Takeaway: Data Task - Amazon Reviews

Key findings

- Some niche categories are having high rating and share of positive reviews. By adding new product in them the overall sales could be boosted

- Longer reviews, which are describing the product in more neutral way, tend to be a driver for a purchase. Customers could be encouraged to make reviews in more unbiased style.

- Reviews with good rating for the main product tends to be repeated for related products which are bought together. Customers should be encouraged to leave reviews and on the bought together products

# What is the relation between the reviews and the helpfulness?

Weak correlation between helpfulness ratio and basic reviews stats but moderate positive connection between the rating and sentiment score

Weak correlation between helpfulness ratio and basic reviews stats but moderate positive connection between the rating and sentiment score

However, few observations on Helpfulness worthy to be mentioned:

- In absolutes, reviews giving higher rating are receiving more votes for helpfulness

- However, share of reviews with votes is higher in lower reviews giving low rating to the products

- Reviews giving higher rating are with higher coefficient of helpfulness

- Longer reviews are more likely to be rated as helpful

- Reviews with positive sentiment are more like to be evaluated as helpful

Correlation Matrix - Only for Reviews with Votes
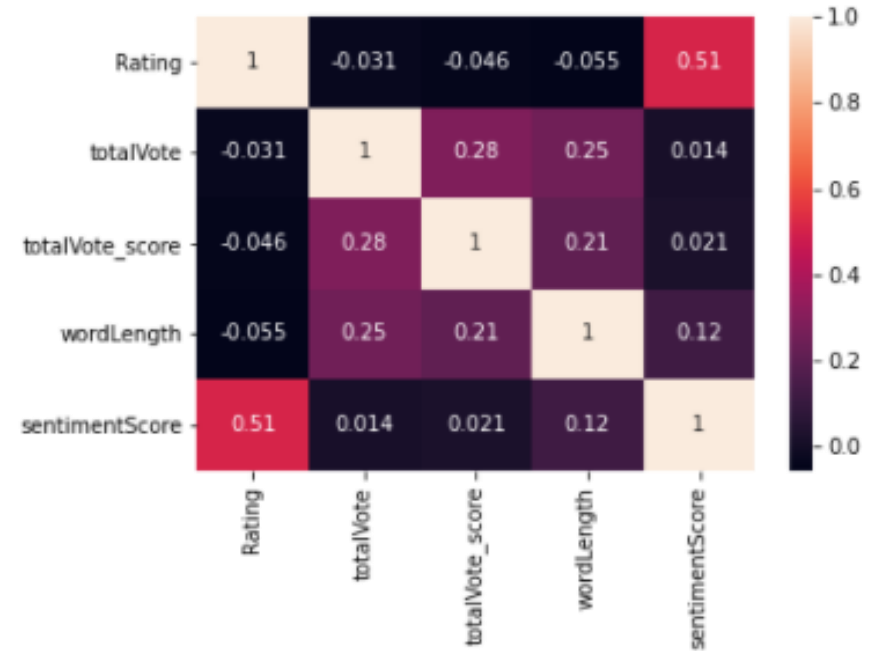
What is the review behavior among different categories?

Some niche categories are having high rating and share of positive reviews. By adding new product in them overall sales could be boosted

Some niche categories are having high rating and share of positive reviews. By adding new product in them overall sales could be boosted

