1. Introduction

Toronto is the most populous city in Canada and the fourth most populous city in North America. It is also one of the most diverse cities in the world. Toronto has a very large pool of diverse people and culture, over 160 languages are spoken here. Toronto has a population of around 6.2 million (2020) which is 16.5% of Canada's population.

Toronto is an important city and filled with people. As 'Food' is one of the basic need of people and I love food, I decided to explore the food options of Toronto and to provide business insights to investors who are trying to invest in Restaurants in Toronto.

Primary my target is to analyze the food options and to guide Investor in which region of Toronto Downtown they should invest in an Indian Restaurant. Secondarily I want to guide more open investors who are open to investing in Restaurants of any cuisine by giving them insights of perfect location and cuisine for their business.

As an investor for Restaurants, we want to invest in a place where the cuisine of a restaurant is in the top 5 common venue for food and have diverse food options in that neighbourhood. If we consider all these things, we can create clusters map and information chart of neighbourhoods of Toronto Downtown where all these conditions are met and can decide to open what type of restaurants in which neighbourhood of Toronto Downtown.

2. Data Collection

To solve the stated purpose of the introduction, the data we will be using are given below:
(These all sets of Data are feature set, at last to narrow down to our goal we focused feature data only related to restaurants)

- List of areas of Toronto with its boroughs, postcodes, and neighbourhood from Wikipedia.
- Foursquare API to get the most explored venues, restaurants, and details of given Borough of Toronto based on latitude and longitude
- Longitude and Latitude of Toronto based on their postcodes

3. Methodology – Walk through from Data Preprocessing to Machine Learning

Firstly, the postal code, borough, neighbourhood data is scrapped from Wikipedia by using pandas, html5lib and beautifulsoup4.

	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

Then this data was cleaned, dropped the rows which have Borough value 'Not assigned' and reset the indexing

	Postal Code	Borough	Neighbourhood
0	M3A	North York	Parkwoods
1	M4A	North York	Victoria Village
2	M5A	Downtown Toronto	Regent Park, Harbourfront
3	M6A	North York	Lawrence Manor, Lawrence Heights
4	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government

After that checked if there are any duplicate values of Postal Code, if there are any duplicate values of postal codes, they are merged into one separated by a comma

Then checked if a cell has a borough but a Not assigned neighborhood, then the neighborhood will be the same as the borough

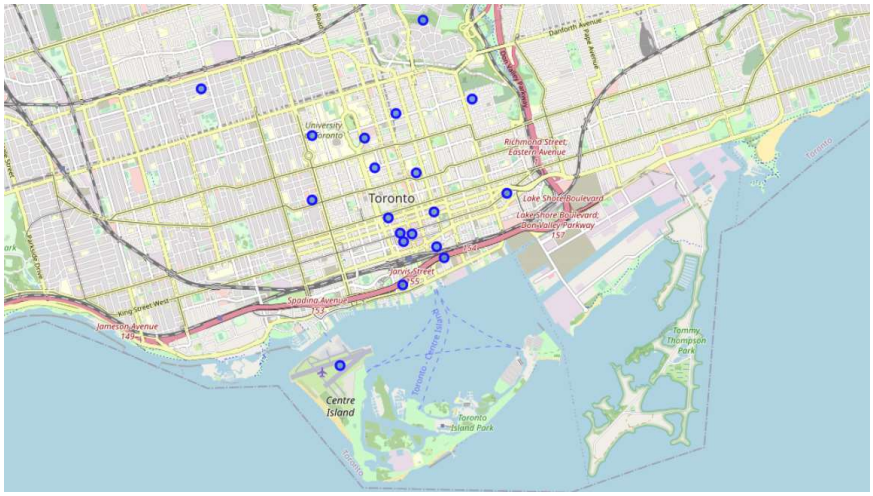
After that Latitude and Longitude values were found through web searching and was coded to merge with our previous data frame, which was cleaned, and the resulting output was like this:

	Postal Code	Borough	Neighbourhood	Latitude	Longitude
0	M3A	North York	Parkwoods	43.753259	-79.329656
1	M4A	North York	Victoria Village	43.725882	-79.315572
2	M5A	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636
3	M6A	North York	Lawrence Manor, Lawrence Heights	43.718518	-79.464763
4	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government	43.662301	-79.389494

As I wanted to explore neighbourhood of Toronto Downtown to find insight of opening new restaurant, so form our data frame I only selected rows which have Borough values ‘Downtown Toronto’.

