Final Project Report

The Best Place for a High-end Restaurant (Toronto)

Ming Zhuang

Introduction:

Toronto is an international city locate at Ontario, Canada. People come from different counties work, study or immigrate to Toronto. This is the reason why this multi-culture city has different style restaurants, such as Chinese restaurants, Japanese restaurant, French restaurant, Mexico Restaurant etc. People in this city are inclusive and willing to attempt new things.

This project is going to find the best place to open a new and high-end western-style restaurant. The cost of the meal in the restaurant would high than fast-food restaurant. The restaurant provides good elaborate food and well design decoration. In order to find the best location that can attract more people, there are some factors that need to consider: 1, average income of neighborhood, 2, Population of neighborhood, and the 3, competition.

Data:

Data I use is from Wikipedia. Demographics of Toronto Neighborhood provide the variety characteristic of each neighborhood, include population, average income, second language beside English. As discuss above, I will use the population and average income.

(The data use
is https://en.wikipedia.org/wiki/Demographics_of_Toronto_neighbourhoods to explore neighborhood average income and population)

In order to access the location of each neighborhood, I use the List of Postal Codes of Canada and Postal Codes with these latitudes and longitudes.

(https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M and http://cocl.us/Geospatial_data to explore location of neighborhood)

To find out the competitor of neighborhood. I use Foursquare to access each neighborhood venues. The venues defined as restaurant would be the competitor

(Foursquare API to explore the competitors)

Data Cleaning:

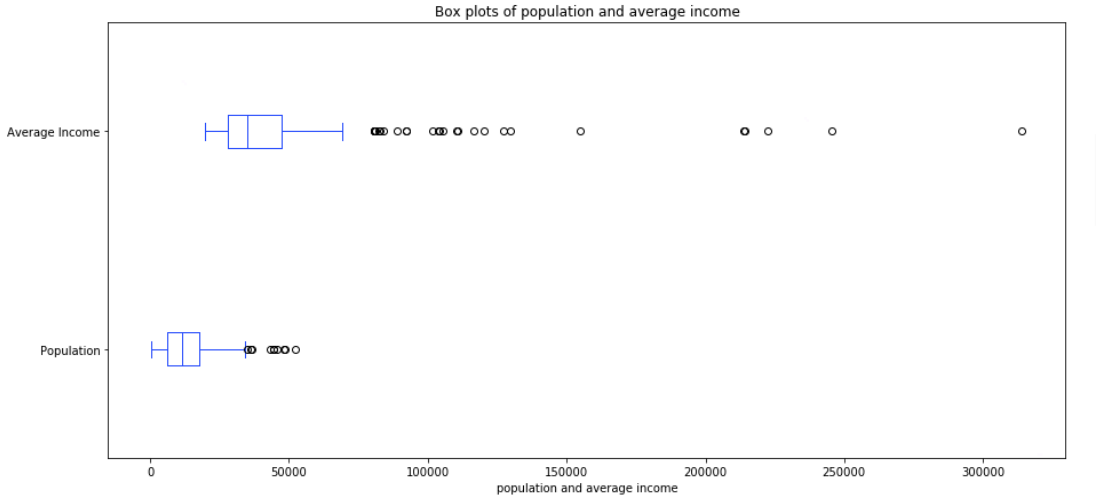
1, For the data of Demographics of Toronto Neighborhood, there are so so many information contains in the table that show at the Wikipedia website. I drop most of them and leave only population and average income. After that, there are no Nan value contain in the table

2, For the data the List of Postal Codes, I drop all the row with Nan value.

3, For the Venues in Foursqure, I only use the venues that contain 'restaurant' in there type of venues and then drop all the fast food restaurant.

