Coffee Shops of America

Where can I find a good coffee shop and how does it compare to the rest of the country?

Michael Shealy

July 5th, 2020

1. Introduction

Like a lot of people, I love coffee. At first, I only drank coffee to wake myself up in the morning (which is still the case). However, I have started to enjoy well-made coffee when I get the chance. There are always the times where I will need to quickly stop by Starbucks because I am late to class, but if I have the time, I love to grab myself a well-made cup.

I try to find coffee shops that I have never been to before, and I will come back to the ones I like the most. However, like I said before, I do not always have the time to visit every shop I want to, and I would like to drink the most good coffee I possibly can.

What can help with this is data analysis. Using data about coffee shops via Foursquare API, I can find what coffee shops around me are the most likely to have the best cups of joe. Not only can this be done with Atlanta (my hometown), but it can be done with almost anywhere in the country!

In this project, coffee shops from around Atlanta and other parts of the country will be compiled and analyzed to determine which places have the best coffee according to other reviews and ratings. Various statistical analyses will be done to get a better idea of which cities and coffee shops have the best chance of having the best coffee.

2. Data and Preprocessing

a. Data Required

The data for this project was obtained using Foursquare API. To achieve the goals of this project, the data used pertained to:

- 50 coffee shops in a 15 km radius from each city center
- The respective geographical coordinates for each coffee shop
- The average rating on a 10-point scale for each coffee shop
- The number of ratings each coffee shop has received on Foursquare

This data will allow us to do the required statistical analyses to achieve our goals. Since rating data requires premium calls (which I only get 500 per day), the rating data was saved to a csv once called upon.

The cities used for this project were:

- Atlanta
- New York City
- Seattle
- San Francisco
- Chicago
- Washington D.C
- Boston
- Denver

Since I had control over what data was extracted from the API, no preprocessing steps really had to be done on each dataframe. Each city dataframe was then compiled into a list holding each of the cities' coffee shops.

b. Libraries Needed

There was a mixture of basic and complex libraries needed for this project. Some well-known libraries used that are needed for almost any project were:

- NumPy
- Pandas
- Matplotlib
- Math

There were also certain libraries needed to extract info from Foursquare API and to plot any location data. The libraries needed were:

- Requests
- Json
- Folium

3. Methodology

In this project, we will look at a couple of different significant ways to show the patterns of coffee shops and quality coffee shops. The list is as follows:

1) We will look at the distribution of coffee shops in Atlanta and the ratings of coffee shops in different areas. This will be done by making a grid of the metropolitan area and showing the

average rating of coffee shops in each grid. The amount of ratings each coffee shop has is important so that a better idea of how quality a coffee shop is can be obtained.

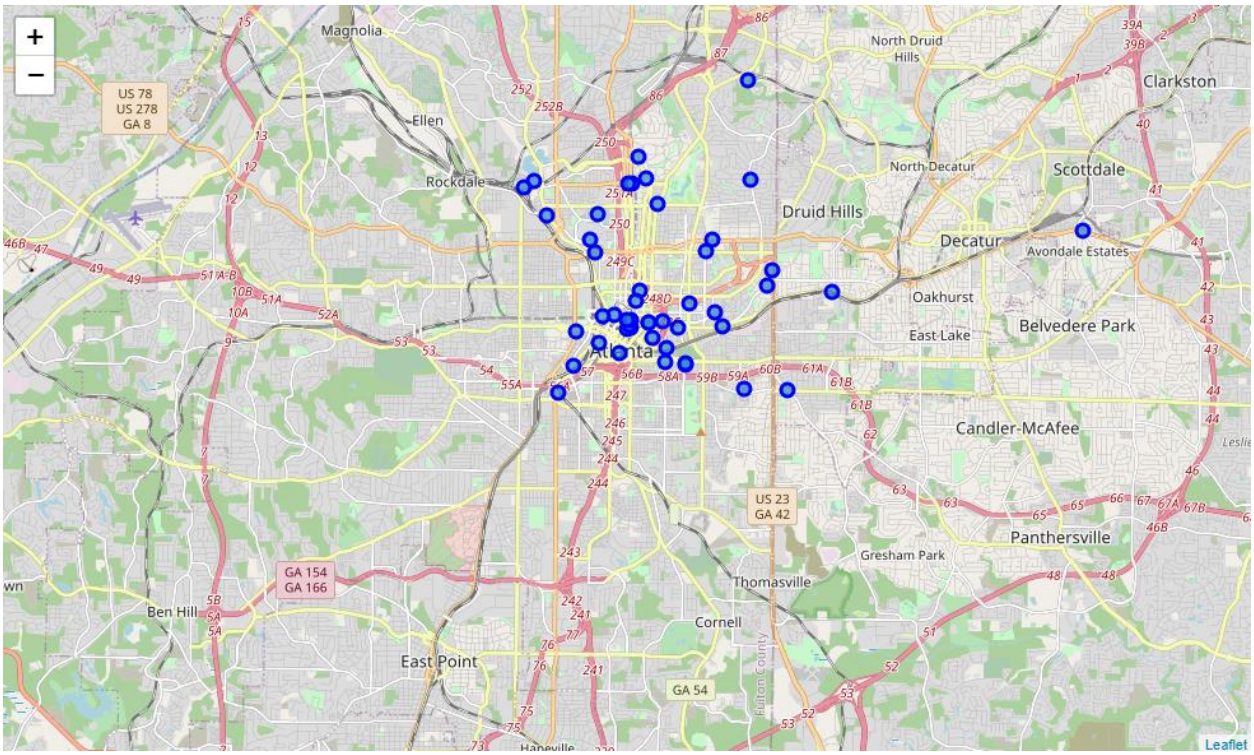
2) Next, we will compile the ratings of coffee shops in each of the cities listed above, and the average rating of coffee shops in each city will be determined.

