IBM DATA SCIENCE PROFESSIONAL CERTIFICATE CAPSTONE

ARNAB PUSHILAL – 2020

TOPIC: Locating the best type of Fried chicken spot to open around Delhi

INTRODUCTION:

The inspiration behind this problem is pretty simple. I'm in love with fried chicken. So I kind of thought of a way to incorporate this final report in and around that.

The idea is to find a few fried chicken spots and around town and classify them with whatever parameters I can find.

BUSINESS PROBLEM:

Identifying critical parameters to open a new friend chicken spot in Delhi.

Data Used:

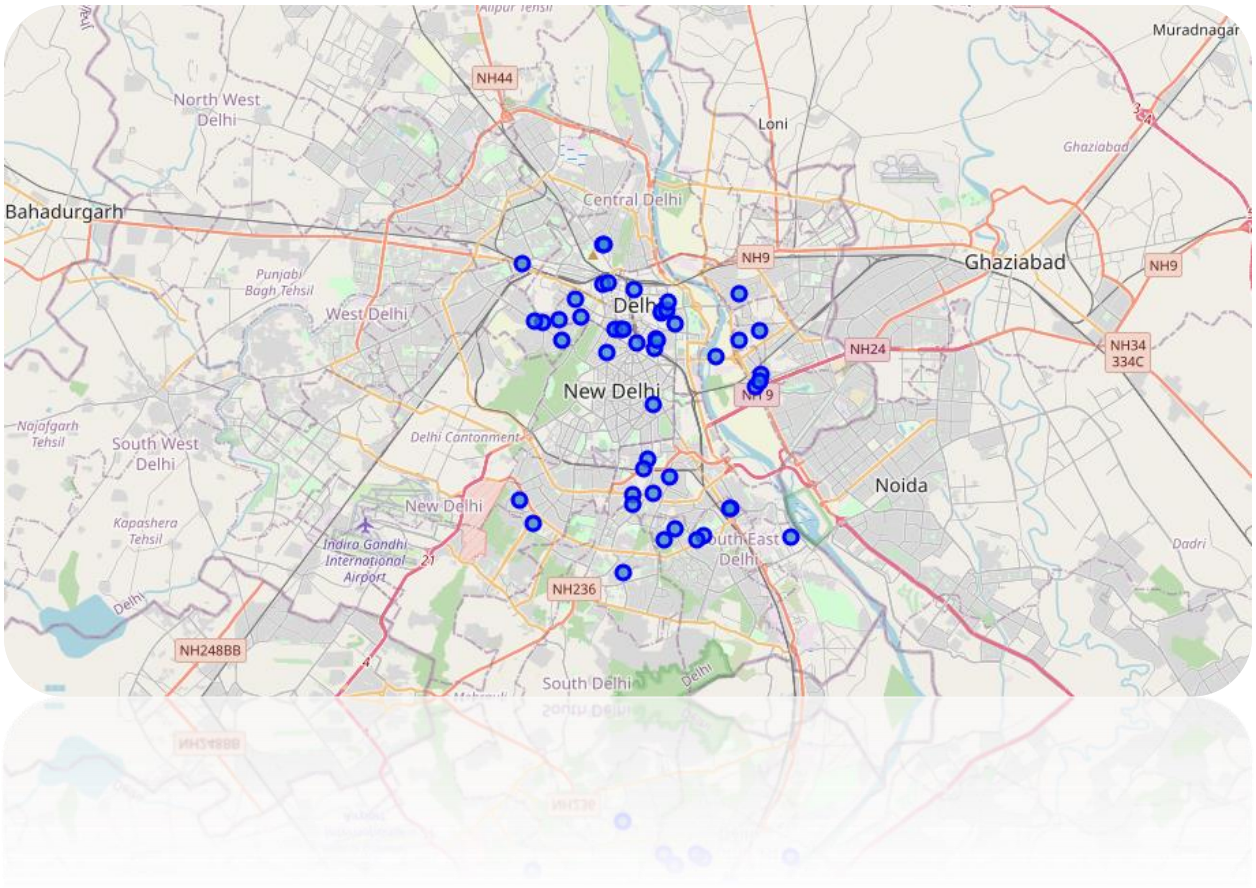
The Data is retrieved from Foursquare API for locations and names of the places [constraint in the project, the project could have been done with the help of Zomato API itself]. The other parameters like Price, Average rating from users are obtained from the ZOMATO API (a popular food delivery service in India).

