

PIZZA, PIES, PROFIT... SOUP, SALAD, SUCCESS...

How can restaurants improve their
consumer ratings on Yelp?

Eric Amaso
Ranabir Bhakat
Helen Tam

Today's Roadmap



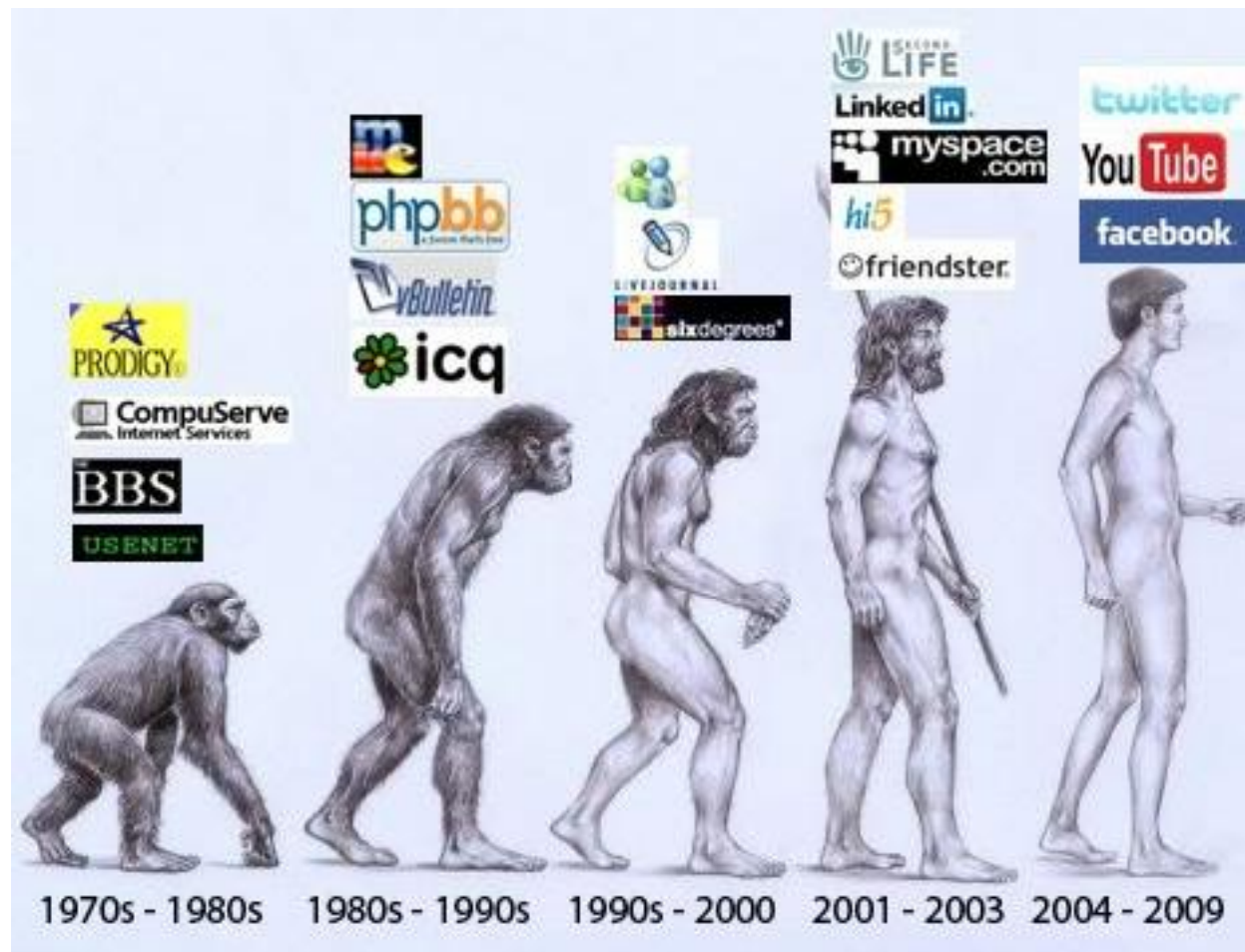
✓ Introduction

.. Methodology

.. Model Construction

.. Myth Busters & Conclusion

Social media has evolved over time



...and is becoming increasingly important in the way we communicate with each other



So why did we choose Yelp? And why should restaurants care about their Yelp ratings?

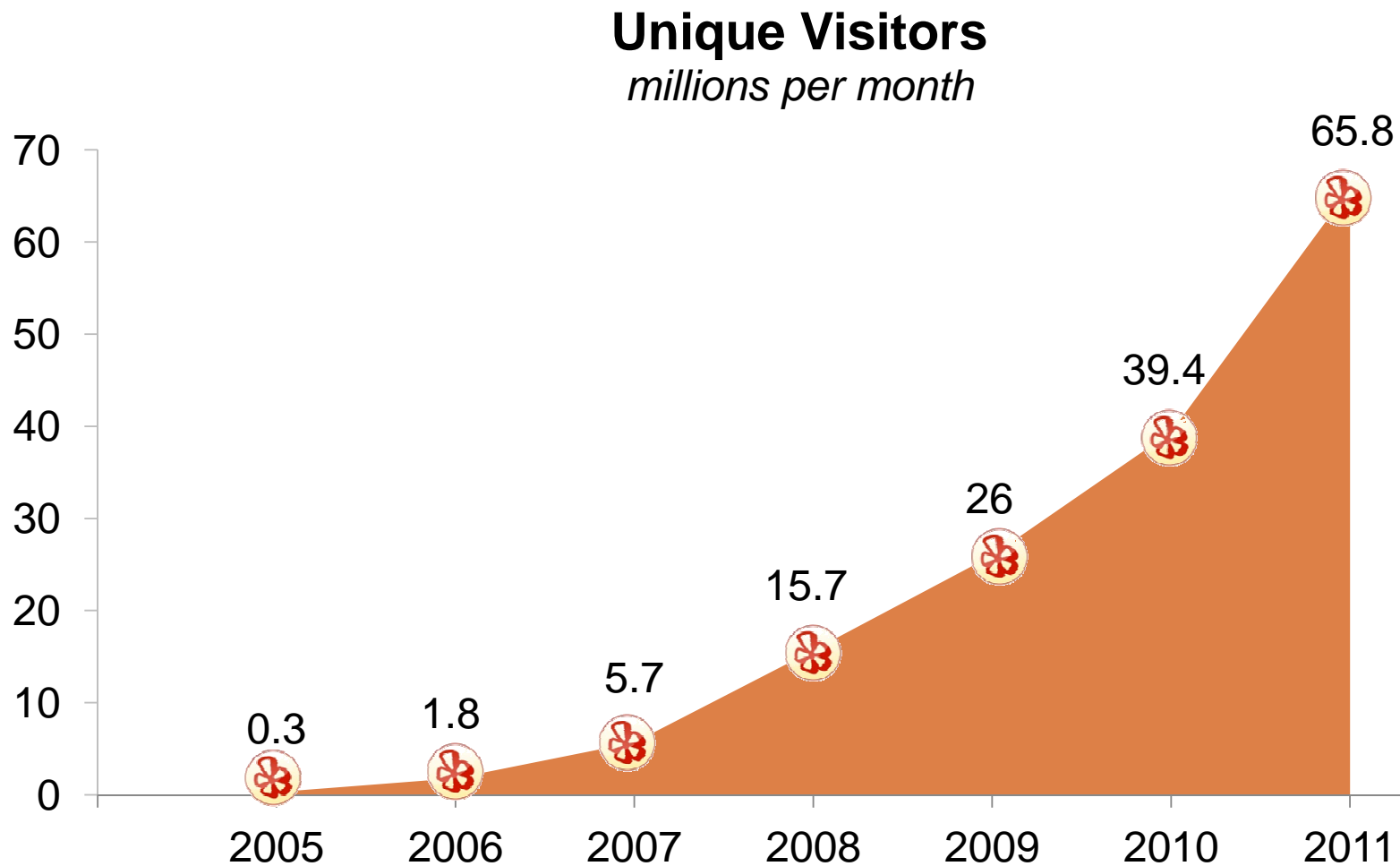
The site connects people with businesses

