TASK

Sentiment Analysis on Amazon Reviews Using Python

Introduction

Customers' behaviors have been greatly transformed by the impact of the internet The retail e-commerce world of the online marketplace does not make products feasible to customers and there are so many new products emerging every day. Hence, customers need to rely largely on product reviews to make better decisions on purchases. However, searching and comparing text reviews can be frustrating for users. Hence, we need a better numerical rating system based on the reviews which will make customers' purchase decisions easy.

During their decision-making process, consumers want to find useful reviews as quickly as possible using a rating system. Therefore, models able to predict the user rating from the text review are critically important. Getting an overall sense of a textual review could in turn improve the consumer experience. Also, it can help businesses to increase sales, and improve their product by understanding customers' needs.

Data Collection and Data Summary

The Amazon Review Data.

The dataset used was obtained from the Candidate Analyst Google Drive folder, the Case_amazon.json: Sample from 5-core file with 500.000 reviews (JSON). It contains the reviews and ratings given by different users.

The dataset is based on Amazon customer reviews. We would like to investigate the impact of customers' reviews on the platform by carrying out exploratory data analysis on this text data.

Customers reviews are customer reviews are a form of customer feedback on electronic commerce and online shopping sites. Customer reviews are forms of customer feedback.

The data set has information on reviews such as Review_ID, asin, reviewName, helpful, reviewText, overall, summary, unixReviewTime, and reviewTime. The main column that we are interested in is whether the customer churned or not. For this data analysis project on reviews, our main focus will be on features such as ratings, text, and helpful.

There are 500000 rows and 9 columns with all of the variables being object type, except overall and unixReviewTime which are both int64.

Problem Statement

Questions

1. What are the characteristics of a good/helpful review?

2. What is the review behavior among different categories? (Optional depending on which data you used, some subsets do not contain different product categories)
3. How would you characterize the relationship between good reviews and increased revenues for Amazon? (Can be answered theoretically, without data insights from data analysis)
4. Which group of reviewers is more valuable to the business?
5. Any ideas in order to improve the reviews on Amazon?

DATA CLEANING

SUMMARY OF THE METHODS AND VISUALIZATIONS DONE DURING DATA CLEANING

The first set of issues I noticed while examining the dataset using the info function is that some variable names are not very descriptive. For example, overall, which changed to OverallRating. Also, the reviewTime variable which is date was stored as

an object data type instead of a date. This was corrected before making progress in the analysis. Also, I concatenated summary and reviewText, and renamed it review_text, then delete the redundant columns using drop.
The helpful column was split into negative_feedback and positive_feedback, and it was dropped to avoid redundancy.

