Toronto Cuisine Analysis

i) Introduction

This project was conducted for the IBM Data Science capstone final project. The topic was chosen based on the previous labs which was where most of the concepts were picked up from. The location was unfortunately also repeated as Geopy.geocode package was behaving unexpectedly for other countries and the accuracy of the coordinates were quite questionable. This project is for analysis of different cuisines/eateries around the toronto area and to get an idea of what kind of establishment are prevalent here.

ii) Business Problem

The analysis is to help people to start their own Food related establishment in the Toronto area. Toronto being one of the major cities in Canada and even the most populous city in Canada so starting a restaurant can be a major risk as the land/rent is very expensive and options are plenty so if one does not take the right approach it will lead to shut down even if the food is good because one is not able to compete with other in the same business

This analysis project is meant to look into the various eateries in and around Toronto. This project will help to determine the location of eateries in their particular neighbourhoods and get the details of restaurants. This would help us understand what kind approach one could take to be competitive in the area

iii) Data Required

For this project we will be using the postal code and district in from wikipedia which we will scrape for a data set. Then we will use geopy to get the coordinates of the place. Once we have the coordinates of the neighbourhood then we start by pinpointing the locations on the map using the folium package.

After the initial inspection we are now ready to dive into the data of the actual eateries around the area. For this we will be using the foursquare api to get details about the restaurants/eateries in their respective areas. Once we have made our custom URL we can search for various location and details like their distance from their particular area and the type/cusine they have which will help us narrow down the what kind of cuisine can be profitable in the particular area

The data gathering step by step

1. Scrape the wikipedia page for postal code

|	Postal_Code	Borough	District
2	M3A	North York	Parkwoods
3	M4A	North York	Victoria Village
4	M5A	Downtown Toronto	Regent Park, Harbourfront
5	M6A	North York	Lawrence Manor, Lawrence Heights
6	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government

2. Use Geocoder to get the coordinates

	Postal_Code	Borough	District	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	43.65426	-79.360636
5	M6A	North York	Lawrence Manor, Lawrence Heights	43.718518	-79.464763
6	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government	43.662301	-79.389494

