OpenTable.com Web Scrape

Ben Berger

Introduction

- I used Scrapy to scrape OpenTable.com on July 24, 2020
- Manhattan restaurants currently taking reservations
- Data scraped included restaurant name, genre, neighborhood, price category, rating, number of reviews and whether it had takeout/delivery options.

delivery	expensive	genre	location	num_reviews_list	rating_list	rest_name	reivews_rating_log
Takeout	4	Steakhouse	Union Square	312.0	4.4	Strip House Speakeasy	25.269214
Neither	2	American	Upper West Side	93.0	4.6	Trump Grill	20.849958
Both	2	Mexican	Hell's Kitchen	1005.0	4.5	Tacuba Hell's Kitchen	31.107343
Neither	2	Seafood	Chelsea	473.0	4.6	Cull and Pistol	28.331839
Takeout	4	Japanese	Midtown West	79.0	4.8	Yakitori Tori Shin - Select Counter	20.973350
Neither	3	Italian	Greenwich Village	519.0	4.4	Gradisca	27.508377
Takeout	2	Burgers	Union Square	681.0	4.4	5 Napkin Burger - Union Square	28.703674
Both	2	Thai	Union Square	397.0	4.2	Spice - Union Square	25.132532
Takeout	2	French	Greenwich Village	344.0	4.6	Bar Six	26.866952

Why?

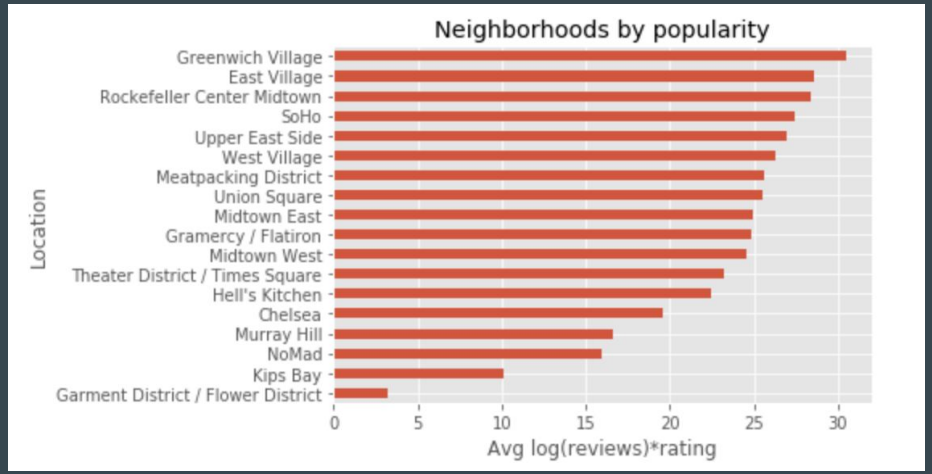
- Notoriously competitive industry in a notoriously competitive place
- Taking a snapshot of Manhattan restaurants: can we identify any patterns or consistencies among popular or successful restaurants?
- If you were starting a restaurant in the city, what characteristics or direction could help give you an edge in a naturally uncertain industry?

How?

- Basic analysis of certain restaurant characteristics vs. analysis of restaurants with no reviews/ratings (nulls)
- Relationship between popularity and uniqueness; do restaurants do better when they're the only one of their type in a given area?

