IBM DATA SCIENCE PROFESSIONAL CERTIFICATE

Capstone Project

Business/Tech Park in Bengaluru

By: Krishna Ramadas

Sept 2019

Introduction:

Bengaluru aka Bangalore is one of the fast-growing metro cities in INDIA. Since it is also known as the "Silicon Valley" of India, most of the IT enthusiast and professionals find Bengaluru as their dream destination. All major IT companies, be it service providers or Product companies have their office in Bengaluru. Not just MNCs but several Start-Ups too have their blooming roots in Bengaluru.

Business/Tech parks are changing the outlook of the city and even expanding the rarely populated parts of the city into highly populous destinations. Finding an ideal place to construct a business/tech park to rent it on a lease or to sell to the interested parties is a challenging venture. The Park has to be built in such a place where transport and other amenities are easily available. There are several business/tech parks in Bengaluru, some being very popular and pricey, yet there are some which are vacant and even never occupied.

The main goal of this project is to find an apt location for the construction of a business/tech park and also to determine whether this will be a success or failure.

Business Problem

This capstone project aims to analyze and select an apt location for the construction of business/tech Park in Bengaluru city, the capital of Karnataka, India. With the help of Data Science methodology and Machine Learning algorithms like clustering, this project looks forward to answering the business problem: Where would you recommend a property developer, if he/she is looking to construct a business/tech park in the city of Bengaluru?

Target audience

This project mainly targets the investors and potential property developers' who are interested to invest in such a multi-million dollar venture. This also targets the prospective clients of such property developing companies who

are keen to know the possibility and feasibility of such a giant project and is ready to collaborate with the company.

Data:

- List of neighborhoods in Bengaluru city (without their PIN codes). The project is confined to the capital city of Bengaluru
- Latitude and Longitude data of these of the neighborhoods in Bengaluru city to plot maps and also to get the venue data (In case if the geo coordinates were not available easily, List of places in Bengaluru District along with PIN codes)
- Venue data, specifically data related to business/tech parks in Bengaluru city. This data is necessary to cluster the regions and hence find the ample location
- The latest per square feet property (land) rate in Bengaluru for each of the locations from property retail sites to know the average price of the land needed and also the prospective selling or leasing price of the business park's commercial space.

Data Source and Extraction Methodology

The Wikipedia page https://en.wikipedia.org/wiki/List_of_neighbourhoods_in_Bangalore contains the list of neighbourhoods in Bengaluru city. Data scraping methods will be used to obtain the required data. Then we will obtain the geo coordinates (latitude and longitude) of the neighbourhoods using the Python geocoder package for visualizing map.

If the geocoder fails as a backup we already have readily available geo coordinates data of India from with the data for Bengaluru's neighbourhood can be extracted.

The website called geonames.org host a free download page which comprises of all the postal codes along with the latitude and longitude data of the countries of the world. All the postal code data of INDIA is downloaded directly into the Local computer as a zip file from the following link address https://download.geonames.org/export/zip/IN.zip. Data

wrangling methods will be used to get the required data regarding the area of interest that is, of Bengaluru city

Then we will use the Foursquare API to extract venue data of these locations in Bengaluru city. Foursquare has one of the largest databases of 105+ million places and is used by over 125,000 developers. Foursquare API will provide venue data according to categories, here in our case its business parks and Tech parks. This project lets us explore our data science skills, from data collection, cleaning, wrangling, using API (Foursquare) to using Machine learning model (K-Means Clustering) and Data visualization.

Thus we can filter down to three or four ample locations for the construction of our Business/tech park

The data required for knowing the average price of the property (land) for the construction and also for the prospective rending or selling price of the commercial space, the business park hoists extracted from this property retail site https://www.99acres.com/search/project/buy/commercial/bangalore?preference=S&np_search_type=NL&city=20&area_unit=1&budget_min=undefined&budget_max=undefined&res_com=C&refSection=GNB&refSection=GNB

Web scraping will be used to extract data, most probably BeautifulSoup module from Python library will be used to get data from this site to compare the prices of the resulting ample locations for the business park and thus we can determine whether the business/tech park will be a success or not