Nearly
25 million
reviews had
been posted on
Yelp by the end
of 2011



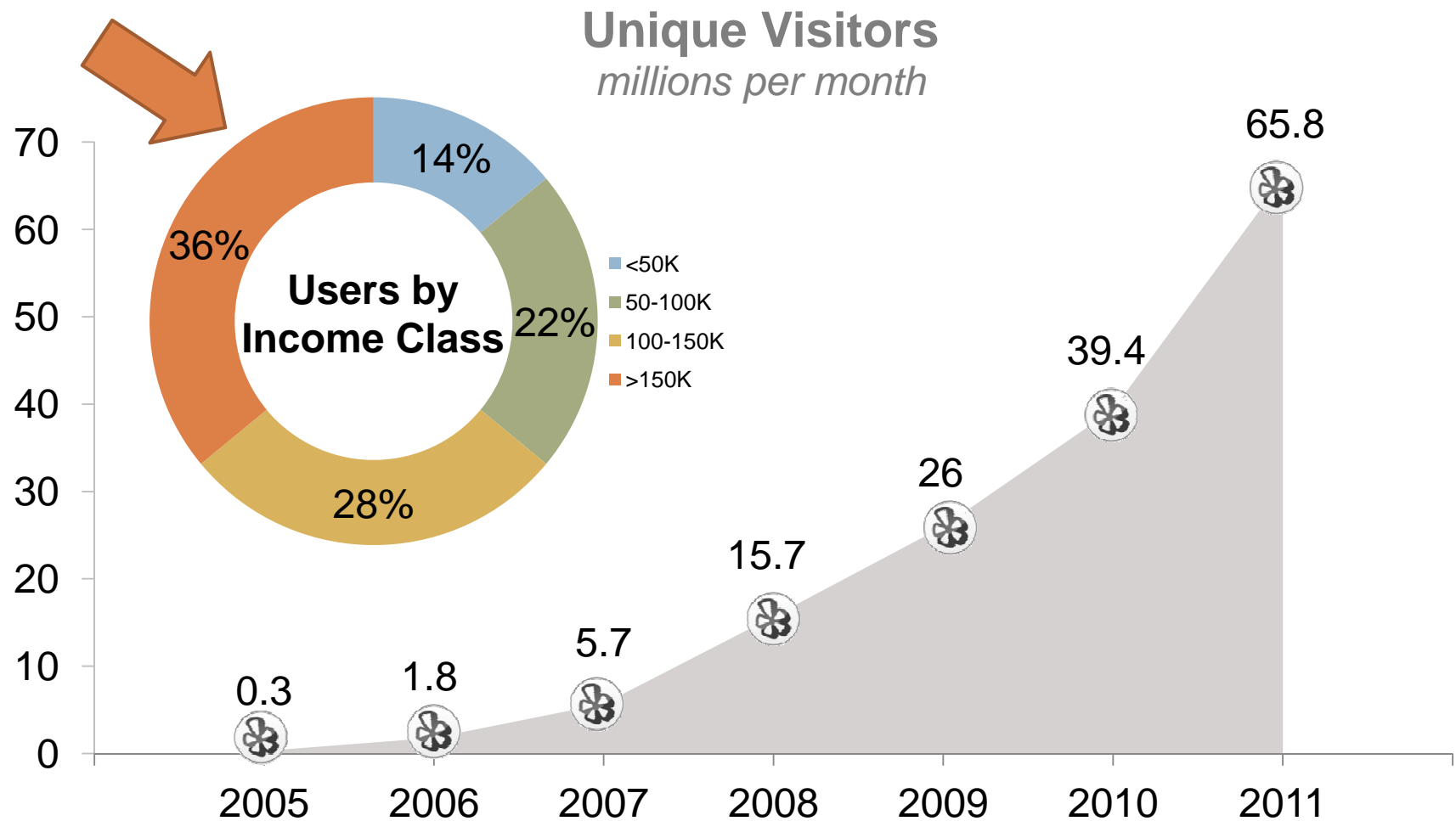
More than
600,000
local businesses
had official
claimed their
listing on the
website

Yelp's user base has more than **doubled** since 2009...



