ANALYSIS OF THE PALATE OF ATLANTA NEIGHBORHOODS

1. INTRODUCTION

Atlanta is the capital, most populous, most ethnically diverse and largest city of the U.S. state of Georgia. With an estimated 2018 population of 498,044, it is also the 37th most populous city in the United States. The city serves as the cultural and economic center of the Atlanta metropolitan area, home to 5.9 million people and the ninth-largest metropolitan area in the nation. As of 2010, Atlanta is the seventh-most visited city in the United States, with over 35 million visitors per year.

2. BUSINESS PROBLEM

Food diversity is naturally an important part of such an ethnically diverse city. Restaurant business is hence one of the lucrative business opportunities in Atlanta, the city being home to a large and diverse population and with a heavy tourist foot fall.

The purpose of this project is to analyze the various neighborhoods of Atlanta to assist restaurateurs to decide on the perfect location for by being able to answer the following questions -

- What are the most top 10 populous neighborhoods in Atlanta?
- What are the top 5 favorite cuisines in Atlanta neighborhoods?
- What are the locations suitable for these cuisines?

3. DATA PREPARATION

To perform the analysis, following data sources will be used:

3.1. WIKIPEDIA - ATLANTA NEIGHBORHOOD

- **Link:** https://en.wikipedia.org/wiki/Table_of_Atlanta_neighborhoods_by_population
- **Description:** The wiki link has been scraped to extract the Atlanta neighborhood vs. population details. Neighborhoods that have population more than 2500 has been selected for the analysis.

Neighborhood	Population
Atlanta University Center	5703
Bolton	2996
Campbellton Road	4709
Candler Park	3291
Center Hill	3058

Note: For the analysis, top 15 most populated neighborhoods have been filtered out using pandas.

3.2. FOURSQUARE API

- **Link:** https://developer.foursquare.com/docs/places-api/endpoints/
- **Description:** The Places API offers real-time access to Foursquare's global database of rich venue data and user content. This will be used to make RESTful API calls to retrieve data about venues in different neighborhoods of Atlanta, via the following endpoints:
    - *venues (search)* - This endpoint returns a list of venues near the current location,
    - *venues (explore)* - This endpoint returns a list of recommended venues near the current location

3.3. RETRIEVE COORDINATES OF ATLANTA NEIGHBORHOODS VIA GEOPY CLIENT

The coordinates of the Atlanta neighborhoods have been obtained using geocoder class of Geopy client.

4 coordinates (Old Fourth Ward, Pine Hills, Home Park, Peachtree Heights West) are incorrectly retrieved, which may have occurred due to spelling differences of the names. So, these coordinates are replaced with values acquired from google search. After little more playing around with pandas, a well-arranged data-frame as below was obtained —

	Neighborhood	Population	Lat	Lng
0	Midtown	16569	33.781127	-84.386360
1	Atlanta Downtown	13411	33.748309	-84.392118
2	Old Fourth Ward	10505	33.764000	-84.372000
3	North Buckhead	8270	33.839814	-84.379559
4	Pine Hills	8033	33.837500	-84.351600
5	Virginia-Highland	7800	33.782656	-84.353691
6	Grant Park	6771	33.735862	-84.370932
7	Georgia Tech	6607	33.776033	-84.398841
8	Kirkwood	5897	33.756217	-84.323258
9	Atlanta University Center	5703	33.751543	-84.413597
10	Collier Heights	5593	33.769829	-84.507154
11	East Atlanta	5033	33.740106	-84.344925
12	Home Park	4941	33.786200	-84.401900
13	Grove Park	4929	33.773440	-84.445485
14	Peachtree Heights West	4767	33.836300	-84.387800

4. METHODOLOGY

A study will be done by exploring the 15 most populated Atlanta neighborhoods in order to find out the top-5 favorite cuisines in those areas. Then the correlation of the food venues with other factors, such as nearness to Arts & Entertainment centers, Travel & Transport, Professional & Other Places etc. will be analyzed to find out the most important factors affecting each of the cuisine types. Eventually the neighborhoods will be clustered based on the frequency of the food venues, the presence of associated other venues and population. Finally, a recommender for the best location of the choice of restaurant will be prepared upon careful analysis of each cluster.

