# CS591 Data Mining Term project Proposal

**Team Members:**
Renqing Gao - gladius@bu.edu
Amal Kadi - akadi@bu.edu

**Dataset:**
We will be using the data from the Yelp Dataset Challenge
(https://www.yelp.com/dataset_challenge).  The set has ready data of multiple cities in
multiple parts of the world (cities in: UK, Germany, Canada & US). The dataset consists
of json objects of businesses, reviews, users, tips, and check-ins.

**Scope:**
We will mainly be concerned with analyzing the reviews of the users and the businesses
along with the timing of check-ins.  The data we will be working on has the reviews of
different users to different business along with the dates of these reviews. Also, we
have some details on the businesses like opening hours, address, categories,
attributes, and average rating. The check-in data has the number of check-ins during
each hour of the week for each business.

**Applications & Expected analysis and results:**
 From the points below we can infer what is important in a business for people in that
area, so a future business owner can make use of these attributes.
- Get cultural trends and differences between businesses in different parts of the
  world; things like opening hours, available parking, payment methods.
- Infer time patterns for each business and area from check-in times.
- Analyze seasonal trends for a business based on the dates of the reviews, or get
  businesses that people prefer in different seasons.  We might also apply
  clustering algorithm to the business so that we can find out seasonal trends in
  different cities or even in different districts.
- Clustering to show users that are similar to each other (provided reviews of the
  same businesses/type of businesses) and giving recommendation based on that.

**Note:**
This proposal is based on techniques and ideas we covered in class, as we progress in
the course we might switch to new/more techniques.