

Ecuador's top Cities and their Parish Segmentation Characteristics

Xavier Jácome Piñeiros

September 20th, 2019

1. Introduction

1.1. Background

Ecuador is a small country located in South America. According to the Ecuadorian national statistics institute named “Instituto Nacional de Estadísticas y Censos”, INEC, the country has around of 17 million of habitants. The Ecuadorian constitution of 2008 established the country as a constitutional republic which political-administrative division comprises from highest to lowest hierarchy, 24 provinces, 221 cantons and 1500 parishes. The most important cities of Ecuador are Quito and Guayaquil. Quito is the political capital and Guayaquil is the main port city of the nation.

The parish is the lowest political-territorial division in Ecuador. There are two types of parishes: urban and rural. The urban parish is one that is circumscribed within the metropolis or city. It consists of the entire necessary infrastructure to be a main city. The rural parish are those that are separated from the main city or metropolis. The capital city, Quito, it is divided into 70 parishes. In the other hand, the main port city, Guayaquil, it is divided into 24 parishes.

In the economical field, the World Bank data shows that Ecuador produced a GDP of 108.398 Billion dollar in 2018. Quito and Guayaquil contribute largely to the total amount of the GDP of the country. In line with this, El Comercio, a much known newspaper in Ecuador, cited in an article that Guayaquil is the main city that contribute to the total GDP of the country followed by Quito. This information shows that Quito and Guayaquil are not only the most important cities of Ecuador due to the political-administrative division of the country, but also for the largely contribution of them to the GDP of the nation.

1.2. Problem

The national statistics institution and other public institutions have different type of data sources that are recollected and prepared for generating some limited analysis. Nevertheless, despite of having information of the parishes and the venues of each parish, there is no analysis related to know how similar or dissimilar parishes due to the common venues people visit are. This type of analysis is important for different type of audiences like companies that want to know where to locate new stores and for public authorities that want to coordinate cities development.

1.3. Objective

The main objective of the project is to determine how the parishes of Quito and Guayaquil cities are grouped because of the venues type people visit in each parish area.

To achieve the main objective of the project, it is necessary to fulfill with these two specific objectives:

- Determine the type of common venues that are present in the radius area of each parish.
- Segment the parishes due to the type of common venues people visit in each parish.

2. Data acquisition and cleaning

2.1. Data sources

To achieve the objectives of the project, two different data sources were used. The first data source was a website (URL: www.codigopostalecuador.com/guayaquil-876) which has information about the latitude and longitude of the cities and parishes of Ecuador. In this website, it is also available the postal code of each political-administrative division of the country. This first data source was very important to the project due to locate the venues that are in the area of each parish.

The second data source was the Foursquare database. It allows developers to obtain information in which using its API empowers users to explore millions of points-of-interests by powering rich location data features like venue search. With Foursquare API, it is possible to know the type of venues located in the radius of the geographical coordinates of a parish. This kind of data was necessary to complete both of the specific objectives of the project. The version of the API used in the project was 20180605.

2.2. Data cleaning

To scrape data from the first data source, pandas method `read_html()` was used. First, all the data cleaning process was applied to Guayaquil data that was assigned into a data frame named `guayaquil`. Therefore, all the steps were applied to Quito's data that was assigned into a data frame named `quito`.

The first step was detecting nulls entries. Neither Guayaquil data nor Quito had null entries.

```
# Dataframe info
guayaquil.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 25 entries, 0 to 24
Data columns (total 4 columns):
PostalCode    25 non-null object
Site          25 non-null object
Latitude      25 non-null float64
Longitude     25 non-null float64
dtypes: float64(2), object(2)
memory usage: 880.0+ bytes
```

Figure 1. Guayaquil data frame information.

The next step was to count unique values of each variable. It allows the data analyst to denote if there are duplicate data. Once applied the method, it was shown that in Guayaquil were two parishes named Pascuales and two entries that have the same latitude and longitude.

```

: for columnas in guayaquil.columns:
    print("----"+columnas+"----")
    print(guayaquil[columnas].value_counts())
    print('\n')

----Site----
Pascuales                2
Olmedo (San Alejo)        1
Ximena                    1

for columnas in guayaquil.columns:
    print("----"+columnas+"----")
    print(guayaquil[columnas].value_counts())
    print('\n')

----Latitude----
-2.23591    2
-2.06667    1
-2.20836    1

----Longitude----
-80.07755    2
-80.23183    1
-79.89745    1

```

Figure 2. Example of unique values analysis applied in Guayaquil data frame.

