MOVING TO ATLANTA, GEORGIA. ANALYSIS OF THE JOURNEY.

Capstone Project: The Battle of Neighborhoods

Introduction

The main problem is the difficulty that people have to take decision when they want to migrate to other state or countries. In this particular case I will migrate to Atlanta, Georgia in some months so creating this analysis will help me to take a better decision selecting a City in Atlanta, rent an apartment, select a good school and university for my daughter and son. One important point is consider Fulton County in Atlanta as a cardinal point to start the analysis around the cities.

For that reason, in the solution to develop analysis of features for a people migrating to Atlanta and search a best city as a comparative analysis. In that case showing list of venues available by cities which allow to show the bunch of categories to consider when rent a house. The features include median housing price and better school according to ratings, to taking decision. Also provide a plot information with index in schools, universities and house prices in Atlanta cities.

Data Section

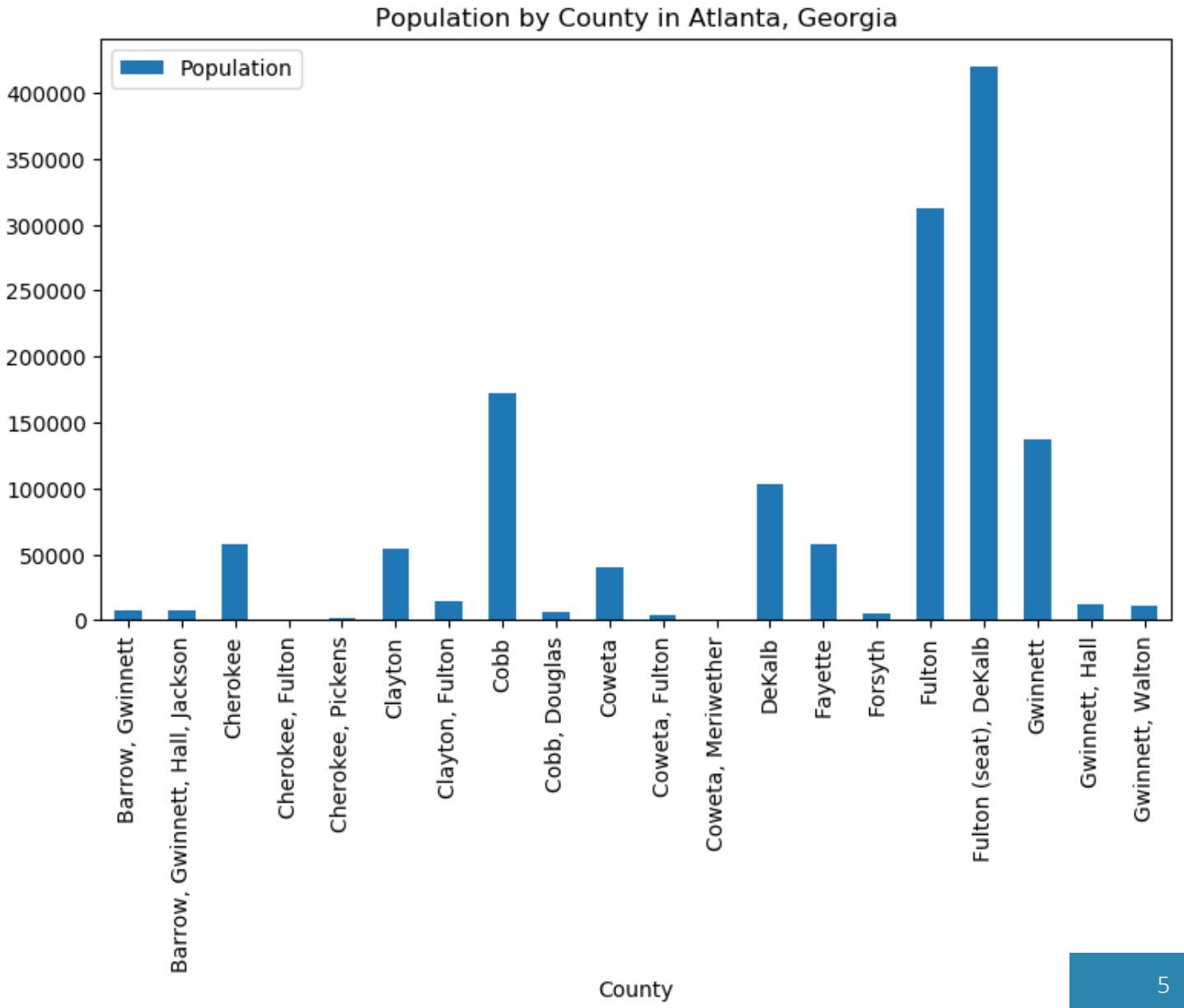
- *1.- Cities and Counties Data:*
- I found the list of Cities and Counties from Wikipedia : https://en.wikipedia.org/wiki/List_of_municipalities_in_Georgia_(U.S._state) . I
- I scrapped the data from the wiki, then cleaned and reduced to be applied creating choropleth map .
-
- *2.- Coordinates for Georgia Cities:*
- I created own table with Coordinates (Latitude and Longitude) using Google Maps. That info will be merged with data from Cities and Counties. [2]
- In this case , since the Geocoder is not allowing to retrieve information related coordinates , we wil create extract coordinates directly using Google Map.
-

