

Chinese Restaurant Data Analysis

Ruyi Yan, Yilun Zhang, Luwei liang

Mar 7, 2019

Overview

- 1 Introduction
- 2 Data Processing – Business Data
- 3 Basic Findings from Business Data
- 4 Simple text-analysis
- 5 Future Work

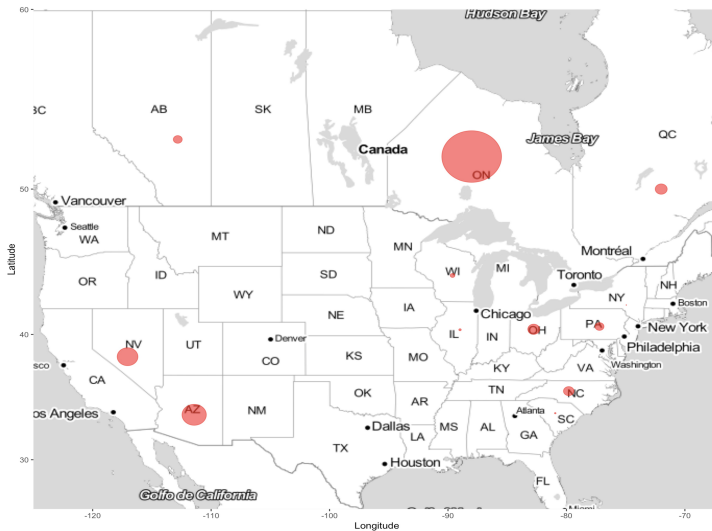
- Why we choose Chinese restaurants?
 - We are more familiar with them.
 - Relatively small number of second-category tags.

Introduction

- 3557 Chinese restaurants in 13 states.

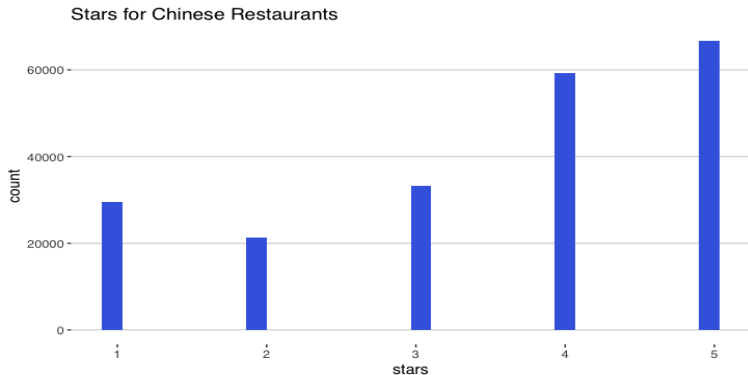
State	Number of Restaurants
AB	185
AZ	531
IL	40
NC	221
NV	451
NY	1
OH	251
ON	1324
PA	187
QC	254
SC	21
WI	90
XGM	1

Introduction



Introduction

- 209897 reviews, most of them are 4 or 5 stars (60%).

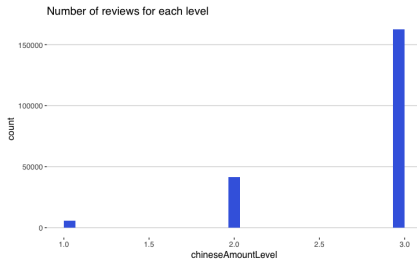
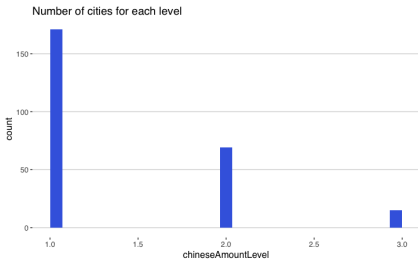


- Delete unimportant and useless features: 'post-code', 'name', 'latitude', 'longitude', 'is-open', 'categories'.

