillations during the process of revising the messages:

$$r_{t+1}(j,k) = \lambda \cdot r_t(j,k) + (1-\lambda) \cdot r_{t+1}(j,k)$$

$$a_{t+1}(j,k) = \lambda \cdot a_t(j,k) + (1-\lambda) \cdot a_{t+1}(j,k)$$

The Silhouette Coefficient is calculated for each individual sample and is made up of two scores, which are as follows [21]:
a: The average distance between a given point and all of the other points that belong to the same category.
b: The approximate difference between a sample and all of the other points that make up the subsequent closest cluster.

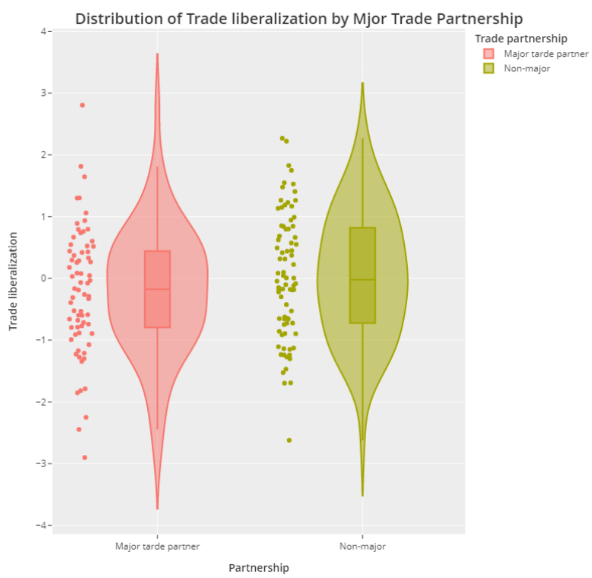
$$s = \frac{b-a}{max(a,b)}$$

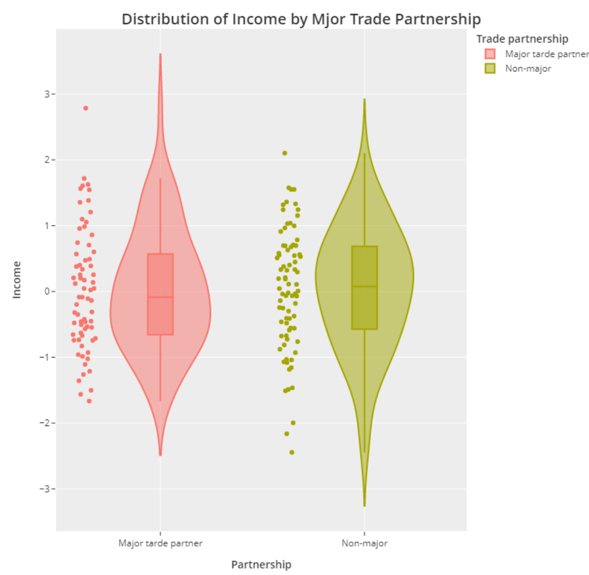
DBSCAN

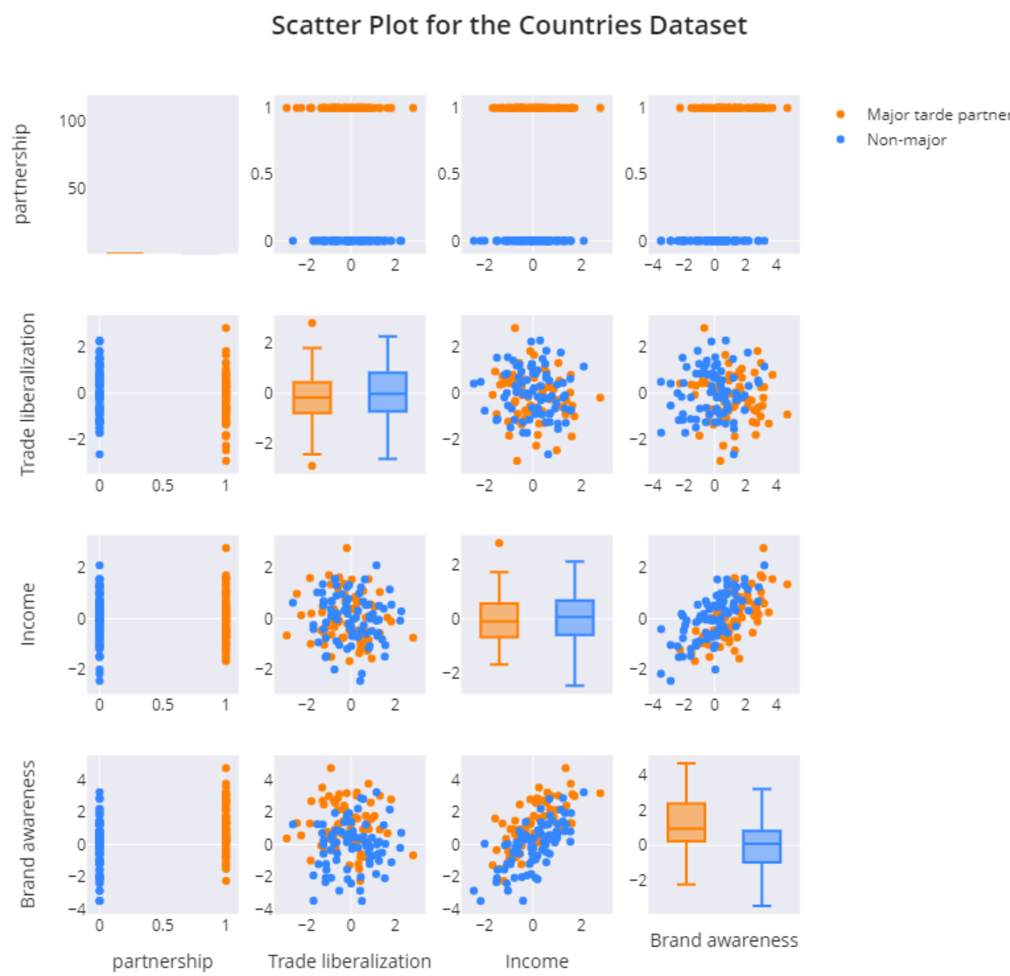
In the DBSCAN approach, the cluster number is determined not by making an estimate but rather by specifying two model parameters called epsilon and n. This is in contrast to other methods, which make an approximation. Epsilon is the distance used to detect points or assess the density in the vicinity of a specific sample. n is the minimum number of points (a threshold) that must be clustered together for a region to be considered dense [22], [23].

Initially, algorithms randomly select a location (one sample) x from the incoming data and store it in a group 1. The algorithm then identifies and counts how many observations are located at a distance of delta from x. If this number is larger than or equal to the needed minimum, it will be designated a core point and will unite all of these neighbors into a single group 1. Then, it will examine every element in cluster 1 to determine each member's distinct neighbors. If a member of cluster 1 has n or more neighbors, they will be added to the cluster. The extension of group 1 will proceed until there are no new samples to add. In the later case, a random sample of data that does not correspond to any group will be selected and assigned to cluster 2. This procedure will continue until all of the observations have been grouped into clusters. As an outlier, the sample points which do not correspond to any group are detected.

