Exploring the Atlanta Metropolitan area with a purpose of finding a suitable location to start a new health food business

Pavan Nagaraju

October 05, 2019

1. Introduction

1.1 Background

Metro Atlanta, designated by the United States Office of Management and Budget as the Atlanta–Sandy Springs–Roswell, GA Metropolitan Statistical Area, is the most populous metro area in the US state of Georgia and the ninth-largest metropolitan statistical area (MSA) in the United States. Its economic, cultural and demographic center is Atlanta, and has an estimated 2018 population of 5,949,951 according to the U.S. Census Bureau. Metro Atlanta is home to a culturally, economically and ethnically diverse population and hence a lot of start-up businesses thrive here catering to the needs of this phenomenal city that is demonstrating amazing growth over the past few years.

1.2 Problem

A group of individuals ventured out with a desire to establish a new restaurant that specializes in organic health food and beverages in Atlanta Metropolitan area. They understand that this city has potential for any new food business to grow but they are in search of a specific location that would maximize the sustainability and profitability of such a venture.

1.3 Interest

This group of entrepreneurs is very much interested in an informed analysis of the demographics, psychographics of different cities within the metro area with a suggested list of handful of cities that would be ideal for starting such a business.

2. Data Acquisition and Cleaning

2.1 Data Sources

Data is obtained from 3 different sources for this analysis.

Income data:
https://en.wikipedia.org/wiki/List_of_Georgia_(U.S._state)_locations_by_per_capita_income

This Wikipedia page lists the per capita income of different cities in Metro Atlanta area as shown below:

1. Berkeley Lake, Georgia – $69,439
2. Johns Creek, Georgia – $65,994
3. Dunwoody, Georgia – $62,523
4. Vinings, Georgia – $61,083
5. Chattahoochee Hills, Georgia – $52,835
6. Milton, Georgia – $49,628
7. Avondale Estates, Georgia – $42,605
8. Alpharetta, Georgia – $42,431
9. Woolsey, Georgia – $42,177
10. Roswell, Georgia – $40,106
11. St. Simons, Georgia – $37,256
12. Braselton, Georgia – $35,921
13. Atlanta, Georgia – $35,662
14. Druid Hills, Georgia – $34,829
15. Isle of Hope, Georgia – $34,067
16. North Decatur, Georgia – $33,739
17. North Druid Hills, Georgia – $33,288

Crime Rate data:

https://www.alarms.org/safest-cities-in-georgia/

This page lists the top safest cities in the stage of Georgia along with crime statistics in the below format:

Safest Cities in Georgia, 2019

Show 100 entries Search:

Rank	City	Violent Crimes	Property Crimes	Violent Crime Rate	Property Crime Rate
1	Holly Springs	12	144	1.030	12.357
2	Braselton	12	94	1.115	8.735
3	Milton	9	298	0.228	7.566
4	Johns Creek	25	575	0.294	6.761
5	Kennesaw	72	480	2.108	14.054
6	Duluth	33	557	1.108	18.694
7	Peachtree City	14	591	0.397	16.742
8	Roswell	108	1,616	1.130	16.903
9	Grovetown	9	261	0.650	18.857
10	Bainbridge	47	379	3.850	31.045
11	Fayetteville	34	468	1.916	26.368
12	Acworth	40	579	1.758	25.454
13	Loganville	14	321	1.178	27.004
14	Lawrenceville	73	947	2.341	30.369
15	Dallas	18	275	1.346	20.564
16	Norcross	80	522	4.689	30.598
17	Woodstock	35	601	1.084	18.611
18	Dalton	108	981	3.155	28.661

Locational data:

Foursquare.com

Using developer account with Foursquare.com I used their APIs to retrieve locational information related to the cities in Metro Atlanta area.

2.2 Data Cleaning

Data is scraped from the above sources into dataframes. The per capita income data is provided in a specific format, so string manipulation had to be used to split the data elements and separate them into meaningful columns in the dataframe by removing unwanted characters/strings. The crime rate data was relatively clean and available as html table. However, the problem with this dataset is that it contains cities outside the metro Atlanta area as well. So, per capita dataset was used as a reference to remove unwanted data records and keep only relevant city records.

Foursquare API returns data in JSON format so JSON utility packages were used to load it into dataframes accordingly.

3. Exploratory Data Analysis

First, crime data is sorted with lowest crime record first and then picked top 20 safest places from this data set.

