The Battle of the Neighbourhoods — Opening a Bubble Tea franchise in Singapore

A quest to find the best neighbourhood to start a Xing Fu Tang Bubble Tea franchise in Singapore.

**Introduction: Business Problem**

Originating from Taiwan in the early 1980s, **bubble tea** is a flavoured/milk tea beverage served with tapioca balls. In recent times, this tea drink has taken the world by storm. It's wildly popular among the youths of Asia, and more recently in the North America. This holds true for the Southeast-Asian city-state of **Singapore** as well.

When tightened circuit-breaker measures were announced in Singapore to curb the spread of COVID-19 on 21 April 2020, Singaporeans [went out in droves](https://www.scmp.com/news/asia/southeast-asia/article/3080933/coronavirus-singapores-bubble-tea-fans-rush-beat-circuit) to get their fix of bubble tea. Searches for the drink on Google spiked, and for the few bubble tea shops that were still allowed to operate during the lockdown, they find themselves [running out](https://www.todayonline.com/singapore/desperate-bubble-tea-some-shops-remain-open-run-out-pearls-within-hours-opening) of tapioca pearls within hours of opening.

As the retail sector gradually recovers from the COVID-19 pandemic in 2021, the mainstays of the Singaporean bubble tea scene should anticipate a return to normalcy and use the opportunity to scale mindfully. This includes one of Singapore's well-known bubble tea brand, **Xing Fu Tang**. At the writing of this report, Xing Fu Tang operates 10 outlets in Singapore and is garnering [positive reviews](http://topten.sg/food/7916).

For this business problem, we'll use data science tools to fetch, visualise and analyse geolocation, demographic, and commercial data.

We begin by locating bubble tea stores in Singapore using the Foursquare API, and checking if they are often located near shopping malls and Mass Rapid Transit (MRT) stations. We'll then import and visualise population, income and household dwelling data of Singapore's Planning Areas/Subzones[1].

With the above, we'll use a clustering model to find Subzones with similar characteristics, finding patterns in the data that'll show us where best to open a new Xing Fu Tang outlet. Promising areas should also be situated away from existing Xing Fu Tang outlets as well.

[1] Singapore's first-level and second-level census divisions, respectively.

## Data

Based on the problem defined, we look to group Subzones based on their following features:

* [Planning Area Boundaries](https://data.gov.sg/dataset/master-plan-2014-planning-area-boundary-no-sea) – To define Planning Area
* [Subzone Boundaries](https://data.gov.sg/dataset/master-plan-2019-subzone-boundary-no-sea) – To define Subzones
* [Location of existing Xing Fu Tang outlets](https://xingfutangsg.com/#stores) – To see where Xing Fu Tang already has a presence [1]
* [Location of bubble tea shops from other franchises (Foursquare API)](https://developer.foursquare.com/docs/places-api/) – To understand bubble tea shop location trends and to score clusters negatively
* [Location of Mass Rapid Transit (MRT) Stations](https://www.kaggle.com/yxlee245/singapore-train-station-coordinates) – To map MRT station locations and relate them to bubble tea shop location trends
* [Location of Shopping Malls (Foursquare API)](https://developer.foursquare.com/docs/places-api/) – To map shopping mall locations and relate them to bubble tea location trends
* [Population of Target Demographic (20 - 44 years old)](https://www.singstat.gov.sg/find-data/search-by-theme/population/geographic-distribution/latest-data) – To relate target demographic population in Subzones to bubble tea shop locations [2] [Reference](https://www.statista.com/statistics/1126754/china-age-distribution-of-fresh-tea-beverage-consumers/) [3]
* [Median Income of Residents](https://www.singstat.gov.sg/find-data/search-by-theme/population/geographic-distribution/latest-data) – To relate median income and bubble tea shop locations [4]
* [Aggregation of Dwelling Types](https://www.singstat.gov.sg/find-data/search-by-theme/population/geographic-distribution/latest-data) – To relate household dwelling types and bubble tea shop locations  [2][5]

[1] Locations scraped from official website

[2] Population Trends > Singapore Residents by Planning Area, Subzone, Age Group, Sex and Type of Dwelling, June 2011-2020

[3] Assumption: Share of bubble tea consumers in Singapore in 2021 is similar to that of China in 2019

[4] General Household Survey 2015 > Resident Working Persons Aged 15 Years and Over

[5] e.g. HDB Flats, Landed Property etc.

## Python Modules

* [Pandas](https://pypi.org/project/pandas/) — Data manipulation and analysis library
* [Folium](https://pypi.org/project/folium/) — Mapping module that visualises data on Leaflet.js maps
* [Beautiful Soup](https://pypi.org/project/beautifulsoup4/) — Used for web scraping
* [Geopandas](https://pypi.org/project/geopandas/) — Adds support for geographic data to pandas objects
* [geopy.geocoders.Nominatim](https://geopy.readthedocs.io/en/stable/#nominatim) — Function that returns OpenStreetMap data when given an address
* [Shapely](https://pypi.org/project/Shapely/) — Geometric object manipulation
* [sklearn](https://scikit-learn.org/) — Machine learning library
* [Matplotlib](https://pypi.org/project/matplotlib/) — Data visualisation library
* [Branca](https://pypi.org/project/branca/) — Folium element/colour manipulation
* [requests](https://pypi.org/project/requests/) — To send HTTP requests
* Miscellaneous: math, os