Amazon Reviews Analysis

Richard Nkusi, Ray Pan, Yifei Wang, Yuki Ying

Business Problem

Our business problem for this project is to identify insights and make recommendations to both Amazon as a platform and to potential Amazon sellers. Given a dataset that contains Amazon reviews in the beauty category from years 2000-2015, we want to identify trends from reviews that can help Amazon improve its user experience and help sellers better market its products in the vast digital platform. More specifically, we hope to advise Amazon on the type of reviews that customers will find most useful and advise sellers on the topics that past consumers care the most about based on the past reviews. Our desired outcome is to help Amazon improve its reviews ecosystem and help sellers optimize their offerings through topic modeling and sentiment analysis.

Data Wrangling

Our dataset is obtained from Kaggle's Amazon US Customer Reviews Dataset (Appendix 1). With the provided data files, we have identified to focus on the reviews in the beauty category because as personal consumers of beauty products, we found product reviews to be one of the most influential points in determining our purchase. An ER model (Appendix 2) was built to help us better understand our data structure, and we found the following statistics through initial exploration. First, there are a total number of 5,115,666 unique reviews, a total number of 2,816,378 unique customers, and a total number of 588,817 unique products. The average star rating for all reviews is 4.187, and the rating distribution is skewed towards 5 stars (Appendix 3). We also noticed that the dataset contained oddly formatted categories that are not Beauty (Appendix 4). We then cleaned the data so that only the reviews under Beauty were analyzed.

After gaining an initial understanding of our data, we then loaded the data into AWS S3 and transferred it into our HDFS storage. We then converted our file into parquet format, as we have previously learned that data processing is faster through parquet.

Limitations of this dataset include the fact that there are no product categories for us to draw analysis from. Beauty is a big category for us to analyze from; even broad categories of skincare and makeup can have very different implications for the reviews. This will make it difficult for us to make meaningful suggestions for sellers. As for challenges, one main challenge will be formatting, as we are dealing with unstructured text with review data. There will also be the challenge of ambiguity and context dependency when analyzing topics and sentiments.

Insights and Analytics

The first step of our analytical process was to draw raw insights from the existing data. One observation is that there is an overall increase in the average rating over the years (Appendix 5). Reviews that participated in the Vine program were identified as more helpful (Appendix 6), but so were non-verified purchases (Appendix 7). As we found the observation of non-verified purchase reviews to be counter-intuitive, we ran additional queries and observed the following: non-verified purchase reviews have, on average, a lower star rating than verified purchase reviews, and lower rating views on average have a higher helpful ratio.

We then ran query analyses looking at review length. Reviews that participated in the Vine program are on average three times longer (Appendix 8), and reviews that are non-verified are on average two times longer (Appendix 9). We also found that review length is not correlated with star rating (-0.051), but is weakly correlated with helpfulness (0.238).

The second step of our process was topic modeling. We first created a word cloud to display distinct keywords that appeared in the review (Appendix 10). We then defined a function to clean the reviews and splitted the reviews into individual words and removed stop words. After transforming the bigrams into a vector of token counts and weighting the features based on the frequency, we developed an LDA model to perform the topic modeling.

We initially built an unigram for the topic words (Appendix 11). However, we found that the unigram results do not provide many meaningful insights. We then decided to build a trigram, as we are dealing with reviews and it may be more insightful if we examine longer topic words. We identified 7 topics with 20 iterations during the learning process (Appendix 12). Overall, the topic modeling results reflect a wide range of customer experience with beauty products, from satisfaction with product longevity and effectiveness to dissatisfaction with product performance and value. There are two categories of beauty products that are more popular which are hair and skin products. We also performed the topic modeling on the product titles to recommend which key words the seller could put to their product titles. We only chose products that have a star rating of 4 or 5 to conduct the topic modeling on product titles (Appendix 13).

The third step was sentiment analysis. We analyzed the review body to determine whether the reviews are of positive or negative emotional tone. To achieve this, we used two sets of csv files that contain positive and negative words that can appear in a review body and proceeded to count in each review body the number of positive words and negative words; the sentiment score is then calculated by ratio of difference of negative words from positive words and their sum (Appendix 14). Although we found that a higher star rating is associated with a higher sentiment ratio, it is important to note that even the average sentiment for 1 star ratings is around 0, suggesting an overall positive leaning in the sentiment of the reviews.

