

A map of Ecuador showing various geographical features and cities. Overlaid on the map are numerous colored dots (purple, red, orange, yellow) representing different cities and their parish segmentation characteristics. The dots are concentrated in the central and southern regions of the country. A large red semi-transparent rectangle is overlaid on the map, containing the title text.

Ecuador's top Cities and their Parish Segmentation Characteristics

Xavier Jácome Piñeiros

"The greatest value of a picture is when it forces us to notice what we never expected to see" John W. Tukey



Web

<https://www.linkedin.com/in/xavierjacomep/>



BACKGROUND

Ecuador is a small country located in South America with around of 17 million of habitants.

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 capital city, Quito, it is divided into 70 parishes. In the other hand, the main port city, Guayaquil, it is divided into 24 parishes.

The World Bank data shows that Ecuador produced a GDP of 108.398 Billion dollar in 2018.



Problem

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.



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.

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.



Data Sources

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.

The second data source was the Foursquare database.



Data Cleaning

1. Identifying null values and drop them.
2. Identifying duplicates and consider just correct data.
Incorrect data was dropped.
3. Save the clean datasets into a data frames.

METHODOLOGY

How the analysis was conducted?

EXPLORATORY DATA ANALYSIS TECHNIQUES

- 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.

MACHINE LEARNING TECHNIQUES

- K-Means algorithm was applied, a non-supervised machine learning algorithm, which is a cluster analysis algorithm.
- 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.

EXPLORATORY

Frequency tables of the first five element with the highest frequencies of the venues type.