	Postal Code	Borough	Neighbourhood	Latitude	Longitude
0	M5A	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636
1	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government	43.662301	-79.389494
2	M5B	Downtown Toronto	Garden District, Ryerson	43.657162	-79.378937
3	M5C	Downtown Toronto	St. James Town	43.651494	-79.375418
4	M5E	Downtown Toronto	Berczy Park	43.644771	-79.373306

Then I used folium to create a map of Downtown Toronto and to mark the neighbourhood in our map using blue points.



After done with primary preprocessing of data I used Foursquare API to explore venues of boroughs. I made limit 100 venues and radius to 500 meters for each borough from their longitude and latitude.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Regent Park, Harbourfront	43.65426	-79.360636	Roselle Desserts	43.653447	-79.362017	Bakery
1	Regent Park, Harbourfront	43.65426	-79.360636	Tandem Coffee	43.653559	-79.361809	Coffee Shop
2	Regent Park, Harbourfront	43.65426	-79.360636	Cooper Koo Family YMCA	43.653249	-79.358008	Distribution Center
3	Regent Park, Harbourfront	43.65426	-79.360636	Body Blitz Spa East	43.654735	-79.359874	Spa
4	Regent Park, Harbourfront	43.65426	-79.360636	Impact Kitchen	43.656369	-79.356980	Restaurant

Then I repeated the process each all location and grouped the data by their neighbourhood.

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Berczy Park	58	58	58	58	58	58
CN Tower, King and Spadina, Railway Lands, Harbourfront West, Bathurst Quay, South Niagara, Island airport	16	16	16	16	16	16
Central Bay Street	64	64	64	64	64	64
Christie	16	16	16	16	16	16
Church and Wellesley	77	77	77	77	77	77
Commerce Court, Victoria Hotel	100	100	100	100	100	100
First Canadian Place, Underground city	100	100	100	100	100	100
Garden District, Ryerson	100	100	100	100	100	100
Harbourfront East, Union Station, Toronto Islands	100	100	100	100	100	100
Kensington Market, Chinatown, Grange Park	65	65	65	65	65	65
Queen's Park, Ontario Provincial Government	34	34	34	34	34	34
Regent Park, Harbourfront	45	45	45	45	45	45
Richmond, Adelaide, King	100	100	100	100	100	100
Rosedale	4	4	4	4	4	4
St. James Town	87	87	87	87	87	87
St. James Town, Cabbagetown	45	45	45	45	45	45
Stn A PO Boxes	95	95	95	95	95	95
Toronto Dominion Centre, Design Exchange	100	100	100	100	100	100
University of Toronto, Harbord	35	35	35	35	35	35

Then I did hot coding, and I processed the data frame again to keep the columns that have venues and contains the word ‘Restaurant’ on it. And the resulting df was something like this:

	Neighborhood	Afghan Restaurant	American Restaurant	Asian Restaurant	Belgian Restaurant	Brazilian Restaurant	Caribbean Restaurant	Chinese Restaurant	Colombian Restaurant	Comfort Food Restaurant	Doner Restaurant	Dumpling Restaurant	Eastern European Restaurant	Ethiopian Restaurant	Falafel Restaurant	Fast Food Restaurant	Filipino Restaurant	French Restaurant	German Restaurant	Glute fr Restaura
0	Regent Park, Harbourfront	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
1	Regent Park, Harbourfront	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
2	Regent Park, Harbourfront	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
3	Regent Park, Harbourfront	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
4	Regent Park, Harbourfront	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
5	Regent Park, Harbourfront	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
6	Regent Park,	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	

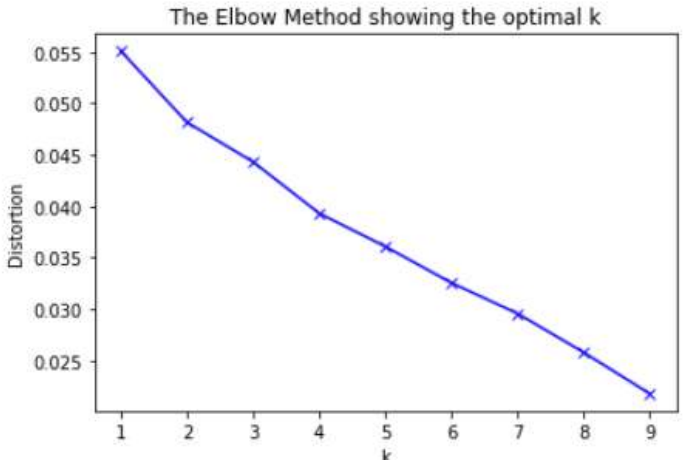
After that I did more processing to the data like grouping by neighbourhood, shuffling columns, adding columns, calculations etc. to find the top 10 visited/popular restaurants in the neighbourhood of Toronto Downtown which was grouped by neighbourhood.

```
----Central Bay Street----
                          venue  freq
0  Italian Restaurant              0.05
1  Japanese Restaurant             0.03
2  Vegetarian / Vegan Restaurant   0.02
3  Portuguese Restaurant           0.02
4  Indian Restaurant               0.02
5  Middle Eastern Restaurant       0.02
6  French Restaurant               0.02
7  Modern European Restaurant      0.02
8  Falafel Restaurant              0.02
9  Korean Restaurant               0.02


----Christie----
                          venue  freq
0  Italian Restaurant              0.06
1  Restaurant                      0.06
```

