EXPLORE YOUR NEW TOWN

INTRODUCTION

I am a prospective Masters Student at the University of South Florida, Tampa.

I am taking up this capstone project to cater needs of International students, since I am one from India. This project will help us to get to know in and out of the neighbourhood of a city where we would be living in future i.e. before arrival on campus. So, that one can be psychologically prepared before actually being there for real and its also a good thing to know a place before you have actually been there. Not only International students even other fellow countrymen from a different location can be benefitted by this project.

Since, I am going to be on USF, Tampa campus this project deals with the neighbourhood of Tampa.

DATA ANALYSIS

There are actually two data sets used in this project:

1. https://geodata.myflorida.com/datasets/ufl::geoplan-civic-centers-stadiums-and-other-large-capacity-facilities-boundaries-in-florida-2019/data
2. Foursquare API data

The first data set is taken from geodata.myflorida.com which has the information of all the civic centers, stadiums and other large capacity facilities in Florida which had to be refined a lot, to bring it around to Tampa, even though it was very informative, this data set did not actually cater to daily essentials information. Hence the second data set was chosen which is a Foursquare API data which was leveraged using the venues API call and further transformed into data frame, which actually helped a lot in the exploratory data analysis.

Let’s have an in-depth analysis of these data sets

1.1 FIRST DATA SET

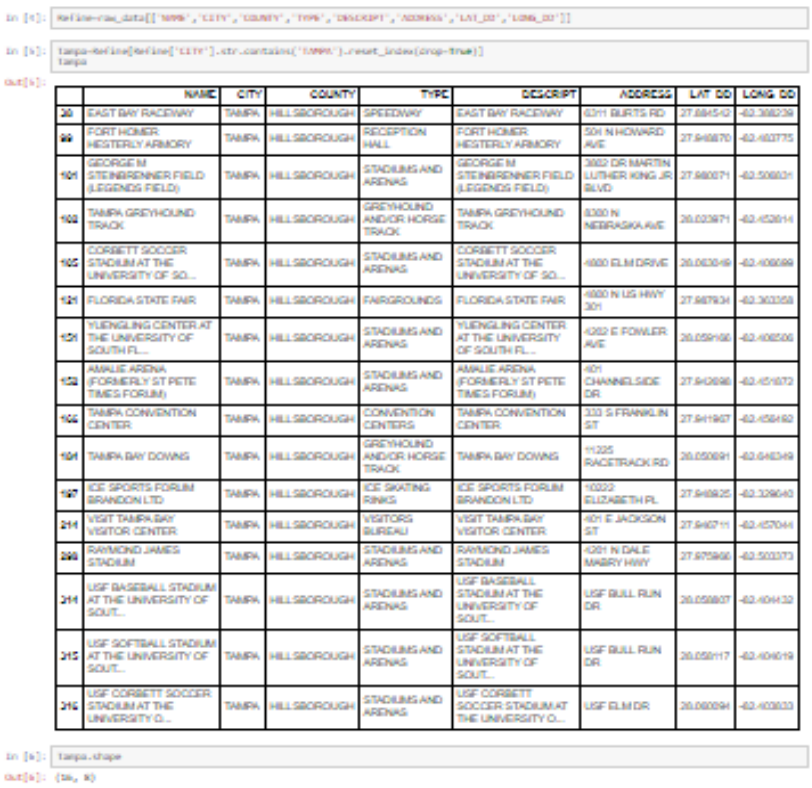
- In the first data set the data comprises of the entire information of Florida consisting of 355 rows and 27 columns.
- Since, our area of interest is only Tampa let's refine it by using column selection technique and see the results brought down to Tampa. Which has 16 rows and 8 columns.

```
In [2]: raw_data=pd.read_csv('https://opendata.arcgis.com/datasets/[illegible].csv')
        raw_data
Out[2]:
```

