**Coursera Capstone: Best location for City Waste Management in Kochi city**

1. Data

This project will make use of Foursquare API for two data collection purposes. One, to find the venues surrounding the waste management plant of one of the world’s cleanest cities and two, to generate a dataset that comprises the neighborhoods and venues in the city of Kochi, Ernakulam district, southern India.

East Calgary Facility location will be obtained from the *geopy* package and Foursquare API will be then used to mine the nearby venues within 5000 metres radius. Then the East Calgary Facility datapoint will be created by representing the facility using the venue categories of each venue mined in the form of average value of one-hot encoded score. The East Calgary Facility datapoint is shown in Figure.2.1 (due to space constraint only few columns are depicted):

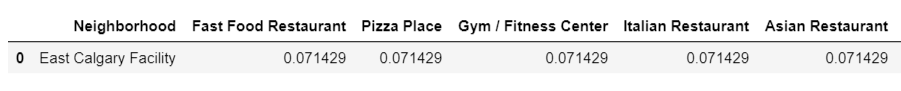


Figure.2.1: East Calgary Facility datapoint

Kochi city will be defined using the neighborhoods of Ernakulam district and the neighborhoods will be webscraped using BeautifulSoup from the website ‘<https://worldpostalcode.com/india/kerala/ernakulam/>’. The neighborhood names scraped will be then used to mine the geolocation data of each of these using the *geopy* package. A small excerpt from the scraped data is shown in Figure.2.2. The webscraped dataset will be then processed to eliminate any missing values and then fed to Foursquare API to obtain nearby venues for each neighborhood. Similar to the East Calgary Facility datapoint, neighborhoods of Ernakulam will also be processed and few part of it is represented in Figure.2.3.

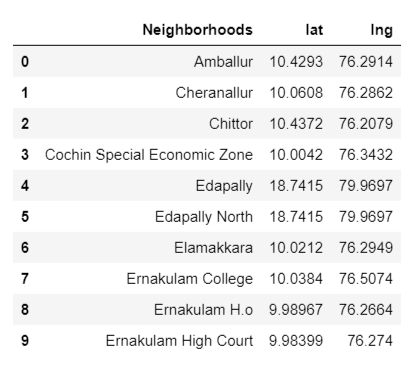


Figure.2.2: Webscraped data of Ernakulam’s neighborhoods

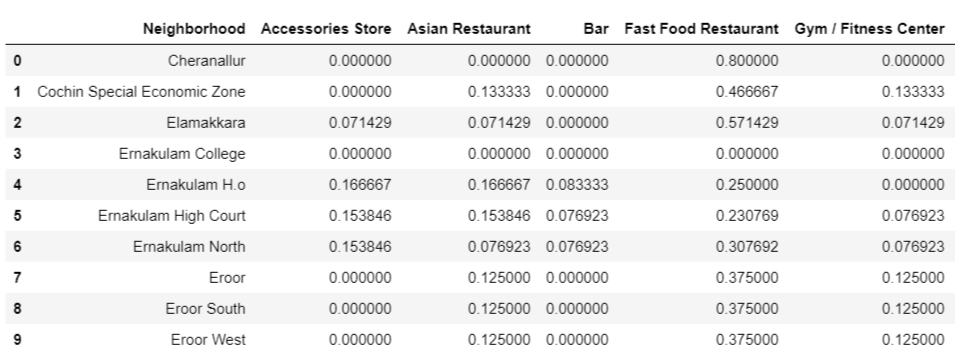


Figure.2.3: Ernakulam district neighborhood data

The East Calgary Facility datapoint will be concatenated with the Ernakulam district neighborhood dataframe and K-means clustering with 4 clusters will be performed to find the suitable location choices for Kochi city waste management.