Source: Yelp, Quantcast, Statista

...and is skewed towards **high-income** classes



Source: Yelp, Quantcast, Statista

Today's Roadmap



- .. Introduction

-  Methodology

- .. Model Construction

- .. Myth Busters & Conclusion

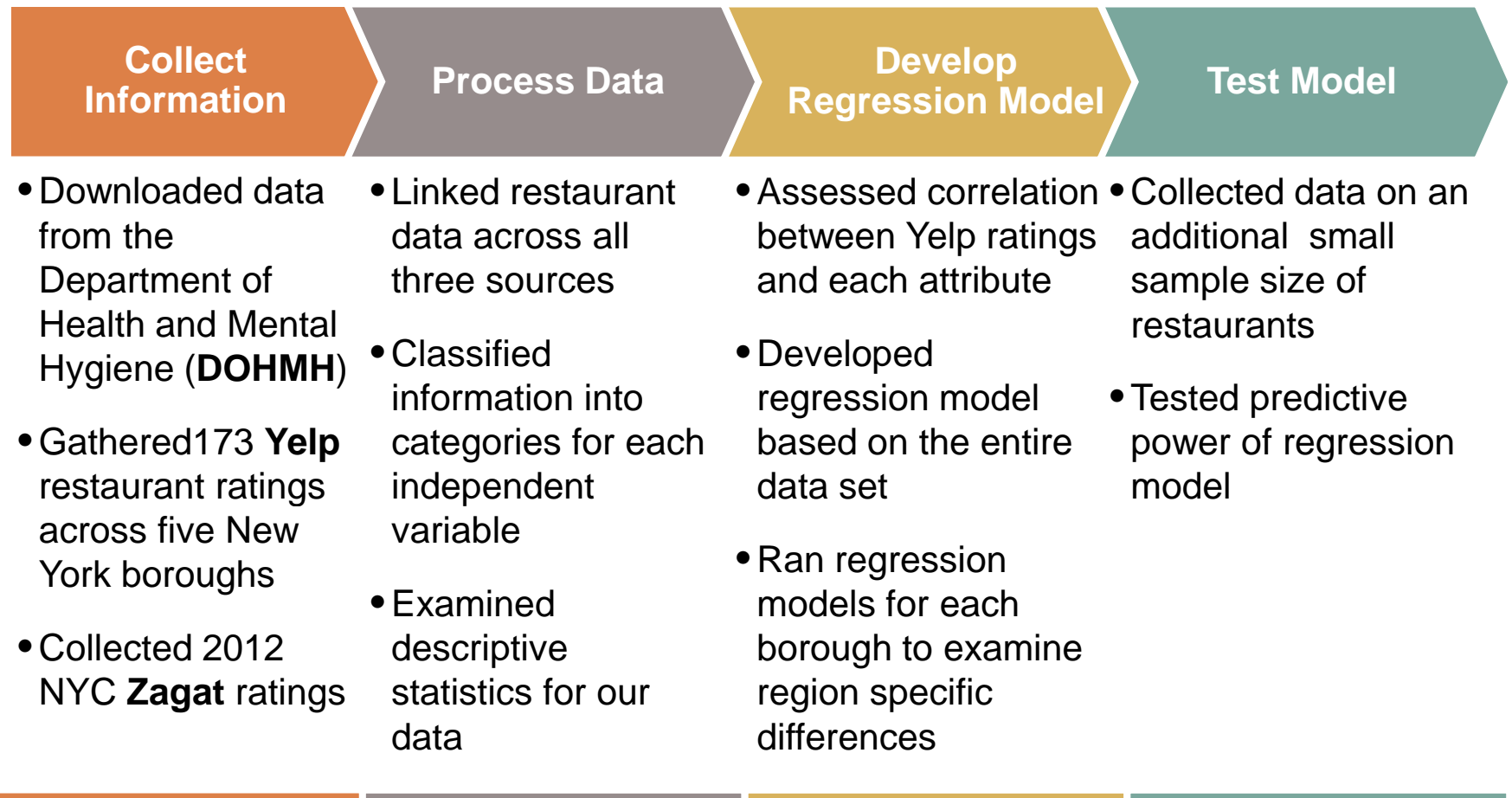
Out of 28 variables, which ones best predict Yelp ratings?

A word cloud of 28 variables used to predict Yelp ratings. The words are arranged in a circular pattern around the word "Good". The variables include:

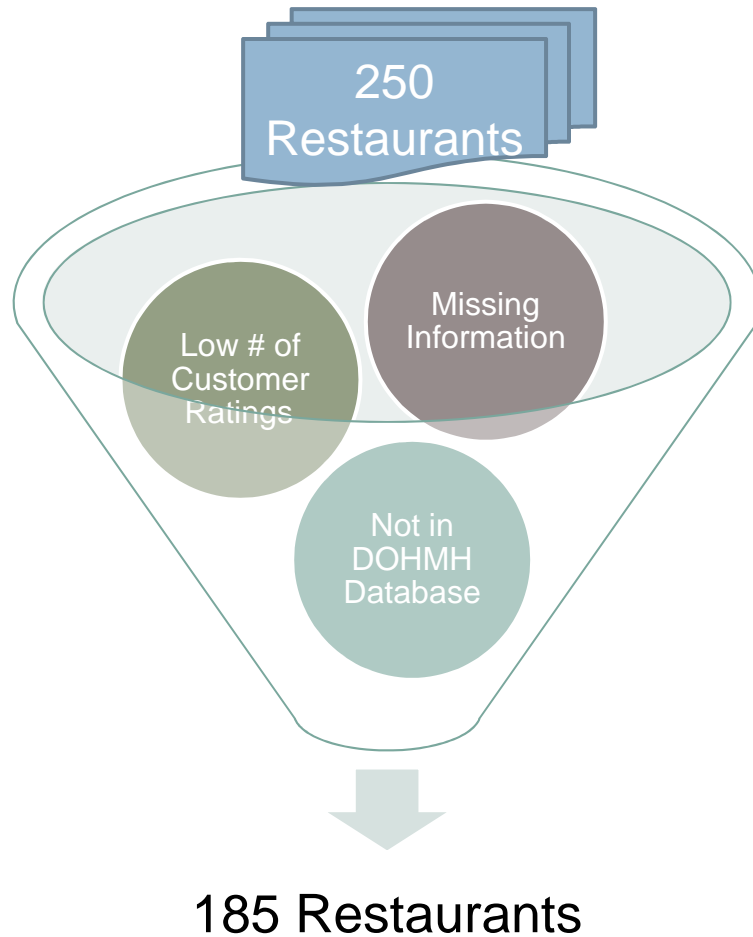
- Reservations
- Transit
- Grade
- Rating
- Score
- Reviews
- Stations
- Inspection
- Card
- Caters
- Cuisine
- Range
- Kids
- Dinner
- Credit
- Nearby
- TV
- Parking
- Waiter
- Lunch
- Alcohol
- Outdoor
- Take
- Borough
- Sanitation
- Ambience
- Delivery
- Breakfast
- Level
- Service
- Attire
- Groups
- Accessible
- Noise
- Seating
- Mass
- Wheelchair
- Price
- Wi-Fi



We used the following approach to develop and test our regression model



We encountered some challenges collecting the data



- .. Many restaurants did not have all their information posted in Yelp
- .. Some restaurants had fewer than 10 customer ratings
- .. Others could not be identified in the DOHMH database

