Summary

The Amazon is one of the biggest online shopping platforms nowadays. Since it offers all kinds of goods for consumer to choose and compare, this way of shopping is gradually replacing the traditional street markets.

The Sunshine Company wants to release three products on Amazon: Hair dryer, Microwave oven and baby pacifier. And we are aim to make some suggestion as a reference for Sunshine Company. We use the sales data from Amazon for these three products with the description of goods, star rating and reviews from year 2002 to 2015. We base the sales of three products model on two main factors:

- Star rating
- Reviews

We identify the verified purchase and predict the reason of that. To predict the consequence of the unverified purchases, we will give some suggestion based on this. Meanwhile, all the efficient data would be used later.

Finding the distribution of the star rating, the average star of each product is likely to achieve in the future can be conclude. we found that the star rating of all three goods is all quite polarized, and mostly consumer would mark as 5-star if the product does not have problem.

Next step it finding the trend of number of purchases at different star rating, so that we plot the data from 2013 to 2015 and we can see an obvious incline for all goods at different star rating level. But the number of goods which been marked at certain level do not have too much change when time goes by.

To evaluation how would review or star rating affects other buyers, we calculating the percentage of helpfulness of the review and results showed that the polarized comment has the largest impact for future consumer on decision making.

We use the efficient data at our first stage to find the real thought from previous consumer. By extracting common words they used in all comments, the characteristics of each product which would have a big impact on star rating.

Contents

1. Overview

1.1 background introduction

Recently, there is increasing amount of online shopping in our daily life. Amazon, which is the one of best online marketplace in the world, offers variety of goods from different department. Before deciding to make the payment, there are some comments about the product from previous consumers which are likely to have some impact on decision making. Therefore, the star rating and comments are quite important for all sellers.

Sunshine Company wants to introduce and sell three products: microwave oven, baby pacifier and hair dryer. We will analysis the data for these three specific goods (from Amazon) to give some suggestion about how to get a great marketing achievement.

1.2 problem restatement

- How to extract the efficient data? What cause some data invalid?
- How does the star rating distribute?
- What cause the star rating different?
- How would number of purchases change due to different star rating?
- How would star rating affect future consumers?
- What quality should each product improve?

2. Assumption

To simplify the problems, we make the following basic assumptions, each of which is properly justified.

1) All unverified purchases should be removed.
2) Most products would be marked as 5-star.
3) The number of sales would increase when time goes by.
4) People would affect by the reviews of 1-star and 5-star rate most.
5) There exist some common factors which always appear in reviews.

3. Notation

Abbreviation	Description	Unit
r	Percentage of verified purchase	%
p	Percentage of helpfulness	%

4. Analysis and Modelling

4.1 Extract efficient data

1) Removing the unverified purchased

We try to extract all the verified orders and removing those who bought it with a deep discount or just wrote a comment without pruchase.

For the first kind of orders, since the consumer got a large discount and they would pay less than others, they would not fair to take the measure of the goods itself. Even if some of them realized the existence of discount, they cannot give a fair star rate or comment as they could not predict the feeling of buying goods at a higher price.

And for the second part, there are some seller who found that comment and star rating have a strong correlation with future sells, so they may write some comments by themselves to improve the reputation. Meanwhile, some competitor may also use the extreme stratagem to reduce others orders. For instance, they would give a low star rate to mislead consumers avoid those goods. Therefore, these comments and star rating should not be considered.

2) The rate of verified purchased
   We simply analysis the percentage of verified purchase.

   r = number of verified purchase /number of total purchase

   This ratio shows the how much misleading information would have, and all ratio are shown below.

 | | Hair dryer | Microwave oven | Baby pacifier |
   |-----|------------|----------------|---------------|
   | r | 85.536% | 67.802% | 81.965% |

   This table clearly illustrate that comments of hair dryer and baby pacifier are mostly reflect the true information, that is to say, when consumer search this type of goods, they are able to get the real thought from previous buyer. In this case, the factors which affect consumer to make the decision is the quality of goods and comments of using this product.

