

# What Are OpenTable Diners Talking About?

**Sudeep Das, Ph.D. @datamusing**  
**Data Scientist**  
**OpenTable**



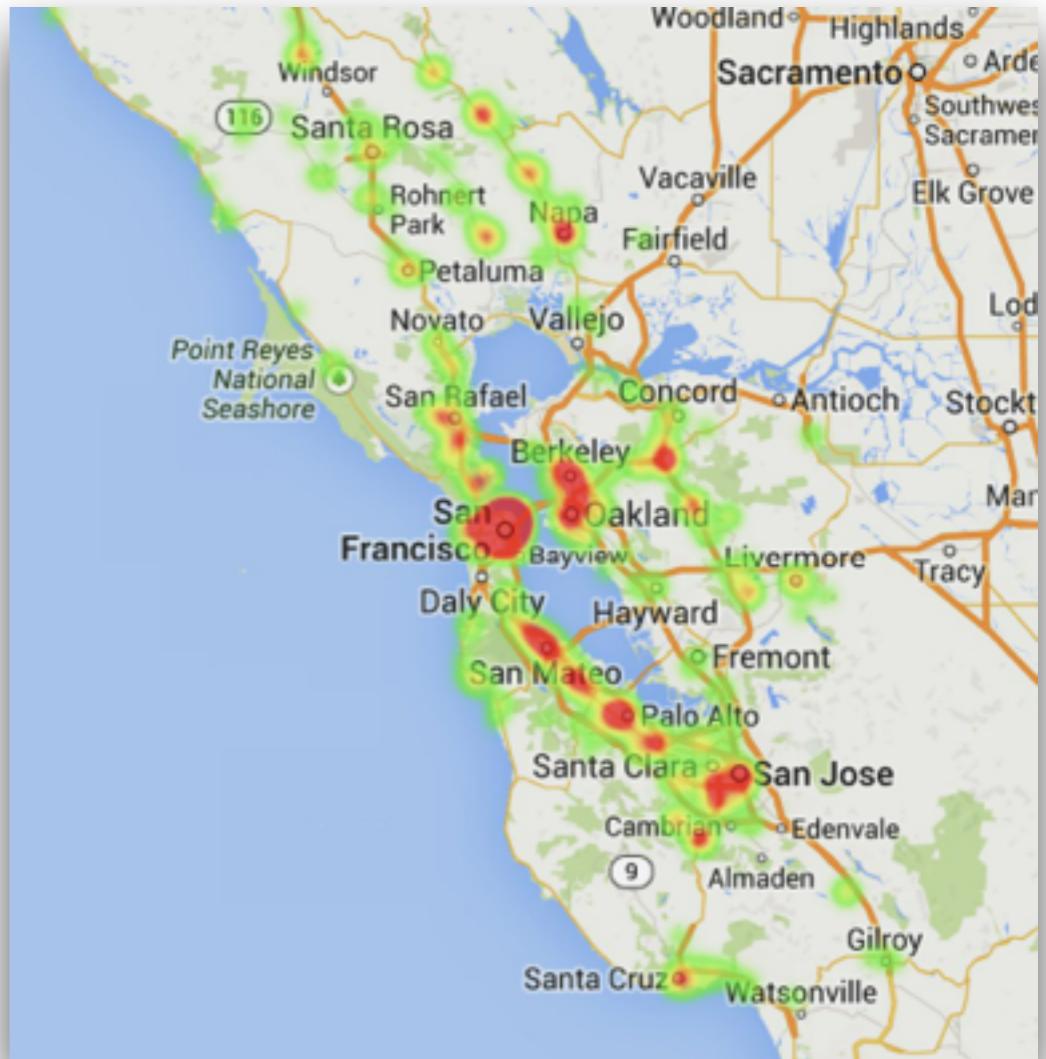
# OpenTable in numbers

- Our network connects diners with more than **31,000 restaurants** worldwide.
- Our diners have spent more than **\$20 billion** at our partner restaurants.
- OpenTable seats more than **15 million** diners each month.
- Every month, OpenTable diners write more than **400,000** restaurant reviews.

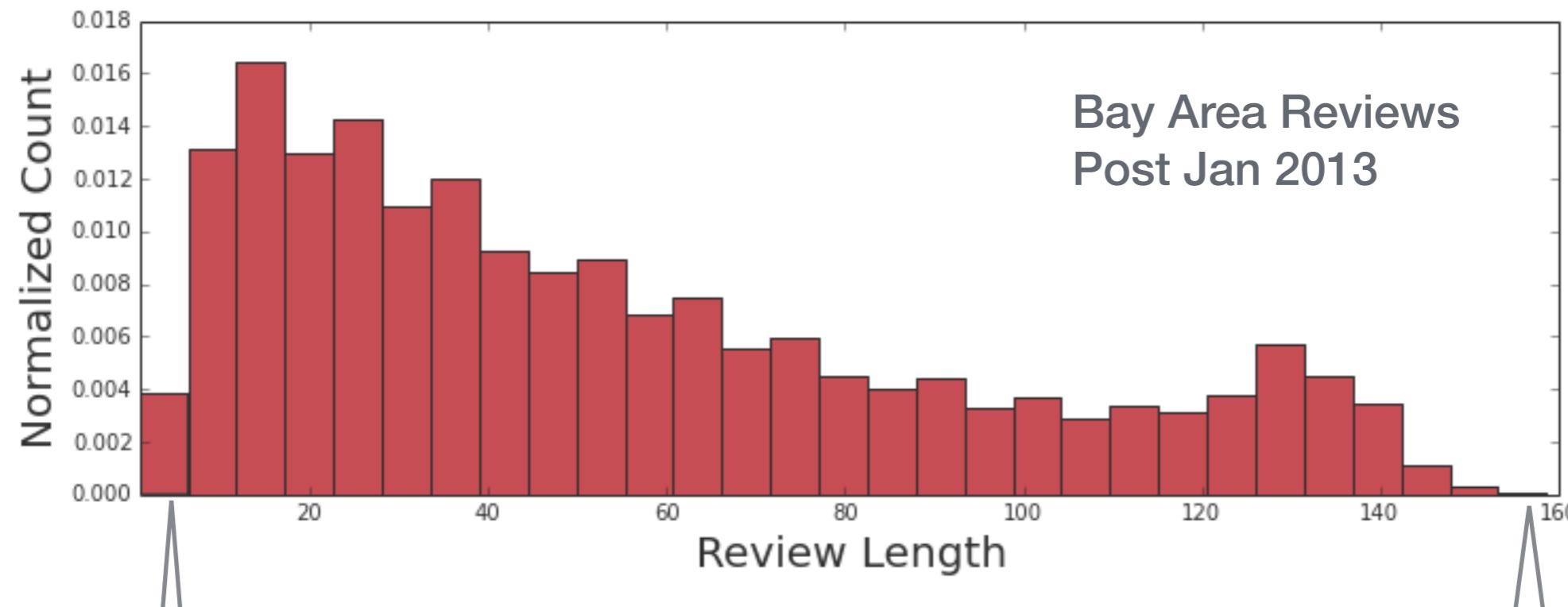


# We've got reviews!

- Our network connects diners with more than **31,000 restaurants** worldwide.
- Our diners have provided more than **25 million reviews**.
- For our first set of reviews, we looked at the **San Francisco Bay Area** where we have technology in about half of all reservation-taking restaurants and seated about 40% of all diners in this market.



