**Proposal of Project 2--ETL**

By KC Oh, David Owens, Ling Lv

**Background**

Review is one of metrics to measure the popularities of restaurants. Positive reviews indicate that a restaurant offers amazing customer service, an attractive atmosphere and top-quality of food. The good reviews can also attract millions of people to a restaurant’s existence and produce an uplift in sales. There’s are different review sources that customer can access, among them, google reviews and yelp are two trustable sources of reviews.

**Specific objectives of the project**

Georgia Institute of Technology (Georgia Tech) is one of the centers of education, business, and tourism in Atlanta. The objective of this project is to extract data from google reviews and yelp about restaurants within 1 mile walking distance of Georgia Institute of Technology (Georgia Tech). The data include the restaurants name, restaurant type, number of reviews, reviews stars, ect. Finally, we will inner join these two data sources by the restaurant name, then calculate the total reviews and average review stars of these two review sources average score. We will store the final dataset into csv and Mongodb/postgresql.

**Material and Methods:**

The method we are going to use is web scraping.

**Deliverables:**

Web scraping code

csv including data extracted from google

csv including data extracted from google

sql query code

csv file containing the final dataset

**Resources:**

Google review

yelp