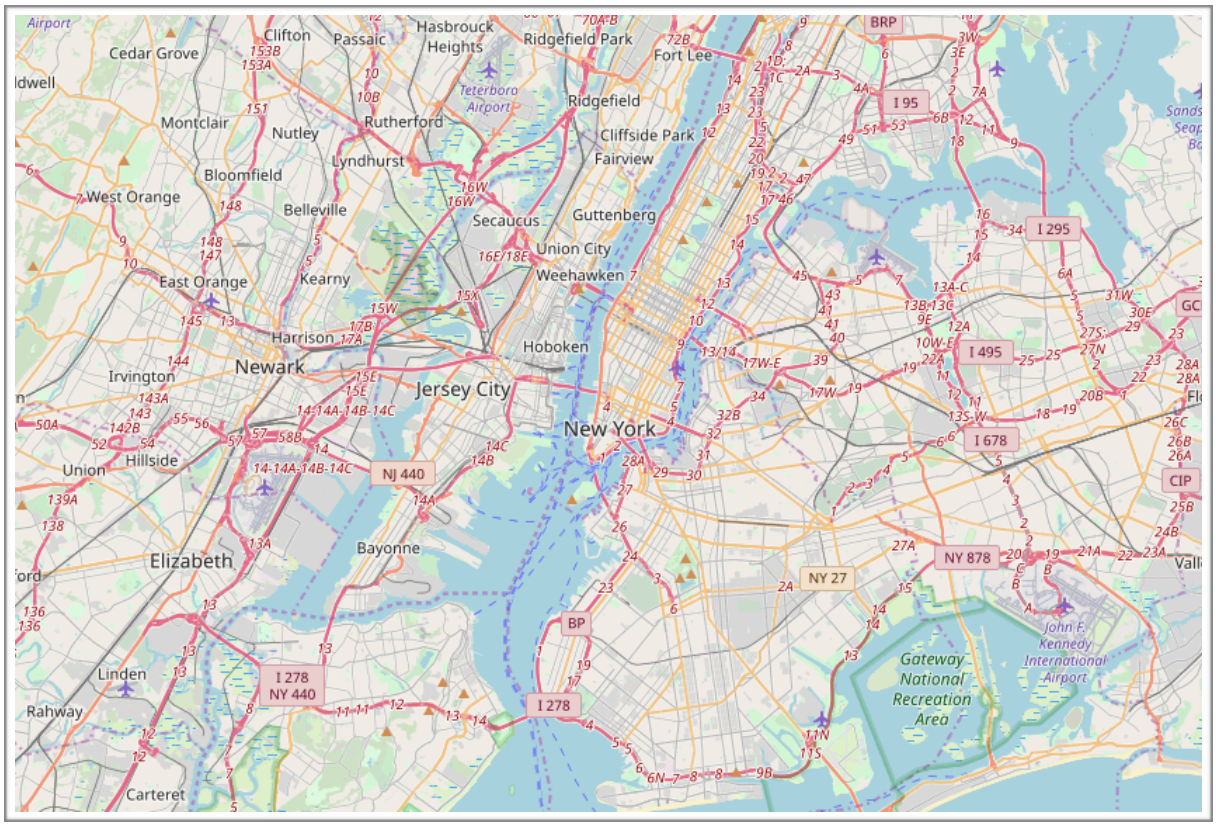
Capstone

The Battle of Neighbourhoods



To Determine Best Place to Setup a Restaurant With an Appropriate Theme

INTRODUCTION

As the most populated city in the United States, New York City is a great place to open a restaurant. According to the stats the city's numbers have reached a record high bringing an increase of more than 5% in the last few years to about 8.6 million.

With the growth in the population, the number of restaurants in NYC keeps growing as well. According to the National Restaurant Association, there were nearly 45,000 eating and drinking establishments, nearly 26,700 open restaurants in the city alone. Though the competition is high among New York City restaurants, there is great potential for success. These numbers also bring in challenges to the process of setting up a restaurant business.

