NYCDSA Web Scraping Project:

Analyzing Geographic and Demographic Influence in Yelp Reviews

Sammy Dolgin

10/14/19

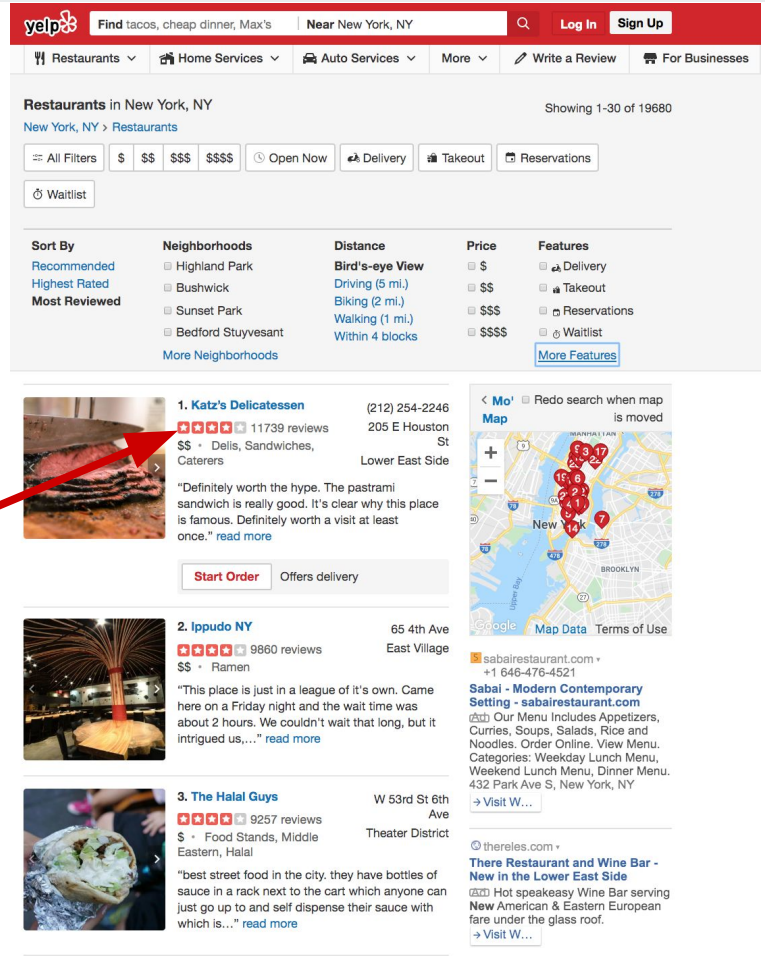
Background

- I moved to New York City on September 21st, 2019, after spending my whole life in the Midwest (Chicago, St. Louis).

- After hearing about how good the food is here, I was surprised that Yelp reviews for all the “top” restaurants seemed to be around a 3.5 or 4.

- I suspected that this was not a reflection of the food, but rather a tendency for New Yorkers to be more critical of their restaurants.

Background


yelp
Find tacos, cheap dinner, Max's
Near New York, NY
Log In
Sign Up
Restaurants
Home Services
Auto Services
More
Write a Review
For Businesses
Restaurants in New York, NY
Showing 1-30 of 19680
New York, NY > Restaurants
All Filters
$
$$
$$$
$$$$
Open Now
Delivery
Takeout
Reservations
Waitlist
Sort By
Recommended
Highest Rated
Most Reviewed
Neighborhoods
Highland Park
Bushwick
Sunset Park
Bedford Stuyvesant
More Neighborhoods
Distance
Bird's-eye View
Driving (5 mi.)
Biking (2 mi.)
Walking (1 mi.)
Within 4 blocks
Price
$
$$
$$$
$$$$
Features
Delivery
Takeout
Reservations
Waitlist
More Features
1. Katz's Delicatessen
11739 reviews
$$ • Delis, Sandwiches, Caterers
(212) 254-2246
205 E Houston St
Lower East Side
"Definitely worth the hype. The pastrami sandwich is really good. It's clear why this place is famous. Definitely worth a visit at least once." read more
Start Order
Offers delivery
2. Ippudo NY
9860 reviews
$$ • Ramen
65 4th Ave
East Village
"This place is just in a league of it's own. Came here on a Friday night and the wait time was about 2 hours. We couldn't wait that long, but it intrigued us,..." read more
3. The Halal Guys
9257 reviews
$ • Food Stands, Middle Eastern, Halal
W 53rd St 6th Ave
Theater District
"best street food in the city. they have bottles of sauce in a rack next to the cart which anyone can just go up to and self dispense their sauce with which is..." read more
Mo' Map
Redo search when map is moved
New York
BROOKLYN
Google
Map Data
Terms of Use
sabairestaurant.com
+1 646-476-4521
Sabai - Modern Contemporary Setting - sabairestaurant.com
Our Menu Includes Appetizers, Curries, Soups, Salads, Rice and Noodles. Order Online. View Menu. Categories: Weekday Lunch Menu, Weekend Lunch Menu, Dinner Menu.
432 Park Ave S, New York, NY
Visit W...
thereles.com
There Restaurant and Wine Bar - New in the Lower East Side
Hot speakeasy Wine Bar serving New American & Eastern European fare under the glass roof.
Visit W...

