



Analysis on Pennsylvania Restaurant

Chaoran Wang, Lu Li, Qiaoyu Wang, and Yuhan Meng

Outline:

- Topic classification
- Rating for different topics
- Analysis of working hour
- Significant attributes for different topics
- Shiny App

Topic classification:

stars	useful	funny	cool
4	0	0	0
4	4	0	0
5	9	6	9
1	1	0	0
5	0	0	0
4	1	0	1

Lots of 'useful', 'funny' and 'cool'

They may have influence on the weight of each review

Some of people prefer viewing a review as useful to making their own review

Topic classification:

'useful' especially takes a quite large proportion



Take only 'useful' into consideration

$$\text{Conversion rate} = \frac{\text{number of review}}{\text{number of 'useful'}} \approx \frac{\text{willingness of making a review}}{\text{willingness of making a 'useful'}}$$

$$\text{New frequency} = (1 + \text{number of useful} * \text{Conversion rate}) * \text{Previous frequency}$$

Topic classification:

categories
Sandwiches, Salad, Restaurants, Burgers, Comfort Food
Nightlife, Bars, Polish, Modern European, Restaurants, Vegan
Japanese, Sushi Bars, Restaurants
American (Traditional), Restaurants
Beer, Wine & Spirits, Food
Restaurants, Caribbean
Fast Food, Sandwiches, Restaurants
Food, Specialty Food, Health Markets

The categories is complex and hard to distinguish easily

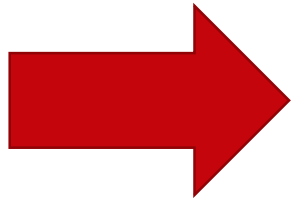
We need some new categories for our further research

Topic classification:



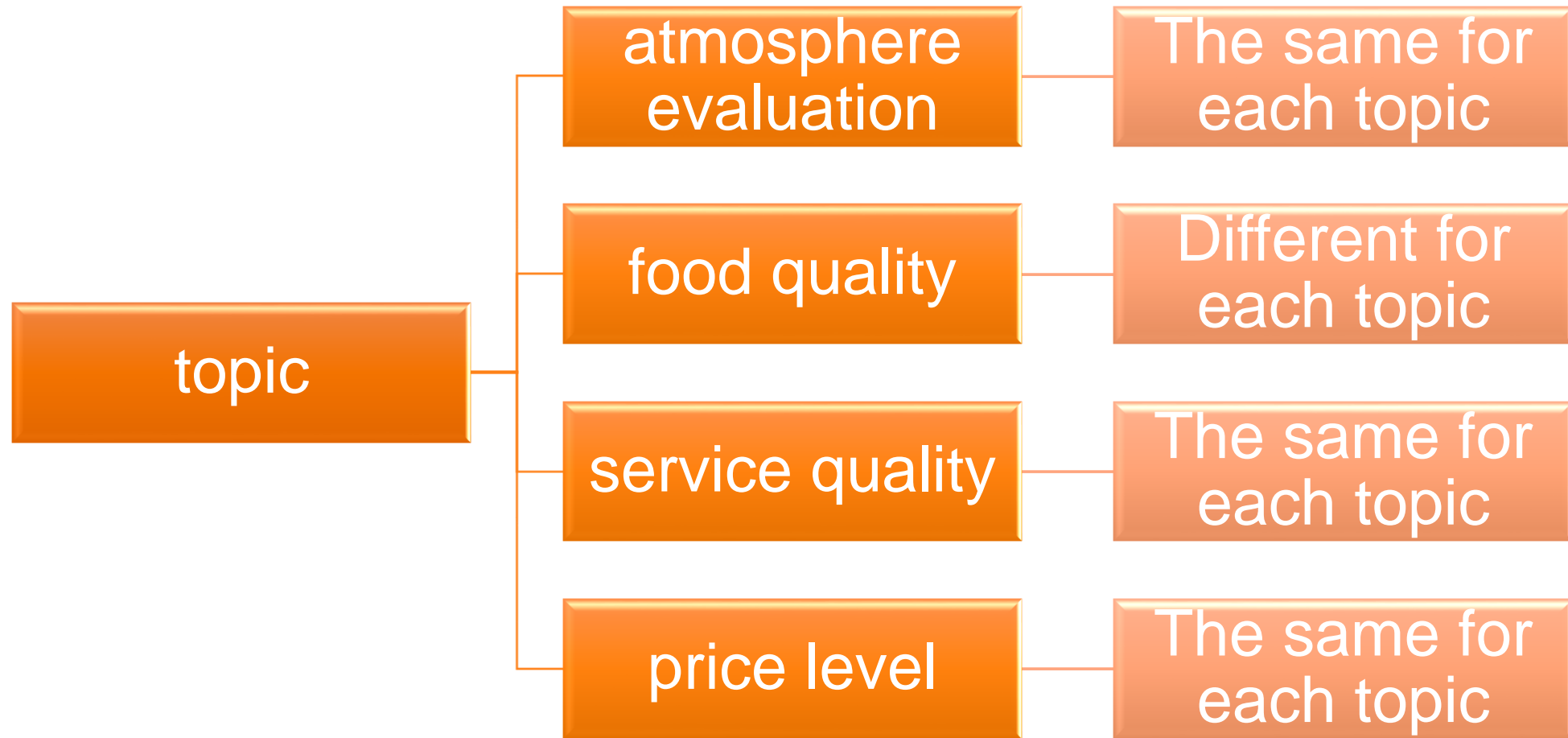
Topic order	LDA result	Topic we name it
1	sanwich get ti go fri coffe friend no breakfast tri	Brunch
2	bar ti drink beer servic menu us restaur tabl realli	Bar
3	cream ice chocol flavor tri cake coffee get also love	Dessert
4	pizza sauc ti salad restaur get chees go tri servic	Fast food
5	chicken restaur taco ti fri dish get also go sauc	Foreign flavor

