Atlanta Metropolitan Area Venues and Housing Prices Data Analysis

Joan Santus Sánchez Pérez-Cortés
August 22, 2020

1.Introduction

1.1 Background and problem:

My wife and I moved to the Atlanta Metropolitan Area over 2 years ago. When we arrived we had no idea of what activities to do or where to go on a casual day and to be honest we are still not close to knowing all that there is to do here yet. I decided to use this project to look at popular venues around the Atlanta Metropolitan Area to take a look at house prices and which venues are popular to help us, people visiting us and hopefully other people moving here where to move and go visit new places around the area as well as possible business opportunities.

The Atlanta Metropolitan Area is designated by the United States Office of Management and Budget as the Atlanta-Sandy Springs- Alpharetta, GA Metropolitan Statistical Area. It is the most populous metro area in the US state of Georgia and the ninth-largest in the United states. Its economic, cultural and demographic center is Atlanta, and it has an estimated 2019 population of 6,020,364 according to the U.S. Census Bureau. [1]

Metro Atlanta is a huge area made up of 29 counties, each with its own cities. For this project I decided to pick the top cities and compare venues and house prices. Since the area is very big I'm considering that my target audience for this analysis has a vehicle to move around the area.

2. Data

2.1 Data acquisition

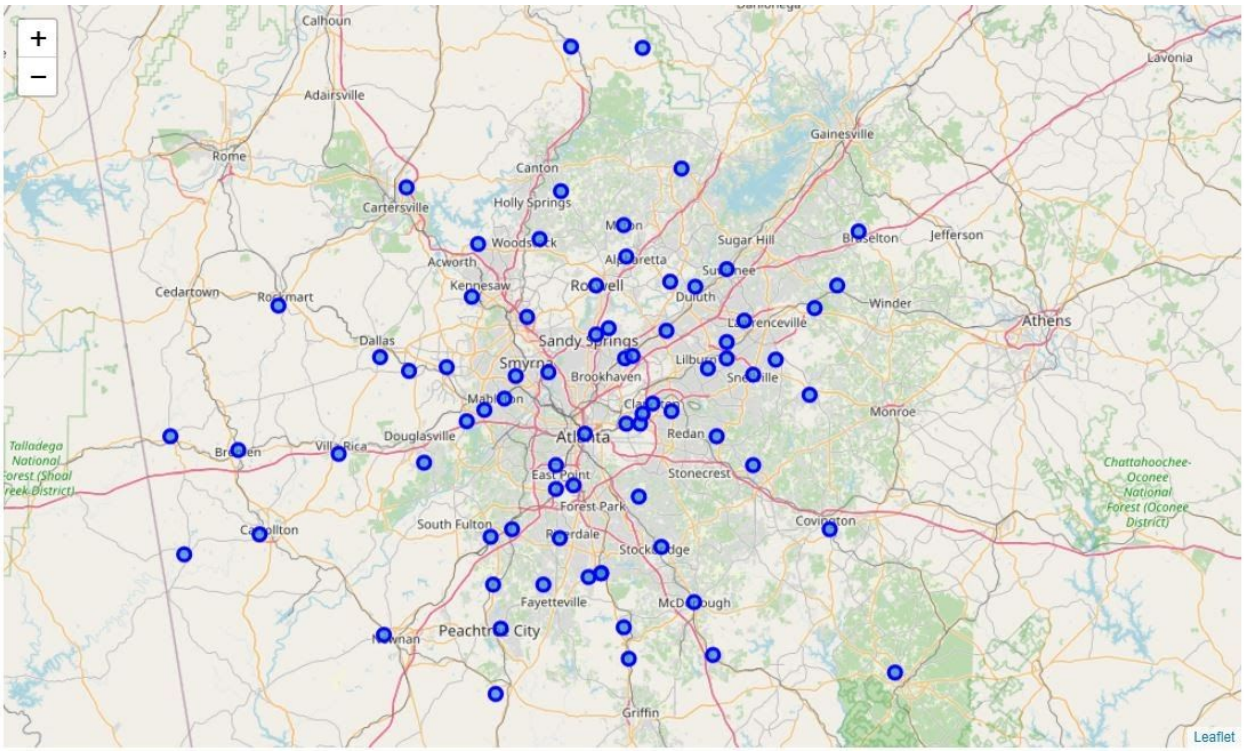
The data I ended up using for this project comes from 4 different places. While doing my research on Metro Atlanta I found out that the whole area is divided in complicated ways. Some counties have cities and neighborhoods and others don't. Some have big populated areas that aren't really cities and others have small populated areas considered cities. I decided to do some research and use this website's criteria to pull a list of cities and use it as a starting point.