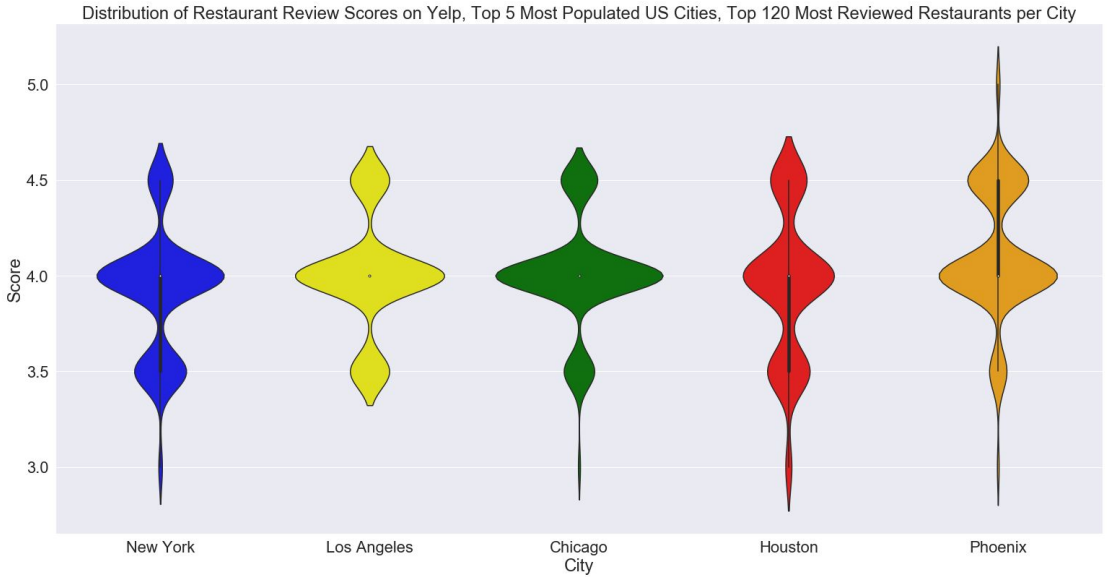
The Question:

Do geographic, demographic, and/or cultural factors influence the way cities review their restaurants?

Approach #1:

Comparing review ratings between the top 5 most populated U.S. cities

Data used: Reviews for the top 120 most reviewed restaurants within each of the top 5 most populated U.S. cities (New York City, Los Angeles, Chicago, Houston, Phoenix)



City	$\bar{x}$ of Scores	σ of Scores
New York	3.918103	0.322652
Houston	3.948276	0.380956
Los Angeles	4.000000	0.294884
Chicago	4.004310	0.283577
Phoenix	4.116379	0.332041

- New York having the lowest mean score is in line with my initial hypothesis
- Some differences in distribution but shapes are generally similar
- Distribution exhibits “normal” behavior but avoiding ANOVA due to ordinal data

City	4.5 Stars or Greater
New York	12.07%
Chicago	15.52%
Los Angeles	17.24%
Houston	20.69%
Phoenix	31.90%

