Yelp data analysis

Group 5: Ouyang Xu, Shuren He, Ziyue Zheng

- **Raw data**

More than 8 millions of review data

More than 50 thousands of business data

- **Our goal**

To analyze how customers judge a sandwich restaurant. And What are the essential qualities of a good sandwich shop

- **Data clean process**

1. Filtering all restaurants which categories include sandwich.
2. Filtering out the review data of the sandwich restaurants.
3. Join these two parts of data and do further analysis.

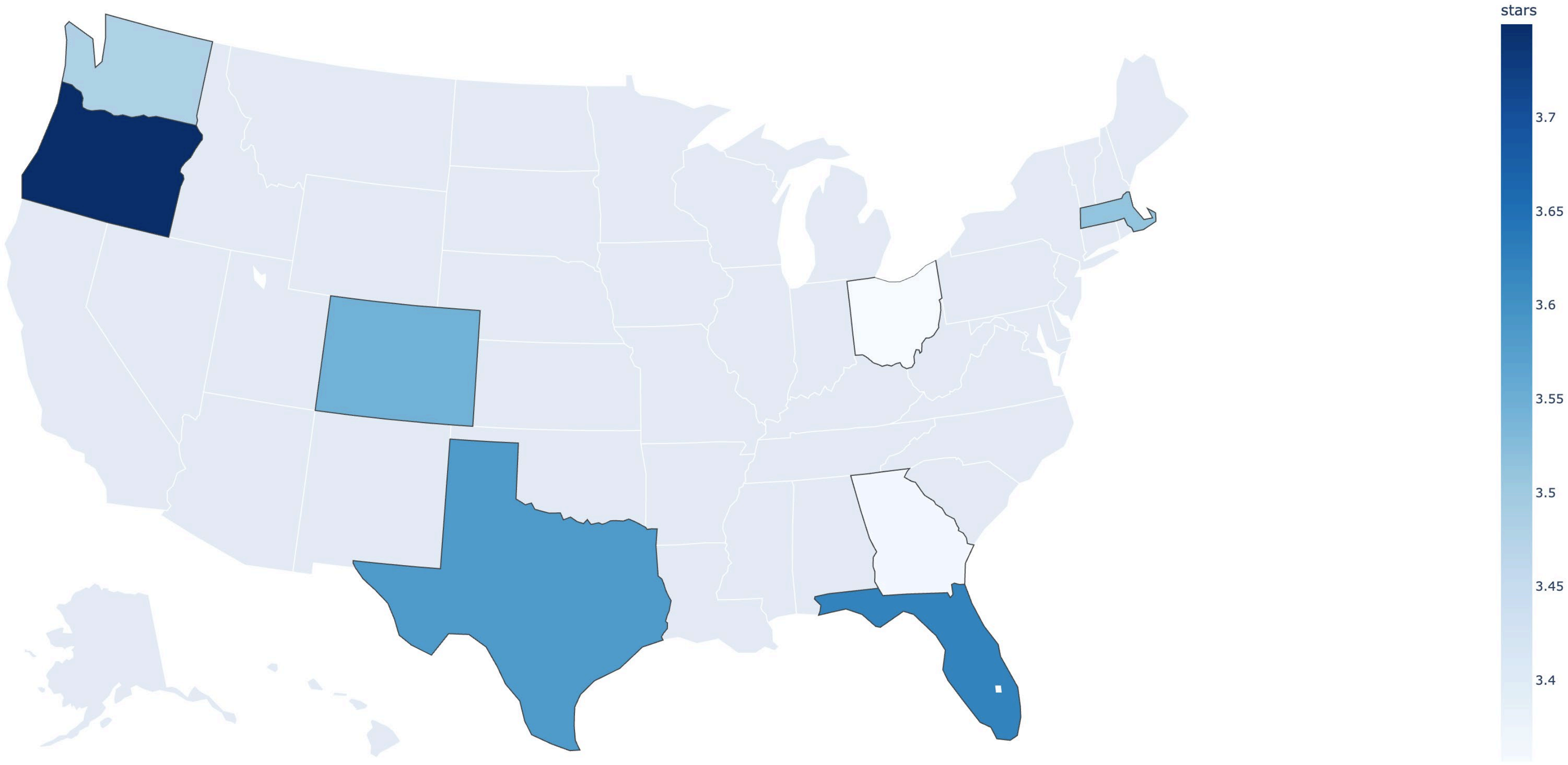
- **Result**

4795 business & approximately 1 million reviews

STATE	NUMBER OF SHOP	NUMBER OF REVIEW	AVERAGE SCORE
BC	654	29480	3.52
CO	135	12768	3.54
FL	1154	94128	3.62
GA	856	91671	3.36
MA	1891	197817	3.51
OH	719	39346	3.36
OR	1025	116833	3.75
TX	735	112443	3.58
WA	103	6459	3.48