Topic classification



- Rating for different topics
- Analysis of working hour
- Significant attributes for different topics

Rating for different topics:




Rating for different topics:

Order the
frequency.



- Place:50189
- Food:48057
- Good:45121
- Sandwich:38962
- ...

Find nouns and
adjectives from the
first 500 most
frequency words.



Classify
nouns into 4
aspects

- atmosphere:['place','area','spot',...]
- service=['servic','staff', 'wait','NOTwait'...]
- priceword=['price']
- food :
- Bar topic:['bacon','beer','sausage', 'drink'...]
- Desert topic:['cream','cupcake','bread'...]
- Foreign flavor topic=['thai','sushi','rice','taco'...]
- ...

Rating for different topics:

- Cut the sentences to find the adjectives around nouns.
- Sum up the the frequency of different adjective for each aspect for each restaurant .

Eg:

Review1 for Pizza hut :The wings is so good but pizza is okay.

Review2 for Pizza hut : The pizza is so good. I love this salad whose flavor is good.
But the order is slow.

wings good || pizza okay || pizza good || salad flavor good

Wing:good:1
Pizza:okay:1,
good:1
Salad:good:1
Order:slow:1

Pizza hut :food {good:1,okay:1}
atmosphere{}
service{slow:1}
price{}

Rating for different topics:

- Construct linear model to compute the weight for adjectives.
- Keep the adjectives with absolute coefficient larger than 0.01.

Y (Stars)	NOTbad	NOTgood	...	awesome
3.5	19	11	...	5
4.5	0	0	...	6
4	4	3	...	13

$Lm(Y \sim X)$

Y: the star level for each business,

X: frequency matrix only about adjective words.

Rating for different topics:

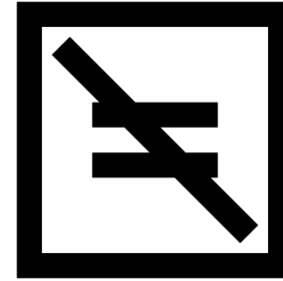
- Compute the score for each business in four aspects separately and use the min-max normalization.
- Find the top ten restaurants in each aspect in each topic.

names	stars	food_scores
NU Jewish Bistro	4.00	1.00
Jamison's On West Liberty	4.5	0.99
Aurochs Brewing	5.0	0.98
Arsenal Cider House & Wine Cellar	5.0	0.94
Full Pint Brewing Company	4	0.94

The top five food quality in Bar topic

Analysis of working hour :

