WHAT AFFECTS RESTAURANT REVIEWS?

Aleksandr Hovsepyan Lina Arustamyan Mariam Poghosyan

August 2020

College of Business and Economics
Instructor: Hrant Davtyan
Course: Data Scraping

Abstract

With the increasing role of tourism and hospitality industry the reviews on TripAdvisor are of great importance for a country like Armenia. When choosing a country to visit tourists look at some specific factors and one of them is national cuisine. The objective of this study is to examine the factors affecting customer reviews. With this objective in mind we decided to scrape information about the top restaurants in Yerevan from TripAdvisor. Even though we do not get significant results because of the specific nature of data, this study leaves room for contemplation for people who want to attract more tourists to Armenia, especially restaurant owners.

Keywords: Armenia, Customer reviews, Restaurant, Scraping, TripAdvisor

Contents

1 Introduction

Nowadays, every company needs to be present in "online world" and to correctly position itself in a highly competitive environment. While deciding which country to visit or what to eat people don't ask their relatives or friends anymore, as they can simply do their own research, read dozens of online reviews and after that make their decisions.

The tourism and hospitality is one of the major industries in the world with an annual economic impact of around 6.5 trillion USD. When making a decision which country to visit people usually look at some key characteristics of the country, for example the ease of getting visa, the number of sightseeing places, the hotel prices and convenience, an of course the number and quality of restaurants. Restaurants and traditional cuisine are an important factor of a country's culture and national heritage, and while deciding which country to visit people usually go to TripAdisor and have a look at the list of restaurants and cafes in a specific country and their reviews.

The objective of this study is to understand the relationship between different restaurant attributes and the number of their reviews. The study will be a start-up point for further investigation and analysis.

2 Literature Review

In their paper "Social Media and Customer reviews of Indian restaurants" Rishi Surya and Sangeetha Gunasekar examine the factors that influence restaurant ratings. They take the data from TripAdvisor.com, for NCR Delhi, India for the year 2014. They study the impact of taste of food, value for money, service quality and atmosphere of the restaurants on the overall rating of the restaurant. They find that value for money and food taste have a greater influence on restaurant ratings, than the atmosphere and service quality.

Another research conducted by Xiangbin Yan, Jing Wang and Michael Chau "Customer revisit intention to restaurants: Evidence from online reviews" analyzes the revisit rate of customers. They used over 10000 restaurant reviews, collected from online life community in China. They conclude that satisfaction of food quality, price and value, service quality, and atmosphere are the antecedents of revisit intention of restaurants, where service quality is the determinant factor.

3 Data

The data used in the analysis was collected by scraping top restaurants in Yerevan from TripAdvisor website. For scraping this website, we used the following techniques: Scrapy framework, Regular expressions, CSS approach, and Yandex Maps API, supported by classes and functions. After cleaning the data, we did the descriptive statistics and regression analysis to find out what affects number of reviews. We scraped the following information about restaurants; the names of the restaurants, ratings, number of reviews, cuisine type, special menu requirements, price range of menu and addresses.

Dependent Variable

The dependent variable in this model is the number of reviews. The data ranges from 10 to 1156. We decided to take the natural logarithm of reviews as the numbers for it are higher compared with the independent variables.

Independent Variables

One of the independent variables used in the model is restaurant ratings. The data ranges from 3 to 5 stars. The coefficient for this variable is expected to be negative, meaning that with an increase in rating the number of reviews decreases. The logic is that usually when people are not satisfied with the service, they tend to write reviews. However, when they are satisfied not everyone is ready to leave a review.

Another independent variable is the price range of the menu in dollars. We decided to split the price range into two categories: minimum and maximum price. The sign of the coefficient for these variables is expected to be positive, meaning that a higher price may lead to more number of reviews.

When people pay a high price for a special product or service, they generally have higher expectations and they are more likely to leave a review about the restaurant they have visited.

The next variable is distance, which measures the distance of the restaurant from the Republic Square. Firstly, using Yandex Maps API we got the longitude and latitude data for the restaurants, and afterwards we have calculated the distance. We expect that the restaurants which are more far from city center have less visitors, and as a result they will have smaller number of reviews.