I also got a list of coordinates for all of Georgia's cities in this database and lastly I got a list of house prices from Zillow. I'm also going to be using foursquare to get my data for the venues in each city.

2.2 Data Cleaning

To create my main dataframe I used my list of cities and started merging it with my other data sets and started filling out information that's missing or taking out cities that I can't get their information.

My first dataset is the list of cities with the coordinate dataset, where I'm missing the coordinates of 15 cities. I used simple google searches to find the missing coordinates and add them to its corresponding city as well as deleting the ZIP code since I won't be using it. After making this dataset I tested it in a Geopy map and fixed the coordinates of cities I could see weren't correct.



I then combined this dataset with my housing prices dataset, using just the latest month data which is March 2020, and then taking out the cities that were missing housing prices. This dataset ended up looking something like this:

	City	Latitude	Longitude	Mar 2020 Price
0	Acworth	34.097998	-84.61830	258900
1	Alpharetta	34.080035	-84.21929	426700
2	Atlanta	33.844371	-84.47405	299300
3	Auburn	34.022974	-83.83416	193700
4	Austell	33.797456	-84.60621	178800

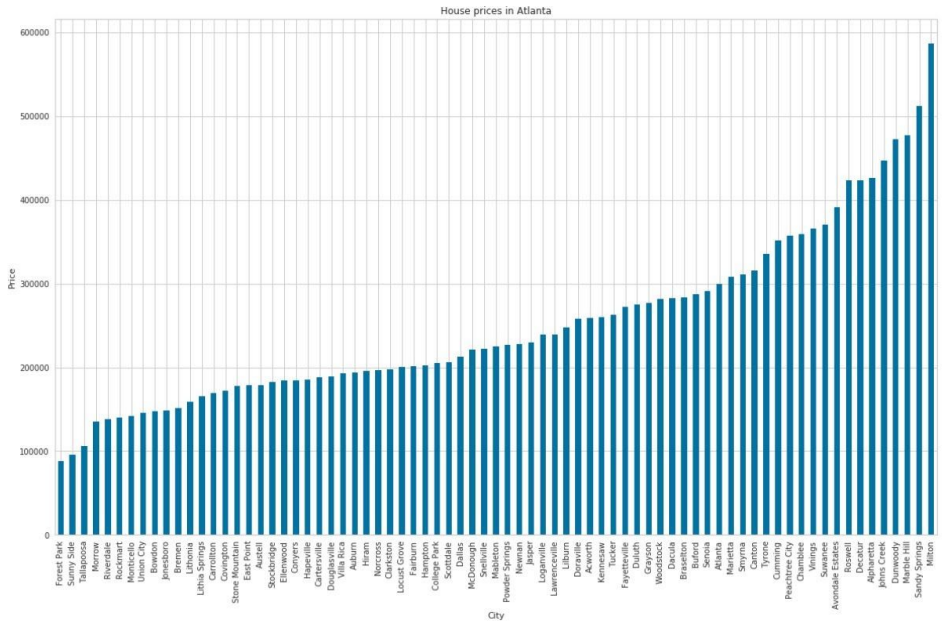
Lastly I pulled my list of venues using Foursquare using a limit of 100 venues as a limit with a range of 2000 meters. This is how my dataset looks after this and now I'm ready to start analyzing everything.

	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	34.065704	-84.677274	Sandwich Place
3	Acworth	34.065933	-84.67688	Lake Acworth Beach	34.061058	-84.682936	Beach
4	Acworth	34.065933	-84.67688	Oak Barrel	34.066145	-84.678016	Wine Shop

3. Methodology

3.1 Housing Prices

The first analysis I did is for the housing prices. I first used a simple bar plot using matplotlib to visualize the prices of all the cities to give me a clear idea on how to categorize them.

