

A photograph of a gym interior. In the foreground, a black 8kg kettlebell sits on a black mat with a circular pattern. The background is blurred, showing people exercising. A large orange diagonal graphic element is overlaid on the right side of the image.

How to improve your gym business?

Thurs Group 7
Yunhui Qi; Qiming Hao; Wanwan Su

Goal:



Explore gym business--- Give suggestions to gym owners to improve their business

1. Do explanatory data analysis
2. Build models to predict stars from reviews
3. Extract features from data and do association rules mining



- 01 > Data cleaning
- 02 > Analysis of reviews
- 03 > Analysis of stars
- 04 > Future plan



Data Cleaning - preparation



- Select the 'Gym' data from data set:
Review_train: 35033 rows 26.1MB
- Split the attributes of business data and add it to the data as columns, such as 'WheelchairAccessible';
- Keep the attributes having missing values less than 35%:
'BusinessParking', 'ByAppointmentOnly', 'GoodForKids';

Data Cleaning - from review to words

Worst Gyms I met. And I
would not come again.

['Worst', 'Gyms', 'I',
'met', 'And', 'I', 'would',
'not', 'come', 'again']

['worst', 'gyms', 'i', 'met', '**and**', 'i',
'would', 'not', 'come', 'again']

['worst', 'gym', 'meet',
'would', 'not', 'come']

Change n't into not

Remove punctuation

Convert to lower case

Lemmatization

Worst Gyms I met.
And I wouldn't come
again.

Split into words

Remove non-alphabetic

Remove stop words
(except for not no nor)

['Worst', 'Gyms', 'I', 'met', '.',
'And', 'I', 'would', 'not', 'come',
'again', '.']

['**Worst**', '**Gyms**', 'I', 'met',
'**And**', 'I', 'would', 'not', 'come',
'again']

['worst', '**gyms**', '**met**', 'would', 'not',
'come']

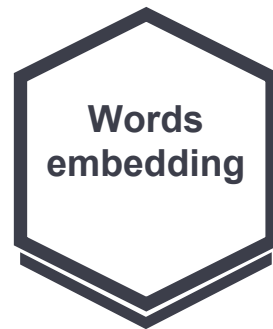
Analysis of reviews wrt words



- ★ Word cloud for overall reviews
- ★ Top 5 frequent words in reviews of different stars



- ★ Top 3 keywords in all review
- ★ Top 3 keywords for reviews of different stars



- ★ Use word2vec in python to create feature vectors for all words

Analysis of reviews wrt words



stars	*	**	***	****	*****
frequency	not	not	not	not	not
	no	like	like	like	great
	membership	no	room	great	love
	month	people	machine	room	like
	tell	room	equipment	machine	work



