Analysis of Restaurant Ratings and Prices in Beijing and Los Angeles

Zhixing Lu (Amy)

Stanford Online High School

Redwood City, CA 94063, USA

amylucky@ohs.stanford.edu

Abstract

This project investigates the relationship between restaurant ratings and prices and compares these statistics between two cities: Beijing and Los Angeles. The purpose of this investigation is to aid customers in restaurant selection and inform them what cuisine prices to expect when browsing restaurant ratings. 50 restaurants from each city were selected through stratified and systematic sampling. In the first part of the study, Spearman Rank Correlation Coefficient tests were conducted to test the correlation between restaurant ratings and prices. The tests show a significant positive relationship. Moreover, regression lines were constructed to display the correlation between the two variables. Customers may use these regression lines to estimate the average prices of a meal given the restaurant's rating. In the second part of the project, two-sample t-tests were used to make comparisons between Beijing and Los Angeles. These two cities were selected due to their similar urbanization, modernization, and income levels. Results show significant discrepancies, indicating that restaurants in Los Angeles have higher ratings and prices than those in Beijing. This difference can be attributed to several factors, including labor cost, food culture, and consumption.

Keywords
Restaurants, Prices, Rates, SRCC, Two-sample t-tests

1. Introduction
The idiom "Bread is the staff of life" demonstrates the significance of scrumptious meals to people's lives. With the development of technology and the growth of food industries in the 21st century, people are starting to rely on social media for their choices of dining places. As a result, reviews, ratings, and pricing became essential assets for restaurants. Several studies have examined how reviews influence consumer restaurant selection (Neflike 2017). Others have investigated the determinants of a restaurant's reputation: menu prices (Dong Hong *et al.* 2018). However, most of these studies are restricted to a particular country and use rough evaluations of ratings and price levels (on scales of 1 to 5). This study will use detailed data to examine the correlation between restaurant ratings and prices and compare these statistics between cities with different backgrounds.

2.Methods
The data for this project -- ratings and prices of restaurants -- were collected from an application named DianPing. Stratified, combined with systematic sampling, was used to select the 50 restaurants at each location.

2.1 DianPing

DianPing is the largest interactive platform in China that provides information for dining, shopping, and entertaining. DianPing has more than 200 million users, 17 million reviews, and 14 million business records, making it a reliable source to refer to. DianPing is chosen over Yelp because, although not being as universally used as the latter, DianPing has more detailed information for ratings and average meal prices. Ratings, from 0 to 15, are computed by summing the separate ratings of three indicators: taste, service, and dining environment, each on a scale from 0 to 5. Average meal prices are calculated automatically by the application through its records of expenditures at each restaurant.

2.2 Hypotheses

For the first part of the project, the null hypothesis is: that there is no relationship between cuisine prices and rates of restaurants. The statistical alternative hypothesis is: that there is a positive relationship between the two variables, that the higher the rating of a restaurant, the more expensive a meal would cost. We expect the alternative hypothesis to be true because higher ratings often reflect better quality, which is associated with higher prices.

For the second part of the project, the null hypothesis is: that there is no discrepancy between the ratings and prices in Beijing and Los Angeles. The statistical alternative hypothesis is: that Los Angeles has higher restaurant prices and ratings than Beijing. We make this assumption because, according to NUMBEO (2022), “Restaurant Prices in Los Angeles, CA are 152.66% higher than in Beijing.” Thus, if there is a relationship between restaurant prices and ratings, both statistics should be significantly higher in Los Angeles.

2.3 Data Collection

50 restaurants in Beijing (Figure 1) and 50 in Los Angeles (Figure 2) were sampled. Since restaurant prices are highly dependent on their type -- fast food, foreign food, classical cuisine, *etc.* -- sampling was stratified upon the categories. The relative proportions of the subgroups were maintained, avoiding under or over-representation of any category. Within each subgroup, systematic sampling was implemented by randomly generating a number between 1 and 10, and starting from that number, selecting every 10^{th} restaurant on the list from DianPing.