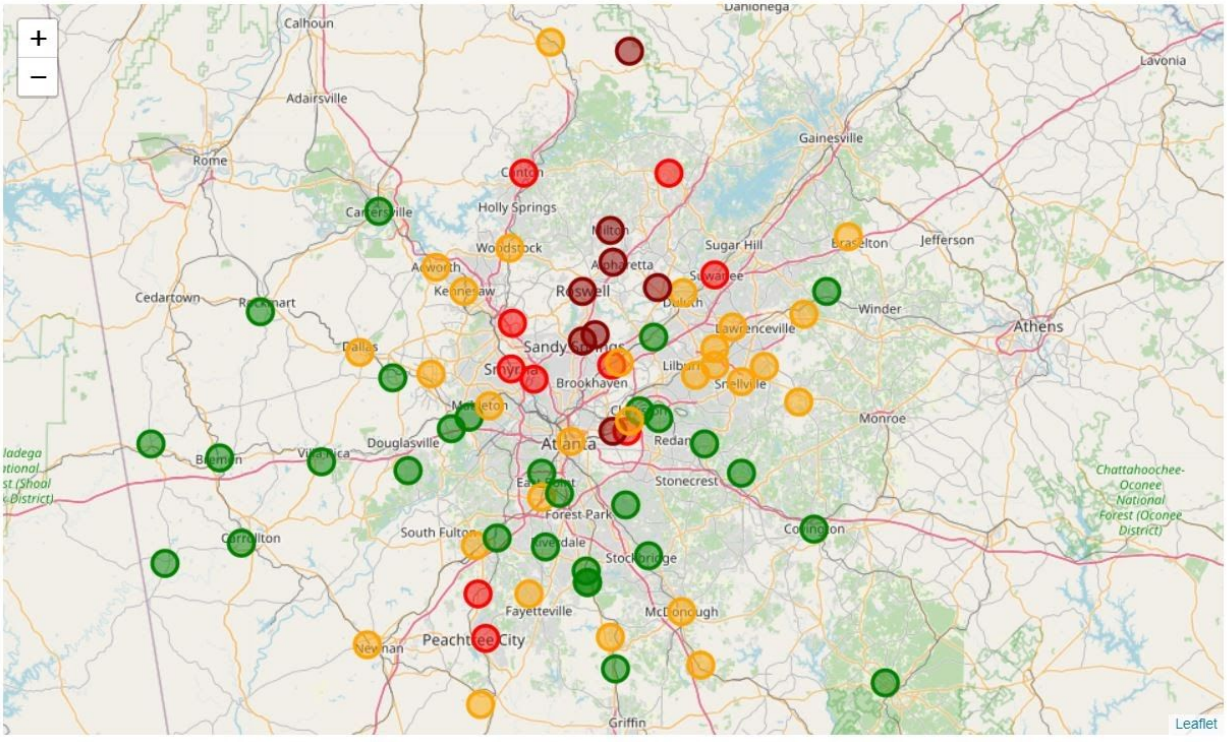


As we can see we can split the housing prices into 4 categories according to their price range.

- Category 1: Equal or lower than 200,000.00$
- Category 2: Equal or lower than 300,000.00$ and higher than 200,000.00$
- Category 3: Equal or lower than 400,000.00$ and higher than 300,000.00$
- Category 4: Higher than 400,000.00$

