Predicting best neighborhood for a new restaurant

Table of contents:

1.Introduction Section:

In this section, we would define the scope of this project and the core idea of this project. Primary question for any business owner would be selecting the location for their operations. Here I would consider a restaurant owner who is looking for expanding in the city of Toronto. The investors/owners would need certain proofs to consider a particular neighborhood in Toronto for their new venture which would result in profits. One can understand the neighborhoods by exploring the venues around and can get an idea on the type of locality.

Business problem:

Challenge is to find a suitable location in Toronto for a new restaurant along with recommendations regarding the theme and cuisine that would comply with the demands of the restaurant owner.

Target audience:

Target audience for this project would be

- Business executives for exploring options in and around a particular location
- Investors/ potential owners of restaurants, coffee shops, gyms
- Real estate dealers
- Data science enthusiasts

2.Data Section:

Data that is required to solve above business problem:

- List of neighborhoods in Toronto
- Geospatial data regarding the neighborhoods

- Venues around neighborhood

Data sources:

- "List of Postal code of Canada: M" https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M) wiki page to get all the information about the neighborhoods present in Toronto
- "https://cocl.us/Geospatial_data" csv file and geocoder to get geospatial data of neighborhoods
- Foursquare API calls for getting information regarding venues in Toronto

Data cleaning/preprocessing:

1. Data from wiki regarding the neighborhoods is extracted into data frame which include neighborhood name, postal code and borough.
2. Only cells with assigned borough are processed ignoring others
3. Neighborhoods with unassigned boroughs are assigned with the same name as borough.
4. Latitude and longitude data is obtained from CSV file "https://cocl.us/Geospatial_data" and added to the data frame to complete the data frame with latitude and longitude details
5. Foursquare API call is utilized to obtain venues around each neighborhood and the data is merged into a dataframe along with venue categories

Data:

	Postal code	Borough	Neighborhood
0	M1A	Not assigned	NaN
1	M2A	Not assigned	NaN
2	M3A	North York	Parkwoods
3	M4A	North York	Victoria Village
4	M5A	Downtown Toronto	Regent Park / Harbourfront

	Postal Code	Latitude	Longitude
0	M1B	43.806686	-79.194353
1	M1C	43.784535	-79.160497
2	M1E	43.763573	-79.188711

index	Postalcode	Borough	Neighborhood	Latitude	Longitude
2	M3A	North York	Parkwoods	43.753259	-79.329656
3	M4A	North York	Victoria Village	43.725882	-79.315572
4	M5A	Downtown Toronto	Regent Park , Harbourfront	43.654260	-79.360636
5	M6A	North York	Lawrence Manor , Lawrence Heights	43.718518	-79.464763

	name	categories	lat	lng
0	Toronto Islands	Park	43.622112	-79.378495
1	Porter Lounge	Airport Lounge	43.630680	-79.395756
2	Bestival Toronto	Music Venue	43.623140	-79.391107
3	Pizza Pizza	Pizza Place	43.622237	-79.378935
4	Toronto Island BBQ & Beer	BBQ Joint	43.623440	-79.378381

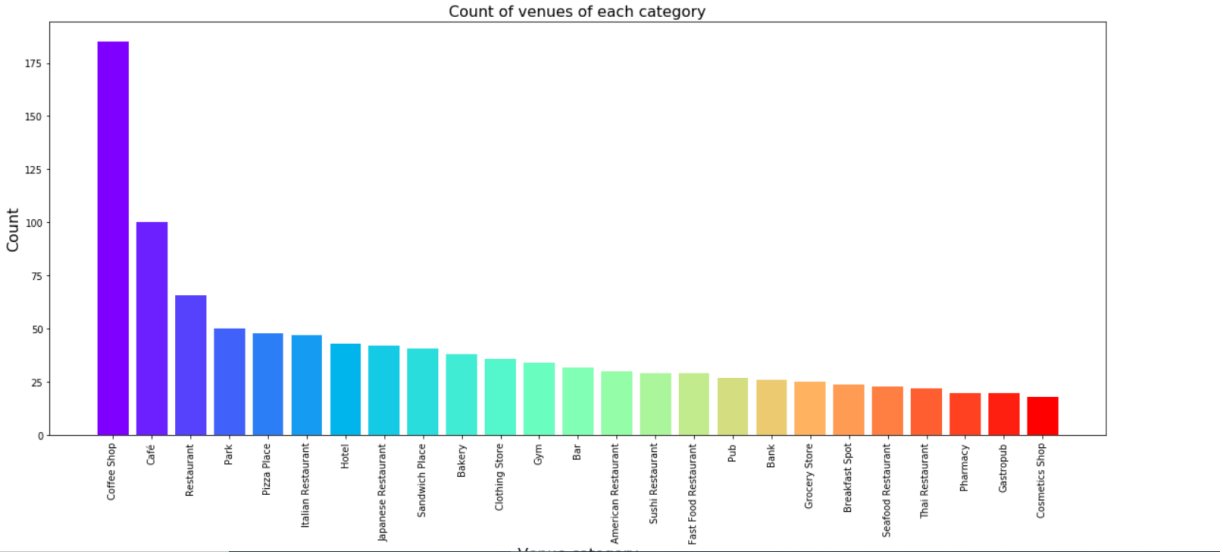
With the data obtained which has neighborhood information along with venues around,we would perform next steps to converge to a solution

