

Week 5 Assignment

June 2, 2020

1 Capstone Project

1.1 Background & Problem Description:

A young couple recently moved from Paris to Toronto and they are looking for a right location to open their first juice bar in the city. As Toronto is one of the most diverse and populous cities in Canada, there are lots of business opportunities in various industries, be it banking & finance, tourism, arts & music etc. People from all over the world fly to Toronto on daily basis for conferences / business meetings / traveling, thus giving rise to lots of famous food & drink places. Though there are plenty of options in front of this couple, making decisions is not easy. The market in Toronto is highly developed and thus, comes with intense competition and high investment cost.

Being attracted by the potential return but scared of choosing the wrong neighborhoods, they decided to hire a team of data analysts to conduct detailed research on the Toronto market and suggest a location that provides space for business growth and has a nice environment that support business operations.

To solve the problem, we need to look at both demand side and supply side. Demand side refers to the group of potential customers, supply side refers to the resources needed to run the juice bar e.g. ingredients to make the drinks. The former is tied to the revenue of the juice bar, while the latter plays a key role in controlling the cost of operations.

Below are the factors we need to consider in our analysis:

- where's the supply comes from? Venues in each neighborhood e.g. grocery stores that will support the running of juice bar business.
- where's the demand comes from? What's the level of competition in each neighborhood e.g. looking at the number of juice bars in each neighborhood to determine the saturation of market.

At the end of project, we want to help this couple find the right location to open their juice bar based on market saturation and supply of resources to run the business.

1.2 Data Description & Approach

Below are the data sources we will collect data from to solve the problem.

1. Wikipedia:

We will extract a list of Canada postal codes from Wikipedia so we can build a dataframe of the postal code of each neighborhood along with the borough name and neighborhood name. In total, there are 5 boroughs and 306 neighborhoods in the dataset.

The data sample is provided below: Postalcode Borough Neighborhood 2 M3A North York Parkwoods 3 M4A North York Victoria Village 4 M5A Downtown Toronto Regent Park, Harbourfront 5 M6A North York Lawrence Manor, Lawrence Heights 6 M7A Downtown Toronto Queen's Park, Ontario Provincial Government 8 M9A Etobicoke Islington Avenue, Humber Valley Village 9 M1B Scarborough Malvern, Rouge 11 M3B North York Don Mills 12 M4B East York Parkview Hill, Woodbine Gardens 13 M5B Downtown Toronto Garden District, Ryerson

2. http://cocl.us/Geospatial_data

In order to utilize the Foursquare location data, we need to get the latitude and the longitude coordinates of each neighborhood. The website above will provide us the geospatial data needed to input into Foursquare.

The data sample is provided below: Postal Code Latitude Longitude 0 M1B 43.806686 -79.194353 1 M1C 43.784535 -79.160497 2 M1E 43.763573 -79.188711 3 M1G 43.770992 -79.216917 4 M1H 43.773136 -79.239476 5 M1J 43.744734 -79.239476 6 M1K 43.727929 -79.262029 7 M1L 43.711112 -79.284577 8 M1M 43.716316 -79.239476 9 M1N 43.692657 -79.264848

3. Foursquare

Once we have the latitude and longitude coordinates, we can make use of Foursquare API to find out more about the venues in each neighborhood. We will look at the top venues in each neighborhood and search for key words such as grocery stores, juice bars etc.

The data sample is provided below: Neighborhood Neighborhood Latitude Neighborhood Longitude Venue Venue Latitude Venue Longitude Venue Category 0 The Beaches 43.676357 -79.293031 Glen Manor Ravine 43.676821 -79.293942 Trail 1 The Beaches 43.676357 -79.293031 The Big Carrot Natural Food Market 43.678879 -79.297734 Health Food Store 2 The Beaches 43.676357 -79.293031 Grover Pub and Grub 43.679181 -79.297215 Pub 3 The Beaches 43.676357 -79.293031 Upper Beaches 43.680563 -79.292869 Neighborhood 4 The Danforth West, Riverdale 43.679557 -79.352188 MenEssentials 43.677820 -79.351265 Cosmetics Shop 5 The Danforth West, Riverdale 43.679557 -79.352188 Pantheon 43.677621 -79.351434 Greek Restaurant 6 The Danforth West, Riverdale 43.679557 -79.352188 Cafe Fiorentina 43.677743 -79.350115 Italian Restaurant 7 The Danforth West, Riverdale 43.679557 -79.352188 Dolce Gelato 43.677773 -79.351187 Ice Cream Shop 8 The Danforth West, Riverdale 43.679557 -79.352188 La Diperie 43.677530 -79.352295 Ice Cream Shop 9 The Danforth West, Riverdale 43.679557 -79.352188 Moksha Yoga Danforth 43.677622 -79.352116 Yoga Studio

