Reviewbox

The Effect of Reviews on Amazon Sales Rank

Project Lead and Manager: Khaled

Customer Success Manager: Tejasvini

Technical Lead Analyst: Iris

Technical mentor coordinator: Yvette

Business mentor coordinator: Jordon

Meetings Coordinator: Daviid

Event planner: Nora

Table of Contents

Background

In the e-commerce industry, sales rank is a key metric for sellers since it indicates the performance of products. Being at the top of sales rank means more traffic, more sales, more revenues, and a better reputation. On e-commerce platforms such as Amazon, sales rank shows how well the product is selling when compared to its competitors. For example, if a brand sells blinders in the Office Products category, then the blinders will be ranked against all other office items in that category. Sales rank gives an idea of how popular the product is on the e-commerce platform. A product's sales rank is usually a number. The smaller the number, the better the product sells.

The algorithm of sale rank is not revealed by Amazon, leaving the public with limited knowledge about how sales rank was calculated. Since sales rank can be an important factor affecting a brand's reputation and the customers' buying decisions, computer engineers and data scientists have tried many methods to figure out how sales rank was calculated. Assumptions have been made according to previous research. The most widely discussed features include search relevancy, conversion rate and customer satisfaction. However, little has been touched on the impact of review recency on sales rank.

Sales ranks on Amazon update on an hourly basis throughout the day, and in this manuscript, we address the following questions regarding a products' Amazon sales rank: Does the sales rank fluctuate a lot, or does it stay relatively stable? What is the impact of recent reviews on sales rank? How's the trend of sales rank developing over time? Is sales rank influenced by seasonality?

Problem Statement

Sales rank indicates the relative popularity of a product and the location of the product on the search page, and therefore can greatly influence customers' decisions when choosing which product to buy rather than its competitors. The main objective of this project is to identify the relationship between reviews / ratings and sales ranks, and the importance of review features on sales rank.

Approaching the Problem

Sales Score project is an interesting research-oriented project, because currently there is no clear market solution to exploring the relationship between sales rank and reviews. However, as Amazon sales rank is such a mysterious and hot topic in the ecommerce business, many IT companies providing services to online retailers published plenty of online articles, which attempt to explain sales rank by discussing various key drivers that might appear in Amazon's algorithm. In short, Amazon sales rank captures the popularity of products sold online, and although it is tied closely with sales volume, other factors like product quality, availability and reviews might also play a role, and the most recent information is weighted more heavily than historical data by the algorithm (Hamrick, 2018).

Besides reading other companies' blogs and online articles, we also look at academia and looked for research papers that analyze online product reviews. One paper has documented the methodology of quantifying the relationship between reviews and actual sales volume (Forman et al., 2018), which is unfortunately not the sales rank that our project targets. While sales rank reflects sales volume, the challenge presented is how to establish a bridge between reviews and sales rank, and more specifically, whether we can build a model that uses reviews to predict the sales rank without sales volume.

To approach the sales rank problem, we also considered the behavior of Amazon consumers when they browse the website and how reviews may influence their purchase decisions. We observe that there are two options in the "Customer reviews" page on the Amazon web: Top Reviews and Recent Reviews, with the former one the default setting. The top reviews are the ones with high helpful votes, and they were often created years ago.

Fig 1: Sample review

Although top reviews are the default, it does not imply that recent reviews are less important. Recent reviews are also preferred by some customers who want to obtain the latest info on the product. More importantly, Amazon algorithm puts higher weight on the star-ratings of recent reviews than historical reviews to calculate the average star rating of the product. Many shoppers who have no habit of reading the reviews may just refer to the average star rating to make purchase decisions.