Focusing in the duplicates found in the last step, it was decided to drop the entries number 3 and number 14 because both entries were not only duplicated but also did not have the right coordinates. Therefore, the data frame index was reset.

3	090115	Chongon	-2.23591	-80.07755
4	090151	Chongon.	-2.23591	-80.07755
13	090116	Pascuales	-2.07104	-79.93162
14	090154	Pascuales	-2.06667	-79.93333

```
guayaquil.drop(index={3,14},inplace=True)
```

```
guayaquil.reset_index(inplace=True)
```

```
guayaquil.drop(columns=guayaquil.columns[0],inplace=True)
```

Now, let's check if everything is ok with the cleaned dataframe.

```
guayaquil.head()
```

	PostalCode	Site	Latitude	Longitude
0	090101	Ayacucho	-2.20548	-79.89029
1	090102	Bolivar (Sagrario)	-2.19923	-79.88992
2	090103	Carbo (Concepcion)	-1.81207	-80.23183
3	090151	Chongon.	-2.23591	-80.07755
4	090104	Febres Cordero	-2.20340	-79.93424

Figure 3. Sequence of steps applied to clean Guayaquil data frame.

In Quito, it was found that six parishes presented the same coordinates, which allowed understanding that there were five outliers. Therefore, it was necessary to know which the correct parish with these coordinates was in order to drop the others.

```

: for columnas in quito.columns:
    print("----"+columnas+"----")
    print(quito[columnas].value_counts())
    print('\n')

----Latitude----
-0.16563      6
-0.24461      1
-0.16540      1

----Longitude----
-78.51045      6
-78.45775      2
-78.47811      1

```

Figure 4. Example of unique values analysis applied in Quito data frame.

As it is visible, two parishes had the same longitude. It was validated that them were right.

So, entries 4, 26, 27, 38, 41 and 63 where dropped. Then, the index was reset.

```

: quito[quito['Longitude']!= -78.51045]

:

```

	PostalCode	Site	Latitude	Longitude
4	170129	Belisario Quevedo	-0.16563	-78.51045
26	170106	El Salvador	-0.16563	-78.51045
27	170125	Eloy Alfaro	-0.16563	-78.51045
38	170139	La Argelia	-0.16563	-78.51045
41	170141	La Ferroviaria	-0.16563	-78.51045
63	170144	Ponceano	-0.16563	-78.51045

```

: quito.drop(index={4,26,27,38,41,63},inplace=True)

: quito.reset_index(inplace=True)

: quito.drop(columns=quito.columns[0],inplace=True)

```

Figure 5. Sequence of steps applied to clean Quito data frame.

In order to stablish that the data coordinates were right, two maps were created to show the points of each parish. To print the maps, folium library was used.