Analysis of reviews wrt words

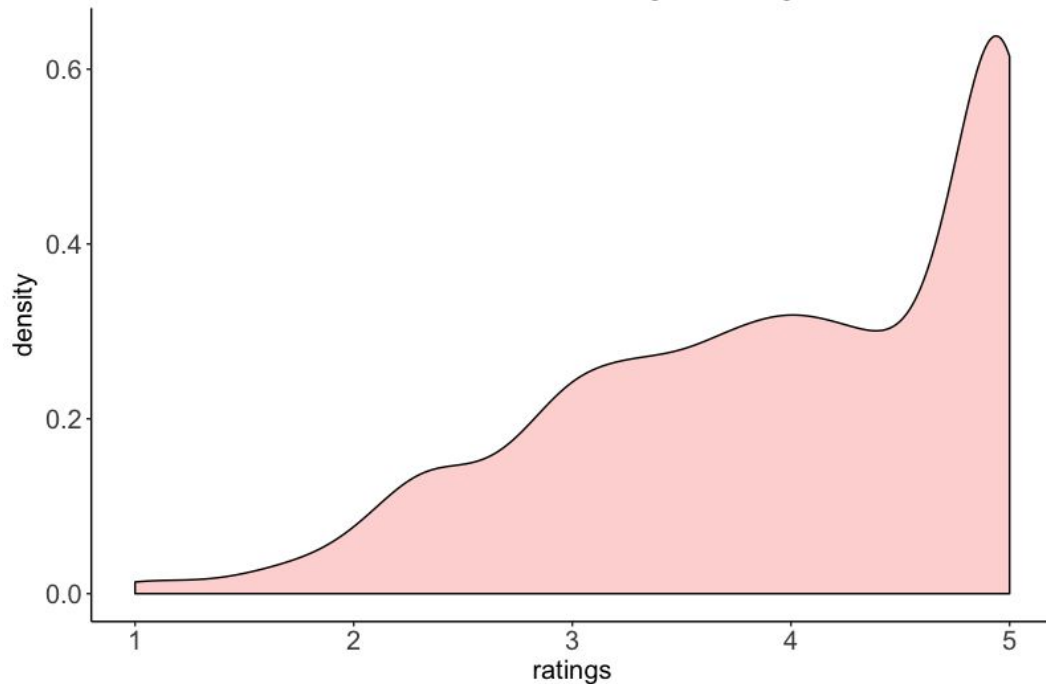


TFIDF	wigwam	funk	normal	valued	love	great
	disgusted	parking	wifi	helpful	convenient	hospitable
	invested	terribly	fix	flawless	overall	affordable
stars	*	**	***	****	*****	all