3) Finally, further statistical analyses will be done using the 'describe' method. Other plots of the data will be used to show how ratings and number of ratings can show which coffee shops have the best chance to have quality coffee.

4. Analysis and Results

4.1 Distribution of Coffee Shops in Atlanta

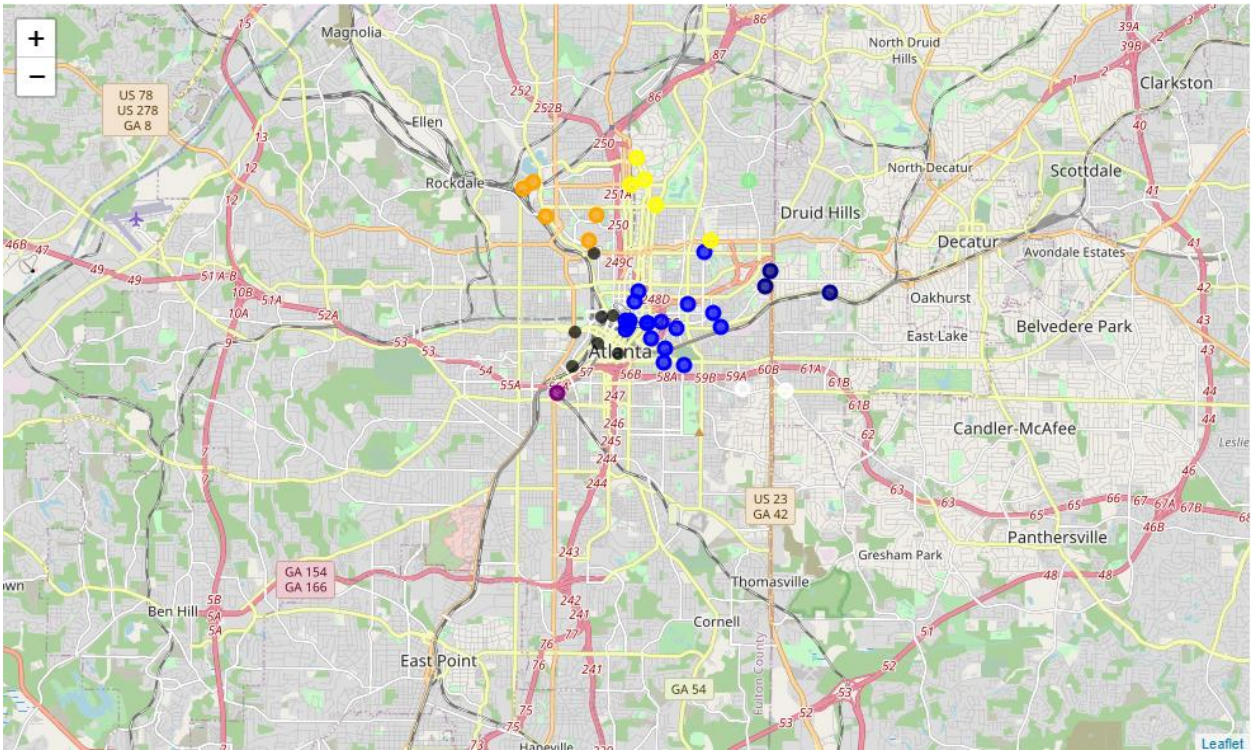
First, the distribution of coffee shops in Atlanta was shown using a folium Map:



Looking at the map above, coffee shops are more concentrated in the center of the city. There may be just as many coffee shops in other areas of the metropolitan area, but they are more spaced out. In terms of conveniently finding coffee shops, it will be easier to find a lot of different coffee shops if you go to the city center than if you go a little outside the main part of the city.

