

Gyms Data Analysis Part 1

STAT 628 Module 3

—

Group 6

Chunyuan Jin
Chen Qian

Hao Pan
Haoxiang Wei

1.

Introduction & Motivation

Introduction & Motivation

Aims:

- Construct a model
- Give suggestions
- Improve ratings

Why Gyms?

- Keeping fit -- popular
- Data amount -- moderate
- 2440 gyms
- 46159 reviews on gyms

2.

Data Preprocessing

Remaining Features

- Necessary for merge
- Possibly useful
- Transformation

(eg: *Friends*

One's Friends' IDs→Friends Amount)

Deleted Features

- Meaningless
- Hard to deal with
- Repeat information

(Merge reviews, business
and user dataset)

3.

Text Processing

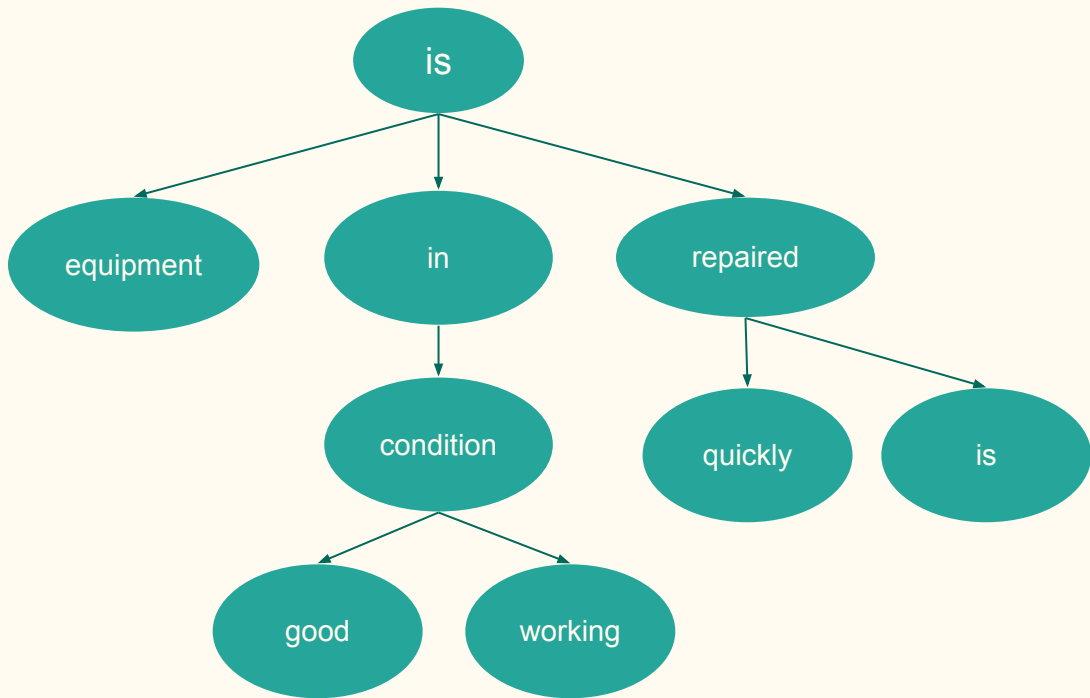
1. Split each review into sentences
2. For each sentence, based on dependency to extract the essence of each sentence

Raw sentence

The equipment is in good working condition, or is repaired quickly.

Essence of this sentence

```
{'root': ['is'],  
'is': ['equipment', 'in', 'repaired'],  
'in': ['condition'],  
'repaired': ['is', 'quickly'],  
'condition': ['good', 'working']}
```



3. Drop those meaningless “sentences” (like a link or even simply a “a++++”)

```
----Pay
----Too
-The
-This
=)
-Around
Y
This
it
...
...
;)
http://reebokcrossfitramsay.com/getting-started/!Hope
;)
http://reebokcrossfitramsay.com/getting-started/!Hope
there-
there-
Smh
```

4. Why we store the data this way

```
{'root': ['is'],
 'is': ['equipment', 'in', 'repaired'],
 'in': ['condition'],
 'repaired': ['is', 'quickly'],
 'condition': ['good', 'working']},
```


4.

