
yelp
Intro to Topic Modeling

Scott Triglia

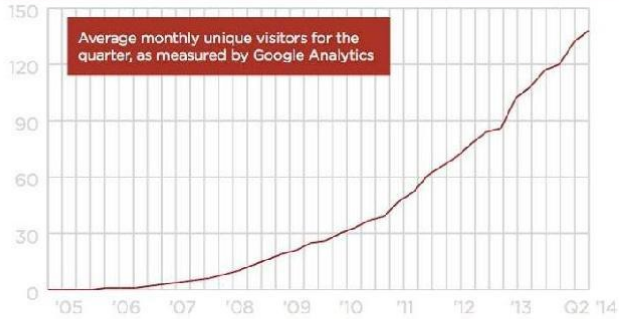
- Been a software engineer on the Search team for 3 years
- Harvey Mudd College and UC Irvine
- Recommendations, Geocoding, Data Quality

yelp

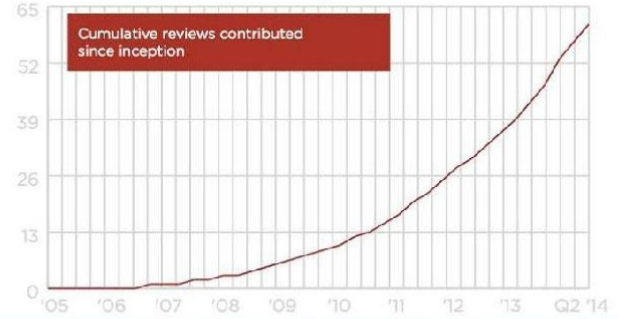
AN INTRODUCTION TO YELP

Metrics as of June 30, 2014

138 MILLION MONTHLY VISITORS



61 MILLION REVIEWS



YELP MOBILE METRICS*

68 MILLION

The approximate number of average monthly unique visitors who used Yelp via their mobile device.

500,000

The number of phone calls made to businesses every day through Yelp's mobile apps.

400,000

The number of directions generated to businesses every day through Yelp's mobile apps.

61%

The percentage of searches on Yelp that came from mobile devices across the globe.

*As of Q2 2014

US Demographics

Age:

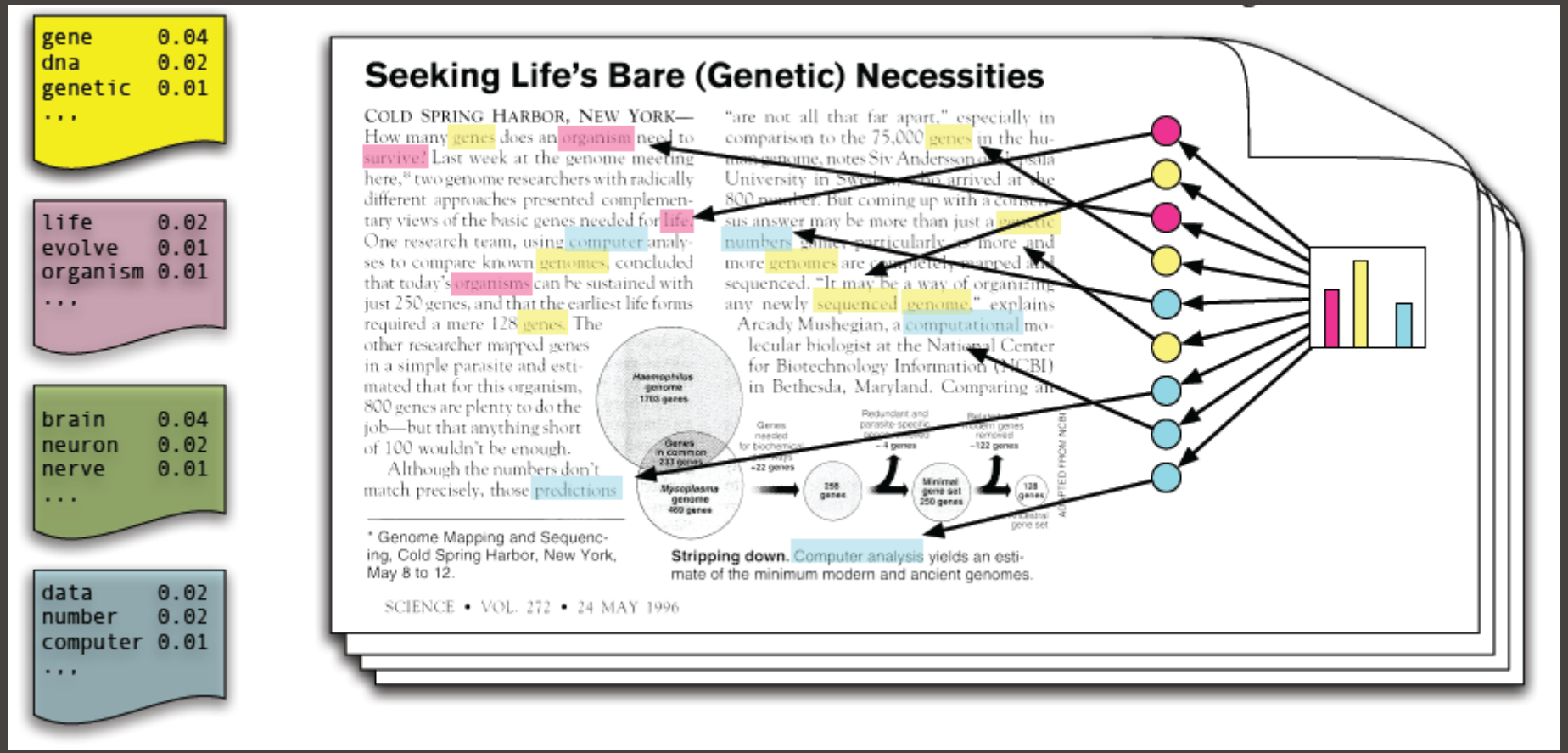
Education:

Income:

Source: comScore. Age and Income data via Media Metrix report as of June 2014. Education data via Plan Metrix report as of May 2014.

REVIEWED BUSINESSES IN EVERY CATEGORY

Today: Topic Modeling!



How do we find the topics discussed in a piece of text?

First off, let’s talk about modeling text at Yelp

Two sample projects today:

Highlights
Automatic Categorization

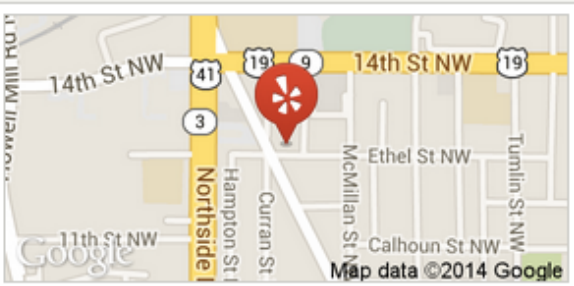
Highlights

Antico Pizza

1577 reviews | Details

Write a Review | Add Photo | Share | Bookmark

$$ · Pizza | Edit



1093 Hemphill Ave NW Edit
Atlanta, GA 30318
Westside / Home Park, Georgia Tech
Get Directions
(404) 724-2333
littleitalia.com

This thing

"This pizza is the **real deal** its superior ingredients are what sets this place apart." in 16 reviews

"Piles of ingredients from Italy and their prized pizza ovens & their epic trip from **Naples**." in 39 reviews