Average stars by state

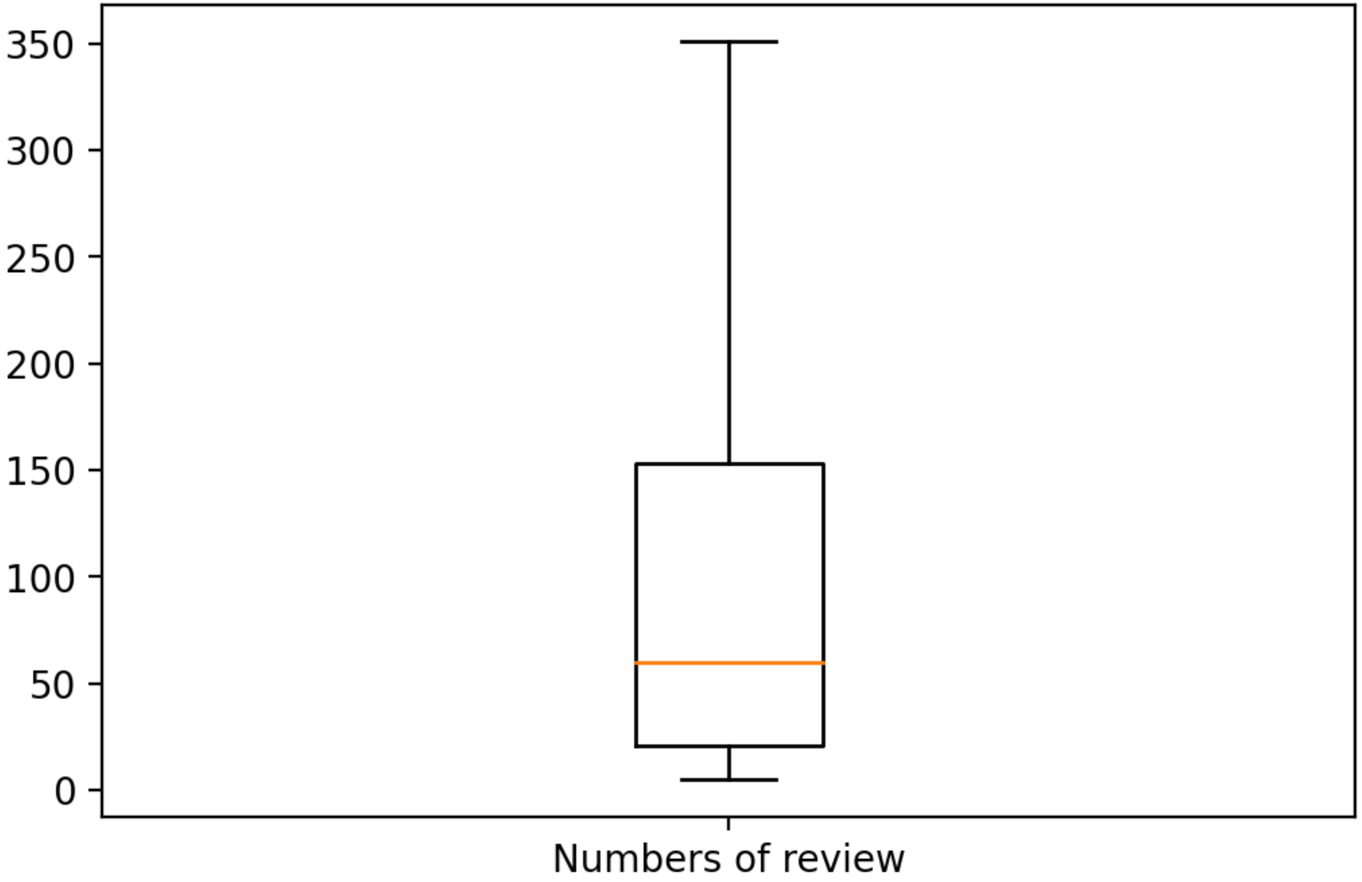
Interactive web page: stars.html



- Number of reviews for each restaurant

