Final Project Report — STAT/CS 287
TEAM NAME: Red Viper
DATE: 12/15/2021

Group Members: Sarah Grobe, Bailey Kretzler, Ben Temkin

Title: Amazon Review Sentiment is Correlated to GDP and General Satisfaction Across Time

---

Project title

Amazon Review Sentiment is Correlated to GDP and General Satisfaction Across Time

Abstract

Using Amazon review data downloaded from a repository, we were able to classify reviews into which aspect of a product they were primarily reviewing, as well as record the related sentiment. This allowed us to see which aspects of a product have the biggest and smallest impact on a review, as well as the most positive and most negative impact. It was ultimately found that a product's performance has the largest impact on whether or not a review will be written about the product, as well as the sentiment towards that product and in turn what its overall rating may be. We also correlated sentiment to socioeconomic factors such as gross domestic product and general satisfaction with the state of the country. Here we found that sentiment showed a positive linear relationship with satisfaction and a negative linear relationship with GDP. Further, these relationships varied by product category highlighting its importance for contextualizing sentiment analysis. We provide a sentiment analysis that may be useful to both producers and consumers that are trying to make marketing or purchasing decisions respectively.

Introduction

Amazon has become a central hub for e-commerce where countless products are reviewed hundreds of times with varying degrees of helpfulness. Many consumers rely on these reviews when making online purchasing decisions. Further, companies depend on product reviews to make product improvements and marketing decisions. Where companies lack the resources to respond to all reviews, sentiment and product aspect analysis can be used as a robust tool to assess the polarity and context of a review. Further, associating review sentiment with product aspects can help identify ways marketing or product design can be improved. By also analyzing the breakdown of these reviews by product aspect, sellers are able to gain insight into which product aspects are most important to consumers, that is, which aspects motivate them the most to leave positive or negative reviews. Sentiment has been also shown to be associated with the review helpfulness as perceived by consumers [Zeng et.al. 2020], which is linked to product profitability [Ghose and Ipeirotis 2011] suggesting importance in understanding consumer purchasing habits. Therefore economic trends may also be helpful in understanding sentiment of Amazon reviews.

Previous studies identified links between profits and a product's review sentiment and rating. An analysis on camcorder and digital camera reviews revealed that certain positive evaluations of these products has a positive impact on sales [Archak et.al. 2011]. Further, in the same study, higher numerical rating showed a positive relationship with product sales [Archak et.al. 2011]. Interestingly, Archak et.al. also found that incorporating text data into their model reduced the effect of average rating on sales, suggesting an impact of review sentiment on consumer decisions, and thus profits. In a similar study, negative product reviews were associated with improved product sales if reviews were highly informative [Ghose and Ipeirotis 2011]. Together, these articles suggest an impact of review sentiment and context on a product's profits, affirming the importance of understanding review sentiment for market research.

Sentiment analysis of product reviews is a well established method for understanding consumer behavior and perception of products. A sentiment analysis performed on Amazon review data focused on visualizing sentiment to provide consumers with insight beyond simply the 1 through 5 star ratings [Bhatt et.al. 2015]. These visualizations were designed to be easily analyzed and consistent across product categories to allow for reasonable comparison.,[Bhatt et.al. 2015]. Such visualizations highlight one way sentiment analysis can be used to benefit consumers, and shows an analysis similar to the one we do here. However, we intend to go a bit more in depth by analyzing sentiment for product types and for certain aspects of those products.

Another sentiment analysis on amazon reviews of tablet computers observed the relationship between user reviews and product sales [Li et.al. 2019]. It was found that just within the single product category, different product aspects were discussed in positive and negative reviews. Positive reviews focused mostly on product functionality, while negative reviews often complained about logistical product aspects such as the user interface and customer support. Additionally, they found that both the numerical and written review played a role in the success of a product. Here Li et.al show that the product aspects prevalent in positive and negative reviews are distinct and ascertainable through sentiment analysis. Therefore, it is likely that the product aspects which consumer reviews focus on differ across product categories and importantly, are not always based solely on functionality. Thus, there is value in knowing what these aspects are on a per-category basis. Given knowledge of product aspects on a per-category basis, developers and manufacturers can meaningfully improve current and future product designs. As such, we feel it is important to discern both review sentiment and the primary product aspects being referenced in the review.

It is also worth noting that reviews and their sentiment are heavily rooted in sociology, and thus the socioeconomic context in which reviews are written can influence sentiment. Specifically, it is interesting to understand how gross domestic product (GDP) correlates to review sentiment across time. Increases in GDP across time have been correlated to levels of e-commerce in Germany and Romania [Pantelimon, Georgescu, & Posedaru 2020]. It is then not unreasonable to assume that fluctuations in GDP or other economic indicators would alter the level of eCommerce and potentially sentiment towards products. It is possible that consumers more harshly review certain products when fluctuations in GDP are observed. The relationship between review sentiment and a population's general satisfaction with their country may be also useful in understanding trends in sentiment. Specifically, if people are decreasingly satisfied with their country they may again be more likely to review things hashly due to an overall negative outlook. Examining such trends can enhance the power of sentiment analysis and provide new tools to companies conducting market research.

Given the current literature, it is clear that sentiment analysis is a useful tool for both consumers and producers when making purchasing and marketing decisions respectively. Sentiment has been correlated to overall profits and methods have been developed to make this information more accessible to consumers. Further, there is an apparent relationship between product aspects emphasized in reviews and product profits and review sentiment. However, limited studies have utilized product aspects on multiple product categories, limiting the scope of these studies. Many studies also fail to account for the socioeconomic context of reviews which may be an important predictor of purchasing habits and consumer satisfaction. Therefore, we hope to examine the following 3 questions: 1) which product aspects are most prevalent in Amazon reviews across categories? 2) how are these product aspects associated with review sentiment? 3) how is review sentiment across time related to GDP and general satisfaction with the state of the country? Through answering these questions we hope to fill some of these knowledge gaps and improve the use of sentiment analysis for both producers and consumers.

Datasets to be used

Amazon review data is fortunately widely available. We have identified a repository of 24 datasets for multiple product types [He and McAuley 2016]. Each dataset contains the reviewer ID, the product ID, the helpfulness rating of the review, the product review text, the numerical rating of the product, the summary of the review, the review time of posting in unix time, and the date of the review posting. We drew our subsets from the de-duplicated dataset, which contains 82.83 million reviews. This deduplication accounts for copy and paste reviews and reviews from users with multiple accounts making it preferable for sentiment analysis. Metadata for these reviews was also drawn from this repository, allowing us to include product categories to each review in our dataset.

Quarterly GDP data was obtained via datahub and satisfaction data was obtained from Gallup's repository of polling data. The GDP data contained the current GDP level for each quarter of a year from 1949 to 2017. Satisfaction data contained the proportion of respondents who express satisfaction, dissatisfaction, or no feeling regarding the current state of the country. This data ranged from 1971 to 2021 but did not contain every month.

