Is Amazon giving preferential treatment to its own brands over competing brands?

Julia Gonzalez, Joaquin Ossa, Daniela Velez

April 11, 2024

1 Introduction

In the digital marketplace, Amazon has emerged as a global retail giant, offering a vast array of products to a diverse and discerning consumer base. With its extensive product listings and dynamic ecosystem, Amazon has become a focal point for both sellers and buyers alike. As the e-commerce market continues to evolve, the manner in which products are presented and promoted on the platform raises intriguing questions about fairness, transparency, and the impact on consumer choice.

This paper explores the dynamics of product positioning on Amazon, with a specific focus on the potential influence of the platform's own branded products and sponsored listings. Our research aims to investigate whether Amazon's proprietary products are consistently given elevated prominence, possibly overshadowing other products with superior sales performance and higher customer ratings.

In pursuit of these objectives, our methodology involves scraping and analyzing a diverse dataset of Amazon product listings, customer reviews and other variables that are relevant for the market place in positioning these items. This analysis researches into different product categories such as Kitchen supplies, Home Goods and Decoration, Fashion, Fitness and Outdoor, Pet Supplies, and Office and work products. By analyzing data scraped from Amazon's product listings, reviews, and other relevant variables, we seek to understand how product visibility in search results may be influenced by Amazon's self-brand products positioning.

2 Literature Review

This paper contributes to the recent literature about self-preferencing at Amazon Chiara Farronato (2023). The study uses real consumer searches who utilize Chrome browser for their online purchases and introduces an indicator to distinguish search results for Amazon-branded products. The findings reveal that only 19.7% of searches display at least one Amazon-branded product. The study consistently demonstrates that the presence of an Amazon brand is a strong predictor of enhanced search ranking. Notably, the influence of Amazon branding on search prominence is quantified to be about 30% to 60% of the impact attributed to paid sponsorships, highlighting the substantial effect of Amazon's branding on its products' search rankings.

We are contributing to these results by incorporating in our regression analysis a precise measure of sentiment scores derived from product reviews. By doing so, we obtain a more continuous quantification of the real value that consumers attribute to each listed product. We aim to better reflect the real costs incurred by consumers when certain products are presented more prominently than others, despite having similar value propositions.

To incorporate sentiment scores, we will be using TextBlob library. It is worth noting that Python has many libraries designed for sentiment analysis, among which TextBlob stands out due to its simplicity and efficiency compared to many others. The value of TextBlob lies in its ability to leverage into an extensive range of lexical resources, including NLTK, consolidating the sentiment analysis process. It employs a lexicon-based approach, which determines the sentiment of a sentence by analyzing the semantic orientation and intensity of each word. This method utilizes a pre-defined dictionary to categorize words as having a negative or positive connotation. Each word is assigned a sentiment score, and the overall sentiment is derived by calculating the average of these scores. This approach underscores the nuanced and comprehensive nature of TextBlob's sentiment analysis capabilities Loria (nd).

When conducting sentiment analysis with TextBlob, it leverages a pre-trained Naive Bayes classifier alongside a rule-based sentiment analysis method. This process produces an outcome known as polarity, which is measured on a scale ranging from -1 to 1, where values below zero denote negative sentiments, and values above zero represent positive sentiments Shah (2020).

3 Data Gathering

We obtained data for nearly six thousand distinct products across various categories drawn from twenty Amazon searched. For this process our initial approach consisted of using the web scraping library, *BeautifulSoup* to navigate and scrape information from

Amazon's product pages. However, after numerous attempts, we encountered challenges associated with the high volume of requests. Despite changing the user agent in the header section in each search or adjusting the frequency of requests and optimizing the structure of our requests to emulate more natural user behavior, our user agents were being blocked. As the impediment to extract the data persisted, the final solution was to use an API.

We integrated the Oxylabs API into our toolkit. This API offered a structured and efficient solution for navigating the challenges that we had faced in web scraping and it improved our data collection efficiency.