Methodology:

Initially, I queried ‘fried chicken’ and obtained results from the foursquare api in and around Delhi-NCR. The radius was kept kind of large to get enough results for a conclusion. Relevant results were stored on a Data Frame. With the latitude & longitude of these places, markers were plotted on a map using folium.

ut[17]:

	name	location.lat	location.lng	Category
0	Md. Hussain Fried Chicken	28.658860	77.220394	Indian Restaurant
1	Kentucy Fried Chicken	28.566110	77.162663	Fast Food Restaurant
2	KFC (Kentucky's Fried Chicken)	28.555874	77.169290	Burger Joint
3	Kentucky Fried Chicken	28.548447	77.235174	Fast Food Restaurant
4	Arabian Fried Chicken	28.549245	77.299158	Fast Food Restaurant
5	Chicken Inn	28.607891	77.229752	Indian Restaurant
6	Classic Chicken	28.631204	77.206324	Fried Chicken Joint
7	Punjabi Chicken	28.635285	77.221800	Indian Restaurant
8	Ilyas Mast Butter Chicken	28.632982	77.230461	BBQ Joint
9	Mitra Da Chicken	28.641410	77.213382	Restaurant
10	Sonu Chicken Corner	28.640913	77.210664	Fried Chicken Joint
11	Sonu Chicken Corner	28.641256	77.214803	Indian Restaurant
12	Star chicken point	28.636901	77.231037	null
13	Ilyas mast butter Chicken	28.636597	77.231492	Indian Restaurant
14	punjabi chicken corner	28.636541	77.184021	Snack Place
15	Republic Of Chicken	28.583693	77.226723	Fried Chicken Joint
16	Republic Of Chicken	28.579972	77.224716	Fried Chicken Joint
17	Chawla Chicken	28.646664	77.193129	Fish & Chips Shop
18	Aslam Chicken Corner	28.648283	77.233648	Indian Restaurant
19	Repulic Of Chicken	28.645309	77.182078	Steakhouse
20	Chicken Plaza	28.648880	77.233676	Indian Restaurant



Next Data from Zomato API is collected using the same latitude longitudes of the obtained results from the foursquare API

Out[21]:

	name1	location.lat1	location.lng1	AvgCostfor2	PriceRange	AverageUserRating	Address
0	Haji Mohd. Hussain	28.6489571542	77.2335048392	300	1	4.0	113, Matia Mahal Road, Bazaar Matia Mahal, Jam...
1	AFC - Arabian Fried Chicken	28.6459340000	77.1739030000	250	1	3.4	29/1, East Patel Nagar, New Delhi
2	KFC	28.6299526213	77.2201360390	450	1	3.9	6 & 7, Scindia House, Outer Circle, Connaught ...
3	AFC - Arabian Fried Chicken	28.6459340000	77.1739030000	250	1	3.4	29/1, East Patel Nagar, New Delhi
4	Arabian Fried Chicken	28.5492298696	77.2992223129	300	1	3.9	A112/2, Ground Gloor, Shaheen Bagh, Jamia, Jas...
5	Chicken Inn	28.6080283245	77.2297360003	1500	3	4.2	13-15, Pandara Road Market, New Delhi
6	Classic Chicken	28.6338741605	77.2062177211	650	2	3.7	8/4, Doctor Lane, Gole Market, New Delhi
7	Kinbuck 2	28.6321597190	77.2166951001	1500	3	4.1	P-10/90, 1st & 2nd Floor, Outer Circle, Connau...
8	Butter Chicken Company	28.5503766808	77.2351517156	500	2	3.9	M-55, 2nd Floor, M Block Market, Greater Kaila...
9	Mitra Di Chaap	28.6041611587	77.2923233360	400	1	4.0	189, Pratap Nagar, Near Mayur Vihar Phase 1, N...
10	Sonu Parantha Corner	28.5844097600	77.3143066200	200	1	3.4	C-61, Sector -2, Noida
11	Sonu Parantha Corner	28.5844097600	77.3143066200	200	1	3.4	C-61, Sector -2, Noida
12	Aqua - The Park	28.6288649465	77.2163078561	4500	4	4.2	The Park, 15, Parliament Street, Connaught Pla...
13	Butter Chicken Company	28.5503766808	77.2351517156	500	2	3.9	M-55, 2nd Floor, M Block Market, Greater Kaila...

