# **Exploring cities in India using APIs**

#### IBM Professional Certificate in Data Science Capstone

This project is a part of the IBM Professional Certificate in Data Science Capstone on Coursera.

**1. Introduction/Business Problem**

The goal of the project is to explore various cities in India to find the city with the best distributions of restaurants, and then explore that particular city to find the types of restaurants and their ratings, and the area with the best restaurants.

The data collected from this study can be used by entrepreneurs dealing with food related businesses or popular food chains looking to establish a new restaurant in India, or even tourists wanting to try out food in India.

**2. Data**

To provide the necessary information, the following data will be used:  
- Cites, Latitude, Longitude: <http://www.tageo.com/index-e-in-cities-IN.htm>  
- Restaurants: Foursquare API, Zomato API

**3. Methodology**

* I scraped data regarding Indian cities from <http://www.tageo.com/index-e-in-cities-IN.htm> using BeautifulSoup and arranged it into a table.
* I plotted these cities on a map using Folium.
* I used Foursquare API to get a list of venues, types, and their location for each city.
* After extracting the necessary data, I added weights for each food related venue. Indian Restaurant:7, Hotel:6, Ice Cream Shop:6, Bar:3, Seafood Restaurant: 5, Coffee Shop:4, Diner: 3, Fast Food Restaurant: 3, Club House:9, Café: 10
* I calculated the city with the greatest mean weight, and retrieved Ludhiana.
* I then used Foursquare API again to get a list of venues within a 100 KM radius of the city Ludhiana, and plotted it on a map using folium.
* I shifted my focus to just restaurants. By using the Zomato API, I retrieved restaurant details for the city of Ludhiana, such as average price, rating, location, etc.
* I began analysing the ratings and the prices.
* From the plot, we can see that a majority of the ratings are between 3.25 and 4.0. The map below depicts the ratings, with dark green being the best and orange/red being the worst. I also plotted this on the map.
* I visualised the prices. We can see that a majority of the restaurants have an average price of 100–200, with price range being 50-400.
* I arranged the data into two clusters based on price range and rating. The venues for Cluster 1 have mean price range of 1.00 and rating spread of around 3.56. The venues for Cluster 2 have mean price range of 1.33 and rating spread around 3.49.

**4. Results and Discussion**

* After our analysis, based on the weights applied, we have selected Ludhiana as our target city.
* In Ludhiana, we saw that a majority of the venues have a rating of 3.5 and average price per person of 100–200.
* The results show clustering near Ludhiana junction. There is also not much difference between the two clusters, and the city seems to have evenly spread out restaurants of all types.

**5. Conclusion**

Therefore, we can conclude that a food enthusiast visiting India might have a good experience in Ludhiana, with good food at decent prices. Even new food shop owners can establish restaurants here as there is demand for all types of food.