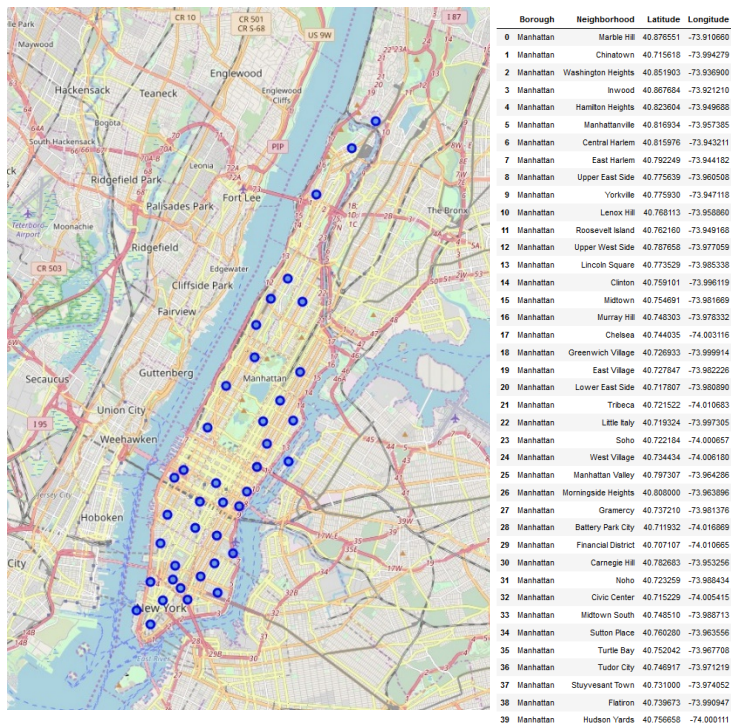


## Introduction

When opening any type of restaurant, we need to look at the location and the needs of the people and the appetite present. Thus, the axiom of location, location, location is vital for any business in the food service industry. We will specifically be looking at the neighborhoods of New York and identifying places for an ideal location for a new fried chicken restaurant in Manhattan. We will be taking into account the density of current fried chicken places and providing some insight as to what would be a region that would allow for a new place to flourish. I think that anyone looking at opening a new restaurant would benefit from this type of analysis. While there are many factors that one would have to consider, this analysis would be focusing on the idea that it would be advantageous to open a place that is near but not in the identical location that have a large number of fried chicken places.

We will take a look at Manhattan specifically and break down which neighborhoods would be the ideal place to start a fried chicken joint.



## Data

The data set used for this analysis come from several sources. The location data is from Wikipedia which allow us to have the exact set of neighborhood names, latitudes and longitude for each neighborhood. We also have complied data from Foursquare which provides venue information in each neighborhood. In this case we specifically are looking at Fried Chicken venues as well as college residence halls. The focus is that we can identify clusters of fried chicken places and hypothesize that those areas have

people who want to have fried chicken and then overlay where the colleges are as a means of providing them with late night fried chicken.

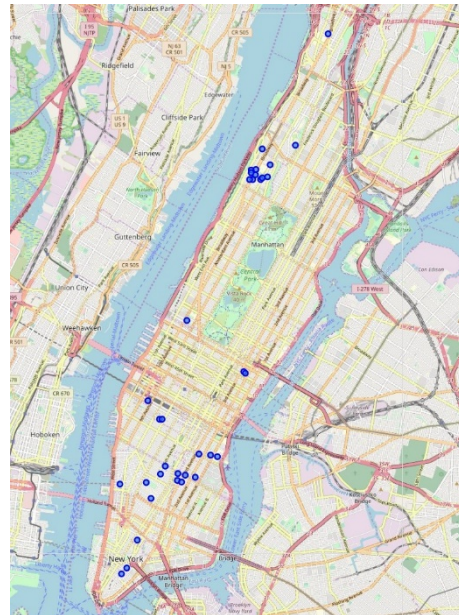
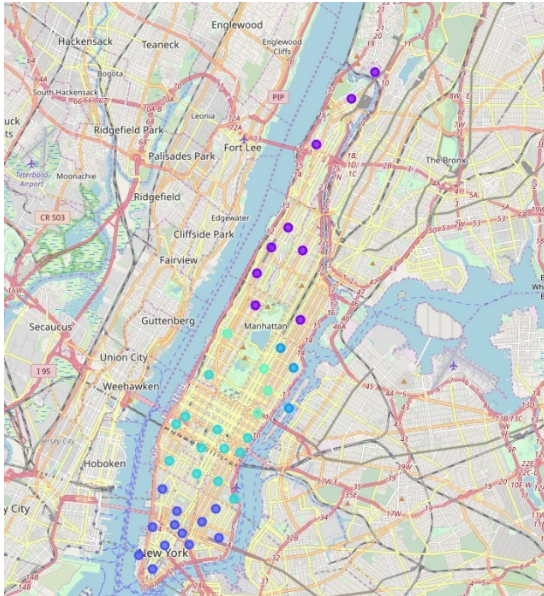
The following are the Python packages I used:

1. Pandas - Library for Data Analysis
2. NumPy – Library to handle data in a vectorized manner
3. JSON – Library to handle JSON files
4. Geopy – To retrieve Location Data
5. Requests – Library to handle http requests
6. Matplotlib – Python Plotting Module
7. Sklearn – Python machine learning Library
8. Folium – Map rendering Library

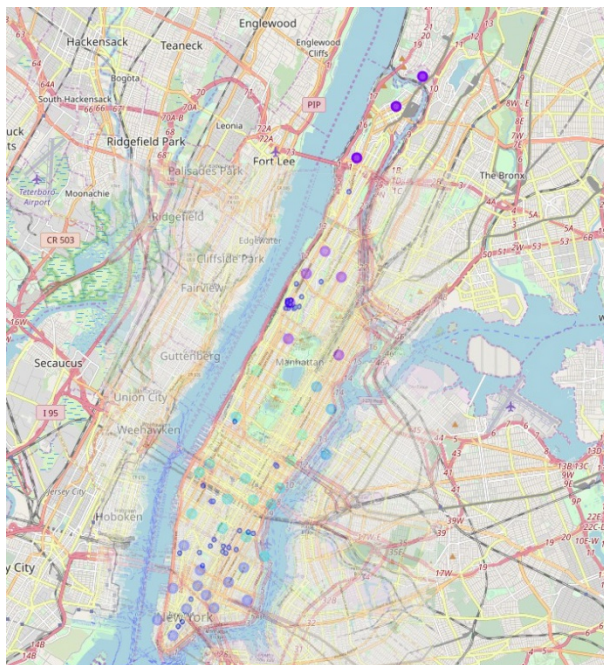
## Methodology

1. Download the New York Neighborhood location information and put that into a panda data frame with GeoPy and Nomiantim library which converts an address into latitude and longitude.
2. Ping the Foursquare API using zip codes of the New York city neighborhoods to match the location information (Latitude and Longitude) with venue data that Foursquare has.
3. Foursquare API search feature localizing it to just 600 meters and focused on Fried Chicken Venues and college residents halls. The Foursquare api allows you to specific what venues with a “categoryId”. This ID can be found on the Foursquare API documentation.
4. Graph the locations of the retrieved fried chicken venues with Folium and Matplotlib. We utilized Folium to display the map as well as showcasing dots of blue of where each neighborhood is located.
5. Using the python scientific libraries: Pandas, NumPy and Scikit-learn. We can analyze the data we have collected. First we do a “one-hot encoding” which is applied to the data we have pulled from the api and groups the data for analysis. Once the data has been grouped we can get a frequency of how often each fried chicken place appears in each neighborhood.
6. Utilizing the K means clustering abilities of Sklearn we can see neighborhoods in Manhattan that would be ideal for opening a new Fried Chicken venue. We utilized kmeans as a way to figure out what is the best kvalue to use and how any clusters should be utilized and we discover that 5 clusters are what are needed.
7. Now that we know where the current Fried Chicken places are, we look at overlaying college residence halls and identity areas that don’t have a solid representation in our clusters to pinpoint an area that would be beneficial to starting a fried chicken place.

## Results



With our K means we were able to identify 5 groups that our fried chicken joints were close together to set. This brought out some pretty interesting clusters and I believe that if we were to overlay housing data, we would see the pricing and rental reasons for the clusters that formed. But that type of analysis would be reserved for another project. In our specific case we can see that the college housing are neatly tied into some of where our fried chicken clusters are presently located. Overall the purple clusters as the region between the blue and green



clusters had the best options for where to apply and put down a new fried chicken venue. In addition with the overlay of college dorms we can clearly see what areas have already shown to have great value and are the place that other shops laid down roots. The addition of a new fried chicken places is perfectly warranted and execution by the staff and management would be the metric for success as I believe the location is more than prime.

## Discussion

My observations from the data analysis that I have completed is we do see some pretty clear correlations between the locations of the college housing and then areas of where fried chicken places are present. I am wondering if we to include more 24 hour food ventures into our data if we would see a similar style of clustering of restaurants.

We do see a strong concentration in Morningsides Heights and the East Village of the k means clusters of fried chicken joints as well as a strong concentration of dorms. I believe that if we were to take a look at other fast food options we would see a strong presence for them as well. My recommendation to any proprietor that wants to open up a new fried chicken joint would be best served to open one up in either of those hot spots.

It should be stated that this type of analysis while cursory is useful because it at least provides an overview of what is happening and it's not giving any type of causation it gives one a bit of time to pause and think about the implications of how to best choose something can be done easily with information that is freely available on line.

## Conclusion

There is a lot that can be done to analyze the information that exists on the web. We can with more information, time and with a specific client in mind; it would be possible to identify the best location for their needs whether it's a restaurant or an office space. There is also always room for improvement as there can be better model training, better training data as well as more update to date information.