The Battle of the Neighborhoods

Where can I run a restaurant in NY?

ANALYSIS REPORT

Winston Suarez

Data Analyst

Logo Name

TABLE OF CONTENTS

BACKGROUND

New York City (NYC)

NYC is the iconic city of in the United States. With an estimated 2019 population of 8,336,817 distributed over about 302.6 square miles (784 km2), New York is also the most densely populated major city in the United States. Located at the southern tip of the U.S. state of New York, the city is the center of the New York metropolitan area, the largest metropolitan area in the world by urban landmass. With almost 20 million people in its metropolitan statistical area and approximately 23 million in its combined statistical area, it is one of the world's most populous megacities. New York City has been described as the cultural, financial, and media capital of the world, significantly influencing commerce, entertainment, research, technology, education, politics, tourism, cousin, art, fashion, and sports.

There are hundreds of distinct neighborhoods throughout the boroughs, many with a definable history and character. If the boroughs were each independent cities, four of the boroughs (Brooklyn, Queens, Manhattan, and the Bronx) would be among the ten most populous cities in the United States (Staten Island would be ranked 37th); these same boroughs are coterminous with the four most densely populated counties in the United States: New York (Manhattan), Kings (Brooklyn), Bronx, and Queens.

Cousin NYC

New York City's food culture includes an array of international cuisines influenced by the city's immigrant history. Central and Eastern European immigrants, especially Jewish immigrants from those regions, brought bagels, cheesecake, hot dogs, knishes, and delicatessens (or delis) to the city. Italian immigrants brought New York-style pizza and Italian cuisine into the city, while Jewish immigrants and Irish immigrants brought pastrami and corned beef, respectively. Chinese and other Asian restaurants, sandwich joints, trattorias, diners, and coffeehouses are ubiquitous throughout the city. Some 4,000 mobile food vendors licensed by the city, many immigrant-owned, have made Middle Eastern foods such as falafel and kebabs examples of modern New York street food. The city is home to "nearly one thousand of the finest and most diverse haute cuisine restaurants in the world", according to Michelin. As of 2019, there were 27,043 restaurants in the city, up from 24,865 in 2017.The Queens Night Market in Flushing Meadows–Corona Park attracts more than ten thousand people nightly to sample food from more than 85 countries.

PROBLEM

As NYC is a city with a very high multicultural component and visited by people from all over the world, it is thought that it can be a city of opportunities, because has a large concentration of habitants and visitors focused on four of its most traditional Boroughs. There is an investor fervent and enthusiastic lover of American culture who believes that he can delight the palates of New Yorkers with a different flavor of Colombian food. Using statistical, demographic and geopolitical information, aim is to help the investor determine which of the most traditional boroughs would be the ideal to install a gourmet Colombian food restaurant.

DATA

- Neighborhood has a total of 5 boroughs and 306 neighborhoods. In order to segment the neighborhoods and explore them, we will essentially need a dataset that contains the 5 boroughs and the neighborhoods that exist in each borough as well as the latitude and longitude coordinates of each neighborhood.

https://geo.nyu.edu/catalog/nyu_2451_34572

- New York Population
- New York City Demographics
- Cuisine of New York city

https://en.wikipedia.org/wiki/New_York_City

https://en.wikipedia.org/wiki/Economy_of_New_York_City

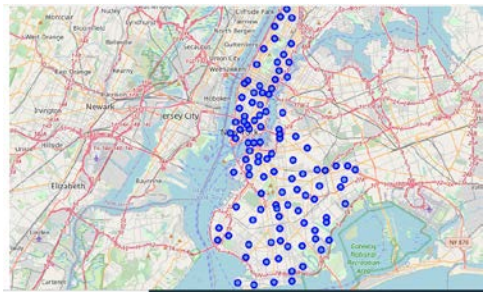
https://en.wikipedia.org/wiki/Portal:New_York_City

METHODOLOGY

A. Main data with more components are the Borough, Neighborhood, Latitude and Longitude

	Borough	Neighborhood	Latitude	Longitude
0	Bronx	Wakefield	40.894705	-73.847201
1	Bronx	Co-op City	40.874294	-73.829939
2	Bronx	Eastchester	40.887556	-73.827806
3	Bronx	Fieldston	40.895437	-73.905643
4	Bronx	Riverdale	40.890834	-73.912585

B. Then was used folium package in Python to visualize geographic details of Bronx, Brooklyn, Manhattan, Queens and Staten Island.



