Navigating the Return to In-Person Dining in NYC

An analysis of data scraped from OpenTable

By Matthew Fay

Table of Contents

COVID-19 Pandemic Disruption of the Restaurant Industry

- Estimated $240 billion in losses due to COVID-19 pandemic
- 100,000+ businesses in the industry closed, either temporarily or permanently

National Restaurant Association, 2021

Lasting impacts of the pandemic into 2021 and industry recovery

- Integration of technology:
  - Explosion in usage of online ordering and delivery services
  - Websites and online reviews are the second-largest source of new customers at 35%, behind only recommendations from friends and family at 49% (Toast, 2020)

- Pent-up demand remains high:
  - In April 2020, 83% of adults said they were not eating on-premise at restaurants as often as they'd like, up from 45% in January 2020 (National Restaurant Association, 2021)

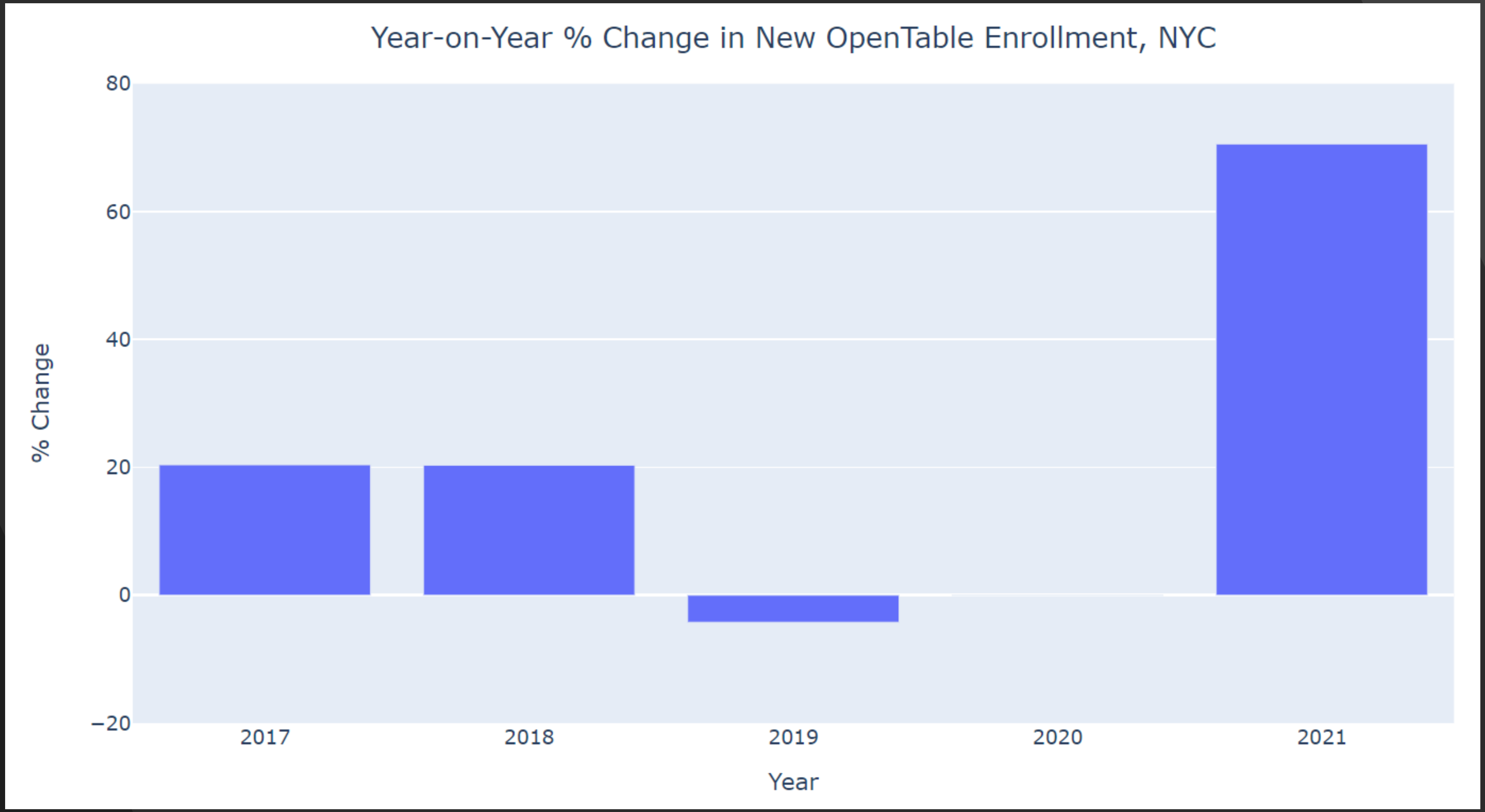
- Global dining reservations as of July 2021 are approaching pre-pandemic levels (OpenTable State of the Industry Report, 2021)

As the return to in-person dining continues, how can a restaurant maximize customer influx through online reservation systems?

