

Business Analysis of Yelp reviews for cinema

Group-16

Chufan Zhou, Mingyu Wang, Xiao Li

December 8, 2022

Introduction

- ▶ Analyze data from Yelp and get some useful and analytic insights.
- ▶ Focus on the cinema industry and three aspects including movies, food and ambiance.
- ▶ Give data-driven action recommendation/plan to the business owners to improve their ratings on Yelp in the future.

Data Preprocessing

- ▶ Filtering out all business whose categories contains “cinema” and get the corresponding reviews.
- ▶ Tokenize the text, make the segmentation lowercase without numbers and punctuations.
- ▶ Exclude stopwords.
- ▶ perform word stemming and lemmatization.

Recommendations Based on Statistical Analysis

ANOVA method for attributes indicators

- ▶ We select 5 attributes indicators as our candidate variables to check whether they are related to user's rating.
- ▶ For each variable, we ignore the NAs and consider the stars as response to conduct ANOVA between False and True value. Set $\alpha = 0.05$ and finally we obtain 3 effective variables(Table).

Table: Analysis of variance

attributes	mean of True	mean of False	Pvalue
BikeParking	3.93	3.45	0.044
OutdoorSeating	4.25	3.67	0.00408
GoodForKids	3.57	4.02	0.0035

Feature Extraction

- ▶ Pick the words that appear in at least 5 percent of comments.
- ▶ Sum each word's rank in “distance score” with its rank in “absoluted stars score”, then get top several words as important factors.

$$\text{distance} = \sum_{i=1}^5 |O_i - S_i| \quad \text{stars} = \frac{1}{5} \sum_{i=1}^5 S_i - A$$

- ▶ O_i refers to overall distribution of star rating
- ▶ S_i refers to distribution of star rating of specific word.

Sentimental Analysis for Review Text Segmentation

- ▶ Our next part considers making use of the review texts by doing sentimental analysis sentence-wisely with syuzhet package.
- ▶ There will be three parts needed to consider: movie, ambiance and food. For every part, we choose key words as below

Table: key words

Parts	key words
movie	price,expensive,cheap,imax
ambiance	dirty,clean,crowd,park,service,bathroom
food	food,popcorn,snack,drink,water,beer,soda

Conclusion

- ▶ We give suggestions based on two aspects, one is from ANOVA of the business indicator variable, the other is from sentimental analysis of reviews.
- ▶ We expect the business owners can get clear barplots of their performance and some corresponding advice easily by Shiny.

Thank you!