1.3 Methodology

First, we extracted the raw data from Wikipedia website.

```
In [1]: from bs4 import BeautifulSoup
import requests
import pandas as pd
```

```
In [2]: List_url = "https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M"
source = requests.get(List_url).text
table=BeautifulSoup(source, 'xml').find('table')
```

Next, we created a dataframe from the raw data including the following columns - postal code, borough, and neighborhood information. Observations with unassigned borough are ignored in our analysis. We also performed some data cleaning such as index reset to make the data more analysis-friendly.

```
In [3]: column_names = ['Postalcode', 'Borough', 'Neighborhood']
df = pd.DataFrame(columns = column_names)
```

```
In [4]: for tr_cell in table.find_all('tr'):
        row_data=[]
        for td_cell in tr_cell.find_all('td'):
            row_data.append(td_cell.text.strip())
        if len(row_data)==3:
            df.loc[len(df)] = row_data
```

```
In [5]: df=df[df['Borough']!='Not assigned']
df.head(10)
```

```
Out[5]:
```

	Postalcode	Borough	Neighborhood
2	M3A	North York	Parkwoods
3	M4A	North York	Victoria Village
4	M5A	Downtown Toronto	Regent Park, Harbourfront
5	M6A	North York	Lawrence Manor, Lawrence Heights
6	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government
8	M9A	Etobicoke	Islington Avenue, Humber Valley Village
9	M1B	Scarborough	Malvern, Rouge
11	M3B	North York	Don Mills
12	M4B	East York	Parkview Hill, Woodbine Gardens
13	M5B	Downtown Toronto	Garden District, Ryerson

```
In [6]: df1=df.groupby('Postalcode')['Neighborhood'].apply(lambda x: "%s" % ', '.join(x))
df1=df1.reset_index(drop=False)
df1.rename(columns={'Neighborhood':'Neighborhood_joined'},inplace=True)
```

```
In [7]: df2 = pd.merge(df, df1, on='Postalcode')
df2.drop(['Neighborhood'],axis=1,inplace=True)
df2.drop_duplicates(inplace=True)
df2.rename(columns={'Neighborhood_joined':'Neighborhood'},inplace=True)
```

```
In [8]: df2.shape
```

```
Out[8]: (103, 3)
```

```
In [9]: df2.head(10)
```

```
Out[9]:
```

	Postalcode	Borough	Neighborhood
0	M3A	North York	Parkwoods
1	M4A	North York	Victoria Village
2	M5A	Downtown Toronto	Regent Park, Harbourfront

3	M6A	North York	Lawrence Manor, Lawrence Heights
4	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government
5	M9A	Etobicoke	Islington Avenue, Humber Valley Village
6	M1B	Scarborough	Malvern, Rouge
7	M3B	North York	Don Mills
8	M4B	East York	Parkview Hill, Woodbine Gardens
9	M5B	Downtown Toronto	Garden District, Ryerson

Building a dataframe of postal code of each neighborhood along with the borough name and neighborhood name is insufficient for our analysis. In order to utilize the Foursquare location data, we need to get the latitude and the longitude coordinates of each neighborhood as well.

```
In [10]: def get_geocode(postal_code):
# initialize your variable to None
lat_lng_coords = None
while(lat_lng_coords is None):
    g = geocoder.google('{}', Toronto, Ontario'.format(postal_code))
    lat_lng_coords = g.latlng
    latitude = lat_lng_coords[0]
    longitude = lat_lng_coords[1]
    return latitude,longitude
```

```
In [11]: df3=pd.read_csv('http://coc1.us/Geospatial_data')
df3.head(10)
```

```
Out[11]:   Postal Code  Latitude  Longitude
0         M1B  43.806686  -79.194353
1         M1C  43.784535  -79.160497
2         M1E  43.763573  -79.188711
3         M1G  43.770992  -79.216917
4         M1H  43.773136  -79.239476
5         M1J  43.744734  -79.239476
6         M1K  43.727929  -79.262029
7         M1L  43.711112  -79.284577
8         M1M  43.716316  -79.239476
9         M1N  43.692657  -79.264848
```

```
In [12]: df3.rename(columns={'Postal Code':'Postalcode'},inplace=True)
merged_data = pd.merge(df3, df2, on='Postalcode')
```

```
In [13]: merged_data.head(10)
```

```
Out[13]:   Postalcode  Latitude  Longitude  Borough \
0         M1B  43.806686  -79.194353  Scarborough
1         M1C  43.784535  -79.160497  Scarborough
2         M1E  43.763573  -79.188711  Scarborough
3         M1G  43.770992  -79.216917  Scarborough
4         M1H  43.773136  -79.239476  Scarborough
5         M1J  43.744734  -79.239476  Scarborough
```

