

Team Name: **Java Brains**

Dataset Name: Amazon Tablet Review

Members:

John Nance 013904174

Robert Jimenez 015906343

Smit Patel 016280860

Emma Perez 025300897

Amazon started selling Amazon Tablets in 2011, since then they introduced different types of tablets. We devolved a python code on specific data given to us on six different types of Fire tablets. The data was mostly on Amazon Tablet reviews and ratings. We hope that our code will give Amazon a clear understanding of the feedback they receive from customers buying these tablets. This way Amazon can tell which specific tablet is the best rated and reviewed, as well as which is the worst. Hence, Amazon can improve based on reviews and ratings, and meet the consumer demands.

Data

The dataset that we chose had 12,412 entries comprised of different variables. These variables were Product ID, Product Name, Star Rating, Reviews, Review Title, Brand, Categories, Review Date, and User Name. All of these variables are important to the creation of the spreadsheet, but Product ID, Product Name, Star Rating, Reviews, and Review Title are vital towards answer the questions that Amazon may have about their products when it comes to the customers. Product ID is unique for each of the six products. For Product Name some of the products have the same name, but are different based on product ID. This could mean that each of the Product ID with the same Product Name could be earlier or later releases. Star ratings range from zero to five stars. Reviews are detailed descriptions of feedback from the customers. Review Title is just the title of the review giving a glimpse of whether the review will be good or bad. Although these variables were not included in our code, if they were the, Review Date is important because it shows how relevant the Review is and if any updates have been done since

then. The User Name allows Amazon to reply to specific customer needs and shows credibility for the review.

Problem

The problem we were trying to solve is; what products had the best reviews, what products had the worst reviews, and what products highest and lowest average star rating. With the data calculated we could recommend the best course of action. Whether, it be recommended that Amazon not sell the product anymore or how Amazon can improve the product. We also wanted to see if customers were consistent with their reviews, meaning was the customers contradicting themselves when writing their reviews.

Analysis

We used a library called TextBlob to find the sentiment polarity and subjectivity. We also used another library called Statistics and found the average, minimum, maximum, and the standard deviation of the sentiment polarity. As well as the average, minimum, maximum, and standard deviation of the sentiment subjectivity. For subjectivity, if the number is closer to one, the text is very subjective meaning that it is very opinionated. But if the number is closer to zero, the text was not subjective meaning it was vague. When we calculated the subjectivity we found that most of the customers entries were subjective meaning that they were opinionated. The average subjectivity was between a .40 and .60 (Figure 1.1 and 1.2 on page 6). For polarity, if it is over zero the review is positive and if it is under zero the review is negative. When we calculated the sentiment polarity, all of the products at most were positive because the numbers were in between .40 and .60 (Figure 1.3 and 1.4 on page 7). Since the average sentiment polarity and subjectivity were calculated for both the review title and the review itself, we were able to

compare both calculations to see if the customer were contradicting themselves. Fortunately the customers were very consistent with their reviews meaning the customers were not contradicting themselves.

When we calculated the average star rating, we had to make sure that any review with a star rating of zero and below or above five was not calculated, since star ratings are between one and five. There were some reviews that did not have any star ratings. Some reasons for this could be that the customer forgot to leave a star rating, or the customer did not want to leave a star rating. After the averages of the star ratings were calculated we were able to see results that were formulated. We found that the average star rating for each product were between four to five rating with AVphgVaX1cnluZ0-DR74 which is the Fire Tablet in magenta being the lowest rating at a 4.459 and AVqklj9snnc1JgDc3khH which is the Fire Tablet in Tangerine at a 5 rating. Even though the Fire Tablet in Tangerine had the highest rating, there were only four entries and they all had a rating of five. If this product had a lot more reviews compared to some with over a thousand reviews than the tablet in tangerine would be the best rated tablet. Figure 1.5 on page 7 shows the average star rating of each product.

A Word Cloud was created for Text Review, Review Title, and both review text and title. For the variable review text, customers detailed their review with many positive words like: love, use, tablet, easy, good, great, etc (Figure 1.6 on page 8). The Word Cloud for the variable review title comprised of same positivity as the review text like: tablet, use, love, easy, good etc (Figure 1.7 on page 8). The last word cloud was on both the text and title review, also has the same positive words as the other word clouds except for prefect and great price (Figure 1.8 on

page 8). These Word Clouds quickly gives Amazon and us clear indicators and more qualitative results that shows how much Amazon customers like their tablet.