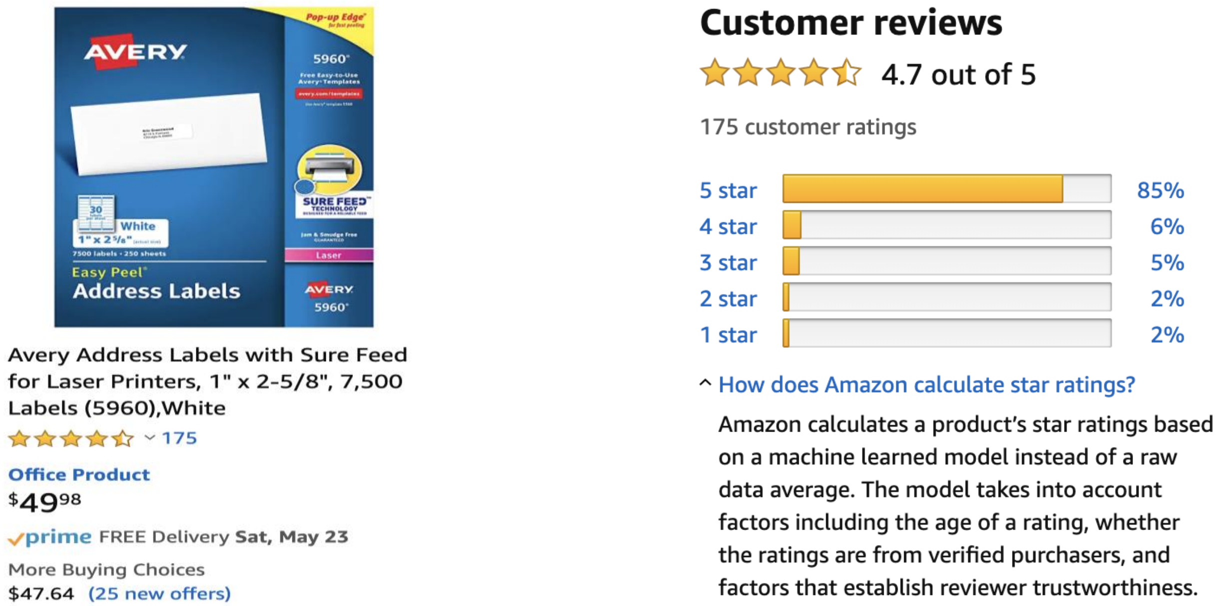


Fig 2 : Review attributes

Methodology

Key Features

After initial research, we first selected the variables that affect the sales volume, and then we put these variables into the machine learning model to investigate if they have a further impact on sales rank. The key features are categorized into four groups: star-rating of reviews, quantity of reviews, quality of reviews and control variables. We list all the variables we have considered in a table (Table 1)

Most features are aggregated into different time frames; for example, we count the numbers of reviews that appear in a one-week span and also in one-month period. The reason for taking a different time frame is because we are curious whether recent or older reviews have a larger effect on sales rank. Besides time frame, we also created lagged variables for certain features, because we consider that reviews may have a momentum impact on the sales rank. Amazon usually posts the customer reviews 72 hours after they are created. Therefore, it is likely that today's sales rank may only reflect last week's reviews and sales performance.

Categories of Metrics	Metrics
Star-rating of reviews	• Average star rating of reviews in one week & its lagged variables • Average star rating of reviews in two weeks & its lagged variables • Average star rating of reviews in one month • Average star rating of reviews in two months • Average star rating of reviews in one quarter • Average star rating of all reviews
Quantity of reviews	• # of reviews in one week & its lagged variables • # of reviews in two weeks & its lagged variables • # of reviews in one month • # of reviews in two months • # of reviews in one quarter • # of all reviews to the date • # of 5 star-rating reviews in one week & its lagged variables • # of 5 star-rating reviews in two weeks & its lagged variables • # of 5 star-rating reviews in one month • # of 5 star-rating reviews in two months • # of 5 star-rating reviews in one quarter • # of 5 star-rating reviews to the date • # of 1 star-rating reviews in one week & its lagged variables • # of 1 star-rating reviews in two weeks & its lagged variables • # of 1 star-rating reviews in one month • # of 1 star-rating reviews in two months • # of 1 star-rating reviews in one quarter • # of 1 star-rating reviews to the date
Quality of reviews	• % of non-anonymous customers for all reviews • Average word count of the recent 10 reviews • Average word count of all the reviews to the date • % of verified reviews in one month • % of verified reviews in one quarter • # of reviews with image in one month • # of reviews with image in one quarter • # of reviews with image to date • # of vine reviews in one month • # of vine reviews in one quarter • # of vine reviews to date
Control variables	• Average weekly price • Average monthly price

Table 1 : Independent variables

Dependent Variable

The sales rank we analyzed was the rank of the most narrowly-defined product category. The movement of the sales rank in a small category which contains fewer items provides more relevant information on the competition environment the products are at.

The tricky part of the project is how to model the sales rank into a dependent variable. At first, we treated the delta of log rank[1] as the prediction target. It measures the growth of logarithm of sales rank over a period. However, the machine learning models using this dependent variable resulted in negative R-squared scores. Therefore, we altered the direction of our research and adopted **delta of delta log rank** as the new dependent variable. The first "**delta**" of the term-delta of delta log rank means the difference between the delta log rank of the product in question and the delta log rank of its competing product set. The product set consists of 10 items within the same category and they have similar sales ranks as the product in interest. In other words, delta of delta log rank depicts the difference in sales rank growth of a product and its competitors. The purpose of analyzing this variable is that we want to look at which review attributes will lead to better sales rank improvement against competitors.