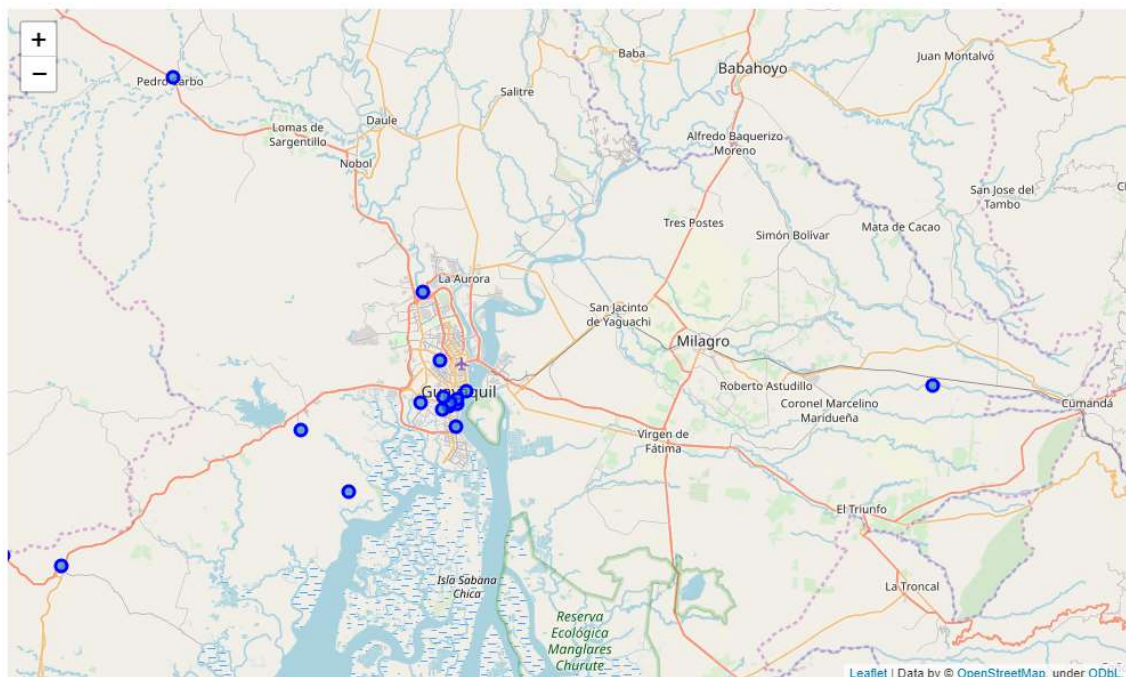


Figure 6. Guayaquil parishes map.

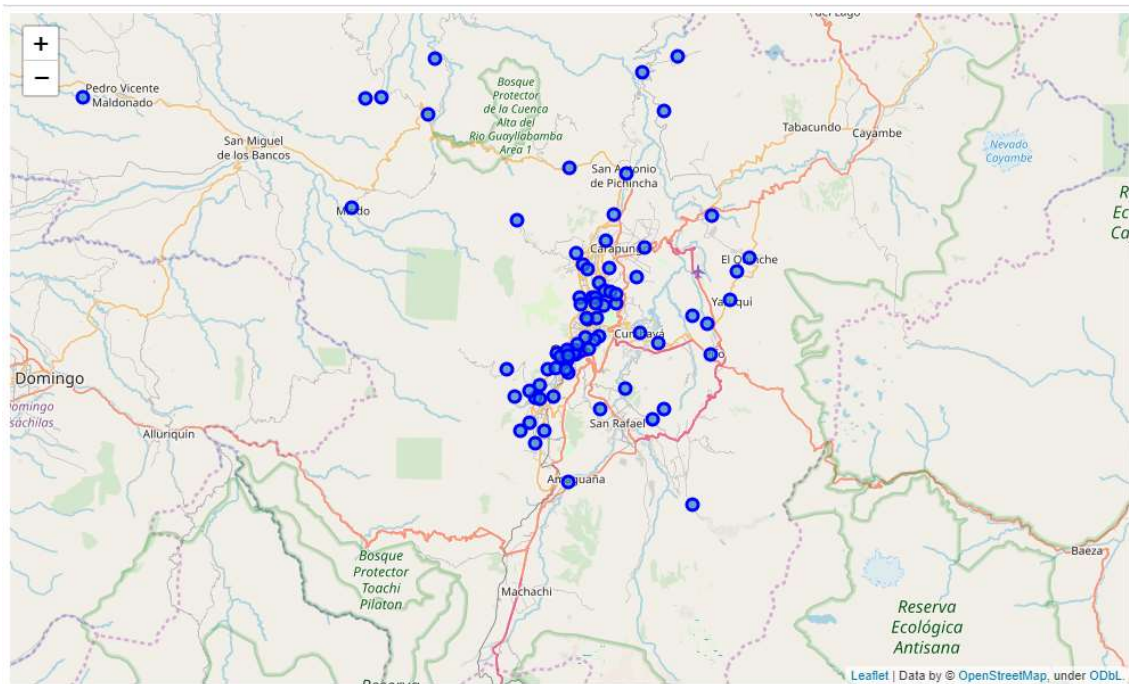


Figure 7. Quito parishes map.

Finally, using the Foursquare API, the most common venues types of each parish were obtained and then using the pandas library the data was assigned into two data frames. It was necessary to start doing the machine learning analysis.

```
neighbourhoods_venues_sorted.head()
```

	Site	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Ayacucho	Seafood Restaurant	Supermarket	Soccer Stadium	Juice Bar	Dessert Shop	Hotel Bar	Hotel	Hot Dog Joint	Health Food Store	Garden
1	Bolivar (Sagrario)	Locksmith	Market	Furniture / Home Store	Pet Store	Pharmacy	Restaurant	Theme Park	Diner	Hot Dog Joint	Health Food Store
2	Carbo (Concepcion)	Movie Theater	Multiplex	Theme Park	Diner	Hotel Bar	Hotel	Hot Dog Joint	Health Food Store	Garden	Furniture / Home Store
3	Chongon.	Playground	Theme Park	Latin American Restaurant	Hotel Bar	Hotel	Hot Dog Joint	Health Food Store	Garden	Furniture / Home Store	Food Truck
4	Febres Cordero	BBQ Joint	Seafood Restaurant	Dessert Shop	Hotel Bar	Hotel	Hot Dog Joint	Health Food Store	Garden	Furniture / Home Store	Food Truck

```
neighbourhoods_venues_sorted_quito.head()
```