Results

Our results concluded that Product ID: AVqklj9snnc1JgDc3khu Product Name: Fire HD Tablet in Tangerine; had the highest average star rating at five; however, it only has two ratings. Product ID: AVqklhwDv8e3D10-lebb Product Name: All new Fire HD 8 in Magenta; had the best reviews based on subjectivity and polarity. Since product AVphgVaX1cnluZ0-DR74 had 9,209 entries, AVqklhwDv8e3D10-lebb had 2,813 entries, AVqVGZNvQMIgsOJE6eUY had 210 entries, product AVpfBEWcilAPnD_xTgb7 had 11 entries, product AVpfwS_CLJeJML43DH5w had four entries, and product AVqklj9snnc1JgDc3khU had two entries. With these number of entries presented, it shows that comparing all the products together were illogical. The reason why it is illogical is because you cannot compare 9,209 entries to 2 entries.

Conclusion

Overall, from our findings we recommend that Amazon encourage their customers to write reviews where they are lacking reviews. Although we were able to find the average polarity, subjectivity, and star rating for each product, it is still not a clear depiction of what Amazon could improve based on the product. When Amazon receives more entries from the products that are lacking reviews, Amazon could then be able to decipher on what products they could invest more money in like advertise or manufacture more of that model, improve, or discontinue.

