**Determining the best locations in New York City for an Italian Restaurant**

Chris Arnold

November 5th, 2020

1. **Introduction**
   1. **Background**

New York City is known for great Italian cuisine and competition is fierce. Although the competition is fierce, not all areas of New York City have an adequate number of restaurants nearby, therefore could be filled with a great Italian restaurant. I own a consulting company that helps prospective business owners find the best locations for their business.

* 1. **Problem**

In this project, I will be focusing on a client that wants to open up an Italian restaurant in New York City, but not sure the best location. They are not looking at any particular area within NYC, but they want to start their business where there is limited competition nearby.

1. **Data acquisition and cleaning**

I will be utilizing Foursquare location data to understand where existing Italian restaurants are currently located. I am searching for a location that has a limited number of Italian restaurants in the neighborhood. I will be using the Foursqaure API to determine location data and venue type, so Client secret and Client ID will be required.

I will also be utilizing NYC neighborhood location data that includes longitude and latitude for each neighborhood to be able to identify which neighborhoods will be best fit for my client to open their Italian restaurant.

I will first take the NYC data and put into a dataframe that will give me the Longitude/Latitude of each neighborhood. From there, I will utilize Foursquare data to understand where other Italian Restaurants are in comparison to the neighborhood's Longitude/Latitude to understand which neighborhoods have the least density of Italian Restaurants.

1. **Methodology**

I first started by gathering New York City neighborhood location data which was formatted in a JSON file and converted that data to a pandas dataframe and plotted the data on a map. Next, I called the Foursquare API to gather location and venue type data, which also needed to be converted to a pandas dataframe. I included a search criteria for ‘Italian’ in the Foursquare date to limit the amount of data that needed to be reviewed. I then merged the neighborhood data with the Foursquare data. I then created a new dataframe which looked at the top 10 types of venue per neighborhood, which would assist in the clustering. I use k-Means clustering to cluster neighborhoods by venue type.

1. **Results**

It was clear after reviewing the clustering results that Cluster 3 represented the best location for an Italian restaurant followed by Cluster 2. Cluster 3’s most common venues are Ice Cream Shops and there is very few Italian restaurants present. One Neighborhood, Rockaway Beach, had Pizza Places as their third most common type of restaurant, but no other type of Italian restaurant.

1. **Discussion and Conclusion**

The clustering data was very helpful in identifying areas of New York City that has a low density of Italian Restaurants. One way to be able to help my clients even more would be to gather average venue ratings, so we could not only look at where Italian restaurants are located, but also find areas that do not have any high rated restaurants to fill the gap.