	FID	PARCELID	NAME	ADDRESS	CITY	ZIPCODE	COUNTY	TYPE	OWNER	OPER
0	1	01323100500010	AMERICAN AIRLINES ARENA	601 BISCAYNE BLVD	MIAMI	33132	MIAMI-DADE	STADIUMS AND ARENAS	MIAMI-DADE COUNTY	COUNT
1	2	74434321170010000	PALM BEACH COUNTY CONVENTION CENTER	650 OKEECHOBEE BLVD	WEST PALM BEACH	33401	PALM BEACH	CONVENTION CENTERS	PALM BEACH COUNTY	COUNT
2	3	1079230010030	HOMESTEAD MIAMI SPEEDWAY	1 SPEEDWAY BLVD	HOMESTEAD	33035	MIAMI-DADE	SPEEDWAY	CITY OF HOMESTEAD	MUNICI
3	4	1 32 43 01 A00 0001.0000	HENDRY COUNTY MOTOR SPORTS PARK (SUGARLAND MOT...	9985 W US HWY 27	CLEWISTON	33440	HENDRY	SPEEDWAY	HENDRY COUNTY SPEEDWAY LLC	PRIVAT
4	5	514010024902	PINES ICE ARENA	12425 TAFT ST	PEMBROKE PINES	33028	BROWARD	ICE SKATING RINKS	FLORIDA ICE ARENAS INC	PRIVAT
...	...	...	...	...	...	...	...	...	...	...
350	351	134625000000010010	SOCCER COMPLEX AT FLORIDA GULF COAST UNIVERSITY	FGCU LAKE PKWY E	FORT MYERS	33965	LEE	STADIUMS AND ARENAS	FLORIDA GULF COAST UNIVERSITY	STATE
351	352	134625000000010010	TENNIS COMPLEX AT FLORIDA GULF COAST UNIVERSITY	FGCU LAKE PKWY E	FORT MYERS	33965	LEE	STADIUMS AND ARENAS	FLORIDA GULF COAST UNIVERSITY	STATE
352	353	3040070000050	FIU BASEBALL STADIUM (INFINITY INSURANCE PARK)...	SW 17TH ST	MIAMI	33174	MIAMI-DADE	STADIUMS AND ARENAS	FLORIDA INTERNATIONAL UNIVERSITY	STATE
353	354	0622220000020	BISCAYNE BAY AQUATIC CENTER AT FLORIDA INTERNA...	3000 NE 151ST ST	NORTH MIAMI	33181	MIAMI-DADE	STADIUMS AND ARENAS	FLORIDA INTERNATIONAL UNIVERSITY	STATE
354	355	4109208030000	MORCOM AQUATIC CENTER AT FLORIDA STATE UNIVERSITY	2500 POTTSDAMER ST	TALLAHASSEE	32310	LEON	STADIUMS AND ARENAS	FLORIDA STATE UNIVERSITY	STATE