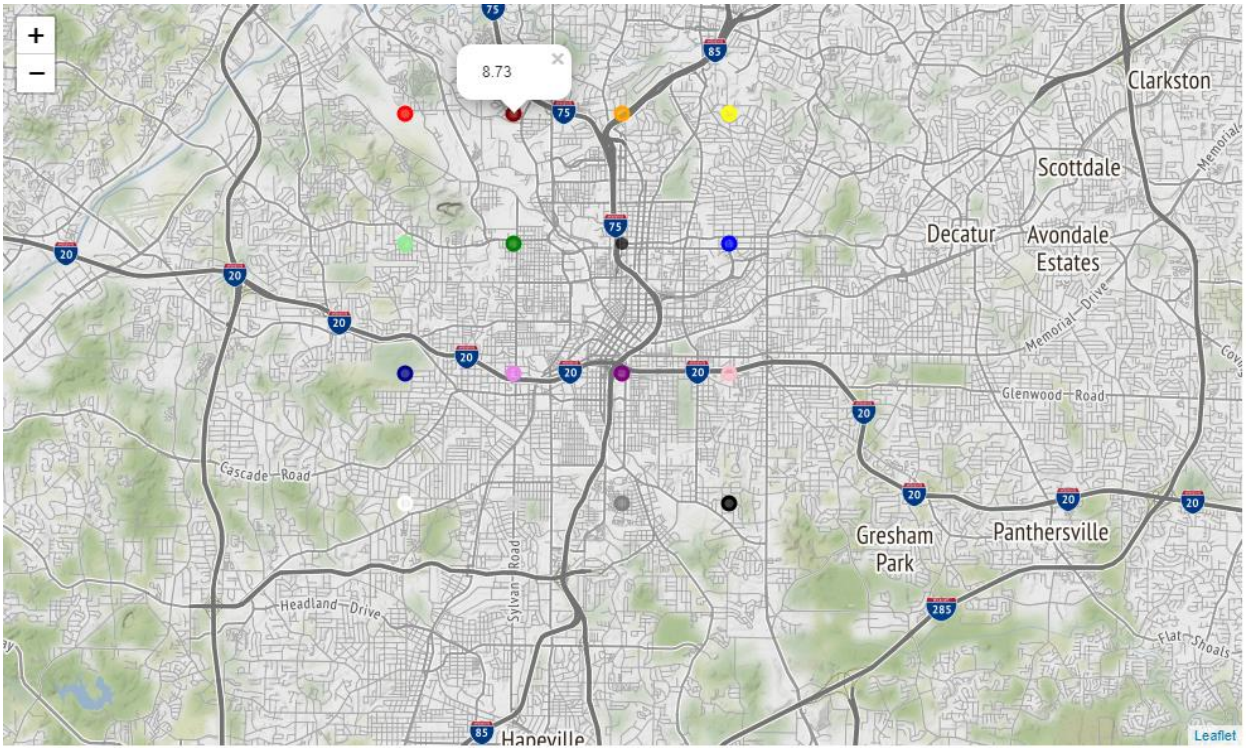
The next step was to divide Atlanta into a 4x4 grid and find the average rating of coffee shops in each block in the grid. This was originally done by finding the furthest coffee shop from the city center and building the grid using this distance, however outliers were found that threw off the

balance of the grid blocks. To overcome this, 8 coffee shops (of which most had no ratings) were removed from the dataset and the process was repeated. The map below shows the “grid block” each coffee shop is associated with:



To determine the average rating of coffee shops in each grid piece, the average rating of each coffee shop was multiplied by the percent of reviews that coffee shop had compared to the total reviews of coffee shops in that grid block. The sum of each grid block was taken to get a weighted average rating. The results are shown below:

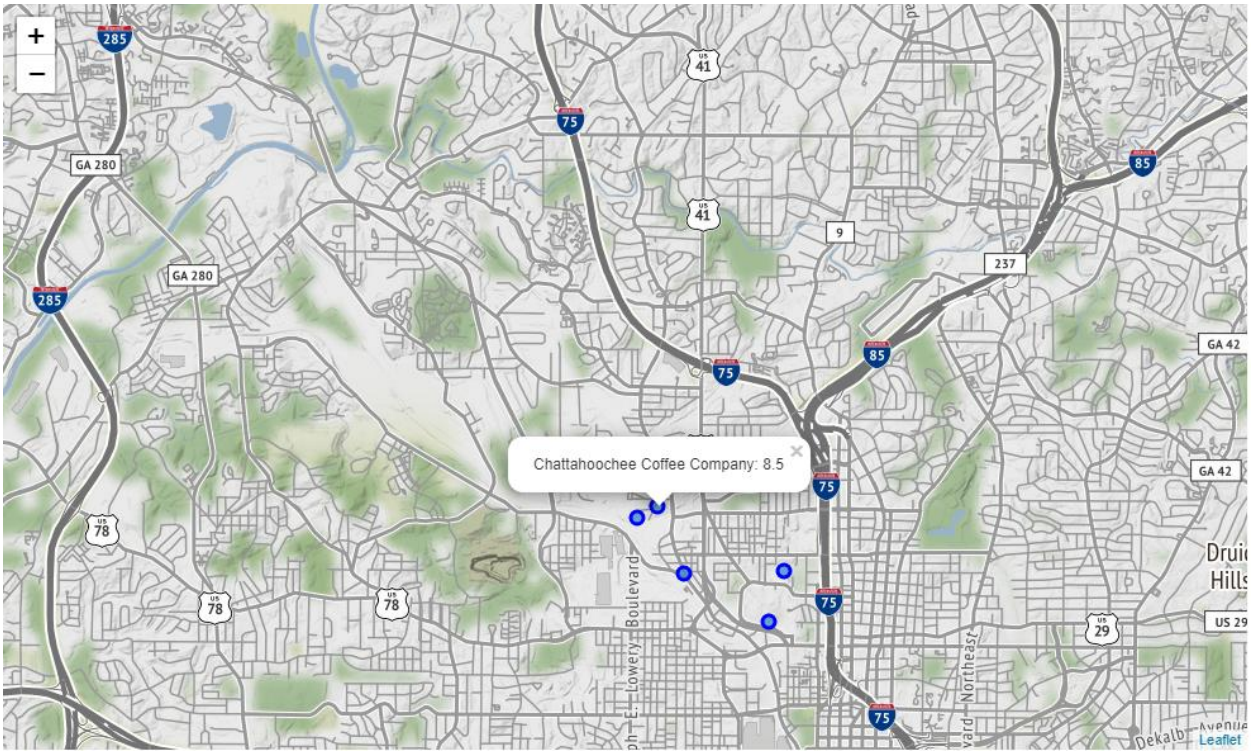
	Grid Class	Avg Rating
0	1.0	0.00
1	2.0	8.73
2	3.0	7.92
3	4.0	7.94
4	5.0	0.00
5	6.0	7.30
6	7.0	8.67
7	8.0	7.74
8	9.0	0.00
9	10.0	0.00
10	11.0	0.00
11	12.0	8.28
12	13.0	6.00
13	14.0	0.00
14	15.0	0.00
15	16.0	0.00



From the table and map above, the two highest rated grid segments are segments 2 and 7. This corresponds to the area where I-75 and I-85 split and the area just below Midtown Atlanta. A lot of the grid blocks have a zero rating because either there were no coffee shops in that area, or the few coffee shops in that area did not have any ratings.

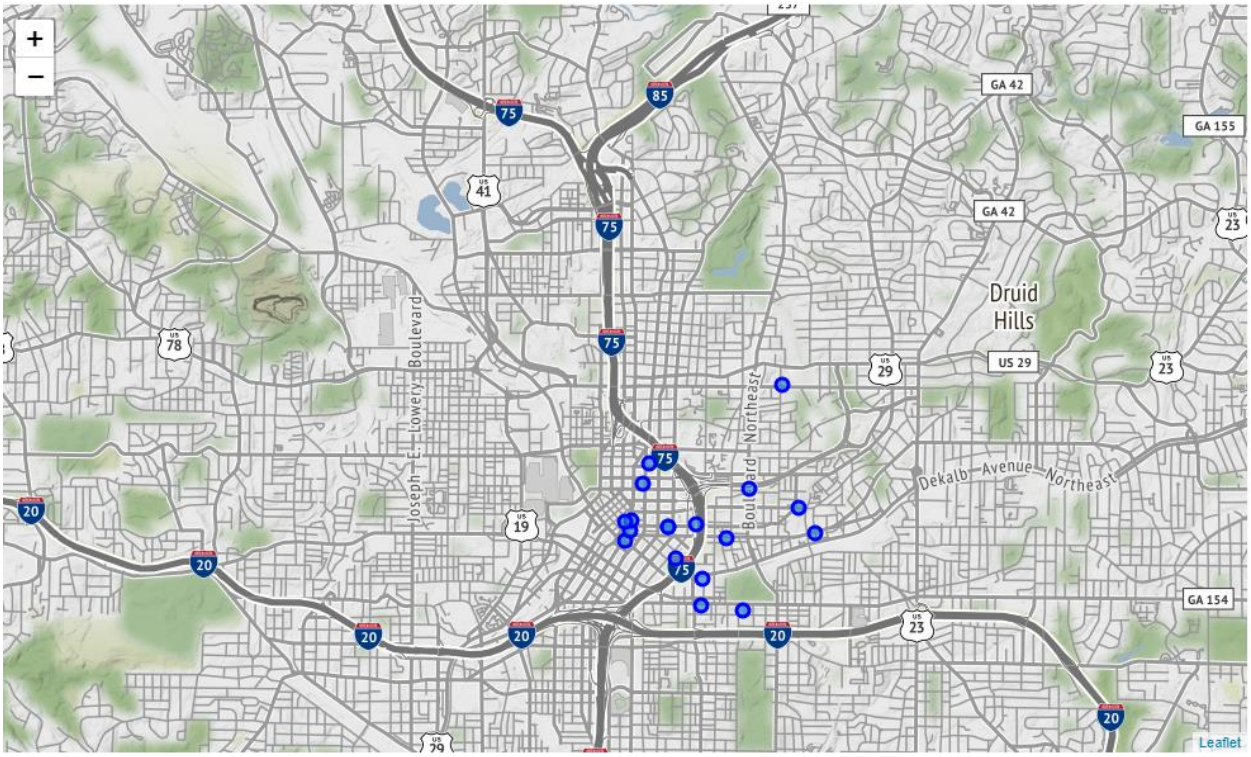
Zooming into segments 2 and 7 give us a better idea of why these two places have the best rated coffee shops:

Segment 2:



There are 5 coffee shops in this area, and 3 have ratings. Octane Coffee and Chattahoochee Coffee Company have very good ratings, and they have a significant number of ratings as well. There is also a decent chance that since there's good coffee shops here, then the shops that do not have reviews also have good coffee.

Segment 7:

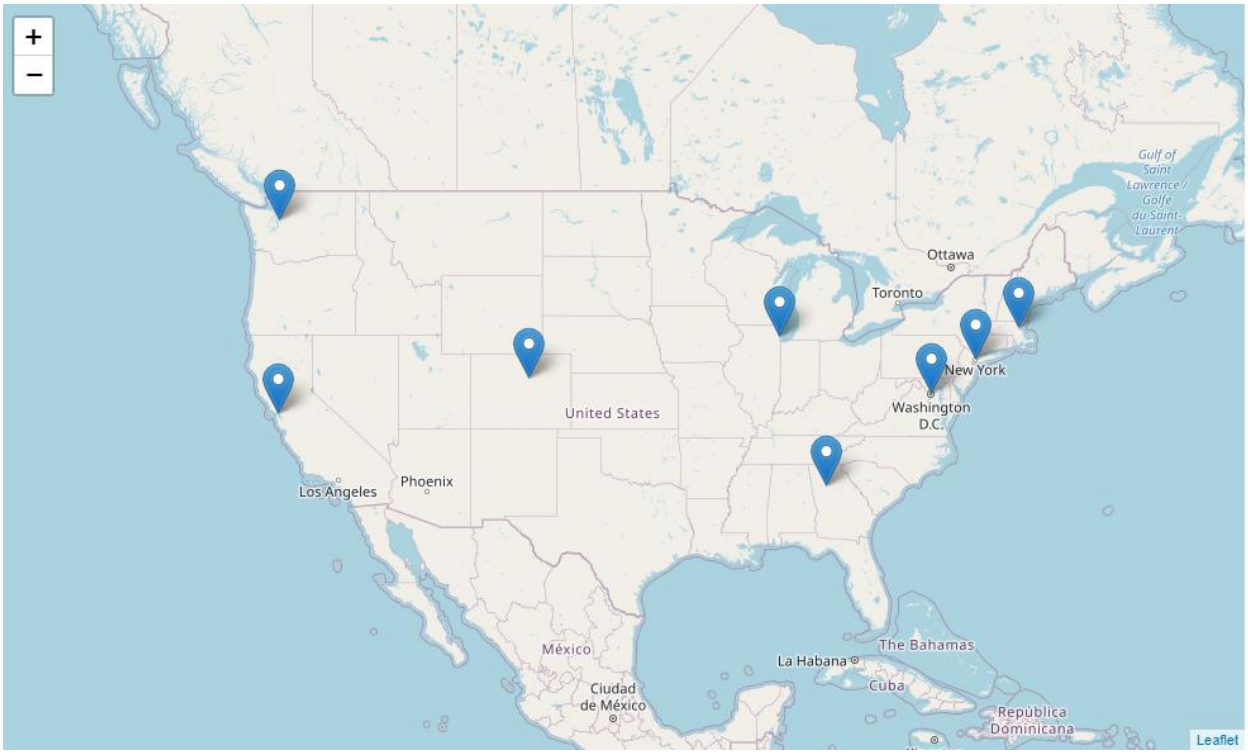


There are A LOT of coffee shops in this grid segment. They are all very close to each other, so if I wanted to visit multiple places at once, this is a great area to go to! It looks like Dancing Goats Coffee Bar has the highest rating with a 9.1. I have actually been to this place and can confirm it has amazing coffee!

