CLUSTER ANALYSIS OF RESTAURANTS TO VISIT, ATLANTA, GA

Submitted by: Snehal Bogar

Date: 21st June 2019

1. Introduction

1.1 Background

Whenever a visitor visits a major city, they tend to spend more time to identify the restaurants which fits their requirement of customer satisfaction and budget, at a given venue. Even after searching for long time, it's not necessary that he will end up at the desired place. By using data science, we can explore data
In order to divide the restaurants in Atlanta in 3 clusters which will help visitor to effectively and efficiently choose a restaurant.

1.2 Interest

This project helps to search venues in Atlanta, USA based upon their rating and average prices. By using Jupyter notebook, we will explore various venues in the city of Atlanta, USA ibn order to help visitors select the restaurants as per their choice. We have used below APIs:

- Foursquare API and
- Zomato API

Whenever a tourist is visiting a city they start looking for places to visit during their stay. They primarily look for places based on the venue ratings across all venues and the average prices such that the locations fit in their budget.

2. Data Acquisition and Data cleaning

Let's use the folium library to create a complete map zoomed on Atlanta. We'll also plot a marker on the coordinates we just identified above. This would give us a relatively good look at the center point we will be considering. The LATITUDE and LONGITUDE of ATLANTA are 33.7490, -84.3894.

2.1 Data Acquisition

Foursquare API:

We will use Foursquare API to fetch venue details around longitude and latitude of Atlanta. We'll call the API over and over till we get all venues from the API within the given distance. The maximum venues this API can fetch is 100, so we will fetch all venues by iteratively calling this API and increasing the offset each time.

Foursquare API requires ClientID, and client secret to function which can be accessed after creating a developer account. We will set the radius as 5 Kilometers. The version is a required parameter which defines the date on which we are browsing so that it retrieves the latest data.

Zomato API:

The Zomato API allows using its search API to search for any given venue based on certain search filters such as query, latitude, longitude and more. Zomato also requires a Zomato user key which can be accessed with a developer account.

We'll use the name, lat, and lng values of various venues fetched from Foursquare API to use the search API and get more information regarding each venue.

The query will be the name of the venue. The start defines from what offset we want to start, so we'll keep it at 0. The count defines the number of restaurants we want to fetch. As we have the exact location coordinates, we'll fetch only one. We will supply the latitude and longitude values. We will set the sorting criteria as real_distance so each time we get the venue we're searching based on location coordinates.

2.2 Data Cleaning

The data from multiple resources might not always align. Thus, it is important to combine the data retrieved from multiple resources properly.

We'll first plot the two data points on the map. We'll then try to combine data points that have their latitude and

longitude values very close to one another. From the remaining selected venues, we will inspect the venues to ensure that any remaining mismatched venues are also removed from the final dataset of venues before we begin any analysis.

We will first plot the Foursquare data on the map and then Zomato data on map.
Foursquare Venues on map:

Zomato Venues on map:

We can see that there are many venues identified by both Foursquare and Zomato. There is a lot of overlapping between the two APIs. However, there are others where the data does not match just like the red dots in the right on the second map.

To combine the two datasets, we'll have to check that the latitude and longitude values of each corresponding venue match. Thus, we will round both the latitude and longitude values up to 4 decimal

places. Then, calculate the difference between the corresponding latitude and longitude values and see if the difference is less than 0.0004 which should ideally mean that the two locations are same.

lot of restaurants aren't mentioned with their Rating and Pricing ('price_for_two'). Hence, we will remove those restaurants from the dataframe.
As we have combined data from both Foursquare and Zomato API, multiple columns are duplicating. Hence, we will remove them too.