Results



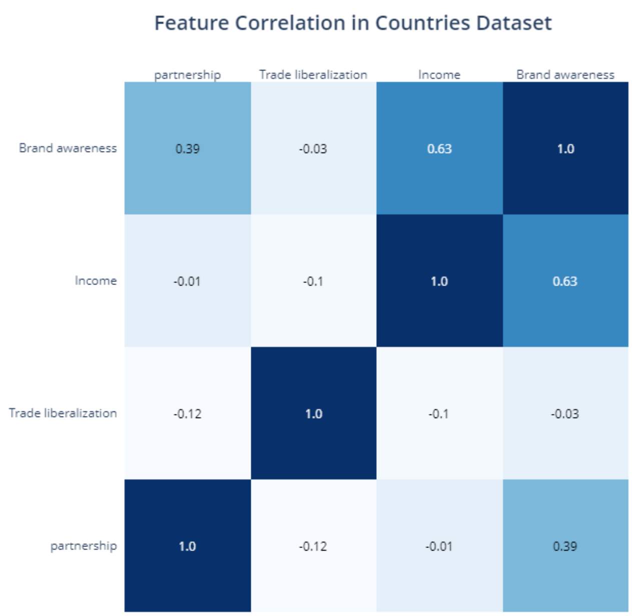


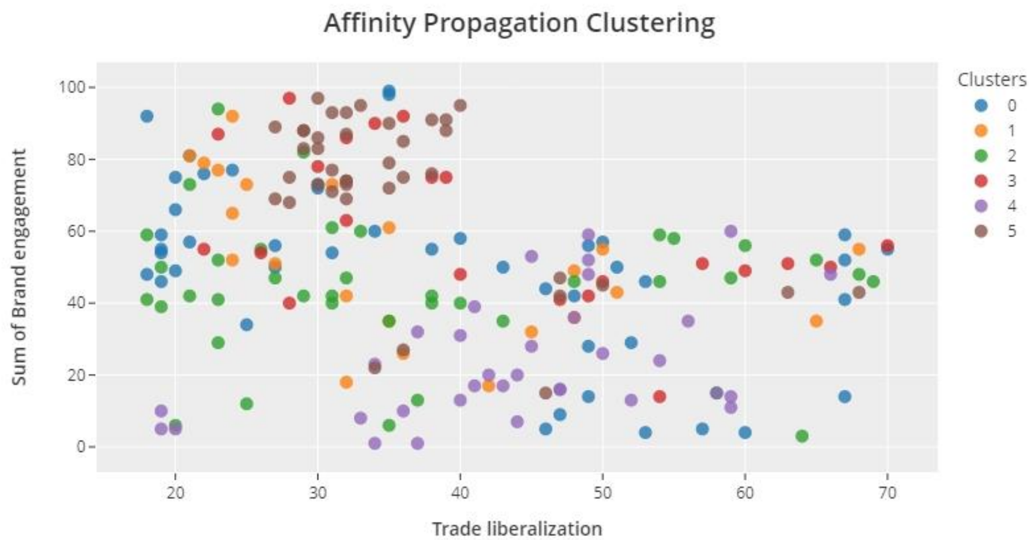


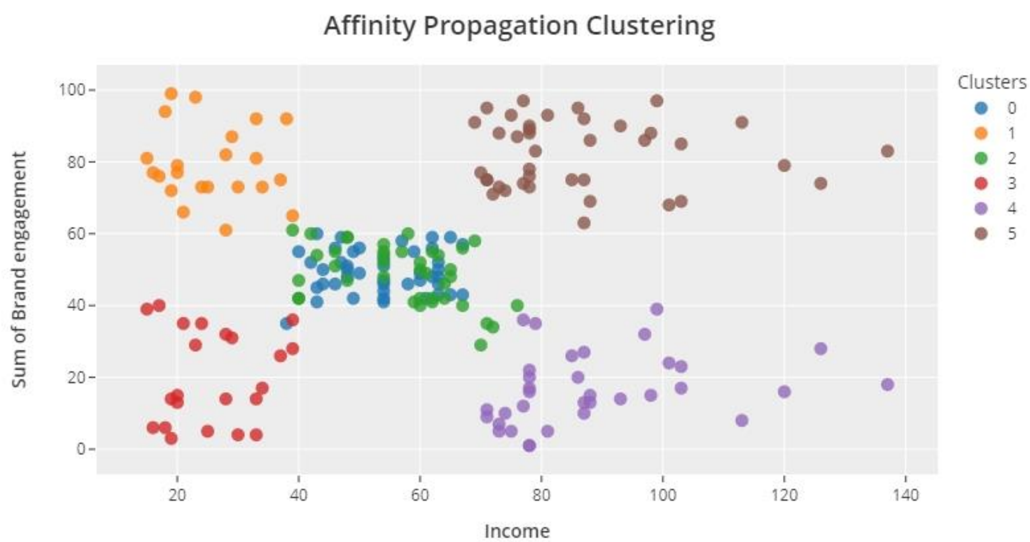
We found that brand awareness is higher in major trade partner countries. According to the results, however, a higher trade liberalization in a foreign country does not increase the brand awareness. This might be due to the fact that consumers now days are well aware of good overseas brands through social media even they never consumed the brands. The scatter plot shows that a higher gross national income is associated with a higher brand awareness.

