Report #3: Best City for investing a Chinese Restaurant

Nicolas Fan

In this project, I suppose an investor came to consult me that he has around $250,000 free money in his bank account and he decided to open a Chinese restaurant at North America area. What should he do to ensure the business go successful, which is defined as short breakeven time and high investment return rate? Since I am a data scientist, I will use the Yelp Dataset to analysis and return a decent and achievable business suggestion plan to him.

The Yelp dataset is a subset of Yelp businesses, reviews, and user data for use in personal, educational, and academic purposes. It is available as JSON files. (<https://www.yelp.com/dataset>) The dataset contains 3 main business files: business.json, review.json and user.json. Business.json contains business data including location data, attributes, and categories. Review.json contains full review text data including the user\_id that wrote the review and the business\_id the review is written for. User.json includes the user's friend mapping and all the metadata associated with the user.

Currently, my algorithm will mainly focus on business.json file. Although it contains 192,609 businesses and takes 8.69 gigabytes uncompressed, most of the data is irrelevant to my project based on the wrong location. To clean the data, I would first delete lines with “state” attributes not equal to “MA” and then select the lines with “city” attributes equal to Boston area (Boston, Cambridge, Medford, Allston, Quincy, Malden and so on). After that, I need to select all the Chinese restaurant by attributes “categories”. Since there are different style Chinese restaurants, Yelp defined 13 different categories for them (Cantonese/ Congee/ Dim Sum/ Fuzhou/ Hainan/ Hakka/ Henghwa/ Hokkien/ Hunan/ Pekinese/ Shanghainese/ Szechuan/ Teochew).

After the date getting cleaned, we will have a dataset of Boston area Chinese restaurant. And then, I am going to find all the successful restaurant under $250,000 initial investment and analyze the characteristics and commonalities of these restaurants. To find successful restaurant, I will collect merchants with high ratings (4 or 5 stars) and the number of comments is greater than a certain value (e.g. review\_count > 50 ). To decide whether these restaurant’s initial investment is lower than $250,000, I will look at the location, search the rent price, device and employees needed.

{"business\_id":"-jdNqqzF1Dbve04oEd4jww","name":"OK OK Diner","address":"1128 Queen Street E","city":"Toronto","state":"ON","postal\_code":"M4M 1K8","latitude":43.6623207,"longitude":-79.3351048,"stars":3.0,"review\_count":62,"is\_open":1,"attributes":{"RestaurantsDelivery":"False","Ambience":"{'romantic': False, 'intimate': False, 'classy': False, 'hipster': False, 'divey': False, 'touristy': False, 'trendy': False, 'upscale': False, 'casual': True}","NoiseLevel":"u'average'","GoodForKids":"True","Caters":"False","RestaurantsTakeOut":"True","RestaurantsAttire":"u'casual'","Alcohol":"u'beer\_and\_wine'","OutdoorSeating":"True","RestaurantsGoodForGroups":"False","WiFi":"u'no'","HasTV":"False","RestaurantsReservations":"False","BikeParking":"True","BusinessParking":"{'garage': False, 'street': True, 'validated': False, 'lot': False, 'valet': False}","RestaurantsPriceRange2":"2"},"categories":"Restaurants, Breakfast & Brunch, Diners","hours":{"Monday":"9:0-16:0","Tuesday":"9:0-16:0","Wednesday":"9:0-16:0","Thursday":"9:0-16:0","Friday":"9:0-16:0","Saturday":"9:0-16:0","Sunday":"9:0-16:0"}}

Example row of the file business.json

**The characteristic of each variable**

business\_id used internally at Yelp company

name name of the business

address address of the business

city where is the business

state which state is the business in

postal\_code postal code of the business

latitude latitude of the location

longitude longitude of the location

stars Yelp uses stars to represent the overall rating of a business, ranged from 1 to 5

review\_count How many reviews do this business have

is\_open Is this business open or not

RestaurantsDelivery Do the restaurants provide delivery

Ambience a set of Booleans about ambience description

RestaurantsPriceRange2 Yelp uses 1 to 4 to represent price range but not sure about why it is called Range2

Categories Categories of the restaurant

BusinessParking a set of Booleans about whether the business has different parking

Hours When the business is running

To figure out how to make investor’s Chinese restaurant business successful, I am aiming to solve following sub-questions:

1. What’s the most popular style of Chinese restaurant at Boston?
   1. High stars percentage/ good review\_count percentage
   2. My first guess is Szechuan, bubble tea and hot pot.
2. Where should the investor open the restaurant?
   1. Are there places gathering good Chinese restaurant other than China town
   2. We all know street like Newburry street is the most popular for business, but does the high rent really pay off? Would a street next to university better? Would a far-away but low rent price plaza better?
3. What’s these restaurants’ estimate revenue?
   1. Based on the RestaurantsPriceRange2 and review\_count
4. What’s these restaurants’ early investments?
   1. Based on the address and floor area