Capstone Project

The Battle of the Neighborhoods (Week 2)

Applied Data Science Capstone by IBM/Coursera

By.. Rungrat Jaroenratanakorn

Date: 14 Jul 2021

Introduction: Business Problem

In this project we will try to find an optimal location for a restaurant. Specifically, this report will be targeted to stakeholders interested in opening restaurant in Alberta, Canada.

Since there are lots of restaurants in Alberta, we will try to detect locations that are not already crowded with restaurants. We would also prefer locations as close to city center as possible. Additional, maybe specific "Thai restaurant", if we have data for considering.

Data

Based on definition of our problem, factors that will influence our decision are:

- number of existing restaurants in the neighborhood (any type of restaurant)
- distance of neighborhood from city center

Following data sources will be needed to extract/generate the required information:

- The neighborhood dataset is available on this Wikipedia page.
- Get the information about Borough, Neighborhood, latitude, Longitude from latitude from https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_T

Alberta - 157 FSAs [edit]

Postal Code	Borough	Neighborhood	Latitude	Longitude
T1A	Medicine Hat	Central Medicine Hat	50.036460	-110.679250
T2A	Calgary	Penbrooke Meadows, Marlborough	51.049680	-113.964320
T3A	Calgary	Dalhousie, Edgemont, Hamptons, Hidden Valley	51.126060	-114.143158
T4A	Airdrie	East Airdrie	51.272450	-113.986980
T5A	Edmonton	West Clareview, East Londonderry	53.5899	-113.4413
T6A	Edmonton	North Capilano	53.5483	-113.408
T7A	Drayton Valley	Not assigned	53.2165	-114.9893
T8A	Sherwood Park	West Sherwood Park	53.519	-113.3216
T9A	Wetaskiwin	Not assigned	52.9741	-113.3646
T1B	Medicine Hat	South Medicine Hat	50.0172	-110.651
T2B	Calgary	Forest Lawn, Dover, Erin Woods	51.0318	-113.9786
T3B	Calgary	Montgomery, Bowness, Silver Springs, Greenwood	51.0809	-114.1616
T4B	Airdrie	West Airdrie	51.2816	-114.0153
T5B	Edmonton	East North Central, West Beverly	53.5766	-113.4608
T6B	Edmonton	SE Capilano, West Southeast Industrial, East Bonnie Doon	53.5322	-113.4404
T7B	Not assigned	Not assigned	Not assigned	Not assigned
T8B	Sherwood Park	Outer Southwest	53.4482	-113.2706
T9B	Not assigned	Not assigned	Not assigned	Not assigned
T1C	Medicine Hat	North Medicine Hat	50.0556	-110.6822

- number of restaurants and their type and location in every neighborhood will be obtained using Foursquare API

Methodology

1. Get the data using BeautifulSoup

```
In [2]: res = requests.get("https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_T")
soup = BeautifulSoup(res.text, 'lxml') #if you find any problem with "lxml" then try using "html.parser" instead
table = soup.find("table",class_="wikitable")
len(soup.find_all('table'))
```

2. Cleansing and structuring data to below format

Out[4]:

	Postal Code	Borough	Neighborhood	Latitude	Longitude
1	T1A	Medicine Hat	Central Medicine Hat	50.036460	-110.679250
2	T2A	Calgary	Penbrooke Meadows, Marlborough	51.049680	-113.964320
3	T3A	Calgary	Dalhousie, Edgemont, Hamptons, Hidden Valley	51.126060	-114.143158
4	T4A	Airdrie	East Airdrie	51.272450	-113.986980
5	T5A	Edmonton	West Clareview, East Londonderry	53.5899	-113.4413
6	T6A	Edmonton	North Capilano	53.5483	-113.408
8	T8A	Sherwood Park	West Sherwood Park	53.519	-113.3216
10	T1B	Medicine Hat	South Medicine Hat	50.0172	-110.651
11	T2B	Calgary	Forest Lawn, Dover, Erin Woods	51.0318	-113.9786
12	T3B	Calgary	Montgomery, Bowness, Silver Springs, Greenwood	51.0809	-114.1616
13	T4B	Airdrie	West Airdrie	51.2816	-114.0153
14	T5B	Edmonton	East North Central, West Beverly	53.5766	-113.4608

From the dataframe, we found 11 boroughs and 97 neighborhoods.

```
In [6]: print('The dataframe has {} boroughs and {} neighborhoods.'.format(
            len(df['Borough'].unique()),
            df.shape[0]
        )
)
```

```
The dataframe has 11 boroughs and 97 neighborhoods.
```

We focus Edmonton borough in Alberta City. Edmonton is the capital city of the Canadian province of Alberta.

```
In [7]: df.groupby(['Borough']).count()
```

Out[7]:

Borough	Postal Code	Neighborhood	Latitude	Longitude
Airdrie	2	2	2	2
Calgary	34	34	34	34
Edmonton	38	38	38	38
Fort McMurray	3	3	3	3
Grande Prairie	3	3	3	3
Leduc	1	1	1	1
Lethbridge	3	3	3	3
Medicine Hat	3	3	3	3
Red Deer	3	3	3	3
Sherwood Park	6	6	6	6
Spruce Grove	1	1	1	1

3. Use geopy library to get the latitude and longitude values of Edmonton, Alberta.

```
: #!conda install -c conda-forge geopy --yes # uncomment this line if you haven't completed the Foursquare API lab
  from geopy.geocoders import Nominatim # convert an address into latitude and longitude values
```

```
: address = 'Edmonton, Alberta'

  geolocator = Nominatim(user_agent="ny_explorer")
  location = geolocator.geocode(address)
  latitude = location.latitude
  longitude = location.longitude
  print('The geograpical coordinate of Alberta are {}, {}.'.format(latitude, longitude))
```

```
The geograpical coordinate of Alberta are 53.535411, -113.507996.
```

4. Use Folium. Folium is a great visualization library. Feel free to zoom into the above map and click on each circle mark to reveal the name of the neighborhood and its respective borough.
   However, for illustration purposes, let's simplify the above map and segment and cluster only the neighborhoods in Alberta.
5. Segment the neighborhoods in Alberta using Foursquare API
   - Define Foursquare Credentials and Version

```
CLIENT_ID = '5EVFNTDVNLBCQMWSMQJI0SBKP3XW4FX2BWSZM3CJOE5RRZ3A' # your Foursquare ID
CLIENT_SECRET = 'EO1QHGNMT4LRH52ERD2HRGHYMTKMHRJ3QR3GHO424WV22LIP' # your Foursquare Secret
ACCESS_TOKEN = 'N25HUQKMW1Q3G5J5XWKKOTRFTTQ42OQ55P1OSQCNZC3GCAR5' # your FourSquare Access Token
VERSION = '20180604'
LIMIT = 30
print('Your credentails:')
print('CLIENT_ID: ' + CLIENT_ID)
print('CLIENT_SECRET:' + CLIENT_SECRET)
```

```
Your credentails:
CLIENT_ID: 5EVFNTDVNLBCQMWSMQJI0SBKP3XW4FX2BWSZM3CJOE5RRZ3A
CLIENT_SECRET:EO1QHGNMT4LRH52ERD2HRGHYMTKMHRJ3QR3GHO424WV22LIP
```

   - Get the top venues that are in Edmonton within a radius of 500 meters.
     We found 7 venues were returned by Foursquare.

	name	categories	lat	lng
0	Madhatter Coffee Roastery	Coffee Shop	50.039163	-110.677189
1	Local Public Eatery Medicine Hat	Pub	50.039218	-110.676133
2	Dairy Queen	Ice Cream Shop	50.037920	-110.680125
3	Esplanade Arts And Heritage Centre	Theater	50.039907	-110.680308
4	Subway	Sandwich Place	50.040062	-110.676073
5	TD Canada Trust	Bank	50.039671	-110.675277
6	Riverside Veterans' Memorial Park	Park	50.039924	-110.675801

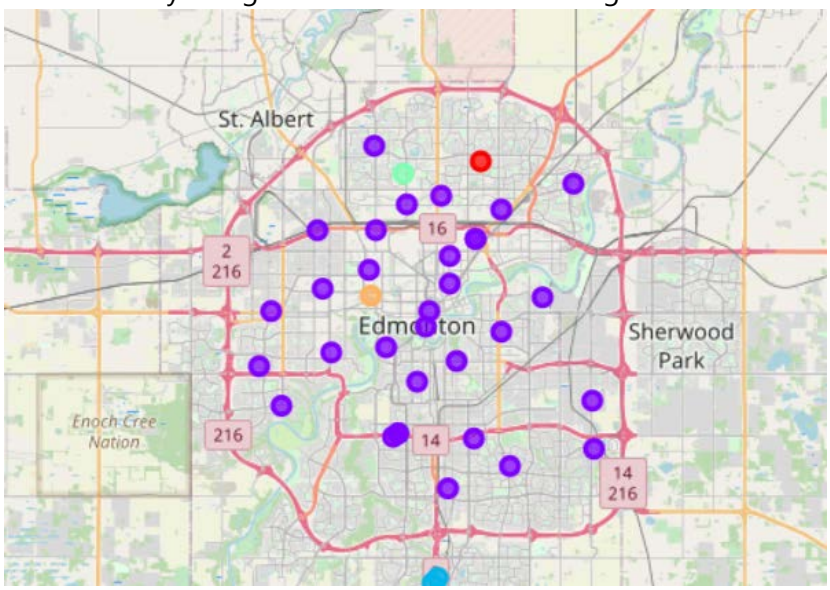
   - Next, group rows by neighborhood and by taking the mean of the frequency of occurrence of each category

