

## Coursera Capstone

### IBM Data Science Capstone

#### Segregating New Delhi Restaurants based on Location and Rating:



By: Nagarjuna Pendekanti

December 2019

## 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>

File

Home

Insert

Page Layout

Formulas

Data

Review

View

Help

Cut

Copy

Format Painter

Clipboard

Calibri

11

A

A

</

## 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 atleast 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

AutoSave

FileHomeInsertPage LayoutFormulasDataReviewViewHelp

Calibri11AaBWrap TextGeneralConditional FormattingFormat as TableCell StylesInsertDeleteFormatCellsEditingIdeas

ClipboardFontAlignmentNumberStylesCells

ShareComments

|    | A          | B           | C         | D | E                      | F            | G           | H         | I        | J           | K          | L          | M                 | N          | O            | P         | Q          | R         | S          | T           | U     | V |
|----|------------|-------------|-----------|---|------------------------|--------------|-------------|-----------|----------|-------------|------------|------------|-------------------|------------|--------------|-----------|------------|-----------|------------|-------------|-------|---|
| H1 |            |             |           |   | Longitude              |              |             |           |          |             |            |            |                   |            |              |           |            |           |            |             |       |   |
| 1  | Restaurant | Restaurant  | Country C |   | Address                | Locality     | Locality V  | Longitude | Latitude | Cuisines    | Average Ci | Currency   | Has Table booking | Has Online | Is deliverin | Switch to | Price rang | Aggregate | Rating col | Rating text | Votes |   |
| 2  | 18287358   | Food Clou   |           | 1 | New Delhi Aaya Naga    | Aaya Naga    | Aaya Naga   | 0         | 0        | Cuisine Va  | 500        | Indian Rup | No                | No         | No           | No        | 2          | 0         | White      | Not rated   | 2     |   |
| 3  | 18351495   | Cakeophou   |           | 1 | New Delhi Alaknanda    | Alaknanda    | Alaknanda   | 0         | 0        | Bakery      | 400        | Indian Rup | No                | No         | No           | No        | 1          | 3.4       | Orange     | Average     | 19    |   |
| 4  | 18361522   | Punjabi Ch  |           | 1 | New Delhi Shop 6, G    | ri Alaknanda | Alaknanda   | 0         | 0        | Fast Food   | 400        | Indian Rup | No                | No         | No           | No        | 1          | 0         | White      | Not rated   | 0     |   |
| 5  | 18375413   | Rama Desi   |           | 1 | New Delhi 1A, Block 1  | Ashok Vih    | Ashok Vih   | 0         | 0        | North Indi  | 650        | Indian Rup | No                | No         | No           | No        | 2          | 0         | White      | Not rated   | 0     |   |
| 6  | 18469682   | Delhi-6 Th  |           | 1 | New Delhi 2987/7, Df   | Chandni Cl   | Chandni Cl  | 0         | 0        | Biryani, Mi | 300        | Indian Rup | No                | No         | No           | No        | 1          | 0         | White      | Not rated   | 0     |   |
| 7  | 310678     | Shri Bhujia |           | 1 | New Delhi 117, Moti    | Chandni Cl   | Chandni Cl  | 0         | 0        | Street Foo  | 100        | Indian Rup | No                | No         | No           | No        | 1          | 0         | White      | Not rated   | 2     |   |
| 8  | 18429374   | Roller's Hq |           | 1 | New Delhi K-1/38, EP   | Chittaranji  | Chittaranji | 0         | 0        | North Indi  | 500        | Indian Rup | No                | No         | No           | No        | 2          | 3.2       | Orange     | Average     | 13    |   |
| 9  | 18459885   | Shepherd's  |           | 1 | New Delhi E 778, Mar   | Chittaranji  | Chittaranji | 0         | 0        | Fast Food   | 250        | Indian Rup | No                | No         | No           | No        | 1          | 3.1       | Orange     | Average     | 9     |   |
| 10 | 18457499   | Tasty Toni  |           | 1 | New Delhi Chittaranji  | Chittaranji  | Chittaranji | 0         | 0        | North Indi  | 1000       | Indian Rup | No                | No         | No           | No        | 3          | 3.1       | Orange     | Average     | 5     |   |
| 11 | 18481320   | Food Natic  |           | 1 | New Delhi Shop 22, N   | Chittaranji  | Chittaranji | 0         | 0        | North Indi  | 400        | Indian Rup | No                | No         | No           | No        | 1          | 0         | White      | Not rated   | 0     |   |
| 12 | 18471314   | Me and M    |           | 1 | New Delhi Chittaranji  | Chittaranji  | Chittaranji | 0         | 0        | Bakery, De  | 500        | Indian Rup | No                | No         | No           | No        | 2          | 0         | White      | Not rated   | 0     |   |