Discussion

In our model-free insights, we found that non-verified views were more likely to have a lower rating and be identified as helpful. This suggests potential authenticity issues for Amazon. In the future, we would advise Amazon to look into developing a verification system so that it ensures the people reviewing are authentic reviewers.

In our model-driven insights, we looked at topic modeling and sentiment analysis. In topic modeling, we found that consumers tend to review the beauty product relating to their personal experiences after usage, and common topics were longevity, usefulness, and the price of the product. Hair and skin were also common topics in the reviews. From this, we would recommend sellers on Amazon to send post-purchase reminders after 1 month to gather consumer insights. We would also advise sellers to consider carefully if they are entering the beauty category to sell hair care and skin care products, as there seems to be high competition in the Amazon market.

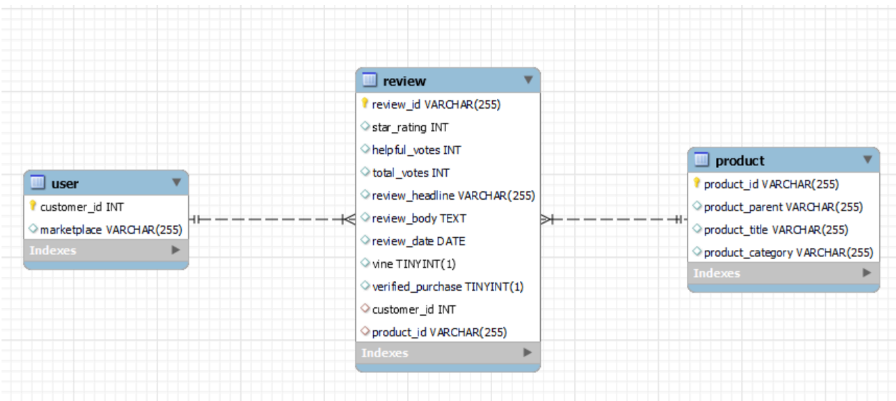
Our sentiment analysis provides us with the insight that 4-5 star ratings are really similar in sentiment ratio, and that the 1-3 star ratings get incrementally more negative in sentiment as the rating decreases. From this, we would advise sellers to focus on addressing concerns mentioned in the 1-3 star rating reviews for a higher opportunity of enhancing customer satisfaction.

One drawback of our analyses was that we were not able to provide more specific recommendations to Amazon and sellers. This is primarily tied with the limitation of our dataset of beauty being a broad category - we were not able to perform sub-category specific analysis. However, we believe the models that we have built can easily be distributed to Amazon and its sellers for more tailored analysis of each shop. Ultimately, we believe our findings and insights can help Amazon better its review ecosystem and help the sellers deploy better strategies.

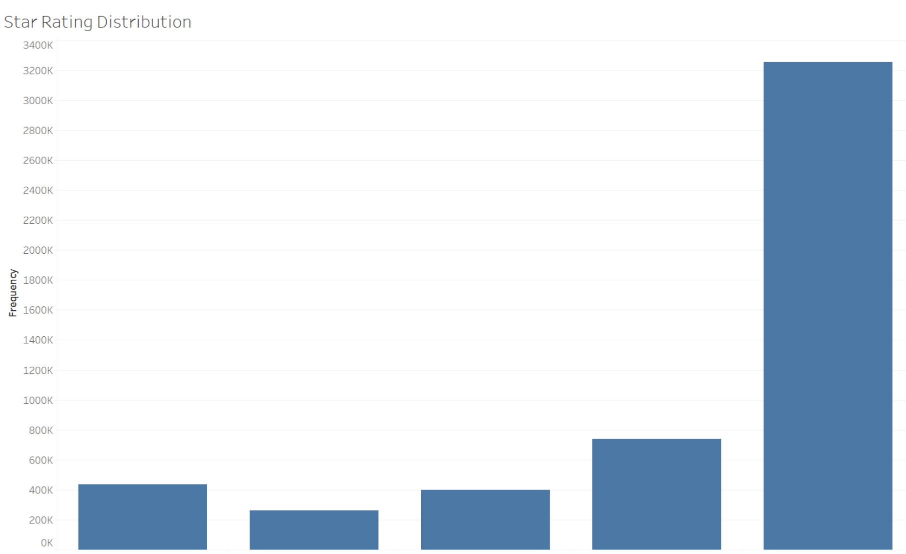
Appendix

Appendix 1: Link to dataset
https://www.kaggle.com/datasets/cynthiarempel/amazon-us-customer-reviews-dataset?rvi=1&select=amazon_reviews_us_Beauty_v1_00.tsv

