

Amazon Product Analyzation Project

The goal of this project is to analyze the sales on Amazon in order to make more precise, more customized and better recommendations to customers and improve the products selling on the platform. The data set we want to analyze is the 5-core subset with the size of 14.3gb. 5-core subset makes sure all users and items have at least 5 reviews (75.26 million reviews).

We want to achieve these two goals based on previous customers' reviews.

We want to find the best product in each category to recommend to customers, which is the product that has the highest average rating among all. This includes filtering the data, because we only want the data that is verified, which means the field "verified" is true. Then we calculate the average rating of every product and the average rating of this whole category. Note that the average score of each product should be a weighted average. For example, reviews with pictures or more words should be weighted more than reviews with plain text. If there is a tie, we choose the one with more reviews.

To make a more customized recommendations, we look at all the reviews given by a customer to find out which categories this customer often shops from. Then Amazon can put more products from the particular categories into this customers recommendation feed.

In order to improve the products on the platform, we first divide the reviews in each category into two parts: GOOD and BAD. GOOD means this review gave a rating of 4 or 5. Otherwise, the review would be categorized as BAD. We would then extract key words from reviews of both GOOD and BAD, so that we can tell what words are mentioned the most in the GOOD reviews and BAD reviews. From the frequently mentioned keywords, the business owner can find out what customers typically like and don't like about their products. Also, Amazon can see what aspects customers tend to focus on more when shopping from different categories. For example, people would probably care more about the looks instead of quality when shopping from Arts Crafts and Sewing category, while they would care much more about quality when shopping from Office Products Category.

Group Member:

Zhenmin He, 301589203

Hershil Piplani, 301591594

Houli Huang, 301593424