

# Predicting the most profitable neighborhood to open a fast-food restaurant in Atlanta

Merve Usta

3/8/2021

## 1. Introduction/ Business Problem

An entrepreneur who owns a fast food restaurant wants to open the second branch of his restaurant in Atlanta. Since he wants to increase his profit, he must open his restaurant in a crowded population and in a neighborhood with low competition in this sector. To find a solution, he applies to a consulting firm who can help with this. I work as a data analyst in the information technology department of this company. In this project, I will try to analyze data that I have and find the most effective solution by using the machine learning clustering algorithm which is 'k-means' to solve the problem of our customer. First, in data processing part of this project, I will determine the top 10 most crowded neighborhoods of Atlanta by cleaning my data. Next, I will visualize my data using Folium library and analyze it.

## 2. Data Section

### 2.1. Data Sources

First, I gathered in a table the population data of the neighborhoods in the city of Atlanta that went to Wikipedia records in 2010 by building the code to scrape the Wikipedia page. The raw data I have is shown below.

	Neighborhood	Population (2010)	NPU
0	Adair Park	1331	V
1	Adams Park	1763	R
2	Adamsville	2403	H
3	Almond Park	1020	G
4	Ansley Park	2277	E
5	Ardmore	756	E
6	Argonne Forest	590	C
7	Arlington Estates	776	P
8	Ashview Heights	1292	T
9	Atlanta University Center	5703	T

## 2.2. Data Cleaning

In this section, since there was no need for 'NPU' information, I removed from the table. I checked some missing and duplicate values. Since the data of the first 10 crowded neighborhoods in Atlanta is sufficient for analysis, I sorted the data I already have by population. I manually created the coordinate data of these neighborhoods in csv format. By processing the data, I gathered the data I needed in a table and made it ready for visualization and analysis. In the methodology section, I will make more detailed inferences using this data frame.

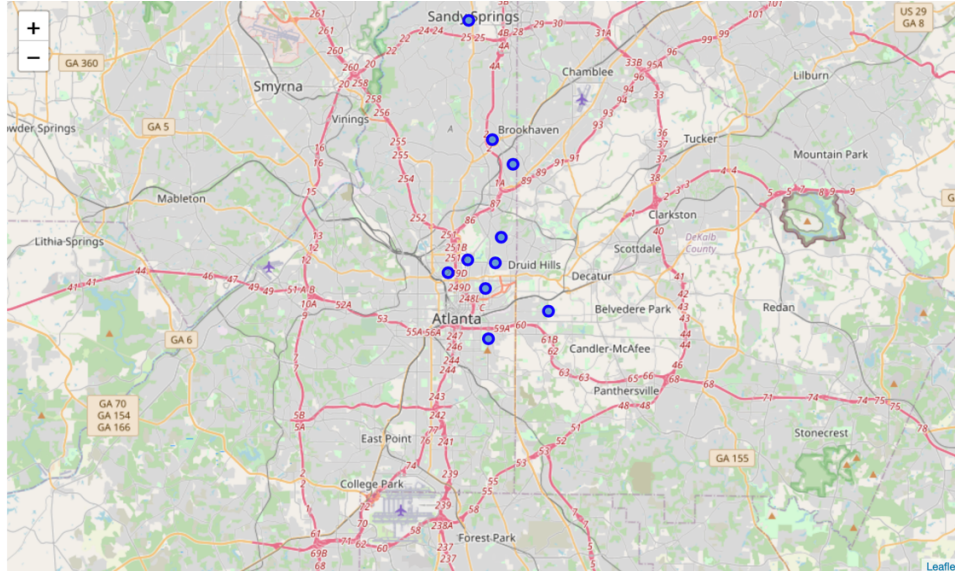
	Neighborhood	Population (2010)	Latitude	Longitude
0	Midtown	16569	33.783020	-84.382332
1	Downtown	13411	33.921520	-84.381912
2	Old Fourth Ward	10505	33.766430	-84.370407
3	North Buckhead	8270	33.852700	-84.365400
4	Pine Hills	8033	33.838715	-84.350830
5	Morningside/Lenox Park	8030	33.796200	-84.359500
6	Virginia-Highland	7800	33.781700	-84.363500
7	Grant Park	6771	33.737200	-84.368200
8	Georgia Tech	6607	33.775600	-84.396300
9	Kirkwood	5897	33.753300	-84.326200