355 rows × 27 columns

1.1 Before Refining



	NAME	CITY	COUNTY	TYPE	DESCRIPT	ADDRESS	LAT_DD	LONG_DD
20	EAST BAY RACEWAY	TAMPA	HILLSBOROUGH	SPEEDWAY	EAST BAY RACEWAY	6311 BURTS RD	27.884542	-82.398239
99	FORT HOMER HESTERLY ARMORY	TAMPA	HILLSBOROUGH	RECEPTION HALL	FORT HOMER HESTERLY ARMORY	501 N HOWARD AVE	27.946870	-82.463775
101	GEORGE M STEINBRENNER FIELD (LEGENDS FIELD)	TAMPA	HILLSBOROUGH	STADIUMS AND ARENAS	GEORGE M STEINBRENNER FIELD (LEGENDS FIELD)	3802 DR MARTIN LUTHER KING JR BLVD	27.980071	-82.506831
102	TAMPA GREYHOUND TRACK	TAMPA	HILLSBOROUGH	GREYHOUND AND/OR HORSE TRACK	TAMPA GREYHOUND TRACK	8300 N NEBRASKA AVE	28.023971	-82.452814
103	CORBETT SOCCER STADIUM AT THE UNIVERSITY OF SO...	TAMPA	HILLSBOROUGH	STADIUMS AND ARENAS	CORBETT SOCCER STADIUM AT THE UNIVERSITY OF SO...	4000 ELM DRIVE	28.063049	-82.406099
121	FLORIDA STATE FAIR	TAMPA	HILLSBOROUGH	FAIRGROUNDS	FLORIDA STATE FAIR	4800 N US HWY 301	27.987834	-82.363358
124	YUENGLING CENTER AT THE UNIVERSITY OF SOUTH FL...	TAMPA	HILLSBOROUGH	STADIUMS AND ARENAS	YUENGLING CENTER AT THE UNIVERSITY OF SOUTH FL...	4202 E FOWLER AVE	28.059106	-82.406506
132	AMALIE ARENA (FORMERLY ST PETE TIMES FORUM)	TAMPA	HILLSBOROUGH	STADIUMS AND ARENAS	AMALIE ARENA (FORMERLY ST PETE TIMES FORUM)	401 CHANNELSIDE DR	27.942696	-82.451872
162	TAMPA CONVENTION CENTER	TAMPA	HILLSBOROUGH	CONVENTION CENTERS	TAMPA CONVENTION CENTER	333 S FRANKLIN ST	27.941967	-82.458162
164	TAMPA BAY DOWNS	TAMPA	HILLSBOROUGH	GREYHOUND AND/OR HORSE TRACK	TAMPA BAY DOWNS	11225 RACETRACK RD	28.050891	-82.640349
197	ICE SPORTS FORUM BRANDON LTD	TAMPA	HILLSBOROUGH	ICE SKATING RINKS	ICE SPORTS FORUM BRANDON LTD	10222 ELIZABETH PL	27.949825	-82.329640
211	VISIT TAMPA BAY VISITOR CENTER	TAMPA	HILLSBOROUGH	VISITORS BUREAU	VISIT TAMPA BAY VISITOR CENTER	401 E JACKSON ST	27.946711	-82.457044
212	RAYMOND JAMES STADIUM	TAMPA	HILLSBOROUGH	STADIUMS AND ARENAS	RAYMOND JAMES STADIUM	4201 N DALE MABRY HWY	27.975968	-82.503373
214	USF BASEBALL STADIUM AT THE UNIVERSITY OF SOUT...	TAMPA	HILLSBOROUGH	STADIUMS AND ARENAS	USF BASEBALL STADIUM AT THE UNIVERSITY OF SOUT...	USF BULL RUN DR	28.058837	-82.404132
215	USF SOFTBALL STADIUM AT THE UNIVERSITY OF SOUT...	TAMPA	HILLSBOROUGH	STADIUMS AND ARENAS	USF SOFTBALL STADIUM AT THE UNIVERSITY OF SOUT...	USF BULL RUN DR	28.058117	-82.404019
216	USF CORBETT SOCCER STADIUM AT THE UNIVERSITY O...	TAMPA	HILLSBOROUGH	STADIUMS AND ARENAS	USF CORBETT SOCCER STADIUM AT THE UNIVERSITY O...	USF ELM DR	28.060094	-82.403033

1.2 After refining

1.2 SECOND DATASET

Even though this data set had a lot of information it didn't actually deal with what exactly was needed so the Foursquare API's venues API calls were made to collect the actual data required

Out[14]:

	name	categories	lat	lng
0	Marshall Student Center (MSC)	Student Center	28.064070	-82.413415
1	Campus Recreation Center	College Gym	28.060172	-82.407672
2	Yuengling Center	College Basketball Court	28.059502	-82.406510
3	USF Tampa Bookstore	College Bookstore	28.063426	-82.412561
4	Chick-fil-A	Food Service	28.063330	-82.413839
5	Publix USF	Grocery Store	28.068375	-82.412075
6	Juniper Dining	College Cafeteria	28.060057	-82.418344
7	Pollo Tropical	College Cafeteria	28.060245	-82.409885
8	Champion's Choice	College Cafeteria	28.059905	-82.407270
9	USF Music Concert Hall	Concert Hall	28.064872	-82.417933
10	USF Theatre Center	Theater	28.064030	-82.414485
11	Felicitous Coffee & Tea House	Coffee Shop	28.055086	-82.400154
12	USF Library	College Library	28.059823	-82.412274
13	Starbucks	Coffee Shop	28.067079	-82.426604
14	USF Track and Field Complex	College Track	28.062679	-82.406174
15	Jamba Juice	Juice Bar	28.063744	-82.413311
16	Wawa	Convenience Store	28.053240	-82.426880