Increasing working time



High star

the best Working time schedule



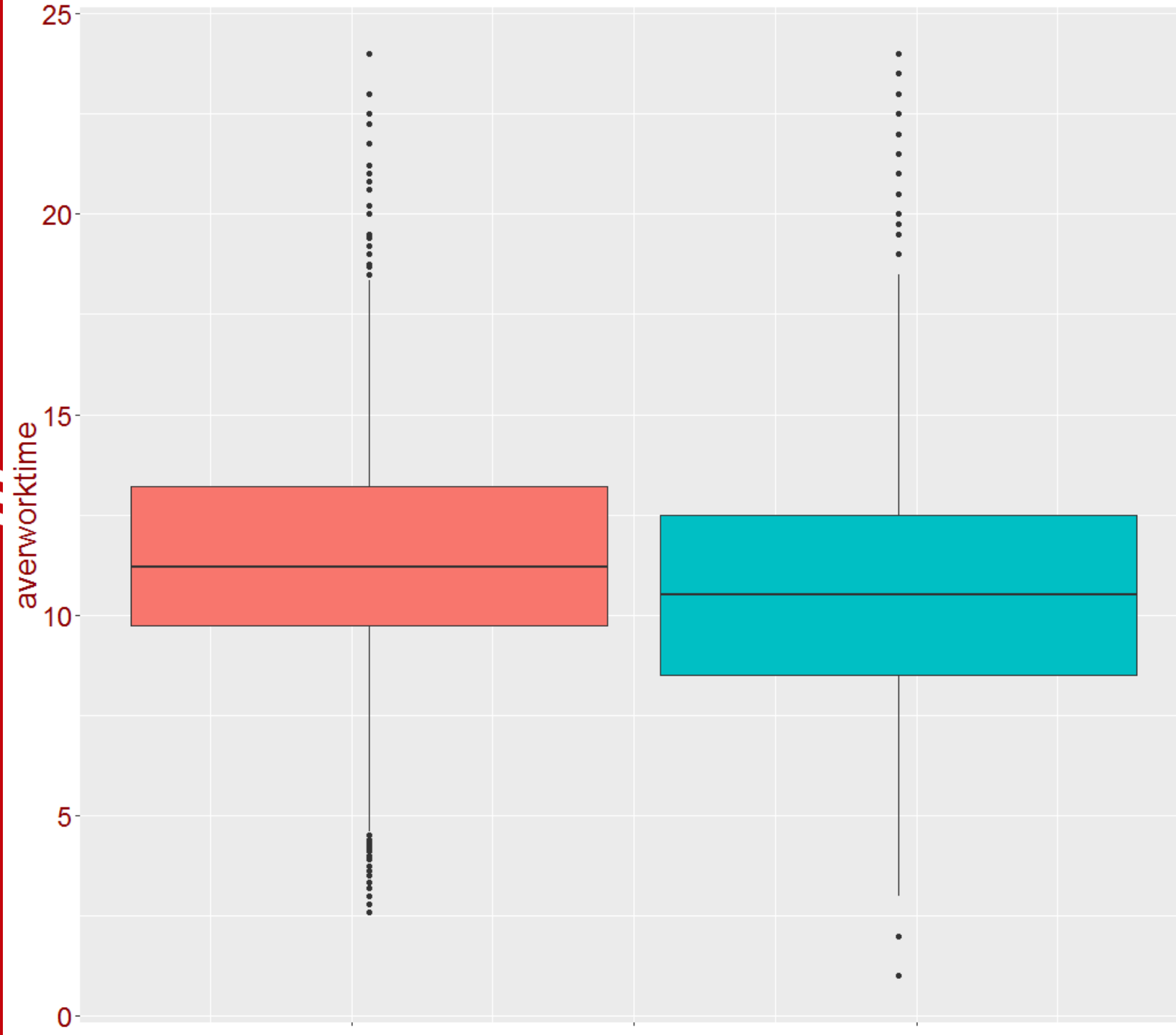
Brunch