Analysis of stars - Distribution of Average Ratings



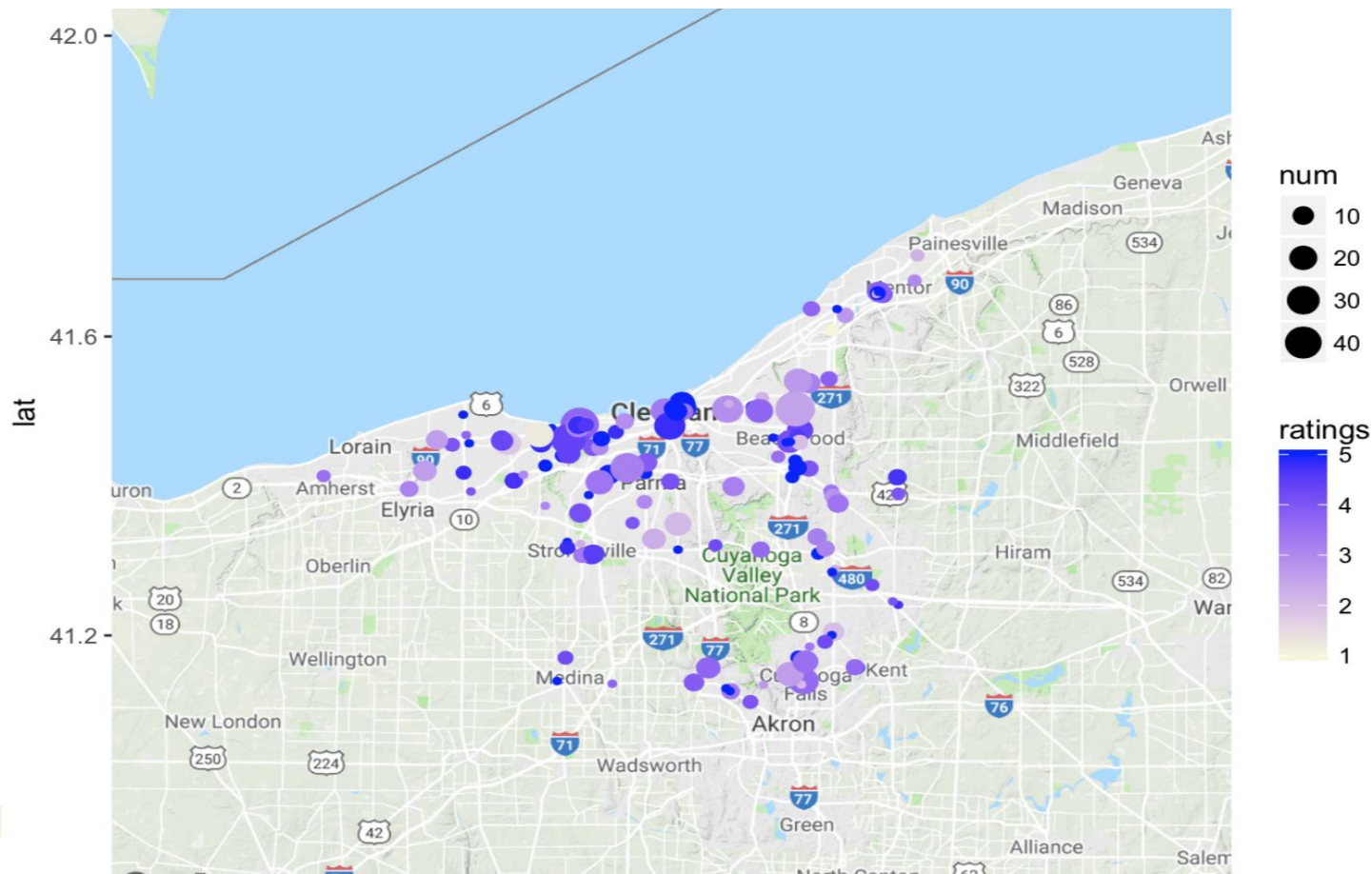
Distribution of Average Ratings



- Calculate average ratings for each business
- Most businesses in gym field have pretty high ratings



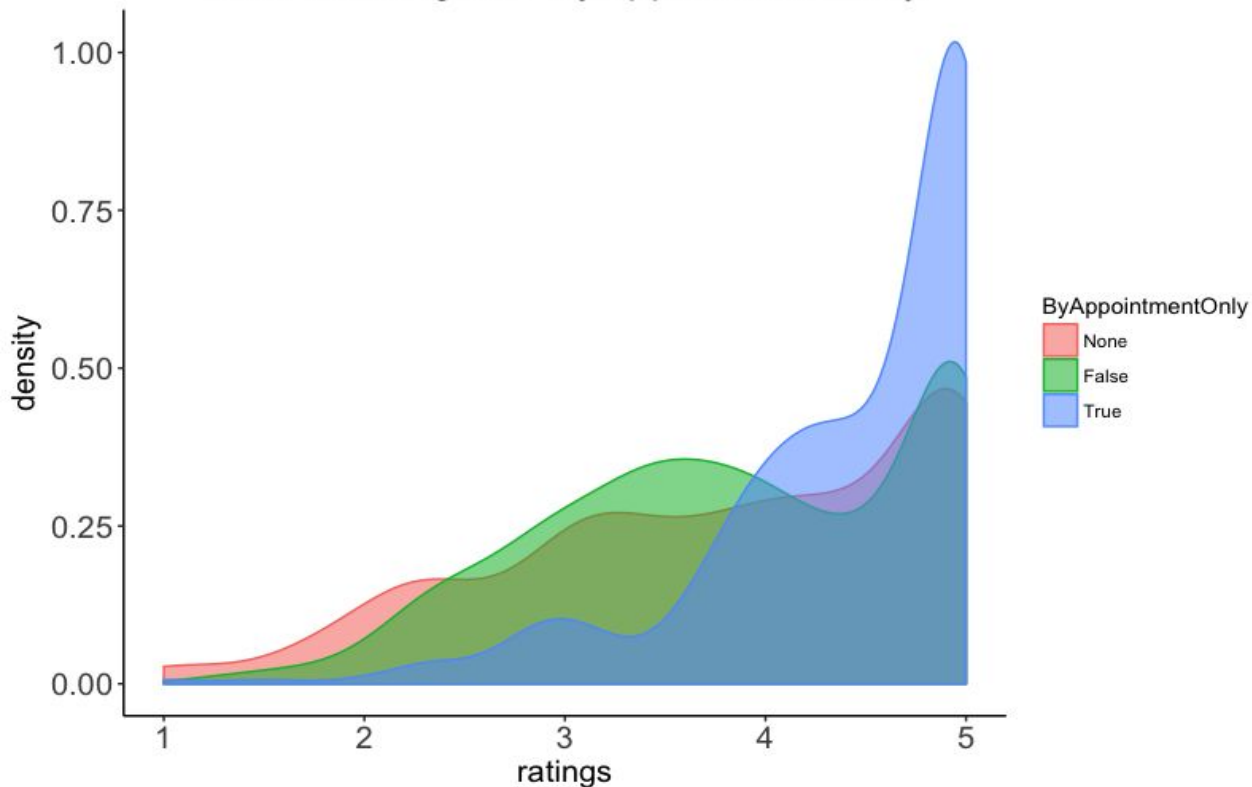
Distribution of gym businesses in Cleveland