Here we can see there are a few duplicates with the latitude and longtitudes that we have provided to the api. Hence we need to remove them. I do this based on address.

Both Datas are merged. Time to get the features ready to be shoved into the model!

```
:  DataMerged
```

9]:

	AvgCostfor2	PriceRange	AverageUserRating	name	location.lat	location.lng	Category
0	300	1	4.0	Md. Hussain Fried Chicken	28.658860	77.220394	Indian Restaurant
1	250	1	3.4	Kentucy Fried Chicken	28.566110	77.162663	Fast Food Restaurant
2	450	1	3.9	KFC (Kentucky's Fried Chicken)	28.555874	77.169290	Burger Joint
3	300	1	3.9	Arabian Fried Chicken	28.549245	77.299158	Fast Food Restaurant
4	1500	3	4.2	Chicken Inn	28.607891	77.229752	Indian Restaurant
5	650	2	3.7	Classic Chicken	28.631204	77.206324	Fried Chicken Joint
6	1500	3	4.1	Punjabi Chicken	28.635285	77.221800	Indian Restaurant
7	500	2	3.9	Ilyas Mast Butter Chicken	28.632982	77.230461	BBQ Joint
8	400	1	4.0	Mitra Da Chicken	28.641410	77.213382	Restaurant
9	200	1	3.4	Sonu Chicken Corner	28.640913	77.210664	Fried Chicken Joint
10	4500	4	4.2	Star chicken point	28.636901	77.231037	null
11	400	1	3.8	punjabi chicken corner	28.636541	77.184021	Snack Place

‘Category’ being a categorical variable(duh!) is one-hot-encoded into labels.

Name is dropped as it is of no use to us as far as features go.

Let's see some graphs!(2)

First I get the data frame grouped by Category

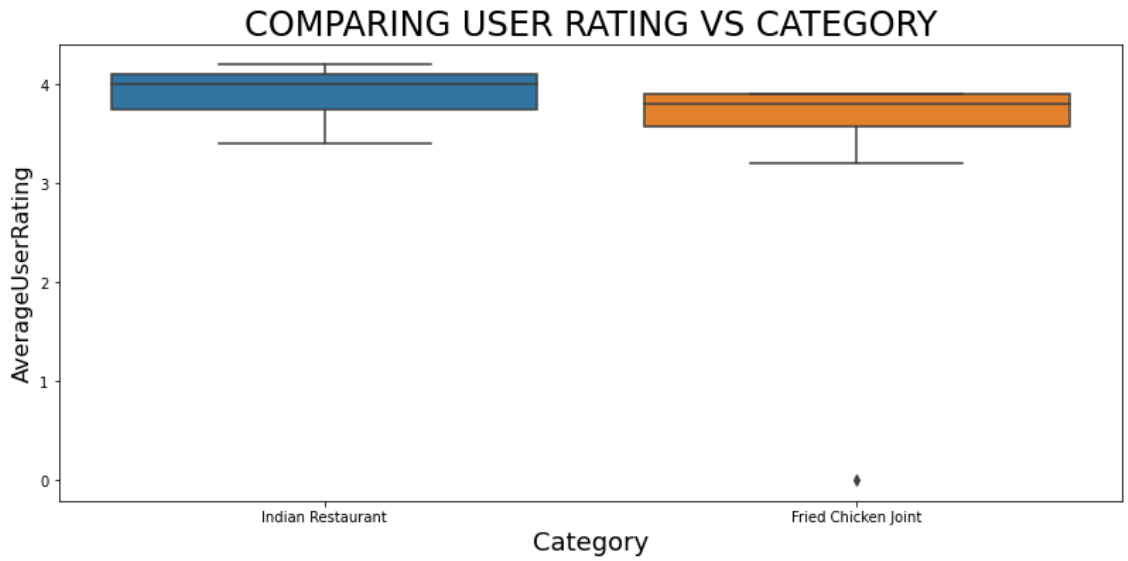
```
Category
Afghan Restaurant       1
Asian Restaurant        1
BBQ Joint               2
Burger Joint            1
Chinese Restaurant      1
Fast Food Restaurant    2
Fish & Chips Shop       1
Food Truck              1
Fried Chicken Joint     8
Indian Restaurant       7
Restaurant              1
Snack Place             1
null                    3
Name: Category, dtype: int64
```

