

Coursera Capstone

IBM Data Science Capstone

Segregating New Delhi Restaurants based on Location and Rating:



By: Venkata Seetharam

Pendekanti

Introduction:

For many connoisseurs , it will be helpful to locate a restaurant with a good rating and near by location in any city.I have chosen to segregate restaurants in New Delhi city based on location,ratings, cuisine type.

Business Problem:

To choose a restaurant based on location, rating and cuisine type. By proper selection of a restaurant which will save enormous amount of commuting time and money for a customer and at the same time rewarding good restaurants. This will be a win-win situation for both customers and restaurant owners.ie all stake holders will be benefited .This will also help to fill lacuna ie if particular cuisine type restaurant with appropriate rating is not available in that particular locality, any prospective investor or an Entrepreneur can start a new venture in that locality.Also to detect any pattern formation in clusters based on cuisine type ,locality and ratings.

Target Audience of this project:

This project is particularly useful to visitors from other parts of the country and for foreign visitors to locate appropriate restaurant suitable to their tastes. Also as mentioned above will be useful for future hoteliers.

Data

To solve the problem, we will need the following data:

First using Foursquare API, list of restaurants in New Delhi are obtained . This data gives restaurant's Name,Address,Longitude,Latitude,Cuisine,etc.This data is supplemented by Zomato data set which is published in Kaggle Competitions. This data set has extensive coverage of Restaurants in New Delhi with many details.