# Reviews come in all shapes and sizes!



**"SUPERB!"**

This really is a hidden gem and I'm not sure I want to share but I will. :) The owner, Claude, has been here for 47 years and is all about quality, taste, and not overcharging for what he loves. My husband and I don't often get into the city at night, but when we do this is THE place. The Grand Marnier Souffle' is the best I've had in my life - and I have a few years on the life meter. The custard is not over the top and the texture of the entire dessert is superb. This is the only family style French restaurant I'm aware of in SF. It also doesn't charge you an arm and a leg for their excellent quality and that also goes for the wine list. Soup, salad, choice of main (try the lamb shank) and choice of



# Learning Topics

## Topics

gene	0.04
dna	0.02
genetic	0.01
...	

life	0.02
evolve	0.01
organism	0.01
...	

brain	0.04
neuron	0.02
nerve	0.01
...	

data	0.02
number	0.02
computer	0.01
...	

## Documents

### Seeking Life's Bare (Genetic) Necessities

COLD SPRING HARBOR, NEW YORK—How many genes does an organism need to survive? Last week at the genome meeting here,\* two genome researchers with radically different approaches presented complementary views of the basic genes needed for life. One research team, using computer analyses to compare known genomes, concluded that today's organisms can be sustained with just 250 genes, and that the earliest life forms required a mere 128 genes. The other researcher mapped genes in a simple parasite and estimated that for this organism, 800 genes are plenty to do the job—but that anything short of 100 wouldn't be enough.

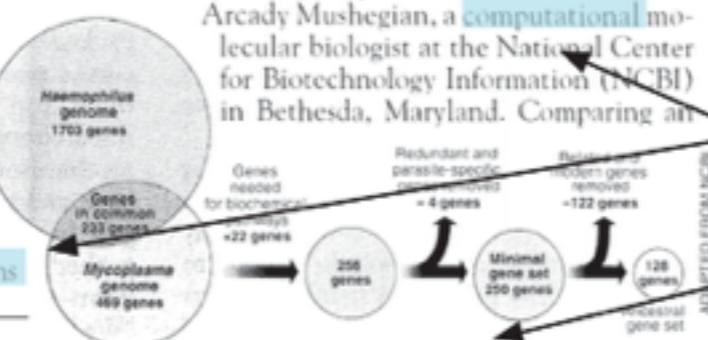
Although the numbers don't match precisely, those predictions

\* Genome Mapping and Sequencing, Cold Spring Harbor, New York, May 8 to 12.

SCIENCE • VOL. 272 • 24 MAY 1996

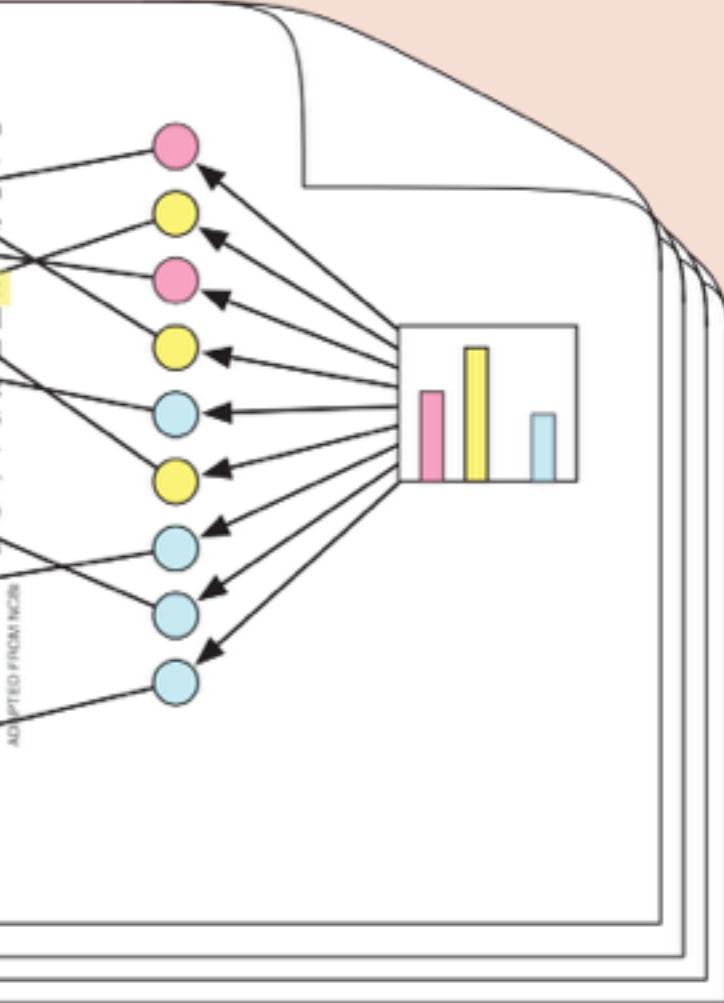
"are not all that far apart," especially in comparison to the 75,000 genes in the human genome, notes Siv Andersson of Uppsala University in Sweden, who arrived at the 800 number. But coming up with a consensus answer may be more than just a genetic numbers game, particularly as more and more genomes are completely mapped and sequenced. "It may be a way of organizing any newly sequenced genome," explains

Arcady Mushegian, a computational molecular biologist at the National Center for Biotechnology Information (NCBI) in Bethesda, Maryland. Comparing all



Stripping down. Computer analysis yields an estimate of the minimum modern and ancient genomes.

## Topic proportions and assignments



# From diner-talk to insights!

Analyze the global corpus of reviews in a geographic region to learn **topics**

---

Classify topics into **categories** (food, ambiance, service etc...)

