“Coursera IBM Data Science Specialization”

Capstone Final Project Report

“Business Opportunities for Greek Restaurants in Toronto, CA”

FINAL REPORT

January, 2021

Report Owner: **Paris Pantazis**

Github: https://github.com/ppantazis1

Intro

The City of Toronto, is one of the most populous city in Canada. It is multicultural. It provides lot of business opportunities and business friendly environment. It has attracted many different players into the market. It is a global hub of business and commerce. This also means that the market is highly competitive. As it is highly developed city so cost of doing business is also high. Thus, any new business venture or expansion needs to be analyzed carefully. The insights derived from analysis will give good understanding of the business environment which help in strategically targeting the market. This capstone project is focused on providing such an analysis for a Greek restaurant business in Toronto.

Business Problem

Toronto's food culture includes an array of international cuisines influenced by the city's immigrant history. Greek restaurants have become popular in Canada. Starting a Greek restaurant can be a great business opportunity, but you need to distinguish yourself from others to enjoy long-term success.

My client wants to open his business in Toronto area, so we define potential neighborhood based on the number of Greek restaurants which are operating right in each neighborhood. Toronto has full potential but also is a very challenging district to open a business because of high competition. New Greek restaurants should be open in an area that inadequate neighborhood in this way the bar can attract more customers. Therefore, this analysis necessary to ensure that we have enough customers and that we are not so close to other GR restaurants.

By using data science methods and machine learning methods such as clustering, this project aims to answer the business question: "In Toronto, if an entrepreneur wants to open a Greek restaurant, where should they consider opening it?"

Target Audience

Entrepreneurs who wants to find the optimal location in Toronto to open a Greek restaurant, in a way to minimize risk and maximize Return on Investment (ROI).

Data

To proceed with the analysis, below data was used:

- List of neighborhoods in Toronto, Canada.
    - *https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada*

- Latitude and Longitude of these neighborhoods.

- *Geocoder package or,*
- *http://cocl.us/Geospatial_data (Existing csv file from IBM)*

- Venue data related to Greek restaurants.
  - *Using Foursquare API to get venue data related to these neighborhoods*

Methodology

1. Get the list of neighborhoods in Toronto, Canada via web scraping (Wikipedia) by utilizing pandas html table scraping method as it is easier and more convenient to pull tabular data directly from a web page into data frame. Apply exploratory analysis and feature engineering to the original df.

Out[14]:

	Postal Code	Borough	Neighbourhood
0	M1B	Scarborough	Malvern, Rouge
1	M1C	Scarborough	Rouge Hill, Port Union, Highland Creek
2	M1E	Scarborough	Guildwood, Morningside, West Hill
3	M1G	Scarborough	Woburn
4	M1H	Scarborough	Cedarbrae

2. Retrieve their coordinates to utilize Foursquare to pull the list of venues near these neighborhoods. For this purpose, I used the csv file provided by IBM team (see Data section) to match the coordinates of Toronto neighborhoods, since it wasn't possible via the Geocoder package.

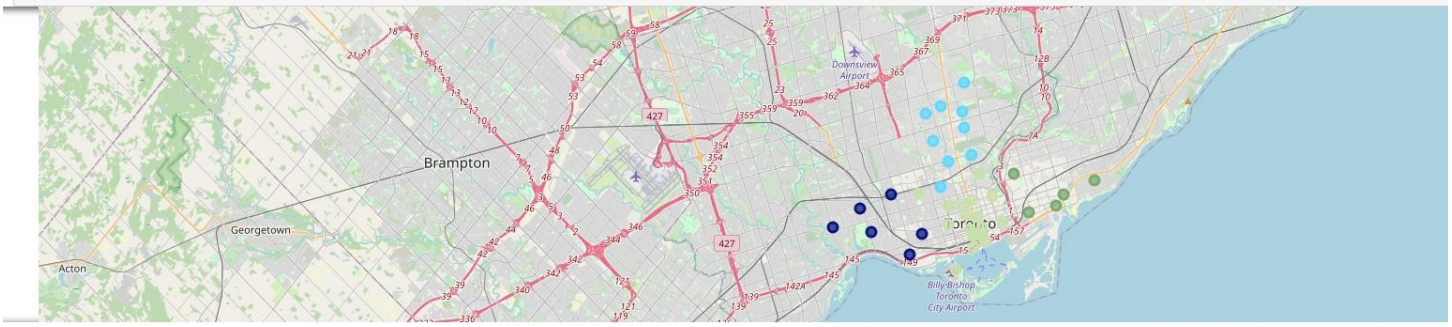
In [20]:
```
#Read CSV file from link and load into dataframe
url_csv = 'http://cocl.us/Geospatial_data'
df_coordinates = pd.read_csv(url_csv)
df_coordinates.head()
```

