APPLIED DATA SCIENCE CAPSTONE

THE BATTLE OF NEIGHBORHOODS

INTRODUCTION

- What makes some businesses successful and others not so much?
- The location and the venues around may have an impact on the customer base
- A Businessman is looking to open a fine dining south asian restaurant in the flourishing community of Surrey, British Columbia
- He needs to know what kind of restaurants are already popular in this community
- To solve this Business Problem, we need a dataset of the restaurants that are present in this community
- We need to use Foursquare api to find the popular venues around the restaurants.

DATA ACQUISITION AND CLEANING

- The data we need is a dataset of restaurants operating in this community of Surrey, British Columbia
- City of Surrey open data

https://data.surrey.ca/dataset/restaurants

- Foursquare APi to make requests

https://foursquare.com

- The Dataset had seven columns and 1346 rows

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	TRACKINGNUMBE	NAME	PHYSICALADDRESS	PHYSICALCITY	FACTYPE	LATITUDE	LONGITUDE								
2	SWOD-APSP3X	104 Sushi & Co.	10422 168 St	Surrey	Restaurant	49.19206	-122.756								
3	SDFO-88ESVK	5 Star Catering	5640 188 St	Surrey	Restaurant	49.10545	-122.701								
4	NDAA-8RNNVR	555 Pizza Ltd	9192 120 St	Surrey	Restaurant	49.17027	-122.89								
5	SYOG-5M5942	7-Eleven #26365	7986 120 St	Surrey	Restaurant	49.14819	-122.89								
6	GEDS-6KHW57	7-Eleven #26517	17638 60 Ave	Surrey	Restaurant	49.11132	-122.734								
7	GEDS-7YZN48	7-Eleven #34339	14410 108 Ave	Surrey	Restaurant	49.1986	-122.823								
8	SHEN-9SGNQT	7-Eleven #34501	6021 168 St	Surrey	Restaurant	49.11179	-122.757								
9	SDFO-8AULF9	7-Eleven #34633	15961 Fraser Hwy	Surrey	Restaurant	49.15789	-122.779								
10	SDFO-8L8SZK	7-Eleven #34907	13993 Fraser Hwy	Surrey	Restaurant	49.17928	-122.835								
11	ACAK-8XKMAF	7-Eleven #35770	6828 128 St	Surrey	Restaurant	49.12696	-122.868								
12	NDAA-9L2T78	7-Eleven #36022H	19188 72 Ave	Surrey	Restaurant	49.13347	-122.692								
13	HPAR-AE2S2K	7-Eleven #37876H	7195 King George B	Surrey	Restaurant	49.13306	-122.846								
14	SPLH-ASNP2T	7-Eleven #38206	19161 Fraser Hwy	Surrey	Restaurant	49.11961	-122.692								
15	SHEN-A2CTME	7-Eleven (Canada Store #372	9989 152 St	Surrey	Restaurant	49.18393	-122.801								
16	HPAR-AFGRCA	7-Eleven (Canada Store #378	14445 64 Ave	Surrey	Restaurant	49.11931	-122.823								
17	HPAR-AFGPRP	7-Eleven (Canada Store #378	10376 152 St	Surrey	Restaurant	49.19107	-122.801								
18	NDAA-99WMTF	7-Eleven Canada Store #366	9583 132 St	Surrey	Restaurant	49.17682	-122.857								
19	HPAR-AE2Q87	7-Eleven Store # 37868	6422 120 St	Surrey	Restaurant	49.11967	-122.89								
20	SWOD-AQAKLU	7-ElevenStore and Gas Bar ((	10416 King George	Surrey	Restaurant	49.19205	-122.845								
21	A016730	North Surrey Secondary Sch	15945 96 Ave	Surrey	Restaurant	49.17707	-122.78								
22	A018238	Northview Golf & Country Cl	6857 168 St	Surrey	Restaurant	49.12597	-122.758								