---

**Map** topics back to restaurants

---

For each restaurant and a topic, surface **relevant reviews**

---

Use learned topic weights to **find restaurants** relevant to a topic/category

---



# Topics can be leveraged to power global to local feature generation

## ULTRA LOCAL

Most talked about menu items at a restaurant:

*“If you are at Slanted Door in San Francisco, try their Cellophane Noodles with Dungeness Crab...”*

## LOCAL

Regional Dishes:

*“If you are in London, be sure to try a Meat Fruit Tipsy Cake ...”*

## GLOBAL

Dish Tags

*“Find all restaurants that serve Lemon Ricotta Pancakes”*



# We used a purely Python stack...

During the exploratory stage, we applied the **Python** stack

pandas

scikit-learn

gensim

py-nltk

We applied two main topic modeling methods:

- Latent Dirichlet Allocation (**LDA**) (Blei et al. 2003)
- Non-negative Matrix Factorization (**NMF**) (Aurora et al. 2012)

*Currently porting over the techniques to Apache Spark.*



# We used a purely Python stack...

pandas

scikit-learn

gensim

py-nltk

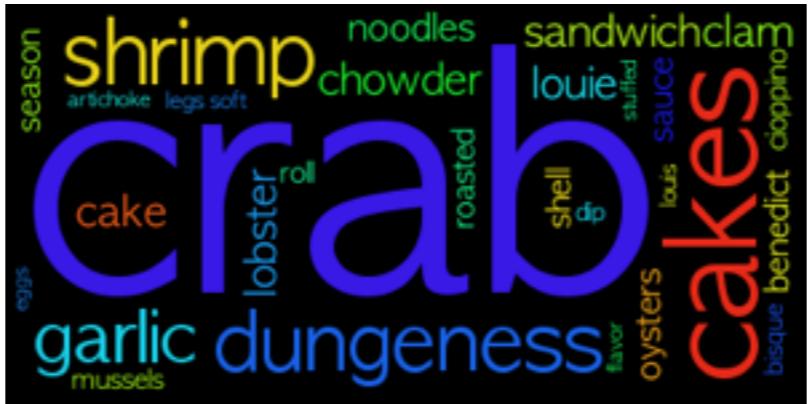
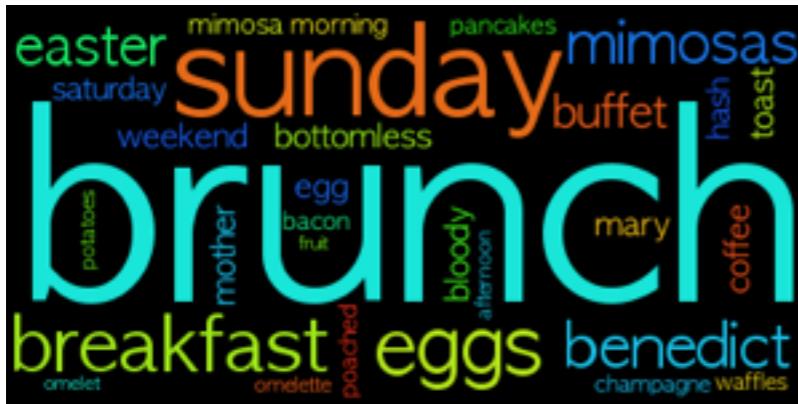
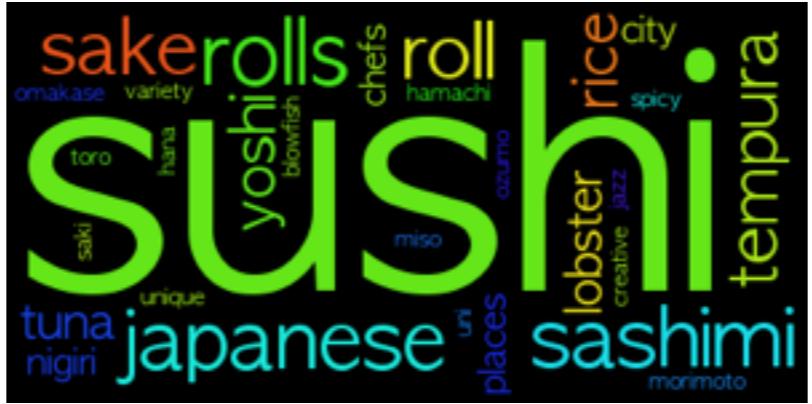
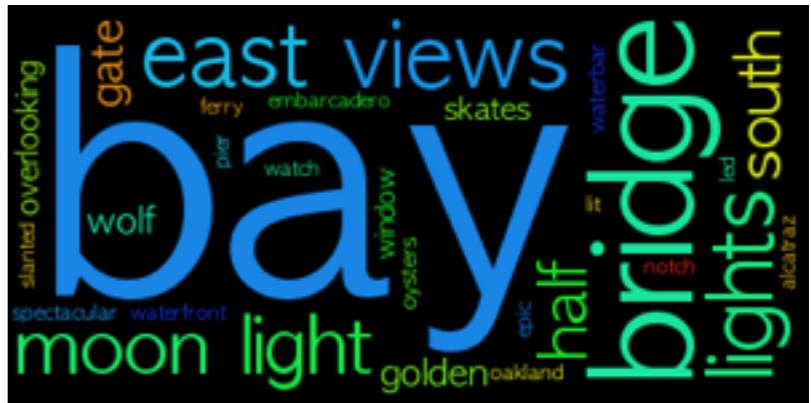
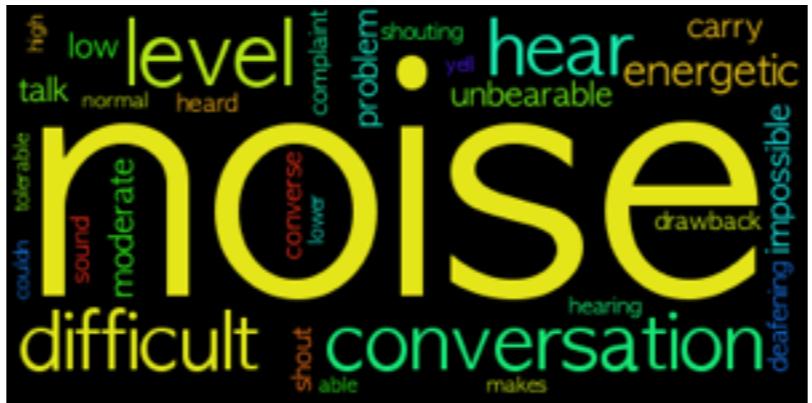
```
vectorizer = CountVectorizer(max_df=0.9,  
                            max_features=10000,  
                            stop_words="english")
```

```
counts = vectorizer.fit_transform(the_reviews.revs.ReviewText.values)
```

```
decomp = decomposition.NMF(n_components=self.n_topics)  
.fit(vectorized_reviews.counts)
```



# Some topics from the SF Bay Area



# Sudeep Das



# Not all topics are created equal!

The global topics cluster into six major categories:

Food	Drinks	Ambiance	
Value	Service	Special occasion	Generic

---

We found **200** to be a good number of topics to learn.

We manually label the first crop of topics, and subsequently use a classifier to relabel any new crop of topics against the labeled ones.

