Capstone Project - Applied Data Science

Open a new Iranian/Persian restaurant in Toronto, Canada

1. Introduction

A client wants to open his first restaurant in the city of Toronto. He would like me to study where he has more chance to succeed in his new business.

A. Background

My client is originally from Iran and new in Canada. He decided to live in Toronto and try to find the best place to open a food business. He wants to offer iranian specialities to his customers. He is a chef so do not fear about the quality of his cuisine. As a new immigrant, he does not know the habits of people and does not want to be in a area where the competition is too high. He does not fear about how long is he going to commute to go to work . One of his concern is not be close to another iranian cuisine around the city. He does not take the rental price into consideration so the financial aspect will not be considered for the study. He does not want to do a bad investment that is a reason why his choice will be based on facts and study.

Toronto is Canada's largest city and a world leader in such areas as business, finance, technology, entertainment and culture. Its large population of immigrants from all over the globe has also made Toronto one of the most multicultural cities in the world.

B. Data

- StatCan or wiki to get the iranian population living in Toronto, if possible breaking down per Borough
- Use the client location and Foursquare to search the iranian/ persian restaurants
- use the demography of Toronto per Borough to analyze where it is more possible to establish a new business

C. Methodology

> 1. Build the dataframe of Toronto neighborhood, merge with coordinates
> 2. Use foursquare to find all Persian restaurants around the city
> 3. Get my client address and explore areas with less iranian restaurants
> 4. Propose to the client one of two areas where he can get more chance to succeed

2. Assumption

My client lives in a crowded area where he can see many middle east immigrants. As we have done working in the first week of the Capstone project with the Toronto neighborhood location, I will use the same data as data entry. Also, I could not found a recent demographic population in the all Toronto so, I decided to use the 2016 demographic data available at https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/hlt-fst/pd-pl/comprehensive.cfm The Toroto geojson file has been created using a shapefile. Thanks to YKCZoli who has an excellent tutorial on how to do this a this url: https://gist.github.com/YKCzoli/b7f5ff0e0f641faba0f47fa5d16c4d8d Those 2 files have been saved into my github and been used in the notebook.

So Let's get started by installing and importing libraries

3. Data preparation and demonstration

In [ ]:

```
!conda install -c conda-forge geopy --yes
!conda install -c conda-forge folium=0.5.0 --yes
!conda install -c conda-forge geocoder
```

In [79]:

```
import numpy as np # library to handle data in a vectorized manner

import pandas as pd # library for data analsysis
pd.set_option("display.max_columns", None)
pd.set_option("display.max_rows", None)

import json # library to handle JSON files

from geopy.geocoders import Nominatim # convert an address into latitude and longitude values
import geocoder # to get coordinates

import requests # library to handle requests
from bs4 import BeautifulSoup # library to parse HTML and XML documents

from pandas.io.json import json_normalize # tranform JSON file into a pandas dataframe

# Matplotlib and associated plotting modules
import matplotlib.cm as cm
import matplotlib.colors as colors

# import k-means from clustering stage
from sklearn.cluster import KMeans

import folium # map rendering library
```

In [80]:

```
path = 'https://raw.githubusercontent.com/Teninmarie/Coursera_Capstone/master/T1201EN.CSV'
df_pop = pd.read_csv(path)
df_pop=df_pop[['Geographic code','Province or territory', 'Population, 2016']]
df_pop=df_pop[(df_pop['Province or territory']== "Ontario")]
df_pop2 = df_pop.sort_values(by=['Population, 2016'], ascending=False)
df_pop2.head()
```

Out[80]:

	Geographic code	Province or territory	Population, 2016
657	K0K	Ontario	111372.0
822	L5M	Ontario	104868.0
890	L9T	Ontario	103839.0
650	K0A	Ontario	103474.0
803	L4N	Ontario	98663.0

In [81]:

```
import requests # library to handle requests
url ='https://raw.githubusercontent.com/Teninmarie/Coursera_Capstone/master/Toronto.geojson'
r = requests.get(url)
#print(r.json())
```

Scrapping Data from website, put it into a dataframe and merge with the Toronto geospatial coordinates

In [82]:

```
#Assign the url to a variable called url
url ='https://en.wikipedia.org/w/index.php?title=List_of_postal_codes_of_Canada:_M&oldid=906439794'
response = requests.get(url)
soup = BeautifulSoup(response.content, 'html.parser')
table =soup.find_all('table', class_ = 'wikitable sortable')
table = table[0]
col1 = []
col2 = []
col3 = []
for row in table.find_all('tr'):
    cell = row.find_all('td')
    if (len(cell))> 0:
        col1.append(cell[0].text)
        col2.append(cell[1].text)
        col3.append(cell[2].text.rstrip('\n'))
```