Analysis of stars - Ratings w.r.t. attributes



Dist. of ratings for ByAppointmentOnly

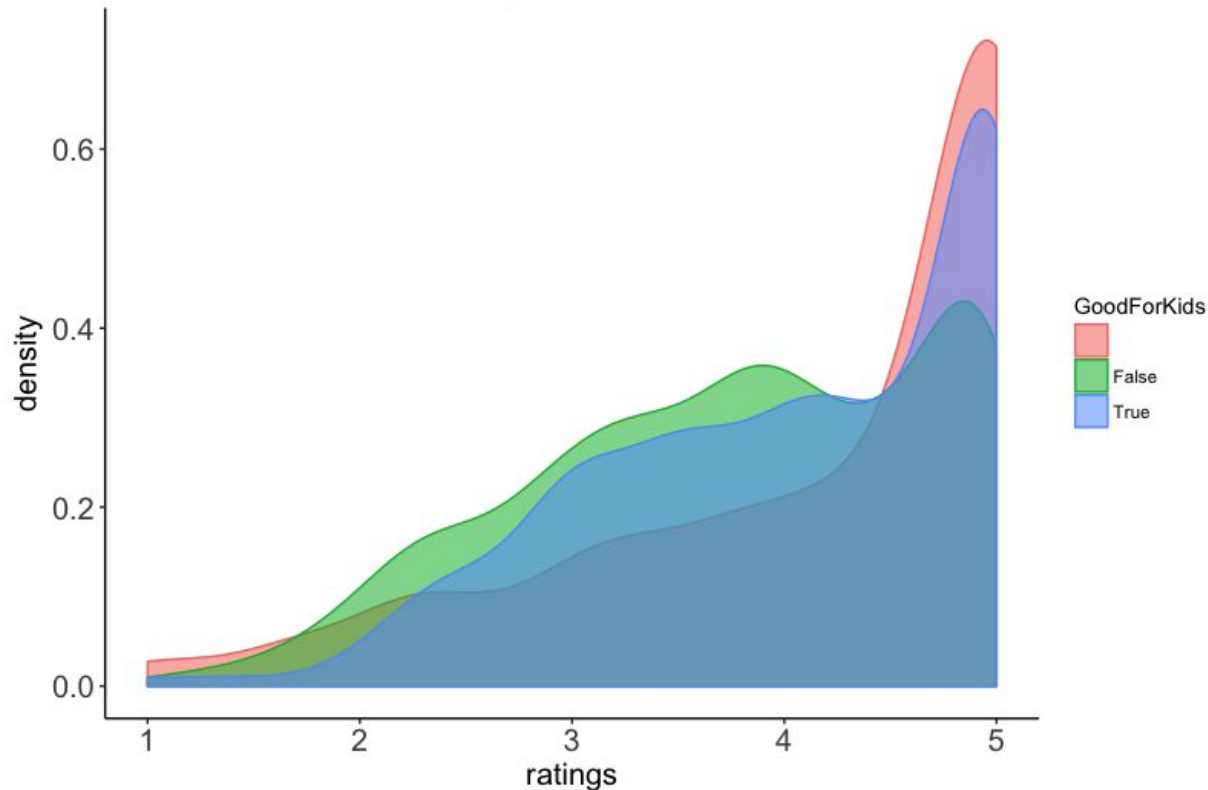


Those require appointment only tends to have higher ratings

Analysis of stars - Ratings w.r.t. attributes

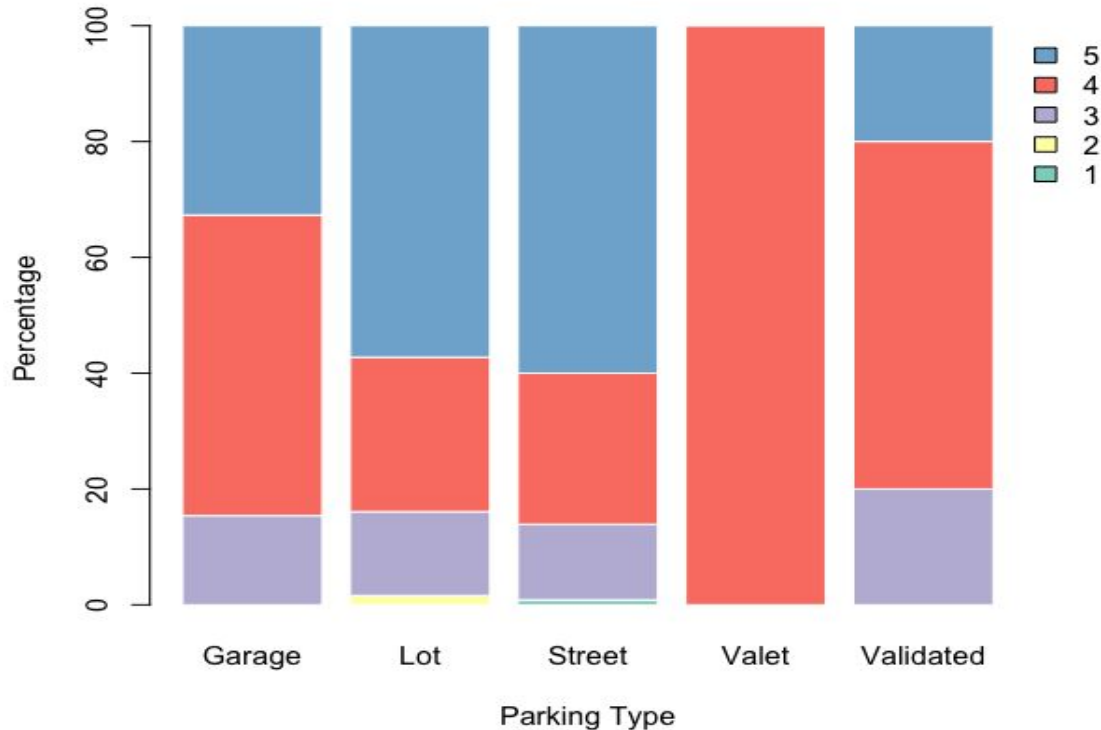


Dist. of ratings for GoodForKids