Bar

Dessert

Fast food

Foreign flavor

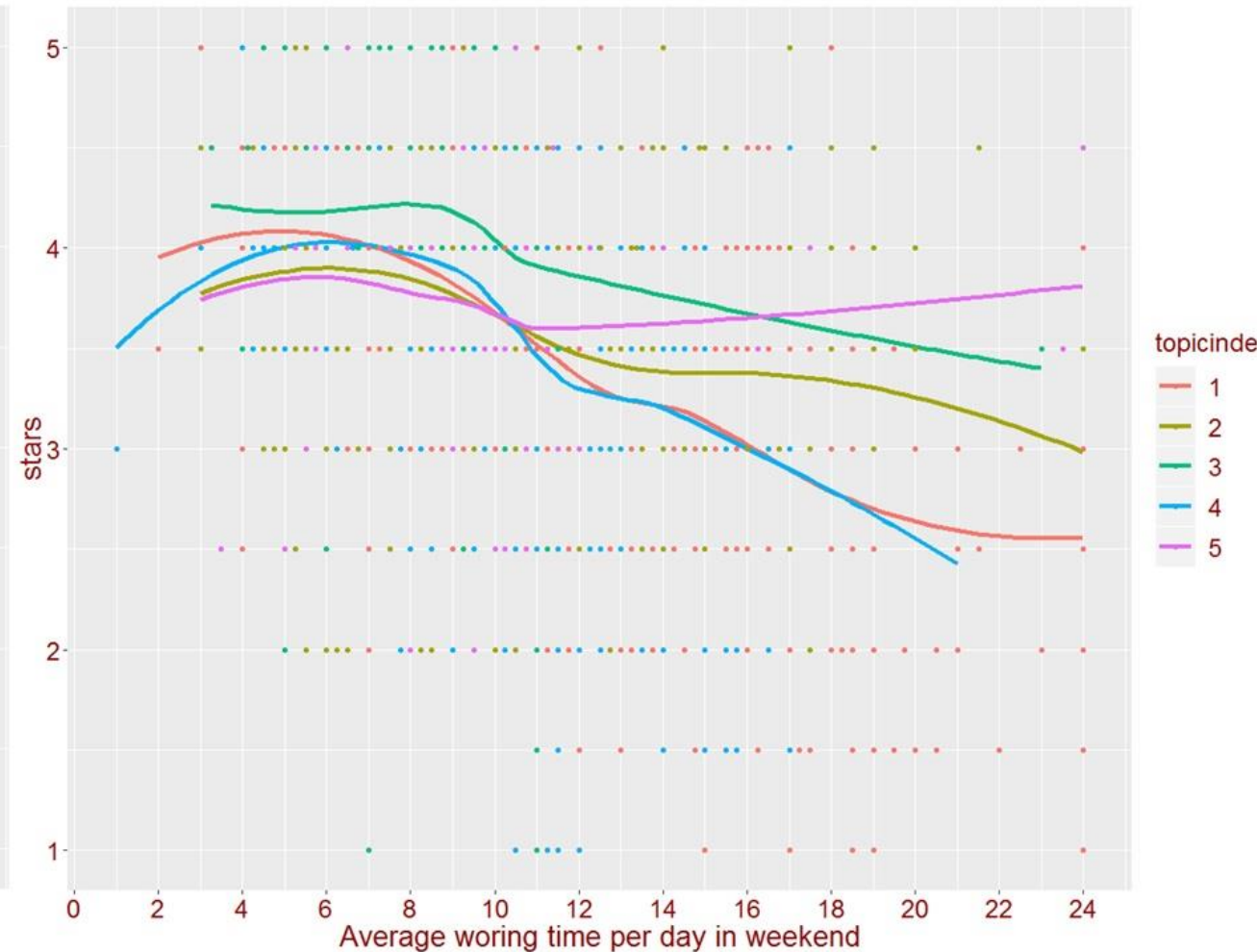
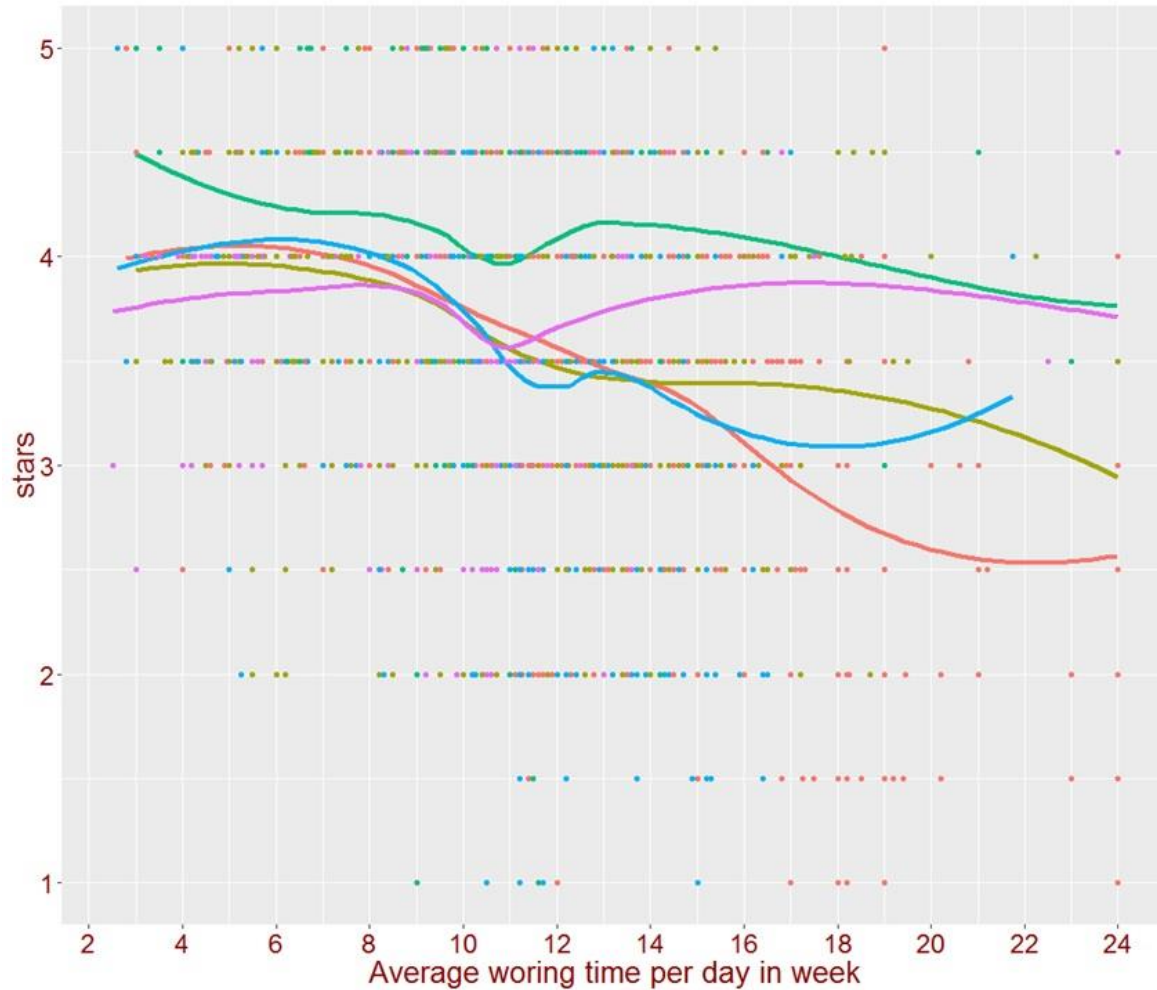