- *3.- Foursquare API Data:*
- I used "**Foursquare API**" to extract data from the most common venues of each city in Atlanta. It will help us to provide the options in the comparative analysis. [3]
- In this case we will use the credentials already created to consume the venues.
-
- *4.-School, Colleges, University ranking and price of house rental Data*
- I created own tables for School, Colleges, University Ranking extracting from different websites It will help us to provide ranking analysis. [4]
- There are not too many public datas related to Education and rental prices of apartments. Therefore, I will create own tables to collect that information for our analysis.

Methodology

To compare the similarities of cities, cities were explored, segmented and grouped into clusters to find similarities between cities in Atlanta, Georgia. To be able to do that, we need to cluster data which is a form of unsupervised machine learning: k-means clustering algorithm.

- First at all , we start collecting cities in “Georgia” so results give us a view of how the population is currently in Atlanta.



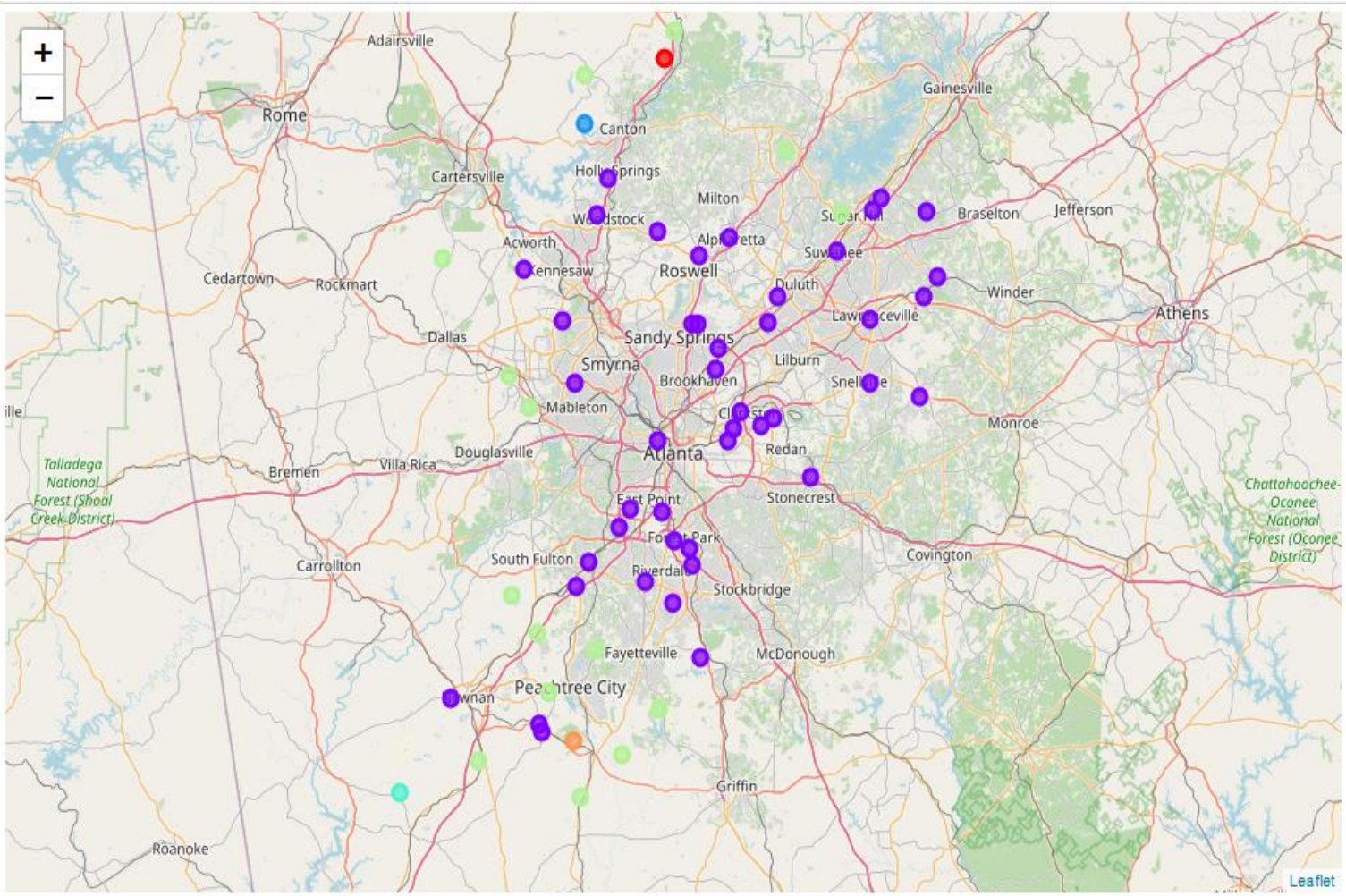
- Using coordinates from Atlanta, we can show using folium the map indicating the coordinates from each city in Atlanta and how they are placed.
- .- We define our Foursquare API credentials and we started exploring cities to obtain their corresponding venues and categorize them.

