New York Kitchen Unit - Work from Home Food Delivery Kitchen

Kevin Masterson

1. Introduction - Business Problem

1.1 Background: Professionals with culinary and hospitality experience have found themselves out of work recently due to the pandemic and are looking for opportunities. An investment firm are taking opportunity in adversity and have commissioned this report. The aim of this report is to source a new venue to open a kitchen to meet the needs of a changed food market in the last 12 months. We have identified a few key aspects which are crucial to success in the current climate (2020 - 2021) and modern working from home routines of many New Yorkers.

1.2 Problem: Breaking into the "Working from home" food market.

Sourcing a location for a new food kitchen which is close enough to residential areas where the vast majority of office workers are now working from home. The location must be accessible by road for delivery drivers and easily reached by customers for pick up. Hence a kitchen in an industrial or remote area is not suitable. However, since we aim primarily to be an online service, a main street unit is not necessary and could save some costs with a side street unit. Our focus will be on finding a location which is non central and more residential in the New York City area. Since we need to cater for delivery drivers, we do not necessarily need to preference a location near a subway which might be higher rent on a unit.

1.3 Interest

New York has a reputation of being one of the foodie capitols of the world with a mix of styles and backgrounds from across the globe. Since sitting in has become a health issue as of late, creating more space for culinary experimenting and utilising support businesses like Uber Eats to serve our customers with a limited distance range. We expect professionals in the food industry will be eager to collaborate on this type of project as we prioritise the kitchen over the dining requirements allowing for more freedom (within budget). Professionals need to arrive ready with an idea to hit the homes of residents in New York twice daily, lunch and dinner, delivery and pick up only service.

A quick Google search tells us that Statin Island is a more laid-back leafy suburb with significantly lower population than the other boroughs, while Manhattan is the economic centre of the city. Thus, we will rule out these two boroughs, due to low population and too business central, and choose two of the remaining three.

This analysis will compare Queens and The Bronx neighbourhood against each other for an area to open a food kitchen for lunch and dinners. The food kitchen will be pick up or delivery only due to restrictions during a pandemic and will likely be closer to residential areas where people are work from home. Though since New York has a lot of high-rise apartments this could still be very central - we shall see what our findings tell us.

2. Data acquisition and cleaning

2.1 Neighbourhood data

The data for New York Neighbourhoods was sourced from a pre-made json file. Once the data was loaded into a Pandas dataframe we were able to find the coordinates of each neighbourhood and determine which areas we will look at for the analyses in the next stage of the project. This data has been prepared for this task whereby we could select the attributes we wished to use. In our selection, we chose the borough, neighbourhood and the latitude and longitude coordinates as the base for building our data tables. This data is relevant for plotting the map of New York using the Folium library in python. The geocoder library in python has been used to obtain latitude and longitude data for various neighbourhoods in the selected boroughs of New York.

We were then able to use geocoder and folium map to transmute our data into a map of New York city to help visualise our data accumulated so far. Have a visual helps us to see the layout of the city and make pre analysis assumptions of where we may like to be situated in the heart of a borough or near a coastal area for example. The data will tell use more.

2.2 Venue Data

The venue data has been extracted using the Foursquare API. This data contains venue recommendations for all neighbourhoods in New York and is used to study the popular venues of different neighbourhoods. Using

FourSquare API, we were able to match the neighbourhood names and coordinates with the top 100 venues in a neighbourhood to determine the type of area we were selecting. This is a valuable data source to understand the lay of the city's venues from a distance and automates a vast amount of research for us. We were able to pair the FourSquare API data with k-means clustering algorithm to determine which neighbourhoods had similar attributes and could be classified together for our analysis. This selection of clustering helped us answer some pre analysis assumptions we made in when the folium map was created.

New York neighbourhood residential data is available online:
https://furmancenter.org/stateofthecity/view/citywide-and-borough-data#the-bronx

The data is available to download at: https://furmancenter.org/coredata/userguide/data-downloads

3.0 Methodology

We were able to utilise the New York census population data to help us determine which of the boroughs best fits the population description we are looking for in our target market (i.e., working from home). This is a valuable exercise where we can find solid facts which are attributable to our target market and help steer our decision making in terms of directing our search for kitchen unit in a thriving area of the city. There were many data points collected in the census, too many to recall in this report. Hence, we selected a few graphs were created to build the picture we are looking for.