   However, microwave oven selling only have 67.802% efficient comment. This would cause a huge misleading to buyer. For example, if someone wants to buy a microwave oven online, and there are 1/3 comments show that this oven cannot work properly, this person is likely to give up this purchase. What if all these 1/3 negative comments are not efficient? This would bring a big trouble for sellers.

4.2 Star rating distribution
   Here we use the whole dataset to analysis the star rating. (i.e. all the information that any buyer can get before purchase)
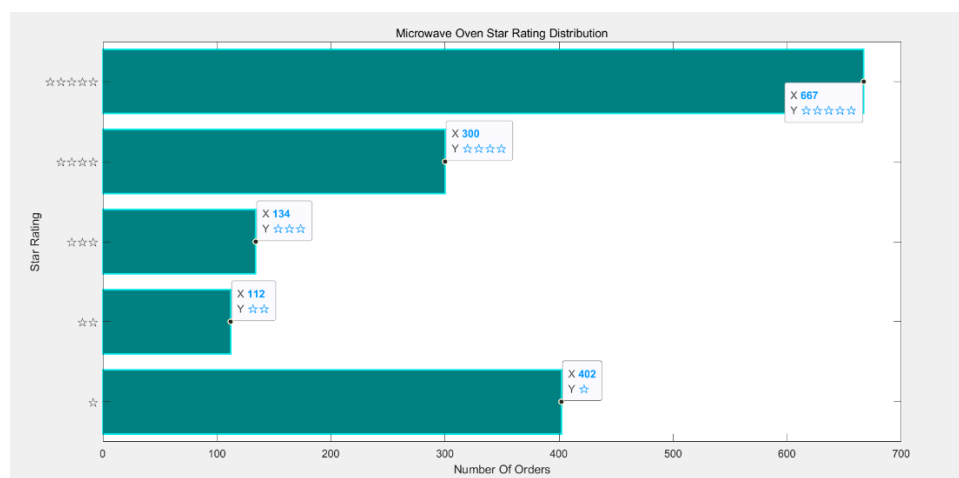   1) Star rating distribution of Hair dryer
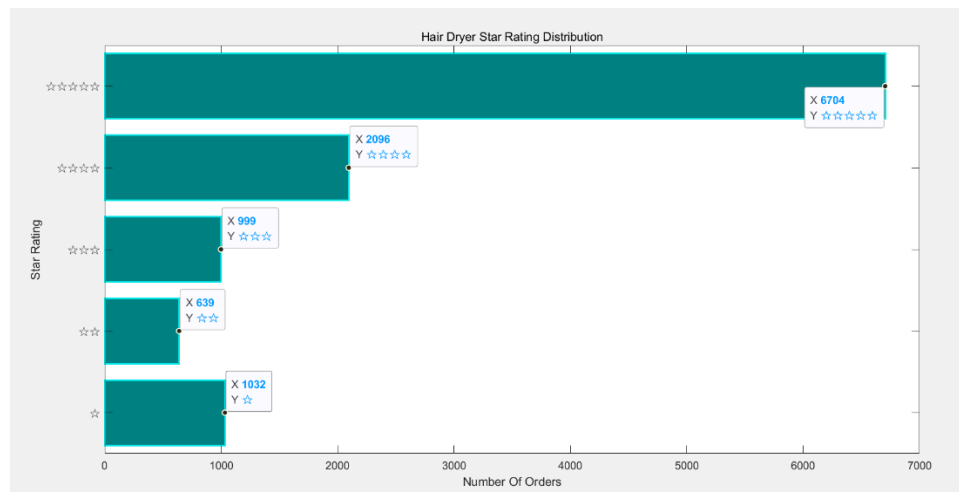
This bar chart shows how the star rate of hair dryer sold on Amazon from year 2002 to 2015 distributed.

The star rate lain at 5 stars took the largest proportion which have nearly 7000 votes. It seems most consumers were quite satisfied with the hair dryer they bought. Meanwhile, the 4-star option took the second highest proportion of total purchase. These two facts tell us that most orders for hair dryers sold on Amazon reached the expectation.

