

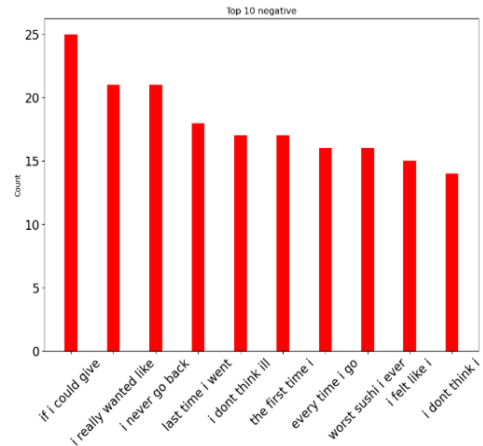
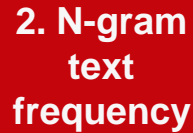


Final Presentation

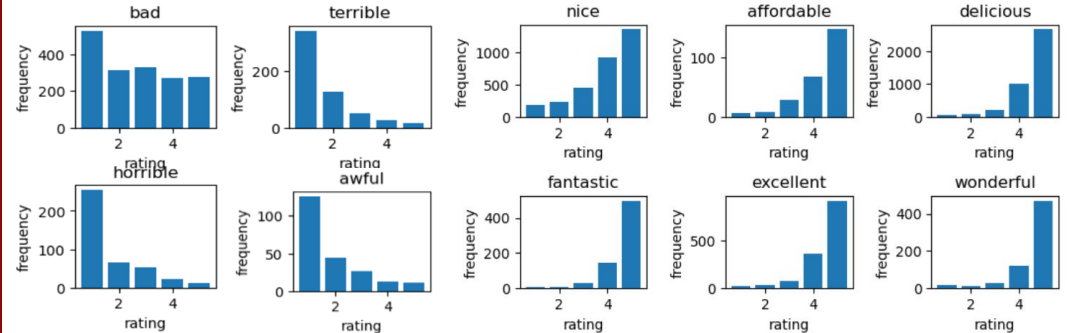
**How various features of East Asian Restaurants in CA
influence their review stars?**

Group 5
12/06/2022

1. Word cloud



3. Word sanity check



Shiny App: Word Cloud View

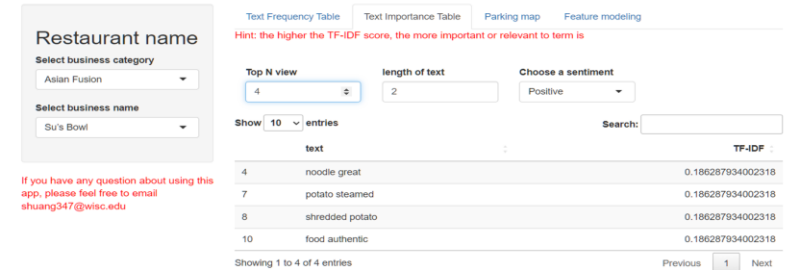
Older version

Newer version



Hint: the more a specific word appears in a source of textual data, the bigger and bolder it appears in the word cloud

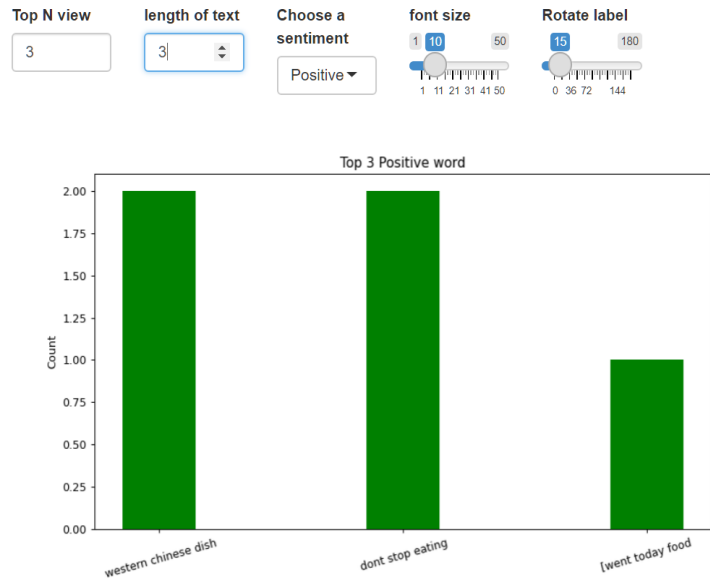
Your customer review and satisfaction analysis