The census data was successful in helping us narrow down further which borough we end up selecting, alongside the FourSquare API data and k-means clustering algorithm which further points to narrower area for our search. These data sources are the positive fuel and also the limitations of this analysis. This is an important fact to keep in mind when performing any analytical review.

3.1 Exploratory Data Analysis

3.1.1 Population

New York census data found that Queens was much more populated than the Bronx. Both Boroughs have grown along with the city, on average, over the last 20 year. Interesting to note there was a large visible drop in population in Queens from 2008 – 2010, with only a minor dip in the Bronx. This could be related to the crash of 2007/8, as we shall see in further graphics, Queens is an expensive area to reside in while the Bronx was not affected as much but the crash. This has created a clue which will help us find our target market or could be representative of the potential crash expected to come as the global economy may contract further when it comes out of the pandemic.

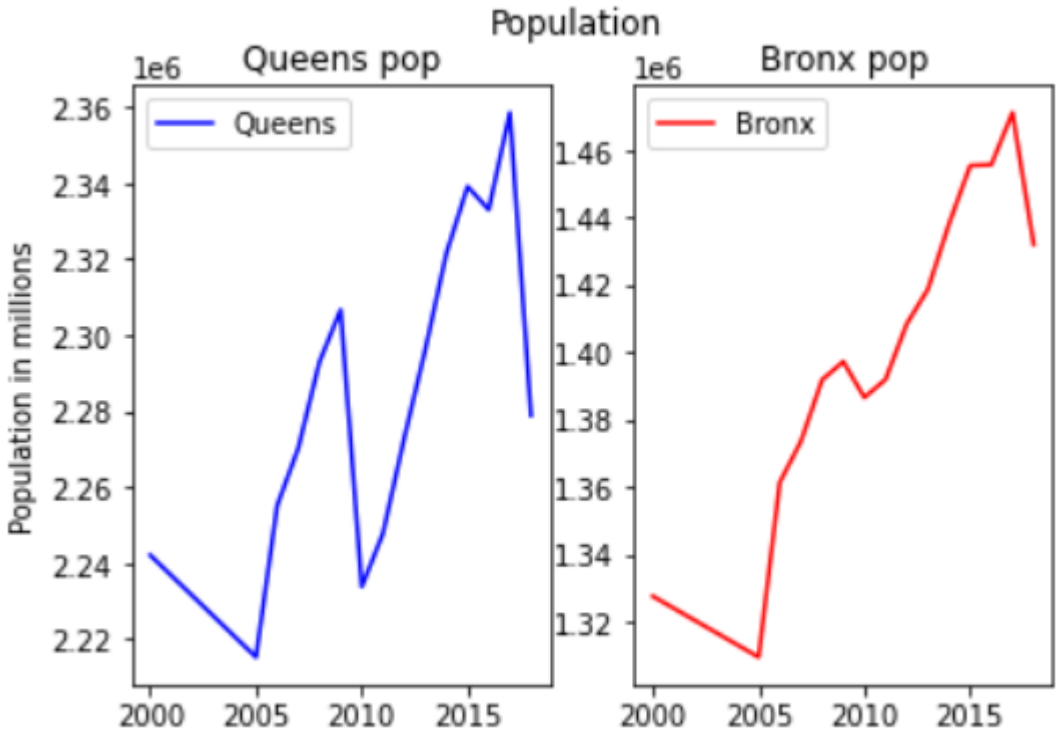

Figure 1 – population data

3.1.2 Income Data

We are seeking professionals who would usually be in the office and are now working from home due to the risk of spreading or contracting Covid-19. In order for to have a successful business venture we should be situated where these professionals are located and still willing to spend. The census income data is, unfortunately, only representative of the past and does not reflect the more recent employment market in each borough. As we have seen above, there was a decline of Queens residents of approx. 80 thousand in the last economic crash.

From the below data, we can see the median base level of Queens income is higher. On average the median household's total income of all members of the household aged 15 years or older. This tells us that Queens may be a bit more upmarket for our target market, with more residents over all and on average a higher income in the market. The home owned income is not too different in either borough, however the median of all households is enough to help us make an educated decision that Queens is a preferred are for our kitchen unit.

