# Stat222 Research Proposal

Hui-Fang Ko, Hsin-Wei Tsao

Feb 3, 2014

## 1 Introduction

## 2 Resources (Dataset)

### 2.1 Restaurant Scores in San Francisco

This is a dataset provided by the San Francisco of Department of Public Health at `https://data.sfgov.org/Public-Health/Restaurant-Scores/stya-26eb`. The Health Department has developed an inspection report and scoring system. After conducting an inspection of the facility, the Health Inspector calculates a score based on the violations observed. So this dataset contains information of 6073 restaurants in San Francisco area and their score in the inspection conducted by the department.

### 2.2 Yelp API

With Yelp API, we can get the consumers' reviews of those restaurants listed in the dataset above. Moreover, we can get some details about the restaurants including price, utilities(ex: Wi-Fi), etc..

### 2.3 Others

More information about the area (by zip code) including average housing price, criminal rate, etc..

## 3 Overall Research Questions

1. Is there any relationship(positive or negative) between the restaurant scores by Health Department and consumers' preferences on Yelp? Also, do they have the same distribution?

2. How does location influence on restaurants' score?

# 4 Approach

We are going to use Chi-Squared Test and Linear Regression Model to exam the relation between two kinds of scores. For the model of restaurant location, we will try some General Lineal Models like Logistic Model.

# 5 Anticipated results

1. Those two scores might have positive relationship, but might not have same distribution.
2. We believe that locations have good influence on restaurants. So maybe we could build a food map recommending nice restaurants with both high ratings on Yelp and nice, clean environment.

# 6 Reference


1. 10 things to Know About Choosing a Restaurant Location `http://restaurants.about.com/od/location/a/10-Things-To-Know-About-Choosing-A-Restaurant-Location.htm`
2.