Report

Capstone Final Asignment:

The Battle of Neighborhoods

by Carlos wilfredo Alayon

Table of Content

Introduction

Description of the problem

A client friend, want to start a Venezuelan fast food restaurant in Toronto, Canada. It will be a restaurant specialized in "arepas" that are made based on ground corn flour.

They are circular and flattened and once cooked, on a metal plate, they are filled with meat, chicken, cheese, ham, pork and many other things. It is part of the traditional gastronomy of Venezuela and its neighboring country Colombia.
Venezuelan arepas restaurants have proliferated in recent years in South America and now also in Europe, the United States and Canada. Our client want to know if Toronto would be a good place to start their restaurant and what would be a good location for his business in that city.

Description and use of the data to solve the problem

The restaurant is designed for all public, but it is reasonable to think that Venezuelan residents in the city would be the restaurant's main clients, at least at the beginning of the business, so it should obtain data from the estimate of Venezuelans living in Toronto and the immigration figures of Venezuelans in recent years to this city, as a way to justify the project.

According to the above, we will investigate a web page with immigration data of Venezuelans to the city of Toronto in Canada in the last 10 years, to present this information in the report.

Number of Venezuelan citizens who got permanent residence in Canada by year

Year	Total admited Venezuelan citizen	Total admited permanent resients	Percentage admited permanent residents
2012	1,373	257,895	0.5%
2013	1,022	258,953	0.4%

The data on neighborhoods and neighborhoods of Toronto will be obtained from the page:
'https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M')

Which will be appropriate pre-processed (cleaned and grouped conveniently) to be able to use it and finally filtered only for the data associated with Toronto.
The Geospatial coordinates will be obtained from the file Geospatial_Coordinates.csv. located on the web page:

http://cocl.us/Geospatial_data/Geospatial_Coordinates.csv

Postcode	Borough	Neighborhood	Latitude	Longitude
M4E	East Toronto	The Beaches	43.676357	-79.293031
M4K	East Toronto	The Danforth West,Riverdale	43.679557	-79.352188
M4L	East Toronto	The Beaches West,India Bazaar	43.668999	-79.315572
M4M	East Toronto	Studio District	.659526	-79.340923
M4N	Central Toronto	Lawrence Park	.728020	-79.38879

A merge will be made with this data from different sources to have the consolidated data in a single dataframe. The geopy library will be used to get the latitude and longitude values of candidate sites to locate the restaurant of our clients and the Foursquare API to explore the neighborhoods and boroughs in Toronto from the possible locations of our client's restaurant.

The location of arepa restaurants in Toronto will be determined, but the algorithms will be run taking into account the locations of Latin American food restaurants: Mexican, Colombian, Peruvian, etc.

The k-means algorithms will be applied to determine and examine the cluster and the recommended location or locations to set up the restaurant.

Methodology

This work aims to select a location for a local Venezuelan arepas in the city of Toronto, Canada. For which an initial study was made of the Venezuelan emigration to Canada and particularly to Toronto, since Venezuelans would be the main clients, at least when they start the food business.

Subsequently, the data was collected, the Toronto postal list with the neighborhoods and their geospatial coordinates. The data was also preprocessed, which included cleaning, mixing and geographic grouping of the data.

We use Foursquare API to explore the neghborhoods and boroughs, 2 Explore and analyze venues in Toronto. They were grouped by neighborhood and by taking the mean of the frequency of occurrence of each category. We printed each neighborhood along with the top 5 most common venues and final put into a pandas dataframe.

For clustering Neighborhoods, we execute the k-means algorithm, using 5 clusters. We create a dataframe that includes the cluster as well as the top 10 venues for each neighborhood. The usters clusters were examined.
On the other hand, we analyze Cluster Neighborhoods of food venues and Examine Clusters with only food venues in each Neighborhood. We aggregate the data by neighborhood taking the mean of the frequency of occurrence of each category.