- The attributes like the ‘tracking number’ are irrelevant to our project and hence was dropped
- The attribute city had the same value ‘Surrey’ for all entries and the attribute ‘Type’ had the same value ‘Restaurant’ for all the entries and hence these are irrelevant to our project as well

```
In [11]: df.head()
```

Out[11]:

	NAME	ADDRESS	CITY	LATITUDE	LONGITUDE
0	104 Sushi & Co.	10422 168 St	Surrey	49.192059	-122.756256
1	5 Star Catering	5640 188 St	Surrey	49.105449	-122.701315
2	555 Pizza Ltd	9192 120 St	Surrey	49.170274	-122.889988
3	7-Eleven #26365	7986 120 St	Surrey	49.148189	-122.890221
4	7-Eleven #26517	17638 60 Ave	Surrey	49.111322	-122.734271

METHODOLOGY

- We start by exploring the first row of our dataframe or the first restaurant in the dataset
- We need to know what other venues are popular around this particular restaurant
- We need Foursquare developer access to have credentials like the client id and client secret
- We proceed to retrieve the nearby venues for the rest of the restaurants in the dataset

- The nearby venues for the first restaurant in our dataset

```
nearby_venues1.columns = [col.split(".")[-1] for col in nearby_venues1.columns]

nearby_venues1.head()
```

Out[43]:

	name	categories	lat	lng
0	The Clayton Public House	Pub	49.104655	-122.701060
1	Sunrise Golf Centre	Golf Course	49.105024	-122.701071
2	Mavericks Taphouse And Grill	Pub	49.104966	-122.701028

The above cell shows that there are three popular venues near the first restaurant ' 104 Sushi and Co' . There are two pubs and a golf course. Probably not a good idea to open a restaurant near noisy unpleasant pubs

- We proceed to retrieve the nearby venues for the rest of the restaurants in the dataset

RESULTS

- The resulting clusters have five clusters each clustering the community based on the nearby venues
- The first cluster is all the restaurants around yoga clubs and sandwich places. The restaurants around fish markets are grouped in another cluster
- . Looking at the clusters, it is evident that it's a good idea to open the restaurant around any restaurant in the first cluster as they are in a busy and happening community around yoga centers and coffee shops

Cluster 1

This cluster is a cluster of yoga studios, stores, pizza and sandwich places, juice bars but mostly inclues the fish market nearby. What an useful insight !

```
In [66]: surrey_venues_sorted.loc[surrey_venues_sorted['Cluster Labels'] == 0, surrey_venues_sorted.columns[[1] + list(range(5, :
```

109	Blenz Coffee (Central City South)	Pizza Place	Greek Restaurant	Breakfast Spot	Supermarket	Restaurant	Event Service	Filipino Restaurant
115	Boiling Point	Chinese Restaurant	Yoga Studio	Filipino Restaurant	Fast Food Restaurant	Farmers Market	Falafel Restaurant	Fabric Shop
116	Bolivar Heights Gas Bar	Fish Market	Fish & Chips Shop	Filipino Restaurant	Fast Food Restaurant	Farmers Market	Falafel Restaurant	Fabric Shop
119	Bon Ga Korean Restaurant	Yoga Studio	Event Service	Fish Market	Fish & Chips Shop	Filipino Restaurant	Fast Food Restaurant	Farmers Market
123	Booster Juice #232	Pizza Place	Greek Restaurant	Breakfast Spot	Supermarket	Restaurant	Event Service	Filipino Restaurant
130	Boston Pizza #59	Sandwich Place	Salon / Barbershop	Korean Restaurant	Middle Eastern Restaurant	Bank	Market	Donut Shop

DISCUSSION

- If we were to cluster only based on the restaurants present in the community, it would have yielded a result that has not much useful insights
- The use of foursquare api to retrieve the venues nearby has boosted our search for the perfect spot
- This gives us an insight into the customer base that's likely to visit each community.

CONCLUSION

- The results of the clustering algorithm proves that the first cluster which is grouped around yoga clubs and cafes would prove to be a better choice for the businessman
- I conclude this project with the observation that the restaurant would be successful if opened around the pizza places in this community of Nordel way, fraser way and 100 avenue.