Capstone Project - The Battle of Neighborhoods)

Business Problem

Everyone has their own taste of restaurant environment, someone like romantic, someone like the simple settings while other may prefer a quiet place to dine alone. Those preferences varied from customer to customer.

My client, the dominant hot pot restaurant chain in China is looking to expand operation in North America. To start with, expand the first location in New York City (NYC).

Problem Statement

Recommend to the client the choice of neighborhood to expand the chained restaurant. Data science tools will be used to discuss the following two problems: 1. NYC population and demographic characteristics. 2. Direct competitors in the locations.

Data Source

Data source 1: Wikipedia:
https://en.wikipedia.org/wiki/New_York_City;
https://en.wikipedia.org/wiki/Demographics_of_New_York_City

Web scraping techniques will be used to get population density data and demographic information from Wikipedia. From preliminary findings, the population density ranked from Manhattan, Brooklyn, Bronx, Queens and Staten Island. While on the other hand, Queens has the highest Asian population density, followed by Manhattan, Brooklyn, Staten Island and Bronx. In this scenario, Asian customers are our target customer.

Data source 2: Geocoding.

Geographical coordinates will be obtained using Python Geocoder which address/neighborhood of interest will be represented in latitude and longitude.

Date Source 3: Foursquare API.

These data source will be used to explore various neighborhoods in NYC and other hot pot restaurants venues within the neighborhood.

Finally, data extracted from the above data sources will be merged into one data frame which will be further analyzed.

Methodology section:

- Collect the new york city data from Wikipedia
- Using FourSquare API we will find all venues for each neighborhood.
- Filter out all venues that are Hotpot Restaurants from New York Open Data.

Analysis

So, I did the analysis and got exciting results. To see full code click here.

Results

In order for a well-known, name-branded hotpot restaurant to successfully expand their footprint in North America, NYC is already pre-determined due to its diversity and Asian population density.

jurisdiction	%_white	%_black_or_african_amercian	%_Asian	%_other	%_mixed_race	%_hispanic_latino_of_other_race
Queens	44.1	20.0	17.6	12.3	6.1	25.0
Manhattan	54.4	17.4	9.4	14.7	4.1	27.2
Bronx	29.9	35.6	3.0	25.7	5.8	48.4
Staten Island	77.6	9.7	5.7	4.3	2.7	12.1
NYC Total	44.7	26.6	9.8	14.0	4.9	27.0

Population Date from Wikipedia

Borough	Amount_of_ Restaurant	Amount_of_Hotpot_ Restaurant	%_Asian_ Population
Queens	11447	121	17.6
Manhattan	10800	91	9.4

Consolidated Restaurants and demographic data

Our analysis shows that although there is a great number of restaurants in NYC (~40000 out of 5 boroughs), however, only three boroughs already have at least one hotpot restaurant, which are Manhattan, Queens and Brooklyn. Although each one of those boroughs all have roughly 10,000 restaurants, in terms of hotpot, Queens has the highest count, 121, while Brooklyn only has 69 in total. Thus, Brooklyn is eliminated from the selection and our final choices become Queens and Manhattan.

Discussion

Throughout this project and searching for data, I notice that even though New York City is a highly diverse region, in terms of asian population (out target customers for the client), only Manhattan, Queens and Brookyln have a resonable amount of asian population and director hotpot restaurant compeition. This could also be a trememdous business opportunity to expand into non-asian customers or data is not well documented for boroughs of Staten Island and Bronx.

Conclusion

In this study, I analyzed the population & demographic data in New York City as well as, restaurant and hotpot restaurant data. I identified key factors which should be used when in location selection, which is the area where the target customers live in and popularity of the restaurant type in that area, where data is acquired from FourSquare and New York Open Data.