Price Category	City
1	29
2	28
3	10
4	8

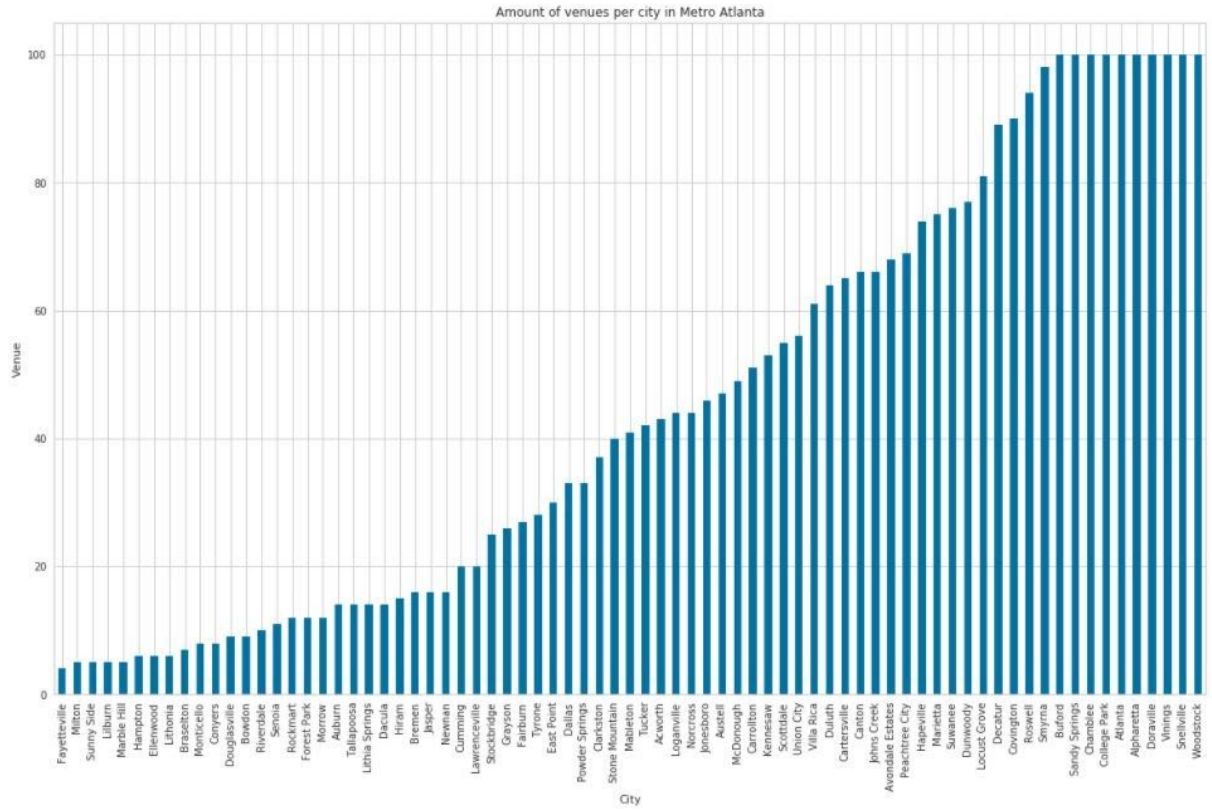
Once I had these categories, I added a new column to my dataframe assigning each city its price category and then I map all the cities again and give them colors to represent their category, from green for category 1 up to dark red for category 4.



This map shows us where the most expensive areas to live in Metro Atlanta are as well as the cheapest. At a glance we see that the north-eastern part of Metro Atlanta would be the wealthiest area too.

3.2 Venues Analysis

The amount of venues I got from my Foursquare data are 3362 with 313 unique categories in the latest notebook run. I compared the cities with the amount of venues each has in a bar plot to give me a better idea of how the distribution is.