Lastly, we use two dummy variables. The first one is the cuisine type used to describe whether or not the restaurant offers Armenian food. Cuisines that do not offer Armenian food were marked "1", and the rest were marked "0". The coefficient for the dummy variable is expected to be negative, meaning that restaurants which do not offer Armenian food will have less reviews compared with the ones that offer Armenian food. Tourists usually look for a restaurant which offers traditional cuisine that is typical to its culture and history, so they will be eager to write reviews about their experience.

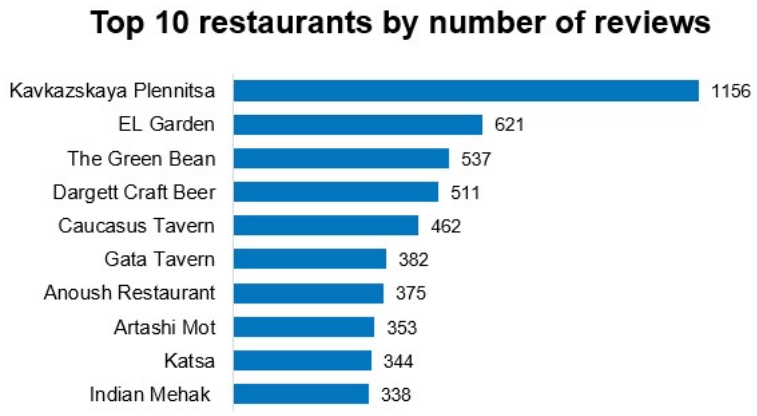
The second dummy variable is special diet availability. We divided restaurants into two categories; the ones that offer vegetarian food and the ones that do not. Cuisines that offer vegetarian food were marked "1", and the rest were marked "0". The effect of this dummy variable on the number of reviews is not so straightforward. On one hand people that have special diet requirements will be happy and ready to leave a review when they find a restaurant which meets their expectations, on the other hand we are not sure that the number of tourists who visit Yerevan and have special diet

requirements is so high.

The descriptive statistics for the variables are provided in the Appendix.

4 Results

We decided to have a look on top 10 restaurants by prices and reviews. We have dropped "Karas" as it had extraordinary high price. As we can see from the charts (for original charts made in Python see Appendix) the top 10 restaurants by price and by reviews are completely different, meaning that our initial assumption that the restaurants which have higher prices will have more reviews may be incorrect.



$$lnReviews = 2.92 + 0.07Rating + 0.01minprice$$
$$- 0.0024maxprice + 0.0002distance$$
$$- 0.12NotArmenian + 1.28Vegetarian$$

As we can see from the results the only significant variable is special diet requirement. The coefficient of the vegetarian is positive, meaning that as compared to restaurants that are not vegetarian friendly, restaurants that are vegetarian friendly usually have more reviews.

The other coefficients were not significant, however the signs of minprice and cuisine type variables were as expected.

5 Conclusion

Even though the R squared is 37% the regression results are not reliable, as the data is biased. This may be because of several reasons. First of all, when we look at the price ranges of some restaurants, they are not logical. Second, the vast majority of the restaurants have a rating of 5, and the data for ratings is not normally distributed.

Taking into account the fact that other researchers have done similar studies and got significant results, we can conclude that we did not get significant results because of the specific nature of our data. However, this may be a starting point for future research and other interesting variables can be added into the model.

