For this week, you will required to submit the following:

1. **A description of the problem and a discussion of the background. (15 marks)**

It is always so hard to have an own business, especially when it comes to having a shop at a good location. Bubble tea is a famous beverage and food in Asia. One of my friend would like to start up a bubble tea shop in Germany but don't know exactly how to do. It motivates me to analyse by using what I have learned from this specialisation. Therefore, I will talk about how shall we find a good location for starting up a bubble tea shop in this project. This project will focus on locations in cities in Germany.

As assumed, the distribution of venues nearby the bubble tea shops in big german cities should play an important role, which deserve to be analysed, and see how the other cities, which haven't had bubble shops yet, are similar to the cities in venues distribution. If they have high similarity, we can assume that open a new bubble tea shop in these cities should be a good decision.
This project also can be seen as a good way for thinking how to choose and compare difference places for business.

2. **A description of the data and how it will be used to solve the problem. (15 marks)**

In this project, I will use data from

i. Wikipedia, in order to get the rank of cities in population, and its location
ii. Foursquare, in which the current running bubble tea shops in Germany, their information and venues near them can easily be found. Besides, we will also fetch venues information of some cities in order to realise if they are potent to host a bubble tea shop.

Step 1. create a cities list by population with latitude and longitude. This will be the basis info for teaching data from foursquare

Step 2. search all the bubble tea shops in Germany with important information, such as name, coordinate, city and their ratings and reviews on Foursquare. The number of ratings and

reviews are very important, because we see them as a pointer of popularity. Here we will select the top 15 biggest cities as reference.

Step 3. label them in a reasonable way. Here I set four classes. The ZERO means the hottest shops. The THREE means the worst.

Step 4. search venues nearby them, in order to know how venues nearby the spots look like.

Step 5. wrangle venues for machine learning process

Step 6. feed venues and labels to machine learning algorithm, to get the classification label, here we will use decision tree due to its high accuracy so far.

Step 7. create another data frame, in which the other top 15 cities with venues nearby are included and seen as test set for machine learning.

Step 8. after processing by machine learning, visualise them on a map