Facility Location Planning in India

Abhay Sobhanan

May 21, 2020

Introduction

India is considered as one of the growing economies with high potential for development and success. The huge number of English-speaking youth and their educational skills attract many multi-national companies. Such companies also benefit tax reductions from the Indian government in addition to low infrastructural and running costs compared to the western countries. The given problem is commonly known as facility location planning. It is a computationally hard problem with increasing constraints present in the model. In this report, we will only look at a small data and categorize it using an unsupervised machine learning algorithm. Location selection problems are not only limited to commercial field applications, even a person looking to move to a new country or a place can utilize this problem to find out the right city or neighborhood for his stay.

Interest to stakeholders:

Selection of the right location for your business or new house is a crucial decision which highly influences your professional and/or personal growth over the coming years. The search for a perfect location requires many additional details such as the company profile, location specific requirements, presence of competitors, etc. However, this analysis can be looked at as a preliminary shortlisting of preferable cities for your team. Instead of wasting a considerable amount of time in contacting different helpers across the country, shortlisting through data analysis will result in the benefit of low-cost and time.

Now, are you a person looking for a notable difference in living environment? This study will also help you to sort out major cities across multiple Indian states to find your future home. Personal preferences for facilities highly vary, apart from social and cultural interests. With the advancement in development, your favorite hangout place can vary among a garden, a beach or dine-out and shopping facilities. You can obtain an initial idea from this analysis which groups different cities based on their popular venues.

Data

Data Requirement:

For the described problem, we require a list of Indian cities and their corresponding geo-coordinates to plot and acquire other related information. If latitude and longitudes of a city is unavailable in the dataset, we will use *pgeocode* Python package to obtain the details. Using Foursquare API, we will extract the list of most popular venues at these locations with a limit of 20 and radius = 5 km.

Data Acquisition:

We extract the data of Indian cities and their geographical coordinates. The data is scrapped from the website: "https://simplemaps.com/data/in-cities". The data is readily available for download as CSV file. Else, we may use *BeautifulSoup* python package to scrap the web data. For each selected city for the analysis, we obtain the venues list via https://api.foursquare.com/v2/venues using a Client ID and Secret provided by the platform.

Data Cleaning:

Initially, the dimensions of the Pandas dataframe containing location and their geo-coordinates had the dimensions: 212 rows and 9 columns. We rename and select the columns of our requirements such as: Location (City), Latitude, Longitude, State and Population. Later, the rows with NA entries for Population were dropped to filter out the main cities with census data. Please note that the Foursquare venues data for the Indian cities are not very accurate and contains less entries. In addition, due to the Covid-19 lockdown rules laid out by the central and state governments, many public places and transport services have stopped functioning. These changes will be reflected in the obtained data. Since these changes cannot be altered at the present time, we will go ahead with the analysis after the preliminary modifications.