Display the TOP 10 Most Common venues near by City

```
num_top_venues = 10

indicators = ['st', 'nd', 'rd']

# create columns according to number of top venues
columns = ['City']
for ind in np.arange(num_top_venues):
    try:
        columns.append('{}{} Most Common Venue'.format(ind+1, indicators[ind]))
    except:
        columns.append('{}th Most Common Venue'.format(ind+1))

# create a new dataframe
city_venues_sorted = pd.DataFrame(columns=columns)
city_venues_sorted['City'] = georgia_grouped['City']

for ind in np.arange(georgia_grouped.shape[0]):
    city_venues_sorted.iloc[ind, 1:] = return_most_common_venues(georgia_grouped.iloc[ind, :], num_top_venues)

city_venues_sorted.head()
```

	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
0	Acworth	Park	Baseball Field	Garden	Discount Store	Farm	Farmers Market	Fast Food Restaurant	Fish & Chips Shop	Flea Market	Flower Shop
1	Alpharetta	Clothing Store	American Restaurant	New American Restaurant	Coffee Shop	Café	Fast Food Restaurant	Asian Restaurant	Mexican Restaurant	Sushi Restaurant	Deli / Bodega
2	Atlanta	Art Gallery	Coffee Shop	Gas Station	Pizza Place	Gym	American Restaurant	Asian Restaurant	Intersection	Trail	Restaurant
3	Auburn	Discount Store	Gym / Fitness Center	Convenience Store	Pharmacy	Breakfast Spot	Gas Station	Fast Food Restaurant	Train Station	Factory	Farm
4	Austell	Moving Target	Home Service	Food & Drink Shop	Dessert Shop	Business Service	Food Court	Food	Fondue Restaurant	Fabric Shop	Food Service

- Finally, we can apply Machine Learning – K-means that help us to make the clustering of the cities and we can se the visualization how the cities has been distributed.

