Capstone Project: The Battle of Neighborhoods

MOVING TO ATLANTA, GEORGIA. ANALYSIS OF THE JOURNEY.

Business Proposal

Introduction:

Georgia is a state in the Southeastern Region of the United States. Georgia is the 24th-largest in area and 8th most populous of the 50 United States. Its 2019 estimated population was 10,617,423, according to the U.S. Census Bureau. Atlanta, a “global” city, is both the state's capital and its largest city.

The Atlanta metropolitan area, with an estimated population of more than 6 million people in 2019, is the 9th most populous metropolitan area in the United States and contains about 57% of Georgia's entire population and become the cultural and economic center of the Southeast. Georgia is divided in counties and in cities.[1]

Current facts related with “Covid-19” has affected all the Globe impacting mainly the economy of the countries. This situation has changed lifestyle of people and make them to look for better and new ways to live.

Many people is looking for new jobs and definitively it will accelerate the migration. Into this journey, some new needs are appearing but also new opportunities of development. There are needs of new locations to live, to work and to start a new business. Also, for sure, information that can help to take good decisions.

Atlanta has its fingers in global commerce, finance, research, and education to name a few; and now, the ATL is taking center stage for being a great place to live, work and why not to start a business.

The purpose of this Project will be centered in explorer all the facilities available around the list of cities in Atlanta, Georgia to provide an deep analysis and reports that allow people and myself to take the best decision when I will go to Atlanta.

Problem Definition:

As we describe above, 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:

The following data was considered in the analysis:

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

l 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

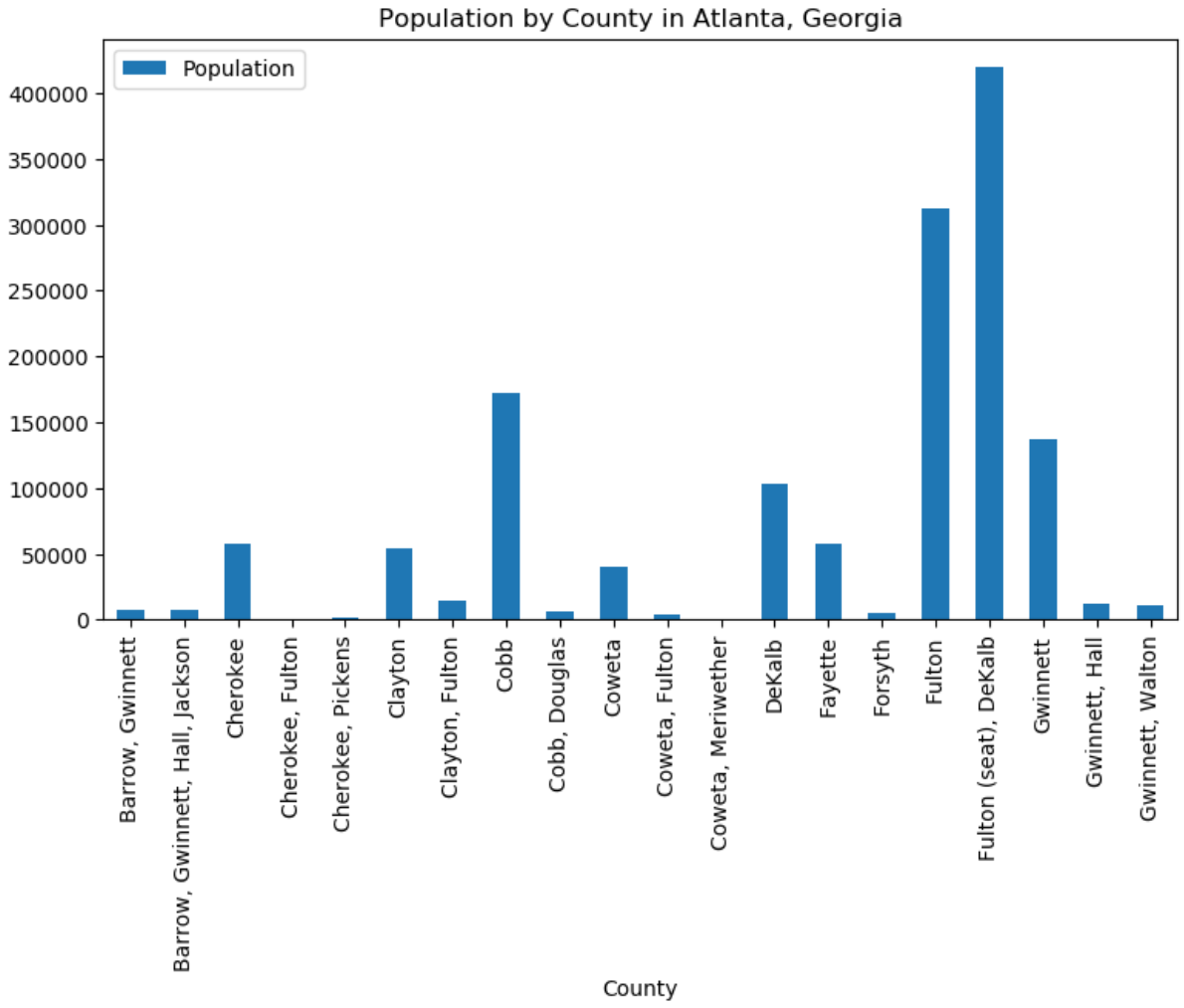
I created own tables for School, Colleges, University Ranking extracting from different websites It will help us to provide ranking analysis. [4]