For the rest three classes, 3-star and 1-star orders have almost same amount, with 999 and 1032 respectively. That is to say, the number of purchases which marked as not satisfied or disappointed with that order still not that much. In addition, the order with 2 stars had the smallest amount with only 639.

Generally speaking, the star rate of hair dryer sold on Amazon has average score 4.1 stars. It would likely to be satisfied with the product if someone decided to buy a hair dryer on Amazon.

2) Star rating distribution of Microwave oven

Microwave oven had a much less sales on Amazon comparing with hair dryer. This bar chart illustrates the distribution of star rating of microwave oven in that period.

Firstly, the number of sales which belongs to satisfied (4 stars and above) occupied majority of all sales. The 5 stars orders reached 667, while there were 300s orders marked 4 stars. if sellers could fix those small problem as soon as possible, the satisfaction will improve a lot.

As for the star rating at 3 or below, they take up less than a half of all comments. Besides the most one (1-star), 2-star and 3-star comments actually only occupied a small ratio. They are all around 100 votes. However, the number of 1-star votes is more than a half of 5-star comments. It reached 402 votes at 2015, which producer should concern about the quality of these microwave oven to reduce the number of 1-star rating. The total mean star rate is 3.4.

3) Star rating distribution of Baby pacifier
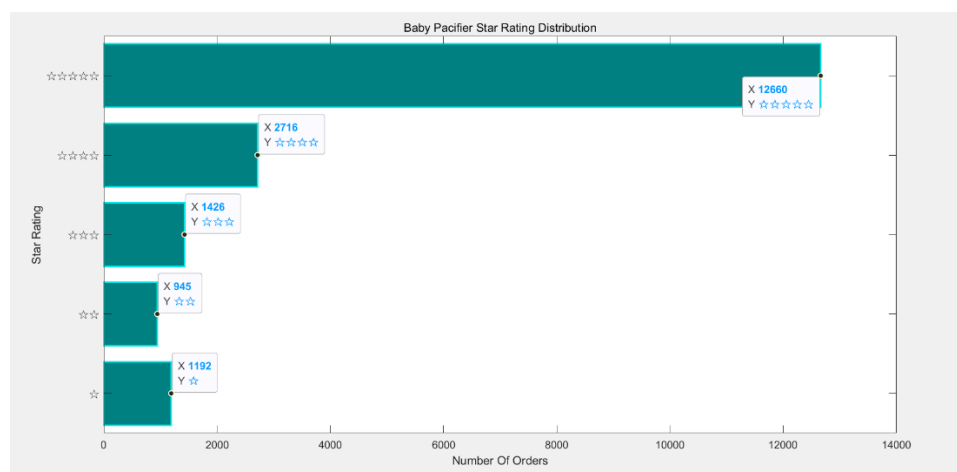
This graph shows the total baby pacifier star rating from the year 2003 to 2015, with 4.3 stars in average. We separated this data into 5 groups to analysis the amount of different rating comments difference.

The baby pacifier had the largest amount of sales among three products. And it also gained the largest number and proportion of 5 stars votes, which went over 12660 orders. The second most went to 4-star comments. It is clear that people always satisfied when they choose to buy baby pacifier.

As for the rest parts, they only took up a small amount of total orders. This gives an information that sellers do not need to concern too much about if the negative comments would cause they reputation become bad. Finding the common factors in those negative comments and having some small improvement would bring the average star rating higher easily.

4.3 The number of sales changing with respect to different star rating

1) The star rating changes from 2013 to 2015 for Hair dryer:

☆： The trade is almost stable which was around 25 sales until December

2014, which soared rapidly and got a peak in February 2015 that was 45 sales. Although the next turning point appeared after two months ago which has dropped back to 20 sales, after that it climbed to around 30 sales.

☆☆: During these three years, it has a significant oscillation which is stable,

at the same time, even though it was always reached the peak, the most was 25 sales which was not too much during this period.

☆☆☆： This bar chart trade is similar to the one star, both of the most sales

were 45.