In addition to this, we used the scatterplot matrix that was associated with the dataset in order to investigate the distribution of the different data points that correspond to the various features. However, scatter plots make it difficult to understand and recognize the connections that exist between the various components of the dataset. Because of this, drawing inferences from the data is a challenging endeavor.

There is a weakly positive correlation between a few of the characteristics and a number of others. For example, given that the correlation between brand awareness and gross income is just 0.01, one may be led to conclude that these two aspects have no effect on one another. In addition to this, the dataset includes several very weak instances of negative associations. For instance, the connection between Brand awareness and Trade liberalization is -0.330, which means that nations with low levels of Trade liberalization also have low levels of Brand awareness.




Affinity Propagation Clustering
Sum of Brand engagement
Trade liberalization
Clusters
0
1
2
3
4
5


Affinity Propagation Clustering
Sum of Brand engagement
Income
Clusters
0
1
2
3
4
5

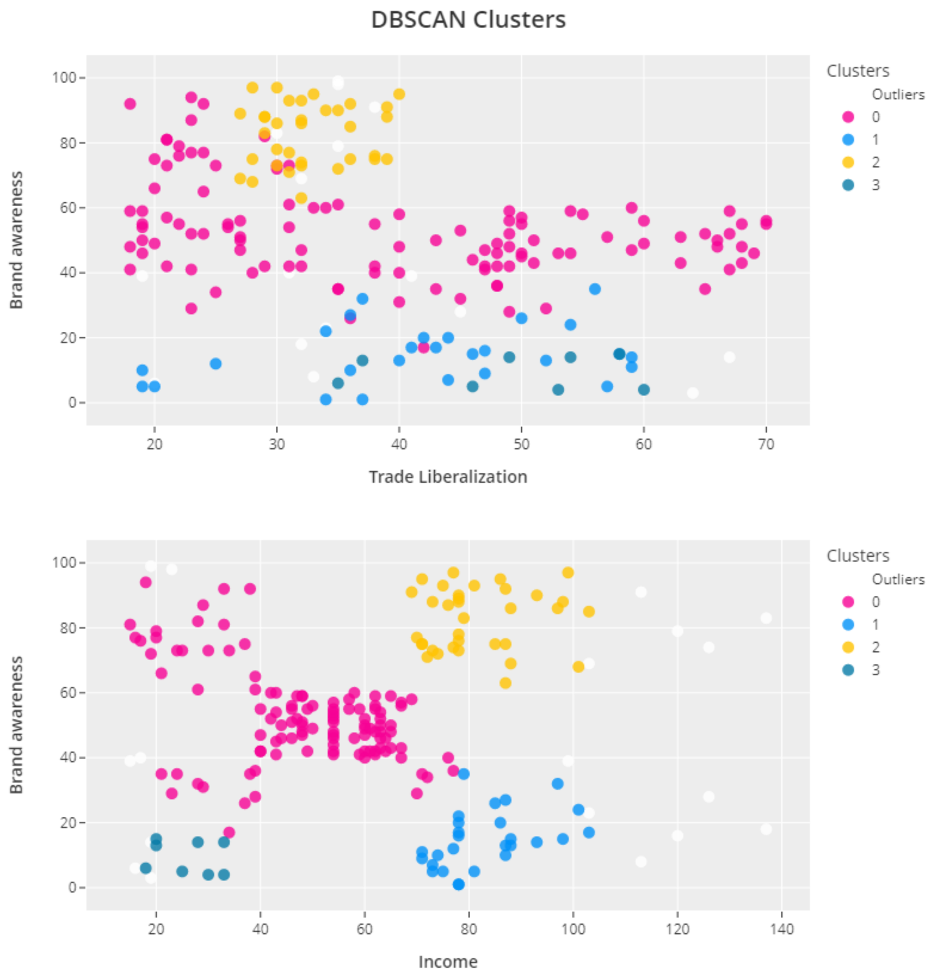
The conclusion drawn from the Affinity Propagation Clusters is as follows: Countries that are classified as belonging to Cluster 0 have a typical level of Trade liberalization, a typical level of Gross income (40-60), and a typical degree of Trade liberalization Brand awareness (40-60). Cluster 0 and Cluster 1 both represent a group of nations that are practically exactly the same as one another, and the distinction between the two clusters is almost impossible to tell apart. Custer 2 places in the same category the nations who have a score of less than 40 for their gross national income and have a high degree of brand recognition. (60-100). Cluster 3 enables us to pinpoint nations that not only have a low Gross National Income (GNI) (below 40), but also have a low degree of consumers' familiarity with various brands (0-40).