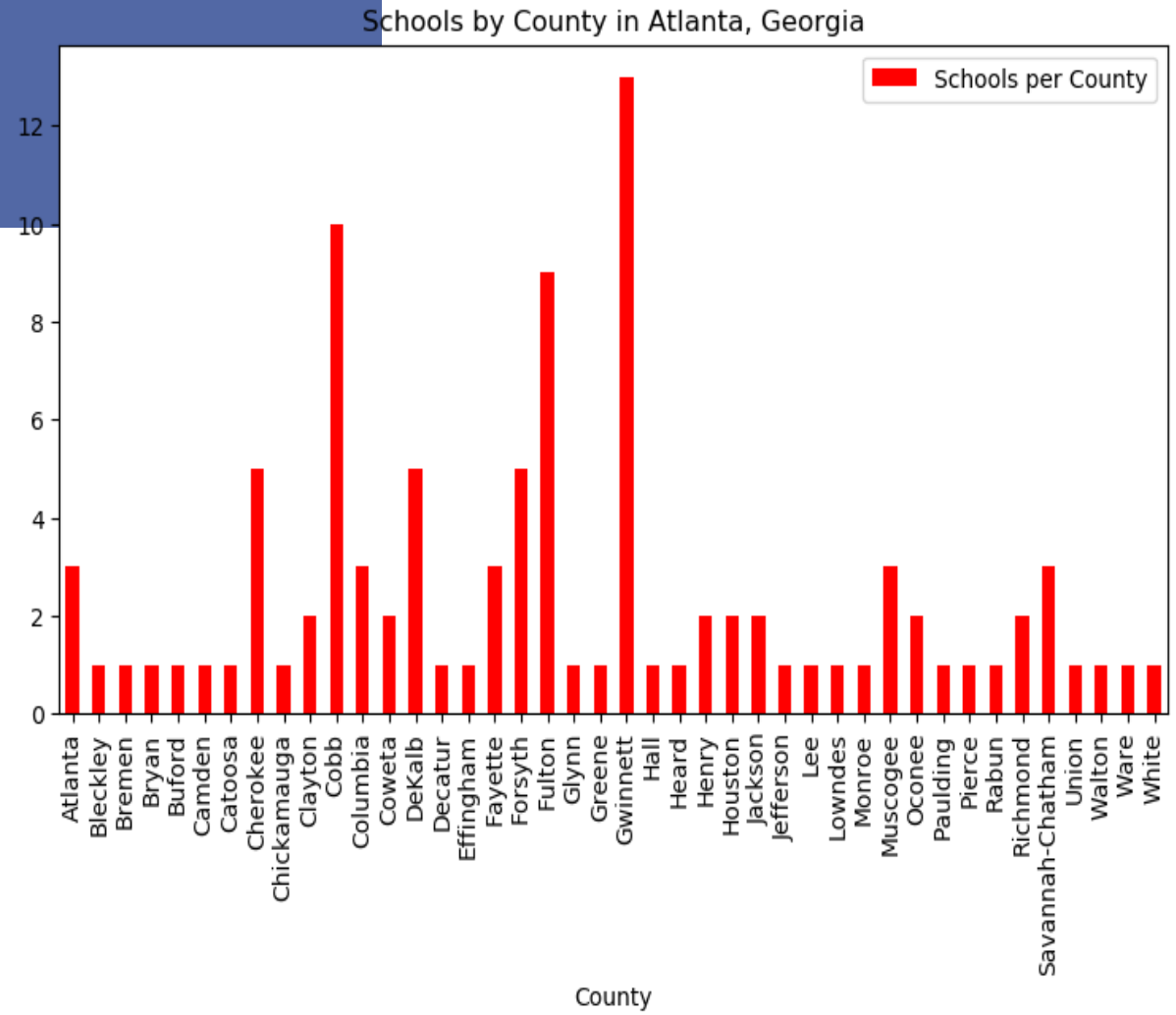


Results Section

After the analysis done, we can review the list of possibilities available in each cluster that can be considered when we want to select a place.