Finally we printed each neighborhood along with the top 5 most common venues. We made Clustering Neighborhoods of food venues and Run k-means to cluster the neighborhood into 5 clusters. We created a new dataframe relationed with food venues includes the cluster as well as the top 10 venues for each neighborhood. Also let's visualize the resulting clusters of the food venues. Finally, we conducted a Results Analysis that included Customer Analysis, Competitors Analysis and Venues Analysis

In the Development of work the following steps were carried out

Step 1. install and Import required libraries
Step 2 Data Collection and Preprocessing
Step 2.1 Data Collection about Venezuelan inmigration to Toronto, Canada
Step 2.2 Get List of postal codes of Canada
Step 2.3 Add column headers, cells cleaning, data merging and grouping by Boroughs

Step 3 Using the Foursquare API to explore the neighborhoods in Toronto and segment them
Step 3.1 Get credentiails
Step 3.2 Explore venues in neighborhoods in Toronto¶
Step 3.3 Analyze venues in each Neighborhood
Step 3.4 let's group rows by neighborhood and by taking the mean of the frequency of occurrence of each category
Step 3.5 Let's print each neighborhood along with the top 5 most common venues
Step 3.6 Let's put that into a pandas dataframe

Step 4 Let's Cluster Neighborhoods
Step 4.1 Run k-means to cluster the neighborhood into 5 clusters
Step 4.2 Let's create a new dataframe that includes the cluster as well as the top 10 venues for each neighborhood

Step 4.4 Examine Clusters

Step 5 Analyse Cluster Neighborhoods of food venues
Step 5.1 Examine Clusters with only food venues
Step 5.2 Analyze venues in each Neighborhood
Step 5.3 let's group rows by neighborhood and by taking the mean of the frequency of occurrence of each category
Step 5.4 Print each neighborhood along with the top 5 most common venues
Step 5.5 Let's put that into a pandas dataframe

Step 6 Let's Cluster Neighborhoods of food venues
Step 6.1 Run k-means to cluster the neighborhood into 5 clusters
Step 6.2 Let's create a new dataframe relationed with food venues includes the cluster as well as the top 10 venues for each neighborhood
Step 6.3 Finally, let's visualize the resulting clusters

Data Collection and Preprocessing

Data Collection about Venezuelan inmigration to Toronto, Canada, I was carried out some preliminary investigation of the Venezuelan immigration to Canada was carried out, finding that Venezuelans are one of the more than ten Latino groups living in Canada. The first Venezuelan immigrants arrived in the United States in the 1960s, but more than half arrived after the general strike in Venezuela from 2002 to 2003 and it have been in constant increase since the years 2001 to 2013. The province of Ontario is home to the majority of Canadians of Venezuelan descent, concentrating in Toronto, the capital and most populous city of the province, as well as the most populated in the country.

Table No. 1. Number of Venezuelan citizens who got permanent residence in Canada by year

Year	Total admited Venezuelan citizen	Total admited permanent residents	Percentage admited permanent residents
2001	572	250.639	0.2%
2002	554	229.048	0.2%
2003	710	221.349	0.3%
2004	1.224	235.823	0.5%
2005	1.211	262.242	0.5%
2006	1.192	251.640	0.5%
2007	1.335	236.753	0.6%
2008	1.239	247.245	0.5%
2009	1.353	252.172	0.5%
2010	998	280.688	0.4%
2011	1.452	248.749	0.6%
2012	1.373	257.895	0.5%
2013	1.022	258.953	0.4%

Table No. 2. Recent immigrants by selected places of birth

Characteristics	Total	Male	Female
Other places of birth in Americas	8390	3540	4845
United States	4015	2085	1925
Jamaica	3530	1840	1685
Mexico	2065	995	1070
Brazil	1915	930	980
Colombia	1430	670	760
Venezuela	1055	480	575
Cuba	770	345	430
Haiti	680	315	360

Using the Foursquare API to explore the neighborhoods in Toronto and segment them