Methods

To examine how Amazon review sentiment and prevalent product aspects varied across product categories, we utilized exploratory data analysis and sentiment analysis on reviews by category and product aspect. These analyses were first done on a subset of data to improve analysis efficiency and then conducted on the whole data set once the analysis was known to be robust and useful. Further, to determine if sentiment was related to social and economic indicators, we examined the relationships between GDP, overall sentiment, and satisfaction with the country's direction. These relationships were analyzed using linear regression on the sentiment, satisfaction and GDP data across time.

Data Subsetting

Deduplicated Amazon review data ranging from the year 1997 to 2014 was downloaded from an online repository maintained by the McCauley Lab at the University of California San Diego. The raw review data were joined with metadata available from the same repository in order to gain information about the products that the reviews were written for. From this joined data set, only data fields of product category, helpfulness rating, review text, and review date (month and year) were retained. In order to develop the analysis and gain a general sense of trends in the data, a subset of the data was taken including approximately ten million of the eighty million reviews. The reviews were distributed evenly across the twenty four categories, equalling slightly over four hundred thousand per product category [data-parser.py, lines 58-59]. Reviews were read-in in parallel, by category [data-parser.py, lines 58-59] and checked to ensure that they included the retained metadata as described above [data-parser.py, lines 79-104]. Then the raw reviews, sorted by category, were written to a temporary json [data-parser.py, lines 108-113] to be read in later when the reviews in each category were sorted by month and year [data-parser.py, lines 168-187]. As reviews are sorted by date within each category the required meta data is extracted [data-parser.py, lines 138-142] to reduce the size of the output file containing the sorted reviews [data-parser.py, lines 193-197] which will be used for further data exploration and analysis in place of the raw data.

Data Exploration

The subset data was loaded in as a JSON [EDA_REDVIPER-fast.py, lines 208-209] containing reviews and ratings organized by category and then date as month and year. Distributions for review length, date, and overall product rating were examined to understand variation in the dataset using histograms generated in matplotlib [EDA_REDVIPER-fast.py, lines 217-220]. Yule coefficients were calculated for each category by separating reviews into high and low rated reviews and counting the frequency of words shared amongst these review types [EDA_REDVIPER-fast.py, lines 56-94]. The yule coefficient was then determined as the frequency of the words in high rated reviews minus the frequency of words in low rated reviews divided by the total frequency of the word in both [EDA_REDVIPER-fast.py, lines 149-155]. Product aspects were identified based on the top 100 high and low leaning words for each category (e.g. cheap in low = negative price aspect vs. in high = positive price aspect.) [EDA_REDVIPER-fast.py, lines 156-159].

Product aspect classification

Using the results of the exploratory data analysis (EDA) as a guide, we chose to classify reviews into one of six general product aspects: price, durability, aesthetics, performance, ease of use, and service. Of these six, price, durability, and performance are fairly self-explanatory, mainly referring to the cost, strength, and speed/execution of the product, respectively. The aesthetics aspect refers mainly to the appearance of the product, as well as some other subjective features such as comfort and size (in the case of clothes, for instance). The ease of use category mainly refers to installation, set up, and use of the product. Finally, the service category refers mainly to the customer service associated with the product, such as interactions with the seller or any warranties that are applicable to the product.

The EDA was then also used to create a list of words for each of these six aspects which are indicative of each of that aspect [aspect_classification_charts-fast.py, lines 90-108]. These lists were created using word stems so as to encompass as many related words as possible.

For instance in the case of the “price” aspect, the presence of the stem “pric” is used to classify the review, as searching for this string would account for not only the word “price,” but also other forms such as “prices,” “priced,” or “pricing.” While this makes the classifications more powerful as they are able to search for account for a larger number of words in the reviews, it also adds some uncertainty to our results, as this may result in certain product aspects having more relevant words to search for than others, thereby potentially biasing our analysis by making some product aspects more common than others. However, we should note that this type of bias is, to a certain degree, unavoidable.

Looping through the review text for each of the reviews, the algorithm then searches for the classifier words for each of the five product aspects, keeping a tally of how many appear for each aspect [aspect_classification_charts-fast.py, lines 114-141]. Once all of these classifiers have been searched for in the text, the name of the product aspect with the largest tally (that is, the aspect which had the most relevant words appear in the text) is appended to the review [aspect_classification_charts-fast.py, lines 189-211]. For simplicity, as well as time and space efficiency of the program, each review is being categorized as a whole into one of the five product aspects. We then analyzed the counts and compound sentiment for these five aspects, not only across all data, but also for each of the 24 product categories.

Sentiment analysis

Sentiment analysis was done on all reviews using the VADER sentiment analyzer (Hutto and Gilbert 2014). VADER is specifically designed for use on social media text, however has been tested on a number of other document types, including amazon reviews, and has been shown to perform at or above the level of other sentiment analyzers (Hutto and Gilbert 2014). VADER provides the positive, negative, neutral, and combination valence (referred to as the “compound” sentiment) for a given document, which in our case is amazon reviews. These compound sentiments were evaluated across time for all categories, as well as within each category and date pairing to determine if there were trends specific to a product category or the data more broadly.

Further, sentiment analysis was conducted for reviews according to which product aspect they were reviewing, based on the exploratory data analysis conducted on word counts as described above. Average compound sentiment was recorded for each of the six aspects overall, as well as for the six aspects within each of the 24 product categories [sentiment_over _time_files-fast.py].

Comparing to Economic Indicators

In order to understand how review sentiment correlates to the state of the economy, we compared trends in sentiment to trends in quarterly GDP and Gallup polling data regarding feelings about the country's direction. GDP data and gallup polling (satisfaction) data were available for all years present in the Amazon review data set, however GDP data was only available on a quarterly basis and coverage of polling data was inconsistent. From the GDP and satisfaction data the current level of GDP and proportion of satisfied respondents were used, respectively [data_parser.py: line 199-257].Data were converted to date formats compatible with the review data and joined based on the quarterly data (i.e. year 01, year 04, year 07 and year 10) [linear_regression.py: line 25-67,71-114]. The joined quarterly data was used to determine the linear relationship between all unique combinations of sentiment, satisfaction and GDP [linear_regression.py: line 116-161]. Linear trends were visualized with matplotlib as scatterplots with a regression line [linear_regression.py: line 169-208].

Results

Sentiment Analysis

To understand the results of the sentiment analysis on the review data, we can begin by looking at how the compound sentiment varied over time across all product categories and aspects, as shown in Figure 1 below.