category	N of products	avgRating	avgReviews	avgNegReviews	avgPosReviews	avgRevWithVotes	avgHelpfulness	avgSentiment
Clothing	10.8K	4.2	12.98	1.45	10.26	4.49	78.34	0.67
Shoes	6K	4.24	11.59	1.39	9.4	3.6	76.19	0.69
Jewelry	2.6K	4.35	11.07	1.5	9.13	3.78	80.48	0.71
No Category	1.2K	4.26	11.61	1.49	8.97	4.11	78.16	0.66
Watches	998	4.34	10.24	1.39	8.2	3.94	76.63	0.69
Sports & Outdoors	780	4.37	11.38	1.3	9.04	3.88	76.04	0.68
Toys & Games	244	4.26	9.06	1.37	7.29	2.79	74.21	0.67
Home & Kitchen	178	4.31	11.85	1.56	9.11	3.92	79.98	0.65
Health & Personal Care	95	4.31	10.62	1.68	7.6	4.57	77.82	0.6
Beauty	71	3.95	10.56	1.63	7.63	3.91	70.31	0.6
Arts, Crafts & Sewing	31	4.27	12.16	3.14	9.1	3.19	81.83	0.62
Patio, Lawn & Garden	19	4.34	29.21	3.5	25.53	10.79	75.85	0.69
Automotive	14	3.98	10.29	1	7.43	3.08	76.43	0.6
Kitchen & Dining	14	4.21	15.07	1.67	12.36	5.77	78.76	0.69
Industrial & Scientific	10	4.21	8.5	1.33	5.4	3.56	85.66	0.58
Electronics	9	4.46	8.22	1	7.33	3.89	78.29	0.79
Baby	6	4.39	9.5	-	8	5	61.69	0.77
Cell Phones & Access...	5	3.98	15.4	1.67	12.4	10.8	70.16	0.66
Camera & Photo	3	4.41	6.33	-	6.33	3.67	97.73	0.94
Software	3	4.32	19.33	4	16.33	10.33	66.97	0.77
Computers & Accesso...	2	3.81	6.5	1	5.5	3	75	0.63
Music	1	4.08	12	-	10	4	34.09	0.68
Home Improvement	1	4.58	36	-	32	9	85.71	0.79

- From the more mass product categories, 'Jewelry' is with best share of positive reviews and highest helpfulness ratio

- 'Arts, Crafts & Sewing' is with relatively high share of negative reviews and in same time high helpfulness ratio

- Patio, Lawn & Garden is very contradictory, relatively highest reviews but both positive and negative

- Category 'Industrial &' is with lowest sentiment score but having almost the second highest helpfulness ratio

https://datastudio.google.com/s/phw9ATpxERQ

Is there a relationship between price and reviews?

Higher priced products tend to have better reviews rating

Higher priced products tend to have better reviews rating

- Correlation between price and N of words

- Strongest correlation between price and N of words in the reviews with positive sentiment

- Filtering product with price 100$ or less (majority of reviews are in the category) are confirming the findings

- Overall highest average rating is for products in 300$ - 400$ range

- Higher priced products are having less reviews but more positive sentiment

- Higher priced products are better rating from reviews

- Helpfulness ratio in better for more expensive products
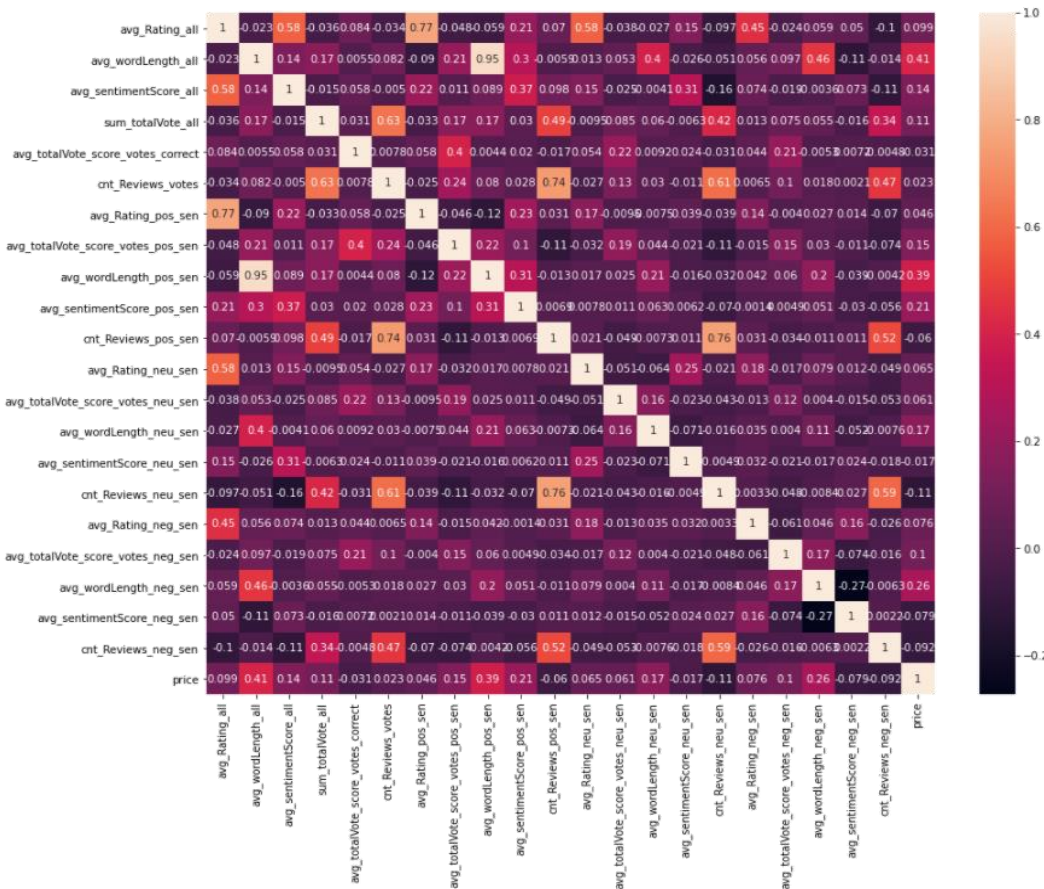
Which group of reviewers is more valuable to the business?