6	M1K	43.727929	-79.262029	Scarborough
7	M1L	43.711112	-79.284577	Scarborough
8	M1M	43.716316	-79.239476	Scarborough
9	M1N	43.692657	-79.264848	Scarborough

	Neighborhood
0	Malvern, Rouge
1	Rouge Hill, Port Union, Highland Creek
2	Guildwood, Morningside, West Hill
3	Woburn
4	Cedarbrae
5	Scarborough Village
6	Kennedy Park, Ionview, East Birchmount Park
7	Golden Mile, Clairlea, Oakridge
8	Cliffside, Cliffcrest, Scarborough Village West
9	Birch Cliff, Cliffside West

```
In [14]: geo_data=merged_data[['Postalcode','Borough','Neighborhood','Latitude','Longitude']]
geo_data.head(10)
```

```
Out[14]:
```

	Postalcode	Borough	Neighborhood \
0	M1B	Scarborough	Malvern, Rouge
1	M1C	Scarborough	Rouge Hill, Port Union, Highland Creek
2	M1E	Scarborough	Guildwood, Morningside, West Hill
3	M1G	Scarborough	Woburn
4	M1H	Scarborough	Cedarbrae
5	M1J	Scarborough	Scarborough Village
6	M1K	Scarborough	Kennedy Park, Ionview, East Birchmount Park
7	M1L	Scarborough	Golden Mile, Clairlea, Oakridge
8	M1M	Scarborough	Cliffside, Cliffcrest, Scarborough Village West
9	M1N	Scarborough	Birch Cliff, Cliffside West

	Latitude	Longitude
0	43.806686	-79.194353
1	43.784535	-79.160497
2	43.763573	-79.188711
3	43.770992	-79.216917
4	43.773136	-79.239476
5	43.744734	-79.239476
6	43.727929	-79.262029
7	43.711112	-79.284577
8	43.716316	-79.239476
9	43.692657	-79.264848

Once the coordinates are incorporated into our dataset, our next step is to explore the neighborhoods in Toronto by using data from Foursquare. Foursquare allow us to find out more about the number of different venues in each neighborhood and thus, evaluating the level of competition and potential business opportunities in each neighborhood.

```
In [15]: Toronto=geo_data[geo_data['Borough'].str.contains("Toronto")]
Toronto.head()
```

```
Out[15]:
```

	Postalcode	Borough	Neighborhood	Latitude	Longitude
37	M4E	East Toronto	The Beaches	43.676357	-79.293031
41	M4K	East Toronto	The Danforth West, Riverdale	43.679557	-79.352188
42	M4L	East Toronto	India Bazaar, The Beaches West	43.668999	-79.315572
43	M4M	East Toronto	Studio District	43.659526	-79.340923
44	M4N	Central Toronto	Lawrence Park	43.728020	-79.388790

```
In [16]: CLIENT_ID = 'XUPAXIZHUH3FSCSDNOI3WBUQGCX5DWMWK3OISKETCQWAZOQX'
CLIENT_SECRET = 'OVZASVKH44Z1OVIQU5B4ZF12RV03YXLRARNBATROPPGJPHZ'
VERSION = '20180604'
```

```
In [17]: def getNearbyVenues(names, latitudes, longitudes):
    radius=500
    LIMIT=100
    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={}&client_version={}&lat={}&lng={}&radius={}&limit={}'
        CLIENT_ID,
        CLIENT_SECRET,
        VERSION,
        lat,
        lng,
        radius,
        LIMIT)

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

        # return only relevant information for each nearby venue
        venues_list.append([
            name,
            lat,
            lng,
            v['venue']['name'],
            v['venue']['location']['lat'],
```

```

        v['venue']['location']['lng'],
        v['venue']['categories'][0]['name']) for v in results])

nearby_venues = pd.DataFrame([item for venue_list in venues_list for item in venue_list])
nearby_venues.columns = ['Neighborhood',
                        'Neighborhood Latitude',
                        'Neighborhood Longitude',
                        'Venue',
                        'Venue Latitude',
                        'Venue Longitude',
                        'Venue Category']

return(nearby_venues)

In [18]: venues = getNearbyVenues(names=Toronto['Neighborhood'],
                                latitudes=Toronto['Latitude'],
                                longitudes=Toronto['Longitude']
                                )

The Beaches
The Danforth West, Riverdale
India Bazaar, The Beaches West
Studio District
Lawrence Park
Davisville North
North Toronto West, Lawrence Park
Davisville
Moore Park, Summerhill East
Summerhill West, Rathnelly, South Hill, Forest Hill SE, Deer Park
Rosedale
St. James Town, Cabbagetown
Church and Wellesley
Regent Park, Harbourfront
Garden District, Ryerson
St. James Town
Berczy Park
Central Bay Street
Richmond, Adelaide, King
Harbourfront East, Union Station, Toronto Islands
Toronto Dominion Centre, Design Exchange
Commerce Court, Victoria Hotel
Roselawn
Forest Hill North & West, Forest Hill Road Park
The Annex, North Midtown, Yorkville
University of Toronto, Harbord
Kensington Market, Chinatown, Grange Park
CN Tower, King and Spadina, Railway Lands, Harbourfront West, Bathurst Quay, South Niagara, Is.
Stn A PO Boxes

