1. Introduction

1.1 Background:

John has been investing Canadian real-estate for years, and he has seen his business growing with the increase of Chinese students or employees who accommodating in apartments that he owns. Starting from 2019, John decided to expand his investment by buying another apartment in Toronto, aiming at Chinese tenants with middle-class background.

Over the years of experience, John knows that apartments with Chinese grocery stores or Chinese restaurants nearby are more popular with Chinese tenants. In addition, a safe and quiet neighborhood near the main street is also preferable.

1.2 Problem:

The major features of the apartment that John is looking for are as following:

- With Chinese grocery stores nearby
- With Chinese restaurants nearby
- Better with a mall or city-center environment

1.3 Interest:

Selecting location has always been of my personal interest, and looking for a better investment choice will benefit in the long-term. Moreover, with the increasing popularity of Airbnb, more and more people are choosing investing apartments as their first choice, it is very interesting to see which features affect the future of an investment choice most.

2. Data Description

Data used in this report were mainly acquired from Wikipedia and Foursquare, which include the venue, service categories, location, ratings and tips.

2.1 Data Scraping

Raw data were scraped from wikipedia (https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M) which were later cleaned and transformed into pandas data-frame for better analysis.

2.2 Location Analysis

The main location source used in the analysis is Foursquare, which provide the neighborhood information of venue, service categories, location, ratings and tips.

3. Methodology

Methodology section which represents the main component of the report where you discuss and describe any exploratory data analysis that you did, any inferential statistical testing that you performed, if any, and what machine learnings were used and why.

3.1 Summary

K–means clustering was applied to find the location similar to city centers and malls, while the other two requirements were solved by the venue and ratings information gained from Foursquare API.

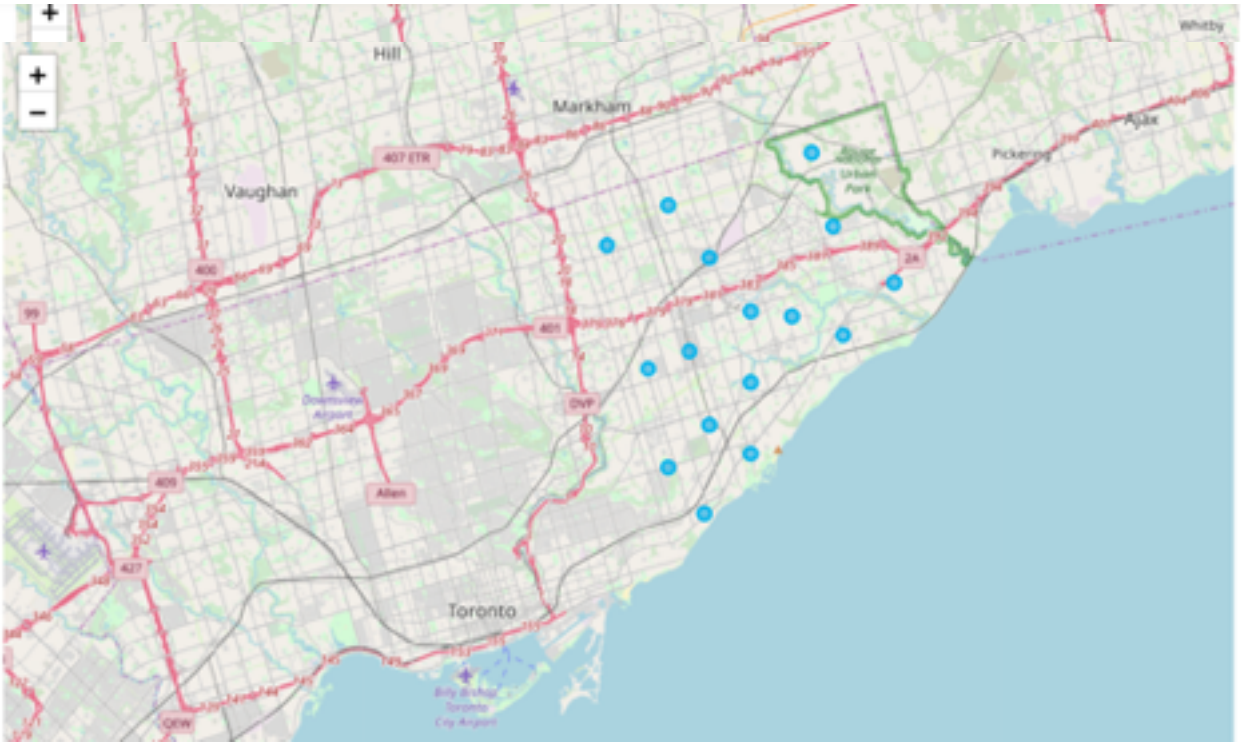
To be specific, K–means clustering enables neighborhoods to be clustered into different segmentations, which also allows each segmentation to be simulated as the city center. In addition, the best k value and the centroids were identified in K–means clustering. Being the most popular unsupervised machine learning algorithm, k–means clustering is the best option to define the similarity between city center and a certain area in Toronto. And information acquired from Foursquare provides the details of an area with Chinese grocery stores and restaurants, which acted as the indicator of choosing an area or not. Another convenient feature of using Foursquare API is that Foursquare API returns venues and details of a particular shop.