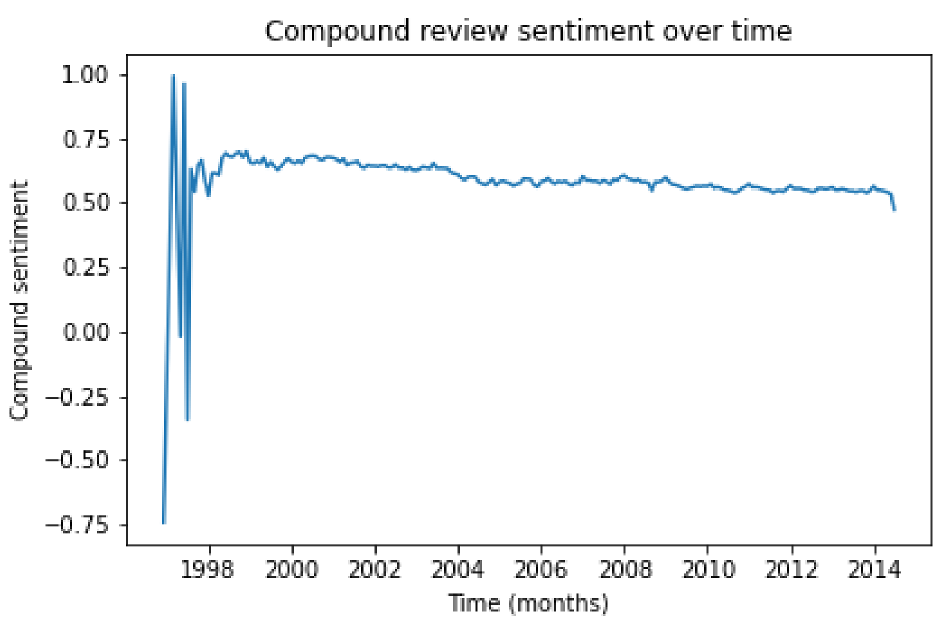

Figure 1. Plot of average compound review sentiment over time across all categories and aspects.

The most notable feature of the above figure is the wide range of values within the first few years. This is most likely due to the relatively low number of Amazon review data from the late 1990s, due to the internet and Amazon itself being less accessible and less widely used. As time goes on, however, the variation seems to decrease significantly, which corresponds with the large increase in review data over this time period. Looking at the trend overall, there appears to be very little change in sentiment over the years. In other words, Amazon reviews tend to skew more positive than negative, with a nearly constant compound sentiment of around 0.6.

Sentiment of the product reviews was then broken up by the six product aspects within each of the 24 categories, in order to help gain a sense of which aspects were most important to reviewers and had the strongest associated sentiment.

Category	Price	Durability	Aesthetics	Performance	Ease of use	Service
Overall	25%	11%	25%	30%	7%	3%
Amazon_Instant_Video	27%	8%	21%	36%	4%	4%
Apps_for_Android	17%	3%	16%	43%	20%	2%
Automotive	25%	10%	17%	29%	18%	2%
Baby	19%	11%	35%	23%	10%	2%
Beauty	29%	7%	27%	28%	5%	4%
Books	19%	12%	20%	39%	7%	3%
CDs_and_Vinyl	22%	7%	21%	43%	3%	3%
Cell_Phones_and_Accessories	27%	13%	18%	34%	5%	3%
Clothing_Shoes_and_Jewelry	25%	11%	47%	13%	2%	2%
Digital_Music	21%	7%	23%	41%	5%	3%
Electronics	30%	8%	19%	34%	6%	2%
Grocery_and_Gourmet_Food	36%	8%	23%	20%	5%	9%
Health_and_Personal_Care	27%	9%	19%	37%	5%	3%
Home_and_Kitchen	23%	12%	29%	26%	7%	2%
Kindle_Store	24%	9%	17%	38%	9%	2%
Movies_and_TV	26%	10%	24%	33%	2%	4%
Musical_Instruments	34%	12%	19%	26%	6%	3%
Office_Products	28%	11%	24%	29%	6%	3%
Patio_Lawn_and_Garden	21%	13%	26%	30%	8%	3%
Pet_Supplies	21%	13%	32%	26%	5%	2%
Sports_and_Outdoors	24%	14%	29%	25%	7%	2%
Tools_and_Home_Improvement	24%	11%	21%	32%	10%	2%
Toys_and_Games	23%	18%	32%	19%	6%	3%
Video_Games	25%	7%	24%	34%	7%	3%

Table 1. Distribution of reviews by category

In the above table we can see the volume of reviews for each aspect within each of the 24 categories. Using this information, we can try to gauge which aspects, if any, had the most reviews written about them, and therefore which aspects are most important to reviewers within each category. For example, the CDs and Vinyl category had 43% of its reviews written about the performance of the product, indicating that this aspect seems to compel the most consumers to write a product review.

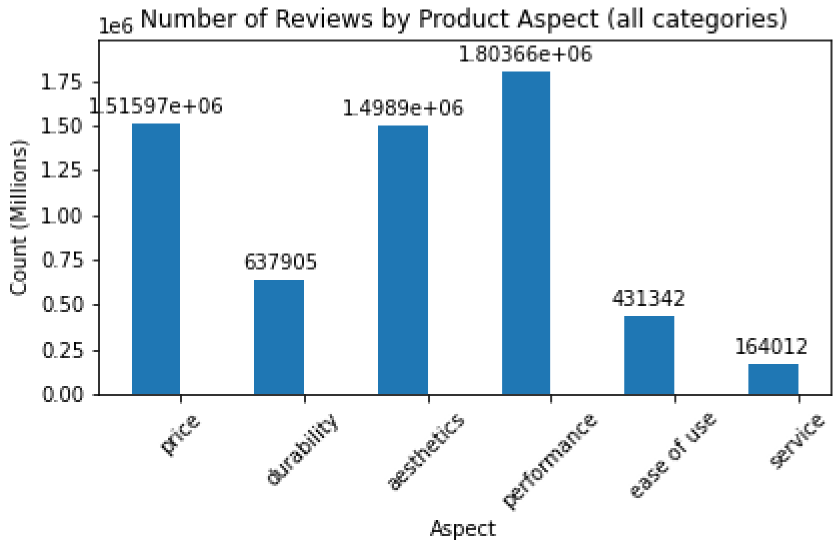

Figure 2. A visualization of the number of reviews for each product aspect across all categories. This graph represents the first row in Table 1 above.

Using the figure above, we can gain a better sense of the breakdown of each aspect across all categories. We can see that in general, performance had the most reviews written about it, while service had the fewest. It therefore seems that performance is the most motivating factor when a consumer is deciding whether or not to leave a review; in turn, this could indicate that how well (or poorly) a product performs will have the largest impact on the product's overall rating, which can have a significant impact on whether or not a prospective buyer will actually buy the product. Similarly, customer service seems to have the lowest effect.

We can then go further and look at the sentiment analysis results for the six aspects within the 24 categories, and ultimately compare these with the counts and percentages above.

