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

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

Answer:

description of the problem:

How to use Foursquare location data to evaluate the customers satisfaction of each shop. We want to know the customers satisfaction level of each shop comparing to the average level or shops nearby, for example, lower rating and lower popularity; or same rating and lower popularity.

discussion of the background:

Evaluating the customers satisfaction of each shop is an essential problem for business owners and operators. It is a very important factor to make business decision such as opening a new shop or market strategy. Anyone in the business management team of Starbucks or other coffee shops (such as owner, manger), or need making a business decision (such as determining if and how to marketing for a shop) will be interested in this report.

There are many methods to evaluate customers satisfaction (such as a survey for customers) but all methods have different kinds of limitations. On the other hand, we usually compared the interested shop with other shop to get "relative" satisfaction level. But each shop is different than others and their customers satisfaction, such as number of reviews are usually influenced by many factors (for example, shops may have different customer populations). Simply comparison between different shops can be biased or misleading.

With the development of the big-data, more user generated review data (such as Foursquare or Yelp) were collected for each shop. Those data provided a direct way to compare different kinds of scores (such as user rating) between shops in the same category. When integrating with location information (nearby shops), many bias factors can be adjusted (for example, shops in the same location have the same customer population), and the score between shops will be comparable and a direct reflection of customers satisfaction of the shop. This can be an important criterion for shop evaluation.

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

Answer:

Description of the:

The same New York neighborhood information as in the course will be used to extract neighborhood information. Foursquare location data for shops in each neighborhood will be queried. Then all

coffee shops in the shop list will be selected and Foursquare API will be used to extract all information/scores related to each shop, including Rating, count of Rating, Count of Likes, and Count of Tips.

how it will be used to solve the problem

Data will be summarized in mean, median, and IQR for each group. Data will be visualized by different methods including boxplot, histogram, and heatmap. Log transformation will be performed in count data to make them distribute more normally. The scores for each Starbucks shop will be compared to other coffee shops to identify Starbucks shops with significantly worse or better rating, or higher or lower popularity (count of Rating, Count of Likes, and Count of Tips). To adjust other factors such as location, we defined two types of "other coffee shops" to compare with each Starbucks shop: (1) Other coffee shops within the same Neighborhood; (2) Five other nearest coffee shops; t test will be used to identify if the Starbucks shop has a significantly lower or higher score (for example, rating) than "other coffee shops". If test P value was less than 0.05 in comparing with any of the two types of "other coffee shops", the score will be considered as significantly different. Then all Starbucks shops will be classified based on their score difference in rating and popularity.