

NYC's restaurants analysis

NYCDSA Web Scraping project

Leonardo Quintero
Jan 2019



Is there a relation between restaurant's prices and positive reviews broken down by cuisine?

Potential interest target

- People looking for good restaurants in town

Purpose of the Project

- Make it easy to decide which restaurants to visit and which ones to avoid in NYC.

Questions

- Does a high price guarantee quality of food?
- Are restaurants in higher end neighborhoods mean higher quality or better experiences?
- Is there a predominance of cuisine type by neighborhood?
- What neighborhoods are worth visiting in order to have a good culinary experience?



Web Scraping

Tripadvisor.com with Scrapy

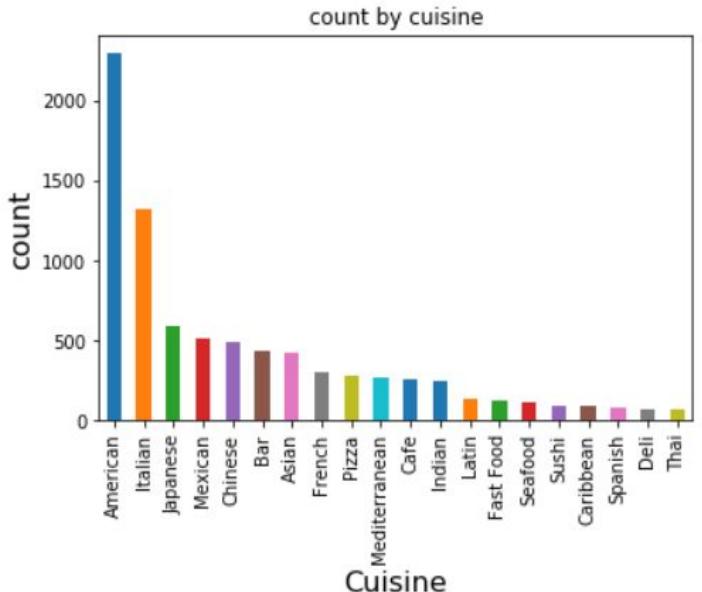
Obtaining the URL from the first page “next” button and creating a for loop to iterate over all subsequent pages

At 30 restaurants per page, 12446 restaurants were scraped including informations as shown below.

avg	cuisine	name	price	rank	reviews
0 4.5	Asian,Restaurants	Taiyaki NYC	\$	#1 of 10,534 Restaurants in New York City	300
1 4.5	American,Steakhouse,Latin,South American,Glute...	K Rico	-	#1 of 530 Quick Bites in New York City	2,562
2 4.5	Asian,Korean,Gluten Free Options,Restaurants,D...	Jungsik		#2 of 530 Quick Bites in New York City	247
3 4.5	Italian,Pizza,Vegetarian Friendly,Vegan Option...	B Side Pizzeria	-	#2 of 10,534 Restaurants in New York City	340
4 4.5	Indian,Vegetarian Friendly,Vegan Options,Glute...	Indian Accent		#3 of 10,534 Restaurants in New York City	441

The screenshot shows the TripAdvisor homepage for New York City. At the top, there's a green header bar with the TripAdvisor logo, a search bar, and navigation links for Travel feed, Hotels, Things to do, Restaurants, Flights, Vacation Rentals, Shopping, and more. Below the header, a breadcrumb trail shows United States > New York (NY) > New York City > New York City Restaurants. To the right, it says "Best Restaurants in New York City, NY". The main content area is titled "Restaurants in New York City". On the left, there's a sidebar with a map, selection filters (Restaurants, Quick Bites), establishment type filters (Restaurants, Quick Bites, Dessert), and a reservations section. The main content area features sections for "Browse New York City by Food" (with categories like Pizza, Steakhouse, Italian, Chinese, Seafood, Japanese, Mexican, Sushi) and "Moderately Priced" (with examples like Piccola Cucina Osteria, Piccola Cucina, SottoCasa Pizzeria, and Song E Napule). Each listing includes a thumbnail image, name, price range, reviews, and a link to the full restaurant profile.