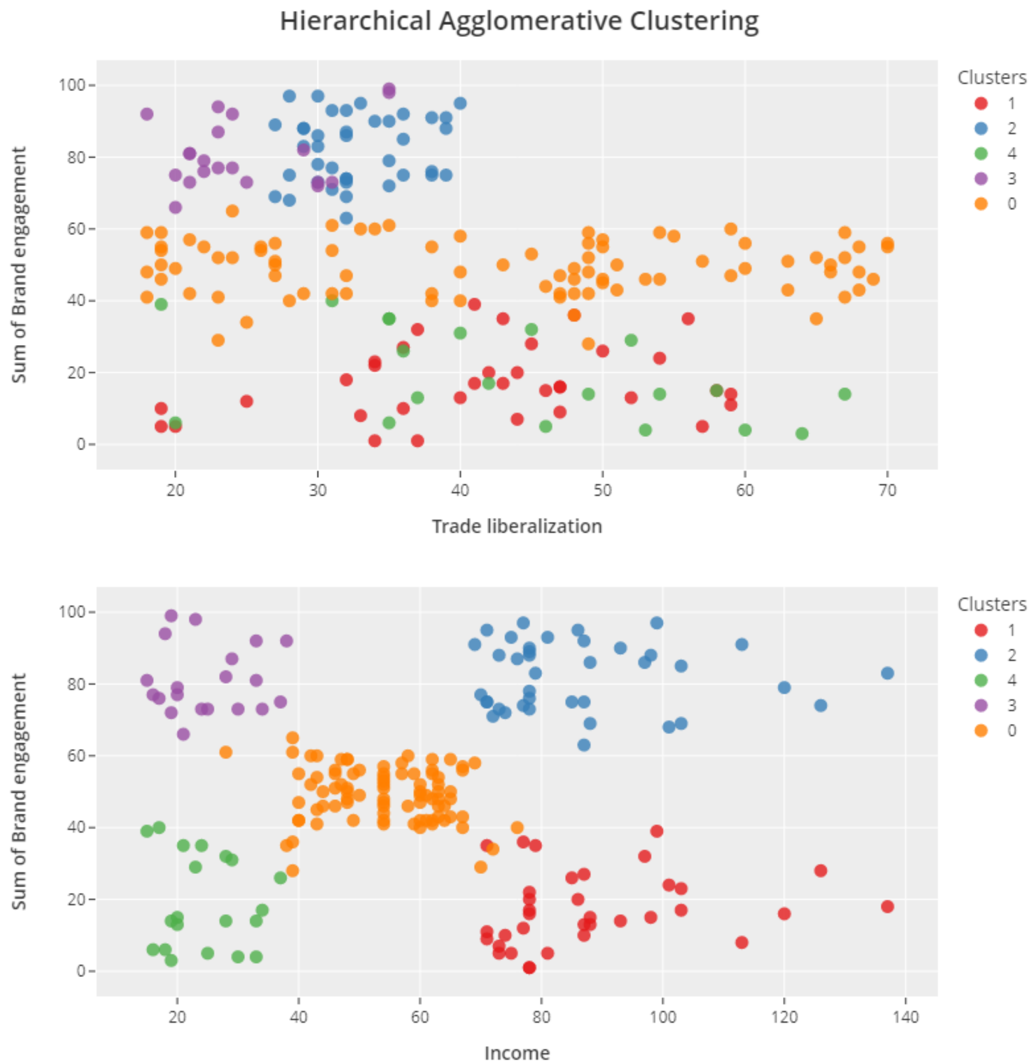
Cluster 4 is the exact opposite of Cluster 3 due to the high levels of gross income e (>80) and high levels of involvement in brand (>60) that are present in the countries that make up Cluster 4. Cluster 5 is not the least important, as it highlights nations that have high levels of Gross income (>80), but low levels of brand awareness (40).

