Customer Satisfaction and Sales Performance on E-commerce Platform

Thuy Vu

Edmonds College

CS 200 Capstone Project I

June 14th, 2024

Dr. Julio Garibay

Table of Contents

Executive Summary

As the internet has become an essential part of daily life, e-commerce has also become a profitable business model. There have been a large number of studies that were dedicated to acquiring insights into customers' attitudes on different aspects of e-commerce to gain and maintain customer satisfaction. Many researchers believe that customer satisfaction may affect the business. While customer satisfaction is certainly important as it informs strategic business decisions, it is unclear if customer satisfaction has a direct relationship with sales performance. This paper is dedicated to testing the relationship between customer satisfaction and sales performance using correlation analysis. The data used in this paper was collected and published by Olist Store, a large e-commerce platform in Brazil. The rating scores from the customers are used as a measurement of customer satisfaction, and the total sales of each product are used as a measurement of sales performance. The results show that there is a weak negative correlation between customer satisfaction and sales performance. In other words, customer satisfaction and sales performance do not always move in tandem. It can be concluded that customer satisfaction may not always be equal to customer loyalty.

Introduction and Literature Review

E-commerce

In the contemporary world, "internet has become the main technology for communication and inclusion" (Choudhary & Tanwar, 2023, as cited in Cunha et al., 2023). Hence, a new business form was introduced with the development of the internet: electronic commerce, or e-commerce for short. All the stages in e-commerce can be done online (Cunta et al., 2023). By definition, e-commerce is the process of transactions done electronically through the Internet, including buying, selling products, and providing customer service (Chintagunta et al., 2011, Guo et al., 2021, as cited in Cunta et al., 2023).

E-commerce has grown tremendously over the years (Liu et al., 2021). According to a report from Statista.com, e-commerce sales were estimated to reach 2.77 trillion US dollars in 2018 globally, which would make up 11.6 percent of the total retail sales of the year (Statista, 2018, as cited in Liu et al., 2021). With e-commerce's rapid growth, the industry also has developed innovative business models, such as the flash sale model, to boost sales performance.

The Growth of E-commerce in Brazil

In Brazil, the number of internet users has grown tremendously. According to CGI.br, there were around 94.2 million Internet users who would make up 55% of the number of citizens who were older than 10 (2014, as cited in Oliveira et al., 2017). Oliveira et al. stated that the National Broadband Plan which included expanding fiber internet and internet access has contributed to the growth of internet usage; as a result, consumers' shopping habits have changed, and online shopping has become extremely popular (2017). As reported by IBGE, compared with 2014, traditional retail sales increased 4.5% in 2015; meanwhile, e-commerce sales increased 16% in the same period (2015, as cited in Oliveira et al., 2017). About 30.7% of

the Brazilian population, or 61.6 million people, shopped online at least once in 2014 (E-bit, 2015, as cited in Oliveira et al., 2017). According to E-bit, 70% of customers fall in the range of 25 to 49 years old; 43% would have a monthly income of 790 US dollars, and 52% of the customers would have pursued at least one post-education degree (2016, as cited in Oliveira et al., 2017).

In 2014 alone, Brazilians made 103.4 million online orders which cost 9.42 billion US dollars in total with an average value of 91 US dollars for each order; popular categories of online products in Brazil include fashion and accessories, cosmetics, perfumery, personal and healthcare, appliances, books and magazines, telephony and mobile, IT, interior design, electronics, sporting goods, and toys and games (Oliveira et al., 2017).

Literature Review

Glowa stated that "conventional wisdom suggests that monitoring customer satisfaction makes good business sense for two reasons. The first is that satisfied customers are likely to continue to buy from/or continue to conduct business with a firm, while dissatisfied customers are likely to take their business elsewhere. The second is that satisfied customers tell others about their positive experience, while dissatisfied customers may tell even more people about their negative experiences" (2014). It was reported that between the early 1970s and 1990, there were over 15,000 articles on customer satisfaction that were published (Peterson & Wilson, 1992, as cited in Glowa, 2014).

With the exponential growth of e-commerce, there have been multiple studies about it. Oliveira et al. studied online shoppers' attitudes toward automatic delivery stations using a multinomial logit model (2017). Cunha et al. studied the factors that affect customer satisfaction before and after the COVID-19 pandemic using multiple statistical methods including correlation

analysis (2023). Both of these studies serve one purpose - to gain an understanding of customers' attitudes and maintain customer satisfaction. Based on these studies, a series of research questions and hypotheses are presented here. They are concerned with customers' attitudes, customer satisfaction, and overall sales performance. The research question is: what is the relationship between customer satisfaction and sales performance? This study uses the dataset from Olist. The correlational analysis was selected to determine the relationship between the variables. According to Bewick et al., correlation quantifies the strength of the linear relationship between a pair of variables (2003). Therefore, correlation analysis is a suitable statistical method for this paper.

