QMSS GR5059, Spring 2017
Team 4
Course Project

1. **Project Description**

   For the final project of QMSS G5069, Team 4 will utilize the Yelp user dataset in order to uncover insights about consumer preferences in Yelp business perspective. More specifically, we want to investigate if consumer preferences vary geographically.

2. **Insight**

   In this project, we will examine for insights answering such as "do people who live in cities with a higher population density value service speed more than people who live in rural locations?"

3. **Research Strategy**

   3.1) Literature Review

   As an intermediary step, we will also look at what makes a review useful. By understanding the characteristics of the most helpful reviews it will help us to give more relevant weights to certain reviews that may be more impactful businesses.

   3.2) Research Method

   In order to answer this question, we will employ different data and text mining techniques, starting with exploratory analysis to look at how business ratings by categories vary across geographic locations. We will then seek to build a predictive model with rating as the dependent variable, to see if there is an interaction effect between geographic location and the coefficient on attributes of the customer's experience (e.g. reviews highlighting speed, taste, accuracy etc.).

4. **Data**

   Team 4 will use a public data set ("challenge dataset") provided by yelp for "Yelp Dataset Challenge".

   Within the challenge dataset, Yelp provides 5 types of information – business, consumer, review text, check-ins and tips. We are likely to use the business dataset for general business attributes, and the review and tips texts for text analysis.

   The advantages of the Yelp data are that it is relatively clean and provides a large enough corpus for text analysis (reviews and tips). However, due to limited regional coverage of data -- data only covers 11 cities in 4 countries that the demographics might be relatively homogeneous--, we expect insufficient variation in consumer preferences for analysis.

5. **Output**

   Through investigating consumer preferences, our output will allow businesses to understand quantitatively the impact of different attributes/topics/keywords of reviews and how it affects their overall rating on the Yelp platform. We may also uncover intermediary insights regarding characteristics of useful reviews.

   Regarding research parameters, we will refer to QMSS GR 5069's suggested timeline, announced by professor Morales.