

Capstone Project - The Battle of Neighborhoods

Speedy Coffee mobile locations



Introduction:

This report must “leverage the Foursquare location data to explore or compare neighborhoods or cities of your choice or to come up with a problem that you can use the Foursquare location data to solve” with this in mind, this report focuses on using the FourSquare API to retrieve location and venue data for a stated business type in a stated location.

Description:

Speedy Coffee is a (fictitious) mobile coffee start-up in Toronto looking to optimise the location of their mobile coffee carts.

Speedy Coffee have opted to select the financial district in Toronto to launch their service on the advice of a business consultant, however, they are unsure where to locate their carts within the financial district to maximise sales, and wish to know the location of any existing coffee shops.

Battle of the Neighborhoods has been commissioned to help Speedy Coffee determine the location of the existing coffee shops within the financial district.

Speedy Coffee will then use this data to determine different potential strategies for their coffee cart placement within Toronto's financial district.

Data:

As Speedy Coffee wishes to launch in Toronto, Battle of the Neighborhoods was able to leverage area data for Toronto acquired for a recent project, this data detailed the postal codes for all areas in Toronto with the postal prefix 'M', and was sourced from the following location:

[List of postal codes of Canada: M](#)

An example of the source data is shown below:

Postcode ↕	Borough ↕	Neighbourhood ↕
M1A	Not assigned	Not assigned
M2A	Not assigned	Not assigned
M3A	North York	Parkwoods
M4A	North York	Victoria Village
M5A	Downtown Toronto	Harbourfront
M6A	North York	Lawrence Heights
M6A	North York	Lawrence Manor
M7A	Downtown Toronto	Queen's Park
M8A	Not assigned	Not assigned
M9A	Queen's Park	Not assigned

Battle of the Neighborhoods will use the location data to query the Foursquare API and extract any venue information relevant to Speedy Coffee's requirement.

FourSquare is a location 'search and discovery' service used for locating businesses and venues for a given location, here is a link to FourSquare's website for more information:

[FourSquare](#)

Battle of the Neighborhoods will use the above two sets of data to, ultimately generate a map for Speedy Coffee showing the locations of any coffee shops within Toronto's financial district.

Methodology:

This section of the report demonstrates the methodology used to arrive at the final result.

Where appropriate you will be presented with an example of the actual data used, and informed of the tools that were used to perform the relevant processing of each step of the process.

The basic flow of the methodology was as follows:

1. Retrieve area, neighbourhood, and postcode data for the city of Toronto, Canada.
2. Clean the data in 1. to focus only on the area of interest (AOT) to the client.
3. Determine the geolocation latitude and longitude coordinates of the AOT from 2.
4. Retrieve the venue information from FourSquare for the AOT with the geo-data from 3.
5. Clean the data from FourSquare in 4. by selecting only the venues of interest (VOI).
6. Determine the coordinates of the VOI from 5.
7. Map the VOI by their geolocation coordinates from 6.

The remainder of this Methodology section will explain the process used, and demonstrate any relevant input or output data from each of the seven stages outlined above.

1: Neighbourhood, area and postcode data for Canada post codes prefix 'M' was obtained from Wikipedia using the following location:

[Canada Post Codes 'M'](#)

As Battle of the Neighborhoods had recently acquired this data for an earlier project, we used our saved data. Here is a screenshot of some of the original data sourced from Wikipedia:

Postcode ↕	Borough ↕	Neighbourhood ↕
M1A	Not assigned	Not assigned
M2A	Not assigned	Not assigned
M3A	North York	Parkwoods
M4A	North York	Victoria Village
M5A	Downtown Toronto	Harbourfront
M6A	North York	Lawrence Heights
M6A	North York	Lawrence Manor
M7A	Downtown Toronto	Queen's Park
M8A	Not assigned	Not assigned
M9A	Queen's Park	Not assigned

2: As we can see the data in the preceding screenshot is a little unmanageable and doesn't easily show us the information we require.