4.1. DATA CLEANING

In order to segment the neighborhoods of Atlanta based on Food venues, the Parent category of each venue is obtained from Foursquare API (venue → categories endpoint)

	Parent Category	id	Venue Category
0	Arts & Entertainment	56aa371be4b08b9a8d5734db	Amphitheater
1	Arts & Entertainment	4fceea171983d5d06c3e9823	Aquarium
2	Arts & Entertainment	4bf58dd8d48988d1e1931735	Arcade
3	Arts & Entertainment	4bf58dd8d48988d1e2931735	Art Gallery
4	Arts & Entertainment	4bf58dd8d48988d1e4931735	Bowling Alley

The neighborhoods are then explored to find out the recommended venues near each location (within its 500 meters of the vicinity) using Foursquare API (venue → explore endpoint). This dataframe is then merged with the above dataframe to mark the parent category for each venue.

The merged dataframe has all the required information. The size of this dataframe is determined, and it is found that there is a total of 172 venues. The final dataframe is as follows –

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category	Parent Category
0	Midtown	33.781127	-84.38636	Cafe Agora	33.780932	-84.384460	Mediterranean Restaurant	Food
1	Midtown	33.781127	-84.38636	Mac's Beer & Wine	33.780916	-84.387992	Liquor Store	Shop & Service
2	Midtown	33.781127	-84.38636	Savi Provisions	33.781010	-84.384278	Gourmet Shop	Shop & Service
3	Midtown	33.781127	-84.38636	Dancing Goats Coffee Bar	33.780810	-84.386653	Coffee Shop	Food
4	Midtown	33.781127	-84.38636	Brazilian Wax by Andreia	33.780646	-84.387739	Spa	Shop & Service

Now that we have 2-level of details for each venue, the neighborhood dataframe is further filtered out to obtain only the food related venues –

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category	Parent Category
0	Midtown	33.781127	-84.38636	Cafe Agora	33.780932	-84.384460	Mediterranean Restaurant	Food
1	Midtown	33.781127	-84.38636	Dancing Goats Coffee Bar	33.780810	-84.386653	Coffee Shop	Food
2	Midtown	33.781127	-84.38636	Sweet Hut Bakery & Cafe	33.780315	-84.384100	Bakery	Food
3	Midtown	33.781127	-84.38636	Marlow's Tavern	33.780110	-84.387566	New American Restaurant	Food
4	Midtown	33.781127	-84.38636	Ecco Midtown	33.778827	-84.385988	Mediterranean Restaurant	Food

4.2. FEATURE ENGINEERING

Now, each neighborhood is analyzed individually to understand the most common cuisine being served.

The above process is taken forth by using 'one hot encoding' function of python 'pandas' library. One hot encoding converts the categorical variables (which are 'Venue Category') into a form that could be provided to ML algorithms.

Upon converting the categorical variables, 'Neighborhood' column is added back.

Further, the number of venues of each category in each neighborhood is counted. The top 5 Food 'Venue Categories' are found by counting their occurrences. This analysis is depicted below which shows that 'American Restaurant', 'Seafood Restaurant', 'Sandwich Place', 'Mediterranean Restaurant', 'Gastropub' are among the top 5.

	count	mean	std	min	25%	50%	75%	max
American Restaurant	13.0	0.692308	1.436698	0.0	0.0	0.0	1.0	5.0
Seafood Restaurant	13.0	0.461538	0.877058	0.0	0.0	0.0	1.0	3.0
Sandwich Place	13.0	0.769231	1.012739	0.0	0.0	0.0	1.0	3.0
Mediterranean Restaurant	13.0	0.384615	0.869718	0.0	0.0	0.0	0.0	3.0
Gastropub	13.0	0.384615	0.767948	0.0	0.0	0.0	0.0	2.0

