The Battle of Neighborhoods: Recommended System Based on the Neighborhoods Similarity

Fernandes Macedo Ribeiro

November 15, 2020

1. Introduction

1.1. Business Understanding

New York City is one of the most famous in the world and with the biggest population in the United States of America. It has a population nearly in 2020 of approximately 8,622,357 million residents. The population density is 27,755.25 people per square mile, with an area of 468.19 square miles. With this density population, it provides many business opportunities, attracting in many different players into the market.

With this basic information, we can gain insights derived from analysis that will give a good understanding of the business environment, which can help in strategically targeting the market.

A restaurant is a business, which prepares and serves food and drink in a famous and excellent cuisine. Food culture include an array of international cuisines influenced by the city's immigrant history. With this information, we will analysis public data to find insight, to start a new way to start open a restaurant.

1.2 Project Objectives

Population density is a key aspect of analyzing data on cities, and it plays an important part in issues such as transport, infrastructure and living standards for residents. Some people prefer smaller, less crowded cities with open spaces, while others would rather have the hustle and bustle of big city life.

In New York, not just Pizzerias or Coffee Shops are famous, but also there is fine Michelin starred restaurants with the most diverse type of food. With this diversity of restaurants with different categories like Chinese, Indian, French, Brazilian, etc. So here, the main problem is to find the following answer for the question below:

- List and visualize all major parts of New York City that has great restaurants.
- What is the best location in New York City for Brazilian?
- Which areas have potential Restaurant Market?
- Which all areas lack Brazilian Restaurants?
- Which is the best place to stay if you prefer Brazilian Food?

1.3 Interest

A fictitious restaurant chain in Brazil, who wants to start to expand their operation of a new restaurant in New York City, requested this project. The objective is to locate and

recommended to the stakeholders which neighborhood of New York City will be best choice to start a restaurant. However is fundamental the stakeholders understand the rationale of the recommendations made.

1.4 Methodology

Every project request, need a good methodology to solve and understand better the business problem. This project will be approach by CRISP-DM (Cross Industry Standard Process for Data Mining) that I believe to be the best lifecycle to develop this project.

The lifecycle of this project will include almost every step defined in the data mining projects that include:

- Business Understanding
- Data Understanding
- Data Preparation & Cleaning
- Modeling
- Evaluation
- Deployment
- Conclusions

2. Data Understanding

2.1 The Data Set

New York City's have a collection of restaurants from across the city that defined by a diverse culture with a lot of diverse type of food, each belonging to different categories like Chinese, Indian, French, Brazilian, etc. For this project, we will use an open data as following below:

New York City data that contains list Boroughs, Neighborhoods along with their latitude and longitude.

- Data set : https://cocl.us/new_york_dataset
- Description: This data set contains the required information. Moreover, we will use this data set to explore various neighborhoods of New York City.

  Chinese, Indian, French, Brazilian restaurants in each neighborhood of New York City.

- Data set : Foursquare API
- Description: By using this API, we will get all the venues in each neighborhood. We can filter these venues to get Chinese, Indian, French, Brazilian restaurants.

  GeoSpace data set
- Data set : https://data.cityofnewyork.us/City-Government/Borough-Boundaries/tqmj-j8zm
- Description: By using this geo, space data we will get the New York Borough boundaries that will help us visualize choropleth map.

Table 1. Simple feature selection during data cleaning.

	Borough	Neighborhood	Latitude	Longitude
0	Bronx	Wakefield	40.894705	-73.847201
1	Bronx	Co-op City	40.874294	-73.829939
2	Bronx	Eastchester	40.887556	-73.827806
3	Bronx	Fieldston	40.895437	-73.905643
4	Bronx	Riverdale	40.890834	-73.912585

3. Data Preparation & Cleaning

I had use the Foursquare API to get the top 100 venues with in a radius of 1000 meters for a given latitude and longitude. And with data um use an Explanatory Data Analysis to understand the data.
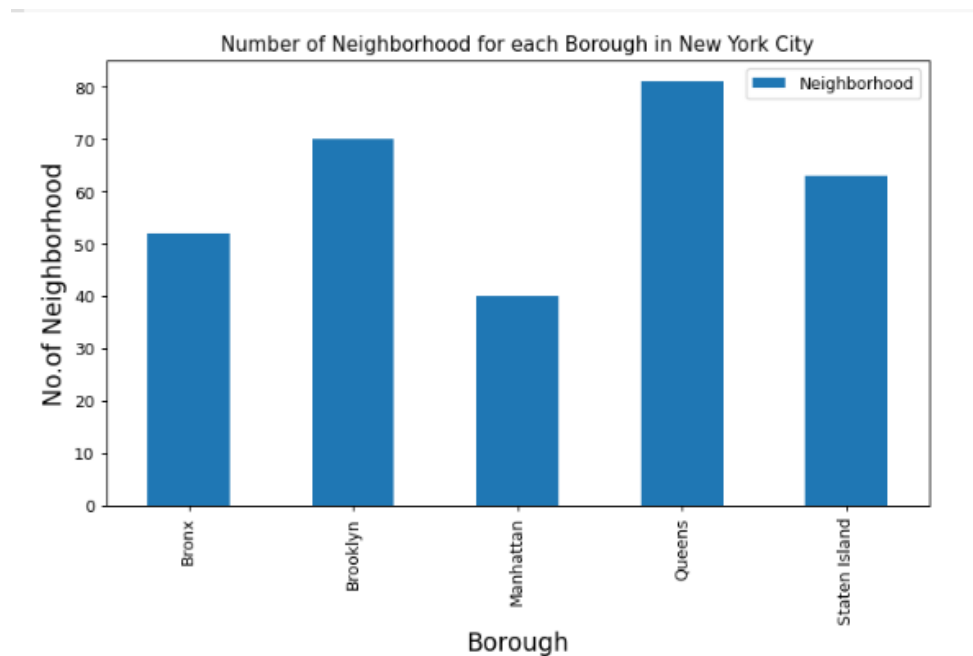

Figure 1: Number of Neighborhood for each Borough in New York City

4. Modeling

In this part of the project, it was not necessary to include machine learning models, as the result was enough for the stakeholders to make the right decision on where to open the new Brazilian restaurant.

5. Evaluation

I have answered all posed business questions in the introduction of this project. Now it is up to the stakeholders to decide whether the new restaurant will open, only with the analysis of this data set, otherwise more data will be needed so that even a machine learning model is implemented.

6. Deployment

In this project, we are going to implement this feature in the future using simple linear regression algorithm with scikit-learn for simplicity, we will use Flask as it is a very light web framework.

7. Conclusions

In this study, I analyzed the relationship between the neighborhoods of New York City to find insights. So now, we can answer the questions asked above in the Business Understand section of the notebook.

For our analysis the answers to the above questions are:

- Greenwich Villag (Manhattan), West Village (Manhattan), Astoria (Queens) are some of the best neighborhoods for Brazilian Cuisine.

- Manhattan have potential to Brazilian Restaurant Market.

- Bedford Stuyvesant ranks last in average rating of Brazilian Restaurants.

- Manhattan is the best place to stay if you prefer Brazilian Cuisine.