# Best Next Move Location

# Contents

# Introduction

Often when business or even private individuals look for new ways to provide growth, they seek it overseas. Although the idea of entering a new country can sound attractive, it can also be intimidating at the same time. There are many things to consider when taking a business to a new country, and some of the most intuitive ones are:

- *"What are the threats of making this move"?*
- *"What are the benefits of making this move"?*
- *"Do we have the resources to expand to that location"?*

And probably the most important questions are:

- *"Is there even demand for what we have to offer"?*
- *"How does the competition look like"?*

We can try to lower the risks of this kind of move by locating the best potential locations to expand to, or at least, categorize the locations we are looking into so we can understand our situation better. This means that if we can set our potential locations for expansion into clusters of similar locations, we can then either choose the best location to expand to, or even better, develop a unique strategy for each cluster.

Each country's culture has evolved on the course of time, some have had a longer time to evolve, and some have evolved out of some other county's culture. Further, countries differ in their economy performance, GDP for example, which can also be a potential factor to investigate when taking a business overseas. Knowing how a country differs from your own beforehand can be advantageous, as you can plan the business strategy ahead of time.

Finding a tool that can shed some light on this question on how to set our potential locations into groups of similarity, can benefit a business when deciding on a target location to expand. A tool of this sort can even help private individuals that just want to find a new location to move to, or just want to see how the location they have in mind is similar or different compared to their home location.

Here, we will focus on the question from a perspective of making the "Next Best Move" in another country, but it is also possible to take the model we will use here and fit it to a more micro approach, like finding a location within a specific city. To formalize the challenge we are looking at here, we will try to answer the question:

- ***"How can we cluster a set of countries of interest into groups of countries that are similar" ?***

Moreover, we will make our case more specific by looking at the challenge from a perspective of a restaurant chain that is looking into expanding the chain to a new country. Furthermore, we will take the point of view of a restaurant chain that its home location is in the United States, and the chain wants to build a strategy for countries within Central America, South America and Europe.

# Data Description

where you describe the data that will be used to solve the problem and the source of the data.

To try and answer our question we will set up a dataset that has information on:

- *Foursquare API (Foursquare, n.d.):*
  - ⇒ *Top venues per location*

- *The World Bank*[1]*:*
  - ⇒ *Doing Business indicators (World Bank, n.d.-a):*
    - *Ease of Doing Business*
    - *Getting Credit*
    - *Enforcing Contracts*
    - *Starting a Business*
  - ⇒ *GDP per capita (World Bank, n.d.-b)*
- *Wikipedia:*
  - ⇒ *Capital city of each country in the dataset ("List of national capitals", 2019)*
- *Nominatim:*
  - ⇒ *Coordinates for each capital city*

The Foursquare API will be used to retrieve data on the top venues per location of choice in other to determine how the competition and demand looks like at the potential locations. The composition of the top venues' types can teach us two things, (1) who competes within a target location, and (2) what kind of places in the leisure industry are demanded at target

[1] To see a comprehensive definition of each of these indicators please see Appendix I where you can also see graph compering a select of five countries for each indicator.

location. We can argue the first by simply saying that a location that is characterized by having many restaurants will have a more fierce competition. For the latter we will assume that the composition of the top ranked venues at a target location show us what the people are interested in, hence, our demand.

To complete analysis dataset, some indicators from The World Bank were used. These indicators include GDP per capita and some "Doing Business" indicators (as listed at the beginning of this section). The GDP per capita in a country can teach us about the economic strength of a country and about the potential of the populations' ability to spend on leisure activities.

The country dataset has a total of 74 countries across the America and Europe continents, summary statistics are available in **Table 1**.