The following venues were found:

The Beaches
The Danforth West,Riverdale
The Beaches West,India Bazaar
Studio District
Lawrence Park
Davisville North
North Toronto West
Davisville
Moore Park,Summerhill East
Deer Park,Forest Hill SE,Rathnelly,South Hill,Summerhill West
Rosedale
Cabbagetown,St. James Town
Church and Wellesley
Harbourfront,Regent Park
Ryerson,Garden District
St. James Town
Berczy Park
Central Bay Street
Adelaide,King,Richmond
Harbourfront East,Toronto Islands,Union Station
Design Exchange,Toronto Dominion Centre
Commerce Court,Victoria Hotel
Roselawn
Forest Hill North,Forest Hill West
The Annex,North Midtown,Yorkville
Harbord,University of Toronto
Chinatown,Grange Park,Kensington Market
CN Tower,Bathurst Quay,Island airport,Harbourfront West,King and Spadina,Railway Land
s,South Niagara
Stn A PO Boxes 25 The Esplanade
First Canadian Place,Underground city
Christie
Dovercourt Village,Dufferin
Little Portugal,Trinity
Brockton,Exhibition Place,Parkdale Village
High Park,The Junction South
Parkdale,Roncesvalles
Runnymede,Swansea
Business reply mail Processing Centre969 Eastern

Table No. 3. Toronto Neighborhoods

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	The Beaches	43.676357	-79.293031	Starbucks	43.678798	-79.298045	Coffee Shop
1	The Beaches	43.676357	-79.293031	Grover Pub and Grub	43.679181	-79.297215	Pub
2	The Beaches	43.676357	-79.293031	Glen Stewart Ravine	43.676300	-79.294784	Other Great Outdoors
3	The Beaches	43.676357	-79.293031	Upper Beaches	43.680563	-79.292869	Neighborhood
4	The Beaches	43.676357	-79.293031	Beaches Fitness	43.680319	-79.290991	Gym / Fitness Center

Table No. 4. Top 10 most common venues for each Neighborhoods of Toronto

	Postcode	Borough	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	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	M4E	East Toronto	The Beaches	43.676357	-79.293031	2	Gym / Fitness Center	Coffee Shop	Other Great Outdoors	Pub	Dessert Shop	Event Space	Ethiopian Restaurant	Electronics Store	Eastern European Restaurant	Dumpling Restaurant
1	M4K	East Toronto	The Danforth West,Riverdale	43.679557	-79.352188	2	Greek Restaurant	Coffee Shop	Ice Cream Shop	Bookstore	Italian Restaurant	Dessert Shop	Brewery	Bubble Tea Shop	Café	Pub
2	M4L	East Toronto	The Beaches West,India Bazaar	43.668999	-79.315572	2	Park	Pet Store	Italian Restaurant	Burrito Place	Sushi Restaurant	Burger Joint	Pub	Ice Cream Shop	Fish & Chips Shop	Hotel
3	M4M	East Toronto	Studio District	43.659526	-79.340923	2	Café	Coffee Shop	Italian Restaurant	Gastropub	American Restaurant	Bakery	Yoga Studio	Sandwich Place	Bar	Fish Market
4	M4N	Central Toronto	Lawrence Park	43.728020	-79.388790	2	Park	Swim School	Bus Line	Dim Sum Restaurant	Women's Store	Dessert Shop	Event Space	Ethiopian Restaurant	Electronics Store	Eastern European Restaurant

Tabla No. 5. Analyse of each Neighborhoods Cluster of venues

Cluster 1

```
toronto_merged.loc[toronto_merged['Cluster Labels'] == 0, toronto_merged.columns[[1] + list(range(5, toronto_merged.shape[1]))]]
```

