

Coursera Capstone Report - Week 5

Table of Contents

1. Introduction	2
Problem description.....	2
Background discussion.....	2
2. Data.....	3
Pre-Processing	3
3. Methodology.....	4
Libraries and initial data import.....	4
Maps and locations.....	4
Top venues and distributions.....	6
4. Results.....	7
5. Discussion	7
6. Conclusion.....	7

1. Introduction

Problem description

Over the course of the last few weeks, we have seen how we can manipulate geographical data to visualize it and extrapolate information. Using the same data set, along with location queries fulfilled by foursquare, we can now enhance the dataset we obtained in Week 3. The problem which we have chosen to tackle is deciding on what type of business we would like to invest in, based on the most popular types of venues in East Toronto.

Background discussion

Toronto is Canada's most populous city and is located on Lake Ontario's northwestern shore. Within the city of Toronto is the borough of Toronto and one of its regions is called East Toronto, or more commonly referred to as the East End. It used to be a borough of its own before being annexed to Toronto more than 100 years ago. Today, it is a lively place full of many cultural sites and thriving atmosphere. It currently consists of 5 neighborhoods.

2. Data

For the project we are using the data obtained in Week-3 which has been cleaned and organized, as well as data regarding venues in East Toronto from Foursquare. The data was sourced from https://cocl.us/Geospatial_data.

Pre-Processing

The data was loaded into the notebook using the `pd.read_csv` command, which allowed us to tidy it by removing postal codes not associated with places of interest to us. This left us with categorized information about the boroughs and neighborhoods in Toronto. Similar to what was done in previous assignments we could easily select a specific region of the studied area. The decision was made for that region to be East Toronto. Through working with the data, we found that it has a very low number of neighborhoods - 5. This proved very useful as the *k* neighbors was very low, as the points were all very distinct, hence it could be taken as two ways: either have one large cluster, which would defeat the purpose of the study, or have 5 clusters, or in other words each neighborhood studied on its own.

	Postal Code	Borough	Neighborhood	Latitude	Longitude
0	M4E	East Toronto	The Beaches	43.676357	-79.293031
1	M4K	East Toronto	The Danforth West, Riverdale	43.679557	-79.352188
2	M4L	East Toronto	The Beaches West, India Bazaar	43.668999	-79.315572
3	M4M	East Toronto	Studio District	43.659526	-79.340923
4	M7Y	East Toronto	Business Reply Mail Processing Centre 969 Eastern	43.662744	-79.321558

Figure 1 - List of neighborhoods and respective coordinate

3. Methodology

Libraries and initial data import

As per usual we start the project by importing the necessary libraries and packages. One that stands out, as it did in Week 3 is BeautifulSoup4 which we use to pull the data from the xml address. Along with that we have geopy and geocoder which are later used for the location and mapping data. After which we need to link the app credentials, which was made on foursquare. This allows us to make calls to the foursquare API.

```
[6]: CLIENT_ID = '5FWPE5F4YXCJ152HQM23DLB0WGS4ACRLVLB1MRLDQDXS1'
      CLIENT_SECRET = 'AXKUYETLGBD31HFI0XVB1DQH2TH0340Q1TS13HWB2IXTWBAV'

      VERSION = '20180604'
      LIMIT = 30
```

Maps and locations

Similar to previous weeks we can use the Folium library to visualize the location data on a map. Here, we show all the postal code coordinates with their respective neighborhoods.

```
[11]: address = 'Toronto, Ontario Canada'
      geolocator = Nominatim()
      location = geolocator.geocode(address)
      latitude = location.latitude
      longitude = location.longitude

[12]: map_toronto = folium.Map(location=[latitude, longitude], zoom_start=11)

      for lat, lng, borough, neighborhood in zip(toronto_neighborhoods['Latitude'],
                                                  toronto_neighborhoods['Longitude'],
                                                  toronto_neighborhoods['Borough'],
                                                  toronto_neighborhoods['Neighborhood']):
          label = '{} , {}'.format(neighborhood, borough)
          label = folium.Popup(label, parse_html=True)
          folium.CircleMarker(
              [lat, lng],
              radius=4,
              popup=label,
              color='blue',
              fill=True,
              fill_color='#87cefa',
              fill_opacity=0.5,
              parse_html=False).add_to(map_toronto)
      map_toronto
```

Here, we can see the map of all the postal code coordinates we already had:

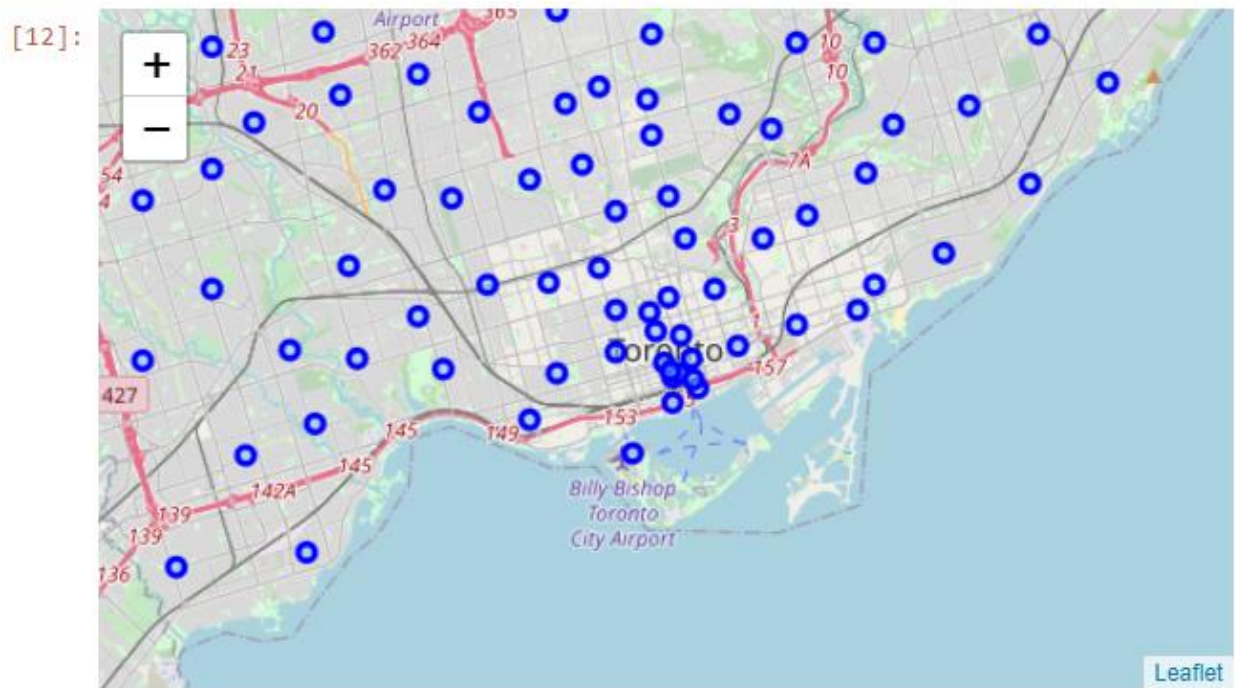


Figure 2 - Locations of all postal codes

Similarly, we can include the locations of the postal codes we decided to work with, i.e. East Toronto:



Figure 3 - Locations of East Toronto postal codes

Top venues and distributions

In this part we explore the data to find the top venues for each area as well as their frequency of occurrence.

Business Reply Mail Processing Centre			Studio District		
	venue	freq		venue	freq
0	Garden	0.14	0	Coffee Shop	0.12
1	Auto Workshop	0.14	1	Café	0.08
2	Fast Food Restaurant	0.14	2	American Restaurant	0.04
3	Brewery	0.14	3	Ice Cream Shop	0.04
4	Farmers Market	0.14	4	Gay Bar	0.04
5	Light Rail Station	0.14	5	Gastropub	0.04
6	Park	0.14	6	Fish Market	0.04
7	Movie Theater	0.00	7	Pet Store	0.04
8	Ice Cream Shop	0.00	8	Diner	0.04
9	Indian Restaurant	0.00	9	Latin American Restaurant	0.04

The Beaches			The Beaches West, India Bazaar		
	venue	freq		venue	freq
0	Other Great Outdoors	0.25	0	Fast Food Restaurant	0.13
1	Trail	0.25	1	Park	0.07
2	Playground	0.25	2	Pub	0.07
3	Park	0.25	3	Ice Cream Shop	0.07
4	Liquor Store	0.00	4	Italian Restaurant	0.07
5	Greek Restaurant	0.00	5	Fish & Chips Shop	0.07
6	Hotel	0.00	6	Liquor Store	0.07
7	Ice Cream Shop	0.00	7	Movie Theater	0.07
8	Indian Restaurant	0.00	8	Pet Store	0.07
9	Italian Restaurant	0.00	9	Hotel	0.07

[31]:

	Neighbourhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Business Reply Mail Processing Centre 969 Eastern	Garden	Light Rail Station	Auto Workshop	Fast Food Restaurant	Farmers Market
1	Studio District	Coffee Shop	Café	American Restaurant	Convenience Store	Clothing Store
2	The Beaches	Trail	Playground	Park	Other Great Outdoors	Yoga Studio
3	The Beaches West, India Bazaar	Fast Food Restaurant	Park	Pet Store	Ice Cream Shop	Burrito Place
4	The Danforth West, Riverdale	Greek Restaurant	Ice Cream Shop	Restaurant	Yoga Studio	Indian Restaurant

4. Results

The results of the data analysis are as follows:

- Two categories of venues stand out as most common: hospitality and parks
- The beaches neighborhood stands out with most outdoor venues
- The Danforth West neighborhood has the highest frequency of food hospitality venues
- The Studio district stands out for its abundance of cafes and bars
- The India Bazaar shows a great deal of diversity in its shopping venues

5. Discussion

The analysis shows the clear diversity of venues in some parts of the region such as the India Bazaar and the clear distinction between the outdoor and hospitality venues such the Beaches.

The benefits of the analysis is that it clearly lists the most popular types of venues in order to give a more clear picture to the investors on what the points of interest are. In reflection to that the drawbacks are that due to working with a small area, in comparison to all of Toronto, the picture that is built might not be entirely representative of the region. Additionally, this is only a preliminary analysis of the geographical locations of the venues and not a summative assessment of the business prospects of the region. At this stage of the analysis that might not be viable.

Furthermore, additional analysis of neighboring areas might be needed in order to clarify the prospectus data.

6. Conclusion

In conclusion, we get a small glimpse into what the area holds as prospects but is not indicative enough to signal any particular strength for the types of venues.

Based on this information the next steps for the investors will be to move past the preliminary analysis of the area and focus on the different types of venues throughout, as well as statistical data which could shine more light onto the profitability of the area by taking factors such as average rent price, population flow and sustainability.