**Table 1 - Country data summary statistics**

| | Getting Credit | Enforcing Contracts | Starting a Business | Ease of Doing Business | GDP/capita |
|---|---|---|---|---|---|
| mean | 12.5 | 61.6 | 86.9 | 70.3 | 22,346 |
| std | 4.0 | 11.3 | 10.0 | 9.8 | 24,036 |
| min | 2.0 | 25.9 | 36.4 | 40.7 | 827 |
| max | 20.0 | 81.3 | 99.6 | 85.3 | 116,640 |
| 25% | 10.0 | 55.8 | 84.1 | 61.8 | 6,071 |
| 50% | 13.0 | 63.6 | 88.8 | 73.1 | 12,164 |
| 75% | 15.8 | 70.1 | 93.2 | 76.8 | 29,818 |
| count | 74 | | | | |

To cluster our countries into groups we need a location in each country that we will explore. To do this we will use each country's capital city as a proxy for that country's properties. The information on each country's capital city is available on the Wikipedia page "List of national capitals" (*"List of national capitals", 2019*).

Lastly, to be able to explore each capital city's top venues using the Foursquare API we need the coordinates for each capital city in our dataset. The coordinates were retrieved using from Nominatim the GeoPy client.

In **Figure 1** you can see a country profile featuring the United States. This country profile shows a summary of the data that is available in our dataset.