	Borough	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
13	Downtown Toronto	0	Coffee Shop	Bakery	Café	Park	Mexican Restaurant	Breakfast Spot	Pub	Theater	Shoe Store	Brewery
35	West Toronto	0	Breakfast Spot	Gift Shop	Restaurant	Cuban Restaurant	Movie Theater	Dessert Shop	Dog Run	Bar	Eastern European Restaurant	Coffee Shop

Cluster 2

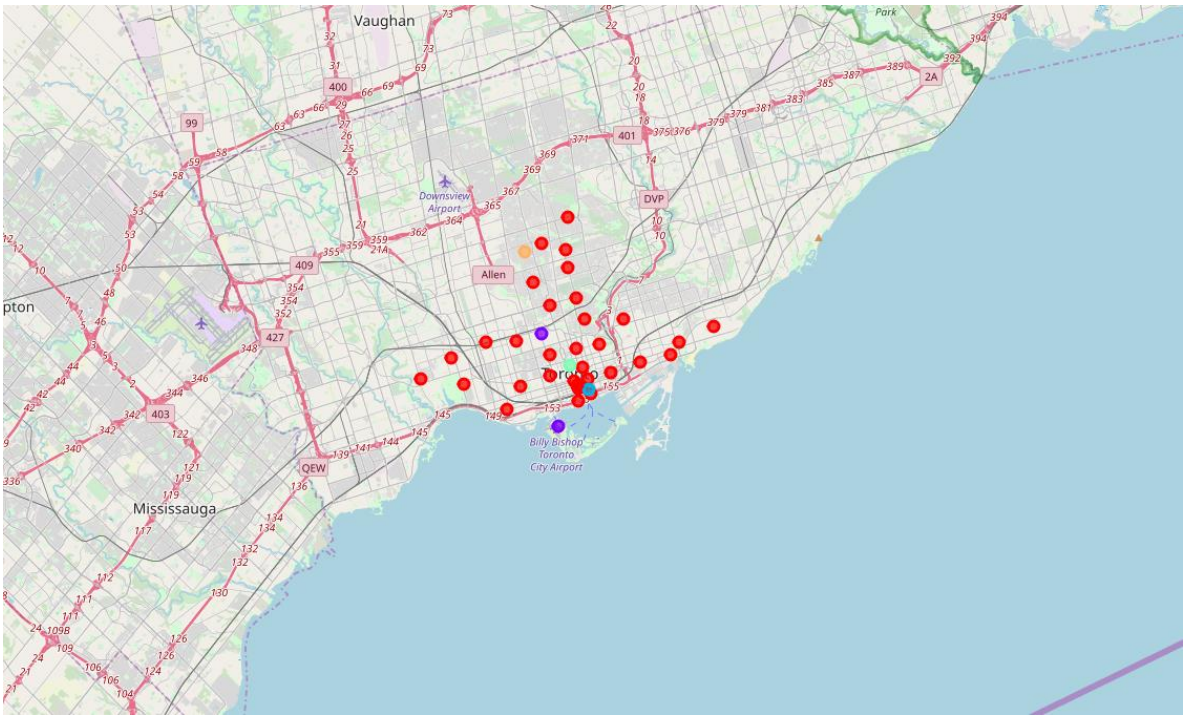
```
toronto_merged.loc[toronto_merged['Cluster Labels'] == 1, toronto_merged.columns[[1] + list(range(5, toronto_merged.shape[1]))]]
```

	Borough	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
17	Downtown Toronto	1	Coffee Shop	Café	Italian Restaurant	Japanese Restaurant	Sandwich Place	Bar	Bubble Tea Shop	Burger Joint	Salad Place	Falafel Restaurant
27	Downtown Toronto	1	Airport Lounge	Airport Terminal	Airport Service	Harbor / Marina	Sculpture Garden	Boutique	Plane	Boat or Ferry	Airport Gate	Airport

Cluster 3

```
toronto_merged.loc[toronto_merged['Cluster Labels'] == 2, toronto_merged.columns[[1] + list(range(5, toronto_merged.shape[1]))]]
```

