Data:

The “Snacks On Wheels” project need location data and trending venues nearby information for its analysis.

In this project, we will be required to explore, segment, and cluster the neighborhoods in the city of Toronto. For the Toronto neighborhood data, a Wikipedia page exists that has all the information we need to explore and cluster the neighborhoods in Toronto. We will be required to scrape the Wikipedia page and wrangle the data, clean it, and then read it into a *pandas* dataframe so that it is will be in a structured format for further analysis.

1. Source: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M,
2. This data contain Postcode ,Borough & Neighbourhood field .
3. Adjust Column header , Ignore cells with a borough "Not assigned”
4. If a cell has a borough but a Not assigned neighborhood, then the neighborhood will be the same as the borough
5. If more than one neighborhood exist in one postal code area , combine them with comma separated

	Postcode	Borough	Neighbourhood
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
5	M1J	Scarborough	Scarborough Village
6	M1K	Scarborough	East Birchmount Park, Ionview, Kennedy Park
7	M1L	Scarborough	Clairlea, Golden Mile, Oakridge
8	M1M	Scarborough	Cliffcrest, Cliffside, Scarborough Village West
9	M1N	Scarborough	Birch Cliff, Cliffside West

Once the data is in a structured format, we need to add Geospatial Data (latitude and longitude) based on postal code. Even though geospatial data is available with google location service , due to frequent changes of google API , this project used the location data which is available on https://cocl.us/Geospatial_data

	PostalCode	Borough	Neighbourhood	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
5	M1J	Scarborough	Scarborough Village	43.744734	-79.239476
6	M1K	Scarborough	East Birchmount Park, Ionview, Kennedy Park	43.727929	-79.262029
7	M1L	Scarborough	Clairlea, Golden Mile, Oakridge	43.711112	-79.284577
8	M1M	Scarborough	Cliffcrest, Cliffside, Scarborough Village West	43.716316	-79.239476
9	M1N	Scarborough	Birch Cliff, Cliffside West	43.692657	-79.264848

Once the current location information of “Snacks on wheel trailer” is available (postal code or borough – User will input this data or use GPS) , which is used to get the “trending venues nearby” data.

Foursquare API is used to get the trending venues nearby information. These are venues that have the highest foot traffic when the call to the database is made. Therefore the results vary depending on when the call is made. So in the morning we might find that trending venues are coffee shops or office spaces whereas in the evening trending venues can be malls, museums or parks. And to get the trending venues, we simply use the trending endpoint along with the credentials and the latitude and the longitude coordinates of the place of interest. When we make the

call to the foursquare database, we get a JSON file of the trending venues that are nearby. In the JSON file, for each trending venue, we get mostly its name, unique ID, location, and category.

The Number of results to return is limited by top 10 , with in the radius of 2000 meters. The distance between current location and trending location is calculated for further analysis. If distance is not available in Foursquare response we uses the 'haversine' formula to calculate the great-circle distance between two points – that is, the shortest distance over the earth's surface – giving an 'as-the-crow-flies' distance between the points (ignoring any hills they fly over, of course!).

Haversine formula: $a = \sin^2(\Delta\varphi/2) + \cos\varphi1 \cdot \cos\varphi2 \cdot \sin^2(\Delta\lambda/2)$

$c = 2 \cdot \text{atan2}(\sqrt{a}, \sqrt{(1-a)})$

$d = R \cdot c$

Where φ is latitude, λ is longitude, R is earth's radius (mean radius = 6,371km);

Note that angles need to be in radians to pass to trig functions!

Foursquare Response sample:


```
{ "meta": { "code": 200, "requestId": "5ac51dde351e3d4df64064f8" }, "response": { "venues": [ { "id": "5735dc3f498e1ac6a088f324", "name": "Union Fare", "location": { "address": "5 E 17th St", "crossStreet": "btwn 5th Ave & Union Sq W", "lat": 40.737697, "lng": -73.991402, "labeledLatLngs": [ { "label": "display", "lat": 40.737697, "lng": -73.991402 } ], "distance": 1802, "postalCode": "10003", "cc": "US", "city": "New York", "state": "NY", "country": "United States", "formattedAddress": [ "5 E 17th St (btwn 5th Ave & Union Sq W)", "New York, NY 10003", "United States" ] }, "categories": [ { "id": "4bf58dd8d48988d157941735", "name": "New American Restaurant", "pluralName": "New American Restaurants", "shortName": "New American", "icon": { "prefix": "https://ss3.4sqi.net/img/categories_v2/food/newamerican_", "suffix": ".png" }, "primary": true } ] } ] } }
```