## 3. Methodology

In the methodology section, I used the data frame that I have created as a database. This data has components such as neighborhood, population, latitude and longitude. To analyze and solve my problem, these information about Atlanta's neighborhoods was sufficient.

### 3.1. Creating a map of Atlanta with neighborhoods

I used Folium library to visualize the data I have. I have created a map that shows the ten most populous neighborhoods of Atlanta, which I have added visually below. This visualization helped me gain a visually general knowledge of the neighborhoods.



### 3.2. Defining Foursquare Credentials and Version

I utilized the Foursquare API to explore the most crowded neighborhoods and segment them. Using the latitude and longitude information of ten crowded neighborhoods, I have reached the data by setting a 100-venue limit within 500 meters of these neighborhoods. I started to explore first neighborhood which is Midtown in our dataframe. As a result, 72 venues were returned by Foursquare. Here is the table of venues of Midtown.

	name	categories	lat	lng
0	Exhale	Spa	33.783294	-84.383368
1	Loews Atlanta Hotel	Hotel	33.783366	-84.383188
2	Café Intermezzo	Café	33.783136	-84.383470
3	Street Food Thursdays (& Mondays)	Food Truck	33.784558	-84.382534
4	Einstein's	New American Restaurant	33.784143	-84.382086

### 3.3. Exploring Neighborhoods in Atlanta

I created a function to repeat the same process to all the neighborhoods have high population in Atlanta. To run this function, I wrote the code for each neighborhood and created new dataframe which contains venues in these neighborhoods of Atlanta. When I checked the size of this dataframe, I came to the conclusion that there are 240 venues in different sectors in these ten crowded neighborhoods.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Midtown	33.78302	-84.382332	Exhale	33.783294	-84.383368	Spa
1	Midtown	33.78302	-84.382332	Loews Atlanta Hotel	33.783366	-84.383188	Hotel
2	Midtown	33.78302	-84.382332	Café Intermezzo	33.783136	-84.383470	Café
3	Midtown	33.78302	-84.382332	Street Food Thursdays (& Mondays)	33.784558	-84.382534	Food Truck
4	Midtown	33.78302	-84.382332	Einstein's	33.784143	-84.382086	New American Restaurant

It is also seen in the table below how many venues I have obtained in several different categories for each neighborhood.

	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Neighborhood						
Downtown	35	35	35	35	35	35
Georgia Tech	12	12	12	12	12	12
Grant Park	9	9	9	9	9	9
Kirkwood	15	15	15	15	15	15
Midtown	72	72	72	72	72	72
Morningside/Lenox Park	3	3	3	3	3	3
North Buckhead	56	56	56	56	56	56
Old Fourth Ward	2	2	2	2	2	2
Pine Hills	3	3	3	3	3	3
Virginia-Highland	33	33	33	33	33	33

Totally, I found 106 unique categories can be curated from all the returned venues. As seen in the table, it is observed that the number of venues in different categories is highest in the Downtown area.

### 3.4. Analyzing Each Neighborhood

In this part of my project, I have analyzed each neighborhood separately by writing some functions to get me to the more accurate result. Then I created a table which shows list of the top 10 venue categories for each neighborhood, as seen below.