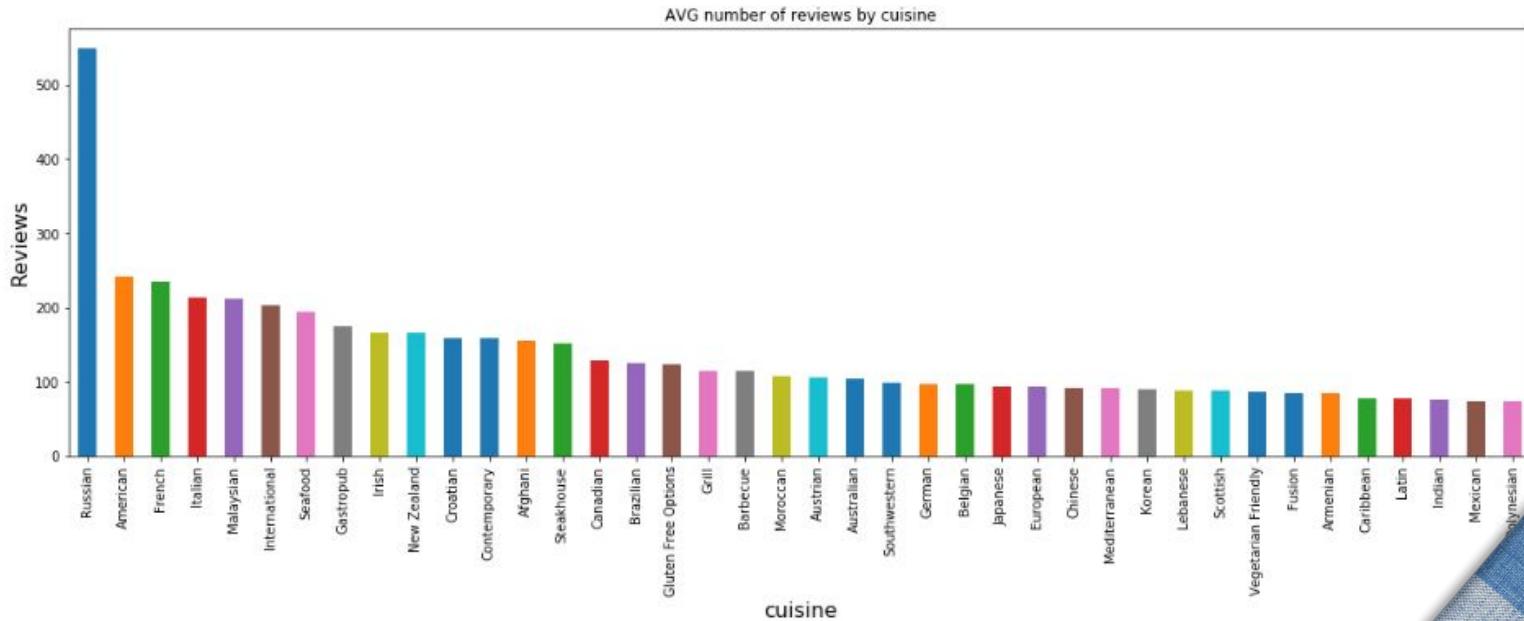
The most popular type of cuisine in NY seems to be American, followed by Italian, Japanese and Mexican.