	Site	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Alangasi	Park	Wings Joint	Cosmetics Shop	Food Truck	Food Court	Food & Drink Shop	Food	Flea Market	Fast Food Restaurant	Farmers Market
1	Alfaro (Chimbacalle)	Science Museum	South American Restaurant	Food Truck	Food Court	Food & Drink Shop	Food	Flea Market	Fast Food Restaurant	Farmers Market	Wings Joint
2	Atahualpa (Chabaspamba)	Hotel	Bar	Cosmetics Shop	Mexican Restaurant	Cajun / Creole Restaurant	Food Court	Music Venue	Historic Site	Food Truck	Department Store
3	Benalcazar	Italian Restaurant	Coffee Shop	Hotel	Bakery	Ice Cream Shop	Movie Theater	Pizza Place	Fast Food Restaurant	Japanese Restaurant	Sandwich Place
4	Calderon (Carapungo)	Pizza Place	Chinese Restaurant	Park	Wings Joint	Farmers Market	Food Court	Food & Drink Shop	Food	Flea Market	Fast Food Restaurant

Figure 8. Example of data frames where common venues information were assigned.

3. Methodology

3.1. Exploratory data analysis techniques

The first analysis was conducted by an exploratory data analysis technique. It was created a for process that shows the first five element with the highest frequencies of the venues type per Site (parish). The process allows having a better understanding of the more common type of venue in each parish. For example, in Ayacucho, a Guayaquil's parish, the most frequent venue type is Seafood Restaurant, with 57% of this type of venues in the radius of the area.

```
num_top_venues = 5

for hood in guayaquil_grouped['Site']:
    print("----"+hood+"----")
    temp = guayaquil_grouped[guayaquil_grouped['Site'] == hood].T.reset_index()
    temp.columns = ['venue', 'freq']
    temp = temp.iloc[1:]
    temp['freq'] = temp['freq'].astype(float)
    temp = temp.round({'freq': 2})
    print(temp.sort_values('freq', ascending=False).reset_index(drop=True).head(num_top_venues))
    print('\n')
```

```
----Ayacucho----
      venue  freq
0  Seafood Restaurant  0.57
1      Supermarket  0.14
2      Juice Bar  0.14
3   Soccer Stadium  0.14
4      BBQ Joint  0.00
```

```
----Bolívar (Sagrario)----
      venue  freq
0      Locksmith  0.17
1      Market  0.17
2      Pet Store  0.17
3      Pharmacy  0.17
4  Furniture / Home Store  0.17
```

```
num_top_venues = 5

for hood in quito_grouped['Site']:
    print("----"+hood+"----")
    temp = quito_grouped[quito_grouped['Site'] == hood].T.reset_index()
    temp.columns = ['venue', 'freq']
    temp = temp.iloc[1:]
    temp['freq'] = temp['freq'].astype(float)
    temp = temp.round({'freq': 2})
    print(temp.sort_values('freq', ascending=False).reset_index(drop=True).head(num_top_venues))
    print('\n')
```

```
----Alangasi----
      venue  freq
0      Park  1.0
1  American Restaurant  0.0
2      Resort  0.0
3  Other Great Outdoors  0.0
4  Pakistani Restaurant  0.0
```

```
----Alfaro (Chimbacalle)----
      venue  freq
0   Science Museum  0.67
1  South American Restaurant  0.33
2   American Restaurant  0.00
3      Pharmacy  0.00
4  Pakistani Restaurant  0.00
```

Figure 9. Exploratory data analysis.

3.2. Machine learning techniques

In order to fulfill the second objective of the project, it was essential to focus the analysis under the non-supervised machine learning methodology. The reason for focusing on this type of machine learning analysis is that the problem has no label data and it is required to find the patrons that define the groups of parishes. In order to all the aforementioned reasons, K-Means algorithm was applied, a non-supervised machine learning algorithm, which is a cluster analysis algorithm.

The K-Means algorithm is a type of cluster analysis technique that groups objects into k groups based on their characteristics. Clustering is done by minimizing the sum of distances between each object and the centroid of its group or cluster. The K-Means algorithm solves an optimization problem, the function being optimized (minimizing) the sum of the quadratic distances of each object to the centroid of its cluster. In the present project, it was decided to generate five clusters per city.