Final dataframe looks like this:

	categories	Venue	latitude	longitude	price_for_two	price_range	rating	address
0	Greek Restaurant	Nick's Food To Go	33.7476	-84.3798	15	1	4.2	240 Martin Luther King Jr Drive SE, Atlanta 30312
1	Korean Restaurant	Blossom Tree	33.7555	-84.3891	20	2	3.9	64 Peachtree Street, Downtown, Atlanta 30303
2	Burger Joint	Grindhouse Burgers	33.7540	-84.3799	15	1	4.4	209 Edgewood Avenue SE, Suite 117/18 30303
3	Farmers Market	Miss D's New Orleans Style Candy	33.7543	-84.3799	25	2	3.4	209 Edgewood Avenue SE, Atlanta 30303
4	Arepa Restaurant	Arepa Mia	33.7541	-84.3800	10	1	4.0	209 Edgewood Avenue SE, Downtown, Atlanta 30303
5	Latin American Restaurant	Alma Cocina	33.7589	-84.3872	40	3	4.2	191 Peachtree Street NE, Downtown 30303

3. Methodology

This project aims at identifying the venues in Atlanta, based upon their rating and price_for_two. This would enable any visitor to identify the venues he/she wants to visit based on their rating and cost preference.

Step I: we retrieved the data from two APIs (Foursquare and Zomato). We extract venue information from the center of Atlanta up to 5 Km radius. The latitude and longitude values are then used to fetch venue rating and price from Zomato.

Step II: Secondly, we then explored the data retrieved from the two APIs on the map and identified the top category types. The data from the two sources is carefully combined based on the name, latitude and longitude values from the two sources. The final dataset would include the rating and price values for each venue.

Step III: We'll analyze the data that we created based on the ratings and price of each venue. We'll identify places where many venues are located so that any visitor can go to one place and enjoy the option to choose amongst many venue options. We'll also explore areas that are high rated and those that are low rated while also plotting the map of high- and low-priced venues. Lastly, we'll cluster the venues based on the available information of each venue. This will allow us to clearly identify which venues can be recommended and with what characteristics.

Finally, we'll discuss and conclude which venues to be explored based on visitor requirement of rating and cost.