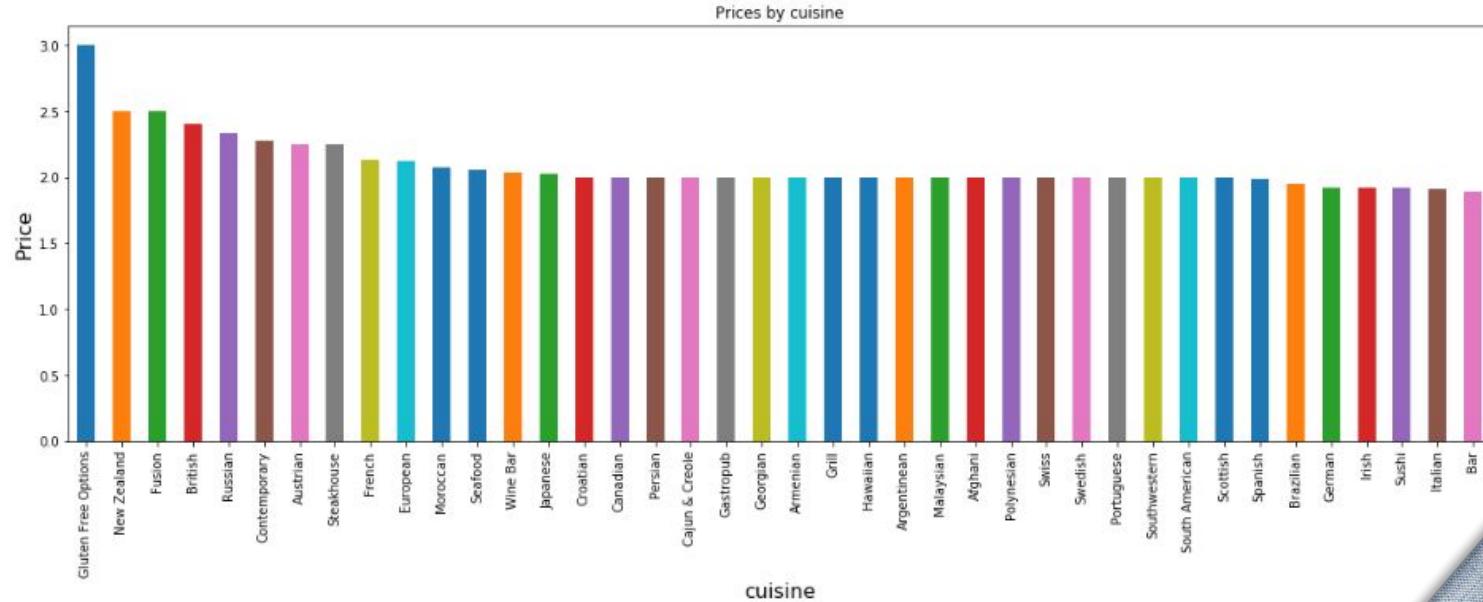
This could be cautiously used as a parameter for ethnicity segmentation in the city of NY.



But the most reviewed cuisine is Russian

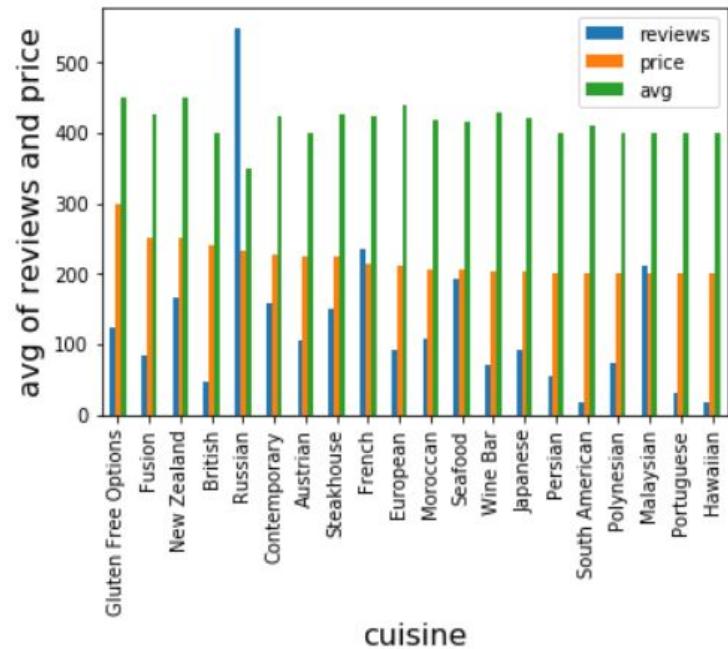


What are the most expensive types of cuisine in NY?

