

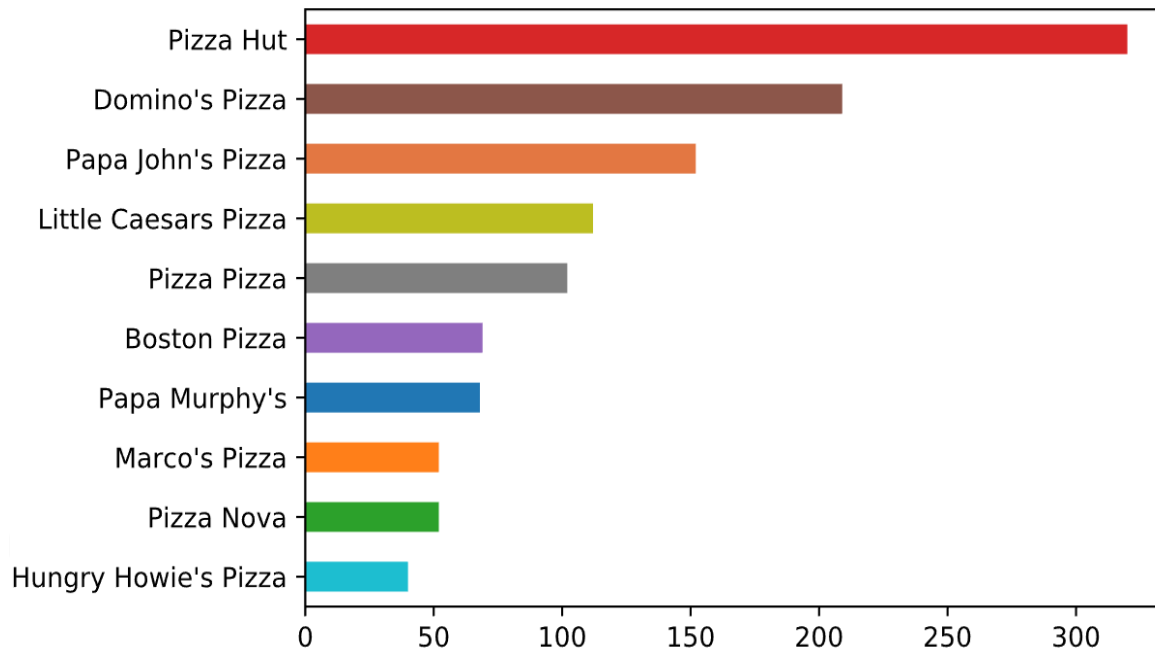
# Yelp Pizza Restaurants' Review Analysis

*Group2*

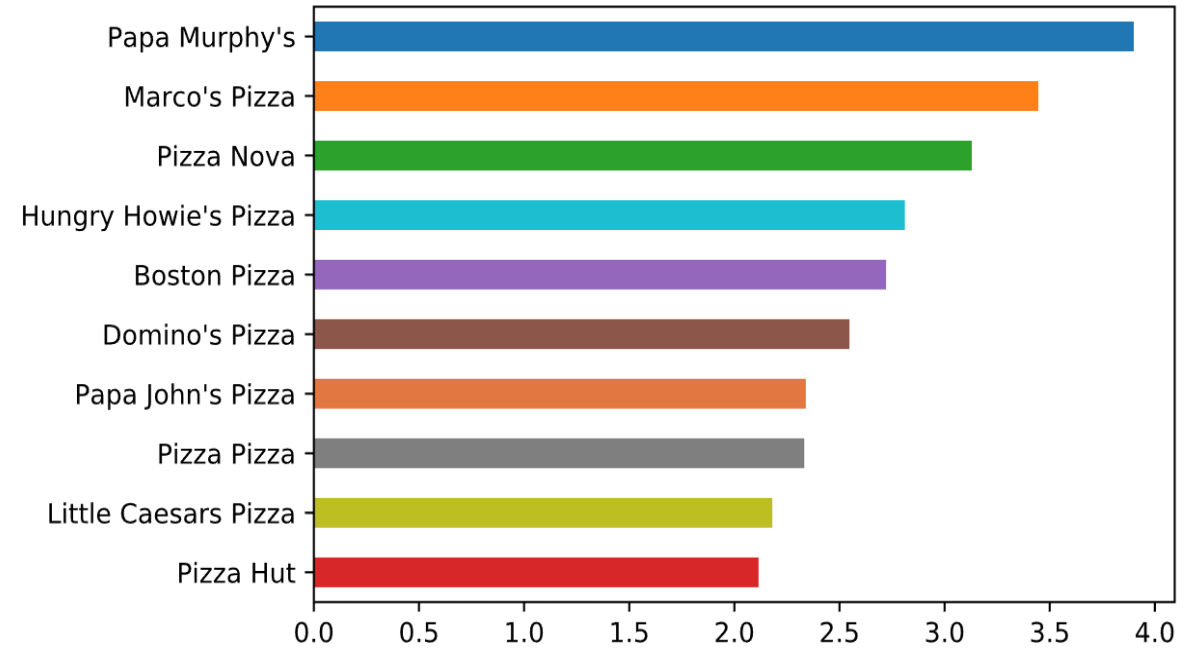
*Jiacheng Miao, Hao Qin, Shirley Zhang, Xiyue Wang*