- Scraped information from every restaurant in New York City currently active on OpenTable: ~1100 restaurants with 25 features
- Scraped number of bookings at each restaurant in NYC over 1 week in July 2021

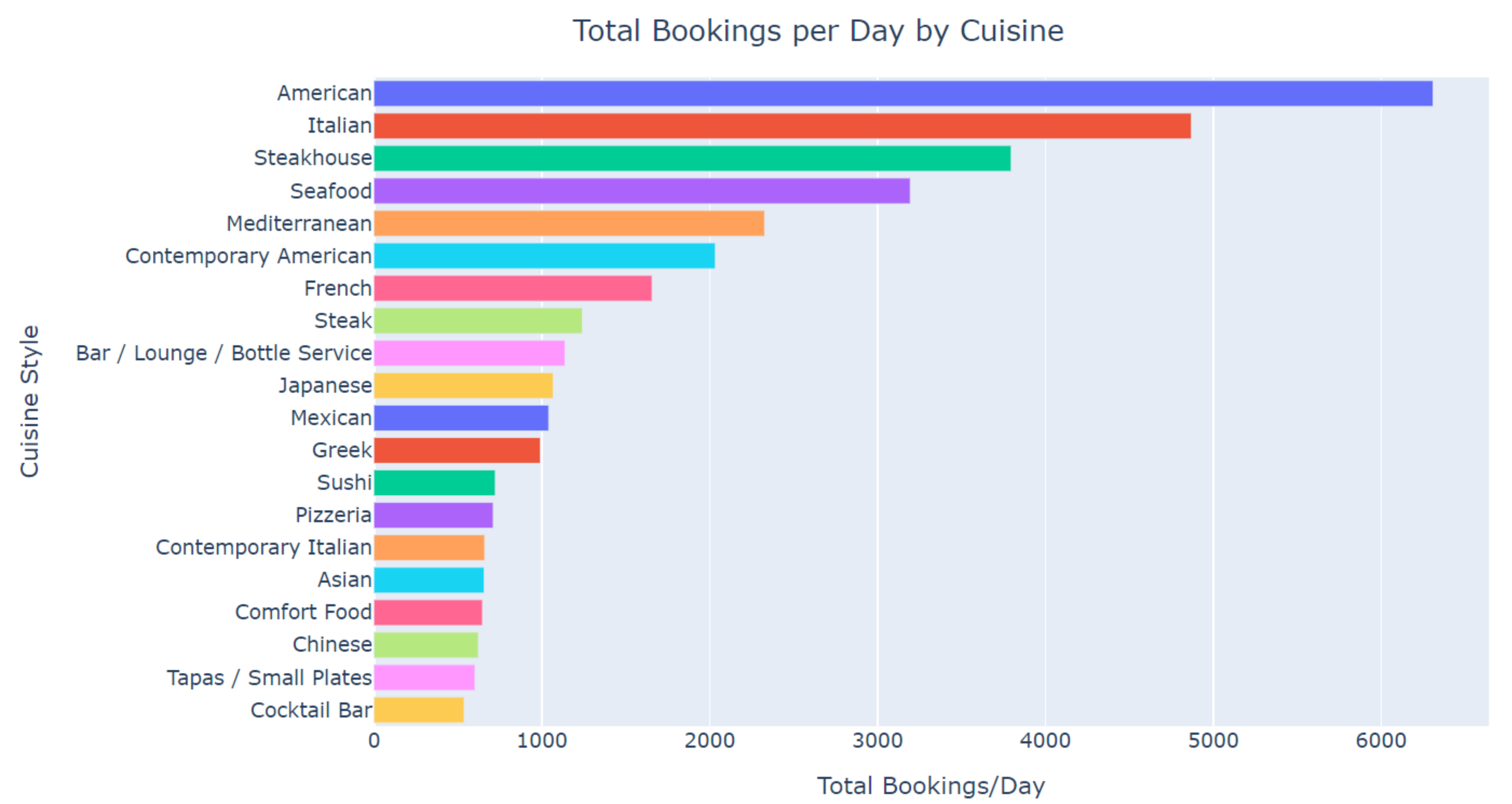
- Analysis of 3 groups of factors and their effects on OpenTable bookings:
  - Types of food
  - User reviews and promoted status
  - Management of COVID-19 safety concerns

