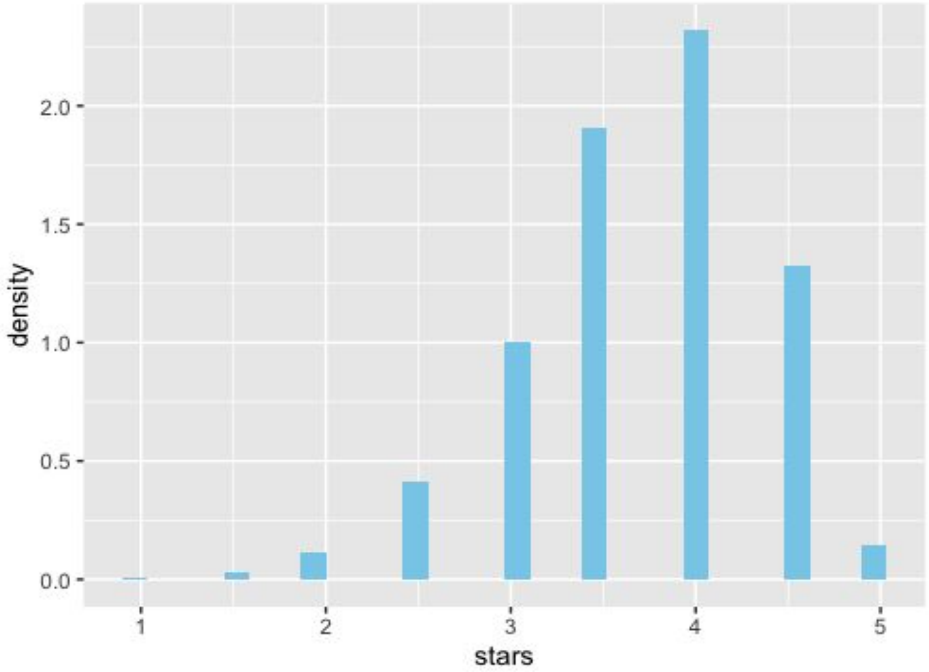
Business Analysis for Sushi Restaurants

Group 13

Outline

- Data selection
- Attributes analysis
- Text analysis
- Suggestions

Data selection



We filtered the restaurants from business.json by categories with SUSHI or JAPANESE and got 2459 restaurants. Based on their business ID, we extracted the reviews of these restaurants.

Attributes Analysis - Data Processing

1. Enumerated the attributes provided by cleaned “sushi.csv” data, totally 39 attributes.
2. Chose the following 7 attributes that most restaurants provided for the analysis.
3. Converted the ‘None’ as missing data (NA).
    - Outdoor Seating (Service): True or False
    - Restaurant Delivery (Service): True or False
    - Wheelchair Accessible (Service: True or False
    - Bike Parking (Service): True or False
    - Noisy Level: Loud or Not loud
    - Free WiFi (Service): True or False
    - Alcohol (Service): True or False

Attributes Analysis - Logistic Regression Model

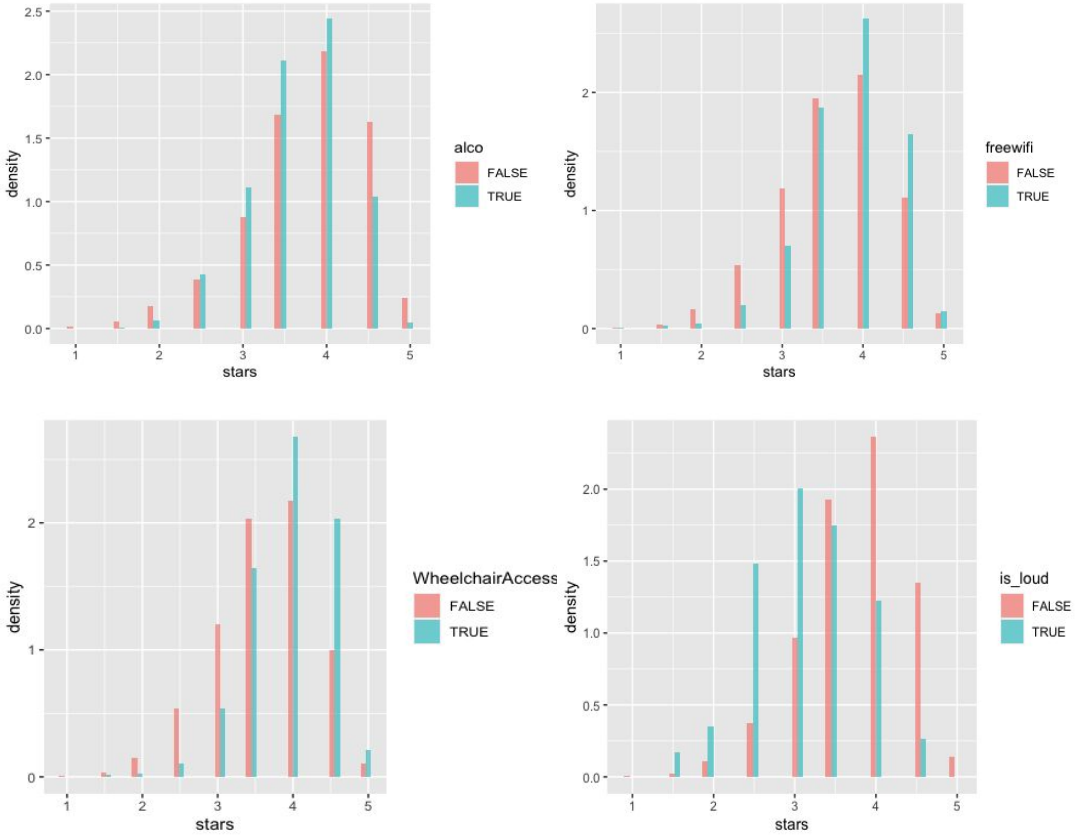
1. Two restaurants categories: a good restaurant (numeric as 1, star >= 4); otherwise, a not good restaurant (numeric as 0).

