

Problem Statement: Our objective is to certify whether the ratings published by Amazon inc. for its products grouped by similar product categories under best sellers with the same published rating are reliable or does the user need to be suspicious about the published ratings for the group of top 3 products with similar ratings in a given product category.

Project Scope: Amazon Inc. on its website states that customer reviews for a product don't use a simple average of the star ratings of all the reviews but instead, its rating system considers things like how recent a review is and analyzes the text of the reviews to verify the trustworthiness of the reviews. However, this doesn't imply that two products in the same product category with the same average rating would be exactly the same in terms of quality or the consistency of the reviews. One approach is to analyze if the fraction of 5 star ratings across the top three bestseller products with the same average rating score published by Amazon Inc. are similar. To do this study, 1000 reviews corresponding to the most recent reviews ordered by date published are selected for each product. The words in each review are analyzed by an open source sentiment analyzer and a positive sentiment score is calculated for each review. This review score is mapped to the 5 star rating and the number of reviews belonging to each of the ratings (1 star, 2 star, 3 star, 4 star, 5 star) is tabulated for each of the three products. This step is clearly illustrated in the table below under the hypothesis test paragraph. This methodology is repeated for three different product categories offered by Amazon. The results of this analysis will provide a decision framework for the buyer to decide whether they could trust the review by Amazon to determine if products in same category with the same average rating are infact similar or whether the buyer should do more research by looking into specific review descriptions.

Null Hypothesis: The fraction of reviewers favoring 5 star ratings is the same across all three brands in the same product category with same published average amazon rating.

Hypothesis Test: The Chi square test methodology shown in example 14.4, pg 731 of the textbook "Mathematical Statistics With Applications" is used to evaluate whether the null hypothesis is acceptable with 95% confidence. In this test, the critical value with alpha of 0.05 and $(r-1)*(c-1)$ or $(5-1)*(3-1) = 8$ degrees of freedom for the Chi square statistic is calculated. The Chi square statistic is used in lieu of other statistical tests such as the t-test or the z-test given the fact that we are dealing with proportions and categorical data.

# Reviews/Brand	Brand 1	Brand 2	Brand 3	
1 Star	100	50	150	300
2 Star	100	100	50	250
3 Star	200	250	200	650
4 Star	200	200	150	550
5 Star	400	400	450	1250
Total # of Reviews	1000	1000	1000	3000

Table 1. An Illustration of the test sample with number of reviews for each rating category

$$\text{Chi square score} = \sum_{i=1}^5 \sum_{j=1}^3 (n_{ij} - E(n_{ij}))^2 / (E(n_{ij}))$$

$$E(11) = 300 \cdot 1000 / 3000$$

$$E(12) = 300 \cdot 1000 / 3000$$

$$E(13) = 300 \cdot 1000 / 3000$$

$$E(22) = 250 \cdot 1000 / 3000$$

.

.

.

$$E(53) = 1250 \cdot 1000 / 3000$$

Tools Used: Pyspark, pandas, matplotlib, numpy, beautiful soup web scrapping tools, scipy stats libraries, NLTK natural language processing libraries.

Project Methodology:

Pages similar to the following pages are scrapped on Amazon:

https://www.amazon.com/product-reviews/B00F4CEH NK/ref=acr_dp_hist_5?ie=UTF8&filterByStar=five_star&reviewerType=all_reviews#reviews-filter-bar

The following open source natural language sentiment analyzer is used on each of the latest 1000 reviews ordered by date:

<https://towardsdatascience.com/detecting-bad-customer-reviews-with-nlp-d8b36134dc7e>

A total of 3 product categories are reviewed and the analysis is presented on whether or not the user should just accept Amazon's ratings or do further analysis of the reviews before purchasing the product.