3. Mapping out the districts
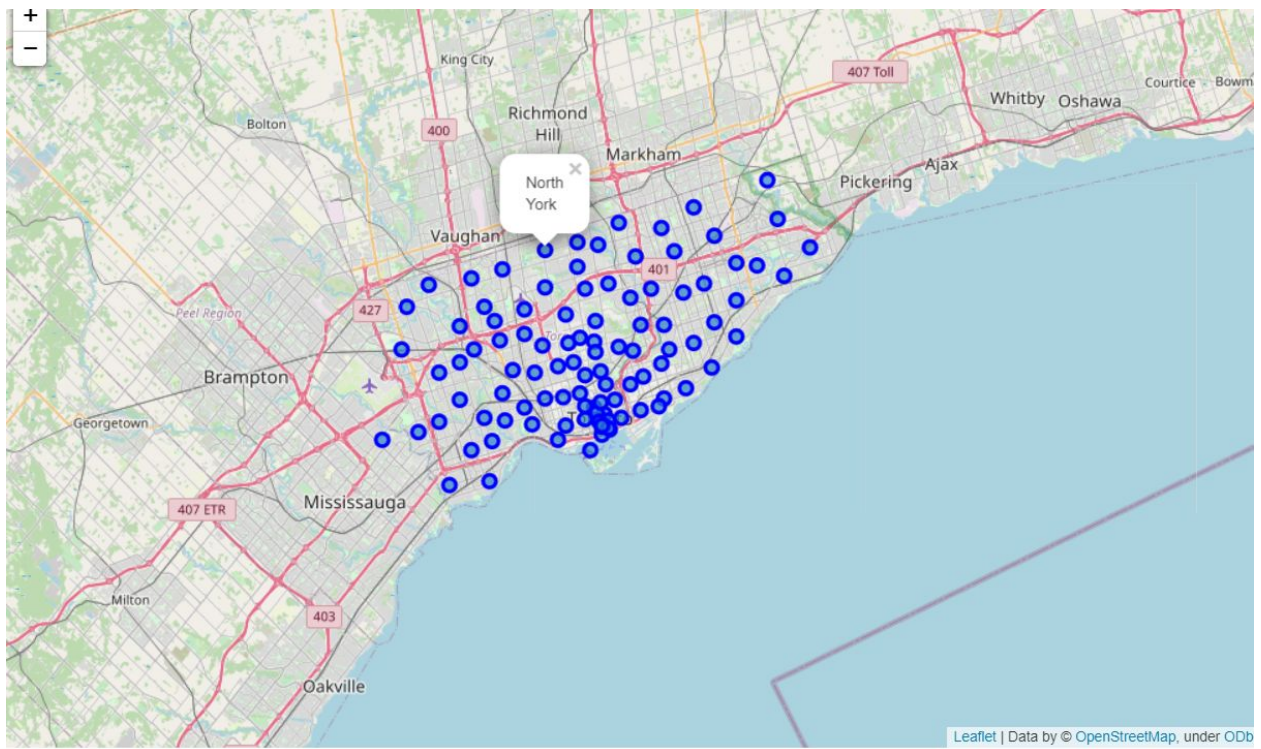
4. Mapping toronto district
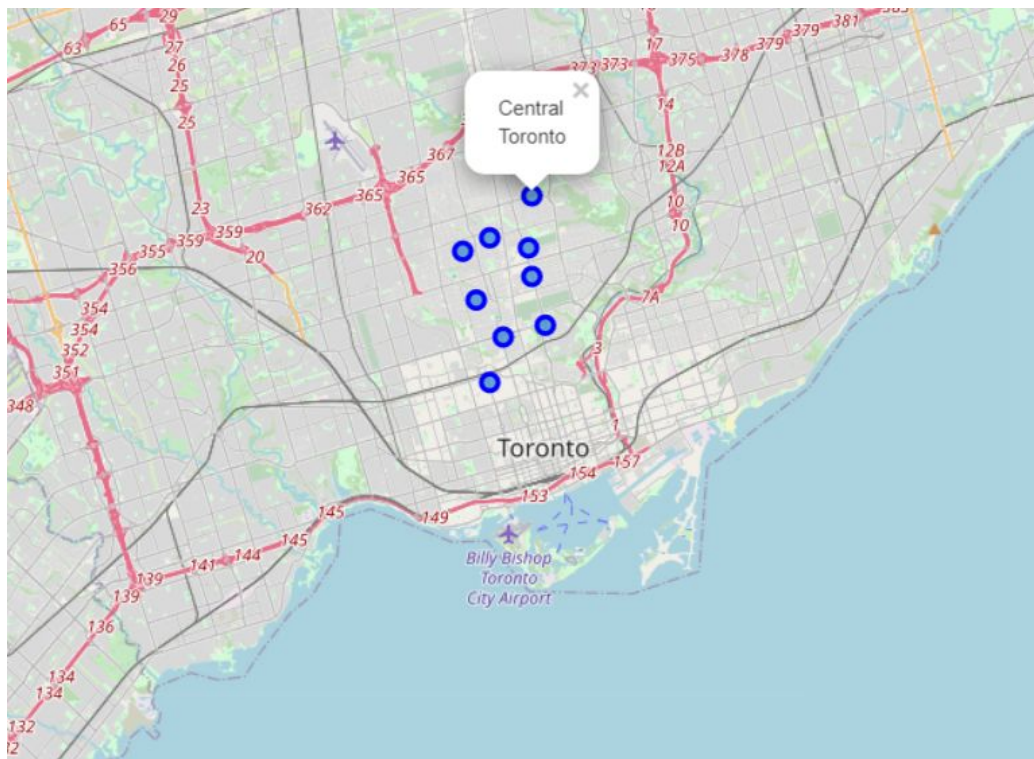
5. Making foursquare api string for finding out eateries in the area

	Borough	lati	long	api_url
0	Central Toronto	43.653482	-79.383935	https://api.foursquare.com/v2/venues/explore?c...
1	Downtown Toronto	43.656322	-79.380916	https://api.foursquare.com/v2/venues/explore?c...
2	East Toronto	43.721789	-79.374027	https://api.foursquare.com/v2/venues/explore?c...
3	East York	43.699971	-79.33252	https://api.foursquare.com/v2/venues/explore?c...
4	Etobicoke	43.643556	-79.565633	https://api.foursquare.com/v2/venues/explore?c...
5	Mississauga	43.590338	-79.645729	https://api.foursquare.com/v2/venues/explore?c...
6	North York	43.754326	-79.449117	https://api.foursquare.com/v2/venues/explore?c...
7	Scarborough	43.773077	-79.257774	https://api.foursquare.com/v2/venues/explore?c...
8	West Toronto	43.653482	-79.383935	https://api.foursquare.com/v2/venues/explore?c...
9	York	46.088526	-66.930803	https://api.foursquare.com/v2/venues/explore?c...