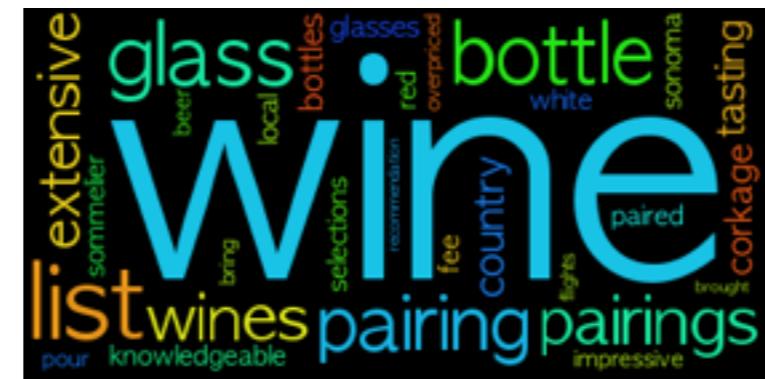


# Topics and categories

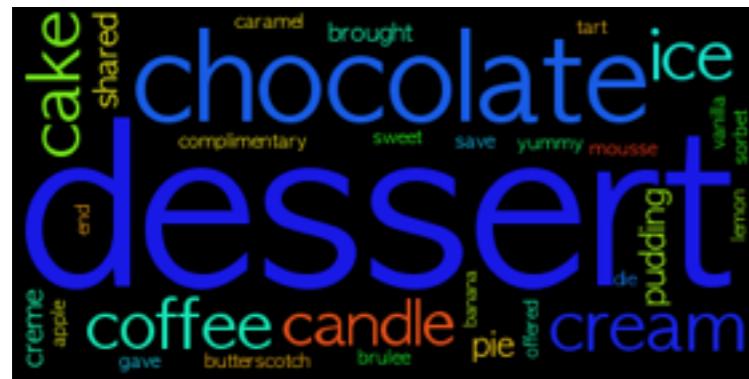
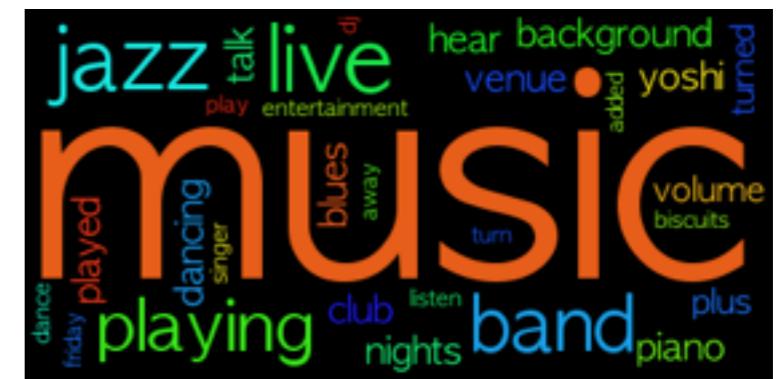
## Food



# Drinks

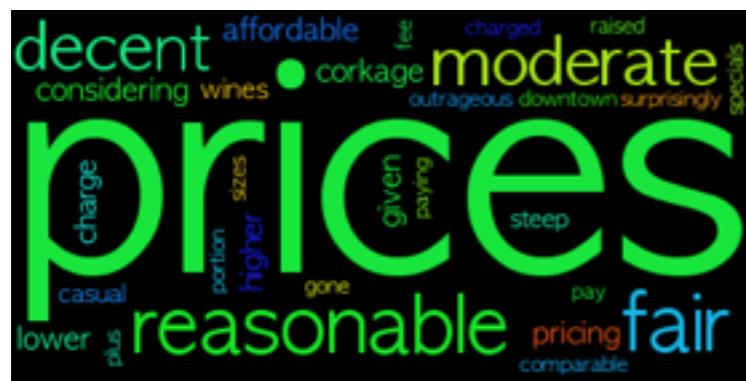


# Ambiance

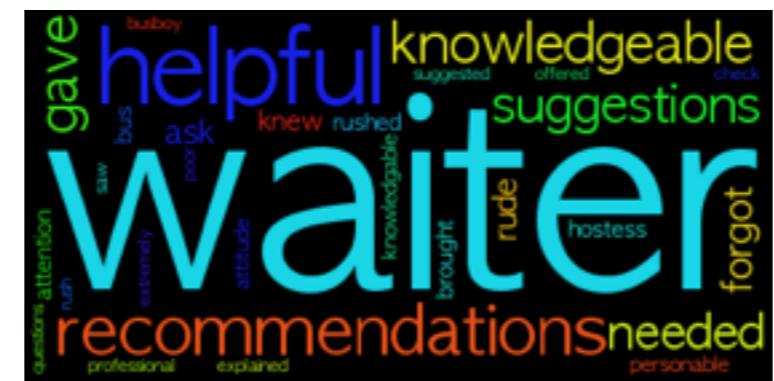


# Topics and categories

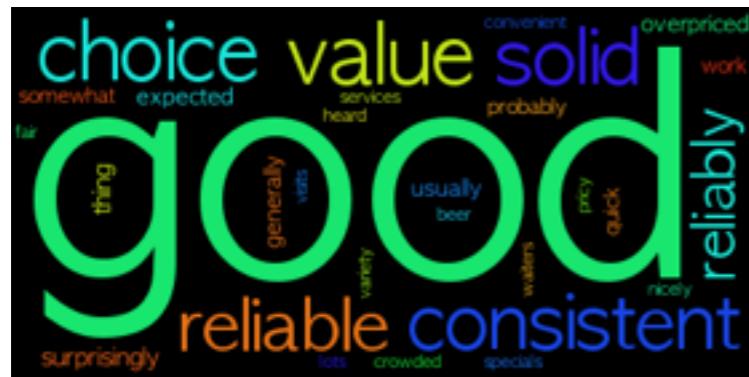
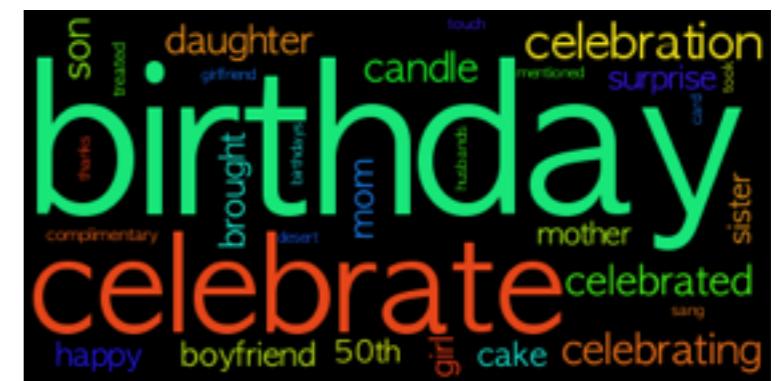
# Value



## Service



# Special occasion



We ignore very generic topics that has to do with just going to a restaurant.



