

Capstone Final Project – IBM Data Science



Background:

As a part of the final Capstone Project, we get a sense of what data scientists go through in real life. Objectives of the final assignments were to define a business problem, look for data in the web and, use Foursquare location data to compare different neighborhoods within wards (municipalities) of Calgary (choice of city depends on the students) to figure out which neighborhood is suitable for starting a restaurant business('idea' also depends on individual students). As prepared for the assignment, I go through the problem designing, data preparation and final analysis section step by step. Detailed codes and images are given in Github.

Introduction:

Calgary is the one of the top 10 cities in the list of most live able cities of the world; definitely one of the best places to start up a new business.

During the daytime, especially in the morning and lunch hours, office areas provide huge opportunities for restaurants. Reasonably priced (one lunch meal 8\$) shops are usually always full during the lunch hours (11 am — 2 pm) and, given this scenario, we will go through the benefits and pitfalls of opening a breakfast cum lunch restaurant in highly densed office places. Usually the profit margin for a decent restaurant lie within 15–20% range but, it can even go high enough to 35%, as discussed here.

We will go through each step of this project and address them separately. I first outline the initial data preparation and describe future steps to start the battle of neighborhoods in Calgary.

Target Audience:

What type of clients or a group of people would be interested in this project?

- Business personnel who wants to invest or open a restaurant. This analysis will be a comprehensive guide to start or expand restaurants targeting the large pool of office workers in Tokyo during lunch hours.
- Freelancer who loves to have their own restaurant as a side business. This analysis will give an idea, how beneficial it is to open a restaurant and what are the pros and cons of this business.
- New graduates, to find reasonable lunch/breakfast place close to office.
- Budding Data Scientists, who want to implement some of the most used Exploratory Data Analysis techniques to obtain necessary data, analyze it, and, finally be able to tell a story out of it.

Data:

1. We use the Calgary neighborhood wiki page to scrap the table. For scrapping and getting the table into data frame, Beautiful soap and pandas library are used. The table and the code for fetching the data is shown below

← → ↻ en.wikipedia.org/wiki/List_of_neighbourhoods_in_Calgary ☆ ⚙ 📄

Name ^[9]	Quadrant	Sector ^[10]	Ward ^[11]	Type ^[10]	2012 Population Rank	Population (2012) ^[9]	Population (2011) ^[9]	% change	Dwellings (2012) ^[9]	Area (km ²) ^[10]	Pop den
Abbeyle	NE/SE	Northeast	10	Residential	82	5,917	5,700	3.8	2,023	1.7	3.
Acadia	SE	South	9	Residential	27	10,705	10,615	0.8	5,053	3.9	2.
Albert Park/Radisson Heights	SE	East	10	Residential	75	6,234	6,217	0.3	2,709	2.5	2.
Altadore	SW	Centre	11	Residential	39	9,116	8,907	2.3	4,486	2.9	3.
Alyth/Bonnybrook	SE	Centre	9	Industrial	208	16	17	−5.9	14	3.8	
Applewood Park	SE/NE	East	10	Residential	69	6,498	6,404	1.5	2,215	1.6	4.
Arbour Lake	NW	Northwest	2	Residential	26	10,836	10,762	0.7	3,918	4.4	2.
Aspen Woods	SW	West	6	Residential	92	5,271	4,469	17.9	2,281	3.8	1.
Auburn Bay	SE	Southeast	12	Residential	60	7,193	5,769	24.7	2,808	4.5	1.
Aurora Business Park	NE	North	3	Industrial	237	0	0	—	0	2.4	
Banff Trail	NW	Centre	7	Residential	113	3,837	3,582	7.1	1,950	1.5	2.
Bankview	SW	Centre	8	Residential	94	5,221	4,754	9.8	3,528	0.7	7.
Bayview	SW	South	11	Residential	182	682	648	5.2	249	0.4	1.
Beddington Heights	NW/NE	North	4	Residential	20	11,585	11,457	1.1	4,348	3.2	3.
Bel-Aire	SW	Centre	11	Residential	188	424	449	−5.6	159	0.3	1.
Beltline	SW/SE	Centre	8	Residential	3	19,681	19,556	0.6	14,456	2.9	6.
Bonavista Downs	SE	South	14	Residential	177	925	947	−2.3	370	0.5	1.
Bowness	NW	Northwest	1	Residential	24	11,012	10,700	2.9	5,313	5.6	1.
Braeside	SW	South	11	Residential	81	5,940	5,978	−0.6	2,544	2	2.

```
import requests
from bs4 import BeautifulSoup
import pandas as pd

WIKI_URL = "https://en.wikipedia.org/wiki/List_of_neighbourhoods_in_Calgary"

req = requests.get(WIKI_URL)
soup = BeautifulSoup(req.content, 'lxml')
table_classes = {"class": ["sortable", "plainrowheaders"]}
wikitable = soup.findAll("table", table_classes)
right_table = soup.find('table', class_='wikitable sortable')
#print(right_table)
A=[]
B=[]

for row in right_table.findAll('tr'):
    cells=row.findAll('td')
    if len(cells)==12:
        A.append(cells[0].find(text=True))
        B.append(cells[4].find(text=True))

df=pd.DataFrame(A,columns=['Neighborhood'])
df['Type']=B
df.head()
```

- Using geopy library, get the latitude and longitude of the data. The first five rows of the Calgary neighborhood with their latitude and longitude are shown below:

	Neighbourhood	Latitude	Longitude
0	Abbeydale	51.058836	-113.929413
1	Acadia	51.102527	-110.213738
2	Albert Park/Radisson Heights	51.044845	-113.990195
3	Altadore	51.015104	-114.100756
4	Alyth/Bonnybrook	51.016669	-114.024294

Methodology:

Using Foursquare API and the data prepared from the above section, we explore the concentration of restaurants across different neighborhoods in Calgary. All the venue categories were obtained for each neighborhood.

	Neighbourhood	Neighbourhood Latitude	Neighbourhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Abbeydale	51.058836	-113.929413	Subway	51.059215	-113.934836	Sandwich Place
1	Abbeydale	51.058836	-113.929413	Mac's	51.059376	-113.934425	Convenience Store
2	Abbeydale	51.058836	-113.929413	roadside pub	51.059277	-113.934529	Wings Joint
3	Albert Park/Radisson Heights	51.044845	-113.990195	7-Eleven	51.047371	-113.989308	Convenience Store
4	Albert Park/Radisson Heights	51.044845	-113.990195	New Dynasty	51.047540	-113.990004	Asian Restaurant

Once all the Venue Categories were obtained, only 'Restaurant' categories were extracted from the dataset to be aligned with the business problem that we are trying to solve.

```
calgary_venues_restaurant = calgary_venues.loc[calgary_venues['Venue Category'].str.contains("Restaurant")]
```

	Neighbourhood	Neighbourhood Latitude	Neighbourhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
4	Albert Park/Radisson Heights	51.044845	-113.990195	New Dynasty	51.047540	-113.990004	Asian Restaurant
8	Altadore	51.015104	-114.100756	Pegasus Restaurant	51.010767	-114.100119	Greek Restaurant
12	Alyth/Bonnybrook	51.016669	-114.024294	Dutch Cash & Carry	51.020801	-114.022336	Restaurant
26	Aspen Woods	51.043119	-114.210185	Edo Japan	51.041333	-114.210438	Japanese Restaurant
29	Aspen Woods	51.043119	-114.210185	Diner Deluxe Aspen	51.039636	-114.209193	Restaurant