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.

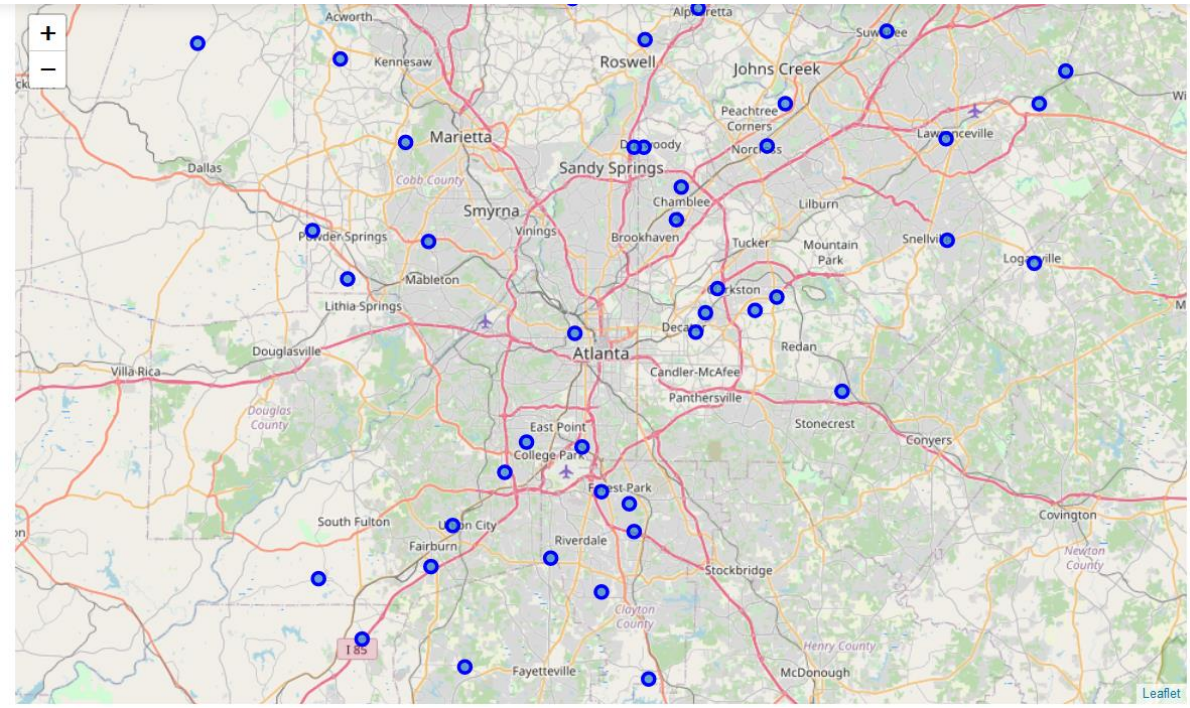
Methodology:

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.