Restaurant enrollment on OpenTable is consistent with resurgence of in-person dining in 2021



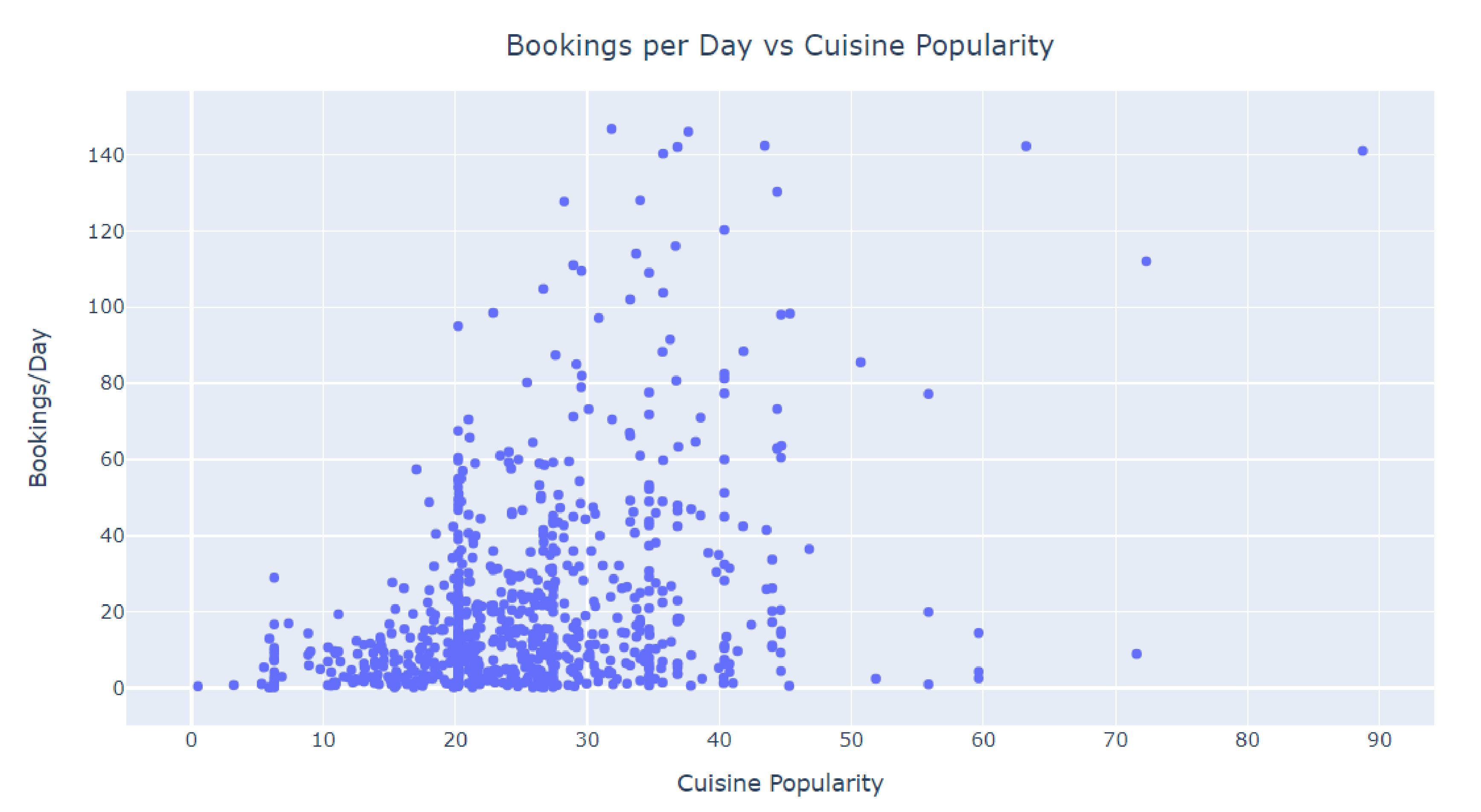
- New restaurant enrollment on OpenTable in 2019 and 2020 in NYC was stagnant at best
- 2021 has seen a 70% increase from 2020 in new restaurants as of July
- This data does not account for restaurant closure or leaving the OpenTable platform; the drop in 2020 was likely much starker than reflected here