3303 restaurants work
in both weekday and
weekend

week
weekend

Paired t-test:
p-value < 2.2e-16

Analysis of working hour :



locally weighted scatterplot smoothing method

Analysis of working hour :

Topic id	Topic name	Working time in Weekdays	Working time in Weekends
1	Brunch	4-7 hours	4-7 hours
2	Bar	4-7 hours	5-8 hours
3	Dessert	3-6 hours	6-9 hours
4	Fast food	5-8 hours	5-8 hours
5	Foreign flavor	6-9 hours	5-8 hours

Analysis of working hour :

Suggestions on Working period

experience from business with 4 or 5 stars

Topic id	Topic name	Working period in Weekdays	Working period in Weekends
1	Brunch	12,13,14,15	12,13,14
2	Bar	18,19,20	18,19,20
3	Dessert	13,14,15,16	13,14,15
4	Fast food	18,19	18,19
5	Foreign flavor	12,13,14,18,19,20	17,18,19,20

Analysis of working hour :

significant time period for high star business

Topic id	Topic name	Working period in Weekdays	Working period in Weekends
1	Brunch	6,12,13,14,15	12,13,14
2	Bar	16,18,19,20	16,18,19,20
3	Dessert	9,13,14,15,16,18	9,10,13,14,15,20
4	Fast food	9,10,18,19	18,19
5	Foreign flavor	10,11,12,13,14,18,19,20	10,17,18,19,20,24