Appendix 2: ER Model Diagram



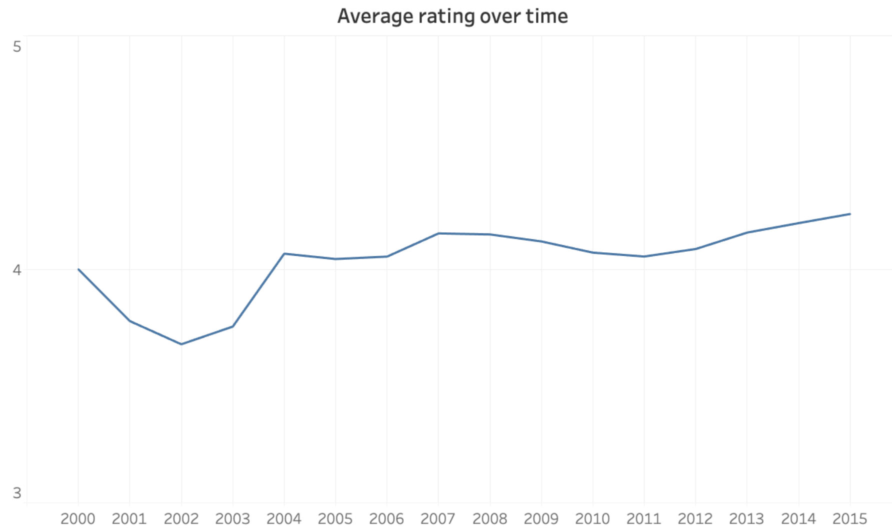
Appendix 3: Rating distribution



Appendix 4: Category entries check

```
+------------------+-------+
|  product_category|  count|
+------------------+-------+
|            Beauty|5115452|
|              null|    213|
|I don't get this ...|      1|
+------------------+-------+
```

Appendix 5: Time-series of average star rating



Appendix 6: Helpful ratio of reviews by Vine program participation

vine	total_reviews	helpful_votes_ratio_vine	helpful_votes_ratio_non_vine
Y	33309	2.374253204839533	null
N	5082143	null	1.7045639998717077

Appendix 7: Helpful ratio of reviews by verified purchase

verified_purchase	total_reviews	helpful_votes_ratio_verified	helpful_votes_ratio_non_verified
Y	4230196	1.4003497710271582	null
N	885256	null	3.1834497591657103

Appendix 8: Review length by Vine program participation

vine	average_review_length
Y	848.3096460416103
N	249.22392821248934

Appendix 9: Review length by verified purchase

verified_purchase	average_review_length
Y	214.6114842979396
N	437.16919289492654

Appendix 10: Word Cloud of beauty products

Appendix 11: Topic Modeling of beauty products using unigram

```
+-----+----------------------------------------------------------------+
|topic|terms                                                           |
+-----+----------------------------------------------------------------+
|0    |[good, not, hair, love, great, more, using, like, really, after]|
|1    |[not, hair, love, really, am, like, great, me, used, after]     |
|2    |[great, not, skin, like, me, even, hair, face, been, time]      |
+-----+----------------------------------------------------------------+
```

