The Caribbean venues

Bernardo Silva

April 30, 2020

1 INTRODUCTION

1.1 Background

The main goal of my project is to analyze The Caribbean (figure 1), which is a region of the Americas that consists of the Caribbean Sea, its islands (some surrounded by the Caribbean Sea and some bordering both the Caribbean Sea and the North Atlantic Ocean) and the surrounding coasts.

This region has more than 700 islands performing 35 countries/territories, diverse ethnic groups and religions. Tourism is one of the Caribbean's major economic sectors. Tourist attraction of the region are those generally associated with a maritime tropical climate: Scuba diving and snorkelling on coral reefs, cruises, sailing, and game fishing at sea. On land; golf, botanical gardens, parks, limestone caves, wildlife reserves, hiking, cycling and horseback riding.

1.2 Problem

Given the high diversity among the different islands in this region it is not an easy task to find information in only one place that can portray an insight of what to expect in the different territories.
An assessment will be performed exploring the characteristic avenues of the Caribbean showing the similarities and differences between the different places.

1.3 Interest

This project is indicated to everyone that is interested in traveling to the Caribbean, as well as to traveling agencies. Providing them good information to make a choice or suggestion of a route given the client preferences.
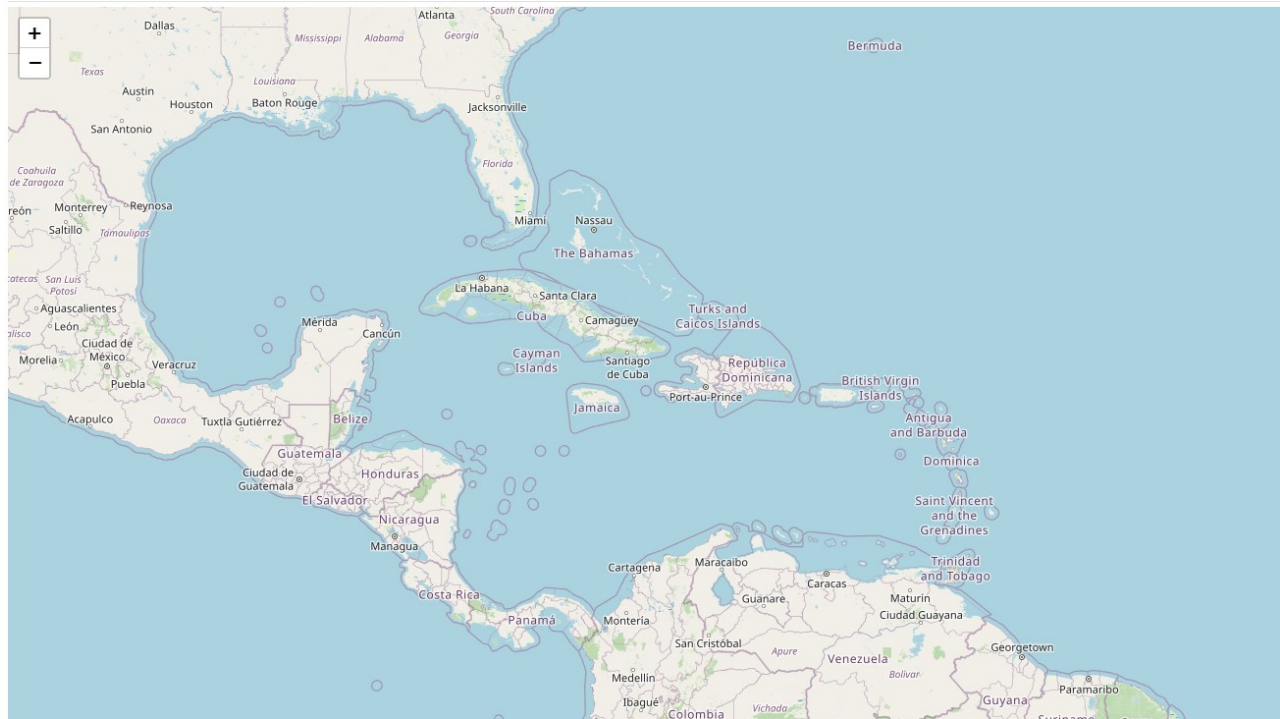
Figure 1. The Caribbean region.

2 DATA

2.1 *Acquisition*

In order to get a meaningful data, 284 cities ranked by population in The Caribbean were chosen based on the followin website: `https://population.mongabay.com/population/region/caribbean`.

