

Research on the Influencing Factors of Book Star Ratings -Taking Amazon Platform as an Example

Rong Song*

Faculty of Science, North China University of Technology, Beijing, 100000, China

* Corresponding Author Email: 2015141094176@stu.scu.edu.cn

Abstract. Books are one of the main ways for people to acquire knowledge, so usually reading books is very important. Reading books can not only enrich people's knowledge base but also make them more cheerful and help people resist loneliness. With the development of the times, reading has become an indispensable part of people's lives, people have gradually shifted from buying books from bookstores to buying books from online shopping platforms. This study utilized data set information from Amazon Kindle books found on the Kaggle website and filtered out 3933 suitable data. Then, this article will use frequency analysis and linear regression analysis to analyze Price, Reviews, Editor's Pick, Kindle Unlimited, and Good Reads Choice. The results of this research are Editor's Pick, Kindle Unlimited, and Good Reads Choice, recommended by editors, have a substantial positive impact on star ratings. The price will have a significant negative effect on star ratings. However, Reviews will not have an impact on Star ratings Based on the analysis obtained from research, some suggestions can be made for people to choose high-star-rated books.

Keywords: Book star; star ratings; influencing factors.

1. Introduction

Books are one of the main ways for people to acquire knowledge, so reading books is a very important thing. Reading to Learn is a department that explores learning from texts, including structured issues and practices and implications for practitioners, policymakers, and researchers [1]. Reading is frequently a means of learning something new, whether it's a fact from a textbook, a story from a novel, or instructions from a manual [2]. Evidence showed that young people are greatly aided by suggestions that they derive from the contents of their reading [3]. Reading comprehension is necessary for academic success. Cognitive resilience, attention, and empathy can be significantly improved through reading [4]. Romances, vampire stories, horror novels, dystopian fiction, and fantasy provided their readers with four distinct types of pleasurable experiences: the pleasurable of play, the intellectually pleasurable, the socially pleasurable, and the work-pleasant [5].

Because people have a great need for reading, books have a huge market. Meanwhile, Publisher revenues are still dominated by physical books, with 69.5% of all net earnings in 2013 coming from them [6]. With the development of the times, people are gradually shifting from buying books from bookstores to buying books from online shopping platforms. Many businesses sell solely online, even though an online non-sale incurs little monetary cost. Additionally, retailers with both an offline and online presence can reap the benefits of increasing their online sales. Online reviews on Amazon or Good reads, liking on Facebook, or tweeting are crucial as much of the market moves online [7].

In recent years, search engines have become an essential tool for promoting and selling products. The relevance of a result to a query can be instantly determined by information search engines like Google based on the ranking of search results [8]. Before making an online purchase decision, the vast majority of online consumers will search for information about the product they want to purchase. Online reviews are one of the most important pieces of information that help consumers make purchasing decisions [9]. Before making a purchase decision, consumers often look for other people's thoughts on a product. Examine the consumer ratings shared by past patrons and ask their family or friends for recommendations. Forrester Research conducted an online survey, and 64% of those who responded wanted to read reviews and ratings on the e-commerce websites they visited [10].



Nowadays, many products have user generated content that has become a form of online consumer reviews and ratings. The qualitative assessment of the product experience of an online reviewer is then utilized to guide and influence the quantitative evaluation of the product [11].

Consumers are particularly interested in online consumer reviews, which are a form of electronic word of mouth (eWOM), because of their impact on their purchasing decisions. Buying products from an online retail market can be challenging for consumers who rely on seller information when making purchase decisions. Thus, consumers look to other consumers' online reviews for more detailed product information. Consumer-oriented information provides indirect product experiences that aid in making purchase decisions. Information that is focused on consumers may have more credibility and relevance than information that is focused on sellers. Therefore, online consumer reviews can be utilized as an instrument to establish consumer trust [12]. Retail websites provide consumers with a place to express their evaluations of products, including digital star ratings (usually ranging from 1 to 5 stars) and consumer reviews of product usage. For a long time now, leading online retailers like Amazon.com have been allowing consumers to provide product reviews, and other retailers have recently started offering this option to their customers. Online customer reviews typically have numerical star ratings that range from one to five stars. Thus, it is important to have a clear understanding of the factors that influence book star ratings.

