

# A better city for a new Chinese restaurant

## 1. Introduction

### 1.1 Background

Toronto and New York are the most populous city in Canada and United States, and the diverse population of these two cities reflect their current and historical role as an important destination for immigrants.

In Toronto, more than 50 percent of residents belong to a visible minority population group, and over 200 distinct ethnic origins are represented among its inhabitants. In 2016, 51.5 percent of the residents of the city proper belonged to a visible minority group, compared to 49.1 percent in 2011, and 13.6 percent in 1981. The largest visible minority groups were South Asian (338,960 or 12.6 percent), Chinese (332,830 or 12.5 percent), and Black (239,850 or 8.9 percent). In 2016, the three most commonly reported ethnic origins overall were Chinese (332,830 or 12.5 percent), English (331,890 or 12.3 percent) and Canadian (323,175 or 12.0 percent). Common regions of ethnic origin were European (47.9 percent), Asian (including middle-Eastern – 40.1 percent), African (5.5 percent), Latin/Central/South American (4.2 percent), and North American aboriginal (1.2 percent). In conclusion, Toronto is an important destination for immigrants, especially for Chinese.

New York also has the same situation. New York contains the highest total Asian population of any U.S. city proper. The city's population in 2010 was 44% white (33.3% non-Hispanic white), 25.5% black (23% non-Hispanic black), and 12.7% Asian. Asians constituted the fastest-growing segment of the city's population between 2000 and 2010. Asian Americans in New York City, number more than one million, greater than the combined totals of San Francisco and Los Angeles. The Chinese population constitutes the fastest-growing nationality in New York, In 2012, 6.3% of New York City was of Chinese ethnicity, New York State have become the top destinations for new Chinese immigrants, respectively, and large-scale Chinese immigration continues into New York City and surrounding areas, with the largest metropolitan Chinese diaspora outside Asia, including an estimated 812,410 individuals in 2015.

### 1.2 Problems

It is not easy to start a new career in a different country. Although there has a lot of Chinese people in the two cities above, it does not mean that it will be very successful if you start to manage a Chinese restaurant everywhere. it already has a lot of Chinese restaurant in these cities, so it is necessary to analysis every part of these cities, in order to make sure choose a better city for the future career successfully.

### 1.3 Benefits

The Chinese population constitutes the fastest-growing nationality in these cities, some of them may want to begin a new life as become a new Chinese restaurant owner. It is very important to choose which city they need to start, and which city can make their restaurant more successful.

## 2. Data acquisition and cleaning

### 2.1 Data sources

Neighborhood in New York City has a total of 5 boroughs and 306 neighborhoods. this dataset exists for free on the web, here is the link to the dataset: [here](#).

In Toronto ,I use [here](#) in order to obtain the data that is in the table of postal codes and [here](#) is a link to a csv file that has the geographical coordinates of each postal code.

In addition, the document about city size and population density in New York City is [here](#), and in Toronto is [here](#).

### 2.2 Data cleaning and Feature selection

#### Neighborhood in New York city

The data of neighborhood in New York city has a total of 5 boroughs and 306 neighborhoods. In order to segment the neighborhoods and explore them, it is essential to need a dataset that contains the 5 boroughs and the neighborhoods that exist in each borough as well as the latitude and longitude coordinates of each neighborhood. all the data I need is in the features key, which is a list of the neighborhoods, so it is necessary to transform data['feature']['properties']['borough'] , data['feature']['properties']['name'] and data['feature']['geometry']['coordinates'] into a pandas data frame. The result is as show below:

	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

```
neighborhoods.shape
```

```
(306, 4)
```

Then, we can use this data frame to explore all venues in New York city by utilizing the Foursquare API.

### Neighborhood in Toronto

It is essential to ignore which borough is Not assigned. And if more than one neighborhood can exist in one postal code area (combined into one row with the neighborhoods separated with a comma), I separate them in two rows. What is more, if a cell has a borough but a Not assigned neighborhood, the neighborhood will be the same as the borough.

In order to utilize the Foursquare location data, it is necessary to get the latitude and the longitude coordinates of each neighborhood. There has a csv file that has the geographical coordinates of each postal code and I combined two files together by using the same postal code. The result as below:

	Postal Code	Borough	Neighborhood	Latitude	Longitude
0	M3A	North York	Parkwoods	43.753259	-79.329656
1	M4A	North York	Victoria Village	43.725882	-79.315572
2	M5A	Downtown Toronto	Regent Park	43.654260	-79.360636
3	M5A	Downtown Toronto	Harbourfront	43.654260	-79.360636
4	M6A	North York	Lawrence Manor	43.718518	-79.464763

```
: neighborhoods.shape
```

```
: (208, 5)
```

Then, we can use this data frame to explore all venues in Toronto by utilizing the Foursquare API.

### New York Population

we use 1960-2020 to forecast the future 10 years' population in New York city.

Year	Population	Growth Rate
2020	18,804,000	-0.01%
2015	18,648,000	0.31%

Year	Population	Growth Rate
2010	18,365,000	0.31%
2005	18,087,000	0.31%
2000	17,813,000	0.86%
1995	16,943,000	1.04%
1990	16,086,000	0.48%
1985	15,827,000	0.29%
1980	15,601,000	-0.22%
1975	15,880,000	-0.39%
1970	16,191,000	0.94%
1965	15,177,000	1.39%
1960	14,164,000	1.40%

Population of Toronto

we use 1960-2020 to forecast the future 10 years' population in Toronto.

Year	Population	Growth Rate
2020	6196731	0.011
2015	5867292	0.013
2010	5499233	0.0178
2005	5035232	0.0179
2000	4607142	0.0188
1995	4197157	0.0197
1990	3806957	0.0256
1985	3355459	0.0221
1980	3008032	0.0166
1975	2770072	0.0179
1970	2534788	0.0391

Year	Population	Growth Rate
1965	2092902	0.0371
1960	1744328	0.0503