Determine the new locations of coffee shops to open

Pham Thái Sơn 2020

Introduction

1. **Background:** as populations grow their demand for food and drinks will largely increase. That is why the global number of coffee shop openings has been growing by double digits every year. However, opening a new coffee shop requires a lot of money, effort, human power and time. But up until now, most businesses just find a random location, determined by luck and human observation which can sometimes be incorrect. This will predict and determine where the population is largest, and where there will be few competition (other existing coffee shops in those areas).
2. **Problem:** the data used will be a combination of population by neighborhood, the locations of existing coffee shops in Toronto from FourSquare API, then combine them together and draw a map to visualize where the potential neighborhoods are.
3. **Interest:** those who will find this report useful are: business owners who wish to expand their coffee shops in Toronto, new business owners who want to open coffee shops and face less competition and high customer traffic areas as possible.

Data

1. Data List:

To execute this report there are 2 important data that is need:
1/ Coordinates of each Toronto neighborhood
2/ The population data that match each neighborhood.
3/ The data and locations of existing coffee shops in Toronto, match to each neighborhood. This data will be taken from Foursquare API database
How we will use the data to execute:
We will use the neighborhood populations data first, then we will use it to draw the chologreph map of Toronto populations. Then we will get data of existing coffee shops in Toronto areas and create a cluster and display them on the chologreph map of Toronto populations. The right neighborhood to build the new coffee shop is the one with the highest populations and the fewest number of existing coffee shops.

2. Data Source & Cleaning:

For the 1st data data we scrap from the Wikipedia page: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M for this data the problem is that they existed as a table on Wikipedia. After scraping using

read_html(). We have to identify the locations of the table we want using [location]. Then we will obtain the data table containing the table.

Then we use read_csv() function to read the data downloaded from the public source of Toronto city. The data at 1st contains:

Geographic name	Population, 2016	Total private dwellings, 2016	Private dwellings occupied by usual residents, 2016
1	CanadaFootnote 1	35151728.0	15412443.0
2	A0A	46587.0	26155.0
3	A0B	19792.0	13658.0
4	A0C	12587.0	8010.0

We do not need the Private Dwelling and Occupied dwelling so we will drop it as we merge with our existing coordinate table form Wikipedia.

After that we have to merge the data tables together along the ‘Postal Code’ column so the final table will have both the coordinates, the name of the neighborhood and.the populations.

After that we will build the chologreph map based on the populations of each neighborhood and give them coloring. The deeper the color indicates high population density and potential to open a coffee shop.

After that we will use the Foursquare API to search for venues within the radius of the neighborhood. After that, we sort only for the coffee shop venues and then cluster them on the map using folium.Circlemarker.

From there we will identify the potential high population neighborhood but few coffee shops to open our new coffee shops.

Methodology Uses:

For the 1st data data we scrap from the Wikipedia page: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M for this data the problem is that they existed as a table on Wikipedia. After scraping using read_html(). We have to identify the locations of the table we want using [location]. Then we will obtain the data table containing the table.

	Postal Code	Borough	Neighborhood	Latitude	Longitude
0	M3A	North York	Parkwoods	43.753259	-79.329656
1	M4A	North York	Victoria Village	43.725882	-79.315572
2	M5A	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636
3	M6A	North York	Lawrence Manor, Lawrence Heights	43.718518	-79.464763
4	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government	43.662301	-79.389494
5	M9A	Etobicoke	Islington Avenue, Humber Valley Village	43.667856	-79.532242
6	M1B	Scarborough	Malvern, Rouge	43.806686	-79.194353
7	M3B	North York	Don Mills	43.745906	-79.352188
8	M4B	East York	Parkview Hill, Woodbine Gardens	43.706397	-79.309937
9	M5B	Downtown Toronto	Garden District, Ryerson	43.657162	-79.378937
10	M6B	North York	Glencairn	43.709577	-79.445073

*Fig1.1: We have the table with Neighborhood and Lad, Longitud*e

Then we use read_csv() function to read the data downloaded from the public source of Toronto city. The population data have Postal Code and Population as one table.

```
In [63]: url1 = 'https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/hlt-fst/pd-pl/Table.cfm?Lang=Eng&T=1201&SR=1&S=22&O=A&RPP=9999&
         df_census = pd.read_html(url1)[0].dropna()
         df_census
```

Out[63]:

	Geographic name	Population, 2016	Total private dwellings, 2016	Private dwellings occupied by usual residents, 2016
1	CanadaFootnote 1	35151728.0	15412443.0	14072079.0
2	A0A	46587.0	26155.0	19426.0
3	A0B	19792.0	13658.0	8792.0
4	A0C	12587.0	8010.0	5606.0
5	A0E	22294.0	12293.0	9603.0
6	A0G	35266.0	21750.0	15200.0
7	A0H	17804.0	9928.0	7651.0
8	A0J	7880.0	4813.0	3426.0
9	A0K	26058.0	15159.0	11090.0
10	A0L	7643.0	3769.0	3178.0
11	A0M	6200.0	3247.0	2769.0