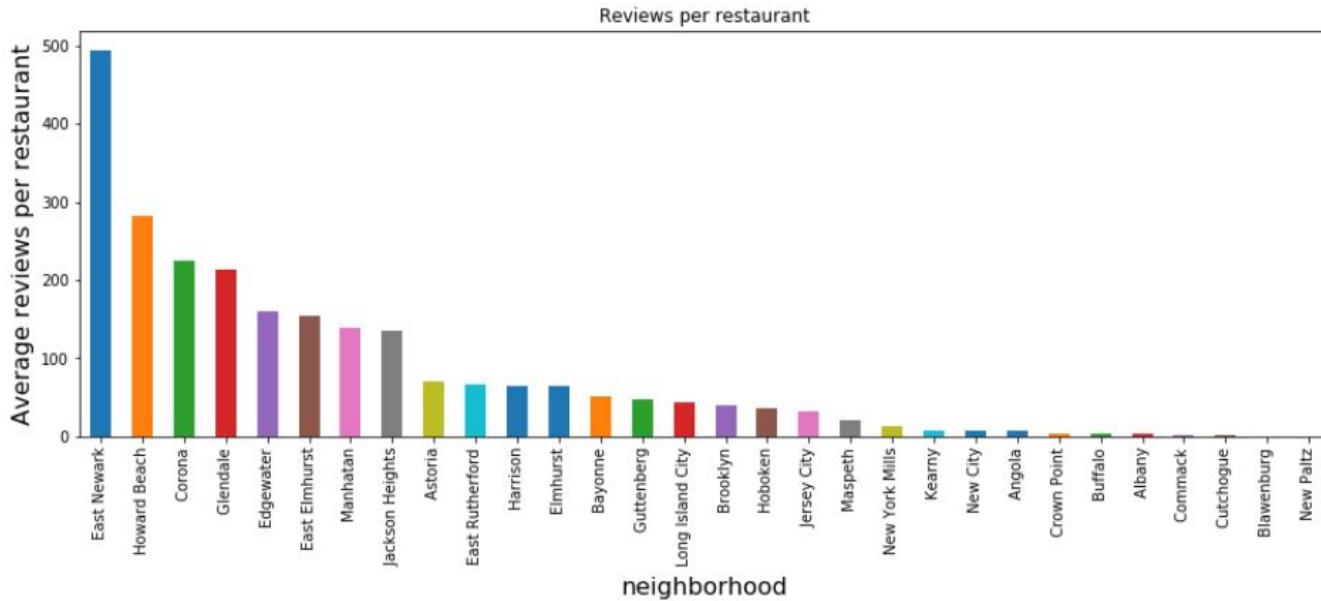




Gluten free options seem to be more expensive options, less likely to be rated but with higher average rating

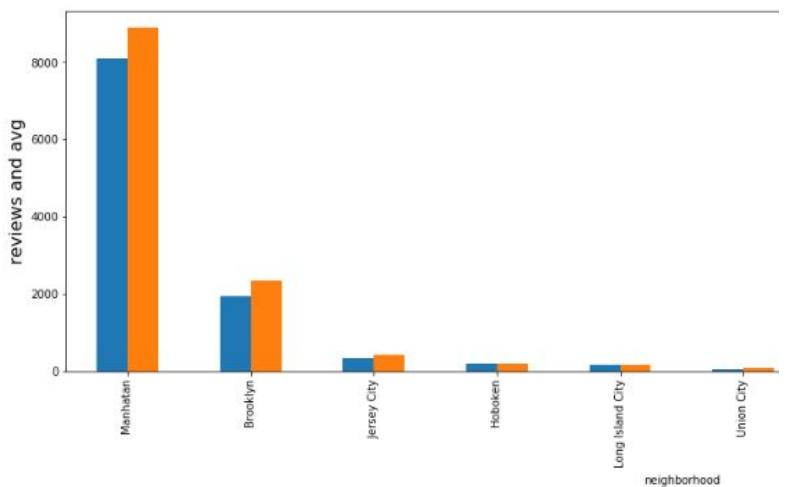


Analysing data by Neighborhood



People seem to be more likely to review restaurants in the surrounding areas, in average, with more reviews per restaurant.

With 1.122.396 reviews, Manhattan seems to be the neighborhood with the highest concentration of restaurants and reviews. Also seems to have the highest ratings.

