Stat222: Weekly Update

Hui-Fang Ko, Hsin-Wei Tsao

Feb 9, 2014

This week, we thought about our model in detail and tried to collect data we need.

The Models We Intend to Build

As we mentioned in the proposal, we are going to build a model to evaluate restaurants. There are few methods we wll use:

1. Linear Regression: We set the customers review score on Yelp as the response variable, and explanatory variables are restaurant score by Public Health Department, area crimal rate, housing price, numbers of reviews, texts in reviews, wifi, price rate, hours, take credit card or not.

2. Clustering: Use the same variables as above.

3. RFM Model: apply the RFM Model to analysis customers' reviews on yelp. The elements we will use are:

   - Recency - How recently did the customer post review?

   - Frequency - How often do they post?

   - Monetary Value - How many texts in reviews?

Data We Need

1. Yelp API: customers' reviews(including scores and texts), wifi available, price rate, hours, take credit, zipcode.

2. Restaurant Scores in San Francisco: The scores, address(zipcode)

3. Others about neighborhood: area crimal rate, housing price.