Capstone Project - Calgary Neighborhoods with Restaurants

Introduction: Business Problem

Calgary is a city in the western Canadian province of Alberta, with a population of 1,285,711 in 2019, making it Alberta's largest city. It is the third-largest municipality in Canada (after Toronto and Montreal), and the largest in western Canada. Calgary has also always been one of the favorite cities for immigration from all over the world. Especially in these years, there are more and more Chinese immigrants moving to Calgary as their home. In 2016, 8.3% of the population in Calgary are Chinese immigrants.

As you can see from the numbers, with that many Chinese immigrants in Calgary, opening a Chinese restaurant in the right neighborhood in Calgary could be a good business. There are some very good Chinese restaurants with great success in Calgary already. As an investor who is looking to open a new Chinese restaurant, choosing the right location or neighborhood is essential. When we think of it by the investor, we expect from them to prefer the neighborhoods where there is relatively lower real estate cost, higher population and lower density of restaurants. We are also particularly interested in areas with no Chinese restaurants in vicinity. We would also prefer locations as close to city center as possible.

By utilizing our Data Science power, we could potentially help the investors/stakeholders to find some promising neighborhoods for opening a Chinese restaurant, so that they are open too more customers and facing less competition.

Data

In order to solve this business problem, we can search for the valuable data from multiple sources, then use our technical data analysis skills to process and clean the data, make them ready for deeper analysis during the later stage:

1. I searched "List of neighborhoods in Calgary" on Wikipedia, 'https://en.wikipedia.org/wiki/List_of_neighbourhoods_in_Calgary'. In the Pycharm IDE, I was able to use "html" python library to read the table off the webpage into "pandas" data frame for processing and cleaning.
2. I collected some information about "median property price" in each neighborhood in Calgary from 'https://www.avenuecalgary.com/calgarys-best-neighbourhoods-2019/the-full-list/' webpage by using python library "BeautifulSoup" and "pandas".
3. By using "geopy" library in python, I was able to find all the latitudes and longitudes coordinates for each given neighborhood including the Calgary City Center.
4. After long period of data processing and cleaning, I was able to make a excel data file called "Final_data.xlsx", which will be used in my "Jupyter Notebook" for my Capstone project.
5. In my project, in order to accurately calculate distances on the map, I need to create grid of locations in Cartesian 2D coordinate system which allows me to calculate distances in meters

(not in latitude/longitude degrees). Then I used “pyproj” library to get the UTM Cartesian coordinates “X” and “Y” for each neighborhood.

6. I used my “Foursquare API” to get the restaurant venues and Chinese restaurant venues in each given neighborhood in Calgary.
7. For data visualization, I used python “folium’ library. Also, in order to create Calgary Choropleth Map, I downloaded the Calgary Neighborhoods Boundaries Geojson File from the Calgary City webpage ‘https://data.calgary.ca/Base-Maps/Community-Boundaries/ab7m-fwn6’.

Methodology