The data obtained from Wikipedia was cleaned and presented using the Pandas library in the following ways:

1. The Wikipedia data was imported into a Pandas dataframe.
2. Any postcodes that were not assigned to a borough or neighbourhood were removed.

These processes resulted in the following dataframe (only the first few rows are shown), as we can see, the 'Not Assigned' entries are no longer present:

	Postcode	Borough	Neighbourhood	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	Harbourfront	43.654260	-79.360636
3	M6A	North York	Lawrence Heights, Lawrence Manor	43.718518	-79.464763
4	M7A	Downtown Toronto	Queen's Park	43.662301	-79.389494

Whilst the above data is more clear, it still does not easily show us the specific area of interest and still gives us much data that is of no interest to the purpose of this report.

The data was further cleaned to only show only the relevant area we wish to focus on;

	Neighbourhood	Latitude	Longitude
0	Commerce Court, Victoria Hotel	43.648198	-79.379817

3: We can see from the preceding screenshot that the geolocation values of interest are:

Latitude: 43.648198
Longitude: -79.379817

We now have everything we need to use the FourSquare API.

4: The geolocation data was used with FourSquare, specifying a maximum distance of 1000 meters from the centre of the financial district.

The results were restricted to the first 100 venues received.

After retrieving venue data from FourSquare we can examine it to see how user-friendly, or otherwise, the data is and determine which next steps we may need to process the data further.

Here is an example of a small portion of the data received by the FourSquare request:

```
'venue': {'id': '501ae947e4b0d11883b910a7',
  'name': 'Equinox Bay Street',
  'location': {'address': '199 Bay St',
    'crossStreet': 'at Commerce Court West, PATH Level',
    'lat': 43.64809974034856,
    'lng': -79.37998869411526,
    'labeledLatLngs': [{'label': 'display',
      'lat': 43.64809974034856,
      'lng': -79.37998869411526}],
    'distance': 17,
    'postalCode': 'M5L 1L5',
    'cc': 'CA',
    'city': 'Toronto',
    'state': 'ON',
    'country': 'Canada',
    'formattedAddress': ['199 Bay St (at Commerce Court West, PATH Level)',
      'Toronto ON M5L 1L5',
      'Canada']},
  'categories': [{'id': '4bf58dd8d48988d176941735',
    'name': 'Gym',
    'pluralName': 'Gyms',
    'shortName': 'Gym',
    'icon': {'prefix': 'https://ss3.4sqi.net/img/categories_v2/building/gym_',
      'suffix': '.png'},
    'primary': True}],
  'photos': {'count': 0, 'groups': []},
  'referralId': 'e-0-501ae947e4b0d11883b910a7-0',
  'reasons': {'count': 0,
    'items': [{'summary': 'This spot is popular',
      'type': 'general',
      'reasonName': 'globalInteractionReason'}]}
```

We can see that we have a great deal of unnecessary data, and the above output retains only to one single venue.

We have data for 100 venues returned from FourSquare, so we can clean and convert this data into an easier to manage, and view, format.

5: As we can see from the preceding output, the data returned by FourSquare is quite extensive and is far more than we require for our analysis. In order to easier use this data we will convert it, from its JSON format, into a more-usable Pandas dataframe (only a small section of the resulting output is shown):

	name	categories	lat	lng
0	Equinox Bay Street	Gym	43.648100	-79.379989
1	Canoe	Restaurant	43.647452	-79.381320
2	Walrus Pub & Beer Hall	Pub	43.647375	-79.379515
3	Mos Mos Coffee	Café	43.648159	-79.378745
4	Pilot Coffee Roasters	Coffee Shop	43.648835	-79.380936
5	Hockey Hall Of Fame (Hockey Hall of Fame)	Museum	43.646974	-79.377323
6	Cactus Club Cafe	American Restaurant	43.649552	-79.381671
7	Beerbistro	Gastropub	43.649419	-79.377237
8	Dineen Coffee	Café	43.650497	-79.378765
9	Adelaide Club Toronto	Gym / Fitness Center	43.649279	-79.381921
10	DAVIDsTEA	Tea Room	43.646658	-79.377641
11	Maman	Café	43.648309	-79.382253

As we can see, this data is much easier for humans to work with than the data as-received from the FourSquare API call.

