Good Location for a Pet Store

1. Introduction

1.1 Background

As more and more people nowadays tend to keep themselves accompanied by some warm creatures, opening a pet store becomes a nice choice for newly coming investors.

Recently, we've been required by Mrs. Ju to choose a good location for her, to open a pet store in Scarborough, Toronto. She once was a housewife with a bachelor degree of economy, having been bored with just staying at home, and finally decided to create her own business. Due to family reasons, she want to locate the store within Scarborough so that it would still be convenient for her to take care of her child, together with her husband.

1.2 Problem

In short, 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.

It is also advised by her that, we should select a place where surrounded by parks, where pet lovers are more likely to gather, and this is the most important. Meanwhile, there would better be supermarkets, cafeterias or shopping malls, which attracts more people to visit there, bringing her more potential customers. What's more, it would be great if there are some places which could indicate residents' high-quality life, such as spa. By the way, a bank might also be of some effect for that customers would be more possible to carry with more currency, which could stimulate their desire for shopping, to some degree.

2. Data Preparation

2.1 Data sources

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.

2.2 Data wrangling and feature selection

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. In view of our customer's requirement, we finally leave these columns of information: "Cafe", "Park", "Bank", "Supermarket", "Clothing shop", "Spa".

3. Data Exploration

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. We also create a map for it, marking those clusters with different colors on the map. 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.

4. Results

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.

	Neighborhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
12	Agincourt	0.0	Clothing Store	Bank	Spa	Supermarket	Park

	Neighborhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Malvern , Rouge	1.0	Bank	Spa	Supermarket	Clothing Store	Park
1	Rouge Hill , Port Union , Highland Creek	1.0	Bank	Spa	Supermarket	Clothing Store	Park
2	Guildwood , Morningside , West Hill	1.0	Bank	Spa	Supermarket	Clothing Store	Park
3	Woburn	1.0	Bank	Spa	Supermarket	Clothing Store	Park
4	Cedarbrae	1.0	Bank	Spa	Supermarket	Clothing Store	Park
6	Kennedy Park , Ionview , East Birchmount Park	1.0	Bank	Spa	Supermarket	Clothing Store	Park
7	Golden Mile , Clairlea , Oakridge	1.0	Bank	Spa	Supermarket	Clothing Store	Park
8	Cliffside , Cliffcrest , Scarborough Village West	1.0	Bank	Spa	Supermarket	Clothing Store	Park
10	Dorset Park , Wexford Heights , Scarborough To...	1.0	Bank	Spa	Supermarket	Clothing Store	Park
11	Wexford , Maryvale	1.0	Bank	Spa	Supermarket	Clothing Store	Park
13	Clarks Corners , Tam O'Shanter , Sullivan	1.0	Bank	Spa	Supermarket	Clothing Store	Park
15	Steeles West , L'Amoreaux West	1.0	Bank	Supermarket	Spa	Clothing Store	Park

	Neighborhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
5	Scarborough Village	2.0	Spa	Bank	Supermarket	Clothing Store	Park

	Neighborhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
14	Milliken , Agincourt North , Steeles East , L'...	3.0	Park	Bank	Spa	Supermarket	Clothing Store

	Neighborhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
9	Birch Cliff , Cliffside West	4.0	Café	Bank	Spa	Supermarket	Clothing Store

5. Discussion

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.

6. Conclusion

In conclusion, it is a process of choosing location for a new pet store in Scarborough, Toronto, making use of machine learning skills. 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. If possible, there would better be questionnaire for residents there to know their attitude about pets.

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.