INTRODUCTION

1.1 Background

Bubble tea is a tea-based drink originated from Taiwan in the 1980's (Martin, Laura C., 2007), it is served typically with tapioca balls called "pearls" or "boba" which adds a unique texture to the drink. There are many varieties of the drink with a wide range of flavours and toppings, with the two most popular varieties being: "black pearl milk tea" and "green pearl milk tea" (Chang, Derrick, 2012). The sweet, creamy milk tea paired with chewy "boba" has proved to be an addictive combination amongst the people.

Taiwan has an established market for bubble tea, having many successful brands throughout the country. In recent years, the popularity of bubble tea has skyrocketed and made its way internationally with many Taiwanese bubble tea brands opening their franchises to the global market, including Malaysia. The people of Malaysia tend to favour bubble tea places as a staple 'go-to' hangout spot, instead of the usual cafes. This overwhelming trend continues to lure more bubble tea brands into Malaysia.

1.2 Business Problem

Throughout the year of 2019, Malaysia has seen a high influx of bubble tea chains entering its market. This project aims to utilise the data science methodology and techniques to analyse, if a bubble tea brand intends to open a chain in Kuala Lumpur, Malaysia, where would be the ideal location to set up a store?

1.3 Target Audience

New bubble tea brands looking to venture into the Malaysian market or even existing bubble tea chains in Malaysia looking to expand their store locations, would be the target audience of this project. The analysis provided by this project will give an idea to interested stakeholders, of the level of market saturation in a particular area in Kuala Lumpur, Malaysia.

DATA

2.1 Data Sources

The following sets of data are needed to carry out this project:

- *List of neighbourhoods in Kuala Lumpur, Malaysia*

  A list of neighbourhood names in Kuala Lumpur is essential to the scope of the project. The source of this data can be found in a Wikipedia page regarding Kuala Lumpur (https://en.wikipedia.org/wiki/Kuala_Lumpur), a sub-page located at the bottom, called 'Kuala Lumpur metropolitan area', contains a list of neighbourhood names that are present in Kuala Lumpur, arranged in table format.

- *Geographical coordinates of the neighbourhoods*

  This project will be utilizing Foursquare location data which requires the latitude and longitude coordinates of each neighbourhood to be used as inputs. The geographical coordinate data can be extracted using Geopy's Geocoder library in Python.

- *Bubble tea shops in each neighbourhood*

  To answer the business problem, analysing the level of market saturation would be the focus of this analysis. Foursquare API will be called to search a venue (neighbourhood) for existing bubble tea shops in the neighbourhood's vicinity. This enables us to collect data on the total number of bubble tea shops in a neighbourhood, as well as number of unique bubble tea shops. Following that, K-means clustering will be done to segregate the neighbourhood into different cluster labels, which would give us a rough idea on the level of market saturation per neighbourhood. The resulting clusters will be superimposed onto Kuala Lumpur's map, as part of the data visualization process.