4.3 DATA VISUALIZATION

The top 5 categories are further plotted individually on bar graph using python 'matplotlib' library. A code block is created which loops and plots the graph of top 10 neighborhoods for a category.
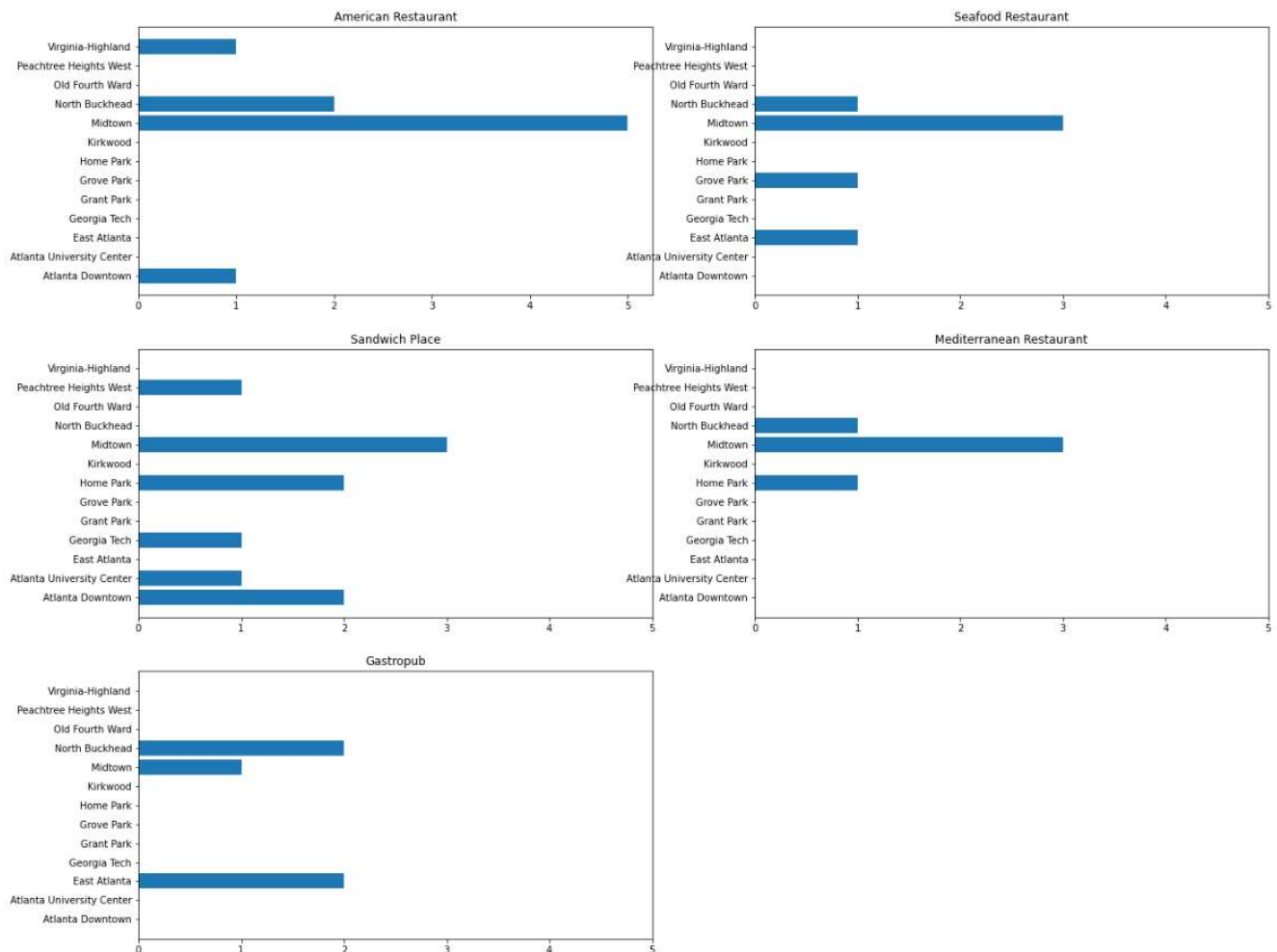
Midtown has the highest concentration of almost all the top5 cuisines.

Next, the other venues that exists in each of the neighborhood are fetched in a dataframe and are merged along with the food venues.

	Neighborhood	Nightlife Spot	Professional & Other Places	Shop & Service	Travel & Transport	American Restaurant	Seafood Restaurant	Sandwich Place	Mediterranean Restaurant	Gastropub	Population
0	Atlanta Downtown	4	0	1	4	1.0	0.0	2.0	0.0	0.0	13411
1	Atlanta University Center	1	0	2	1	0.0	0.0	1.0	0.0	0.0	5703
3	East Atlanta	11	0	7	0	0.0	1.0	0.0	0.0	2.0	5033
4	Georgia Tech	0	0	1	0	0.0	0.0	1.0	0.0	0.0	6607
5	Grant Park	1	0	4	1	0.0	0.0	0.0	0.0	0.0	6771
6	Grove Park	0	0	3	1	0.0	1.0	0.0	0.0	0.0	4929
7	Home Park	1	0	4	0	0.0	0.0	2.0	1.0	0.0	4941
8	Kirkwood	0	0	0	0	0.0	0.0	0.0	0.0	0.0	5897
9	Midtown	7	1	19	6	5.0	3.0	3.0	3.0	1.0	16569
10	North Buckhead	2	0	10	1	2.0	1.0	0.0	1.0	2.0	8270
11	Old Fourth Ward	0	0	2	0	0.0	0.0	0.0	0.0	0.0	10505
12	Peachtree Heights West	0	0	12	0	0.0	0.0	1.0	0.0	0.0	4767
14	Virginia-Highland	5	0	16	0	1.0	0.0	0.0	0.0	0.0	7800