The data was saved in the 'caribbean.csv' file composed of two columns, the name of the city and the name of the county/territory that city belongs to. The library 'pandas' is the used to read the '.csv' file and save it as a dataframe named 'df', figure (2) shows the first 10 rows.

	Country	City
0	Anguilla	The Valley
1	Antigua and Barbuda	Saint John's
2	Aruba	Angochi
3	Aruba	Babijn
4	Aruba	Oranjestad
5	Bahamas	Freeport
6	Bahamas	Nassau
7	Barbados	Bridgetown
8	Bermuda	Hamilton
9	Bonaire	Kralendijk

Figure 2. First 10 rows of dataframe 'df' imported from caribbean.csv file .

2.2 Treatment

The geopy library is used to get the latitude and longitude of every city, and this information is added to the dataframe 'df' as shown in image (3).

	Country	City	Latitude	Longitude
0	Anguilla	The Valley	18.214586	-63.051776
1	Antigua and Barbuda	Saint John's	17.118457	-61.844851
2	Aruba	Angochi	12.520460	-69.964053
3	Aruba	Babijn	12.540947	-69.995816
4	Aruba	Oranjestad	12.526874	-70.035685
5	Bahamas	Freeport	26.535681	-78.695362
6	Bahamas	Nassau	25.078346	-77.338333
7	Barbados	Bridgetown	13.097783	-59.618418
8	Bermuda	Hamilton	32.295608	-64.782705
9	Bonaire	Kralendijk	12.147174	-68.274078

Figure 3. First 10 rows of dataframe 'df' after being joined with the columns Latitude and Longitude .

A map of The Caribbean region is created using the 'folium' library with the chosen cities superimposed on top (figure 4).
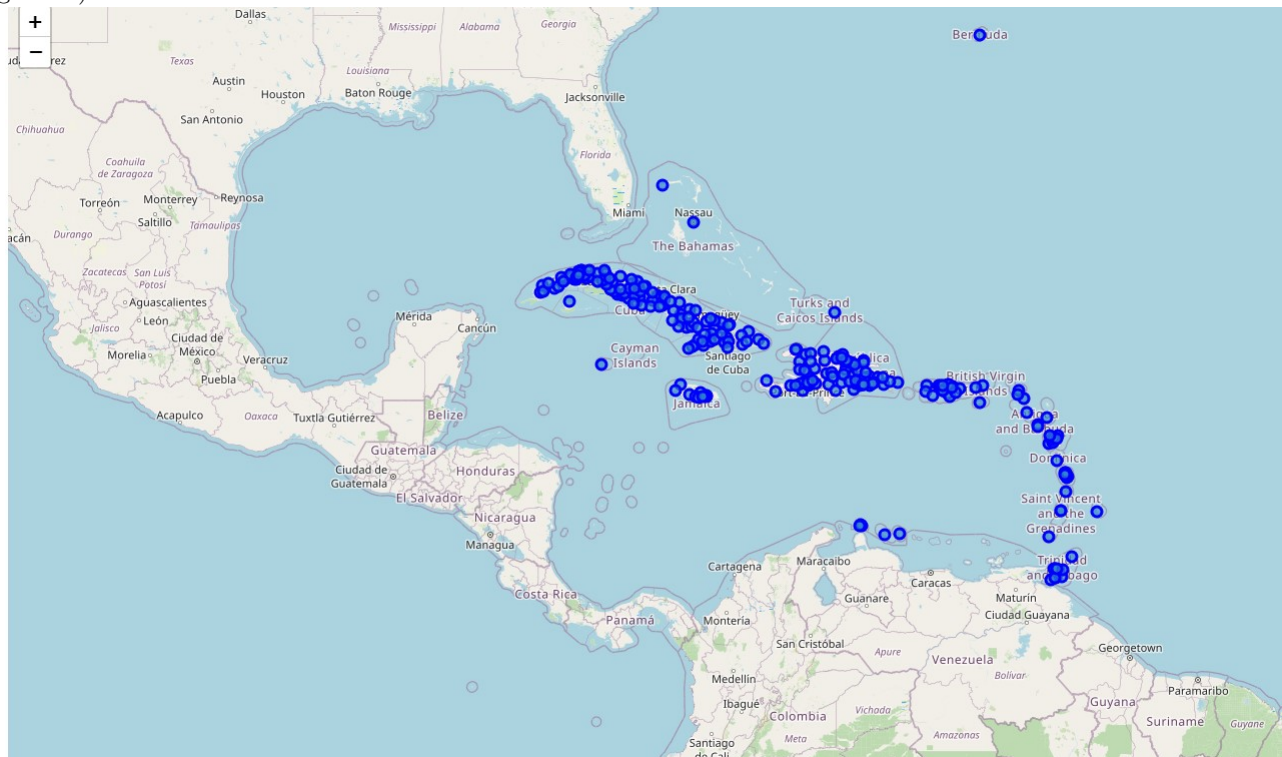