Appendix 12: Topic Modeling of beauty products using trigram

```
|0    |[1, 3, 4, 11, 13, 19, 49, 38, 57, 70]      |[0.011356165460777933, 0.010172945878968175, 0.009559917070164941, 0.00743780705529164,
0.006766849832880295, 0.0048265038407249924, 0.004512425100088468, 0.004225025724599102, 0.003444537785317589, 0.0034209213613736237]    |[g
oes long way, little goes long,  year old, lasts long time,  stars because, give  stars, doesnt last long, gave  stars, do not recommend, be
en  weeks]                     |
|1    |[0, 2, 36, 41, 1, 56, 71, 23, 114, 121]    |[0.07155294603040079, 0.0060438843478034, 0.0038932831342640027, 0.0037011926858203236,
0.0036800472290111313, 0.003099335024323915, 0.0029524841895364586, 0.0027721567871857914, 0.0026674315606936574, 0.002565991717255565]  |[
,  highly recommend, did not like, does good job, goes long way, leaves hair soft, does not last, works great , bit goes long, not purchase
again]                   |
|2    |[5, 6, 14, 37, 43, 66, 110, 103, 48, 102]  |[0.011456830315149148, 0.008527792737963722, 0.006881453022748348, 0.004757343688183043,
0.004287678423432634, 0.0036695316857324875, 0.003271064991883501, 0.0032177099000564474, 0.003095936095159717, 0.003069073403841871]   |[lo
ve love love, im not sure,  years ago, am not sure, works well , years ago , definitely purchase again, not worth money, after reading revie
ws,  not too]            |
|3    |[16, 27, 34, 15, 29, 25, 39, 17, 54, 72]   |[0.007761683530407832, 0.006268408037343434, 0.00584642415259604, 0.0049601918671313685,
0.004637609693037661, 0.004636738300572908, 0.004607695178343379, 0.004253307475857054, 0.004088665798126986, 0.0037782809393750562]    |[do
es not work, works really well, tea tree oil,  years old, leaves skin feeling, recommend anyone who,  works well, makes skin feel, dont know
why, didnt work me]     |
|4    |[26, 42, 45, 79, 91, 8, 120, 111, 224, 139]|[0.006379419257415196, 0.005418757091497364, 0.0048801257641554905, 0.003904639429115318,
0.0035622187658047966, 0.003297641819392486, 0.003263447049663485, 0.0027422518763110222, 0.0024846026237304837, 0.0024302135698399084]|[wor
th every penny, gets job done, do not like, cant beat price, not only does, over  years, cant say enough,  highly recommended, hard time fin
ding, doesnt work well]|
|5    |[24, 30, 46, 33, 51, 22, 105, 132, 59, 138]|[0.0066943908264188385, 0.004644042665995291, 0.004305993300913928, 0.003867991386243242
6, 0.0038661869976823744, 0.003657008359757857, 0.0032908082881374506, 0.0030562223303152547, 0.002957401287978959, 0.002927788727547924]|[d
ont waste money,  works great, vitamin c serum,  years now, cant go wrong,  did not, leaves hair feeling, been able find,  really like, am h
appy purchase]           ||
|6    |[10, 7, 44, 21, 75, 17, 87, 126, 124, 101] |[0.006654266998969794, 0.005411553084670335, 0.004963269794517515, 0.0047959510872266735,
0.0035077192263936303, 0.0032209441099978974, 0.002913354158166176, 0.002881211533110317, 0.00285535458975848, 0.002663490101468199]   |[did
not work, last long time, makes hair feel, does great job, well worth money, makes skin feel,  times week, makes hair look, makes skin look,
does not smell]      |
```

Appendix 13: Topic Modeling result of products title

```
+-----+-----------------------------------------------------------------------+
|topic|terms                                                                  |
+-----+-----------------------------------------------------------------------+
|0    |[hair, oil, oz, set, skin, spray, nail, 100, beauty, art]              |
|1    |[oz, black, brush, color, face, 3, ounce, hair, head, professional]    |
|2    |[ounce, oz, hair, skin, pack, oil, natural, fl, razor, eye]            |
|3    |[oz, 2, set, 8, ceramic, treatment, vitamin, travel, 1, anti]          |
|4    |[oz, hair, cream, body, men, pure, oil, facial, fluid, ounce]          |
+-----+-----------------------------------------------------------------------+
```

Appendix 14: Sentiment analysis output table

star_rating	positive_count	negative_count	ratio
1	11662.0	13562.0	-0.075325
2	9489.0	6793.0	0.165582
3	16020.0	7781.0	0.346162
4	40443.0	13017.0	0.513019
5	184975.0	45399.0	0.605867