By keeping the samples citywide (rather than nationwide), confounding variables *within* each sample, such as regional disparities, dietary habits, and consumption levels would be controlled. This boosts the accuracy of the first part of the project: separately analyzing restaurants' rates and cuisine prices in each location. By selecting Beijing and Los Angeles as the two sites, confounding differences *between* the two cities, such as income, modernization, and urbanization levels would be minimized. This improves the accuracy in measuring the discrepancies in consumption levels and food culture.

Beijing:			Price per Meal	
	Name	Rate (0-15)	RMB	Dollar
Fast Food	KFC	12.5	35	5.25
	McDonalds	12.3	23	3.45
	Yoshinoya	11.9	28	4.2
	Subway	11.8	30	4.5
	Ju'er Family	13.7	47	7.05
	Sangu Bold Cuisine	11.9	43	6.45
	Golden spicy stew	12	44	6.6
Western Rest.	TASTY	14.1	157	23.55
	Pizza Hut	12.9	91	13.65
	New York Restaurant	13.2	119	17.85
Chinese Rest.	Grandma's	12.7	75	11.25
	Lady Goose	13.6	133	19.95
	Nanjing Stall	14.1	74	11.1
	Jinfu Salt Gang	14.6	132	19.8
	Dian Da Chi	14.5	167	25.05
	Little Pear Soup	14	97	14.55
	Tai Er	14.2	106	15.9
	Meizhou Dongpo	13.4	112	16.8
	Yunhai Dish	13.2	103	15.45
	Green Tea Restaurant	13.6	87	13.05
	Wong Kee Wong stew pot	12.7	97	14.55
	Xibei Oat Noodle	13.9	120	18
	Tai Xiang Mi	13.6	172	25.8
	Meizhou XiaoChi	13	46	6.9
	Da Ya Li	13.6	101	15.15
	Ninety-nine Yurts	14.5	271	40.65
	Fish head pancake	14.7	150	22.5
	Stone Cooking	14.5	96	14.4
	Thirteen Grilled Pork	14.6	106	15.9
	Long ago lamb kebab	14.5	127	19.05
	Yi Qun Fa Xiao	14.4	108	16.2
	Shrimp maker	14.1	123	18.45
	Ti Du	14.6	310	46.5
	Si Ji Min Fu	14.4	179	26.85
	Long Ren House	14.5	194	29.1
	Tangcheng Kitchen	13.9	115	17.25
	Green Courtyard Restauran	12.7	101	15.15
	Beiping House	14.4	128	19.2
	Old House Gate	11.1	67	10.05
	Wang Shun Ge	14.4	149	22.35
	Hua Jia Yi Yuan	14	148	22.2
Hot pot & BBQ	Hai Di Lao Hot Pot	14.5	138	20.7
	One spicy pot	12.5	98	14.7
	Sipu Sipu	13.2	76	11.4
	Shrimp Pot	13.5	96	14.4
	Minato Hot Pot	13.9	153	22.95
	Chongqing Gourd Hot Pot	14.1	129	19.35
Japanese Rest.	Ye Cai Village	14.1	200	30
	Niu Jiao Japanese BBQ Poı	14.4	227	34.05
	Ritan crudites	12.9	142	21.3
	Half Step Up Bistro	13.9	101	15.15