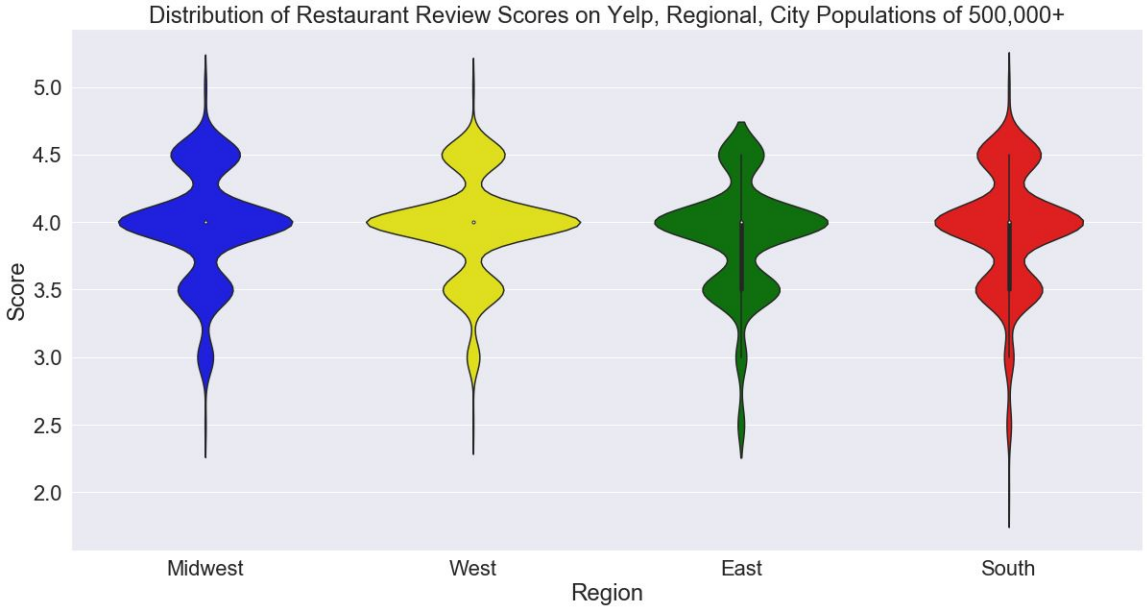
- Scores are very much centered around 4.0 (61.6% of all scores among these cities are a 4.0).
- Lower tendency for New Yorkers to give 4.5 stars or above

Approach #2:

Comparing review ratings between regional clusters of urban populations

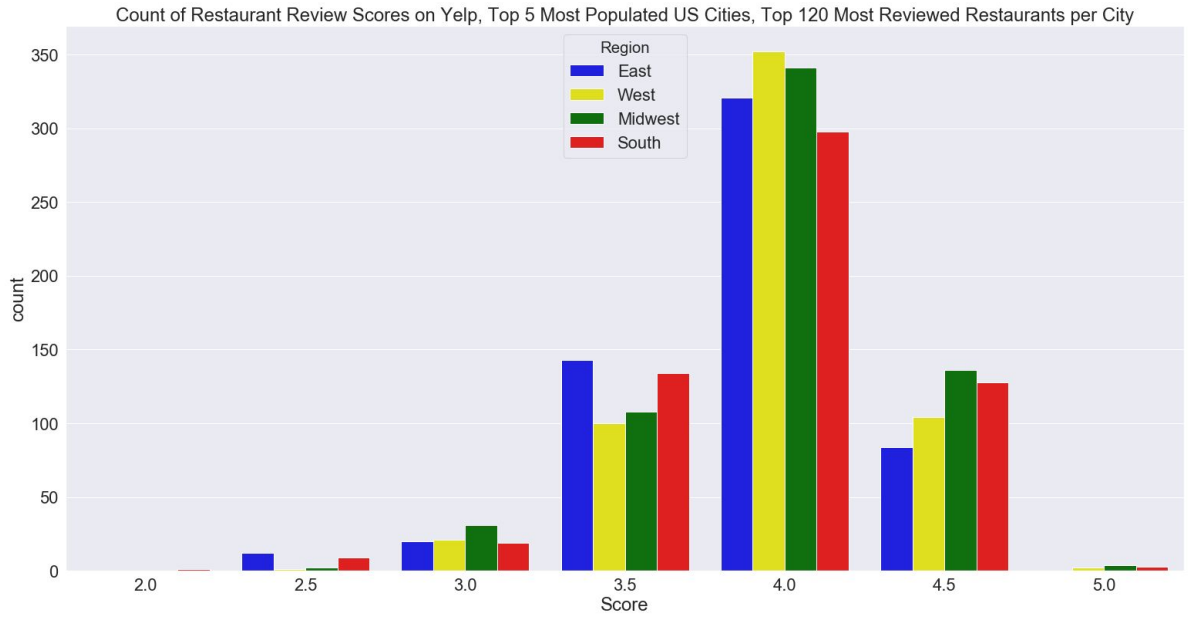
Data used: Reviews for the top 120 most reviewed restaurants within 25 highly populated U.S. cities (population of 500,000 or more), grouped into regions.