- Transfer feature 'city' into magnitude of chinese restaurants .
 - $N_{chinese} \leq 3 \implies \text{Small}(1)$
 - $N_{chinese} \geq 50 \implies \text{Large}(3)$
 - $3 < N_{chinese} < 50 \implies \text{Median}(2)$

Data Processing – Business Data

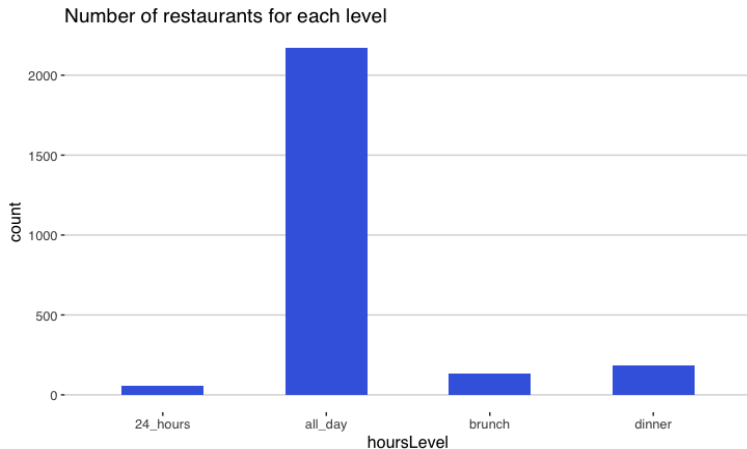
- Transfer feature 'city' into magnitude of chinese restaurants .



Data Processing – Business Data

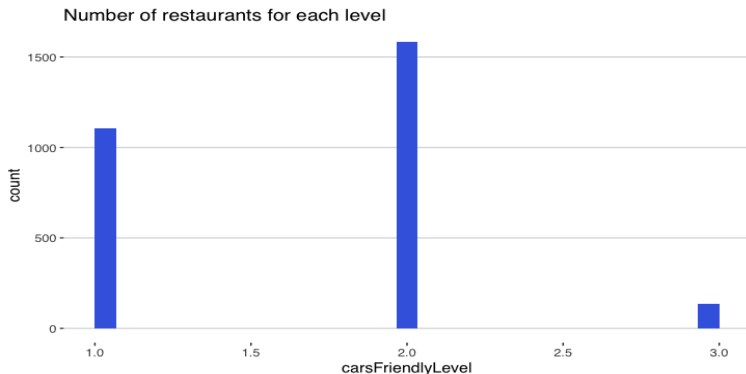
- Divide feature 'hours' into 4 levels: '24_hours', 'all_day', 'lunch', 'dinner'.
 - $hours \geq 15 \implies '24_hours'$
 - $6 \leq hours < 15 \implies 'all_day'$
 - Closing time earlier than 15:00pm $\implies 'lunch'$
 - Opening time later than 15:00pm $\implies 'dinner'$

Data Processing – Business Data



Data Processing – Business Data

- For 'attributes', we select these 16 attributes.
- For 'Business Parking', we classify them into 3 levels according to convenient level.



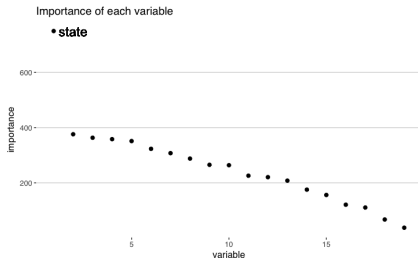
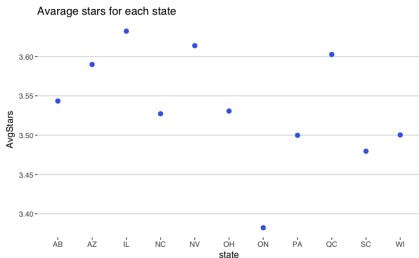
Basic Findings from Business Data

- Some overall findings:

	Negative	Score _N	Positive	Score _P
Alcohol	beer&wine	3.47	None	3.60
BusinessParking	No parking	3.42	Parking friendly	3.57
HasTV	No TV	3.47	Has TV	3.63
NoiseLevel	Load	3.29	Quiet	3.63
WiFi	No WiFi	3.52	Free WiFi	3.62

Basic Findings from Business Data

- State is a really important variable.