	Neighborhood	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	9th Most Common Venue	10th Most Common Venue
0	Berczy Park	Seafood Restaurant	Restaurant	Eastern European Restaurant	Thai Restaurant	Greek Restaurant	Sushi Restaurant	Vegetarian / Vegan Restaurant	Italian Restaurant	Japanese Restaurant	French Restaurant
1	CN Tower, King and Spadina, Railway Lands, Harbourfront West, Bathurst Quay, South Niagara, Island airport	Vietnamese Restaurant	Dumpling Restaurant	Gluten-free Restaurant	German Restaurant	French Restaurant	Filipino Restaurant	Fast Food Restaurant	Falafel Restaurant	Ethiopian Restaurant	Eastern European Restaurant
2	Central Bay Street	Italian Restaurant	Japanese Restaurant	Indian Restaurant	Portuguese Restaurant	French Restaurant	Vegetarian / Vegan Restaurant	Korean Restaurant	Middle Eastern Restaurant	New American Restaurant	Modern European Restaurant
3	Christie	Italian Restaurant	Restaurant	Vietnamese Restaurant	Dumpling Restaurant	German Restaurant	French Restaurant	Filipino Restaurant	Fast Food Restaurant	Falafel Restaurant	Ethiopian Restaurant
4	Church and Wellesley	Japanese Restaurant	Sushi Restaurant	Restaurant	Mediterranean Restaurant	Indian Restaurant	Mexican Restaurant	American Restaurant	Caribbean Restaurant	Ethiopian Restaurant	Fast Food Restaurant
5	Commerce Court, Victoria Hotel	Restaurant	American Restaurant	Seafood Restaurant	Italian Restaurant	Japanese Restaurant	Asian Restaurant	Vegetarian / Vegan Restaurant	Thai Restaurant	Gluten-free Restaurant	New American Restaurant
6	First Canadian Place, Underground city	Japanese Restaurant	Restaurant	American Restaurant	Asian Restaurant	Seafood Restaurant	Sushi Restaurant	Colombian Restaurant	Greek Restaurant	Italian Restaurant	Fast Food Restaurant

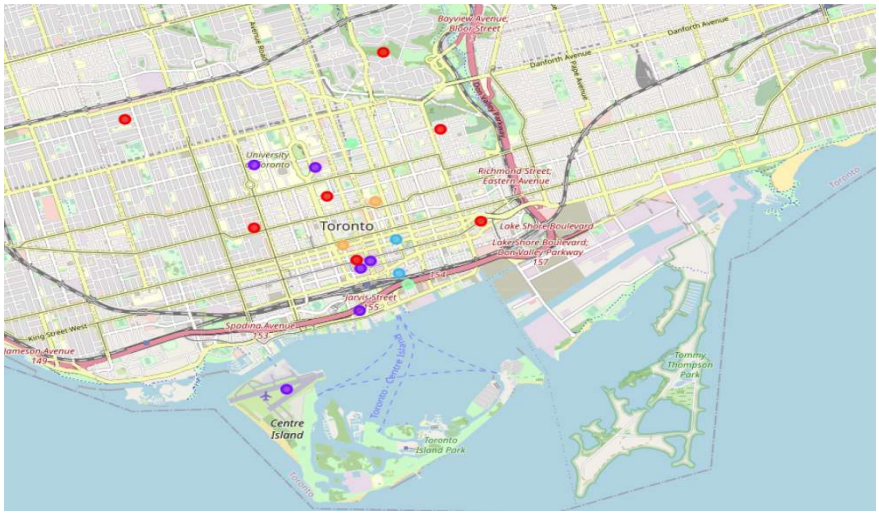
As we have some common restaurant trends in neighbourhoods, so I used unsupervised learning K-means algorithm to cluster the neighbourhoods. K-Means algorithm is common cluster method of unsupervised learning. After analyzing elbow method, it seems that K=5 works best for the clustering in this scenario.



After that I merged the latest cleaned data frame with the cluster labels for each neighbourhood.