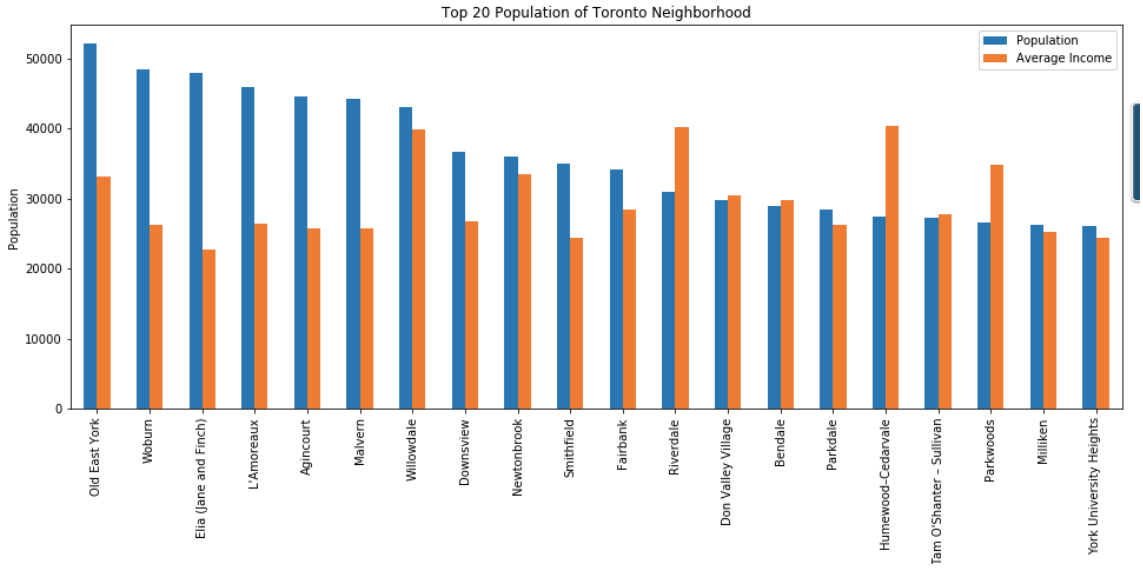
Analysis:

The population and average income is two main factors that need to explore. Firstly, let's look at the distribution of data of population and average income:

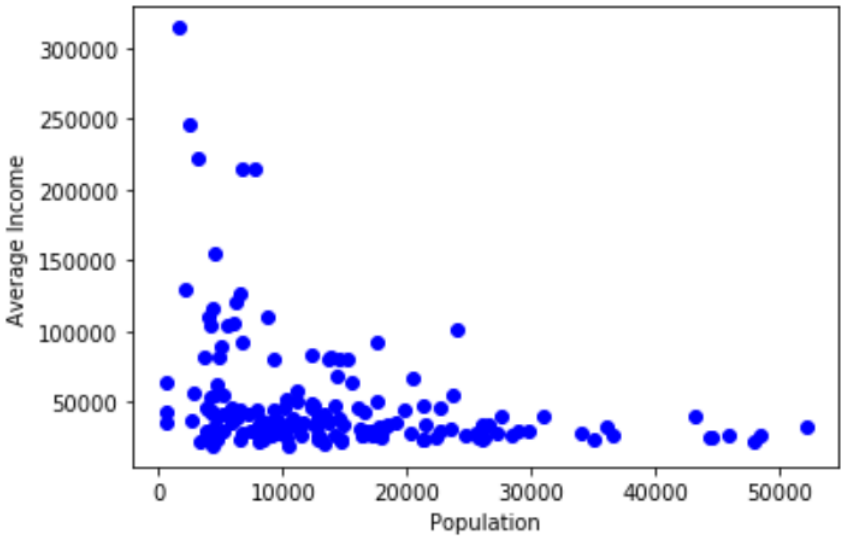


We can see the only few of neighborhood has average income above $100,000. If you have the income that around 75000, you would be the outlier of data. and most of them are at below $50,000 level. Most of population is below 25000.