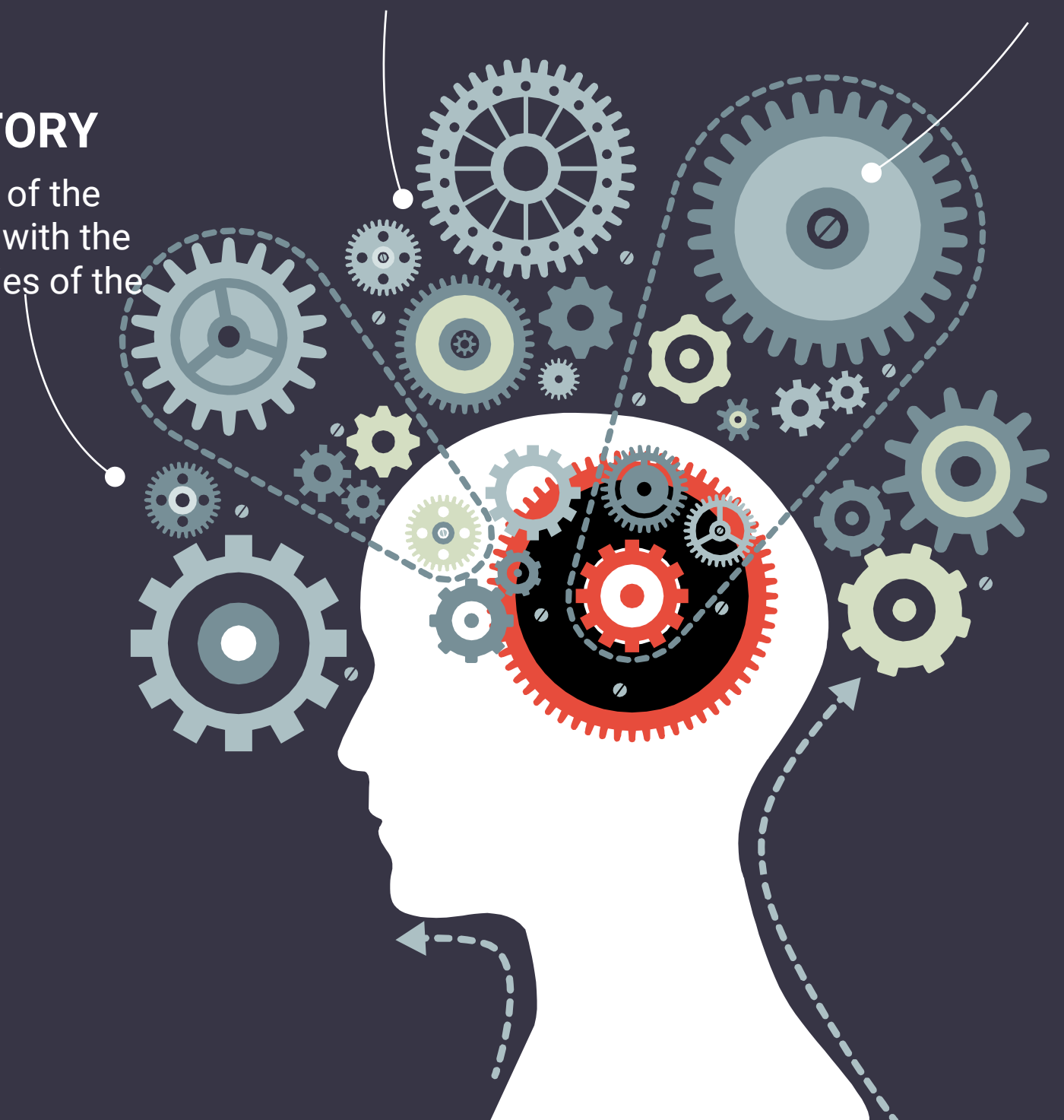
K-MEANS

It was decided to generate five clusters per city.

MAPS

Folium library was used.

Once the clusters were created, it was created a map per city to show every parish with the color that identifies to what cluster it belongs. The maps were created using the folium library.