Figure 1.1: Average Subjectivity of the Review Title for Each Tablet

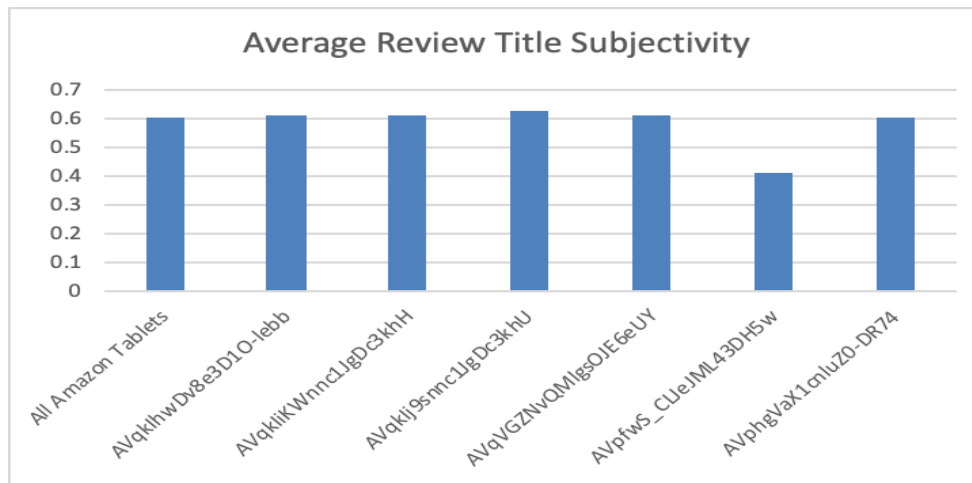


Figure 1.2: Average Subjectivity of the Review Text for Each Tablet

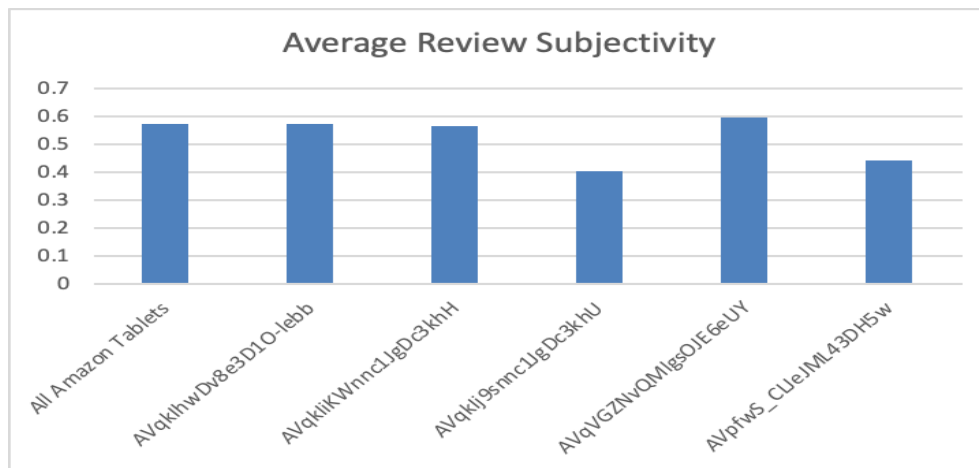


Figure 1.3: Average Polarity of the Review Title for Each Tablet

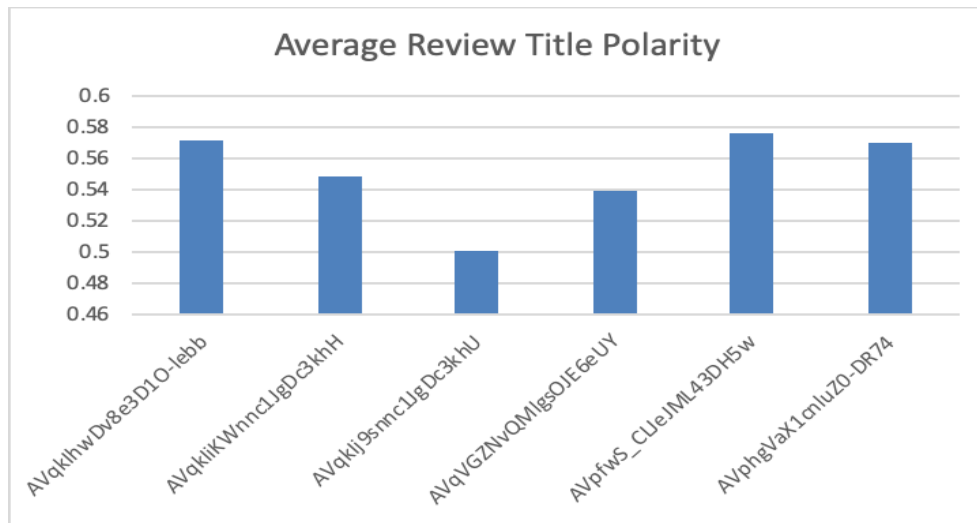


Figure 1.4: Average Polarity of the Review Text for Each Tablet