# Figure 1

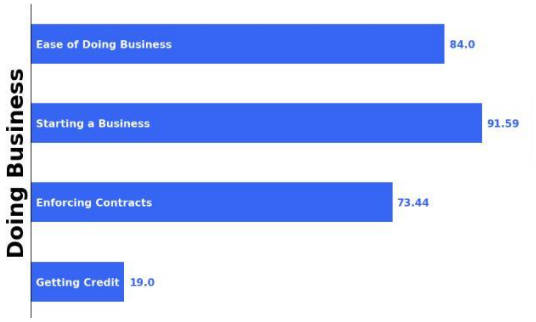
## Country Profile: United States

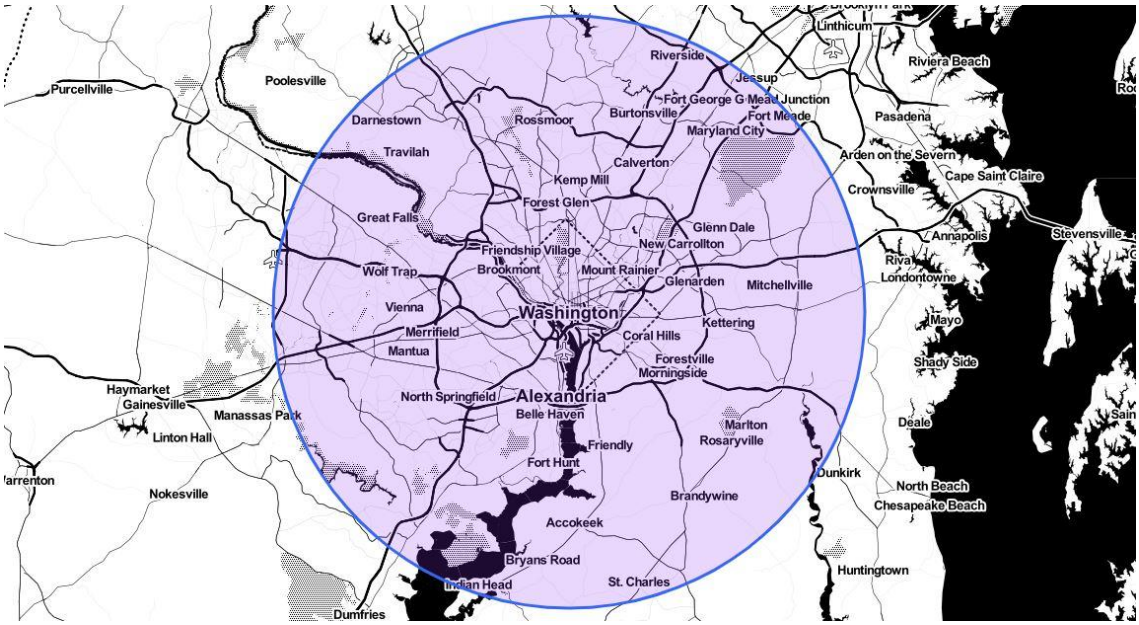
**Capital City:**

**Washington, D.C.**

**GDP/capita:** 62795

**Location:** (38.89, -77.04)

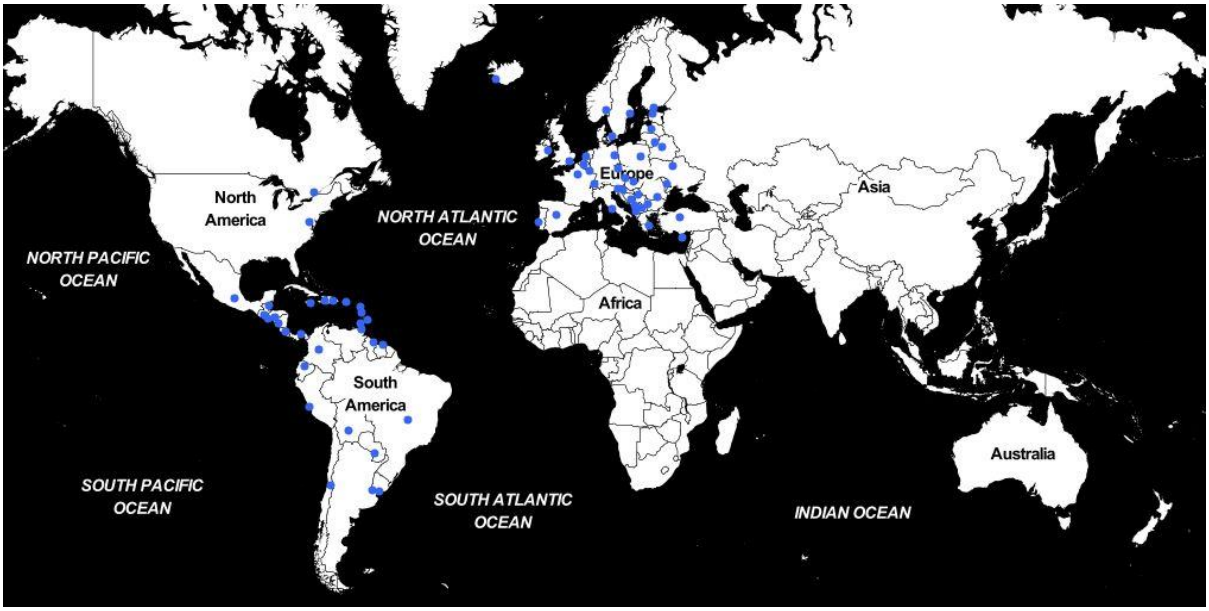


# Methodology

Before we start with the analysis stage, we will take a look at some exploratory measures that have been taken to build up the strategy for our analysis. First, we can see in **Figure 2** a plot of the capital cities in our dataset on a map. During the data preparation stage this plot has been used to do a quality control over the correctness of the data. For example, the coordinates retrieved for Brasília the capital city of brasil were not accurate and have been fixed.

## Figure 2 - Plot of capital cities of the countries in the data set




The chosen statistical method for this analysis is the k-means method. Using the k-means method allows us to split a large group of observations into sensible groups while using many characteristics of each observation to make this determination. Countries in the dataset have been set into groups according to the data collected. Specifically, countries were clustered according to:

- *Ease of Doing Business indicator*
- *GDP per capita*
- *Top venues per capital city*

For the "Doing Business" indicators only the "Ease of Doing Business" has been kept for the analysis, as it is a simple average of all the Doing Business indicators and takes in mind the other 3 indicators that have been show in the data collection stage. The other 3 indicators

are used for making sense of the results later where we want to give each cluster a better meaning.