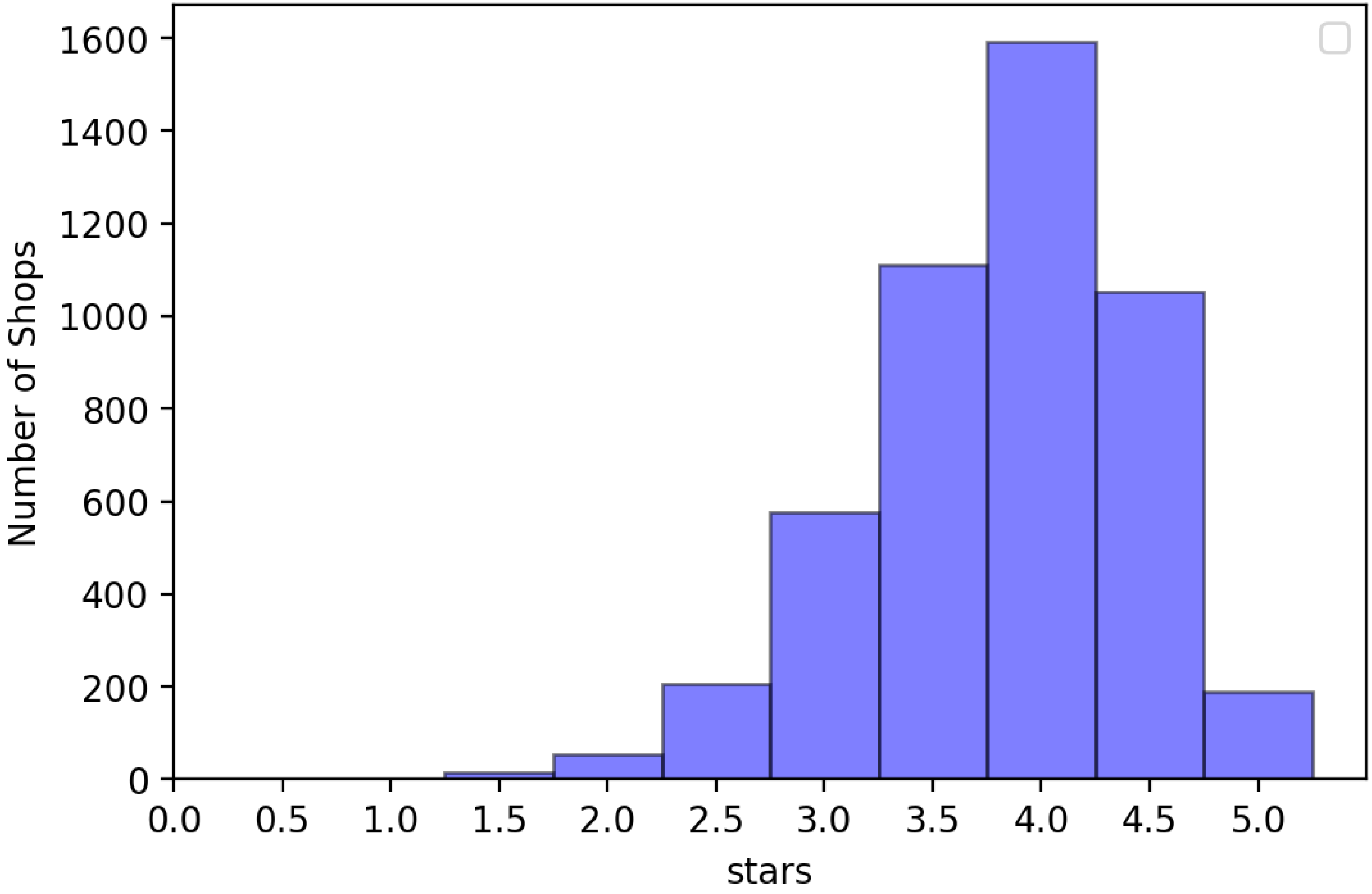
A majority of restaurants have less than 400 reviews.

(However, Panera Bread has near 10,000 reviews.)