Figure 1. Raw Data for Beijing

Los Angeles:			Price per Meal	
	Name	Rate (0-15)	RMB	Dollar
Foreign	qin west chinese cuisine	13	96	14.4
	baekJeong Ktown	13.7	265	39.75
	feast at rieber	14.2	87	13.05
	downtown	9.3	85	12.75
	gypsy's trattoria italiano	12.2	87	13.05
	tacos tumbras a tomas	12	45	6.75
	Sun Nong Dan	14.2	299	44.85
	huli huli hawaiian grill	13.3	79	11.85
	daikokuya	11.3	105	15.75
Seafood and BBQ	bubba gump shrimp	14.3	181	27.15
	red lobster	14.6	372	55.8
	quality seafood	13.2	501	75.15
	park's BBQ	13.8	417	62.55
	the albright	13.7	276	41.4
	the kickin crab	14.2	323	48.45
	rock'N fish	14.4	260	39
	santa monica seafood	14.3	369	55.35
Fast Food and Pizza	in n out burger	14.4	44	6.6
	zpizza	14	83	12.45
	shake shack	14.2	128	19.2
	pink's famous ot dogs	12.6	124	18.6
	cheeseboard pizza	11.5	79	11.85
	dirt dog	12.9	70	10.5
	pieology pizzeria	13.7	75	11.25
Dessert and Cafe	the cheesecake factory	14.4	148	22.2
	the original pantry cafe	13.1	144	21.6
	lady m cake boutique	14.4	144	21.6
	Urth caffe	13.5	164	24.6
	cupertino	13.1	84	12.6
American Dishes	del frisco's grille	14.7	348	52.2
	blu jam (fairfax)	13.8	189	28.35
	hillstone restaurant (Emeritu	14.2	415	62.25
	eataly L.A.	14.5	234	35.1
	the butcher, the baker, the	14.2	254	38.1
	perch	14.5	270	40.5
	eggslut	13.3	93	13.95
	ruth's chris steak house	14	619	92.85
	bottega louie	14.4	284	42.6
	polo lounge	13.4	708	106.2
	republique	14.5	376	56.4
	71Above	14.3	924	138.6
	lawry's the prime rib	14.6	509	76.35
	water grill	14.5	367	55.05
	hard rock cafe	11.4	178	26.7
	dulce los angeles	13.5	76	11.4
	houston's	14.7	447	67.05
	the ivy	14.5	862	129.3
	fritto misto	14.7	226	33.9
	getty center restaurant	14.3	412	61.8
	howlin' Ray's	14	113	16.95

Figure 2. Raw Data for Los Angeles

2. Data and Statistical Analyses

In the first place, influential points -- outliers and leverage points -- were eliminated. After preconditioning data, Spearman Rank Correlation Coefficient (SRCC) tests were used to compute the correlation between restaurant ratings and prices. As the tests show strong affiliations, regression lines were drawn for the two variables. Next, two-sample t-tests were used to compare the restaurant ratings and prices between Beijing and Los Angeles.

3.1 Data Summary and Eliminating Outliers

After collecting the necessary information, all data were imported into an excel spreadsheet.
Before processing data, it is necessary to remove all outliers and leverage points. Observations are considered outliers if they have unusual (> Q3 + 1.5*IQR or < Q1 - 1.5*IQR) outcome values, and leverage points if they have unusual predictor values. These unusual values would significantly bias the correlation tests. In the Beijing sample, there are two outliers: restaurants with average costs of $40.65 and $46.5. Removing them, the observations have a mean rating of 13.59 and a mean cost of $17.562. Plotting the variables on a scatter plot (x-axis: Rate, and y-axis: Average meal price per person at the restaurant), it appears visually that there is a positive monotonic relationship (Figure 3). The x-axis ranges from 11 to 15; the y-axis, $3.45 to $46.5.