No significant differences about this attribute

Analysis of stars - Ratings w.r.t. attributes

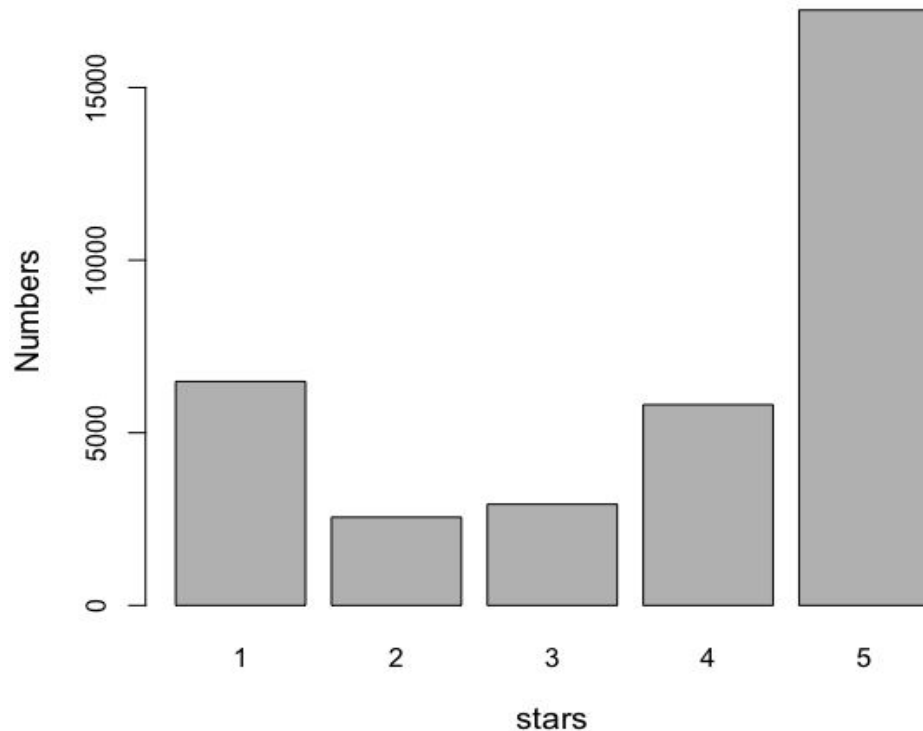


Those have lot parking and street parking tend to have higher ratings

Analysis of stars - Ratings w.r.t. reviews



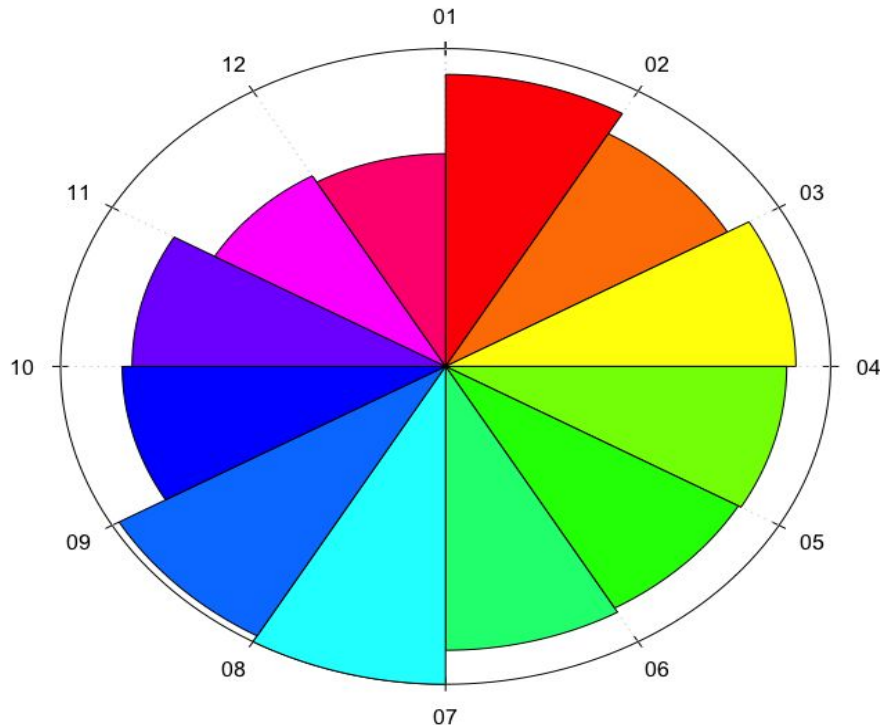
