**THE BATTLE OF NEIGHBORHOODS**

|  |
| --- |
| **Barath Srinivasan** |

**INTRODUCTION/BUSINESS PROBLEM:**

This project is an attempt to leverage Four Square Location Service Provider API to address a unique problem applying learnings from earlier practical sessions. Commonly experienced problem of moving to a new apartment to a different location among considered areas was constructed.

Mr. John, 65 years old, has decided to move to a new apartment in Toronto with his wife, to one among the shortlisted Boroughs:

* Downtown Toronto
* East Toronto
* West Toronto
* Central Toronto

As with old age, they have some pre-existing medical conditions and they need a location in the close vicinity of Hospitals for their treatment to be continued. Mr. John is a **Heart Patient** and his requirement is that the new location should be nearer to **heart speciality clinics**. Mrs. John has been visiting **Toronto General Hospitals** for past few years and she needs atleast one among these hospitals accessible from the new location. Apart from this condition, they are flexible enough to handle other requirements.

They need our help to **suggest them an apt location** as near as possible according to above conditions, so that they can move without any hesitation and also continue to take their treatments in this old age without any hassle.

**DATA USED:**

The Toronto neighborhood data is extracted from their Wikipedia page.

[**https://en.wikipedia.org/wiki/List\_of\_postal\_codes\_of\_Canada:\_M**](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M)

This has all the information we need to explore the neighborhoods in Toronto.

It is critical to convert this data into a pandas dataframe, in a structured format, clean it and make it inferable to carry out the analysis to explore and decide a prompt location according to Mr. and Mrs. Johns conditions.

*Structure of Data used:*

* The above data has three columns – Postal Code, Borough and Neighborhood of Toronto. We can see many NA values in them and they need to be handled. We processed only the cells with an assigned Borough and removed the rows if Borough Column had NA values. More than one neighbourhood can exist in one postal code area. For example, in the table on the Wikipedia page, we notice that **M5A** is listed twice and has two neighbourhoods: **Harbourfront**and **Regent Park**. These two rows were combined into one row with the neighbourhoods separated with a comma.

Table

Description automatically generated

**Figure 1. Sample of cleaned dataset**

**METHODOLOGY:**

The steps applied and any interesting aspects are shared below:

*Data Extraction, Cleaning and Visualization:*

A gist of this was already explained above. There are multiple methods to do this. A common method which was also suggested during practical sessions was the use of **BeautifulSoup** Library. We can scrape any webpage and apply methods available to extract details from BeautifulSoup Object. In this case we can extract the table, process the elements into a list and create a dataframe.

Lesson learnt is that this library has useful methods to extract from complex web pages. But to just extract the table, **read html from pandas library** can be directly applied. This saved a lot of time.

A picture containing text

Description automatically generated

**Figure 2. Read Html method from Pandas**

The Latitude and Longitude coordinates were directly added from CSV provided earlier. This can also be extracted by applying any of the geocoding libraries available. Our finalized dataset is created and it is inferable for data analysis now. To quickly visualize the scope – the dataset is filtered only of interested locations of our requirement and a bar chart was used. From 103 observations only 39 of them match the requirement with highest number of locations in Downtown Toronto.

Chart, bar chart

Description automatically generated

**Figure 3. Visualize locations of interest**

*Map Creation and Plotting:*

The coordinates were determined using **Nominatim function from geopy.** The user agent for accessing the location object was given and the exact location along with State and Country was provided to fetch latitude and longitude coordinates. These values were fed to be plotted in the Map. For this, **Folium** Library was used with appropriate zoom levels to portray the correct location scale and suitable markers were added. This helped to visualize these areas in the Map and firmed up our scope.

Map

Description automatically generated

**Figure 4. Location scope – Plotted using Folium**

*Four Square API – Location Service Provider:*

Four Square is a location service provider which provided nearby areas to be explored, venues in the location, tips and reviews. They can provide this information through search query API based transactions that happen by registering with them a developer account. The security information and tokens are required to establish connectivity and transact with them. A URL was created with location information, connection information and our query. As per our requirement for Mr. and Mrs. John, Hospitals information that was near(500 metres) was queried from Foursquare API in Downtown Toronto which had the maximum number of locations. Filters were applied to get Toronto General and Heart Speciality Clinics from the total venue results. The Neighbourhoods that contained the results were grouped and visualized for clarity.

Graphical user interface, text, application

Description automatically generated

Chart, bar chart

Description automatically generated

**Figure 5. Neighbourhoods satisfying Hospital condition**

The Geo locations of the neighbourhoods were plotted and nearest area to all these places was shortlisted as one of the solutions to this problem

A picture containing text

Description automatically generated

**Figure 6. Location Suggestion – nearest to all of the visit places geolocations**

We can see that **St.Patricks near Central Bay** would be a great location from where both **Heart Speciality Clinics and Toronto General Hospitals** are conveniently accessible. Mr and Mrs.John would be happy to move to this kind of a place.

*Discussion: Determining the shortest distance to geolocations – Intuition:*

Based on the take away from project, we decided that Centroid of these locations – would suffice as a solution.

**Why ?**

If accuracy is the key criteria of the project, then we may need to apply the concept of **Great Arcs - Haversine distance**. This is an interesting concept which considers the curvature of earth as a sphere in picture, which is the actual reality. A sample solution was implemented and checked in stack overflow community.

But as per their discussion and scale of impact, As our locations in this particular area **are not very far from each other**, we can roughly consider earth as a plane and estimated the new location by determining the **centroid** of the surrounding geo location coordinates. This would be the simple yet effective recommendation.

*Conclusion :*

The use of location data to explore a geographical location is quite a handy and effective tool to solve specific problems like above.