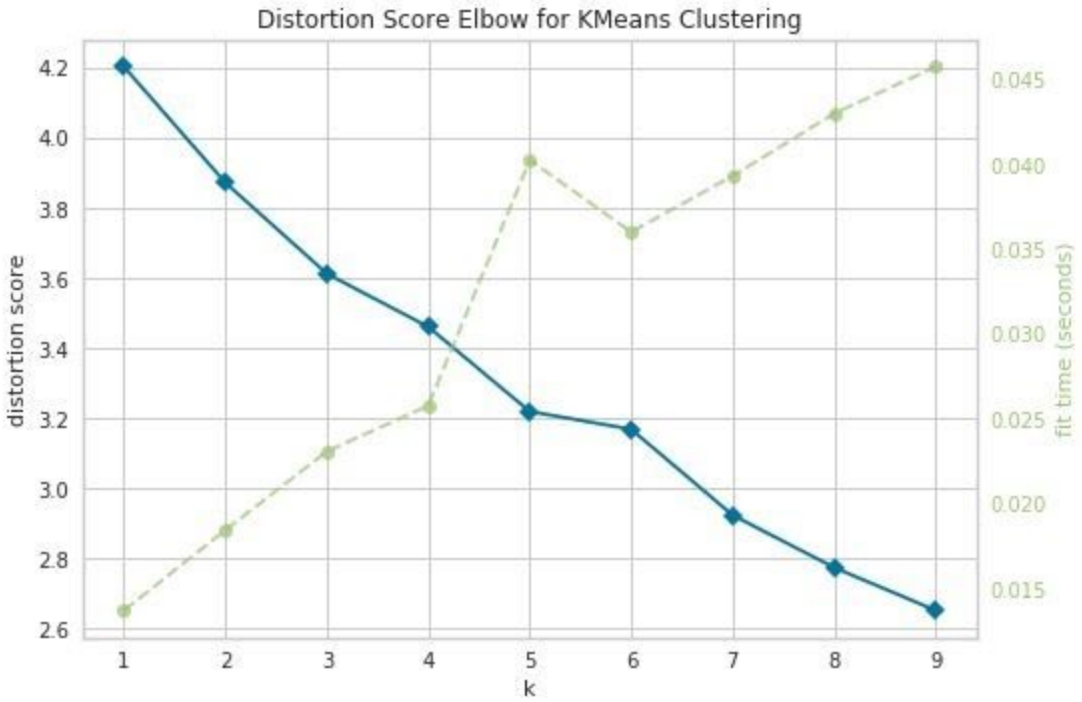


There are 10 cities that reach the 100 venues limit and the cities have a wide range of venues. I took a closer look at my data and created a new dataframe with the top venues for each city to give me a better idea on what they are. I looks like this:

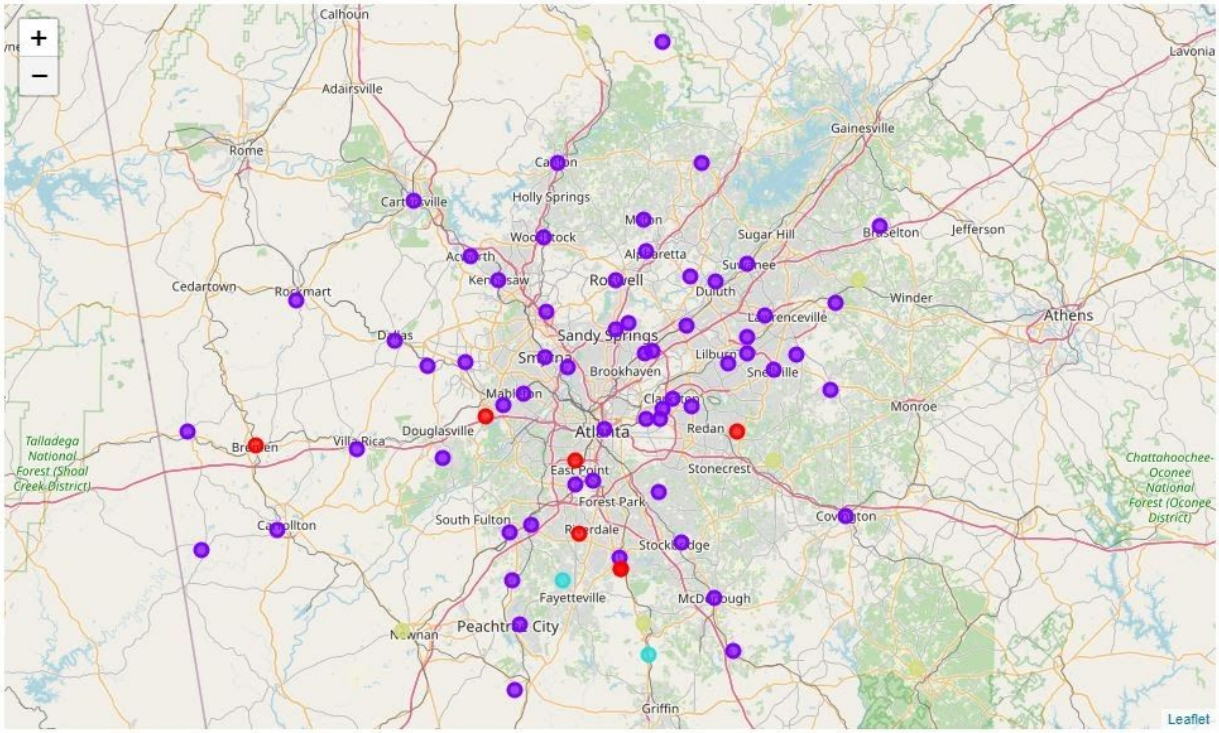
	City	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Acworth	American Restaurant	Fast Food Restaurant	Baseball Field	Sandwich Place	Park	Beach	Hotel	Lake	Pharmacy	Pizza Place
1	Alpharetta	Clothing Store	New American Restaurant	American Restaurant	Coffee Shop	Ice Cream Shop	Fast Food Restaurant	Furniture / Home Store	Plaza	Beer Bar	Café
2	Atlanta	Coffee Shop	History Museum	Hotel	Seafood Restaurant	Caribbean Restaurant	Park	American Restaurant	Bar	Art Gallery	BBQ Joint
3	Auburn	Discount Store	Pharmacy	Construction & Landscaping	Sandwich Place	Baseball Field	Pizza Place	Home Service	Convenience Store	BBQ Joint	Chinese Restaurant
4	Austell	Fast Food Restaurant	Fried Chicken Joint	Gas Station	Rental Car Location	Sandwich Place	Chinese Restaurant	Breakfast Spot	American Restaurant	Boutique	Clothing Store