In [83]:

```
df = pd.DataFrame({'PostalCode': col1,'Borough': col2,'Neighborhood': col3})
df.drop(df[df.Borough == 'Not assigned'].index, inplace=True)
df.head()
```

Out[83]:

	PostalCode	Borough	Neighborhood
2	M3A	North York	Parkwoods
3	M4A	North York	Victoria Village
4	M5A	Downtown Toronto	Harbourfront
5	M5A	Downtown Toronto	Regent Park
6	M6A	North York	Lawrence Heights

In [84]:

```
df_group = df.groupby(["PostalCode", "Borough"], as_index=False).agg(lambda x: ',' .join(x))
df_group.head()
```

Out[84]:

	PostalCode	Borough	Neighborhood
0	M1B	Scarborough	Rouge,Malvern
1	M1C	Scarborough	Highland Creek,Rouge Hill,Port Union
2	M1E	Scarborough	Guildwood,Morningside,West Hill
3	M1G	Scarborough	Woburn
4	M1H	Scarborough	Cedarbrae

In [85]:

```
path='http://cocl.us/Geospatial_data/Geospatial_Coordinates.csv'
df_Geospatial = pd.read_csv(path)
df_Geospatial.rename(columns={"Postal Code": "PostalCode"}, inplace=True)
df_toronto = df_group.merge(df_Geospatial, on="PostalCode", how="left")
df_toronto.head()
```

Out[85]:

	PostalCode	Borough	Neighborhood	Latitude	Longitude
0	M1B	Scarborough	Rouge,Malvern	43.806686	-79.194353
1	M1C	Scarborough	Highland Creek,Rouge Hill,Port Union	43.784535	-79.160497
2	M1E	Scarborough	Guildwood,Morningside,West Hill	43.763573	-79.188711
3	M1G	Scarborough	Woburn	43.770992	-79.216917
4	M1H	Scarborough	Cedarbrae	43.773136	-79.239476

Define Foursquare credentials

In [86]:

```
#@ hidden_cell
CLIENT_ID = 'your Foursquare ID' # your Foursquare ID
CLIENT_SECRET = 'your Foursquare Secret' # your Foursquare Secret
VERSION = '20190604'
```

Convert my client adress to longitude, latitude and check how many restaurants he has within 10 kms

In [87]:

```
address = '26 Grover Hill Avenue, Richmond Hill, ON'

geolocator = Nominatim(user_agent="foursquare_agent")
location = geolocator.geocode(address)
latitude = location.latitude
longitude = location.longitude
print(latitude, longitude)
```

```
43.8852770434783 -79.4112332173913
```

Search all the iranian retaurants

In [88]:

```
search_query = 'persian'
radius = 100000
LIMIT = 100
url = 'https://api.foursquare.com/v2/venues/search?client_id={}&client_secret={}&ll={},{}&v={}&query={}&radius={}&limit={}'.format(CLIENT_ID, CLIENT_SECRET, latitude, longitude, VERSION, search_query, radius, LIMIT)
```

In [89]:

```
import requests
results = requests.get(url).json()
print ('There are {} Iranian/Persian around your address.'.format(len(results['response']['venues'])))
```

```
There are 30 Iranian/Persian around your address.
```

There are 30 Iranian restaurants based on the data retrieve from foursquare website. hen, let see the details of this data into a dataframe.

In [90]:

```
venues = results['response']['venues']
dataframe = json_normalize(venues)
dataframe.head()
```

Out[90]:

	categories	hasPerk	id	location.address	location.cc	location.city	location.cou
0	[{'id': '4bf58dd8d48988d115941735', 'name': 'M...	False	4bba721e1261d13a8da8ea98	10711 Yonge St	CA	Richmond Hill	Can
1	[{'id': '52f2ab2ebcbc57f1066b8b2a', 'name': 'C...	False	571ff3d8498ee3255e52bc5b	9251	CA	Richmond Hill	Can
2	[{'id': '4bf58dd8d48988d1c4941735', 'name': 'R...	False	588cf74e5289302f30e711e1	2015 Avenue Rd	CA	North York	Can
3	[{'id': '52f2ab2ebcbc57f1066b8b2a', 'name': 'C...	False	5787ab82498ec9461f0fea79	9-2100 Steeles Ave W	CA	Concord	Can
4	[{'id': '4bf58dd8d48988d18f941735', 'name': 'F...	False	51d875ac454a0fb48cbac2b4	10133 Yonge St.	CA	Richmond Hill	Can

