

2019

Optimal Pizzeria Location Determination in The Bronx



Jon Wages

IBM Data Science Professional Certification

Applied Data Science Capstone

11/27/2019

Introduction

New York has a long history of producing bold and original styles of pizza. The first pizzeria to be opened in the United States was Lombardi's founded by Gennaro Lombardi in Little Italy, Manhattan in 1905. An Italian immigrant and pizzaiolo, or pizza maker, from Naples, Lombardi brought his favorite dish to America and started a pizza revolution that would leave a permanent mark on the culture of New York. As time passed, the pizza Lombardi created became a fusion of Italian cuisine and American tastes leading to the iconic New York-style pizza many enjoy to this day. Possessing a thin, hand-tossed crust and served in large foldable slices to be eaten on the go, New York-style pizza is as innovative and flavorful as the city where it originated.

New York has become a popular destination for tourists and pizza connoisseurs alike, and with tourism comes business opportunity. An estimated 62.8 million people visited New York in 2017 and that number is expected to continue to increase. According to research from Pizza Today, Americans spend nearly \$38 billion on pizza; equating to over 3 billion pizzas sold each year. Furthermore, the average American will eat 46 slices of pizza this year alone. With statistics like that, many entrepreneurs are eager to get a "slice" of the disposable income of New York tourists and residents.

Business Opportunity & Audience

The purpose of this data science capstone project is to analyze the neighborhoods within The Bronx, New York to determine which neighborhood or group of neighborhoods would be the best in which to open a pizzeria. I will use the methodology and techniques which I have learned over the course of pursuing my IBM Data Science Professional Certification in order to address this issue. This research would be of use to the many entrepreneurs and restaurateurs seeking to expand their existing pizzeria footprint or those looking to disrupt the competitive New York pizza restaurant scene.

Data Sources & Utilization

I will be using the below listed sources to solve this business problem:

- Foursquare API Data
 - This will provide venue data which can be utilized to determine not only which neighborhoods within The Bronx have pizzerias, but also to determine the number and frequency of pizzerias within each neighborhood.
- Location Data
 - Location data will be obtained from the New York University Spatial Data Repository which contains a list of New York boroughs and neighborhoods.
 - Link to the Repository: https://geo.nyu.edu/catalog/nyu_2451_34572
- Geospatial Data for Boroughs and Neighborhoods within New York
 - This information is housed in the above link and will be utilized to plot a map of the Bronx and provide a visual representation of neighborhood clusters according to the frequency of pizzerias within them.

Methodology

What follows is a description of the steps I took to determine the optimal pizzeria location in The Bronx, New York. I began by calling the location data from New York University's Spatial Data Repository into the notebook. As this set of data used a JSON format, I first pulled the data into a pandas data frame. This allowed me to easily compile the location data according to Borough, Neighborhood, Latitude, and Longitude. After creating a data frame as a starting point, I narrowed the focus of the data frame to only neighborhoods within the Bronx.

Once I was able to exclude all other boroughs in New York from the data frame, I then pulled a list of the top 100 venues within 500 meters of the center of each neighborhood by calling on the Foursquare API. This allowed me to obtain a detailed list of each kind of venue within each neighborhood and also provided me with the number of each type of venue present. I then grouped the venues within each neighborhood by type and the neighborhood in which they were located. This facilitated the ability to obtain the frequency of each venue within each unique neighborhood.

Once I had determined the venue type frequency within each neighborhood, I filtered the frequency and neighborhood data to focus on, what Foursquare calls, “Pizza Places” or as I have called them in this document “Pizzerias.”

The next step was to cluster this data using K-means Clustering. Effectively, K-means Clustering groups together data points based on their relative similarity to other data points while also ensuring appropriate distance between data points that are dissimilar. I chose to divide the neighborhoods into four separate clusters. This helped to identify four unique types of neighborhoods based on the frequency of pizzerias within them.