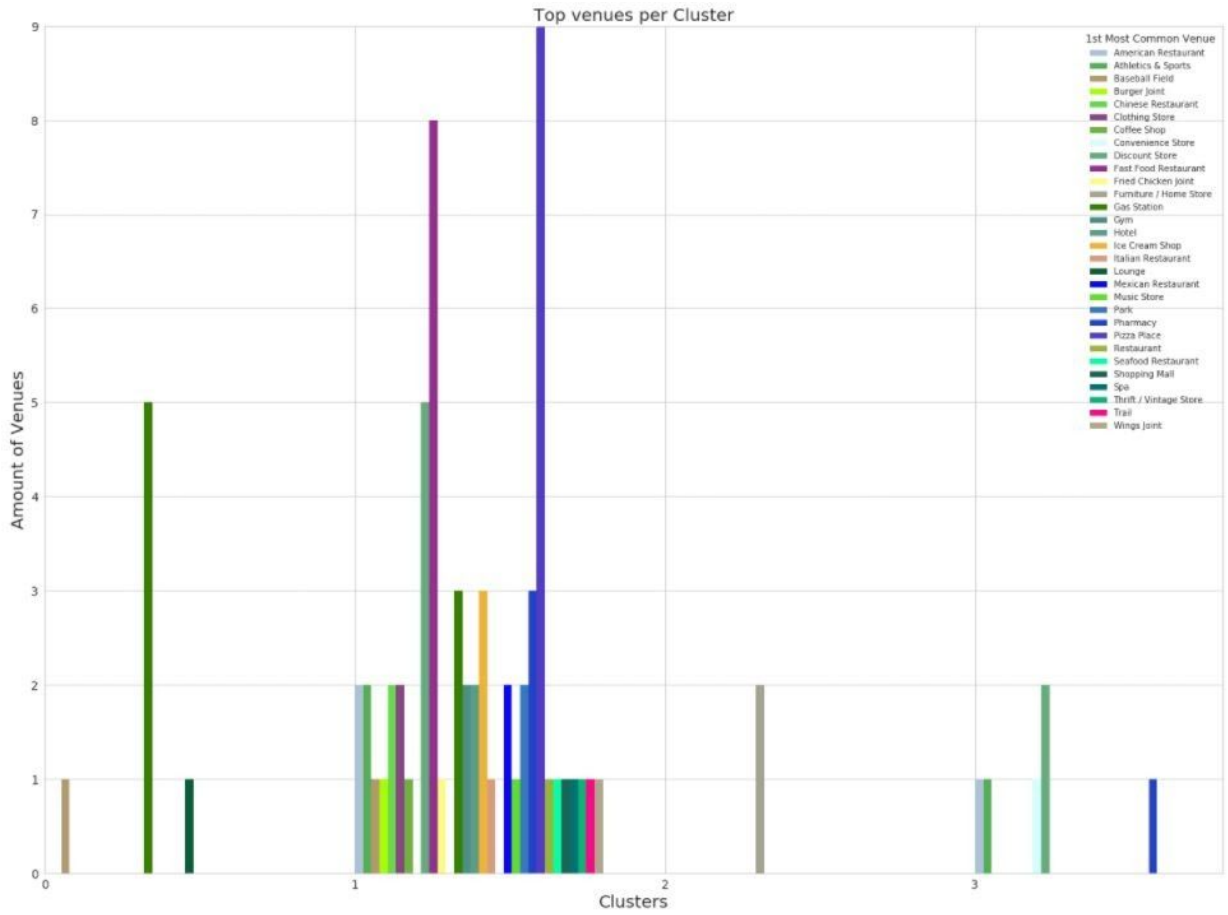
After I created my new dataset I used K-means to cluster the cities. I use the KElbowVisualizer from yellowbrick to visualize the best K for my clusters. The graph didn't show an elbow shape,

