CAPSTONE PROJECT

By Manish Bhatia

Unification of Restaurants based on Cuisine type in Downtown TORONTO, Canada and suggesting optimum cluster to Open an Indian Restaurant

Introduction: -

Toronto city in Canada is one of the most multicultural & lively cities in the world, this is the reason Toronto comes in Top 10 best cities to live in and attract a lot of tourists from all over the Globe. The city is famous of beautiful lakes, sceneries and outdoor life. The one major characteristic of this city is that it is having more than 50% population which are migrated from outside of Canada.

This has resulted in diverse food choices of people who are living here or the tourists who usually visit in seasonal time. Due to this there are around 7500 restaurants of different cuisine styles are opened like German, Italian, Japanese, Indian, Asian and so on.

Description of Problem Statement: -

As it becomes hard to conclude the choices of which area/restaurant to visit in a long list of restaurants or which are the **best restaurant venues nearby** or a suggestion to open an **Indian restaurant** in which neighborhood where it could attract more people and could possibly generate more revenue.

In this project we will analyze most famous area '**Downtown Toronto'** of Toronto city using its demographics characteristics and will visualize which specific cuisine restaurants are most common in certain neighborhoods using **clustering** methodology so that tourists could plan the trip based on their food choices and also we will try to identify the neighborhoods where an Indian restaurant could be opened based on existing restaurant number and food style priorities.

Description of Data: -

Based on problem statement, below sources were utilized to extract the desired data

1. **Wikipedia**: - **'https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M'**
   This data will provide information about the 'Postal codes', 'Boroughs' and 'Neighborhoods' of Toronto city which can help in filtering 'Downtown Toronto' from the Boroughs and its associated neighborhoods.

	Postal Code	Borough	Neighbourhood
0	M1A	Not assigned	Not assigned
1	M2A	Not assigned	Not assigned
2	M3A	North York	Parkwoods
3	M4A	North York	Victoria Village
4	M5A	Downtown Toronto	Regent Park, Harbourfront

2. **Geospatial Data: - 'http://cocl.us/Geospatial_data**
   With the help of geospatial data, we can extract information about 'Latitude' and 'Longitude' of each and every postal code of Toronto city and with the help of merging technique , the location information can be imposed on Neighborhoods on the table extracted from Data source 1.

	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

3. **Foursquare API:-**
   https://api.foursquare.com/v2/venues/explore?&client_id={}&client_secret={}&v={}&ll={},{}&radius={}&limit={}'.format(CLIENT_ID, CLIENT_SECRET, VERSION, neighborhood_latitude,neighborhood_longitude,radius,LIMIT)

   This API will help in retrieving information about the most famous venues in 'Downtown Toronto' and the category of those venues like Restaurants, Bars, park etc.

   The main advantage of this API is that the data can be extracted based on geographical parameters and which can be modified to extract desired information in a certain radius and desired number of venues are also changeable.

4. **Demographic information from WIKI: - https://en.wikipedia.org/wiki/Demographics_of_Toronto_neighbourhoods**
   This data from wiki webpage is to get demographic information of specific neighborhoods , this table can help in studying living population in certain neighborhoods and in which areas the most spoken language was either 'Hindi' or 'Punjabi' so that the proposal of Indian restaurants could be decided and chances of higher revenue could get increased.

How Data will be used to solve the problem: -

The Toronto neighborhood data will be extracted form wiki page1 and this will provide an association of all the neighborhoods in 'Downtown Toronto' area of Toronto city. After extracting this information, the table will be mapped with Geospatial data which will provide latitude and longitude of each neighborhood in Downtown Toronto. Once it is done then by using Foursquare API extracted table will be utilized to extract venues and their categories in each neighborhood. Thus, using this API Categories of different restaurants will be filtered out for each neighborhood and we can extract information of most famous restaurants in the neighborhoods and also can categories them by rank.

After categorizing, the data can help in proper clustering of restaurants in each neighborhood, this could be a valuable information for tourists to filter out the neighborhoods where specific cuisine restaurant is most famous.

Using Folium map, we will try to plot each neighborhood on Toronto map and will visualize the distribution of clusters of restaurants of different cuisine. This data will also support in extracting information about the clusters where Indian restaurants are most famous or in specific neighborhoods Indian restaurants are not in top 15 venues.

Thus by imposing demographic information like 'population' and the spoken language in any neighborhood we can finalize a cluster where other restaurants are most famous but the locality is of people who belongs to Indian or Asian community and proposal of an Indian restaurant could be favorable.