Can't be answered due insufficient data. However, we evaluate which reviews are more valuable to the business - total revenue of products bought after viewing.

Longer reviews are more likely leading to a purchase

- Reviews with neutral sentiment are having good correlation to the revenue from products bought after reviewing – mainly helpfulness ratio and N of words in the review

- Similar, but not that strong connection is observer for the reviews with neutral sentiment

- Expected, reviews with negative are having weak and even negative correlation to the revenue from products bought after reviewing

NOTE: *Very few from products bought after viewing are matched with price*
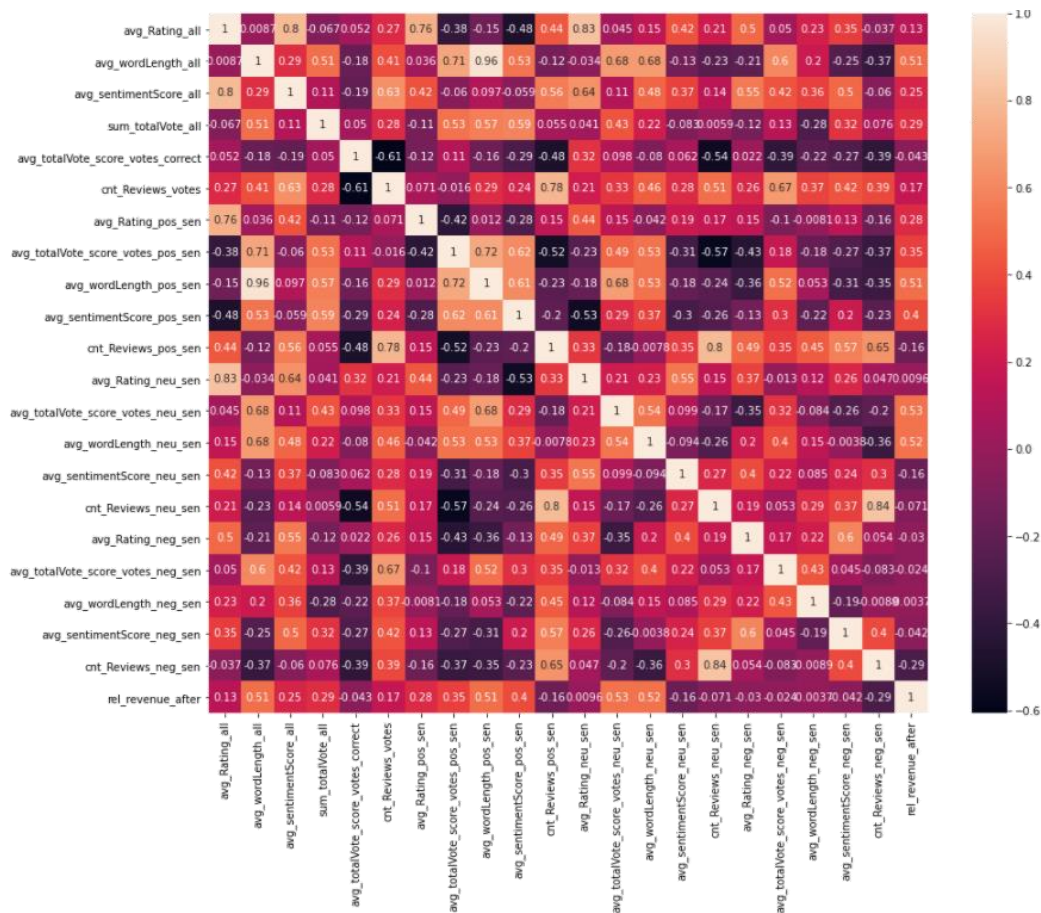
Is there a relation between reviews from products which are bought together?