A visualization is now created to find out the correlation of each of the top 5 food categories with the existence of other venues and population in each neighborhood.

|	American Restaurant	Seafood Restaurant	Sandwich Place	Mediterranean Restaurant	Gastropub
Nightlife Spot	0.440121	0.536426	0.178391	0.3289	0.659203
Professional & Other Places	0.900885	0.869626	0.661831	0.90354	0.240772
Shop & Service	0.714149	0.610402	0.262772	0.619139	0.384436
Travel & Transport	0.826309	0.696572	0.723214	0.706442	0.153684
American Restaurant	1	0.849568	0.577143	0.902908	0.418322
Seafood Restaurant	0.849568	1	0.411362	0.840366	0.580558
Sandwich Place	0.577143	0.411362	1	0.676834	-0.0906648
Mediterranean Restaurant	0.902908	0.840366	0.676834	1	0.383906
Gastropub	0.418322	0.580558	-0.0906648	0.383906	1
Population	0.803488	0.552962	0.578346	0.630224	0.126448

4.4 MACHINE LEARNING

k-means is an unsupervised machine learning algorithm which creates clusters of data points aggregated together because of certain similarities. This algorithm will be used to count neighborhoods for each cluster label for variable cluster size.

The k-Means algorithm with the number of clusters = 3 is used to cluster the neighborhoods. A dataframe is created that depicts neighborhoods assigned to different clusters:
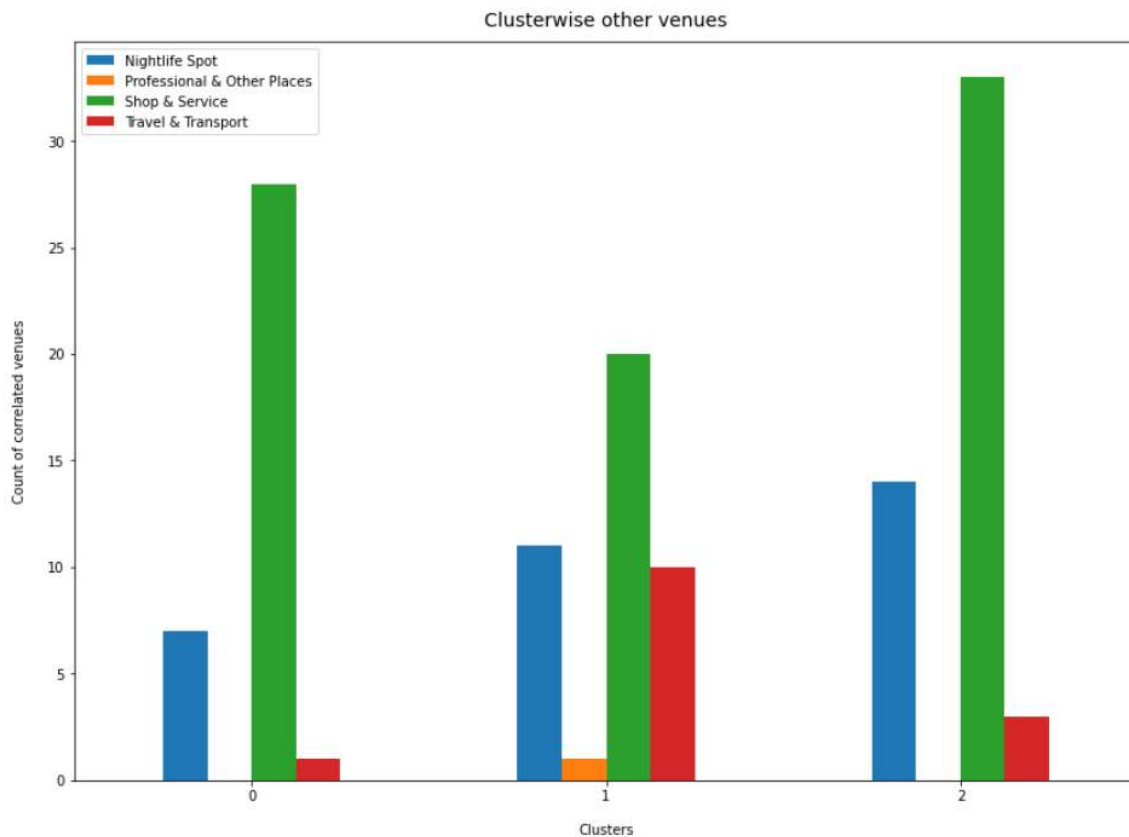
Cluster	Neighborhood
1	Atlanta Downtown
2	Atlanta University Center
2	East Atlanta
2	Georgia Tech
2	Grant Park
2	Grove Park
2	Home Park
2	Kirkwood
1	Midtown
0	North Buckhead
0	Old Fourth Ward
2	Peachtree Heights West
0	Virginia-Highland