6. Parse the results to get eatrie details

	id	name	latitude	longitude	distance_from_location	address	cuisine	Borough
0	537773d1498e74a75bb75c1e	Eggspectation Bell Trinity Square	43.65314384	-79.38198017	161	[483 Bay Street (Albert Street), Toronto ON M5...	Breakfast Spot	Central Toronto
1	4ae7b27df964a52068ad21e3	Japango	43.65526772	-79.38516507	222	[122 Elizabeth St. (at Dundas St. W), Toronto ...	Sushi Restaurant	Central Toronto
2	4e5d8181a8092f63968617ee	Crepe Delicious	43.65453649	-79.38088886	271	[220 Yonge St., Toronto ON M5B 2H1, Canada]	Fast Food Restaurant	Central Toronto
3	57bcd3b7498e652a678d0378	Poke Guys	43.65489528	-79.38505238	181	[112 Elizabeth St (at Dundas St W), Toronto ON...	Poke Place	Central Toronto
4	59246b5aad1789316b35d66c	JOEY Eaton Centre	43.65540380	-79.38192890	268	[1 Dundas St W, Toronto ON M5G 1Z3, Canada]	Restaurant	Central Toronto
...	...	...	...	...	...	...	...	...
273	4b744336f964a520d8d02de3	Somethin' 2 Talk About	43.65839479	-79.38533766	558	[78 Gerrard St W, Toronto ON M5G 1J5, Canada]	Middle Eastern Restaurant	West Toronto
274	4ad4c05df964a52059f620e3	Canoe	43.64745207	-79.38132002	703	[66 Wellington St West (at Bay Street), Toront...	Restaurant	West Toronto
275	4b114234f964a520637923e3	Pumpernickel's Deli	43.64883150	-79.38197006	541	[100 King St. W (at First Canadian Place Food ...	Deli / Bodega	West Toronto
276	506a3830e4b0677a1f9d8eda	Ninki Izakaya	43.65022782	-79.38486260	369	[133 Richmond Street West (Richmond And Univer...	Japanese Restaurant	West Toronto
277	4ad9f607f964a520691c21e3	Manpuku まんぷく	43.65361241	-79.39061276	538	[105 McCaul St. Unit 29-31 (at Dundas St. W.),...	Japanese Restaurant	West Toronto

