# Coursera Capstone Project

This Jupyter Notebook was created for the Coursera Capstone Project for the Data Science Professional Certification.

## Introduction

### Location recommendation for an Asian Restaurants in the Charleston (South Carolina) Area

For this project, I had the idea to do some analysis in regards to the presence of what Asian Restaurants in Charleston, SC Area. My wife and I really enjoy Asian cuisine, but we usually have a hard time finding a really good Asian restaurant close to our area.

Finding an Asian restaurant gets even more difficult, when we want to narrow down to a specific kind of Asian cuisine, i.e. Korean food.

As one might know, Charleston Areas are big and sometimes we need to drive a fairly amount of miles to our favorites restaurants. Therefore, I had the idea of creating this project (which is also similar to one of the ideas suggested in the instructions):

#### "Which Charleston Area/Location would it be recommended to open an Asian Restaurant?"

My object with this report is to provide information in regards the Asian Restaurant presence in the Charleston Area. By the end of this report, I will provide enough insight for someone who is researching for a location to open an Asian Restaurant in the area

Note:   
For the scope of this project, I will narrow down to East (Chinese, Japanese, Korean, ...) and Southeast (Thai, Vietnamese, Filipino,...) cuisines.

Kind reminder that Asian cuisine can also be South Asian (Indian, Pakistani, Bangladeshi, ...), West Asian (Arab, Turkish, Mesopotamian, ...) and other regions, but this is not the focus for this analysis.

## Data

A) In order to identify all Charleston Areas, I will be using the charleston.com web site as a reference of all Areas located in Charleston (<https://www.charleston.com/areas>). The "Areas", in this case, could represent a city, a town or even a neighborhood.

B) Once the Areas are identified, I will be using the geopy.geocoders in order to find the main coordinate (latitude and longitude) of each area.

C) Later, I will use the Foursquare API to extract Food Venues around each Area.

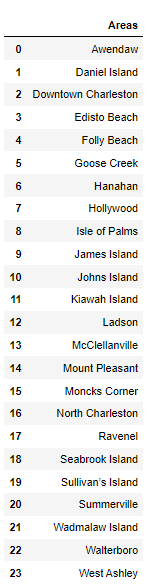
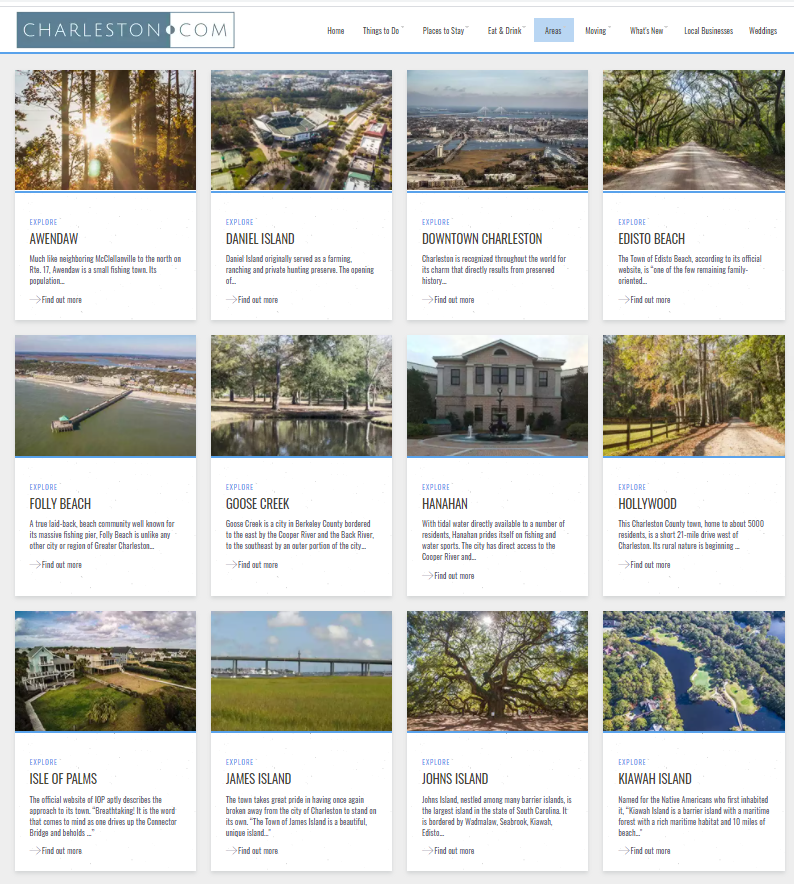
D) Finally, an analysis will be made with all the food locations with the objective of answering the above question.