☆☆☆☆： Similar to the front charts its tendency growth and down were

steadily, and had same turning point. The only one difference was the highest number which was about 90 sales.

☆☆☆☆☆： This line chart illustrates the tendency was slightly different

from other ones. Firstly, the buyers wrote five stars rating was the most which was around 350 sales. After the first season of 2015 the trend was smooth as well.
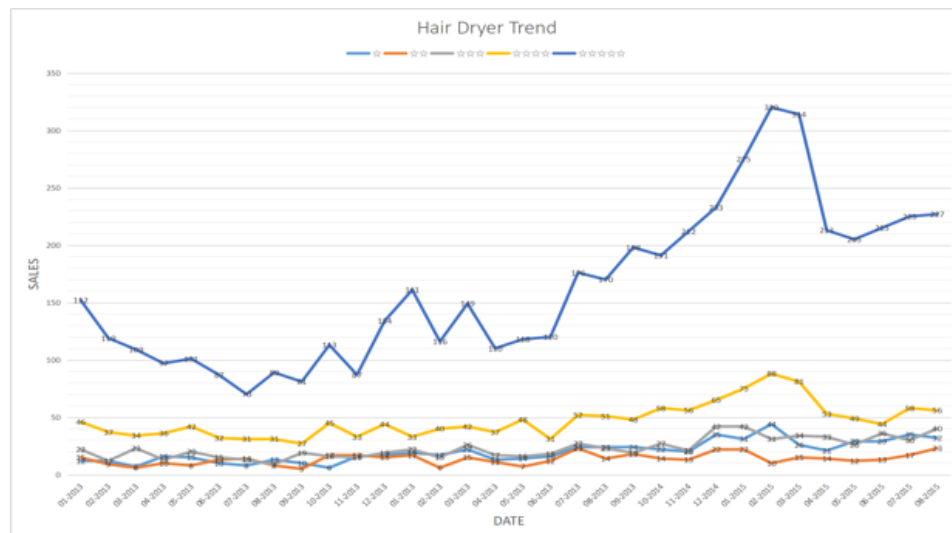
Summary:

According to the line chart, it shows all the stars tendency are increasing and reaching a peak in the end of 2014 and the beginning of 2015. This growth due to the Amazon overseas policy, the number of consumers increased rapidly. However, after the start of policy, it back to a normal number of sales but it still more than before. Meanwhile, those three goods stars rating got the minimum value in the third season of 2013. Also, most of buyers were satisfied by the hair dryer. Therefore, the star rating is no significant association with sales after 2015

2) The star rating changes from 2013 to 2015 for Microwave Oven:
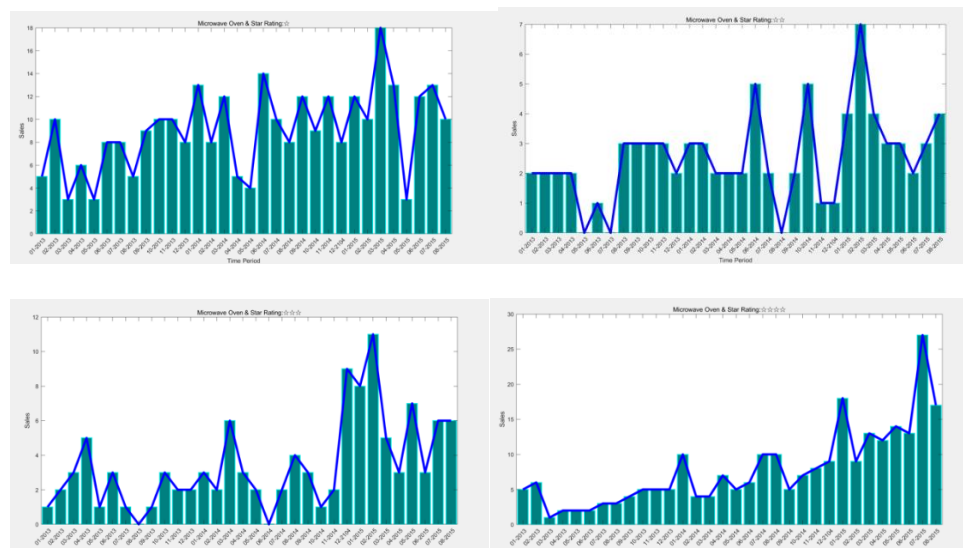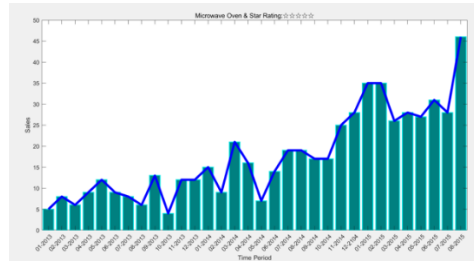
☆：The oscillation is very obvious，and the most was 18 sales. Actually, it oscillated in two months cycle. The tendency shows the sales were affected by the star rating, and through the whole trend line chart, the influence power just followed by the five stars rating.