278 rows × 8 columns

7. Map out the eateries
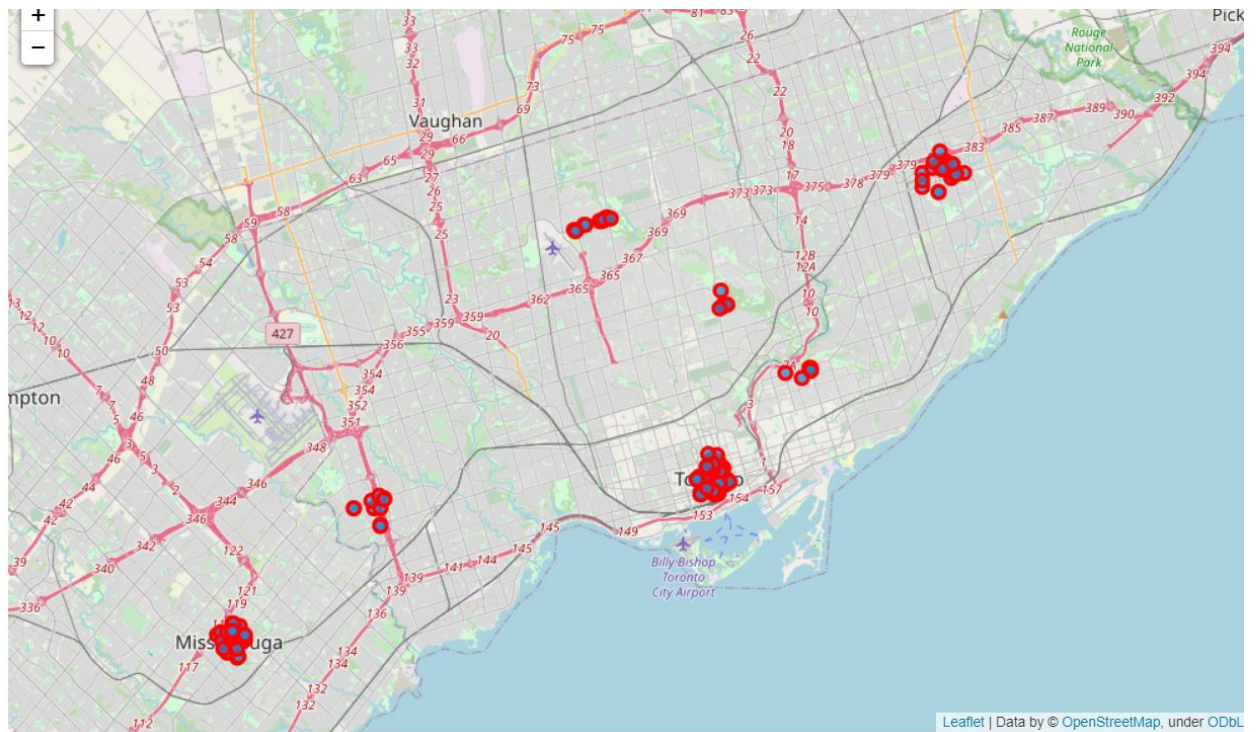
iv) Methodology

Now we have got all the data and we have cleaned it up so now we can proceed with our analysis process. We could try to use the groupby() function to try to get an idea of the kind of data we are working with but it does not have that much clarity

```
In [357]: df_det.groupby('cuisine').count()
Out[357]:
```

cuisine	id	name	latitude	longitude	distance_from_location	address	Borough
American Restaurant	13	13	13	13	13	13	13
Asian Restaurant	8	8	8	8	8	8	8
Bakery	10	10	10	10	10	10	10
Bistro	1	1	1	1	1	1	1
Breakfast Spot	10	10	10	10	10	10	10
Burger Joint	2	2	2	2	2	2	2
Burrito Place	3	3	3	3	3	3	3
Café	23	23	23	23	23	23	23
Chinese Restaurant	11	11	11	11	11	11	11
Creperie	1	1	1	1	1	1	1
Deli / Bodega	8	8	8	8	8	8	8
Dim Sum Restaurant	1	1	1	1	1	1	1
Diner	2	2	2	2	2	2	2
Donut Shop	1	1	1	1	1	1	1
Falafel Restaurant	2	2	2	2	2	2	2
Fast Food Restaurant	10	10	10	10	10	10	10
Fish & Chips Shop	1	1	1	1	1	1	1
Food Court	7	7	7	7	7	7	7
French Restaurant	1	1	1	1	1	1	1
Fried Chicken Joint	1	1	1	1	1	1	1
Gastropub	5	5	5	5	5	5	5

1. we will have to reduce our area of search to only areas around toronto so we can get a much more accurate analysis.
2. Once we make a new data frame with only with toronto main district we will start encoding our dataset

3. We will use 'onehot' encoder to encode the different types of cuisines

	Neighborhood	American Restaurant	Asian Restaurant	Bakery	Breakfast Spot	Burger Joint	Burrito Place	Café	Chinese Restaurant	Creperie	...	Pizza Place	Poke Place	Ramen Restaurant	Restaurant	Sandwich Place
0	Central Toronto	0	0	0	1	0	0	0	0	0	...	0	0	0	0	0
1	Central Toronto	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0
2	Central Toronto	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0
3	Central Toronto	0	0	0	0	0	0	0	0	0	...	0	1	0	0	0
4	Central Toronto	0	0	0	0	0	0	0	0	0	...	0	0	0	1	0

5 rows × 35 columns

4. Then we will use a combination of groupby() and mean() function to group together the results of different district in just 4 districts of toronto