Distribution of Stars in Reviews



Analysis of stars - Reviews distribution by month



Number of reviews for each month of the year



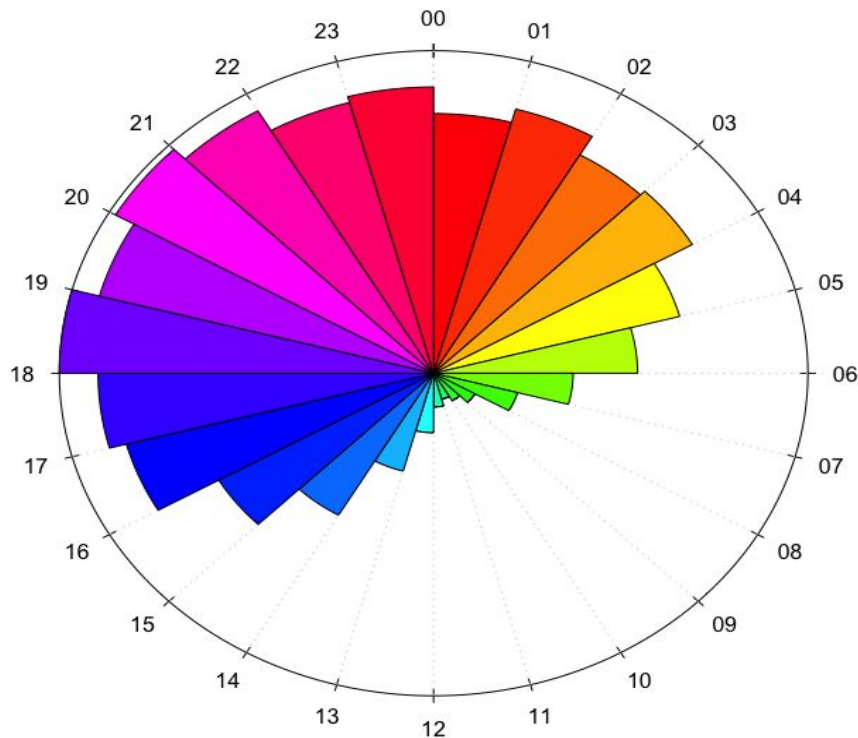
- People tend to start working out in January
- People tend to stop working out in September
- The number of reviews reaches to its maximum in July