Strongest similarity is observer on rating, sentiment and number of words in a review

Strongest similarity is observer on rating, sentiment and N of words

- Rating are correlating strongest – 0.39

- Followed by average number of words (0.36) and the sentiment score (0.3)

- Negative correlations is observed on:
  - Rating of the main product and number of reviews with negative sentiment on the related products

  - Number of words in main product reviews and number of reviews with positive sentiment on the related products

  - Helpfulness of reviews for the main product and number of reviews with negative sentiment on the related products
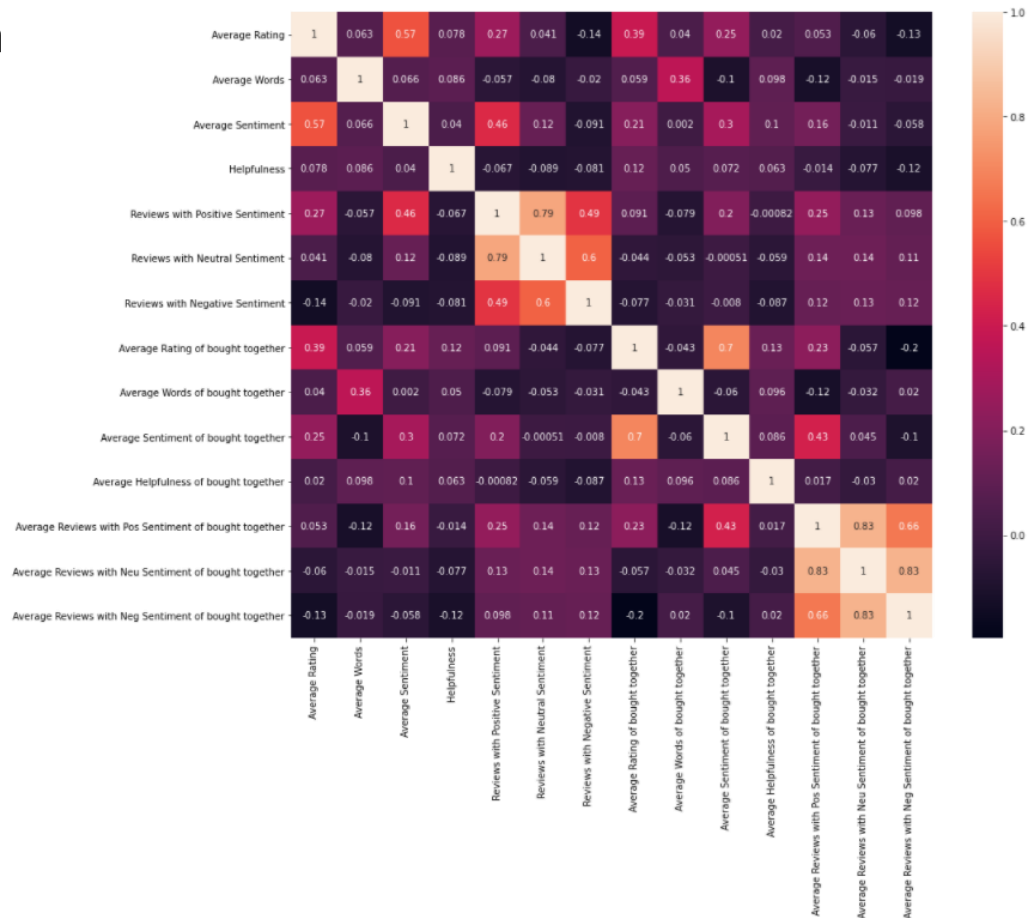
Direction for next steps