	Neighborhood	American Restaurant	Asian Restaurant	Bakery	Breakfast Spot	Burger Joint	Burrito Place	Café	Chinese Restaurant	Creperie	...	Pizza Place	Poke Place	Ramen Restaurant	Restaurant	Sar
0	Central Toronto	0.06	0.04	0.04	0.06	0.00	0.02	0.12000000	0.04	0.00	...	0.00	0.02	0.02	0.08000000	
1	Downtown Toronto	0.02	0.02	0.02	0.04	0.02	0.02	0.08000000	0.04	0.02	...	0.02	0.02	0.06	0.08000000	
2	East Toronto	0.00	0.00	0.00	0.00	0.00	0.00	0.16666667	0.00	0.00	...	0.00	0.00	0.00	0.16666667	
3	West Toronto	0.06	0.04	0.04	0.06	0.00	0.02	0.12000000	0.04	0.00	...	0.00	0.02	0.02	0.08000000	

4 rows × 35 columns

5. Based on which we will give some primitive findings about the the kinds of places which appear more frequently

```
Central Toronto
                   cuisine  freq
0                     Café  0.12
1   Japanese Restaurant     0.08
2              Restaurant   0.08
3   Seafood Restaurant      0.06
4   American Restaurant     0.06


Downtown Toronto
                   cuisine  freq
0   Japanese Restaurant     0.08
1                     Café  0.08
2              Restaurant   0.08
3               Gastropub   0.06
4      Ramen Restaurant     0.06


East Toronto
                   cuisine  freq
0             Food Court    0.33
1          Deli / Bodega   0.17
2   Falafel Restaurant     0.17
3                     Café  0.17
4              Restaurant   0.17


West Toronto
                   cuisine  freq
0                     Café  0.12
1   Japanese Restaurant     0.08
2              Restaurant   0.08
3   Seafood Restaurant      0.06
4   American Restaurant     0.06
```

6. Now we will convert these findings to a dataframe to much clearly represent the findings

	Neighborhood	1st Most Common Cuisine	2nd Most Common Cuisine	3rd Most Common Cuisine	4th Most Common Cuisine	5th Most Common Cuisine
0	Central Toronto	Café	Japanese Restaurant	Restaurant	American Restaurant	Breakfast Spot
1	Downtown Toronto	Japanese Restaurant	Restaurant	Café	Seafood Restaurant	Ramen Restaurant
2	East Toronto	Food Court	Café	Falafel Restaurant	Restaurant	Deli / Bodega
3	West Toronto	Café	Japanese Restaurant	Restaurant	American Restaurant	Breakfast Spot