Figure 4. The Caribbean region with cities.

Foursquare API is used to explore each city and find the top 100 venues that are in radius of 2500 meters. An example of the Valley city is shown in figure (5).

	City	City Latitude	City Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	The Valley	18.214586	-63.051776	Shoal Bay Beach	18.254033	-63.031663	Beach
1	The Valley	18.214586	-63.051776	Veya	18.201734	-63.079456	Caribbean Restaurant
2	The Valley	18.214586	-63.051776	Gwen's Reggae Bar & Grill	18.252853	-63.034531	Caribbean Restaurant
3	The Valley	18.214586	-63.051776	Roy's Bayside Grill	18.197958	-63.091111	Steakhouse
4	The Valley	18.214586	-63.051776	Tasty's	18.196671	-63.084617	Caribbean Restaurant
5	The Valley	18.214586	-63.051776	Sandy Ground Anguilla	18.201715	-63.090445	Beach
6	The Valley	18.214586	-63.051776	Da Vida	18.219870	-63.066997	Caribbean Restaurant
7	The Valley	18.214586	-63.051776	Pumphouse	18.202669	-63.090122	Pub
8	The Valley	18.214586	-63.051776	Jacala	18.184076	-63.135961	French Restaurant
9	The Valley	18.214586	-63.051776	Sun Shine Shack	18.177536	-63.113524	Bar

Figure 5. The Valley venues.

3 METHODOLOGY

3.1 *Classifier*

The classifier *k-means clustering* is applied in order to partition the data into non-overlapping subsets, 6 *clusters*, minimizing the intra-cluster distance and maximizing the inter-cluster distance based on the top 10 Venue Categories of each city. Similar groups may then be found.

3.1.1 Procedure

From the *pandas* library, the function *get dummies* is used to convert all venue values into zeros (if the venue is not applied to the respective city) and ones (if venue is applied). The result may be inspected in figure 6.

	Cities	Adult Boutique	Airport	Airport Food Court	Airport Lounge	Airport Service	Airport Terminal	American Restaurant	Amphitheater	Antique Shop	Aquarium	Arcade	Arepa Restaurant	Argentinian Restaurant	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant	Athletics & Sports	BBQ Joint	Bagel Shop	Bakery	Bank	Bar	Baseball Field	Baseball Stadium
0	The Valley	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	The Valley	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	The Valley	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	The Valley	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	The Valley	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
5	The Valley	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
6	The Valley	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
7	The Valley	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
8	The Valley	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
9	The Valley	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

Figure 6. Venues hot encoding dataframe.

A function called *return_most_common_venues* is defined to find the top 10 most common venues from each city and then the data is saved in a dataframe, figure 7.

	Cities	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	Abreus	Plaza	Cafeteria	Hotel	Pier	Theater	Zoo Exhibit	Empanada Restaurant	Exhibit	Factory	Farm
1	Aguada de Pasajeros	Caribbean Restaurant	Zoo Exhibit	Eastern European Restaurant	Empanada Restaurant	Exhibit	Factory	Farm	Farmers Market	Fast Food Restaurant	Fish & Chips Shop
2	Aguadilla	Caribbean Restaurant	Coffee Shop	Beach	Sandwich Place	Bakery	Pizza Place	Surf Spot	Café	Ice Cream Shop	Resort
3	Alamar	Hotel	Cuban Restaurant	Beach	Park	Nightclub	Hotel Bar	Plaza	Cocktail Bar	Restaurant	Café
4	Alquízar	BBQ Joint	American Restaurant	College Arts Building	Vegetarian / Vegan Restaurant	Plaza	Zoo Exhibit	Fish Market	Exhibit	Factory	Farm
5	Amancio	Park	Antique Shop	Zoo Exhibit	Fish Market	Exhibit	Factory	Farm	Farmers Market	Fast Food Restaurant	Fish & Chips Shop
6	Angochi	Beach	Resort	Caribbean Restaurant	Hotel	Grocery Store	Seafood Restaurant	Scenic Lookout	Food Court	Brazilian Restaurant	Chinese Restaurant
7	Arecibo	Caribbean Restaurant	Restaurant	Coffee Shop	Bar	Department Store	Bakery	Pharmacy	BBQ Joint	Sandwich Place	Pizza Place
8	Arima	Bakery	Pizza Place	Indian Restaurant	Café	Scenic Lookout	Food Truck	Gym	Ice Cream Shop	Bar	Grocery Store
9	Arroyo Naranjo	Boat or Ferry	Southern / Soul Food Restaurant	Cuban Restaurant	Park	Brewery	Factory	Farm	Farmers Market	Fast Food Restaurant	Fish & Chips Shop

