

APPLIED DATA SCIENCE CAPSTONE

Places to Visit In Delhi

PROJECT BY-

Harsh Ajay Kalra

INDEX

INTRODUCTION	03
BUSINESS PROBLEM	04
DATA ANALYSIS	05
DATA COLLECTION	05
DATA PRE-PROCESSING AND WRANGLING	05
DATA FILTERING	06
DATA CLEANING	07
MODEL FITTING	09
CLUSTERING	09
COLOR CODING	10
DATA VISUALIZATION	11
VISUALIZING TARGET CLUSTER	11
OBSERVATION	12

INTRODUCTION

Delhi, the capital of India, has a rich history. The city is dotted with spellbinding mosques, forts, and monuments left over from the Mughal rulers that once occupied the city. The contrast between rambling Old Delhi and well-planned New Delhi is immense, and it's interesting to spend time exploring both. If you feel in need of some relaxation, just head to one of Delhi's flourishing landscaped gardens.



The influence of religious diversity can be seen in the city along with the cultural impact of the Mughal, the ancient Indian and the British. There are many beautiful gardens in the city, away from pollution and busy city life that provide opportunities to walk leisurely during greenery.

The capital city is divided into two sections popularly known as Purani Dilli or Old Delhi and Nayi Dilli or New Delhi. Old Delhi is popular for its ancient culture and monuments along with its overcrowded gastronomical lanes. Let's have a quick peek into the different aspects of Delhi below.

People of all ages can enjoy there as it has something for everyone to complement their interests. There is nothing like best time to visit this majestic land but the period between mid-October to mid-March is considered as an ideal time for those who are not familiar with the summers of Rajasthan. It is because the weather remains pleasant at this time.

One of the country's largest urban agglomerations, Delhi sits astride (but primarily on the west bank of) the Yamuna River, a tributary of the Ganges (Ganga) River, about 100 miles (160 km) south of the Himalayas. The national capital territory embraces Old and New Delhi and the surrounding metropolitan region, as well as adjacent rural areas. To the east the territory is bounded by the state of Uttar Pradesh, and to the north, west, and south it is bounded by the state of Haryana.

Business Problem:

As described above area and population of Delhi is huge. So, the problem here arises that traffic in Delhi is up to great extent. Congestion on Delhi roads has worsened. This has led to lot of travel time due to affected speed of vehicle which acts a grip of a worsening congestion and pollution crisis. Visiting places in Delhi has now become a challenge. Therefore, to visit the capital of our country we need to make a proper business plan to get rid of traffic issues and visit maximum places. This idea can solve the problems faced by travelers visiting places of Delhi. In this project we will divide Delhi into regions which will cover maximum of places. By the end of this project, we will be able to visualize which region we have to visit in order to cover maximum number of places in less duration of time.

DATA ANALYSIS

Data Collection/Gathering: The data used was the location of schools that was acquired using the **foursquare** website. To gather the data, foursquare API was used along with the foursquare credentials Client ID and Client Secret. A 'search' query was made in the **IBM Watson Studio** with **Python kernel**, so as to search the schools. Through the website, coordinates of schools in a particular location were generated and processed.

Data Preprocessing and Wrangling: Using the modules of python, only valid and usable data was selected from the JSON file generated by foursquare and data-frame was created using '**Pandas**'. Since the project required only the locations, the '**venues**' section under the '**response**' section was selected. The generated data-frame still had numerous of unwanted data inside the '**venues**' section, which needed to be filtered and cleaned.