Today's Roadmap



- .. Introduction

- .. Methodology

- ✓ Model Construction

- .. Myth Busters & Conclusion

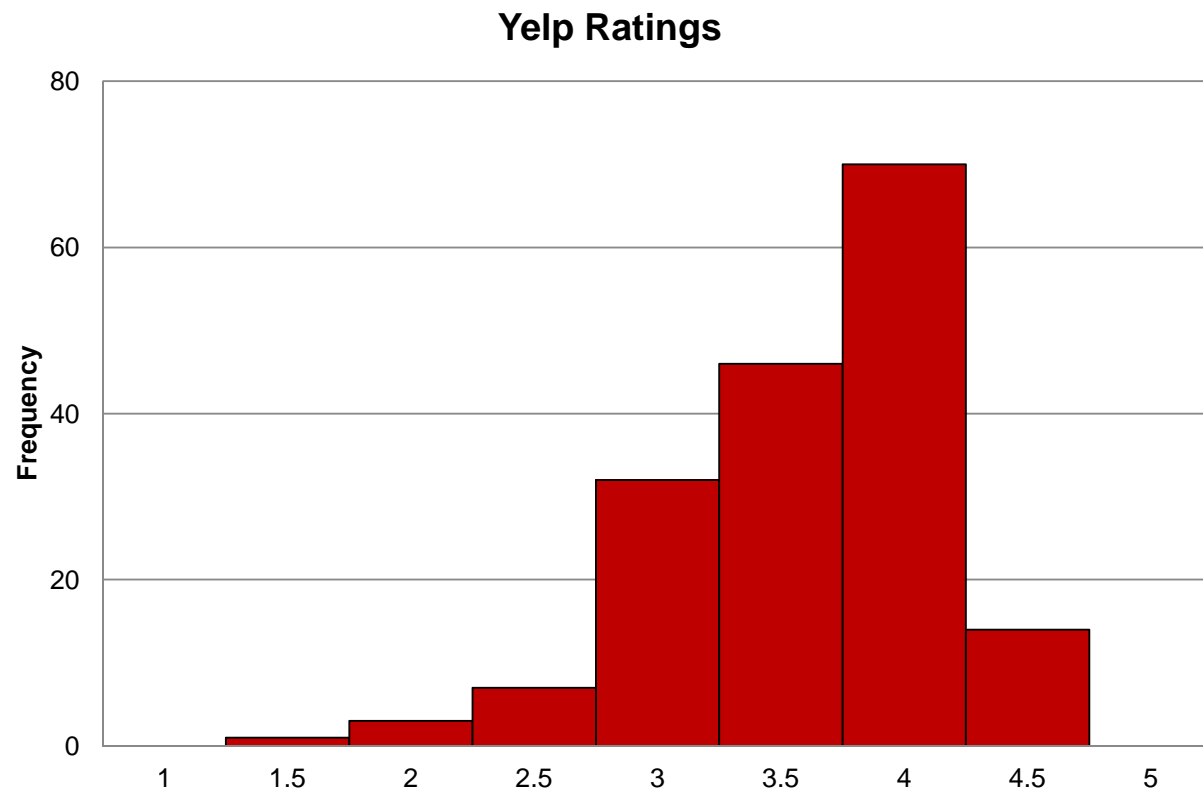
Our data is biased towards higher Yelp ratings

Descriptive
Statistics

Correlation
Analysis

Model
Refinement

Model Testing



We collected restaurant data across a diverse number of dimensions

Descriptive
Statistics

Correlation
Analysis

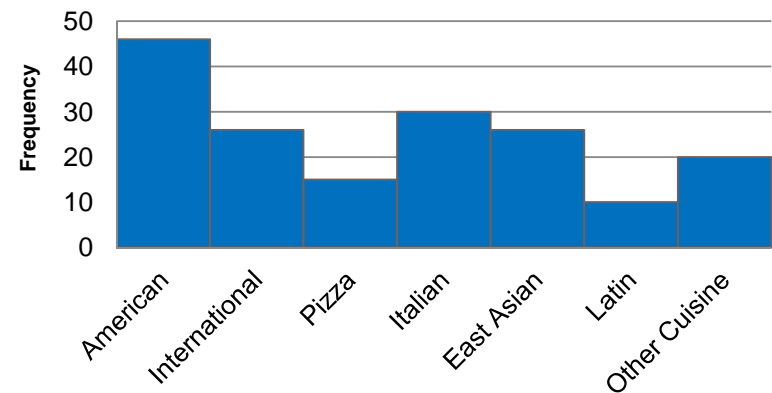
Model
Refinement

Model Testing

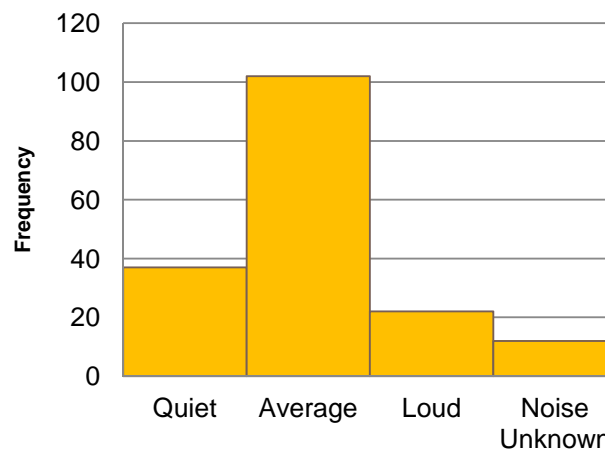
Price Range



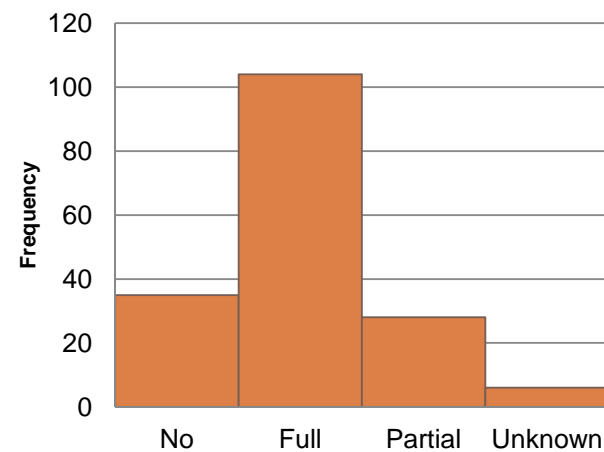
Cuisine Type



Noise Level



Alcohol



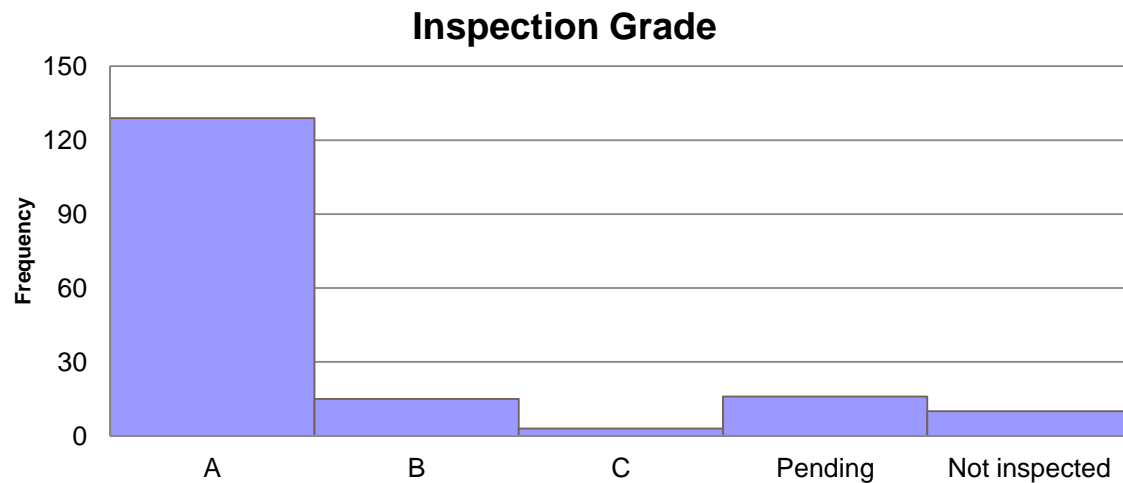
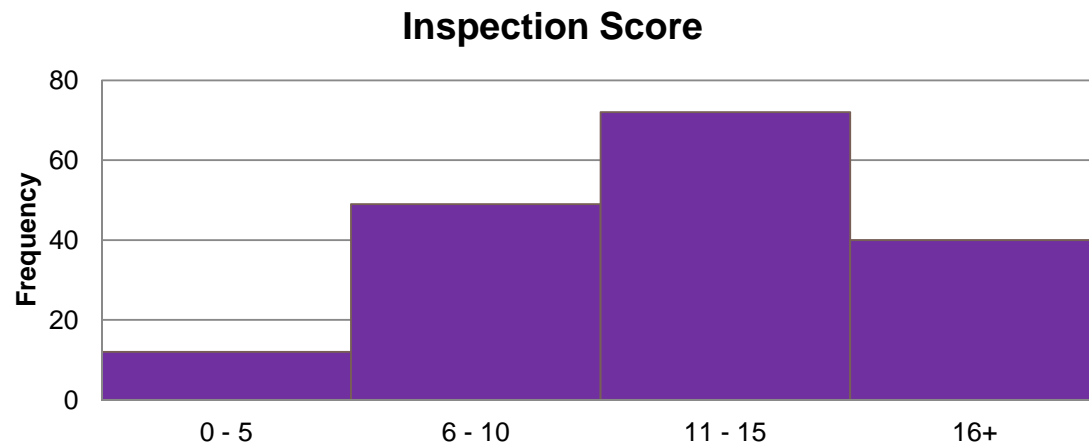
We were also able to integrate data from DOHMH into our sample

