REPORT:

Capstone Project - Covid Test Centers Segmenting and Clustering in Toronto

Introduction: Business Problem

COVID-19 has impacted the whole world at unprecedented levels.
Governments and Health care departments are put under extreme pressure to come up with various tasks in the battle against this pandemic. One of many such projects is setting up test centers to assess patients.
In this project, we'll be segmenting and clustering the Covid-19 Test centers in city of Toronto.
By plotting the test centers on the city map gives us a visual of how they are spread across this vast city with a population of 3 million.

At the end using Foursquare and 'K nearest neighbors' data science concepts we find the 10 most common venues around each of the test centers. This will help in understanding what kind of businesses or amenities exist around each test center.

Stakeholders for the insights achieved can be many;
a) Government officials while planning more test centers in other neighborhoods than the present ones.
b) Patients to quickly identify what kind of services/amenities exist around each test center e.g Parking centers.
c) Businesses to take extra precautions for possible exposure due to increased influx of people in the neighborhood.

Data:

To achieve the solution, we begin with the official datasets provided by Govt of Onatrio in their website that list the Covid test centers across the Province.

URL: https://data.ontario.ca/dataset/8ba078b2-ca9b-44c1-b5db-9674d85421f9/resource/04bede2c-5e30-4a05-b890-cd407043485e/download/assessment-centre-locations.csv

Along with test center name file contains city name, postal_code, latitude and longitude details that we'll be using when plotting maps (using Folium) and when segmenting and clustering while finding 10 most common venues around each test center present in Toronto (using Four Square API).

Pandas Data Frame;

	location_name	city	address	postal_code	latitude	longitude
0	Kirkland and District Hospital	Kirkland Lake	145 Government Road East	P2N 3P4	48.153552	-80.014725
1	Collingwood Health Centre	Collingwood	186 Erie Street	L9Y 4T3	44.501231	-80.204437
2	Midland Assessment Centre	Midland	845 King Street, Unit 3	L4R 0B7	44.734750	-79.870959
3	Milton District Hospital	Milton	725 Bronte Street South	L9T 9K1	43.497418	-79.868476
4	Oakville Trafalgar Memorial Hospital	Oakville	3001 Hospital Gate	L6M 0L8	43.450869	-79.763927

Methodology

In the first step, we have collected the required data which contained below fields;

1) **location_name**
2) **city**
3) **address**
4) **postal_code**
5) **latitude**
6) **longitude**

We also performed necessary data wrangling for smooth analysis further.

In the Second step in our analysis we will start using Four Sqaure API for finding the nearby venues around each test center.

In third and final step we will focus on plotting our findings on maps using Folium and then use K nearest neighbor clustering, thereby finding the n most common venues around each test center.

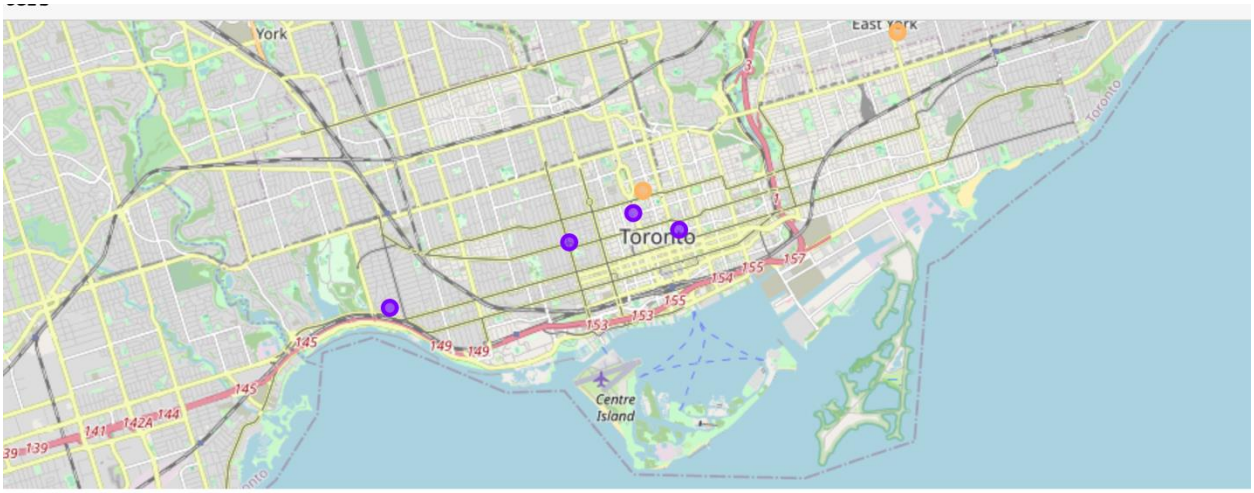
Using **Foursquare API** we got all the nearby venues;