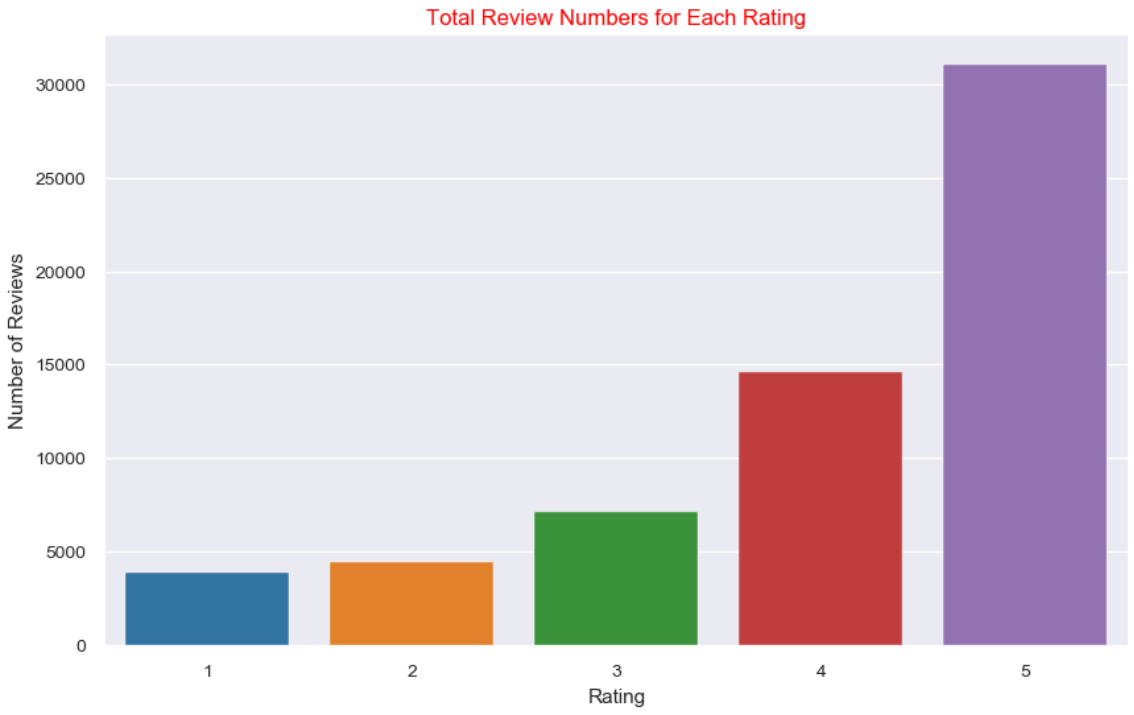
Handling Duplicates, Missing Values

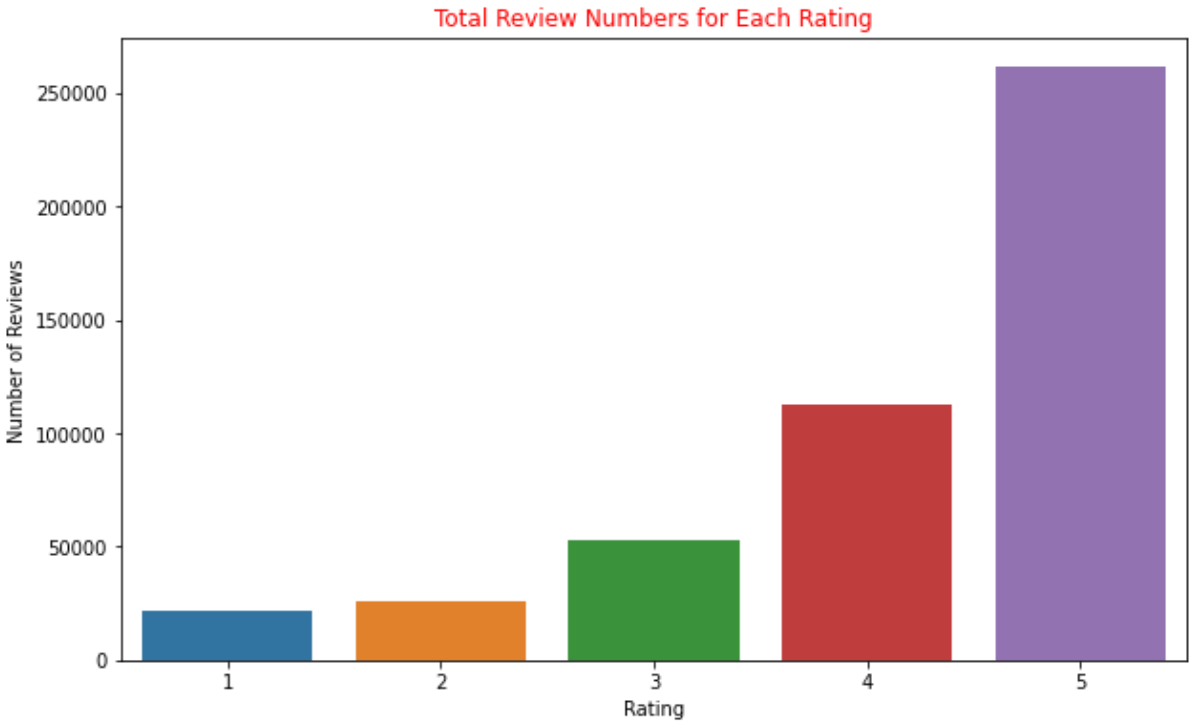
1. missing values in “reviewerName”, “refviewText”, and “summary” were dropped.
2. “reviewText” and “summary” were concatenated and were and saved as the review_text feature
3. The “helpful” feature was split into positive and negative feedback.
4. Ratings greater than or equal to 3 were categorized as “good” and less than 3 were classified as “bad”.
5. The “helpful ratio” was calculated based on positive feedback/total feedback for that review
6. Dropped duplicates based on “asin”, “reviewerName”, ”unixReviewTime”. After dropping duplicates, the dataset consisted of 475008 rows and 11 features from 475340 rows before removing duplicates.
7. ReviewTime was converted to Datetime ‘%m %d %Y format.
8. Columns were renamed for clarity purposes.

