Business Closing Prediction based on Yelp Dataset

# Abstract

Founded in 2004, Yelp is a local-search service powered by crowd-sourced review forum. The company trains small businesses on how to respond to reviews, hosts social events for reviewers, and provides data about businesses. Often, a review describes various dimensions about a business and the experience of the user with respect to those dimensions. For instance, Yelp reviews judge the quality of services that are provided by local businesses, and sometimes people refer to before hanging out with friends. One thought would be to discover the impact of reviews on a business activity.

This project will focus on how to predict whether a business will remain open or going to close based on the Yelp dataset.

I will use a subset of the Yelp Challenge Round 12 dataset <<https://www.yelp.com/dataset>> and apply text classification and sentiment analysis to predict if a business is likely to close or not. I will use various tools like R, SQL, Python and Spark to find predictors of a business closure and train a model to make the prediction.