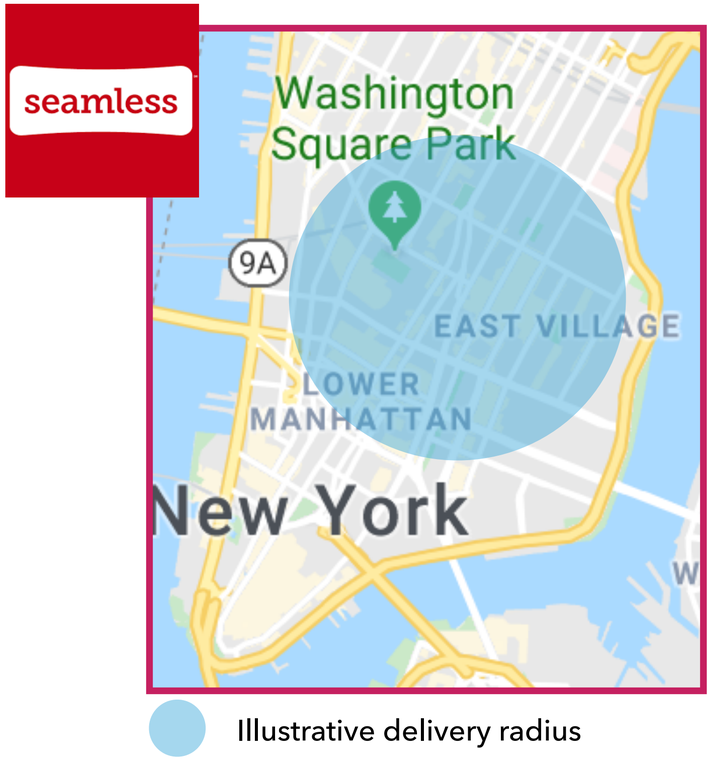
WEB SCRAPING SEAMLESS.COM

EAST VILLAGE RESTAURANT RATING PREDICTORS

10/28/2019

The Seamless site was scraped using Selenium to obtain data for 1,251 restaurants that deliver to the East Village in Manhattan

Dataset



- 1,513 restaurants that deliver to my address in Manhattan
- Complete cases of restaurants with >30 ratings yielded 1,013 results

Variables Collected

Variable	Description
Name	name of the restaurant
Rating	restaurant star rating (1-5 stars)
Number of Ratings	number of times the restaurant was rated
On-Time Percentage	proportion of users reporting that orders were delivered on time from a restaurant
Accuracy Percentage	proportion of users reporting that orders were accurate from a restaurant
Quality Percentage	proportion of users reporting that “food was good” from a restaurant
Price	restaurant price level
Category	restaurant cuisine category; may contain multiple categories (e.g., ethnicity, food type)

The data was obtained for the purpose and answering three key questions regarding the impact of collected variables on restaurant rating

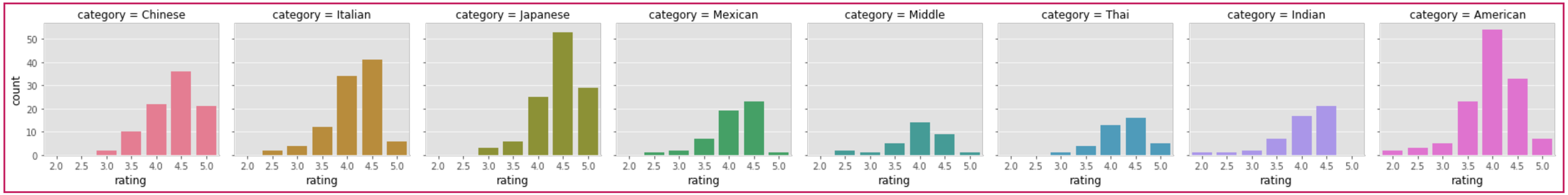
1. Are restaurant ratings significantly different by cuisine type (e.g., American, Japanese, Middle Eastern, etc.)?

2. How are order accuracy, timeliness, and food quality correlated with restaurant ratings, if at all?

3. Does the correlation of order accuracy, timeliness, and food quality correlated with restaurant ratings vary by cuisine type? How?

