Recommendation System- Starting a New Restaurant Business

Abhinay Dornipati

I. Introduction/Business Problem

A. Description & Discussion of the Background

Toronto is the provincial capital of Ontario and the most populous city in Canada, with an estimated population of 2,956,024 (2018) and an estimated population of 6,341,935 in the Toronto Region (2018.) Located on the shores of the western end of Lake Ontario, Toronto is an international center of business, finance, arts, and culture. Its large population of immigrants from around the globe has also made Toronto one of the most multicultural and cosmopolitan cities in the world.

Toronto is one of Canada's leading tourism destinations. In 2017, the Toronto-area received 43.7 million tourists, of which 10.4 million were domestic visitors and 2.97 million were from the United States, spending a total of $8.84 billion. Toronto has an array of tourist attractions, and a rich cultural life.

Restaurant sales in Canada have grown over the past five years from 61 billion Canadian dollars in 2010 to 85 billion Canadian dollars in 2017. That accumulates to an increase of approximately 24 billion Canadian dollars in the Canadian restaurant industry. The same trend holds good for the city of Toronto.

Chicago-based foodservice database, marketing, and analytics firm, CHD Expert has new data indicating that independent restaurant operators are making an impact in Toronto's foodservice market landscape. The new data reveals that 69.3 percent of restaurants in Toronto are independent restaurants, whereas only 63.4 percent of nationwide restaurants are considered independent. The focus was to explore the Canadian restaurant landscape, specifically focusing on Toronto's independent restaurant operator growth, and popular menu types. Toronto's landscape is distinctly different than the rest of the country. There are more European menu type restaurants in Toronto than in the rest of Canada, making up 5.3 percent of the Toronto landscape, compared to 3.2 nationwide. Data reveals that 9.6 percent of restaurants in Toronto are Asian, whereas Asian restaurants only account for 6.6 percent nationwide. One of the USA's favorite menu type 'hamburger,' does not have as strong of a presence in Canada and is even less predominant is Toronto. As a nation, approximately 6.1 percent of Canadian restaurants serve a 'hamburger' menu type.

Considering Toronto's support for the independent restaurant operators, it is quite evident that starting a restaurant business would earn you more money in comparison to other businesses. As rightly said, "Nothing Comes For Free", with the profitable business comes also the high competition. The recommendation system developed here, serve as one of the guides to start a new restaurant business serving a specific type of cuisine like Chinese, American, Indian etc. Also the companies can get benefitted by knowing the best location to start their business, taking into account the competition in that locality.

B. Problem Description

Restaurant business involves preparing the food and drinks for the customers in return for the money. The kind of cuisine each restaurant serves depends on many factors like; location in

which it is present, the expertise of the chefs, preferences of the customers etc. Similar is the case with the restaurants in the city of Toronto. Because of the diversity of population living in Toronto, lot of tourists coming in from various parts of the world, there are various kinds of restaurants present. This makes Toronto famous for its wide range of excellent cuisine. Therefore, it is quite evident that, survival in such a competitive market demands a very good strategy from the new companies who are willing to start a restaurant business in the region of Toronto. Accordingly, various factors have to be investigated to decide upon the location and the type cuisine to be served in the restaurant. Few of them are;

1. Toronto Population and demographics
2. Competitors in that locality
3. The cuisine served by the competitors
4. Study about the popular venues around that locality – for example, Tourist places, Parks, Theaters etc.
5. Study of the locations or neighborhoods based on the Boroughs in which they are present.

And the list goes on….

Choice of an optimal location to start the business is and first and most important step even for a well-funded companies like ABCs. If this strategy turns out to be successful for them, they can use the same strategy for their future endeavors.

Target Audience:

For instance, to recommend an optimal location, ABC Company Ltd has appointed me to lead the Data Science Team. The goal is to locate and recommend which neighborhood of Toronto would be the best choice to start a restaurant serving a specific type of cuisine. Of course, all the factors have to be considered while making a recommendation, this would also increase the confidence of the management.

Success Criteria:

A good recommendation of the locality or neighborhood for the restaurant to ABC Company Ltd would be the success criteria for this project. An ideal recommendation would be the one that is made with the consideration of competition, lack of such restaurant in that location.