"Order the **Diavola** if you dont mind spicy peppers and delicious cured meats

Today **11:30 am - 11:30 pm**
Open now

Full menu

$$$$ Price range **$11-30**

Work here? Claim this business

Highlights | Query | Metrics | **Entities** | Survey

San Francisco, CA | San Tung Chinese Rest | Go

#12677427 ★★★★☆

Every time I walked past this place at night, it was always packed and sometimes out the door with people. A friend and I came here last month to check out what the food hype was here which made it so popular. It looked like a typical Chinese eatery to me until we sat down and looked at the menu. The menu came with a good description in English and in Chinese. Plus it came with pictures! I'm a visual person and I like to see pretty pictures. Then again we were also observing what other patrons had ordered. Other than a Chinese cuisine style restaurant, most of the items on the menu are northern style. Most of the Chinese restaurants around the SF Bay Area are southern style and come in large portions. This place also gave us large portions, but the service was excellent. I didn't have to constantly ask them to refill my tea pot, they checked the pot every 10 minutes to see if it was half full. Maybe they knew I was coming haha. Onward to the food... While we looked at the pretty menu, my friend and I were served tea, pickled vegetables and roasted peanuts. I don't know many places which give you this kind of service! Anyways, we ordered the dry fried chicken wings which was delicately infused with spices and marinated in a tangy sauce. We were kinda wondering what the Szechuan Tea Smoked Duck would taste like, so we ordered that too. It was a pricey $12 but it was well worth what we ordered. As for veggies, we went with our favorite Eggplant sauteed in a garlic sauce and Chinese mustard greens, plus rice of course. You can't go wrong with ordering mustard greens. There's not many Chinese eateries that I know of which serve tea infused or dry fried meats. I guess these two items really stood out for me on the menu. If you would like to come and eat at San Tung's, I highly recommend coming here at the ass crack of dawn and.. I'm just kidding. It's better to come here early for dinner or lunch since this place fills up fast. Parking is a bit hard to find around here, but I suggest parking 5-6 blocks away and walking to the restaurant. Therefore if you are feeling a bit bloated, you can walk it off as well.

#31751318 ★★★★☆

Alright, let's get the obvious out of the way - the chicken is good. Crazy good. Worth the wait good. Stood outside in the cold SF summer just for this. However, I do have to say that though the quality of the chicken is great, the sauce is not as unique as one may think. The rest of the dishes were good. String beans are delicious and recommended. The dumplings are nice - but not up to par. Surprisingly, the hot and sour soup was really really good, and this is coming from someone who isn't normally partial to it. Definitely worth it if you have been waiting in the dead cold of summer. 2 orders of dried fried chicken. string beans. dumplings. hot and sour soup. rice = $50 between 3 people. Street parking, both free and metered.

#45538417 ★★★★★

THE CHICKEN WINGS OMGGGGG!!!!! They're so delicious!! It's sweet but not too sweet, just a hint of spice, and that satisfying crunch. I prefer bones because I like munch on something, but some people think bones are too messy but who cares?! IT'S THE SAME WINGS!! Service could be better, and the wait is long (especially those cold days)

