Project Proposal

Yuanyuan Lin

11/23/2019

Topic Description

I would do Yelp Star Rating Prediction based on the grading from customers.

Dataset

Yelp Data Analysis Project contain multiple datasets("user","business","tip").

Content

1.Data Exploration(EDA)

2.EFA

Since there are too many variables in "user" dataset in predicting star rating for resturants. Exploratory Factor Analysis mignt be used to reduce the dimensionality of dataset.

3.Clustering Analysis for Restaurant Star Rating Segmentation

Using K-means clustering to divide a reastaurant base into groups of individuals that are similar in specific ways relevant to market, such as restaurant category, review counts and stars.

4.Text analysis(sentiment analysis) on customers review

Text analysis of customers review include finding the most common words in all the reviews. Using the most common words to get the corresponding restaurants list.Then, I would implement sentiment analysis using tidy data principles.

5.Mapping for yelp restaurants

Using maps to locate yelp restaurants in OH and IN. Compare the density of restaurants between two states.

6.Topic analysis on restaurants description text

Using NLP,especially LDA could help in identifying major topics on restaurant description

7. Shiny app

The shiny app is going to present OH,IN restaurant mapping, major EDA plots from my analysis project. Viewer is able to explore the specific content of clustering analysis for restaurant star rating segmentation, and sentiment analysis on customers review.

8.Presentation

Presentation is going to be made using r ioslide. Presentation includes graphs from EDA and modeling.