Direction for next steps

Areas for improvement of the analyses
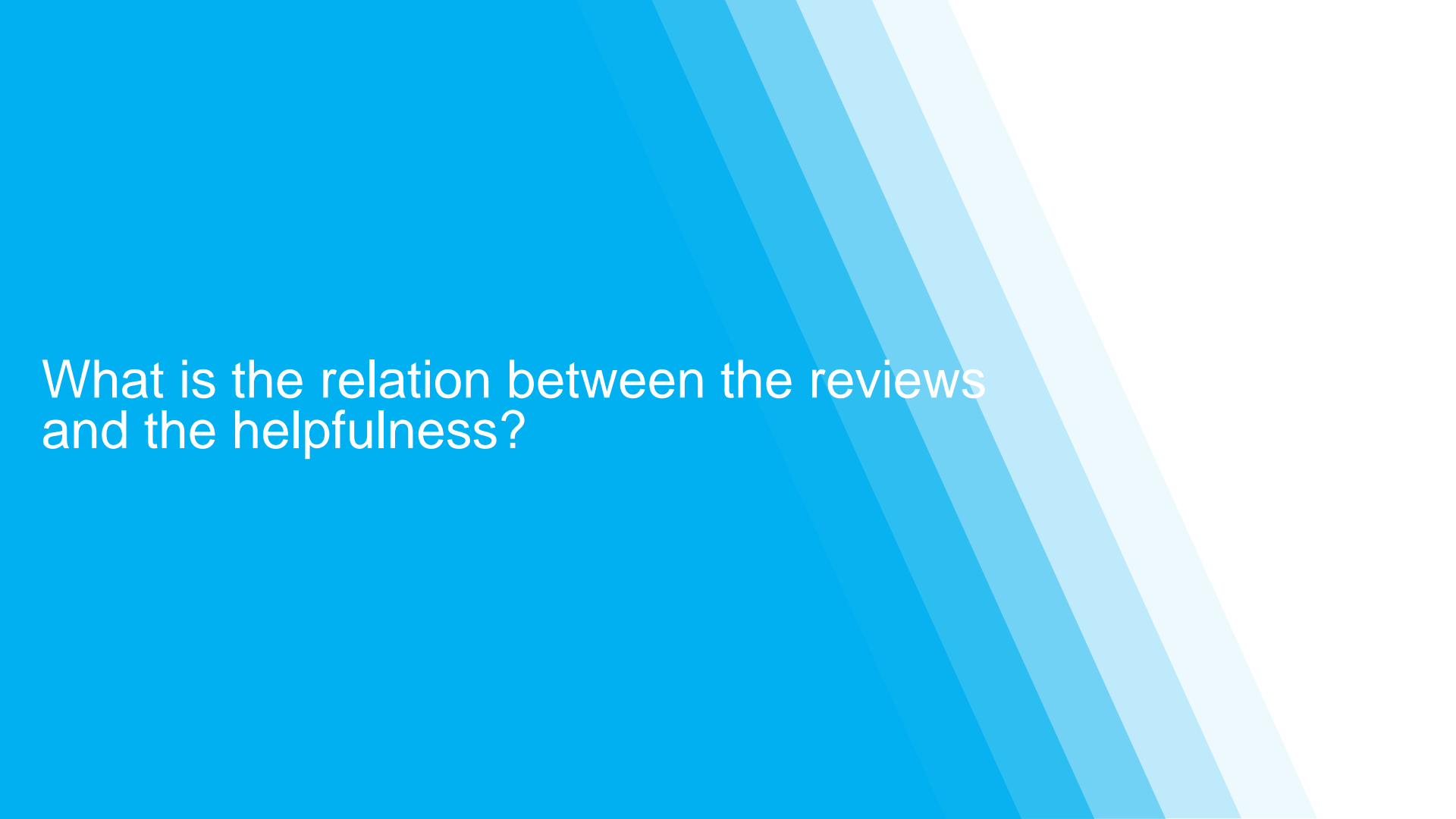
- Analyze and exclude the outliers