East	West	Midwest	South
New York, NY	Los Angeles, CA	Chicago, IL	Houston, TX
Philadelphia, PA	Phoenix, AZ	Columbus, OH	San Antonio, TX
Washington DC	San Diego, CA	Indianapolis, IN	Dallas, TX
Boston, MA	San Jose, CA	Detroit, MI	Austin, TX
Baltimore, MD	San Francisco, CA	Milwaukee, WI	Nashville, TN



Region	$\bar{x}$ of Scores	σ of Scores
East	3.883621	0.406965
South	3.941723	0.430911
West	3.968103	0.361488
Midwest	3.974277	0.401464

- Like before, cities in the East region exhibiting the lowest mean
- Wider spread of scores within the Southern region



Region	4.5 Stars or Greater
East	14.48%
West	18.28%
South	22.13%
Midwest	22.51%

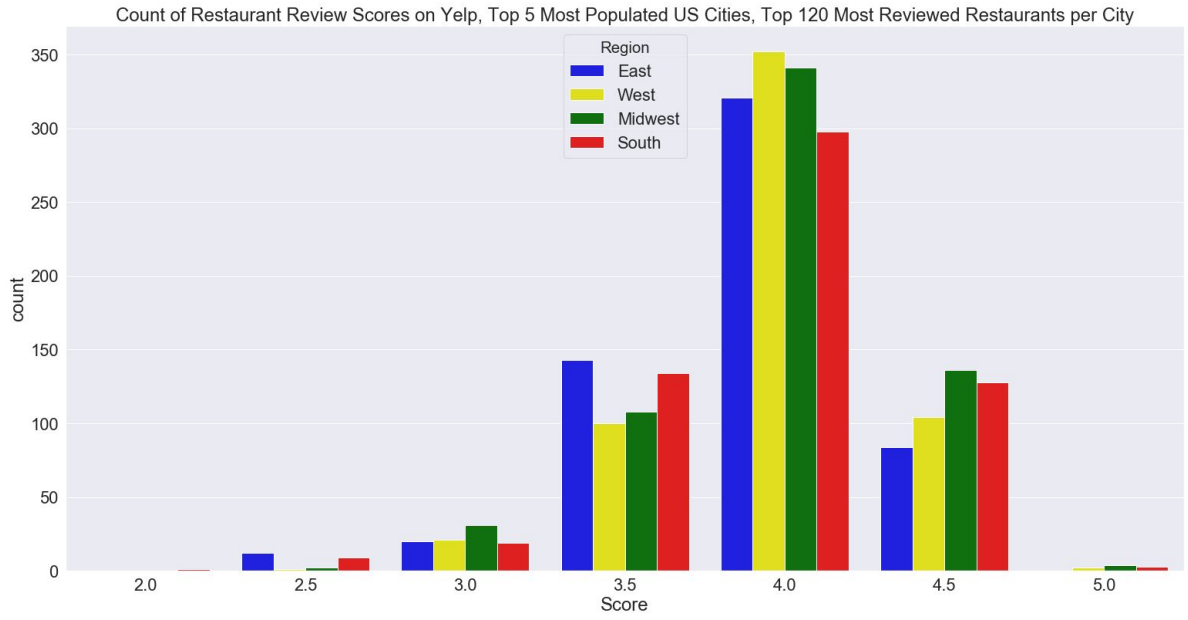
Two-Proportion Z-Test

P_1 = Proportion of East Region Reviews 4.5 stars or greater

P_2 = Proportion of Non-East Region Reviews 4.5 stars or greater

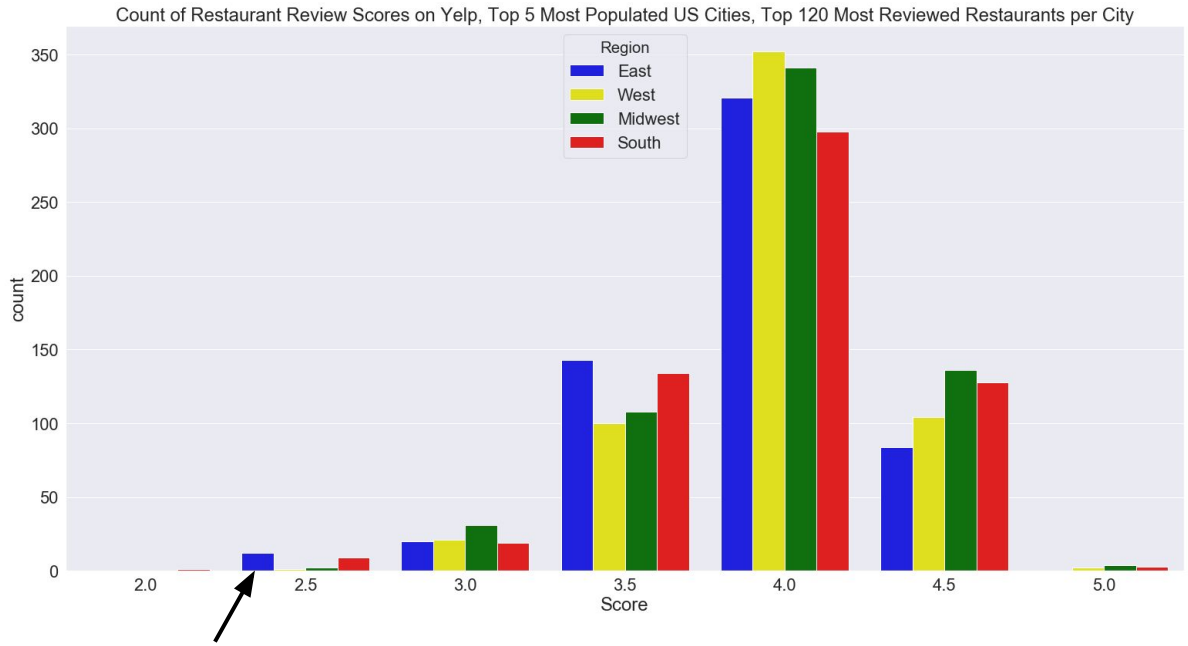
$H_0: P_1 - P_2 = 0$
$H_1: P_1 - P_2 < 0$

Z-Statistic = -3.22
p-value = 0.0013