| 13 | 18466971   | Cafe Coffe  |           | 1 | New Delhi MBD Hous     | Daryaganj    | Daryaganj   | 0         | 0        | Cafe        | 450        | Indian Rup | No                | No         | No           | No        | 1          | 0         | White      | Not rated   | 0     |   |
| 14 | 18449652   | Delhi 6 Fox |           | 1 | New Delhi 3787, Netz   | Daryaganj    | Daryaganj   | 0         | 0        | North Indi  | 300        | Indian Rup | No                | No         | No           | No        | 1          | 0         | White      | Not rated   | 0     |   |
| 15 | 18345506   | Hilal Hotel |           | 1 | New Delhi Gali Mada    | Daryaganj    | Daryaganj   | 0         | 0        | Mughlai     | 400        | Indian Rup | No                | No         | No           | No        | 1          | 0         | White      | Not rated   | 0     |   |
| 16 | 18492527   | Taj Chicke  |           | 1 | New Delhi 1205, Belf   | Daryaganj    | Daryaganj   | 0         | 0        | North Indi  | 400        | Indian Rup | No                | No         | No           | No        | 1          | 0         | White      | Not rated   | 0     |   |
| 17 | 18387105   | Cake Cent   |           | 1 | New Delhi D Block, D   | Defence C    | Defence C   | 0         | 0        | Bakery, De  | 300        | Indian Rup | No                | No         | No           | No        | 1          | 3.4       | Orange     | Average     | 13    |   |
| 18 | 18471239   | Da Pizza F  |           | 1 | New Delhi Shop 2/AS    | Dilshad Ga   | Dilshad Ga  | 0         | 0        | Italian     | 400        | Indian Rup | No                | No         | No           | No        | 1          | 0         | White      | Not rated   | 1     |   |
| 19 | 311712     | Wah Ji Wa   |           | 1 | New Delhi E-6, Jyoti C | Durga Puri   | Durga Puri  | 0         | 0        | North Indi  | 350        | Indian Rup | No                | No         | No           | No        | 1          | 2.9       | Orange     | Average     | 8     |   |
| 20 | 18377458   | Froshier Fr |           | 1 | New Delhi B 142/1, B   | East of Kai  | East of Kai | 0         | 0        | Healthy Fc  | 100        | Indian Rup | No                | No         | No           | No        | 1          | 3         | Orange     | Average     | 7     |   |
| 21 | 18361752   | 17 Degree   |           | 1 | New Delhi Shop 41/1, E | ast of Kai   | East of Kai | 0         | 0        | North Indi  | 260        | Indian Rup | No                | No         | No           | No        | 1          | 0         | White      | Not rated   | 1     |   |
| 22 | 18380379   | Aurangzeb   |           | 1 | New Delhi D-130, Eas   | East of Kai  | East of Kai | 0         | 0        | Mughlai, C  | 500        | Indian Rup | No                | No         | No           | No        | 2          | 0         | White      | Not rated   | 1     |   |
| 23 | 18432652   | Delhi Food  |           | 1 | New Delhi 9A, C Bloc   | East of Kai  | East of Kai | 0         | 0        | Chinese, N  | 400        | Indian Rup | No                | No         | No           | No        | 1          | 0         | White      | Not rated   | 0     |   |
| 24 | 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           | No        | 1          | 0         | White      | Not rated   | 2     |   |
| 25 | 305402     | Dessert Ca  |           | 1 | New Delhi D-148, Eas   | East of Kai  | East of Kai | 0         | 0        | Bakery, De  | 500        | Indian Rup | No                | No         | No           | No        | 2          | 0         | White      | Not rated   | 0     |   |
| 26 | 18415664   | Firangi Mit |           | 1 | New Delhi 30, Kallash  | East of Kai  | East of Kai | 0         | 0        | Bakery      | 250        | Indian Rup | No                | No         | No           | No        | 1          | 0         | White      | Not rated   | 0     |   |
| 27 | 18427222   | Heera Chic  |           | 1 | New Delhi Ground Fk    | East of Kai  | East of Kai | 0         | 0        | Fast Food   | 300        | Indian Rup | No                | No         | No           | No        | 1          | 0         | White      | Not rated   | 0     |   |
| 28 | 18425767   | Janaab      |           | 1 | New Delhi East of Kai  | East of Kai  | East of Kai | 0         | 0        | Awadhi, M   | 600        | Indian Rup | No                | No         | No           | No        | 2          | 0         | White      | Not rated   | 1     |   |
| 29 | 18331598   | Sher-e-Pur  |           | 1 | New Delhi 76, 1st Flo  | East of Kai  | East of Kai | 0         | 0        | North Indi  | 450        | Indian Rup | No                | No         | No           | No        | 1          | 0         | White      | Not rated   | 0     |   |