Olist is a large Brazilian "e-commerce platform that connects small and medium-sized businesses to customers" with a single contract; "the platform operates as a marketplace, where merchants can list their products and services and customers can browse and purchase them online" (Meta, 2024; Olist & Sionek, 2021). As this paper studies the relationship between customer satisfaction and sales performance of Olist, Olist can gain some benefits from the insights about the relationship, and this study may also provide some insights about the relationship between customer satisfaction and sales performance in general. From those insights, it can inform better business decisions.

Methodology

Data Source

The dataset was obtained from Kaggle; it is a Brazilian public dataset regarding e-commerce. The dataset was created by Olist Store, and it contains information on one hundred thousand orders of Olist from the year of 2016 to 2018 (Olist & Sionek, 2021). The

comprehensiveness of this dataset allows multidimensional views from “order status, price, payment, freight performance to customer location, product attributes, and finally reviews from the customers"; an extra dataset of geolocation which includes zip codes, latitude, and longitude was also attached in this dataset for further analysis (Olist & Sionek, 2021).

As this dataset is a real commercial dataset, the dataset has been anonymized by the authors due to ethical issues. Any information that consists of personal information and could violate the privacy of any party was replaced with other texts, such as names of Game of Thrones great houses (Olist & Sionek, 2021).

For research purposes, only two out of nine CSV files are used as these CSV files contain the essential information for this paper. The two CSV files are named “olist_order_items_dataset” and “olist_order_reviews_dataset”.

Research Hypothesis

The research hypotheses of this paper concern the correlation between customer satisfaction measured by customer rating scores and sales performance. The null hypothesis is that there is no correlation between customer satisfaction and sales performance, and the alternative hypothesis is that there is a correlation between customer satisfaction and sales performance.

Independent Variable

The independent variable in this study is customer satisfaction with the products. It is measured based on customers’ rating scores for each order. The rating scores of the orders consist of five distinct values which are 1, 2, 3, 4, and 5. All of the ratings are whole numbers. The rating scores for the products are extracted from the rating scores of the orders and calculated for the average rating scores for each product. The average rating scores of the

products are used as the measurement of customer satisfaction for each product, and they are expressed as decimal numbers from the range of 1 to 5.

Dependent Variable

The dependent variable in this study is the sales performance. This variable is calculated based on the occurrences of each product in all of the orders, which is also understood as the total sales for each product. The sales performance is expressed as whole numbers. The range of the sales are from 1 to 527 sales.

Data Cleaning, Manipulation, and Analysis

The tool used for data cleaning, manipulation, and analysis in this paper is RStudio, a powerful tool in the field of data analytics.

For data manipulation, the files were imported into RStudio. Then, the counts or the sales as well as the average rating scores of each product were calculated and displayed in two different tables. After that, these two tables were joined together. As a result, there are three columns in the final table which are “product_id”, “total_sales”, and “average_rating”. For data cleaning, all rows that contain empty cells were omitted. In total, 162 rows, or approximately 0.5%, of the dataset were omitted, and the remaining 32,789 rows remained in the dataset. The dataset is clean and ready to be analyzed. The statistical method used for this paper is correlation analysis to measure the strength of the relationship between customer satisfaction measured by average rating scores and sales performance.