# Last time



Plot 1: Top 10 Pizza restaurants count



Plot 2: Top 10 Pizza restaurants average star

Most restaurants, lowest stars => Further Analysis

# Translation

```
set(language)  ##Language types
```

```
{'af', 'da', 'en', 'es', 'fr', 'hr', 'it', 'pt', 'sk', 'so', 'sw'}
```

Afrikaans	Danish	Spanish	French	Croatian
Italian	Portuguese	Slovak	Somali	Swahili



English

# Which adjectives are important?

- X feature matrix (Nxp)  
N: Number of reviews (17951)  
p: Number of selected adj words (100)
- Eg: A row of matrix

array([1., 0., 0., 0., 0., 0., 0., 1., 1., 0., 0., 3., 0., 0., 0., 0., 0.,  
0., 0., 1., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0.,  
0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0.,  
0., 0., 0., 0., 0., 0., 0., 0., 1., 0., 1., 0., 0., 0., 0., 0.,  
0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0.,  
0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0.] )

# Logistic Regression

- Binary Y labels :

$$Y = \begin{cases} 0 & \text{if } star = 1,2,3 \\ 1 & \text{if } star = 4,5 \end{cases}$$

Accuracy: 80.23%

# Linear Regression

- Y labels : Star for each review

MSE: 3.933

# User Weight

- Which variables did us take into consideration?

Review Counts, Amount of Followers,

Number of useful/funny/cool votes sent by the user,

Number of plain/funny/cool compliments received by the user

- How to compute user weights based on these variables?

## Entropy Weight Method

- Based on amount of information to determine the index's weight, then use them to compute user's weight and help with next steps.
- More than 86% user's weight almost haven't been changed. This method only influenced points which should be highlighted.

# User Weight

- High - Weighted Users

Name	Review count	Fans	Cool	Funny	Useful	Compliment Cool	Compliment Fun	Compliment Plain
Fox	6407	1574	82128	86122	89418	5429	5429	4379
Bruce	10022	528	61383	49785	74829	2502	2502	2749
Stefany	4321	641	30846	22790	44667	1635	1635	1497
Joyce	763	738	32269	24344	34316	5462	5462	3686
Jennifer	6314	715	12995	8823	20336	1606	1606	3629
Dave	2678	305	36081	21461	43324	1859	1859	1353
Karen	2340	411	30626	19182	34815	3411	3411	2312

# Local Weighted Linear Regression(LWLR)

- Formula:  $\hat{\beta} = (X'WX)^{-1}(X'WY)$
  - W: User weight matrix
- MSE: 3.683

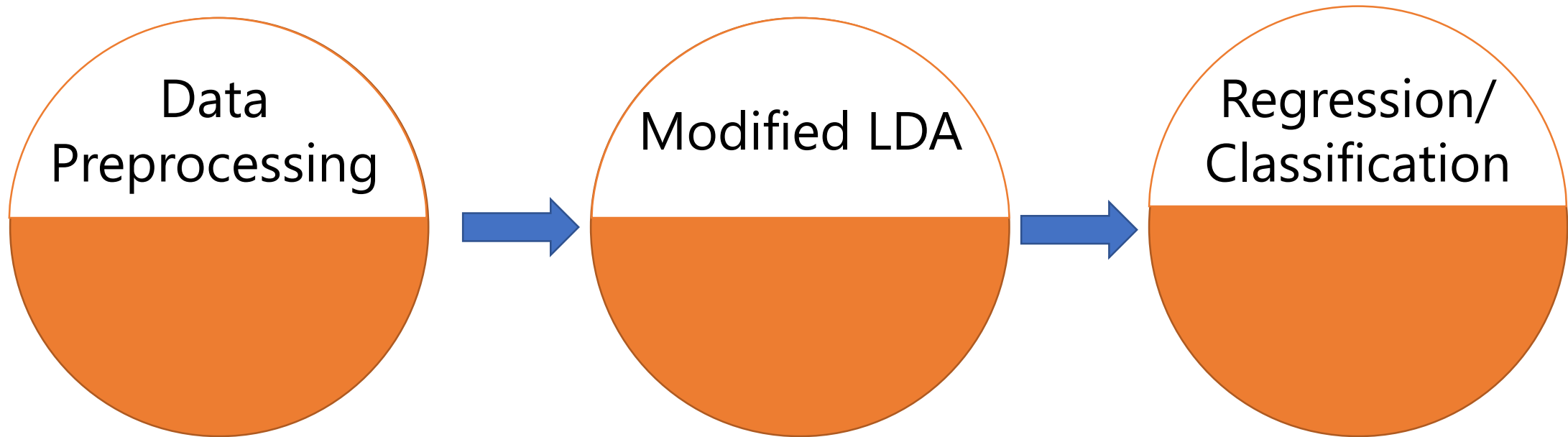
notready	-0.2841758047183900
neverordered	-0.3023918838235440
later	-0.3279864851082150
neverreally	-0.6195962103424990

notgoingback	-0.2975867135047550
neverordered	-0.3013204369443360
later	-0.3303707270438090
neverreally	-0.7563027667217800