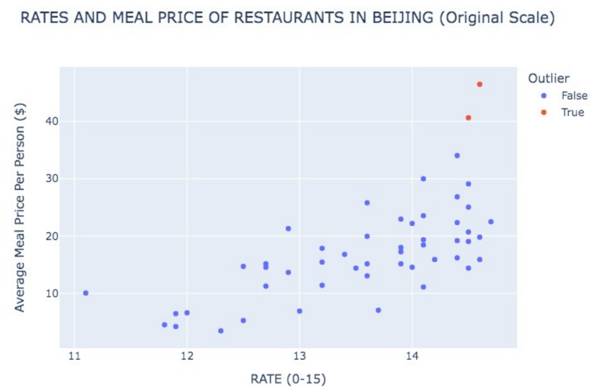

Figure 3. Scatter Plot of Rates and Prices of Restaurants in Beijing

The same steps are performed on the sample in Los Angeles. This sample has two outliers with meal prices of $138.6 and $129.3. There are also four leverage points with extraordinarily low ratings: 9.3, 11.3, 11.4, and 11.5. Removing these points, the sample has a mean rate of 14.2 and a mean price of $35.16. Plotting the points on a scatter plot, a positive correlation was observed (Figure 4). However, the observations are more sporadic than the Beijing sample, and the distribution is more skewed towards higher rates. Leaving out the outliers, the x-axis ranges from 12 to 15, and the y-axis, $6.6 to $106.2.

Figure 4. Scatter Plot of Rates and Prices of Restaurants in Los Angeles

3.2 Statistical Analysis: SRCC and Regression Line

To test the strength and significance of the relationship between restaurants' ratings and prices, Spearman Rank Correlation Coefficient (SRCC) tests were conducted. SRCC was chosen rather than the Product Moment Correlation Coefficient (PMCC) test because it is suitable for ordinal data, measures monotonicity, and is resistant to outliers. After running the test, the results show a r_s value of 0.68 and a p-value < 0.0001. Since the p-value is less than 0.05, and r_s is between 0.60 and 0.79, the correlation is positive, strong, and statistically significant. The results of this test provide evidence to reject the null hypothesis, supporting the alternative statistical hypothesis.

To make estimations of prices based on ratings, a regression line is drawn. However, the original data does not strictly follow a linear trend and is more variable toward large values. Therefore, when the regression line is drawn from the original data, the residual plot (Figure 5) is not randomly dispersed (having a fan-shaped pattern). To solve this issue, the dependent variable -- prices -- should be converted to a log scale.

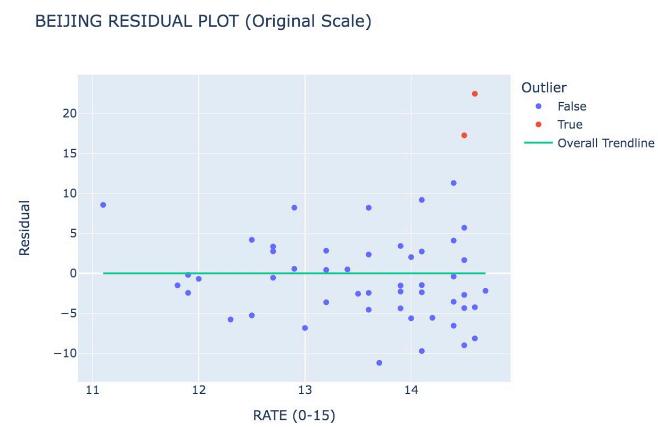

Figure 5. Residual Plot for the Correlation of Restaurants in Beijing

Reshaping the y-axis into a logarithmic scale, the observations now follow an approximately linear trend (Figure 6). Under this condition, the least-squares method can be used to construct the regression line. Results return a slope of 0.20 and a y-intercept of -1.53. Therefore, the equation for the line of best fit under logarithmic scale is:

$$\hat{y} = 0.20X - 1.53$$

In this formula, X is the rating, and ŷ is the estimated meal price in logarithmic form. For example, Pizza Hut has a rating of 12.9. Using this formula, ŷ is 1.05. Converting to standard form ($10^{1.05} = 12.02$), the regression line estimates a meal price at Pizza Hut to be $11.22. The reality is $13.65. This difference between the estimation and the real value is known as the residual.