Category	Price	Durability	Aesthetics	Performance	Ease of use	Service
Overall	0.62	0.46	0.62	0.52	0.66	0.40
Amazon_Instant_Video	0.60	0.45	0.63	0.55	0.68	0.48
Apps_for_Android	0.56	0.45	0.61	0.46	0.61	0.40
Automotive	0.56	0.38	0.49	0.46	0.60	0.29
Baby	0.67	0.55	0.69	0.57	0.75	0.36

Beauty	0.65	0.42	0.63	0.54	0.65	0.40
Books	0.60	0.50	0.60	0.55	0.69	0.45
CDs_and_Vinyl	0.78	0.70	0.80	0.78	0.78	0.68
Cell_Phones_and_Accessories	0.51	0.27	0.55	0.39	0.58	0.22
Clothing_Shoes_and_Jewelry	0.67	0.51	0.63	0.57	0.73	0.47
Digital_Music	0.75	0.68	0.78	0.76	0.78	0.68
Electronics	0.61	0.42	0.63	0.45	0.61	0.27
Grocery_and_Gourmet_Food	0.68	0.49	0.66	0.64	0.76	0.53
Health_and_Personal_Care	0.57	0.41	0.53	0.46	0.65	0.31
Home_and_Kitchen	0.63	0.45	0.64	0.56	0.76	0.33
Kindle_Store	0.59	0.55	0.63	0.59	0.72	0.45
Movies_and_TV	0.64	0.45	0.61	0.54	0.62	0.50
Musical_Instruments	0.70	0.52	0.69	0.59	0.71	0.50
Office_Products	0.55	0.44	0.58	0.45	0.59	0.26
Patio_Lawn_and_Garden	0.54	0.38	0.52	0.41	0.63	0.27
Pet_Supplies	0.60	0.48	0.59	0.45	0.69	0.35
Sports_and_Outdoors	0.62	0.46	0.60	0.53	0.66	0.35
Tools_and_Home_Improvement	0.55	0.37	0.53	0.47	0.59	0.28
Toys_and_Games	0.67	0.60	0.70	0.57	0.78	0.49
Video_Games	0.64	0.44	0.64	0.51	0.62	0.45

Table 2. Average compound sentiment of aspects by category

In the table above we can see which product aspect had the strongest sentiment, both positive and negative, for each of the 24 product categories. For instance, within the category Toys and Games, ease of use had the highest compound sentiment, while service had the lowest. This indicates that reviewers of products in this category generally felt most positive about ease of use and most negative about service, compared with the other product aspects.

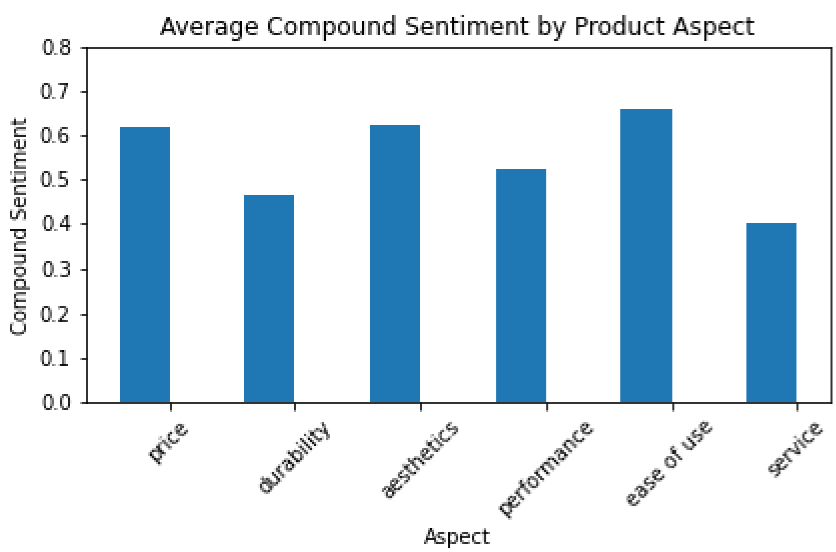

Figure 3. The average compound sentiment for all six product aspects across all product categories. Average compound sentiment ranges from -1 to 1. This is a visualization of the first row in Table 2 above.

Looking at the above figure, we can see a clear split of three product aspects with generally higher sentiment (price, aesthetics, and ease of use), and three with generally lower sentiment (durability, performance, and service). Interestingly, the ease of use aspect second lowest volume of reviews (as shown in Figure 2 above) while having the highest compound sentiment, indicating that those who are compelled to write a review about this aspect generally do so because they feel positively about it. It should also be noted that the compound sentiment ranges from -1 to 1. Given this range, all compound sentiments fall within roughly 10% of it, between about 0.4 and 0.6. This shows that across all six product aspects, those who leave product reviews generally feel positively about the product, with a roughly consistent ratio of positivity to negativity.

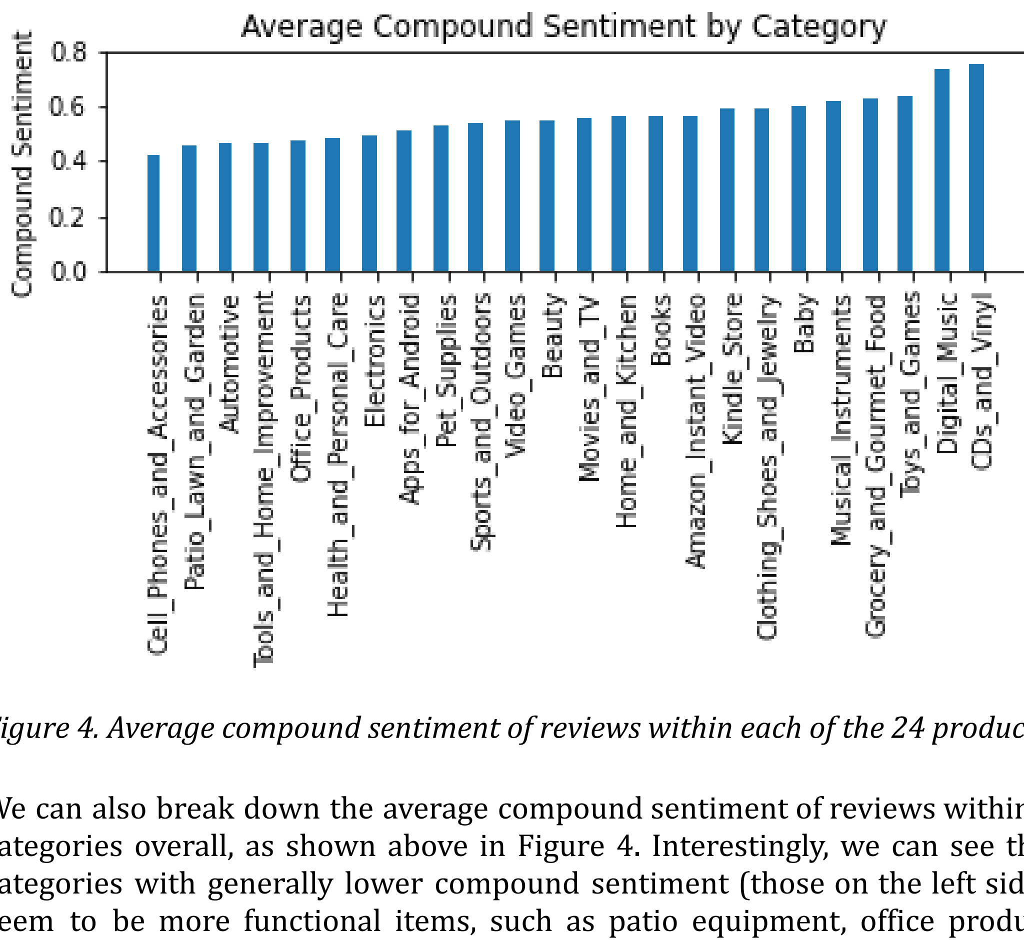

Figure 4. Average compound sentiment of reviews within each of the 24 product categories.