Link : <https://www.kaggle.com/shrutimehta/zomato-restaurants-data>

FileHomeInsertPage LayoutFormulasDataReviewViewHelp

Cut

Copy

Paste

Format Painter

Clipboard

Calibri

11

A

A

Font

Paragraph

Alignment

Number

Styles

Cells

Editing

Views

Autosum

Find & Select

Share

Comments

General

Conditional Formatting

Format as Table

Cell Styles

Insert

Delete

Format

Autosum

Find & Select

Share

Comments

Clipboard

Font

Paragraph

Alignment

Number

Styles

Cells

Editing

Views

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V								
1	Restaurant	Restaurant	Country	Ci	Address	Locality	Locality	in	Longitude	Latitude	Cuisines	Average	Currency	Is	Table	booking	Has	Online	Is	delivery	Switch	to	price	range	Aggregate	Rating	cat	Rating	text	Votes
2	18287358	Food	Close	1	New Delhi Aya	Naga	Aaya	Naga	Aaya	Naga	0	0	Cuisine	Va	500	Indian	Rup	No		No	No		2	0	White		Not	rated		2
3	18216944	Burger	in	1	New Delhi	84	Near	H	Adichini	Adichini	77.19692	28.53538	Fast	Food	350	Indian	Rup	No	Yes	No	No		1	3.2	Orange		Average		46	
4	1131333	Days	of	in	1	New Delhi	81/3	1st	F	Adichini	Adichini	77.19747	28.53549	North	Indi	1500	Indian	Rup	Yes	Yes	No	No		3	3.4	Orange		Average		46
5	53884127	Deli	Ka	Oh	1	New Delhi	66/A	Grou	Adichini	Adichini	77.19805	28.53755	South	Indi	500	Indian	Rup	No	No	No	No		2	2.6	Orange		Average		11	
6	582	Governor		1	New Delhi	84	Adja	Adichini	Adichini	77.19692	28.53552	South	Indi	500	Indian	Rup	Yes	No	No	No		2	3.4	Orange		Average		238		
7	18416405	Mrebaan		1	New Delhi	A- 96	Sri	Adichini	Adichini	77.19812	28.53813	Mughlai		400	Indian	Rup	No	No	No	No		3	3.1	Orange		Average		8		
8	304243	Say	Chesi	1	New Delhi	88/3	Adch	Adichini	Adichini	77.19816	28.53745	South	Indi	400	Indian	Rup	Yes	No	No	No		3	2.7	Orange		Average		64		
9	3554	Scusa		1	New Delhi	88/4	Adch	Adichini	Adichini	77.19795	28.53747	South	Indi	450	Indian	Rup	Yes	No	No	No		3	2.6	Orange		Average		113		
10	18169673	Monozot		1	New Delhi	Sri	Aerobiz	Adichini	Adichini	77.19804	28.53839	Pizza		300	Indian	Rup	Yes	No	No	No		3	2.7	Yellow		Good		66		
11	948	Waves		1	New Delhi	A-4	Servo	Adichini	Adichini	77.19881	28.53867	North	Indi	1500	Indian	Rup	Yes	Yes	No	No		3	3.5	Yellow		Good		141		
12	2853	Delhi	Darb	1	New Delhi	84	Main	R	Adichini	Adichini	77.19697	28.53676	Chinese	N	500	Indian	Rup	No	No	No	No		2	2.2	Red		Poor		77	
13	18433900	Chouteu		1	New Delhi	84	Aurobi	Adichini	Adichini	77.19679	28.53596	Mediterr		1600	Indian	Rup	Yes	No	No	No		3	4.2	Green		Very	Good		43	
14	18425159	Nariyal	Ca	1	New Delhi	A-4	3rd	F	Adichini	Adichini	77.19915	28.53844	Continents		1200	Indian	Rup	Yes	No	No	No		3	4.2	Green		Very	Good		46
15	110058	Ranton	o	1	New Delhi	94-A/11	Ac	Adichini	Adichini	77.19816	28.53739	Parsi		1500	Indian	Rup	Yes	Yes	No	No		3	4.2	Green		Very	Good		965	
16	18314443	BorShale		1	New Delhi	Aditya	Me	Aditya	Me	Aditya	Me	77.30063	28.60575	Finger	Foo	600	Indian	Rup	No	No	No		3	3.2	Orange		Average		31	
17	18255341	Chavala	o	1	New Delhi	Shop	5G	Kai	Me	Aditya	Me	77.3012	28.60605	North	Indi	800	Indian	Rup	Yes	No	No		2	3.4	Orange		Average		32	
18	187	Dominic	o	1	New Delhi	15/14/16	Aditya	Me	Aditya	Me	77.30163	28.60645	Pizza	Fast	700	Indian	Rup	No	No	No		3	2.8	Orange		Average		83		
19	308235	Sufatar		1	New Delhi	33-9C	Grc	Aditya	Me	Aditya	Me	77.3016	28.60626	North	Indi	1000	Indian	Rup	Yes	No	No		3	3.7	Yellow		Good		288	
20	18258757	Bella	Italia	1	New Delhi	Ground	Fx	Aerocity	Aerocity	77.0879	28.55446	Fast Food		550	Indian	Rup	No	No	No	No		2	3.2	Orange		Average		33		
21	307383	Clever	Fox	1	New Delhi	Red	Fox	Aerocity	Aerocity	77.12064	28.55169	North	Indi	900	Indian	Rup	No	No	No	No		2	3.2	Orange		Average		26		
22	18186746	165	Natal	1	New Delhi	Multilevel	Aerocity	Aerocity	77.0879	28.55446	Continents		600	Indian	Rup	Yes	No	No	No		2	0	White		Not	rated		2		
23	18317895	Tpot		1	New Delhi	Basement	Aerocity	Aerocity	77.0879	28.55446	Cafe		800	Indian	Rup	No	No	No	No		2	0	White		Not	rated		0		
24	110491	on	4	64	Re	1	New Delhi	112	Aggar	Aggarwal	77.13482	28.60014	Street	Foo	1000	Indian	Rup	Yes	No	No		3	2.6	Orange		Average		49		
25	18337789	Romfing		1	New Delhi	Shop	31	A	Aggarwal	Aggarwal	77.13456	28.60001	Street	Foo	200	Indian	Rup	No	No	No		3	3.3	Orange		Average		13		
26	18225371	Selkie	Lox	1	New Delhi	214-215	2	Aggarwal	Aggarwal	77.13477	28.60899	North	Indi	1500	Indian	Rup	Yes	No	No	No		3	3.2	Orange		Average		64		
27	3370	BTW		1	New Delhi	G-71	6	Aggarwal	Aggarwal	77.11779	28.70013	Street	Foo	400	Indian	Rup	No	No	No	No		3	3.4	Orange		Average		167		
28	307439	Chiff	Press	1	New Delhi	G-89	50	A	Aggarwal	Aggarwal	77.11764	28.70064	North	Indi	900	Indian	Rup	Yes	No	No	No		2	3.3	Orange		Average		99	
29	3509	Pizza	Hut	1	New Delhi	17	Mange	Aggarwal	Aggarwal	77.11726	28.70399	Italian	Piz	5000	Indian	Rup	Yes	No	No	No		3	2.6	Orange		Average		125		