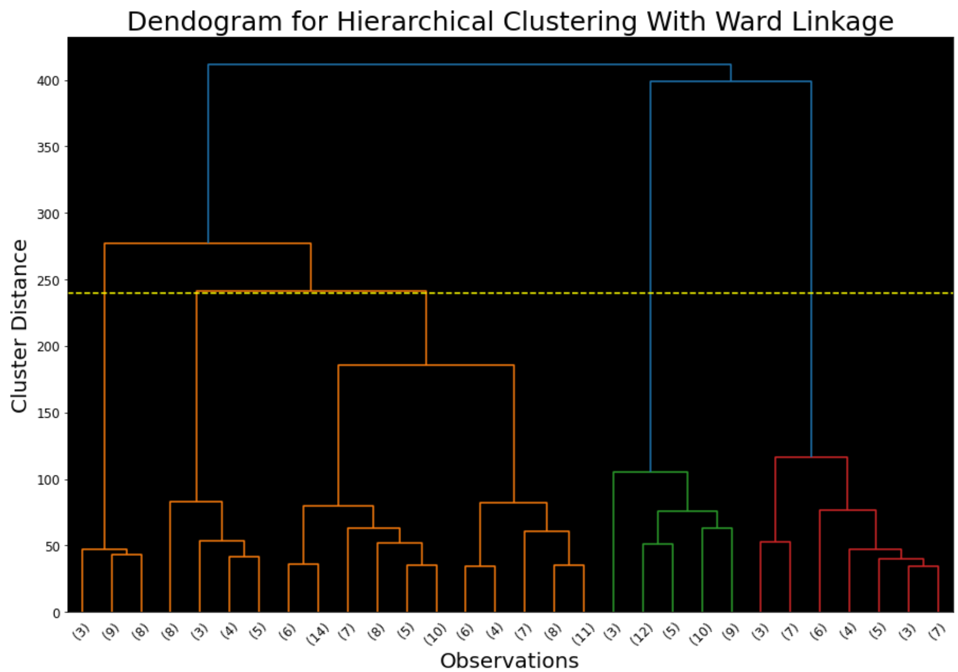
Outliers in the dataset, as determined by DBSCAN Clusters, are shown by white circles. These are observations that do not fit into any of the three categories, and they are referred to as outliers. Cluster 0 is made up of the nations that have a Gross income that is lower than 60 and a Brand awareness that is higher than 40. These nations have both of these requirements met. Cluster 1 countries have an extremely low level of digital competence (below 40) and a low degree of familiarity with the brands. Cluster 1 countries also have a low level of brand awareness (below 20). The class of nations that make up Cluster 2 are those that have a high degree of familiarity with the brands (more than 60) and a high level of digital proficiency (more than 70). Cluster 3 countries have a high degree of digital competence (above 70), but a low level of awareness with the brands in their industry (below 40).