We can also break down the average compound sentiment of reviews within each of the 24 categories overall, as shown above in Figure 4. Interestingly, we can see that the product categories with generally lower compound sentiment (those on the left side of the graph) seem to be more functional items, such as patio equipment, office products, and home improvement tools. By contrast, those on the right side of the graph have higher compound sentiment, and seem to be primarily leisure or entertainment related items, such as toys and music. Relating this back to the analysis of product aspects which showed performance as being most important to reviewers, it makes sense that more functional items would be rated more critically, as the performance of something like a drill or a chainsaw is more important and significant than the performance of a CD, for instance.

Finally, we can dive deeper into these review sentiments by looking at a few notable product categories in particular.

Reviews for the Beauty category of products tended to fall in the mid range of compound sentiment compared with the other product categories, as shown in Figure 4 above. Using the following two figures, we can take a closer look at the distribution of reviews and their sentiment across the six product aspects.

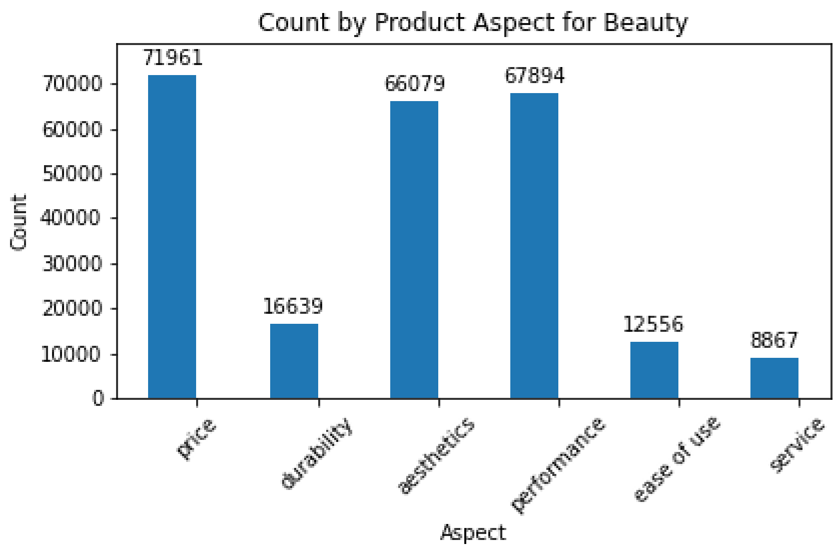

Figure 5. Volume of reviews for each of the six product aspects within the Beauty products category.

We can see a very sharp contrast in the volume of reviews for the various product aspects, as shown in Figure 5 above. The aspects of price, aesthetics, and performance have very high review counts, indicating that a lot of reviewers see these aspects as the most important for this category, and aer most motivated to write reviews about these aspects. Comparatively, the aspects of durability, ease of use, and service seem to be the least important aspects to reviewers, as the vast majority of reviews in this category were not related to these aspects.

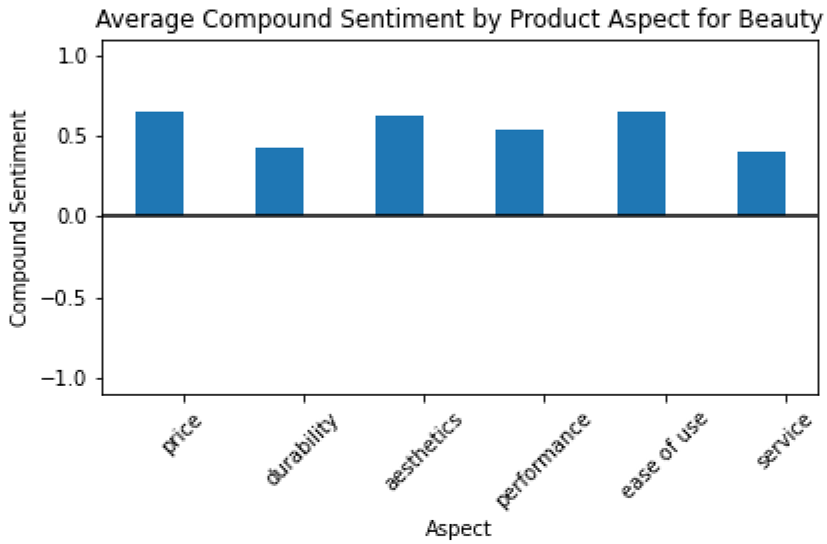

Figure 6. Average compound sentiment for each of the six product aspects for all reviews within the Beauty category. This is a visualization of the Beauty row in Table 2 above.

The results of Figure 5 show an interesting connection to those of Figure 6 above. Generally, the product aspects with relatively higher compound sentiment (price, aesthetics, and performance) match those which also had a high volume of reviews. Durability and service had generally lower volumes of reviews coupled with generally lower overall sentiment. This shows that users who were compelled to write about these seemingly less important aspects such as durability and performance would more often have a negative sentiment towards it, compared with the other aspects of higher volumes. The interesting exception to this is the ease of use aspect, which had a relatively lower review count but relatively higher compound sentiment, indicating a similar idea: it seems those who were compelled to leave a review about the ease of use of a product in this category were driven by a more positive sentiment towards it.

Products falling in the Baby category generally had a higher compound review sentiment, as shown in Figure 4 above. Looking closer at the breakdown of counts and sentiment for the six aspects within this category may also provide us with more insight about higher rated products.