### A) Charleston areas

As mentioned above, we will use the web site <https://www.charleston.com/areas> as a reference for the Charleston Areas.

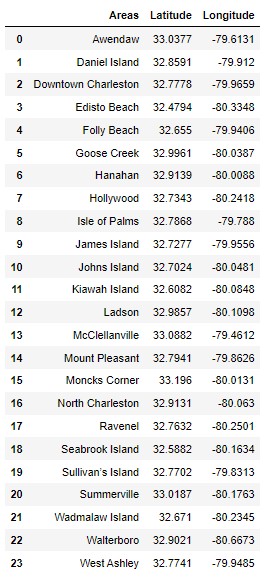
By opening the website and *inspecting each element*, it is possible to identify in which *section* the Charleston Areas are listed. Once the *section* is located, we can iterate through all *a* elements and create a list of all areas.

We can conclude that Charleston have 24 different Areas.

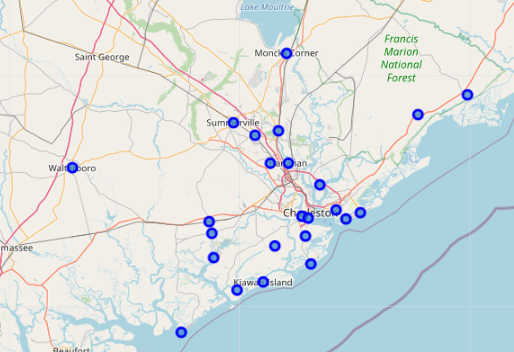
 

### B) Coordinates

Now that we have all Charleston Areas, we will identify the Coordinates (Latitude and Longitude) for each location using geopy.geocoders.



With Charleston Areas and their respective coordinates, we can also try to plot these coordinates in a map. On the below map, we can identify all the 24 Areas from Charleston.



### C) Foursquare API

The next step is to apply all the identified coordinates into Foursquare API in order to gather Food Venues.

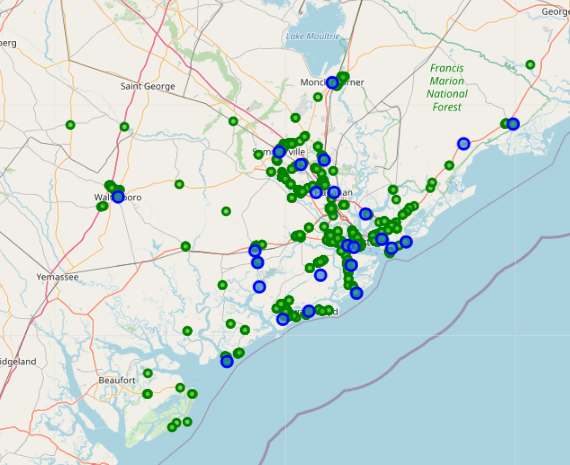
We will create a function, which explore each Charleston Area. The function will create a URL for Foursquare, which get Venues under the section "food" within a radius of 25,000 meters (~15 miles).

We will send the GET request and examine the results.



### D) Final data collection

With this, we complete the data collection. which can be presented in the below map.



## Methodology and Results

Now that we were able to identify the Food Venue locations in the Charleston Area, the first step that I would like to take is to cluster all the venues and identify the regions with higher density of restaurants.  
Depending on the business strategy, new business owners might want to open a restaurant close to other food locations, which naturally attract many people (reducing on the Marketing cost). On the other hand, some entrepreneurs might prefer locations far away from other restaurants.

In order to create the aforementioned clusters, I will apply the DBSCAN method.  
With this method, we will be able to create arbitrary shape clusters and the Density-based clustering is perfect to locate regions with high density vs. low density.

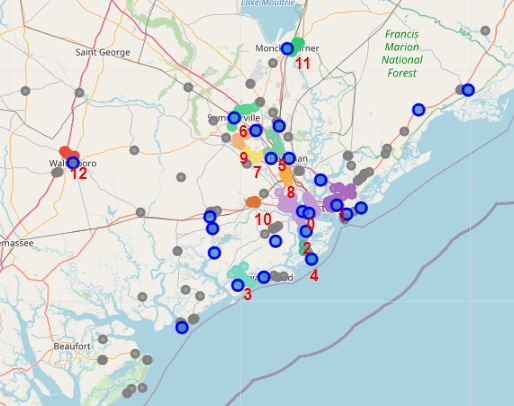
Once these clusters are identified, we can now segment which clusters have a higher or lower Asians cuisine presence.  
For this step, we will work on identifying what is considered "Asian cuisine" and understand which clusters have more or less Asians restaurants.