Descriptive
Statistics

Correlation
Analysis

Model
Refinement

Model Testing



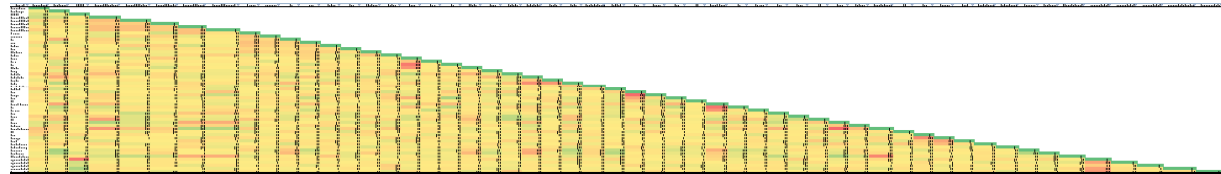
Next, we identified which factors were most highly correlated with Yelp ratings

Descriptive Statistics

Correlation Analysis

Model Refinement

Model Testing



	Yelp Rating
Yelp Rating	1.000
Outdoor Seating	-0.268
Borough 4 (Queens)	0.246
American	-0.215
East Asian	0.171
WiFi	-0.165
Price Range	-0.164
Pizza	0.144
Partial Bar	0.122
Credit Card	-0.121
Dinner Unknown	0.121
Borough 1 (Manhattan)	-0.117
Other Cuisine	-0.117
Formal	0.117
Garage	-0.115
Wheelchair Accessible	-0.111
Review Count	0.111
Casual	-0.107
Inspection Grade A	0.104

	Yelp Rating
Loud	-0.104
No Inspection Grade	-0.099
Inspection Grade B	-0.094
Romantic	0.089
Quiet	0.086
Other Attire	0.086
Noise Unknown	0.084
Dinner	-0.083
Full Bar	-0.079
Take Out	-0.076
No Dinner	-0.075
Kids	0.074
Street	0.071
SCORE	-0.069
Waiter Service	-0.066
Inspection Grade C	-0.064
Borough 3 (Brooklyn)	-0.060
Valet	0.059

	Yelp Rating
Inspection Grade Pending	0.059
International	0.058
Borough 5 (Staten Island)	-0.057
Mass Transit	0.055
Latin	0.052
TV	-0.050
Casual Ambience	-0.048
Delivery	-0.044
Average	-0.044
Group	-0.031
Dressy	0.029
Borough 2 (The Bronx)	-0.022
No Alcohol	-0.012
Caters	0.010
Classy	-0.010
Alcohol Unknown	-0.009
Italian	-0.007
Reservations	-0.002

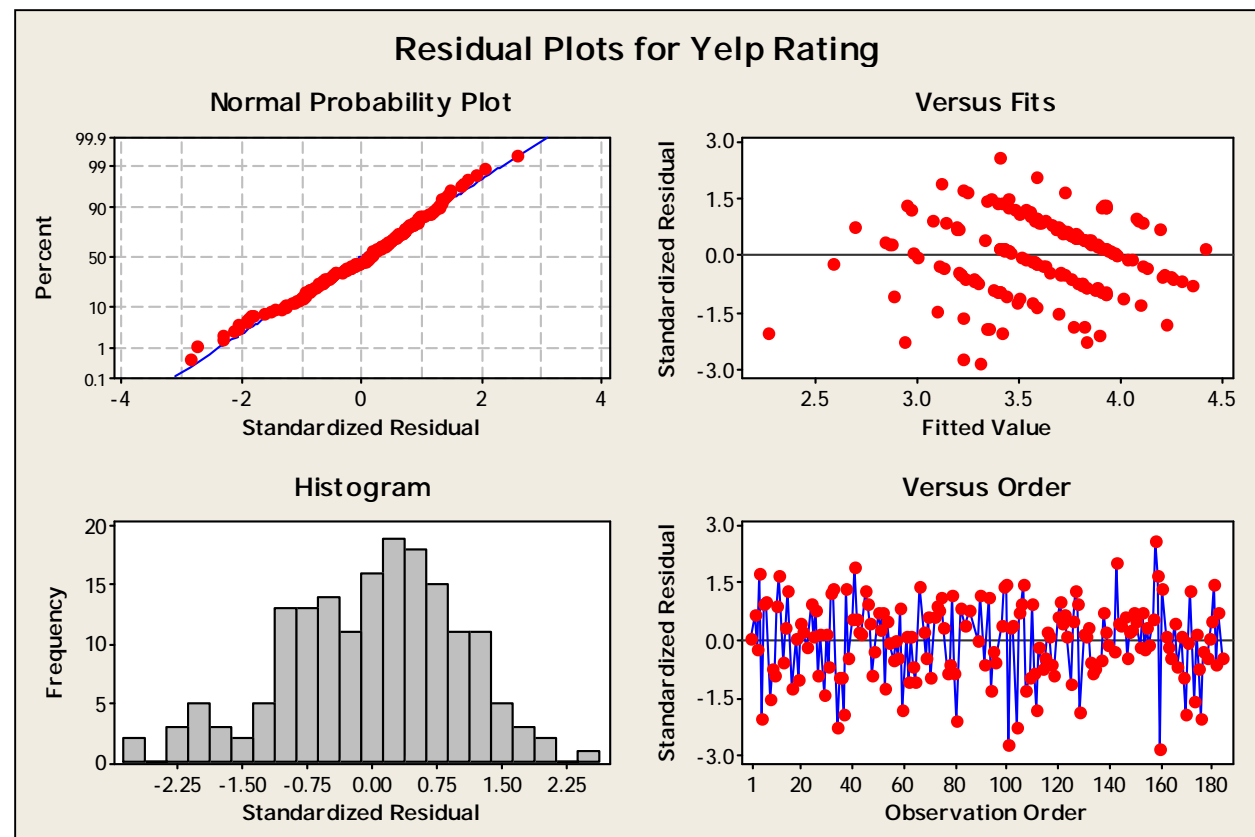
Residuals were normally distributed across our sample