Figure 7. Top 10 most common category venues.

Now that the top 10 most common venues were found, running *k-means clustering*, the cities are partitioned into 6 clusters. This results in a distribution of subsets classified from [0 to 5] which is added into the data in column *Cluster Labels*. Merging the main information results in the final dataframe named *caribbean merged*, as shown in figure 8.

	Country	City	Latitude	Longitude	Cluster Labels	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	Anguilla	The Valley	18.214586	-63.051776	5.0	Beach	French Restaurant	Resort	Caribbean Restaurant	Bar	Hotel	Italian Restaurant	Seafood Restaurant	Beach Bar	Hotel Bar
1	Antigua and Barbuda	Saint John's	17.118457	-61.844851	5.0	Resort	Beach	Caribbean Restaurant	Hotel	Bar	Café	Middle Eastern Restaurant	Sandwich Place	Cocktail Bar	Mediterranean Restaurant
2	Aruba	Angochi	12.520460	-69.964053	5.0	Beach	Resort	Caribbean Restaurant	Hotel	Grocery Store	Seafood Restaurant	Scenic Lookout	Food Court	Brazilian Restaurant	Chinese Restaurant
3	Aruba	Babijn	12.540947	-69.995816	5.0	Beach	Resort	Caribbean Restaurant	Hotel	Grocery Store	Seafood Restaurant	Pub	Bar	Brazilian Restaurant	Chinese Restaurant
4	Aruba	Oranjestad	12.526874	-70.035685	5.0	Beach	Resort	Caribbean Restaurant	Hotel	Grocery Store	Seafood Restaurant	Bar	Fast Food Restaurant	Restaurant	Bakery
5	Bahamas	Freeport	26.535681	-78.695362	5.0	Resort	Caribbean Restaurant	Beach	Restaurant	Pizza Place	Mobile Phone Shop	Hotel Bar	Bar	Latin American Restaurant	Fast Food Restaurant
6	Bahamas	Nassau	25.078346	-77.338333	5.0	Beach	Resort	Hotel	Caribbean Restaurant	Italian Restaurant	Restaurant	Ice Cream Shop	Café	Coffee Shop	Sushi Restaurant
7	Barbados	Bridgetown	13.097783	-59.618418	5.0	Beach	Caribbean Restaurant	Café	Bar	Resort	Sandwich Place	Seafood Restaurant	Hotel	Asian Restaurant	Food Truck
8	Bermuda	Hamilton	32.295608	-64.782705	5.0	Restaurant	Beach	Resort	Café	Hotel	Seafood Restaurant	Golf Course	Harbor / Marina	Bar	Coffee Shop
9	Bonaire	Kralendijk	12.147174	-68.274078	5.0	Resort	Beach	Café	Hotel	Surf Spot	Caribbean Restaurant	Bar	Beach Bar	Grocery Store	BBQ Joint

Figure 8. Final dataframe, caribbean-merged.

4 RESULTS

The resulting clusters may be visualized overlap with the Caribbean map as depicted in figure 9.