	id	name	categories	referralId	hasPerk	location.address	location.lat	location.lng	location.labeledLatLngs	location.distance	loc...
0	4c17720230d30f472c7836a9	Chawri Bazaar चावड़ी बाजार	[[{"id": '4bf58dd8d48988d112951735', 'name': 'H...'}]]	1611762300	V-	False	Chawri Bazar, Chandni Chowk	28.649771	77.227364	[[{"label": 'display', 'lat': 28.64977098977666...}]]	572
1	4c16685f71d00f47bf2efab6	New Delhi Railway Station (NDLS)	[[{"id": '4bf58dd8d48988d129951735', 'name': 'T...'}]]	1611762300	V-	False	Paharganj-Ajmeri Gate	28.642028	77.219625	[[{"label": 'display', 'lat': 28.64202821789463...}]]	1102
2	53b2b7f5498e396ed85b7220	sufiahandicraft	[[{"id": '4bf58dd8d48988d116951735', 'name': 'A...'}]]	1611762300	V-	False	NaN	28.653348	77.222825	[[{"label": 'display', 'lat': 28.65334762634776...}]]	201
3	5280a63211d26b82c4ba65c7	Spice Market	[[{"id": '4bf58dd8d48988d1f9941735', 'name': 'F...'}]]	1611762300	V-	False	NaN	28.657287	77.222595	[[{"label": 'display', 'lat': 28.65728706392764...}]]	623
4	501c273be4b0e3507bd52252	Khan omelet corner	[[{"id": '4bf58dd8d48988d1df931735', 'name': 'B...'}]]	1611762300	V-	False	Fateh puri	28.654721	77.221952	[[{"label": 'display', 'lat': 28.65472075072693...}]]	334

Data Filtering: There are many irrelevant columns in the dataset. So remove the irrelevant columns as it will be of no use. This is known as filtering of data. So we removed columns like categories, hasPerk, id, location.cc, referral id, etc

	name	categories	address	lat	lng	labeledLatLngs	distance	postalCode	cc	neighborhood	city	state	country	formattedAddress	crossStreet
0	Chawri Bazaar चावड़ी बाज़ार	Hardware Store	Chawri Bazar, Chandni Chowk	28.649771	77.227364	[{"label": "display", "lat": 28.64977098977666...	572	110006	IN	Central Delhi	New Delhi	Delhi	India	[Chawri Bazar, Chandni Chowk, New Delhi 110006...	NaN 4c177202
1	New Delhi Railway Station (NDLS)	Train Station	Paharganj-Ajmeri Gate	28.642028	77.219625	[{"label": "display", "lat": 28.64202821789463...	1102	110001	IN	Central Delhi	New Delhi	Delhi	India	[Paharganj-Ajmeri Gate, New Delhi 110001, Delh...	NaN 4c1668
2	sufiahandicraft	Antique Shop	NaN	28.653348	77.222825	[{"label": "display", "lat": 28.65334762634776...	201	NaN	IN	NaN	NaN	NaN	India	[India]	NaN 53b2b7f54
3	Spice Market	Food & Drink Shop	NaN	28.657287	77.222595	[{"label": "display", "lat": 28.65728706392764...	623	NaN	IN	NaN	NaN	NaN	India	[India]	NaN 5280a632
4	Khan omelet corner	BBQ Joint	Fateh puri	28.654721	77.221952	[{"label": "display", "lat": 28.65472075072693...	334	NaN	IN	NaN	Delhi	Delhi	India	[Fateh puri (Fateh puri), Delhi, Delhi, India]	Fateh puri 501c273be
...
115	Meghraj & Sons	Snack Place	NaN	28.656643	77.223566	[{"label": "display", "lat": 28.65664342714284...	570	NaN	IN	NaN	NaN	NaN	India	[India]	NaN 52be69e2
116	Gole hatt	Snack Place	NaN	28.657232	77.222830	[{"label": "display", "lat": 28.65723224327310...	620	NaN	IN	NaN	NaN	NaN	India	[India]	NaN 4fd5800de
117	Heralal Chat corner	Fast Food Restaurant	Chowri Bazaar	28.649891	77.228808	[{"label": "display", "lat": 28.649891, "lng":...	701	NaN	IN	NaN	NaN	NaN	India	[Chowri Bazaar, India]	NaN 51fbad8b
118	Chainaram	Snack Place	Fathepuri	28.656618	77.223476	[{"label": "display", "lat": 28.65661805874002...	565	110006	IN	NaN	Delhi	Delhi	India	[Fathepuri, Delhi 110006, Delhi,	NaN 4edcb2b0

Checking for null values.

```
[ ] df_filter.isnull().sum()
```

```

name          0
categories    13
address       64
lat           0
lng           0
labeledLatLngs 0
distance      0
postalCode    84
cc            0
neighborhood  115
city          68
state         67
country       0
formattedAddress 0
crossStreet   95
id            0
dtype: int64

```

Data Cleaning: To fit the model, one needs to get rid of the null values. Hence, the firstly, the columns with null, none or NaN values, were identified. The cleaning of data was done by removing the columns with NaN or null values.