	Rank	City	Violent Crimes	Property Crimes	Violent Crime Rate	Property Crime Rate
1	1	Holly Springs	12	144	1.030	12.357
2	2	Braselton	12	94	1.115	8.735
3	3	Milton	9	298	0.228	7.566
4	4	Johns Creek	25	575	0.294	6.761
5	5	Kennesaw	72	480	2.108	14.054
6	6	Duluth	33	557	1.108	18.694
7	7	Peachtree City	14	591	0.397	16.742
8	8	Roswell	108	1,616	1.130	16.903
9	9	Grovetown	9	261	0.650	18.857
10	10	Bainbridge	47	379	3.850	31.045
11	11	Fayetteville	34	468	1.916	26.368
12	12	Acworth	40	579	1.758	25.454
13	13	Loganville	14	321	1.178	27.004
14	14	Lawrenceville	73	947	2.341	30.369
15	15	Dallas	18	275	1.346	20.564
16	16	Norcross	80	522	4.689	30.598
17	17	Woodstock	35	601	1.084	18.611
18	18	Dalton	108	981	3.155	28.661
19	19	Suwanee	32	520	1.585	25.760
20	20	Decatur	40	531	1.711	22.714

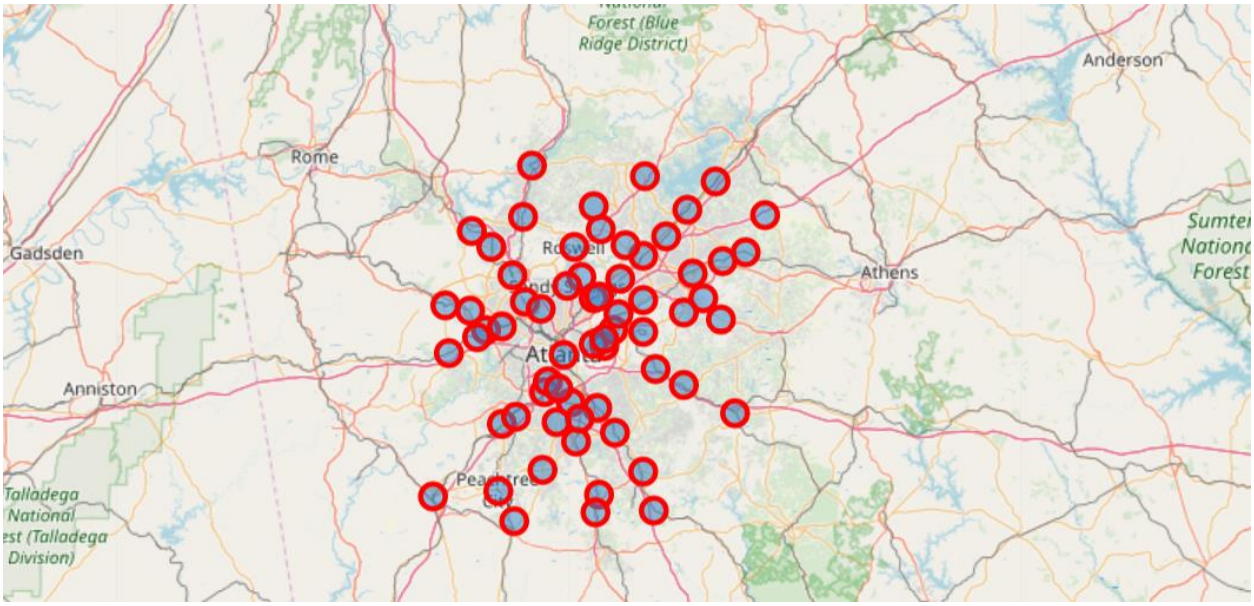
Simarly, sorted the per capita income data in descending order and picked top 20 records.

	City	Per Capita Income
0	Berkeley Lake	69439
1	Johns Creek	65994
2	Dunwoody	62523
3	Vinings	61083
4	Chattahoochee Hills	52835
5	Milton	49628
6	Avondale Estates	42605
7	Alpharetta	42431
8	Woolsey	42177
9	Roswell	40106
10	St. Simons	37256
11	Braselton	35921
12	Atlanta	35662
13	Druid Hills	34829
14	Isle of Hope	34067
15	North Decatur	33739
16	North Druid Hills	33288
17	Tybee Island	32406
18	North Atlanta	32087
19	Peachtree City	31667

Then merged these two datasets to present the cities that are high in income as well as low crime rate.

	City	Per Capita Income	Rank	Violent Crimes	Property Crimes	Violent Crime Rate	Property Crime Rate
0	Johns Creek	65994	4	25	575	0.294	6.761
1	Dunwoody	62523	64	63	2,003	1.277	40.612
2	Milton	49628	3	9	298	0.228	7.566
3	Roswell	40106	8	108	1,616	1.130	16.903
4	Braselton	35921	2	12	94	1.115	8.735
5	Atlanta	35662	37	4,504	22,991	9.357	47.764
6	Peachtree City	31667	7	14	591	0.397	16.742
7	Suwanee	29712	19	32	520	1.585	25.760
8	Decatur	29363	20	40	531	1.711	22.714
9	Duluth	29185	6	33	557	1.108	18.694
10	Smyrna	27637	30	172	1,544	2.987	26.817
11	Chamblee	27481	33	135	951	4.727	33.301
12	Kennesaw	27165	5	72	480	2.108	14.054
13	Fayetteville	26551	11	34	468	1.916	26.368
14	Snellville	26131	32	39	697	1.952	34.881
15	Woodstock	25586	17	35	601	1.084	18.611
16	Marietta	23409	35	215	2,209	3.488	35.834
17	Holly Springs	22992	1	12	144	1.030	12.357
18	Lilburn	22503	36	36	479	2.803	37.294
19	Douglasville	22283	57	244	2,099	7.253	62.396
20	Rincon	22023	45	12	351	1.188	34.749