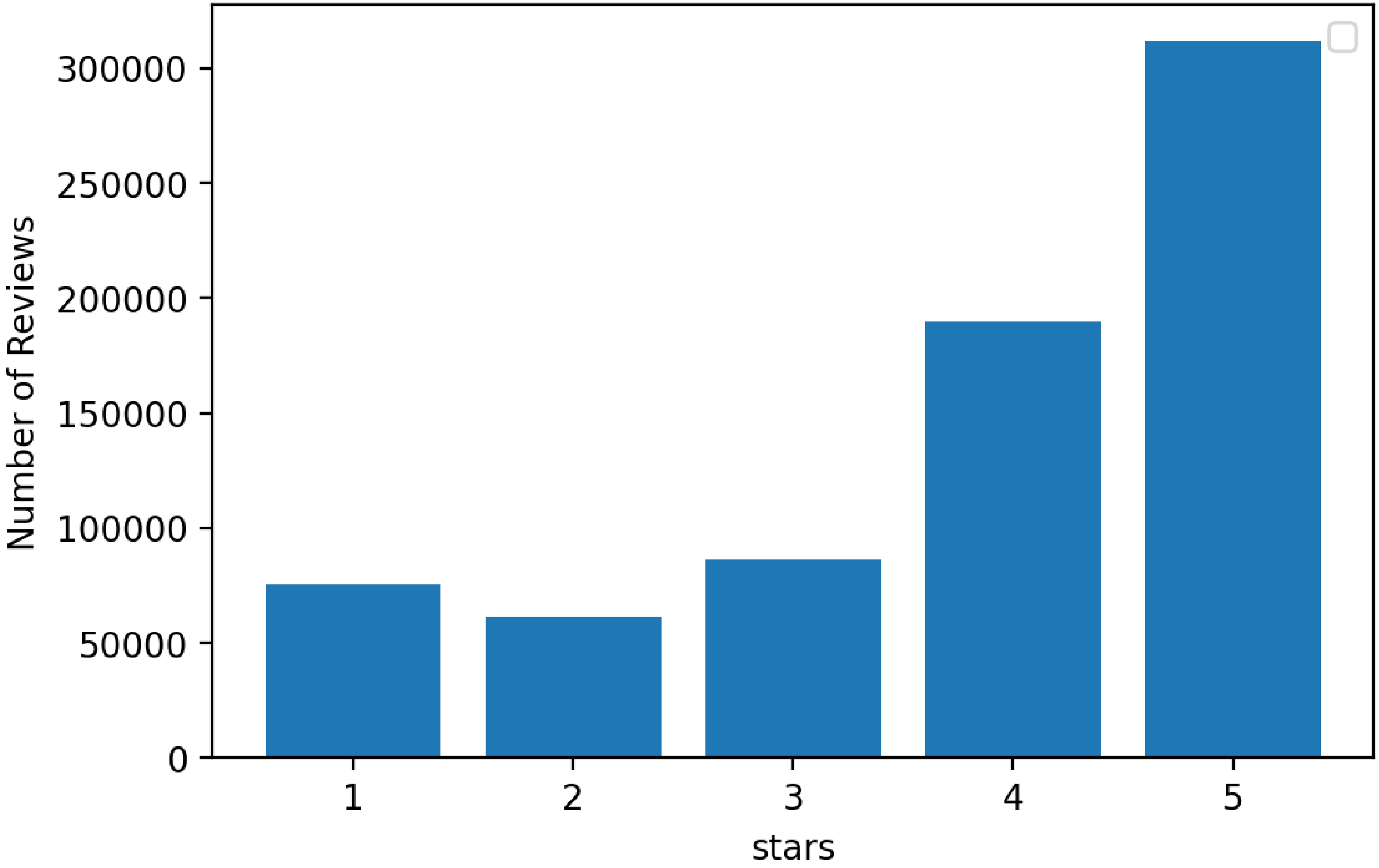
- The distribution of shop scores

Most shops are rated around 4. Both very low and very high are rare.



- The distribution of reviews scores

Customer are ‘friendly’, they are willing to give 4 or 5 scores



Customers are “mean”