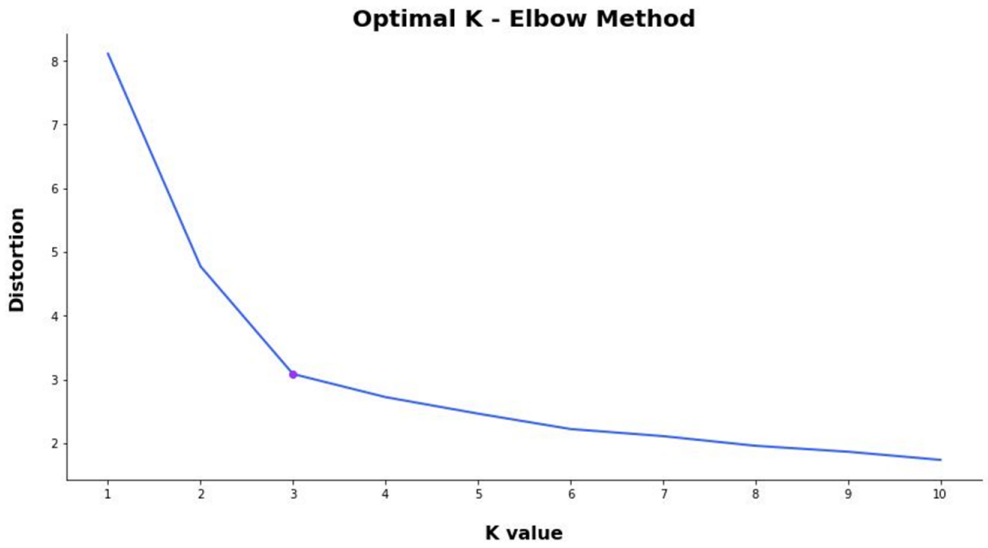
The top venues results were grouped by capital city and then normalized so the sum of the frequencies of the different venues in a capital city sum up to 1. Further, the Ease of Doing Business indicator and the GPD per capita value were normalized according to **Equation 1**, this was done to keep the clustering data at the same scale.

## Equation 1

$$\frac{Value_{Original} - Value_{min}}{Value_{max} - Value_{min}}$$

Before the application of the k-means method, the optimal K value was determined according to the "Elbow Method". This means that the k-means analysis has been done for different K values between 1 and 10 and then plotted against the distortion value of each K value. The distortion values were calculated as the sum of squared distances of the samples from their closest cluster center. The K values for the cluster analysis was set to 3 according to the results shown in **Figure 3**.

**Figure 3**




After splitting the countries in the dataset into 3 groups, their characteristics are determined by grouping the clustering data set by the cluster labels assign to each capital city. This gives each cluster a detailed data frame with each cluster's venue type frequencies. Then the top 10 most frequent venues of each cluster are investigated to see its characteristics. The