	Borough	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
0	East Toronto	2	Gym / Fitness Center	Coffee Shop	Other Great Outdoors	Pub	Dessert Shop	Event Space	Ethiopian Restaurant	Electronics Store	Eastern European Restaurant	Dumpling Restaurant
1	East Toronto	2	Greek Restaurant	Coffee Shop	Ice Cream Shop	Bookstore	Italian Restaurant	Dessert Shop	Brewery	Bubble Tea Shop	Café	Pub



Analyse of Cluster Neighborhoods of fast food venues

Examine Clusters with only food venues filtered by the following categories: *"Food", "BBQ", "Bistro", "Chicken", "Steak", "Taco", "Joint", "Arep"* and we made the analysis of the top ten venues most common to each neighborhood.

Table No. 7. Top 10 most common venues for each Neighborhoods of Toronto

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
Adelaide,King,Richmond	Thai Restaurant	American Restaurant	Steakhouse	Restaurant	Burger Joint	Sushi Restaurant	Asian Restaurant	Japanese Restaurant	Greek Restaurant	Gluten-free Restaurant
Berczy Park	Seafood Restaurant	Steakhouse	Restaurant	Greek Restaurant	Italian Restaurant	Comfort Food Restaurant	Burger Joint	Bistro	Belgian Restaurant	French Restaurant
Brockton,Exhibition Place,Parkdale Village	Italian Restaurant	Falafel Restaurant	Caribbean Restaurant	Vietnamese Restaurant	Doner Restaurant	Fried Chicken Joint	French Restaurant	Food Truck	Food Court	Food & Drink Shop
Business reply mail Processing Centre969 Eastern	Fast Food Restaurant	Restaurant	Vietnamese Restaurant	Dim Sum Restaurant	Fried Chicken Joint	French Restaurant	Food Truck	Food Court	Food & Drink Shop	Food
CN Tower,Bathurst Quay,Island airport,Harbourf...	Airport Food Court	Vietnamese Restaurant	Dim Sum Restaurant	Fried Chicken Joint	French Restaurant	Food Truck	Food Court	Food & Drink Shop	Food	Filipino Restaurant

Analyse of each Neighborhoods Cluster of fast food venues

We examined each cluster and determine the discriminating venue categories that distinguish each cluster based in categories of Food Establisment.

Tabla No. 8. Analyse of each Neighborhoods Cluster of food venues

Cluster 1

```
toronto_food_merged.loc[toronto_food_merged['Cluster Labels'] == 1, toronto_food_merged.columns[[1] + list(range(5, toronto_food_merged.shape[1]))]]
```

	Borough	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	food Cluster Labels
15	Downtown Toronto	1	Italian Restaurant	Burger Joint	Japanese Restaurant	Falafel Restaurant	American Restaurant	Indian Restaurant	Korean Restaurant	Chinese Restaurant	Hawaiian Restaurant	Vegetarian / Vegan Restaurant	1
24	Downtown Toronto	1	Airport Food Court	Vietnamese Restaurant	Dim Sum Restaurant	Fried Chicken Joint	French Restaurant	Food Truck	Food Court	Food & Drink Shop	Food	Filipino Restaurant	3

Cluster 2

```
toronto_food_merged.loc[toronto_food_merged['Cluster Labels'] == 2, toronto_food_merged.columns[[1] + list(range(5, toronto_food_merged.shape[1]))]]
```

	Borough	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	food Cluster Labels
0	East Toronto	2	Greek Restaurant	Italian Restaurant	American Restaurant	Health Food Store	Sushi Restaurant	Indian Restaurant	Caribbean Restaurant	Fast Food Restaurant	Eastern European Restaurant	Ethiopian Restaurant	1
1	East Toronto	2	Fast Food Restaurant	Sushi Restaurant	Steakhouse	Italian Restaurant	Burger Joint	Vietnamese Restaurant	Dumpling Restaurant	Eastern European Restaurant	Ethiopian Restaurant	Falafel Restaurant	1
2	East Toronto	2	American Restaurant	Italian Restaurant	Middle Eastern Restaurant	Comfort Food Restaurant	New American Restaurant	Chinese Restaurant	Seafood Restaurant	Latin American Restaurant	Vietnamese Restaurant	Ethiopian Restaurant	1