```

First Canadian Place, Underground city
 Christie
 Dufferin, Dovercourt Village
 Little Portugal, Trinity
 Brockton, Parkdale Village, Exhibition Place
 High Park, The Junction South
 Parkdale, Roncesvalles
 Runnymede, Swansea
 Queen's Park, Ontario Provincial Government
 Business reply mail Processing Centre, South Central Letter Processing Plant Toronto

In [19]: venues.head(10)

```

Out[19]:
      Neighborhood  Neighborhood Latitude \
0      The Beaches      43.676357
1      The Beaches      43.676357
2      The Beaches      43.676357
3      The Beaches      43.676357
4  The Danforth West, Riverdale      43.679557
5  The Danforth West, Riverdale      43.679557
6  The Danforth West, Riverdale      43.679557
7  The Danforth West, Riverdale      43.679557
8  The Danforth West, Riverdale      43.679557
9  The Danforth West, Riverdale      43.679557

      Neighborhood Longitude      Venue  Venue Latitude \
0      -79.293031      Glen Manor Ravine      43.676821
1      -79.293031  The Big Carrot Natural Food Market      43.678879
2      -79.293031      Grover Pub and Grub      43.679181
3      -79.293031      Upper Beaches      43.680563
4      -79.352188      MenEssentials      43.677820
5      -79.352188      Pantheon      43.677621
6      -79.352188      Cafe Fiorentina      43.677743
7      -79.352188      Dolce Gelato      43.677773
8      -79.352188      La Diperie      43.677530
9      -79.352188      Moksha Yoga Danforth      43.677622

      Venue Longitude      Venue Category
0      -79.293942      Trail
1      -79.297734      Health Food Store
2      -79.297215      Pub
3      -79.292869      Neighborhood
4      -79.351265      Cosmetics Shop
5      -79.351434      Greek Restaurant
6      -79.350115      Italian Restaurant
7      -79.351187      Ice Cream Shop
8      -79.352295      Ice Cream Shop
9      -79.352116      Yoga Studio
  
```



```
In [20]: venues.groupby('Neighborhood').count()
```

```
Out[20]:
```

Neighborhood	Latitude	Longitude
Berczy Park	56	
Brockton, Parkdale Village, Exhibition Place	22	
Business reply mail Processing Centre, South Ce...	17	
CN Tower, King and Spadina, Railway Lands, Harb...	15	
Central Bay Street	66	
Christie	17	
Church and Wellesley	77	
Commerce Court, Victoria Hotel	100	
Davisville	34	
Davisville North	7	
Dufferin, Dovercourt Village	16	
First Canadian Place, Underground city	100	
Forest Hill North & West, Forest Hill Road Park	4	
Garden District, Ryerson	100	
Harbourfront East, Union Station, Toronto Islands	100	
High Park, The Junction South	25	
India Bazaar, The Beaches West	22	
Kensington Market, Chinatown, Grange Park	58	
Lawrence Park	3	
Little Portugal, Trinity	44	
Moore Park, Summerhill East	3	
North Toronto West, Lawrence Park	20	
Parkdale, Roncesvalles	14	
Queen's Park, Ontario Provincial Government	33	
Regent Park, Harbourfront	45	
Richmond, Adelaide, King	94	
Rosedale	4	
Roselawn	2	
Runnymede, Swansea	37	
St. James Town	80	
St. James Town, Cabbagetown	46	
Stn A PO Boxes	94	
Studio District	41	
Summerhill West, Rathnelly, South Hill, Forest ...	16	
The Annex, North Midtown, Yorkville	22	
The Beaches	4	
The Danforth West, Riverdale	42	
Toronto Dominion Centre, Design Exchange	100	
University of Toronto, Harbord	36	