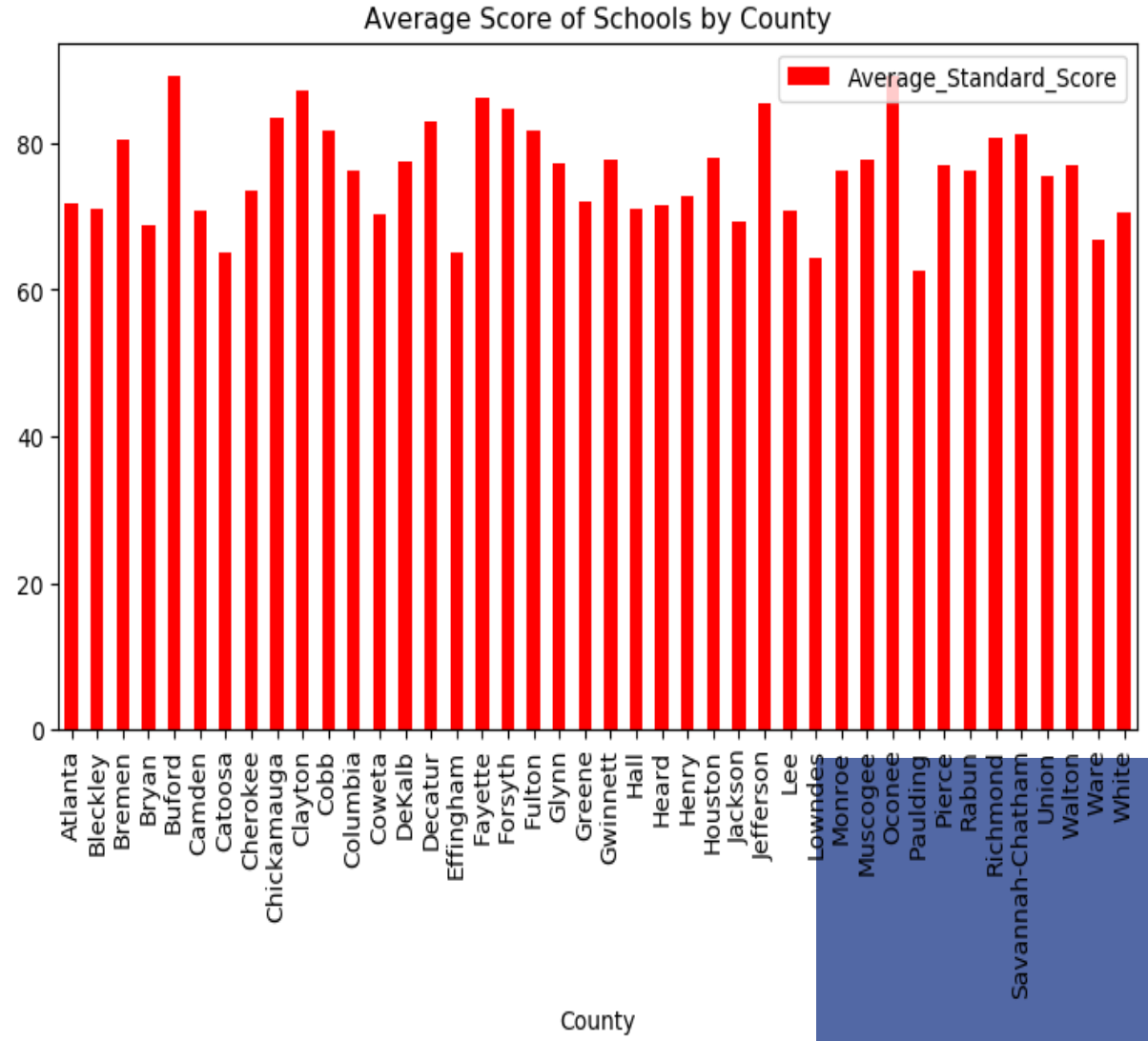
Definitively count with the collection of venues allow to select and consider which will be the most important per each individual person in the moment to decide what city choose to live.