An overview of the New Yorker's palette: Top 20 cuisines by total volume of bookings



Big fish in smaller ponds

- Be specific when labeling your product—targeting niche audiences is a very successful strategy for some businesses
- OpenTable allows 4 cuisine tags, less than 5% of restaurants use all 4
- Ex. A Sardinian restaurant might tag Italian to have broader appeal, but in the sea of NYC Italian restaurants, also tagging Sardinian may capture a large portion of the users who know and seek out Sardinian food

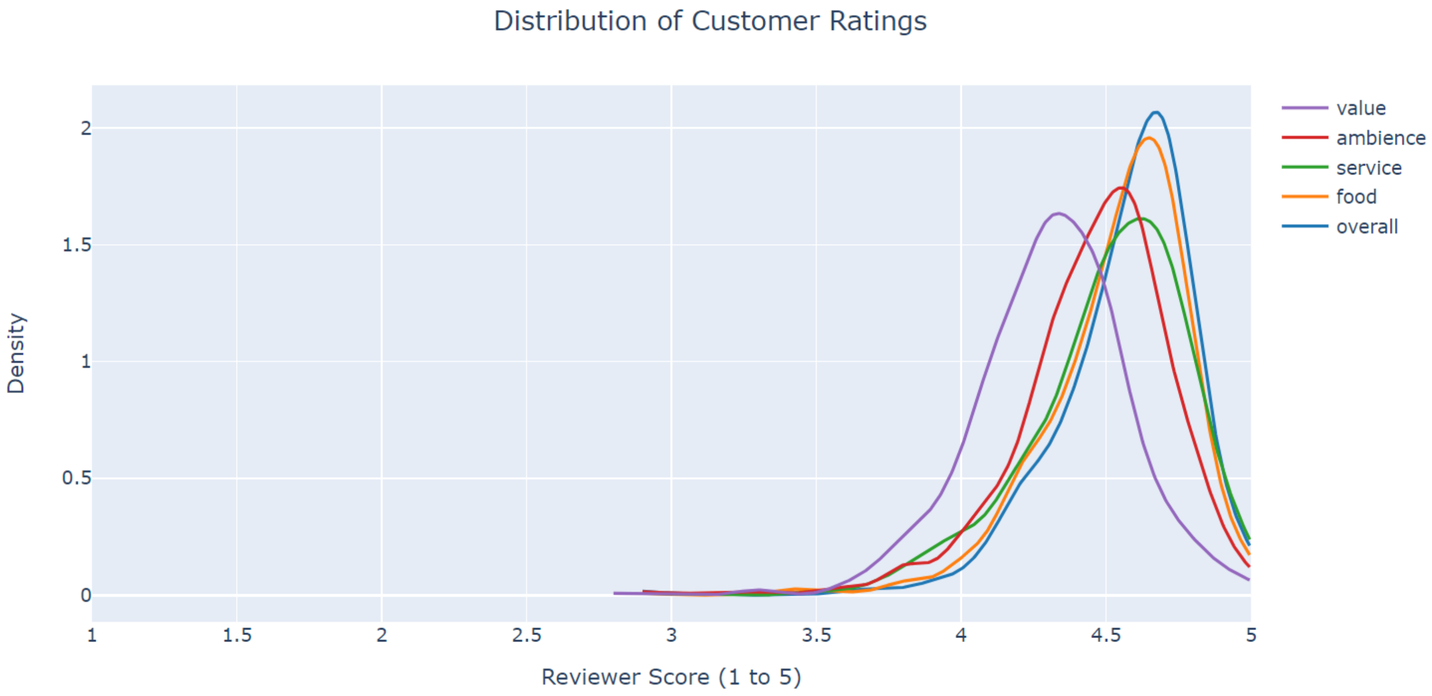


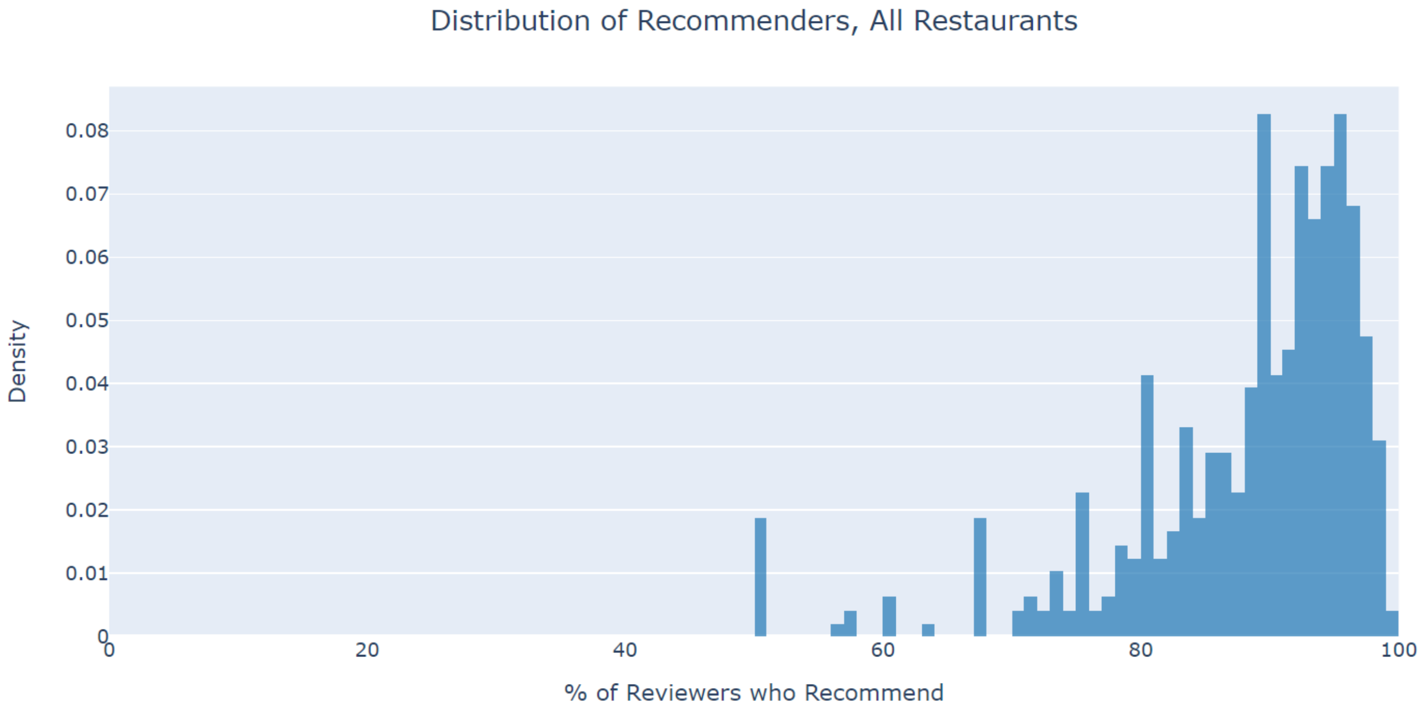
- Combined cuisine popularity for each restaurant was estimated using the geometric mean of bookings/day for all tagged cuisine styles at that restaurant

Customer ratings: the be-all, end-all?

OpenTable user reviews have 5 criteria:

- Food
- Service
- Ambience
- Value
- Would you recommend it to others?





Nearly 100% of NYC restaurants on OpenTable have 4 stars or better in every individual category and overall

Recommendations follow a very similar distribution, clustered around 80-100% recommended

Explanations:

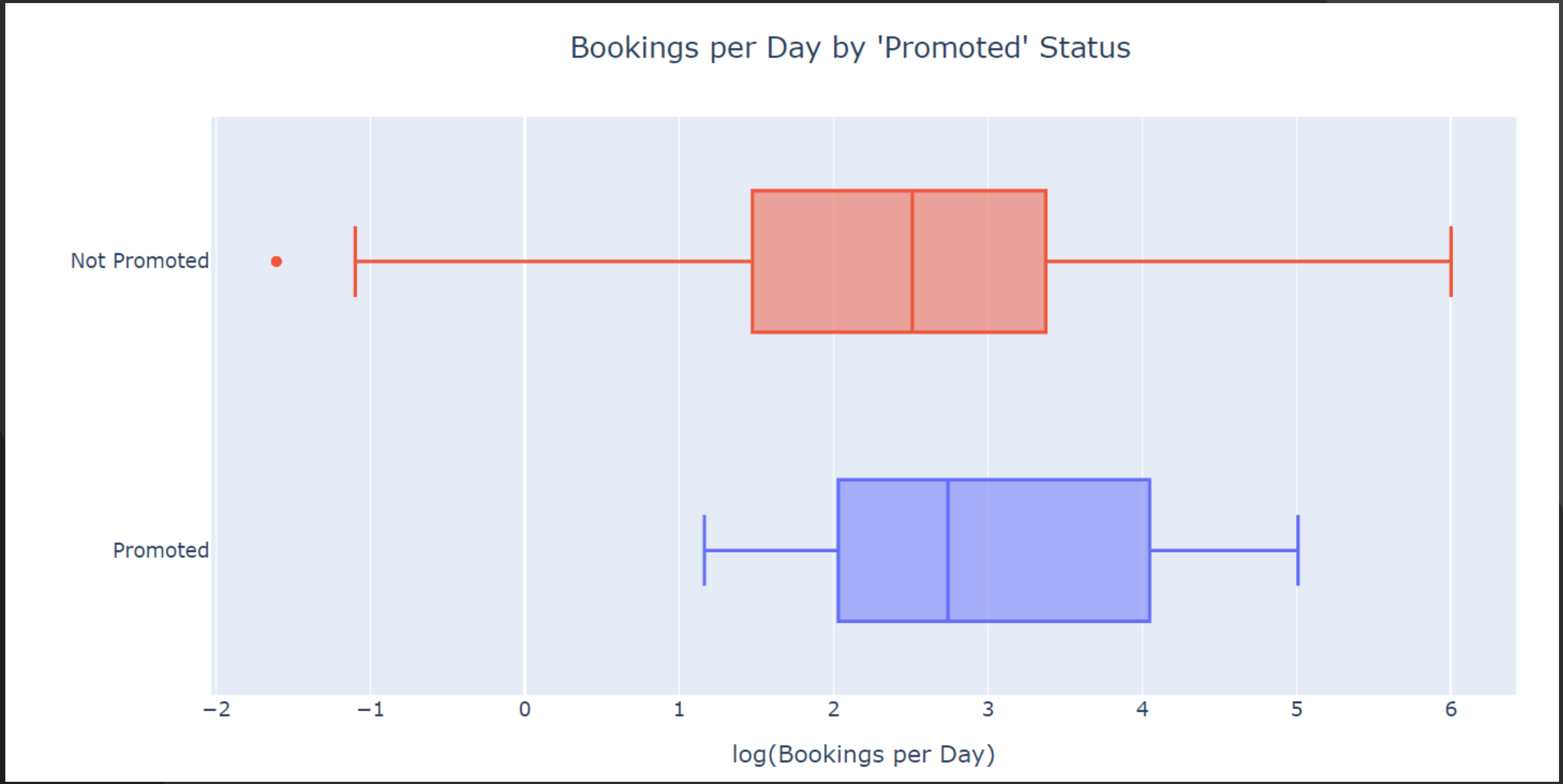
- Bad to mediocre restaurants do not get enough customers to make an online reservation system a good investment
- Only restaurants with high-quality products survived the industry's decline in 2020
- Reviews are curated and some portion of negative ratings are removed
  - All levels of OpenTable membership plans offer 'Review Management'
  - See https://restaurant.opentable.com/products
- Nearly all restaurants in New York City are just really, really good?

What factors are predictive of reservation numbers if not ratings?

Categorical factors: Price

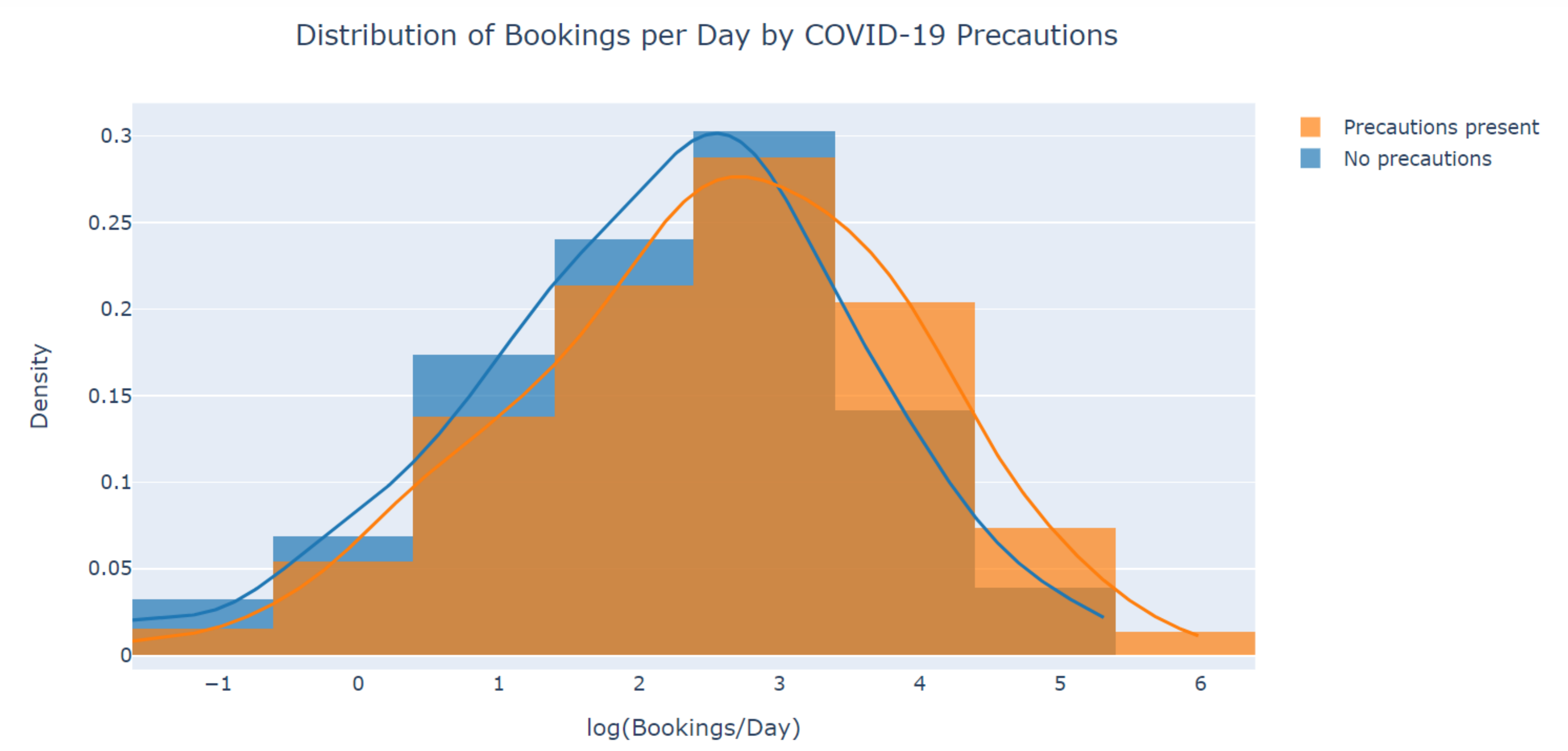
- More expensive restaurants get more bookings, significant at $p<0.05$ level in ANOVA
- Probable selection bias: cheaper/more casual restaurants tend to get more of their business from walk-ins and less from reservations

