Coursera Capstone Project: Applied Data Science

Opening a Chinese Restaurant in Brooklyn, New York

Bingqing Ge
gebingqing@gmail.com

New York University

1 Introduction

New York is the largest city in the United States with a long history of international immigration. Over the last decade the city has been growing faster than the region. The New York region continues to be by far the leading metropolitan gateway for legal immigrants admitted into the United States. The New York City Metropolitan Area contains the largest ethnic Chinese population outside Asia. In the recent decades, American people especially show more and more interests in Chinese Food. However, when it comes to Chinese food, Americans have a stereotype of cheap, dirty, unhealthy, etc. It's hard to people to associate Chinese restaurants with high-end restaurants.

1.1 Business Problem

Our stakeholder is willing to open a high-end Chinese restaurant in New York City with middle-high level prices. Of course, choosing a location for business is one of the stressful and controversial tasks, since there are a lot of criteria that have to be satisfied in order to achieve the highest revenue. Here are some of them:

- the density of other restaurants
- the density of specifically Chinese restaurants
- population density around the location

- solvency of the population around the location
- Neighborhood HH Income, etc.

In this project, we will implement the basic analysis and try to find the most optimal Borough to open the high-end Chinese restaurant according to those criteria. It's obvious, that there are many additional factors, such as distance from parking places or distance from the main streets, but this analysis can be done after choosing the Borough, and thus will not be done within the scope of this project.

2 Data

2.1 Data Description

Based on criteria listed above the following data will be utilized in our analysis:

- the number of restaurants within the certain radius of each borough (Foresquare API)
- the net income per person in each borough. Since the restaurant will have middle-high prices, it is important to consider the solvency of population. Source: Kaggle (https://www.kaggle.com/goldenoakresearch/us-household-income-stats-geo-locations/version/8#)
- the population and the population density of the borough. Source: NYC Open Data (https://data.cityofnewyork.us/City-Government/New-York-City-Population-By-Neighborhood-Tabulatio/swpk-hqdp)
- the coordinates of the borough. Source: Kaggle (https://www.kaggle.com/goldenoakresearch/us-household-income-stats-geo-locations/version/8#)

2.2 The final Data Set to use for analysis will be as below:

Neighborhood	Median Income	Population	Asian	AA	Hispanic	White	Neighborhood Latitude	Neighborhood Longitude	Venue Latitude	Venue Longitude	Likes	Tips	American Restaurant
Bay Ridge	69989.0	123488	0.282	0.031	0.158	0.50900	40.625801	-74.030621	40.624251	-74.030346	40.625000	40.625000	0.000000
Bensonhurst	54513.0	205850	0.389	0.018	0.155	0.40400	40.611009	-73.995180	40.612326	-73.996229	11.750000	11.750000	0.000000
Sunset Park	57870.0	144332	0.327	0.024	0.399	0.22700	40.645103	-74.010316	40.646869	-74.009024	37.166667	37.166667	0.000000
Greenpoint	78069.0	152002	0.070	0.036	0.216	0.64500	40.730201	-73.954241	40.730615	-73.955152	176.666667	176.666667	0.000000
Sheepshead Bay	62443.0	171030	0.052	0.080	0.637	0.62443	40.586890	-73.943186	40.584168	-73.944172	62.909091	62.909091	0.090909
Flatbush	57678.0	150707	0.088	0.289	0.150	0.44200	40.636326	-73.958401	40.635423	-73.961846	9.142857	9.142857	0.000000
Crown Heights	61253.0	141725	0.056	0.557	0.117	0.24100	40.670829	-73.943291	40.670715	-73.944628	5.000000	5.000000	0.000000
East Flatbush	50290.0	140087	0.021	0.871	0.062	0.02700	40.641718	-73.936103	40.641618	-73.936250	10.000000	10.000000	0.000000
Prospect Heights	61253.0	141725	0.056	0.557	0.117	0.24100	40.676822	-73.964859	40.676587	-73.963800	177.900000	177.900000	0.000000
Brownsville	20640.0	111511	0.019	0.713	0.198	0.04300	40.663950	-73.910235	40.663760	-73.911446	1.600000	1.600000	0.000000
Williamsburg	78069.0	152002	0.070	0.036	0.216	0.64500	40.707144	-73.958115	40.709014	-73.956971	305.875000	305.875000	0.000000
Bushwick	51622.0	140474	0.056	0.170	0.539	0.21500	40.698116	-73.925258	40.699513	-73.926408	94.100000	94.100000	0.000000
Bedford Stuyvesant	52897.0	142027	0.027	0.488	0.194	0.26600	40.687232	-73.941785	40.684215	-73.943584	95.142857	95.142857	0.000000
Brooklyn Heights	94327.0	135444	0.090	0.258	0.149	0.47200	40.695864	-73.993782	40.695981	-73.993911	74.800000	74.800000	0.000000
Carroll Gardens	137375.0	116209	0.072	0.061	0.172	0.63900	40.680540	-73.994654	40.680919	-73.994313	192.142857	192.142857	0.000000
Fort Greene	94327.0	135444	0.090	0.258	0.149	0.47200	40.688527	-73.972906	40.689098	-73.971816	377.333333	377.333333	0.000000
Park Slope	137375.0	116209	0.072	0.061	0.172	0.63900	40.672321	-73.977050	40.670396	-73.978731	128.000000	128.000000	0.000000

```
nycdata_new.describe()
```

	Median Income	Population	Asian	AA	Hispanic	White	Neighborhood Latitude	Neighborhood Longitude	Venue Latitude	Venue Longitude	Likes	Tips
count	23.000000	23.000000	23.000000	23.000000	23.000000	23.000000	23.000000	23.000000	23.000000	23.000000	23.000000	23.000000
mean	67338.173913	148251.217391	0.098043	0.305435	0.207087	0.379845	40.657399	-73.958955	40.657465	-73.959119	80.495840	80.495840
std	28293.484119	30360.227639	0.099452	0.275905	0.145533	0.215230	0.038109	0.036380	0.038217	0.036390	103.319799	103.319799
min	20640.000000	111511.000000	0.018000	0.018000	0.062000	0.027000	40.574293	-74.030621	40.575665	-74.030346	0.500000	0.500000
25%	51834.000000	129466.000000	0.047000	0.048500	0.118500	0.225500	40.634347	-73.989591	40.635324	-73.988546	7.071429	7.071429
50%	61253.000000	141725.000000	0.070000	0.258000	0.158000	0.404000	40.663950	-73.958115	40.663760	-73.956971	37.166667	37.166667
75%	78089.000000	152002.000000	0.090000	0.557000	0.207000	0.589215	40.683886	-73.938944	40.682567	-73.939917	111.571429	111.571429
max	137375.000000	215637.000000	0.389000	0.871000	0.637000	0.689000	40.730201	-73.880699	40.730615	-73.879535	377.333333	377.333333