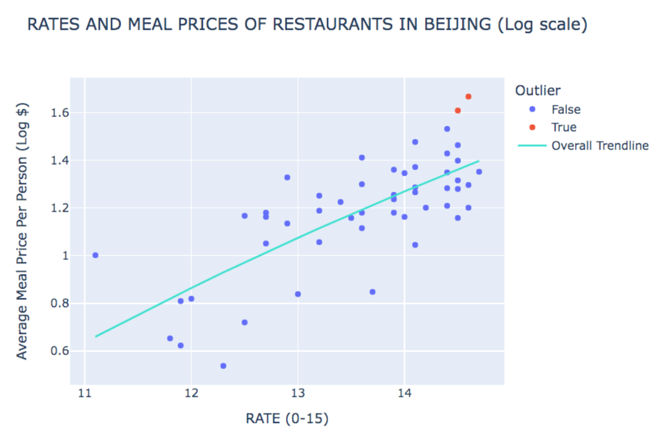

Figure 6. Rates and Prices of Restaurants in Beijing Adjusted in Log Scale

The residual plot in logarithmic scale is plotted (Figure 7). Now, the points are randomly dispersed and do not follow a pattern. Therefore, simple linear regression is appropriate for the data on a log scale. With this regression line, customers would be able to estimate the average meal price of a restaurant in Beijing given its rating.

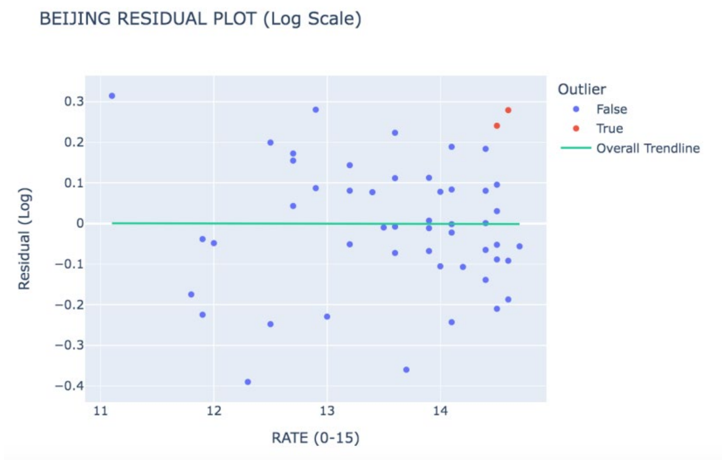

Figure 7. Residual Plot for the Correlation of Restaurants in Beijing Adjusted in Log Scale

Similarly, the Spearman Rank Correlation Coefficient is used for the Los Angeles sample (again, because it is suitable for ordinal data, measures monotonicity, and is resistant to outliers). This time, the test returns a r_s of 0.47 and a p-value of 0.0014. Since the p-value is less than 0.05, and r_s is between 0.40 and 0.59, so there is a moderate, positive, and significant correlation between ratings and prices for restaurants in Los Angeles. Again, the result rejects the null hypothesis and supports the alternative statistical hypothesis

Due to the same reasons -- that the original data does not have a linear correlation and is skewed towards larger values -- the regression line should be drawn from the logarithmic scale. Reshaping the prices into a log scale, the observations tend to follow a more linear trend (Figure 8). Under this condition, the least-squares method can be used to approximate the regression line. Results show a slope of 0.152 and a y-intercept of -0.615. Therefore, the equation for the line of best fit under the logarithmic scale is:

$$\hat{y} = 0.152X - 0.615$$

With this regression line, customers would be able to estimate the average meal price of a restaurant in Los Angeles given its rating.

Figure 8. Rate and Prices of Restaurants in Los Angeles Adjusted in Log Scale

3.3 Comparison Between the Two Locations using Two-sample t-tests

The second part of the project involves comparing the restaurant ratings and prices between Beijing and Los Angeles. The samples have close average ratings, with Beijing, 13.56, and Los Angeles, 13.9. To test whether this difference is significant, the two-Sample t-test was applied. Two-sample t-tests measure whether the unknown population means of two groups are unequal. The null hypothesis is that the two populations — restaurants in Beijing and Los Angeles — have the same average rates. The test returns a t-score of 2.01 and a p-value of 0.047. Since the p-value is less than 0.05 and the t-score is greater than the critical value (≈1.96), the null hypothesis is rejected. Los Angeles's restaurants have a significantly higher rate than Beijing's.