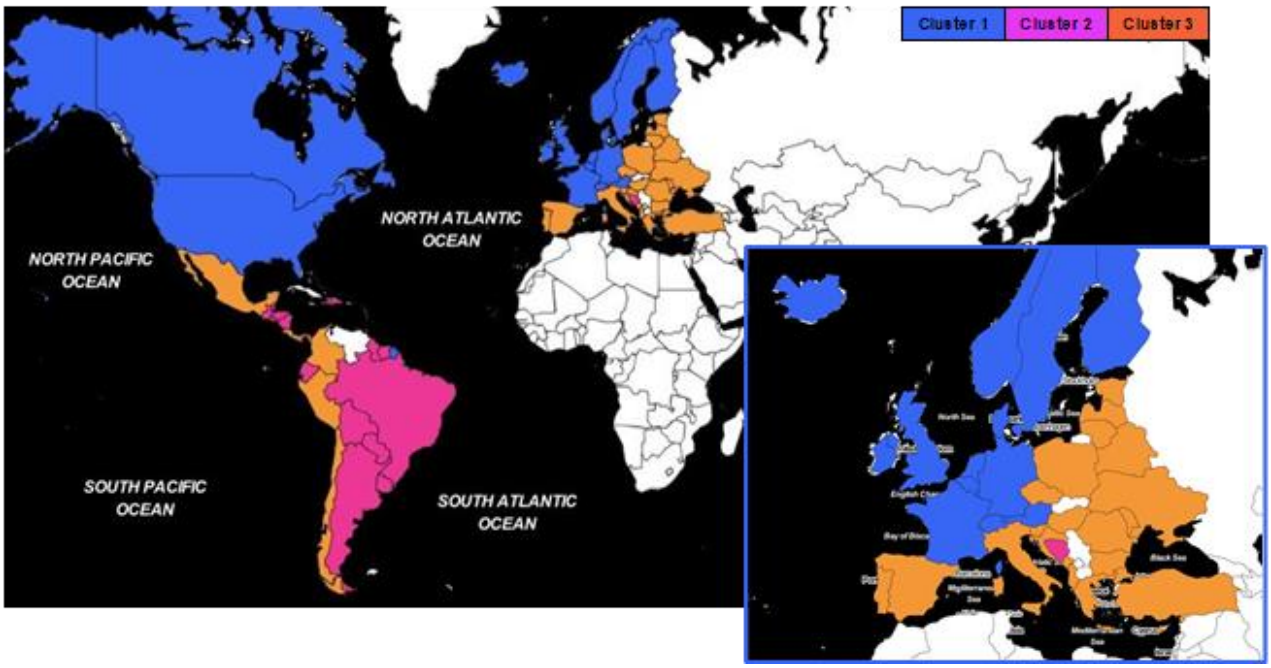
competition and demand characteristics of each cluster are determined according to this list of venues.

To conclude the analysis stage, a summary statistics of each cluster is compared with the other two. This summary statistics includes the Doing Business indicators and the GDP per capita value. Each cluster differs from the other two in these values and according to this difference we can make sense of each cluster and learn what kind of strategy can be formed for each cluster.

# Results

**Figure 4** shows the results of the clustering. We can see our 3 clusters and how they spread between the continents and regions we have focused on in this analysis. To give a better sense to what we are looking at we will look at series of tables that show more information about the 3 clusters. For convenience, **Table 2** shows a list of 68 countries split into three clusters – while collecting coordinates data 6 countries have been dropped automatically (coordinates not found).

**Figure 4 - Countries by cluster**



| Table 2 - Countries by cluster | | | | | |
|---|---|---|---|---|---|
| **Cluster 1** | | **Cluster 2** | | **Cluster 3** | |
| Austria | | Antigua and Barbuda | Nicaragua | Albania | Mexico |
| Belgium | | Argentina | Paraguay | Belarus | Moldova |
| Canada | | Barbados | Suriname | Bulgaria | Montenegro |
| Denmark | | Belize | Trinidad and Tobago | Chile | North Macedonia |
| Finland | | Bolivia | Uruguay | Colombia | Panama |
| France | | Bosnia and Herzegovina | | Costa Rica | Peru |
| Germany | | Brazil | | Croatia | Poland |
| Iceland | | Dominica | | Cyprus | Portugal |
| Ireland | | Dominican Republic | | Czech Republic | Puerto Rico |
| Luxembourg | | Ecuador | | Estonia | Romania |
| Netherlands | | El Salvador | | Greece | Serbia |
| Norway | | Grenada | | Hungary | Slovenia |
| Sweden | | Guatemala | | Italy | Spain |
| Switzerland | | Guyana | | Jamaica | Turkey |
| United Kingdom | | Haiti | | Latvia | Ukraine |
| United States | | Honduras | | Lithuania | |

In **Table 3** we can see the top 10 most frequent venues of each cluster. From this table we can learn about the competition and demand characteristics of each cluster. We can that the "Hotel" Category is at top of clusters 1 and 3 and takes the 3rd place in cluster 2. We can also see that "Coffee Shops" and "Café " categories are the 2nd and 3rd places (not necessary in this order) in clusters 1 and 3 where as "Café" category takes the 1st place in cluster 2 and "Coffee Shop" takes only the 6th place within this cluster.