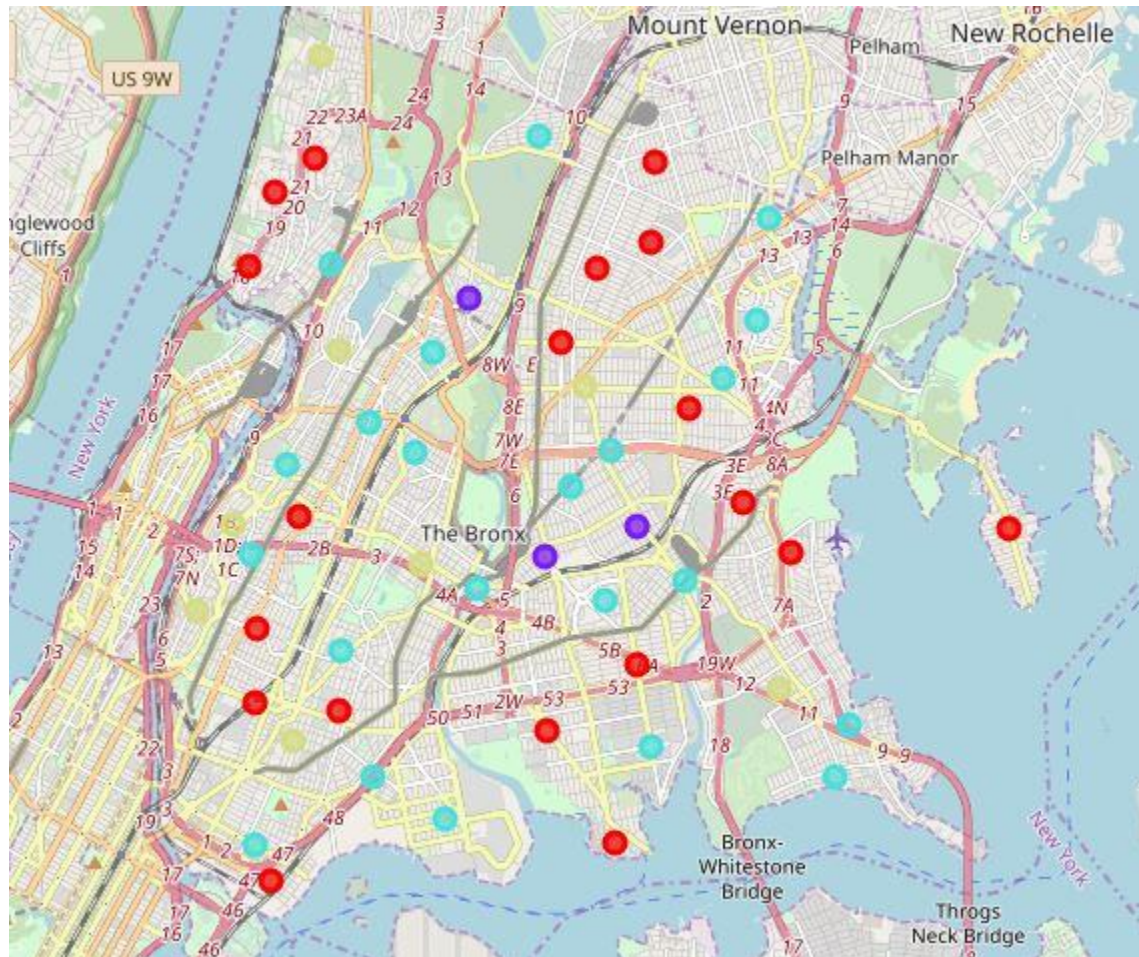
Results

The results of the K-means Clustering of neighborhoods based on pizzeria frequency are detailed below:

- Cluster 1 – Represented by **Red** Dots
 - This cluster contained neighborhoods with no pizzerias or a very low number of pizzerias.
 - There were 19 neighborhoods in this cluster.
- Cluster 2 – Represented by **Purple** Dots
 - This cluster contained neighborhoods with a substantially high number of pizzerias.
 - There were 3 neighborhoods in this cluster.
- Cluster 3 – Represented by **Blue** Dots
 - This cluster contained neighborhoods with a low number of pizzerias.
 - There were 22 neighborhoods in this cluster.
- Cluster 4 – Represented by **Pale Yellow** Dots
 - This cluster contained neighborhoods with a moderate number of pizzerias.
 - There were 8 neighborhoods in this cluster.