### DBSCAN

The Clustering will be based on the venue locations, i.e. Latitude and Longitude.

DBSCAN from sklearn library can runs DBSCAN clustering from vector array or distance matrix. In our case, we pass it the Numpy array Clus\_data and Set to find core samples of high density and expands clusters from them.

In total, according to the DBSCAN methodology, we were able to identify 13 different clusters, which can be visualized on the below map.

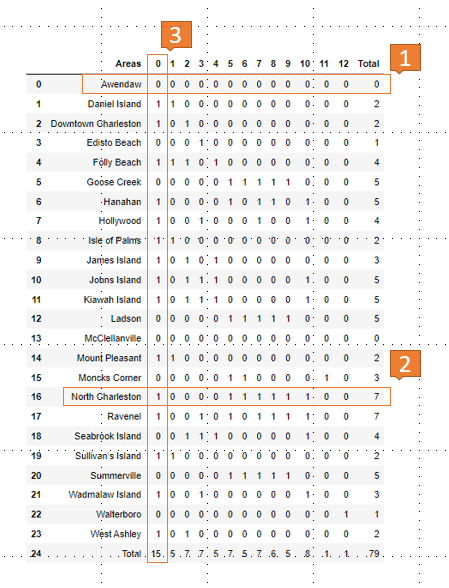


There are some significant number of outliers (marked in gray), but we were still able to identify 13 different clusters (from "0" to "13").

### Areas with higher density of venues

With some data manipulation we can conclude the following.

1. Some Charleston Areas do not have any cluster (Awendaw and McClellanville). These are areas that have food venues, but not a significant number that would create a cluster.
2. On the other hand, we also have the North Charleston area which is close to 7 different Clusters.
3. Finally, the Cluster Number 0 is the cluster which impacts most areas. This Cluster is mostly composed of restaurants from Downton Charleston. Since this cluster is located "in the center" of Charleston, it is natural to expect that multiple areas will be impacted by this cluster.



### Identifying Asian Restaurants Categories

In total, we were able to identify 48 different Categories of Food Venues.

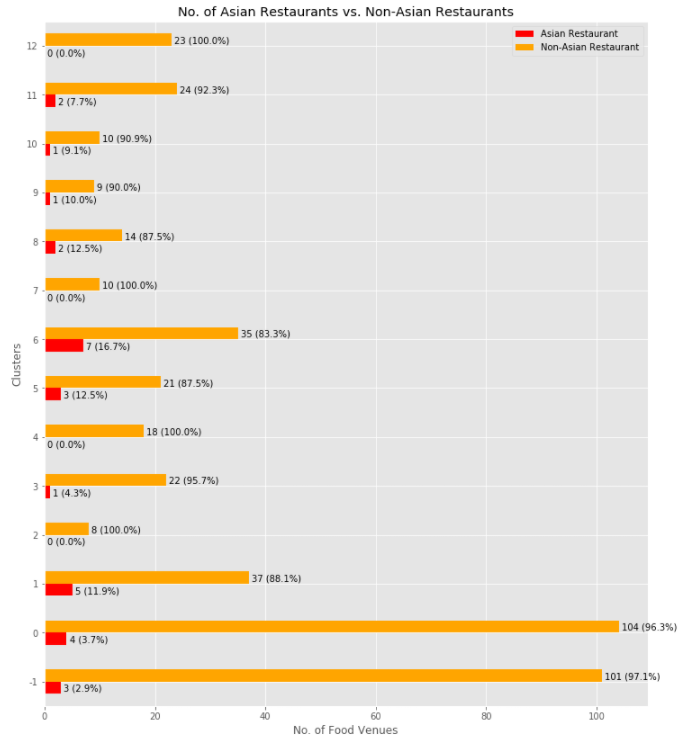
However, we only found the below Asian Cuisines:

* Asian Restaurant
* Chinese Restaurant
* Japanese Restaurant
* Sushi Restaurant
* Thai Restaurant
* Vietnamese Restaurant

### Representation of Asian Restaurants in each Clusters

Finally, we can now calculate how many Asian and Non-Asian restaurant we have in each cluster and calculate the representation (%) of Asian Restaurants among all Food Venues in each clusters from the Charleston Area.