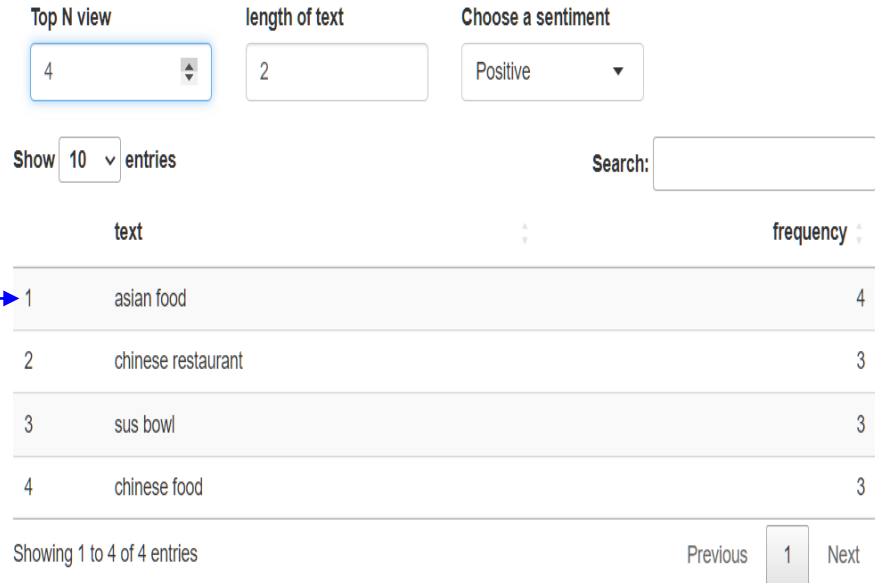
1. Due to python virtual environment on shinyapp.io, python plotting functions can't be shown properly (word cloud)
2. Word cloud view was removed in newer version of shiny

Shiny App: Text Frequency view

Old version

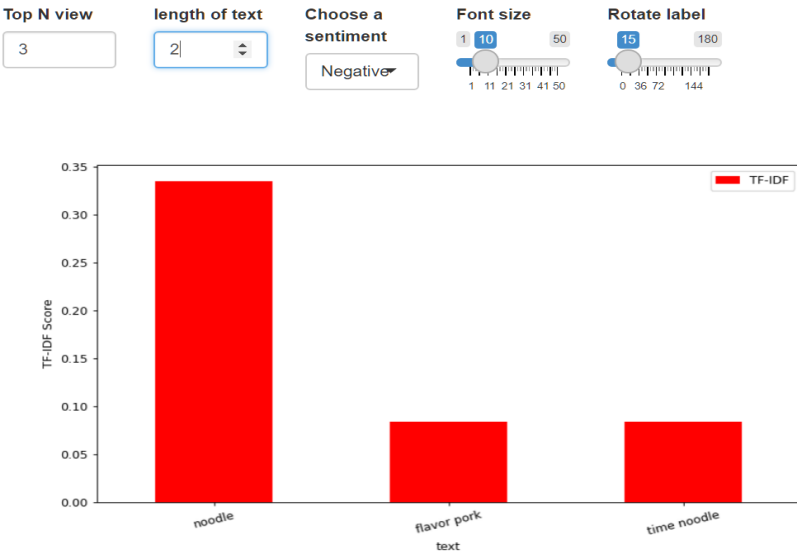


New version



Shiny App: Text Importance View

New version



New version

Top N view: 4

length of text: 2

Choose a sentiment: Positive

Show 10 entries

Search:

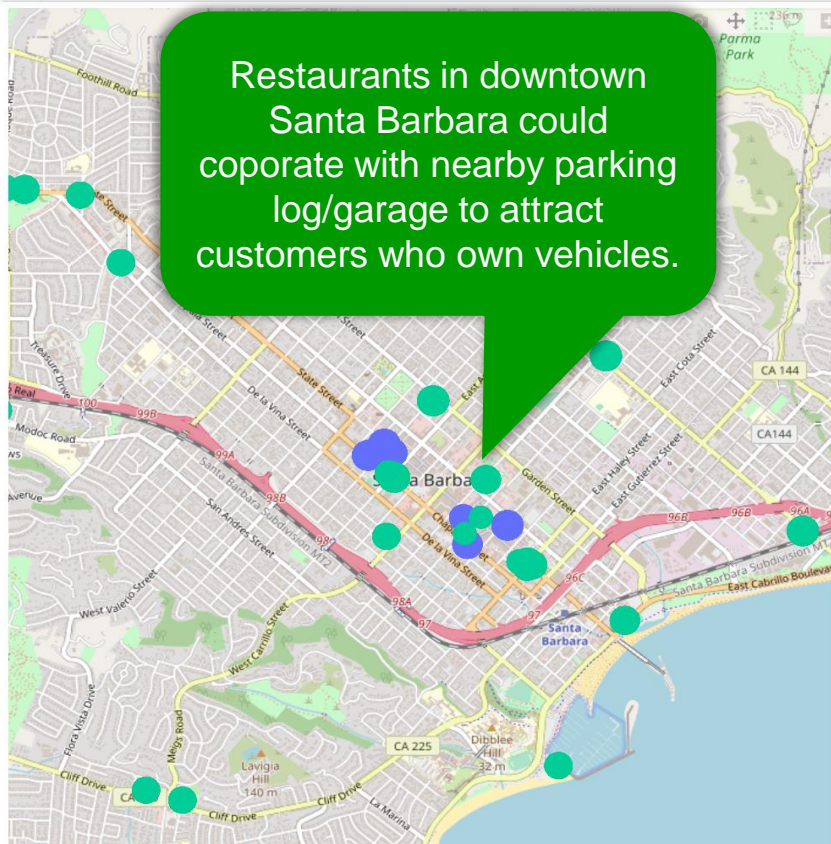
	text	TF-IDF
4	noodle great	0.186287934002318
7	potato steamed	0.186287934002318
8	shredded potato	0.186287934002318
10	food authentic	0.186287934002318

Showing 1 to 4 of 4 entries

Previous 1 Next

Shiny App: Parking Map & Modeling

Restaurants in downtown Santa Barbara could coporate with nearby parking log/garage to attract customers who own vehicles.



garage 
● True
● False

Restaurant name

Select business category

Asian Fusion

Select business name

Su's Bowl

Word Cloud

Text Frequency

Text Importance

Parking map

Feature modeling

Instruction: This part is only related to category & features appeared in the page!!!

WIFI

0

Valet Parking

0

Noise level

0

Has TV

0

Total opening hour

50

Garage Parking

0

Upscale or Classy

0

Dinner provided

0

Parking lot

0

Street Parking

0

Bike Parking

0

Group friendly

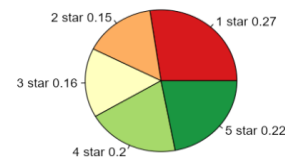
0

Alcohol

0

Reservation provided

0



The predicted probability of getting star 1 is: 27%
The predicted probability of getting star 2 is: 15%
The predicted probability of getting star 3 is: 16%
The predicted probability of getting star 4 is: 20%
The predicted probability of getting star 5 is: 22%