Highlights | Query | Metrics | **Entities** | Survey

San Francisco, CA | San Tung Chinese Rest | Go

Menu Items

#12677427 ★★★★☆

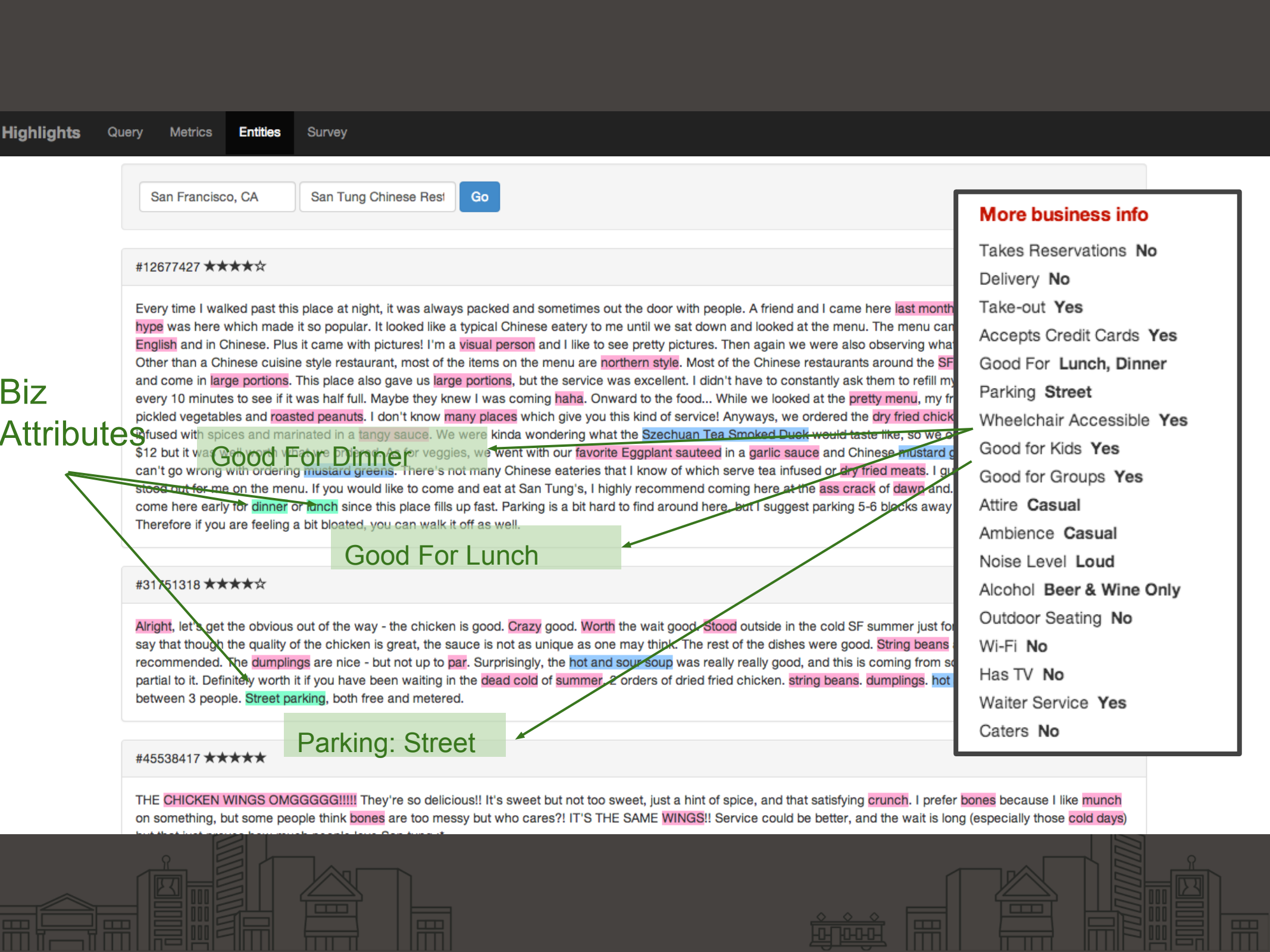
Every time I walked past this place at night, it was always packed and sometimes out the door with people. A friend and I came here last month to check out what the food hype was here which made it so popular. It looked like a typical Chinese eatery to me until we sat down and looked at the menu. The menu came with a good description in English and in Chinese. Plus it came with pictures! I'm a visual person and I like to see pretty pictures. Then again we were also observing what other patrons had ordered. Other than a Chinese cuisine style restaurant, most of the items on the menu are northern style. Most of the Chinese restaurants around the SF Bay Area are southern style and come in large portions. This place also gave us large portions, but the service was excellent. I didn't have to constantly ask them to refill my tea pot, they checked the pot every 10 minutes to see if it was half full. Maybe they knew I was coming haha. Onward to the food... While we looked at the pretty menu, my friend and I were served tea, pickled vegetables and roasted peanuts. I don't know many places which give you this kind of service! Anyways, we ordered the dry fried chicken wings which was delicately infused with spices and marinated in a tangy sauce. We were kinda wondering what the Szechuan Tea Smoked Duck would taste like, so we ordered that too. It was a pricey $12 but it was well worth what we ordered. As for veggies, we went with our favorite Eggplant sauteed in a garlic sauce and Chinese mustard greens, plus rice of course. You can't go wrong with ordering mustard greens. There's not many Chinese eateries that I know of which serve tea infused or dry fried meats. I guess these two items really stood out for me on the menu. If you would like to come and eat at San Tung's, I highly recommend coming here at the ass crack of dawn and.. I'm just kidding. It's better to come here early for dinner or lunch since this place fills up fast. Parking is a bit hard to find around here, but I suggest parking 5-6 blocks away and walking to the restaurant. Therefore if you are feeling a bit bloated, you can walk it off as well.

#31751318 ★★★★☆

Alright, let's get the obvious out of the way - the chicken is good. Crazy good. Worth the wait good. Stood outside in the cold SF summer just for this. However, I do have to say that though the quality of the chicken is great, the sauce is not as unique as one may think. The rest of the dishes were good. String beans are delicious and recommended. The dumplings are nice - but not up to par. Surprisingly, the hot and sour soup was really really good, and this is coming from someone who isn't normally partial to it. Definitely worth it if you have been waiting in the dead cold of summer. 2 orders of dried fried chicken. string beans. dumplings. hot and sour soup. rice = $50 between 3 people. Street parking, both free and metered.