II. DATA

For building the recommendation model, the following data and information is considered for analysis

1. Scrapped Wikipedia using BeautifulSoup, to extract information about 10 Toronto boroughs, also known as local authority districts. Also, the local areas or neighborhoods for each borough are considered for detailed analysis.
2. Foursquare API is used to get information about the available restaurants in a given neighborhood and given borough. The API also provides information about the category of restaurants based on the cuisine each restaurant is serving.
3. The location data or address in the form of latitude and longitude was present for each neighborhood in the form of a .csv file.

III. METHODOLOGY

Business Understanding:

The main objective of this project is to find an optimal location for a new restaurant business in Toronto for ABC Company Ltd based on cuisine and competition in that locality.

A. EXPLORATORY DATA ANALYSIS

Toronto is mainly divided into 10 boroughs also known as local authority districts with 103 neighborhoods. Using Folium library, map of Toronto along with its neighborhoods is created for a better visualization.
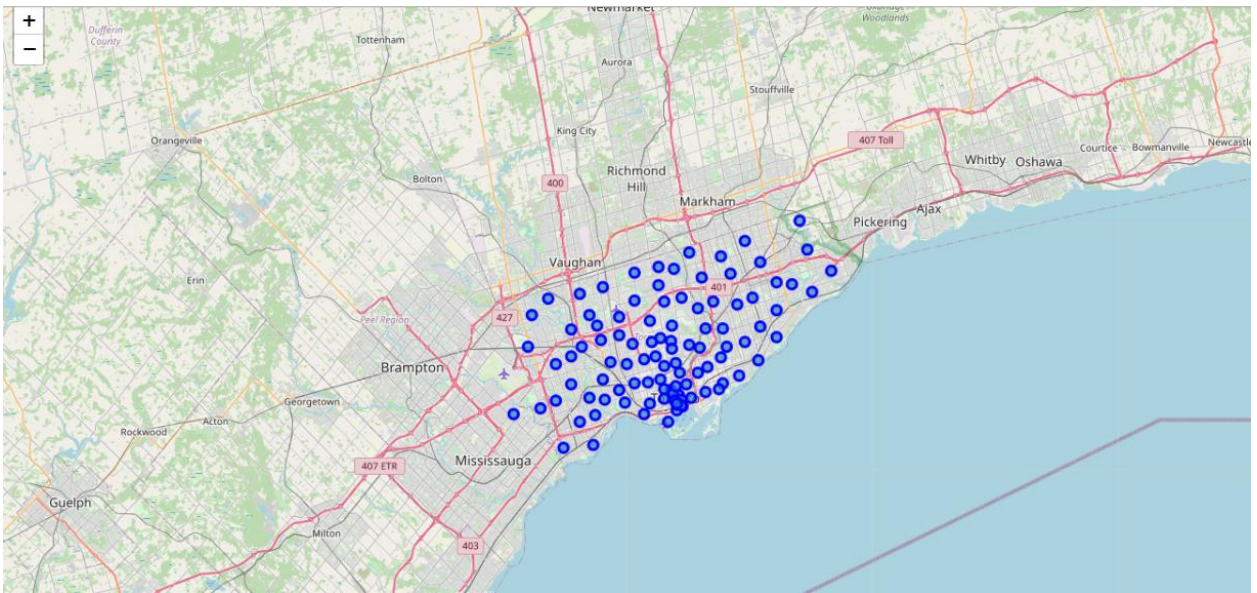
All the blue markers on the map above represents neighborhoods present in the region of Toronto. It can be clearly observed that the city is more congested near the coastal region and widespread as we move away from the coast.

B. PROBLEM APPROACH AND EMPLOYMENT OF K-MEANS CLUSTERING

The problem starts with the scrapping of Wikipedia page containing information about the Toronto boroughs and neighborhoods, using the BeautifulSoup library. The scrapped data is written out into a .csv file, which is later read into a pandas data frame using read_csv method.

Each data point in the data frame consists of information about the neighborhood and its corresponding borough.

In order to plot these neighborhoods on a map using Folium library, we need the location information about them in the form of latitude and longitude. This information about the neighborhoods was readily available in the form of a .csv file. It is later read into a data frame and

joined with the data frame of neighborhoods. The resultant data frame therefore contains all the information to plot the neighborhoods on a map using Folium.