Home

100%

Type here to search

31-12-2019 0755

## 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:

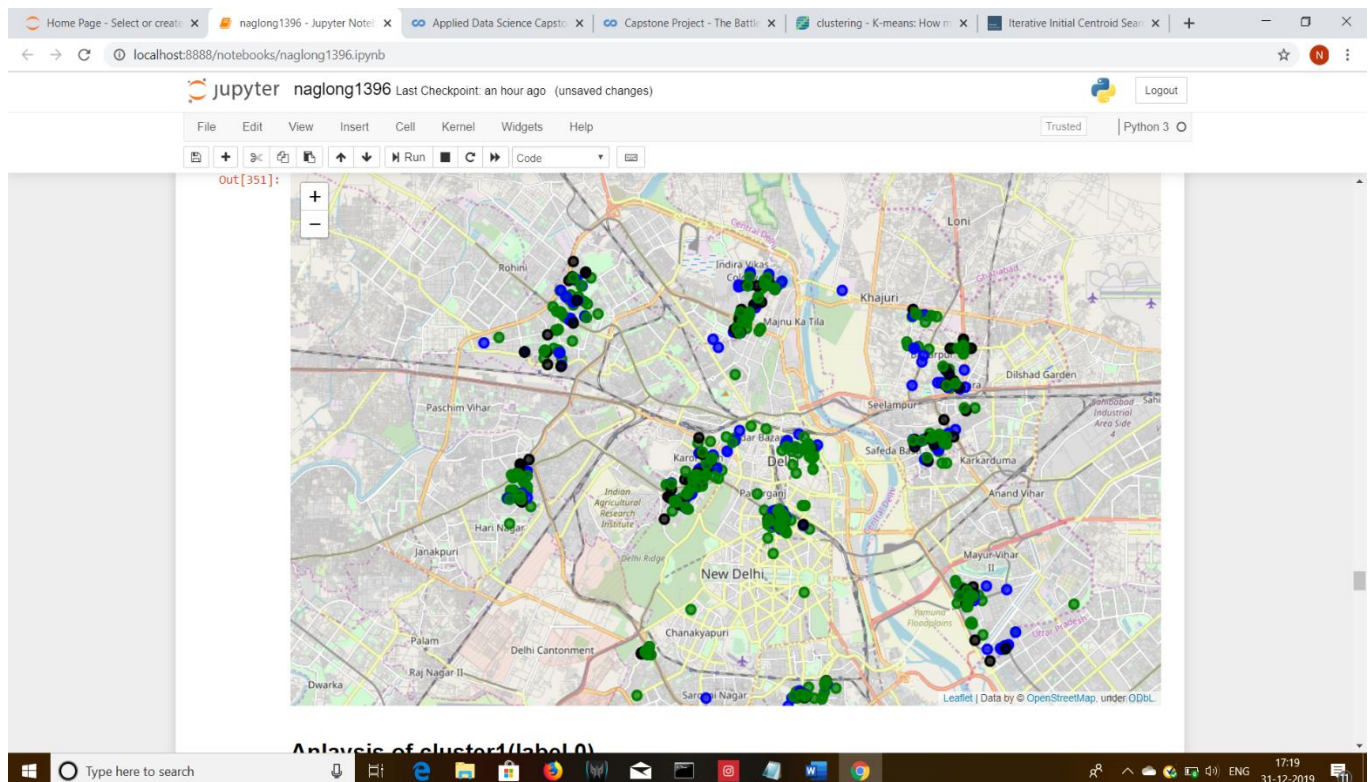
Cluster Analysis was performed on combined data frame. K-means clustering algorithm was deployed with number of clusters equals to three. Main reason for choosing number of clusters to 3, with this number, algorithm will take into account of cuisine type, locality and rating.