#45538417 ★★★★★

THE CHICKEN WINGS OMGGGGG!!!!! They're so delicious!! It's sweet but not too sweet, just a hint of spice, and that satisfying crunch. I prefer bones because I like munch on something, but some people think bones are too messy but who cares?! IT'S THE SAME WINGS!! Service could be better, and the wait is long (especially those cold days)


Highlights
Query
Metrics
Entities
Survey
San Francisco, CA
San Tung Chinese Rest
Go
#12677427 ★★★★☆
Every time I walked past this place at night, it was always packed and sometimes out the door with people. A friend and I came here last month
hype was here which made it so popular. It looked like a typical Chinese eatery to me until we sat down and looked at the menu. The menu can
English and in Chinese. Plus it came with pictures! I'm a visual person and I like to see pretty pictures. Then again we were also observing what
Other than a Chinese cuisine style restaurant, most of the items on the menu are northern style. Most of the Chinese restaurants around the SF
and come in large portions. This place also gave us large portions, but the service was excellent. I didn't have to constantly ask them to refill my
every 10 minutes to see if it was half full. Maybe they knew I was coming haha. Onward to the food... While we looked at the pretty menu, my fr
pickled vegetables and roasted peanuts. I don't know many places which give you this kind of service! Anyways, we ordered the dry fried chick
infused with spices and marinated in a tangy sauce. We were kinda wondering what the Szechuan Tea Smoked Duck would taste like, so we o
$12 but it was well worth it! For veggies, we went with our favorite Eggplant sauteed in a garlic sauce and Chinese mustard g
can't go wrong with ordering mustard greens. There's not many Chinese eateries that I know of which serve tea infused or dry fried meats. I gu
stood out for me on the menu. If you would like to come and eat at San Tung's, I highly recommend coming here at the ass crack of dawn and
come here early for dinner or lunch since this place fills up fast. Parking is a bit hard to find around here, but I suggest parking 5-6 blocks away
Therefore if you are feeling a bit bloated, you can walk it off as well.
#31751318 ★★★★☆
Alright, let's get the obvious out of the way - the chicken is good. Crazy good. Worth the wait good. Stood outside in the cold SF summer just fo
say that though the quality of the chicken is great, the sauce is not as unique as one may think. The rest of the dishes were good. String beans
recommended. The dumplings are nice - but not up to par. Surprisingly, the hot and sour soup was really really good, and this is coming from s
partial to it. Definitely worth it if you have been waiting in the dead cold of summer. 2 orders of dried fried chicken. string beans. dumplings. hot
between 3 people. Street parking, both free and metered.
#45538417 ★★★★★
THE CHICKEN WINGS OMGGGGG!!!!! They're so delicious!! It's sweet but not too sweet, just a hint of spice, and that satisfying crunch. I prefer bones because I like munch
on something, but some people think bones are too messy but who cares?! IT'S THE SAME WINGS!! Service could be better, and the wait is long (especially those cold days)
Biz Attributes
Good For Dinner
Good For Lunch
Parking: Street
More business info
Takes Reservations No
Delivery No
Take-out Yes
Accepts Credit Cards Yes
Good For Lunch, Dinner
Parking Street
Wheelchair Accessible Yes
Good for Kids Yes
Good for Groups Yes
Attire Casual
Ambience Casual
Noise Level Loud
Alcohol Beer & Wine Only
Outdoor Seating No
Wi-Fi No
Has TV No
Waiter Service Yes
Caters No

Highlights | Query | Metrics | **Entities** | Survey

San Francisco, CA | San Tung Chinese Rest | Go

Potentially Interesting Noun Phrases

#12677427 ★★★★☆

Every time I walked past this place at night, it was always packed and sometimes out the door with people. A friend and I came here last month to check out what the food hype was here which made it so popular. It looked like a typical Chinese eatery to me until we sat down and looked at the menu. The menu came with a good description in English and in Chinese. Plus it came with pictures! I'm a visual person and I like to see pretty pictures. Then again we were also observing what other patrons had ordered. Other than a Chinese cuisine style restaurant, most of the items on the menu are northern style. Most of the Chinese restaurants around the SF Bay Area are southern style and come in large portions. This place also gave us large portions, but the service was excellent. I didn't have to constantly ask them to refill my tea pot, they checked the pot every 10 minutes to see if it was half full. Maybe they knew I was coming haha. Onward to the food... While we looked at the pretty menu, my friend and I were served tea, pickled vegetables and roasted peanuts. I don't know many places which give you this kind of service! Anyways, we ordered the dry fried chicken wings which was delicately infused with spices and marinated in a tangy sauce. We were kinda wondering what the Szechuan Tea Smoked Duck would taste like, so we ordered that too. It was a pricey $12 but it was well worth what we ordered. As for veggies, we went with our favorite Eggplant sauteed in a garlic sauce and Chinese mustard greens, plus rice of course. You can't go wrong with ordering mustard greens. There's not many Chinese eateries that I know of which serve tea infused or dry fried meats. I guess these two items really stood out for me on the menu. If you would like to come and eat at San Tung's, I highly recommend coming here at the ass crack of dawn and.. I'm just kidding. It's better to come here early for dinner or lunch since this place fills up fast. Parking is a bit hard to find around here, but I suggest parking 5-6 blocks away and walking to the restaurant. Therefore if you are feeling a bit bloated, you can walk it off as well.