2. Methods

2.1. Data Source

Kaggle's website provided the data set used in this study. The content of this data set is information about Amazon Kindle books. The Amazon Kindle book data set found was obtained in October 2023. This data set contains data from many Kindle books, with 10 influencing factors in the data. After organizing this data, the author obtained data on 3933 books. The quantity and quality of data are sufficient to assist people in conducting research.

2.2. Indicator Introduction

In the analysis, factors such as Price, Reviews, Editor's Pick, Kindle Unlimited, and Good Reads Choice were selected to explore their impact on book star ratings. Frequency analysis and linear regression analysis will be used to analyze these factors.

2.3. Method Introduction

After preliminary selection and simple analysis of the data, this study used linear regression analysis to conduct an in-depth analysis of appropriate factors in these data. In linear regression analysis, books priced between \$0-5 will be calculated as \$5, books priced between \$5.01-10 will be calculated as \$10, books priced between \$10.01-15 will be calculated as \$15, and books priced above \$15 will be calculated as \$20. The number of comments for books with 0-100 comments will be calculated as 100, the number of comments for books with 101-1000 comments will be calculated as 1000, the number of comments for books with 1001-10000 comments will be calculated as 10000, and the number of comments for books with more than 10000 comments will be calculated as 100000. 0 represents false and 1 represents true.

3. Results and Discussion

3.1. Descriptive Analysis

From Table 1, it can be seen that frequency analysis was first applied to the overall data, and it can be seen that all data are not missing in the five factors of Price, Reviews, Editor's Pick, Kindle Unlimited, and Good Reads Choice.

Table 1. Frequency analysis

Name	Option	Frequency	Percentage (%)	Accumulated percentage (%)
Price	5	936	23.80	23.80
	10	933	23.72	47.52
	15	1437	36.54	84.06
	20	627	15.94	100.00
Reviews	100	1414	35.95	35.95
	1000	1416	36.00	71.96
	10000	883	22.45	94.41
	100000	220	5.59	100.00
Editor's Pick	0	2220	56.45	56.45
	1	1713	43.55	100.00
Kindle Unlimited	0	2751	69.95	69.95
	1	1182	30.05	100.00
Good Reads Choice	0	3918	99.62	99.62
	1	15	0.38	100.00
Total		3933	100.0	100.0

From Figure 1, it can be seen that the price of 936 samples is \$5, the price of 933 samples is \$10, the price of 1437 samples is \$15, and the price of 627 samples is \$20. Among the samples, the sample with a price of \$15 has the highest quantity, while the sample with a price of \$20 has the lowest quantity.

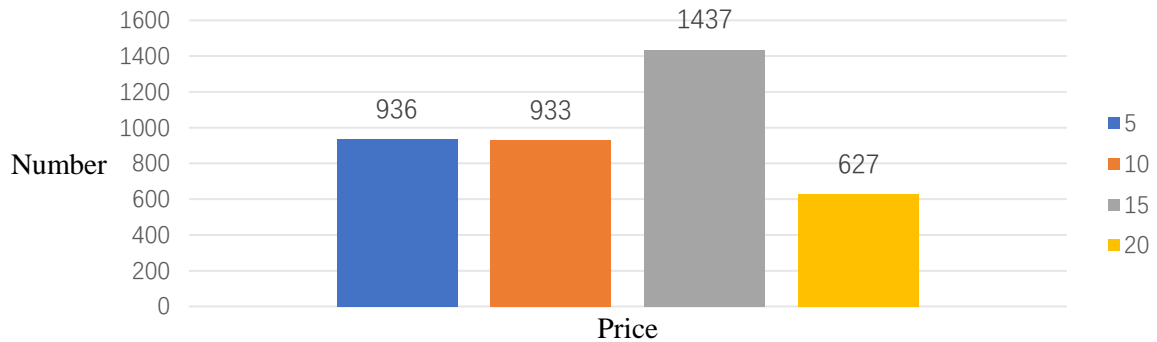


Fig. 1 Histogram of price

From Figure 2, it can be seen that 1414 samples have 100 comments, 1416 samples have 1000 comments, 833 samples have 10000 comments, and 220 samples have 100000 comments. Among the samples, the sample with 1000 comments is the highest, while the sample with 100000 comments is the lowest.