There is little variation in restaurant rating both within and across categories

Restaurant Rating Histograms by Cuisine Category*

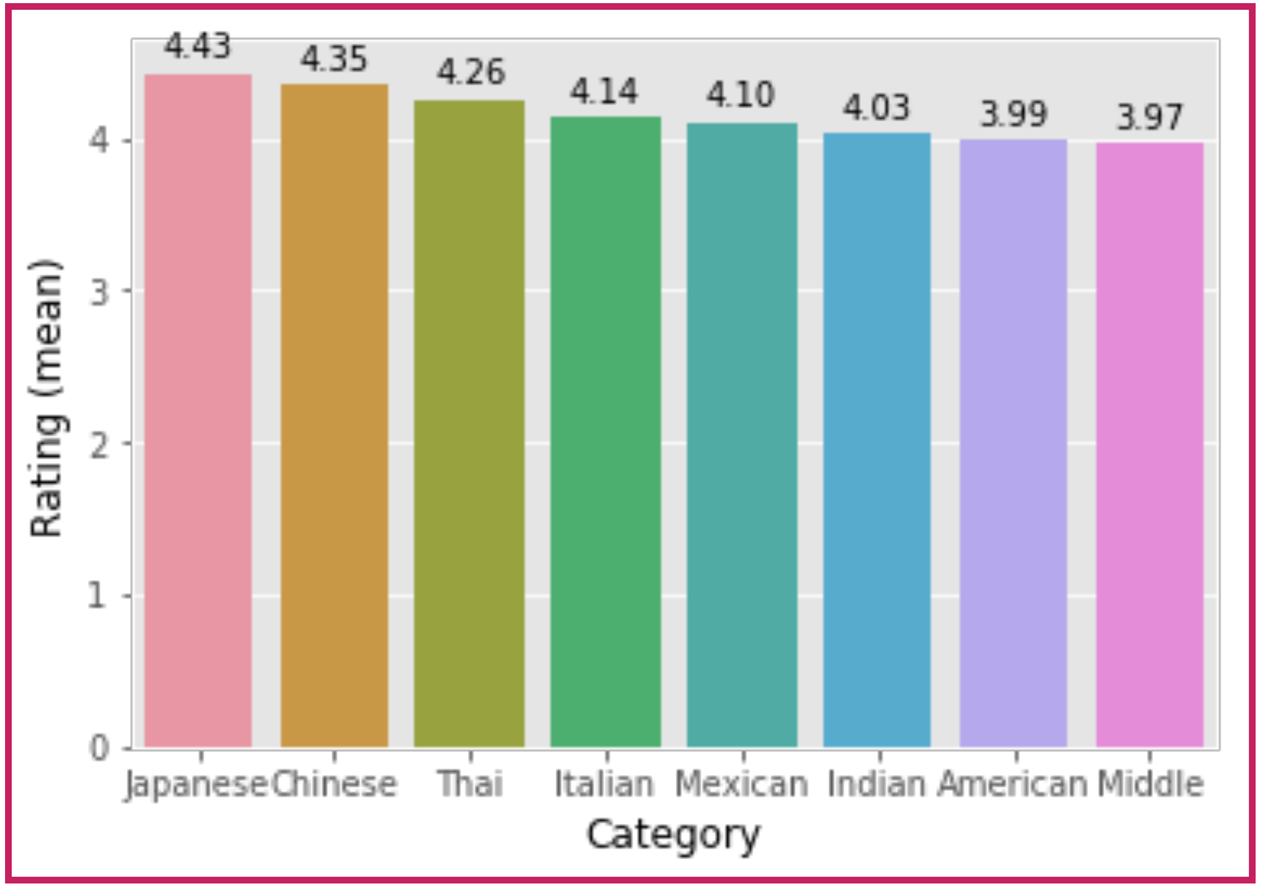


- Ratings data is not true continuous data, but rather pseudo-interval data
- Nonetheless, across cuisine categories ratings appear to be roughly normally distributed with some left skew
- There is little variation in rating within categories - most observations fall in the 4 or 4.5 rating bucket
- This pattern is true across cuisine categories, though some have a larger proportion of “5” ratings

*Restaurants in these cuisine categories account for 606/1,1013 total restaurants in the dataset; only cuisine types with >30 restaurants were included

East Village food delivery consumers rated Asian restaurants most highly

Mean Restaurant Rating by Category



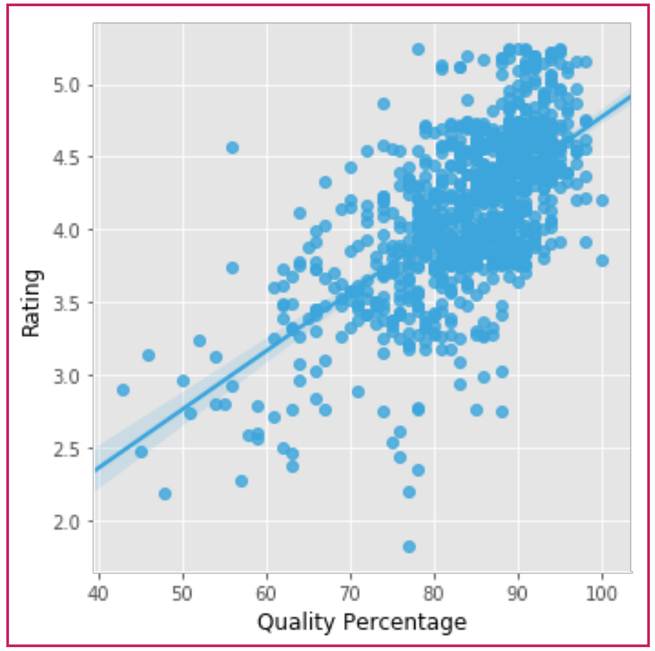
- Differences in mean rating by category were small but statistically significant (in some cases)*
- East Village food delivery consumers tend to rate Japanese, Chinese, and Thai restaurants significantly more highly than Indian, American, and Middle Eastern restaurants (and in some cases Mexican and Italian restaurants)**
- Italian restaurants were also rated significantly more highly than American restaurants**