4.2 Comparing Coffee Shops Across the Country

So far, we have looked at coffee shops in Atlanta alone, how they are distributed, and how different areas in Atlanta have different amount and ratings of coffee shops. So how does Atlanta coffee compare to other cities in the country? To do this, a similar approach was taken compared to the previous section. This time, the cities will not be separated into grid pieces. Instead, we will use all the coffee shops in each list to create an average weighted rating for coffee shops in each city. After restoring the Atlanta coffee shop dataframe to its original state to be the same as the other cities, the averages were calculated and plotted on a folium Map seen below:

	Avg. Weighted Rating
Atlanta	8.40
New York City	8.54
Seattle	8.33
San Francisco	8.79
Chicago	8.52
Washington DC	8.05
Boston	7.83
Denver	8.11



Using the approach described above, San Francisco had the highest average rating with an 8.79, whereas Boston has the lowest average rating with a 7.83. Atlanta had a relatively high average score with an 8.40, which is the 4th highest rating out of the cities listed. One interesting note is that Atlanta had a higher average rating than Seattle, a city known for having good coffee. We will talk about this more in the Discussion section.

4.3 Further Statistical Analysis on City Ratings

Having an overall weighted average rating of each coffee shop in each city is great, but what other statistics can give us a better look at which cities have good coffee shops? We will start with using the ".describe()" method on each dataframe to see some well-known valuable statistics for this insight. This will initially give us different stat tables for each city, so new tables were made for each significant stat and comparing them between cities. Three tables were made from this: a mean table, a standard deviation table, and a max table:

Mean Table:

	City Latitude	City Longitude	Coffee Shop Latitude	Coffee Shop Longitude	Avg Rating	# of Ratings	% of Ratings	Weighted Rating
Atlanta	33.74278	-84.38889	33.764301	-84.380128	3.950	75.58	0.02	0.168060
New York City	40.71280	-74.00600	40.716835	-74.000577	6.064	314.50	0.02	0.170741
Seattle	47.60620	-122.33210	47.611214	-122.330848	6.342	105.16	0.02	0.166603
San Francisco	37.77490	-122.41940	37.774805	-122.419366	6.436	520.48	0.02	0.175891
Chicago	41.87810	-87.62980	41.885681	-87.635718	4.524	110.40	0.02	0.170376
Washington DC	38.90720	-77.03690	38.905513	-77.033848	6.314	89.34	0.02	0.161011
Boston	42.36010	-71.05890	42.357427	-71.062908	3.724	53.02	0.02	0.156515
Denver	39.73920	-104.99030	39.740912	-104.989144	5.042	62.62	0.02	0.162124

Standard Deviation Table:

	City Latitude	City Longitude	Coffee Shop Latitude	Coffee Shop Longitude	Avg Rating	# of Ratings	% of Ratings	Weighted Rating
Atlanta	2.153270e-14	5.742052e-14	0.028563	0.025560	4.028685	149.633306	0.039596	0.346754
New York City	7.177566e-15	4.306539e-14	0.013759	0.013732	3.699739	483.664256	0.030758	0.271220
Seattle	1.435513e-14	8.613079e-14	0.008339	0.010783	3.118941	137.710041	0.026191	0.231007
San Francisco	5.024296e-14	1.579064e-13	0.011035	0.012525	3.390308	845.786972	0.032500	0.291891
Chicago	4.306539e-14	4.306539e-14	0.015424	0.015112	4.097115	244.588451	0.044310	0.388458
Washington DC	4.306539e-14	4.306539e-14	0.007788	0.015042	2.951306	118.015324	0.026419	0.233574
Boston	3.588783e-14	0.000000e+00	0.007583	0.012666	3.822199	97.110392	0.036632	0.293650
Denver	7.177566e-15	4.306539e-14	0.016176	0.016058	3.732488	78.678859	0.025129	0.218631

Maxes Table:

	City Latitude	City Longitude	Coffee Shop Latitude	Coffee Shop Longitude	Avg Rating	# of Ratings	% of Ratings	Weighted Rating
Atlanta	33.74278	-84.38889	33.843913	-84.265889	9.1	718.0	0.189997	1.671977
New York City	40.71280	-74.00600	40.746237	-73.957043	9.4	2087.0	0.132719	1.167924
Seattle	47.60620	-122.33210	47.638268	-122.306504	9.1	668.0	0.127045	1.117992
San Francisco	37.77490	-122.41940	37.796875	-122.392982	9.3	3618.0	0.139026	1.237327
Chicago	41.87810	-87.62980	41.942366	-87.618591	9.2	1591.0	0.288225	2.536377
Washington DC	38.90720	-77.03690	38.928707	-76.994760	9.1	606.0	0.135662	1.220954
Boston	42.36010	-71.05890	42.374590	-71.045925	9.0	376.0	0.141833	1.191399
Denver	39.73920	-104.99030	39.772406	-104.942560	9.1	333.0	0.106356	0.967838