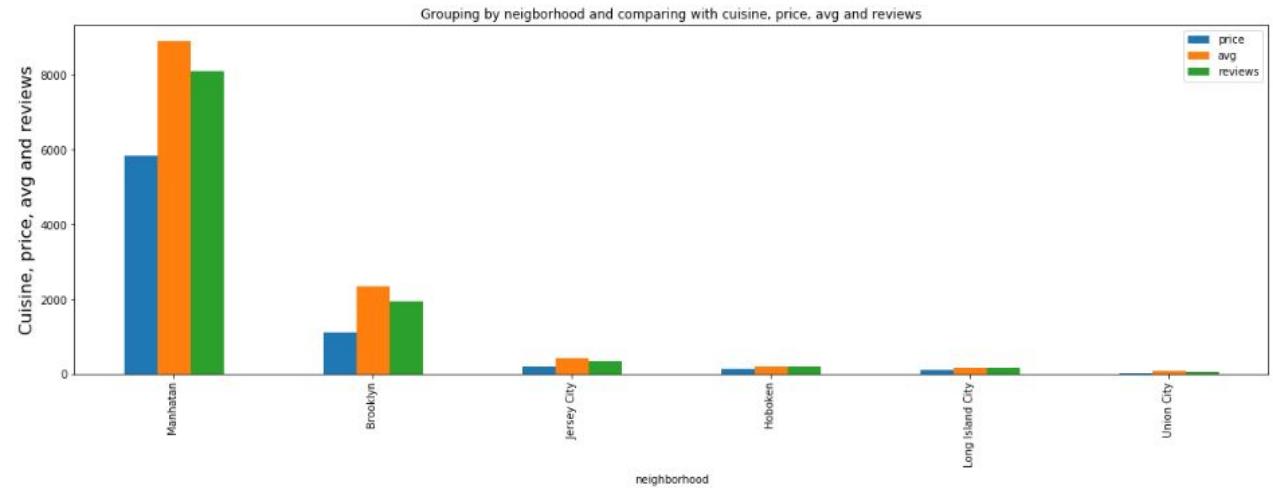


neighborhood	reviews	avg
Manhattan	1122396.0	3654250.0
Brooklyn	76810.0	992000.0
Jersey City	10810.0	168550.0
Hoboken	6735.0	82700.0
Long Island City	6632.0	68650.0
Astoria	5654.0	35300.0
Union City	655.0	33850.0
Sunnyside	1070.0	26150.0
Secaucus	3097.0	11700.0
North Bergen	1385.0	8700.0
Union City	1.0	1.0



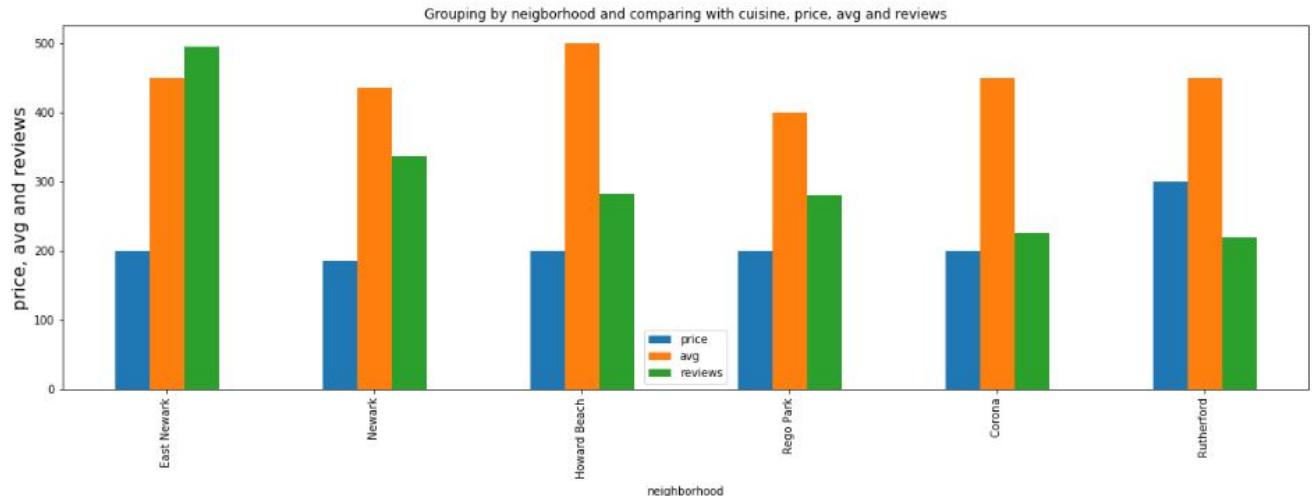


Manhattan seems to have a lower correlation between prices and average rates compared to the rest of neighborhoods





However, in average, the highest concentration of reviews, low prices and best ratings are away from manhattan.



neighborhood	price	avg	reviews
East Newark	2.000000	4.500000	494.000000
Newark	1.857143	4.357143	336.142857
Howard Beach	2.000000	5.000000	282.000000
Rego Park	2.000000	4.000000	281.000000
Corona	2.000000	4.500000	225.333333
Rutherford	3.000000	4.500000	225.000000

Conclusion

- 12,446 restaurants scraped from NYC
- Scrapped data from trip advisor's NYC area, including surrounding areas.
- There is a chance to select a restaurant or cuisine by cuisine
- Vegan or gluten free options are an important part of the market, probably due to being specific types of food.

- Higher prices do not guarantee better quality of food.
- There are better chances of finding a better culinary experience depending on the neighborhood you decide to dine in.