| Table 3 - Top venues by cluster | | | |
|---|---|---|---|
| Rank | **Cluster 1** | **Cluster 2** | **Cluster 3** |
| 1st | Hotel | Café | Hotel |
| 2nd | Coffee Shop | Restaurant | Café |
| 3rd | Café | Hotel | Coffee Shop |
| 4th | Bar | Bar | Bar |
| 5th | Plaza | Pizza Place | Restaurant |
| 6th | Restaurant | Coffee Shop | Plaza |
| 7th | Scandinavian Restaurant | Bakery | Italian Restaurant |
| 8th | Cocktail Bar | Italian Restaurant | Theater |
| 9th | Italian Restaurant | Ice Cream Shop | Cocktail Bar |
| 10th | Park | Fast Food Restaurant | Ice Cream Shop |

Further, we move to the most important category for our analysis "Restaurants". A short reminder, we want to expand our restaurant chain to a new location in a new country, meaning we want to know how that industry is doing at the target location. We can see that the "Restaurant" category takes the 6th, 2nd and 5th places in cluster 1, 2 and 3 respectively. We can also see that there are more categories that rank in the top 10 which are of a specific type of restaurant.

There is one more thing to show in this section before we head on to the discussion section which is the summary statistics of the Doing Business indicators and GDP per capita of each cluster. These can be found in **Table 4**, **Table 5** and **Table 6** for Cluster 1, Cluster 2 and Cluster 3 respectively.

**Table 4 - Cluster 1 summary statistics**

| | Getting Credit | Enforcing Contracts | Starting a Business | Ease of Doing Business | GDP/capita |
|---|---|---|---|---|---|
| mean | 12.4 | 68.8 | 91.7 | 79.3 | 62,018 |
| std | 3.6 | 6.8 | 4.0 | 3.9 | 20,086 |
| min | 3.0 | 57.1 | 83.3 | 69.6 | 41,464 |
| max | 19.0 | 81.3 | 98.2 | 85.3 | 116,640 |
| 25% | 11.0 | 64.2 | 90.2 | 76.8 | 47,582 |
| 50% | 12.5 | 68.9 | 92.9 | 79.6 | 53,816 |
| 75% | 14.0 | 73.6 | 94.3 | 82.1 | 74,595 |
| count | 16 | | | | |

**Table 5 - Cluster 2 summary statistics**

| | Getting Credit | Enforcing Contracts | Starting a Business | Ease of Doing Business | GDP/capita |
|---|---|---|---|---|---|
| mean | 9.4 | 52.2 | 77.0 | 57.4 | 7,998 |
| std | 4.1 | 10.7 | 12.9 | 5.8 | 5,307 |
| min | 2.0 | 25.9 | 36.4 | 40.7 | 868 |
| max | 17.0 | 68.1 | 89.6 | 65.4 | 17,949 |
| 25% | 6.0 | 50.1 | 71.4 | 55.5 | 4,549 |
| 50% | 9.0 | 56.3 | 80.4 | 59.0 | 6,234 |
| 75% | 12.0 | 57.9 | 86.4 | 60.6 | 10,640 |
| count | 21 | | | | |

**Table 6 - Cluster 3 summary statistics**

| | Getting Credit | Enforcing Contracts | Starting a Business | Ease of Doing Business | GDP/capita |
|---|---|---|---|---|---|
| mean | 14.0 | 62.4 | 89.4 | 73.7 | 15,082 |
| std | 2.9 | 9.8 | 4.5 | 4.0 | 8,973 |
| min | 9.0 | 34.3 | 79.9 | 66.6 | 3,095 |
| max | 19.0 | 78.8 | 97.4 | 81.6 | 34,483 |
| 25% | 12.0 | 55.0 | 86.7 | 70.1 | 7,094 |
| 50% | 14.0 | 64.4 | 89.3 | 73.4 | 14,910 |
| 75% | 16.0 | 69.3 | 92.5 | 76.4 | 21,702 |
| count | 31 | | | | |