meaning that there wasn't a clear K number for the clusters. Here is the latest graph created after several attempts. The scores were usually between 3 and 7 K. I did several cluster runs using numbers between 3 and 7. I decided to leave K at 4 at the end and analyze the clusters.



I mapped the clusters in the map and created a bar plot with the most common venue in each city to help me analyze all the venues.

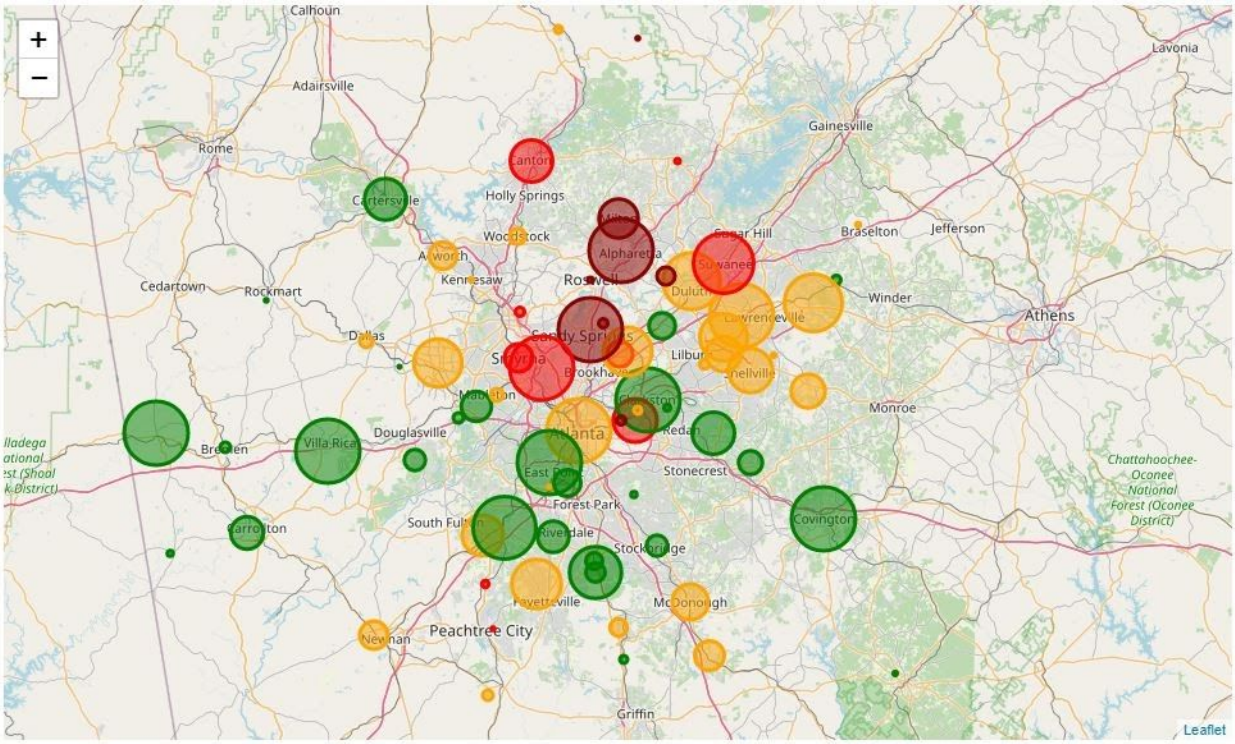




The graph shows that the first cluster is made of gas stations, the second cluster has a lot of fast food restaurants, pizza places and other common stores. The third cluster has furniture/discount stores and the last cluster has discount stores. There's no clear reason to me on why they are clustered this way. After several attempts prior to this one, the results were similar in a way where a single cluster had a disproportionate amount of venues.
I took a closer look at my dataframe to find a common denominator, but it seems the cities don't have a great diversity in venues. I also noticed that some cities started to get less venues as time passed since I've been running my notebook at least once each day for the last couple of weeks.

3.3 Venues and Housing Prices

After I analyzed the venues and the housing prices I decided to visualize them in a geopy map. I decided to display the amount of venues as the size of the marker and keep the previous colors used to display the prices. With this last visualization it gives a good idea of which city to explore according to your needs. I also combined all my data into a final dataframe where I can check all the details for each city analyzed.



4. Results and Discussion

4.1 Results

In the last map it's possible to visualize which cities are the best option according to one's needs. The north-eastern part of Metro Atlanta is where the wealthiest people live since that's where the most expensive houses are, the east, center and west part of Metro Atlanta are the zones where the housing are the cheapest.
We can also see which cities have a bigger amount of venues and therefore there's a bigger amount of business there. The smaller marks represent cities where there aren't many venues and can be a good place for someone who wants to live without too much noise and people's movement.

4.2 Discussion