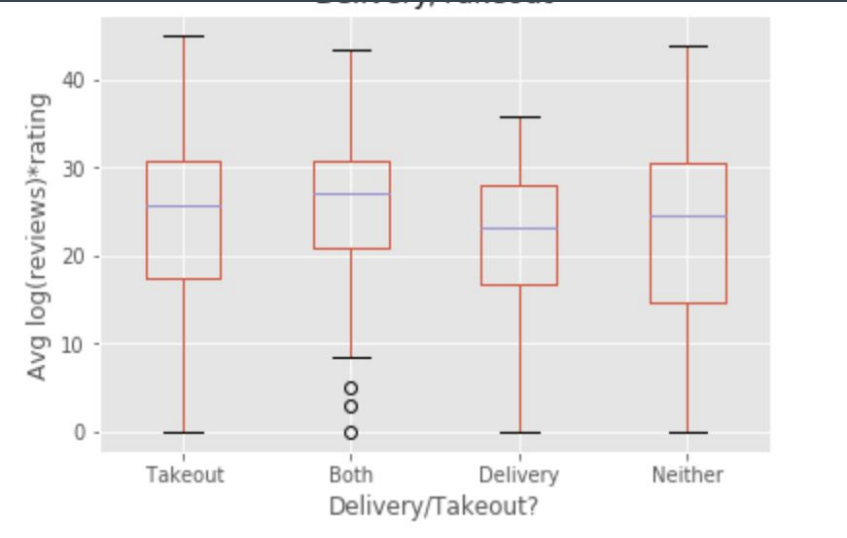
Location

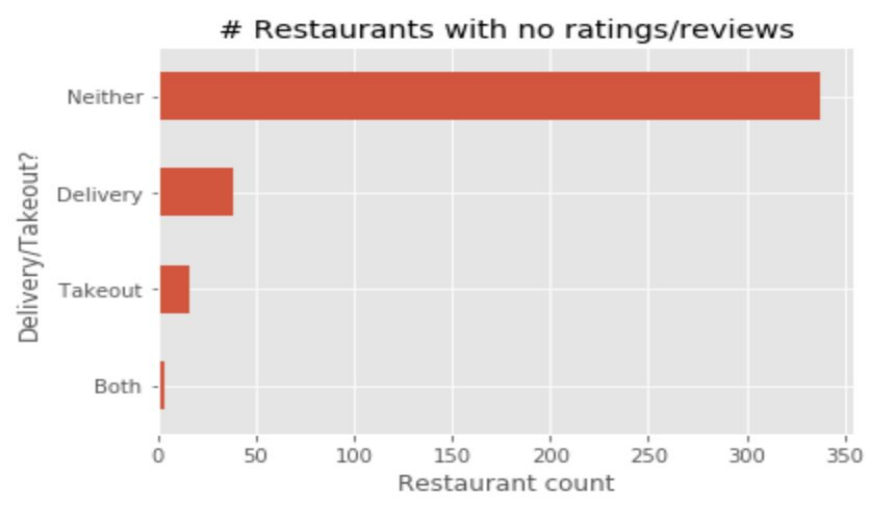
- Midtown West, Midtown East, Murray Hill, Chelsea, Times Square:
- Make up almost 60% of all restaurants in the data set, and 70% of restaurants with zero reviews. Not one in the top 5 of popularity.



Delivery/carryout

- ‘Both’ mean popularity score highest of all four, 25.76 mean and 7.89 standard deviation vs. 22.69 and 10.61 respectively for ‘Neither.’ (means significantly different at .05 significance level)
- 38% of whole data set is ‘Neither’ vs 85% in null set.