#31751318 ★★★★☆

Alright, let's get the obvious out of the way - the chicken is good. Crazy good. Worth the wait good. Stood outside in the cold SF summer just for this. However, I do have to say that though the quality of the chicken is great, the sauce is not as unique as one may think. The rest of the dishes were good. String beans are delicious and recommended. The dumplings are nice - but not up to par. Surprisingly, the hot and sour soup was really really good, and this is coming from someone who isn't normally partial to it. Definitely worth it if you have been waiting in the dead cold of summer. 2 orders of dried fried chicken. string beans. dumplings. hot and sour soup. rice = $50 between 3 people. Street parking, both free and metered.

#45538417 ★★★★★

THE CHICKEN WINGS OMGGGGG!!!!! They're so delicious!! It's sweet but not too sweet, just a hint of spice, and that satisfying crunch. I prefer bones because I like munch on something, but some people think bones are too messy but who cares?! IT'S THE SAME WINGS!! Service could be better, and the wait is long (especially those cold days)

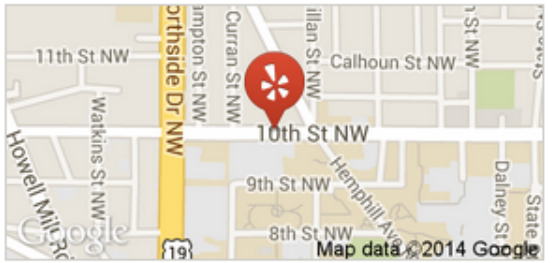
Automatic Categorization

Sublime Doughnuts

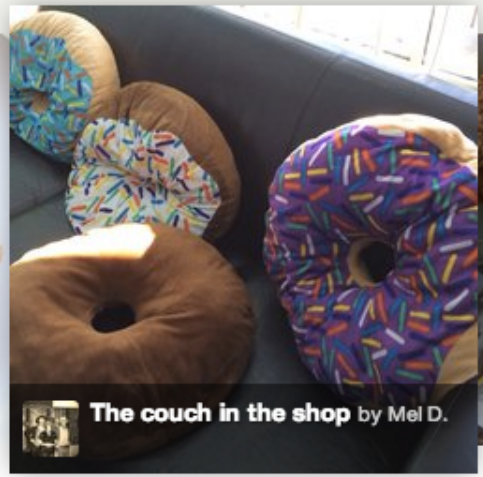
492 reviews Details

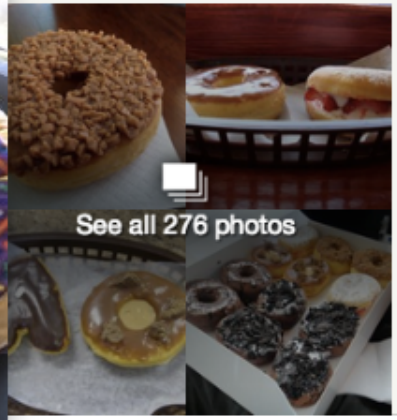
$ · Donuts Edit

Write a Review | Add Photo | Share | Bookmark



535 10th St NW Edit
Atlanta, GA 30318
Westside / Home Park, Georgia Tech
Get Directions
(404) 897-1801
Message the business
sublimedoughnuts.com





Debby L.
Atlanta, GA
Elite '14
1029 friends
959 reviews

8/14/2014

2 check-ins

Listed in ATL - COFFEE, ATL - DESSERTS

They close at 11 pm Tues -Sunday now.

I love donuts.

STAR SHAPED LEMON DONUT
One of the best donut places in Atlanta. They have a starry lemon one that is voted one of the things you must eat when in Atlanta.