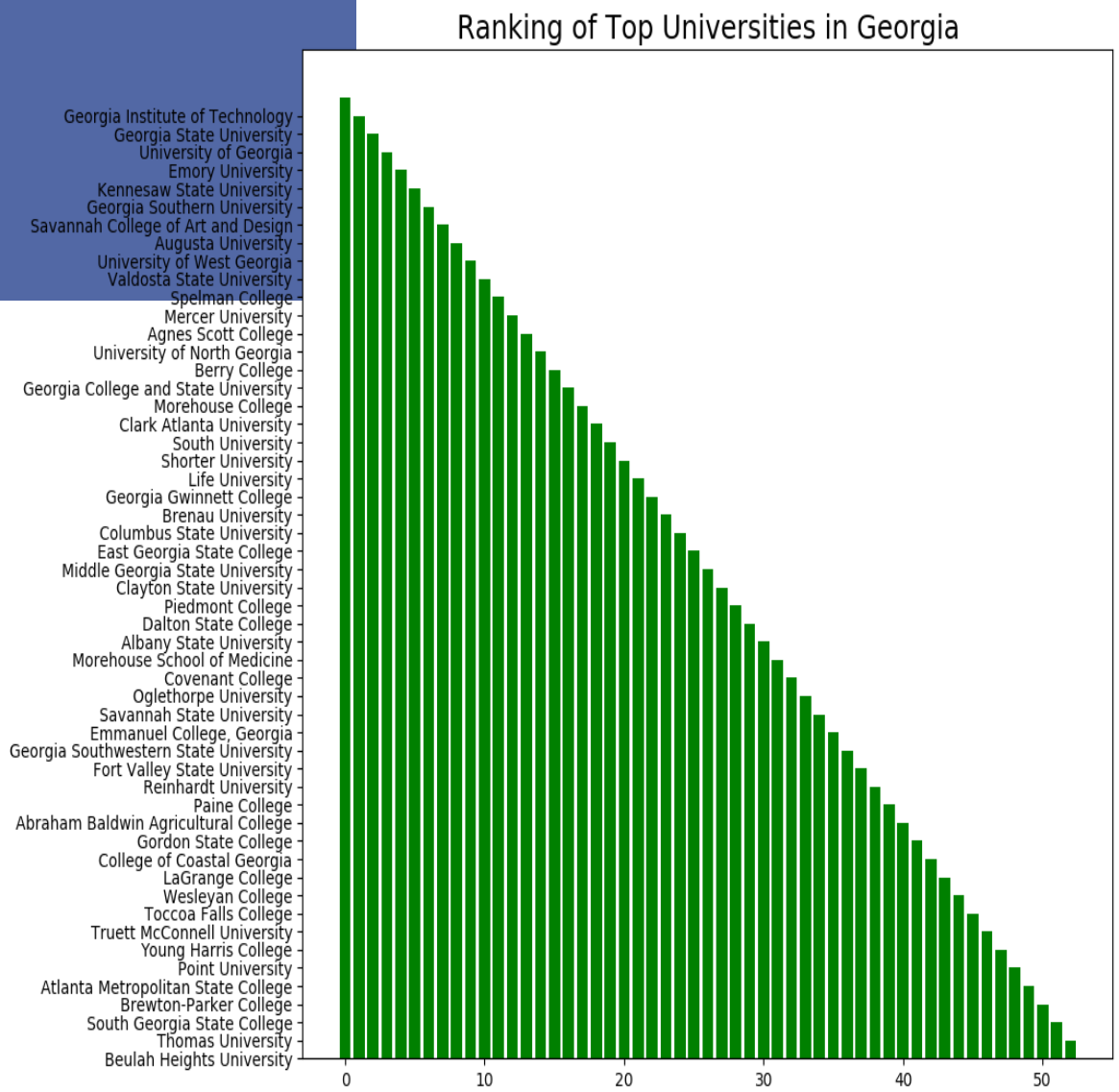
	index	City	County	Population	Land	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue
31	259	Kennesaw	Cobb	29783	9.44	34.026161	-84.687084	1.0	Sandwich Place	Mexican Restaurant	BBQ Joint	Pizza Place
38	297	Marietta	Cobb	56579	23.08	33.948371	-84.612413	1.0	Fast Food Restaurant	Breakfast Spot	Trail	Hardware Store
52	416	Sandy Springs	Fulton	93853	37.64	33.943358	-84.352816	1.0	Italian Restaurant	Mexican Restaurant	Pizza Place	Nail Salon