2. Fitted a logistic regression model, tested the significant effect on the classification (alpha = 0.01).

3. Wheelchair accessible, noisy level, free wifi and alcohol are four significant effects for sushi restaurant ratings (p-value).

Attributes	Coefficient	P-Value
Outdoor Seating (True)	-0.016	0.494713
Restaurant Delivery (True)	0.042	0.040161
Wheelchair Accessible (True)	**0.196**	**< 2e-16**
Bike Parking (Ture)	0.032	0.129031
Noisy Level (Loud)	**-0.305**	**1.74e-08**
Free WiFi (True)	**0.083**	**0.000167**
Alcohol (True)	**-0.111**	**1.21e-07**

Attributes Analysis - Logistic Regression Model



- They are very similar. Note that the 'true' value in the last graph represents a high level of noise.
- More restaurants satisfy these attributes when star >= 4.0, which is the ‘good restaurant’.
- This shows that if a business can provide these services, customers would prefer to give higher a star rating.

Attributes Analysis - General Suggestions

Quieter dining environment ⟶ +0.305, p-value=1.74e-08

Wheelchair accessible ⟶ +0.196, p-value< 2e-16

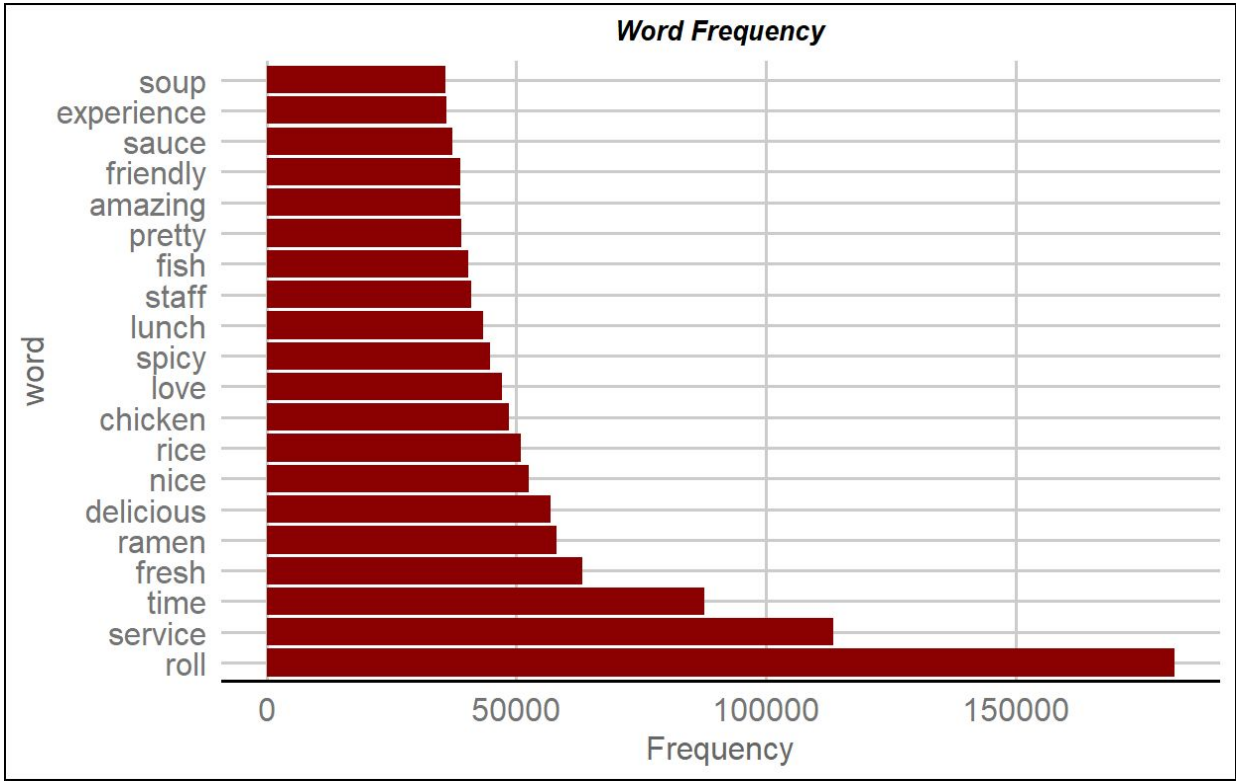
Free wifi ⟶ +0.083, p-value=0.000167

Alcoholic beverages ⟶ +0.111, p-value=1.21e-07