We can learn from Tables 4-6 that on average, Cluster 1 ranks highest on all measures excluding the "Getting Credit" indicator where Cluster 3 ranks higher, and Cluster 3 ranks second on all other measures. We can also see that Cluster 3 is a bit behind Cluster 1 whereas Cluster 2 is well behind the other 2. Taking the GDP per capita for example, Cluster 3 is about a quarter than what the GDP per capita of Cluster 1, and Cluster 2's GDP per capita is a little over a half of Cluster 2.

# Discussion

A section where you discuss any observations you noted and any recommendations you can make based on the results.

The k-means analysis has given us some useful information about the countries we are looking at as a potential location for our restaurant chain to expand to. We can now classify each cluster and discuss the strategies that can be beneficial for our chain.

Cluster 1 – on average, this cluster is characterized with high GDP per capita and ranks high on "Ease of Doing Business", "Starting a Business" and "Enforcing Contract". This means that this Cluster should provide a safer location to expand to because contracts are enforced better and it is easier to start a business. The high GDP per capita gives our chain a good potential for growth because individuals will have more to spend on average. We can see that the minimum value of the GDP per capita in Cluster 1 (about $ 41.5K US) is higher than the next runner up in Cluster 3 (about $ 34.5K US). From the venue ranking for Cluster 1 we can learn that there are different restaurant categories in the top 10 but there is still room for them to rank higher withing the top 10. This may satisfy our chain as it appears that people in these countries show interest in restaurants, but there may still be room for new restaurants.

Cluster 2 – this cluster look like the least attractive one since it ranks lower than the other two clusters, it ranks specially on the GDP per capita measure at an average of almost $ 8k US. Further, the "Restaurant" category is ranked 2$^{nd}$, which leave us in a pretty competitive environment.

Cluster 3 – we will say that this cluster is pretty similar to Cluster 1 but ranks a bit lower than Cluster 1 on average. The top 10 venues within this cluster are also pretty similar to Cluster 1 with a little variance.

Strategies – instead of targeting one country from a specific cluster we can come up with a general strategy per cluster so in the long run we can use this set of strategies when ever looking into expanding to a new country within one of the three clusters. We will also assume that the chain offers a mid-price range dining experience, which can have a great effect on the selected strategy. Here is a recommendation for each cluster's strategy:

- ⇒ *"Premium" Cluster (Cluster 1)* – for these countries the restaurant chain can plan to provide an upper end version of the chain's restaurants. This means that the chain can offer a higher end dining experience including premium food and a more luxurious version of the chain's restaurant.
- ⇒ *"Low-cost" Cluster (Cluster 2)* – the strategy here will be the opposite from the Premium Cluster. The chain can offer the minimalistic dining experience, still making a stand in these countries but at a lower cost for the chain. The chain can even create a new brand name for these countries to avoid harming the chains original brand name.
- ⇒ *"Business as usual" Cluster (Cluster 3)* – this cluster will have the minimal amount of adjustment to the chain's strategy. Our assumption is that the chain competes in the mid-price range restaurants, hence the chain's strategy at the home location will suffice.

Of course these are all just suggestions, there is a lot to look into and to take in consideration before forming any effective strategy. To build a solid strategy we must also give the proper weight for other internal variables like the chain's financials, human capital and any specialties that can benefit in one cluster over another.

# Conclusion

To conclude this report, we can see that we have managed to cluster our potential locations into three groups. Using the k-means method and by using the Foursquare API to find the top venues in each location, we have managed to analyze our potential locations and for a strategy according to analysis results. This analysis can provide a strong tool for different businesses or even private individuals when trying to understand a new location. Although this can be a strong tool more information about a business must be taken into consideration while using this tool to analyze the business potential and form strategies.