In the *'get response'* section of our script, we employ the HTTP *'POST'* method to communicate with the API endpoint, specifying the target URL and sending our search parameters as a JSON object in the request body. Following the submission of our request, we process the response to extract content in JSON format. This content includes various details about the search results, one of which is the ASIN (Amazon Standard Identification Number) — a unique identifier for products listed on Amazon. The ASIN number is stored within a dictionary alongside other several key-value pairs. The characteristics that we extracted are all those that are relevant for Amazon for positioning products in search results ranking. For instance, *is_amazons_choice* -a binary variable- indicates that the product is determined to be as well priced and available to ship immediately based on Amazon criteria. The binary variable *is_sponsored* refers to whether the manufacturer is paying for rankign improvement in the publication of the product listing. The last binary variable *best_seller* indicates the items that have experienced large sales volume in the previous month and therefore, are also placed in higher positions. Other relevant variables that will serve us as control are the *price* of the product, the *rating* -an integer which represents the stars qualification by the customers- and the *reviews_count* -the number of reviews per product- . We also extract all content relating to reviews and the title and *url* of the product, from which we will extract if the Amazon product belongs to any of their Amazon Brands or not.

We also obtained the URL of the review page for each product. After generating CSV files from this extracted data for each search query, a request to the reviews URL will be sent for each product. From these requests we obtain information such as title, author, rating, and content of each review. The content of these descriptions will be used as input for our regression analysis.

The study introduced before (Chiara Farronato (2023)) analyzed a dataset that included 228,281 search results from 3,019 unique searches performed by 184 users. This analysis revealed that Amazon, on average, presents 76 results for each search. This insight suggests that typical users may primarily focus on exploring the search results displayed on the first, and possibly the second page.

In contrast, our data collection methodology is designed to scrape up to 6 pages, thereby capturing approximately 300 results per search with an average of fifty results per page. This approach collects more data than what a typical Amazon user is expected to review. This fact will be taken into consideration in our analysis, running a model that captures all data and another model that only considers results in the first two pages.

4 Data Preprocessing

Once we have completed the scraping phase, we pursued an exploratory analysis of our data to understand how to preprocess it. The scraping was performed on twenty product searches across different categories, with an average of 298 results for search and totaling 5950 observations. Of these, only thirteen searches contained at least one Amazon product, amounting to 3753 observations, which served as the initial dataset.

To begin with, we observed that in two of the thirteen categories, the three first observations did not correspond to a single product but to sponsored brands pages with no reviews, ratings or prices. As our objective is to analyze specific product characteristics and their impact on search rankings, these rows were excluded from our dataset. Our final set of observations amounts to 3747 data points.

In the other eleven product searches, we noticed that for some observations at the top of some pages, corresponding to sponsored products, the scraping process failed to retrieve the prices of these observations. To address this issue, we opted to impute the mean price of products from the same page, this helps to preserve the overall distribution of the dataset.

The *'POS'* column represents the position in the ranking of search results on Amazon. However, we have observed a consistent pattern where the first three products that are sponsored listings, are assigned positions 1, 2, and 3, respectively and subsequently, the organic search results then start again with positions starting from 1 and incrementing upwards. Instead, we use the index to get the Ranking as it reflects the sequential order of search results.

We also dropped the column *'subgroup'* which included values such as *'amazons choice'*, *'paid'* and *'organic'*.After examination, we see that these columns include information already included in *'is amazons choice'*, and *'is sponsored'* and are perfectly correlated.

Further refining our dataset, we converted the binary variables *'is amazons choice'*, *'is sponsored'* and *'best seller'* columns from categorical to numerical format. Additionally, we extracted descriptions of reviews from the *'reviews'* column that will be the input of our sentiment analysis.

We standardized the price column mainly because our sample is formed out of different products of different price ranges and distributions. We wanted to ensure uniform treatment of variables, irrespective of their original scale and distribution, which is crucial for facilitating accurate comparisons across variables . We also ensure the minimization of the influence of outliers.