Visualize Delta of Delta Log Rank Approach

The following graphs will provide you with more detailed information on the selection of competing products. The red line visualizes the sales rank movement against time for the product in interest, which we name as ***product****. In week 1, from the same category, we take 10 products that have the closest sales ranks to the ***product**** and their respective sales rank values are represented by yellow dots. Then, we bundle the 10 products into a product set called ***product set in week 1*** by taking the average of their sales rank, and this product set is visualized by the yellow line. We will repeat this product set selection process every week, because we want to make two groups comparable in their current performances and investigate their difference in the next period. The same logic is applied to week 2 that we pick another product set with similar sales rank as the ***product****, and the ***product set in week 2*** is plotted in green.

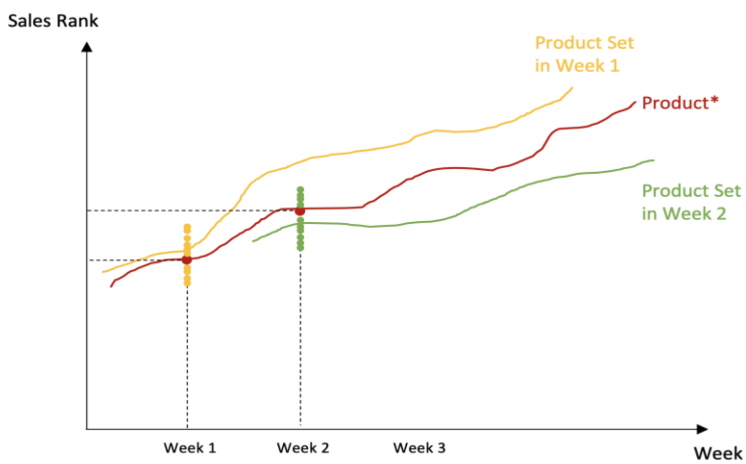


Fig 3: Sales rank trend for product and product set

[1] Delta of log rank = log(Sales Rank_t) - log(Sales Rank_(t-1))

Now we will explain what “delta of delta log rank” is using the graph below. After selecting the comparable ***product set in week 1***, we will calculate its delta log sales rank (yellow Δ symbol), which measures the change in sales rank from week 1 to week 2. The delta log sales rank of ***product**** is also computed, which is denoted by the red Δ symbol. The delta of delta log rank is the difference between the red Δ and the yellow Δ. Therefore, in other words, delta of delta log rank is the difference in growth of sales rank between the product in interest and its competitors with similar performance in this time period.

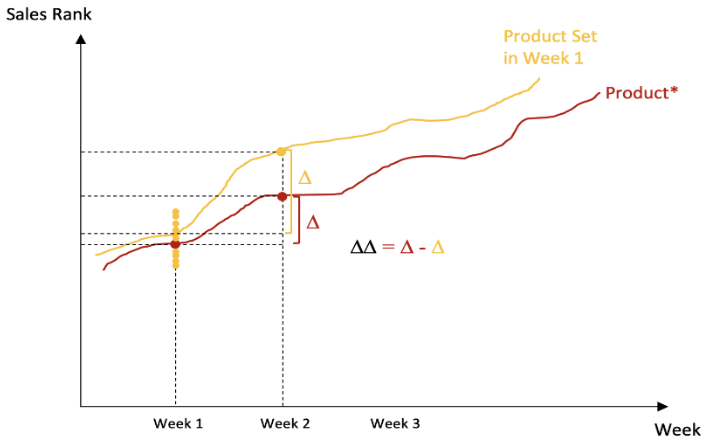


Fig 4a: Delta-Delta Sales Rank

In the “key features” section, we introduce the relevant metrics that can help predict the sales rank. However, they are not directly put into the model. The independent variable eventually employed by the machine learning model is the difference between the key features of ***product**** and its competing ***product set in week 1***. For example, the **difference** of the average rating of the ***product**** and the average rating of the ***product set in week 1*** is used, instead of directly using the average rating of the ***product****. The interpretation is that we are interested in examining how much better the average rating the product* earns in this week than its competitors can contribute to its lower growth of sales rank than its competitors.