	Cluster Labels	Neighborhood	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	9th Most Common Venue	10th Most Common Venue
0	0	Berczy Park	Seafood Restaurant	Restaurant	Eastern European Restaurant	Thai Restaurant	Greek Restaurant	Sushi Restaurant	Vegetarian / Vegan Restaurant	Italian Restaurant	Japanese Restaurant	French Restaurant
1	1	CN Tower, King and Spadina, Railway Lands, Harbourfront West, Bathurst Quay, South Niagara, Island airport	Vietnamese Restaurant	Dumpling Restaurant	Gluten-free Restaurant	German Restaurant	French Restaurant	Filipino Restaurant	Fast Food Restaurant	Falafel Restaurant	Ethiopian Restaurant	Eastern European Restaurant
2	4	Central Bay Street	Italian Restaurant	Japanese Restaurant	Indian Restaurant	Portuguese Restaurant	French Restaurant	Vegetarian / Vegan Restaurant	Korean Restaurant	Middle Eastern Restaurant	New American Restaurant	Modern European Restaurant
3	2	Christie	Italian Restaurant	Restaurant	Vietnamese Restaurant	Dumpling Restaurant	German Restaurant	French Restaurant	Filipino Restaurant	Fast Food Restaurant	Falafel Restaurant	Ethiopian Restaurant
4	3	Church and Wellesley	Japanese Restaurant	Sushi Restaurant	Restaurant	Mediterranean Restaurant	Indian Restaurant	Mexican Restaurant	American Restaurant	Caribbean Restaurant	Ethiopian Restaurant	Fast Food Restaurant

Here is the visualization of the 5 clusters (different colors) in the map



At last I examined and studied each cluster to get to our prediction and answer to the problem which I will be discussing below.

4. Results & Discussion

By looking at the clusters and studying them it can be said that for opening Indian Restaurant we must choose as Indian restaurant in this clusters and neighbourhood is in top 5 popular places:

i. St. James Town, Cabbage town of cluster label 2
ii. Church and Wellesley of cluster label 3
iii. Central Bay Street of cluster label 4

But if we study more in depth the clusters and if we use other external data, we can conclude more places for Indian restaurants based on similarity of cuisines. For example, the neighbourhood and cluster where Mexican/Thai/Asian (not Chinese) Restaurant is in top 5 visiting food place Indian Restaurant might also be hit there as similar spices.

Again, we can get a lot more insights from this cluster based on our requirements, for example if our profit target is not maximum, we can give Indian restaurants to neighbourhoods/clusters where Indian Restaurant is in top 10 visiting food place. So, we can use this data various way according to our requirements. Over here we made our scope small so that it does not get complex. The Data snapshots of clusters are:

	Cluster Labels	Neighborhood	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	9th Most Common Venue	10th Most Common Venue
0	0	Berczy Park	Seafood Restaurant	Restaurant	Eastern European Restaurant	Thai Restaurant	Greek Restaurant	Sushi Restaurant	Vegetarian / Vegan Restaurant	Italian Restaurant	Japanese Restaurant	French Restaurant
5	0	Commerce Court, Victoria Hotel	Restaurant	American Restaurant	Seafood Restaurant	Italian Restaurant	Japanese Restaurant	Asian Restaurant	Vegetarian / Vegan Restaurant	Thai Restaurant	Gluten-free Restaurant	New American Restaurant
6	0	First Canadian Place, Underground city	Japanese Restaurant	Restaurant	American Restaurant	Asian Restaurant	Seafood Restaurant	Sushi Restaurant	Colombian Restaurant	Greek Restaurant	Italian Restaurant	Fast Food Restaurant
12	0	Richmond, Adelaide, King	Restaurant	Thai Restaurant	American Restaurant	Sushi Restaurant	Vegetarian / Vegan Restaurant	Modern European Restaurant	Gluten-free Restaurant	Latin American Restaurant	Mediterranean Restaurant	Colombian Restaurant
14	0	St. James Town	Restaurant	American Restaurant	Moroccan Restaurant	Italian Restaurant	Seafood Restaurant	Comfort Food Restaurant	German Restaurant	Vegetarian / Vegan Restaurant	Japanese Restaurant	Middle Eastern Restaurant
16	0	Stn A PO Boxes	Italian Restaurant	Seafood Restaurant	Restaurant	Japanese Restaurant	Molecular Gastronomy Restaurant	American Restaurant	Thai Restaurant	Sushi Restaurant	Vegetarian / Vegan Restaurant	French Restaurant
17	0	Toronto Dominion Centre, Design Exchange	Restaurant	American Restaurant	Seafood Restaurant	Italian Restaurant	Japanese Restaurant	Asian Restaurant	Sushi Restaurant	French Restaurant	Fast Food Restaurant	Gluten-free Restaurant