For the sentiment analysis of reviews, a multi-step text preprocessing approach was undertaken to refine and standardize the textual data. The strip function was implemented to cleanse the text by removing non-word characters and we also lowercase all the words. Moreover, we used lemmatization which is employed to distill words down to their base or dictionary forms. This step is advantageous over stemming because it maintains the contextual usage of words, which is crucial for accurate sentiment analysis. We used stopword removal from the NLTK library, which has common English stop words filtered out.

Following these preprocessing steps, the TF-IDF (Term Frequency-Inverse Document Frequency) vectorization was applied. This technique transforms the text into a numerical representation, emphasizing the importance of words relative to their frequency across the entire corpus and within individual documents. The threshold for Term Frequency was set to ignore words that appear in more than 3% of the documents. This is the highest threshold for which we have observed common words across documents. The rationale behind this is to further identify and remove context-specific stopwords not previously captured by the generic NLTK list. These are words that, despite their frequency, do not contribute to the sentiment analysis because they are not indicative of positive or negative sentiment such as *battery, chair, hanger, jacket, make, mirror, pan, towel.*

The refined and vectorized text data was then analyzed using the TextBlob library, which yields polarity scores. The resulting polarity score is a float within the range [-1.0, 1.0], where -1 indicates a negative sentiment, 0 a neutral sentiment, and 1 a positive sentiment.

Figure 1 in Annexes illustrates how sentiment scores can yield a more continuous specification of the value that users place on each product, we plotted a scatter plot with a polarity score on the y-axis and star ratings on the x-axis. As expected, the higher the star ratings, the higher the polarity score -demonstrating a positive correlation-.

However, the dispersion of points at certain star rating levels, specially at higher rating levels, reflects the variability in sentiment among reviews with similar ratings. For instance, among those four to five star ratings, generally indicative of a positive review, the sentiment analysis, as shown by the polarity scores, captures a broader range of emotions and nuances ranging from approximately 0.0 and 1.0. This could mean that even though products are rated similarly, the intensity and expression of satisfaction can vary greatly among reviewers. Similarly, for low-rated products, we observe how not necessarily products rated with zero stars exhibit more discomfort than products rated with two-star review. Star ratings provide a quantified evaluation, but they do not capture the full spectrum of consumer feelings. Polarity scores add depth to this understanding by analyzing the language used in reviews, which can reveal nuances such as enthusiasm, disappointment, or indifference that are not as apparent from star ratings alone. For this reason in the next section we run one regression with the stars ranking and another regression with the sentiment score to evaluate how the results vary with respect to the ranking of the products.

5 Exploratory Data Analysis

After conducting exploratory data analysis, we noted that on average, each search lists approximately 4 Amazon Branded products (1,41%) totaling fifty amazon products in the whole sample. Regarding the distribution of product indicators, 62% of the Amazon brand products, 70% of products labeled as Best Seller and 16% of sponsored products appear in the first page of results, while within the first two pages, there are 72%, 82% and 31% of each label respectively. We observed how sponsored products are more uniformly distributed across pages while Amazon branded products and best sellers are found mainly in the first pages.

We also draw from these data that the average price of products listed in page one and two are 65% lower than average price of products listed in page three to six. In table 1 we present mean characteristics at the product level regarding Amazon products and non-amazon products. Amazon products have on average around fifty two thousand reviews while non-amazon products, condition on being reviewed, have around three thousand reviews. Amazon products also have on average half more points of rating. Amazon products are 10% more likely to be labeled as “best seller” and almost 10% more likely to be sponsored. Amazon’s average price is around 65% lower than non-Amazon products and the average ranking position is 66 for Amazon products versus 142 for non-Amazon products. Overall, we observe that Amazon products are higher-rated, have a lower price, and experience larger sales volumes than non-Amazon products. Regarding the latter result, we proceed to analyze whether, despite these facts, Amazon products are consistently placed in higher positions in search results.