The Atlanta neighborhoods are visualized utilizing the python 'folium' library. Following map is generated which shows the desired segmentation of the neighborhoods:
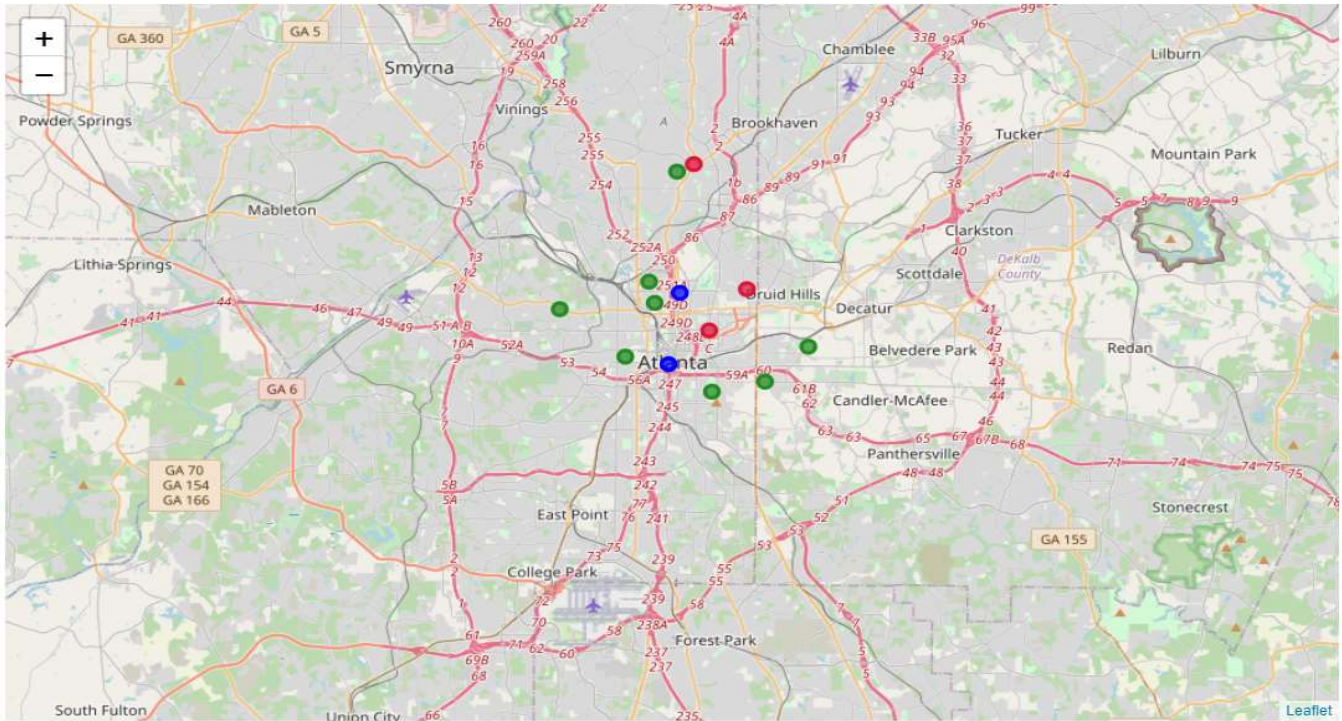
Cluster 0: red, Cluster 1: blue, Cluster 2: green

We can visualize the count of other venues categories in each of the clusters by using the following bar plot –

The above visualization is important because we will compare each cluster and will try to find out what type of food category is suitable in which cluster and in what neighborhood based on the correlation, we found out earlier.