Then moved on to locational data. Using Geocoders package, obtained Latitude and Longitude values for all the cities in the list. Then using Folium package, presented all the cities on the map



Using the Foursquare API, obtained the top 100 venues for each of the city and got the resonse in JSON format. For each city, the venue, its category looked something like this:

	City	City Latitude	City Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Acworth	34.065933	-84.67688	Henry's Louisiana Grill	34.066011	-84.677728	Cajun / Creole Restaurant
1	Acworth	34.065933	-84.67688	Fusco's via Roma	34.065781	-84.677163	Italian Restaurant
2	Acworth	34.065933	-84.67688	Miss L's Sandwich Shop	34.065861	-84.677276	Sandwich Place
3	Acworth	34.065933	-84.67688	Lacey's Drugs	34.065340	-84.676674	Pharmacy
4	Acworth	34.065933	-84.67688	Oak Barrel	34.066145	-84.678016	Wine Shop

When number of venues for each city is obtained to find out the popularity of the city:

City	City Latitude	City Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Acworth	14	14	14	14	14	14
Alpharetta	46	46	46	46	46	46
Atlanta	10	10	10	10	10	10
Auburn	5	5	5	5	5	5
Austell	7	7	7	7	7	7
Avondale Estates	5	5	5	5	5	5
Braselton	8	8	8	8	8	8
Buford	10	10	10	10	10	10
Canton	12	12	12	12	12	12
Chamblee	20	20	20	20	20	20
Clarkston	7	7	7	7	7	7
College Park	17	17	17	17	17	17
Conyers	11	11	11	11	11	11
Covington	12	12	12	12	12	12
Cumming	13	13	13	13	13	13
Dacula	6	6	6	6	6	6
Decatur	57	57	57	57	57	57
Doraville	31	31	31	31	31	31

One observation made is that the cities that are high in income and low crime rate have more number of venues listed by Foursquare API which means these cities are more thriving than others.

One hot encoding was used to extrapolate this data

	City	Accessories Store	American Restaurant	Antique Shop	Arcade	Arepa Restaurant	Art Gallery	Arts & Crafts Store	Asian Restaurant	Athletics & Sports	Automotive Shop	BBQ Joint	Bagel Shop	Bakery	Bank	Bar	Basel Fi
0	Acworth	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
1	Acworth	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
2	Acworth	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
3	Acworth	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
4	Acworth	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	

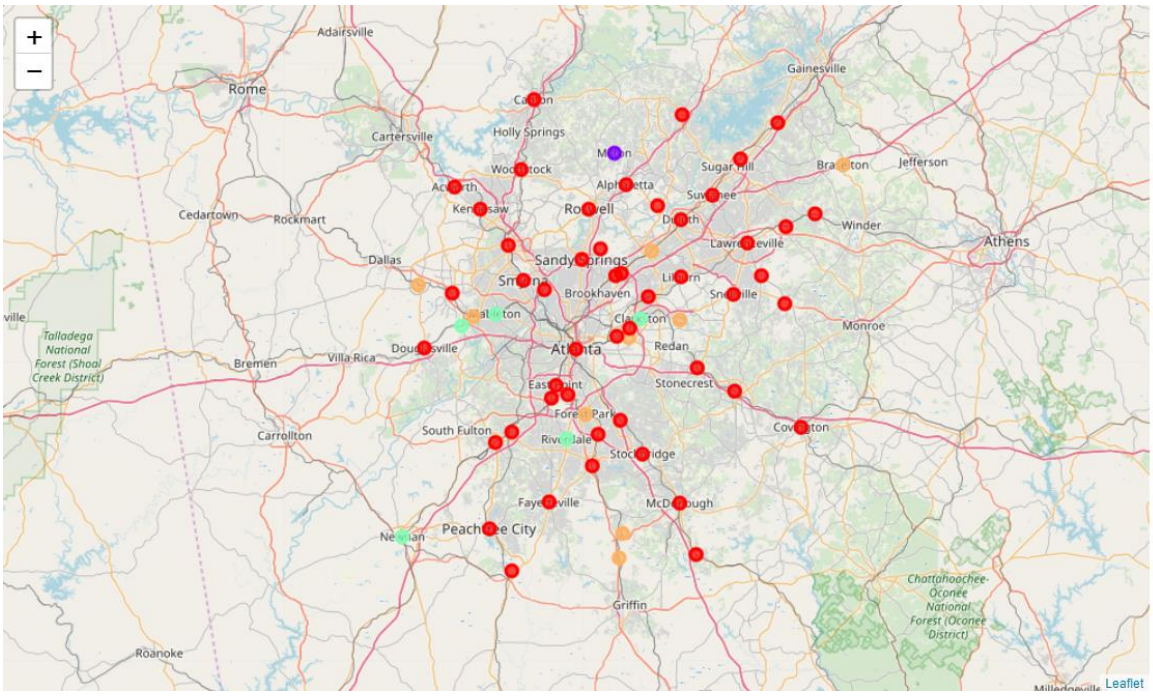
Then frequency of occurring with in each category is determined:

	City	Accessories Store	American Restaurant	Antique Shop	Arcade	Arepa Restaurant	Art Gallery	Arts & Crafts Store	Asian Restaurant	Athletics & Sports	Automotive Shop	BBQ Joint	Bagel Shop	Ba
0	Acworth	0.000000	0.142857	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.07
1	Alpharetta	0.000000	0.043478	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.021739	0.00
2	Atlanta	0.000000	0.100000	0.000000	0.000000	0.000000	0.100000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00
3	Auburn	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.200000	0.000000	0.00
4	Austell	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00
5	Avondale Estates	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00
6	Braselton	0.000000	0.125000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00
7	Buford	0.000000	0.100000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00

K-means algorithm is used for grouping the cities into 5 different clusters using the above data.

	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 Mos Commor Venue
0	Acworth	34.065933	-84.676880	0	American Restaurant	Sporting Goods Shop	Bakery	Pharmacy	Wine Shop	Food Court	Bike Shop	New American Restaurant	Italian Restaurant	Health & Beauty Service
1	Alpharetta	34.070958	-84.274733	0	Clothing Store	Furniture / Home Store	New American Restaurant	Kids Store	Women's Store	Sushi Restaurant	American Restaurant	Plaza	Coffee Shop	Pizza Place
2	Atlanta	33.749099	-84.390185	0	Event Space	Rental Car Location	American Restaurant	Caribbean Restaurant	Intersection	Music Venue	Art Gallery	Breakfast Spot	Food Court	College Basketbal Cour
3	Auburn	34.013662	-83.827350	0	Convenience Store	BBQ Joint	Baseball Field	City Hall	Chinese Restaurant	Dessert Shop	Event Space	French Restaurant	Football Stadium	Food Truck
4	Austell	33.812606	-84.634378	4	Mexican Restaurant	Gift Shop	Pharmacy	Pet Store	Park	Diner	Yoga Studio	Food Truck	Food Court	Food & Drink Shop

Using Folium again, these clusters are presented on the map



Then each cluster is closed examined to determine its characteristics:

	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	American Restaurant	Sporting Goods Shop	Bakery	Pharmacy	Wine Shop	Food Court	Bike Shop	New American Restaurant	Italian Restaurant	Health & Beauty Service
1	Alpharetta	Clothing Store	Furniture / Home Store	New American Restaurant	Kids Store	Women's Store	Sushi Restaurant	American Restaurant	Plaza	Coffee Shop	Pizza Place
2	Atlanta	Event Space	Rental Car Location	American Restaurant	Caribbean Restaurant	Intersection	Music Venue	Art Gallery	Breakfast Spot	Food Court	College Basketball Court
3	Auburn	Convenience Store	BBQ Joint	Baseball Field	City Hall	Chinese Restaurant	Dessert Shop	Event Space	French Restaurant	Football Stadium	Food Truck
7	Buford	Insurance Office	Bistro	New American Restaurant	Bar	Tapas Restaurant	Discount Store	Steakhouse	Italian Restaurant	Hookah Bar	American Restaurant
8	Canton	American Restaurant	Playground	Jewelry Store	Soccer Field	Bookstore	Gastropub	Seafood Restaurant	Sandwich Place	Diner	Southern / Soul Food Restaurant

4. Results

From the data analysis performed, the following cities were found to be suitable for starting an organic health food restaurant business:

Johns Creek

Alpharetta

Sandy Springs

Dunwoody

Milton

5. Discussion

The above cities were belonging to clusters 1 &2 which represent locations with people having a modern lifestyle which indicates that they tend to opt for healthy food more often. Also, these cities are in the top few cities with high per capita income and low crime rate, thus being ideal locations for starting such a business.

6. Conclusion

While Metro Atlanta in general is business-friendly, the above 5 cities seem to have right mix of demographics, income and life-style of people that are the target customer base for a health food business.