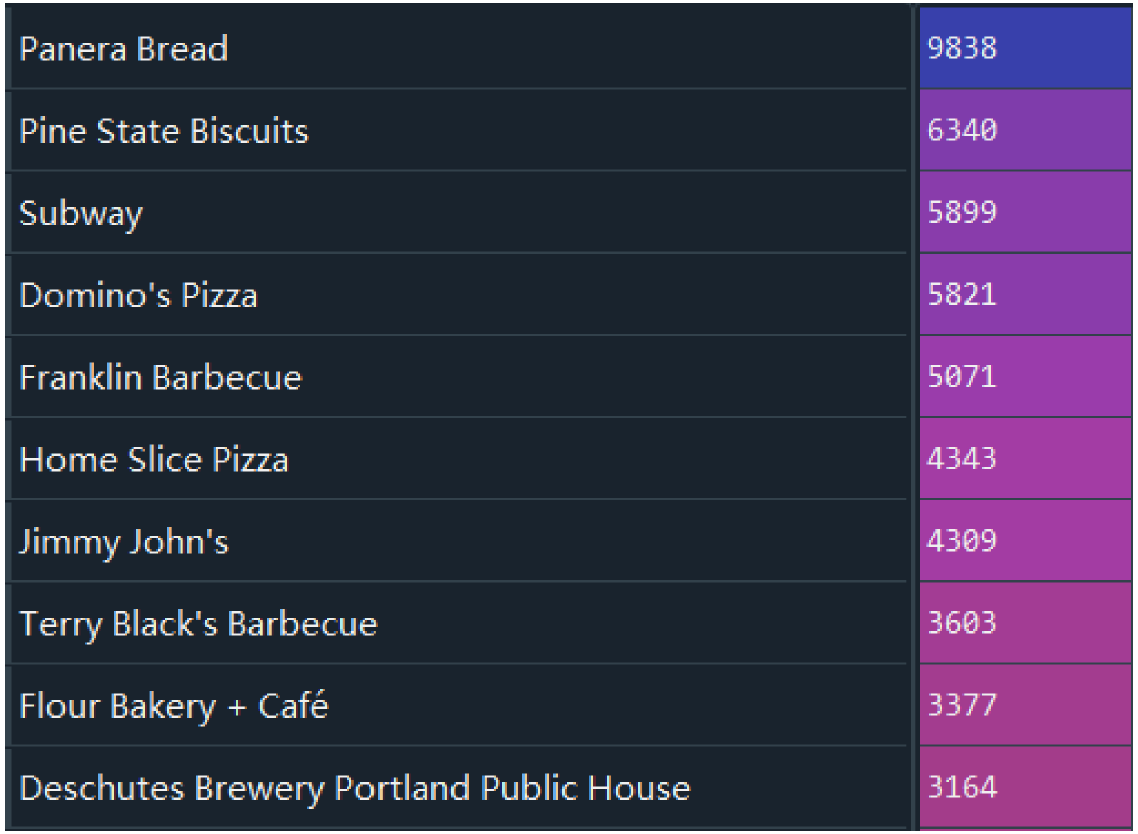
166 customers gave one-star stores with a five-star review

5,190 customers gave five-star stores with a one-star review.

Work TODO: Explore key words of giving good and bad reviews:

1. Good reviews of top restaurants
2. Bad reviews of poor restaurants
3. Good reviews of poor restaurants
4. Bad reviews of top restaurants

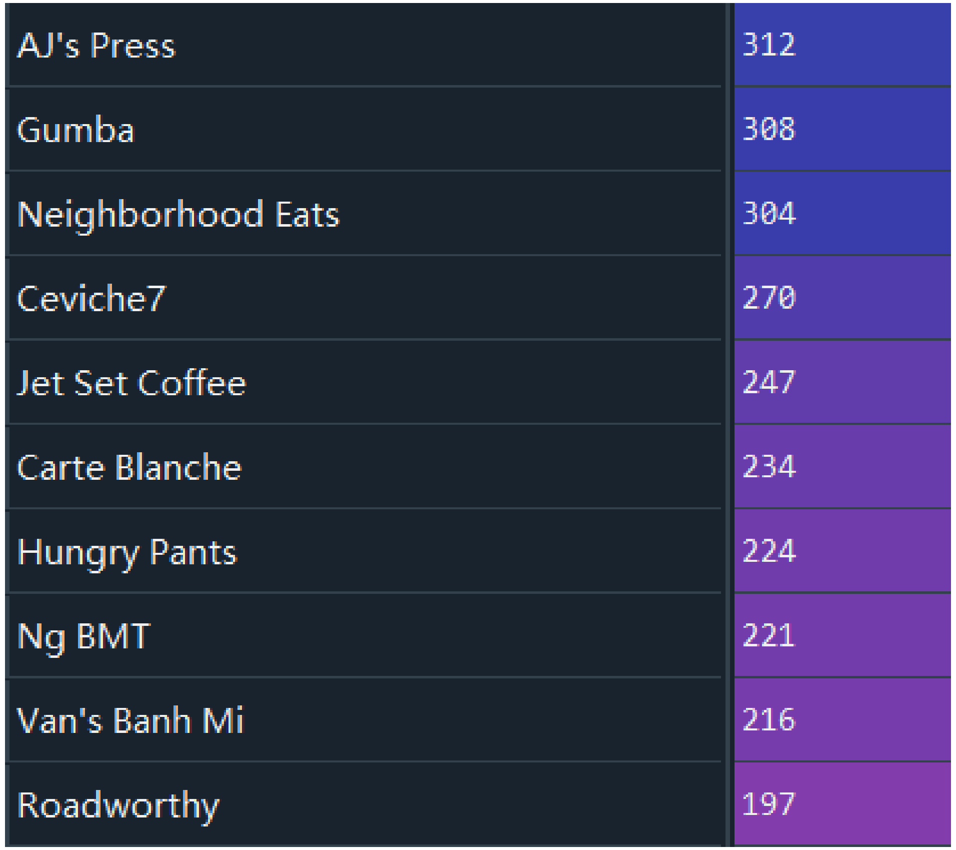
Top 10 most ‘popular’ sandwich restaurants