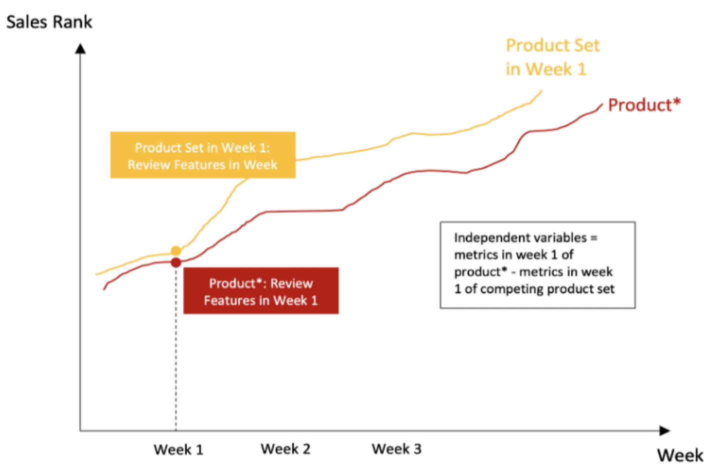


Fig 4b: Delta-Delta Sales Rank

Feature Engineering

Here we take the **average star rating in one week** as an example to illustrate how we performed feature engineering in the Sales Score project. For instance, we want to find the value of the **average star rating in one week** for the product ID B0004Z5QO on May 10th, 2020. This date will be used as a clue to locate the rows of Star_Rating column in the review dataset with the Date values seven days prior to May 10th, 2020. Then the mean value of the sevenstar ratings are computed as the final result.

Sales Rank Dataset

Product_ID	Date
B00004Z5QO	2020-05-10

Review Dataset

Product_ID	Date	Star_Rating
B00004Z5QO	2020-05-03	4
B00004Z5QO	2020-05-04	5
B00004Z5QO	2020-05-05	3
B00004Z5QO	2020-05-06	5
B00004Z5QO	2020-05-07	1
B00004Z5QO	2020-05-08	5
B00004Z5QO	2020-05-09	5

Avg_rating: 4.0

Fig 5: Feature engineering methodology

Modeling

Our objective is to model delta-delta sales rank and see the factors on which the change in sales rank depends on.

Data

The dataset is stratified such that the products in the train and test data do not overlap. The time period of the reviews in both training and testing data is the same. For the modelling, the largest category - Avery was taken into consideration. The aggregated dataset falls in the time period of one year since March 2019

Model

Full model

With Delta of delta sales rank as the dependent variable and 54 review attributes the linear model generated a very low Adjusted R-sq 2.6%. By training the model on Random forest regressor the out of sample R-sq enhanced to 12%. Which meant that the review attributes alone could explain 12% of variation of delta-delta sales rank.

However the main drawback of the model was that enough number of observations was not inputted into the model, i.e, every variables needs to have roughly 15-20 observations but the model took 384 observations for 54 variables, making the results less reliable.
The number of observations as an input to the model was reduced due to high number of null values in the attributes. Therefore, the full model was deemed less reliable.

Reduced model

$$\begin{aligned}\Delta\Delta \log \boldsymbol{rank} = {} & \Delta \boldsymbol{stars}_{\boldsymbol{oneweek}} + \Delta \boldsymbol{stars}_{\boldsymbol{pri4thweek}} + \Delta \boldsymbol{stars}_{\boldsymbol{twoweek}_{\boldsymbol{lag1}}} + \Delta \boldsymbol{stars}_{\boldsymbol{twoweeks}_{\boldsymbol{lag2}}} \\ & + \Delta \boldsymbol{stars}_{\boldsymbol{twomonths}} + \Delta \boldsymbol{stars}_{\boldsymbol{avg}} + \Delta \boldsymbol{numreview}_{\boldsymbol{oneweek}_{\boldsymbol{lag1}}} \\ & + \Delta \boldsymbol{numreview}_{\boldsymbol{twoweek}_{\boldsymbol{lag2}}} + \Delta \boldsymbol{num1star}_{\boldsymbol{twoweeks}_{\boldsymbol{lag1}}} + \Delta \boldsymbol{num1star}_{\boldsymbol{twoweek}_{\boldsymbol{lag2}}} \\ & + \Delta \boldsymbol{num1star}_{\boldsymbol{twomonths}} + \Delta \boldsymbol{num1star} + \Delta \boldsymbol{num5star}_{\boldsymbol{oneweek}_{\boldsymbol{lag2}}} \\ & + \Delta \boldsymbol{num5star}_{\boldsymbol{oneweek}_{\boldsymbol{lag3}}} + \Delta \boldsymbol{num5star}_{\boldsymbol{oneweek}_{\boldsymbol{lag4}}} + \Delta \boldsymbol{num5star}_{\boldsymbol{twomonths}} \\ & + \Delta \boldsymbol{num5star}_{\boldsymbol{onequarter}} + \Delta \boldsymbol{avgWordcount} + \Delta \boldsymbol{namedauthor} \\ & + \Delta \boldsymbol{verified}_{\boldsymbol{onemonth}} + \Delta \boldsymbol{numImage} + \Delta \boldsymbol{numVine} + \Delta \boldsymbol{avg}_{\boldsymbol{monthlyprice}}\end{aligned}$$