Descriptive
Statistics

Correlation
Analysis

Model
Refinement

Model Testing



The full model was statistically significant...but had a low adjusted R-square

Descriptive
Statistics

Correlation
Analysis

Model
Refinement

Model Testing

Source	DF	SS	MS	F	P
Regression	45	23.424	0.521	2.090	0.001
Residual Error	127	31.628	0.249		
Total	172	55.052			



	Full
R-Sq	42.50%
R-Sq(adj)	22.20%

We didn't find any region specific differences,
most likely because our sample size was too small

Descriptive
Statistics

Correlation
Analysis

Model
Refinement

Model Testing

Source	DF	SS	MS	F	P
Regression	16	6.0063	0.3754	1.50	0.181
Residual Error	23	5.7375	0.2495		
Total	39	11.7438			



	Full	Manhattan
R-Sq	42.50%	51.1%
R-Sq(adj)	22.20%	17.2%

However, we were able to improve the adjusted R-sq and simplify our model

Descriptive
Statistics

Correlation
Analysis

Model
Refinement

Model Testing

Source	DF	SS	MS	F	P
Regression	6	2.2108	0.3685	3.3965	0.0093
Residual Error	36	3.9055	0.1085		
Total	42	6.1163			



	Full	Manhattan	Zagat
R-Sq	42.50%	51.1%	36.15%
R-Sq(adj)	22.20%	17.2%	25.50%

Our model accurately predicted 25% of the Yelp ratings in an additional subset of data

Descriptive Statistics

<i>Intercept</i>	<i>Zagat Food Rating</i>	<i>Zagat Decor</i>	<i>Zagat Service</i>	<i>Zagat Cost</i>	<i>Inspection SCORE</i>	<i>Inspection Grade A</i>
1.57332	0.10486	-0.02346	0.01859	-0.00626	0.00205	0.10216

Correlation Analysis

Model Refinement

Model Testing


DBA	Zagat Food Rating	Zagat Decor	Zagat Service	Zagat Cost	SCORE	A	Yelp Rating	Predicted Yelp Rating
AFGHAN KEBAB HOUSE	20	11	18	39	7	1	3.5	3.6
ANGELO OF MULBERRY ST.	23	16	20	46	13	1	3.5	3.8
ARTURO'S	21	14	17	27	12	1	4	3.7
BASTA PASTA RESTAURANT	23	17	21	45	23	0	4	3.7
DOMINICK'S BAR & RESTAURANT	24	11	19	39	6	1	4	4.1
EL MALECON RESTAURANT	20	9	15	22	9	1	3.5	3.7
BAMONTE'S RESTAURANT	24	17	22	46	13	1	3	3.9
FERDINANDO'S RESTAURANT	25	13	18	26	2	1	4	4.2
BEN-BEST DELI & RESTAURANT	23	9	17	24	10	1	3.5	4.1
ARIRANG HIBACHI STEAK HOUSE	20	19	22	38	23	0	3.5	3.4
CAROL'S CAFE	24	19	22	57	10	1	4	3.8
DENINO'S PIZZERIA TAVERN	26	11	19	22	8	1	4	4.4

So what did we find?

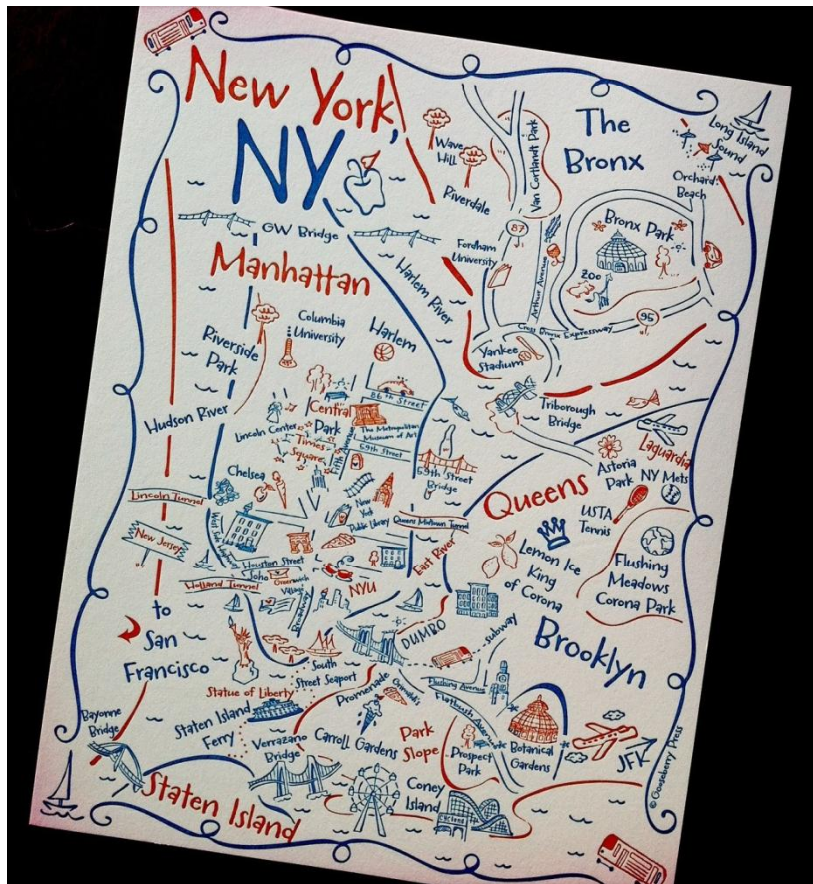
Yelp Rating (Full Model) = 5.33 + 0.000413 Review Count - 0.233 Price Range - 0.0110 SCORE - 0.572 Boro 1 - 0.267 Boro 2 - 0.474 Boro 3 + 0.144 Boro 4 - 0.066 American + 0.059 International + 0.089 Pizza + 0.092 Italian + 0.301 East Asian + 0.229 Latin + 0.126 Mass Transit - 0.122 Casual + 0.623 Formal - 0.152 Dressy + 0.004 Group - 0.202 Take Out - 0.122 No + 0.137 Partial + 0.046 Full - 0.064 Credit Card - 0.133 Street - 0.245 Garage - 0.128 WiFi - 0.497 Casual Ambience - 0.280 Classy + 0.070 Caters - 0.112 Delivery + 0.047 Kids + 0.212 Dinner + 0.303 No Dinner + 0.056 TV - 0.228 Quiet - 0.364 Average - 0.347 Loud - 0.342 Outdoor Seating + 0.138 Reservations - 0.099 Waiter Service + 0.078 Wheelchair Accessible + 0.004 Grade A + 0.040 Grade B + 0.080 Grade C + 0.412 Grade Pending

Yelp Rating (Zagat Model) = 1.57 + 0.00205 SCORE + 0.105 Zagat Food Rating - 0.0235 Zagat Decor + 0.0186 Zagat Service - 0.00626 Zagat Cost + 0.102 A

Today's Roadmap

- 
- .. Introduction
 - .. Methodology
 - .. Model Construction
 - ✓ Myth Busters & Conclusion