Categorical factors: Promoted Status



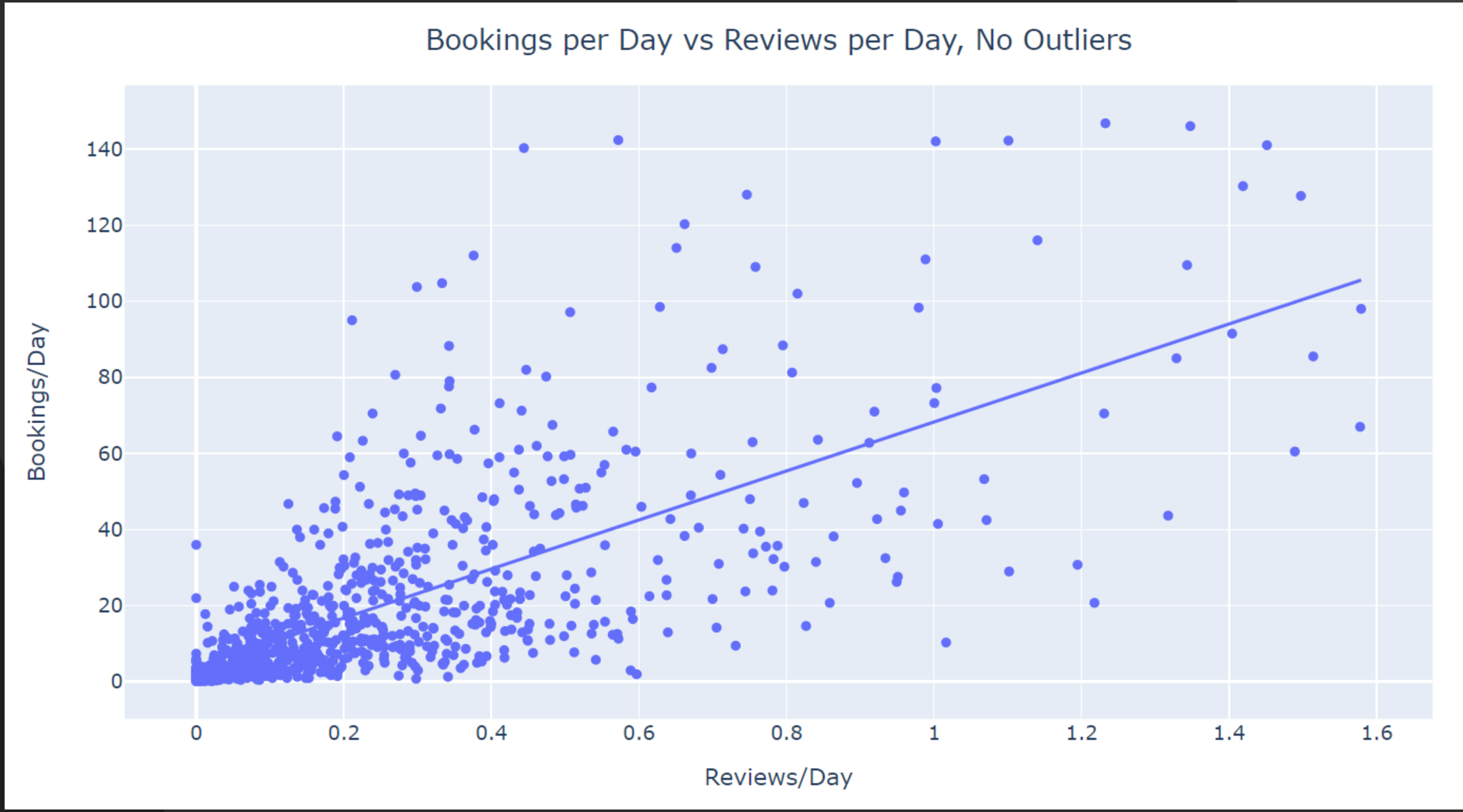
- Promoted status on OpenTable results in higher numbers of bookings, significant at $p<0.05$ level in 2-sample t test
- More on this later

Categorical factors: COVID-19 Precautions



- 4 types of safety measures: distancing, PPE, sanitizing, and screening
- Data was grouped by whether any of the 4 types were used
- Small difference may be because most consumers who are dining in-person again are either vaccinated or do not weigh the risk of COVID-19 very highly
- Significant at $p<0.05$ level in 2-sample t test

Quantitative analysis and modeling



- The best predictor by far for bookings per day was reviews per day
- Linear model: Each review per day averages ~64 bookings per day (r^2 = 0.47, p = 1e-113)
- Much of the variance in the dataset is still unaccounted for in this model