In [91]:

```
# keep only columns that include venue name, and anything that is associated with location
filtered_columns = ['name', 'categories'] + [col for col in dataframe.columns if col.startswith('location.')] + ['id']
dataframe_filtered = dataframe.loc[:, filtered_columns]

# function that extracts the category of the venue
def get_category_type(row):
    try:
        categories_list = row['categories']
    except:
        categories_list = row['venue.categories']

    if len(categories_list) == 0:
        return None
    else:
        return categories_list[0]['name']

# filter the category for each row
dataframe_filtered['categories'] = dataframe_filtered.apply(get_category_type, axis=1)

# clean column names by keeping only last term
dataframe_filtered.columns = [column.split('.')[-1] for column in dataframe_filtered.columns]

dataframe_filtered
```

Out[91]:

	name	categories	address	cc	city	country	crossStreet	distance	formattedAddress	l
0	Persian Palace	Middle Eastern Restaurant	10711 Yonge St	CA	Richmond Hill	Canada	at Elgin Mills Rd	2399	[10711 Yonge St (at Elgin Mills Rd), Richmond ...	[{'labe 43.888
1	Persian Rugs please	Carpet Store	9251	CA	Richmond Hill	Canada	YONGE &16th	3921	[9251 (YONGE &16th), Richmond Hill ON L4C 9T3,...	[{'labe 43
2	Darbar Persian Grill	Restaurant	2015 Avenue Rd	CA	North York	Canada	NaN	16689	[2015 Avenue Rd, North York ON M5M 4A5, Canada]	[{'labe 43
3	Persian Rugs Inc	Carpet Store	9-2100 Steeles Ave W	CA	Concord	Canada	NaN	13025	[9-2100 Steeles Ave W, Concord ON L4K 2V1, Can...	[{'labe 43
4	Royalty Persian Rugs	Furniture / Home Store	10133 Yonge St.	CA	Richmond Hill	Canada	1st rug store north of Major Macenzie	2337	[10133 Yonge St. (1st rug store north of Major...	[{'labe 43.874
5	Pastry Persian Bakery	Bakery	NaN	CA	NaN	Canada	NaN	2723	[Canada]	[{'labe 43.862
6	Persian & Oriental Rug Centre	Furniture / Home Store	8240 Yonge Street	CA	Vaughan	Canada	South of Highway 407	6532	[8240 Yonge Street (South of Highway 407), Vau...	[{'labe 43.827
7	Persian Auto Performance	None	38 Guardsman Ave.	CA	Thornhill	Canada	NaN	7265	[38 Guardsman Ave., Thornhill ON, Canada]	[{'labe 43
8	Turco Persian Rug Company Inc	Laundry Service	452 Richmond St E	CA	Toronto	Canada	at Ontario St	25989	[452 Richmond St E (at Ontario St), Toronto ON...	[{'labe 43
9	persian gourmet food	Restaurant	NaN	CA	NaN	Canada	NaN	10092	[Canada]	[{'labe 43.794
10	Taj persian rugs inc.	Furniture / Home Store	8400 Woodbine Ave.	CA	Markham	Canada	NaN	5940	[8400 Woodbine Ave., Markham ON L3r 4n7, Canada]	[{'labe 43.848
11	Zaytoon Persian Kitchen & Dessert Bar	Middle Eastern Restaurant	8199 Yonge Street	CA	Thornhill	Canada	NaN	6371	[8199 Yonge Street, Thornhill ON L3T 2C6, Canada]	[{'labe 43.829
12	Persian Rug Specialist	Laundry Service	7-1110 Finch Ave W	CA	North York	Canada	NaN	13977	[7-1110 Finch Ave W, North York ON M3J 2T2, Ca...	[{'labe 43.
13	Mediterranean & Persian Kabob House	Mediterranean Restaurant	141 King Road, Unit 3	CA	Richmond Hill	Canada	NaN	7521	[141 King Road, Unit 3, Richmond Hill ON L4E 3...	[{'labe 43
14	Shatter Abbas Persian Cuisine شاطر عباس	Middle Eastern Restaurant	Centerpoint Mall	CA	Toronto	Canada	Yonge St	9994	[Centerpoint Mall (Yonge St), Toronto ON M2M 3...	[{'labe 43
15	Cyrus Persian Restaurant	Middle Eastern Restaurant	NaN	CA	NaN	Canada	NaN	8448	[Canada]	[{'labe 43
16	Treasure Gallery	Antique Shop	9875 Yonge St	CA	Richmond Hill	Canada	at Major Mackenzie Dr	2747	[9875 Yonge St (at Major Mackenzie Dr), Richmo...	[{'labe 43.868
17	Shah Persian Food	BBQ Joint	NaN	CA	NaN	Canada	NaN	10519	[Canada]	[{'labe 43
18	North Restaurant	Middle Eastern Restaurant	27-100 Steeles Ave W	CA	Vaughn	Canada	btwn Hilda & Yonge	9777	[27-100 Steeles Ave W (btwn Hilda & Yonge), Va...	[{'labe 43.798
19	Tabriz Persian Cookhouse	Persian Restaurant	995 bay street	CA	Toronto	Canada	NaN	24558	[995 bay street, Toronto ON M5S 3C4, Canada]	[{'labe 43
20	Sheharzad Persian Grill	Restaurant	NaN	CA	NaN	Canada	NaN	25450	[Canada]	[{'labe 43.656
21	Woven Treasures Persian Rug Gallery	Furniture / Home Store	169 Queen St E	CA	Toronto	Canada	NaN	25914	[169 Queen St E, Toronto ON, Canada]	[{'labe 43.654
22	Cyrus Persian Restaurant	Restaurant	Ritson	CA	Oshawa	Canada	NaN	45279	[Ritson, Oshawa ON, Canada]	[{'labe 43
23	Rayhoon Persian Eatery	Persian Restaurant	420 Pearl St.	CA	Burlington	Canada	NaN	69474	[420 Pearl St., Burlington ON L7R 2N1, Canada]	[{'labe 43.326
24	Nannaa Persian Eatery	Persian Restaurant	1010 King St W	CA	Hamilton	Canada	Marion Ave S	79981	[1010 King St W (Marion Ave S), Hamilton ON L8...	[{'labe 43
25	Paradise Persian Cuisine	Middle Eastern Restaurant	NaN	CA	Hamilton	Canada	NaN	78687	[Hamilton ON, Canada]	[{'labe 43
26	Silk Road Persian Rugs	Furniture / Home Store	428 Gage Ave.	CA	Kitchener	Canada	NaN	101996	[428 Gage Ave., Kitchener ON, Canada]	[{'labe 43.444
27	Canadian Distribution Channel Inc.	Distillery	636 Queensway Court	CA	Peterborough	Canada	at The Queensway	96267	[636 Queensway Court (at The Queensway), Peter...	[{'labe 44.275
28	Altona Kabob	Middle Eastern Restaurant	460 George St. N	CA	Peterborough	Canada	NaN	99091	[460 George St. N, Peterborough ON K9H 3R8, Ca...	[{'labe 44
29	Sorena Kebab House	Persian Restaurant	220 King St North, Waterloo, Unit B	CA	Waterloo	Canada	King&University	100551	[220 King St North, Waterloo, Unit B (King&Uni...	[{'labe 43.475