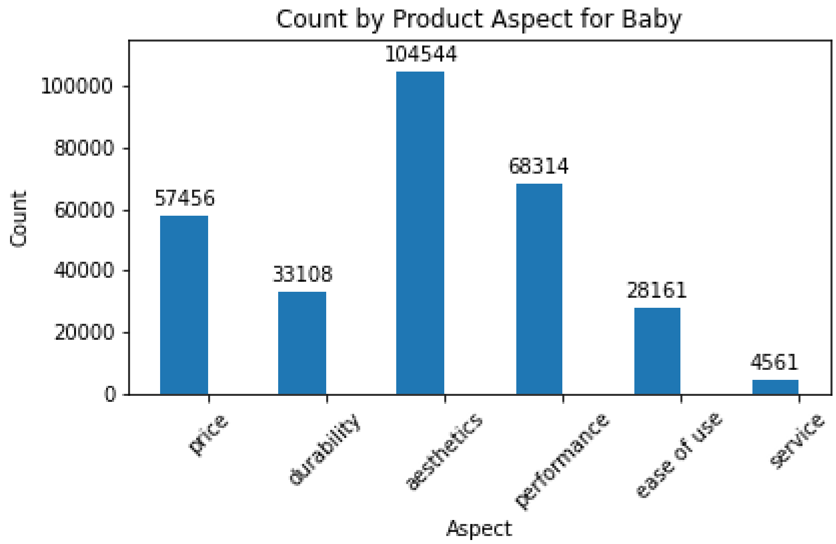

Figure 7. Volume of reviews for each of the six product aspects within the Baby products category.

As shown above in Figure 7, the aesthetics aspect had the highest volume of reviews within this category, indicating it was generally the most important to reviewers. The service aspect had the lowest volume, indicating it was generally least important, and service related to the product was the least compelling reason to leave a review.

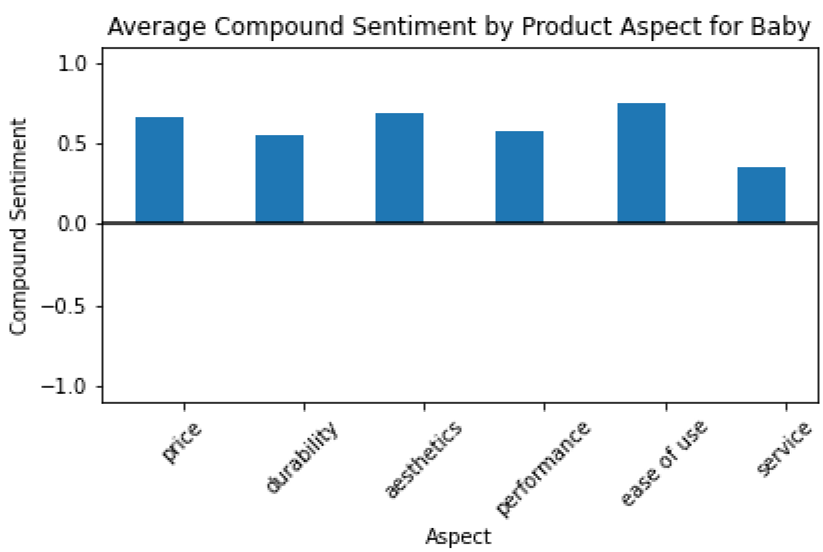

Figure 8. Average compound sentiment for each of the six product aspects for all reviews within the Baby category. This is a visualization of the Baby row in Table 2 above.

As shown in Figure 8 above, the general sentiment for all six aspects was positive. Interestingly, the aspects with the highest and lowest average sentiment, ease of use and service, were also those with the second lowest and lowest volume of reviews, respectively. We could therefore suppose that it is strong positive or negative sentiments which compel reviewers to leave a review about these particular aspects, as opposed to purely its importance to the product.

Automotive - generally low compound sentiment

Finally, we can take a look at the Automotive product category, as reviews in this category had generally lower compound sentiment compared with other categories.

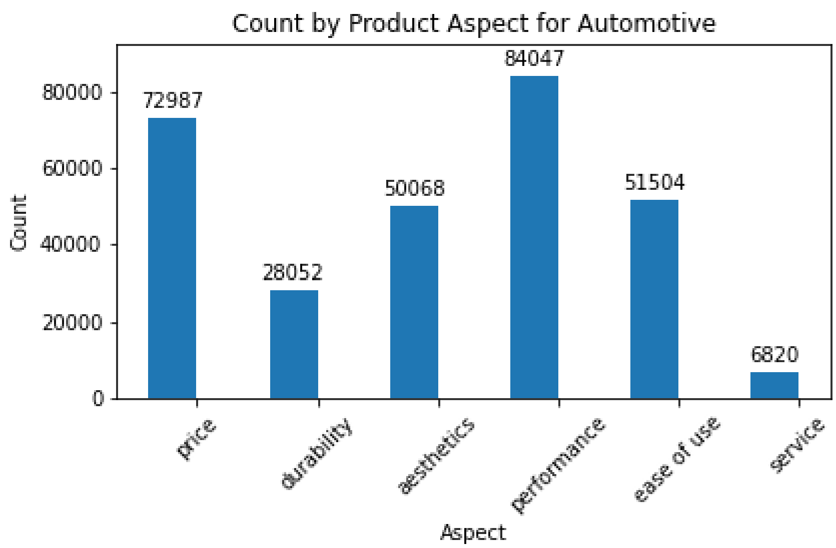

Figure 9. Volume of reviews for each of the six product aspects within the Automotive products category.

We can see from Figure 9 above that performance and price aspects seemed to be the most important to reviewers, as they had the largest volume of reviews. Durability and service have the lowest volume, and therefore be the least compelling aspects for consumers writing reviews.

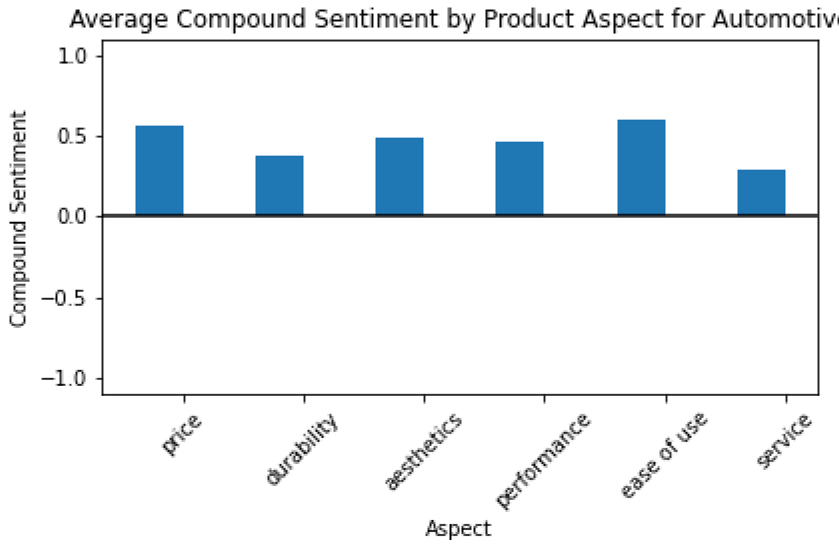

Figure 10. Average compound sentiment for each of the six product aspects for all reviews within the Automotive category. This is a visualization of the Automotive row in Table 2 above.

