The Chinese Restaurant in Downtown Toronto

Peining Li

01/18/2020

1, Introduction: Business Problem

In this project, I try to analysis Chinese Restaurant in Downtown Toronto. Since most people use apps like yelp to search for restaurant, the ratings for those spots are crucial. Price is one of the most important factors to consider when people are choosing restaurant. In this project, I will analysis the relationship between prices and ratings.

2, Data

The database I will use are exclusively from Foursquare. The following data will be utilized:
1, Chinese restaurants in Downtown Toronto
2, ratings of each restaurant
3, prices of each restaurant

At the beginning, I will try to find the geographic relationship for each restaurant and then cluster them using machine learning model. Then I will compare price of each restaurant in each cluster, trying to build a regress model.

2.1 Data source
Data is exclusively from Foursquare databases. In Downtown Toronto, Venus under the category of "Chinese Restaurant" will be used for data analysis

2.1 Data cleaning
From Geocoders, the longitude and latitude of Downtown Toronto is (43.6541737 -79.38081164 513409). Using this geographic location, I got a list of venues using the search query, "Chinese Restaurant". Then, I start to clean the data. At the beginning, I created a data frame to summariz e the result. There were 30 spots at that time. Then I filtered the information for the second time. I only kept the restaurants whose category is Chinese Restaurant. The final result was that there were five Chinese Restaurants in Downtown Toronto. I visualized the result on map.
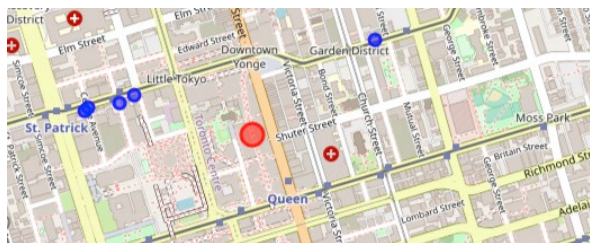
Later, I clustered them according to their geographic location. There were 2 cluster.
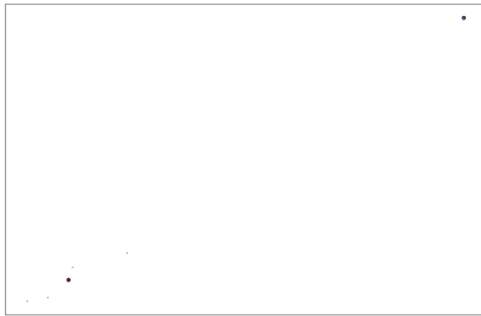

3, Methodology

I used Restaurants' ID to get more information about each restaurant. Then I discovered what keys are available in dataset. The I subtracted "Price" and "Ratings" for each restaurant. Here is the final dataset for information I collected.

	ID	Price tier	Rating
0	4b2027b5f964a520f82d24e3	2	7.1
1	4adf80b9f964a5202f7b21e3	1	6.2
2	4ad4c060f964a5205cf720e3	3	6.6
3	52a7ae41498eed3af4d0a3fa	1	7.4
4	4c69740b8d22c9284d42b745	1	6.1

Then I made a plot for to visualize.
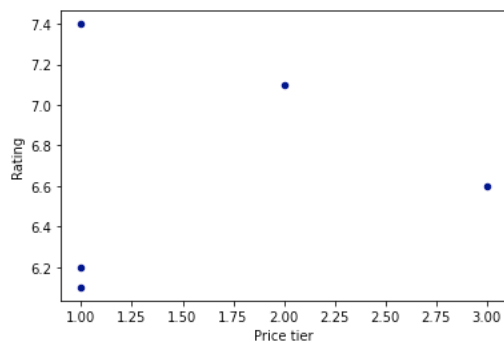
After this, I tried to build a regression model for ratings and price.
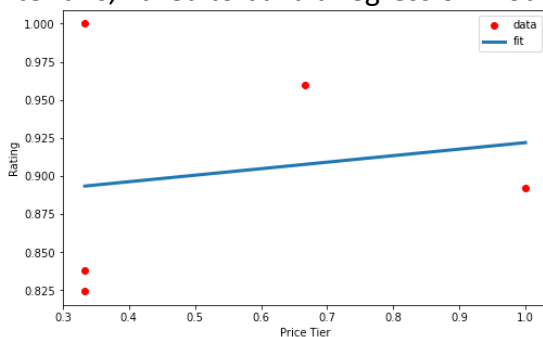
4, Result and Discussion

The final model is $y=0.92x^{0.028}$, which is also a non-linear relationship between price and ratings for Chinese restaurant in Downtown Toronto. According to the graph, it does not seem like a good fit for our data. There are many reasons leading to the result:
1, limited data available for analysis. The number of available Chinese Restaurant in Downtown Toronto is only 5. This number cannot provide statistically significant results. 2, Ratings can be related many other factors. However, there is not enough quantitative date in Foursquare database. This also prevent building a better model for ratings.

the model also is not trained because there is not enough data. In the future, more data is needed for discovering ratings

5, Conclusion

There is no relationship between price and ratings for Chinese Restaurant in Downtown Toronto.