6 Methodology

We will run four different OLS regressions, each covering different considerations. The first two models will include all observations obtained from the six different pages of each search result. Models 3 and 4 will include only observations obtained from the first two pages of the results. As mentioned before, the search results obtained from

Table 1: Comparison of Features between Amazon and Non-Amazon Products

Feature	**Amazon Products**	**Non-Amazon Products**
Reviews Count	51938.3200	2981.357565
Rating	4.4840	3.849770
Best Seller	0.1000	0.014065
Price	39.1614	115.623560
Sponsored	0.0400	0.052204
Ranking	70.7400	149.205572

the study 'Self Preferencing at Amazon: Evidence from search ranking' (Chiara et al., Chiara Farronato (2023)), reveal that the average number of products resulting from a search was 76. Therefore, we find it interesting to observe the results when considering only products appearing in the first two pages, given that this would represent, in most cases, the real cost that consumers would face for certain products being presented more prominently than others. The first two pages contained, on average, 96 product listings per category.

The first and the third model will include as a covariate, the rating of the product obtained directly from the Amazon site, -extracted through scraping-. In model 2 and 4, the sentiment score obtained for each product applying the sentiment analysis techniques will be used instead.

Thus, for model 1 and 3 we ran the following OLS regression:

$$\begin{aligned}\text{Ranking} = \beta_0 &+ \beta_1\text{Amazon Brand Indicator} + \beta_2\text{Sponsored indicator} \\ &+ \beta_3\text{Rating} + \beta_4\text{Amazon's choice indicator} + \beta_5\text{Best seller indicator} \\ &+ \beta_6\text{Price} + \epsilon.\end{aligned}$$

For model 2 and 4 we change β_3 five to *'polarity score'* to be able to observe the how impact of Amazon branded product changes when using polarity scores rather than rating.

We are particularly interested in assessing the relationship between Amazon Brand products and their ranking positions. To accomplish this, we will compare the impact of Amazon Brand products on ranking positions with that of sponsored products across various specifications. Additionally, we will include indicators such as 'Amazon's Choice' and 'Best Seller' to control for products that may receive preferential ranking positions due to their performance.

Additionally, during the preprocessing of the data, we included an indicator variable indicating the product type of each observation, with one for each of the thirteen categories that we have worked with. This category will serve as a grouping variable for computing clustered standard errors for the coefficient estimates. As the total sample is derived from thirteen different product searches, there might be within-group correlations due to the different types of classification strategies that Amazon might be using for different categories. For instance, there could be different ranking placements of products such as best sellers, sponsored products, or Amazon brand products across pages.

7 Results

We observe that in Models 1 and 2, being an Amazon branded product is associated with a 59-position boost when considering the Rating and 68 when considering the sentiment score. In Models 3 and 4, this association is reduced, as expected. Regardless of the rating-related covariate used, we obtain similar coefficients.

Comparing this with the projection for a product if it is sponsored, we observe that when considering all results, Amazon Brand products have approximately a 50% and 134% (model 1 and 2) larger effect in prominence than sponsored products. However, if we only consider the first two pages of results, we notice that Amazon products have between a 62% and 72% (model 3 and 4) impact that sponsored products have on ranking, obtaining a similar result to Chiara Farronato (2023)). As noted during the exploratory analysis, sponsored products have a more uniform distribution across pages, while Amazon products are predominantly placed in the first pages. Thus, we can understand why the impact is larger when considering the results of all 6 pages. In both scenarios, we observe a more significant impact on the prominence of Amazon brand products compared to sponsored products when using the polarity score as a regressor variable instead of the rating. This observation provides insight into the possibility that non-Amazon products may elicit stronger sentiments in reviews, while Amazon products may generate more moderate feedback. As seen in *Figure 3* in Annexes, we find that Amazon products are assigned more stable polarity scores across pages compared to those non-Amazon products, that exhibit polarity scores closer related to rating qualifications.

We also note that in Model 4, polarity score is positively associated with the outcome variable, contrary to what we would expect and to the results obtained in Models 1, 2 and 3. This is explained by the fact that, -as seen in *Figure 2* in Annexes-, even if average rating and polarity scores have a very similar downward trend across pages, in the first two pages, polarity scores seem to have instead a very slight upward trend.