	Cluster Labels	Neighborhood	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	9th Most Common Venue	10th Most Common Venue
1	1	CN Tower, King and Spadina, Railway Lands, Harbourfront West, Bathurst Quay, South Niagara, Island airport	Vietnamese Restaurant	Dumpling Restaurant	Gluten-free Restaurant	German Restaurant	French Restaurant	Filipino Restaurant	Fast Food Restaurant	Falafel Restaurant	Ethiopian Restaurant	Eastern European Restaurant
8	1	Harbourfront East, Union Station, Toronto Islands	Restaurant	Italian Restaurant	Indian Restaurant	Vegetarian / Vegan Restaurant	Sushi Restaurant	Seafood Restaurant	Japanese Restaurant	New American Restaurant	Chinese Restaurant	Dumpling Restaurant
9	1	Kensington Market, Chinatown, Grange Park	Vegetarian / Vegan Restaurant	Mexican Restaurant	Vietnamese Restaurant	Dumpling Restaurant	Caribbean Restaurant	Japanese Restaurant	Comfort Food Restaurant	Doner Restaurant	Filipino Restaurant	Belgian Restaurant
10	1	Queen's Park, Ontario Provincial Government	Japanese Restaurant	Sushi Restaurant	Portuguese Restaurant	Mexican Restaurant	Vietnamese Restaurant	Dumpling Restaurant	French Restaurant	Filipino Restaurant	Fast Food Restaurant	Falafel Restaurant
11	1	Regent Park, Harbourfront	French Restaurant	Restaurant	Mexican Restaurant	Vietnamese Restaurant	Dumpling Restaurant	German Restaurant	Filipino Restaurant	Fast Food Restaurant	Falafel Restaurant	Ethiopian Restaurant
13	1	Rosedale	Vietnamese Restaurant	Dumpling Restaurant	Gluten-free Restaurant	German Restaurant	French Restaurant	Filipino Restaurant	Fast Food Restaurant	Falafel Restaurant	Ethiopian Restaurant	Eastern European Restaurant

```
eighborhoods_venues_sorted[neighborhoods_venues_sorted['Cluster Labels'] == 2]
```

```
neighborhoods_venues_sorted[neighborhoods_venues_sorted['Cluster Labels'] == 2]
```

4]:

	Cluster Labels	Neighborhood	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	9th Most Common Venue	10th Most Common Venue
3	2	Christie	Italian Restaurant	Restaurant	Vietnamese Restaurant	Dumpling Restaurant	German Restaurant	French Restaurant	Filipino Restaurant	Fast Food Restaurant	Falafel Restaurant	Ethiopian Restaurant
15	2	St. James Town, Cabbagetown	Restaurant	Italian Restaurant	Chinese Restaurant	Indian Restaurant	Thai Restaurant	Taiwanese Restaurant	Japanese Restaurant	Caribbean Restaurant	Dumpling Restaurant	Filipino Restaurant

	Cluster Labels	Neighborhood	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	9th Most Common Venue	10th Most Common Venue
4	3	Church and Wellesley	Japanese Restaurant	Sushi Restaurant	Restaurant	Mediterranean Restaurant	Indian Restaurant	Mexican Restaurant	American Restaurant	Caribbean Restaurant	Ethiopian Restaurant	Fast Food Restaurant
18	3	University of Toronto, Harbord	Japanese Restaurant	Restaurant	Comfort Food Restaurant	French Restaurant	Italian Restaurant	Sushi Restaurant	Chinese Restaurant	Dumpling Restaurant	Filipino Restaurant	Fast Food Restaurant

	Cluster Labels	Neighborhood	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	9th Most Common Venue	10th Most Common Venue
2	4	Central Bay Street	Italian Restaurant	Japanese Restaurant	Indian Restaurant	Portuguese Restaurant	French Restaurant	Vegetarian / Vegan Restaurant	Korean Restaurant	Middle Eastern Restaurant	New American Restaurant	Modern European Restaurant
7	4	Garden District, Ryerson	Italian Restaurant	Japanese Restaurant	Fast Food Restaurant	Ramen Restaurant	Middle Eastern Restaurant	Ethiopian Restaurant	Chinese Restaurant	New American Restaurant	Vietnamese Restaurant	Modern European Restaurant

5. Conclusion

As people are getting busy and to socialise also, they are visiting restaurants to get rid of hunger and fulfil their cravings, so restaurant business is getting huge. Investors can do similar analysis to see the trend and to find the perfect cuisine or spot for their restaurants so they can maximize returns fast.