Analysis of working hour :

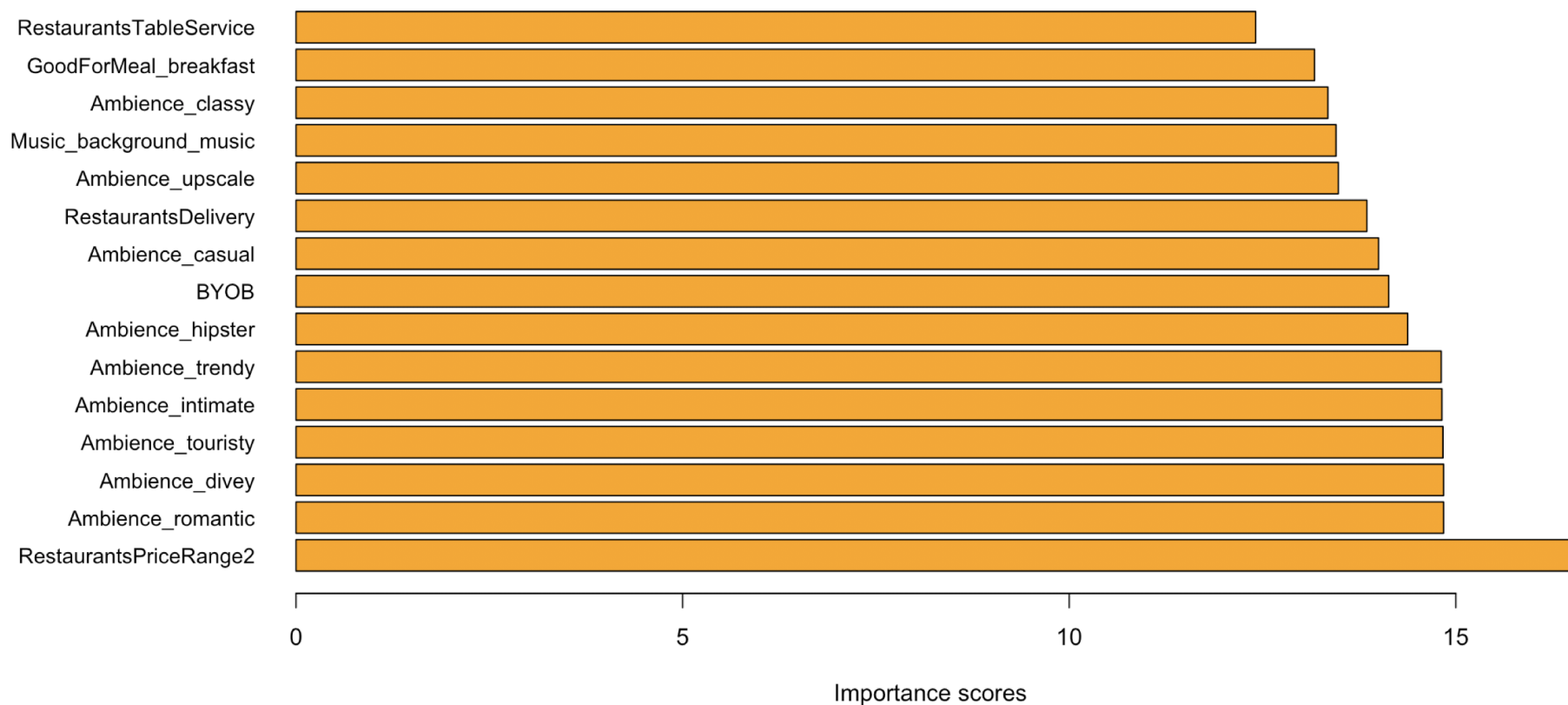
Final Suggestions on Working period

Topic id	Topic name	Working period in Weekdays	Working period in Weekends
1	Brunch	5:00-6:00 and 11:00-15:00	11:00-14:00
2	Bar	15:00-16:00 and 17:00-20:00	15:00-16:00 and 17:00-20:00
3	Dessert	8:00-9:00, 12:00-16:00 and 17:00-18:00	8:00-10:00, 12:00-15:00 and 19:00-20:00
4	Fast food	8:00-10:00 and 17:00-19:00	17:00-19:00
5	Foreign flavor	9:00-14:00 and 17:00-20:00	9:00-10:00, 16:00-20:00 and 23:00-24:00

