SHALL WE OPEN A NEW VIETNAMESE RESTAURANT IN TORONTO

IBM DS Capstone Project

INTRODUCTION

Vietnamese food is not only delicious but also healthy, thus I would like to recommend it to the community in Toronto.

I suppose there may not be enough Vietnamese restaurants in Toronto and it might present a great opportunity for this entrepreneur who is based in Canada. As Vietnamese food is very similar to other Asian cuisines, this entrepreneur is thinking of opening this restaurant in locations where Asian food is popular (aka many Asian restaurants in the neighborhood).

With the purpose in mind, finding the location to open such a restaurant is one of the most important decisions for this entrepreneur and I am designing this project to help him find the most suitable location.

BUSINESS PROBLEM

The objective of this capstone project is to find the most suitable location for the entrepreneur to open a new Vietnamese restaurant in Toronto, Canada. By using data science methods and machine learning methods such as clustering, this project aims to provide solutions to answer the business question: In Toronto, if an entrepreneur wants to open a Vietnamese restaurant, where should they consider opening it?

Target Audience

The entrepreneur who wants to find the location to open authentic Vietnamese restaurant

DATA

To solve this problem, we will need below data:

1. List of neighborhoods in Toronto, Canada.
2. Latitude and Longitude of these neighborhoods.
3. Venue data related to Asian restaurants. This will help us find the neighborhoods that are most suitable to open a Vietnamese restaurant.

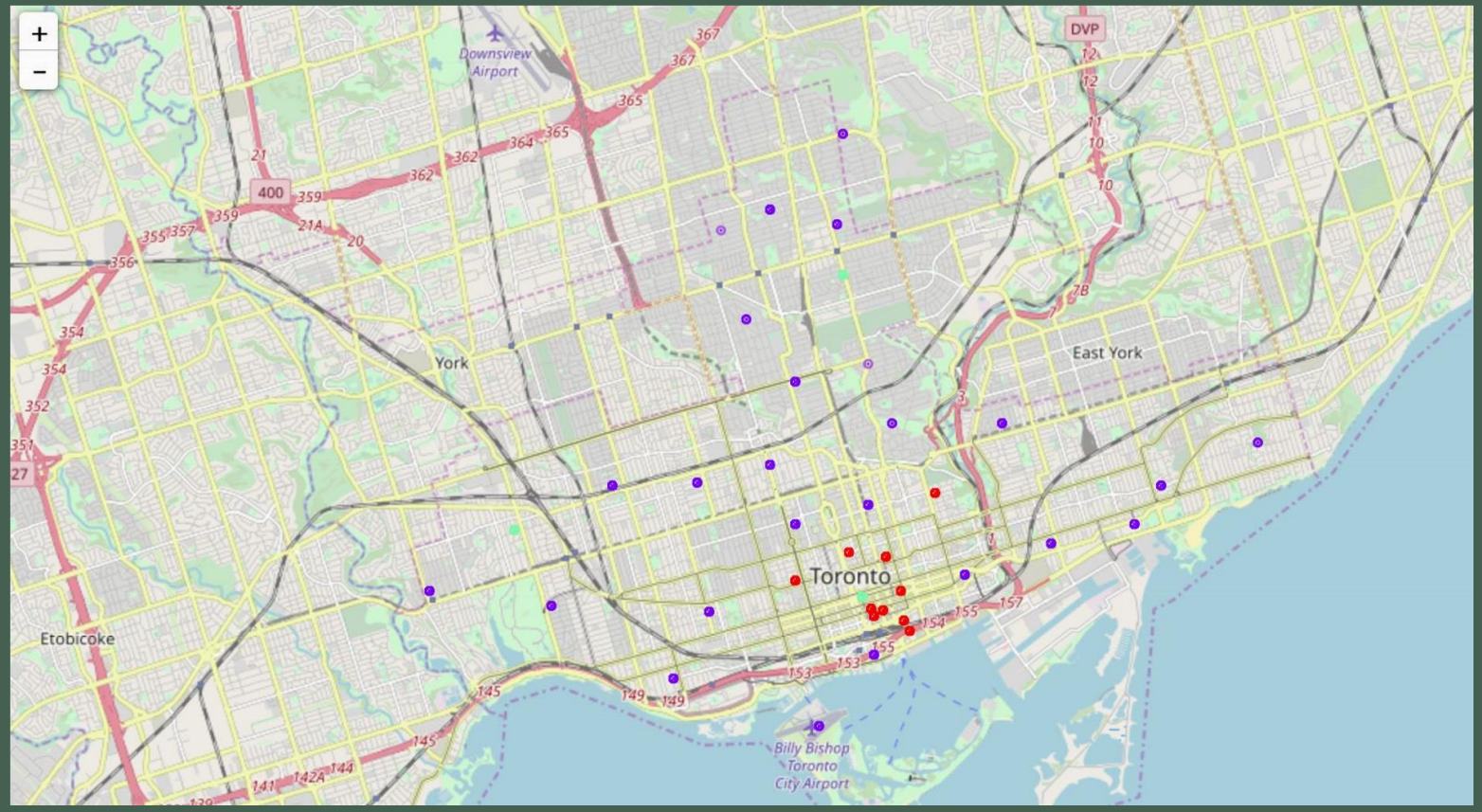
EXTRACTING THE DATA

1. Scrapping of Toronto neighborhoods via Wikipedia
2. Getting Latitude and Longitude data of these neighborhoods via Geocoder package
3. Using Foursquare API to get venue data related to these neighborhoods

METHODOLOGY

1. get the list of neighborhoods in Toronto, Canada. This is possible by extracting the list of neighborhoods from wikipedia page
2. use Foursquare API to pull the list of top 100 venues within 500 meters radius of “Thai restaurants” (Vietnamese food and Thai food are very similar in taste)
3. cluster the neighborhoods in Toronto into 3 clusters based on their frequency of occurrence for “Thai food”.
4. Based on the results (the concentration of clusters), I will be able to recommend the ideal location to open the restaurant

RESULTS
CLUSTERS



RESULTS

The results from k-means clustering show that we can categorize Toronto neighborhoods into 3 clusters based on how many Thai restaurants are in each neighborhood:

1. Cluster 0: Neighborhoods with little or no Thai restaurants
2. Cluster 1: Neighborhoods with no Thai restaurants
3. Cluster 2: Neighborhoods with high number of Thai restaurants

RECOMMENDATIONS

Most of Thai restaurants are in Cluster 2 which is around Adelaide, King, Richmond areas and lowest (close to zero) in Cluster 1 areas which are North Toronto West and Parkdale areas. Also, there are good opportunities to open near Chinatown, St James town as the competition seems to be low.

Looking at nearby venues, it seems Cluster 1 might be a good location as there are not a lot of Asian restaurants in these areas. Therefore, this project recommends the entrepreneur to open an authentic Vietnamese restaurant in these locations with little to no competition.

Nonetheless, if the food is authentic, affordable and good taste, I am confident that it will have great following everywhere =)

FUTURE RESEARCH

Consideration of one factor: the occurrence / existence of Thai restaurants in each neighborhood. There are many factors that can be taken into consideration such as population density, income of residents, rent that could influence the decision to open a new restaurant.

However, to put all these data into this project is not possible to do within a short time frame for this capstone project. Future research can take into consideration of these factors.

In addition, I am relying on the existence of Thai restaurants only for this project but future research can take into consideration of other variables such as existence of Asian restaurants, Asian population level in each neighborhood etc.

CONCLUSION

We have gone through the process of identifying the business problem, specifying the data required, extracting and preparing the data, performing the machine learning by utilizing k-means clustering and providing recommendation to the stakeholder.