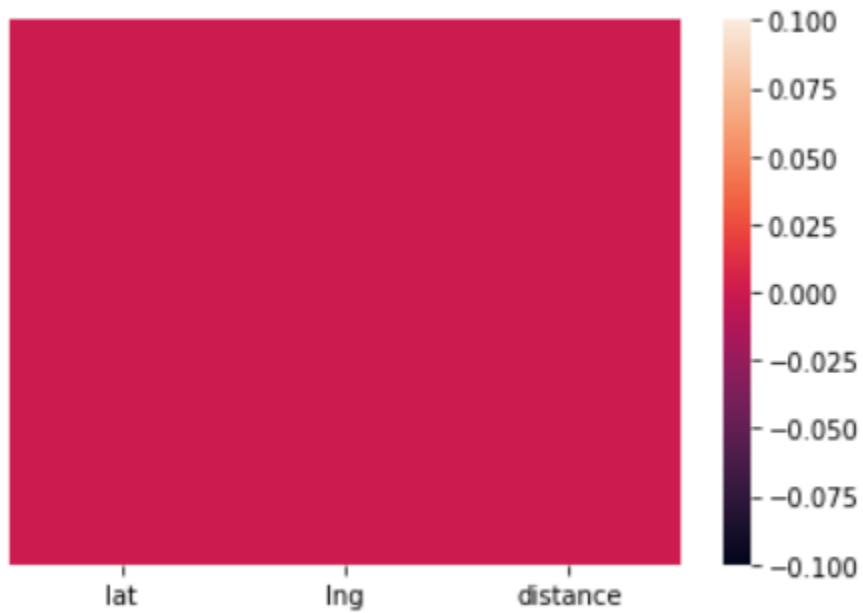
The columns were dropped keeping in mind whether they were really useful for analysis purpose or not. Heat map is generated for visualizing all the null values of the columns.

	lat	lng	distance
0	28.649771	77.227364	572
1	28.642028	77.219625	1102
2	28.653348	77.222825	201
3	28.657287	77.222595	623
4	28.654721	77.221952	334

Now fixing null values for City, as we know city is Delhi so we can replace all the null values in City column by Delhi.

Now generate heat map with no null values for verification.

 <matplotlib.axes._subplots.AxesSubplot at 0x7fd5a6cee8d0>



MODEL FITTING

CLUSTERING: Now we will divide places of Delhi on the basis of their location(East ,West ,South and North) by making clusters of different colors.

We will use K-means clustering technique for this project. In this we will specify value of k i.e. number of clusters. K is chosen 4 here. The model was the fitted and the labels were generated in the form of array.

The dataframe is shown below:

	lat	lng	distance	Labels
0	28.649771	77.227364	572	0
1	28.642028	77.219625	1102	2
2	28.653348	77.222825	201	3
3	28.657287	77.222595	623	1
4	28.654721	77.221952	334	3