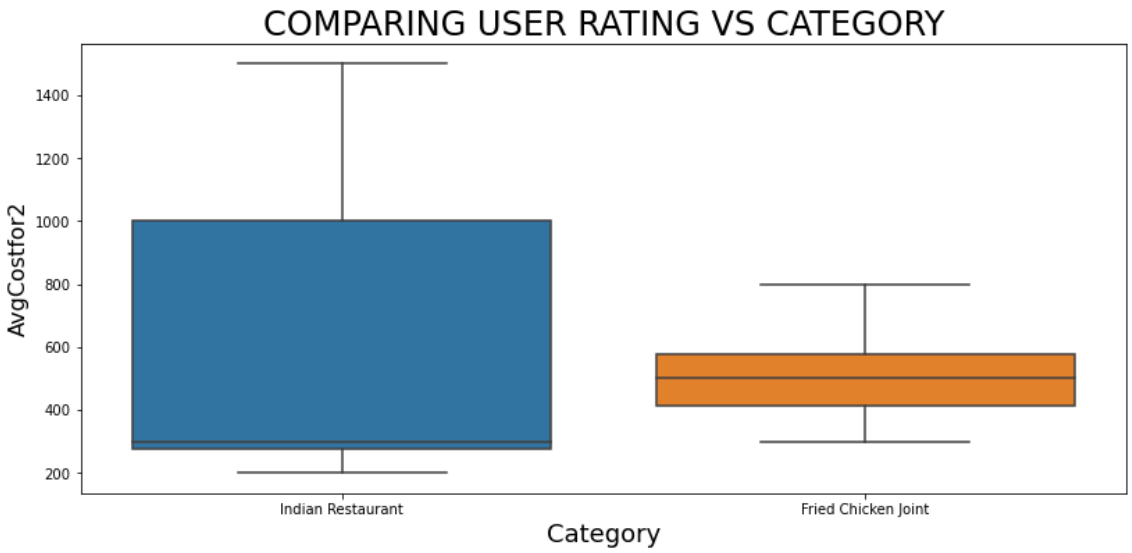
I choose the ones which are greater > 2 and ignore null

```
b=[]
d=[]
DataMerged['Bool']=0
c=a.index
for i,j in zip(a,c):
    if (i >2 ) and (j != 'null'):
        b.append(j)
for i,k in enumerate(DataMerged['Category']) :
    for j in b:
        if k == j :
            DataMerged['Bool'].iloc[i] =True

for k,i in enumerate(DataMerged['Bool']):
    if i == 0:
        DataMerged['Bool'].iloc[k] = False
```

I index it as True or false based on whether they are >2 + not null and I use these features in my boxplot!





On an initial glance it seems opening an ‘Indian restaurant Category is better since it has a higher ratings and cost [in terms of range] is also higher! In the Indian Restaurant Category, we see above in the figure that the Q3 is quite high!

7]:

Category	AvgCostfor2	PriceRange	AverageUserRating
Afghan Restaurant	400.000000	1.000000	3.600
Asian Restaurant	400.000000	1.000000	3.700
BBQ Joint	575.000000	2.000000	3.950
Burger Joint	450.000000	1.000000	3.900
Chinese Restaurant	450.000000	1.000000	3.500
Fast Food Restaurant	275.000000	1.000000	3.650
Fish & Chips Shop	700.000000	2.000000	3.200
Food Truck	3000.000000	4.000000	4.000
Fried Chicken Joint	506.250000	1.625000	3.275
Indian Restaurant	650.000000	1.714286	3.900
Restaurant	400.000000	1.000000	4.000
Snack Place	400.000000	1.000000	3.800
null	1883.333333	2.333333	4.100

Now all that's left is that we min-max scale the data (since K-means is based on Euclidean Distance , we don't want any features contributing more than they should!) and feed them into the Kmeans from sklearn.