*Barlett test for equal variance and visual inspection of rating distribution for normality by category showed that ANOVA assumptions were met. ANOVA results: $F = 9.57$, $p<0.01$
**As determined by pairwise two sided t-tests. Table with pairwise t-test values will be added to a presentation appendix

Delivery quality, timeliness, and accuracy are all correlated with restaurant rating

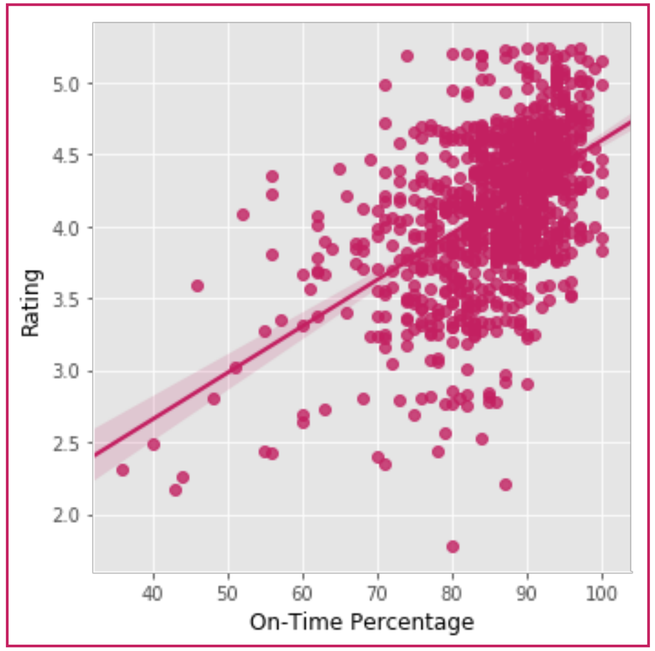
Restaurant Rating Correlation with Order Quality, Accuracy, and Timeliness*