C. Using Foursquare API, explored the neighborhoods and segmented them. Was set a limit of 100 venues with a radius 500 m for each neighborhood given w Latitude and Longitude info.

	Neighborhood	NeighborhoodLatitude	NeighborhoodLongitude	Venue	VenueLatitude	VenueLongitude	VenueCategory
0	Marble Hill	40.876551	-73.91066	Arturo's	40.874412	-73.910271	Pizza Place
1	Marble Hill	40.876551	-73.91066	Bikram Yoga	40.876844	-73.906204	Yoga Studio
2	Marble Hill	40.876551	-73.91066	Tibbett Diner	40.880404	-73.908937	Diner
3	Marble Hill	40.876551	-73.91066	Starbucks	40.877531	-73.905582	Coffee Shop
4	Marble Hill	40.876551	-73.91066	Dunkin'	40.877136	-73.906666	Donut Shop

D. Once all neighborhoods were segmented was grouped information in venue category per type of business.

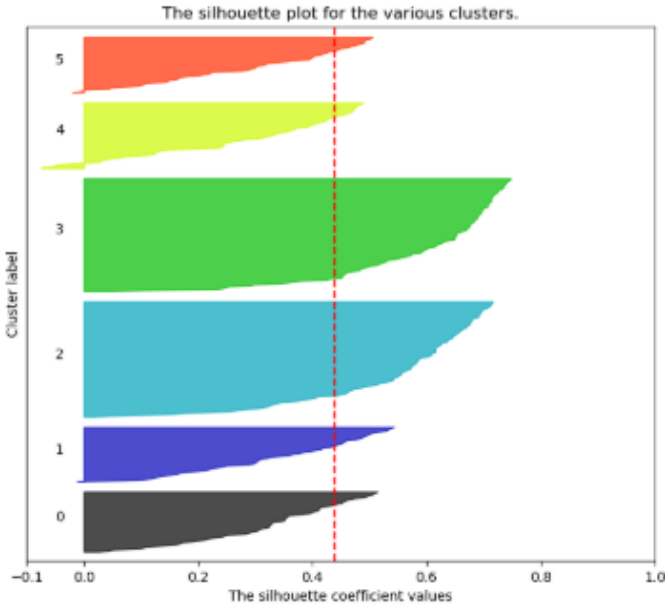
E. Result was filtered only by restaurants founded more than 70 types.

```
BM_venues.groupby('VenueCategory')['Venue'].count().sort_values(ascending=Fals
```

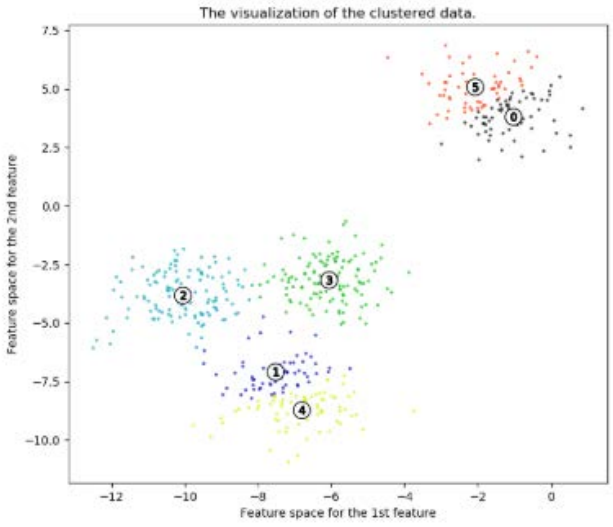
```
]: VenueCategory
Coffee Shop                244
Pizza Place                201
Italian Restaurant         186
Bar                        143
Bakery                     137
Café                       128
American Restaurant        119
Park                       118
Mexican Restaurant         106
Deli / Bodega              103
Sandwich Place              96
Cocktail Bar                93
Chinese Restaurant          91
Grocery Store               91
Ice Cream Shop              87
Wine Shop                   85
Gym / Fitness Center        83
```

