IBM data Science Capstone Project

Week 4: Part - 2
Choosing the best place for living based on the nearby amenities.

Data description

I. Required data to solve the problem:

1. ***Location of interest:*** It is mainly the address of the workplace or the center of interest based on what the whole analysis will be conducted. For this project we are considering **'301 Front St W, Toronto, ON' –** address of CN tower, as our workplace location.
2. ***Apartment data:*** This the data that contains all the nearby Residential Building (Apartment / Condo) with in 5 Km distance from the location of interest. This data mainly contains the latitude and longitude coordinates of those apartments. This is required in order to plot the map and to get the nearby venue data.
3. ***Postal codes of Canada:*** (*https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M*) This data contains the Postal codes, Borough and the Neighborhoods of Toronto city. This data was merged with the '**Apartment data**' on the postal codes to check the Boroughs of each apartments within the radius of the location of interest.
4. ***Geospatial data:*** (*http://cocl.us/Geospatial_data*) This data contains the latitude and the longitude coordinates of all the Postal codes of Toronto city. This data was merged with the '***Postal codes of Canada***' data to get the neighborhoods with corresponding coordinates. This data is useful to find a residence based on the neighborhood amenities instead of the apartments.

II. Reasons of choosing Toronto data:

Toronto city neighborhoods were chosen as the observation target due to the following reasons:

- The availability of higher job opportunity,
- The availability of geo data which can be used to visualize the dataset onto a map.

III. The process of collecting and clean data:

- Location of interest or the workplace data was chosen randomly at the heart of Toronto downtown by searching on google.
- Apartment data was found using foursquare API by searching for the nearby Residential Building (Apartment / Condo) with in 5 km distance. The cleaned apartments dataset is given below:

	Apt_Name	Address	Distance	Formatted_Address	Latitude	Longitude	Postalcode
0	University Apartment	16 willison sq	1427	[16 willison sq (spandina AVe), Toronto ON M5T...	43.652168	-79.398821	M5T
1	Elm Place Apartments	222 Elm St.	1576	[222 Elm St. (Elm & Mccaul), Toronto ON M5T 1K...	43.656243	-79.392139	M5T
2	Epitome Apartments	160 Huron Street	1836	[160 Huron Street, Toronto ON M5T 2B6, Canada]	43.657119	-79.397814	M5T
3	1 Homewood Ave Apartments	1 Homewood Ave	2514	[1 Homewood Ave (Carlton St), Toronto ON M4Y 2...	43.663222	-79.374450	M4Y
4	Allan Plaza Apartments	166 Carlton Street	2568	[166 Carlton Street, Toronto ON M5A 2K5, Canada]	43.663463	-79.373573	M5A

Figure 1 – Apartment dataset

- Scraped Wikipedia page and transformed the data into a structured data frame.
- Geospatial data of Toronto neighborhoods were found from week 3 module of the IBM capstone project course and merged with the 'Postal codes of Canada' data. The Toronto neighborhoods dataset and the final data set for analysis are given below:

	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

Figure 2 – Toronto neighborhoods dataset

	Apt_Name	Address	Distance	Formatted_Address	Latitude	Longitude	Postalcode	Borough
0	University Apartment	16 willison sq	1427	[16 willison sq (spandina AVe), Toronto ON M5T...	43.652168	-79.398821	M5T	Downtown Toronto
1	Elm Place Apartments	222 Elm St.	1576	[222 Elm St. (Elm & Mccaul), Toronto ON M5T 1K...	43.656243	-79.392139	M5T	Downtown Toronto
2	Epitome Apartments	160 Huron Street	1836	[160 Huron Street, Toronto ON M5T 2B6, Canada]	43.657119	-79.397814	M5T	Downtown Toronto
3	1 Homewood Ave Apartments	1 Homewood Ave	2514	[1 Homewood Ave (Carlton St), Toronto ON M4Y 2...	43.663222	-79.374450	M4Y	Downtown Toronto
4	Cromwell Apartments	55 Isabella Street	2869	[55 Isabella Street, Toronto ON M4Y 1M8, Canada]	43.668187	-79.383245	M4Y	Downtown Toronto

Figure 3 – Final apartment dataset in Toronto city

IV. How the data will solve the problem:

- Initially a location is required to look for nearby neighborhoods or residences. **Location of interest** data will provide the nearby residential building (Apartment / Condo) data within 5 Km distance.
- **Apartment data** and the **Postal codes of Canada** data will be useful to get the postal codes, neighborhoods, boroughs, distances from the location of interest and nearby amenities from the apartments to analyze the residential buildings.
- **Geospatial data** will provide the coordinates of the Toronto neighborhoods those will be used to analyze the neighborhood amenities for choosing the best one among others if analyzed by the neighborhoods instead of residences.