3.Methodology and exploratory data analysis:

As a first step, data regarding neighborhoods is extracted from Wikipedia page which includes the neighborhood name, postal code and borough name. Grouped the neighborhoods based on the postal code and borough name. latitude and longitude

information for each postal code is extracted from geospatial csv file. New data frame with added geospatial data is formed.
Using latitude and longitude information of the neighborhoods, folium map is created

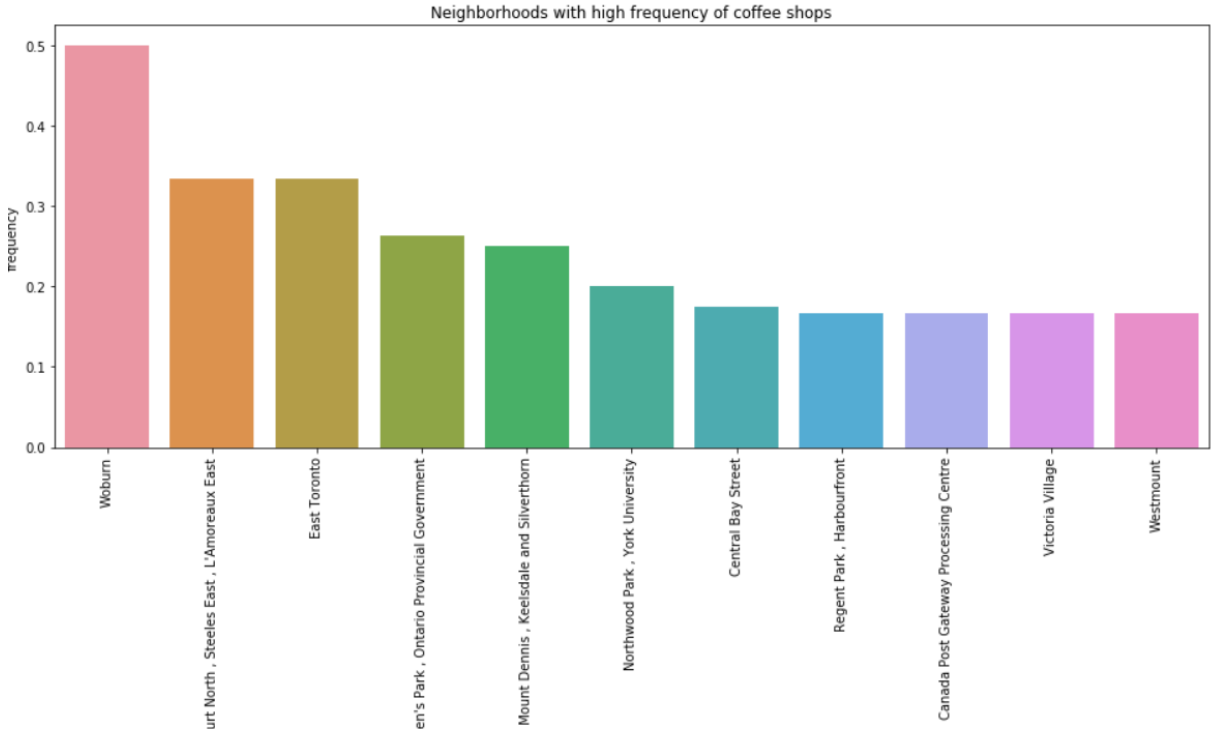
Now Foursquare API is utilized to extract venues and their respective categories for each of the neighborhood of Toronto.
Distribution of venue categories is as follows:

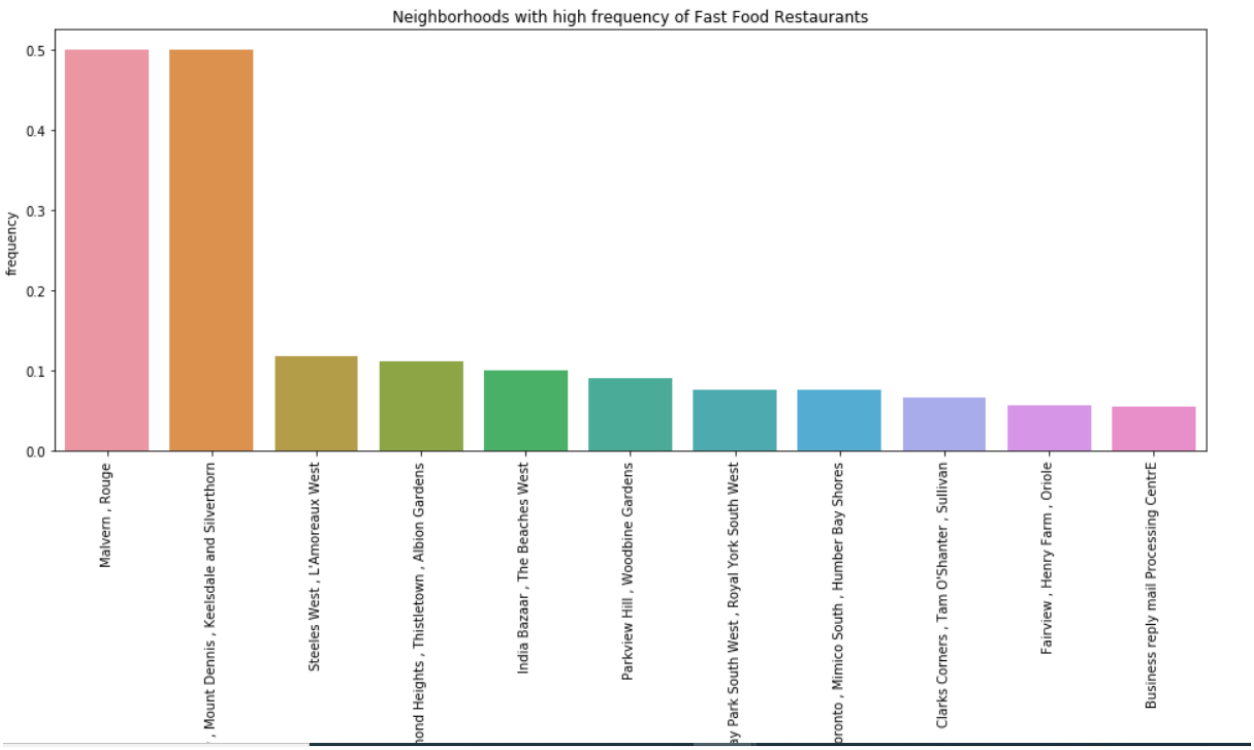


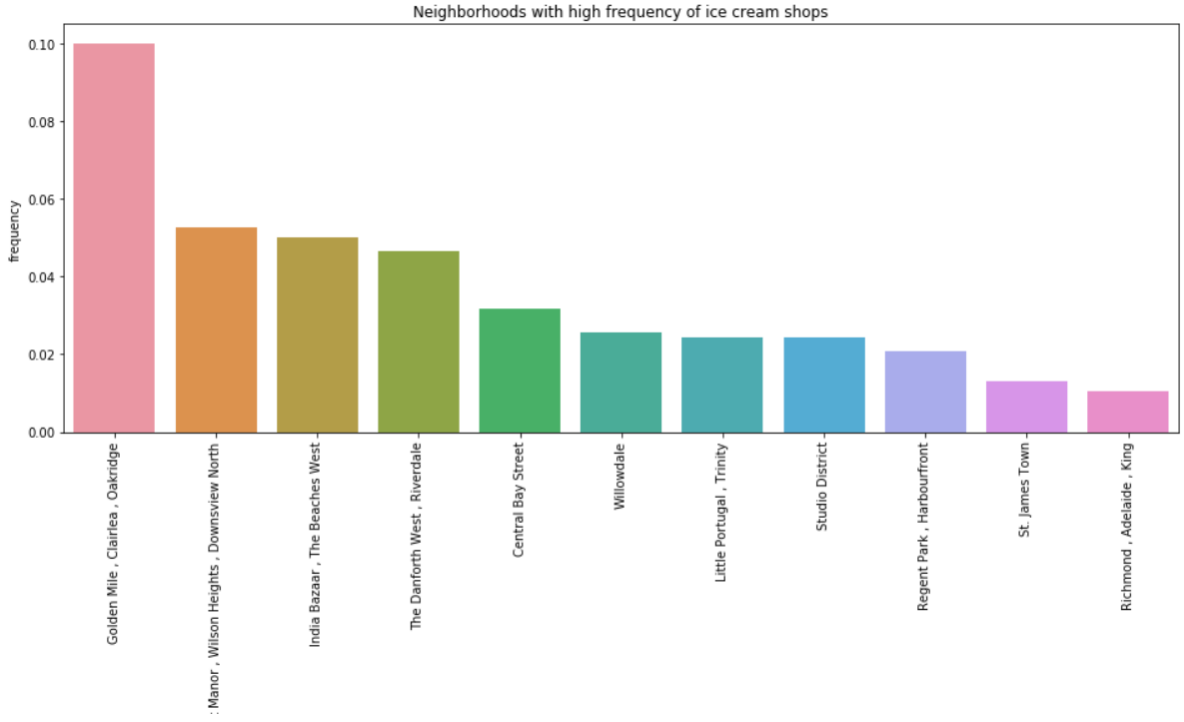
From the distribution plot, its evident that most of the venues are in coffee shop category followed by café and restaurants. From this we can say that population of Toronto has an affinity towards coffee shops and restaurants.

