**IBM Applied Data Science Capstone**

**Exploring Venues in Coimbatore**

**Data Section**

**1. Data Sources**

To get location and other information about various venues in Coimbatore, I used two APIs and decided to combine the data from both of them together.

Using the Foursquare’s exploring API (which gives venues recommendations), I fetched venues up to a range of 4 kilometers from the center of Coimbatore and collected their names, categories and locations (latitude and longitude).

Using the name, latitude and longitude values, I used the Zomato search API to fetch venues from its database. This API allows finding venues based on search criteria (usually the name), latitude and longitude values and more. Given that the data from the two APIs did not align completely, I had to use data cleaning to combine the two datasets properly.

From Foursquare API (<https://developers.zomato.com/api>), I retrieved the following for each venue:

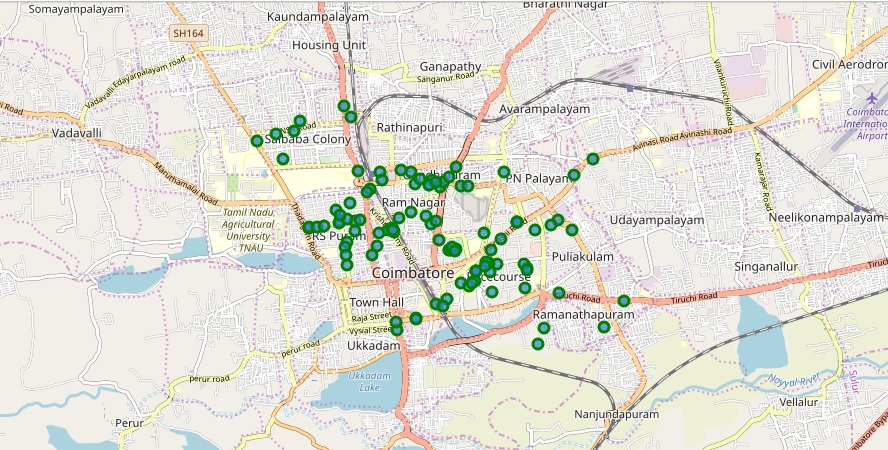
* **Name:** The name of the venue.
* **Category:** The category type as defined by the API.
* **Latitude:** The latitude value of the venue.
* **Longitude:** The longitude value of the venue.

From Zomato API (<https://developers.zomato.com/api>), I retrieved the following for each venue:

* **Name:** The name of the venue.
* **Address:** The complete address of the venue.
* **Rating:** The ratings as provided by many users.
* **Price range:** The price range the venue belongs to as defined by Zomato.
* **Price for two:** The average cost for two people dining at the place. I later convert the same to average price per person by dividing by 2.
* **Latitude:** The latitude value of the venue.
* **Longitude:** The longitude value of the venue.

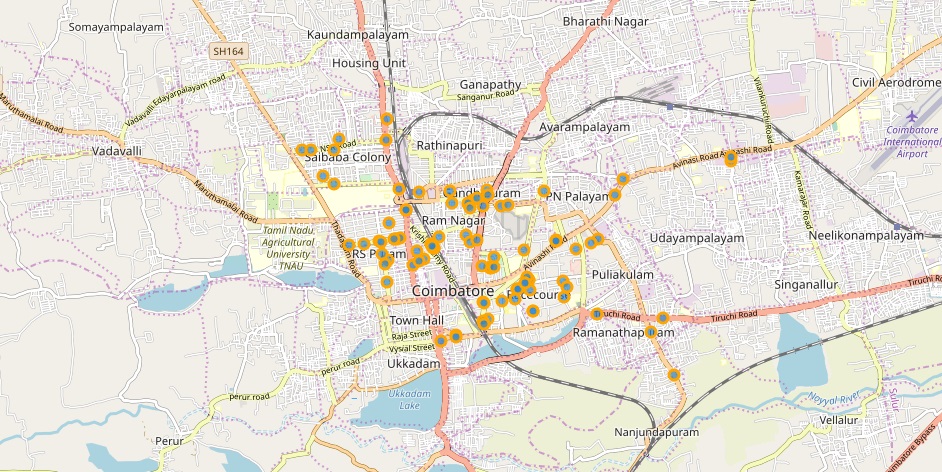
**2. Data Cleaning**

**Venues retrieved from Foursquare API**

****

*Figure 1: Venues retrieved from Foursquare API*

**Venues retrieved from Zomato API**

****

From figure 1 and figure 2, we can clearly see that some venues from the two APIs do not align with each other. Thus, I decided to combine them using their latitude and longitude values.

To combine the two datasets, I had to check that the latitude and longitude values of each corresponding venue match. After careful analysis, I decided to drop all corresponding venues from the two datasets that had their latitude and longitude values different by more than 0.0004 from one another. Thus, I rounded both the latitude and longitude values up to 4 decimal places.

Then, I calculated the difference between the corresponding latitude and longitude values and saw if the difference was less than 0.0004 which should ideally mean that the two locations are the same. This removed many outliers from the two datasets. Once this was done, I observed that there were still some venues which were not correctly aligned.

They can be categorised as follows:

1. There are venues that have specific restaurants/cafes inside them as provided by Zomato API (Pizza Hut inside Brookefiels Mall).
2. Two locations are so close that they have practically same latitude and longitude values (Chin Chinand Bikes & Barrels).

Venues belonging to category 1 are perfect to keep. However, the venues that belong to category 2 should be dropped. After careful inspection and removal, the final dataset had a total of 49 venues with which we can work.

As a final dataset, we’re left with 49 venues with 8 columns as described in figure 3.

**Final data aggregated from both APIs**



*Figure 2: Final data aggregated from both APIs*