	Neighborhood	American Restaurant	Arts & Crafts Store	Asian Restaurant	Bakery	Bank	Bar	Baseball Field	Baseball Stadium	Big Box Store	Bookstore	Breakfast Spot	Brewery	Buffet	Burger Joint	Bus Station	Business Service	Butcher	Café	Casino	Cheese Shop	Chinese Restaurant	Clo
0	West Clareview, East Londonderry	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	
1	West Clareview, East Londonderry	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	

- Create the new dataframe and display the top 10 venues for each neighborhood.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Central Beverly	Park	Grocery Store	Department Store	Furniture / Home Store	Smoke Shop	Construction & Landscaping	Dog Run	Eastern European Restaurant	Electronics Store	French Restaurant
1	Central Bonnie Doon	American Restaurant	Water Park	Trail	Fast Food Restaurant	French Restaurant	Food Truck	Food & Drink Shop	Flower Shop	Filipino Restaurant	Eastern European Restaurant
2	Central Jasper Place, Buena Vista	Café	Sushi Restaurant	Sandwich Place	Health & Beauty Service	Bakery	Pizza Place	Convenience Store	Salad Place	Liquor Store	Fast Food Restaurant
3	Central Londonderry	Health & Beauty Service	Wine Shop	Furniture / Home Store	Diner	Discount Store	Distribution Center	Dog Run	Eastern European Restaurant	Electronics Store	Fast Food Restaurant
4	Central Mistatim	Liquor Store	Warehouse Store	Casino	Electronics Store	Filipino Restaurant	French Restaurant	Food Truck	Food & Drink Shop	Flower Shop	Wine Shop

6. Cluster Neighborhoods by using k-means to cluster the neighborhood into 5 clusters.



7. Examine Clusters, examine each cluster and determine the discriminating venue categories that distinguish each cluster.

Examine Clusters

Now, you can examine each cluster and determine the discriminating venue categories that distinguish each cluster. Based on the defining categories, you can then assign a name to each cluster.

Cluster 0 ¶

```
Alberta_merged.loc[Alberta_merged['Cluster Labels'] == 0, Alberta_merged.columns[[1] + list(range(5, Alberta_merged.shape[1]))]]
```

]:

	Borough	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th M Common Ve
5	Edmonton	0	Bus Station	Record Shop	Buffet	Toy / Game Store	Mexican Restaurant	Filipino Restaurant	French Restaurant	Food Truck	Food & Drink Shop	Flower S
6	Edmonton	0	Playground	Construction & Landscaping	Bus Station	Ski Trail	Distribution Center	Discount Store	Dog Run	Eastern European Restaurant	Fried Chicken Joint	Fast F Restau
14	Edmonton	0	Park	Grocery Store	Department Store	Furniture / Home Store	Smoke Shop	Construction & Landscaping	Dog Run	Eastern European Restaurant	Electronics Store	French Restau
15	Edmonton	0	Business Service	Playground	Bar	Furniture / Home Store	Discount Store	Distribution Center	Dog Run	Eastern European Restaurant	Electronics Store	Fast F Restau
24	Edmonton	0	American Restaurant	Water Park	Trail	Fast Food Restaurant	French Restaurant	Food Truck	Food & Drink Shop	Flower Shop	Filipino Restaurant	Eastern Europ Restau
32	Edmonton	0	Butcher	Comic Shop	Arts & Crafts Store	Bakery	Dog Run	Hockey Arena	Baseball Field	Recreation Center	Grocery Store	Food Tr

Results

Finally, we can find the good location for opening a restaurant in Edmonton City, Alberta, Canada. Edmonton City has 38 neighborhoods. We try to detect locations that are not already crowded with restaurants. We would also prefer locations as close to city center as possible. From Edmonton's neighborhood, we found 7 venues, 129 unique categories. were returned by Foursquare within a radius of 500 meters. And we cluster the Edmonton's neighborhood to 5 clusters by using k-means. The result from clustering, some clusters have many restaurants. Some clusters also have Thai restaurants in most common venue. Some clusters don't have Thai restaurants.

Discussion

We prefer the cluster that don't have Thai restaurants and be close to city center. And more notices, some cluster has many Asian restaurants, maybe we consider separating from the existing Asian restaurants.

Conclusion

Purpose of this project was to identify Edmonton areas close to Alberta center. Clustering of those locations was then performed in order to create zones of interest and addresses of those zone centers were created to be used as starting points for final exploration by stakeholders.

Final decision about optimal restaurant location will be made by stakeholders based on specific characteristics of neighborhoods and locations in every recommended zone, taking into consideration additional factors like attractiveness of each location (proximity to park or water), levels of noise / proximity to major roads, real estate availability, prices, social and economic dynamics of every neighborhood etc.