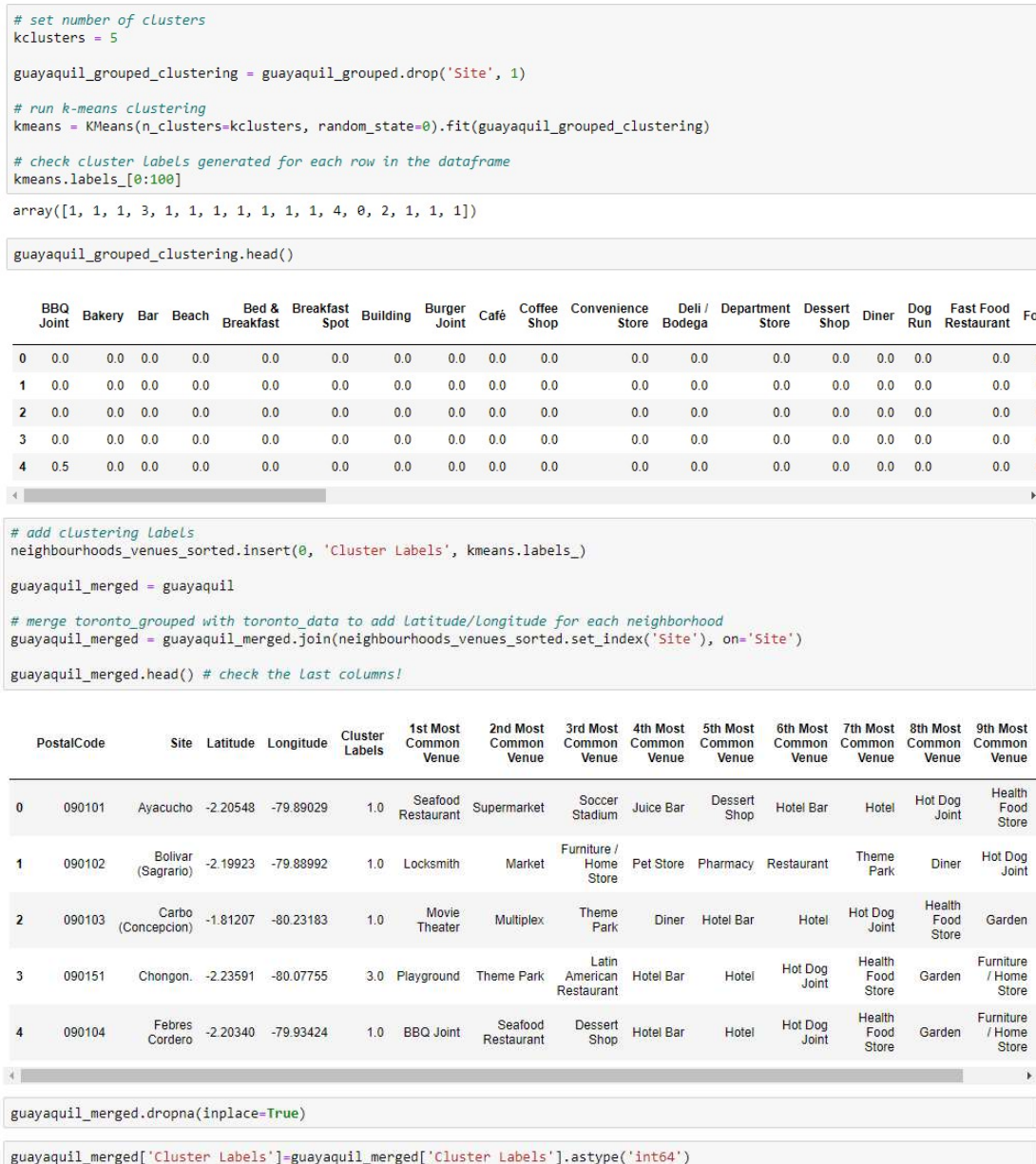


Figure 10. Example of Cluster Analysis applied.

4. Results

The first cluster analysis was developed for Guayaquil. It was applied K-Means algorithm and the results showed an interesting fact. The fact mentioned is that one cluster group all the parishes that are mainly located in the urban area. Other four clusters have a unique parish per cluster

due to the marked difference that their common venues have from the ones that are part of Cluster 2. The results per cluster are in the following lines:

4.1.1. Guayaquil Cluster 1

Guayaquil Cluster 1

```
#Cluster 1
guayaquil_merged.loc[guayaquil_merged['Cluster Labels'] == 0, guayaquil_merged.columns[[1] + list(range(5, guayaquil_merged.shape[1]))]
```

	Site	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
14	Posorja	Port	Theme Park	Latin American Restaurant	Hotel Bar	Hotel	Hot Dog Joint	Health Food Store	Garden	Furniture / Home Store	Food Truck

Figure 13. Guayaquil Cluster 1.