With the above variables as the input the below table shows the model performance. The interpretation is that a linear regression model performs worse than average (i.e, guessing that the dependent variable is the average of all the independent variables combined)

Random forest model does much better explaining 4% of the variance and the boosting regressor model performs twice as well.

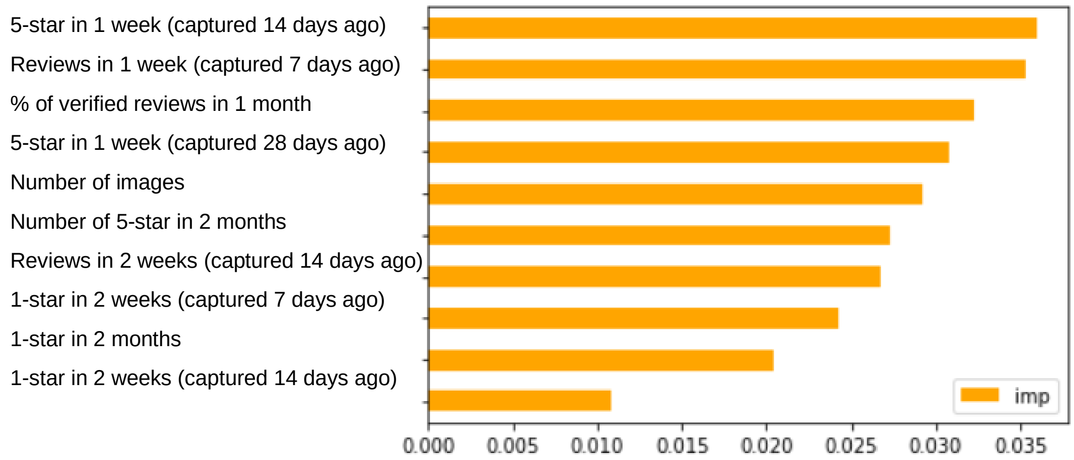
Model	Performance metric
Linear Regression OLS	negative
Random forest regressor	4%
Gradient Boosting Regressor	**8%**

Table: R^2 for each of the models

The gradient boosting regressor shows the best performance for the model. It reduces variance. Random forest build trees in parallel, while in boosting, trees are built sequentially i.e. each tree is grown using information from previously grown trees unlike in bagging where we create multiple copies of original training data and fit separate decision tree on each. Due to which Gradient boosting generally performs better than random forest. Therefore, was selected for interpretation.

Results

Below table and graph shows the top 10 important features observed.



The change in number of five star rating aggregated over a week with a lag of two weeks is the most important attribute in explaining the Sales rank, followed by the change in number of total reviews over a week with a lag of a week, change in verified purchases (in one month) and so on.

We had suspected the impact of the reviews to be realized with a time lag and see it in the results. We see that the five-star related attributes are twice as important of one-star related attributes.

Conclusion

From the model results we can conclude that the delta delta sales rank is impacted by number of one star, five-star ratings, number of reviews, verified purchases and price.

The lag variables indicate the effect on sales rank by taking time into consideration. The model is expected to yield better results using actual sales volume data as currently sales rank was used as a proxy for sales volume

The model explains 8% of variation on the sales rank which is significantly higher than a random guess. Brands can take the lag parameters into consideration for anticipation in change of sales rank and can use this to decide advertising campaign to optimization advertising campaign.

References

1. Hamrick, D. (2018, April 2). *Master The Amazon Sales Rank: Top 10 Things You Need to Know.* Jungle Scout. https://www.junglescout.com/blog/amazon-sales-rank-top-10-things-you-need-to-know/

2. Forman, C., Ghose, A., & Wiesenfeld, B. (2008). Examining the relationship between reviews and sales: The role of reviewer identity disclosure in electronic markets. *Information systems research,* 19(3), 291-313.