From the three tables above, the first thing that pops out is the average rating for each city:

	Avg Rating
Atlanta	3.950
New York City	6.064
Seattle	6.342
San Francisco	6.436
Chicago	4.524
Washington DC	6.314
Boston	3.724
Denver	5.042

These values are significantly different compared to the weighted averages we calculated earlier. This is because there are coffee shops in each database that do not have any ratings, and therefore, have a 0.0 rating overall. We can show that as the number of coffee shops with no ratings increases, the average rating we see above decreases:

	0	Avg Rating
Atlanta	25	3.950
New York City	13	6.064
Seattle	9	6.342
San Francisco	10	6.436
Chicago	22	4.524
Washington DC	8	6.314
Boston	25	3.724
Denver	17	5.042

These results make sense; the 'Avg Rating' mean considers the zero ratings as the same weight as any other rating. This is why we used the weighted average rating previously.

One potential conclusion that could be made is that if one wanted to go to a lot of well-known quality coffee shops, then go to the cities with not only high average weighted ratings, but also go to ones with minimal coffee shops with zero ratings. If one wanted to be the first to find a hidden gem of a coffee shop, then go to a city with high weighted average rating, but also a decent amount of coffee shops with zero ratings. If there are high-rated coffee shops in a city, then maybe the same coffee shops in that city with zero ratings will be good too!

The next thing that stands out is the number of ratings:

	# of Ratings
Atlanta	75.58
New York City	314.50
Seattle	105.16
San Francisco	520.48
Chicago	110.40
Washington DC	89.34
Boston	53.02
Denver	62.62

San Francisco and New York City have significantly higher average number of ratings compared to the other cities. Combined with the fact that these two cities had relatively higher average ratings than the other cities, then that makes those averages even more impressive. The standard deviations on these values also provide some information:

	# of Ratings
Atlanta	149.633306
New York City	483.664256
Seattle	137.710041
San Francisco	845.786972
Chicago	244.588451
Washington DC	118.015324
Boston	97.110392
Denver	78.678859

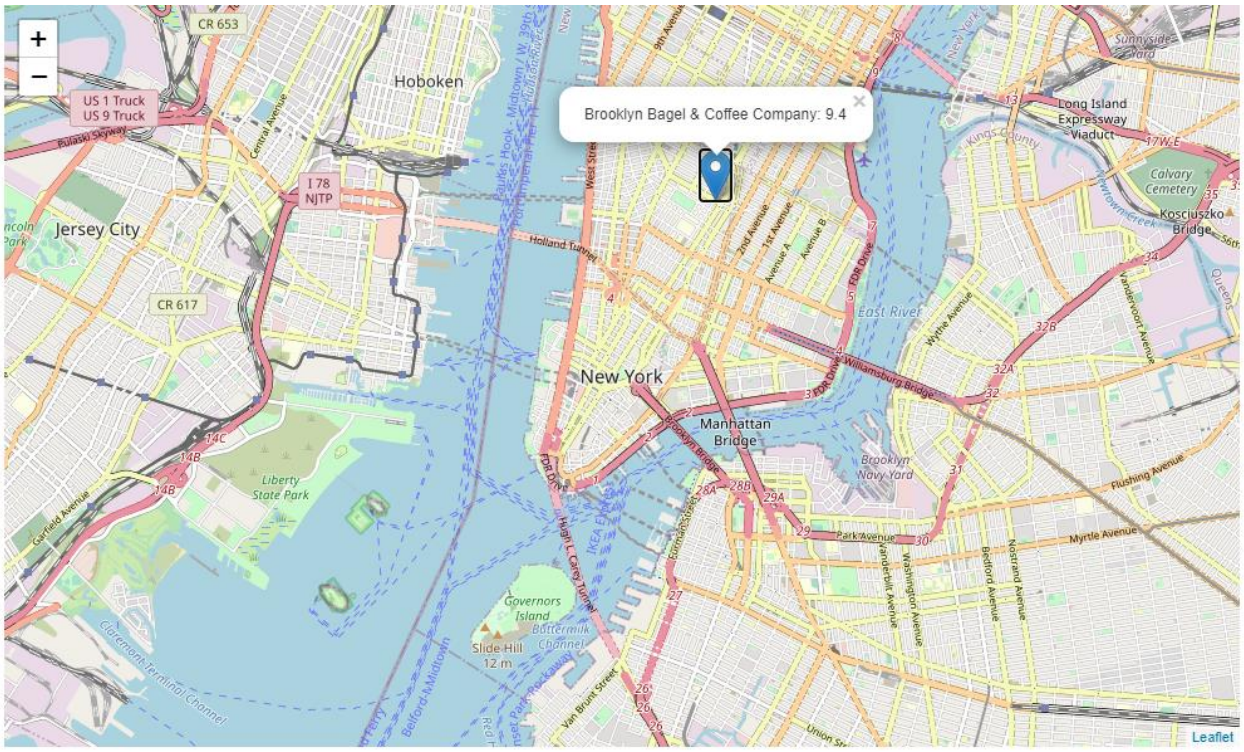
The standard deviations are high but make sense; the standard deviation will increase as the actual values increase in magnitude. Coffee shops with zero ratings will further increase the standard deviation since it will increase the range of distribution of number of ratings.