*plot function by Vincent Zoonekynd

Analysis of stars - Reviews distribution by day



Number of reviews for each hour of the day



Most people prefer to work out at night

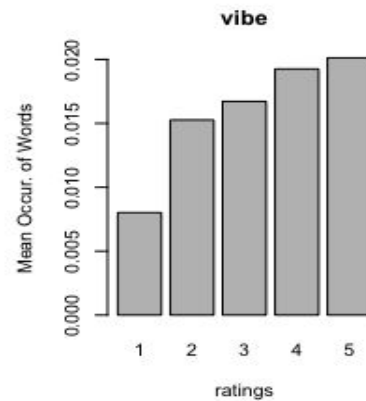
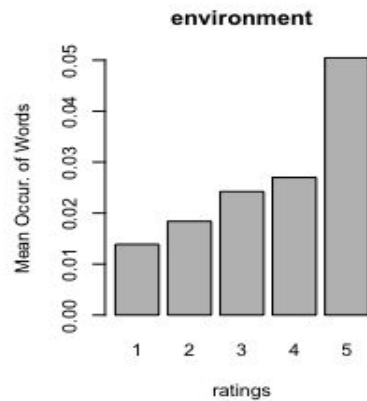
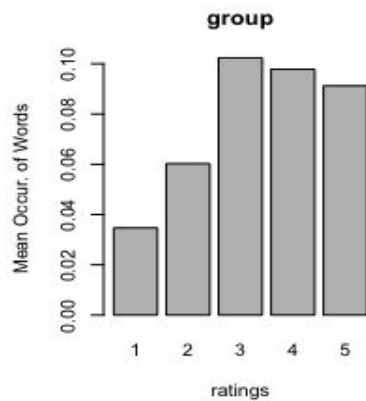
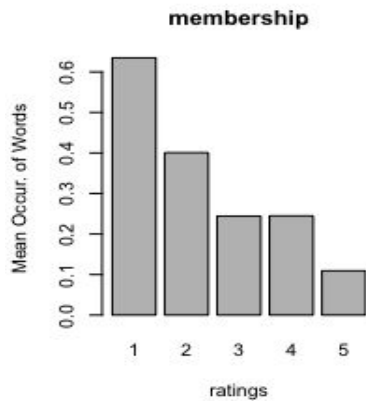
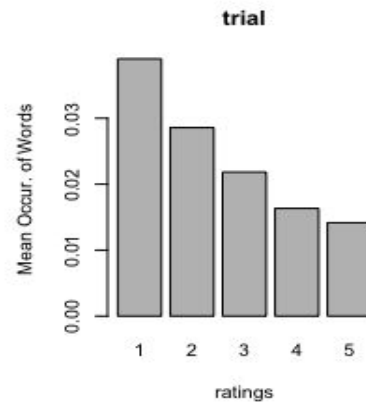
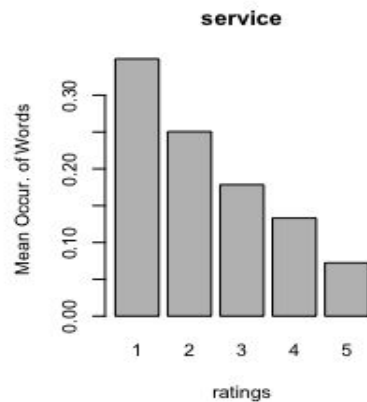
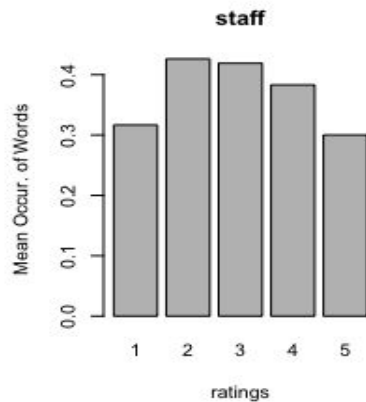
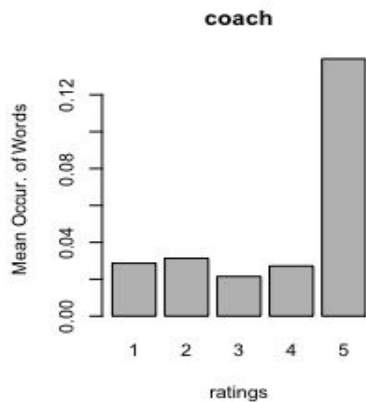
*plot function by Vincent Zoonekynd

Analysis of stars - Mean occurrence of words



- Find some common nouns in gym from background knowledge.
(coach, environment, membership etc.)
- Calculate mean occurrence of each word among ratings.
- Describe potential relationship.

Analysis of stars - Mean occurrence of words



Future Plan



Feature based summary

- Find features from data
- Identify sentences discussing these features
- Identify positive or negative sentence



Prediction Models

- Words embedding and frequency
- Logistic Model
- Neural Network





Thank you !

Q & A