Now, Foursquare API is used to explore the neighborhoods and segment them. The information about the each neighborhood and its corresponding location information is given to explore the venues present around them. The Foursquare API returns 2124 venues with 268 unique categories.

Since, this project deals only with restaurants, all the restaurants are extracted from the venue category list. The result is a data frame which contains 47 unique categories or cuisine serving restaurants in Toronto. Then, one hot encoding is performed on the data frame for each neighborhood. This results shows that there are 499 unique restaurants in Toronto serving 49 different styles of cuisines.

Then the rows in the resulting data frame are grouped by "neighborhood" to determine the frequency of occurrence of each restaurant in a particular neighborhood. Next, a new data frame is created, where each row contains a neighborhood and its corresponding top ten common restaurants based on the cuisine served.

Ultimately, k-means clustering is performed on the data frame to check the pattern for each neighborhood and obtain the information about top ten common restaurants in each neighborhood. Before fitting the model using the data frame, the best value of 'k', i.e, the number of clusters is determined using silhouette_score. For this silhouette_score has to be imported from sklearn.metrics library. A graph has been plotted between the silhouette_score and different values of k. From the graph below it can be observed that, for a value of k=3, the silhouette_score is maximum. Accordingly, the value of k is chosen to be 3 and then the data frame (model) is fit using k-means clustering algorithm.
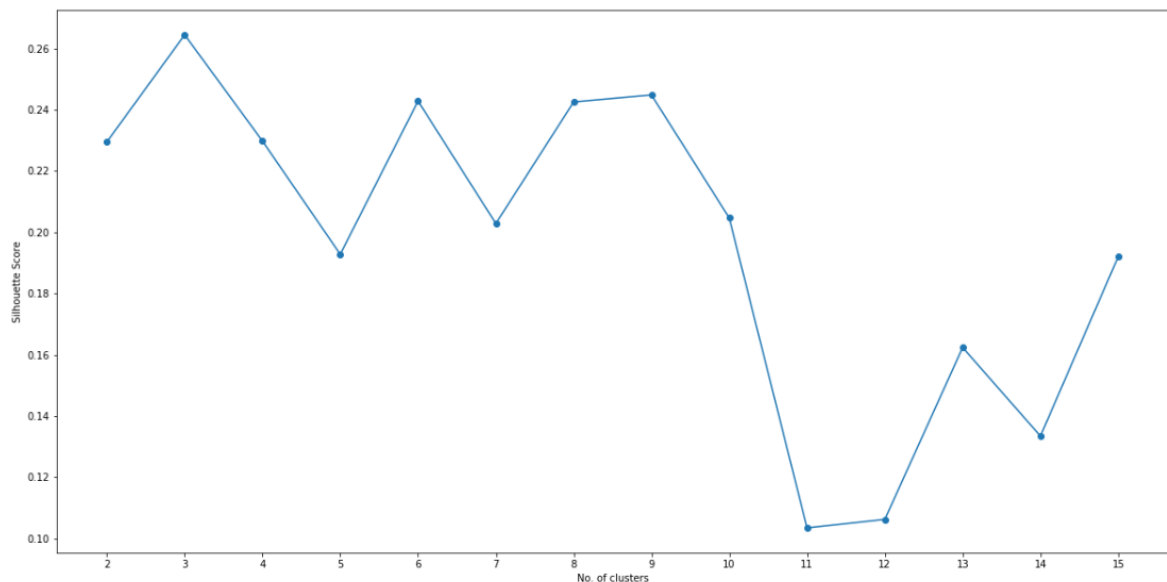
Each neighborhood in the region of Toronto is assigned to one of the clusters based on the top ten restaurants serving same cuisines in the respective neighborhood.