☆☆：Before the June 2014, it owned a steady number of sales except the mid-season of 2013 relatively. After that, the number of sales has a soared increase and rapid decrease in alternative month. Even though, given the microwave oven with one star has still occupied a small part with the highest number was 7.

☆☆☆：Both tendency and number of sales of this chart were similar to the one-star rating chart, which peak value was 11 sales in February 2015.

☆☆☆☆：The number of sales with four stars rating expressed a climbing trend, it shows a stable growing in the three years.

☆☆☆☆☆：This tendency resembles the four stars one, both of the high star rating will have a positive effect significantly.
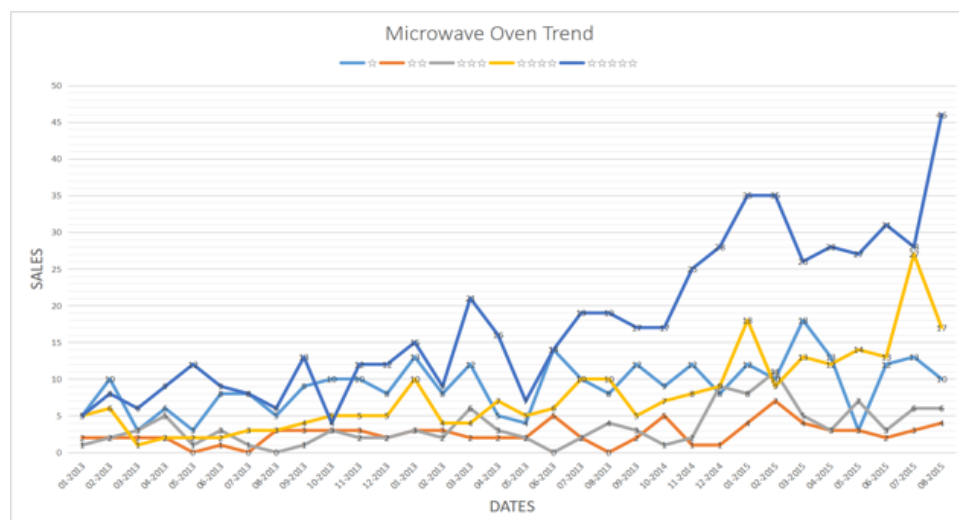
Summary:
According to the line chart, both five stars rating and 1-star rating will have an obvious influence on the sales followed by the four stars rating. Contrastively, the three stars rating has the least influence on the sales. At

the same time, all of the star rating for the microwave oven sale be the helpful suggestions for the consumers.

3) The star rating changes from 2013 to 2015 for Baby Pacifier:

☆：The highest number of sales was 70 in December 2014, because of that the influence of sales almost can be ignored.

☆☆：Although the 2-star rating is the low rating like one star, the sales of the highest number was ten less than the one star with about 60 sales. Meanwhile, the trend is similar to the one star.

☆☆☆：The whole tendency is similar to the front two charts, the only difference was the number of sales as well.

☆☆☆☆：The amount of sales doubled compared with the three stars rating, resemble trend as well.