Then we use merge to the population table with the neighborhood table along the postal code columns to ensure no duplicates. Then drop all na values to make sure.

```
In [65]: df_census1 = df3.merge(df_census, how = 'left', on = 'Postal Code')
         df_census1 = df_census1.dropna()
         df_census1.rename(columns = {'Population, 2016':'Population', 'Private dwellings occupied by us
```

```
In [66]: df_census1
```

Out[66]:

	Postal Code	Borough	Neighborhood	Latitude	Longitude	Population
0	M3A	North York	Parkwoods	43.753259	-79.329656	34615.0
1	M4A	North York	Victoria Village	43.725882	-79.315572	14443.0
2	M5A	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636	41078.0
3	M6A	North York	Lawrence Manor, Lawrence Heights	43.718518	-79.464763	21048.0
4	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government	43.662301	-79.389494	10.0
5	M9A	Etobicoke	Islington Avenue, Humber Valley Village	43.667856	-79.532242	35594.0
6	M1B	Scarborough	Malvern, Rouge	43.806686	-79.194353	66108.0
7	M3B	North York	Don Mills	43.745906	-79.352188	13324.0
8	M4B	East York	Parkview Hill, Woodbine Gardens	43.706397	-79.309937	18628.0
		Downtown				

Use define function to call for venues around 500 radius for each neighborhood. Then arrange them into a table.

```
print('CLIENT_SECRET:' + CLIENT_SECRET)
```

```
Your credentails:
CLIENT_ID: YXDRSTG3ZSYGFGKTXZCQ3BUARRDEUHYOFVA14VOMS0YDZBIV
CLIENT_SECRET:YK3FLCPG1FSSATR1Y5Q1KEYICHTMAFQIUN3BGXVCBUM3WBP5
```

```
def getNearbyVenues(names, latitudes, longitudes, radius=500):

    venues_list=[]
    for name, lat, lng in zip(names, latitudes, longitudes):
        print(name)

        # create the API request URL
        url = 'https://api.foursquare.com/v2/venues/explore?&client_id={}&client_secret={}&v={}&ll={},{}&radius={}&limit={}'.fo
            CLIENT_ID,
            CLIENT_SECRET,
            VERSION,
            lat,
            lng,
            radius,
            LIMIT)

        # make the GET request
        results = requests.get(url).json()["response"]['groups'][0]['items']
```

The table contains the name of the neighborhood and the type of venue within that radius. We also use sort to take only the venue that are “coffee shop”

```
(2117, 7)
```

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Parkwoods	43.753259	-79.329656	Brookbanks Park	43.751976	-79.332140	Park
1	Parkwoods	43.753259	-79.329656	Variety Store	43.751974	-79.333114	Food & Drink Shop
2	Parkwoods	43.753259	-79.329656	Corrosion Service Company Limited	43.752432	-79.334661	Construction & Landscaping
3	Victoria Village	43.725882	-79.315572	Victoria Village Arena	43.723481	-79.315635	Hockey Arena
4	Victoria Village	43.725882	-79.315572	Portugril	43.725819	-79.312785	Portuguese Restaurant

In [107]:
```
df_coffeeshop = toronto_venues[toronto_venues['Venue Category']=='Coffee Shop']
df_coffeeshop.head()
```

Out[107]:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
5	Victoria Village	43.725882	-79.315572	Tim Hortons	43.725517	-79.313103	Coffee Shop
9	Regent Park, Harbourfront	43.654260	-79.360636	Tandem Coffee	43.653559	-79.361809	Coffee Shop
20	Regent Park, Harbourfront	43.654260	-79.360636	Sumach Espresso	43.658135	-79.359515	Coffee Shop
21	Regent Park, Harbourfront	43.654260	-79.360636	Arvo	43.649963	-79.361442	Coffee Shop
23	Regent Park, Harbourfront	43.654260	-79.360636	Rooster Coffee	43.651900	-79.365609	Coffee Shop