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



2.- Using coordinates from Atlanta, we can show using folium the map indicating the coordinates from each city in Atlanta and how they are placed :



3.- We define our Foursquare API credentials and we started exploring cities to obtain their corresponding venues and categorize them.

```
print(georgia_venues.shape)
georgia_venues.head()
```

```
(2357, 7)
```

21]:

	City	City Latitude	City Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Acworth	34.041467	-84.848962	Burnt Hickory Park	34.034527	-84.840158	Park
1	Acworth	34.041467	-84.848962	Dollar General	34.048726	-84.832739	Discount Store
2	Acworth	34.041467	-84.848962	Burnt Hickory Ballfields	34.023908	-84.850913	Baseball Field
3	Acworth	34.041467	-84.848962	Rose Hall Wedding & Event Center	34.025564	-84.839308	Garden
4	Alpharetta	34.073757	-84.280203	Regal Avalon	34.071753	-84.276500	Movie Theater

```
georgia_venues.groupby('City').count()
```

4.- With the venues categorized we can display the TOP 10 most common venues in each city.

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

5.- 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 distribuited.

Cluster Cities

apply k-means to cluster the city into 5 clusters and add in a dataframe.

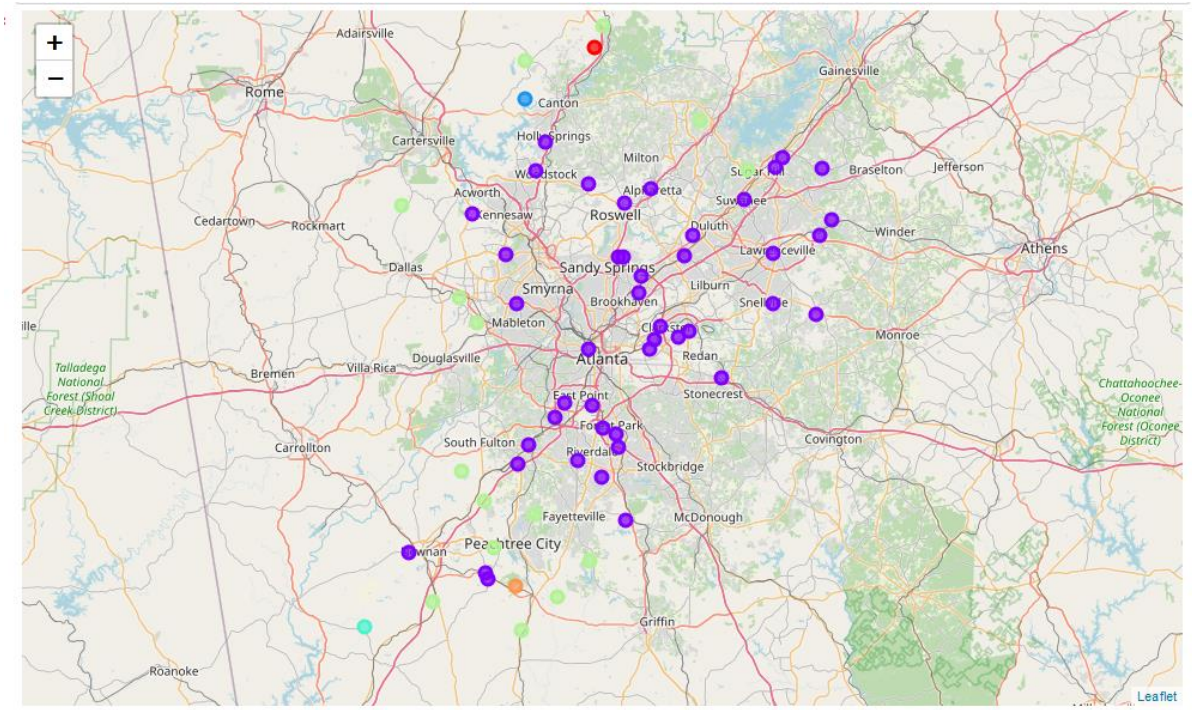
```
from sklearn.cluster import KMeans
import sklearn.cluster.k_means_
km = KMeans(n_clusters=3, init='k-means++', max_iter=100, n_init=1,
  verbose=True)
```

```
kclusters = 6
georgia_grouped_clustering = georgia_grouped.drop('City', 1)
kmeans = KMeans(n_clusters=kclusters, random_state=0).fit(georgia_grouped_clustering)
print(kmeans.labels_[0:10])
print(len(kmeans.labels_))
```

```
[4 1 1 1 4 1 0 1 1 4]
63
```

```
city_venues_sorted.insert(0, 'Cluster Labels', kmeans.labels_)
georgia_merged = df_atlanta
georgia_merged = georgia_merged.join(city_venues_sorted.set_index('City'), on='City')
georgia_merged.dropna(subset=["Cluster Labels"], axis=0, inplace=True)
georgia_merged.head(10) # check the last columns!
#print(georgia_merged.dtypes)
```