Table 1

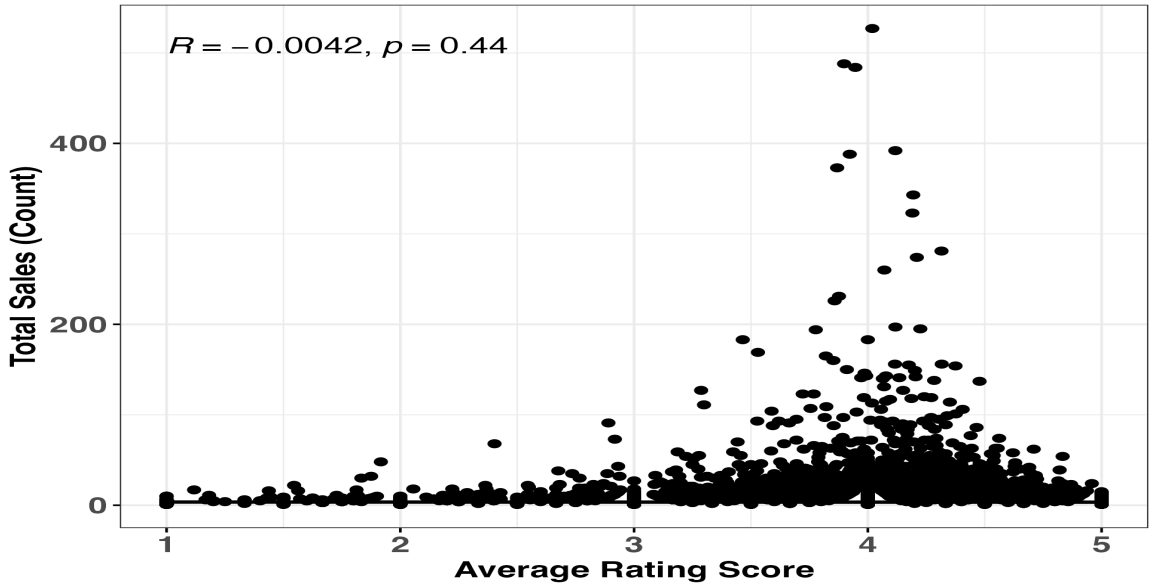
Descriptions of Data

Data Description	Total Sales	Average Rating
Min	1.000	1.000
1st Quartile	1.000	3.611
Median	1.000	4.500
Mean	3.429	4.048
3rd Quartile	3.000	5.000
Max	527.000	5.000

Note. The table describes the characteristics of the dataset used for correlation analysis, which includes the minimum value, the first quartile, the median, the mean, the third quartile, and the maximum values of each column in the dataset. Compiled by the researcher.

Figure 1

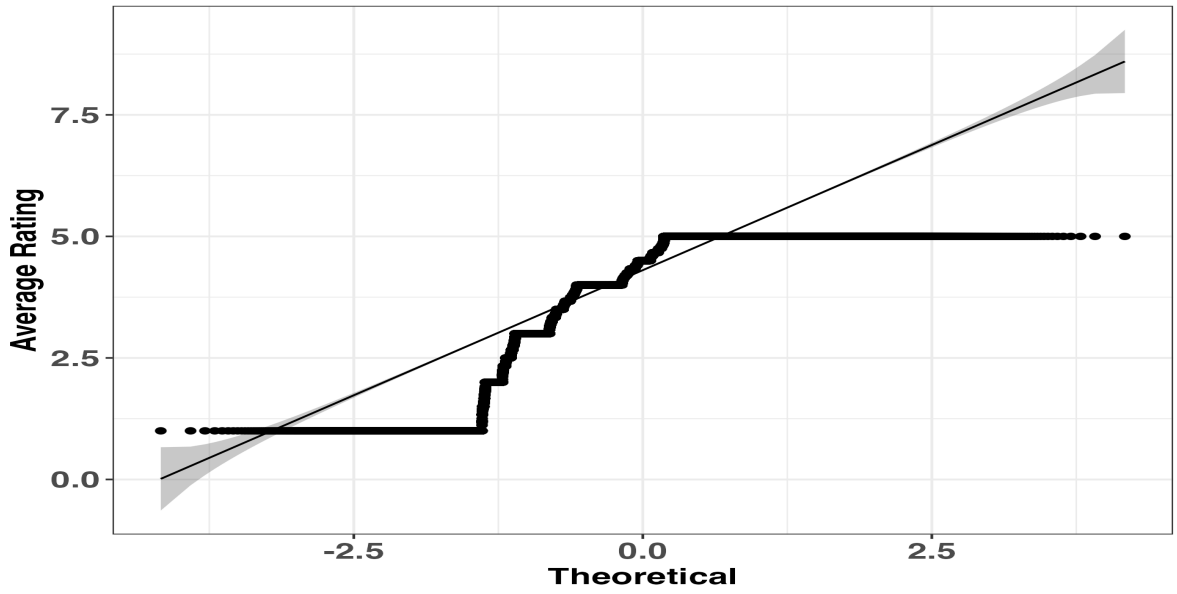
The relationship between average rating scores and total sales.



Note. Owned by the researcher.

Figure 2

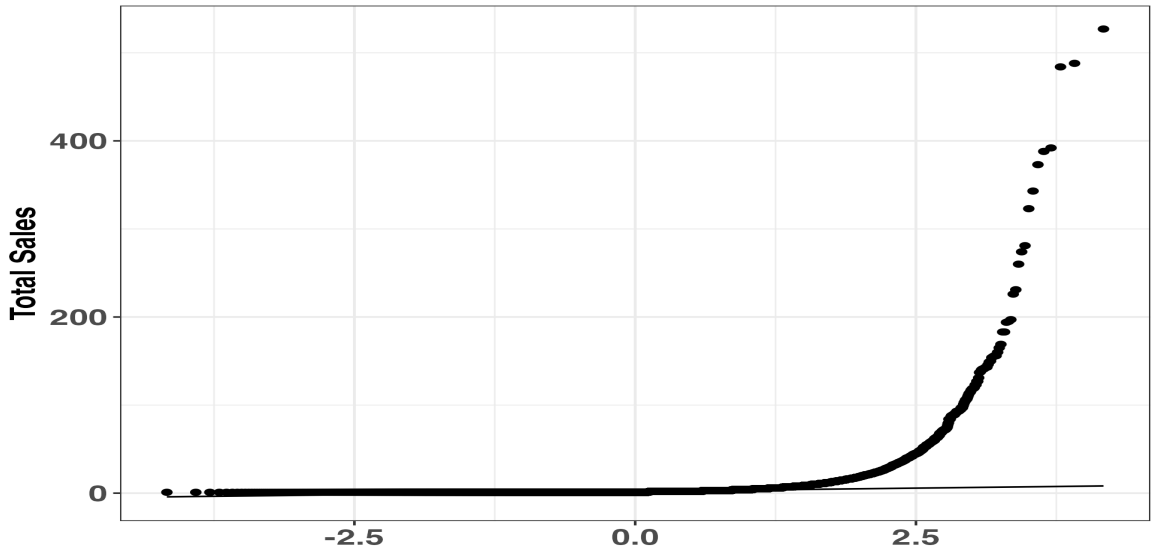
Quantile-Quantile Plot of Average Rating Scores



