

Capstone Project – The Battle of Neighborhood
The Optimal Location for a Planned Mexican
Restaurant in New York City

Xin Zhang

12-3-2020

1. Introduction

1.1 Project Description

This project is to find an optimal location to open a Mexican restaurant in New York City. To find this location, several factors should be considered, including: 1) the number of the same type restaurants nearby; 2) the population density near the new restaurant.

Less Mexican restaurants nearby means less competition, while higher population density means there are more potential customers of the planned restaurant. They're both essential parts of the restaurant business and will be the very first thing that should be considered.

In this project, we should first determine the candidate borough, and then find out the best neighborhood to open the restaurant.

New York City is composed of five boroughs: The Bronx, Brooklyn, Manhattan, Queens, and Staten Island. From the table below, we can get the population density of each borough.

| New York City's five boroughs | | | | | | | | |
|-------------------------------|---------------|------------------------|-----------------------------|--------------------------|---------------------|------------------|--------------------------------|---------------------------------|
| Jurisdiction | | Population | Gross Domestic Product | | Land area | | Density | |
| <i>Borough</i> | <i>County</i> | <i>Estimate (2019)</i> | <i>billions (2012 US\$)</i> | <i>per capita (US\$)</i> | <i>square miles</i> | <i>square km</i> | <i>persons / m²</i> | <i>persons / km²</i> |
| The Bronx | Bronx | 1,418,207 | 42.695 | 30,100 | 42.10 | 109.04 | 33,867 | 13,006 |
| Brooklyn | Kings | 2,559,903 | 91.559 | 35,800 | 70.82 | 183.42 | 36,147 | 13,957 |
| Manhattan | New York | 1,628,706 | 600.244 | 368,500 | 22.83 | 59.13 | 71,341 | 27,544 |
| Queens | Queens | 2,253,858 | 93.310 | 41,400 | 108.53 | 281.09 | 20,767 | 8,018 |
| Staten Island | Richmond | 476,143 | 14.514 | 30,500 | 58.37 | 151.18 | 8,157 | 3,150 |

Obviously, the population density of Manhattan is much higher than the other boroughs. So, I'll focus on Manhattan and try to find optimal restaurant location in

this area.

After that, all neighborhoods in Manhattan will be considered. By comparing the number of exist Mexican restaurants in each neighborhood and the population density of each neighborhood, we'll obtain 5 best location. This result will be used for the restaurant owner and stakeholders to decide the final location.

1.2 Background

Manhattan is the most densely populated of the five boroughs of New York City. Manhattan contains more than 40 neighborhoods, the population is 1.628 million. There are over 25,000 restaurants and bars in Manhattan, of which over 500 are Mexican restaurants. But these Mexican restaurants distribute unevenly. Some neighborhoods barely have such restaurants. Normally they are either very small neighborhoods or dominated by one ethnicity. In some neighborhoods the situation is quite the opposite. There could be more than two Mexican restaurants within a few miles around. So, our goal is to find the optical location for the new Mexican restaurant, where neither isolated nor crowded.

2. Data Acquisition and Cleaning

2.1 Data Sources

- New York City population density by borough can be downloaded [here](#).
- Manhattan neighborhoods data can be download from the [here](#).
- Mexican restaurants data can be found through Foursquare API.
- Manhattan population by neighborhood data is downloaded [here](#).
- New York City geojson file used to generate map can be downloaded [here](#).

2.2 Data Cleaning

The Manhattan neighborhoods dataset can be generated from newyork_data.json file. Create a query through Foursquare, using “Mexican” as key word to get the venues

data. Download Manhattan population dataset.

There are several problems with the venues data.

First, the venues category contains some non-conformance terms, such as “Mediterranean Restaurant”, “Dessert Shop”, “Cheese Shop”. I checked all the venues category and dropped the non-conformance terms.

The second problem is that some neighborhood names in “Neighborhood-Population” dataset is different from the ones in “newyork_data”. For example, some neighborhoods have more than one name, some neighborhoods with large area are divided into several child-neighborhoods. I unified the names to make sure that the datasets can be merged based on neighborhood name.

After merging all the datasets together, the dataframe contains 5 columns and 52 rows.