IBM Data Science Capstone
Robert Webster

BATTLE OF THE NEIGHBORHOODS

NEW YORK CITY

FY 2020

This report discusses the challenges of finding a good neighborhood to move to and start a business considering living expenses, availability of venues that are preferred, etc., and offers a solution by using modern Data Science Tools.

Image courtesy of https://upload.wikimedia.org/wikipedia/commons/3/30/1895_Bien_Map_of_New_York_City_%28w-_Queens_%5E_the_Bronx%29_-_Geographicus_-_NYC-bien-1895.jpg

TABLE OF CONTENTS

Contents

Image courtesy of https://www.goodfreephotos.com/united-states/new-york/new-york-city/face-of-the-statue-of-liberty-in-new-york.jpg.php

BATTLE OF THE NEIGHBORHOODS

Introduction

BUSINESS PROBLEM

I want to start living in New York City (NYC), but I don't know what neighborhood to live in. I love to eat Chinese food so I want to live in a neighborhood that is small, but has a high density of Chinese food restaurants. Eventually, if the neighborhood is well-suited, I may consider opening my own restaurant – I want to open mine in a community of similar minded eaters. (If you want good Chinese food – go to China Town!)

TARGET AUDIENCE / STAKEHOLDERS

This analysis is a good project for data scientists seeking to explore available open-source tools. It is also a useful tool for someone who wants to move to NYC! The simple problem statement above is actually quite complex to resolve, but easily done with the correct tools and code

BACKGROUND

NYC is exciting and interesting and I like tall buildings. I enjoy the cultural diversity. I like to eat a lot of food, so I want there to be lots options to eat near-by. Specifically, I like to eat Chinese food.

Unfortunately (or fortunately – depending on your perspective) NYC is a gigantic city with many neighborhoods and a significant number of venues to choose from. This project would be nearly impossible to do by flipping through a phone book, or trying to search manually though google.

This project seeks to utilize the advantages of modern-day data science tools to solve the problem discussed above.

Data

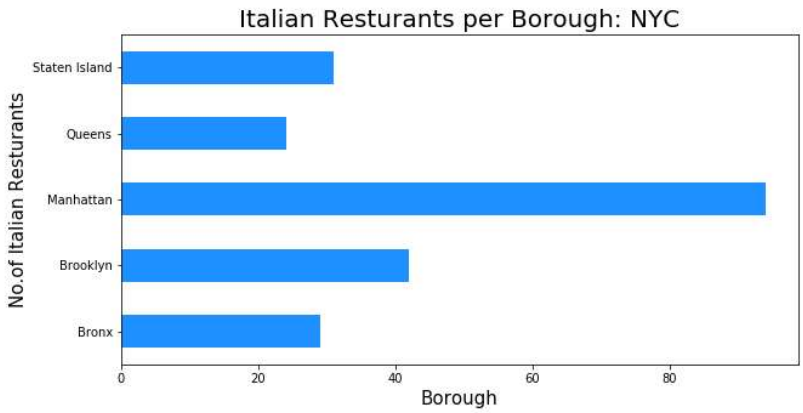
DATA SETS

The analysis begins with an examination of various publically available data sets. It is critical that the data science solution be able to utilize up-to-date sources of data, so the web and services such as Four Square are heavily leveraged.

1. Neighborhoods: This data, including neighborhood name, borough, and location, is extracted by using the Python extension called *Beautiful Soup*. Its parses html data and makes it available to insert into a pandas dataframe.
2. Neighborhoods: The data can also be obtained from a publically available csv file at this location - *https://cocl.us/new_york_dataset*

	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. Specific Venue Data: This is obtained through use of the Four Square API. Four Square provides continuously updated, user-supplied information regarding venues. All data is geo-tagged so that it can be compared to other locations with relative ease.



4. Location Data: Location data about almost any area can be obtained using the Geopy python extension. In this project it is used to find the latitude and longitude of New York City and some neighborhoods within it, but can be used to find any similar location desired.

```
: # Use geopy library to get the latitude and longitude values of New York City.
address = 'New York City, NY'

geolocator = Nominatim(user_agent="ny_explorer")
location = geolocator.geocode(address)
latitude = location.latitude
longitude = location.longitude
print('The geograpical coordinate of New York City are {}, {}.'.format(latitude, longitude))
```

```
The geograpical coordinate of New York City are 40.7127281, -74.0060152.
```

```
Latitude and longitude values of Marble Hill are 40.87655077879964, -73.91065965862981.
```