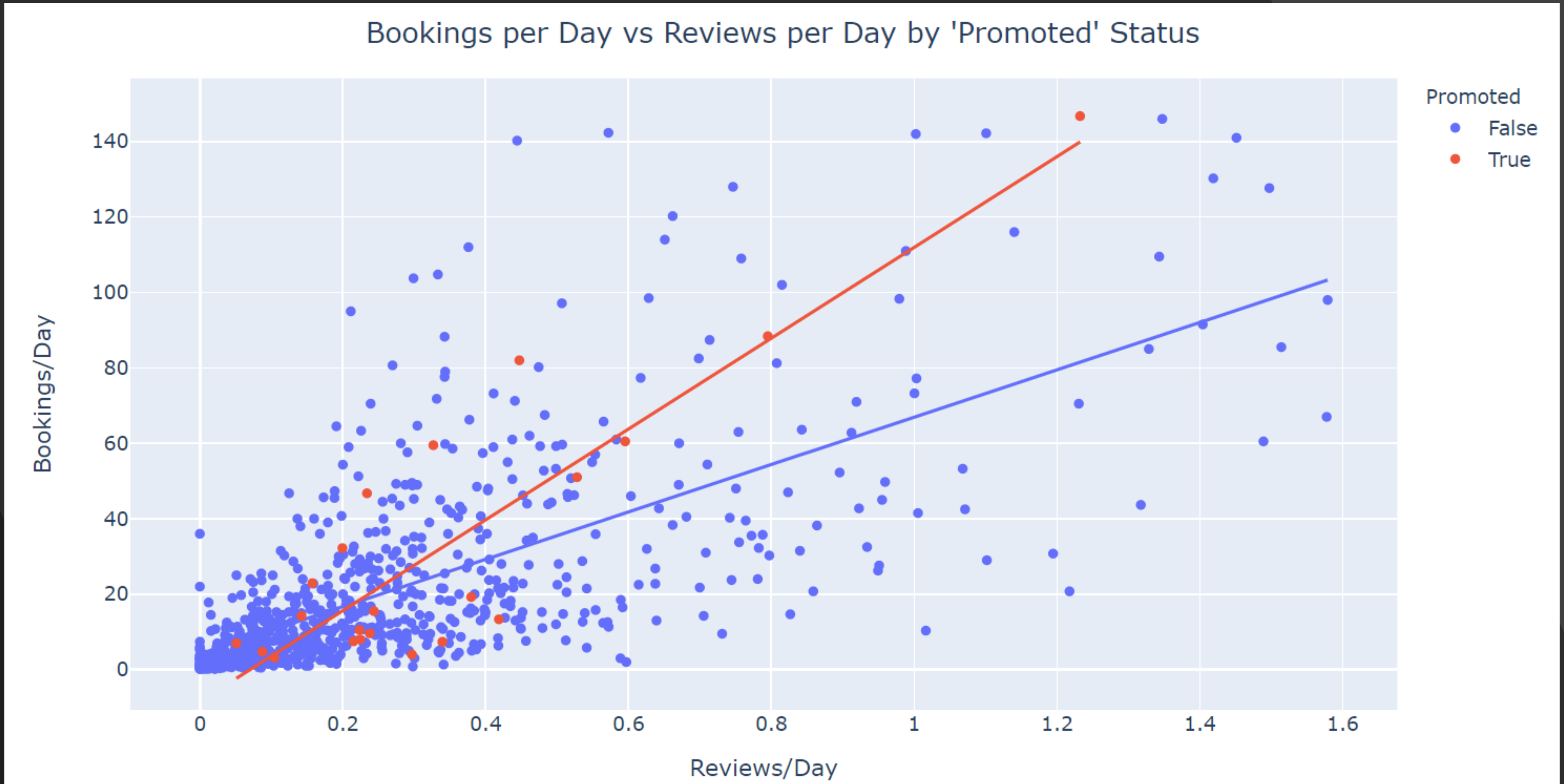
Some commentary on the model:

- Unfortunately, this analysis cannot point to whether the correlation between reviews/day and bookings/day is causative