Amongst others, the challenge is also to be able to setup the right stuff in the right regions with the right cuisines and also be able to survive and sustain.

We work under an assumption that we need to setup a restaurant keeping in mind the locality and also the expenses that would be incurred as the client does not have large sums of money to invest.

DATA

1. Demographics_of_New_York = https://en.wikipedia.org/wiki/Demographics_of_New_York_(state)
2. Cuisine_of_New_York_City = https://en.wikipedia.org/wiki/Cuisine_of_New_York_City
3. New_York_City = https://en.wikipedia.org/wiki/New_York_City
4. Culture_of_New_York_City = https://en.wikipedia.org/wiki/Culture_of_New_York_City
5. Regions_in_New_York_City = https://en.wikipedia.org/wiki/Neighborhoods_in_New_York_City
6. Markets = https://data.cityofnewyork.us/dataset/DOHMH-Farmers-Markets-and-Food-Boxes/8vwk-6iz2/data
7. Foursquare data will be used for neighbourhoods and cross-verification purposes based on the latitude and longitude of NYC.

DATA DESCRIPTION

1. Demographics_of_New_York, New_York_City : To understand the demographics of the city and understanding population densities.
2. Cuisine_of_New_York_City, Culture_of_New_York_City : To understand the variety of cuisines present given the diversity in population cultures and their tastes with regard to their food preferences.
3. Markets, Regions_in_New_York_City : To understand the strategic placement of resources with regard to raw materials for the restaurant and their correspondng neighbourhoods.

CHALLENGES

1. Determining the right location within the city to set up the restaurant.
2. Determining the prevalent cuisine in chosen areas.
3. Looking at the need for a delivery service to increase returns.
4. Probable Competitors
5. Accessibility to required raw materials and resources