napozmate2

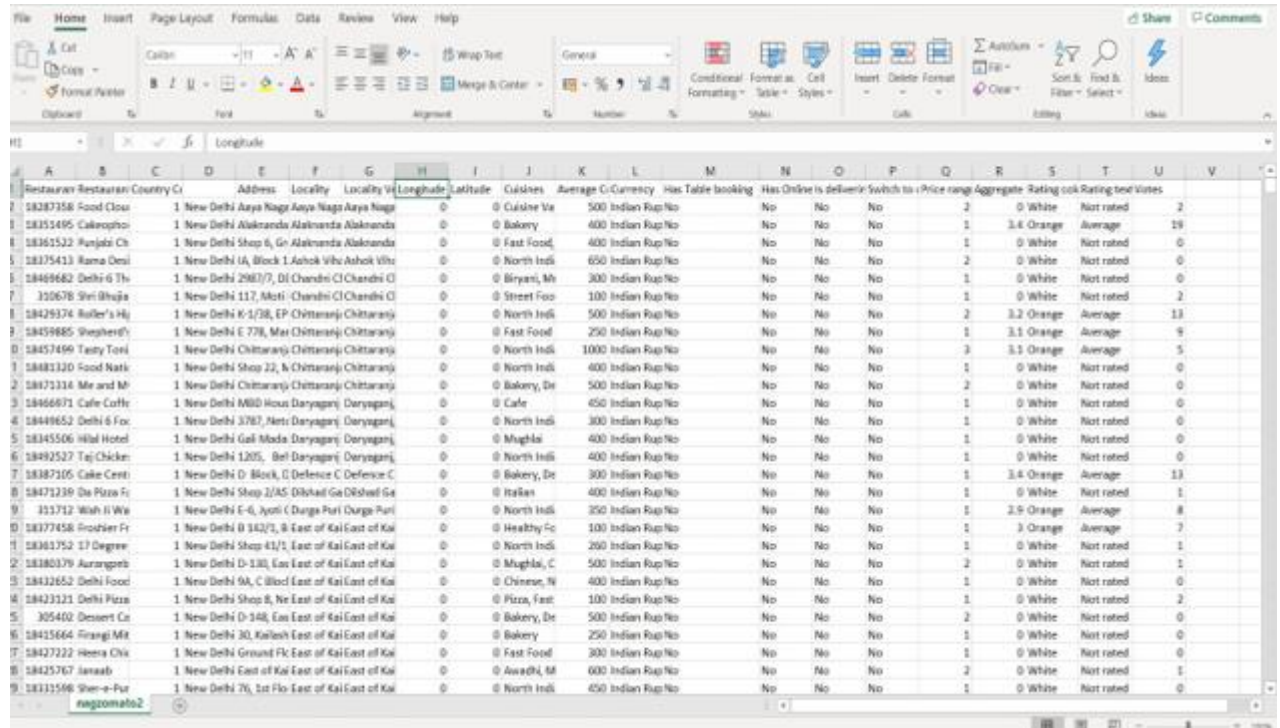
Data Cleaning:

The data downloaded from Kaggle data set contained a lot of null values and longitude and latitude column had few zero values. These rows were deleted and appropriate changes were made into dataset.

Some rating values were not available so, these values were replaced with average ratings of remaining rows

There were more than 50 locations in original Kaggle data set, to make it manageable, only 18 large locations were chosen. The criteria for selecting locality was it should have at least 65 restaurants registered with Zomato. Appropriate changes were made to Foursquare dataset to make it compatible with Kaggle-zomato data set.