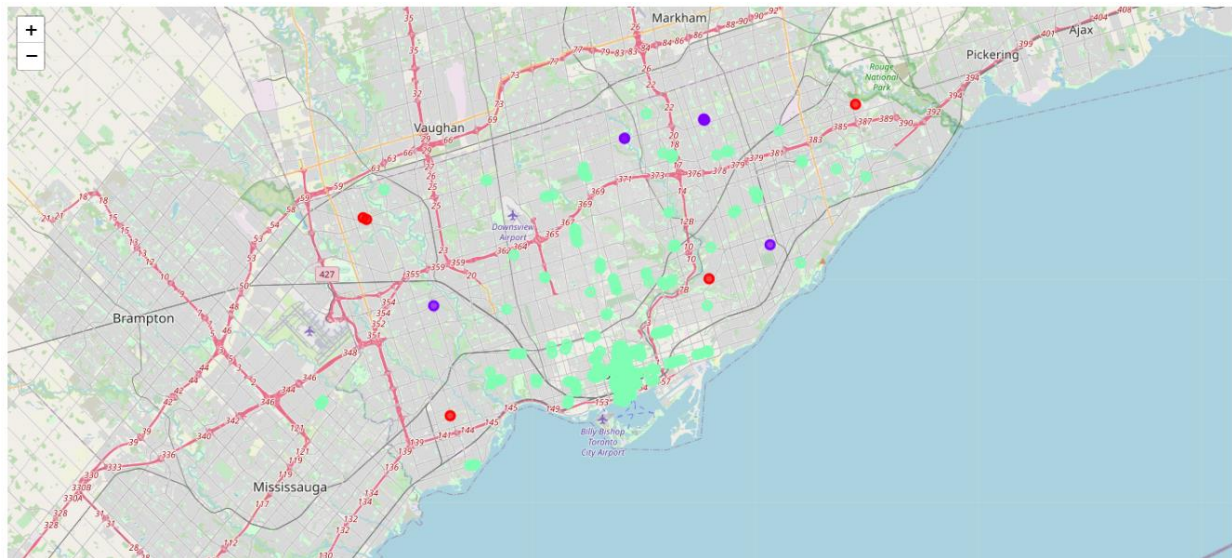|	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
5	Victoria Village	43.725882	-79.315572	Portugril	43.725819	-79.312785	Portuguese Restaurant	0	Portuguese Restaurant	Vietnamese Restaurant	Doner Restaurant	German Restaurant	French Restaurant	Filipino Restaurant	Fast Food Restaurant	Falafel Restaurant	Ethiopian Restaurant	Empanada Restaurant
13	Regent Park / Harbourfront	43.654260	-79.360636	Impact Kitchen	43.656369	-79.356980	Restaurant	0	Restaurant	Asian Restaurant	French Restaurant	Mexican Restaurant	Vietnamese Restaurant	Dumpling Restaurant	German Restaurant	Filipino Restaurant	Fast Food Restaurant	Falafel Restaurant
29	Regent Park / Harbourfront	43.654260	-79.360636	Cluny Bistro & Boulangerie	43.650565	-79.357843	French Restaurant	0	Restaurant	Asian Restaurant	French Restaurant	Mexican Restaurant	Vietnamese Restaurant	Dumpling Restaurant	German Restaurant	Filipino Restaurant	Fast Food Restaurant	Falafel Restaurant
33	Regent Park / Harbourfront	43.654260	-79.360636	Cocina Economica	43.654959	-79.365657	Mexican Restaurant	0	Restaurant	Asian Restaurant	French Restaurant	Mexican Restaurant	Vietnamese Restaurant	Dumpling Restaurant	German Restaurant	Filipino Restaurant	Fast Food Restaurant	Falafel Restaurant
41	Regent Park / Harbourfront	43.654260	-79.360636	Izumi	43.649970	-79.360153	Asian Restaurant	0	Restaurant	Asian Restaurant	French Restaurant	Mexican Restaurant	Vietnamese Restaurant	Dumpling Restaurant	German Restaurant	Filipino Restaurant	Fast Food Restaurant	Falafel Restaurant

IV. ANALYSIS

Neighborhood K-Means clustering based on the mean occurrence of venue category:

All the 3 clusters have unique pattern with regards to the top ten restaurants for a particular neighborhood. The picture below presents the number of neighborhoods assigned to each cluster.

```
2    486
1      8
0      5
```

Cluster 2 stands on the top with highest number of neighborhoods of 486. Whereas, the clusters 1 and 0 have only 8 and 5 number of neighborhoods respectively.