To visualize this difference, a box plot featuring the ratings of the two cities was drawn (Figure 9). Los Angeles's distribution is more compacted towards higher rates, with most observations clustered above 14.

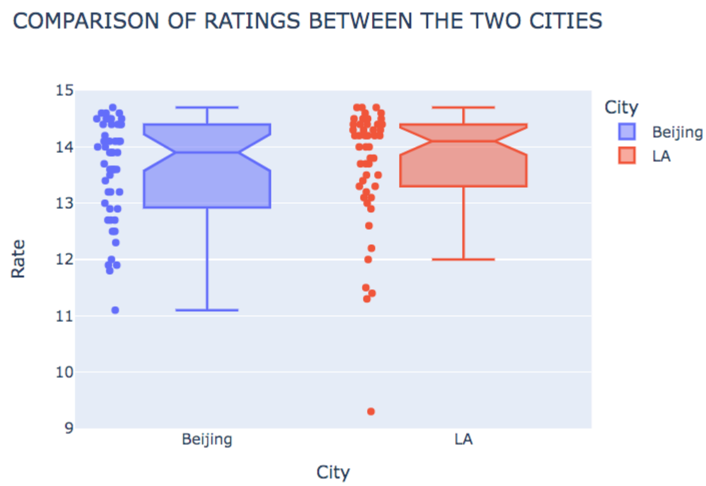

Figure 9. Box Plot Comparing the Restaurant Ratings Between Beijing and Los Angeles

To explain this discrepancy, scatter plots — color-coded by the category of the restaurants — are drawn (Figure 10 for Beijing and Figure 11 for Los Angeles). In Beijing, all fast-food restaurants have ratings lower than 12.5, while the other restaurants are evenly distributed above 13. On the other hand, fast food restaurants in Los Angeles have ratings approximately from 12.5 to 14.5, and most of the formal western restaurants are centered around 14.5.