Results Analysis

a. Customer Analysis

It is expected that potential customers for the business that starts, be Venezuelan and Latin American. Although we do not have the data of distribution of them in the different neighborhoods, we know from the data of Census Profile, 2016 Census Toronto, City, Ontario, Canada *, in that country there is a significant population of Latin Americans and particularly of Venezuelans. Venezuela is among the top 10 countries in the Americas with the largest presence of immigrants in Toronto. Please see tables 1 and 2 at the beginning of the notebook.

b. Competitors Analysis

One way to identify where the ideal type of customer is is to look at the businesses of the competition, since restaurants often seek to be close to their fiercest competitors, for example it is common to see a McDonalds and a Burger

King in the same neighborhood . We also have the option of finding a restaurant location away from the competition, so as not to share the flow of customers with other competitors, but it would be worth asking why no one has considered that area before. It could become more of an investment risk. So our strategy will be to select a cluster, with a number of complementary restaurant options.

Fousquare only reported 1 arepa stablishment, located in the vicinity of Chinatown, Grange Park, Kensington Market. So, here is also the recommendation of the location of our client's restaurant in Downtown Toronto. This would correspond to Downtown Toronto, in cluster No. 2.

```
toronto_venues [toronto_venues ['Venue Category'].str.contains("Arep", case=False)]
```

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
1222	Chinatown,Grange Park,Kensington Market	43.653206	-79.400049	El Arepazo	43.653357	-79.401424	Arepa Restaurant

c. Venues Analysis

An important aspect is foot traffic because it represents the flow of people who could potentially walk around your restaurant. It would be advisable to be close to a subway station for example.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
62	The Beaches West,India Bazaar	43.668999	-79.315572	Subway	43.666238	-79.317019	Sandwich Place
116	Davisville North	43.712751	-79.390197	Subway	43.708378	-79.390473	Sandwich Place
131	North Toronto West	43.715383	-79.405678	Subway	43.716818	-79.400136	Sandwich Place
163	Davisville	43.704324	-79.388790	Subway	43.701742	-79.387600	Sandwich Place
168	Davisville	43.704324	-79.388790	Subway	43.708378	-79.390473	Sandwich Place
225	Cabbagetown,St. James Town	43.667967	-79.367675	Subway	43.665598	-79.368470	Sandwich Place
866	Harbourfront East,Toronto Islands,Union Station	43.640816	-79.381752	Subway	43.639465	-79.383533	Sandwich Place
1136	The Annex,North Midtown,Yorkville	43.672710	-79.405678	Subway	43.674965	-79.406868	Sandwich Place
1140	The Annex,North Midtown,Yorkville	43.672710	-79.405678	Subway	43.675626	-79.410101	Sandwich Place
1141	The Annex,North Midtown,Yorkville	43.672710	-79.405678	Dupont Subway Station	43.674802	-79.406921	Metro Station
1662	Runnymede,Swansea	43.651571	-79.484450	Subway	43.649517	-79.483947	Sandwich Place

As we determined that the ideal location of the establishment is in the Downtown, and we also suggest that it be close to a subway station, then we could

select from the following Neighborhoods: Cabbagetown, St. James Town, Harbourfront, or Regent Park.

Conclusions

In principle the project is feasible and it is expected that it will have enough Venezuelan and Latin American potential clients in Toronto, to start according to the official census of Canada. According to the results, the recommended location for the arepas restaurant is in the Toronto Downtownr, where we could locate only a similar business selling arepas. We also suggest that the business be close to a subway station to make it easier for foot traffic of customers to the restaurant.