Note. Owned by the researcher.

Figure 3

Quantile-Quantile Plot of Total Sales



Note. Owned by the researcher.

To check for the test assumptions before doing a correlation analysis, some graphs were created, as shown in Figure 1, Figure 2, and Figure 3. As observed from Figure 1, the relationship between average rating scores and total sales of the products does not look very linear. As observed from Figure 2 and Figure 3, both total sales and average rating scores may not come from normal distributions. As suggested by the data in Table 1, there might be some skewness in the distribution for both variables. Therefore, Spearman's correlation test will be used in the paper for further analysis. The alpha level (or the significant level) is set to 0.05 or 5%.

Conclusion

After running the Spearman's correlation test, the Spearman rank-order correlation coefficient or rho is approximately -0.19957, and the p-value is smaller than 2.2e-16, which is even smaller than the significant level of 0.05 and very close to zero. The rank-order correlation coefficient suggests a weak negative correlation between average customer rating scores and total sales, which means when total sales increase, the customer rating scores slightly decrease and vice versa. The significant level suggests that the null hypothesis that there is no correlation between customer satisfaction and sales performance is rejected. In conclusion, it is statistically significant that there is a weak negative correlation between customer satisfaction measured by product ratings and sales performance. In other words, customer satisfaction and sales performance do not exactly move in tandem.

Limitations and Recommendations for Future Research

There were several limitations of this study. The first limitation was that there was sampling bias. Sampling was limited to only one e-commerce platform. The dataset was selected

because of its convenience and availability. Therefore, while the analysis might provide some insights for Olist, it might not be generalizable for the population.

The second limitation was that product ratings were extracted from rating scores of whole orders while order rating scores might be affected by other factors such as service, price, and delivery speed as well, not just customer satisfaction with the products. Besides the rating scores, there were some review comments from the customers as well. However, due to the lack of linguistic knowledge of Portuguese, the researcher could only use the scores for analysis. For future research recommendations, some text analysis might be helpful to gain valuable insights into customer satisfaction. Surveys would also be a better method to gain insights into customer satisfaction, in comparison to rating scores.

The third limitation was that the data was collected from the year of 2016 to 2018. There might have been changes from that point as that period of time was before the COVID-19 pandemic. New data collection might be necessary for future studies.

References

Bewick, V., Cheek, L., & Ball, J. (2003). Statistics review 7: Correlation and regression. *Critical Care*, *7*(6), 451–459. https://doi.org/10.1186/cc2401

Cunha, B. M., Lettieri, C. K., Cadena, G. W., & Pereira, V. R. (2023). Analyzing the influence of covid-19 on the e-commerce customer's retail experience in the supermarket industry: Insights from Brazil. *Logistics*, *7*(3), 53. https://doi.org/10.3390/logistics7030053

Glowa, T. (2014). *Measuring customer satisfaction exploring customer satisfaction's relationship with purchase behavior*. BookBaby.

Liu, X., Zhou, Y.-W., Shen, Y., Ge, C., & Jiang, J. (2021). Zooming in the impacts of merchants' participation in transformation from online flash sale to mixed sale e-commerce platform. *Information & Management*, *58*(2), 103409. https://doi.org/10.1016/j.im.2020.103409

Olist, & Sionek, A. (2021, October 1). *Brazilian e-commerce public dataset by Olist*. Kaggle. https://www.kaggle.com/datasets/olistbr/brazilian-ecommerce/data

Olist, the new e-commerce unicorn. Meta. (2024, January 22). https://metait.us/cases/olist/

Oliveira, L. K., Morganti, E., Dablanc, L., & Oliveira, R. L. (2017). Analysis of the potential demand of automated delivery stations for e-commerce deliveries in Belo Horizonte, Brazil. *Research in Transportation Economics*, *65*, 34–43. https://doi.org/10.1016/j.retrec.2017.09.003