	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	5th Most Common Venue	6th Most Common Venue
0	1	Acworth	Cobb	20425	8.24	34.041467	-84.848962	4.0	Park	Baseball Field	Garden	Discount Store	Farm	Farmers Market
1	13	Alpharetta	Fulton	57551	26.91	34.073757	-84.280203	1.0	Clothing Store	American Restaurant	New American Restaurant	Coffee Shop	Café	Fast Food Restaurant
			Fulton										Pizza	American



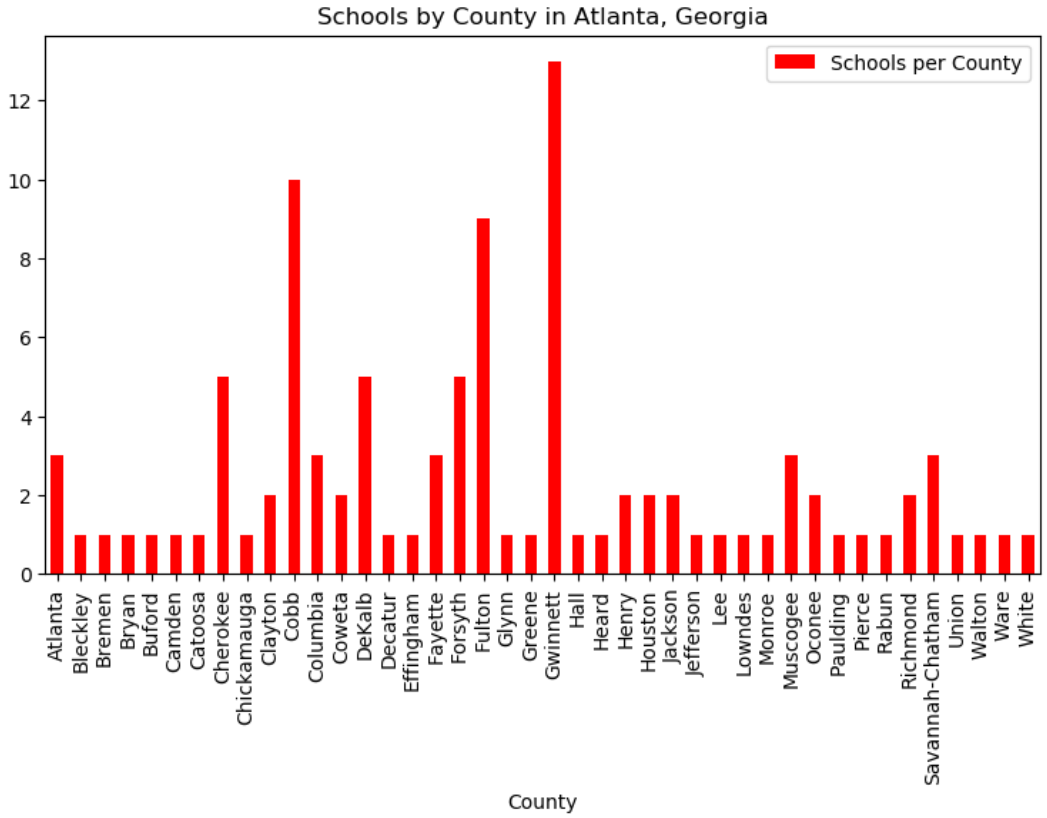
Results Section :