Cluster's color:

- Cluster 0 - Red

- Cluster 1 - Purple

- Cluster 2 - Blue

- Cluster 3 - Soft Blue

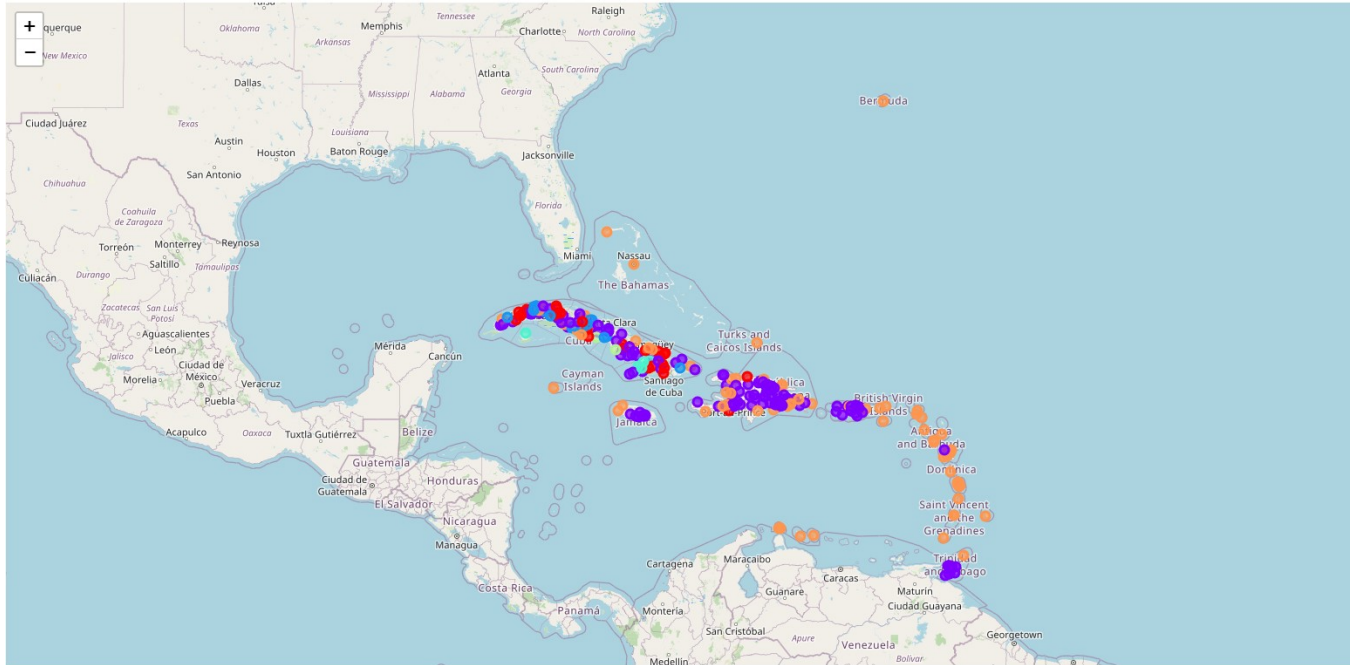- Cluster 4 - Green

- Cluster 5 - Orange

Figure 9. The Caribbean map with clusters.

4.1 Cluster's examination

Each cluster is composed of different countries which have in common the kind of venues they offer the most. The goal now is to determine the discriminating venue categories that distinguish each cluster. For cluster 0, the figure (10) shows the first 5 rows of the table of the top 10 most common venues.

	City	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
18	Artemisa	Hotel	Nature Preserve	Café	Park	Scenic Lookout	Vegetarian / Vegan Restaurant	Fast Food Restaurant	Exhibit	Factory	Farm
19	Báguanos	Furniture / Home Store	Baseball Stadium	Hotel	Zoo Exhibit	Fish Market	Exhibit	Factory	Farm	Farmers Market	Fast Food Restaurant
20	Bahía Honda	Hotel	Nature Preserve	Café	Park	Scenic Lookout	Zoo Exhibit	Fast Food Restaurant	Exhibit	Factory	Farm
21	Banes	Hotel	Beach	Zoo Exhibit	Fishing Spot	Exhibit	Factory	Farm	Farmers Market	Fast Food Restaurant	Fish & Chips Shop
23	Baraguá	Stadium	Baseball Stadium	Hotel	Zoo Exhibit	Fish & Chips Shop	Empanada Restaurant	Exhibit	Factory	Farm	Farmers Market

Figure 10. Cluster 0 table.