The types of venues are similar in almost all the cities. If someone wants to start a business, this map can give a better idea on the socioeconomic level of the people living in the city of choice and help decide on the type of venue. Food related venues are what's most common, pizza places and fast food restaurants being the most popular ones.The city of Atlanta is where the biggest variety of venues are packed together compared to the other cities.

Distance is also a big factor to consider here. In the map it's very clear where the wealthiest consumers live and where the cheapest housing is located. If someone wants to move to an affordable area but wants a business in a wealthy area or vice-versa, traveling time needs to be considered. Traffic is a big issue in Metro Atlanta, so traveling long distances isn't recommended unless someone really enjoys to drive or the venue has unusual business hours.

5. Conclusion

5.1 Conclusion

The Atlanta Metropolitan Area is a huge area with many business opportunities, especially if you are interested in opening a restaurant, as well as a great option to live in. The whole area is divided into cities where one can make easy divisions by looking at the last map and map out the areas of interest. While housing is a little over the average housing price of 230,000.00$,it's a metropolitan area that keeps up growing and it will keep growing in the upcoming years.

5.2 Observations and Disclaimer

This whole analysis was done during the Summer of 2020. The data presented here might change drastically during the upcoming months. I would have compared the price increase in each city but I felt it wouldn't be a good prediction yet.
I ran my notebook everyday during the last weeks and found out the venues were also changing. Many businesses are closing because of the Covid-19 pandemic. This analysis was done in an area where I can go out to do some field research and indeed many businesses were closed from one week to another. I'm going to run the analysis again in a couple of months to see what has changed. The current analysis shows food related venues as the best options but that can change drastically in the next months.

6. Appendix

Websites:
[1] Atlanta metropolitan Area:
https://en.wikipedia.org/wiki/Atlanta_metropolitan_area
Metro Atlanta Cities:
http://atlanta.metroguide.com/list_of_cities.htm
https://www.knowatlanta.com/atlanta-metro-counties-cities/
Georgia Cities Coordinate:
https://public.opendatasoft.com/explore/dataset/us-zip-code-latitude-and-longitude/table/?refine.state=GA
Housing Prices Metro Atlanta:
https://www.zillow.com/atlanta-sandy-springs-roswell-metro-ga_r394347/home-values/