- Data on reviewer level to be enriched with the actual purchases and user's lifetime values

- Get help from a data engineer to simplify and optimize the code

Next step for additional analyses

- Make user segmentation – find the reviewers with most useful reviews and offer them to try and review other products

- Market basket analyze for better understanding of the related products and eventually offer combos of products

- Analyses using 'word of bags' to validate which key words are more likely to trigger a purchase

BACK UP

What is the relation between the reviews and the helpfulness?

Observations on Helpfulness worthy to be mentioned

- In absolutes, reviews giving higher rating are receiving more votes for helpfulness

- However, share of reviews with votes is higher in lower reviews giving low rating to the products

- Reviews giving higher rating are with higher coefficient of helpfulness

- Longer reviews are more likely to be rated as helpful

- Reviews with positive sentiment are more like to be evaluated as helpful
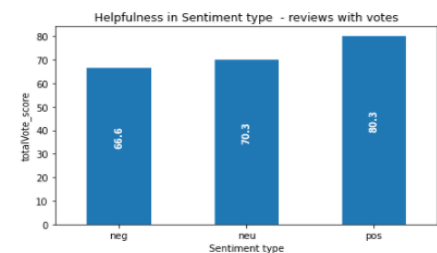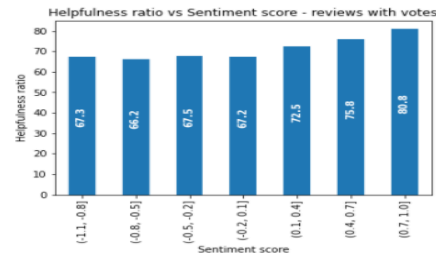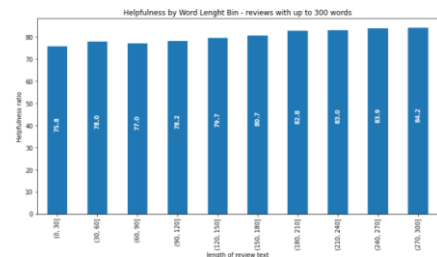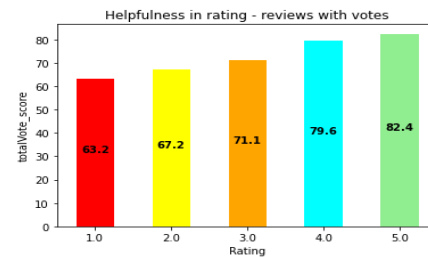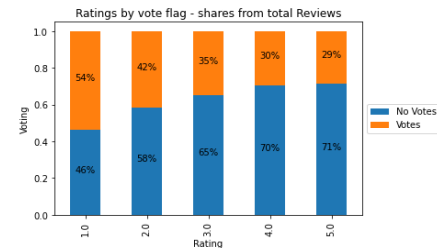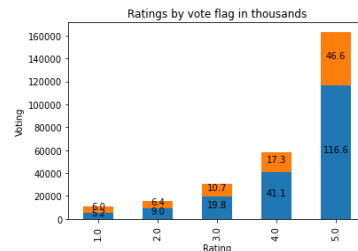
Note: Most of the reviews are up to 300 words. That's why helpfulness vs words length is filtered

What is the review behavior among different categories?