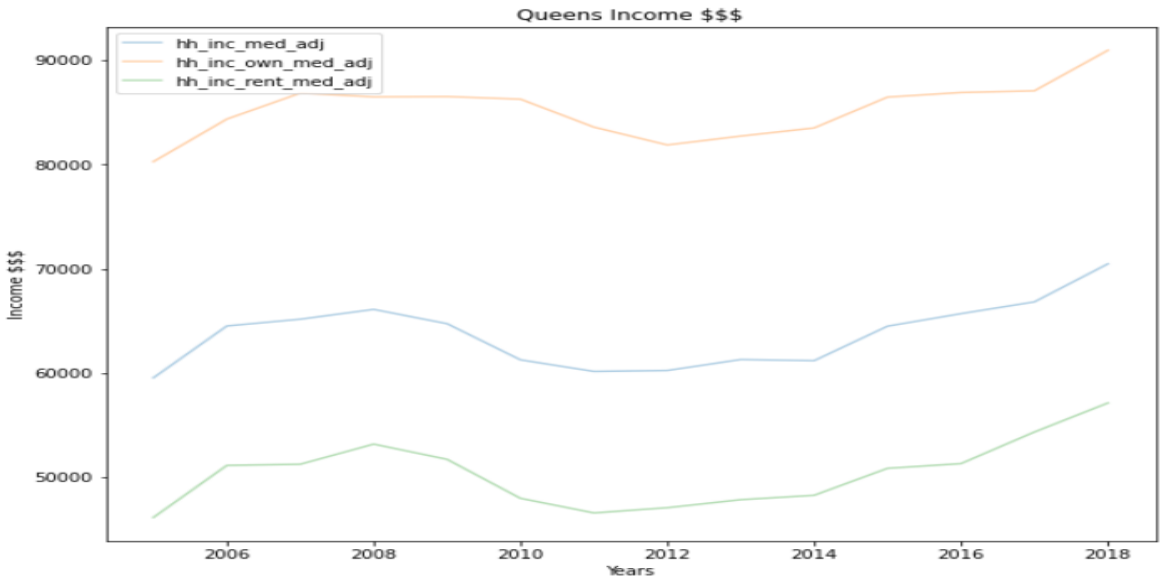

Figure 2 – Queens Income Data

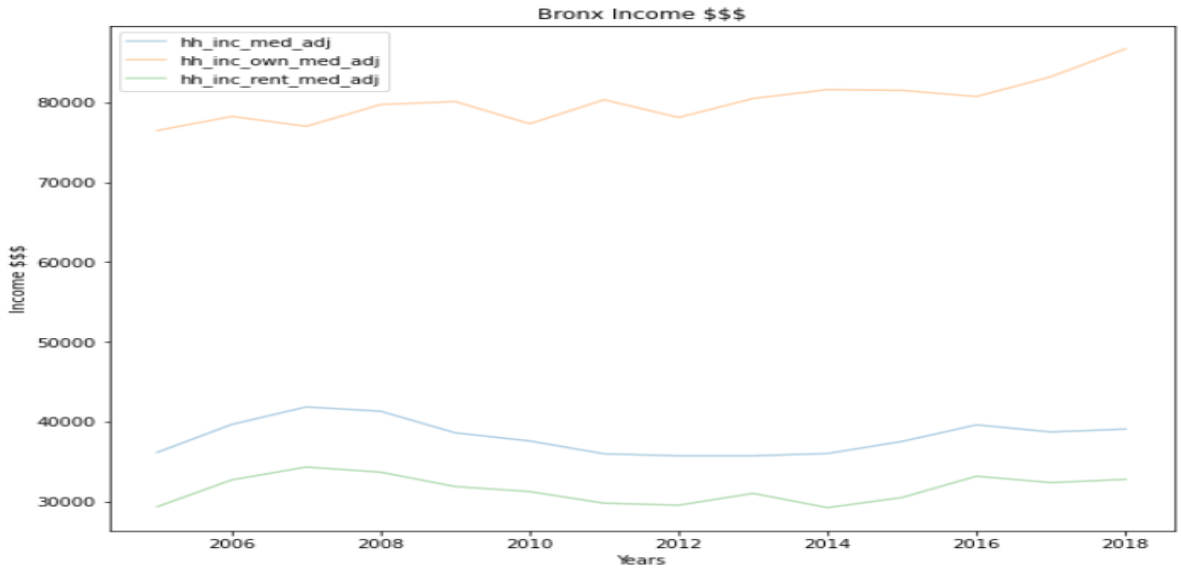

Figure 3 – The Bronx Income Data

hh_inc_med_adj: The median household's total income of all members of the household aged 15 years or older.