# Regional nuances in topics

We ran topic analyses separately for each of four metropolitan areas:

**SAN FRANCISCO**

**CHICAGO**

**NEW YORK**

**LONDON**



# Bay view, ocean view, city view ...

## SAN FRANCISCO

bridge beat window  
like lights romantic scenic  
seat wharf sun pacific  
viewvieW harbor beach  
mountain sunset gorgeous diff  
seats marina sun pacific  
spectacular breathtaking

gate east views  
slanted overlooking ferry  
gate pier watch  
wolf window oysters  
spectacular waterfront  
moon light golden oakland  
half bridge led south  
it notch alcatraz



## CHICAGO

scenic lake windows  
overlooking windy river  
windows bub big  
avenue live kansas  
favorites park riva  
tower pier signature  
burbs window floor north  
spectacular views cte michigan  
spots skyline suburbs

## LONDON

tower paul window  
terrace shard waffle paramount modern  
requested overlooking barbecue  
paul vieW floor park outside tate  
city sun stuning skyline cathedral  
duck river thames bridge

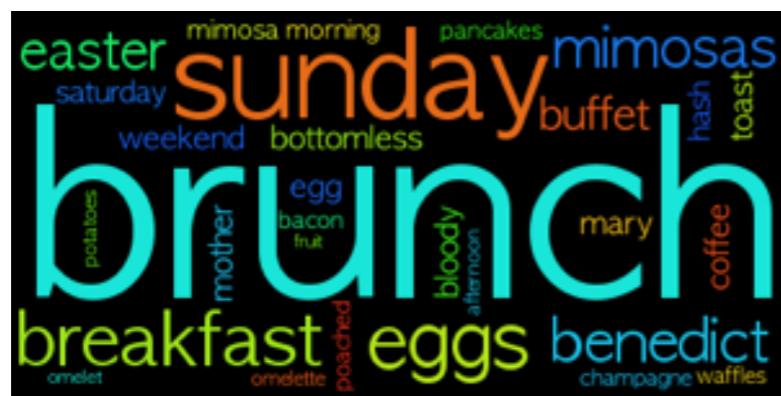
## NEW YORK

river windows robert  
circle lake skyline  
incredible manhattan water outdoor  
sunset skyline hudson beat  
spectacular window scenic columbus  
gorgeous seat outside park central  
outdoor charts views

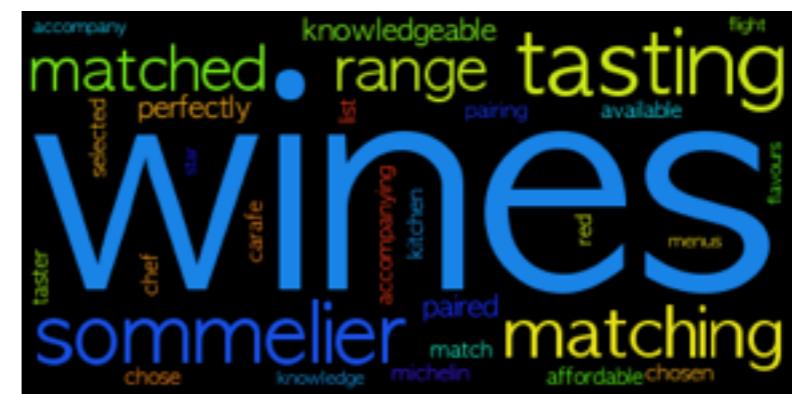
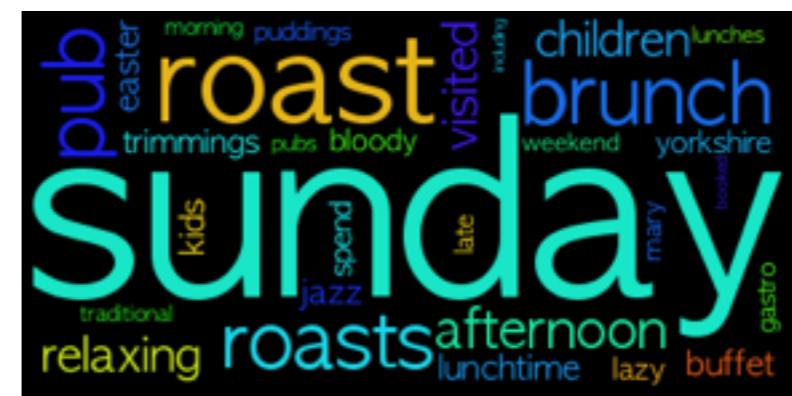


# Sunday Brunch or Sunday Roast?

US CITIES

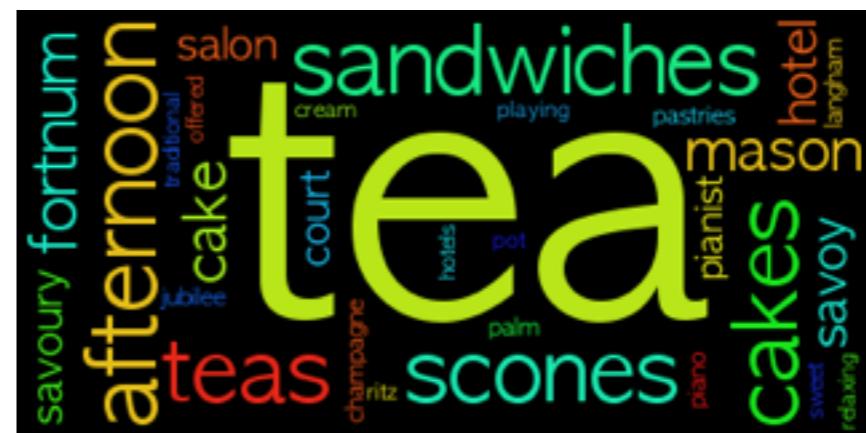
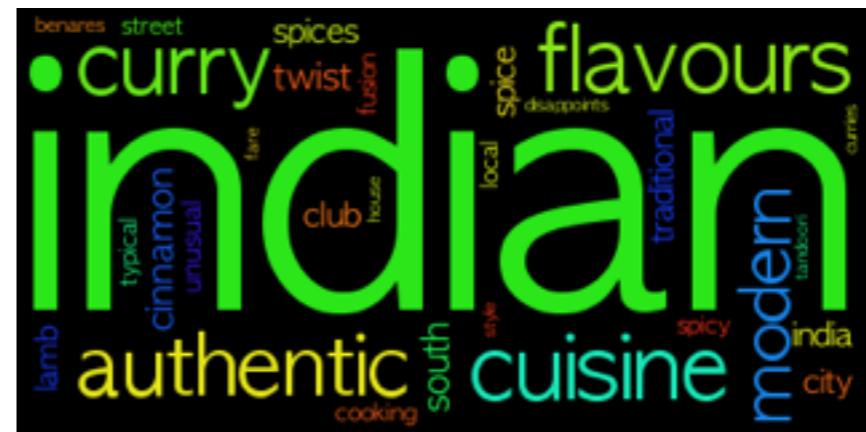
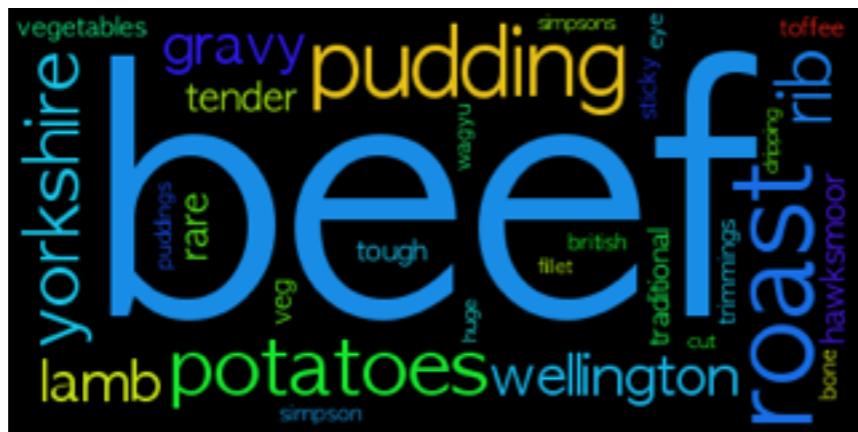