I really like the creamy and custard donuts. The chocolate ones are a little too much for me, but my friends love the toffee one a lot too.

I always have to get a few glazed ones, but they don't have as much here. There is a cinnamon glazed and a chocolate swirl I think.

Debby L.
Atlanta, GA
Elite '14
1029 friends
959 reviews

8/14/2014

2 check-ins

Listed in ATL - COFFEE, ATL - DESSERTS

They close at 11 pm Tues -Sunday now.

I love donuts.

STAR SHAPED LEMON DONUT
One of the best donut places in Atlanta. They have a starry lemon one that is voted one of the things you must eat when in Atlanta.

I really like the creamy and custard donuts. The chocolate ones are a little too much for me, but my friends love the toffee one a lot too.

I always have to get a few glazed ones, but they don't have as much here. There is a cinnamon glazed and a chocolate swirl I think.

80% of your time is spent preparing data!

Useless Words

Debby L.
Atlanta, GA
Elite '14
1029 friends
959 reviews

8/14/2014

2 check-ins

Listed in ATL - COFFEE, ATL - DESSERTS

They close at 11 pm Tues -Sunday now.

I love donuts.

STAR SHAPED LEMON DONUT
One of the best donut places in Atlanta. They have a starry lemon one that is voted one of the things you must eat when in Atlanta.

I really like the creamy and custard donuts. The chocolate ones are a little too much for me, but my friends love the toffee one a lot too.

I always have to get a few glazed ones, but they don't have as much here. There is a cinnamon glazed and a chocolate swirl I think.

Processing

Debby L.
Atlanta, GA
Elite '14
1029 friends
959 reviews

8/14/2014

2 check-ins

Listed in ATL - COFFEE, ATL - DESSERTS

They close at 11 pm Tues -Sunday now.

I love donuts.

STAR SHAPED LEMON DONUT
One of the best donut places in Atlanta. They have a starry lemon one that is voted one of the things you must eat when in Atlanta.

I really like the creamy and custard donuts. The chocolate ones are a little too much for me, but my friends love the toffee one a lot too.

I always have to get a few glazed ones, but they don't have as much here. There is a cinnamon glazed and a chocolate swirl I think.

Why doesn’t Yelp use all these cool models I hear about in classes?

Diverse use cases

Product needs are rarely satisfied by simple application of existing Topic Models

Need to run this over 40M+ businesses!

Performance on benchmarks is rarely the whole story

We need a system we can understand and modify safely!

Today's Challenge!

Take a sample dataset and try to model it!

Github repo

Section 1: Preparation

Goal: Load data from file and clean it up!

- Unpack JSON `(import json)`
- Remove stopwords (and, or)
- Clean up the text itself (\n, word counts)

Bonus: Make this script accept command line arguments to read in an arbitrary data file.

Expert: Use a generator to lazily clean the input instead of doing it all at once.

Section 2: Exploration!

Goal: Summarize a biz from its reviews

- Write a function to find the five most interesting, common words for each business
- How well can you guess categories for a business just from the reviews? What process are you using?

Bonus: Write code to automatically identify Mexican restaurants.

Expert: Write a method to quantify how well your Mexican food classifier is doing.

Section 3: Automatic topic modeling!

- Classify some businesses with my `naive_mexican_classifier`. When does it work well? Can you come up with example data it does poorly on?
- Write your own improved Mexican classifier!
- How can we compare multiple classifiers? Can you quantify how well your handbuilt mexican classifier does against my naive one?

Bonus: Try using the Gensim library, to learn topics automatically!

Expert: Write a method to find similar business using Gensim similarity queries.

If you’re so inclined, you can see how I generated our sample dataset in the `data_generation` folder.

The whole Yelp dataset is very broad! We’ ve only looked at a tiny sliver of the available businesses in this tutorial.

Yelp Dataset Challenge

Academic dataset from Phoenix, Las Vegas, Madison, Waterloo and Edinburgh!

- 1,125,458 Reviews
- 42,153 Businesses
  - 320,002 Business attributes
- 403,210 Tips
- 252,898 Users
  - 955,999 Edge social graph
- 31,617 Checkin Sets

+

Your academic project, research and/or visualizations

submitted by **December 31, 2014**

=

$5,000 prize + $1,000 for publication + $500 for presenting*

yelp.com/dataset_challenge

*See full terms on website

Questions?

Email: striglia@yelp.com
Twitter: @scott_triglia