hh_inc_own_med_adj: The median owner-occupied household's total income of all members of the household aged 15 years or older.

hh_inc_rent_med_adj: The median renter household's total income of all members of the household aged 15 years or older.

3.1.3 Educated Population Data

One important attribute of our working professionals which is available in the census data is education levels. New York is the financial capitol of America and will attract high achievers and high earners. Hence this data attribute has provided a good insight of the two boroughs. The data shows that Queens' college educated population is growing above the 30% mark as of 2018 in an upward trend compared with the downward trend of individuals who have not graduated high school. Interestingly the Bronx data shows the same trend of growing college education and decrease non-educated individuals, however their respective starting points are visibly clear the Bronx is playing catch up to Queens, and this is where we may assume our working professional will be located.

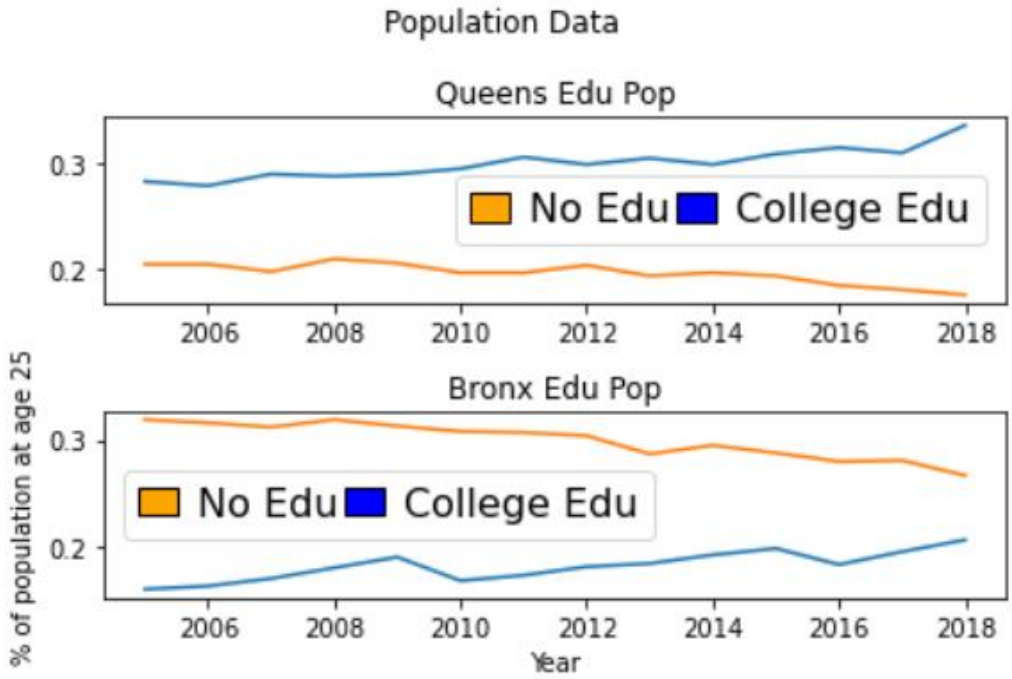

Figure 4 – Educated population data

% population over 25 who have obtained a degree

pop_edu_collp_pct: The percentage of the population aged 25 and older who have attained a bachelor's degree or higher.

pop_edu_nohs_pct: The percentage of the population aged 25 and older who have not graduated from high school or received a GED.

3.1.4 Rent Burden

Queens is a larger population, better educated and higher income area. How does this reflect on the housing economy in the area? The rent burden provides us with some reflection on New York as a city which is rather expensive. Up until this point, we have painted a picture of Queens being a rosy neighbourhood. The rent burden provides information that says the squeezed middle resides here. Across the board of low- and moderate-income renter households, Queens has higher rent burden. Young working professionals are not likely to be house owners yet and potentially make up a portion of our targeted working from home professionals.

