

# IBM Applied Data Science Capstone

Capstone Project - The Battle of Neighborhoods

Opening a bakery in New York City, New York

Author: Alexandros Vellios



## **Introduction**

With an estimated population (2018) of 8,398,748 distributed over about 302.6 square miles New York is also the most densely populated major city in the United States. It is also the most populous city in the United States.

The city has been described by many as the cultural, financial, and media capital of the world, significantly influencing commerce, entertainment, research, technology, education, politics, tourism, art, fashion, and sports.

Given how densely populated and economically developed the city is, space is limited and the cost of doing business high. However the big population makes the area very appealing for new businesses, and the rewards high if one strategically plan the business venture.

In this project we will work for a client who is looking to open a bakery in NYC after he identified a gap in the market around a neighborhood of New York.

## **Business Problem**

Our client, an independent business owner, has come to us to advise them regarding which borough and neighborhoods within this borough in NYC are the best locations to open a bakery.

According to our client we should focus on three features to find the best location:

1. Lack of competition (bakeries)
2. Presence of entertainment businesses near the location
3. Presence of tourist attractions near the location

In order to find the best location we will segment and cluster the different boroughs of NYC and come up with the best possible location for our client to open his business.

## **Target Audience**

The target audience of this report will be our client, who wants to open a bakery in NYC

## **Data Selection**

For this project we will utilise two data sources. The first data source ([https://geo.nyu.edu/catalog/nyu\\_2451\\_34572](https://geo.nyu.edu/catalog/nyu_2451_34572) ) has a total of 5 boroughs and 306 neighborhoods. It also contains the latitude and longitude coordinates of each neighborhood. Example of the data:

	Borough	Neighborhood	Latitude	Longitude
0	Manhattan	Marble Hill	40.876551	-73.910660
1	Manhattan	Chinatown	40.715618	-73.994279
2	Manhattan	Washington Heights	40.851903	-73.936900
3	Manhattan	Inwood	40.867684	-73.921210
4	Manhattan	Hamilton Heights	40.823604	-73.949688

Furthermore, using the geographical coordinates data as input for the Foursquare API, we will extract venue information for each neighborhood. Therefore, we will use the Foursquare API to explore neighborhoods in New York City. The below is an example of the Foursquare API data:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
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

With data collected from the two sources above we will be able to tackle our client's issue.