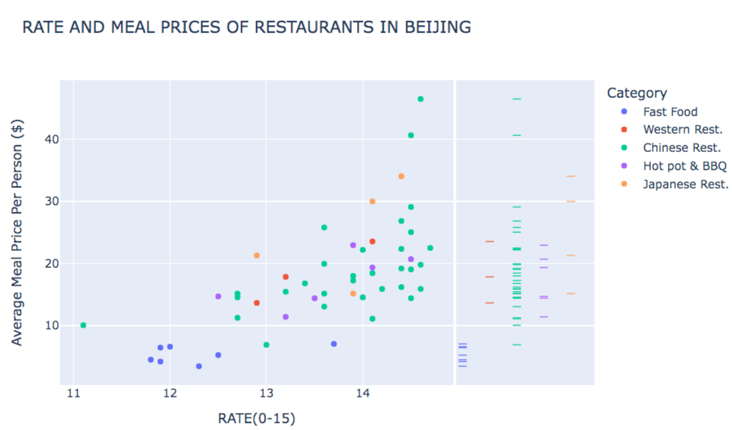

Figure 10. Scatter Plot of Restaurants in Beijing Colored by Restaurant Type

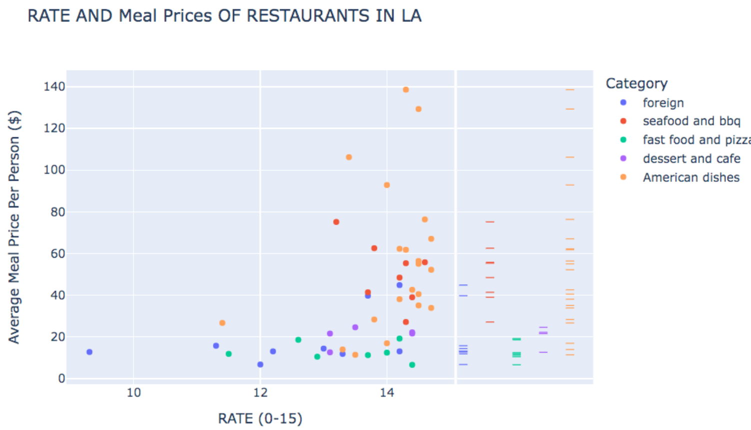

Figure 11. Scatter Plot of Restaurants in Los Angeles Colored by Restaurant Type

The scatter plots reflect the difference between Chinese and Western dietary cultures. "Chinese cuisine puts great importance on the aesthetic sense of the dish, and pursues the harmony in color, aroma, taste, shape, and utensil." (Lin 2000) Therefore, fast food, even if it were appetizing and affordable, might not receive high rates in China, as are ordinary restaurants without distinguishing features and exquisite plating. Moreover, there are significantly fewer fine dining restaurants in Beijing than in Los Angeles. There are only 10 Michelin restaurants (fine dining restaurants with the best foods and services) in Beijing, compared to 24 in Los Angeles, which may partially explain the larger cluster of ratings above 14 in Los Angeles compared to Beijing.

After examining the ratings between the two locations, a comparison of meal prices is carried out. The Los Angeles and Beijing samples have largely different cuisine prices, with Los Angeles averaging $35.16 per meal, and Beijing, $17.562. The two-sample t-test was used again to test the significance of this difference. The null hypothesis is that the two populations -- restaurants in Beijing and Los Angeles -- have the same average costs. The test returns a t-score of 5.59 and a p-value < 0.0001. Since the p-value is less than 0.05 and the t-score is greater than the critical value ($\approx$1.96), the null hypothesis was rejected. Therefore, Los Angeles's restaurants have a significantly higher prices than Beijing's.

Again, to visualize this difference, a box plot of the distribution of the rates of the two cities was drawn (Figure 12). The restaurants in Los Angeles have considerably higher costs and a larger range and variance.

Figure 12. Box Plot that Compares the Restaurant Prices Between Beijing and Los Angeles

The significant difference depicted in the box plots can be explained by discrepancies in dietary habits and labor costs. Chinese meals are comprised of two parts: staple food -- including rice, noodles, or steamed buns -- and vegetable and meat dishes. The meat dishes are generally cooked from pork, which differs from Western meals, where meat, especially beef, and other proteins are the main dish. Los Angeles, specifically, has a large number of seafood restaurants, which are generally more expensive than ordinary restaurants. Moreover, there is a large gap in labor costs between the United States and China. Wage and salary costs for private industry workers averaged $38.07 per hour in the United States (U.S. Bureau of Labor Statistics 2021). On the other hand, manufacturing labor costs in China were estimated to be $5.51 per hour (M. Szmigiera 2021). Both employer and material costs contribute to the significant difference in meal prices between Los Angeles and Beijing restaurants.

In terms of shape and dispersion, Beijing's distribution is approximately normal, with most observations centered around the median; on the other hand, Los Angeles' is heavily right-skewed. Unlike Beijing, Los Angeles has a great number of cheap fast-food restaurants, contributing to the aggregation of points in the first quartile -- from $6 to $15. Additionally, the Los Angeles distribution is more dispersed, partly due to the large variety of Western cuisines, from spaghetti and pizza to lobsters and grilled ribs.

4. Conclusion

In the first part of the project, two Spearman Rank Correlation Coefficient tests were conducted for the samples in Beijing and Los Angeles, and both were statistically significant. This result suggests that there is a positive, moderate to strong, relationship between restaurants' ratings and their cuisine prices. A potential explanation is that higher ratings are accompanied by increases in customer demand, which in turn boosts prices and profit.

In the second part of the project, two 2-sample t-tests were carried out to compare the statistics between the two cities. Both t-tests yield significant results, indicating that restaurants in Los Angeles have higher ratings and prices than in Beijing. There are a wide variety of reasons behind this gap, including differences in consumption, food culture, and labor costs.