Out[20]:

	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. Visualize the map of Toronto using Folium package to verify whether these are correct coordinates.

```
map_toronto
```



4. Next, Foursquare API was used with my personal Client ID/Secret dev credentials to pull the list of top 100 venues within 500 meters radius. From Foursquare, I pulled the names, categories, latitude and longitude of the venues and checked how many unique categories exist. Then, analyze each neighborhood by grouping the rows by neighborhood and taking the mean on the frequency of occurrence of each venue category. This is to prepare clustering task.

```
In [39]: #Number of venues per neighborhood
         toronto_venues.groupby('Neighborhood').count()
```

Out[39]:

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Adelaide, King, Richmond	100	100	100	100	100	100
Berczy Park	56	56	56	56	56	56
Brockton, Exhibition Place, Parkdale Village	23	23	23	23	23	23
Business Reply Mail Processing Centre 969 Eastern	16	16	16	16	16	16
CN Tower, Bathurst Quay, Island airport, Harbourfront West, King and Spadina, Railway Lands, South Niagara	16	16	16	16	16	16
Cabbagetown, St. James Town	47	47	47	47	47	47
Central Bay Street	83	83	83	83	83	83
Chinatown, Grange Park, Kensington Market	87	87	87	87	87	87
Christie	19	19	19	19	19	19
Church and Wellesley	82	82	82	82	82	82
Commerce Court, Victoria Hotel	100	100	100	100	100	100
Davisville	32	32	32	32	32	32
Davisville North	8	8	8	8	8	8
Deer Park, Forest Hill SE, Rathnelly, South Hill, Summerhill West	15	15	15	15	15	15
Design Exchange, Toronto Dominion Centre	100	100	100	100	100	100
Dovercourt Village, Dufferin	16	16	16	16	16	16
First Canadian Place, Underground city	100	100	100	100	100	100
Forest Hill North, Forest Hill West	4	4	4	4	4	4
Harbord, University of Toronto	37	37	37	37	37	37
Harbourfront	47	47	47	47	47	47
Harbourfront East, Toronto Islands, Union Station	100	100	100	100	100	100
High Park, The Junction South	24	24	24	24	24	24
Lawrence Park	3	3	3	3	3	3
Little Portugal, Trinity	53	53	53	53	53	53
Moore Park, Summerhill East	1	1	1	1	1	1
North Toronto West	21	21	21	21	21	21
Parkdale, Roncesvalles	14	14	14	14	14	14
Queen's Park	39	39	39	39	39	39
Rosedale	4	4	4	4	4	4
Roselawn	3	3	3	3	3	3