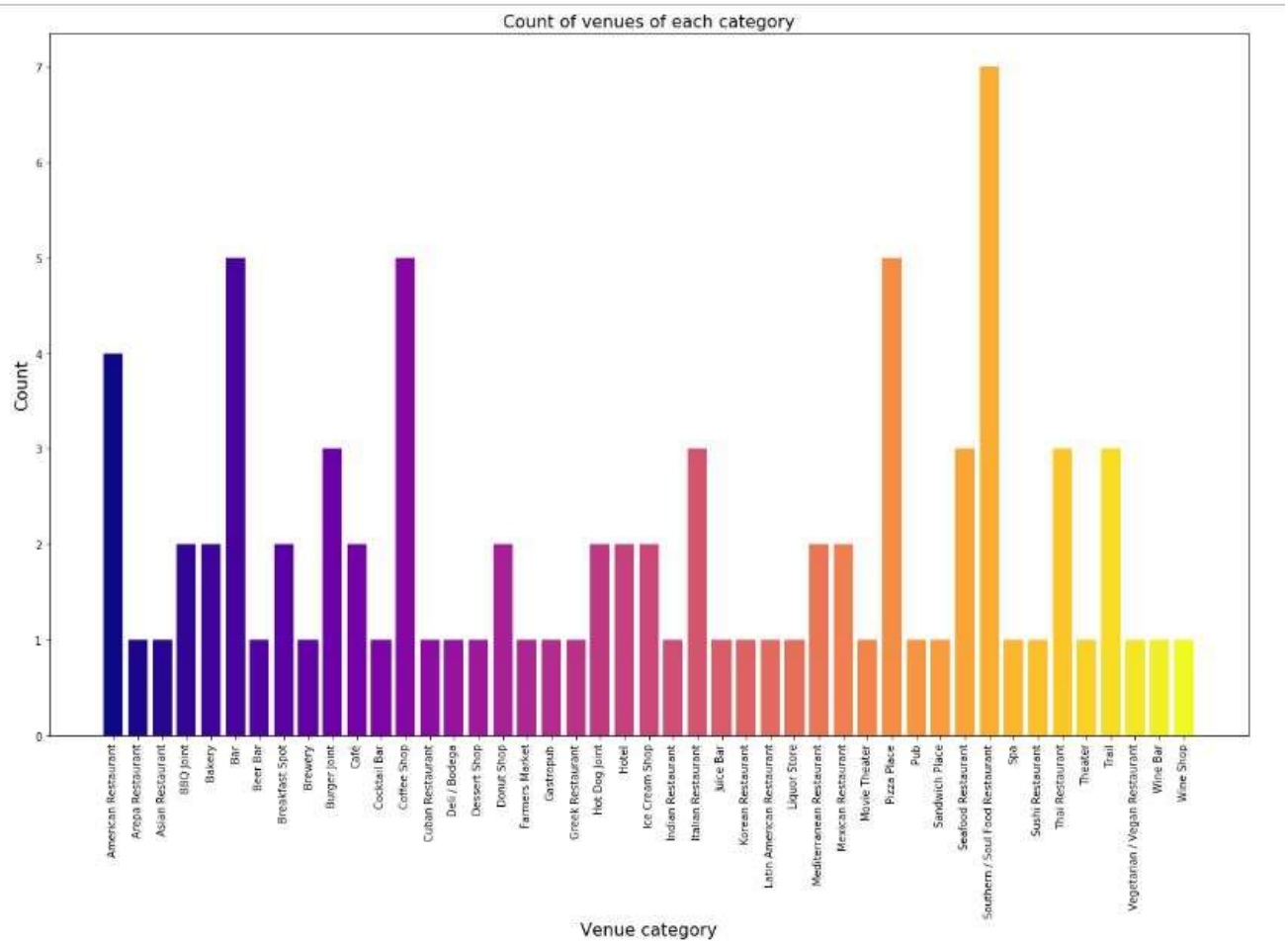
4. Analysis

The complete dataset is now in its final form.

4.1 Venues

We will inspect these venues based on their rating. The rating of a venue is based on user reviews and belongs to a range from 1 to 5. We'll also analyze the venues based on their price per person as well as the price range.

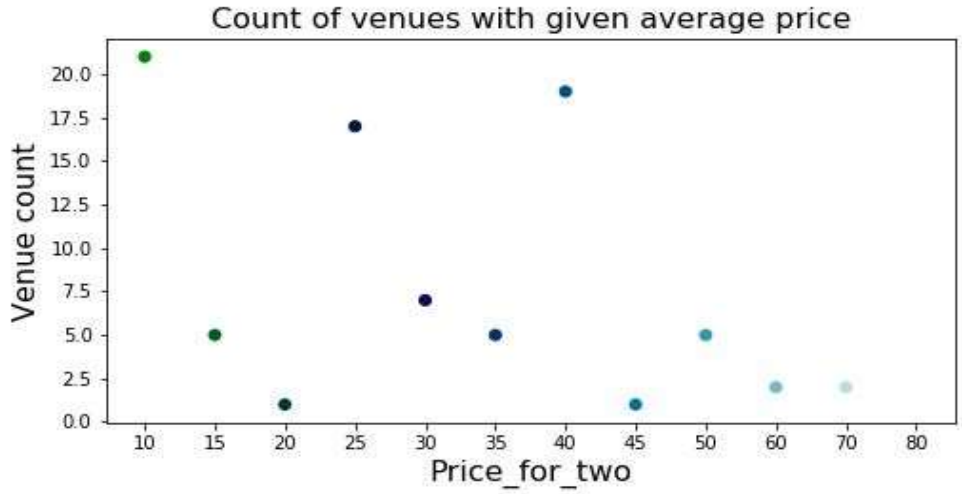
Categories We have various types of venues in the final dataset. We will look at the venues and check which are the majority venue categories in the list.