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.

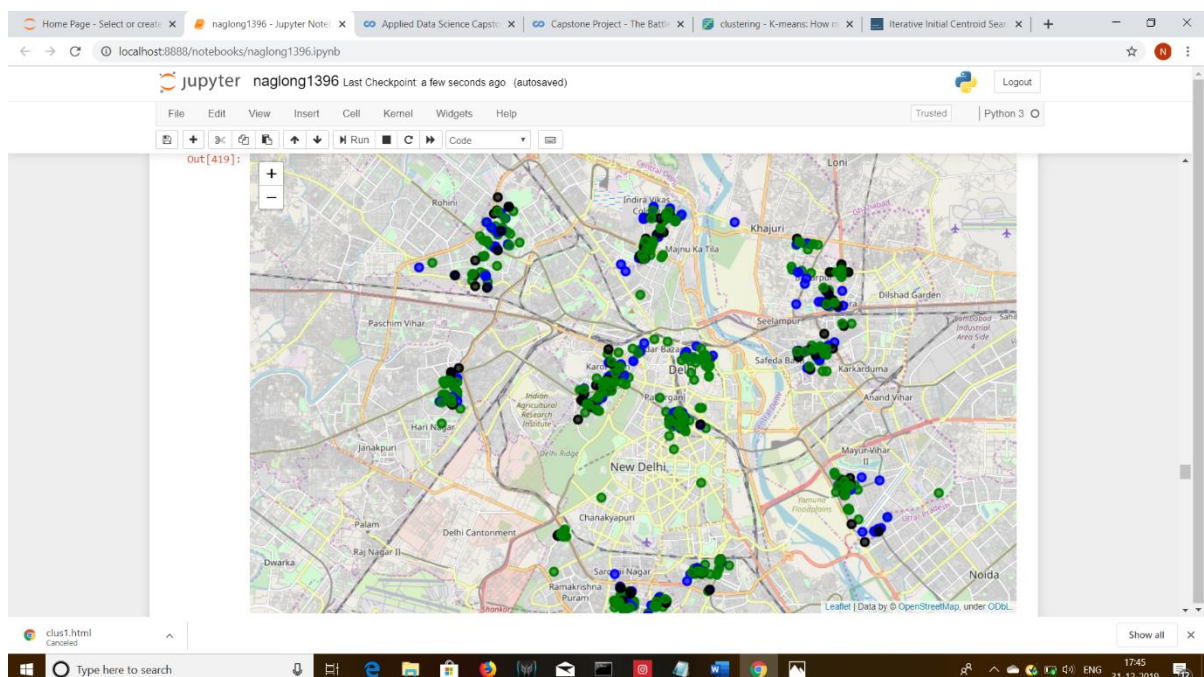
Algorithm picked up Restaurants in cluster 1 with ratings ranging from 2.0 to 4.9