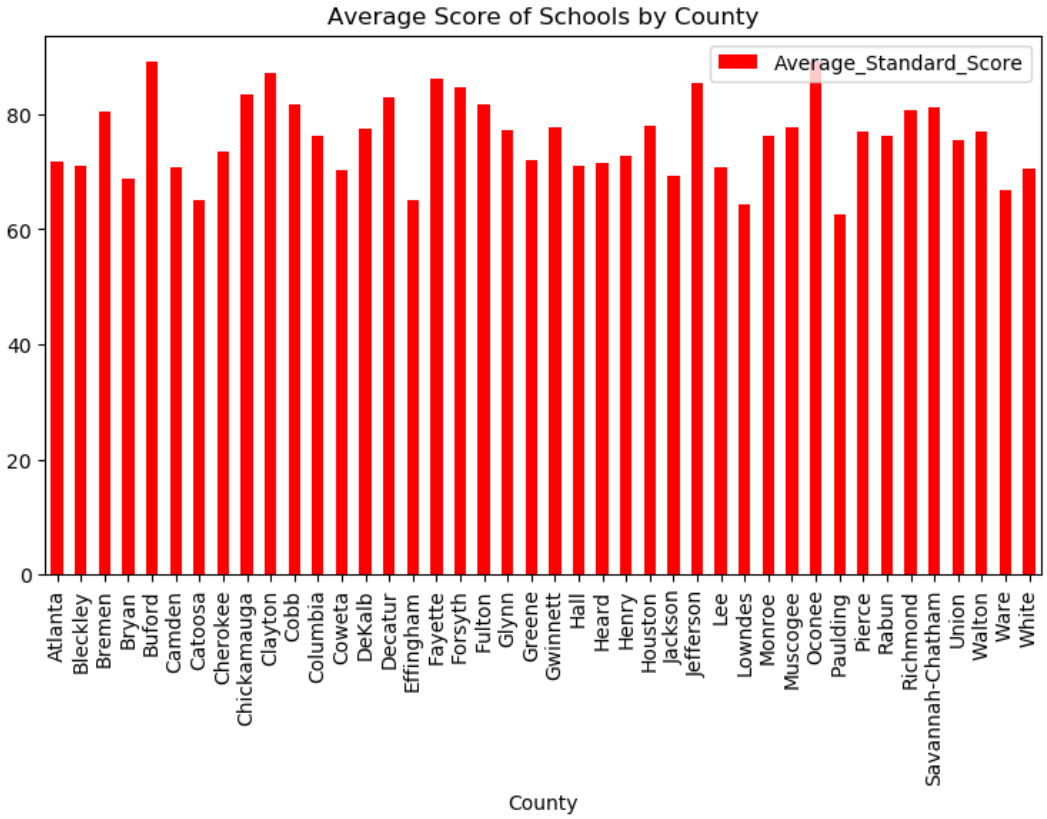
After the analysis done, we can review the list of possibilities available in each cluster that can be considered when we want to select a place.

