

Find home cleaners, burgers, spas...

Near Manhattan, NY



Home Cleaners



Restaurants



Home Services ▾



Delivery



Web Scraping Project

Esther Chang
NYCDSA Fall 2018

Research Overview

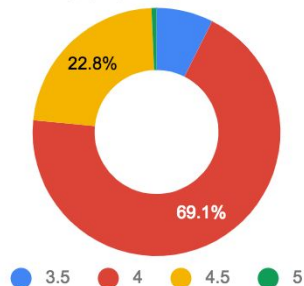
Objective	Use information from Yelp to find trends and insights into NYC restaurants	
Tools	Web scraping using scrapy in Python, Data analysis using dplyr in R	
Data	Restaurants searched under Manhattan, NY	
Information collected	Restaurant Name Type of cuisine (up to 3) Star Rating Price Range Number of reviews Number of photos Neighborhood Address Phone Number	First review date Takes reservation (T/F) Delivery (T/F) Takeout (T/F) 3 review highlights

Research Questions

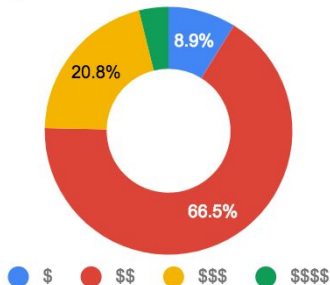
- Which neighborhood has the **most highly rated** restaurants?
- Does the distribution of **price range** differ by neighborhood?
- Is there a correlation between **number of reviews** and star ratings? Or **number of photos**?
- Where do we see the most opening of **new restaurants**?

Exploratory Data Analysis

Yelp Star Rating (5-point)



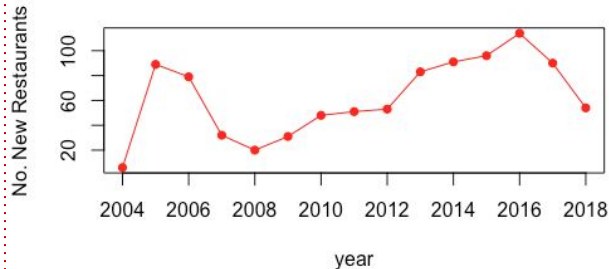
Price Range



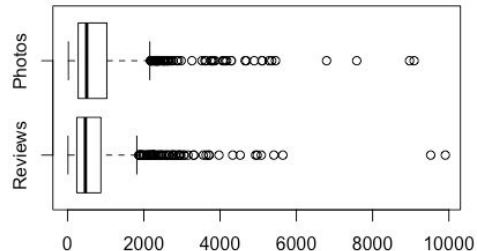
Top 5 Cuisine

1	Italian	16%
2	American (New)	13%
3	Japanese	7%
4	French	6%
5	Chinese	5%

Timeline of Newly Opened Restaurants



Number of Reviews & Photos



Median no. reviews: 467; Most reviewed: Katz's Delicatessen
 Median no. photos: 501; Most photographed: Ippudo NY

Takes Reservation	21%
Delivery	40%
Takeout	40%

Which neighborhood has the **most highly rated** restaurants?

Top 5 Highly Rated Neighborhood

1	Murray Hill (n=20)	★ 4.22
2	Lower East Side (n=54)	★ 4.18
3	Tribeca (n=25)	★ 4.12
4	Gramercy (n=71)	★ 4.12
5	Midtown East (n=88)	★ 4.12

- Restaurants in Murray Hill have the highest average star rating. Although only 20 restaurants were captured in the dataset, this makes sense since Murray Hill is well known for its variety in cuisines and price points.
- Four in five restaurants in LES is \$ or \$\$ with Chinese as the most frequent type of cuisines. This can be explained by many younger crowd living in this area.

Does the distribution of **price range** differ by neighborhood?

Highest Average Price Range*

1	Tribeca (n=25, avg= 2.68)	\$\$: 44%, \$\$\$: 44%, \$\$\$\$: 12%
2	SoHo (n=14, avg=2.64)	\$\$: 43%, \$\$\$: 50%, \$\$\$\$: 7%
3	Brooklyn (n=10, avg=2.50)	\$\$: 60%, \$\$\$: 30%, \$\$\$\$: 10%

* Calculated using \$\$ as numeric 2, \$\$\$ as 3 and \$\$\$\$ as 4

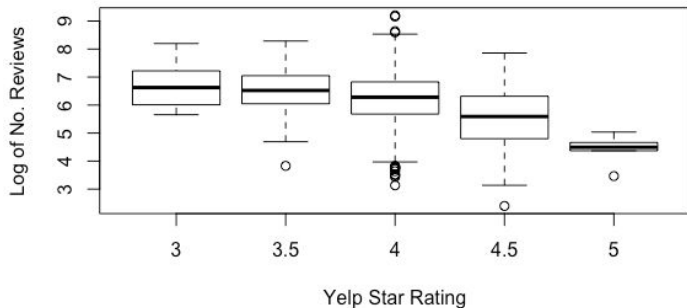
Lowest Average Price Range

1	Chinatown (n=17, avg= 1.53)	\$: 53%, \$\$: 41%, \$\$\$: 6%
2	Little Italy (n=28, avg=2.07)	\$: 7%, \$\$: 79%, \$\$\$: 14%
3	LES (n=54 avg=2.09)	\$: 17%, \$\$: 63%, \$\$\$: 15%, \$\$\$\$: 6%