Surprisingly, we observe that price does not appear to be statistically significant in any of the regressions. In this regard, we must consider that in product searches, results may include single products or offers for sets of products, depending on the provider. Additionally, there is often no indicator of quantity in the results. Consequently, we may encounter a range of prices across the pages, with no clear pattern of increasing or decreasing prices. This highlights the fact that Amazon does not consider the price variable on its own when placing products in search results, but rather uses a combination of different characteristics such as reviews, sales volume, etc. While we observed during the exploratory analysis that the mean price difference between the first two pages and pages 3 to 6 was around 65% lower, it is true that we did not observe a clear upward trend of prices from page 3 to 6, but rather a stagnation.

Table 2: OLS Regression Models Coefficients

	Model 1	Model 2	Model 3	Model 4
const	211.48***	174.76***	69.22***	49.25***
	(5.41)	(9.22)	(5.58)	(1.98)
Amazon Brand Indicator	-59.53***	-68.40***	-18.08***	-18.64***
	(18.51)	(18.96)	(2.88)	(2.95)
sponsored	-39.57***	-29.34***	-29.36***	-25.86***
	(8.98)	(9.57)	(3.24)	(4.16)
rating	-15.24***		-3.91***	
	(1.84)		(1.18)	
polarity score		-106.75**		11.71*
		(48.42)		(6.06)
amazon's choice	-73.51***	-79.23***	-2.65	-3.04
	(13.50)	(12.48)	(2.07)	(2.14)
best seller	-82.93***	-91.34***	-27.08***	-27.75***
	(11.15)	(11.32)	(3.74)	(3.69)
price	-0.36	-0.56	1.14	0.97
	(3.30)	(3.48)	(1.36)	(1.34)
Observations	3747	3747	1247	1247
R-squared	0.107	0.058	0.093	0.085
Adjusted R-squared	0.106	0.056	0.088	0.081
F-statistic	1.48e$^{-0.6}$	8.21e$^{-0.6}$	2.19e$^{-0.6}$	7.35e$^{-0.7}$

Note: $***p < 0.01$, $**\ p < 0.05$, $*p < 0.1$.

8 Limitations and further research

An important consideration in the analysis is the limited presence of Amazon-branded products within the dataset. As highlighted, only 1.41% of the entire sample consists of Amazon-branded products. This significant reduction in sample size for this category may impact the generalized and interpretative power of the results concerning Amazon-branded products. Also, we created some interaction terms between Amazon Brand Indicator and other regressors to further explore joint effects of different variables on the outcome variable. Despite the fact that in our trials they were not statistically significant and were subsequently omitted, future investigations with larger datasets may benefit from considering them. Doing so could enable the capture of more complex relationships between regressors and the outcome variable. Moreover, it would be of interest to explore how the association between product and Ranking evolves over time, considering the dynamics of rating scores. Such analysis could be extended to sponsored products, Amazon products, and non-Amazon and non-sponsored products, providing further insights into the temporal dynamics of search rankings.

9 Conclusion

To conclude, we have observed how Amazon's self brand-products enjoy higher positioning ranking with respect of those non-amazon sponsored products when considering all page results. Narrowing down the analysis to the usual number of products viewed by a consumer -approximately two pages of result- the impact ranges between 62% and 72% of that of sponsored products. Particularly, the effect on positioning has been found to be larger when using polarity scores of text reviews instead of average consumer rating per product. The latter finding potentially reveals that the effect of self-prominence -and thus, cost to consumer- is larger when considering a more nuanced view a product's value to consumer.

References

Chiara Farronato, Alexander MacKay, A. F. (2023). Self-preferencing at amazon: Evidence from search rankings. Accessed: 2024-03-14.

Loria, S. ("n.d."). Blob classes. Accessed: 2024-03-14.

Shah, P. (2020). Sentiment analysis using textblob. Accessed: 2024-03-14.