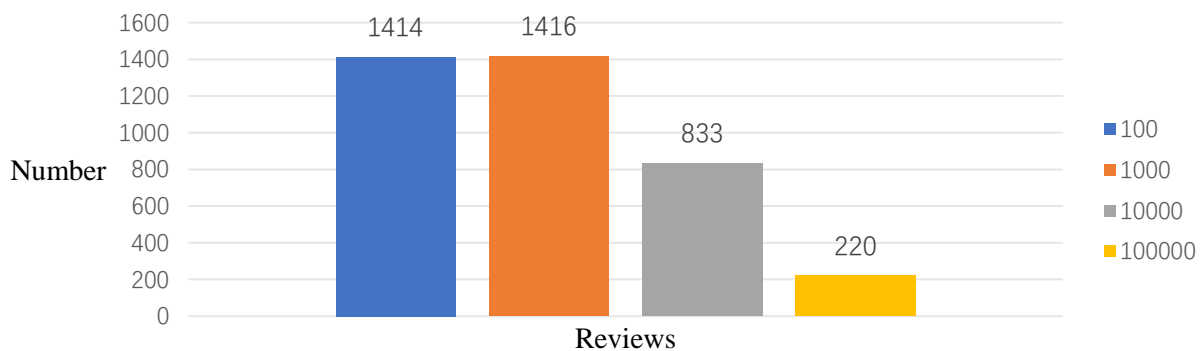


Fig. 2 Histogram of reviews

From Figure 3, it can be seen that the Editor's Picks of 2220 samples are "0", and the Editor's Picks of 1713 samples are "1". The number of samples with Editor's Picks of "0" is the highest. The proportion of samples with or without Editor's Picks is 0.77:1.

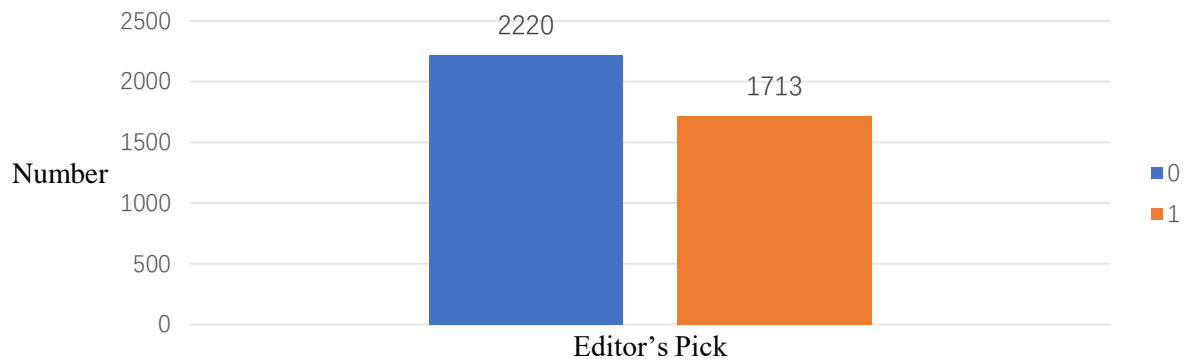


Fig. 3 Histogram of Editor's pick

3.2. Linear Regression Analysis

From Table 2, it can be seen that linear regression analysis is conducted, and the model formula is determined by taking into account Price, Reviews, Editor's Pick, Kindle Unlimited, and Good Reads Choice as independent variables and Star ratings as dependent variables., the model formula is:

$$\text{Star ratings} = 3.813 - 0.003 \times \text{price} + 0.000 \times \text{reviews} + 0.686 \times \text{Editors Pick} + 0.196 \times \text{Kindle Unlimited} + 0.225 \times \text{Good Reads Choice} \quad (1)$$

Table 2. Linear regression analysis

	Nonstandardized coefficient		Standardized Coefficient	t	p	Collinearity diagnosis	
	B	Std. Error	Beta			VIF	Tolerance
Constant	3.813	0.027	-	142.594	0.000**	-	-
Price	-0.003	0.002	-0.034	-2.087	0.037*	1.535	0.651
Reviews	0.000	0.000	0.021	1.547	0.122	1.105	0.905
Editor's Pick	0.686	0.017	0.644	41.093	0.000**	1.450	0.689
Kindle Unlimited	0.196	0.020	0.170	9.599	0.000**	1.861	0.537
Good Reads Choice	0.225	0.113	0.026	1.991	0.047*	1.024	0.977
R ²			0.335				
Adjust R ²			0.334				
F			F (5, 3927)=394.825, p=0.000				
D-W			1.171				

