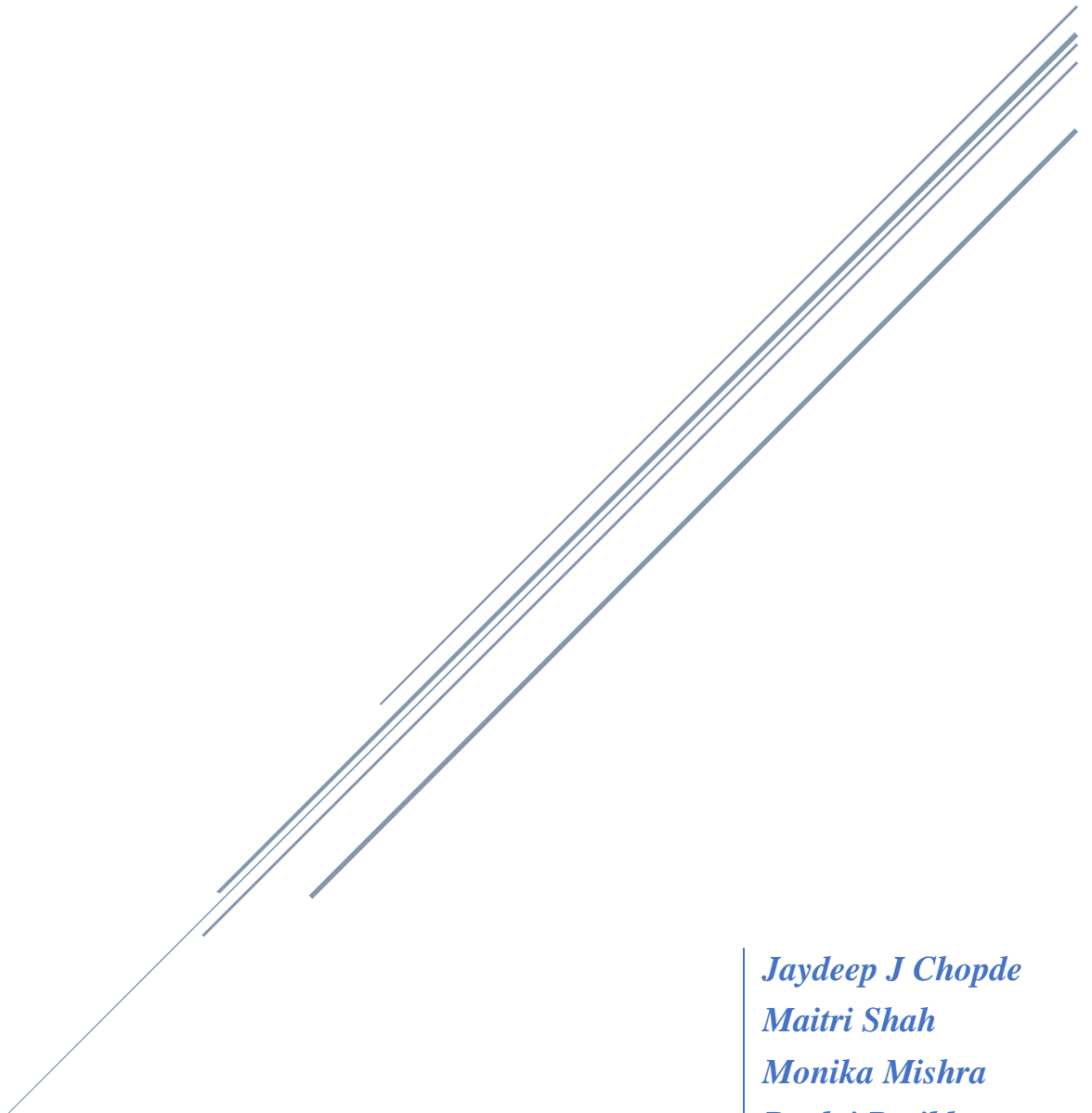


AMAZON REVIEWS ANALYSIS

A study of user reviews



Jaydeep J Chopde

Maitri Shah

Monika Mishra

Pankti Parikh

Rakshit Chandan

Objective:

In today's competitive environment, an online retailer must be exceptionally good in terms of product and services to its target audience. Even after being of quality, certain products end up being reviewed thousand times while others have over millions of reviews from customer base across the world. Through this study is to analyze parameters that affect the performance and success of product in the online store. Of the various parameters that will be analyzed, importance will be given to the quality and quantity of target audience of the product in question. Apart from these, reviews, ratings will be compared to study how they affect the ranking or success of a product. The objective of this study is to understand how certain product make it big in the online retail while others with equal level end up in average reviews. The ultimate goal is to understand if there is a fixed pattern amongst all the highly rated and reviewed products.

Project Description and Motivation of the study:

- This study therefore was designed to determine the particular pattern behind the success of a product on Amazon store. The results obtained from this study would be helpful for sellers to inculcate the specific qualities that would make their product fit the pattern we determine. The analysis of some products would be done of the basis of several parameters. These parameters will then be compared for each product in accordance with their success. Once that is done, a pattern if exists would be established to simplify the results.
- This analysis can be carried out based on the following comparisons.
 - I. Category vs reviews
 - II. Rating vs No. of reviews
 - III. Region vs Ratings.

Dataset: - <https://s3.amazonaws.com/amazon-reviews-pds/tsv/index.txt>

Data Description: -

Schema

	Column name	Data type	Key
1	review_id	string	
2	marketplace	string	
3	customer_id	int	
4	product_id	string	
5	product_parent	int	
6	product_title	string	
7	star_rating	int	
8	helpful_votes	int	
9	total_votes	int	
10	vine	string	
11	verified_purchase	string	
12	review_headline	string	
13	review_body	string	
14	review_date	timestamp	
15	row_num	int	
16	sentiment	string	
17	product_category	string	Partition (0)

Questions:

- I. Is there a relation between certain parameters and their performance, if so, what is that relation?
- II. Does the balance of no. of reviews and ratings affect success, if so what is the perfect balance?
- III. How can review collection be improved?

- IV. What could be a standard pattern, which when followed will ensure success for the app?
- V. Is the pattern similar for product in all the categories?

References

Amazon Customer Reviews Dataset. (n.d.). Retrieved October 6, 2018, from <https://registry.opendata.aws/amazon-reviews/>

Hasson, R. (2018, June 04). How to scale sentiment analysis using Amazon Comprehend, AWS Glue and Amazon Athena | Amazon Web Services. Retrieved October 6, 2018, from <https://aws.amazon.com/blogs/machine-learning/how-to-scale-sentiment-analysis-using-amazon-comprehend-aws-glue-and-amazon-athena/>

Hu, N., Liu, L., & Zhang, J. (2008). Do Online Reviews Affect Product Sales? The Role of Reviewer Characteristics and Temporal Effects. *SSRN Electronic Journal*. doi:10.2139/ssrn.1324190