At first sight, it could be said that the parish included in this cluster has been correctly assigned. It could also be said that the venues information has been gotten in a right way due to the reason that the first most common venue assigned for Posorja was Port. It is correct because in this parish is been established the new port of the city. This cluster could be named as Port Cluster.

4.1.2. Guayaquil Cluster 2

Guayaquil Cluster 2

```
#Cluster 2
guayaquil_merged.loc[guayaquil_merged['Cluster Labels'] == 1, guayaquil_merged.columns[[1] + list(range(5, guayaquil_merged.shape[1]))]
```

	Site	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Ayacucho	Seafood Restaurant	Supermarket	Soccer Stadium	Juice Bar	Dessert Shop	Hotel Bar	Hotel	Hot Dog Joint	Health Food Store	Garden
1	Bolivar (Sagrario)	Locksmith	Market	Furniture / Home Store	Pet Store	Pharmacy	Restaurant	Theme Park	Diner	Hot Dog Joint	Health Food Store
2	Carbo (Concepcion)	Movie Theater	Multiplex	Theme Park	Diner	Hotel Bar	Hotel	Hot Dog Joint	Health Food Store	Garden	Furniture / Home Store
4	Febres Cordero	BBQ Joint	Seafood Restaurant	Dessert Shop	Hotel Bar	Hotel	Hot Dog Joint	Health Food Store	Garden	Furniture / Home Store	Food Truck
5	Garcia Moreno	Sandwich Place	Breakfast Spot	Food	Seafood Restaurant	Theme Park	Dessert Shop	Hotel Bar	Hotel	Hot Dog Joint	Health Food Store
6	Guayaquil	Seafood Restaurant	Restaurant	Stadium	Furniture / Home Store	Coffee Shop	Department Store	Hotel	Hot Dog Joint	Health Food Store	Garden
8	Letamendi	Restaurant	South Indian Restaurant	Food Truck	Pharmacy	Fast Food Restaurant	Convenience Store	Dessert Shop	Hotel	Hot Dog Joint	Health Food Store
9	Morro	Bakery	Sports Club	Rest Area	Theme Park	Dessert Shop	Hotel Bar	Hotel	Hot Dog Joint	Health Food Store	Garden
10	Nueva De Octubre	Pizza Place	Department Store	Hot Dog Joint	Breakfast Spot	Snack Place	Coffee Shop	Diner	Hotel Bar	Hotel	Health Food Store
12	Pascuales	Bakery	Health Food Store	South American Restaurant	Soccer Field	Theme Park	Dessert Shop	Hotel Bar	Hotel	Hot Dog Joint	Garden
16	Roca	Hotel	Seafood Restaurant	Coffee Shop	Plaza	Bar	Café	Juice Bar	Restaurant	BBQ Joint	Bakery
19	Tarqui	Restaurant	Building	Burger Joint	Furniture / Home Store	Dessert Shop	Hotel Bar	Hotel	Hot Dog Joint	Health Food Store	Garden
21	Urdaneta	BBQ Joint	Dessert Shop	Bar	Hotel	Seafood Restaurant	Fast Food Restaurant	Dog Run	Rest Area	Hotel Bar	Hot Dog Joint

Figure 14. Guayaquil Cluster 2.

Guayaquil Cluster 2 mainly contains the urban parishes of the city. It is interesting that the cluster algorithm captured the information pattern that makes these parishes similar to them but dissimilar to others not included due to identifying the types of venues that are common in the urban part of Guayaquil. As it could be seen, the venues type that are commonly available and visit in the urban area are Seafood Restaurants, Supermarkets, Bakeries, Hotels, Coffee Shops, Stadiums, and other. This cluster could be named as Urban Cluster.

4.1.3. Guayaquil Cluster 3

Guayaquil Cluster 3

```
#Cluster 3
guayaquil_merged.loc[guayaquil_merged['Cluster Labels'] == 2, guayaquil_merged.columns[[1] + list(range(5, guayaquil_merged.shape[1]))]
```

	Site	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
15	Puna	Beach	Theme Park	Diner	Juice Bar	Hotel Bar	Hotel	Hot Dog Joint	Health Food Store	Garden	Furniture / Home Store

Figure 15. Guayaquil Cluster 3.