A Appendix

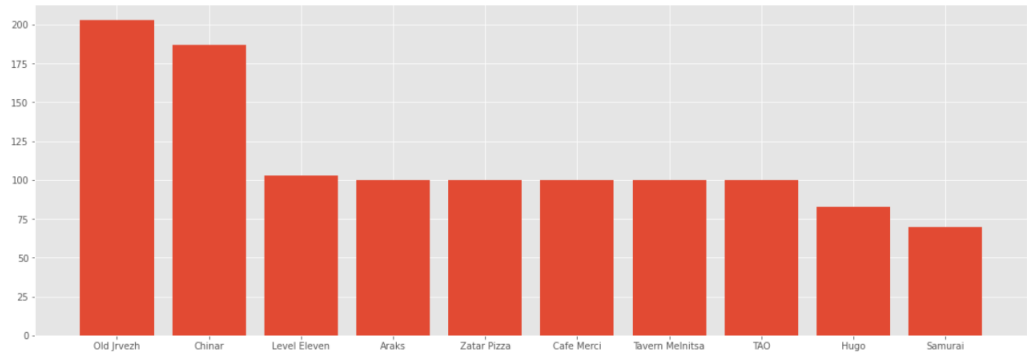


Figure 1: Most Expensive Restaurants

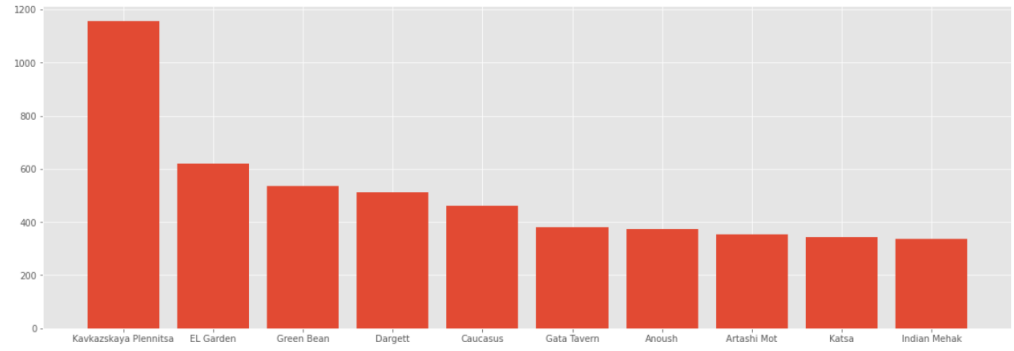


Figure 2: Top 10 Restaurants by Reviews

	rating	review	min_price	max_price	latitude	longtitude	distance	cuisine_NotArmenian	specdiet_Vegetarian
count	155.000000	155.000000	155.000000	155.000000	155.000000	155.000000	155.000000	155.000000	155.000000
mean	4.416129	85.651613	5.645161	27.445161	44.466405	40.292639	11.996247	0.561290	0.470968
std	0.451618	138.569669	4.489153	28.732113	0.573040	1.268378	123.848394	0.497838	0.500774
min	3.000000	10.000000	0.000000	2.000000	37.415685	40.147227	0.007023	0.000000	0.000000
25%	4.000000	18.000000	2.000000	10.500000	44.509663	40.177628	0.488979	0.000000	0.000000
50%	4.500000	38.000000	5.000000	20.000000	44.512546	40.181039	0.535542	1.000000	0.000000
75%	4.500000	85.500000	7.500000	30.000000	44.515466	40.186492	1.003822	1.000000	1.000000
max	5.000000	1156.000000	31.000000	203.000000	44.611092	55.966786	1541.006066	1.000000	1.000000

Figure 3: Descriptive Statistics

OLS Regression Results

Dep. Variable:	np.log(review)	**R-squared:**	0.379
Model:	OLS	**Adj. R-squared:**	0.353
Method:	Least Squares	**F-statistic:**	15.03
Date:	Sun, 02 Aug 2020	**Prob (F-statistic):**	2.19e-13
Time:	15:32:18	**Log-Likelihood:**	-192.95
No. Observations:	155	**AIC:**	399.9
Df Residuals:	148	**BIC:**	421.2
Df Model:	6		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	2.9240	0.699	4.181	0.000	1.542	4.306
rating	0.0707	0.155	0.456	0.649	-0.236	0.377
min_price	0.0139	0.016	0.857	0.393	-0.018	0.046
max_price	-0.0024	0.003	-0.921	0.358	-0.007	0.003
distance	0.0002	0.001	0.324	0.746	-0.001	0.001
cuisine_NotArmenian	-0.1256	0.143	-0.880	0.380	-0.408	0.157
specdiet_Vegetarian	1.2809	0.141	9.103	0.000	1.003	1.559

Omnibus:	2.569	**Durbin-Watson:**	1.311
Prob(Omnibus):	0.277	**Jarque-Bera (JB):**	2.425
Skew:	0.306	**Prob(JB):**	0.297
Kurtosis:	2.971	**Cond. No.**	1.29e+03

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 1.29e+03. This might indicate that there are strong multicollinearity or other numerical problems.

Figure 4: Regression Results

B References

Rishi Surya, Sangeetha Gunasekar (2015). Social Media and Customer reviews of Indian restaurants, Proceedings of the International Symposium on Emerging Trends in Social Science Research(IS15Chennai Symposium) ISBN: 978-1-941505-23-6

Yan, Xiangbin Wang, Jing Chau, Michael. (2013). Customer revisit intention to restaurants: Evidence from online reviews. Information Systems Frontiers. 17.10.1007/s10796-013-9446-5.