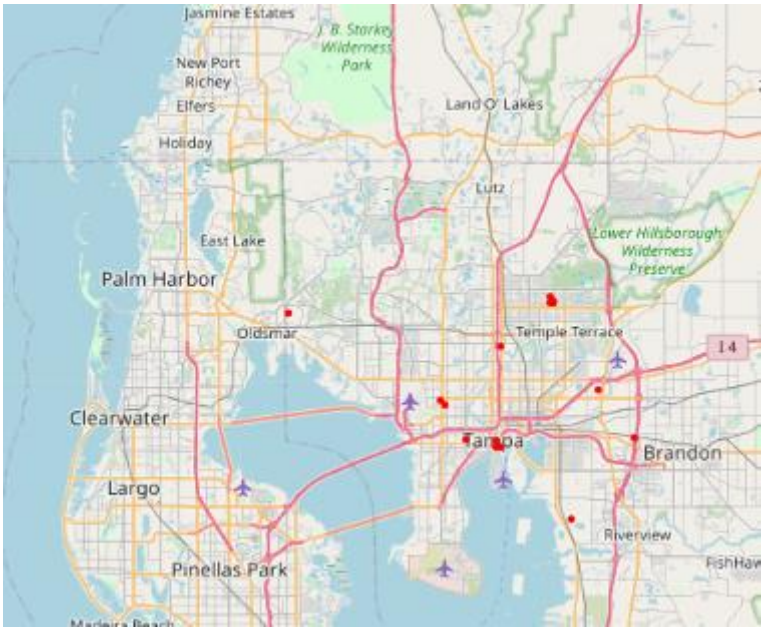
1.3 Foursquare API Data

METHODOLOGY

This section deals with the various methods involved to carryout the exploratory data analysis on the above two data sets

2.1 FIRST DATA SET

The first data set had a pretty much straight forward approach since it had all the necessary data priorly and the only thing left to do was to just plot those places on the map, which is as follows:



2.1 Tampa Neighbourhood

2.2 SECOND DATA SET

The second data set was obtained from Foursquare API servers by making venues API calls, since it's retrieved as a json file it has to be converted into a data frame format first. After this step comes the **Exploratory data analysis** where all the data in this data set has been encoded using one-hot encoding where all the values will be either 0 or 1, then are grouped together based on the frequency calculated and are then further segmented into venues with the most common appearing first and so on.

Out[16]:

	name	American Restaurant	Arcade	Asian Restaurant	BBQ Joint	Bagel Shop	Bakery	Bar	Big Box Store	Bookstore	...	Smoke Shop	Smoothie Shop	Student Center	Sushi Restaurant	Tha Restaurar
0	Marshall Student Center (MSC)	0	0	0	0	0	0	0	0	0	...	0	0	1	0	0
1	Campus Recreation Center	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0
2	Yuengling Center	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0
3	USF Tampa Bookstore	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0
4	Chick-fil-A	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0

5 rows × 71 columns

2.2 One-hot encoding

In [18]:
```
USF_grouped = USF_onehot.groupby('name').mean().reset_index()
USF_grouped
```

Out[18]:

	name	American Restaurant	Arcade	Asian Restaurant	BBQ Joint	Bagel Shop	Bakery	Bar	Big Box Store	Bookstore	...	Smoke Shop	Smoothie Shop	Student Center	Sushi Restaurant	Resta
0	ABC Fine Wine & Spirits	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0
1	ALDI	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0
2	Anarchist Closet	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0
3	Bagels Plus	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0
4	Ben & Jerry's	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0
5	Blaze Pizza	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0
6	Blue Lizard Hookah Lounge	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0
7	Bobacup	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0
8	Burgerfi	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0
9	Burlington	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0

2.3 Grouping

```
neighborhoods_venues_sorted
```

Out[22]:

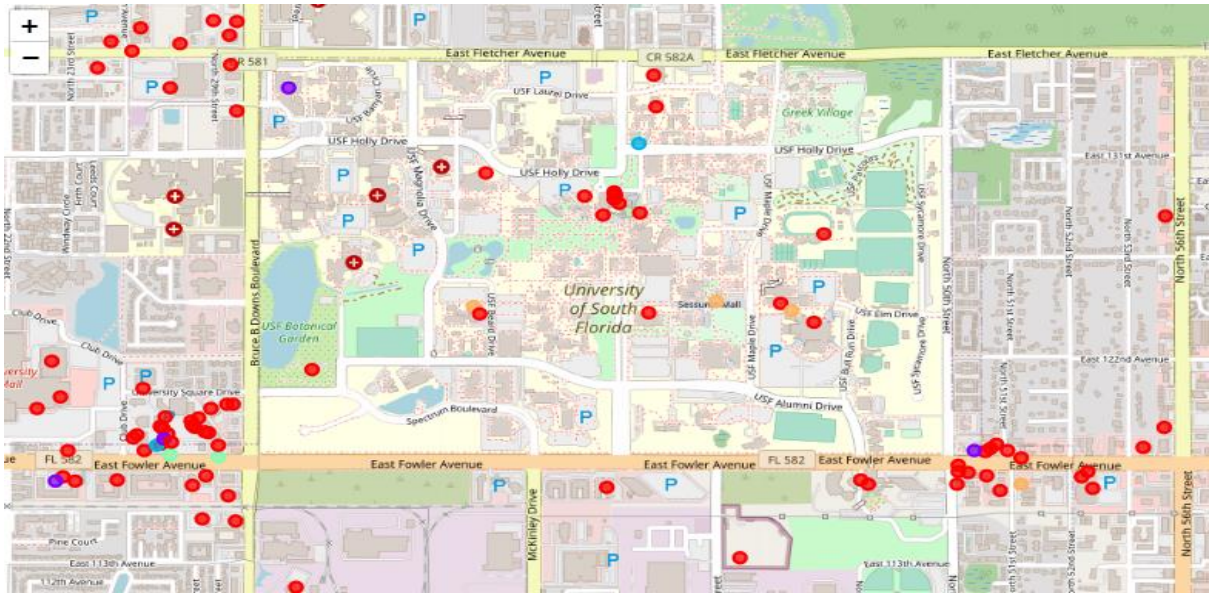
	name	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	ABC Fine Wine & Spirits	Liquor Store	Women's Store	College Gym	College Library	College Track	Concert Hall	Convenience Store	Cosmetics Shop	Department Store	College Bookstore
1	ALDI	Grocery Store	Department Store	College Gym	College Library	College Track	Concert Hall	Convenience Store	Cosmetics Shop	Discount Store	College Bookstore
2	Anarchist Closet	Women's Store	Department Store	College Gym	College Library	College Track	Concert Hall	Convenience Store	Cosmetics Shop	Discount Store	College Bookstore
3	Bagels Plus	Bagel Shop	Women's Store	Department Store	College Gym	College Library	College Track	Concert Hall	Convenience Store	Cosmetics Shop	Discount Store
4	Ben & Jerry's	Ice Cream Shop	Women's Store	Department Store	College Gym	College Library	College Track	Concert Hall	Convenience Store	Cosmetics Shop	Discount Store
5	Blaze Pizza	Pizza Place	Women's Store	Cosmetics Shop	College Cafeteria	College Gym	College Library	College Track	Concert Hall	Convenience Store	Department Store
6	Blue Lizard Hookah Lounge	Hookah Bar	Women's Store	Department Store	College Gym	College Library	College Track	Concert Hall	Convenience Store	Cosmetics Shop	Discount Store
7	Bobacup	Bubble Tea Shop	Women's Store	College Bookstore	College Gym	College Library	College Track	Concert Hall	Convenience Store	Cosmetics Shop	Department Store
8	Burgerfi	Burger Joint	Women's Store	College Bookstore	College Gym	College Library	College Track	Concert Hall	Convenience Store	Cosmetics Shop	Department Store
9	Burlington	Department	Women's	College	College	College	Concert	Convenience	Cosmetics	Discount	College

2.4 Grouping based on common venues

After all these steps of working on with data the data set is further subjected to clustering which actually filters and classifies all these data into different categories which we will further see in results section.

RESULTS

The data set will now be subjected to clustering using k-means clustering method where k is the number of clusters to be formed here in our data set the algorithm has classified the data set into 5 clusters which is as follows:



3.1 Clustered venues

Now let’s have a deeper look into each cluster:

Out[27]:

	categories	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	Student Center	Student Center	Women's Store	Cosmetics Shop	College Cafeteria	College Gym	College Library	College Track	Concert Hall	Convenience Store	Department Store
1	College Gym	College Gym	Women's Store	Department Store	College Library	College Track	Concert Hall	Convenience Store	Cosmetics Shop	Discount Store	College Bookstore
2	College Basketball Court	College Basketball Court	Department Store	College Gym	College Library	College Track	Concert Hall	Convenience Store	Cosmetics Shop	Discount Store	College Bookstore
3	College Bookstore	College Bookstore	Women's Store	Department Store	College Gym	College Library	College Track	Concert Hall	Convenience Store	Cosmetics Shop	Discount Store
4	Food Service	Fast Food Restaurant	Food Service	Women's Store	Cosmetics Shop	College Gym	College Library	College Track	Concert Hall	Convenience Store	Discount Store
5	Grocery Store	Grocery Store	Department Store	College Gym	College Library	College Track	Concert Hall	Convenience Store	Cosmetics Shop	Discount Store	College Bookstore
9	Concert Hall	Concert Hall	Women's Store	Department Store	College Gym	College Library	College Track	Convenience Store	Cosmetics Shop	Discount Store	College Bookstore
10	Theater	Theater	Women's Store	Cosmetics Shop	College Cafeteria	College Gym	College Library	College Track	Concert Hall	Convenience Store	Department Store

