Best Place to live for Food Lover

Final Report
for
IBM Data Science

Weihan Luo

2019-06

 # Introduction: The best place to live for food lover in Toronto

In this project I will trying to find the best area in Toronto for food lover.
I will find out the the restaurants type like Breakfast spot, bakery, dinner, fish market, noodle house, pizza, cafe, sandwich and much more!
I will show which area has the most and best restaurant in Toronto.

 # Data Base

Base on our goal, I will need the following data:

1. Type of resturant (food).
2. Numbers of the resturant.

I will use FourSquare for the geo service and Jupyter Notebook as the tool of the data anlyze.

Main Article

Part One: Find out all the neighborhood in the Scarborough.

In this part, I will use the post code to identify all the neighborhood inside the scarborough.

The data is coming from:

https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

* Wikipedia Site.

Part Two: Collecting data from "FourSquare" and reform the data

After the part one (identify all the neighborhood of Scarborough), I will connect the Jupytor Notebook with FourSquare API to collect the information about venues inside each neighborhood. For each neighborhood I picked the range of 1km as the radius which means the foursquare will gather all the information within 1km from the center of the neighborhood.

Part Three: Analyze the reformed data and set up a new data frame for all the inside the Scarborough.

After all the data I need has been fully collected and reformed, I will start the searching of the desirable data from the new data set. At this point, I will also sort out the type of the restaurant and store in details. For example, for than "Asian dishes" I will list out restaurant type such as "Noodle house", "Shanghai food", "Korean BBQ ", "Cafe" and even restaurants that's focus on dinner and breakfast. After this step, out dataset is ready to process the machine learning step.

Part Four: Machine Learning with K-means clustering.

Under this stage, the neighborhood were dividend to 5 clusters. After the grouping we can see which one has the most restaurant.

Our final decision will be SUM SCORE= NUMBER OF RESTAURANT + RATING