	Neighborhood	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	Downtown	Spa	Restaurant	Bakery	Vegetarian / Vegan Restaurant	Pizza Place	Bar	Middle Eastern Restaurant	Accessories Store	Park	Chinese Restaurant
1	Georgia Tech	Fast Food Restaurant	Sandwich Place	Chinese Restaurant	Food Court	College Theater	Coffee Shop	Music Venue	Restaurant	Bank	Athletics & Sports
2	Grant Park	Zoo Exhibit	Music Venue	Playground	Pharmacy	Park	Pool	Wine Bar	Video Store	Historic Site	Fast Food Restaurant
3	Kirkwood	Pet Store	Pizza Place	Bar	Coffee Shop	Mexican Restaurant	Breakfast Spot	Sandwich Place	Historic Site	Sports Bar	Juice Bar
4	Midtown	American Restaurant	Hotel	Seafood Restaurant	Spa	New American Restaurant	Italian Restaurant	Coffee Shop	Southern / Soul Food Restaurant	Gay Bar	Indian Restaurant

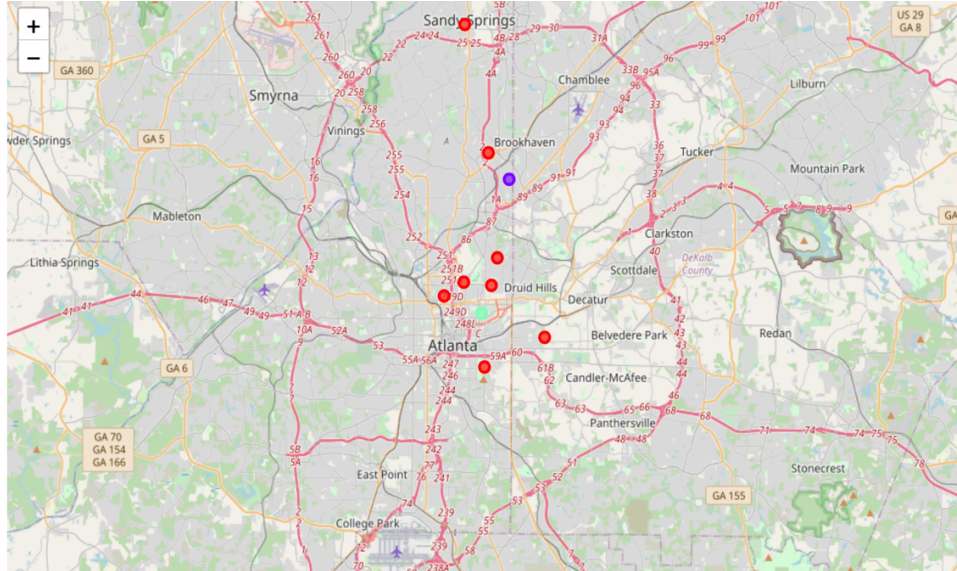
### 3.5. Clustering Neighborhoods

Since there are the same kinds of venues in every neighborhood, it will be easier for us to gather them in a group according to their similarities. For this reason, I used the ‘k-Means’ which is the most popular unsupervised machine learning algorithm to cluster the neighborhoods.

I run k-means to cluster the neighborhood into 3 clusters. Then I created a new dataframe that includes the cluster as well as the top 10 venues for each neighborhood. Here is my merged table with cluster labels for each neighborhood.

	Neighborhood	Population (2010)	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
0	Midtown	16569	33.783020	-84.382332	0	American Restaurant	Hotel	Seafood Restaurant	Spa	New American Restaurant	Italian Restaurant	Coffee Shop	Southern / Soul Food Restaurant
1	Downtown	13411	33.921520	-84.381912	0	Spa	Restaurant	Bakery	Vegetarian / Vegan Restaurant	Pizza Place	Bar	Middle Eastern Restaurant	Accessories Store
2	Old Fourth Ward	10505	33.766430	-84.370407	2	Italian Restaurant	Playground	Furniture / Home Store	Dive Bar	Doctor's Office	Electronics Store	Exhibit	Farmers Market
3	North Buckhead	8270	33.852700	-84.365400	0	Women's Store	Steakhouse	Boutique	Hotel	Italian Restaurant	Coffee Shop	Furniture / Home Store	Kids Store
4	Pine Hills	8033	33.838715	-84.350830	1	Pool	Scenic Lookout	Furniture / Home Store	Dive Bar	Doctor's Office	Electronics Store	Exhibit	Farmers Market

Finally, I created the map below by visualizing the resulting clusters. Each color represents different cluster.



