

Coursera Capstone

IBM Applied Data Science Capstone

Opening a bubble tea shop in Hong Kong, China

By Angela LEUNG
Apr2019

Introduction

Bubble tea (also known as **pearl milk tea**, **bubble milk tea**, or simply **boba**) is a Taiwanese tea-based drink invented in Tainan and Taichung in the 1980s.[1] Boba with its shaky and milky tea, is a craze in Hong Kong, an international city in China, Asia, and this becomes a super sweet market for inventive Hong Kong entrepreneurs. However, in what way the entrepreneurs could to reach the maximum profile for their bubble tea shop investment? Location to run the business is definitely one of importance factor.

Business Problem

The objective of this capstone project is to analyze the best location to start up a Boba tea shop in Hong Kong, China. Using data science methodology and clustering by machine learning, this project would provide a suggestion to answer the business question: Where would be the recommended location to invest a boba shop in Hong Kong, China.

Target Audience of this project

This project is particularly useful to anyone who would invest a boba tea shop or plan to expend a boba tea shop in city Hong Kong, where having a bubble tea craze. This project is timely as the city currently suffering from oversupply of bubble tea.

Data

In this section, the data need to provide solution are listed:

List of neighborhoods in Hong Kong. This defines the scope of this project which is confined to the city of Hong Kong, an international market of China in Asia.

Latitude and longitude coordinates of those neighborhoods. This is required in order to plot the map and also to get the venue data.

Venue data, particularly data related to boba tea shop. This is the data for clustering on the neighborhoods.

Sources of data and methods to collect

The list of neighborhoods in Hong Kong is sourcing from a Wikipedia page (https://en.wikipedia.org/wiki/Districts_of_Hong_Kong). Hong Kong with a total of 18 neighborhoods. Web scrap techniques is used to collect the data from the Wikipedia page under the help of Python requests and beautifulsoup packages.

Next is getting the geographical coordinates of the neighborhoods using Python Geocoder package which will return the latitude and longitude coordinates of the neighborhoods.

Last, the venue data of boba shop will be sourcing with Foursquare API. Foursquare has one of the largest databases of 105+ million places and is used by over 125,000 developers. Foursquare API provide many categories of venue data, but in order to address our key question and align of this project scope, Boba shop is the key interest.

In summarize, this is a project make use of data science skills for analysis, including:

- Web scraping from Wikipedia
- API connection with Foursquare
- Data cleaning
- Data wrangling
- Machine learning on K-means clustering
- Map visualization with Folium