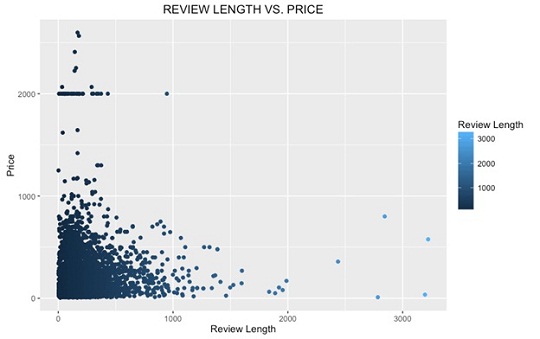
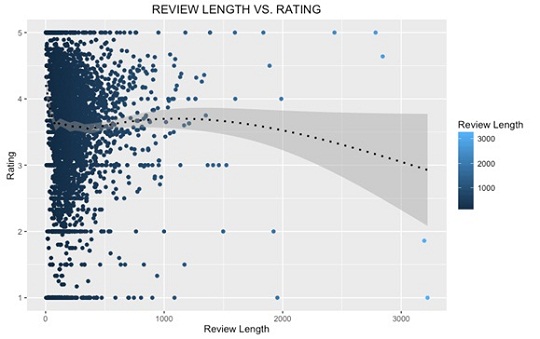
increase in the price.  The correlation is very close to zero and by removing the outliers the correlation remains weak (r = 0.01).



#### **5. RELATIONSHIP BETWEEN REVIEW LENGTH AND RATING**

The plot between average review length and rating will help us find out if the products with detailed reviews attract better rating. Here we can see that there is no correlation between both.



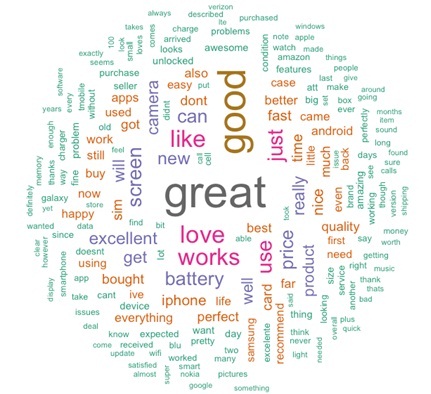
#### **6. RELATIONSHIP BETWEEN PRODUCT PRICE AND RATING**

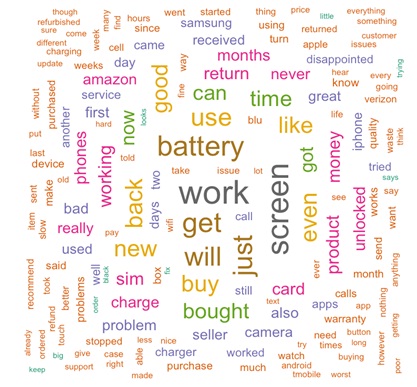
Now we’ll find out if costlier products have better ratings. This plot shows there is some correlation (r = 0.26) between rating and price. When consumers pay more for a product, they also expect better quality and sellers need to meet this expectation. It can be considered that with cost the product quality increases, which in turn leads to higher rating.



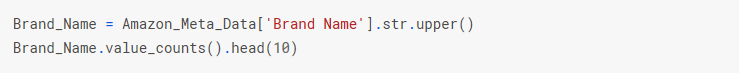
#### **7. WORD CLOUD**

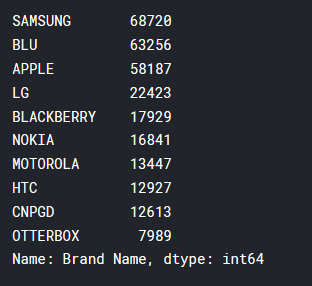
We segregated the reviews according to their ratings – positive reviews (4 or 5 star) and negative reviews (1 or 2 star). In both type of reviews there are certain common words like “work”, “battery” and “screen”. The most frequently used words in positive reviews are: “great”, “good”, “camera”, “price”, “excellent”, etc. In case of negative reviews words such as “return”, “back”, “problem”, “charge” are prevalent.

 **Figure: Word cloud of positive reviews**

 **Figure: Word cloud of negative reviews**

***8. TOP REVIEW COUNTS WITH BRAND***



***RESULT:***

***9. Mean and Median Price In Given Data:***

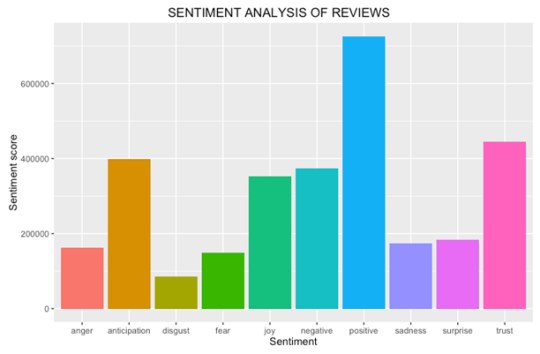




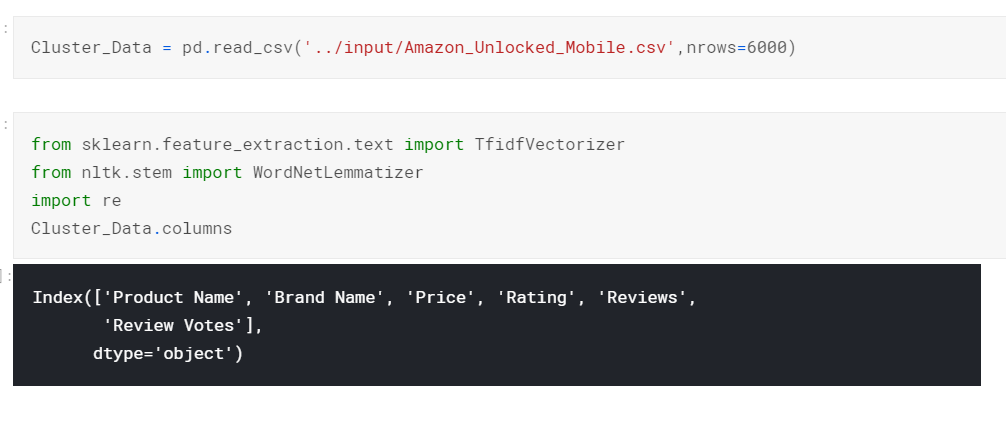


#### **10. SENTIMENT ANALYSIS**

The sentiment analysis shows that the majority of reviews have positive sentiment and comparatively, negative sentiment is close to half of positive. Among the eight emotions, “trust”, “joy” and “anticipation” have top-most scores. High scores for “joy” and “anticipation” could be because of the newly delivered phones. Also, the highest score for “trust” among all the emotions shows that the reviewers are writing the reviews with conviction and they trust the product.



**Customer Review Analysis using** **Kmeans Clustering**



### **CONCLUSION**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Amazon’s product review platform shows that most of the reviewers have given 4-star and 3-star ratings to unlocked mobile phones. The average length of the reviews comes close to 230 characters. We also uncovered that lengthier reviews tend to be more helpful and there is a positive correlation between price & rating. Sentiment analysis shows that positive sentiment is prevalent among the reviews and in terms of emotions, ‘trust’, ‘anticipation’ and ‘joy’ have highest scores.

It’d be interesting to perform further analysis based on the brand (example: Samsung vs. Apple). We can also look at building a model to predict the helpfulness of the review and the rating based on the review text. Corpus-based and knowledge-based methods can be used to determine the semantic similarity of review text. There are many more insights to be unveiled from the Amazon reviews.