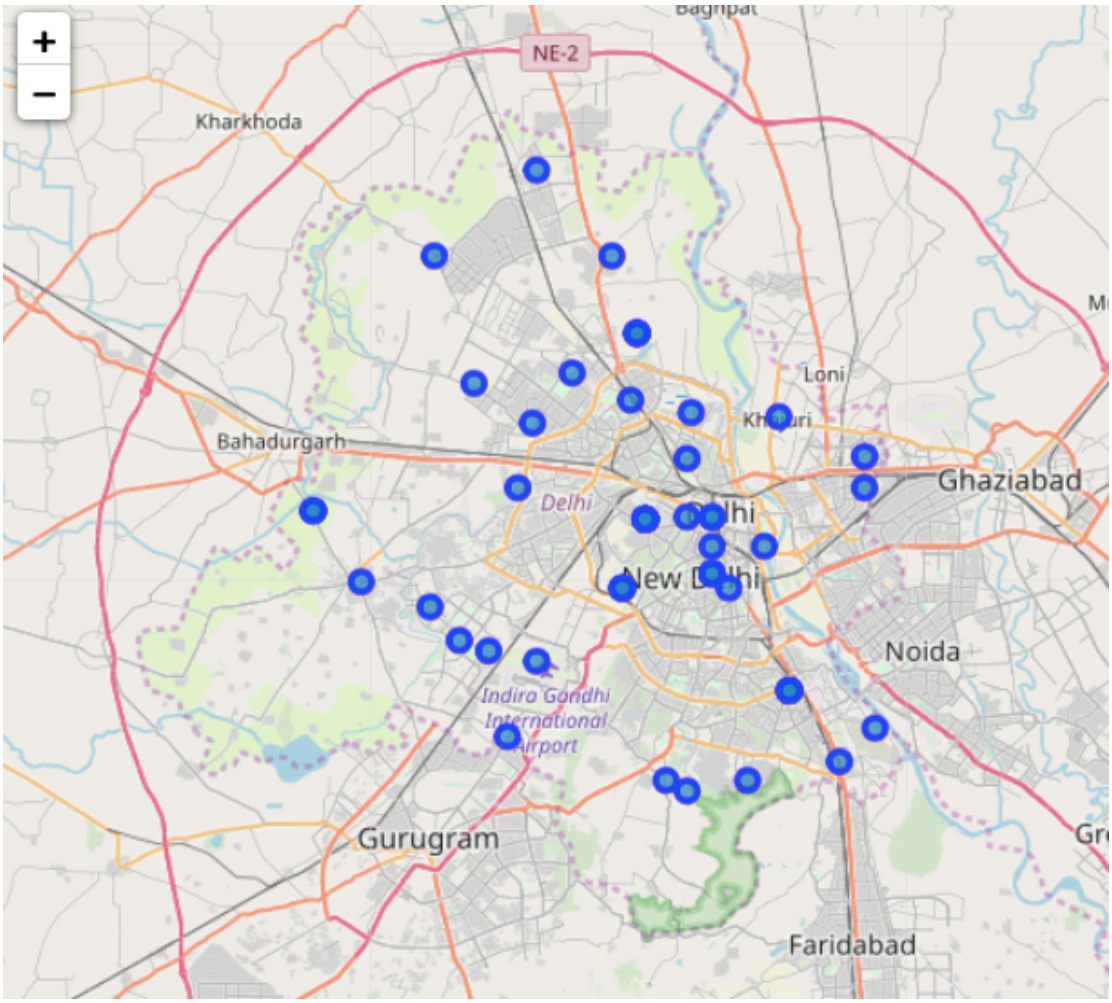
THE MOST APPROPRIATE LOCATION TO OPEN A RESTAURANT IN NEW DELHI (INDIA)

PREPARED FOR
IBM DATA SCIENCE PROFESSIONAL CERTIFICATE
(COURSERA)



PREPARED BY
MEHUL KUMAR

JULY 28TH, 2019

TABLE OF CONTENTS

1. Introduction

1.1 Background

This is the capstone project of IBM Data Science Professional Certificate. In this project, I will use the Foursquare location data to explore neighbourhoods in New Delhi, and to come up with a problem that I can use the Foursquare location data to solve.

1.2 Problem

I will explore the neighbourhoods in New Delhi and answer the question: “Where is the most appropriate location to open a new restaurant in New Delhi”.

1.3 Interest

New Delhi is the capital city of India, a rapidly growing and developing nation with abundant of business opportunities. Opening a restaurant here is an attractive idea for any one who want to extends its business to India. They would be very interested in this project.

2. Data

2.1 Data Sources

My main two data sources are:

- Open Government Data Platform India provides the list of All India Pincode directory with contact details along with Latitude and longitude.

- Venues data from Foursquare

Although the coordinates data was not available for New Delhi. So, I created a dataset for the Longitude and Latitude for all the postal codes available in New Delhi.

2.2 Data Cleaning

New Delhi is an urban district of Delhi which serves as the capital of India. First of all, I need to collect the districts and neighbourhoods data from Open Government Data Platform India provides the list of All India Pincode directory with contact details along with Latitude and longitude. and find the coordinates for every neighbourhood. I did this manually and create a file named DELHI_GEO.csv

The formatted data[1] looks like this:

	PostalCode	Region	Localities	Neighborhood	Latitude	Longitude
0	110001	Central Delhi	Baroda House S.O,Bengali Market S.O,Bhagat Sin...	Connaught Place	28.633300	77.216700
1	110002	Central Delhi	A.G.C.R. S.O,Ajmeri Gate Extn. S.O,Darya Ganj ...	Darya Ganj	28.633300	77.250000
2	110003	Central Delhi	Delhi High Court Extension Counter S.O,Delhi H...	Aliganj	28.650000	77.216700
3	110004	Central Delhi	Rashtrapati Bhawan S.O	Rashtrapati Bhawan	28.650000	77.216700
4	110005	Central Delhi	Anand Parbat Indl. Area S.O,Anand Parbat S.O,B...	Lower Camp Anand Parbat	28.650000	77.200000
5	110006	Central Delhi	Hauz Qazi S.O,Jama Masjid S.O	Bara Tooti	28.650000	77.216700
6	110007	North Delhi	C.C.I. S.O,Delhi University S.O,Gulabi Bagh S....	Birla Lines	28.683300	77.200000
7	110008	Central Delhi	Dada Ghosh Bhawan S.O,Patel Nagar East S.O,Pat...	Patel Nagar	28.650000	77.216700
8	110009	North West Delhi	Dr.Mukerjee Nagar S.O,G.T.B.Nagar S.O,Gujranwa...	G.T.B Nagar	28.709895	77.202561
9	110010	South West Delhi	505 A B Workshop S.O,A F Palam S.O,Aps Colony ...	Delhi Cantt	28.550000	77.266700

2.3 Feature Selection

In this project, I will get data of recommended venues inside 1000 meters radius of every neighbourhood, calculate the top10 most common venues by its category as features. Plus, the 11th feature is if there is a bus/metro station nearby.

Feature Label	Type & Description
1st Most Common Venue	str. Name of the vanue category
2nd Most Common Venue	str
3rd Most Common Venue	str
4th Most Common Venue	str
5th Most Common Venue	str
6th Most Common Venue	str
7th Most Common Venue	str
8th Most Common Venue	str
9th Most Common Venue	str
10th Most Common Venue	str
Station	str. Yes or No

[1] refer: https://github.com/er-mehulkumar/Coursera_Capstone/blob/master/ND_DATASET.ipynb