METHODOLOGY

Exploratory Data Analysis Examples

```
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

----Bolivar (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

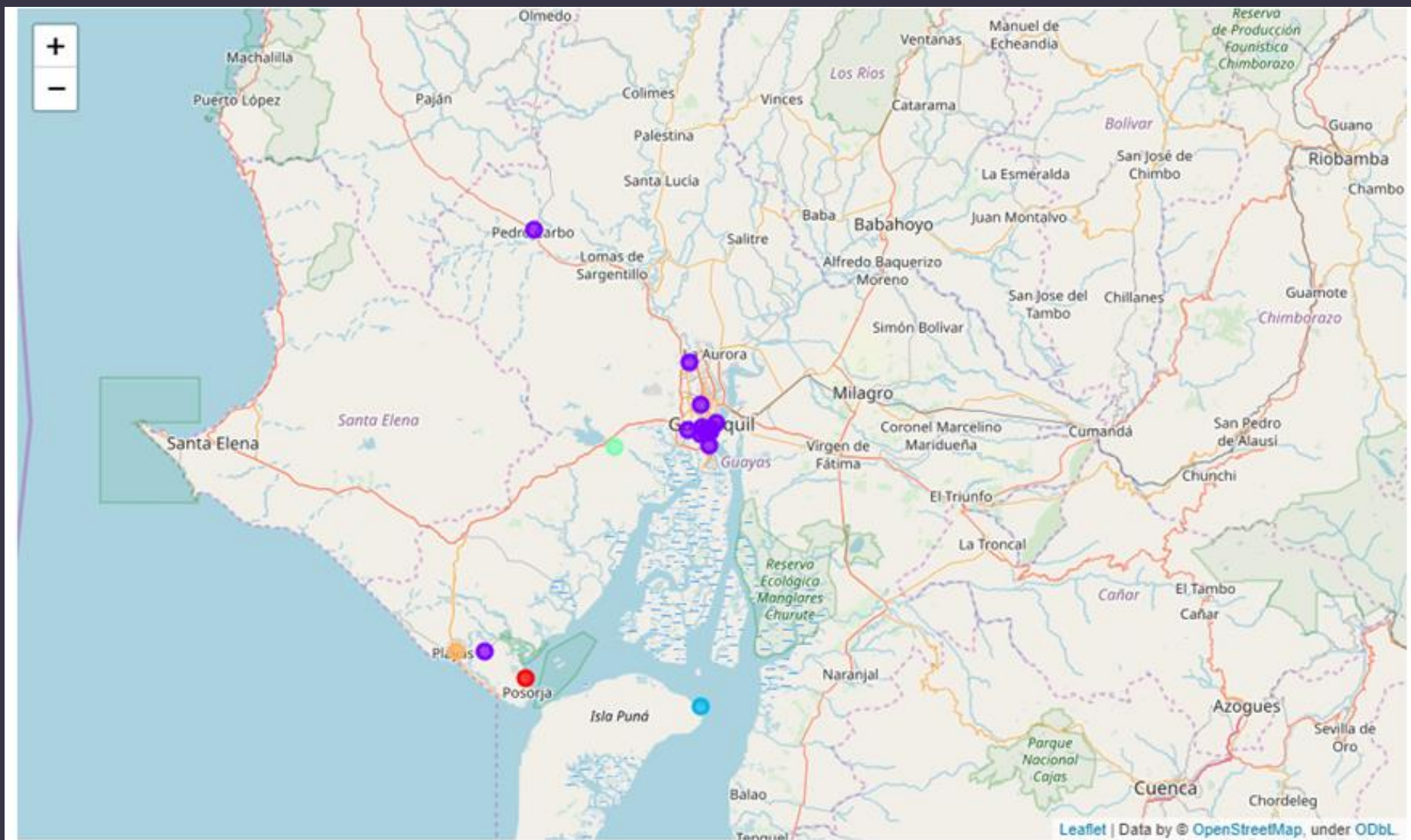
for hood in quito_grouped['Site']:
    print("----"+hood+"----")
    temp = quito_grouped[quito_grouped['Site'] == hood].T.reset_index()
    temp.columns = ['venue', 'freq']
    temp = temp.iloc[1:]
    temp['freq'] = temp['freq'].astype(float)
    temp = temp.round({'freq': 2})
    print(temp.sort_values('freq', 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

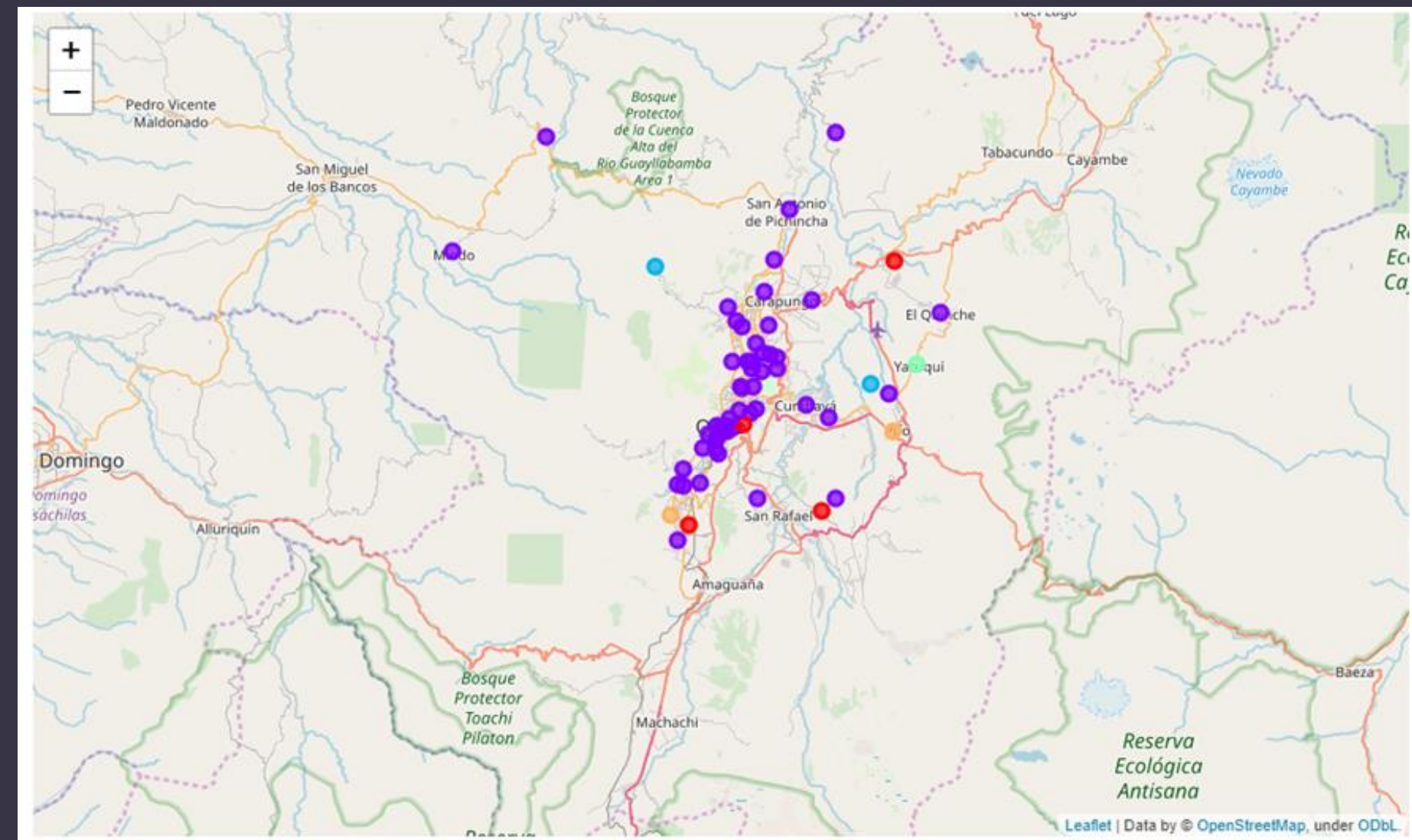
----Alfaro (Chimbacalle)----
      venue  freq
0  Science Museum  0.67
1  South American Restaurant  0.33
2  American Restaurant  0.00
3      Pharmacy  0.00
4  Pakistani Restaurant  0.00
```