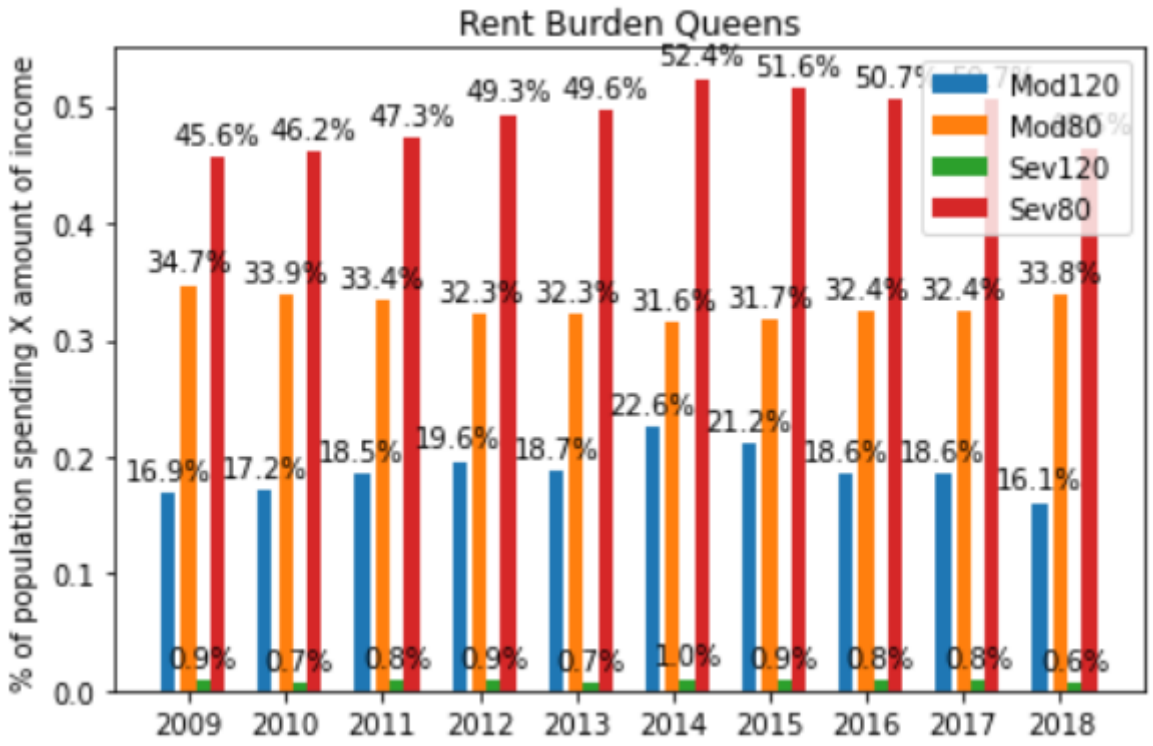

Figure 5– Queens rent burden

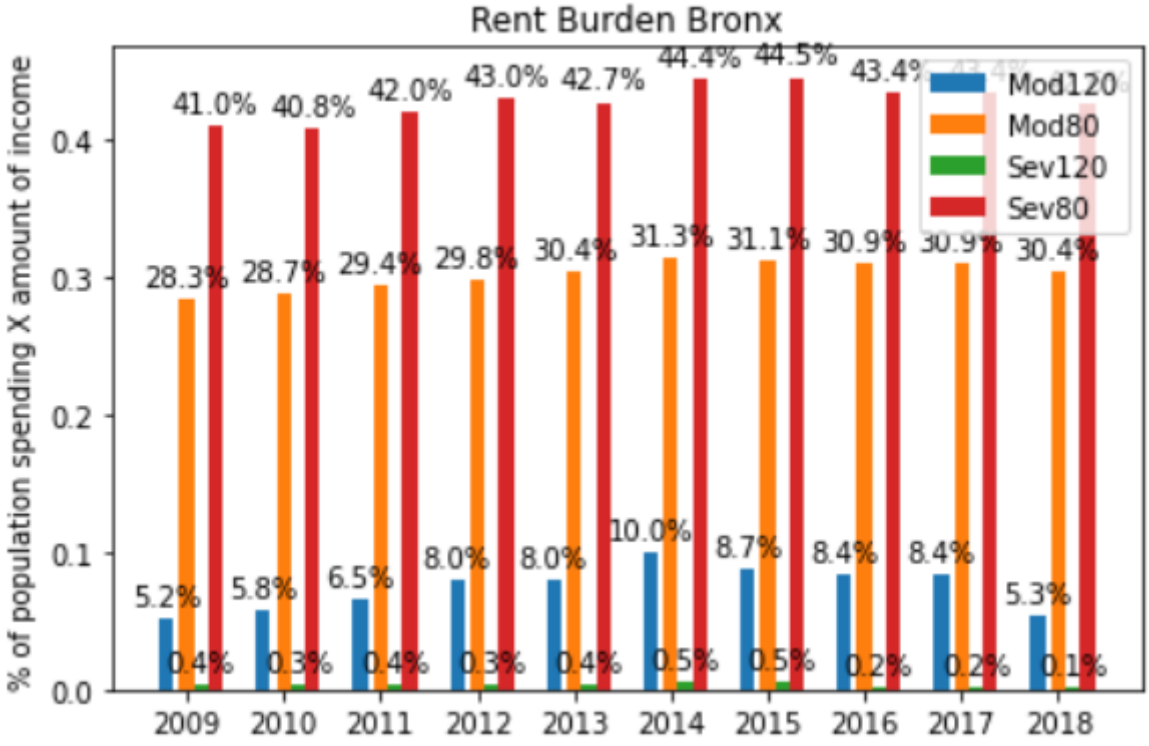

Figure 6 – The Bronx rent burden

rburden_mod_ami_81_120: The share of moderate-income renter households whose gross rent made up at least 30 percent of their monthly pre-tax income.

rburden_mod_ami_le80: The share of low-income renter households whose gross rent made up at least 30 percent of their monthly pre-tax income.

rburden_sev_ami_81_120: The share of moderate-income renter households whose gross rent made up at least 50 percent of their monthly pre-tax income.

rburden_sev_ami_le80: The share of low-income renter households whose gross rent made up at least 50 percent of their monthly pre-tax income.

3.2 FourSquare API Venue Data

With the census analysis complete a decision, for the final report, is to continue with Queens as the area we will review in FourSquare venue data. Additional FourSquare venue analysis of The Bronx can be found in the workings report. The venue data sourced from FourSquare API was successfully collected for top 100 venues in each neighbourhood. Provided the volume of data is large the use of k-means algorithm has been implemented in order to assess the similarity of the areas in each respective Borough.

3.2.1 Silhouette Score

This report will not describe in detail the workings of k-means clustering. What is relevant is knowing we used a silhouette score to determine how many clusters would be of benefit to fit out model. The below graphic shows that the optimal number of clusters in this for k-means clustering is 3. However, since we are not only trying to find a location but eliminate some areas this analysis used 5 clusters to show its efficacy in reducing the possible locations for our kitchen unit (as the report will show this is beneficial).