Business reply mail Processing Centre, South Ce...	17
CN Tower, King and Spadina, Railway Lands, Harb...	15
Central Bay Street	66
Christie	17
Church and Wellesley	77
Commerce Court, Victoria Hotel	100
Davisville	34
Davisville North	7
Dufferin, Dovercourt Village	16
First Canadian Place, Underground city	100
Forest Hill North & West, Forest Hill Road Park	4
Garden District, Ryerson	100
Harbourfront East, Union Station, Toronto Islands	100
High Park, The Junction South	25
India Bazaar, The Beaches West	22
Kensington Market, Chinatown, Grange Park	58
Lawrence Park	3
Little Portugal, Trinity	44
Moore Park, Summerhill East	3
North Toronto West, Lawrence Park	20
Parkdale, Roncesvalles	14
Queen's Park, Ontario Provincial Government	33
Regent Park, Harbourfront	45
Richmond, Adelaide, King	94
Rosedale	4
Roselawn	2
Runnymede, Swansea	37
St. James Town	80
St. James Town, Cabbagetown	46
Stn A PO Boxes	94
Studio District	41
Summerhill West, Rathnelly, South Hill, Forest ...	16
The Annex, North Midtown, Yorkville	22
The Beaches	4
The Danforth West, Riverdale	42
Toronto Dominion Centre, Design Exchange	100
University of Toronto, Harbord	36

	Venue	Venue Latitude \
Neighborhood		
Berczy Park	56	56
Brockton, Parkdale Village, Exhibition Place	22	22
Business reply mail Processing Centre, South Ce...	17	17
CN Tower, King and Spadina, Railway Lands, Harb...	15	15
Central Bay Street	66	66
Christie	17	17
Church and Wellesley	77	77
Commerce Court, Victoria Hotel	100	100

Davisville	34	34
Davisville North	7	7
Dufferin, Dovercourt Village	16	16
First Canadian Place, Underground city	100	100
Forest Hill North & West, Forest Hill Road Park	4	4
Garden District, Ryerson	100	100
Harbourfront East, Union Station, Toronto Islands	100	100
High Park, The Junction South	25	25
India Bazaar, The Beaches West	22	22
Kensington Market, Chinatown, Grange Park	58	58
Lawrence Park	3	3
Little Portugal, Trinity	44	44
Moore Park, Summerhill East	3	3
North Toronto West, Lawrence Park	20	20
Parkdale, Roncesvalles	14	14
Queen's Park, Ontario Provincial Government	33	33
Regent Park, Harbourfront	45	45
Richmond, Adelaide, King	94	94
Rosedale	4	4
Roselawn	2	2
Runnymede, Swansea	37	37
St. James Town	80	80
St. James Town, Cabbagetown	46	46
Stn A PO Boxes	94	94
Studio District	41	41
Summerhill West, Rathnelly, South Hill, Forest ...	16	16
The Annex, North Midtown, Yorkville	22	22
The Beaches	4	4
The Danforth West, Riverdale	42	42
Toronto Dominion Centre, Design Exchange	100	100
University of Toronto, Harbord	36	36

Venue Longitude \

Neighborhood	
Berczy Park	56
Brockton, Parkdale Village, Exhibition Place	22
Business reply mail Processing Centre, South Ce...	17
CN Tower, King and Spadina, Railway Lands, Harb...	15
Central Bay Street	66
Christie	17
Church and Wellesley	77
Commerce Court, Victoria Hotel	100
Davisville	34
Davisville North	7
Dufferin, Dovercourt Village	16
First Canadian Place, Underground city	100
Forest Hill North & West, Forest Hill Road Park	4
Garden District, Ryerson	100

Harbourfront East, Union Station, Toronto Islands	100
High Park, The Junction South	25
India Bazaar, The Beaches West	22
Kensington Market, Chinatown, Grange Park	58
Lawrence Park	3
Little Portugal, Trinity	44
Moore Park, Summerhill East	3
North Toronto West, Lawrence Park	20
Parkdale, Roncesvalles	14
Queen's Park, Ontario Provincial Government	33
Regent Park, Harbourfront	45
Richmond, Adelaide, King	94
Rosedale	4
Roselawn	2
Runnymede, Swansea	37
St. James Town	80
St. James Town, Cabbagetown	46
Stn A PO Boxes	94
Studio District	41
Summerhill West, Rathnelly, South Hill, Forest ...	16
The Annex, North Midtown, Yorkville	22
The Beaches	4
The Danforth West, Riverdale	42
Toronto Dominion Centre, Design Exchange	100
University of Toronto, Harbord	36