Modeling data

Y

X

comment star	Hour per week	HasTV	Alcohol	WiFi	Garage	Dinner	Accept Noise
Star given by customer	Weekly open hours of each restaurant	Is there a TV in the restaurant	Provide Alcohol or not	Is free WiFi available	Any place to park cars	Provide dinner Or not	Is the noise acceptable
Multi- Categorical Ordinal (1,2,3,4,5)	Continuous (16h-102h)	Binary Categorical data (1 indicates yes; 0 indicates no)					

Cumulative Logit Models For Ordinal Responses

Model function:

$$\text{logit}[P(Y \leq j)] = \log\left[\frac{P(Y \leq j)}{1 - P(Y \leq j)}\right] = \log\left[\frac{\pi_1 + \dots + \pi_j}{\pi_{j+1} + \dots + \pi_J}\right] = \alpha_j + \beta x, j = 1, 2, \dots, J - 1$$

$$P(Y \leq j) = \exp(\alpha_j + \beta x) / [1 + \exp(\alpha_j + \beta x)], j = 1, 2, \dots, J - 1$$

Result with 12 significant features:

$$\frac{P(Y \leq j | X = x + 1) / P(Y > j | X = x + 1)}{P(Y \leq j | X = x) / P(Y > j | X = x)} = \exp(\beta) = \exp(0.008) \approx 1$$

variable	coefficient &significance level	variable	coefficient &significance level	variable	coefficient &significance level	variable	coefficient &significance level
Intercept 1	-3.843***	BikeParking1	0.621***	lot1	-0.143**	dinner1	0.268**
Intercept 2	-3.204***	Reservations1	0.133**	valet1	-0.683***	Total_hour	0.008***
Intercept 3	-2.582***	Alcohol1	-0.416***	garage1	-0.877***	acceptable noise1	1.301***
Intercept 4	-1.632***	WiFi1	-0.267***	street1	0.269***	upscale classy1	-0.176**

Significant level: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Cumulative Logit Models under different categories

Difference between Japanese and Chinese restaurants:

Japanese			
variable	coefficient & significance level	variable	coefficient & significance level
Intercept 1	-3.563***	BikeParking1	0.948***
Intercept 2	-2.888***	Good For Groups	0.455***
Intercept 3	-2.247***	garage1	-0.801***
Intercept 4	-1.269***	WiFi1	-0.839***
Total_hour	0.014***	street1	0.461***
upscale classy1	-0.360***		

Significant level: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Chinese			
variable	coefficient & significance level	variable	coefficient & significance level
Intercept 1	-2.768***	BikeParking1	0.692***
Intercept 2	-2.153***	HasTV	-0.225*
Intercept 3	-1.592**	Alcohol1	-0.687***
Intercept 4	-0.627**	lot1	0.585***
Total_hour	0.021***	dinner1	-0.343*
upscale classy1	0.387***		

Significant level: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Goodness of fit test

H_0 : The model fits the data well

H_1 : The model doesn't fit the data well

Goodness of Fit test			
	χ^2	residual df	p-value
12features Model	$4.29 * 10^4$	42740	0.238
Chinese	6594.96	6629	0.614
Japanese	$1.802 * 10^4$	18041	0.527

Conclusion



Working longer will **not improve** ratings

Provide more **attentive service** to make customers give more positive reviews

Different categories of restaurants need to focus on different features to make improvements

Limitation



The diagram consists of several overlapping circles of varying shades of red and pink on a solid red background. The circles contain text describing limitations. The largest circle in the center-top is dark red and contains the text 'Improve model prediction accuracy'. To its right is a light pink circle containing 'Further explore features that are highly correlated with rating stars'. Below the central circle is a light pink circle containing 'Provide customized suggestion for each business'. To the right of that is a dark red circle containing 'Apply vertical research on how specific restaurant can be improved based on customers review'. There are also two smaller circles on the left side of the main cluster.

Improve model prediction accuracy

Further **explore features** that are highly correlated with rating stars

Provide **customized suggestion** for each business

Apply **vertical research** on how specific restaurant can be improved based on customers review



THANK YOU