As we can see the majority venues are **Southern Food, coffee shops and Pizza place**.

4.2 Price_for_two

We will now look the venues based on the price values. We have two price features for our venues, one is price_for_two which defines the cost for two people and the other is price_range which determines the price range as defined by Zomato.



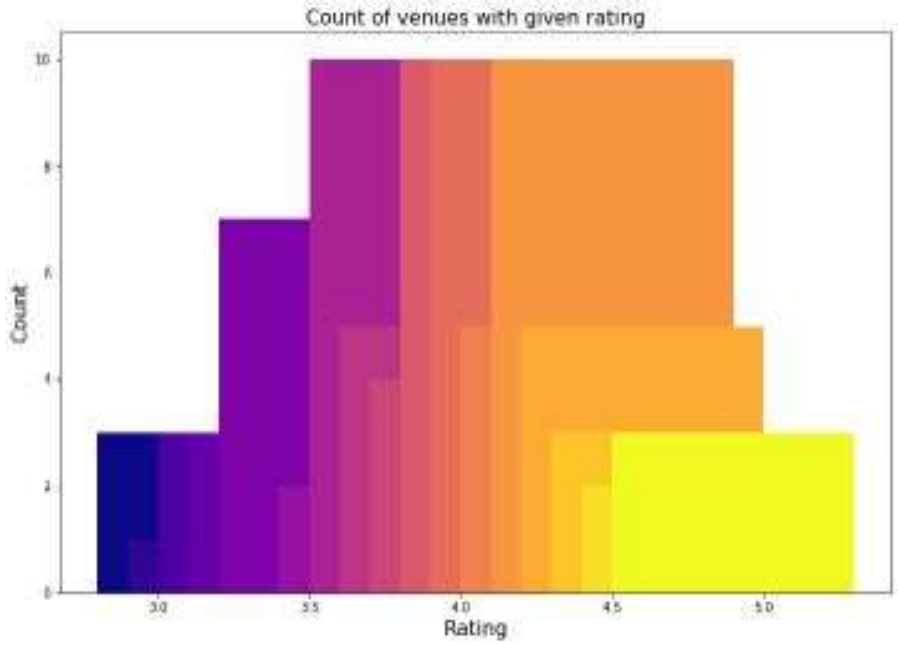
We will first explore the price_for_two using a scatter plot between the price and the count of venues with that average price. We'll size the points based on the price to highlight their price.

When we look at the price values of each venue, we explore that many venues have prices which are offering meal at USD 10, USD 40 and USD 25 for two people

4.3 Rating

Rating of a venue is an important factor on which a visitor decides whether it is worth it to visit the place. To cater to this, we will first see what the average rating for all the venues in the city is. Next, we will plot the venues on the map and color code them.