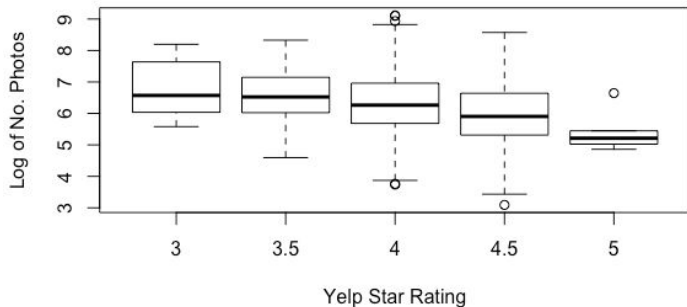
- The three mentioned neighborhoods with high average price range do not have any \$ restaurant - however, given the low sample size, we should take this statement with a grain of salt.
- Neighborhoods with low average price range tend to have higher proportion of \$ and \$\$ restaurants.

Is there a correlation between **number of reviews/ photos** and star ratings?

Boxplot of No. Reviews and Star Rating



Boxplot of No. Photos and Star Rating



- Restaurants with fewer reviews and photos are more likely to have higher ratings. This makes sense because newer restaurants tend to get high ratings initially with a soft opening.
- However this relationship is weak with the correlation(r) of -0.18 and -0.11 respectively. More robust modeling is needed.
- Note that number of reviews and photos were natural-logged for better presentation

Where do we see the most opening of **new restaurants**?

Top 5 Neighborhood with the most opening of new restaurants*

1	Midtown West (n=163)	11
2	Gramercy (n=71)	7
3	Midtown East (n=88)	7
4	East Village (n=81)	6
5	Greenwich Village (n=105)	5

* Restaurant with the first review date in 2018 is used as a proxy.

- Neighborhood with more restaurants tend to have more openings proportionally.
- Hell's Kitchen area in Midtown West particularly has a high rate of new restaurants opening.

Key Findings

- Murray Hill and LES are two foodie destinations for yelpers.
- Certain neighborhoods such as Chinatown, Little Italy and LES tend to have more affordable price range than others.
- High star ratings were weakly associated with fewer number of reviews and photos.
- Midtown West has many new restaurants.

Next Steps & Future Research

- Statistical modeling of the relationship between star ratings and all other variables
- Text analytics and advanced graphics capabilities
- Machine Learning application (i.e. dimension reduction and/or cluster analysis)
- Search and recommendation algorithm based on filters



Find home cleaners, burgers, spas...

Near Manhattan, NY



Home Cleaners



Restaurants



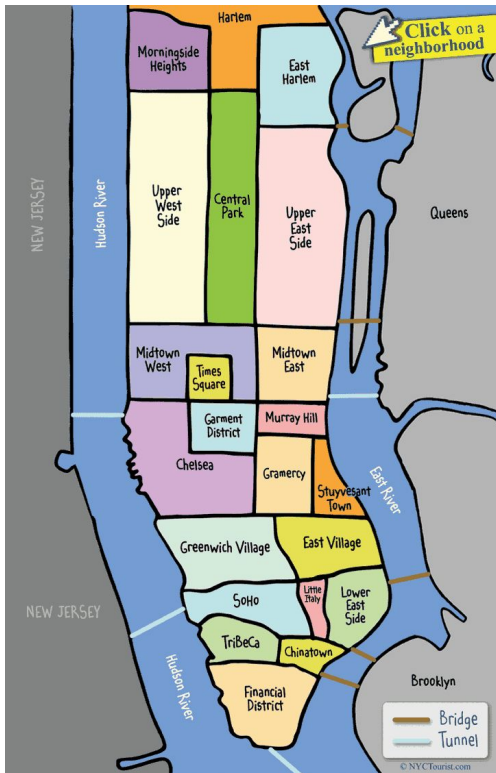
Home Services ▾



Delivery

Appendix

Data: Manhattan Restaurants



*source: nyc tourist.com

Neighborhood	No. of Restaurants
Midtown West	163
Greenwich Village	105
Midtown East	88
East Village/ Stuyvesant Town	81
Gramercy	71
Chelsea	62
Upper West Side	58
Lower East Side	54
Upper East Side	53
Theater District	42
Little Italy	28
Tribeca	25
Financial District	23
Harlem/ Morningside Heights	23
Murray Hill	20
Chinatown	17
SoHo	14
Brooklyn	10