# References


Foursquare. (n.d.). *Foursquare Developer.* Retrieved December 28, 2019, from Foursquare: https://developer.foursquare.com/

Wikipedia contributors. (2019, December). *List of national capitals*. Retrieved December 28, 2019, from Wikipedia: https://en.wikipedia.org/wiki/List_of_national_capitals

World Bank, Doing Business. (n.d.-a). *Doing Business Indicators [Data file]*. Retrieved from The World Bank: https://databank.worldbank.org/source/doing-business

World Bank, World Development Indicators. (n.d.-b). *GDP per capita (current US$) [Data File]*. Retrieved December 28, 2019, from The World Bank: https://databank.worldbank.org/source/world-development-indicators

# Appendix I – The World Bank Indicators

## Getting credit (World Bank, n.d.-a)

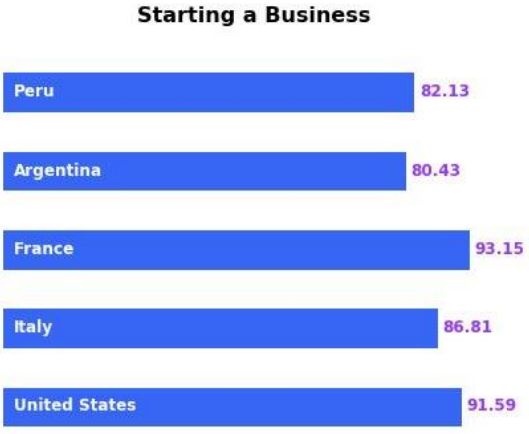
*"The total score for getting credit is the sum of the strength of legal rights index and the depth of credit information index"*[2] (World Bank, n.d.-a).


## Enforcing Contracts

*"The score for enforcing contracts is the simple average of the scores for each of the component indicators: the time and cost for resolving a commercial dispute through a local first-instance court, as well as the quality of judicial processes that promotes quality and efficiency in the court system"*[3] (World Bank, n.d.-a).


## Starting a Business

*"The score for starting a business is the simple average of the scores for each of the component indicators: the procedures, time and cost for an entrepreneur to start and formally operate a business, as well as the paid-in minimum capital requirement"* (World Bank, n.d.-a).




[2] The score is computed Based on the methodology in the DB15-20 studies.

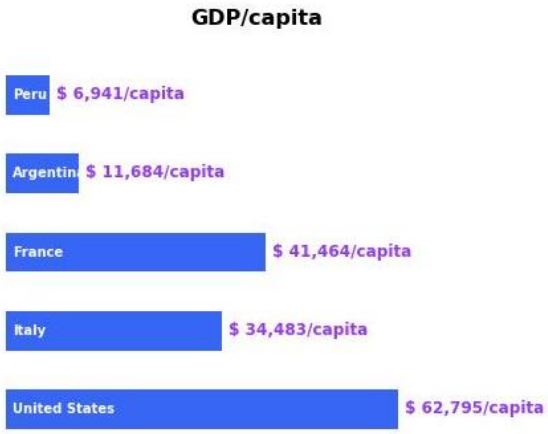
[3] The score is computed based on the methodology in the DB17-20 studies.

## Ease of Doing Business

*"The ease of doing business score is the simple average of the scores for each of the Doing Business topics: starting a business, dealing with construction permits, getting electricity, registering property, getting credit, protecting minority investors, paying taxes, trading across borders, enforcing contracts and resolving insolvency"*[4] (World Bank, n.d.-a).


## GDP per capita

*"GDP per capita is gross domestic product divided by midyear population. GDP is the sum of gross value added by all resident producers in the economy plus any product taxes and minus any subsidies not included in the value of the products. It is calculated without making deductions for depreciation of fabricated assets or for depletion and degradation of natural resources"*[5] (World Bank, n.d.-b).




[4] The score is computed based on the methodology in the DB17-20 studies for topics that underwent methodology updates.

[5] Data are in current U.S. dollars.