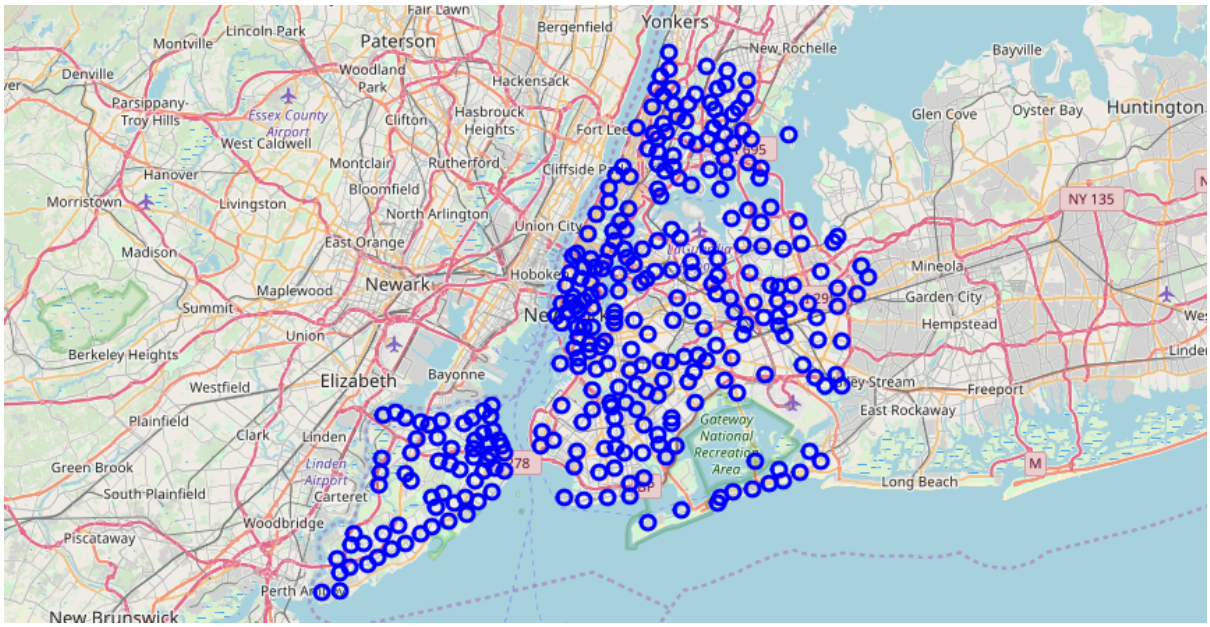
FACTORS TO STUDY

1. Population.
2. Population density and demographics.
3. Closeness to residential areas for delivery purposes.
4. Possible increase in profits due to food delivery and its trade off.
5. Trends indicating likelihood of turning towards a certain cuisine over a period time.
6. Looking for the X factor
7. We shall also examine a few other factors eventually

METHODLOGY

NYC has a total of 5 boroughs and 306 neighbourhoods. We have verified the same in our initial stages of EDA.

We firstly load and explore NYC data and data related to its demographics



Having a look at the demographics data we get the following snippet

Out[108]:

	Composition	2010	1990	1970	1940
0	White	44.0%	52.3%	76.6%	93.6%
1	—Non-Hispanic	33.3%	43.2%	62.9%	92.0%
2	Black or African American	25.5%	28.7%	21.1%	6.1%
3	Hispanic or Latino (of any race)	28.6%	24.4%	16.2%	1.6%
4	Asian	12.7%	7.0%	1.2%	–

This data gives us insight into the trends in population diversity and their increase/decrease over time.

The next set of data we worked on was on the Culture and Cuisine datasets

Working up on the culture of different boroughs in NYC we can get a lot of insight on the form of entertainment enjoyed, the heritage, etc. This can give us a peek into the themes that such regions appreciate and also a slight insight into the financial status of individual based on the expenses involved with such activities.

Let us look at the cuisines and cultures of every borough

Staten_Island Culture

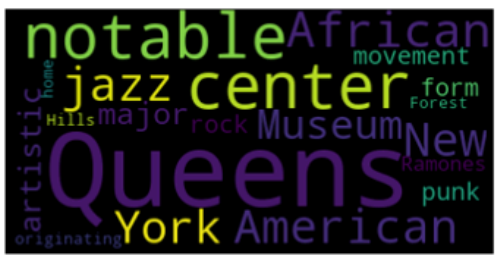
```
Staten_culture = open('stn.txt', 'r').read()
cloudc4 = WordCloud(
    background_color='black',
    max_words=20,
)
cloudc4.generate(Staten_culture)
plt.imshow(cloudc4, interpolation='spline16')
plt.axis('off')
fig = plt.figure()

fig.set_figwidth(50)
fig.set_figheight(50)

plt.show()
```

Staten_island

```
cloud4 = WordCloud(
    background_color='black',
    max_words=5,
)
cloud4.generate(State_island)
plt.imshow(cloud4, interpolation='spline16')
plt.axis('off')
fig = plt.figure()

fig.set_figwidth(50)
fig.set_figheight(50)

plt.show()
```

Queens Culture

```
Queens_culture = open('Qns.txt', 'r').read()
cloudc3 = WordCloud(
    background_color='black',
    max_words=20,
)
cloudc3.generate(Queens_culture)
plt.imshow(cloudc3, interpolation='spline16')
plt.axis('off')
fig = plt.figure()

fig.set_figwidth(50)
fig.set_figheight(50)

plt.show()
```



Queens

```
cloud3 = WordCloud(
    background_color='black',
    max_words=8,
)
cloud3.generate(Queens)
plt.imshow(cloud3, interpolation='spline16')
plt.axis('off')
fig = plt.figure()

fig.set_figwidth(50)
fig.set_figheight(50)

plt.show()
```

Brooklyn Culture

```
Brooklyn_culture = open('bln.txt', 'r').read()
cloudc2 = WordCloud(
    background_color='black',
    max_words=20,
)
cloudc2.generate(Brooklyn_culture)
plt.imshow(cloudc2, interpolation='spline16')
plt.axis('off')
fig = plt.figure()

fig.set_figwidth(50)
fig.set_figheight(50)

plt.show()
```

