This project is for our customer Mrs. Ju. In view of her requirement, we focused on the info of Scarborough, Toronto. Our task is help her find a good place to open a pet store, and we have to satisfy her criteria.

The data we mainly used is the part of Scarborough neighborhoods, which we got from Wikipedia. We also make use of detailed information of venues and instruments there, which we attained with the help of Foursquare.

To explore the data, we use Foursquare-api to get the info of further details of each venues in those neighborhoods, so that we could group them and find features of neighborhoods. Meanwhile we wrangle the data and drop those data which we are not interested in. During the later process, K-means clustering is what we use to perform machine learning. For that it could help us divide those neighborhoods into different clusters with their similarities. After such clustering, I could make it clear that which of them are suitable for our customer Mrs. Ju’s criteria, so that we could make proper decision for her.

As the results shows, 12 of the neighborhoods are more focusing on bank and spa, which are not the most significant factors in our customer’s mind. Neighborhood Agincourt and Scarborough Village are of same problem. While Milliken, Agincourt North, Steeles East, L'... are more focused on park and Birch Cliff , Cliffside West are focused on cafeteria, bank and spa.

From the results, combining the requirements of our customer, we could make the decision that it would be better to locate at Milliken, Agincourt North, Steeles East, L'..., which perfectly match her criteria. However, in this whole process, we find that whatever the cluster is, the neighborhoods there are all equipped with banks. It seems not a factor we need to take into consideration for that it has covered the whole borough. We’re not sure whether it has done side effects to our results.

In conclusion, as my first task, it is not that good. I guess it could merely be defined as “pass”. In a more formal analysis process, I should collect more types of data such as the income and age distribution in those neighborhoods. And K-means is not the best machine learning for this task, I should manage more skills in detail. Dear classmate, thank you for your reading all of my report. It’s an endpoint for our learning process in this course. But I believe it would also be a new start point for both of us. Carry on. Hope you would either achieve where you want to go.