3.2 Cluster 1

In [28]:
```
USF_merged.loc[USF_merged['Cluster Labels'] == 1, USF_merged.columns[[1] + list(range(5, USF_merged.shape[1]))]]
```

Out[28]:

	categories	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
20	Sandwich Place	Sandwich Place	Women's Store	College Basketball Court	College Cafeteria	College Gym	College Library	College Track	Concert Hall	Convenience Store	Cosmetics Shop
24	Sandwich Place	Sandwich Place	Women's Store	College Basketball Court	College Cafeteria	College Gym	College Library	College Track	Concert Hall	Convenience Store	Cosmetics Shop
76	Sandwich Place	Sandwich Place	Women's Store	College Basketball Court	College Cafeteria	College Gym	College Library	College Track	Concert Hall	Convenience Store	Cosmetics Shop
82	Sandwich Place	Sandwich Place	Women's Store	College Basketball Court	College Cafeteria	College Gym	College Library	College Track	Concert Hall	Convenience Store	Cosmetics Shop

3.3 Cluster 2

Cluster 3

In [29]:
```
USF_merged.loc[USF_merged['Cluster Labels'] == 2, USF_merged.columns[[1] + list(range(5, USF_merged.shape[1]))]]
```

Out[29]:

	categories	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
38	Burger Joint	Burger Joint	Women's Store	College Bookstore	College Gym	College Library	College Track	Concert Hall	Convenience Store	Cosmetics Shop	Department Store
41	Burger Joint	Burger Joint	Women's Store	College Bookstore	College Gym	College Library	College Track	Concert Hall	Convenience Store	Cosmetics Shop	Department Store

3.4 Cluster 3

Cluster 4

```
In [30]: USF_merged.loc[USF_merged['Cluster Labels'] == 3, USF_merged.columns[[1] + list(range(5, USF_merged.shape[1]))]]
```

Out[30]:

	categories	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	Restaurant	Restaurant	Women's Store	College Basketball Court	College Cafeteria	College Gym	College Library	College Track	Concert Hall	Convenience Store	Cosmetics Shop
21	Restaurant	Restaurant	Women's Store	College Basketball Court	College Cafeteria	College Gym	College Library	College Track	Concert Hall	Convenience Store	Cosmetics Shop

3.5 Cluster 4

Cluster 5

```
In [31]: USF_merged.loc[USF_merged['Cluster Labels'] == 4, USF_merged.columns[[1] + list(range(5, USF_merged.shape[1]))]]
```

Out[31]:

	categories	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
6	College Cafeteria	College Cafeteria	Department Store	College Gym	College Library	College Track	Concert Hall	Convenience Store	Cosmetics Shop	Women's Store	College Bookstore
7	College Cafeteria	Latin American Restaurant	College Cafeteria	Food Service	Food Court	Flea Market	Fast Food Restaurant	Donut Shop	Discount Store	College Bookstore	Department Store
8	College Cafeteria	College Cafeteria	Department Store	College Gym	College Library	College Track	Concert Hall	Convenience Store	Cosmetics Shop	Women's Store	College Bookstore
85	Latin American Restaurant	Latin American Restaurant	College Cafeteria	Food Service	Food Court	Flea Market	Fast Food Restaurant	Donut Shop	Discount Store	College Bookstore	Department Store

3.6 Cluster 5

DISCUSSION

Hence, Clustering has helped us a lot to differentiate different places around campus into many categories such as:

1.Student Centers

2.Sandwich place

3.Burger joint

4.Restuarant

5.College Cafeteria

CONCLUSION

Therefore with the help of data science techniques we were able to explore the Tampa neighbourhood which can be a life saver for new comers on this locality.