10 Annex

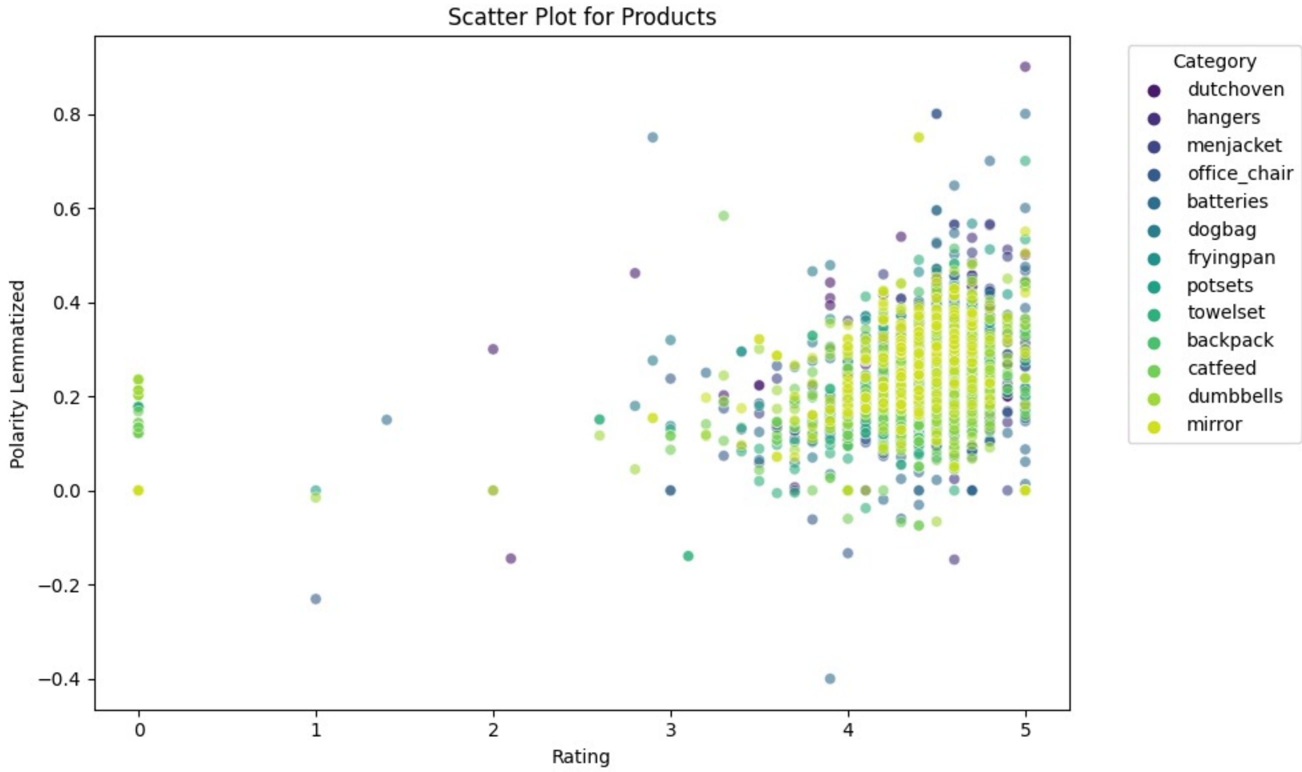

Figure 1: Correlation between rating and polarity score

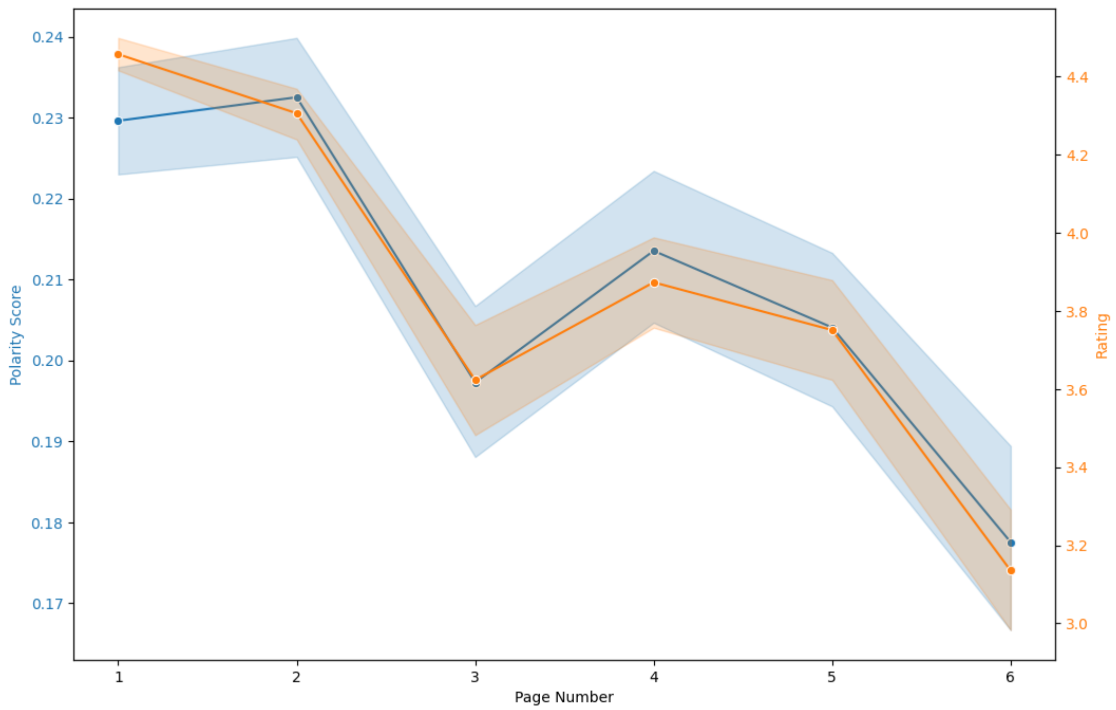