Data Description



The distribution of business average rating

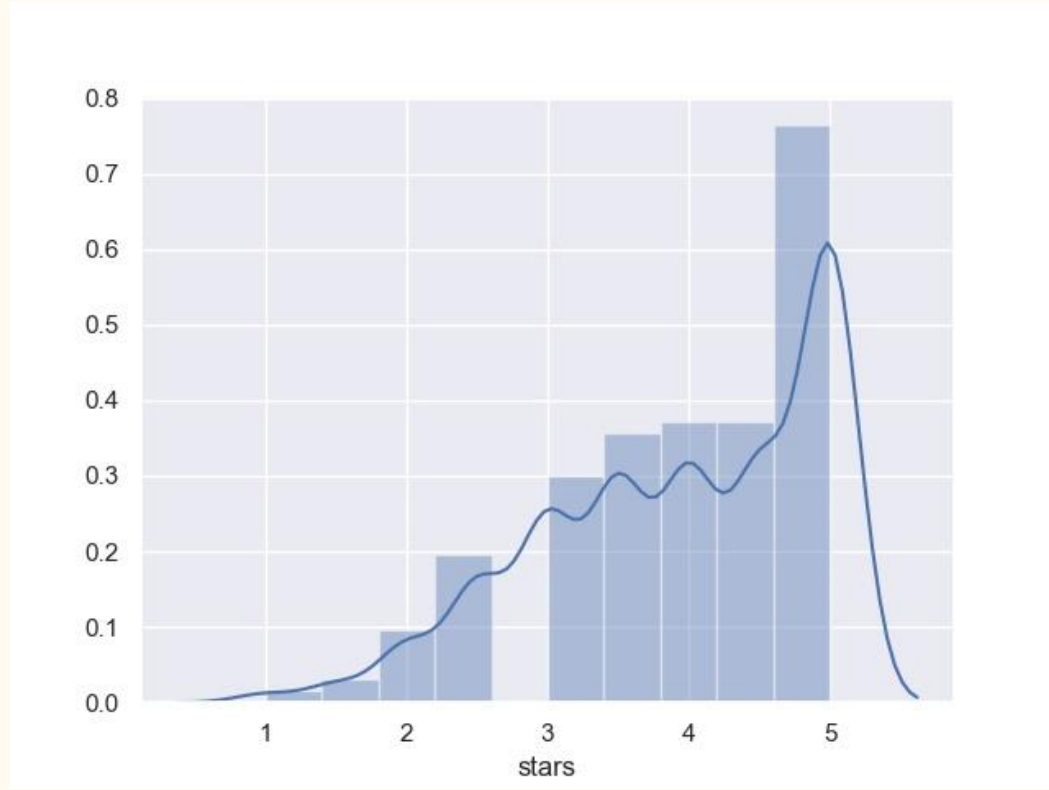


Figure 1.