5. Here, I made a justification to specifically look for "**Greek restaurant**" which is related directly to my analysis.

```
In [52]: #Number of unique venue categories
         print('There are {} uniques categories.'.format(len(toronto_venues['Venue Category'].unique())))

         There are 235 uniques categories.
```

```
In [41]: #print out the list of categories
         toronto_venues['Venue Category'].unique()[:100]

Out[41]: array(['Trail', 'Health Food Store', 'Pub', 'Neighborhood',
                 'Asian Restaurant', 'Greek Restaurant', 'Cosmetics Shop',
                 'Italian Restaurant', 'Ice Cream Shop', 'Brewery', 'Yoga Studio',
                 'Fruit & Vegetable Store', 'Dessert Shop', 'Pizza Place',
                 'Bookstore', 'Restaurant', 'Juice Bar', 'Bubble Tea Shop', 'Diner',
                 'Spa', 'Furniture / Home Store', 'Grocery Store', 'Coffee Shop',
                 'Bakery', 'Caribbean Restaurant', 'Frozen Yogurt Shop',
                 'American Restaurant', 'Liquor Store', 'Gym', 'Burger Joint',
                 'Fish & Chips Shop', 'Park', 'Sushi Restaurant', 'Burrito Place',
                 'Pet Store', 'Steakhouse', 'Fast Food Restaurant', 'Movie Theater',
                 'Sandwich Place', 'Light Rail Station', 'Fish Market', 'Café',
                 'Cheese Shop', 'Gay Bar', 'Seafood Restaurant',
                 'Middle Eastern Restaurant', 'Comfort Food Restaurant',
                 'Thai Restaurant', 'Stationery Store', 'Wine Bar',
                 'Coworking Space', 'Bar', 'Latin American Restaurant',
                 'Gym / Fitness Center', 'Gastropub', 'Bank', 'Convenience Store',
                 'Clothing Store', 'Music Store', 'Swim School', 'Bus Line',
                 'Food & Drink Shop', 'Breakfast Spot', 'Department Store', 'Hotel',
                 'Dance Studio', 'Chinese Restaurant', 'Salon / Barbershop',
                 'Mexican Restaurant', 'Sporting Goods Shop', 'Shoe Store',
                 'Bagel Shop', 'Rental Car Location', 'Indian Restaurant',
                 'Toy / Game Store', 'Gas Station', 'Pharmacy', 'Farmers Market',
                 'Gourmet Shop', 'Tennis Court', 'Supermarket', 'Sports Bar',
                 'Fried Chicken Joint', 'Vietnamese Restaurant',
                 'Health & Beauty Service', 'Playground', 'Japanese Restaurant',
                 'Butcher', 'Jewelry Store', 'General Entertainment',
                 'Taiwanese Restaurant', 'Deli / Bodega', 'Gift Shop', 'Market',
                 'Beer Store', 'Snack Place', 'Theme Restaurant',
                 'Ramen Restaurant', 'Beer Bar', 'Ethiopian Restaurant'],
                dtype=object)
```

```
In [54]: # check if the results contain "Greek Restaurant"
         "Greek Restaurant" in toronto_venues['Venue Category'].unique()

Out[54]: True
```

6. Lastly, I performed the clustering method by using **k-means clustering,** for k=3. I clustered the neighborhoods in Toronto into 3 clusters based on their frequency of occurrence for “Greek restaurant”.

Results

Examine Clusters: **Cluster 0**

```
In [69]: #Cluster 0
         to_merged.loc[to_merged['Cluster Labels'] == 0]

Out[69]:
```

	Neighborhood	Greek Restaurant	Cluster Labels	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
17	Kensington Market, Chinatown, Grange Park	0.0	0	43.653206	-79.400049	Blue Banana Market	43.655669	-79.402551	Furniture / Home Store
25	Richmond, Adelaide, King	0.0	0	43.650571	-79.384568	The Fifth & Terrace	43.649250	-79.389320	Modern European Restaurant
25	Richmond, Adelaide, King	0.0	0	43.650571	-79.384568	Cardio-Go	43.647017	-79.388143	Gym
25	Richmond, Adelaide, King	0.0	0	43.650571	-79.384568	Astarté Fresh Yogurt Bar	43.647596	-79.386419	Café
25	Richmond, Adelaide, King	0.0	0	43.650571	-79.384568	Dineen @CommerceCourt	43.648251	-79.380127	Coffee Shop
...	...	...	...	...	...	...	...	...	...
13	Garden District, Ryerson	0.0	0	43.657162	-79.378937	Chatime 日出茶太	43.655542	-79.384684	Bubble Tea Shop
13	Garden District, Ryerson	0.0	0	43.657162	-79.378937	JOEY Eaton Centre	43.656094	-79.381878	New American Restaurant
13	Garden District, Ryerson	0.0	0	43.657162	-79.378937	Roots	43.653613	-79.380244	Clothing Store
13	Garden District, Ryerson	0.0	0	43.657162	-79.378937	Scaddabush Italian Kitchen & Bar	43.658920	-79.382891	Italian Restaurant
13	Garden District, Ryerson	0.0	0	43.657162	-79.378937	Trattoria Mercatto	43.654453	-79.380974	Italian Restaurant

