

Coursera Capstone

IBM Data Science Capstone Project

Opening a Pub in Mumbai

Yatharth Aggarwal
April, 2020



Introduction

A pub, or public house, is an establishment licensed to sell alcoholic drinks, which traditionally include beer (such as ale) and cider. It is a social drinking establishment and a prominent part of British, Irish, Breton, New Zealand, Canadian, South African and Australian cultures] In many places, especially in villages, a pub is the focal point of the community. In his 17th-century diary, Samuel Pepys described the pub as "the heart of England".

But now it has spread to almost all countries across the globe. Developing countries like India with its huge population and once a colony of Britishers has definitely moved into this trend. Westernization and night life is a common trend amongst both the upper and middle class of the society.

Mumbai is the economic capital of India. The Richest man in Antila to the Poorest man in Dharavi live in this city. Thus this city have a complete diversity among the social and economic groups, making a perfect scenario of every type of recreational facility to be built.

As a result, there are many pubs in the Mumbai and many more are being built. Opening pubs allows property developers to earn consistent rental income. Of course, as with any business decision, opening a new pub requires serious consideration and is a lot more complicated than it seems. Particularly, the location of the pub is one of the most important decisions that will determine whether it will be a success or a failure.

Business Problem

The objective of this capstone project is to analyse Pub clusters in the city and select the best locations in Mumbai, India to open a new pub.

Using data science methodology and machine learning techniques like clustering, this project aims to provide solutions to answer the business question: In Mumbai, India, if a property developer is looking to open a new pub, where would you recommend that they open it?

Target Audience of this project

This project is particularly useful to property developers and investors looking to open or invest in new pubs or similar structure in Mumbai. Commoners are also stake holder for any public space being built around them, thus they are also a target audience.

Data

To solve the problem, we will need the following data:

- List of neighbourhoods in Mumbai. This defines the scope of this project which is confined to the Mumbai city.
- Latitude and longitude coordinates of those neighbourhoods. This is required in order to plot the map and also to get the venue data.
- Venue data, particularly data related to pubs. We will use this data to perform clustering on the neighbourhoods.

Sources of data and methods to extract them

This Wikipedia page (https://en.wikipedia.org/wiki/Category:Suburbs_of_Mumbai) contains a list of neighbourhoods in Mumbai, with a total of 40 neighbourhoods. We will use web scraping techniques to extract the data from the Wikipedia page, with the help of Python requests and beautiful soup packages.

Then we will get the geographical coordinates of the neighbourhoods using Python Geocoder package which will give us the latitude and longitude coordinates of the neighbourhoods.

After that, we will use Foursquare API to get the venue data for those neighbourhoods. Foursquare has one of the largest database of 105+ million places and is used by over 125,000 developers.

Foursquare API will provide many categories of the venue data; we are particularly interested in the Pub category in order to help us to solve the business problem put forward.

This is a project that will make use of many data science skills, from web scraping (Wikipedia), working with API (Foursquare), data cleaning, data wrangling, to machine learning (K-means clustering) and map visualization (Folium). In the next section, we will present the Methodology section where we will discuss the steps taken in this project, the data analysis that we did and the machine learning technique that was used.