Review Cleaning

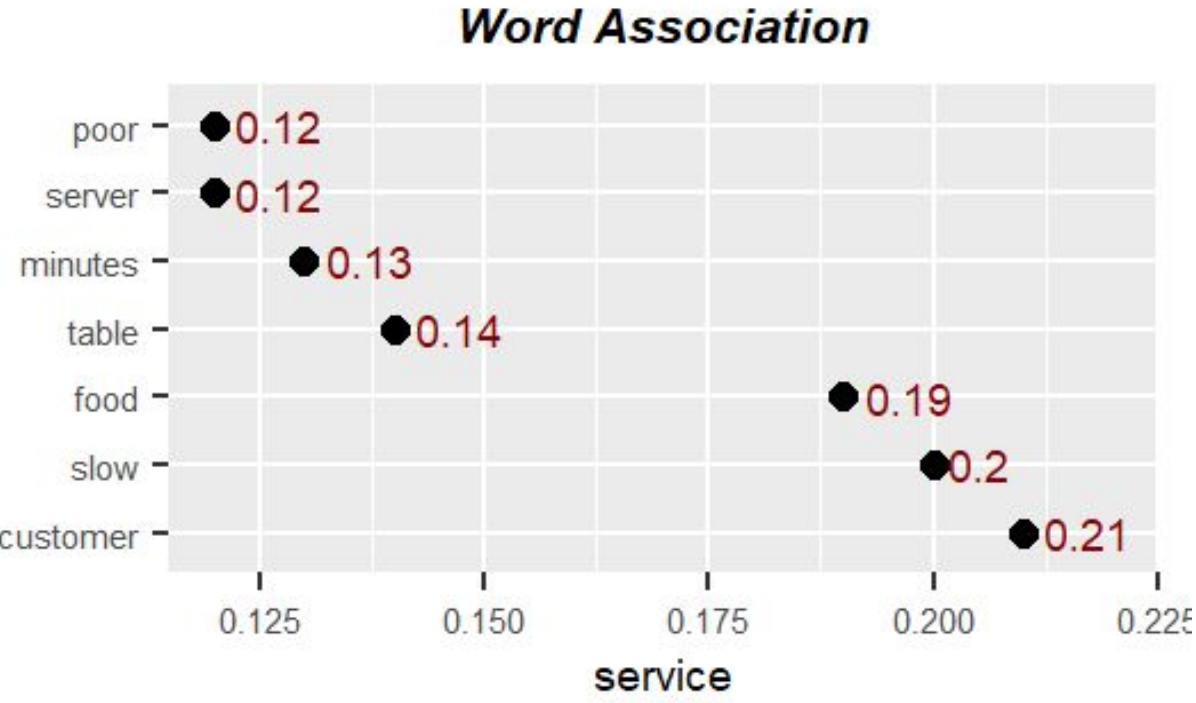
- Turn text into lower case
- Remove punctuations and digits
- Remove stopwords
- Tokenization
- Stemming

Review Analysis



- Rolls are the most ordered food, a good quality of it is the most essential.
- Fried chicken and ramen are popular in sushi restaurants. It would be better to serve nice ramen and fried chicken to attract customers.
- LUNCH shows that lunch might be the busiest time of the day.

Review analysis



From the Word Association plot, some negative words showed their close relationship to service, which means when customers mention service, they tend to complain, especially the slow serving speed. It suggests that improving serving speed is necessary.

Strength and Weakness

	Weakness	Strength
Attribute Analysis	Low r-squares	Simple model (logistic regression): easy to use and interpret
	Dependency	
Text Analysis	Filtering out meaningless words	Providing an intuitive way to navigate important words
	Interpretation of word association	

Conclusion

Actions to take:

	Please provide:
Attribute Analysis	Quiet environment
	Wheelchair
	Wifi
	Alcohol
Text Analysis	Sushi roll
	Fried chicken
	Ramen

Thank You