City	4.5 Stars or Greater	Region
San Jose	6.03%	West
Boston	10.34%	East
San Francisco	11.21%	West
New York	12.07%	East
Washington	12.93%	East
Detroit	14.06%	Midwest
Chicago	15.52%	Midwest
Los Angeles	17.24%	West
Baltimore	17.24%	East
Dallas	19.83%	South
Philadelphia	19.83%	East
San Antonio	19.83%	South
Austin	19.83%	South
Houston	20.69%	South
San Diego	25.00%	West
Indianapolis	25.20%	Midwest
Milwaukee	27.34%	Midwest
Nashville	29.69%	South
Columbus	30.08%	Midwest
Phoenix	31.90%	West

- 3 of the bottom 5 proportions from East Region cities
- No East Region cities with over 20% of reviews as 4.5 stars or above



Region	2.5 Stars or Less
West	0.17%
Midwest	0.32%
South	1.69%
East	2.07%

City	2.5 Stars or Less
Austin	1
Boston	1
Dallas	1
Indianapolis	1
Milwaukee	1
San Jose	1
Philadelphia	2
San Antonio	8
Baltimore	9

- Also wanted to investigate the 2.5 star or less ratings
- East led the pack, but mostly skewed by Baltimore
- Not making any conclusions on this end

Approach #3:

Comparing review ratings between mid-sized and large urban areas

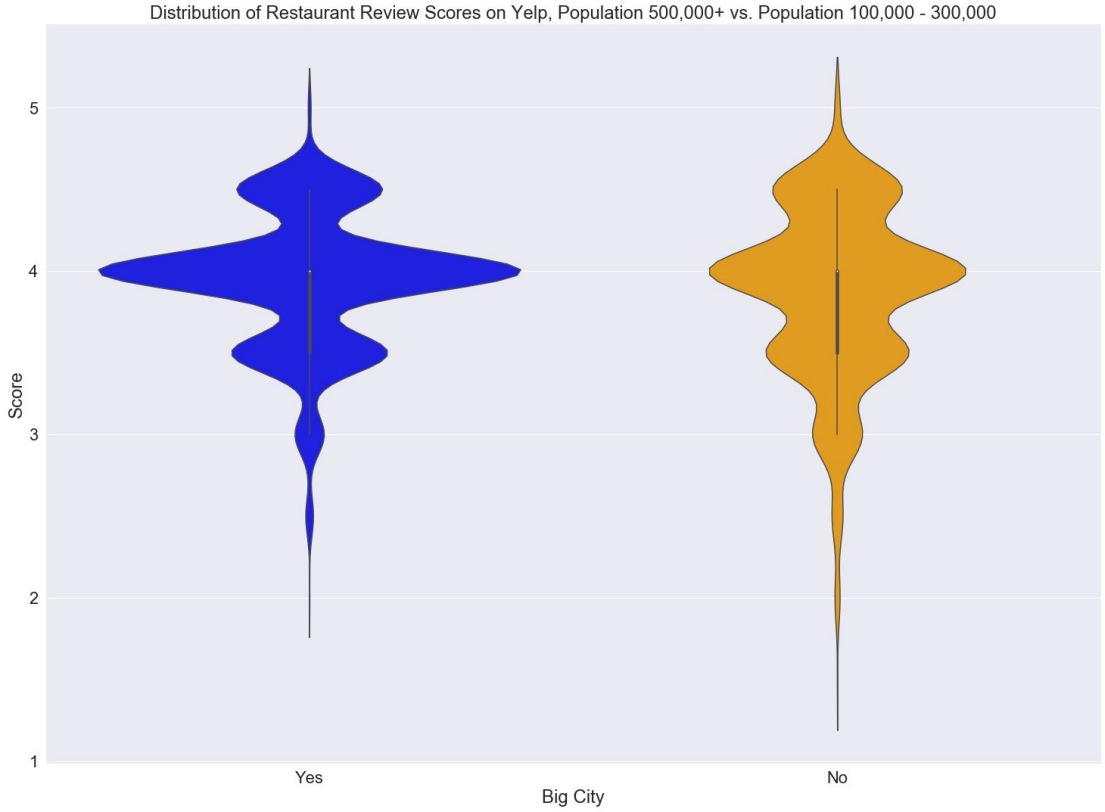
Data used: Reviews for the top 120 most reviewed restaurants within 25 highly populated U.S. cities (population of 500,000 or more) and 25 mid-sized U.S. cities (population 100,000 - 300,000), grouped into size classifications and regions.

Population 500,000+

East	West	Midwest	South
New York, NY	Los Angeles, CA	Chicago, IL	Houston, TX
Philadelphia, PA	Phoenix, AZ	Columbus, OH	San Antonio, TX
Washington DC	San Diego, CA	Indianapolis, IN	Dallas, TX
Boston, MA	San Jose, CA	Detroit, MI	Austin, TX
Baltimore, MD	San Francisco, CA	Milwaukee, WI	Nashville, TN

Population 100,000 - 300,000

East	West	Midwest	South
Newark, NJ	Boise, ID	Lincoln, NE	Baton Rouge, LA
Buffalo, NY	Tacoma, WA	Des Moines, IA	Birmingham, AL
Worcester, MA	San Bernardino, CA	Grand Rapids, MI	Little Rock, AR
Providence, RI	Salt Lake City, UT	Aurora, IL	Tallahassee, FL
Rochester, NY	Salem, OR	Akron, OH	Richmond, VA



Big City	$\bar{x}$ of Scores	σ of Scores
No	3.868866	0.516235
Yes	3.942502	0.402479

- Much higher concentration of 4.0 scores among big cities
- Much more variability among mid-sized cities

Big City	$\bar{x}$ of Scores	σ of Scores
No	3.868866	0.516235
Yes	3.942502	0.402479

Mid-Sized City

City	$\bar{x}$ of Scores
Newark	3.637931
Aurora	3.671875
Lincoln	3.696850
Tacoma	3.797414
Worcester	3.804688
San Bernardino	3.814050
Salem	3.821429
Buffalo	3.848361
Baton Rouge	3.864754
Tallahassee	3.889764
Little Rock	3.914062
Akron	3.916667
Rochester	3.917969
Grand Rapids	3.929688
Richmond	3.940171
Providence	3.951220
Boise	3.953125
Des Moines	3.972656
Birmingham	3.976562
Salt Lake City	4.046218

Big City

City	$\bar{x}$ of Scores
San Jose	3.724138
Baltimore	3.780172
San Antonio	3.840517
Detroit	3.847656
Boston	3.875000
Washington	3.879310
New York	3.918103
Austin	3.922414
San Francisco	3.931034
Milwaukee	3.945312
Houston	3.948276
Philadelphia	3.965517
Dallas	3.982759
Los Angeles	4.000000
Chicago	4.004310
Nashville	4.007812
Columbus	4.036585
Indianapolis	4.043307
San Diego	4.068966
Phoenix	4.116379

Big City	$\bar{x}$ of Scores	σ of Scores
No	3.868866	0.516235
Yes	3.942502	0.402479