* Overall, only ~6% of the Food Venues identified by the code (using the Foursquare API) might be considered Asian Restaurants;
* Most clusters do not have more than 10% representation of Asian restaurants and there are even cases that we found Zero Asian Restaurants in a Cluster (No. 2-'James Island', 4-'Folly Beach', 7-'Between North Charleston and Ladson' and 12-'Walterboro')
* The only cluster with significant Asian representation (above 10%) are the Clusters 1-'Mount Pleasant', 5-'Between North Charleston and Goose Greek' and 6-'Summerville'.
* Finally, the Clusters 1, 5 and 6 not only have a high presence of Asian Restaurants, but it also have a significant number of Food Venues overall (1:108 venues, 5:24 venues and 6:42 venues). The exceptions, would be Clusters 1 and 11, which have 42 and 26 venues, respectively, but not a high present of Asian Restaurants.



## Discussion

Based on the above analysis, one could find or prefer the following conclusions for "Which Charleston Area would it be recommended to open an Asian Restaurant?":

### A) Distant from all clusters

Some restaurant owners would prefer to open a new restaurant away from all Clusters. On the above analysis, we identified 13 different Clusters which should be avoided in this case. Notice that 104 of 465 Food Venus are not located in a cluster, this more than 22% of all Food Venues. In other words, it is common for a restaurant in the Charleston area to be isolated or far away from a Cluster.

### B) In a cluster with Zero presence in Asian Restaurants

We were able to identify four Clusters without any Asian Cuisine presence.

**No. 2** (James Island Area) and **No. 4** (Folly Beach Area) are fairly away from Downtown Charleston (avoiding the competition from that area). In addition to that, since these are clusters, a new owner could also take advantage of the cluster, since people/tourists would be already visiting the areas looking for Food Venues.

**No. 12** (Walterboro) is another Cluster without any Asian Cuisine, but this area is very isolated to all other Charleston Areas. In this case, it would not take advantage of the tourism from the region. On the other hand, there is a whole population from this City/Area that are not close to any Asian Restaurants.

Finally, **No. 7** (Cluster between North Charleston Area and Ladson Area) does not have any Asian Restaurants, but it is surrounded by other Areas with Asian Cuisine. If the objective is to find a cluster with zero presence of Asian Restaurants, this would not be a good choice.

### C) In a cluster close to multiple areas

Some entrepreneurs have high confidence on their product and, in this case, being located near to most areas would be a great choice. Regardless of the competition, the idea is to be in the middle of everything.  
For this case, two regions might be interesting:

Cluster **No. 0** (Downton Charleston), this is the center/heart of Charleston. It is a highly populated area and it attract a lot of tourist all year around. Other Clusters surround this Cluster. Furthermore, this Cluster is also close by different areas such as James Island, West Ashley, North Charleston, Daniel Island and Mount Pleasant.

In the North Charleston Area, this area is surrounded by different Cluster (5, 6, 7, 8 and 9) and very close to other areas such as Hanahan, Goose Greak, Ladson and Summerville. This is an area with a lot of growth and many opportunities.

### D) A unexplored "sub-type" of Asian Cuisine

Surprisingly, the code along with the Foursquare API did not identify many different types of Asian Cuisine. This is a sign that there is not much variety of Asian Cuisine in the Charleston Area. On the other hand, perhaps a different API or different parameter could identify more Asians restaurants.  
Either way, the analysis shown very little representation of Asian Cuisine. This show potential for exploring this kind of cuisine on the Charleston Areas.  
Some of the types of popular Asian food that were not identified are:

* Korean cuisine
* Singaporean cuisine
* Filipino cuisine

In addition, although this was not parts of the scope of the project, we were not able to identify, for example:

* Indian cuisine
* Arab cuisine
* Turkish cuisine

## Conclusion

With this Report/Project, one can conclude that there is potential for opening an Asian Restaurant in the Charleston Area.

There are multiple **Locations** that can be explored or avoided (depending on the business strategy), but overall the Asian Cuisine is not highly present in the Charleston Area (overall only ~6% of the restaurants are Asian). Furthermore, some "sub-types" are little to none present in the area (i.e. Korean Food), which is the biggest advantage that a new restaurant owner could explore.

On a personal note, opening a Korean Restaurant (known to be very popular in other cities/locations) in the North Charleston or Downtown Charleston area is a great opportunity. These two areas are highly populated and centered among most of the Charleston Areas. Moreover, Korean food can stand out among other Asian cuisines for presenting a very different type of food (i.e. Korean Barbecue). Furthermore, the Korean Culture has been gaining popularity in the USA.

Naturally, this is only one aspect of analysis that an investor would take in consideration. Many indicators were not approached such as population/demographic target, infrastructure cost, staffing requirements, marketing/advertisement requirements, etc. However, this is not the scope of this project. The only aspect that we are targeting is location.