Some of duplicate values were also deleted



The rows with longitude,latitude zeros are shown below in original dataset

Restaurant	Country	Address	Locality	Locality V	Longitude	Latitude	Cuisines	Average C	Currency	Has Table booking	Has Online is deliver	Switch to	Price range	Aggregate	Rating	cook	Rating	best	votes
18287358 Food Close	1	New Delhi Aya Naga Aya Naga Aya Naga			0	0	Cuisine Va	500	Indian Rup	No	No	No	2	0	White	Not rated	2		
18251495 Cakeshop	1	New Delhi Alaknanda Alaknanda Alaknanda			0	0	Bakery	400	Indian Rup	No	No	No	1	3.4	Orange	Average	19		
18365522 Punjabi Ch	1	New Delhi Shop 6, G Alaknanda Alaknanda			0	0	Fast Food	400	Indian Rup	No	No	No	1	0	White	Not rated	0		
18375413 Rama Desi	1	New Delhi 1A, Block 1 Ashok Vihar Ashok Vihar			0	0	North Indi	650	Indian Rup	No	No	No	2	0	White	Not rated	0		
18469682 Delhi 6 Th	1	New Delhi 2987/7, D1 Chandni Ch Chandni Ch			0	0	Biryani, Mh	300	Indian Rup	No	No	No	1	0	White	Not rated	0		
330678 Shri Bhajia	1	New Delhi 117, Mori Chandni Ch Chandni Ch			0	0	Street Foo	100	Indian Rup	No	No	No	1	0	White	Not rated	2		
18429174 Ruler's H	1	New Delhi K-1/38, EP Chitrananj Chitrananj			0	0	North Indi	500	Indian Rup	No	No	No	2	3.2	Orange	Average	13		
18459885 Shepherd's	1	New Delhi E 778, Mar Chitrananj Chitrananj			0	0	Fast Food	250	Indian Rup	No	No	No	1	3.1	Orange	Average	9		
18457499 Tasty Tots	1	New Delhi Chitrananj Chitrananj Chitrananj			0	0	North Indi	1000	Indian Rup	No	No	No	1	3.1	Orange	Average	5		
18481320 Food Nati	1	New Delhi Shop 22, h Chitrananj Chitrananj			0	0	North Indi	400	Indian Rup	No	No	No	1	0	White	Not rated	0		
18471114 Me and M	1	New Delhi Chitrananj Chitrananj Chitrananj			0	0	Bakery, De	500	Indian Rup	No	No	No	2	0	White	Not rated	0		
18466971 Cafe Caffe	1	New Delhi MID Hous Daryaganj Daryaganj			0	0	Cafe	450	Indian Rup	No	No	No	1	0	White	Not rated	0		
18449652 Delhi 6 Fox	1	New Delhi 3787, Net Daryaganj Daryaganj			0	0	North Indi	300	Indian Rup	No	No	No	1	0	White	Not rated	0		
18345506 Hill Hotel	1	New Delhi Gul Moha Daryaganj Daryaganj			0	0	Mughlai	400	Indian Rup	No	No	No	1	0	White	Not rated	0		
18492527 Taj Chicken	1	New Delhi 1205, Bel Daryaganj Daryaganj			0	0	North Indi	400	Indian Rup	No	No	No	1	0	White	Not rated	0		
18387105 Cake Cent	1	New Delhi D Block, D Defence C Defence C			0	0	Bakery, De	300	Indian Rup	No	No	No	1	3.4	Orange	Average	13		
18475239 De Pizza Fi	1	New Delhi Shop 2/AS, Dohad Ga Dohad Ga			0	0	Italian	400	Indian Rup	No	No	No	1	0	White	Not rated	1		
351712 Wala Ji Wa	1	New Delhi E-6, Jyoti C Durga Pur Durga Pur			0	0	North Indi	350	Indian Rup	No	No	No	1	2.9	Orange	Average	8		
18377458 Frodler Fi	1	New Delhi H 142/1, B East of Kai East of Kai			0	0	Healthy Fo	100	Indian Rup	No	No	No	1	3	Orange	Average	7		
18361752 17 Degree	1	New Delhi Shop 41/1, East of Kai East of Kai			0	0	North Indi	260	Indian Rup	No	No	No	1	0	White	Not rated	1		
18380139 Aurangzeb	1	New Delhi D-130, East East of Kai East of Kai			0	0	Mughlai, C	500	Indian Rup	No	No	No	2	0	White	Not rated	1		
18432652 Delhi Food	1	New Delhi 9A, C Bloc East of Kai East of Kai			0	0	Chinese, W	400	Indian Rup	No	No	No	1	0	White	Not rated	0		
18423121 Delhi Pizza	1	New Delhi Shop 8, Ne East of Kai East of Kai			0	0	Pizza, Fast	100	Indian Rup	No	No	No	1	0	White	Not rated	2		
305402 Dessert Ce	1	New Delhi D-148, East East of Kai East of Kai			0	0	Bakery, De	500	Indian Rup	No	No	No	2	0	White	Not rated	0		
18415604 Firangi Mi	1	New Delhi 30, Kalash East of Kai East of Kai			0	0	Bakery	250	Indian Rup	No	No	No	1	0	White	Not rated	0		
18427222 Heena Chi	1	New Delhi Ground Fl East of Kai East of Kai			0	0	Fast Food	300	Indian Rup	No	No	No	1	0	White	Not rated	0		
18425767 Jataab	1	New Delhi East of Kai East of Kai East of Kai			0	0	Awadhi, M	600	Indian Rup	No	No	No	2	0	White	Not rated	1		
18315106 Sher-e-Pa	1	New Delhi 76, 1st Flo East of Kai East of Kai			0	0	North Indi	450	Indian Rup	No	No	No	1	0	White	Not rated	0		