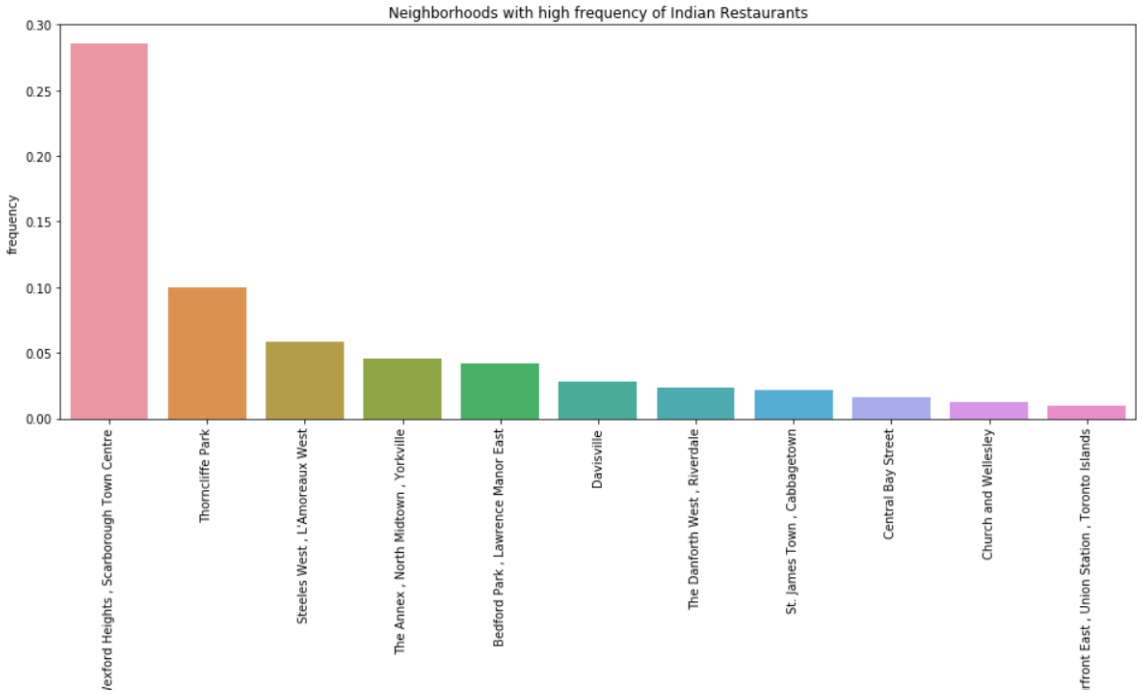
After obtaining the venues, sorting and cleansing of data is done to obtain top ten common venue categories for each neighborhood based on the frequency of occurring of venues in each category.