We can clean the data still further, to improve its ease of use, as we are only interested in coffee shops in the financial district and none of the other business types.

The output of this provides exactly what we need, in an easier to use, and read, format:

	index	name	categories	lat	lng
0	4	Pilot Coffee Roasters	Coffee Shop	43.648835	-79.380936
1	16	Dineen @CommerceCourt	Coffee Shop	43.648251	-79.380127
2	22	Pilot Coffee Roasters	Coffee Shop	43.645018	-79.380415
3	25	Starbucks	Coffee Shop	43.647261	-79.378599
4	40	Starbucks	Coffee Shop	43.646731	-79.383951
5	57	Starbucks	Coffee Shop	43.650159	-79.377793
6	70	M Square Coffee Co	Coffee Shop	43.651218	-79.383555
7	78	Balzac's Coffee	Coffee Shop	43.644373	-79.383065
8	82	Versus Coffee	Coffee Shop	43.651213	-79.375236

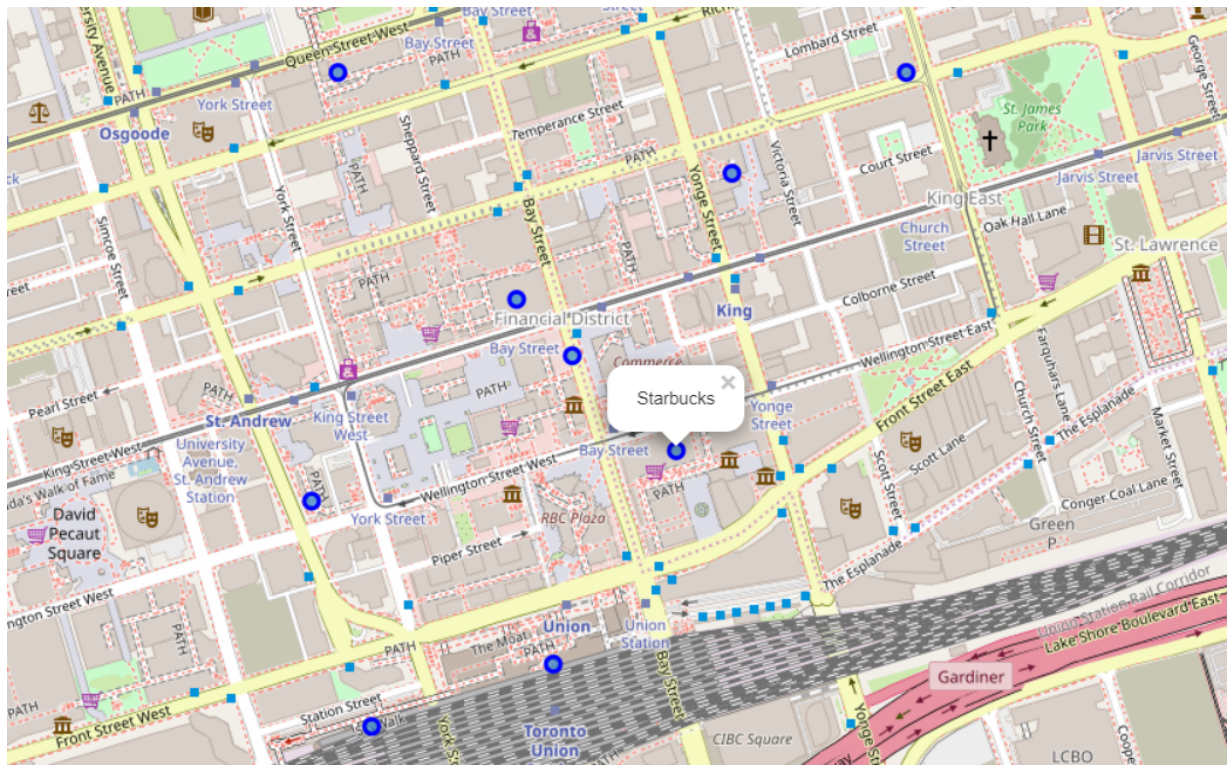
6: Now that we have cleaned and sorted the data we can easily see we have the latitude and longitude data, as well as the venue names, for each of the coffee shops in the area of interest.