Schools by county in Atlanta, Georgia

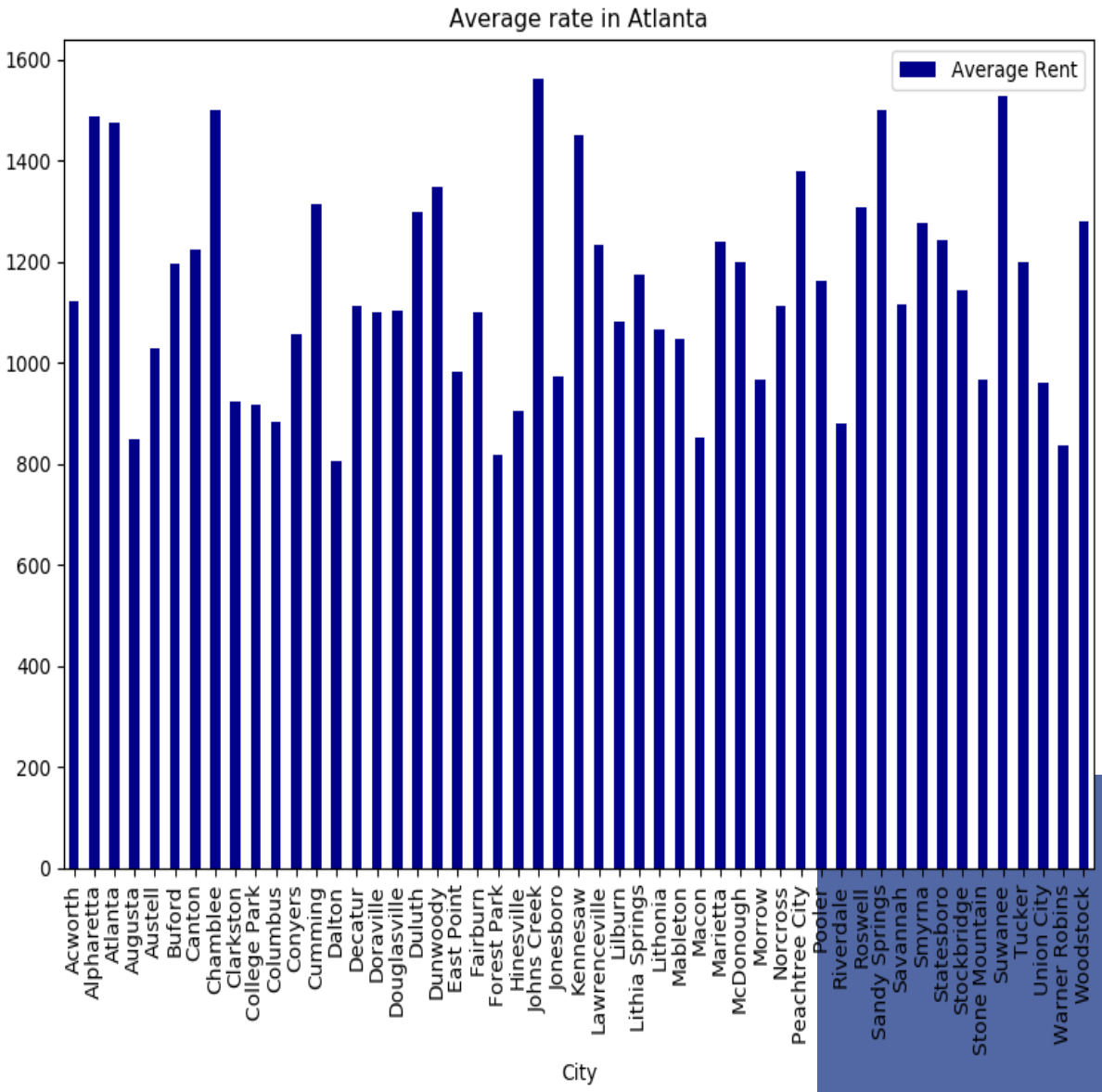
Average score of schools by county





Ranking of top universities in Georgia

Average rent rate in Atlanta



Discussion Section

As we can review results, clustering help us to split the bunch of options of venues obtained. And together with the information of schools, universities and house prices for rent give a better landscape of opportunities.

However, we know that our analysis can be enriched considering more variables like crime data, traffic in the city, etc. Definitively, it is something that I will work in a second wave.

Conclusion

Finally, we can conclude that the main purpose of this project is to develop and show the different opportunities that a person can consider when want to take a decision to choose a city where to live. In our case, we have used K-means cluster algorithm that helped us to group all the venues provided by Foursquare and what put in table as options of places that the user can consider at the moment to take the decision. Also, ranking of schools, universities and house prices show the best sceneries to consider.

Particularly, based on the list of venues, family preferences and the other variables like school, university and house prices our decision is most inclined by Cities as Marietta, Kennesaw and Sandy Springs.