7. Now with the help of the K-means clustering algorithm we can further simplify our findings and group the eateries into distinct clusters so we filter out more unnecessary details
8. Using k-means we find out that the data can be grouped into 2 unique sets
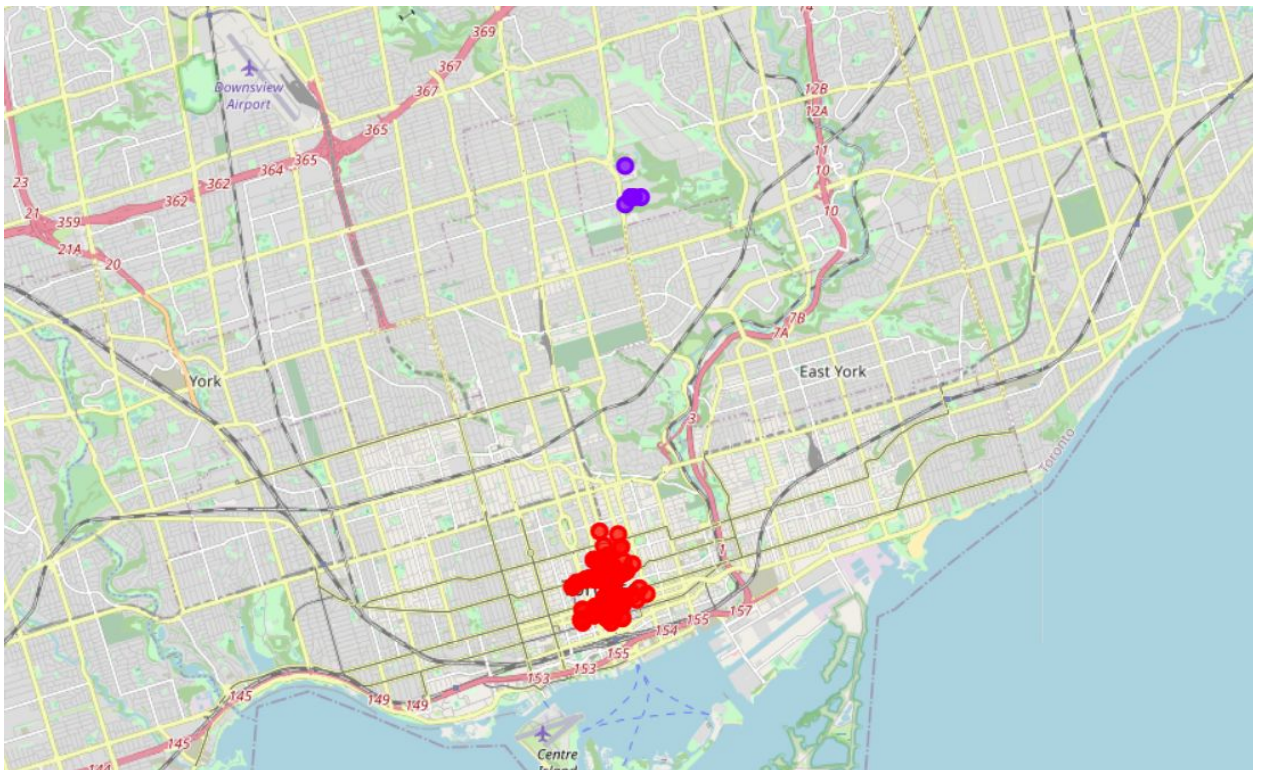
v) Results

From the above steps we are able to get 2 unique cluster with different readings which helps us determine what steps one needs to keep in mind to start a eatery

Cluster 1(RED):

3	Central Toronto	43.65489528	-79.38505238	0	Café	Japanese Restaurant	Restaurant	American Restaurant	Breakfast Spot
4	Central Toronto	43.65540380	-79.38192890	0	Café	Japanese Restaurant	Restaurant	American Restaurant	Breakfast Spot
...	...	...	...	...	...	...	...	...	...
273	West Toronto	43.65839479	-79.38533766	0	Café	Japanese Restaurant	Restaurant	American Restaurant	Breakfast Spot
274	West Toronto	43.64745207	-79.38132002	0	Café	Japanese Restaurant	Restaurant	American Restaurant	Breakfast Spot

Cluster 2(BLUE):

	Neighborhood	latitude	longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
100	East Toronto	43.72181380	-79.37652043	1	Food Court	Café	Falafel Restaurant	Restaurant	Deli / Bodega
101	East Toronto	43.72162952	-79.37642455	1	Food Court	Café	Falafel Restaurant	Restaurant	Deli / Bodega
102	East Toronto	43.72161664	-79.37592540	1	Food Court	Café	Falafel Restaurant	Restaurant	Deli / Bodega
103	East Toronto	43.72178602	-79.37415044	1	Food Court	Café	Falafel Restaurant	Restaurant	Deli / Bodega
104	East Toronto	43.72045203	-79.37793909	1	Food Court	Café	Falafel Restaurant	Restaurant	Deli / Bodega
105	East Toronto	43.72731127	-79.37783462	1	Food Court	Café	Falafel Restaurant	Restaurant	Deli / Bodega

vi) Conclusion

From our findings we can conclude that the 1st cluster has a lot of cafe style eateries where people go for snacks rather than entire meals and cluster 2 has food courts as the prevalent type of eatery which has a variety of food at a cheap price. We can see this trend with the other common venues. But one of the outliers of cheap and fast food are the japanese restaurants present in the first cluster. This could be attributed to the city facing a major lake which could provide a good and cheap source for these restaurants .

If one is planning to open his own establishment it would be sensible to open a fast food/cafe style eatery or exploit the geographical advantage of being next to a freshwater fish supply and open a seafood related eatery.

THANK YOU