

Project title: **Exploratory Analysis of Las Vegas from Yelp Data**

Team members: Chengxiang Zhao u0879155

Yishuai Du u0884588

Xinbo Wang u0930578

Team members give feedback: Salama, Mazen(mazen.salama@utah.edu, u1074476)

Lee, Wonkyung(u1214948@utah.edu, u1214948)

General Comments

The Topic sounds very interesting and suitable for the group project of DS class. You are planning to extract features by crawling the Yelp website and do explanatory analysis in Las Vegas. The five hypotheses are as below:

- (1) Which region in Las Vegas has more local businesses?
- (2) What is the type of businesses type which has the biggest number of stores?
- (3) Which business has the most reviews?
- (4) Which business has the highest star?
- (5) Find out the authenticity of tourist reviews and its reliability

For the analysis plan, you could reveal the features available on the website and the essential features you are going to use on Yelp and from other source data. In our understanding, Yelp is open source data so it seems to be pretty safe for ethical consideration unless you target some specific customer. For the stakeholder analysis, you may think out some people who got influenced by this analysis who could be customers, the current business owners as well as a business start-up.

Data Acquisition and Cleanup

The data acquisition is likely to be realistic but you should check their policy for data scraping. You could get enough data for your plan unless there is some limit for data scraping. It requires pretty high efforts for scraping, clean-up and pre-processing of data. So I hope you endure a long time to be ready for analysis and see the interesting results.

Analysis Methodology

You could apply some visualization methods for exploring the data and letting other students understand the outcome of your project at a glance such as Geospatial Plot.

You planned to make matrix correlation to show the relationship between local businesses and city information such as income, population, education. If you can find enough time, you could finish the multivariate analysis using some machine learning methods, such as decision trees, SVM as you planned it. But it could be helpful to clarify how you define the reliability of the store.