Different neighborhoods have different set of venue categories, distribution can be seen as below.




Neighborhoods with high frequency of Fast Food Restaurants
frequency
0.0
0.1
0.2
0.3
0.4
0.5
Malvern , Rouge
, Mount Dennis , Keelsdale and Silverthorn
Steeles West , L'Amoreaux West
hond Heights , Thistletown , Albion Gardens
India Bazaar , The Beaches West
Parkview Hill , Woodbine Gardens
ay Park South West , Royal York South West
pronto , Mimico South , Humber Bay Shores
Clarks Corners , Tam O'Shanter , Sullivan
Fairview , Henry Farm , Oriole
Business reply mail Processing CentrE


Neighborhoods with high frequency of ice cream shops
frequency
0.00
0.02
0.04
0.06
0.08
0.10
Golden Mile , Clairlea , Oakridge
: Manor , Wilson Heights , Downsview North
India Bazaar , The Beaches West
The Danforth West , Riverdale
Central Bay Street
Willowdale
Little Portugal , Trinity
Studio District
Regent Park , Harbourfront
St. James Town
Richmond , Adelaide , King



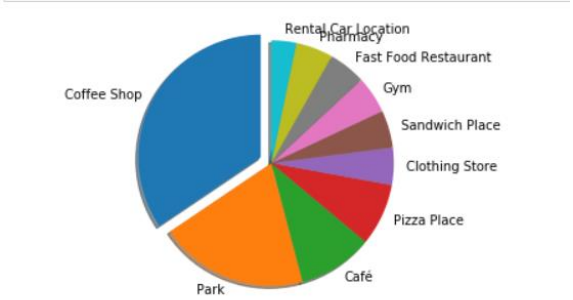
This data can be utilized for competitive analysis as this gives a fair idea on the type of venue categories in each neighborhood and the neighborhoods with high frequency of a particular category.

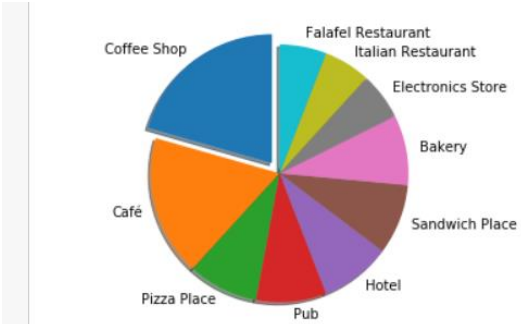
Data is modified to provide top ten categories in each neighborhood.

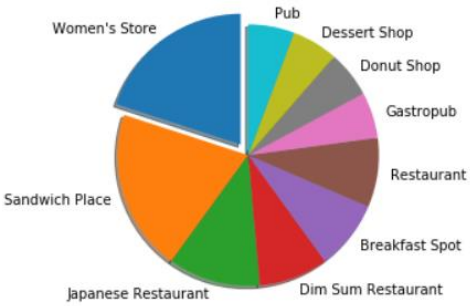
	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	Comm
0	Agincourt	Lounge	Latin American Restaurant	Clothing Store	Breakfast Spot	Skating Rink	Event Space	Ethiopian Restaurant	Electr
1	Alderwood , Long Branch	Pizza Place	Gym	Pub	Sandwich Place	Dance Studio	Athletics & Sports	Skating Rink	C
2	Bathurst Manor , Wilson Heights , Downsview North	Coffee Shop	Bank	Shopping Mall	Bridal Shop	Ice Cream Shop	Supermarket	Middle Eastern Restaurant	Sushi