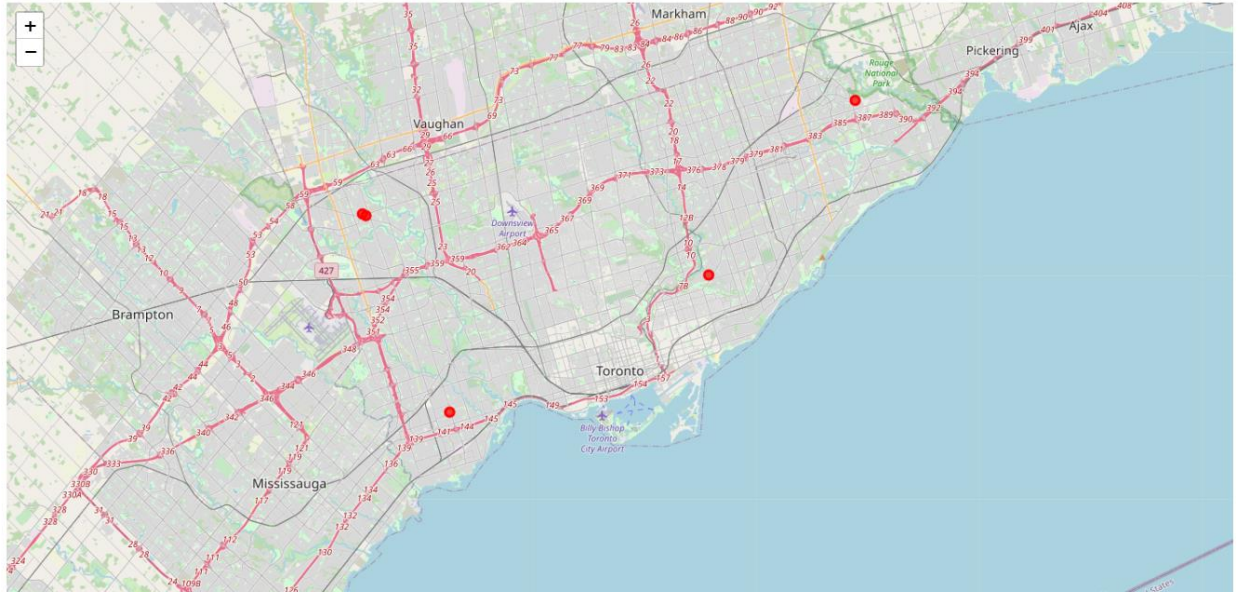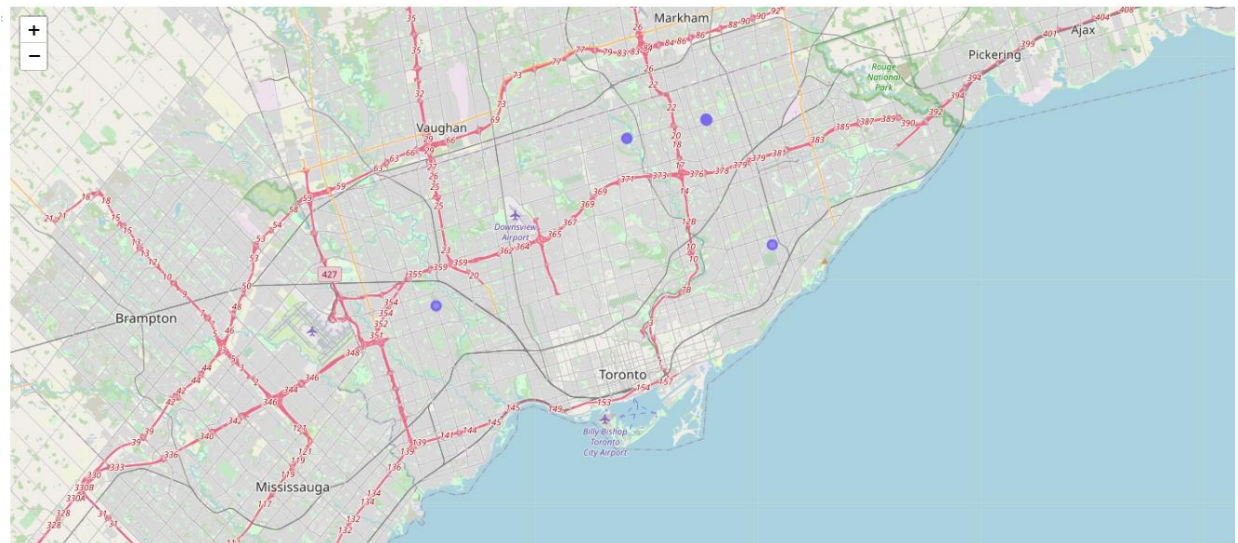The figure below shows the map with clustered neighborhoods:

IV. RESULTS & DISCUSSION

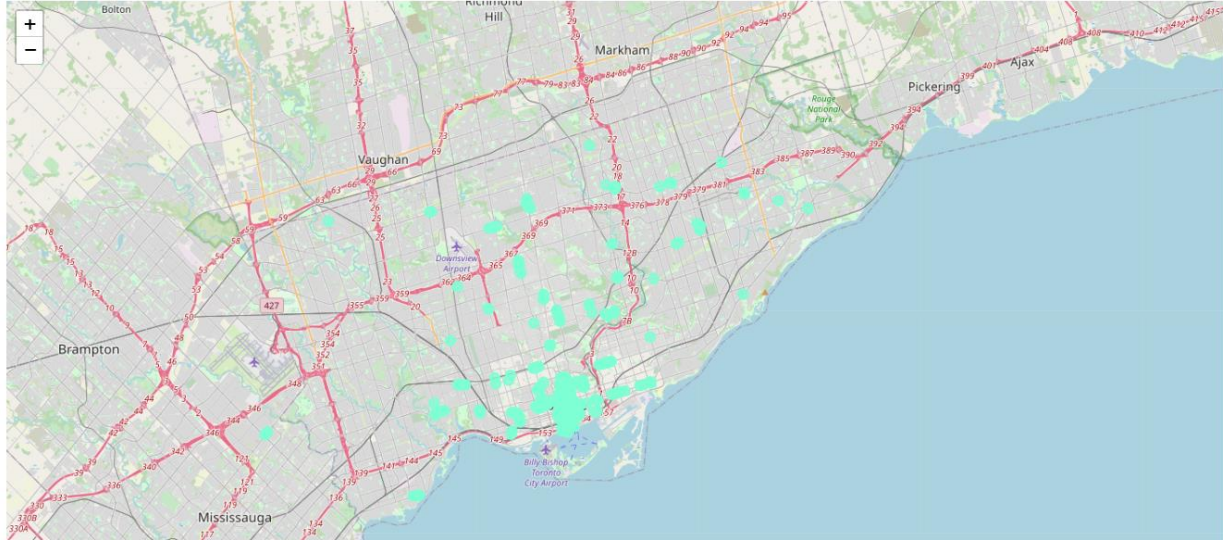Analysis: Each cluster of the neighborhoods in Toronto

Cluster 1:

As already mentioned there only 5 neighborhoods in this cluster.

Cluster 2:

Cluster 2 contains 8 neighborhoods in it.

Cluster 3:

Cluster 3 is the most populous one with 486 number of neighborhoods in it.

Now, examining each cluster and determining the discriminating restaurants that distinguish each cluster, as shown in the figure below.

	Cluster	Most Common Restaurant	Least Common Restaurant	2nd Least Common Restaurant
0	1	Fast Food	Eastern European Restaurant	Empanda Restaurant
1	2	Chinese Restaurant	Empanda Restaurant	Ehiopian Restaurant
2	3	Italian Restaurant	Gulten-free Restaurant	Japanese Restaurant

The results can be analyzed in two ways:

1. If ABC Company Ltd wants to open a restaurant in an already decided location irrespective of the cuisine, then it's good for them to serve a cuisine which is not being served by any of the restaurants in that location. This would bring good profits to the company.

2. In contrast, if the ABC Company Ltd wants to open a restaurant serving a specific cuisine irrespective of the location, then it's preferred to open the restaurant in a location where this kind of cuisine is not served or at least served by least number of restaurants.

V. CONCLUSION

This model is built on limited amount of data available. Therefore, the recommendation made might not be totally accurate. With the availability of more amount of data, this model can make better recommendations. It has been primarily observed that, even though Toronto has large number of restaurants, they are all present only in certain neighborhoods. Few neighborhoods have very less number of restaurants serving only few kinds of cuisine. Therefore, there is a lot of scope to start new restaurants in such locations, serving the cuisines not available till date.

For example, every neighborhood can be checked for number of restaurants available and the type of cuisine being served. Accordingly, a location with lowest competition and thereby lowest risk can be chosen to start a new restaurant business.