Guayaquil Cluster 3 just contains a parish. Puna is an island located in the front of the urban area of Guayaquil. It contains a beach and some kind of venues related to fast food. Nevertheless, Puna has poorer population than Playas, that is also a beach, but it is well separated because Playas has different common venues type due to the economic situation.

4.1.4. Guayaquil Cluster 4

Guayaquil Cluster 4

```
#Cluster 4
guayaquil_merged.loc[guayaquil_merged['Cluster Labels'] == 3, guayaquil_merged.columns[[1] + list(range(5, guayaquil_merged.shape[1]))]
```

	Site	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
3	Chongon.	Playground	Theme Park	Latin American Restaurant	Hotel Bar	Hotel	Hot Dog Joint	Health Food Store	Garden	Furniture / Home Store	Food Truck

Figure 16. Guayaquil Cluster 4.

Guayaquil Cluster 4 contains a rural area of Guayaquil, which is Chongon. It contains different venues type that differ completely from other Guayaquil parishes because of the common venues. For that reason, the cluster algorithm categorizes it in a different cluster.

4.1.5. Guayaquil Cluster 5

Guayaquil Cluster 5

```
#Cluster 5
guayaquil_merged.loc[guayaquil_merged['Cluster Labels'] == 4, guayaquil_merged.columns[[1] + list(range(5, guayaquil_merged.shape[1]))]
```

	Site	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
13	Playas (General Villamil)	Supermarket	Theme Park	Juice Bar	Hotel Bar	Hotel	Hot Dog Joint	Health Food Store	Garden	Furniture / Home Store	Food Truck

Figure 17. Guayaquil Cluster 5.

According to the data source analyzed, Playas is a parish of Guayaquil. Playas is a beach located one hour far from Guayaquil downtown. It has been assigned into a different cluster because of the type of common venues that are in the radius area of the parish. This is because the economic condition of Playas is better than the economic condition of Puna.

4.2. Quito Cluster Analysis

It has been found that Quito's parishes were grouped into an urban cluster and three rural clusters. It shows that the rural parishes have different venues type when focusing on this type

of area. Nonetheless, in the urban area there are no different patten of venues per parishes which is the reason why this parishes where group in just one cluster.

4.2.1. Quito Cluster 1

Quito Cluster 1

```
#Cluster 1
quito_merged.loc[quito_merged['Cluster Labels'] == 0, quito_merged.columns[[1] + list(range(5, quito_merged.shape[1]))]]
```

	Site	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Alangasi	Park	Wings Joint	Cosmetics Shop	Food Truck	Food Court	Food & Drink Shop	Food	Flea Market	Fast Food Restaurant	Farmers Market
22	El Beaterio	Park	Wings Joint	Cosmetics Shop	Food Truck	Food Court	Food & Drink Shop	Food	Flea Market	Fast Food Restaurant	Farmers Market
30	Guayllabamba	Restaurant	Park	Wings Joint	Farmers Market	Food Court	Food & Drink Shop	Food	Flea Market	Fast Food Restaurant	Farm
42	La Vicentina	Park	Wings Joint	Cosmetics Shop	Food Truck	Food Court	Food & Drink Shop	Food	Flea Market	Fast Food Restaurant	Farmers Market

Figure 18. Quito Cluster 1.

Quito Cluster 1 is characterized for having parks, food courts and farms. It shows that this cluster has rural parishes where people can do outdoor distraction activities. This cluster could be named as Parks Cluster.

4.2.2. Quito Cluster 2

Quito Cluster 2

```
#Cluster 2
quito_merged.loc[quito_merged['Cluster Labels'] == 1, quito_merged.columns[[1] + list(range(5, quito_merged.shape[1]))]]
```

	Site	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
1	Alfaro (Chimbacalle)	Science Museum	South American Restaurant	Food Truck	Food Court	Food & Drink Shop	Food	Flea Market	Fast Food Restaurant	Farmers Market	Wings Joint
3	Atahualpa (Chabaspamba)	Hotel	Bar	Cosmetics Shop	Mexican Restaurant	Cajun / Creole Restaurant	Food Court	Music Venue	Historic Site	Food Truck	Department Store
4	Benalcazar	Italian Restaurant	Coffee Shop	Hotel	Bakery	Ice Cream Shop	Movie Theater	Pizza Place	Fast Food Restaurant	Japanese Restaurant	Sandwich Place
6	Calderon (Carapungo)	Pizza Place	Chinese Restaurant	Park	Wings Joint	Farmers Market	Food Court	Food & Drink Shop	Food	Flea Market	Fast Food Restaurant
7	Carcelen	Farmers Market	Soccer Field	Pizza Place	Seafood Restaurant	BBQ Joint	Gym	Wings Joint	Food & Drink Shop	Food	Flea Market
8	Centro Historico	Breakfast Spot	Burger Joint	Pizza Place	Restaurant	Seafood Restaurant	Stadium	Farmers Market	Food & Drink Shop	Food	Flea Market

Figure 19. Quito Cluster 2.