Brooklyn

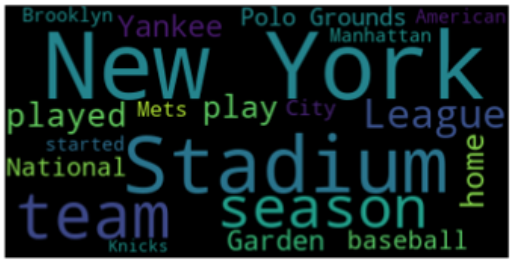
```
cloud2 = WordCloud(
    background_color='black',
    max_words=8,
)
cloud2.generate(Brooklyn)
plt.imshow(cloud2, interpolation='spline16')
plt.axis('off')
fig = plt.figure()

fig.set_figwidth(50)
fig.set_figheight(50)

plt.show()
```

Manhattan

```
cloud5 = WordCloud(
    background_color='black',
    max_words=8,
)
cloud5.generate(Manhattan)
plt.imshow(cloud5, interpolation='spline16')
plt.axis('off')
fig = plt.figure()

fig.set_figwidth(50)
fig.set_figheight(50)

plt.show()
```

Manhattan Culture

```
Manhattan_culture = open('man.txt', 'r').read()
cloudc5 = WordCloud(
    background_color='black',
    max_words=20,
)
cloudc5.generate(Manhattan_culture)
plt.imshow(cloudc5, interpolation='spline16')
plt.axis('off')
fig = plt.figure()

fig.set_figwidth(50)
fig.set_figheight(50)

plt.show()
```



Bronx Culture

```
Bronx_culture = open('bx.txt', 'r').read()
cloudc1 = WordCloud(
    background_color='black',
    max_words=20,
)
cloudc1.generate(Bronx_culture)
plt.imshow(cloudc1, interpolation='spline16')
plt.axis('off')
fig = plt.figure()

fig.set_figwidth(50)
fig.set_figheight(50)

plt.show()
```

The_Bronx

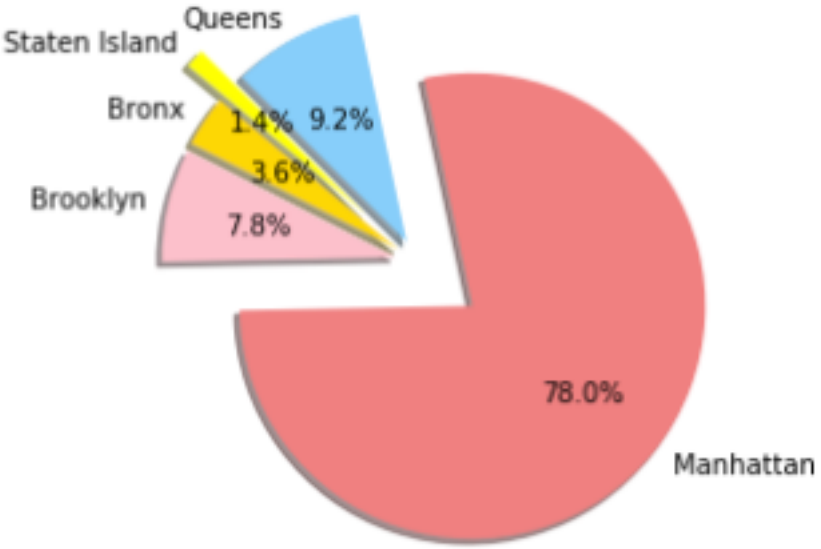
```
cloud1 = WordCloud(
    background_color='black',
    max_words=8,
)
cloud1.generate(Bronx)
plt.imshow(cloud1, interpolation='spline16')
plt.axis('off')
fig = plt.figure()

fig.set_figwidth(50)
fig.set_figheight(50)

plt.show()
```

We can thus distinctly observe the cultures and cuisines of these regions and determine how we would go about finalising them with regard to the heritage of the region.