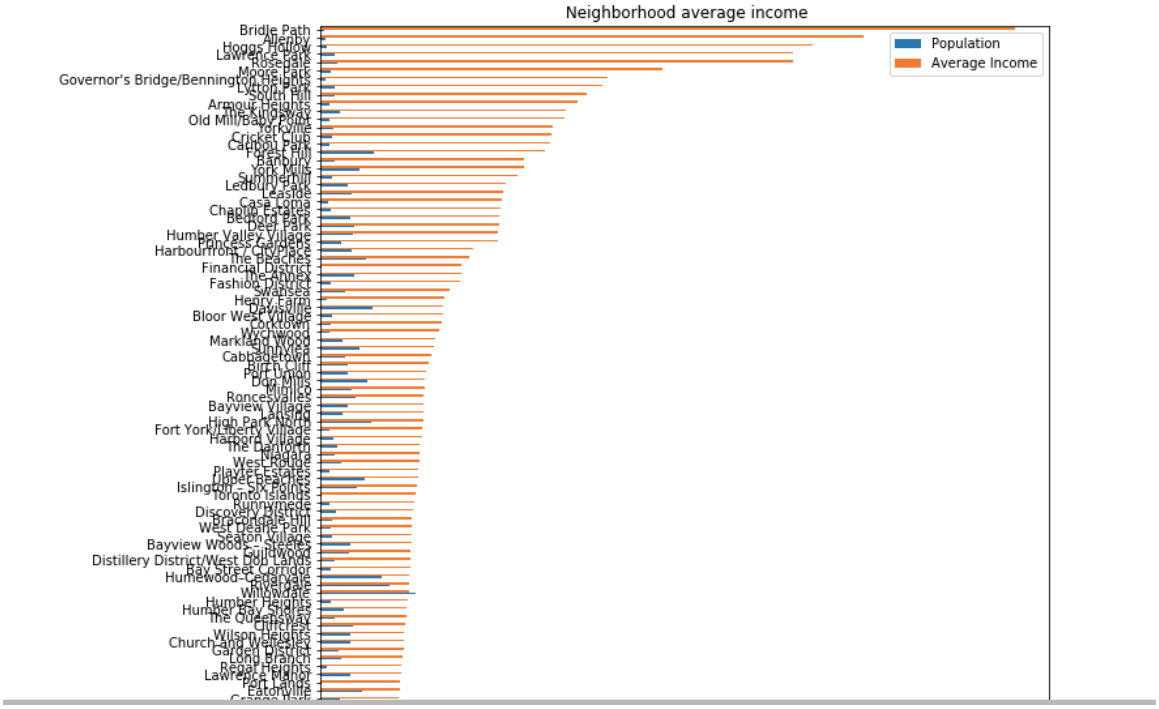
Then, I look at the largest 20 population neighborhood in Toronto:



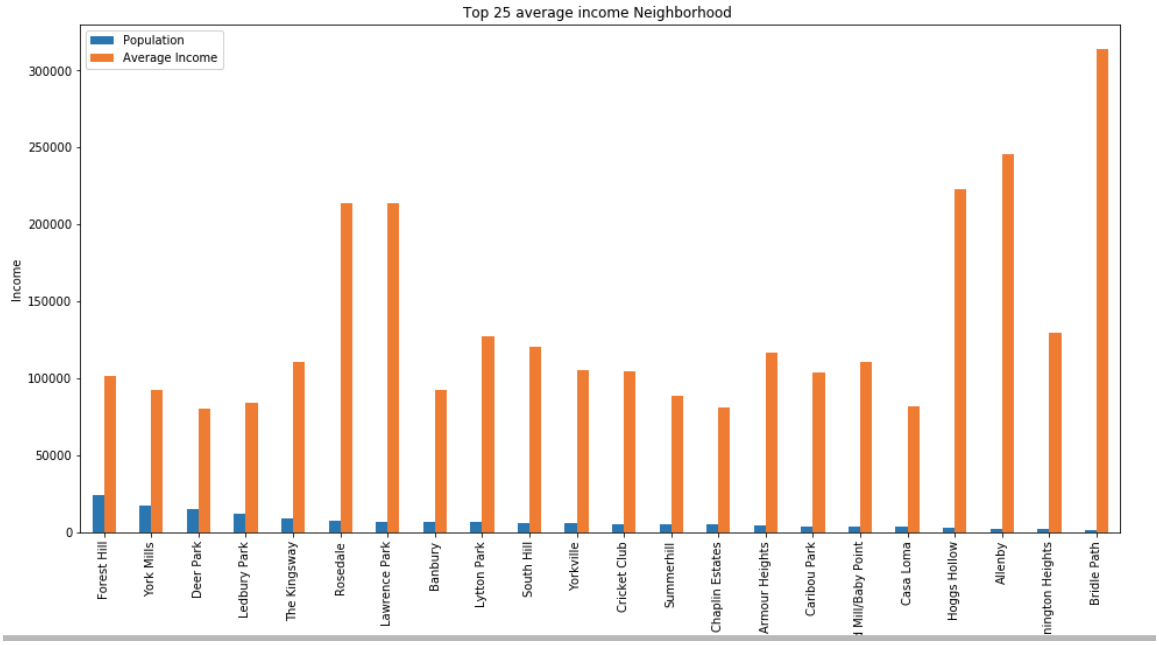
This is hard to find out the common neighborhood. Then I use regression machine learning to explore Toronto neighborhood relationship between population and the income. The coefficient of -1.226, the scatter diagram is:



The coefficient and scatter diagram show that they don't have close relationship, positive or negativeTherefore, I decide that the high income people is the most important target customers. I would leave the population at the last step. Then I make a bar chart to access all neighborhood average income



I find out that top 25 is much higher that rest of the neighbor so I look closer and gain more information of top 25 average income neighborhood.