☆☆☆☆☆：Most of the consumers with the positive attitude to the baby pacifier to give the five stars rating. The peak value was 800 sales in December 2014. Most of consumers were satisfied by this pacifier.

Summary:

The sales had a significant bottom area in the second season of 2014 according to front five charts. However, it reached a peak point at the end of 2014. At the same time, the five stars rating in the baby pacifier sale was the main effect for the sales, and the rest stars rating were with a similar influence on the sales, which only the four stars rating was slightly higher than the other three. All the tendency was similar, the only difference factor among that was the amount, they oscillated on their own amount. Also, it shows most people prefer to give a high star rating for the goods. Therefore, the star rating will not have a significant influence on the sale, because the whole tendency is stable for the recent three years, the wave is due to the number of total sales.

4.4 The helpfulness of star rating

The helpful votes would demonstrate how deep would that specific comment would affect other consumers thoughts. And the percentage of helpfulness of that specific comment can be calculated with the formula below:

P=number of helpful votes / number of total votes

Since some comments did not have any votes, the denominator would be 0, this part of data should be removed. From real world perspective, the number of total votes is 0 means nobody is affected by that specific comment, therefore, it is unnecessary to take those into consideration.

1) Hair dryer
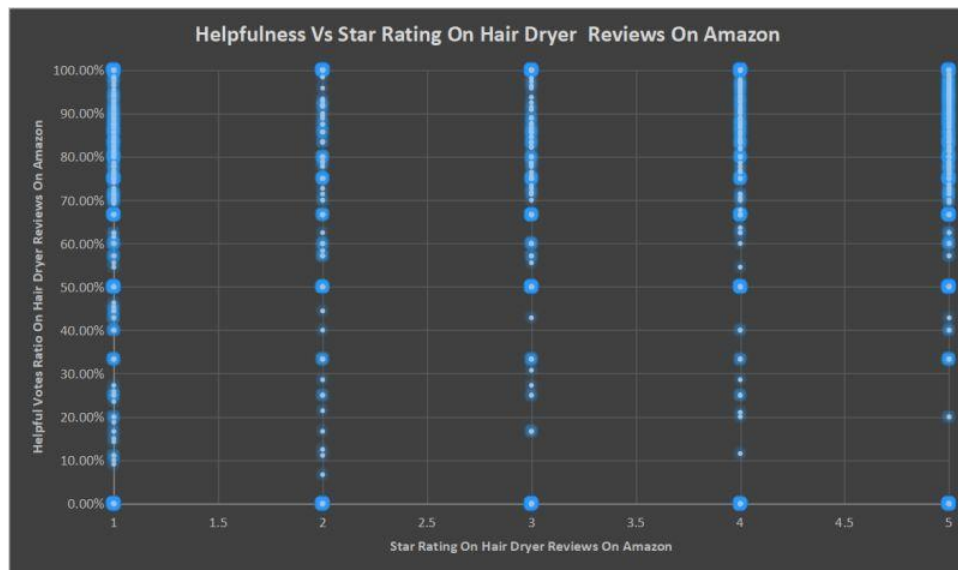
The helpfulness of hair dryer reviews at different star rate are shown on this scatter chart.

By comparing different star rates, it demonstrated that the comments with 1-star products and 5-star products had the great influence on future consumers, meanwhile the comments which the good have been marked as 4 stars is likely to have a bit less than those two stars rates. The density of data of helpfulness are more concentrated on 70% to 100%, that is to say, no matter at which star rate, most comments which have been voted for testing helpfulness have a strong influence on purchase. In addition, there are still existing some comments which were less helpful (under 50% helpfulness) than others, especially for comments of products have been marked as 1-star. Meanwhile, the 5-star and 3-star good's answer of the shopping experience is likely to offer more efficient information than others comparatively.