From this results, we can tell that the information concerning the neighborhood is not properly entered so It will be difficult to know where exaclty are located those restaurants if we do not plot them on the map. Let see how many Persian restaurants are located close (radius less than 5 kms) to the client's address

In [92]:

```
dataframe_filtered[(dataframe_filtered['distance']<5000)].count()
```

Out[92]:

```
name                5
categories          5
address             4
cc                  5
city                4
country             5
crossStreet         4
distance            5
formattedAddress    5
labeledLatLngs      5
lat                 5
lng                 5
neighborhood        1
postalCode          4
state               4
id                  5
dtype: int64
```

There are 5 restaurants in the Richmond area so it is not a good idea to open a new shop offering the same cuisine . Based on this information, he will not be suitable to have a new investment in this community.

We are going to map the population as well as the location of all those persian restaurants and see which is area is propably suitable for a new investment.

In [93]:

```
venues_map = folium.Map(location=[latitude, longitude], zoom_start=10) # generate map centred around my client's address

#
folium.Marker(
    [latitude, longitude],
   # color='red',
    popup='Clients address'
).add_to(venues_map)

# add a red circle marker to represent my client address with a radius equals to 5 kms
folium.features.CircleMarker(
    [latitude, longitude],
    radius= 5,
    color='red',
    popup='5 kms circle around the focus point',
    fill = True,
    fill_color = 'red',
    fill_opacity = 0.6
).add_to(venues_map)

# add the iranian restaurants as blue circle markers
for lat, lng, label in zip(dataframe_filtered.lat, dataframe_filtered.lng, dataframe_filtered.categories):
    folium.features.CircleMarker(
        [lat, lng],
        radius=5,
        color='blue',
        popup=label,
        fill = True,
        fill_color='blue',
        fill_opacity=0.6
    ).add_to(venues_map)


venues_map.choropleth(
   geo_data=r.json(),
    data = df_pop,
    columns=['Geographic code','Population, 2016'],
    key_on='feature.properties.CFSAUID',
    fill_color='YlOrRd',
    fill_opacity=0.4,
    line_opacity=0.4,
    legend_name='Population by FSA'

)

# display map
venues_map
```

Out[93]:

Discussion

Some area seem to not have any restaurant and seem to be crowded based on our dataset. it is obvious that the eastern and wetern part of Toronto are good candidates. However, those information can be considered. For example, how many venues are located in those area and what are the habits of people living in those area. The study can be go deeper and will probably help to reduce the neighborhoods which can be interesting to invest in.