• Significant attributes for different topics:

- Method: Use GUIDE to select out most important attributes in each groups
- Result:

Topic 1:



- Significant attributes for different topics:

- **Method:**

- Use linear regression and boxplot to detect meaningful relationship between star rating and attributes within each group

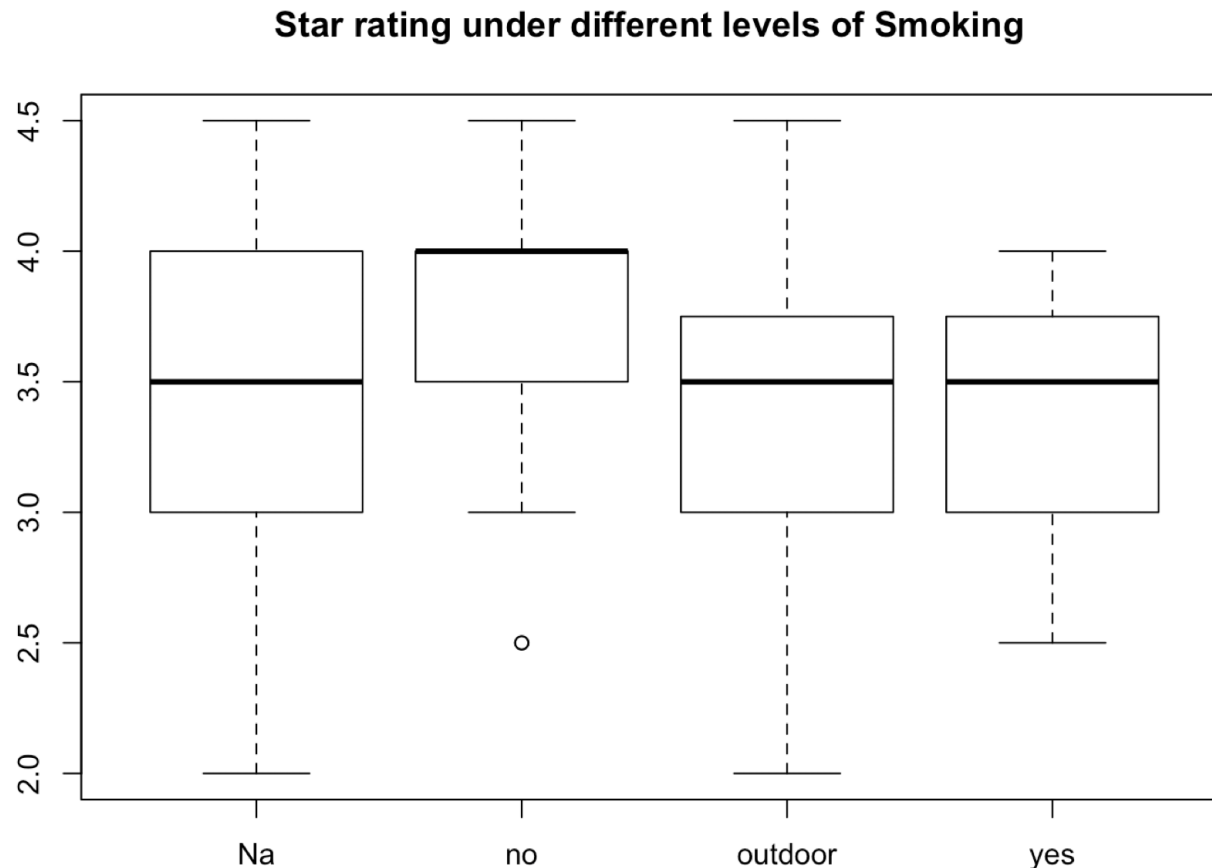
- **Result:**

- Make some suggestions based on the relationship between the star rating and important attributes

- **Suggestions illustration:**

- Significant attributes for different topics:

➤ Suggestions illustration: Bar type restaurant :



- `lm(formula = stars ~ NoiseLevel)`
- Coefficients:

	Estimate
(Intercept)	3.50466
NoiseLevelloud	-0.14275
NoiseLevelquiet	0.02237
NoiseLevelvery_loud	-0.35188

Shiny app

<https://chaoranwang.shinyapps.io/shiny/>