The distribution of review rating

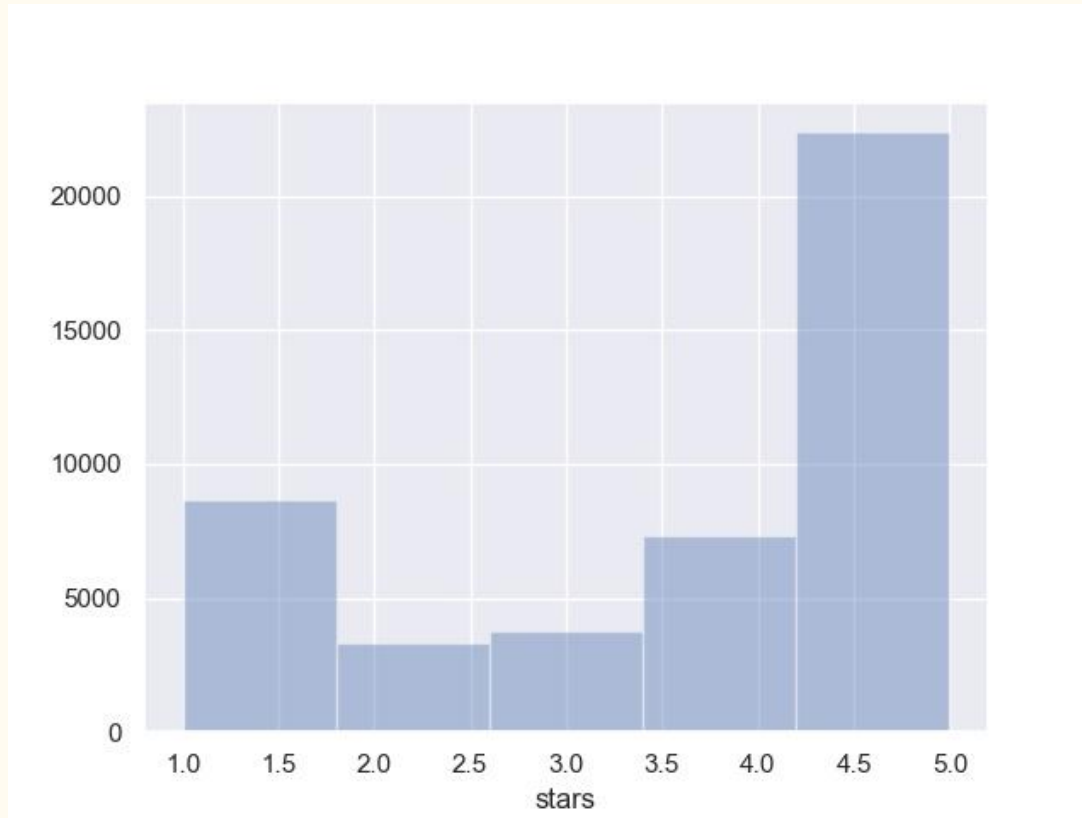


Figure 2.

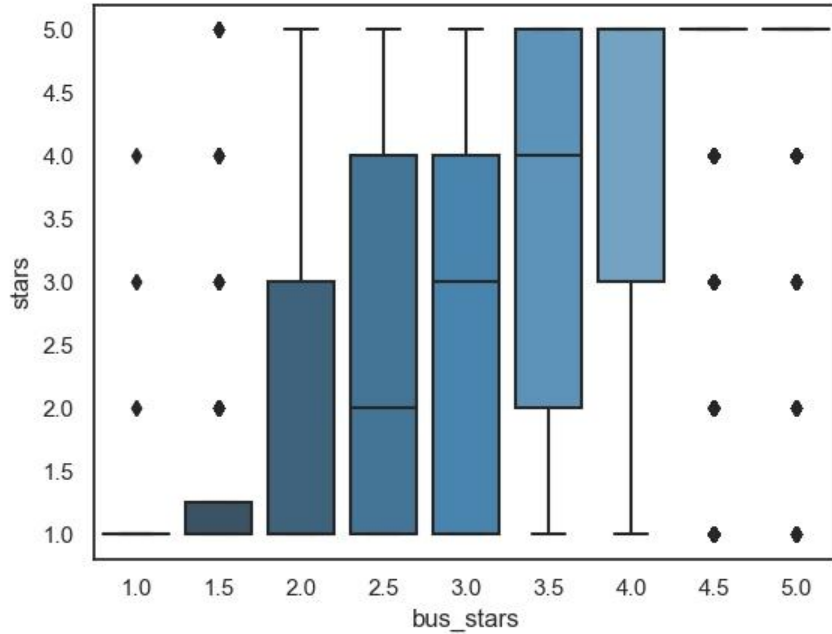


Figure 3.

Box plot of
business rating

Word cloud



Keys:

Time

Class

Trainer

Location

Customer service

Locker room

Figure 4.

Word cloud in high rating and low rating review



Keys for 5 stars:

time\class\trainer\location

love\good\highly recommended\amazing



Keys for 1 and 2 stars:

time\location

staff\member\equipment\customer service

5.

Model Design & Future Work

Model Design

Generate
improving
aspects



Discriminate
the sentiment
of each aspect



Adjust
weight of
each review



Give
Recommendation
for each business



Aggregate
reviews for
each business



Future Work

- Abstract some words such as class, instructor, and location as the improving aspects according to the importance (eg. TF-IDF) of words.
- Discriminate which aspects are mentioned in a review and whether it is positive or negative in the context.
- Take the features such as the number of user's compliments, and the time of reviewing into consideration.

Future Work

- Use these features we abstracted to adjust the weight of each review. For example, the newest review may have a highest weight.
- Create flag variables to record contradictory relationship between stars and reviews sentiment and use such flags to adjust weights of these reviews.
- Aggregate reviews for each gym and provide a final recommendation.