Details of the Data Studio report – some screenshots but should be demonstrated on live
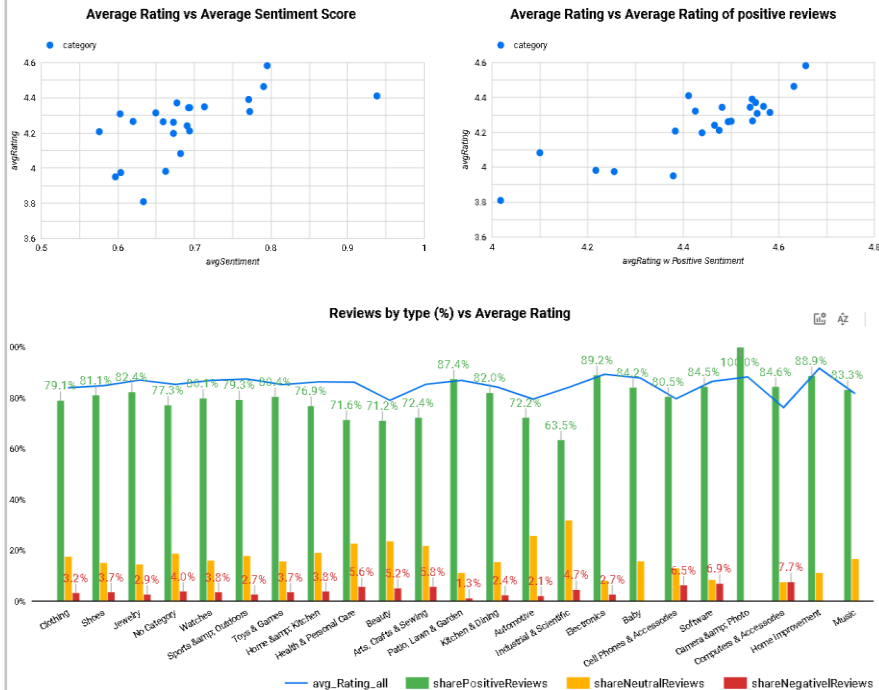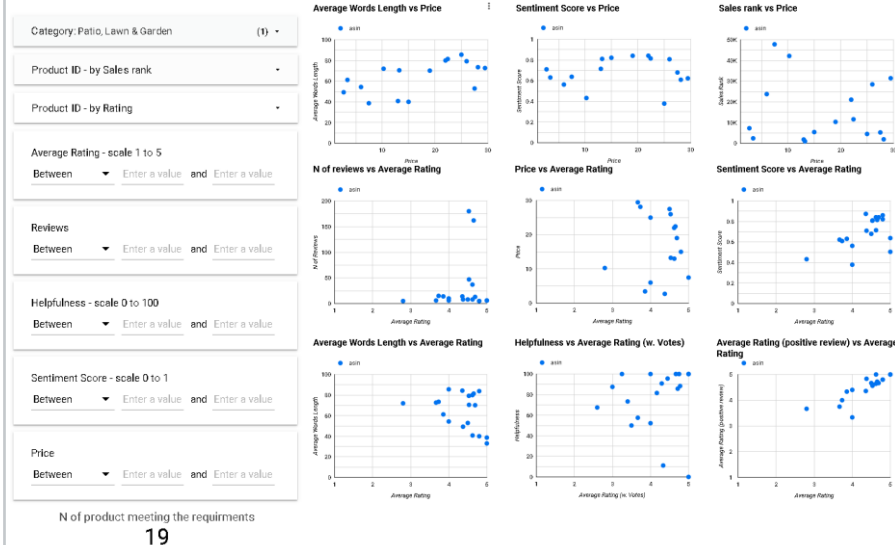
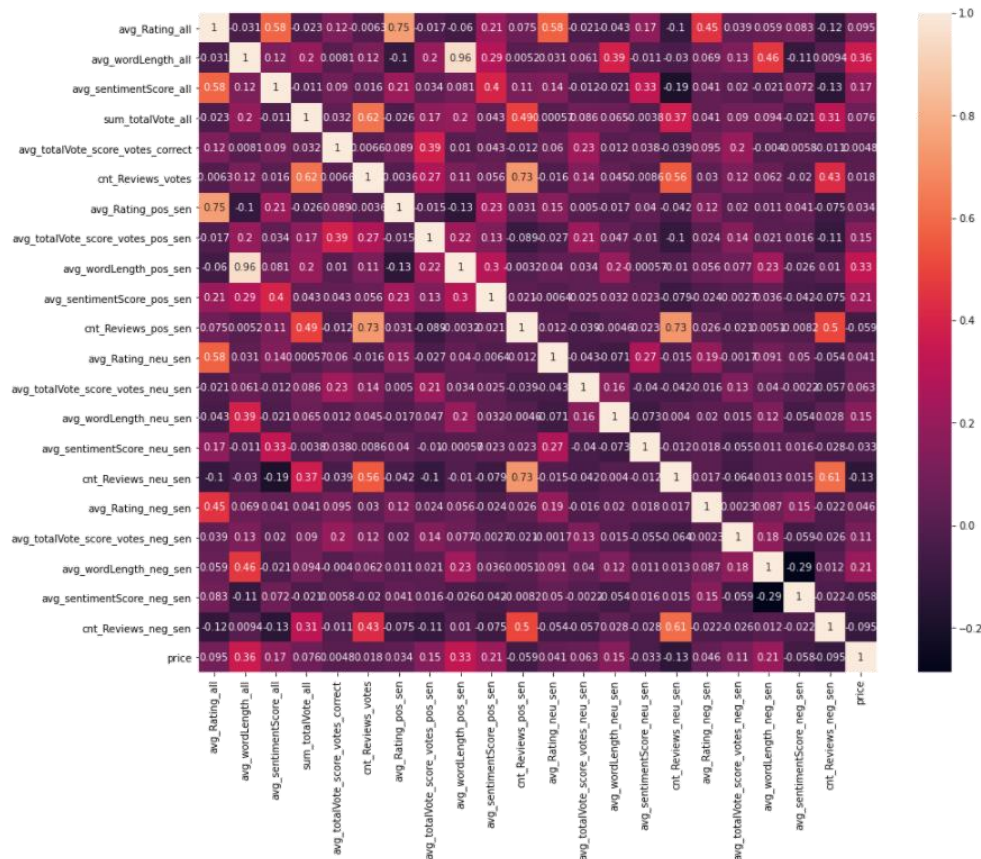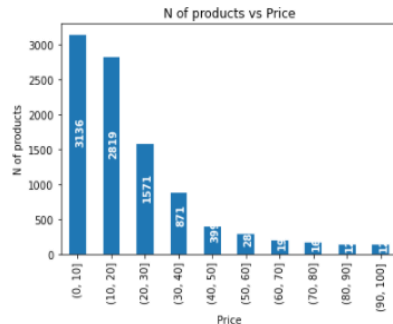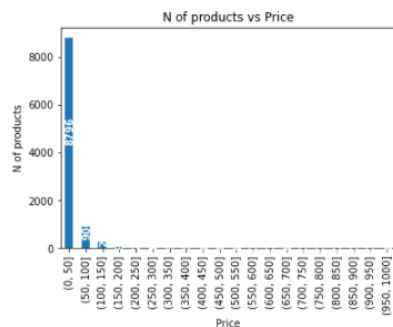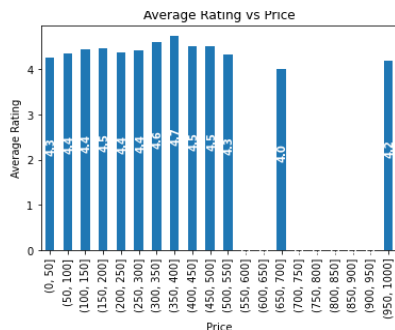Is there a relationship between price and reviews?

Zooming correlation matrix for products with price 100$ or less

More relations between price and reviews stats

- Higher priced products are better rating from reviews

- Higher priced products are having less reviews but more positive sentiment

- Helpfulness ratio in better for more expensive products

- Mid range products (<= 200$) are having longer reviews