

# Yelp Star is not Everything

*Sau-Chin Chen*

## Introduction

Yelp has been collecting the users' review for the businesses from the major cities at North America and Europe since October 2004. This data set *business* collects the information from 61184 businesses across 10 cities. Data set *checkin* stores 45166 durations and frequencies a user stay in every business. Data set *review* summarizes 1569264 reviews for a registered business. Data set *tip* summarizes the 495107 recommendations for a business. Data set *user* collects the information from 366715 registered users.

[Yelp's recommendation engine](#) has been designed to filter the active and available review articles for a business. A couple of attributes were summarized from the data accumulated in the past decade. The data set *business* has 38 variables in which 7 variables are data frame. If these variables constituent the evaluation model how yelp users evaluate a local businesses, will we figure, for a category of business, a global standard across cities or many local standards in each city? In use of [Bayesian network](#), we set up the criterion to answer this question. When the Bayesian network showed that the city is not the child node of any variable, there is no a unified model for all the businesses of a category across cities. Otherwise, there is a unified model to tell people how to evaluate a kind of business based on yelp stars or a specific attribute.

## Method

### Preprocessing Data

### Exploratory Analysis

### Build Model(s)

## Result

## Discussion