LONDON



# Yorkshire Pudding, Curry and Afternoon Teas!

# Some topics that are *dominant* only in London



# Reassigning topics back to restaurants

Each **review** for a given restaurant has certain **topic distribution**

Combining them, we identify the **top topics for that restaurant.**



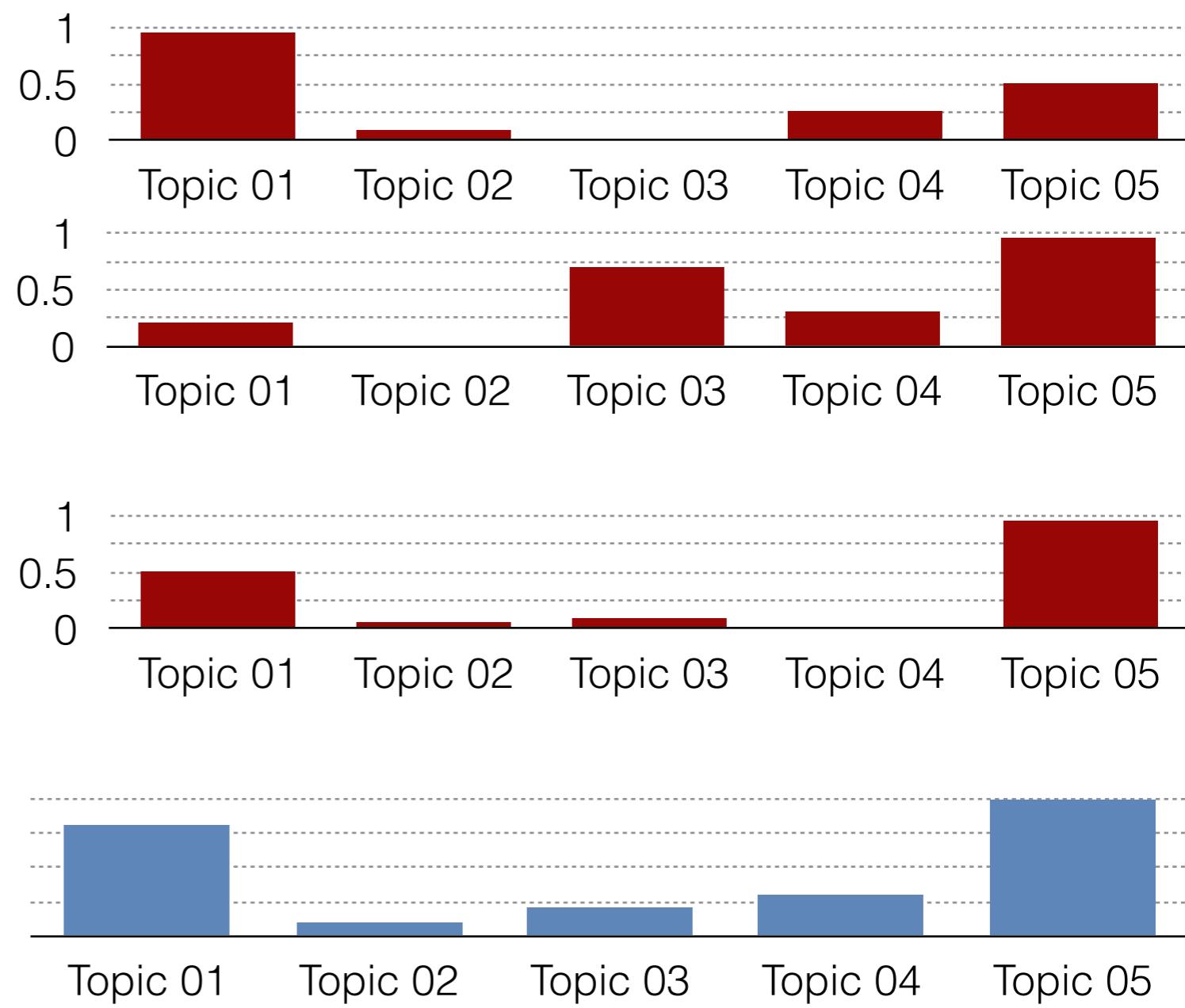
review 1

review 2

:

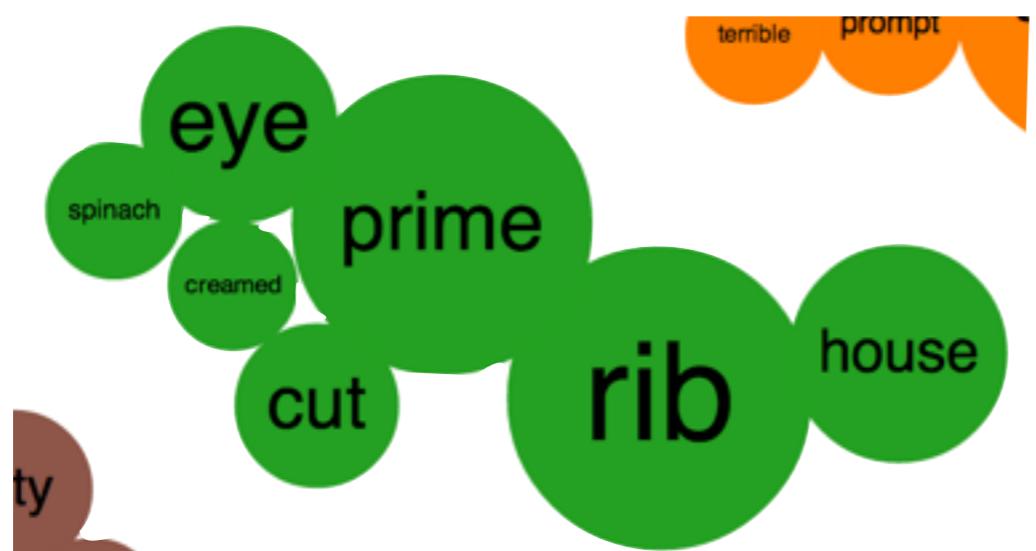
review N

Restaurant



# From Restaurant Topics to Reviews

Once we have identified the top topics for a restaurant, we **score reviews** according to the weight of that topic, and surface the **top reviews**.



Most amazing **prime rib** I have EVER had!!!!

nothing but **prime rib** or a special of the day fish on menu, Excellent **prime rib**. if everyone in your party likes **prime rib** this is the place to go.

ohhhhhhhh all was so good,,, drinks & **prime rib!!!**

Excellent service, best **prime rib** ever

My favorite **prime rib** restaurant.