Venue Category

Neighborhood	
Berczy Park	56
Brockton, Parkdale Village, Exhibition Place	22
Business reply mail Processing Centre, South Ce...	17
CN Tower, King and Spadina, Railway Lands, Harb...	15
Central Bay Street	66
Christie	17
Church and Wellesley	77
Commerce Court, Victoria Hotel	100
Davisville	34
Davisville North	7
Dufferin, Dovercourt Village	16
First Canadian Place, Underground city	100
Forest Hill North & West, Forest Hill Road Park	4
Garden District, Ryerson	100
Harbourfront East, Union Station, Toronto Islands	100
High Park, The Junction South	25
India Bazaar, The Beaches West	22
Kensington Market, Chinatown, Grange Park	58
Lawrence Park	3
Little Portugal, Trinity	44

Moore Park, Summerhill East	3
North Toronto West, Lawrence Park	20
Parkdale, Roncesvalles	14
Queen's Park, Ontario Provincial Government	33
Regent Park, Harbourfront	45
Richmond, Adelaide, King	94
Rosedale	4
Roselawn	2
Runnymede, Swansea	37
St. James Town	80
St. James Town, Cabbagetown	46
Stn A PO Boxes	94
Studio District	41
Summerhill West, Rathnelly, South Hill, Forest ...	16
The Annex, North Midtown, Yorkville	22
The Beaches	4
The Danforth West, Riverdale	42
Toronto Dominion Centre, Design Exchange	100
University of Toronto, Harbord	36

In the following steps, we will convert categorical variables into a form that could be provided to ML algorithms. The categorical value represents the numerical value of the entry in the venues data.

```
In [21]: onehot = pd.get_dummies(venues[['Venue Category']], prefix="", prefix_sep="")
onehot.drop(['Neighborhood'],axis=1,inplace=True)
onehot.insert(loc=0, column='Neighborhood', value=venues['Neighborhood'] )
toronto_grouped = onehot.groupby('Neighborhood').mean().reset_index()
toronto_grouped.head()
```

```
Out [21]:
```

	Neighborhood	Afghan Restaurant	\
0	Berczy Park	0.0	
1	Brockton, Parkdale Village, Exhibition Place	0.0	
2	Business reply mail Processing Centre, South C...	0.0	
3	CN Tower, King and Spadina, Railway Lands, Har...	0.0	
4	Central Bay Street	0.0	

	Airport	Airport Food Court	Airport Lounge	Airport Service	\
0	0.000000	0.000000	0.000000	0.000000	
1	0.000000	0.000000	0.000000	0.000000	
2	0.000000	0.000000	0.000000	0.000000	
3	0.066667	0.066667	0.133333	0.133333	
4	0.000000	0.000000	0.000000	0.000000	

	Airport Terminal	American Restaurant	Antique Shop	Aquarium	...	\
0	0.000000	0.0	0.0	0.0	...	
1	0.000000	0.0	0.0	0.0	...	
2	0.000000	0.0	0.0	0.0	...	

3	0.133333	0.0	0.0	0.0	...
4	0.000000	0.0	0.0	0.0	...

	Trail	Train Station	Vegetarian / Vegan Restaurant	Video Game Store	\
0	0.0	0.0	0.017857	0.0	
1	0.0	0.0	0.000000	0.0	
2	0.0	0.0	0.000000	0.0	
3	0.0	0.0	0.000000	0.0	
4	0.0	0.0	0.015152	0.0	

	Video Store	Vietnamese Restaurant	Wine Bar	Wings Joint	Women's Store	\
0	0.000000	0.0	0.000000	0.0	0.0	
1	0.000000	0.0	0.000000	0.0	0.0	
2	0.000000	0.0	0.000000	0.0	0.0	
3	0.000000	0.0	0.000000	0.0	0.0	
4	0.015152	0.0	0.015152	0.0	0.0	

	Yoga Studio
0	0.000000
1	0.000000
2	0.058824
3	0.000000
4	0.015152

[5 rows x 235 columns]

The most important selection criterion in our analysis is the number of juice bars in each neighborhood. We want to find the neighborhoods with the least number of juice bars, which suggest those neighborhoods have more business opportunities compared to others.

```
In [22]: def return_least_common_venues(row, num_top_venues):
        row_categories = row.iloc[1:]
        row_categories_sorted = row_categories.sort_values(ascending=True)

        return row_categories_sorted.index.values[0:num_top_venues]
```



```
In [24]: import numpy as np

num_venues = 10

indicators = ['st', 'nd', 'rd']

# create columns according to number of top venues
columns = ['Neighborhood']
for ind in np.arange(num_venues):
    try:
        columns.append('{} Least Common Venue'.format(ind+1, indicators[ind]))
    except:
```

```

        columns.append('{}th Least Common Venue'.format(ind+1))

# create a new dataframe
venues_sorted = pd.DataFrame(columns=columns)
venues_sorted['Neighborhood'] = toronto_grouped['Neighborhood']

for ind in np.arange(toronto_grouped.shape[0]):
    venues_sorted.iloc[ind, 1:] = return_least_common_venues(toronto_grouped.iloc[ind

venues_sorted.head()