Data combining:

Data obtained from both foursquare and Kaggle-zomato data set was combined into a new dataframe after cleaning the data as mentioned above.

Duplicate values were deleted and new data frame is obtained.

Types of Cuisines in Data set:

As the combined dataset(Foursquare,Kaggle) had around 1500 restaurant details in various parts of New Delhi,clustering methodology is employed to detect the patterns. This data set contained main cuisine types are:

- American
- Bakery Dessert
- Fast Food
- Others
- Beverages
- Cafe
- Cafe,Bakery,Others
- Chinese
- Continental
- Desserts
- Fast Food
- Ice Cream
- Italian
- Mix
- North Indian
- North Indian, Chinese
- South Indian
- Street Food
- Bar

Restaurant Locations

These Restaurants were spread over various localities of New Delhi and these Restaurants were assigned to Zomato

Restaurants located in following localities

Connaught Place
Rajouri Garden
Hauz Khas
Malviya Nagar
Safdarjung
Pitampura
Mayur Vihar Phase 1
Rajinder Nagar
Satyaniketan
Karol Bagh
Defence Colony
Mahipalpur
Mukherjee Nagar
Kalkaji
Krishna Nagar
Shahdara
Chandni Chowk
Delhi University-GTB Nagar

One Hot Encoding

One Hot encoding is done for cuisine types as they are non-numeric fields. Though there were many cuisine types, main cuisine types were found to be 19 in number as all other were combined into a category called "Others". Reason for inclusion cuisine type column is to examine or detect a pattern in a cluster so that one will come to know whether a pattern exist based on cuisine type in the cluster.

One Hot encoding is also done for Location(column name: Localities) as they are non-numeric fields. Though there were many Locations , main Locations were found to be 18 in number. The reason for inclusion of Localities column is to examine or detect a pattern in cluster based on Locality.

Cluster Analysis:

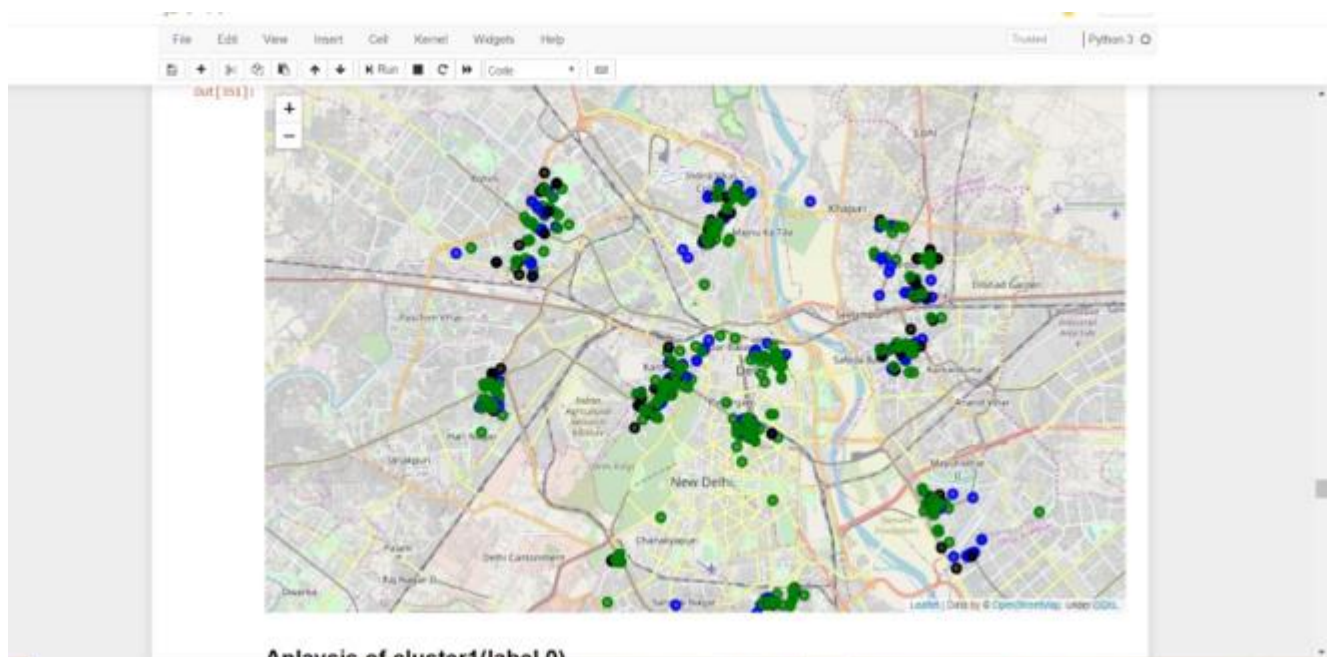
Cluster Analysis was performed on combined data frame. K-means clustering algorithm was deployed with number of clusters equals to three. Main reason for choosing number of clusters to 3, with this number, algorithm will take into account of cuisine type, locality and rating.

1000 iterations were performed with K-Means Clustering algorithm on the above data set to stabilise the centroids of clusters to give stable values.

Analysis of cluster1(label 0):

Algorithm formed first cluster with 896 records which are green in color in the image. cluster1 mostly consists of Restaurants which offer Fast food,Chinese,Café and bakery, Others cuisine types.Cluster1 also picked up most of locations from Chandini Chowk, Connaught Place, Defense colony,Delhi University as displayed in the Jupyter notebook.

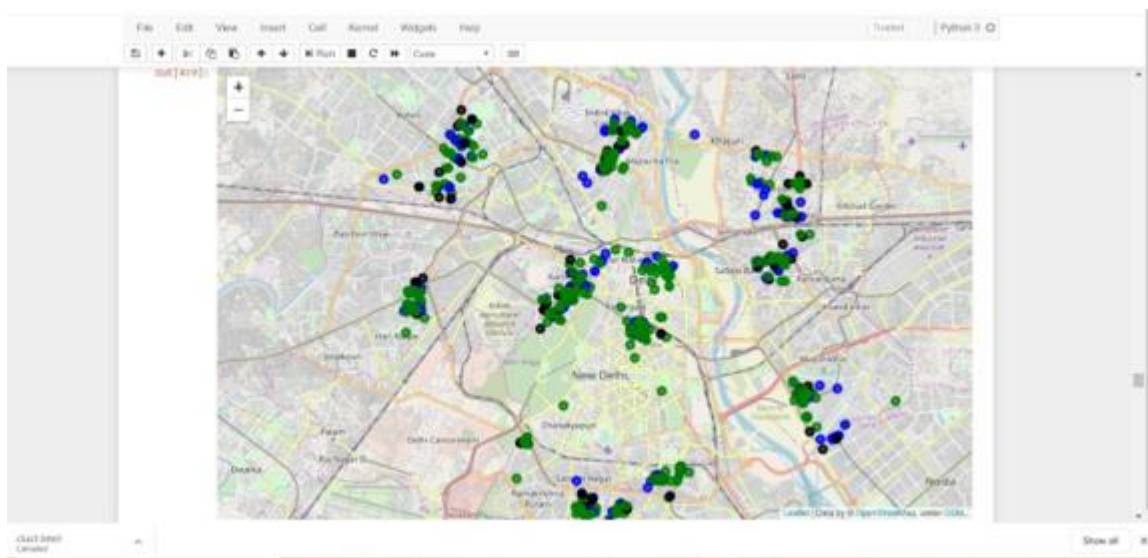
Algorithm picked up Restaurants in cluster 1 with ratings ranging from 2.0 to 4.9