* p<0.05 ** p<0.01

The R-squared value of the model is 0.335, which means Price, Reviews, Editor's Pick, Kindle Unlimited and Good Reads Choice can explain the 33.5% change in Star ratings. The negative impact of Price on Star ratings can be seen by observing its regression coefficient value of -0.003 (t=-2.087, p=0.037<0.05) The value of Reviews in the regression coefficient is 0.000 (t=1.547, p=0.122>0.05), which indicates that Reviews have no effect on Star ratings. Editor's Pick has a significant positive impact on Star ratings, as evidenced by its regression coefficient value of 0.686 (t=41.093, p=0.000<0.01). Kindle Unlimited has a regression coefficient value of 0.196 (t=9.599, p=0.000<0.01), which shows a significant positive influence on Star ratings. Good Reads Choice has a significant

positive impact on star ratings, as indicated by its regression coefficient value of 0.225 ($t=1.991$, $p=0.047<0.05$).

4. Conclusion

This study used frequency analysis and linear regression to analyze the effects of Price, Reviews, Editor's Pick, Kindle Unlimited, and Good Reads Choice on star ratings, respectively. Through data analysis, this article has obtained some statistical results, which indicate that Editor's Pick, Kindle Unlimited, and Good Reads Choice, recommended by editors, have a major positive influence on the star rating. And Price will have a negative impact on Star ratings. However, Reviews will not have an impact on Star ratings. Based on the analysis obtained from research, some suggestions can be made for people to choose high-star-rated books. When choosing high-star books, people can choose books with Editor's Pick, Kindle Unlimited, and Good Reads Choice during the screening process, thereby improving the efficiency of selecting high-star books and helping people better participate in reading behavior. Meanwhile, future research can explore other factors that may affect book star ratings, in order to gain a more comprehensive understanding of the factors that affect book star ratings.

References

- [1] Armbruster, Bonnie B. Reading to Learn: Readings about Reading to Learn. *The Reading Teacher*, 1993, 598-600.
- [2] Weingarten Samuel. Reading Can Help Gifted Adolescents. *The Reading Teacher*, 1956, 219-225.
- [3] Merga K M, Moon B. The Impact of Social Influences on High School Students-Recreational Reading. *The High School Journal*, 2016, 99(2): 122-140.
- [4] Ritchey D K, Palombo K, Silverman D R, et al. Effects of an Informational Text Reading Comprehension Intervention for Fifth-Grade Students. *Learning Disability Quarterly*, 2017, 40(2): 68-80.
- [5] Panagiotelis, Anastasios, et al. From Amazon to Apple: Modeling Online Retail Sales, Purchase Incidence, and Visit Behavior. *Journal of Business & Economic Statistics*, 2014, 32: 14-29.
- [6] McDermott Ian, Erin C. Dunigan. Art Book Publishing: Past, Present, Future. *Art Documentation: Journal of the Art Libraries Society of North America*, 2013, 239-252.
- [7] Ghose, Anindya, et al. Examining the Impact of Ranking on Consumer Behavior and Search Engine Revenue. *Management Science*, 2014.
- [8] Kulwinder K, Tejinderpal S. Impact of Online Consumer Reviews on Amazon Books Sales: Empirical Evidence from India. *Journal of Theoretical and Applied Electronic Commerce Research*, 2021, 16(7): 2793-2807.
- [9] Sun, Monic. How Does the Variance of Product Ratings Matter? *Management Science*, 2012, 696-707.
- [10] Shrihari S, Raji S. Social Influence Effects in Online Product Ratings. *Journal of Marketing*, 2012, 76(5): 70-88.
- [11] Baek H, Ahn J, Choi Y. Helpfulness of Online Consumer Reviews-Objectives and Review Cues. *International Journal of Electronic Commerce*, 2012, 17(2): 99-126.
- [12] Mudambi M S, Schuff D. What Makes a Helpful Online Review? A Study of Customer Reviews on Amazon.com. *MIS Quarterly*, 2010, 34(1): 185-200.