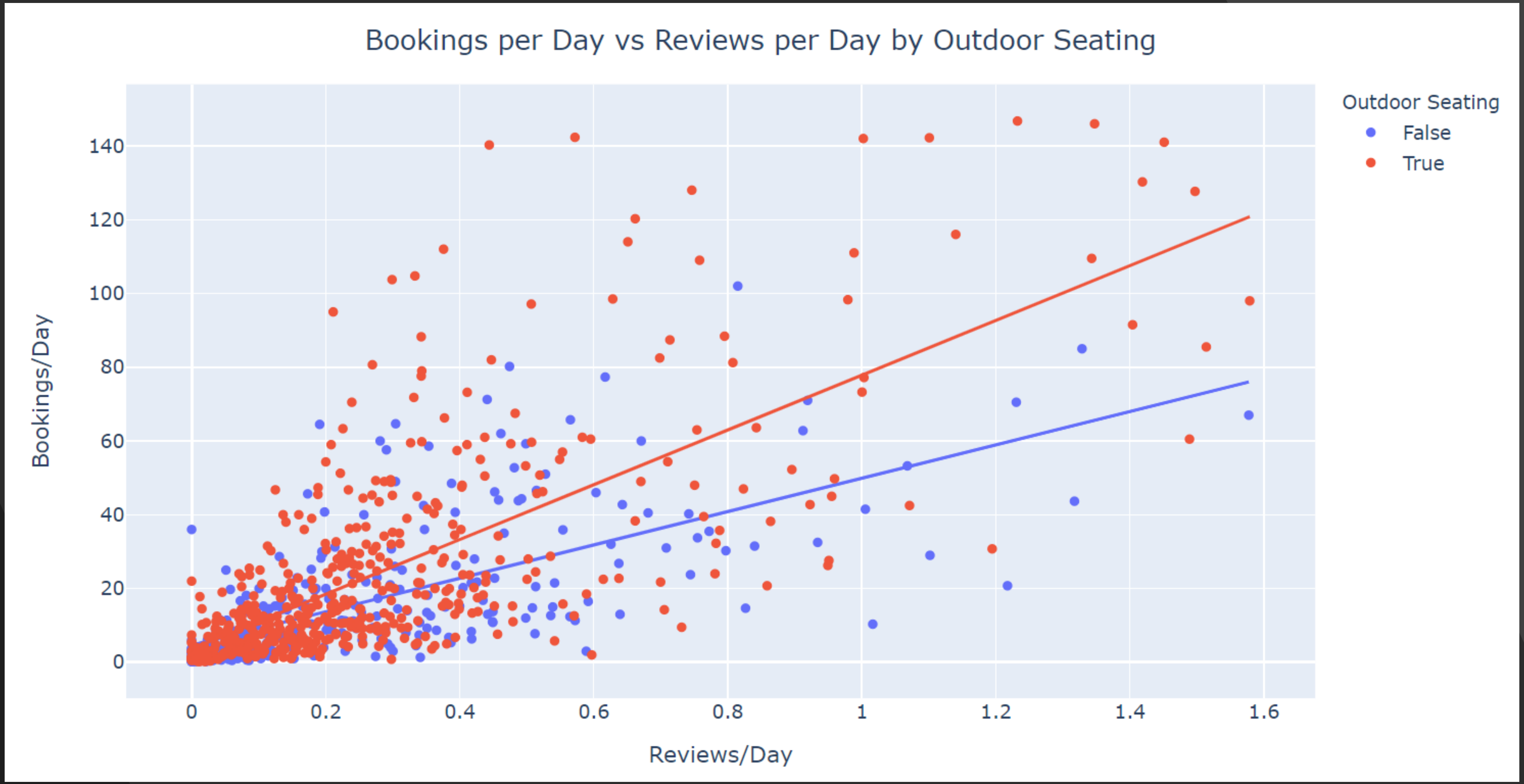
- As a business, you could test the causality of the relationship
  - Ex. Offer incentives for customers to leave reviews and investigate whether increasing your reviews/day rate increases your bookings/day rate

- Confounding variables:
  - Physical seating capacity of each restaurant—bigger space means more potential customers who can leave reviews and also the ability to take more bookings/day
  - Ideally, would have data to control for this and model (bookings/day) vs (reviews/day/seat)

- Layering categorical analyses on top of this model provides interesting insights:



- Promoted restaurants get almost twice as many bookings per review per day as non-promoted restaurants
- Promoted: slope = 120 $r^2 = 0.78$ p = 6e-8
- Non-promoted: slope = 63 $r^2 = 0.47$ p = 1e-107



- Restaurants that offer outdoor dining options receive about 65% more bookings per review per day than restaurants with indoor seating only
- Outdoor seating: slope = 74 $r^2 = 0.52$ p = 8e-82
- Indoor only: slope = 45 $r^2 = 0.44$ p = 1e-37

Dep. Variable:	avg_bookings	**R-squared:**	0.520
Model:	OLS	**Adj. R-squared:**	0.516
Method:	Least Squares	**F-statistic:**	122.0
Date:	Mon, 26 Jul 2021	**Prob (F-statistic):**	4.73e-121
Time:	21:11:03	**Log-Likelihood:**	-3421.1
No. Observations:	795	**AIC:**	6858.
Df Residuals:	787	**BIC:**	6896.
Df Model:	7		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	-11.2658	2.046	-5.506	0.000	-15.282	-7.249
C(outdoor)[T.True]	5.2176	1.399	3.728	0.000	2.471	7.965
C(any_precautions)[T.True]	2.5384	1.413	1.797	0.073	-0.235	5.311
C(promoted)[T.True]	6.6054	3.895	1.696	0.090	-1.040	14.251
C(price_tier)[T.2]	2.9469	1.431	2.060	0.040	0.139	5.755
C(price_tier)[T.3]	-0.1629	2.226	-0.073	0.942	-4.533	4.207
reviews_per_day	57.1456	2.602	21.960	0.000	52.038	62.254
cuisine_popularity	0.4295	0.072	5.968	0.000	0.288	0.571

Omnibus:	240.432	**Durbin-Watson:**	1.924
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	917.455
Skew:	1.386	**Prob(JB):**	5.99e-200
Kurtosis:	7.473	**Cond. No.**	165.

- Multiple linear regression incorporating:
  - Reviews/day
  - Cuisine popularity (see pg 9)
  - Price range
  - Promoted status
  - Outdoor seating
- Only marginally better than single linear regression on reviews/day
  - $r^2 = 0.52$
  - $p = 5e\text{-}121$
- Again, important factors still unaccounted for in this analysis
- Including any category of user ratings or recommendations weakened the model

Conclusions

- Be specific about labeling your cuisine, especially if you serve less-known regional or specialized foods
- The rate at which customers leave reviews, not the quality of those reviews, is the best predictor of bookings generated through the platform
- Promoted status increases bookings per review per day by almost 100%
    - Profitability of promoted status depends on your profit per booking vs the cost of buying promoted status
- Consumers in the post-pandemic NYC environment value outdoor seating

Improvements and future directions

- Longer timeframe
  - Bookings per day data are averages from 5 days in July 2021
  - Ideally, pre-pandemic data could be obtained for comparison
- Speed and scalability
  - First attempt at web scraping
  - I have since learned about methods much more efficient than the ones used here
- Little evidence for causality between reviews/day and bookings/day
  - Obtaining more data to control for physical capacity of restaurants
- Mapping and spatial analysis
  - Is location a factor in a restaurant’s success online?
  - Rent prices vs increased bookings tradeoffs by neighborhood/zip code