Basic Findings from Business Data

- 'Characteristic' of chinese restaurants:

	A=3 or 4 stars	B=1 or 2 stars	A/B
Chinese restaurants	89	3067	0.03
All restaurants	2715	38953	0.07

Outdoor Seats	A=FALSE	B= TRUE	A/B
Chinese restaurants	2630	340	7.7
All restaurants	23482	15133	1.5

Has TV	A=FALSE	B= TRUE	A/B
Chinese restaurants	1610	1190	1.35
All restaurants	17261	18606	0.92

Has WiFi	A=FALSE	B= TRUE	A/B
Chinese restaurants	1819	697	2.61
All restaurants	16807	15448	1.08

Simple text-analysis

- A simple text-analysis on QQ-Express
 - 59 reviews in total
 - Create the word frequency sparse matrix with $p=1329$

bad	good	wrong	fat
1	2	1	1
0	3	0	0

Simple text-analysis

- A simple text-analysis on QQ-Express
 - Perform a simple logistic regression with $Y = \text{review\$stars}$
 - The top 20 positive words:

price	quick	lunch	nice	good	an	highly
2.3783480	1.9992634	1.7940573	1.7322821	1.5253039	1.5221779	1.4696290
on	very	rice	it's	many	favorite	this
1.3195870	1.0231201	1.0158467	0.9341363	0.8460795	0.8168955	0.7558957
they're	ma	peanut	tofu	portions	cheap	
0.7149960	0.6793929	0.5752047	0.5533853	0.4724339	0.2245626	

Simple text-analysis

- A simple text-analysis on WI OH

- Top 30 positive words for WI:

hunan	seasoning	grab	head	eater	fortune	wonderful	breakfast
14.102031	13.465977	12.629091	12.494890	12.469823	11.566556	11.394696	11.145384
surprisingly	efficient	cozy	perfect	delicious	hip	friday	sichuan
11.080850	11.077103	10.998931	10.807234	10.523581	10.521020	10.418608	10.118084
fantastic	exactly	chive	inexpensive	polite	broths	awesome	choices
9.966765	9.911748	9.765545	9.690291	9.617348	9.561046	9.361583	9.320079
scene	comfortable	11	college	convenient	crazy		
9.184885	9.095556	9.041229	8.822231	8.774138	8.579816		

- Top 30 positive words for OH:

welcoming	reminds	finding	juicy	tay	decorated	superb	constantly
30.03936	27.83956	26.92101	26.39497	22.45939	21.78852	19.50859	19.00057
baby	crunch	ive	incredible	chocolate	carrots	complaints	zen
19.00017	18.88252	18.68885	17.84578	17.50056	17.29875	17.26471	16.53449
enjoying	quiet	fault	wide	salted	professional	phenomenal	chose
16.47570	15.95108	14.99555	14.82546	14.42603	14.21570	14.18688	14.02222
additional	kinds	pleased	yum	flat	worker		
13.88430	13.45716	13.45136	13.40480	13.39362	13.38716		

Future Work

- For each review, evaluate the score it got in these 3 aspects: Price, Service/Environment, Taste

"Decent fast inexpensive food. Large portions. Good value. It was cafeteria style so the service is fast. The lo mein was pretty good. The general zhos chicken was OK (small chicken, big breading). Pepper chicken and bbq chicken good. Not fancy but fills you up."

⇒

Price	Service/Environment	Taste
4	4	3

- For restaurants, figure out why they got high(low) stars (Good service? Great taste or All of them?). For states, figure out which aspect do people value most.
- Possible methods/direction: Part-of-speech tagging, syntactic parsing, ...

Thank you