Panera Bread	9838	2.55
Pine State Biscuits	6340	4.2
Subway	5899	2.6
Domino's Pizza	5821	2.39
Franklin Barbecue	5071	4.5
Home Slice Pizza	4343	4.5
Jimmy John's	4309	2.76
Terry Black's Barbecue	3603	4.5
Flour Bakery + Café	3377	4.25
Deschutes Brewery Portland Public House	3164	4.0

- Top 10 most popular 5-star sandwich restaurants

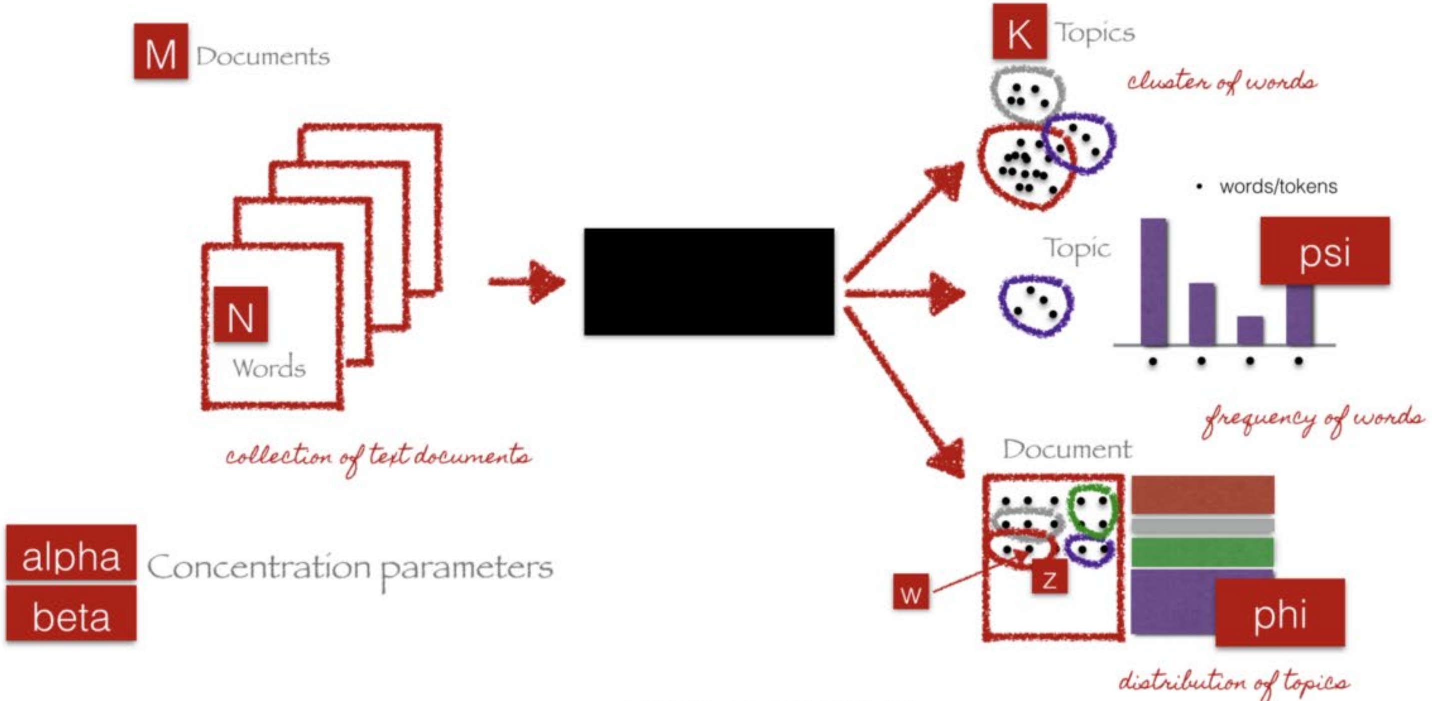
5-star restaurants are all small size

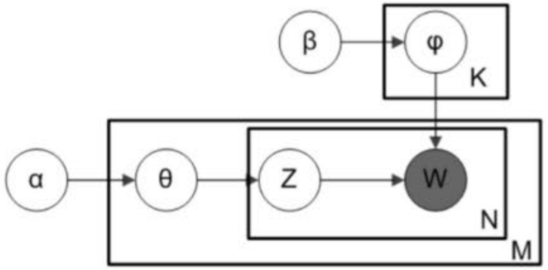


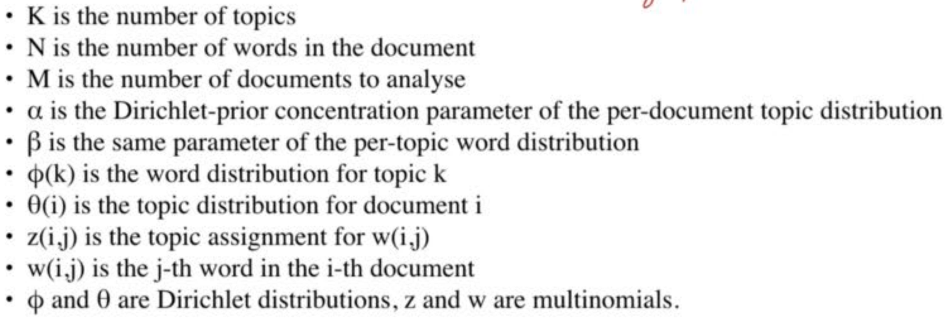
AJ's Press	312
Gumba	308
Neighborhood Eats	304
Ceviche7	270
Jet Set Coffee	247
Carte Blanche	234
Hungry Pants	224
Ng BMT	221
Van's Banh Mi	216
Roadworthy	197

LDA Model Theoretical Overview

- LDA is a generative probabilistic model that assumes each topic is a mixture over an underlying set of words, and each document is a mixture of over a set of topic probabilities.