1438 rows × 9 columns

Examine Clusters: **Cluster 1**

```
In [70]: #Cluster 1
to_merged.loc[to_merged['Cluster Labels'] == 1]
```

Out[70]:

	Neighborhood	Greek Restaurant	Cluster Labels	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
36	The Danforth West, Riverdale	0.186047	1	43.679557	-79.352188	Simone's Caribbean Restaurant	43.678655	-79.346582	Caribbean Restaurant
36	The Danforth West, Riverdale	0.186047	1	43.679557	-79.352188	Urban Nails	43.676668	-79.356602	Spa
36	The Danforth West, Riverdale	0.186047	1	43.679557	-79.352188	Marvel Coffee Co.	43.678630	-79.347460	Coffee Shop
36	The Danforth West, Riverdale	0.186047	1	43.679557	-79.352188	Astoria Shish Kebob House	43.677596	-79.351738	Greek Restaurant
36	The Danforth West, Riverdale	0.186047	1	43.679557	-79.352188	Momo Hut And Gardens	43.677491	-79.351516	Tibetan Restaurant
36	The Danforth West, Riverdale	0.186047	1	43.679557	-79.352188	Dough Bakeshop	43.676643	-79.356846	Bakery
36	The Danforth West, Riverdale	0.186047	1	43.679557	-79.352188	Sher-E-Punjab	43.677308	-79.353066	Indian Restaurant
36	The Danforth West, Riverdale	0.186047	1	43.679557	-79.352188	Bulk Barn	43.676790	-79.355865	Grocery Store
36	The Danforth West, Riverdale	0.186047	1	43.679557	-79.352188	Kalyvia	43.677973	-79.351208	Greek Restaurant
36	The Danforth West, Riverdale	0.186047	1	43.679557	-79.352188	IL FORNELLO on Danforth	43.678604	-79.346904	Italian Restaurant
36	The Danforth West, Riverdale	0.186047	1	43.679557	-79.352188	Sushi Friends	43.677614	-79.351641	Japanese Restaurant
36	The Danforth West, Riverdale	0.186047	1	43.679557	-79.352188	Cafe Frappe	43.678126	-79.348434	Coffee Shop
36	The Danforth West, Riverdale	0.186047	1	43.679557	-79.352188	Factory Girl	43.676693	-79.356299	American Restaurant
36	The Danforth West, Riverdale	0.186047	1	43.679557	-79.352188	Book City	43.677413	-79.352734	Bookstore
36	The Danforth West, Riverdale	0.186047	1	43.679557	-79.352188	Bar Oak	43.677931	-79.348724	Lounge
36	The Danforth West, Riverdale	0.186047	1	43.679557	-79.352188	Menchie's	43.678309	-79.348105	Frozen Yogurt Shop
36	The Danforth West, Riverdale	0.186047	1	43.679557	-79.352188	LCBO	43.676816	-79.356047	Liquor Store
36	The Danforth West, Riverdale	0.186047	1	43.679557	-79.352188	Kitchen Stuff Plus	43.678613	-79.346422	Furniture / Home Store
36	The Danforth West, Riverdale	0.186047	1	43.679557	-79.352188	Starbucks	43.678879	-79.346357	Coffee Shop