Myth #1: It's all about Location, Location, Location



.. Manhattan schmattan...it won't give you a leg up in Yelp ratings

.. Adj R-Sq: 5.4%

	Coefficients	P-value
Intercept	3.6504	0.0000
Borough 1 (Manhattan)	-0.1629	0.1106

Myth#2: You have to be classy



Myth#2: You have to be classy



- .. Nope. Not at all! You can save the candles for later.
- .. Adj R-Sq: -0.5%

	<i>Coefficients</i>	<i>P-value</i>
Intercept	3.6171	0.0000
Classy	-0.0504	0.7425

Myth #3: You need to be well stocked for HaPpY HoUR



- Or not...
Unless it involves beer pong + Juran
à *p-value 0.0000*
- Adj R-Sq: 0.6%

	Coefficients	P-value
Intercept	3.6884	0.0000
Full Bar	-0.1259	0.1523

Myth #4: The kitchen's gotta be clean

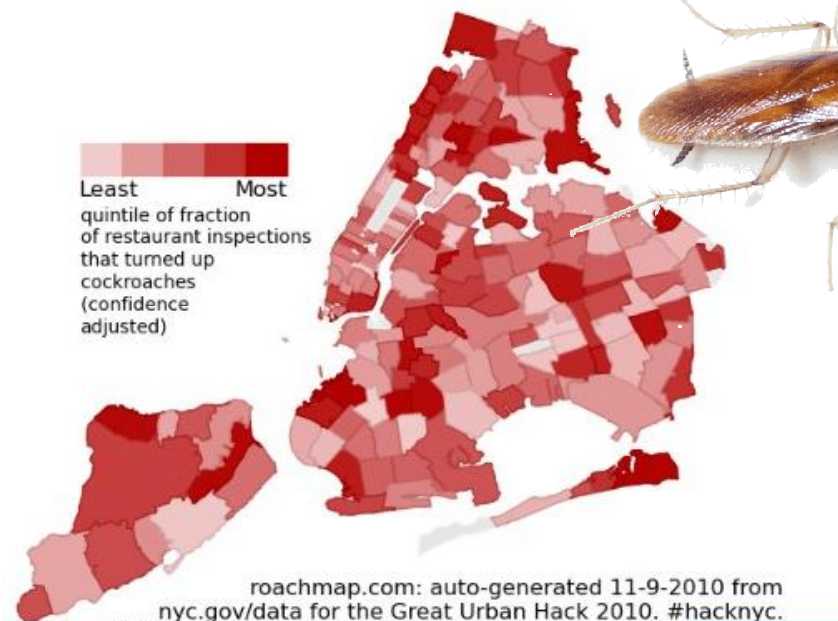
Online 'Roach Map' Lets Users Track City's Pests

Updated November 10, 2010 1:52pm

November 10, 2010 1:50pm | By Olivia Scheck, DNAinfo Reporter/Producer

0 share print RECOMMEND

Where are the roaches in the past four weeks?



...but cockroaches provide extra protein don't they?

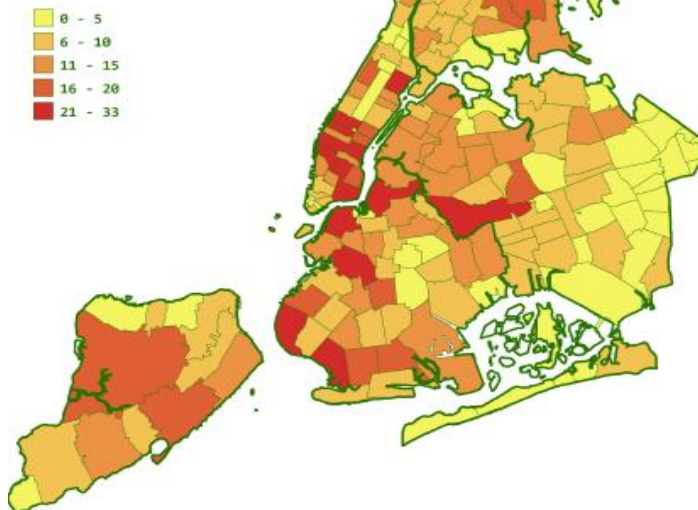
Adj R-Sq: -0.2%

	Coefficients	P-value
Intercept	3.6729	0.0000
Inspection SCORE	-0.0044	0.3993

Myth #5: Forget the pizza, go for the sushi!



Pizza Places in New York
by Zip Code



.. New Yorkers are into sushi, but they still love their pizza

.. Adj R-Sq: 4.2%
Model p-value: 0.0020

	<i>Coefficients</i>	<i>P-value</i>
Intercept	3.3750	0.0000
American	0.0163	0.9106
International	0.3173	0.0504
Pizza	0.4917	0.0086
Italian	0.2750	0.0803
East Asian	0.4904	0.0027
Latin	0.3750	0.0755

Myth #6: It's NY – price at a premium



- .. No!
- .. Adj R-Sq: 2.0%

	<i>Coefficients</i>	<i>P-value</i>
Intercept	3.8508	0.0000
Price Range	-0.1204	0.0367

Myth #7: Make sure your food tastes good



- Well, obviously.
- Adj R-Sq: 19.5%

	<i>Coefficients</i>	<i>P-value</i>
Intercept	2.2171	0.0000
Zagat Food Rating	0.0680	0.0018

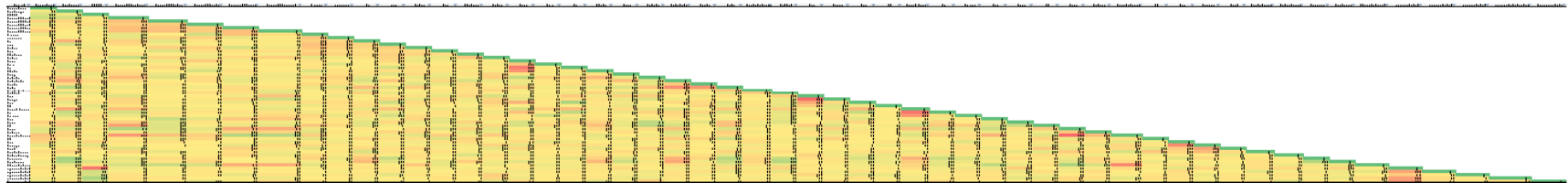
So what really matters when it comes to getting good Yelp ratings?