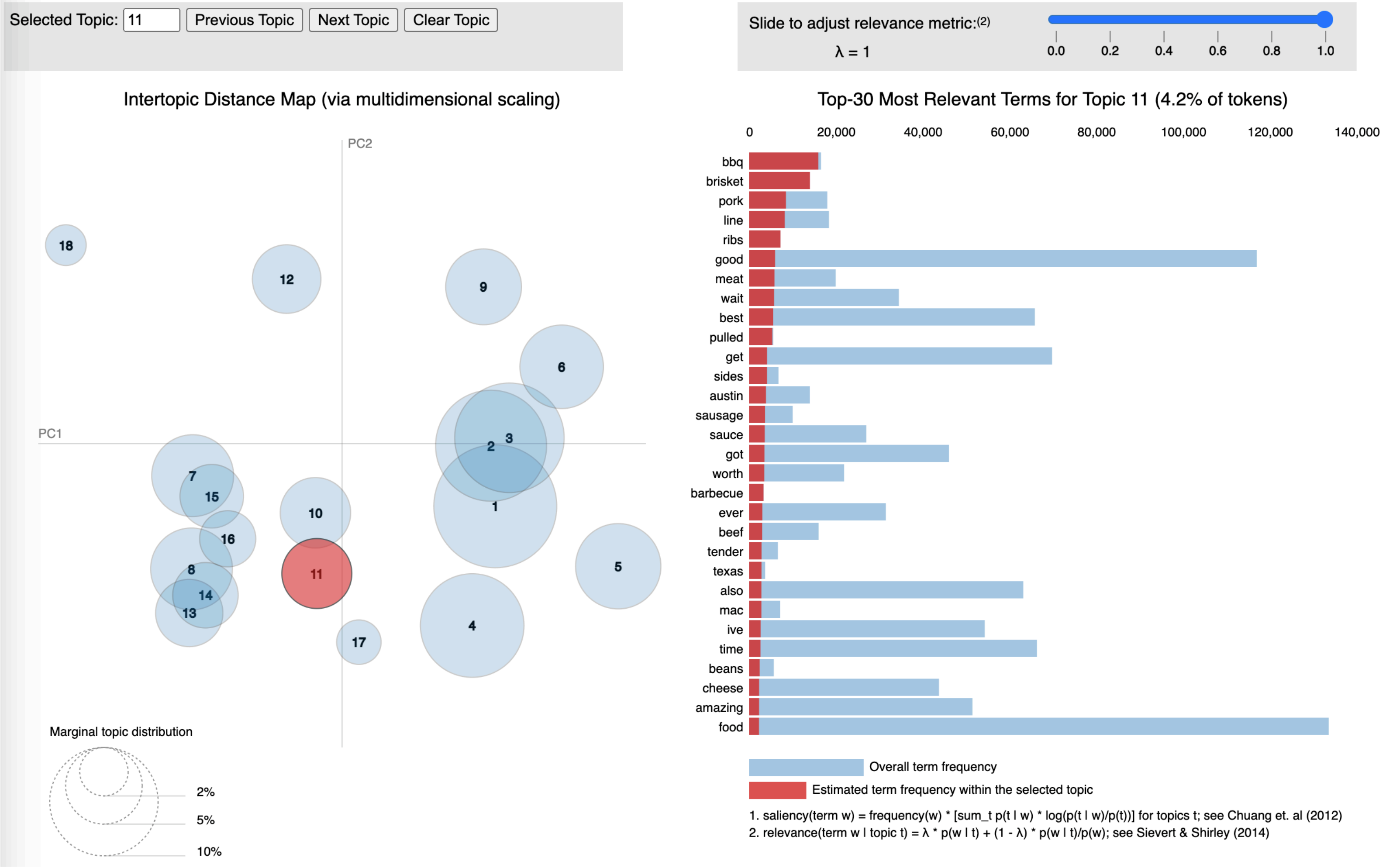




- K is the number of topics
- N is the number of words in the document
- M is the number of documents to analyse
- α is the Dirichlet-prior concentration parameter of the per-document topic distribution
- β is the same parameter of the per-topic word distribution
- ϕ(k) is the word distribution for topic k
- θ(i) is the topic distribution for document i
- z(i,j) is the topic assignment for w(i,j)
- w(i,j) is the j-th word in the i-th document
- ϕ and θ are Dirichlet distributions, z and w are multinomials.

LDA Model Result

- **Interactive web page: vis_lda.html**
- **Preview:**



Selected Topic: 11

Slide to adjust relevance metric:(2)

λ = 1

1. saliency(term w) = frequency(w) * [sum_t p(t | w) * log(p(t | w)/p(t))] for topics t; see Chuang et. al (2012)
2. relevance(term w | topic t) = λ * p(w | t) + (1 - λ) * p(w | t)/p(w); see Sievert & Shirley (2014)

tagxedo.com

tagxedo.com

tagxedo.com

sandwich
good
place
food
great
time
REALLY
cheese
SERVICE
order
little
chicken
love
try
wait
BEST
IM
friendly
LOT
RESTAURANT
PRICE
dont
table
taste
IVE
fresh
looking
tasty
coffee
burger
recommend
SERVED
LOCATION
bit
eat
come
pizza
FRIES
lunch
sauce
came
work
options
busy
quick
favorite
small
hot
think
SWEET
SOUP
STARS
atmosphere
right
AREA