The conclusion drawn from hierarchical clustering is that nations that are classified as belonging to Cluster 0 have an average Trade liberalization Gross income (ranging from 40 to 60) and an average Trade liberalization Brand awareness (40-60). Countries that are classified as belonging to Cluster 1 have a high Gross income (more than 70), yet they have a poor Brand awareness (fewer than 40). Countries that are classified as belonging to Cluster 2 have a high degree of gross income and awareness in more than sixty percent of all of their available possibilities. Cluster 3 countries have a high degree of awareness in brands (more than 60). Last but not least, nations that are classified as belonging to Cluster 4 have a level of Brand awareness that is lower than 40 and a Gross income level that is lower than 40,000.

The results of this could guide how a multinational companies will be able to create personalized brand and profit strategies plans for each of the identified macro clusters.


DBSCAN Clusters
Clusters
Outliers
0
1
2
3
Brand awareness
Trade Liberalization
Brand awareness
Income


Hierarchical Agglomerative Clustering
Sum of Brand engagement
Trade liberalization
Clusters
1
2
4
3
0
Sum of Brand engagement
Income
Clusters
1
2
4
3
0



Conclusion

International segmentation assists the company in organizing the variability that occurs between customers and countries and in identifying subgroups that can be targeted effectively and efficiently. Foreign segmentation is used for a variety of objectives, including understanding international consumers, recognizing similarities and/or variances across international marketplaces.

An international clustering may need a multi-domestic strategy in which each country represents a different segment - this technique is suitable when items are manufactured locally and are adapted to meet local needs. A pan-global strategy, on the other hand, is used when multinational organizations understand that clusters of customers from various nations may share more in common than consumers from the same country. This allows firms to use a framework to sell the same goods and/or market in other nations by pursuing the same client groups in different locations.

Organizations may segment overseas markets employing one or a mix of various characteristics. Segmenting global markets on the grounds of the variables selected suggests that segments should comprise of clusters of nations. Global segmentation systems often categorize nations based on a variety of macroeconomic, geopolitical, and legal, as well as cultural and social aspects.

Finally, it should be noted that macro-segmentation ignores consumer distinctions inside each nation and across country markets that are grouped together, as well as the presence of segments that extend beyond the limits of a certain geographic area. As a result, the organization may be losing customers or revenues by failing to meet the needs of customer segments throughout different geographic parts. Macro-segmentation leads to inaccurate national stereotyping, which neglects within-country variability. Ignoring similarities in requirements across national lines leads in governments missing economies of scale advantages that may be obtained by fulfilling the needs of a larger population across national boundaries.