3.2 Procedures

I. 55 neighborhoods were found during the process of gaining city segmentation, and a map of Toronto and Scarborough was generated as below:

II. 10 venues were generated after accessing Foursquare location data, and onehot technique contributed to the process.

III. 55 neighborhoods were segmented into 5 categories by applying K–means clustering.



IV. City center was identified to cluster 2, and the other clusters were set into Scarborough:

V. 3 neighborhoods were selected since they have Chinese restaurants as one of their top venues, which was from filtering the data-frame by searching for Chinese restaurants and

grocery stores.

M1K	Scarborough	East Birchmount Park, Ionview, Kennedy Park
M1P	Scarborough	Dorset Park, Scarborough Town Centre, Wexford Heights
M1W	Scarborough	L'Amoreaux West, Steeles West

VI. Filtering each segmentation by distance from each centroids. Specifically, 5 grocery stores in districts East Birchmount Park, Ionview, Kennedy Park, however, only one has the rating details, which was Ghadir Mid-Easton Grocery with a rating of 8.1. As for the district Dorset Park, Scarborough Town Centre, Wexford Heights, only one grocery store was found with no ratings. For L'Amoreaux West, Steels West, The Low Carb Grocery is found but it only has a rating of 7.2.

VII. Finally, when exploring the Chinese restaurants around the rest neighbors, which were with only similar restaurants rated as 6.48, after removing "L'Amoreaux West, Steels West"

3.3 Evaluation

The selected two neighborhoods were quite similar, and they were not far from each other, therefore, this neighborhood that they locate on could be considering as the preferred area for investment.

4. Results

The two districts, i.e. ‘Dorset Park, Scarborough Town Centre, Wexford heights’, and “L’Amoreaux West, Steels West” can be selected in terms of Restaurants and Grocery Stores. Moreover, they both locate in with the cluster 2, which is Scarborough, and can be considered as the ideal place meeting the requirements of John.

5. Discussion

As mentioned above, the selected two neighborhoods were quite similar, and they were not far from each other, therefore, this neighborhood that they locate on could be considering as the preferred area for investment.

And the limitations of this analysis include the possible neglect of of neighbors with more Chinese grocery stores and restaurants since it considered more frequency than the amount.

Lastly, the transport factor was not considered in the investment plan, which can be very important in reality. The adoption of 10km parameter of search venues may lead to long distance of traveling and cause inconvenience for tenants.

6. Conclusion

In this project, we first identified the business problem, which was followed by the data acquisition and preparation. As for machine learning , we utilized K-means clustering and Foursquare API. To conclude, “Dorset Park, Scarborough Town Centre, Wexford heights”, “L’Amoreaux West, Steels West” meets the requirement of our client but better choices may exist. The finding of this project can also help the relevant investors with similar requirements.