

# Harlem, NYC: What Type of Restaurant Should You Build?

Maria DiMedio

July 2020

## I. Introduction

### A. Background

New York City is world renowned for its food scene. The availability of different types of cuisine is what draws many consumers and tourists from across the globe. One of the hubs of this food scene is Harlem. Known for both its diversity and its accessibility, you can get affordable food of almost any type of cuisine there. That's what makes it difficult for new restaurant owners to compete – what cuisine should you cater to if there are already so many?

### B. Business Problem

One of the most challenging things for restaurant investors and small business owners is to understand what the food scene in the location they are interested in has capacity for. In this analysis, I will uncover which types of cuisines are the most saturated in Harlem, and suggest which cuisines are not represented and could be a good opportunity for the right investor.

### C. Interest

Having lived in Harlem for a short time, my favorite memories are centered around the food I tasted. I tried many things there for the very first time. I also watched several new restaurants try to open, some more successfully than others. When many of these restaurants opened, it seemed like only new ideas were succeeding over ones that had many direct competitors. For example, new Italian or Dominican restaurants never seemed to last, and that may have been because there were already so many that were neighborhood favorites.

## II. Data

### A. Sources

I will be using a few data sources for this analysis. First, I will be using Foursquare's API and restaurant database to collect data on the types of restaurants in Harlem. I will be pulling results from all businesses in the geographic region that Harlem encompasses, with specific focus on the 'Venue Category' feature. Secondly, the geographic locations of each neighborhood can be pulled from this public [dataset](#), containing the latitude and longitude of every zip code in the United States. Third, the neighborhoods which correspond with each zip code in Manhattan were pulled from this [repository](#) published by the NY State Department of Health.

### B. Data Cleaning and Feature Selection

The US Zip Code dataset contained the latitude, longitude, zip, and state features needed to run the Foursquare query API. However, it did not indicate which neighborhoods each zip code corresponded to. The repository for the NY State of Health contained the zip and neighborhood features, so I could combine the two data sets to get every feature I need.

My first step was to performed a join operation on the US Zip Codes data with the Manhattan Neighborhoods Zip Code data set. The result was a merged data set of every zip code, latitude, and

longitude value for every neighborhood in Manhattan. Then, I narrowed the dataset to just Harlem, which encompasses two sub-neighborhoods: “Central Harlem” and “East Harlem”. I did this by filtering the combined data set ‘Neighborhood’ feature for any name containing ‘Harlem’.

Next, using Foursquare’s API capabilities, I collected data on businesses near every zip code in Harlem. For this analysis, I needed to know how many of every type of restaurant there are in each zip code. Therefore, I ran the query to comb a radius of 500 meters from the center point of each zip code, for the top businesses nearby. Because I am only interested in the restaurant businesses, I ran a unique value filter on the ‘Venue Category’ column and narrowed my data down to only the businesses who are categorized relating to food service.

Once I had the dataset I wanted, I then used the one-hot encoding technique to get the frequency of each type of restaurant. By doing this, I was then able to get a count of each type of restaurant that is in Harlem segmented by style of cuisine.

### **III. Methodology**

#### **A. Data Analysis**

#### **B. Modeling**

### **IV. Results**

### **V. Discussion**

### **VI. Conclusions**