Voila we have our labels!

RESULTS

DF WITH LABELS

	AvgCostfor2	PriceRange	AverageUserRating	name	location.lat	location.lng	Category	label
0	300	1	4.0	Md. Hussain Fried Chicken	28.658860	77.220394	Indian Restaurant	2
1	250	1	3.4	Kentucy Fried Chicken	28.566110	77.162663	Fast Food Restaurant	0
2	450	1	3.9	KFC (Kentucky's Fried Chicken)	28.555874	77.169290	Burger Joint	0
3	300	1	3.9	Arabian Fried Chicken	28.549245	77.299158	Fast Food Restaurant	0
4	1500	3	4.2	Chicken Inn	28.607891	77.229752	Indian Restaurant	2
5	650	2	3.7	Classic Chicken	28.631204	77.206324	Fried Chicken Joint	1
6	1500	3	4.1	Punjabi Chicken	28.635285	77.221800	Indian Restaurant	2
7	500	2	3.9	Ilyas Mast Butter Chicken	28.632982	77.230461	BBQ Joint	0
8	400	1	4.0	Mitra Da Chicken	28.641410	77.213282	Restaurant	0

GROUPING BY LABELS TO SEE SOME PATTERN

Out[36]:

label	AverageUserRating	AvgCostfor2	PriceRange
0	3.275000	506.250000	1.625000
1	3.875000	618.750000	1.625000
2	3.846667	923.333333	1.733333

1) 0--> MEDIUM USER RATING BUT LOW PRICE RANGE

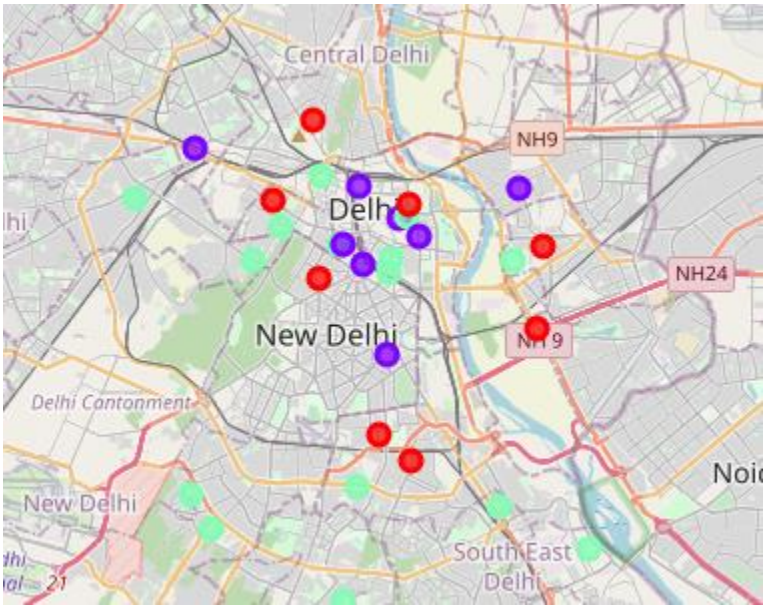
2) 1--> MEDIUM PRICE RANGE AND LOW USER RATING

3) 2--> HIGH COST + HIGH RATING

GROUPING LABELS BY CATEGORY

```
Category              label
Afghan Restaurant     2        1
BBQ Joint             2        2
Burger Joint          2        1
Chinese Restaurant    2        1
Fast Food Restaurant  2        2
Fish & Chips Shop     2        1
Food Truck            2        1
Fried Chicken Joint   0        8
Indian Restaurant     1        8
Restaurant            2        2
Snack Place           2        1
null                  2        3
Name: label, dtype: int64
```

CLUSTER MAP OF THE RESTURANTS:



Green-2

Red- 0

Violet -1

DISCUSSION

Label 0 seems like the best option if you have a lower budget. (Fried chicken spots)

1 has the lowest rating. So no point being in this category. (Indian Restaurants)

We can see clusters of Green (2) [Food truck, BBQ joint etc.] around southeast and central delhi – being posh places, where people like to spend more. So if you have budget you should go for these locations.

CONCLUSION

Several parameters of the problem were analyzed and discussed successfully, and the question where and what type of fried nchicken Restaurant should you open was answered.