pork
big
hour
SPOT
FRIEND
need
going
special
review
MEAT
TRIED
night
YOURE
BAR
used
better
super
experience
Copyright 2021
tagxedo.com

PLACE
FOOD
sandwich
great
good
DELICIOUS
service
time
love
best
order
make
try
CHEESE
DONT
RESTAURANT
REALLY
DEFINITELY
PIZZA
fries
coffee
IVE
SAUCE
Sweet
right
review
table
came
tried
meal
day
flavor
thing
eat
SALAD
WANT
PEOPLE
super
going
BURGER
visit
better
EXCELLENT
CUSTOMER
WORK
pork
little
lot
worth
years
FRIENDS
NEW
special
TASTE
EXPERIENCE
SERVED
DRINKS
OPEN
helpful
line
pretty
free
toast
QUALITY
ATTENTIVE
PARKING
HAPPY
reasonable
enjoy
dinner
long
beer
selection
toppings
SEASONED
SPOT
SMALL
SERVER
bagel
live
FEEL
HOME
HUSBAND
DIFFERENT
DISAPPOINTED
AREA
DIDNT
sausage
counter
biscuits
portions
potato
highly
options
busy
soup
bacon
french
beef
brunch
ONION
STARS
SURE
pickles
COOKED
youre
walk
local
inside
person
expect
need
austin
crispy
items
tasty
SEATING
STOP
KNOW
egg
quick
decided
wonderful
INCREDIBLE
WELCOMING
lobster
HOUSE
dishes
CLEAN
offer
perfectly
usually
VEGGIE
NIGHT
THERES
including
EMPLOYEES
pastries
outside
OWNER
brisket
grilled
TODAY
CHECK
SHOP
ingredients
fantastic
absolutely
choice
especially
DESSERT
chocolate
STREET
COMFORTABLE
IMPRESSED
slice
LARGE
FAMILY
THOUGHT
CREAM
PORTLAND
Copyright 2021
tagxedo.com

Thank you!