Reference

[1] J.-B. E. M. Steenkamp and F. Ter Hofstede, "International market segmentation: issues and perspectives," *International Journal of Research in Marketing*, vol. 19, no. 3, pp. 185–213, Sep. 2002.
[2] V. R. Errunza and D. P. Miller, "Market Segmentation and the Cost of the Capital in International Equity Markets," *Journal of Financial and Quantitative Analysis*, vol. 35, no. 4, pp. 577–600, Dec. 2000.
[3] S. Formica and M. Uysal, "Market Segmentation of an International Cultural-Historical Event in Italy," *J. Travel Res.*, vol. 36, no. 4, pp. 16–24, Apr. 1998.
[4] H. Müller and U. Hamm, "Stability of market segmentation with cluster analysis – A methodological approach," *Food Qual. Prefer.*, vol. 34, pp. 70–78, Jun. 2014.
[5] F. A. Palumbo and I. Teich, "Market segmentation based on level of acculturation," *Marketing Intelligence & Planning*, vol. 22, no. 4, pp. 472–484, Jan. 2004.
[6] O. Furrer, B. S.-C. Liu, and D. Sudharshan, "The Relationships between Culture and Service Quality Perceptions: Basis for Cross-Cultural Market Segmentation and Resource Allocation," *J. Serv. Res.*, vol. 2, no. 4, pp. 355–371, May 2000.
[7] M. Wedel and W. A. Kamakura, "Market segmentation: Conceptual and methodological foundations," 2000.
[8] L. G. Debo and L. B. Toktay, "Market segmentation and product technology selection for remanufacturable products," *Management*, 2005.
[9] C. Gaston-Breton and M. M. Oscar, "International market selection and segmentation: a two-stage model," *Int. Mark. Rev.*, vol. 28, no. 3, pp. 267–290, Jan. 2011.
[10] F. Ter Hofstede, J.-B. E. M. Steenkamp, and M. Wedel, "International Market Segmentation Based on Consumer–Product Relations," *J. Mark. Res.*, vol. 36, no. 1, pp. 1–17, Feb. 1999.
[11] P. Stefanatou, G. Konstantakopoulos, E. Giannouli, N. Ioannidi, and V. Mavreas, "Patients' needs as an outcome measure in schizophrenia," *Eur. Psychiatry*, vol. 33, no. S1, pp. S453–S453, Mar. 2016.
[12] S. P. Stefanatou Pentagiotissa, G. K. George Konstantakopoulos, E. G. Eleni Giannouli, S. V. Silia Vitoratou, and V. M. Venetsanos Mavreas, "The relationship between patients' needs and psychopathology in schizophrenia: Do patients and therapists agree?," *Eur. Psychiatry*, vol. 30, p. 1376, Mar. 2015.
[13] P. Ghauri and P. Cateora, *EBOOK: International Marketing*. McGraw-Hill Education, 2014.
[14] W. J. Keegan and N. K. Bhargava, *Global Marketing Management*, 7th ed. Delhi, India: Pearson Education, 2010.
[15] J. M. Gil, A. Gracia, and M. Sánchez, "Market segmentation and willingness to pay for organic products in Spain," *The International Food and Agribusiness Management Review*, vol. 3, no. 2, pp. 207–226, Jun. 2000.
[16] D. W. Cravens and N. Piercy, "Strategic marketing," 2006. [Online]. Available: https://www.academia.edu/download/47094280/MAR402_Course_Outline_F2010-11.pdf.

[17] J. Hoek, P. Gendall, and D. Esslemont, "Market segmentation: A search for the Holy Grail?," *Journal of Marketing Practice: Applied Marketing Science*, vol. 2, no. 1, pp. 25–34, Jan. 1996.
[18] S. Achar, "Early Consequences Regarding the Impact of Artificial Intelligence on International Trade," *American Journal of Trade and Policy*, 2019.
[19] Maimon and Rokach, "Clustering methods," *Data Min. Knowl. Discov.*, 2005.
[20] K. Wang, J. Zhang, D. Li, X. Zhang, and T. Guo, "Adaptive Affinity Propagation Clustering," *arXiv [cs.AI]*, 08-May-2008.
[21] P. Bafna, D. Pramod, and A. Vaidya, "Document clustering: TF-IDF approach," in *2016 International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT)*, Chennai, India, 2016, pp. 61–66.
[22] D. Birant and A. Kut, "ST-DBSCAN: An algorithm for clustering spatial–temporal data," *Data Knowl. Eng.*, vol. 60, no. 1, pp. 208–221, Jan. 2007.
[23] A. Smiti and Z. Elouedi, "DBSCAN-GM: An improved clustering method based on Gaussian Means and DBSCAN techniques," in *2012 IEEE 16th International Conference on Intelligent Engineering Systems (INES)*, 2012, pp. 573–578.