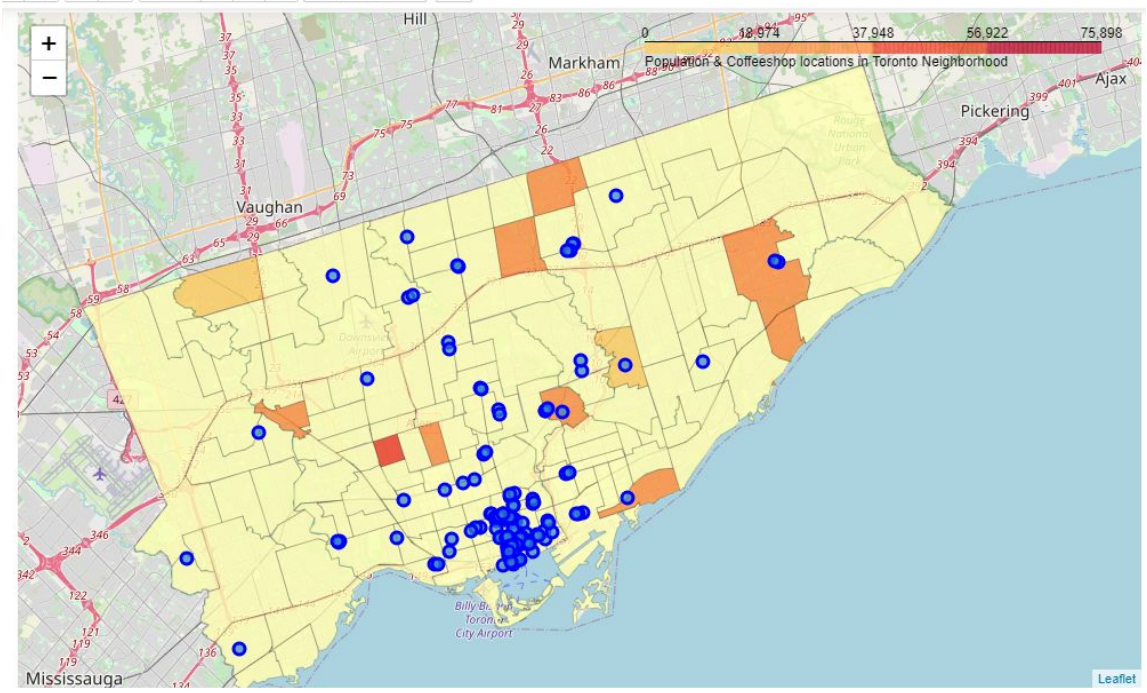
Then we use folium and Circlemaker to draw up the map that contains the information about the population of each neighborhood. As well as the location of each existing coffee shop in the neighborhood.

```
# draw choroleth map of population density in Toronto.
toronto_map.choropleth(
    geo_data=world_geo,
    data=df_census1,
    columns=['Neighborhood', 'Population'],
    key_on = 'feature.properties.HOOD',
    fill_color='YlOrRd',
    threshold_scale = threshold_scale,
    fill_opacity = 0.8,
    line_opacity=0.2,
    legend_name='Population & Coffeeshop locations in Toronto Neighborhood'
)
```

```
# add markers to map
for lat, lng, neighborhood in zip(df_coffeeshop['Venue Latitude'], df_coffeeshop['Venue Longitude'], df_coffeeshop['Neighborhood'
    label = '{}'.format(neighborhood)
    label = folium.Popup(label, parse_html=True)
    folium.CircleMarker(
        [lat, lng],
        radius=5,
        popup=label,
        color='blue',
        fill=True,
        fill_color='#3186cc',
        fill_opacity=0.7,
        parse_html=False).add_to(toronto_map)
```

The Result

After the results are done we have the map which indicates the population density and Toronto as well as the location of existing coffee shops. We want to open our new coffee shop in a location with high population density and few coffee shops.



Then we use groupby to merge the category of all neighborhoods together. While we also use the function count to count all the time coffee shops appear in that neighborhood. This will effectively show how many coffee shops each neighborhood has.

Then we use the function sort_values to sort for the high populations neighborhood, while also having the lowest number of coffee shops. In this table we see that the potential neighborhood are:

	Neighborhood	Number of Coffee Shop	Population
0	Willowdale, Willowdale East	2	75897.0
1	Fairview, Henry Farm, Oriole	5	58293.0
2	Steeles West, L'Amoreaux West	1	48471.0
3	Kennedy Park, Ionview, East Birchmount Park	1	48434.0
4	Dufferin, Dovercourt Village	1	44950.0
5	Regent Park, Harbourfront	7	41078.0
6	Brockton, Parkdale Village, Exhibition Place	2	40957.0
7	Willowdale, Willowdale West	1	40792.0
8	Don Mills	2	39153.0
9	Eringate, Bloordale Gardens, Old Burnhamthorpe...	1	38291.0

The Discussion

From the results above we see that we should prioritize the neighborhoods that have the highest populations and the fewest coffee shops.

It is worth to notice that this report only indicates the potential neighborhood to open a new coffee shop. The business owner needs to check for the law, regulations and available lands to open the coffee shops.

It is also worth noting that there are neighborhoods with quite high population density while there is no coffee shop in that neighborhood to serve them. This might come from the fact that home areas are separated with downtown, entertainment areas and so the population will potentially move to those downtown to enjoy their coffee. However, there is still room for convenience near your home coffee shops.

From the method above we can also use it to find out where to open: Pet stores, Indian restaurants, etc…

The Conclusion

Business owners and coffee shops owners should open more coffee shops in the following neighborhoods. Prioritize low coffee shop neighborhood such as:

Willowdale, Steeles West, Kennedy Park, Dufferin

	Neighborhood	Number of Coffee Shop	Population
0	Willowdale, Willowdale East	2	75897.0
1	Fairview, Henry Farm, Oriole	5	58293.0
2	Steeles West, L'Amoreaux West	1	48471.0
3	Kennedy Park, Ionview, East Birchmount Park	1	48434.0
4	Dufferin, Dovercourt Village	1	44950.0
5	Regent Park, Harbourfront	7	41078.0
6	Brockton, Parkdale Village, Exhibition Place	2	40957.0
7	Willowdale, Willowdale West	1	40792.0
8	Don Mills	2	39153.0
9	Eringate, Bloordale Gardens, Old Burnhamthorpe...	1	38291.0