We will use each venue's latitude and longitude data to visualise the coffee shops on a map of the financial district of Toronto, as Speedy Coffee will use this map for defining their strategy:

	index	name	categories	lat	lng
0	4	Pilot Coffee Roasters	Coffee Shop	43.648835	-79.380936
1	16	Dineen @CommerceCourt	Coffee Shop	43.648251	-79.380127
2	22	Pilot Coffee Roasters	Coffee Shop	43.645018	-79.380415
3	25	Starbucks	Coffee Shop	43.647261	-79.378599
4	40	Starbucks	Coffee Shop	43.646731	-79.383951
5	57	Starbucks	Coffee Shop	43.650159	-79.377793
6	70	M Square Coffee Co	Coffee Shop	43.651218	-79.383555
7	78	Balzac's Coffee	Coffee Shop	43.644373	-79.383065
8	82	Versus Coffee	Coffee Shop	43.651213	-79.375236

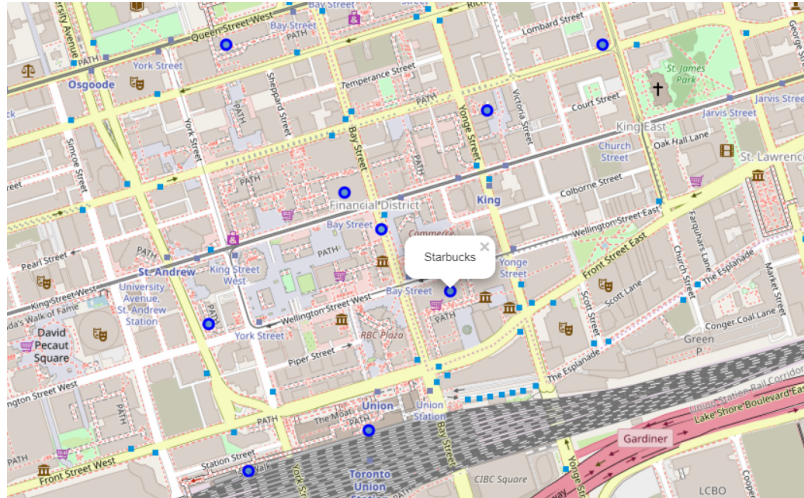
7: Now all we need to do is to map out the coffee shop locations.

Using Leaflet via Folium, the resulting screenshot shows the nine coffee shop locations:



Results:

As we can see from the map (a smaller version is shown below) there are only nine coffee shops in the data returned from FourSquare in the financial district of Toronto, and they are generally reasonably-well spaced, giving potentially many strategies for Speedy Coffee:



Discussion:

We can make some observations from the results as follows:

There are many options for Speedy Coffee to site their mobile coffee carts in and around the financial district, in a variety of location strategies.

The data received from FourSquare shows us there are nine coffee shops in the area of interest to Speedy Coffee, however, this data may only cover above-ground venues.

This may be relevant to Speedy Coffee as a [Wikipedia article](#) states that this area contains the largest underground shopping complex in the world (with 371,600 square metres/4,000,000 sq ft of retail space). This underground complex is known as the 'Path'.

Summary:

The 'Path' may well host other coffee shops within it and this should be investigated further. The results of such an investigation should be considered in addition to the findings of this report, as it is entirely possible that many more coffee shops exist in this area.

It could also be suggested that perhaps any underground coffee shops are not necessarily in direct competition with Speedy Coffee's mobile units, or it is possible that Speedy Coffee also intend to utilise the Path in their strategy.

These types of strategic decisions are unknown to the author and are outside of the intended scope of this report.