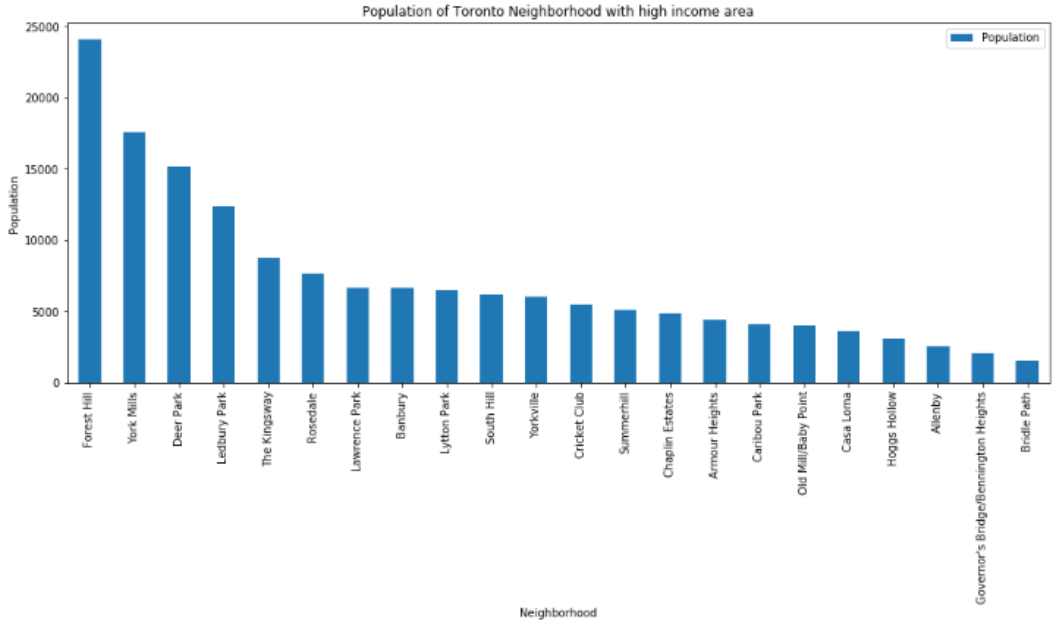


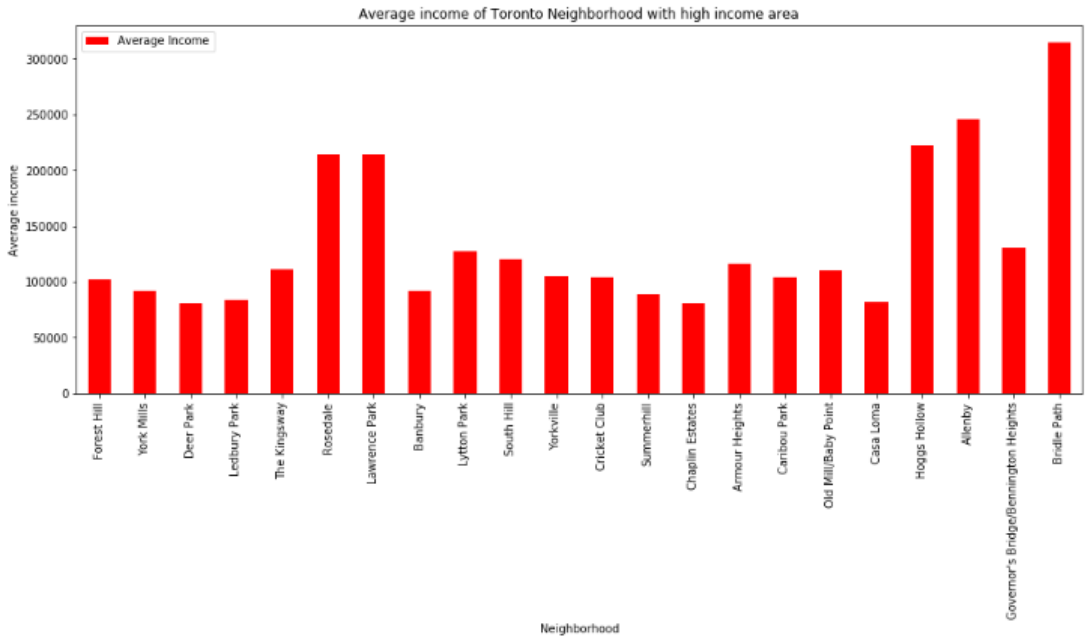
Then I access all venues of this 25 neighborhood, I set the radius to 1200 to explore wider range. At this point, I did not include the fast-food restaurant since I believe fast food restaurant would not target the same customer with a high-end restaurant. Then I found:

Neighborhood	index
Bedford Park	11
Leaside	3
Moore Park	1

There are 11 restaurants at Bedford Park, 3 restaurants at Leaside and 1 at Moore Park. In order to avoid directly compete, I would not consider this 3 neighborhood.

Then, I make the bar charts of population and average income that are not include Bedford Park, Leaside, and Moore Park.





Discuss:

If only look at the population of top 25 average income neighborhood (exclude Bedford Park, Leaside, and Moore Park), Forest Hill, York Millis, Deer Park and Ledbury Park have over 10 thousand populations. But they have relative low income at this 25 neighborhood. But I believe that it is good enough to open a restaurant nearby. Then if

we look at the Rosedale and Lawrence Park, they have very high average income and the population is around 5000, I think this two Neighborhood also is considerable.

Results:

After access population, average income and competitors of each neighborhood, I would consider Forest Hill, York Millis, Deer Park and Ledbury Park firstly and then Rosedale and Lawrence Park.

Conclusion:

This report is to find out the best location for the high-end restaurant in Toronto, I use data that include the population, average income and the venues to explore a place that have big population and high average income and the less competitors. The place I find out include Forest Hill, York Millis, Deer Park and Ledbury Park firstly and then Rosedale and Lawrence Park.

In this report I use the python to clean and processing all data and make bar chart to explore the data each neighborhood. Also, I use regression find out the relationship between the population and average income. The package include pandas, sklean, matplotlib.