METHODOLOGY

F. Modeling: In order to identify locations to run and install a new restaurant, was running a clustering method as **The Silhouette Coefficient.**



Silhouette analysis can be used to study the separation distance between the resulting clusters. The silhouette plot displays a measure of how close each point in one cluster is to points in the neighboring clusters and thus provides a way to assess parameters like number of clusters visually. This measure has a range of [-1, 1].



Silhouette coefficients (as these values are referred to as) near +1 indicate that the sample is far away from the neighboring clusters. A value of 0 indicates that the sample is on or very close to the decision boundary between two neighboring clusters and negative values indicate that those samples might have been assigned to the wrong cluster.

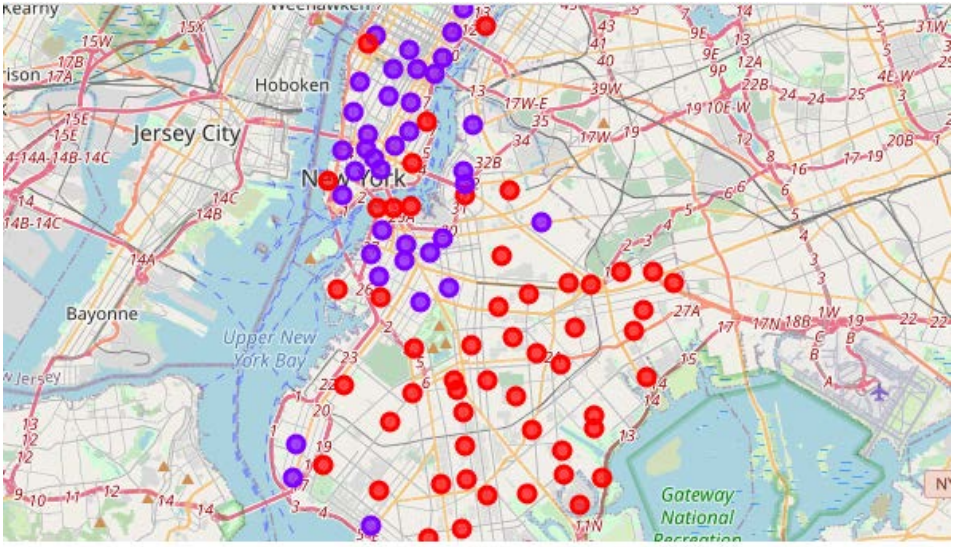
RESULTS

Determine the optimal value of K for our dataset

```
from sklearn.cluster import KMeans

from sklearn.metrics import silhouette_score

BM_grouped_clustering = BM_restaurant_grouped.drop('Neighborhood', 1)

for n_cluster in range(2, 10):
    kmeans = KMeans(n_clusters=n_cluster).fit(BM_grouped_clustering)
    label = kmeans.labels_
    sil_coeff = silhouette_score(BM_grouped_clustering, label, metric='euclidean')
    print("For n_clusters={}, The Silhouette Coefficient is {}".format(n_cluster, sil_coeff))
```

```
For n_clusters=2, The Silhouette Coefficient is 0.413242604992111
For n_clusters=3, The Silhouette Coefficient is 0.2871308397292759
For n_clusters=4, The Silhouette Coefficient is 0.21124201987737967
For n_clusters=5, The Silhouette Coefficient is 0.22091647952002086
For n_clusters=6, The Silhouette Coefficient is 0.20731252555972404
For n_clusters=7, The Silhouette Coefficient is 0.181862581099662
For n_clusters=8, The Silhouette Coefficient is 0.18371905152160772
For n_clusters=9, The Silhouette Coefficient is 0.16215569416739328
```