2) Microwave oven
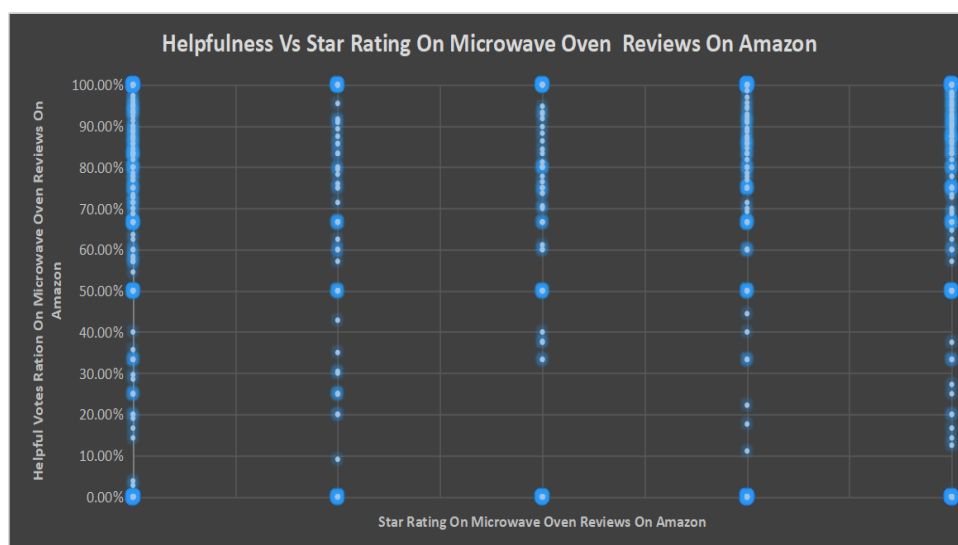
The reviews of microwave oven have almost same spread out with hair dryer's based on helpfulness of comments on Amazon.

Firstly, the reviews of 1 star, 4 star and 5 stars microwave oven had a large density at 50% and above, especially at 100%. That is to say, people can always get the precise answer about the shopping experience and product information which exactly what they concern about. Besides that, the chart also illustrated that the comments with less helpfulness is less likely to appear in any comment. This fact tells that most comments on Amazon which describes the usefulness of microwave oven would always help future consumer to make the decision.

However, there is a set of data here shows something rare with two other charts. Reviews with 2 stars have no specific information, as the data is evenly spread along the different percentage level.

3) Baby pacifier
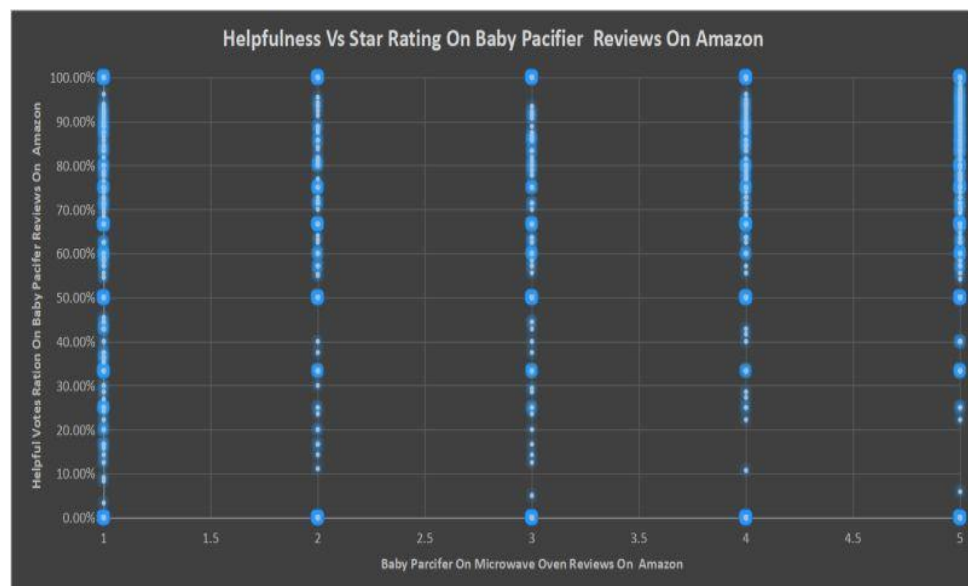
This scatter chart demonstrated the relationship of helpfulness and star rate of baby pacifier, which is slightly different pattern with hair dryer.