A map displaying these clusters has been placed on the next page for your reference.

Cluster Map



Discussion

The data from the cluster analysis reveals four distinct kinds of neighborhoods in relation to the frequency of pizzerias within them. Cluster 2 (Purple) contains the highest number of pizzerias and is also the cluster with the fewest neighborhoods. Cluster 4 (Pale Yellow) contains a moderate number of pizzerias and contains the second fewest neighborhoods. Cluster 3 (Blue) contains a low number of pizzerias and contains the highest number of neighborhoods. Cluster 1 (Red) contains neighborhoods with either no pizzerias or very few pizzerias.

Cluster 2 and Cluster 4

Entrepreneurs and Restaurateurs seeking to start or expand their existing pizzeria business should avoid selecting a neighborhood within either Cluster 2 or Cluster 4. These clusters contain an elevated number of pizzerias relative to the other neighborhoods within The Bronx. By choosing a neighborhood within either of these clusters, a business owner would be immediately faced with a high level of competition and may not be able to fully capitalize on their investment into the pizza business.

Cluster 1

Cluster 1 contains the lowest frequency of pizzerias out of any of the clusters provided by my analysis. While the initial reaction of some may be to think this would be the best cluster in which to place a pizzeria, it is important to remember the pizza culture of New York. With pizza being a large part of New York life and history, it is possible neighborhoods within Cluster 1 do not have many pizzerias due to their inability to support new restaurants. This could be either due to a lack of interest within these neighborhoods or limiting socioeconomic factors outside the scope of this report. In short, Cluster 1 is a pizzeria desert from a data analysis standpoint.

Cluster 3

Cluster 3 contains neighborhoods with a lower frequency of pizzerias comparative to Clusters 2 and 4. This is indicative of lower levels of pizza business competition and makes Cluster 3 potentially more attractive than Cluster 2 and Cluster 4. Cluster 3 also has a higher frequency of pizzerias than Cluster 1 which points to the ability of neighborhoods within Cluster 3 to sustain pizzerias. Additionally, Cluster 3 contains the highest number of neighborhoods which allows business owners to have a range of neighborhoods to choose from when starting or expanding their business and would potentially allow early long-term planning regarding expansion into numerous other neighborhoods within this cluster.

Recommendation

Out of all the neighborhoods within the Bronx analyzed by this report, I would recommend pizza-centric entrepreneurs and restaurateurs focus their attention on neighborhoods within Cluster 3 when planning to either start or expand their pizza business.

A table of neighborhoods within Cluster 3 is included on the next page for your review.

Cluster 3 Table

Low Pizzeria Frequency Neighborhoods	
Baychester	Kingsbridge
Bedford Park	Longwood
Belmont	Mott Haven
Bronxdale	Mount Eden
Castle Hill	Parkchester
Claremont Village	Pelham Parkway
Co-op City	Throgs Neck
Eastchester	University Heights
Edgewater Park	West Farms
Fordham	Westchester Square
Hunts Point	Woodlawn

Conclusion

In this report, I have provided a brief history of pizza culture within New York, introduced a business opportunity concerning optimal pizzeria location determination in The Bronx, listed sources of data which were used to address this business opportunity, walked the reader through my methodology and approach to solving for where the best location would be to open a pizzeria, presented the results of my analysis, and provided my recommendation (Cluster 3) based on the results from my analysis.

This report will be of use to those looking to either open a pizzeria or expand their pizza business; however, this report does not address socioeconomic or population-based factors which could contribute to the success or failure of a new pizzeria. Further research and analysis on these topics would be needed to address these concerns.