As always, the **prime rib** is done to perfection.

I am always on the look out for great **prime rib** or even a place that can do **prime rib** better



# Time for a little demo!

Python Flask + D3.js

Deployed with Docker

<http://restaurant-topics.herokuapp.com/>



# Using Topics to Find Relevant Restaurants!

We can now turn the topic analysis over its head!

---

Once a topic is chosen, we can **rank restaurants** according to its relevance to that topic!

---

That, now, becomes a way to **find restaurants** according to a topic!

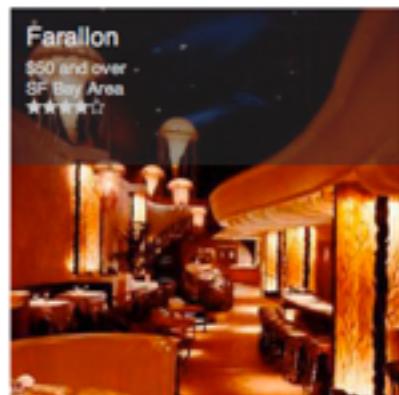
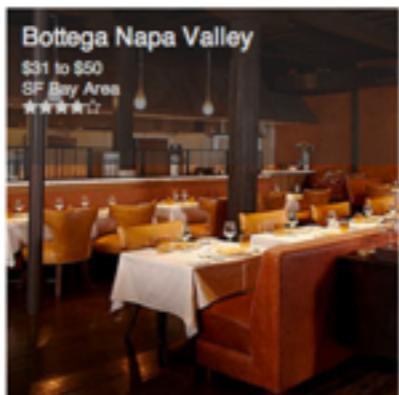
---

*“Find me all restaurants with a great wine list”*



# Using Topics to Find Relevant Restaurants!

Top restaurants in the **wine list topic**



*“Find me all restaurants with a great wine list”*



# From N-grams and topics to complex dish tags ...

An n-gram is a contiguous sequence of n items from a given sequence of text. ... (e.g. to be, be or, or not, not to, to be )

---

Most dishes are usually **1-grams** (“tiramisu”) **2-grams** (“pork cutlets”) or **3-grams** (“lemon ricotta pancake”)

---

For each restaurant, we perform an N-gram analysis of the reviews within the scope of food topics and surface **candidate dish tags**

---

*We were able to generate several thousands of dish tags using this methodology!*



# From N-grams and topics to complex dish tags ...

- braised short rib
- soft shell crabs
- rib eye steaks
- chilean sea bass
- sweet potato fries
- beef short rib
- crab cake appetizers
- flourless chocolate cakes
- ny strip steaks
- mixed green salad
- king crab leg
- roasted red peppers
- ice cream dessert
- bread pudding dessert
- lump crab meat
- bacon wrapped scallop
- vanilla bean ice cream
- goat cheese appetizers
- ice cream sundaes
- blue cheese salad
- steak au poivre
- roasted brussel sprouts
- grilled pork chops
- foie gras appetizer
- squid ink pasta
- wild mushroom risotto
- cheese stuffed olives
- braised beef short ribs
- tuna tartare appetizer
- potato leek soup
- creme brulee dessert
- seafood pasta dish
- short rib entree
- twice baked potato
- filet mignon
- tea ice cream
- lemon butter sauce
- dry aged steaks
- homemade potato chips
- white bean soup
- butter ice cream
- mahi mahi special
- blue cheese olives
- grilled octopus appetizers
- duck confit salad
- green tea ice cream



# From dish tags to most talked about menu items

For each restaurant we find the dish tags with the strongest signal

---

We perform a fuzzy matching with the menu data (both name and description) from SinglePlatform

---

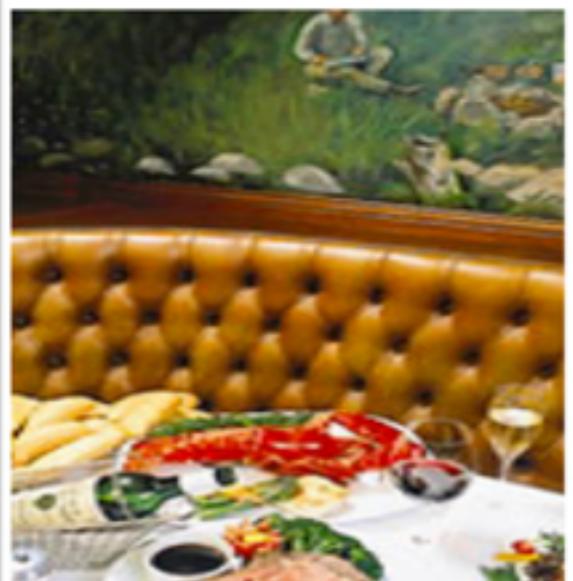
This lets us automagically disambiguate the various ways people refer to the same dish!

---

For e.g “crab noodles”, “noodles with dungeness”, “cellphone noodles with crab” all refer to the same dish “**Cellophane Noodles with Dungeness Crab**” on the menu.



# Most talked about menu Items



BAY AREA

## HARRIS'

- the harris steak a thick bone-in new york
- filet mignon
- harris steak tartare
- steak diane a paillard of filet mignon grilled and finished with a demi glace of cognac and shallots
- creamed spinach



BAY AREA

## PRESIDIO SOCIAL CLUB

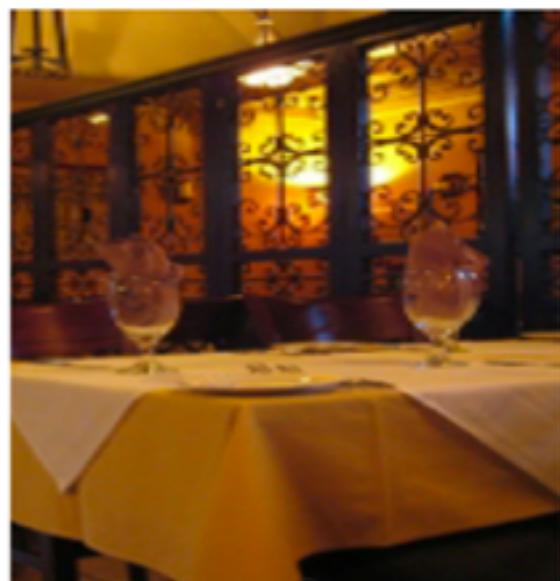
- lemon meringue cake light, fluffy, and delicious
- meatloaf shaved carrots, mashed potatoes
- pork porterhouse chop spatzle, bacon, apricot mustard
- butter lettuce salad radishes, champagne vinaigrette, fines herbs
- fish of the day daily preparation
- market vegetable salad little gem lettuce, pecorino cheese
- psc gruyere cheese toasts tomato fondue