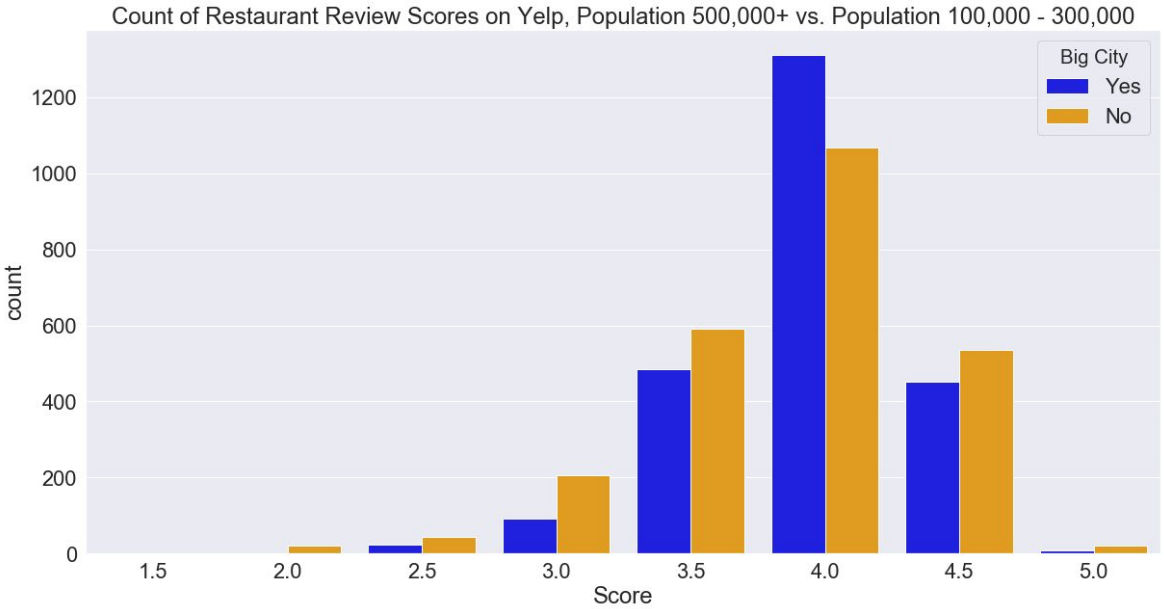
- The lowest standard deviation of scores among the mid-range cities is still higher than the standard deviation of 12 highly populated cities
- Clear increase in score variance with smaller populations (in line with standard statistical theory)

Mid-Sized City

City	σ of Scores
Salt Lake City	0.390521
Richmond	0.405577
Des Moines	0.417664
Grand Rapids	0.433510
Rochester	0.447114
Buffalo	0.449469
Providence	0.467881
Boise	0.475533
Baton Rouge	0.481204
Tallahassee	0.487598
Birmingham	0.489493
Akron	0.520577
Little Rock	0.521617
Salem	0.542086
Lincoln	0.549738
San Bernardino	0.555774
Aurora	0.559237
Tacoma	0.559419
Worcester	0.579283
Newark	0.721069

Big City

City	σ of Scores
Chicago	0.283577
Los Angeles	0.294884
San Diego	0.301421
San Francisco	0.315516
New York	0.322652
Phoenix	0.332041
Boston	0.360404
Dallas	0.360742
Washington	0.364875
Houston	0.380956
Philadelphia	0.382919
Indianapolis	0.383300
Austin	0.398930
Columbus	0.405751
San Jose	0.418823
Detroit	0.423515
Nashville	0.425489
Milwaukee	0.453483
Baltimore	0.548868
San Antonio	0.550095



Big City	3.5 Stars or Less	4.0 Stars Exactly	4.5 Stars or Greater
No	34.71%	42.92%	22.37%
Yes	25.32%	55.27%	19.42%

χ^2 Contingency Test

P_1 = Proportion of Reviews 3.5 stars or less

P_2 = Proportion of Reviews 4.0 stars exactly

P_3 = Proportion of Reviews 4.5 stars or more

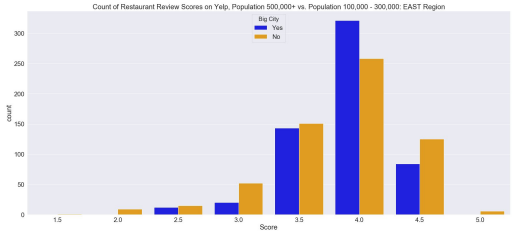
$H_0: (P_1:P_2:P_3)_{\text{big city=no}} = (P_1:P_2:P_3)_{\text{big city=yes}}$

$H_1: (P_1:P_2:P_3)_{\text{big city=no}} \neq (P_1:P_2:P_3)_{\text{big city=yes}}$