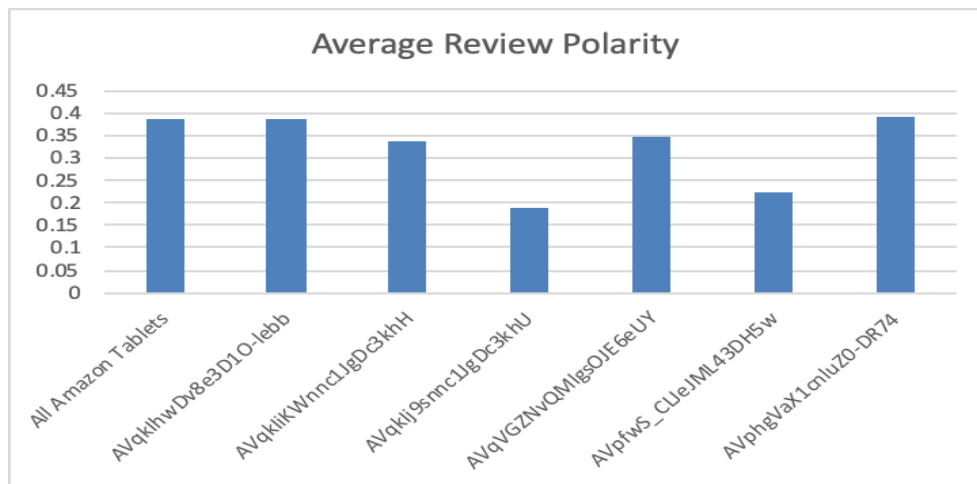


Figure 1.5: Average Star Rating For All and Each Product

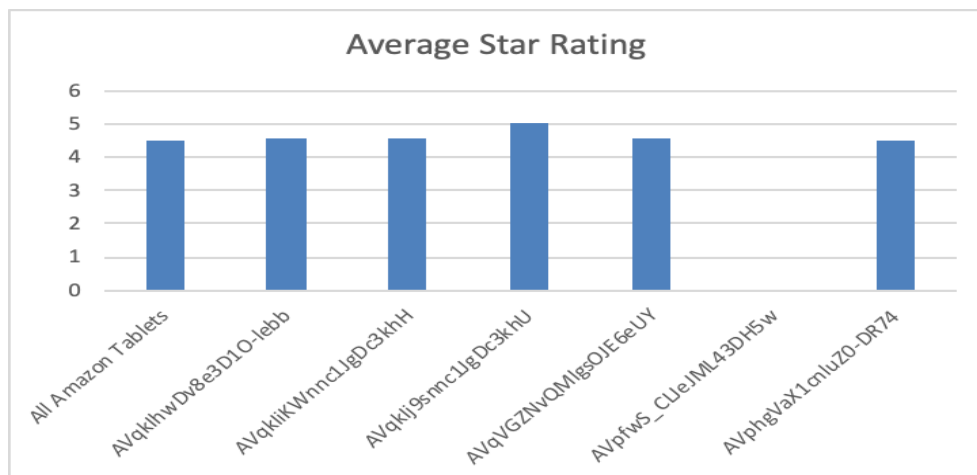


Figure 1.6: Word Cloud for Review Text

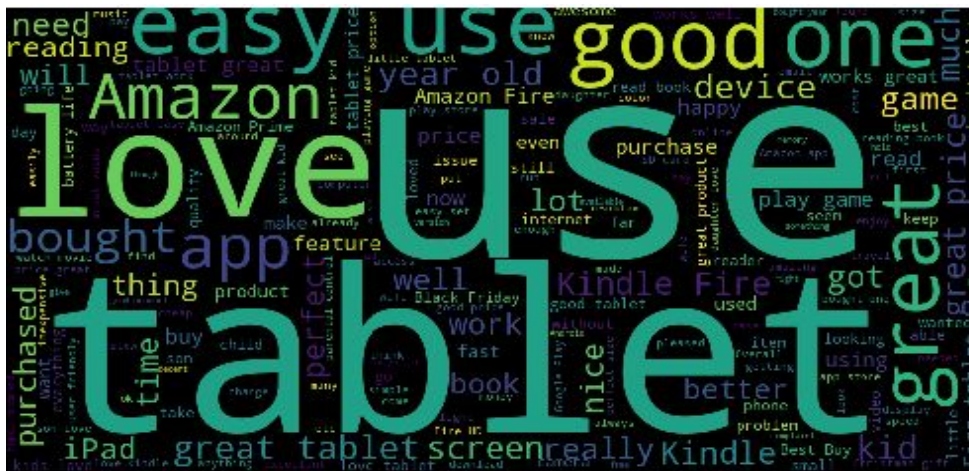


Figure 1.7: Word Cloud for Review Title

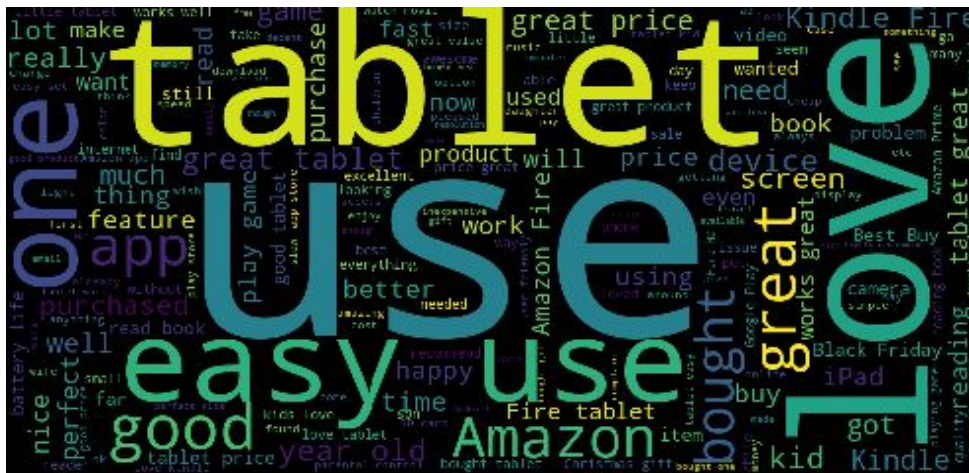


Figure 1.8: Word Cloud for Review Text and Title