Analysis of cluster 2(label1)

Algorithm formed second cluster with 306 records which are blue in color in the image . cluster2 consists of Restaurants which offer only North Indian .Cluster2 also picked up most of the locations with no location has a major presence as displayed in the Jupyter notebook.

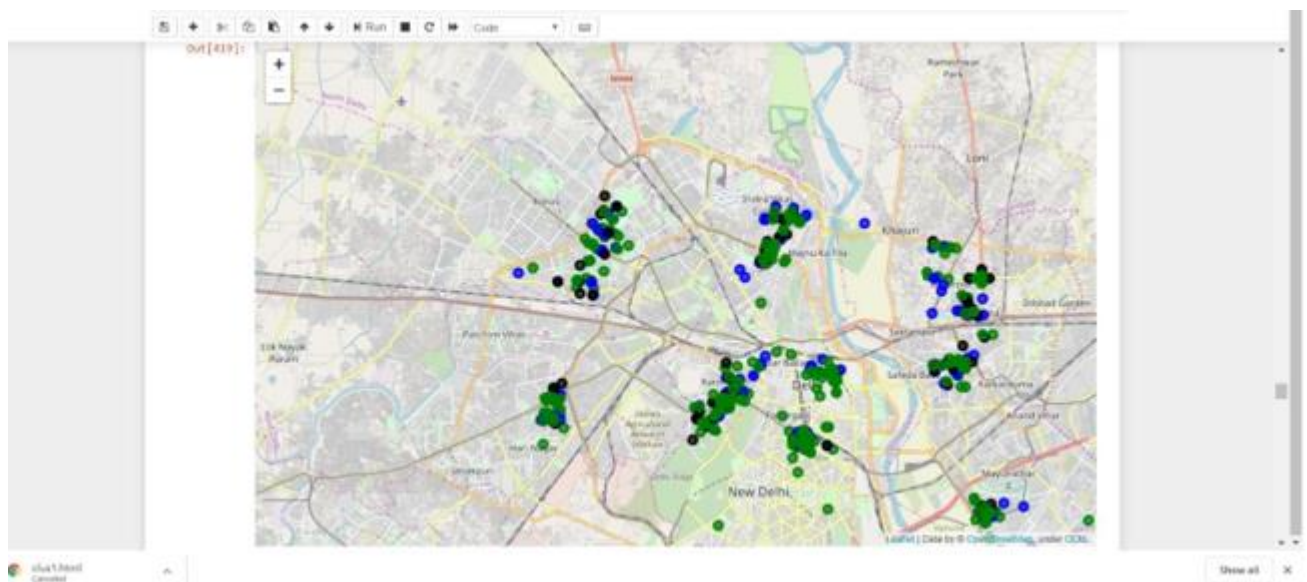
Algorithm picked up Restaurants in cluster 2 with ratings ranging from 2.4 to 4.8



Analysis of cluster 3(label2)

Algorithm formed third cluster with 210 records which are black in color in the image. Cluster3 consists of Restaurants which offer only Chinese and North Indian together. Cluster3 also picked up most of the locations with no location has a major presence as displayed in the Jupyter notebook.

Algorithm picked up Restaurants in cluster 1 with ratings ranging from 2.0 to 4.4



Conclusion:

K-means clustering algorithm in this particular case is giving more weightage to cuisine column as compared to locality and ratings columns while forming clusters.

As shown above cluster 2 only picked up restaurants which offer North Indian and cluster 3 only picked up

Restaurants which offer North Indian,Chinese combined. The clusters formed above are not location specific nor ratings specific.