5. RESULTS

Cluster 0

This cluster has a good number of Shop & Services and some Nightlife spots. The Travel & Transport venues are very negligible and there are no Professional places. Based on the correlation map that was created earlier, we can conclude that Cluster 0 is best suited for Seafood restaurant, followed by Gastropub.

The following neighborhoods are part of Cluster 0:

	Neighborhood	Nightlife Spot	Professional & Other Places	Shop & Service	Travel & Transport	American Restaurant	Seafood Restaurant	Sandwich Place	Mediterranean Restaurant	Gastropub	Population	Cluster
10	North Buckhead	2	0	10	1	2.0	1.0	0.0	1.0	2.0	8270	0
11	Old Fourth Ward	0	0	2	0	0.0	0.0	0.0	0.0	0.0	10505	0
14	Virginia-Highland	5	0	16	0	1.0	0.0	0.0	0.0	0.0	7800	0

Further **Virginia-Highland** can be considered the best location for starting **Seafood Restaurant** or **Gastropub** business, as this neighborhood has Nightspots and Shop Service venues centers, all of which have warm correlation with Seafood / Gastropub cuisines. But Virginia-Highland has no Seafood / Gastropub restaurant, hence there will be no competition at all.

Cluster 1

This cluster has a good number of Shop & Services, moderate numbers of Nightlife spots and Travel & Transport. The Professional & Other places venues are very negligible. Based on the correlation map that was created earlier, we can conclude that Cluster 1 is best suited for Sandwich places, followed by Mediterranean, American Restaurant and Seafood.

	Neighborhood	Nightlife Spot	Professional & Other Places	Shop & Service	Travel & Transport	American Restaurant	Seafood Restaurant	Sandwich Place	Mediterranean Restaurant	Gastropub	Population	Cluster
0	Atlanta Downtown	4	0	1	4	1.0	0.0	2.0	0.0	0.0	13411	1
9	Midtown	7	1	19	6	5.0	3.0	3.0	3.0	1.0	16569	1

Further **Atlanta Downtown** can be considered the best location for starting **Seafood Restaurant** or **Mediterranean restaurant** business, as this neighborhood has 4 Travel & Transport venues, which has warm correlation to Seafood / Gastropub cuisines. But Atlanta Downtown has no Seafood / Gastropub restaurant, hence there will be no competition at all.

Cluster2

This cluster has the highest number of Shop & Services, moderate Nightlife spots, low Travel & Transport and no Professional places. Based on the correlation map that was created earlier, we can conclude that Cluster 2 is best suited for American Restaurant and Seafood restaurants.

	Neighborhood	Nightlife Spot	Professional & Other Places	Shop & Service	Travel & Transport	American Restaurant	Seafood Restaurant	Sandwich Place	Mediterranean Restaurant	Gastropub	Population	Cluster
1	Atlanta University Center	1	0	2	1	0.0	0.0	1.0	0.0	0.0	5703	2
3	East Atlanta	11	0	7	0	0.0	1.0	0.0	0.0	2.0	5033	2
4	Georgia Tech	0	0	1	0	0.0	0.0	1.0	0.0	0.0	6607	2
5	Grant Park	1	0	4	1	0.0	0.0	0.0	0.0	0.0	6771	2
6	Grove Park	0	0	3	1	0.0	1.0	0.0	0.0	0.0	4929	2
7	Home Park	1	0	4	0	0.0	0.0	2.0	1.0	0.0	4941	2
8	Kirkwood	0	0	0	0	0.0	0.0	0.0	0.0	0.0	5897	2
12	Peachtree Heights West	0	0	12	0	0.0	0.0	1.0	0.0	0.0	4767	2

Further **Grant Park** and **Grove Park** can be considered the best location for starting **American Restaurant** business, as this neighborhood has moderate Shop & Service venues and1 Travel & Transport venues each, which has warm correlation with American Restaurant cuisines. But Grant Park and Grove Park has no American restaurant, hence there will be no competition at all.

Finally, we arrive at the following summary on location recommendation for each type of cuisines

Cuisine Category	Recommended location
American Restaurant	Grant Park , Grove Park
Seafood Restaurant	Virginia-Highland , Atlanta Downtown
Mediterranean Restaurant	Atlanta Downtown
Gastropub	Virginia-Highland

6. CONCLUSION

This study provides a detailed insight to the restaurateurs on the choice of cuisines that are most famous in the Atlanta neighborhoods. Also, this can be used as a recommender for the location when deciding on where to start the business so as to get the best out of it.