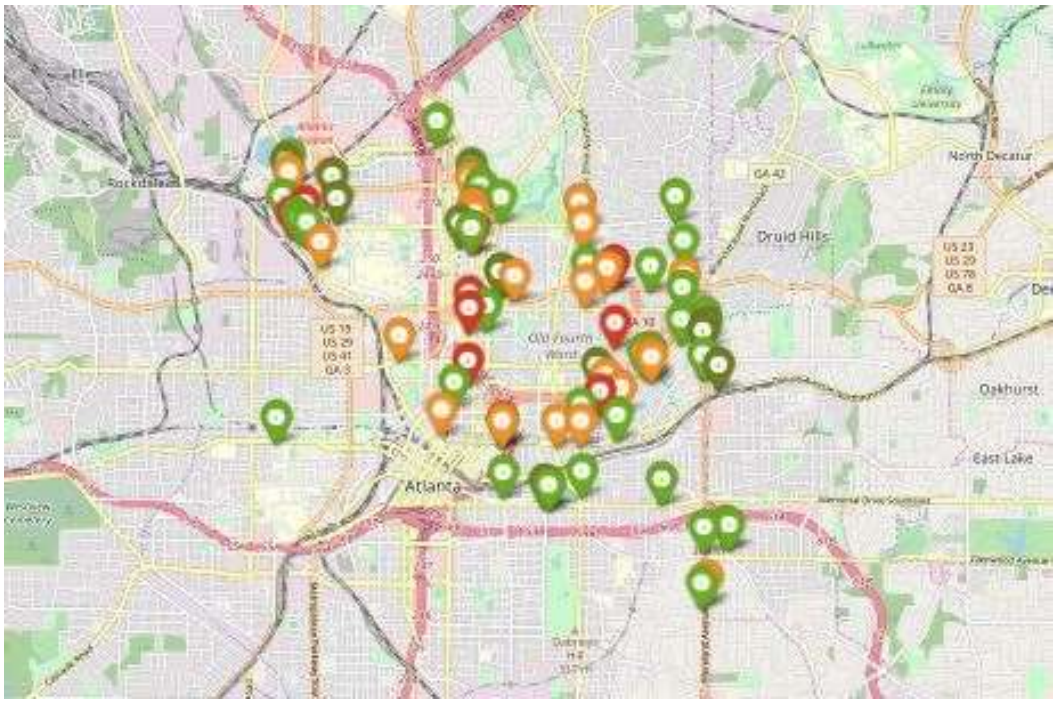
We'll first identify the various rating values and plot them as a bar plot with their counts to see the most common rating.



From the plot above, majority venues have their rating between 3.5 to 4.5.

Let's create bins for various ratings and plot them in different colors on the map. The ratings will be divided between 4 bins:

3.0 to 3.5
3.5 to 4.0
4.0 to 4.5
4.5 to 5.0



The map has the location of all the venues. It appears that many venues are located near about North East and North West Avenue with rating above 4.

Cluster Analysis

We will now cluster all these venues based on their price range, location and more to identify similar venues and the relationship amongst them. We'll cluster the venues into two separate groups.

Sample restaurants in Cluster 0

```
Cluster 0
```

Out[78]:

	cluster_labels	categories	Venue	latitude	longitude	price_for_two	price_range	rating	address	rating_bin
0	0	Farmers Market	Miss D's New Orleans Style Candy	33.7543	-84.3799	25	2	3.4	209 Edgewood Avenue SE, Atlanta 30303	Low
1	0	Latin American Restaurant	Alma Cocina	33.7589	-84.3872	40	3	4.2	191 Peachtree Street NE, Downtown 30303	Good
2	0	Breakfast Spot	Ria's Bluebird	33.7465	-84.3737	30	3	4.6	421 Memorial Drive SE, Grant Park 30312	Very good
3	0	Cocktail Bar	The Sound Table	33.7541	-84.3718	25	2	3.9	483 Edgewood Avenue SE, Old Fourth Ward, Atlan...	Okay
4	0	Mexican Restaurant	Agave Restaurant	33.7478	-84.3679	40	3	4.5	242 Boulevard SE 30312	Good
5	0	Pizza Place	Ammazza	33.7541	-84.3681	40	3	4.0	591 Edgewood Avenue SE, Old Fourth Ward, Atlan...	Okay

These venues for cluster 0 have mean price range meal of 4.00 and rating spread around 4.23.

Sample restaurants in Cluster 1

Cluster 1

	cluster_labels	categories	Venue	latitude	longitude	price_for_two	price_range	rating	address	rating_bin
0	1	Hotel	Sear	33.7616	-84.3851	70	4	3.4	265 Peachtree Center Avenue 30303	Low
1	1	Seafood Restaurant	Six Feet Under Pub & Fish House-Grant Park	33.7462	-84.3730	50	4	4.8	437 Memorial Drive SE, Grant Park 30312	Very good
2	1	Italian Restaurant	BoccaLupo	33.7547	-84.3627	60	4	4.1	753 Edgewood Avenue NE, Inman Park, Atlanta 30307	Good
3	1	Italian Restaurant	Sotto Sotto	33.7630	-84.3577	80	4	4.4	313 N Highland Avenue NE, Inman Park, Atlanta ...	Good
4	1	Spa	Cafe Intermezzo	33.7834	-84.3832	50	4	4.5	1065 Peachtree Street NE, Atlanta 30309	Good
5	1	Seafood Restaurant	The Optimist	33.7799	-84.4108	70	4	4.5	914 Howell Mill Road, Westside 30318	Good