BAY AREA

## THE SPINNAKER - SAUSALITO

- sea bass grilled served with strawberry mango salsa
- soup of the day or clam chowder
- caesar salad the spinnaker's special version of this delightful salad. prepared at your table for a minimum of two persons
- sea bass saute with white truffle oil, portobello and crimini mushrooms, roast garlic and roma tomatoes
- crab cakes dungeness and rock crab



BAY AREA

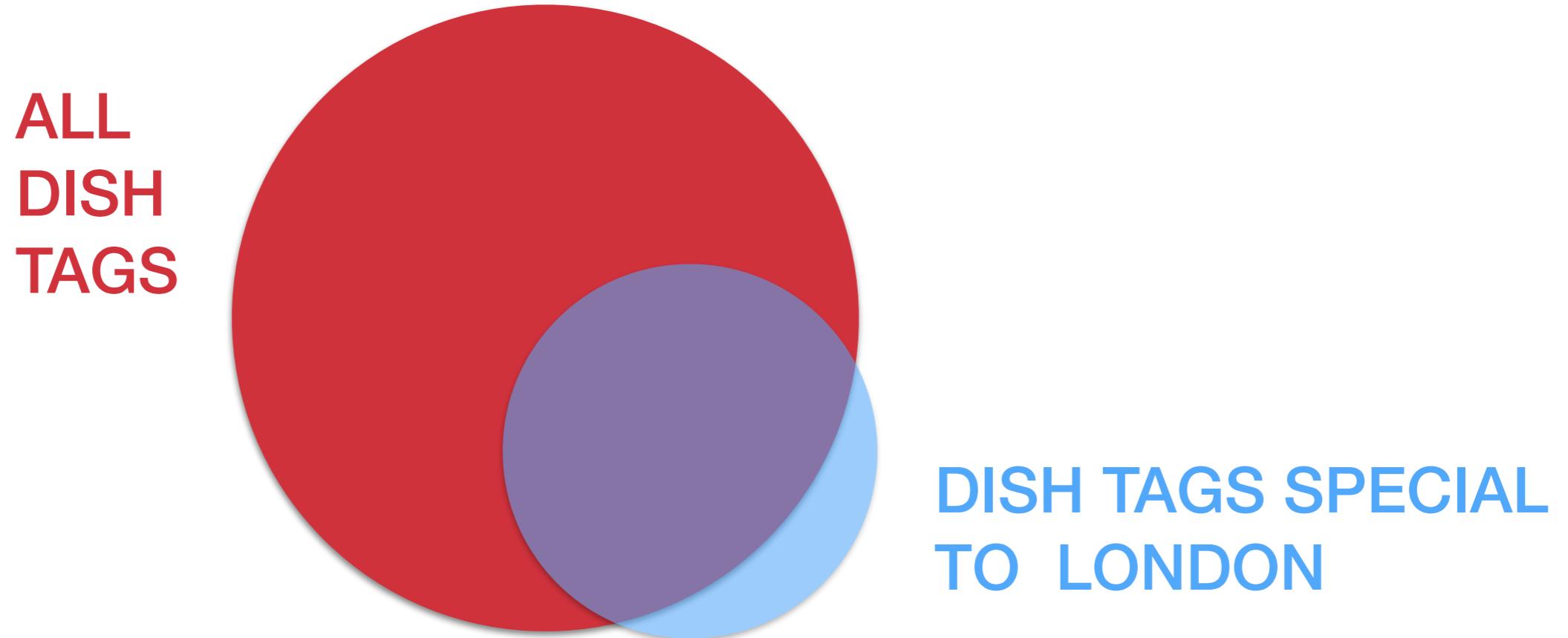
## A BELLAGIO

- bistecca al rosмарino new york strip steak with fresh rosemary, garlic and olive oil
- risotto porcini risotto in porcini mushroom sauce
- vongole alla marinara clams in a garlic and white wine sauce served with fresh tomatoes, parsley
- coscia dagnello lamb shank with onions, carrots, celery in a marinara sauce over polenta
- gnocchi gorgonzola homemade potato gnocchi in a velvety gorgonzola cream sauce



# Finding regional dishes!

Because we performed dish tag discovery *by metro area*, we were able to also surface the most unique dishes per metro.



# Finding regional dishes!



## One City, Nine Dishes!

Did you know you could get "Flower Pot Bread" in Toronto? We had some fun applying Natural Language Processing tools on fine dining restaurant reviews from the OpenTable review corpus. Using a combination of topic modeling and n-gram analysis we were able to find many unique dishes for some of the prominent metros using our services. Here, we show the top nine picks for each city! Hover on a city's circle to reveal the dishes!



roasted pig face



miso butterscotch budino



finger pork chop

## CHICAGO



600 day prosciutto



da pork chop



tomato gin soup



pot roast nachos



hot mess fries



oven roasted pig



# Finding regional dishes!



## One City, Nine Dishes!

Did you know you could get "Flower Pot Bread" in Toronto? We had some fun applying Natural Language Processing tools on fine dining restaurant reviews from the OpenTable review corpus. Using a combination of topic modeling and n-gram analysis we were able to find many unique dishes for some of the prominent metros using our services. Here, we show the top nine picks for each city! Hover on a city's circle to reveal the dishes!



squid and mackerel burger



roast beef yorkshire



beef dripping fries

### LONDON



triple cooked chips



savoy afternoon tea



steak ale pudding



shell crab burger



meat fruit tipsy cake



salted caramel rolos



# Finding regional dishes!



## One City, Nine Dishes!

Did you know you could get "Flower Pot Bread" in Toronto? We had some fun applying Natural Language Processing tools on fine dining restaurant reviews from the OpenTable review corpus. Using a combination of topic modeling and n-gram analysis we were able to find many unique dishes for some of the prominent metros using our services. Here, we show the top nine picks for each city! Hover on a city's circle to reveal the dishes!



duck poutine pizza



spicy spanish fries



kobe beef tacos



TORONTO



flower pot bread



breton snow crab



tapioca coconut pudding



crispy duck egg



smoked mackerel oatcakes



crab dip triscuits



# Evaluating Spark+MLlib...

```
import org.apache.spark.mllib.recommendation.{ALS, Rating, MatrixFactorizationModel}

val ratings = sc.textFile("./mllib/data/chicago2.mtx").map { line =>
    val fields = line.split(" ")
    // format: (timestamp % 10, Rating(userId, movieId, rating))
    (Rating(fields(0).toInt, fields(1).toInt, fields(2).toDouble))}

als.setIterations(100)
als.setLambda(1)
als.setRank(20)
als.setBlocks(2)
als.setNonnegative(true)
val model = als.run(ratings)
model.productFeatures.take(3)
model.userFeatures.map{ case (id, vec) => id + " " + vec.mkString(" ") }
    .saveAsTextFile("./userFeatures2")
model.productFeatures.map{ case (id, vec) => id + " " + vec.mkString(" ") }
    .saveAsTextFile("./productFeatures2")
```



# Future Directions

**Sentiment Analysis on Reviews**

**Trending Topic and Sentiments**

**And a lot more ...**