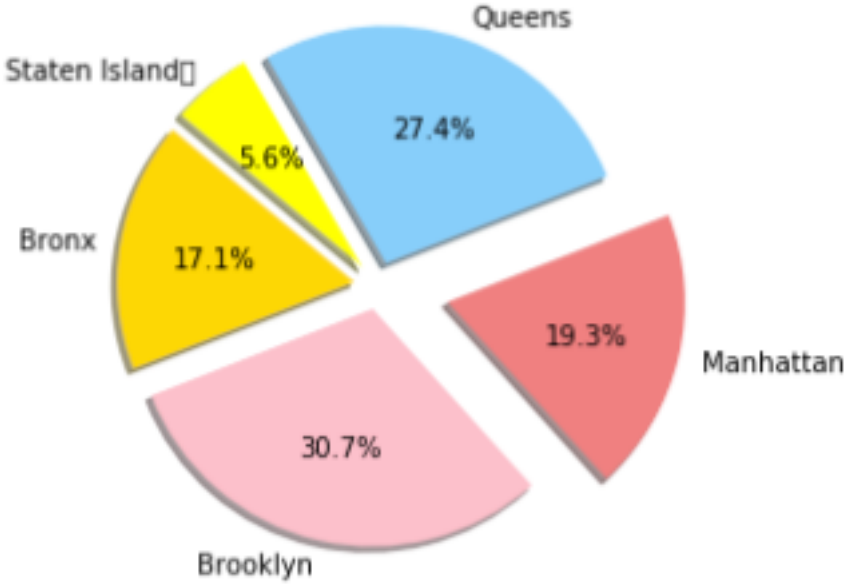
The final dataset we inspected involved looking at the financials like the gap of a region along with population densities to understand the standard of living and also look at the availability of resources with respect to the farmers market data.

We further cleaned the farmers market data by removing food box data as we don't consider them to of use with regard to our restaurant setup.

The below given pie chart gives us insight into the amount of contribution to the GDP made by individual boroughs.
We see that Manhattan contributes 78%, Queens contributes 9.2%, Brooklyn contributes 7.8%, The Bronx contributes 3.6%, and Staten Island contributes 1.4%.