χ^2 Statistic = 274.31

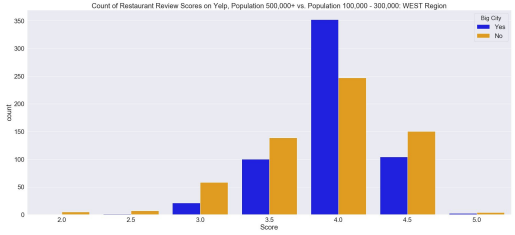
p-value = $2.71 * 10^{-60}$

East



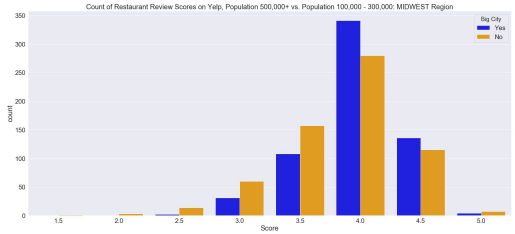
Big City	3.5 Stars or Less	4.0 Stars Exactly	4.5 Stars or Greater
No	36.95%	41.82%	21.23%
Yes	30.17%	55.34%	14.48%

West



Big City	3.5 Stars or Less	4.0 Stars Exactly	4.5 Stars or Greater
No	34.26%	40.49%	25.25%
Yes	21.03%	60.69%	18.28%

South



Big City	3.5 Stars or Less	4.0 Stars Exactly	4.5 Stars or Greater
No	36.89%	43.96%	19.15%
Yes	22.67%	54.82%	22.51%

Midwest

Big City	3.5 Stars or Less	4.0 Stars Exactly	4.5 Stars or Greater
No	30.71%	45.34%	23.95%
Yes	27.53%	50.34%	22.13%

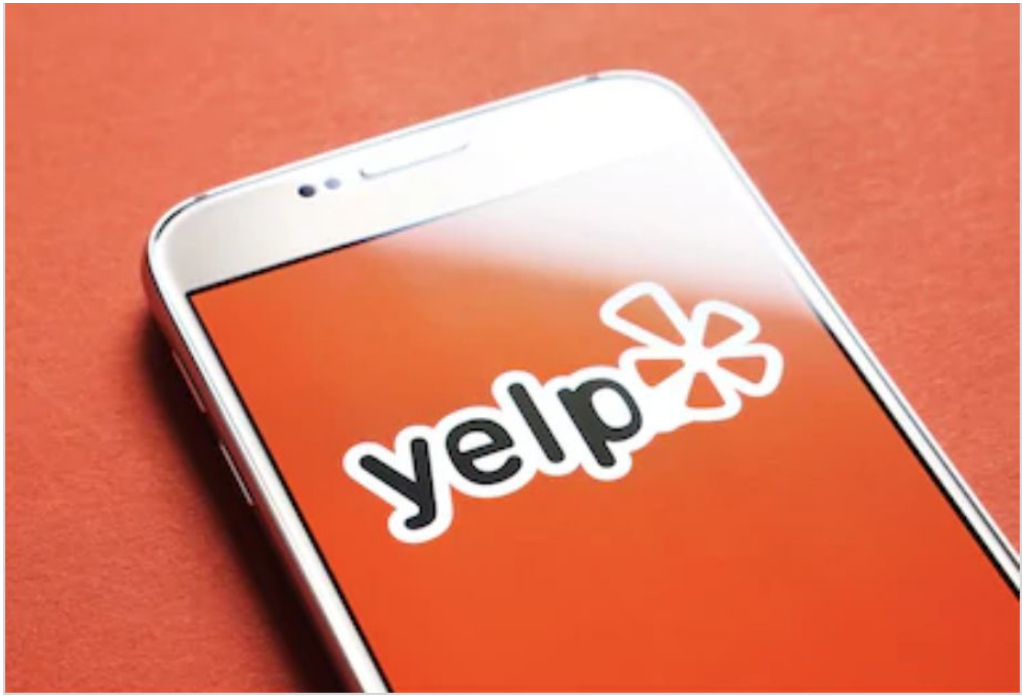
Conclusions

1. Geographic, demographic, and cultural factors do appear to have influence on the way cities review their restaurants.

2. Major U.S. cities (population of 500,000 or above) in the Northeastern region of the U.S. are less likely than other regions to give a popular restaurant (top 120 most reviewed) a score of 4.5 stars or above.

3. Restaurants in mid-sized U.S. cities (population of 100,000 to 300,000) generally experience a wider range of review scores than major U.S. cities, which center more closely around a 4.0 rating.

Thank you!



NYC DATA SCIENCE ACADEMY