As shown on the clustered map, the red colored clusters are more than the others because the number of venues with common characteristics is higher in these neighborhoods.

## 4. Results

### Examine Clusters

I examined each cluster and determined the discriminating venue categories that distinguish each cluster. Based on the defining categories, I assigned a name to each cluster.

#### Cluster 1

```
9]: Atlanta_merged.loc[Atlanta_merged['Cluster Labels'] == 0, Atlanta_merged.columns[[1] + list(range(5, Atlanta_merged.
```

9):

	Population (2010)	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	16569	American Restaurant	Hotel	Seafood Restaurant	Spa	New American Restaurant	Italian Restaurant	Coffee Shop	Southern / Soul Food Restaurant	Gay Bar	Indian Restaurant
1	13411	Spa	Restaurant	Bakery	Vegetarian / Vegan Restaurant	Pizza Place	Bar	Middle Eastern Restaurant	Accessories Store	Park	Chinese Restaurant
3	8270	Women's Store	Steakhouse	Boutique	Hotel	Italian Restaurant	Coffee Shop	Furniture / Home Store	Kids Store	Accessories Store	Toy / Game Store
5	8030	Playground	Trail	Park	Zoo Exhibit	Furniture / Home Store	Doctor's Office	Electronics Store	Exhibit	Farmers Market	Fast Food Restaurant

#### Cluster 2

```
Atlanta_merged.loc[Atlanta_merged['Cluster Labels'] == 1, Atlanta_merged.columns[[1] + list(range(5, Atlanta_merged.
):
```

	Population (2010)	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
4	8033	Pool	Scenic Lookout	Furniture / Home Store	Dive Bar	Doctor's Office	Electronics Store	Exhibit	Farmers Market	Fast Food Restaurant	Food Court

#### Cluster 3

```
Atlanta_merged.loc[Atlanta_merged['Cluster Labels'] == 2, Atlanta_merged.columns[[1] + list(range(5, Atlanta_merged.
:
```

	Population (2010)	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
2	10505	Italian Restaurant	Playground	Furniture / Home Store	Dive Bar	Doctor's Office	Electronics Store	Exhibit	Farmers Market	Fast Food Restaurant	Food Court

As seen in the visuals above, if we focus on restaurant data in three different clusters, the number of restaurants belonging to different cultures such as Italian and Chinese in the downtown of Atlanta is quite high. American restaurants, seafood restaurants and steakhouses are among the most preferred ones. In the cluster 1 table, although fast food restaurants are in the top ten on the most common venues list, they are not among the top three most common places.

## 5. Discussion

Atlanta is a city developed in terms of art, culture and food and beverage tourism. Opening a restaurant in such a city can be advantageous as well as risky. Since there are many people from different cultures living in this city, opening a place that is open to innovations may attract the attention of customers. If I have to interpret the concrete data I worked on in my project, fast food restaurants in the downtown of Atlanta are less than other restaurants. For this reason, opening a new fast-food restaurant in the neighborhoods close to downtown may be logical in terms of investment, but it should be kept in mind that the rent of the venue here is high.

## 6. Conclusion

In this project, I analyzed the relationship between the population and restaurant selection in the most crowded neighborhoods of Atlanta. I created a data frame that contains population, latitude and longitude information for each neighborhood by cleaning data. I built the clustering method using the k mean algorithm to predict the most common venue in Atlanta. I did visualization using the folium library to support my predictions. Although opening a new place or restaurant in a big city such as Atlanta depends on many sociological

and economic factors, I believe that I have reached the most accurate result with the available data. By using the analysis methods, I have used, this project can be studied in more detail and extensively if more features are achieved regarding these neighborhoods.