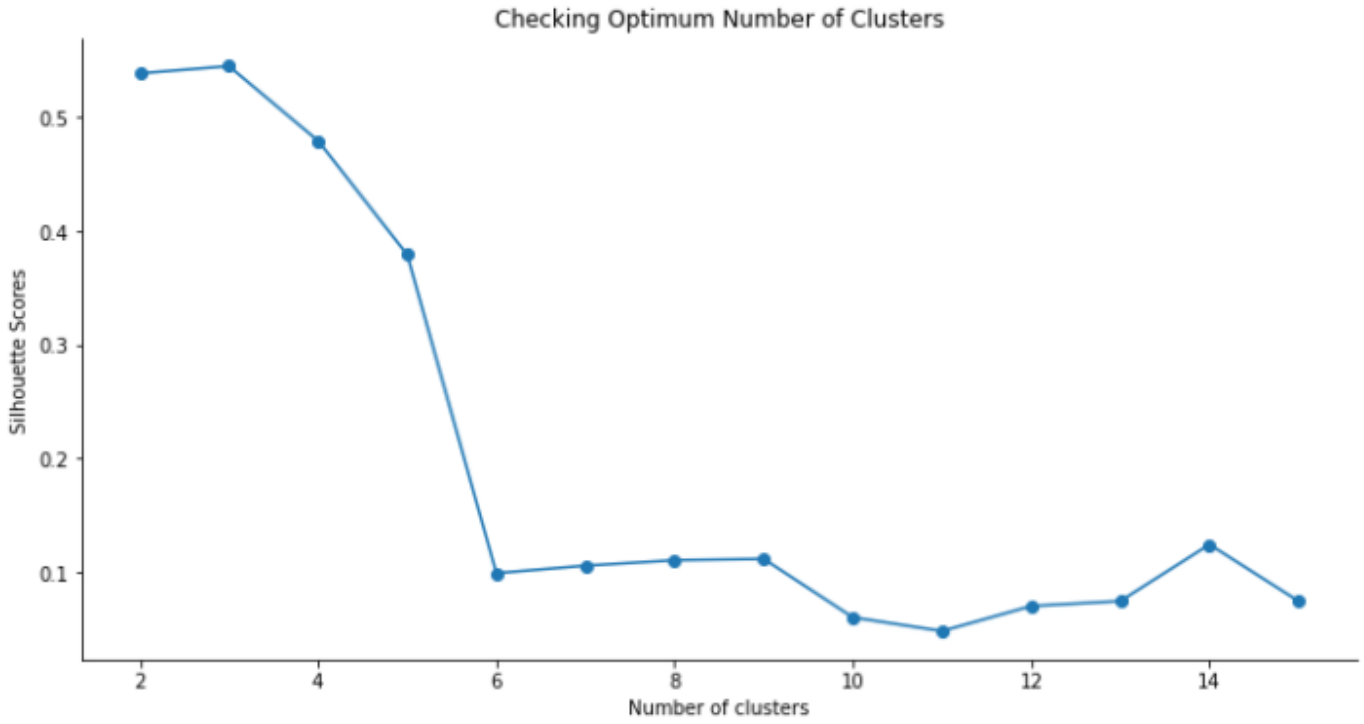

Figure 7 – Queens' silhouette score

3.2.2 k-means Clustering

	Neighbourhood	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	Astoria	Bar	Middle Eastern Restaurant	Hookah Bar	Greek Restaurant	Seafood Restaurant	Café	Deli / Bodega	Indian Restaurant	Bakery	Mediterranean Restaurant
1	Woodside	Grocery Store	Thai Restaurant	Bakery	Filipino Restaurant	Pizza Place	Latin American Restaurant	American Restaurant	Pub	Donut Shop	Chinese Restaurant
2	Jackson Heights	Latin American Restaurant	Peruvian Restaurant	South American Restaurant	Bakery	Mobile Phone Shop	Thai Restaurant	Mexican Restaurant	Grocery Store	Pharmacy	Pizza Place
3	Elmhurst	Thai Restaurant	Mexican Restaurant	Bubble Tea Shop	Vietnamese Restaurant	Chinese Restaurant	Gym / Fitness Center	Seafood Restaurant	Sushi Restaurant	Fast Food Restaurant	Malay Restaurant
4	Howard Beach	Italian Restaurant	Pharmacy	Chinese Restaurant	Sandwich Place	Fast Food Restaurant	Japanese Restaurant	Clothing Store	Gym	Mexican Restaurant	Seafood Restaurant

Table 1 – Queens cluster 3

The algorithm has clustered together a large part of Queens as the venue data classify them with similarity. The decision to include extra clustering's deemed beneficial to eliminate further areas of the borough from the investigation.

