Coursera

IBM Applied Data Science Capstone Project

Opening a New Milk Tea & Street Food Store in Toronto, Canada

Yawen Liu

July 2021

1

2

[1] https://d1ralsognjng37.cloudfront.net/88d3e1e7-8c07-4058-8a24-a052ec9d1c4b.webp
[2] https://www.torani.com/sites/default/files/recipes/illustration/GettyImages-1265715517-min.jpg

Introduction / Business Problem

Nowadays, milk tea (or bubble tea) has become an essential drink, especially in younger age people's daily life. Due to its versatility (i.e. flavors, toppings, kinds of milk), convenience, and successful social media marketing, milk tea becomes a very popular affordable drink around the world, specifically in a city with a large young population like Toronto, Canada. Many people like to grab a milk tea and enjoy it while shopping, work, and hangout.

However, with the development of the milk tea shop market, stores selling milk tea only are not competitive anymore. Milk tea stores selling street food (i.e. popcorn chicken, lunch bento box) are more favored by consumers as consumers can get both food and drinks conveniently without going to another food store. It is preferred by people who enjoy Asian flavors and want to have a faster order process. Opening a milk tea & street food store would be very profitable and promising from a business standpoint with the appropriate selection of store locations in megacities.

One of my friends just moved to Toronto and wants to start her own business which is opening a milk tea & street food store to serve people who enjoy milk tea and Asian street food. As Toronto is the most populous city in Canada with a highly diversified and strong economy, on one hand, there will be no problems in finding consumers. On the other hand, the competition in the milk tea industry is fierce as well. To succeed, the location of the store is of vital importance. **As a result, in this project, we will utilize the previously learned Data Science & Analysis tools to explore which area(s) is(are) a better location for my friend to open her milk tea & street food store**.

After a thorough discussion with my friend, we decided to **focus on two areas when selecting the location of the store to attract consumers as much as possible**:

1. The store is expected to be around/near shopping malls
2. The store is expected to be far from similar milk tea stores

Data

To solve this problem, we will need the following data to obtain insights:

- **List of Neighborhoods in Toronto, Canada**. For this project, the scope is confined to the city of Toronto, Canada which is the largest city in Canada and a world leader in business, finance, technology, entertainment, and culture with a large population.
- **Geospatial Data of Toronto's Neighborhoods**. The geospatial data include the Latitude and Longitude coordinates of each neighborhood in Toronto. This data is required to plot the map for visualization and to obtain the nearby venue data.
- **Nearby Venue data of Each Neighborhood**. This data will be utilized to perform clustering particularly related to shopping malls and similar milk tea (bubble tea) places. We can make decisions based on the result of clustering.

Methodology

We will go through the following steps to finish this project:

1. **The List of Neighborhoods in Toronto, Canada** data will be obtained from the following Wikipedia page: https://en.wikipedia.org/w/index.php?title=List_of_postal_codes_of_Canada:_M&oldid=945633050.
   We will use the web scraping technique to extract data from the Wikipedia page. The Beautifulsoup package will be used for web scraping with the help of Python requests.

2. After web scrapping, we will use Python libraries like Pandas and Numpy to go through the data cleaning process for further analysis.

3. When the data cleaning is finished, we will use the Python Geocoder package to obtain the geospatial information (i.e. Latitude, Longitude coordinate) of the neighborhoods on the list.

4. Once we have the geospatial information of neighborhoods, the Foursquare API will be used to obtain the venue data for neighborhoods on the list. This step is expected to give us detailed venue information (i.e. name, category) as we specified. For each neighborhood in this project, we will get the top 100 venues that are within a radius of 2000 meters. We are going to filter “Shopping Mall” and “Bubble tea” as venue categories for neighborhoods.

5. With the help of further data wrangling from obtained JSON data, we will go through the data clustering process and visualize the map data using the Folium package. We will use K-means clustering for data clustering. In this project, we are going to cluster the neighborhoods into 3 clusters based on

their occurrence frequency for “Shopping mall” and “Bubble tea”.

6. The data clustering results are expected to give us insights on which neighborhoods have more “Shopping malls” and which neighborhoods have fewer “Bubble tea” places as competitors. By analyzing the result, we can narrow down the location selection of the new milk tea and street food shop.