	Afghan Restaurant	African Restaurant	American Restaurant	Arepa Restaurant	Argentinian Restaurant	Asian Restaurant	Australian Restaurant	Austrian Restaurant	Brazilian Restaurant	Cajun / Creole Restaurant	Cambodian Restaurant	Cantonese Restaurant	Caribbean Restaurant	Caucasian Restaurant	R
cluster0	3.469447e-18	-1.387779e-17	0.459016	1.387779e-17	0.032787	0.081967	1.387779e-17	6.938894e-18	3.469447e-18	-1.387779e-17	3.469447e-18	-1.387779e-17	0.491803	0.016393	
cluster1	2.040816e-02	6.122449e-02	1.857143	8.163265e-02	0.122449	0.285714	8.163265e-02	4.081633e-02	2.040816e-02	6.122449e-02	2.040816e-02	6.122449e-02	0.326531	0.040816	



CONCLUSION & DISCUSSION

Cluster 1

```
BM_merged[BM_merged['Cluster_Labels'] == 1].reset_index(drop=True)
```

3]:

	Borough	Neighborhood	Latitude	Longitude	Total	Cluster_Labels
0	Brooklyn	Bay Ridge	40.625801	-74.030621	27	1
1	Brooklyn	Greenpoint	40.730201	-73.954241	21	1
2	Brooklyn	Prospect Heights	40.676822	-73.964859	26	1
3	Brooklyn	Bushwick	40.698116	-73.925258	21	1
4	Brooklyn	Brooklyn Heights	40.695864	-73.993782	22	1
5	Brooklyn	Cobble Hill	40.687920	-73.998561	26	1
6	Brooklyn	Carroll Gardens	40.680540	-73.994654	24	1
7	Brooklyn	Fort Greene	40.688527	-73.972906	23	1
8	Brooklyn	Park Slope	40.672321	-73.977050	16	1
9	Brooklyn	Bath Beach	40.599519	-73.998752	17	1
10	Brooklyn	Clinton Hill	40.693229	-73.967843	35	1
11	Brooklyn	Downtown	40.690844	-73.983463	27	1
12	Brooklyn	Boerum Hill	40.685683	-73.983748	15	1
13	Brooklyn	North Side	40.714823	-73.958809	23	1

Cluster 0

```
BM_merged[BM_merged['Total'] == 0].reset_index(drop=True)
```

4]:

	Borough	Neighborhood	Latitude	Longitude	Total	Cluster_Labels
0	Brooklyn	Crown Heights	40.670829	-73.943291	0	0
1	Brooklyn	Mill Island	40.606336	-73.908186	0	0
2	Brooklyn	Manhattan Beach	40.577914	-73.943537	0	0
3	Brooklyn	Dyker Heights	40.619219	-74.019314	0	0
4	Brooklyn	Bergen Beach	40.615150	-73.898556	0	0
5	Brooklyn	Midwood	40.625596	-73.957595	0	0
6	Manhattan	Stuyvesant Town	40.731000	-73.974052	0	0

- **Cluster 1 combines restaurants in all boroughs of the neighborhoods of Brooklyn and Manhattan, denoting on the analysis that this area is highly saturated.**
- **Cluster 0 shows area is not so saturated with 6 neighborhoods in Brooklyn and one neighborhood in Manhattan that could be adjusted to install a restaurant.**
- **Colombian food restaurants were not identified in the analyzed area, it is required to obtain more statistical, marketing and demographic information to identify if there will interest in this type of food.**
- **The investor needs to adjust the focus of the type of food to offer, knowing that most frequents sites for people in this area are: coffee shops, pizza place, Italian restaurants, Bars, American restaurants and Mexican restaurants**
- **Moreover, I only used k-means The Silhouette Coefficient method without any comparison with other techniques, I will practice more complex data to do more advanced analysis such as hierarchy clustering.**