36	The Danforth West, Riverdale	0.186047	1	43.679557	-79.352188	Tsaa Tea Shop	43.677769	-79.351304	Bubble Tea Shop
36	The Danforth West, Riverdale	0.186047	1	43.679557	-79.352188	Leonidas Chocolates Cafe	43.678118	-79.349485	Café
36	The Danforth West, Riverdale	0.186047	1	43.679557	-79.352188	Demetres	43.677683	-79.351608	Dessert Shop
36	The Danforth West, Riverdale	0.186047	1	43.679557	-79.352188	7 Numbers	43.677062	-79.353934	Italian Restaurant
36	The Danforth West, Riverdale	0.186047	1	43.679557	-79.352188	Don Valley Trail	43.676331	-79.353923	Trail
36	The Danforth West, Riverdale	0.186047	1	43.679557	-79.352188	IQ Living	43.678477	-79.347811	Furniture / Home Store
36	The Danforth West, Riverdale	0.186047	1	43.679557	-79.352188	Pantheon	43.677621	-79.351434	Greek Restaurant
36	The Danforth West, Riverdale	0.186047	1	43.679557	-79.352188	MenEssentials	43.677820	-79.351265	Cosmetics Shop
36	The Danforth West, Riverdale	0.186047	1	43.679557	-79.352188	Cafe Fiorentina	43.677743	-79.350115	Italian Restaurant
36	The Danforth West, Riverdale	0.186047	1	43.679557	-79.352188	La Diperie	43.677702	-79.352265	Ice Cream Shop
36	The Danforth West, Riverdale	0.186047	1	43.679557	-79.352188	Dolce Gelato	43.677773	-79.351187	Ice Cream Shop
36	The Danforth West, Riverdale	0.186047	1	43.679557	-79.352188	Moksha Yoga Danforth	43.677622	-79.352116	Yoga Studio
36	The Danforth West, Riverdale	0.186047	1	43.679557	-79.352188	Messini Authentic Gyros	43.677704	-79.350480	Greek Restaurant
36	The Danforth West, Riverdale	0.186047	1	43.679557	-79.352188	Valley Farm Produce	43.677999	-79.349969	Fruit & Vegetable Store
36	The Danforth West, Riverdale	0.186047	1	43.679557	-79.352188	Mezes	43.677962	-79.350196	Greek Restaurant
36	The Danforth West, Riverdale	0.186047	1	43.679557	-79.352188	The Auld Spot Pub	43.677335	-79.353130	Pub
36	The Danforth West, Riverdale	0.186047	1	43.679557	-79.352188	Rikkochez	43.677267	-79.353274	Restaurant
36	The Danforth West, Riverdale	0.186047	1	43.679557	-79.352188	Pizzeria Libretto	43.678489	-79.347576	Pizza Place
36	The Danforth West, Riverdale	0.186047	1	43.679557	-79.352188	The Big Carrot Organic Juice Bar	43.677438	-79.352683	Juice Bar
36	The Danforth West, Riverdale	0.186047	1	43.679557	-79.352188	Christina's On The Danforth	43.678240	-79.349185	Greek Restaurant
36	The Danforth West, Riverdale	0.186047	1	43.679557	-79.352188	Re: Reading	43.678507	-79.347678	Bookstore