## 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.

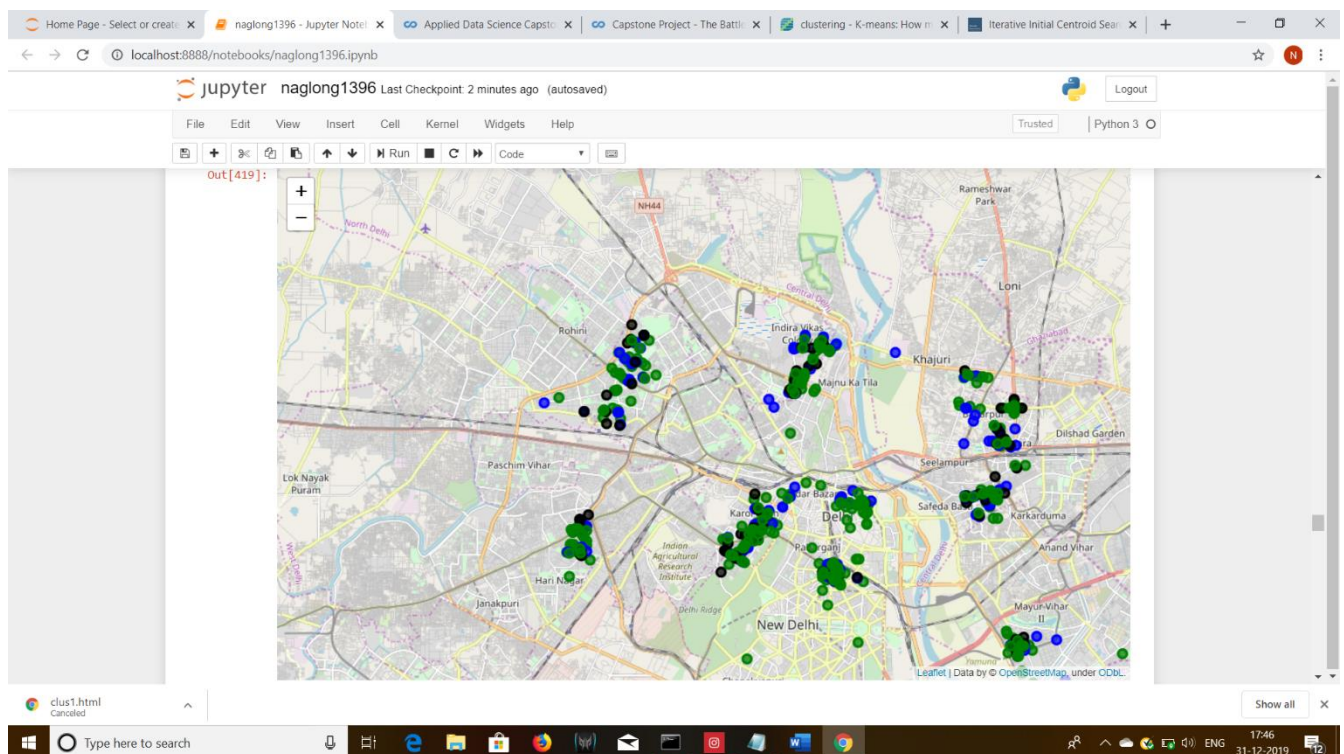
Algorithm picked up Restaurants in cluster 2 with ratings ranging from 2.4 to 4.8



## 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.