Interestingly, Figure 10 above seems to show a slightly different trend from the previously analyzed product categories. The performance aspect, while having the highest volume of reviews as shown in Figure 9, has one of the lower compound sentiments compared with others. This shows that performance of a product in this category is not only important to reviewers, it also has potential room for improvement for products within this category. Similarly, improved performance of an Automotive product could lead to improved product ratings overall, as a larger number of users will be more compelled to write more positive reviews of the product.

Review Sentiment and State of the Economy

In general, GDP increases across time whereas sentiment and satisfaction both decrease across time. Sentiment across time shows a significant linear relationship with both gross domestic product(GDP) as well as overall satisfaction. As GDP increases, it appears that sentiment associated with overall product reviews decreases (Figure 11a). Alternatively, as satisfaction increases so does sentiment indicating a positive linear relationship between these variables (Figure 11b). The relationships of GDP and satisfaction to sentiment reflect the relationship of GDP to satisfaction which is a negative linear relationship where satisfaction decreases as GDP increases. When these relationships are examined across different product categories, they are generally consistent. In some cases the slope appears steeper and/or the p-value indicates a non-significant relationship, indicating that the relationship of GDP or satisfaction to sentiment may be dependent on the product in question. Specifically, Movie and TV products as well as Automotive products showed a significant positive linear relationship between GDP and Sentiment, contrary to the overall trend (Figure 12a,b). It is also worth noting the irregularities in the relationship between sentiment and GDP for the electronics category. While a significant linear regression was fit to this relationship, the points show strong deviation from the line around GDP of 13000 (Figure 13a). When looking at the sentiment across time graphs for the electronics category, a similar trend is observed with a dip in sentiment around 2006-2007 which may explain the observed trend in sentiment by GDP (Figure 13b)

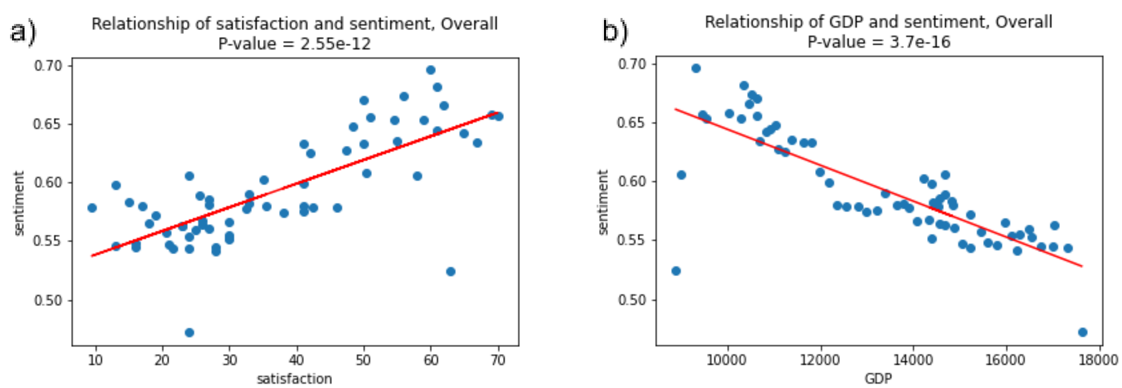

Figure 11 Scatter plots depicting the linear relationship between a) GDP and sentiment and b) satisfaction and sentiment. Points represent monthly averages of sentiment and either variable, red lines depict the fit of the linear regression.

Typically, satisfaction shows a significant positive relationship to sentiment - that is, as satisfaction increases sentiment also increases (Figure 11a). However for some product categories, there is a significant or near significant (alpha = 0.05) negative relationship of sentiment to satisfaction. Specifically, Movie and TV products as well as Automotive products again show a relationship that diverges from the overall relationship where sentiment decreases as satisfaction increases (Figure12 c,d). It is important to note here that Movie and TV products do not meet the p-value cut off of an alpha of 0.05. However P-value cut-offs are arbitrary and there still seems to be sufficient evidence against the null slope hypothesis. Further, electronics again show considerable divergence from the regression line due to the fluctuation in sentiment around 2006/2007.

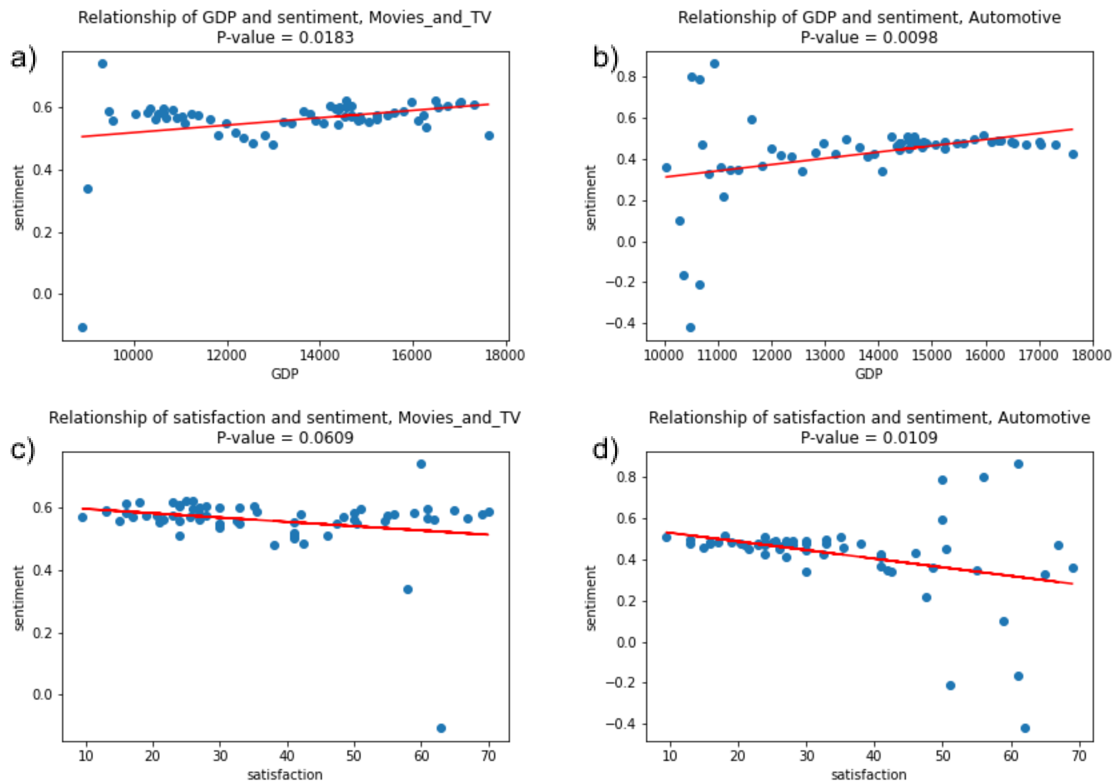

Figure 12: Scatterplots depicting the linear regression for a) GDP and sentiment for the Movie and TV product category, b) GDP and sentiment for the Automotive product category, c) satisfaction and sentiment for the Movie and TV product category and d) satisfaction and sentiment for the Automotive product category. Points represent monthly averages of sentiment and either variable, red lines depict the fit of the linear regression.