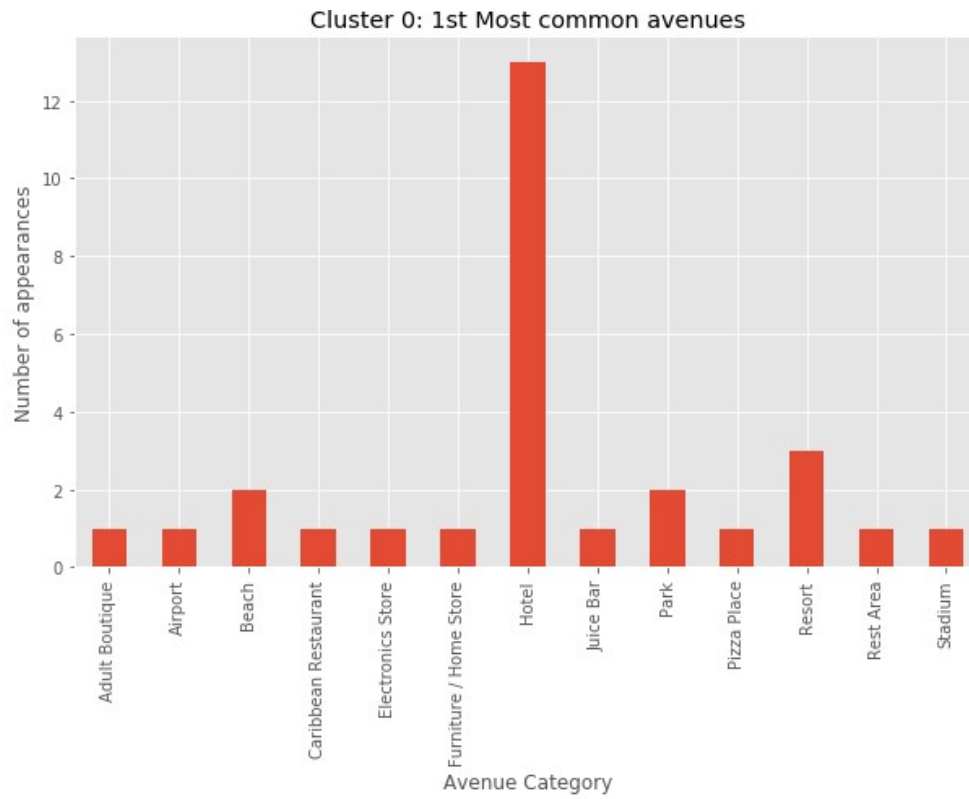In order to get a better visualization of the data, a bar graph was used. Some examples are shown, figures (11 to 14).

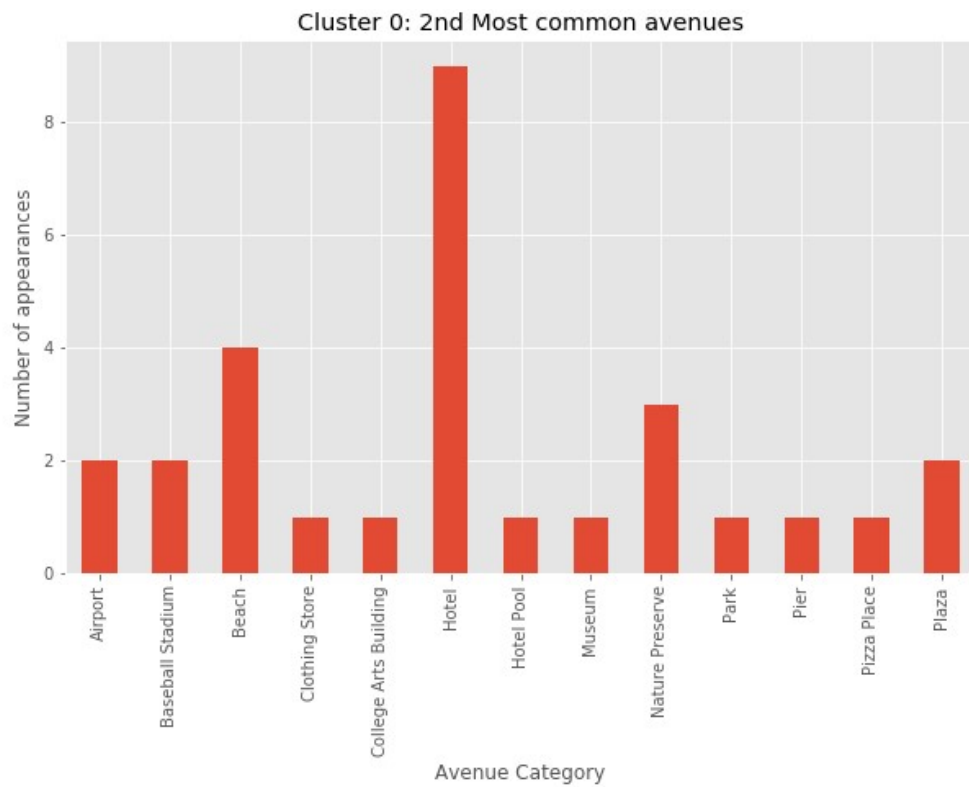**Figure 11.** Cluster 0: 1st most common venues.