Quito Cluster 2 characterizes for having some different venues type. This cluster shows a typically pattern of an urban area in any of the main Ecuadorian cities. This cluster contains the parishes that has stadiums, food trucks, cafeterias, historic sites, etc. around. This cluster could be named as Urban Area Cluster.

4.2.3. Quito Cluster 3

Quito Cluster 3

```
#Cluster 3
quito_merged.loc[quito_merged['Cluster Labels'] == 2, quito_merged.columns[[1] + list(range(5, quito_merged.shape[1]))]]
```

	Site	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
51	Nono	Recreation Center	Cosmetics Shop	Food Truck	Food Court	Food & Drink Shop	Food	Flea Market	Fast Food Restaurant	Farmers Market	Farm
59	Puemblo	Recreation Center	Athletics & Sports	Farm	Food Truck	Food Court	Food & Drink Shop	Food	Flea Market	Fast Food Restaurant	Farmers Market

Figure 20. Quito Cluster 3.

Quito Cluster 3 mainly has recreation centers and food courts. It denotes that these parishes are located in an area where people do different kind of sports and can get some fast food. This cluster could be named as Recreation Area Cluster.

4.2.4. Quito Cluster 4

Quito Cluster 4

```
#Cluster 4
quito_merged.loc[quito_merged['Cluster Labels'] == 3, quito_merged.columns[[1] + list(range(5, quito_merged.shape[1]))]]
```

	Site	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
78	Yaruqui	Cafeteria	Farm	Wings Joint	Fast Food Restaurant	Food Truck	Food Court	Food & Drink Shop	Food	Flea Market	Farmers Market

Figure 21. Quito Cluster 4.

Quito Cluster 4 has only a parish, Yaruqui. It is known that Yaruqui is a rural parish of Quito. Due to the type of common venues that this parish has, it is clear that this is a farm area with restaurants, but they are mainly fast food restaurants. Therefore, this cluster could be named as Farm Cluster.

4.2.5. Quito Cluster 5

Quito Cluster 5

```
#Cluster 5
quito_merged.loc[quito_merged['Cluster Labels'] == 4, quito_merged.columns[[1] + list(range(5, quito_merged.shape[1]))]]
```

	Site	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
36	La Ecuatoriana	Fried Chicken Joint	Seafood Restaurant	French Restaurant	Food Court	Food & Drink Shop	Food	Flea Market	Fast Food Restaurant	Farmers Market	Farm
55	Pifo	Seafood Restaurant	Wings Joint	French Restaurant	Food Court	Food & Drink Shop	Food	Flea Market	Fast Food Restaurant	Farmers Market	Farm

Figure 22. Quito Cluster 5.

Quito Cluster 5 has also two rural parishes like Quito Cluster 4 that have farms around. The difference between Quito Cluster 4 and Cluster 5 is that in Quito Cluster 5, it could be found specialized restaurants, like French restaurants, and that in Quito Cluster 4 there is no common cafeteria venues. This cluster could be named as Farm & Specialized Restaurants Cluster.

5. Discussion

Based on the results, a recommendation for a future analysis is to explore the segment groups that area in the urban cluster of both cities at the neighborhood level. Analyzing the neighborhood parishes that the present project has grouped into the urban cluster let it knows

the groups that have similar composition according to the common venues and has a better characterization of the urban cluster. Nevertheless, it should be built a data frame that contains the neighborhoods and the location of the centroid latitude and longitude of each one.

6. Conclusion

In this project, the two most important cities of Ecuador were analyzed. Putting in contrast the characteristics of the clusters obtained per each city, it was shown a clearly patten that denotes that urban parishes have different type of common venues when comparing them to the rural ones. In both cities, the urban parishes were located in one cluster due to the set of venues types they have. On the other hand, the rural parishes were located in different clusters according to their common type of venues. In Quito, some clusters contain parishes that have parks or farms around them. In contrast, Guayaquil rural parishes showed to have very different type of venues than Quito like beach, port, and others. To conclude, it could be inferred with the results of this project that the common venues in rural areas differ from urban areas due to geographic location. In cities like Quito that are located between mountains it could be found farms, parks and recreation areas, not the same in port cities like Guayaquil where it could be found beach and port.