In Houston, My friends and I are looking to open a Chinese healthy breakfast bar, and the main goal of the project is to find the optimal location.

1. Background and Vision.

I am concerned that obesity is an increasing trend in the US. Obesity prevalence in Texas is greater than the national average. Obesity and its complications bring enormous health and economic burden to our community. Obesity is generally caused by eating too much fat and sugars and moving too little. Traditional Chinese breakfast foods are usually low in sugar and fat. In order to promote and advocate a healthy life style, I want to open a Chinese healthy breakfast bar in Houston, my hometown. I will focus on developing low sugar snacks and soy milk.

The main goal of the project is to find the optimal location. Since Houston is an immigrant city and has a diverse culture, people here should be open to healthy changes in life styles. Houston’s neighborhoods are very diverse and each has a unique character. Therefore, it is expected that different areas will have different levels of demand for Chinese Breakfast Foods. Using Foursquare data, I would like to identify the neighborhoods in which Chinese foods are popular.

2. A description of the data and how it will be used to solve the problem.

I will explore the foursquare API to extract the data for all Chinese restaurants within a 6 mile radius around my home address. The responses and reviews for each restaurant in different locations will be collected. The responses contain the ratings, the price tiers, and the number of likes. The response data will be integrated with latitude and longitude coordinates values to cluster the current Chinese restaurants by using unsupervised k-means methods. Each cluster would theoretically group together restaurants by geographic proximity and various metrics of performance. By analyzing the average characteristics of each cluster, I hope to be able to pinpoint the area with most potential.

1. A link to your Notebook on your Github repository, showing your code. (**15 marks**)

With map view

<https://nbviewer.jupyter.org/github/guoweilleexh123/data-science-capstone-project/blob/master/final%20project-breakfast%20bar%20%281%29.ipynb>

Github

https://github.com/guoweilleexh123/data-science-capstone-project/blob/master/final%20project-breakfast%20bar%20(1).ipynb

1. A full report consisting of all of the following components (**15 marks**):

* Introduction where you discuss the business problem and who would be interested in this project.
* Data where you describe the data that will be used to solve the problem and the source of the data.
* Methodology section which represents the main component of the report where you discuss and describe any exploratory data analysis that you did, any inferential statistical testing that you performed, if any, and what machine learnings were used and why.
* Results section where you discuss the results.
* Discussion section where you discuss any observations you noted and any recommendations you can make based on the results.
* Conclusion section where you conclude the report.

3. Your choice of a presentation or blogpost. (**10 marks**)