Examine Clusters: **Cluster 2**

```
In [71]: #Cluster 2
to_merged.loc[to_merged['Cluster Labels'] == 2]
```

Out[71]:

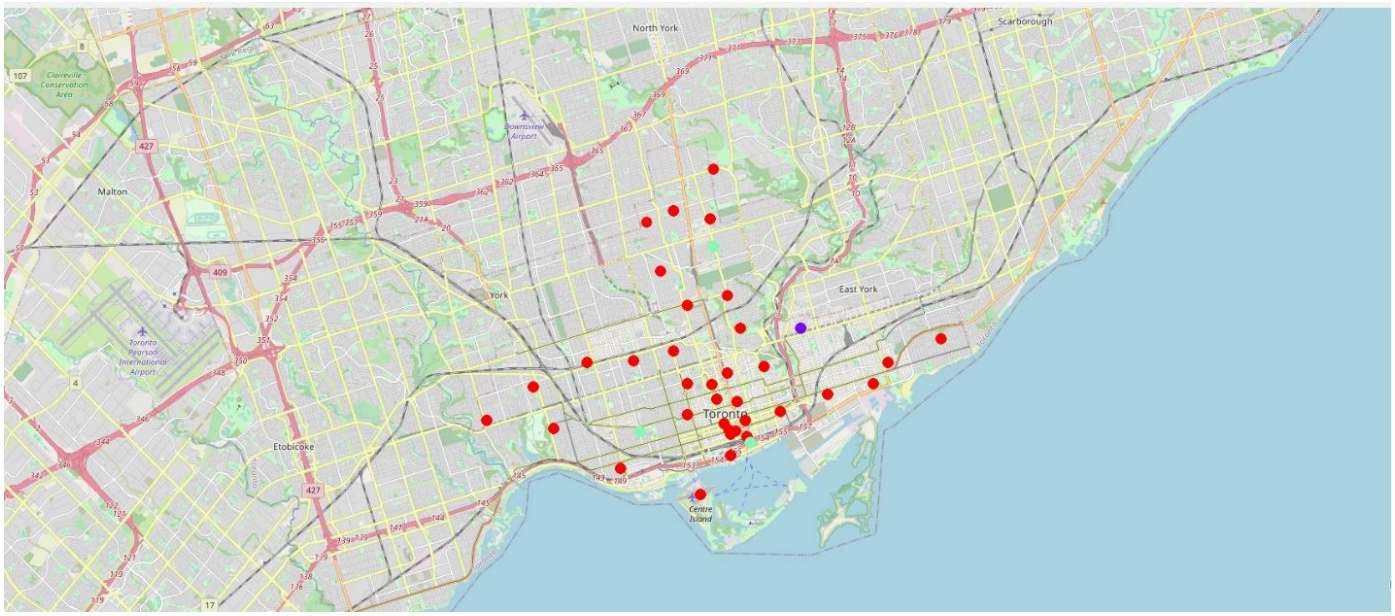
	Neighborhood	Greek Restaurant	Cluster Labels	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Berczy Park	0.017241	2	43.644771	-79.373306	Crepe It Up!	43.648736	-79.371623	Creperie
0	Berczy Park	0.017241	2	43.644771	-79.373306	Alexandro's World Famous Gyros	43.641663	-79.375214	Greek Restaurant
0	Berczy Park	0.017241	2	43.644771	-79.373306	St. Lawrence Market Plaza	43.649169	-79.372330	Art Gallery
0	Berczy Park	0.017241	2	43.644771	-79.373306	Pravda Vodka Bar	43.648516	-79.374732	Cocktail Bar
0	Berczy Park	0.017241	2	43.644771	-79.373306	Olympic Cheese	43.648702	-79.371541	Cheese Shop
...	...	...	...	...	...	...	...	...	...
8	Davisville	0.030303	2	43.704324	-79.388790	Pizza Pizza	43.706138	-79.389292	Pizza Place
8	Davisville	0.030303	2	43.704324	-79.388790	Apple Tree Farmer's Market	43.700326	-79.389760	Farmers Market
8	Davisville	0.030303	2	43.704324	-79.388790	Petro-Canada	43.702269	-79.387955	Gas Station
8	Davisville	0.030303	2	43.704324	-79.388790	Shoppers Drug Mart	43.707806	-79.389893	Pharmacy
0	Berczy Park	0.017241	2	43.644771	-79.373306	The Keg Steakhouse + Bar - Esplanade	43.646712	-79.374768	Restaurant

134 rows × 9 columns

The results from k-means clustering show that we can categorize Toronto neighborhoods into 3 clusters based on how many Greek restaurants are in each neighborhood:

- Cluster 0: Neighborhoods with no Greek restaurants
- Cluster 1: Neighborhoods high number of Greek restaurants
- Cluster 2: Neighborhoods with little or no Greek restaurants

The results are visualized in the above map with Cluster 0 in red color, Cluster 1 in purple color and Cluster 2 in light green color.



Recommendations

Most of the Greek restaurants are concentrated in cluster 1. This represents a great opportunity and high potential areas to open New Greek restaurants in clusters 0 and 2 as there is very little to no competition from existing identical business. Also, the recommendation is to avoid neighborhoods in cluster 1 which already have high concentration of reek restaurants and suffering from intense competition.

Limitations and Suggestions for Future Research

In this project, the only considered factor is the existence of Greek restaurants in each neighborhood. There are many factors that can be taken into consideration such as population density, income of residents, rent that could influence the decision to open a new restaurant. Future research can take into consideration of these factors. In addition, I am relying on the existence of Greek restaurants only for this project but

future research can be expanded to take into consideration other types of ethnic cuisines as well (ie. Italian, Thai, Indian, etc.)

Conclusion

In this project, we have gone through the process of identifying the business problem, specifying the data required, exploratory analysis and feature engineering over the data, performing the machine learning by utilizing k-means clustering and providing recommendation to the stakeholder.