5. Future Research

For this study, differences between restaurant rates and prices in Beijing and Los Angeles were proved significant. However, several factors might have influenced the accuracy of the result. Firstly, DianPing is a Chinese app used solely by Chinese subscribers. Data for both Beijing and Los Angeles were collected from this application to ensure uniformity of the rating benchmarks. However, it could cause other problems. Los Angeles locals are largely likely to have different tastes from Chinese tourists, so the ratings might be biased for the Los Angeles sample. Moreover, purchasing power was not taken into consideration. Although all measurements were converted into dollars, the same amount of money might have different values in different markets. To solve this problem, future research should compare cities in the same country. This controls purchasing power and turns the focus on the effects of the urbanization level. Additionally, there would no longer be biases in the user population. DianPing can still be used when investigating samples in China, while Yelp would be more suitable for Western countries.

Lastly, a sample size of 50 is not large enough to represent a population of several million restaurants. Since manually collecting data is tedious and inefficient, future investigations may expand sample sizes by incorporating datasets from Yelp. Obtaining comprehensive data, more factors such as location, advertising, etc., can be examined.

6. References

"Chinese Food Cultural Profile." *EthnoMed*, 21 Apr. 2020, ethnomed.org/resource/chinese-food-cultural-profile/.

"Cost of Living." *Cost Of Living Comparison*, www.numbeo.com/cost-of-living/comparison.jsp.

Cypris, written by Maya, et al. "Is China's Fine-Dining Market Lagging in the World Market?" *Innovated in China*, 26 Dec. 2017, innovatedinchina.cn/2017/12/15/china-fine-dining/.

"ECT News Releases." *U.S. Bureau of Labor Statistics*, U.S. Bureau of Labor Statistics, www.bls.gov/eci/.

Feng, Zhaoyin. "What I Learned Eating at 8,000 Chinese Restaurants." *BBC News*, BBC, 24 Nov. 2021, www.bbc.com/news/world-us-canada-59356176.

Gunden, Nefike. "How Online Reviews Influence Consumer Restaurant Selection." *Digital Commons @ University of South Florida*, digitalcommons.usf.edu/etd/6707/.

Luca, Michael. "Reviews, Reputation, and Revenue: The Case of Yelp.com." *SSRN*, 16 Sept. 2011, papers.ssrn.com/sol3/papers.cfm?abstract_id=1928601.

"Occupational Employment and Wages in Los Angeles-Long Beach-Anaheim - May 2020 : Western Information Office." *U.S. Bureau of Labor Statistics*, U.S. Bureau of Labor Statistics, 23 June 2021, www.bls.gov/regions/west/news-release/occupationalemploymentandwages_losangeles.htm.

Published by M. Szmigiera, and Mar 30. "Manufacturing Labor Costs per Hour: China, Vietnam, Mexico 2016-2020." *Statista*, 30 Mar. 2021, www.statista.com/statistics/744071/manufacturing-labor-costs-per-hour-china-vietnam-mexico/.

"Spearman's Rho Calculator." *Spearman's Rho Calculator (Correlation Coefficient)*, www.socscistatistics.com/tests/spearman/default2.aspx.

"Two-Sample T-Test." *JMP*, www.jmp.com/en_ch/statistics-knowledge-portal/t-test/two-sample-t-test.html.

Yim, Eun Soon, et al. "Determinants of a Restaurant Average Meal Price: An Application of the Hedonic Pricing Model." *International Journal of Hospitality Management*, Pergamon, 2 Mar. 2014, www.sciencedirect.com/science/article/abs/pii/S027843191400019X.

Zhu, Dong Hong, et al. "Good Discounts Earn Good Reviews in Return? Effects of Price Promotion on Online Restaurant Reviews." *International Journal of Hospitality Management*, Pergamon, 3 July 2018, www.sciencedirect.com/science/article/abs/pii/S027843191731023X.

Zimmermann, Kim Ann. "American Culture: Traditions and Customs of the United States." *LiveScience*, Purch, 21 Dec. 2021, www.livescience.com/28945-american-culture.html.