Visualizing this data provides information regarding the top venue categories in all neighborhoods combined. Plots below represent the distribution of categories as

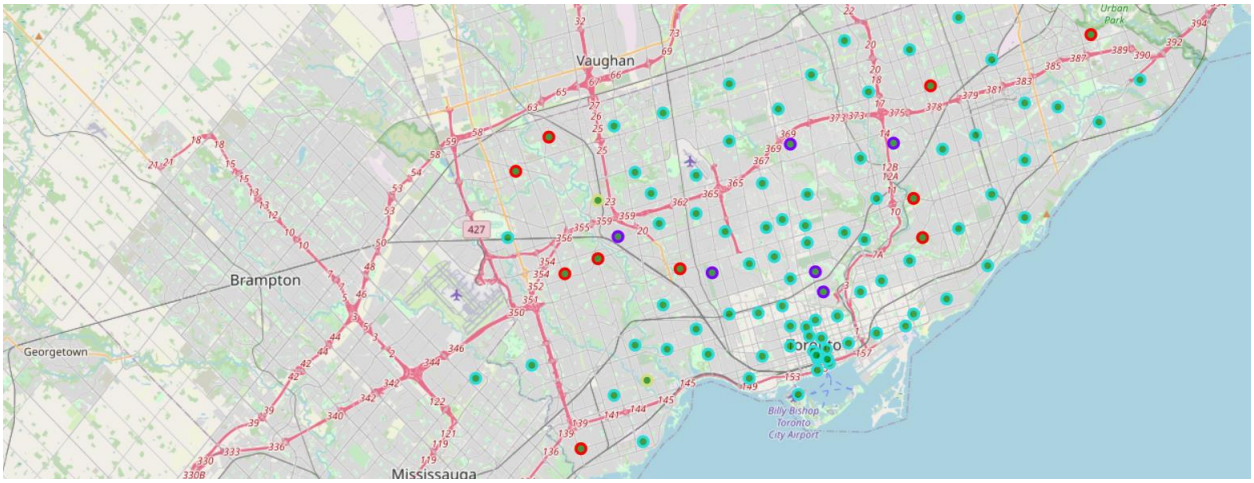
1. First most common venue category
2. Second most common venue category
3. Third most common venue category
4. Fourth most common venue category







Post the data analysis, Clustering is carried out on the neighborhood based on the venue details in the neighborhood.



4.Results and Discussion:

Based on the data obtained, it is evident that the Toronto people have more interest in coffee shops and they are available in almost every neighborhood. Even Cafes are most prominent in many neighborhoods. So if a new investment is to be thought, we can consider it to be a coffee shop or a café as they appear to be most present venues in Toronto. But tis would not be a best choice incase of revenue generation as already there are a lot of stores in every neighborhood. So, the investor should rethink investing in a coffee shop if they want to have a high revenue generation.

Case1: If coffee shop is a preferred investment

In this case, we would go ahead with cluster 2 where most of the neighborhoods have a coffee shop as first or second most common venue.

In cluster2 we would select those neighborhood which do not have coffee shop as its first common venue as this would be the location where there is a possibility of new coffee shop being successful. This is based on the assumption that the neighborhoods in this cluster are similar in other ways.

Some of the preferred neighborhoods would be with postal codes, M3B,M1C and M6B

Further analysis can be done based on population density, number of shops in the area and competitors to select a final location

Case2: if coffee shop is not a preferred investment

In this case, we need to come up with alternate investment opportunity. I would go ahead with the restaurant plan as next best common venue in Toronto is restaurant.

For this case, a location in cluster 0 can be selected as this cluster defines a market condition with lots of food places in which majority of neighborhoods have pizza place or fast food restaurant as their first common venues.

We can select a neighborhood in this cluster with low number of pizza place or fast food restaurants.

Further analysis in terms of ethnicity of the neighborhood would be useful in determining the type of restaurant along with the theme

As of now, we can recommend neighborhoods with postal codes M8W and M9V

5.Conclusion:

The purpose of this project was to explore the neighborhoods of Toronto and recommend a best investment opportunity for a restaurant owner. Based on the obtained results, we concluded that the coffee shop is most preferred as a common venue and restaurant is a best option for revenue generation. Further analysis can be done by considering population and ethnicity of the neighborhoods for determining the best type of restaurant.