These venues for cluster 1 have mean price range of 4.00 and rating spread around 4.23

Sample restaurants in Cluster 2

Cluster 2

	cluster_labels	categories	Venue	latitude	longitude	price_for_two	price_range	rating	address	rating_bin
0	2	Greek Restaurant	Nick's Food To Go	33.7476	-84.3798	15	1	4.2	240 Martin Luther King Jr Drive SE, Atlanta 30312	Good
1	2	Korean Restaurant	Blossom Tree	33.7555	-84.3891	20	2	3.9	64 Peachtree Street, Downtown, Atlanta 30303	Okay
2	2	Burger Joint	Grindhouse Burgers	33.7540	-84.3799	15	1	4.4	209 Edgewood Avenue SE, Suite 117/18 30303	Good
3	2	Arepa Restaurant	Arepa Mia	33.7541	-84.3800	10	1	4.0	209 Edgewood Avenue SE, Downtown, Atlanta 30303	Okay
4	2	Coffee Shop	Octane Coffee / Little Tart Bakeshop	33.7461	-84.3730	10	1	4.3	437 Memorial Drive, Atlanta 30312	Good
5	2	American Restaurant	Atlanta Breakfast Club	33.7647	-84.3955	10	1	3.9	249 Ivan Allen Junior Boulevard, Atlanta 30313	Okay

These venues for cluster 2 have mean price range of 1.04 and rating spread around 3.97

5. Results and Discussion

After collecting data from the Foursquare and Zomato APIs, we collected data for of 233 different venues. However, not all venues from the two APIs were identical. Hence, we had to inspect their latitude and longitude values as well as names to combine them and remove all the outliers. This resulted in a total venue count of 86.

We identified that from the total set of venues, majority of them were Southern restaurants, Coffee shops and Pizza centers.

When we look at the price values of each venue, we explore that many venues have prices which are offering meal at USD 10, USD 40 and USD 25 for two people. However, the variation in prices is very large, given the complete range starts from USD 10 and goes until USD 70.

While the complete range of ratings range from 1 to 5, the majority venues have ratings was between 3.5 to 4.5. This means that most restaurants provide good quality food which is liked by the people of the city, thus indicating the high rating. When we plot these venues on the map, we discover that there are clusters of venues around Northeastern and Northwestern Avenue near Georgia University. These clusters also have very high ratings (more than 3.9).

Finally, through clusters we identified that there are many venues which are relatively lower priced but have an average rating of 3.97. On the other hand, there are few venues which are high priced and have average rating of 4.23.

If you're looking for low priced places with relatively high rating, you should check Peachtree Street. If you're looking for the best places, with the highest rating but might also carry a high price tag, you should visit downtown restaurants near Highland Avenue. If you're looking to explore the city and have intermediate requirements on budget and rating to decide upon the places you want to visit, you should try Spring street.

A company can use this information to build up an online website/mobile application, to provide users with up to date information about various venues in the city based on the search criteria (name, rating and price).

6. Conclusion

The purpose of this project was to explore the places that a person visiting Atlanta can go for a meal. The venues have been identified using Foursquare and Zomato API and have been plotted on the map. The map reveals that there are three major areas a person can visit: Peachtree Street, Highland Avenue or Spring street. Based on the visitor's venue rating and price requirements, he/she can choose amongst the three places.