```

```

Out [24]:

```

	Neighborhood	1st Least Common Venue \
0	Berczy Park	Afghan Restaurant
1	Brockton, Parkdale Village, Exhibition Place	Afghan Restaurant
2	Business reply mail Processing Centre, South C...	Afghan Restaurant
3	CN Tower, King and Spadina, Railway Lands, Har...	Afghan Restaurant
4	Central Bay Street	Afghan Restaurant

	2nd Least Common Venue	3rd Least Common Venue \
0	Lake	Latin American Restaurant
1	Liquor Store	Lounge
2	Liquor Store	Lounge
3	Lounge	Malay Restaurant
4	Juice Bar	Lake

	4th Least Common Venue	5th Least Common Venue	6th Least Common Venue \
0	Light Rail Station	Lingerie Store	Malay Restaurant
1	Malay Restaurant	Market	Martial Arts Dojo
2	Malay Restaurant	Market	Martial Arts Dojo
3	Market	Martial Arts Dojo	Mediterranean Restaurant
4	Latin American Restaurant	Light Rail Station	Lingerie Store

	7th Least Common Venue	8th Least Common Venue	9th Least Common Venue \
0	Market	Martial Arts Dojo	Mediterranean Restaurant
1	Mediterranean Restaurant	Men's Store	Mexican Restaurant
2	Mediterranean Restaurant	Men's Store	Mexican Restaurant
3	Men's Store	Mexican Restaurant	Middle Eastern Restaurant
4	Liquor Store	Lounge	Malay Restaurant

	10th Least Common Venue
0	Men's Store
1	Middle Eastern Restaurant
2	Middle Eastern Restaurant
3	Miscellaneous Shop
4	Market

```

In [63]: venues_sorted_boolean = venues_sorted.apply(lambda row: row.astype(str).str.contains(

```

```

In [65]: venues_sorted_juice = venues_sorted[venues_sorted_boolean]

```

```
venues_sorted_juice.head()
```

Out [65]:

	Neighborhood	1st Least Common Venue	\
4	Central Bay Street	Afghan Restaurant	
7	Commerce Court, Victoria Hotel	Afghan Restaurant	
11	First Canadian Place, Underground city	Afghan Restaurant	
14	Harbourfront East, Union Station, Toronto Islands	Afghan Restaurant	
37	Toronto Dominion Centre, Design Exchange	Afghan Restaurant	

	2nd Least Common Venue	3rd Least Common Venue	4th Least Common Venue	\
4	Juice Bar	Lake	Latin American Restaurant	
7	Juice Bar	Korean Restaurant	Lake	
11	Jazz Club	Jewelry Store	Juice Bar	
14	Intersection	Irish Pub	Jazz Club	
37	Jazz Club	Jewelry Store	Juice Bar	

	5th Least Common Venue	6th Least Common Venue	7th Least Common Venue	\
4	Light Rail Station	Lingerie Store	Liquor Store	
7	Light Rail Station	Lingerie Store	Liquor Store	
11	Korean Restaurant	Lake	Latin American Restaurant	
14	Jewelry Store	Juice Bar	Korean Restaurant	
37	Korean Restaurant	Lake	Latin American Restaurant	

	8th Least Common Venue	9th Least Common Venue	10th Least Common Venue
4	Lounge	Malay Restaurant	Market
7	Lounge	Malay Restaurant	Market
11	Light Rail Station	Lingerie Store	Liquor Store
14	Latin American Restaurant	Light Rail Station	Lingerie Store
37	Light Rail Station	Lingerie Store	Liquor Store

```
In [66]: def return_most_common_venues(row, num_top_venues):
row_categories = row.iloc[1:]
row_categories_sorted = row_categories.sort_values(ascending=False)

return row_categories_sorted.index.values[0:num_top_venues]
```

```
In [70]: num_venues = 10
```

```
indicators = ['st', 'nd', 'rd']

# create columns according to number of top venues
columns = ['Neighborhood']
for ind in np.arange(num_venues):
    try:
        columns.append('{}-{} Most Common Venue'.format(ind+1, indicators[ind]))
    except:
        columns.append('{}th Most Common Venue'.format(ind+1))
```



```
# create a new dataframe
venues_sorted2 = pd.DataFrame(columns=columns)
venues_sorted2['Neighborhood'] = toronto_grouped['Neighborhood']

for ind in np.arange(toronto_grouped.shape[0]):
    venues_sorted2.iloc[ind, 1:] = return_most_common_venues(toronto_grouped.iloc[ind, 1:], venues_sorted2.columns[1:])

venues_sorted2.head()
```

```
Out[70]:
```