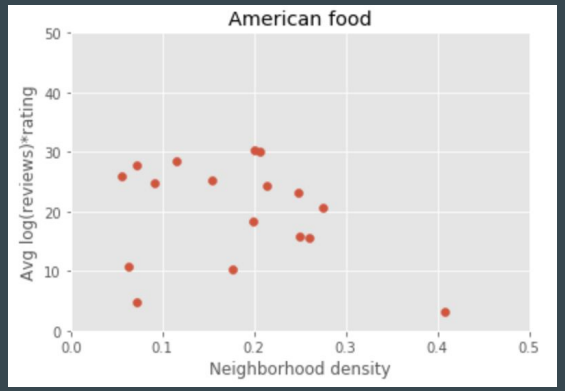


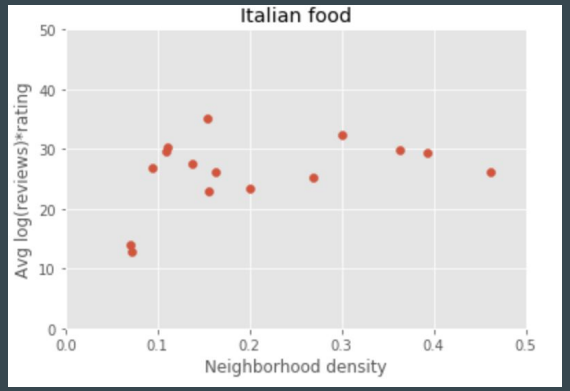


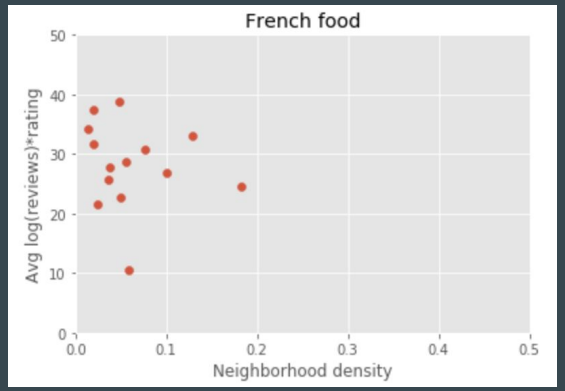
Characteristic density

- To address the question of performance relative to how many similar competitors exist around you, I found the number of a restaurant with a certain trait as a proportion of total restuarants in a neighborhood and plotted that density against popularity.
- I did the analysis first with three of the most commons genres of restaurants: American, Italian and French.
- Next with the three different price categories.

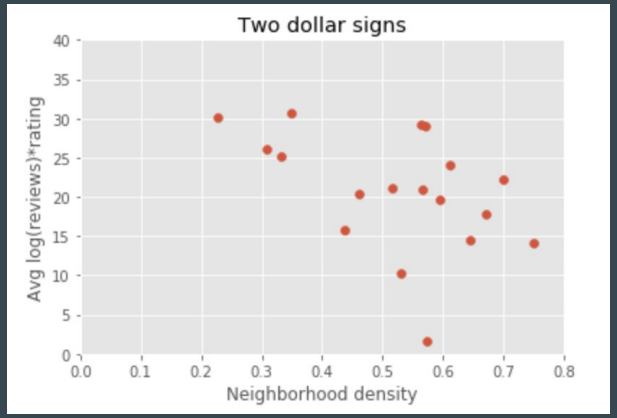
Genre

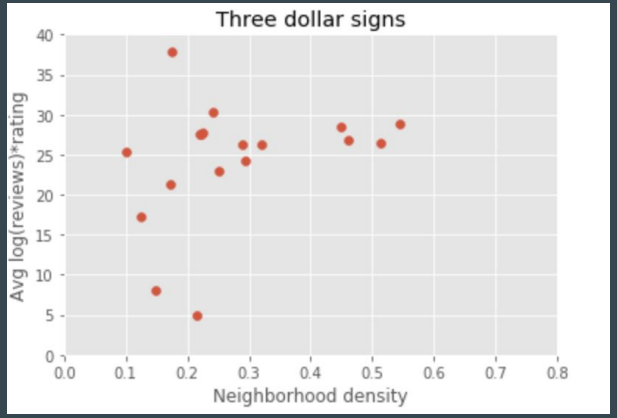

American food
Avg log(reviews)*rating
Neighborhood density
0
10
20
30
40
50
0.0
0.1
0.2
0.3
0.4
0.5

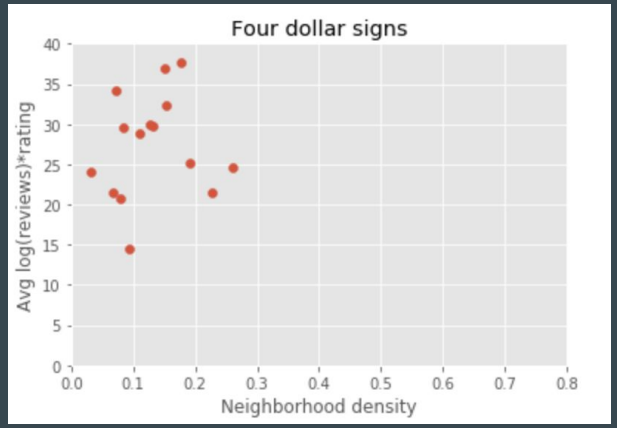

Italian food
Avg log(reviews)*rating
Neighborhood density
0
10
20
30
40
50
0.0
0.1
0.2
0.3
0.4
0.5


French food
Avg log(reviews)*rating
Neighborhood density
0
10
20
30
40
50
0.0
0.1
0.2
0.3
0.4
0.5

Price category







Conclusions

- Of course, none of this data is conclusive on its own. But it raises questions and encourages further research into the idiosyncratic world of food!
- Particularly in this post COVID world, are office heavy neighborhoods like Midtown as attractive places to set up shop as they once were? The data hints at them potentially being overcrowded and now possibly dated.
- Restaurant dining is a shadow of its former self. Our findings here heavily point at restaurants being more successful when their food is available in non sit down ways.
- Is it better to be unique, or join in on a popular trend? Unclear. But it seems that relationship changes depending on the characteristic you're analyzing.