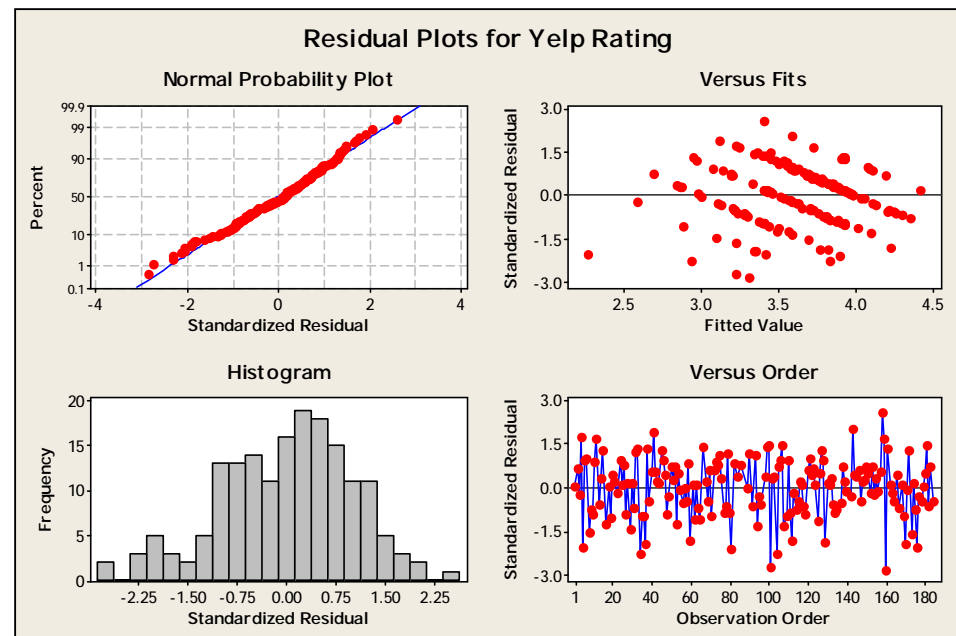


Appendix

Full Model Regression on Yelp Data



Yelp Rating	
Yelp Rating	1.000
Outdoor Seating	-0.268
Borough 4 (Queens)	0.246
American	-0.215
East Asian	0.171
WiFi	-0.165
Price Range	-0.164
Pizza	0.144
Partial Bar	0.122
Credit Card	-0.121
Dinner Unknown	0.121
Borough 1 (Manhattan)	-0.117



Full Regression Model

Yelp Rating = 5.33 + 0.000413 Review Count - 0.233 Price Range - 0.0110 SCORE - 0.572 Boro 1 - 0.267 Boro 2 - 0.474 Boro 3 + 0.144 Boro 4 - 0.066 American + 0.059 International + 0.089 Pizza + 0.092 Italian + 0.301 East Asian + 0.229 Latin + 0.126 Mass Transit - 0.122 Casual + 0.623 Formal - 0.152 Dressy + 0.004 Group - 0.202 Take Out - 0.122 No + 0.137 Partial + 0.046 Full - 0.064 Credit Card - 0.133 Street - 0.245 Garage - 0.128 WiFi - 0.497 Casual Ambience - 0.280 Classy + 0.070 Caters - 0.112 Delivery + 0.047 Kids + 0.212 Dinner + 0.303 No Dinner + 0.056 TV - 0.228 Quiet - 0.364 Average - 0.347 Loud - 0.342 Outdoor Seating + 0.138 Reservations - 0.099 Waiter Service + 0.078 Wheelchair Accessible + 0.004 Grade A + 0.040 Grade B + 0.080 Grade C + 0.412 Grade Pending

Source	DF	SS	MS	F	P
Regression	45	23.4239	0.5205	2.09	0.001
Residual Error	127	31.6282	0.249		
Total	172	55.052			

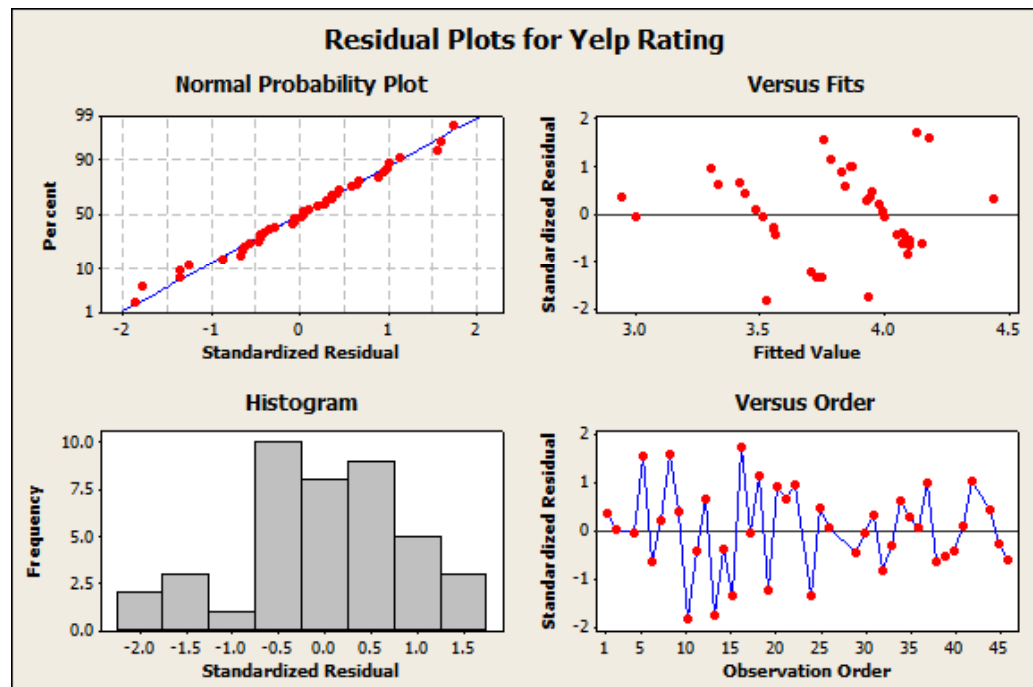
S = 0.49904 R-Sq = 42.50% R-Sq(adj) = 22.20%

P-value <5%

Zagat Model Regression on Yelp Data

	Yelp Rating	Inspection SCORE	Zagat Food Rating	Zagat Decor	Zagat Service	Zagat Cost	Inspection Grade A
Yelp Rating	1						
Inspection SCORE	-0.0517	1.0000					
Zagat Food Rating	0.4661	-0.1899	1.0000				
Zagat Decor	-0.1383	-0.2664	0.3561	1.0000			
Zagat Service	0.0878	-0.2874	0.6142	0.8045	1.0000		
Zagat Cost	0.0361	-0.2447	0.6460	0.6223	0.7672	1.0000	
Inspection Grade A	0.1004	-0.8027	0.1726	0.1230	0.2427	0.2763	1.0000

	Yelp Rating
Yelp Rating	1.000
Zagat Food Rating	0.466
Zagat Decor	-0.138
Inspection Grade A	0.100
Zagat Service	0.088
Inspection SCORE	-0.052
Zagat Cost	0.036



Zagat Regression Model

$$\begin{aligned} \text{Yelp Rating} = & 1.57 + 0.00205 \text{ SCORE} + 0.105 \text{ Zagat Food Rating} \\ & - 0.0235 \text{ Zagat Decor} + 0.0186 \text{ Zagat Service} - 0.00626 \text{ Zagat Cost} \\ & + 0.102 A \end{aligned}$$

Source	DF	SS	MS	F	P
Regression	6	2.2108	0.3685	3.4	0.009
Residual Error	36	3.9055	0.1085		
Total	42	6.1163			

$$S = 0.329371 \quad R\text{-Sq} = 36.1\% \quad R\text{-Sq}(\text{adj}) = 25.5\%$$

P-value <5%

P-value <10%