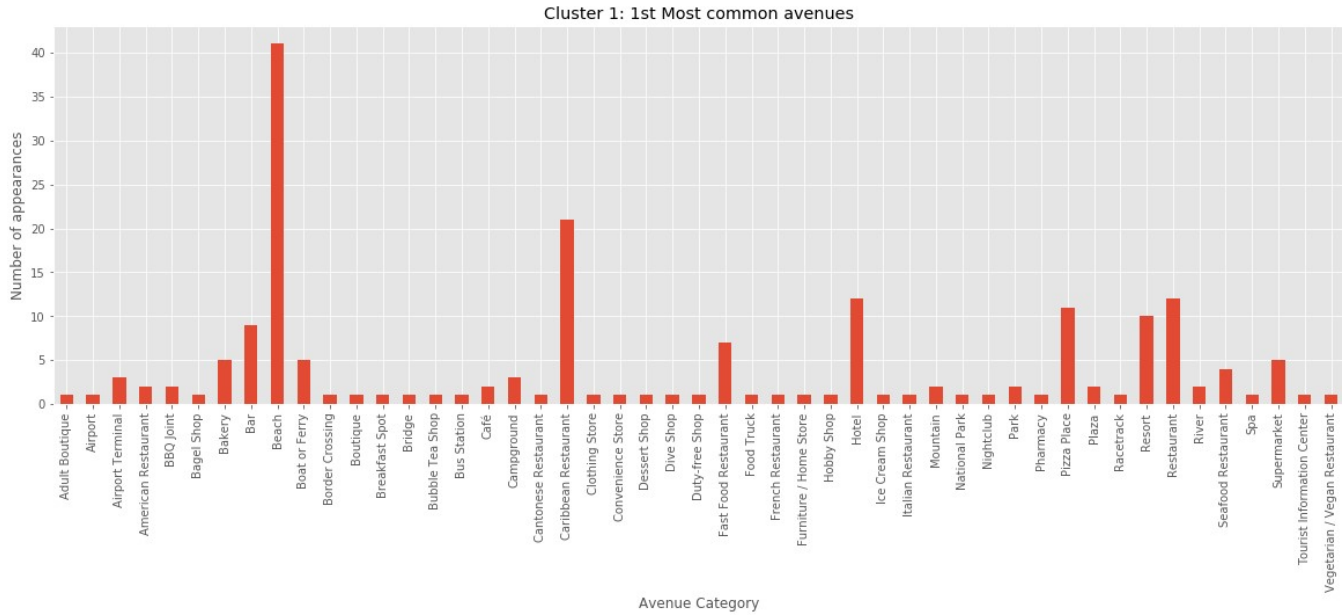**Figure 12.** Cluster 0: 2nd most common venues.