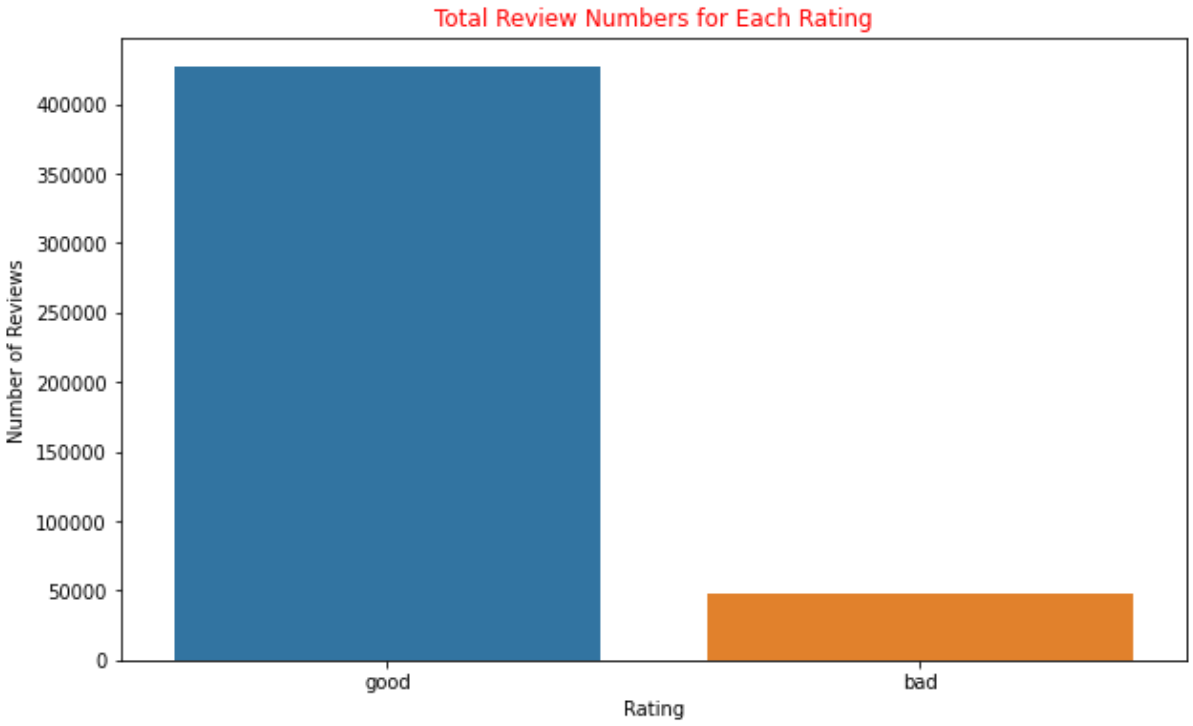
Problem Solution

1. What are the characteristics of a good/helpful review?

From the analysis of the dataset, it is clear that the length of reviews of one of the characteristics of good/helpful reviews.







Good reviews have a higher length of reviews, while bad reviews have a lower length of reviews.

Findings show that review good/helpful review is significantly influenced by the **length, valence, timeliness, and deviation rating of the reviews**. The results also underlie that a review submitted by a reviewer who has more followers and experience is more affected by review characteristics.

The reading time of helpful reviews is twice that of non-helpful reviews which means people find longer reviews helpful.

2. What is the review behavior among different categories? (Optional depending on which data you used, some subsets do not contain different product categories)

BN: I could not work on this. Although, I planned to do so before my system crashed in the early hours of yesterday.

3. How would you characterize the relationship between good reviews and increased revenues for Amazon? (Can be answered theoretically, without data insights from data analysis)

Generally speaking, reviews are very important for customers, they’re also very important for businesses. Reviews directly affect brands' reputation (for better or worse), increase or decrease sales, and they can be the final nudge that either converts a customer or convinces them to never give your brand a second thought. More measurably, the number of reviews you have impacts your SEO, as online reviews are factored into Google search results.

Reviews and ratings on Amazon, a digital marketplace are a very powerful and useful way to help funnel those potential customers to the store so that they can start browsing, clicking, and buying. Thus, increasing sales and revenues in the long run.

User reviews, particularly on sites like Amazon, mean a great deal to shoppers. A product that has just one review is 65% more likely to be purchased than a product that has none, according to Power Reviews CEO Matt Moog. He added that one-third of online shoppers refuse to purchase products that have not received positive feedback from customers.

Power Reviews also found that positive reviews increase sales by 20% on sites that implement them into the shopping experience well. And with online sales around the world hitting $2 trillion per year, the impact of reviews is immense.

BloomReach, a marketing research firm, found that Amazon product reviews are the most popular and trusted.

Resources

https://www.mayan.co/blog/are-amazon-reviews-important-for-your-business

https://pattern.com/blog/why-are-reviews-on-amazon-important/

https://www.businessinsider.com/amazon-reviews-greatly-impact-online-shopping-sales-2017-3?r=US&IR=T

BN: I could not proceed with my analysis from here. This is a result of my system crashing in the early hours of yesterday. I lot almost all my visuals. I sent emails twice yesterday to this effect to clemens.kuehn.external@zalando.de. I could not extract most of my visuals from the jupyter notebook I used before it crashed.

4. Which group of reviewers is more valuable to the business?

Valuable Reviewers are measured using **Rank by Helpful Votes**.

Quality, quantity, and consistency – are the main key components to consider if you want to improve your Reviewer Rank

5. Any ideas in order to improve the reviews on Amazon?

Identify what reviews mean: I believe this is the first step to improving reviews on amazon.

Improve fulfillment: Answer questions and address concerns from reviews

Improve listing quality.

Provide good customer service.

Communicate professionally.

Resources

https://courses.analyticsvidhya.com/courses/Intro-to-NLP

https://www.analyticsvidhya.com/blog/2020/04/beginners-guide-exploratory-data-analysis-text-data/

THIS REPORT WAS WRITTEN BY: Ayomide Fadeyi