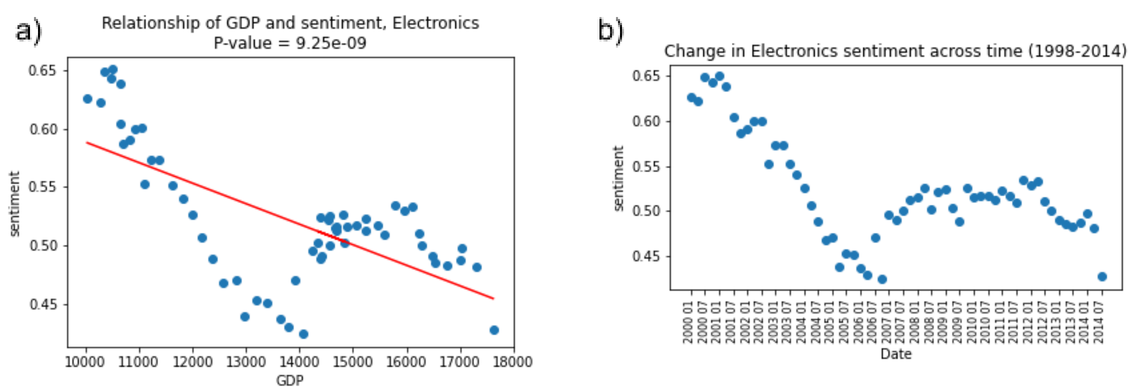

Figure 13: Scatter plots depicting a) the linear regression between GDP and sentiment for the electronics product category and b) the sentiment across time for reviews in the electronics category. Points represent monthly averages of sentiment and either variable, the red line depicts the fit of the linear regression.

It is possible that the relationship of both GDP and satisfaction to sentiment in the Movie and TV products as well as Automotive product categories is driven by review sentiment across time. Both of these product categories show little change in sentiment over time. Since GDP and satisfaction increase and decrease across time respectively, their relationship to an unchanging sentiment will follow this same trend. The unchanging sentiment of reviews associated with these products may be linked to consumer expectations surrounding them. For example, you often buy a movie or TV product either because you know you like it or you want to watch it and oftentimes you understand what you are getting. Further these may be framed as luxury purchases, which consumers may be less likely to review harshly or review at all. Automotive products, alternatively are more likely to be needs or repeat purchases meaning, again, you know what to expect.

The trend observed in sentiment of reviews associated with Electronics, however, has no clear logical explanation and thus must be rooted in some external factor. For example, quality of the product may have changed due to changes in availability of materials used in manufacturing. Alternatively, new electronics may have been released opening up new complaints and product comparisons which may have impacted the assessment of reviews associated with the service or quality product aspects.

Conclusion

Overall, our results indicate influences of both product aspects and socioeconomic trends on review sentiment. Additionally, these influences vary across product categories due to differences in the nature of the product. While the performance aspect seemed to be most important to reviewers across all categories and therefore holds the largest weight in a product's overall rating, individual categories have certain aspects that have a higher weight, such as the importance of aesthetics and price in beauty products specifically. Further, the relationship of GDP and overall satisfaction with the direction of the country and review sentiment vary by category. Such variation indicates that product characteristics and significance to our lives may be meaningful factors to consider when interpreting sentiment analysis.

While our results are promising there are several logistical improvements that could be made and many directions to take future work. The most notable improvement would be to replace our review classifier, ideally with a neural network, to improve the accuracy and range of the aspect classification. One possible limitation of our current design is that we only attribute one aspect to each review, instead each review could be classified as all aspects with a high enough confidence . Another improvement could be made to our data architecture to allow us to more easily handle larger quantities of data. In that same vein, including more reviews, specifically more recent ones, might give a better picture of the present day consumer's concerns. Another interesting direction would be to take the economic status of the population into account more directly. Specifically, wage changes and unemployment over time could be used instead of generalized economic indicators like GDP. Additionally, if possible, separating reviews by country might also lead to interesting findings about the purchasing mindsets of different groups of peoples. Looking at sentiment over time for each category might give us a more in depth view of how consumer values have changed over time.

Our analysis provides an important starting point for making sentiment analysis more useful to both consumers and producers. We provide relevant insights to the impact of product aspects to overall and with category review sentiment. Further, we connect overall and within review sentiment to trends in socioeconomic data which may be used as predictors of consumer needs in future analyses. Lastly we provide helpful figures that allow both consumers and producers to understand the potential strengths and weaknesses of a product based on six product attributes. These visualizations could be helpful for producers in making marketing decisions or product improvements or consumers in making purchasing decisions.

Bibliography

Archak, N., Ghose, A., & Ipeirotis, P. G. (2011). Deriving the pricing power of product features by mining consumer reviews. *Management science*, *57*(8), 1485-1509.

Bhatt, A., Patel, A., Chheda, H., & Gawande, K. (2015). Amazon Review Classification and Sentiment Analysis. *International Journal of Computer Science and Information Technologies*, *6*(6), 5107–5110.

Ghose, A., & Ipeirotis, P. (2011). Estimating the Helpfulness and Economic Impact of Product Reviews: Mining Text and Reviewer Characteristics. IEEE Transactions on Knowledge and Data Engineering, 23(10), 1498-1512.

He, R., & McAuley, J. (2016). *Amazon product data*. Amazon review data. Retrieved November 29, 2021, from https://jmcauley.ucsd.edu/data/amazon/.

He, R., & McAuley, J. (2016, April). Ups and downs: Modeling the visual evolution of fashion trends with one-class collaborative filtering. In *proceedings of the 25th international conference on world wide web* (pp. 507-517).

Hutto, C.J. & Gilbert, E.E. (2014). VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. *Eighth International Conference on Weblogs and Social Media (ICWSM-14)*. Ann Arbor, MI, June 2014.

Li, X., Wu, C., & Mai, F. (2019). The effect of online reviews on product sales: A joint sentiment-topic analysis. *Information & Management*, *56*(2), 172-184.

Pantelimon, F., Georgescu, T., & Posedaru, B. (2020). The Impact of Mobile e-Commerce on GDP: A Comparative Analysis between Romania and Germany and how Covid-19 Influences the e-Commerce Activity Worldwide. *Informatica Economica, 24*(2/2020), 27-41.

Zeng, Z., Zhou, Z., & Mu, X. (2020). User review helpfulness assessment based on sentiment analysis. Electronic Library, 38(2), 337-351.