

APPLIED DATA SCIENCE CAPSTONE – REPORT

OPENING A NEW MICROBREWERY IN BANGALORE, INDIA

MAY 2020

Introduction

Bangalore is one of the largest cities in India. The city has population of around 12 million and the city is expanding rapidly over the last couple of decades due to the Information Technology industry. Due to this reason the city is also known to be the 'Silicon Valley of India'. But, over the ages Bangalore is popularly known across India to be the Pub Capital of country. And it continues to be so. With the rapid increase in population, especially many young people moving here job prospects there has been a growth in the Pub and Restaurant scene in the city. Now, Craft Beer is trending all over the world, and Bangalore is too it is a same case. As of 2019, there has been more than 60+ microbreweries spread across the city and the number is going to keep on increasing. According to the current trends it is likely that more Microbreweries are going to pop up across the city.

Business Problem

In this capstone project, we are going to analyse and select the best neighbourhoods in Bangalore, India to open a Microbrewery. We will use Data Science Methodologies and Machine Learning techniques such as Clustering to find the best location to set up a Microbrewery. The results will help someone planning to set up a Microbrewery find the perfect location. There are many socio-economic, geographical and political reasons of setting up microbreweries in certain parts of the city.

Data

We will use Data Scraping techniques to obtain the list of all the wards(neighbourhoods) in Bangalore, India from the Wikipedia page: https://en.wikipedia.org/wiki/List_of_wards_in_Bangalore. Wards are actually local authority areas(like small neighbourhoods) used for electoral purposes. In Bangalore the civic authority BBMP(Bruhat Bengaluru Mahanagara Palike) administers these 198 wards. In the table we also get the Assembly Constituency(for state electoral) and Lok Sabha Constituency(for national electoral) information as well. This is helpful in finding the correct geolocation since some place might have same names and this can help in differentiating. We will incorporate data cleaning and data wrangling on this data obtained and finally, get the geolocation data (latitude and longitude) using the python geocoder.

After this, we will use Foursquare API to get all the venue information of each neighbourhood. We will filter this venue data to with the relevant venue information which relates to a Microbrewery.

We are going to form Clusters based on the venue information and visualize it on a map of Bangalore using Folium API.