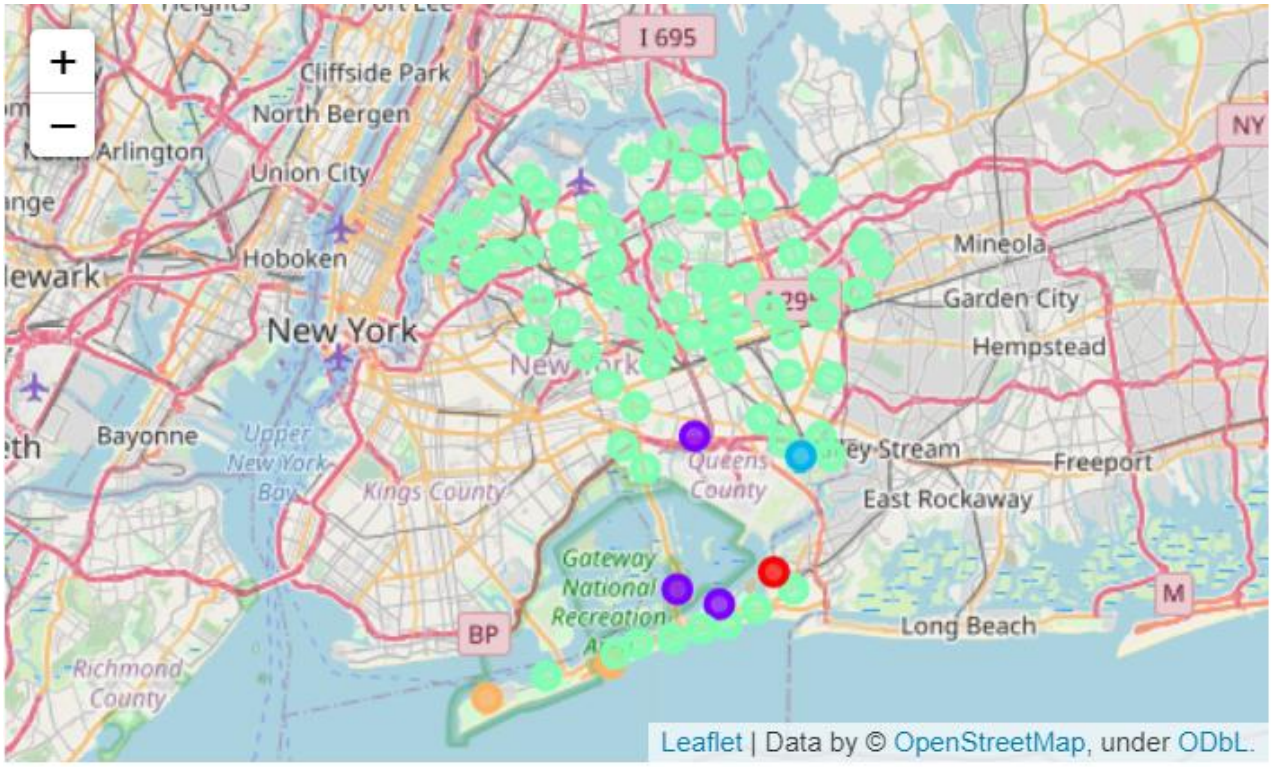

Figure 8 – Queens map of cluster locations

4. Results

From reviewing the clusters in Queens and The Bronx, the k-means clustering algorithm is telling us that both boroughs are fairly well distributed in terms of restaurants, cafes and supermarkets with cluster 3 in Queens and cluster 1 in The Bronx containing most neighbourhoods. Each of these clusters would be the ideal area to open a kitchen for the respective boroughs as there is already evidence of a food culture here from FourSquare API venue data. The other clusters are less suitable as we will discuss below.

What it does help us with is eliminating a number of clusters where the neighbourhoods most common venue is recreational or public service areas. Queens Clusters 0,1, 2 and 4 main venue is recreational, while the Bronx has cluster 0, 2, 3 and 4 are either public service or recreational areas. It might not appear like much on the map, but it is helping with eliminating some neighbourhoods for our kitchen unit.

Visibly from the Folium map, Cluster 3 in queens is widespread with the remaining clusters in a more coastal location and located close to each other. As Queens cluster 3 is our preferred located given:

1. Higher population
2. Higher income
3. Higher Educated
4. And well spread-out food culture as per the clustering results

Referring to the census data which directed our attention to Queens over the Bronx. It is beneficial to search for multiple data sources to help increase accuracy in business decision making.

5. Discussion

This analysis has proved insightful. Further analysis may include more clustering or classification model to determine what venue food types are more common if there was a need for a type of food in an area. As per this business analysis, we are to open the kitchen for professionals who want to enter the business and are not strictly concerned with this in our research as commissioned by the investment firm.

6. Conclusion

Queens Cluster 3 – central area away from the coast in a built-up area would be suitable location to find a space for a kitchen unit to open. This should provide our investors with prime area in a central location for the working from home food market to reach future customers in an evolving food market.