	Neighborhood	1st Most Common Venue	\
0	Berczy Park	Coffee Shop	
1	Brockton, Parkdale Village, Exhibition Place	Café	
2	Business reply mail Processing Centre, South C...	Light Rail Station	
3	CN Tower, King and Spadina, Railway Lands, Har...	Airport Lounge	
4	Central Bay Street	Coffee Shop	

	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	\
0	Cocktail Bar	Pub	Beer Bar	
1	Coffee Shop	Breakfast Spot	Grocery Store	
2	Auto Workshop	Park	Comic Shop	
3	Airport Service	Airport Terminal	Boat or Ferry	
4	Café	Italian Restaurant	Japanese Restaurant	

	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	\
0	Bakery	Seafood Restaurant	Cheese Shop	
1	Stadium	Burrito Place	Restaurant	
2	Pizza Place	Restaurant	Burrito Place	
3	Harbor / Marina	Coffee Shop	Sculpture Garden	
4	Sandwich Place	Burger Joint	Department Store	

	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Café	Restaurant	Japanese Restaurant
1	Climbing Gym	Performing Arts Venue	Bakery
2	Brewery	Skate Park	Smoke Shop
3	Rental Car Location	Bar	Boutique
4	Salad Place	Bubble Tea Shop	Restaurant

1.4 Results & Discussion

Based on the tables above, we are able to locate the top 5 neighborhoods with 'Juice Bar' in their top 10 least common venues, namely Central Bay Street, Commerce Court, First Canadian Place, Harbourfront East and Toronto Dominion Center.

The first two neighborhoods have 'Juice Bar' as their 2nd least common venue, while the rest have 'Juice Bar' as their 4th or 6th least common venue. However, we cannot make decision based on the ranking of least common venue alone - we also need to consider the presence of markets / grocery stores that will support the running of juice bar business.

For instance, Central Bay Street and Commerce Court have 'Market' in their list of least common venues, which make them less attractive locations to open a juice bar.

Now we are left with three options - First Canadian Place, Harbourfront East, and Toronto Dominion Center. Since these neighborhoods are pretty similar in terms of their least common venue list, we need to dive further into their venues i.e. we also need to look at their most common venue profiles.

```
In [71]: venues_sorted2[(venues_sorted2.Neighborhood=="First Canadian Place, Underground city")]
```

```
Out[71]:
```

	Neighborhood	1st Most Common Venue	\
11	First Canadian Place, Underground city	Coffee Shop	
14	Harbourfront East, Union Station, Toronto Islands	Coffee Shop	
37	Toronto Dominion Centre, Design Exchange	Coffee Shop	

	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	\
11	Café	Hotel	Gym	
14	Aquarium	Hotel	Café	
37	Café	Hotel	Restaurant	

	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	\
11	Japanese Restaurant	Restaurant	Salad Place	
14	Fried Chicken Joint	Scenic Lookout	Restaurant	
37	Seafood Restaurant	Japanese Restaurant	Italian Restaurant	

	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
11	Asian Restaurant	American Restaurant	Steakhouse
14	Italian Restaurant	Brewery	Sporting Goods Shop
37	American Restaurant	Salad Place	Beer Bar

As we can see from the results, all three neighborhoods have 'Coffee Shop' and 'Cafe' in their top 10 most common venue. However, while First Canadian Place and Toronto Dominion Center have 'Cafe' as their 2nd most common venues, Harbourfront East only has 'Cafe' as its 4th most common venue. On top of that, Harbourfront East also have more tourist attractions such as Aquarium and Scenic Lookout.

It also does not have 'Salad Place' as one of the common venues, which suggest less substitutes in the market (both salad places and juice bars target at people who want a healthy lifestyle, and salad places usually sell juices as well).

1.5 Conclusion

In conclusion, based on our analysis, we would suggest the young couple to choose Harbourfront East as their juice bar location. The neighborhood is full of tourist attractions such as aquarium and scenic lookout, and it's not occupied with too many restaurants. It also has lots of hotels and sporting goods shop, which will also support the growth of business.