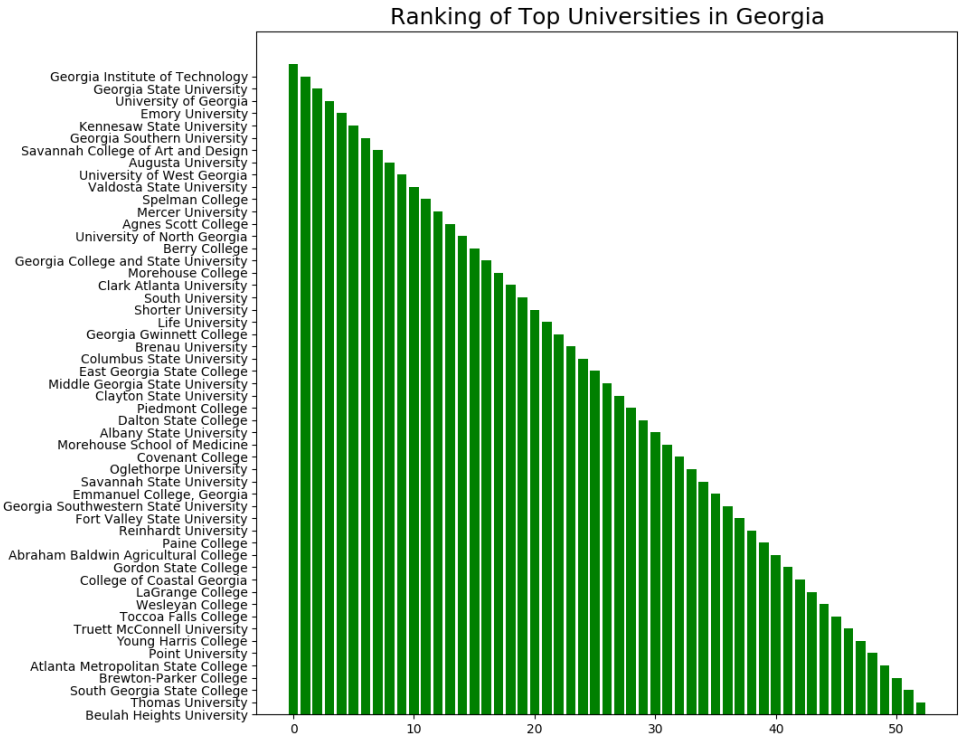
Definitively count with the collection of venues allow to select and consider which will be the most important per each individual person in the moment to decide what city choose to live.

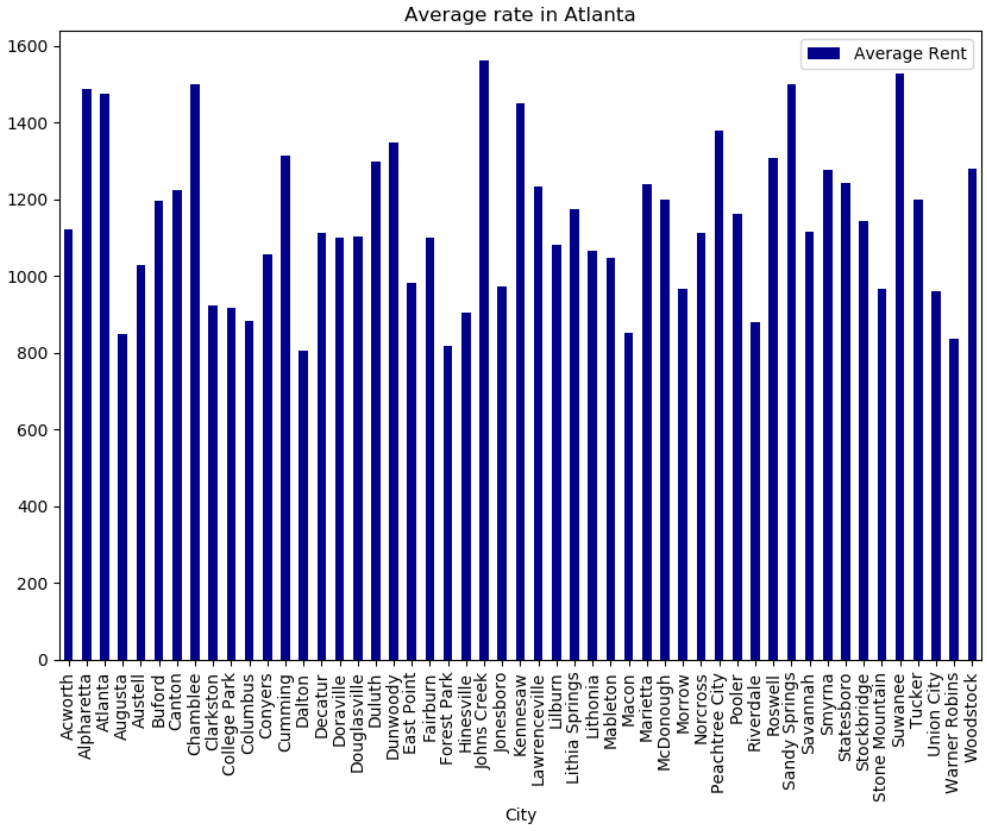
	index	City	County	Population	Land	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue
31	259	Kennesaw	Cobb	29783	9.44	34.026161	-84.687084	1.0	Sandwich Place	Mexican Restaurant	BBQ Joint	Pizza Place
38	297	Marietta	Cobb	56579	23.08	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

Also, the plots allow to have a view of the ranking of the schools and universities and house prices that will complement the variables to consider in the decision time.