One important conclusion that can be confirmed from this is something we can generally expect: the cities and coffee shops with higher ratings tend to have a greater number of ratings as well. This makes sense since if a coffee shop is known to have good coffee, more people will go there, which leads to more higher reviews.

Finally, let us find the highest rated coffee shop in each city using the maxes table:

	Avg Rating
Atlanta	9.1
New York City	9.4
Seattle	9.1
San Francisco	9.3
Chicago	9.2
Washington DC	9.1
Boston	9.0
Denver	9.1

All the highest rated coffee shops in each city are close in rating, but one coffee shop in New York City ends up having the highest rating with a 9.4. Let us see which one it is and where it is in New York City:



The highest rated coffee shop out of the ones test was a Brooklyn Bagel & Coffee Company right in the middle of Manhattan! Although this coffee shop only had 91 reviews, it still got great reviews from the people that went there.

5. Discussion

In the first part of the analysis, coffee shops in Atlanta were analyzed and divided into different grid segments of the city. It was noticed that a majority of the coffee shops in Atlanta can be found in the center of the city, and the proximity of coffee shops in this part of Atlanta was much closer than any other part of the city. The average weighted rating of each segment of a 4x4 grid of Atlanta was calculated, and the highest ratings came from the center of the city and just off the split of I-75 and I-85. If I wanted to find good coffee shops in Atlanta, those are the two areas of the city I would most likely want to go to. One improvement to this part of the analysis would be to either gather more coffee shops from around the city or ensure that all the coffee shops had ratings. Since this would require an account on Foursquare API that would cost money, this was avoided for this analysis, but would be useful for further insight.

The second part of this analysis involved taking the weighted average of all the obtained coffee shops for 8 major cities and plotted on a folium Map. From this analysis, San Francisco had the highest weighted average rating, whereas Boston had the lowest. Compared to other major cities, Atlanta is fairly average when it comes to quality coffee. One important note that this analysis does not consider is how people in each city rate the coffee in the city they live in. Although Seattle is known for having excellent coffee, it still had a lower rating than Atlanta. Demographic factors could play an important part in how ratings are given to coffee shops; citizens in some cities could be more critical of coffee at certain shops, whereas citizens in other cities could be content with getting a warm cup of coffee in the morning.

Finally, the data from each city was further evaluated using the '.describe()' method. A strong correlation between actual average rating and the number of coffee shops with zero ratings was found, and also expected. There was also a strong positive correlation between average rating and number of ratings for each city, which makes sense since better coffee shops will have more customers and thus, more ratings. The overall highest rated coffee shop out of the entire dataset was found to be in New York City, right in the middle of Manhattan. The weighted average rating could be fine tuned in the future to reward coffee shops with more reviews. A coffee shop with a slightly lower rating but hundreds of more reviews could end up being a higher quality coffee shop than one that has few reviews but have higher ratings.

6. Conclusion

In this project, coffee shops from around Atlanta and other parts of the country were compiled and analyzed to determine which places have the best coffee according to other reviews and ratings. Various statistical analyses were done to get a better idea of which cities and coffee shops have the best chance of having the best coffee.

The distribution of coffee shops in Atlanta were analyzed, and "hotspots" of coffee shops in the city were found. The average rating of coffee shops in each city was found and compared to determine which cities had the highest rated coffee shops. Finally, correlations were found

between ratings and other statistics to find valuable insights on how ratings and number of ratings are distributed between coffee shops.

Overall, some great areas of Atlanta were found so that I can find some great places to get a good cup of coffee. I can tell that these places are some of the best in the country by comparing these coffee shops to other cities, and if I get the chance to travel to some of these cities, I will know exactly where to go to get a cup of coffee and start my day.