Table 1: The weight for each adjective words (left is without user weight adjustment)



# LDA combined with Regression



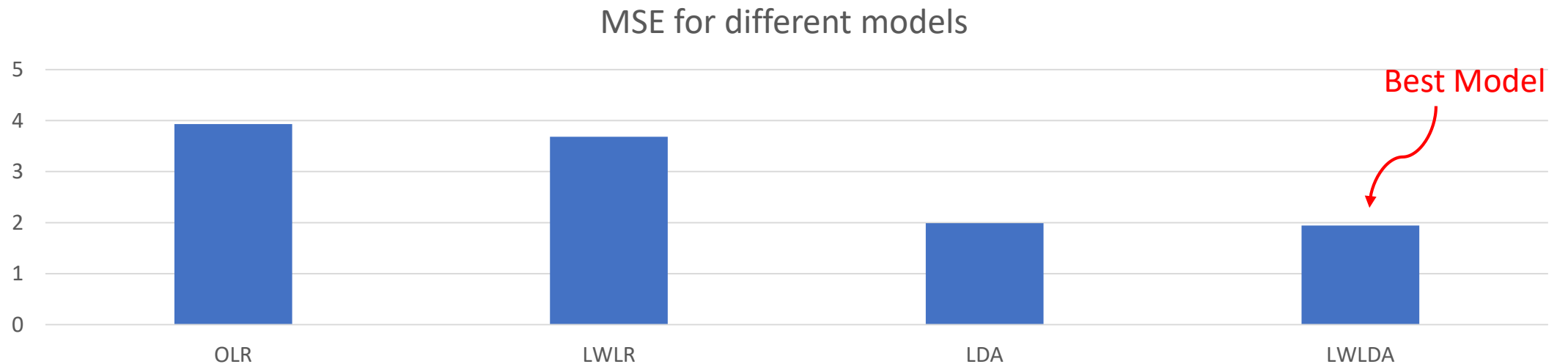
# Local Weighted LDA

- Local weighted idf term

$$widf_{\{i,j\}} = w_j * lg\left(\frac{|D|}{\{j: t_i \in d_j\}}\right)$$

$w_i$  is the weight corresponding to the  $j$ **th** user,  $D$  is the total number of all reviews and the denominator is the number the review which includes the  **$i$ th** word.

# Model compare



Model	OLR	LWLR	LDA	LWLDA
MSE	3.933	3.683	1.990	1.934

Table2: MSE of 3 models.(OLR:Original Linear Regression.LWLR:Local Weighted Linear Regression.LDA:Latent Dirichlet Allocation.LWLDA:Local Weighted Latent Dirichlet Allocation)

# Contingency Table Test

The sentiment table of noun word "pizza"		
	Pizza Hut	Other 9 Pizza Chain
Positive	74	322
Negative	108	200

Table3: The sentiment table of noun word 'pizza'

chi-square test :

P-value = 1.318e-06

The brand effect is significant

# Business Suggestion

Noun and "Fast" pairs (in every 1278 nouns)		
Noun.	PizzaHut	Other 9 Pizza Chain
Delivery	39	31
Service	32	17
Pizza	7	5



Efficient, good at delivery, high-speed

Noun and "Delicious" pairs (in every 1278 nouns)		
Noun.	PizzaHut	Other 9 Pizza Chains
Pizza	15	35
Food	5	32
Crust	1	17



More kinds of nouns appear in other 9:  
gralic, dessert, pepperoni, cream, jalapeño, ravioli,  
breadstick, sangria, cheesecake, burger

# Business Suggestion

SWOT Analysis	
Strength	High work efficiency, delivery on time
Weakness	Poor selection menu, not so tasty
Opportunities	Brand impact, well-funded to carry out business plan
Threats	Hard to diffuse new technology or new strategy to all chains

*“Our focus on digital efforts is clearly to make certain we are appealing to younger audiences, which will be the future of the overall business.”*

**— Steve Richie**  
*Papa John’s CEO*

# The strength and weakness

## *Strength:*

- Introducing highlighted users
- Accurate sentiment detection model

## *Weakness:*

- Words are exchangeable
- Have not mine enough information from short reviews

# Thank You

Shinny link:

<https://jiacheng-miao.shinyapps.io/Module3/>