	Test_Center	Test_Center Latitude	Test_Center Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Humber River Hospital Assessment Centre	43.754813	-79.525921	Best Western Plus Toronto North York Hotel & S...	43.756887	-79.528779	Hotel
1	Humber River Hospital Assessment Centre	43.754813	-79.525921	Holiday Inn Express Toronto-North York	43.756232	-79.527348	Hotel
2	Humber River Hospital Assessment Centre	43.754813	-79.525921	Tim Hortons	43.754344	-79.527024	Coffee Shop
3	Humber River Hospital Assessment Centre	43.754813	-79.525921	Hwy 400 at Finch W.	43.754399	-79.526967	Intersection
4	Humber River Hospital Assessment Centre	43.754813	-79.525921	Perkins	43.756567	-79.527475	American Restaurant

K nearest neighbor clustering

	location_name	city	address	postal_code	latitude	longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue
100	Humber River Hospital Assessment Centre	Toronto	2111 Finch Avenue West	M3N 1N1	43.754813	-79.525921	2	Hotel	American Restaurant	Coffee Shop	Intersection	Diner	Eastern European Restaurant	Donut Shop	Doner Restaurant
102	Michael Garron Hospital - Outpatient Clinic	Toronto	825 Coxwell Avenue	M4C 3E7	43.689910	-79.324858	4	Coffee Shop	Café	Park	Pizza Place	Farmers Market	Dance Studio	Sandwich Place	Diner
103	Market Place Temporary Assessment Centre	Toronto	4 The Market Place	M4C 5M1	43.695869	-79.292138	3	Golf Course	Metro Station	Convenience Store	Park	Creperie	Curling Ice	Dance Studio	Deli / Bodega
104	Mount Sinai Hospital	Toronto	600 University Avenue	M5G 1X5	43.657575	-79.390096	1	Coffee Shop	Café	Japanese Restaurant	Italian Restaurant	Ramen Restaurant	French Restaurant	Art Gallery	Bar

Plotting our findings on Toronto map using **Folium;**



Results and Discussion

Here we are with all the results for analysis.

First using the datasets we narrowed down to COVID test centers available in city of Toronto.
This we achieved by narrowing down from all the locations listed in the province of Ontario's dataset.
As observed we have 5 locations across the city in different neighborhoods.
In order to understand proximity of each of those we plotted them on map and this gave us an insight of a visual representation of their presence.

Next using Four Square we identified most common business amenities available around each of the test center location.
For this we used the latitude and longitude of the respective test centers and used the Four Square's features to fetch each of the venues around it.
Finally using K-nearest-neighbors data science algorithm we segmented and clustered each of the above mentioned locations.

Conclusion

Purpose of this project was to identify COVID test center locations across Toronto. And the results of this will help the stake holders as mentioned below;
a) Government officials while planning more test centers in other neighborhoods than the present ones.
b) Patients to quickly identify what kind of services/amenities exist around each test center e.g Parking centers.
c) Businesses to take extra precautions for possible exposure due to increased influx of people in the neighborhood.
Considering the results, we achieved by plotting test centers on Toronto map and by finding the most common venues around each test center we can conclude that we have addressed the business problem we started with.