Figure 2: Average Rating and Polarity Score across pages

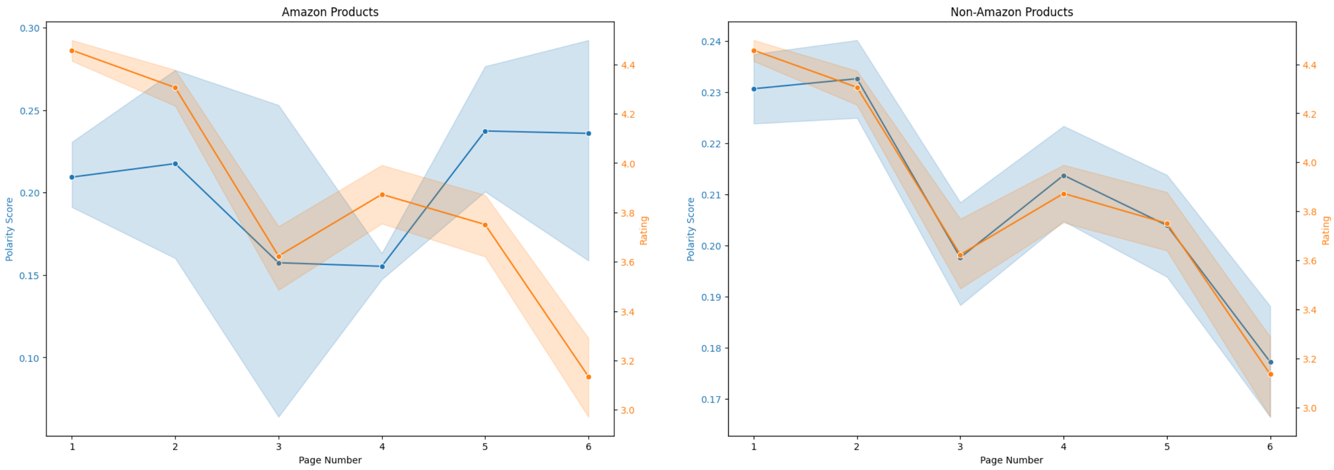

Figure 3: Average Rating and Polarity Score across pages for Amazon and non-Amazon products

.