Rating v. Quality



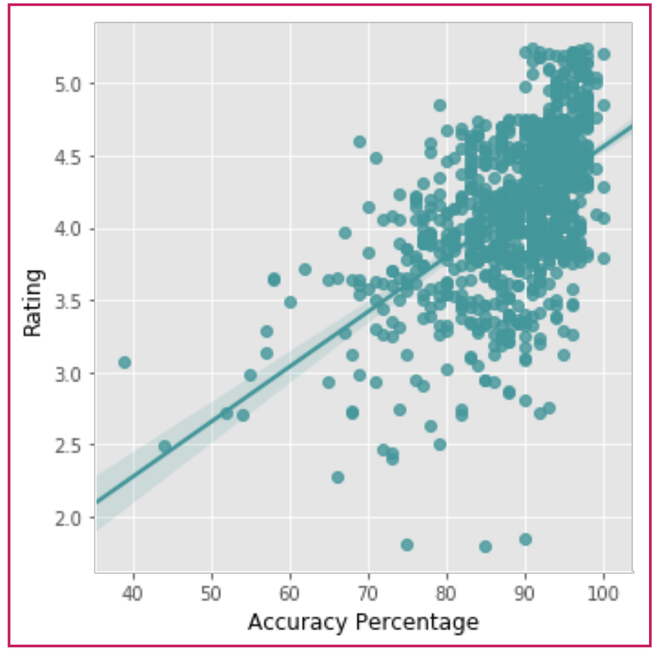
Pearson correlation coefficient = 0.66

Rating v. Timeliness



Pearson correlation coefficient = 0.54

Rating v. Accuracy



Pearson correlation coefficient = 0.56

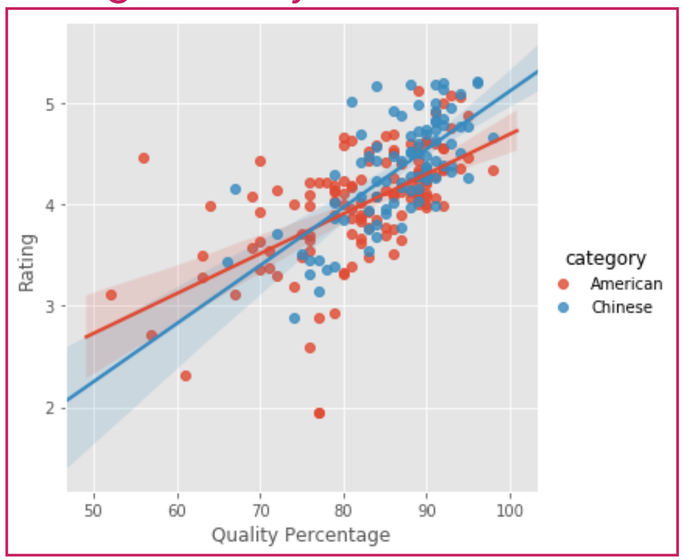
- Delivery quality, timeliness, and accuracy all show moderate-strong positive correlation with restaurant rating
- However, quality, timeliness, and accuracy show a high degree of multi-collinearity (see appendix), suggesting they are not independent explanatory factors. Additional work is planned to develop a multivariate regression model isolating best predictor(s).

*Data slightly stylized in that jitter was added along the y-axis

Order timeliness, accuracy, and quality (less so) are all more highly correlated with restaurant rating for Chinese vs. American restaurants

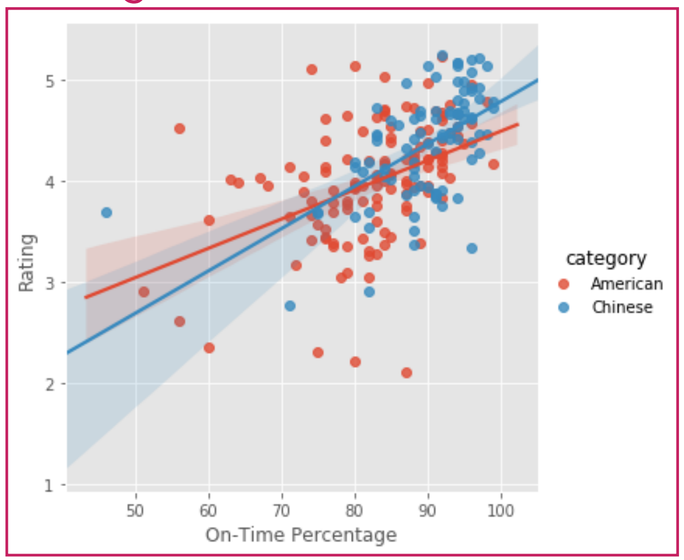
Restaurant Rating Correlation with Order Quality, Accuracy, and Timeliness by Cuisine Type*

Rating v. Quality



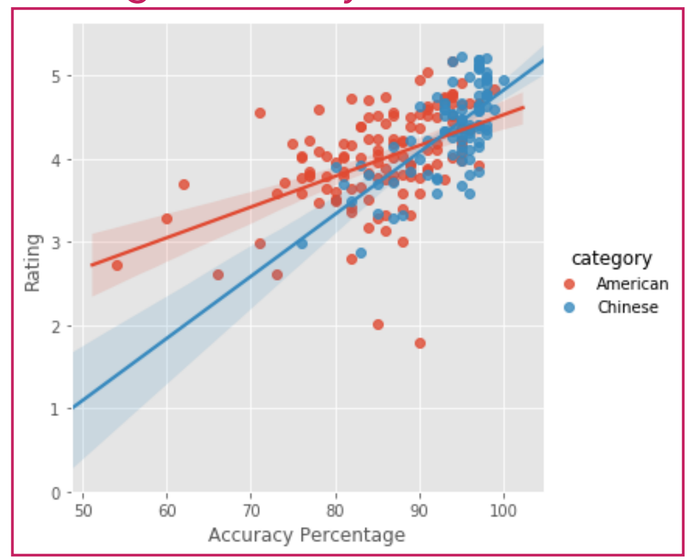
American: Pearson correlation coefficient = 0.6
Chinese: Pearson correlation coefficient = 0.71

Rating v. Timeliness



American: Pearson correlation coefficient = 0.45
Chinese: Pearson correlation coefficient = 0.61

Rating v. Accuracy



American: Pearson correlation coefficient = 0.51
Chinese: Pearson correlation coefficient = 0.71

- Delivery quality, timeliness, and accuracy all are more highly correlated with ratings in Chinese restaurants vs. American, perhaps due the wide variability among these variables in lower-rated American restaurants
- Correlation of quality and ratings was more slightly more similar in American vs. Chinese restaurants than correlations of order timeliness and accuracy with ratings

*Data slightly stylized in that jitter was added along the y-axis

Appendix

Order quality, timeliness, and accuracy are all strongly correlated with each other

	ontime_perc	accuracy_perc	quality_perc
ontime_perc	1.000000	0.689684	0.706180
accuracy_perc	0.689684	1.000000	0.679322
quality_perc	0.706180	0.679322	1.000000