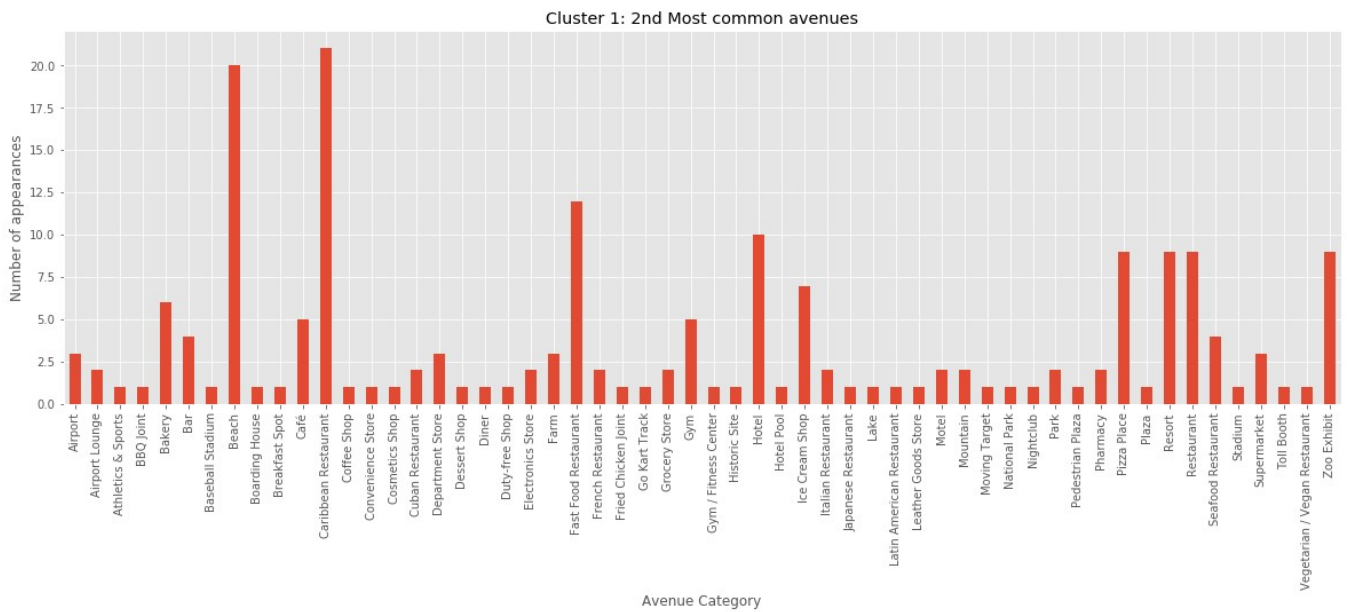**Figure 13.** Cluster 1: 1st most common venues..

Figure 14. Cluster 1: 2nd most common venues..

5 DISCUSSION

Inspecting the several bar graphs obtained in the Notebook code (some examples presented in the Results section), each cluster may be mainly characterized by:

- **Cluster 0** - big hotel's offer, zoo exhibits and parks. It has 29 cities.

- **Cluster 1** - big offer of beaches, Caribbean Restaurants and pizza places. Restaurants and bars are definitely a big investment in the cities that belong to this cluster. It is also the cluster with the higher number of cities, 190.

- **Cluster 2** - offer in Airports, zoo exhibition. It is a small cluster with 6 cities.

- **Cluster 3** - offer of Bed & Breakfast and Zoo Exhibitions. It is the smallest cluster with only 2 cities.

- **Cluster 4** - Cuban Restaurant, hotel offer and beaches. This cluster has 47 cities.

- **Cluster 5** - offer of Burger joint, zoo exhibition and cricket ground. It has 4 cities.

It may be inspected in the map of figure (9) that cluster 4 (green color) which is characterized by the big offer of Cuban restaurants, so expected to be concentrated in Cuba, turns out to be spread as well over the Dominican Republic.

Clusters 2 and 3 are small. Still, from the number 2 (blue) which is characterized by categorical avenue *Airport*, in Cuba for instance, may be an indication of cities apart from the urban zones. For the *Isla de la Juventude* the dot is the main city, which may be an indication of a place with low industrialization.

The most spread cluster over the Caribbean region is the number 1 (Purple), characterized by the great offer of beaches and Caribbean Restaurants. Which was also expected, since the Caribbean is formed by islands.

Cuba is the country were great number of cities has as most common venue the offer of hotels, as can be seen by the red circles (cluster 0) in the map (figure 9)

One particular conclusion that may be obtained from this project that, aim to show the difference between group's of cities distributed all over the Caribbean, is that *Zoo exhibitions* stands out by being a general characteristic of all the cities assessed in this project, since every cluster shows a indication of it as one of the most common venues.

6 CONCLUSION

This project gives a good insight of the kind of activities practiced all over the Caribbean region just by looking at the map in figure 9. It may be a good start point to filter places if a person desires to fly to the Caribbean but is not sure were to go. However, is a very general view of the region, so further search may be done before jumping to conclusions.

Some interesting insights about this region, as the ones referred in the *Discussion* section, toke shape while assessing this only document and further perspectives may be found from it.

The information for this project is based on the Foursquare API, which provided enough details to make this work, however, the question about its database and update remains. Some cities may not have great part of the information on their database for instance.

The cities were chosen based on the population rank, it will be interesting to add more cities to the data used in this project. For instance, from The Bahamas only two cities participated which was kind of a disappointment and it would've been interesting to have more from there.