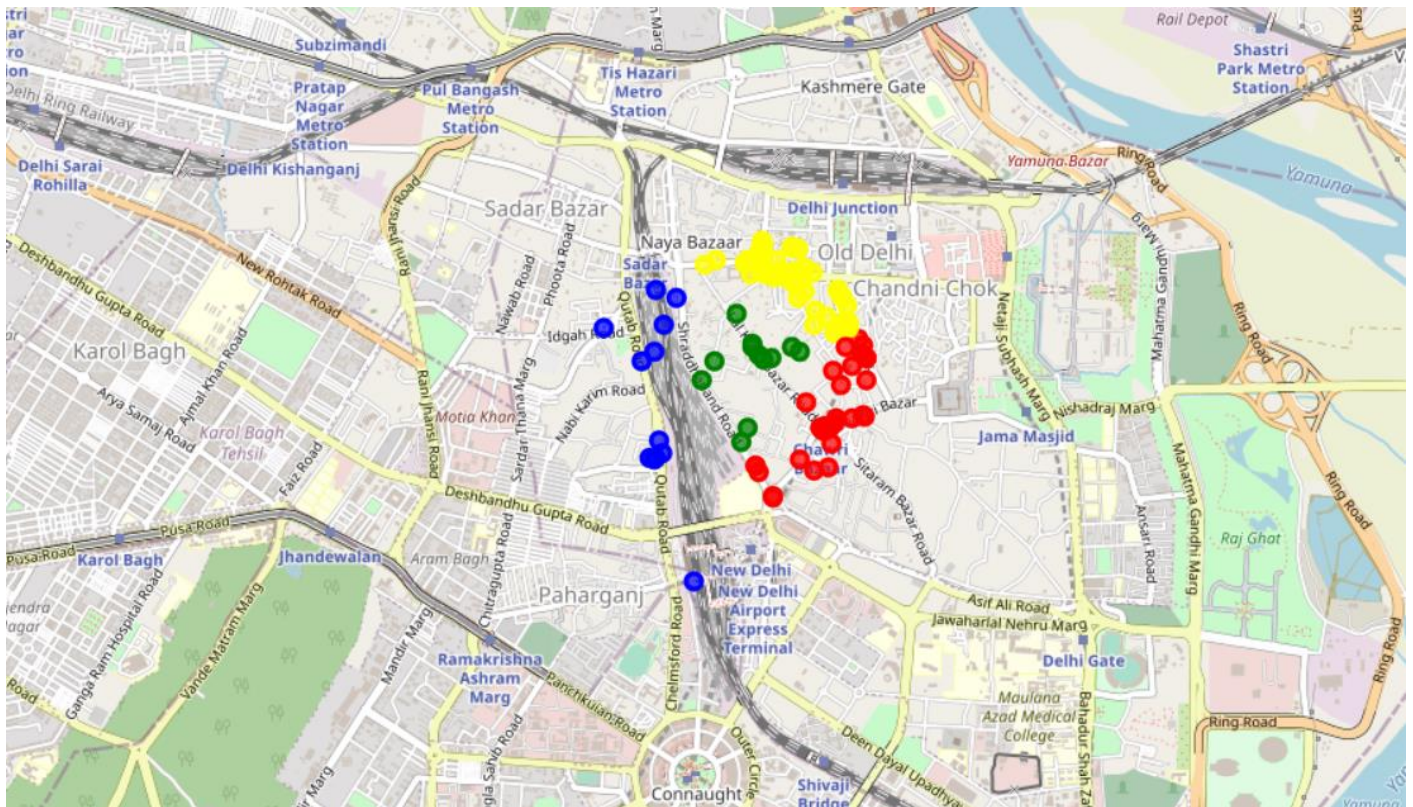
Color code: The color codes used in this project are shown below:

Cluster	Color Code
0	Red
1	Yellow
2	Blue
3	Green

DATA VISUALIZATION

All the 4 clusters were visualized on a map centered on Delhi. The color coding was applied while visualizing for differentiating between the clusters.

Visualizing target cluster: Now we visualize the places on the basis of cluster which is located on the basis of latitude and longitude of each area. There are four colors: Red, Yellow, Blue and Green.