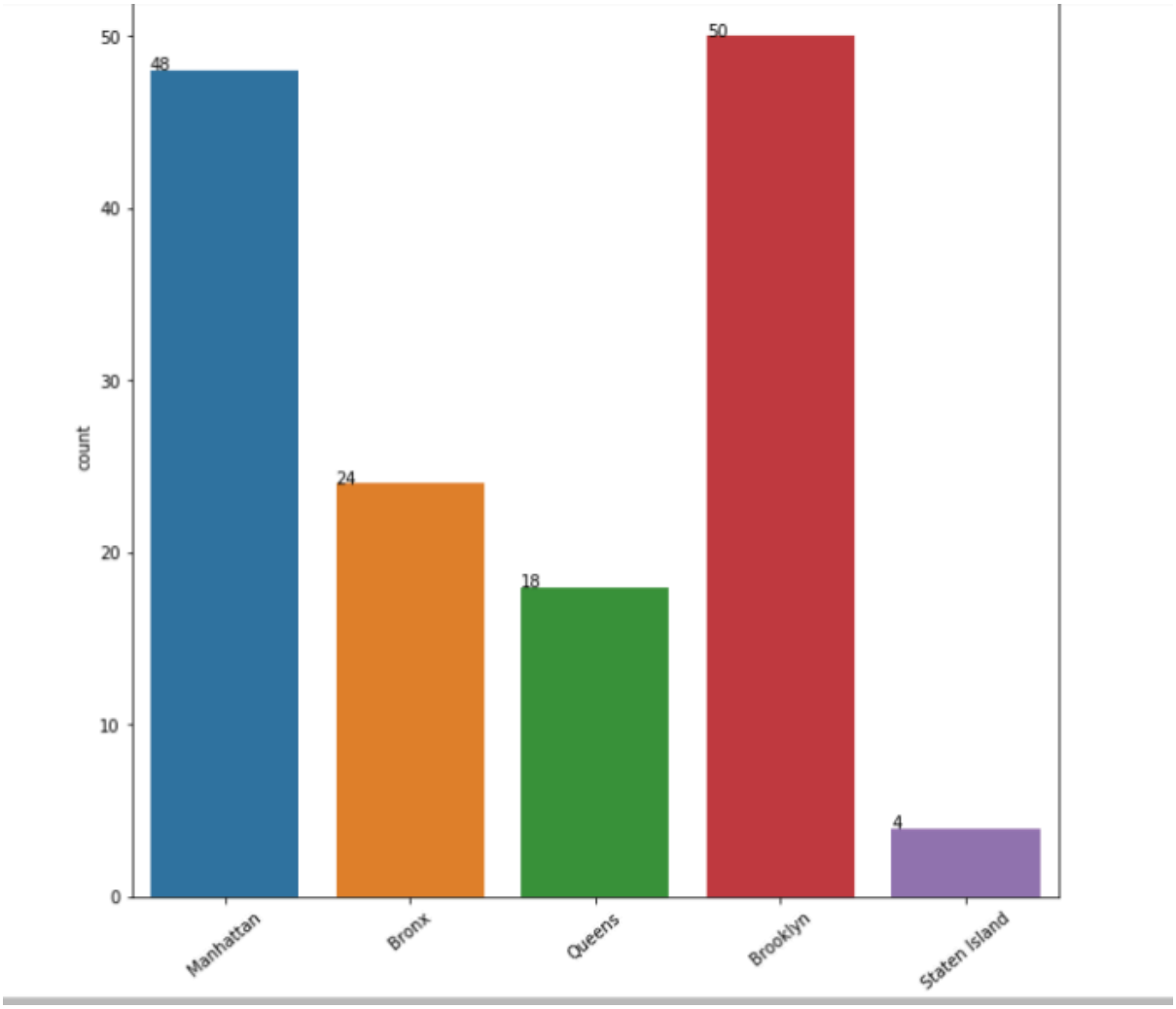
Furthermore,



The population densities in relatively high GDP regions are
Manhattan 19.3%
Queens 27.4%
Brooklyn 30.7%

We need not consider the others as their contribution is low in terms of GDP and also the overall population density.
We ignore them as such data is indicative of relatively slow development and setting up a business with higher profit margins would take a much longer time to kick off in such regions.

On performing analysis of areas with Food markets we get regions of interest as follows



Manhattan has 48 food markets
Queens has 18 food markets
Brooklyn has 50 food markets
Bronx has 24 food markets
Staten island has 4 food markets

RESULTS

Finally we cluster the 3 most probable options of neighbourhoods and their restaurants data using K-Means clustering. i.e Manhattan, Brooklyn, Queens

DISCUSSION

Given the constraints of investment, we would not advice starting up in manhattan where even though there is a very high contribution to the GDP the cost of living would also be high and there by the setup cost would be at an all time high

Queens would be the current best option to setup as it shows great potential with regard to its growth and also has a fairly decent contribution to the GDP.

Brooklyn on the other hand would qualify as the next best option for setting up a restaurant

CONCLUSION

Given the limited data we can conclude Queens to be the best option followed by Brooklyn to setup the first restaurant. Themes for these regions could be jazz, punk with cuisines like India, Italian and television and films as a theme with cuisines like Italian, Puerto Rican for Queens and Brooklyn regions respectively.

While we have made claims of certain places being optimal for setup, there could be high possibility of bias due to scarcity of data in other regions. There could also be an inclusion of a lot more factors given the restaurant vs residential area distances and many other factors even though small could make a great difference for precise analysis.