METHODOLOGY

Machine Learning Analysis Examples



Guayaquil clusters



Quito clusters

```
# set number of clusters
kclusters = 5

guayaquil_grouped_clustering = guayaquil_grouped.drop('Site', 1)

# run k-means clustering
kmeans = KMeans(n_clusters=kclusters, random_state=0).fit(guayaquil_grouped_clustering)

# check cluster labels generated for each row in the dataframe
kmeans.labels_[0:100]

array([1, 1, 1, 3, 1, 1, 1, 1, 1, 1, 1, 1, 4, 0, 2, 1, 1, 1])
```

```
guayaquil_grouped_clustering.head()
```

	BBQ Joint	Bakery	Bar	Beach	Bed & Breakfast	Breakfast Spot	Building	Burger Joint	Café	Coffee Shop	Convenience Store	Deli / Bodega	Department Store	Dessert Shop	Diner	Dog Run	Fast Food Restaurant	Foc
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0
1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0
2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0
3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0
4	0.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0

RESULTS

Guayaquil Cluster Analysis



Cluster 1

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.



Cluster 2

Guayaquil Cluster 2 mainly contains the urban parishes of the city. The venues type that are commonly available and visit in the urban area are Seafood Restaurants, Supermarkets, Bakeries, Hotels, Coffee Shops, Stadiums, and other.



Cluster 3

Guayaquil Cluster 3 just contains Puna, an island located in the front of the urban area. 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. They have different common venues type due to the economic situation.



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.



Cluster 5

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.

RESULTS

Guayaquil Cluster Analysis Examples

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

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

RESULTS

Quito Cluster Analysis



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.



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.



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.



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.



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 5 there is no common cafeteria venues.

RESULTS

Quito Cluster Analysis Examples

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

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

DISCUSSION

Recommendations for Future Analysis

- 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.



CONCLUSION

Recommendations for Future Analysis

- Putting in contrast the characteristics of the clusters obtained per each city, it was shown a clearly patter 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.

- 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.