Total Count of places to visit in each cluster:

Total venue in:

Cluster 1: 46

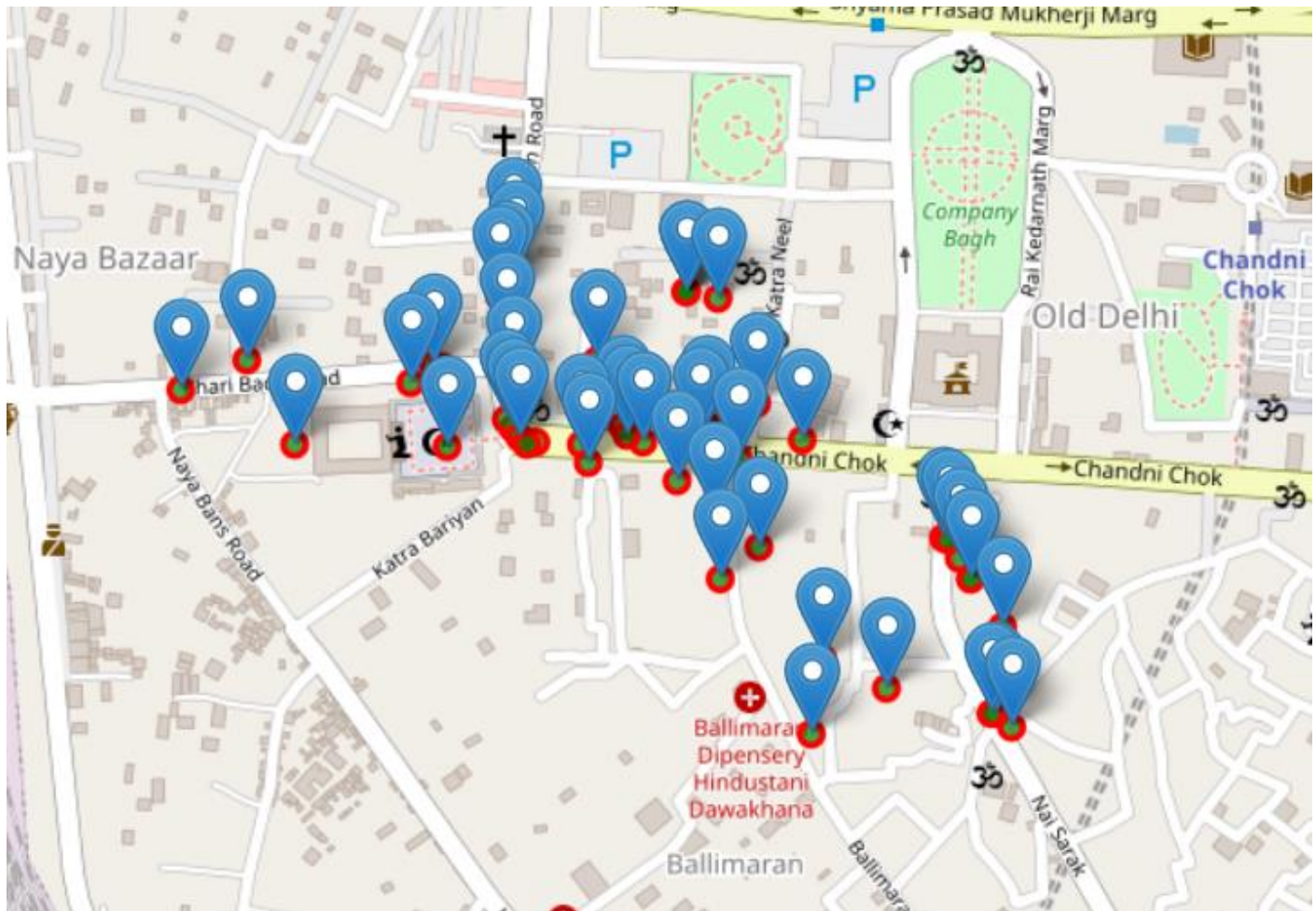
Cluster 2: 13

Cluster 3: 12

Cluster 4: 49

Observation: Hence we can see that Cluster 4 has the maximum places, therefore we will visit places of Cluster 4 to visit maximum places in less amount of time.

Places of Cluster 4 can be visualized using Folium map as shown below:



Project URL:

<https://drive.google.com/file/d/1z3K7AefLgGQVpBTvXCffv0EmP4MamfYp/view?usp=sharing>