Schools by County in Atlanta, Georgia
Schools per County
0
2
4
6
8
10
12
Atlanta
Bleckley
Bremen
Bryan
Buford
Camden
Catoosa
Cherokee
Chickamauga
Clayton
Cobb
Columbia
Coweta
DeKalb
Decatur
Effingham
Fayette
Forsyth
Fulton
Glynn
Greene
Gwinnett
Hall
Heard
Henry
Houston
Jackson
Jefferson
Lee
Lowndes
Monroe
Muscogee
Oconee
Paulding
Pierce
Rabun
Richmond
Savannah-Chatham
Union
Walton
Ware
White
County


Average Score of Schools by County
Average_Standard_Score
0
20
40
60
80
Atlanta
Bleckley
Bremen
Bryan
Buford
Camden
Catoosa
Cherokee
Chickamauga
Clayton
Cobb
Columbia
Coweta
DeKalb
Decatur
Effingham
Fayette
Forsyth
Fulton
Glynn
Greene
Gwinnett
Hall
Heard
Henry
Houston
Jackson
Jefferson
Lee
Lowndes
Monroe
Muscogee
Oconee
Paulding
Pierce
Rabun
Richmond
Savannah-Chatham
Union
Walton
Ware
White
County


Ranking of Top Universities in Georgia
Georgia Institute of Technology
Georgia State University
University of Georgia
Emory University
Kennesaw State University
Georgia Southern University
Savannah College of Art and Design
Augusta University
University of West Georgia
Valdosta State University
Spelman College
Mercer University
Agnes Scott College
University of North Georgia
Berry College
Georgia College and State University
Morehouse College
Clark Atlanta University
South University
Shorter University
Life University
Georgia Gwinnett College
Brenau University
Columbus State University
East Georgia State College
Middle Georgia State University
Clayton State University
Piedmont College
Dalton State College
Albany State University
Morehouse School of Medicine
Covenant College
Oglethorpe University
Savannah State University
Emmanuel College, Georgia
Georgia Southwestern State University
Fort Valley State University
Reinhardt University
Paine College
Abraham Baldwin Agricultural College
Gordon State College
College of Coastal Georgia
LaGrange College
Wesleyan College
Toccoa Falls College
Truett McConnell University
Young Harris College
Point University
Atlanta Metropolitan State College
Brewton-Parker College
South Georgia State College
Thomas University
Beulah Heights University
0
10
20
30
40
50


Average rate in Atlanta
Average Rent
1600
1400
1200
1000
800
600
400
200
0
Acworth
Alpharetta
Atlanta
Augusta
Austell
Buford
Canton
Chamblee
Clarkston
College Park
Columbus
Conyers
Cumming
Dalton
Decatur
Doraville
Douglasville
Duluth
Dunwoody
East Point
Fairburn
Forest Park
Hinesville
Johns Creek
Jonesboro
Kennesaw
Lawrenceville
Lilburn
Lithia Springs
Lithonia
Mableton
Macon
Marietta
McDonough
Morrow
Norcross
Peachtree City
Pooler
Riverdale
Roswell
Sandy Springs
Savannah
Smyrna
Statesboro
Stockbridge
Stone Mountain
Suwanee
Tucker
Union City
Warner Robins
Woodstock
City

Discusion 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.

References:

- [1] Georgia — Wikipedia : https://en.wikipedia.org/wiki/Georgia_(U.S._state)
- [2] Google Map : https://www.google.com/maps
- [3] Forsquare API : https://developer.foursquare.com/
- [4] Schools, University, Rental Houses websites :
    - https://www.schooldigger.com/go/GA/schoolrank.aspx?level=3#
    - https://www.rentcafe.com/blog/rental-market/local-rent-reports/georgia-rent-report-october-2019/
    - https://www.4icu.org/us/georgia/