The reviews of goods which have been marked as 5-star would bring the most useful information as usual. All the data points are concentrated on 60% to 100% region. The 4-star goods' reviews also have the same distribution, and comments with 0% to 50% helpfulness are a little bit more than 5-star good' s. In addition, the comments of 2-star baby pacifier are more likely to influence others.

Meanwhile, the comments for 1-star and 3-star products have resemble spread with the helpfulness of reviews, which have the almost same amount of reviews with all different percentage of helpfulness. The only difference is that 1-star goods have more voted reviews than 3-star baby pacifier.

4) Conclusion

Generally speaking, for any products, the future consumer is more likely to read the extreme positive or negative comments and vote those comments as helpful information. At the same time, the polarity of reviews on goods may give a better description on shopping experience and quality of goods. Therefore, more people will be willing to believe these comments.

4.5 The quality of products
1) The characteristics of Hair dryer:
   Y:9811;85.536% N:1659;14,464%

Star rating	☆	☆☆	☆☆☆	☆☆☆☆	☆☆☆☆
Percentage	7.532% (739)	5.188% (509)	8.755% (859)	18.388% (1804)	60.147% (5901)

The relationship between star rating and reviews:
I.   For all consumers:
     ☆:
     a) Quality problems:
     i. components problem
     ii. the air is too hot
     iii. the sound is annoying
     iv. the machine does not work well which only last few weeks or months
     b) region using problem:

     i. most of in Europe area does not work，especially in Italy and UK.

     c) the diffuser mode is unsatisfied.
     ☆☆:
     a) get quality problem (same problems as ☆) after using a while
     ☆☆☆:
     a) the attachments and plug problems appeared
     ☆☆☆☆:
     a) good for travel
     b) easy to use
     ☆☆☆☆☆:
     a) similar reviews as four stars

II. For vine consumers:
     (i.e. vine: Customers are invited to become Amazon Vine Voices based on the trust that they have earned in the Amazon community for writing accurate and insightful reviews. Amazon provides Amazon Vine members with free copies of products that have been submitted to the program by vendors. Amazon doesn't influence the

opinions of Amazon Vine members, nor do they modify or edit reviews.)

☆:0

☆☆:0

☆☆☆:

a) right weight

b) not the best one that the buyers have ever used

☆☆☆☆:

a) powerful

b) good for thick hair

☆☆☆☆☆:

a) fast to dry hair

b) easy to pick

III.   About the association between the specific description and star rating:

a) for one and two stars with positive reviews
   6.344% (106/1671)

b) for four and five stars with negative reviews
   1.284% (113/8801)

2) The characteristics of microwave oven:
   Y:1095;67.802% N:520;32.198%

Star rating	☆	☆☆	☆☆☆	☆☆☆☆	☆☆☆☆☆
Percentage	10.594% (117)	5.205% (57)	8.493% (93)	22.831% (250)	53.324% (583)

The relationship between star rating and reviews：

I.   For all consumers:

☆:

a) quality problems
   i)   components problem
   ii)   good work only last few weeks or months
   iii)   the heat part does not work well and has bad smell

☆☆:

most of buyers recognize it does not work very well but it is a nice one

☆☆☆:

this is a high cost-performance ratio

☆☆☆☆:

a) it has a good price

b) great size

c) easy to use

☆☆☆☆☆:
similar as four stars reviews

II.   For vine consumers:
None.

III.   About the association between the specific description and star rating:
   a) for one and two stars with positive reviews
      4.864% (25/514)
   b) for four and five stars with negative reviews
      1.034% (10/967)

3) The characteristics of baby pacifier:
   Y:16261;81.965% N:3578;18.035%

Star rating	☆	☆☆	☆☆☆	☆☆☆☆	☆☆☆☆☆
Percentage	5.498% (894)	4.838% (785)	7.589% (1234)	14.396% (2341)	67.714% (11011)

The relationship between star rating and reviews:
I.   For all consumers:
☆:
   a) the seal not good
   b) for kids cannot breathe well
   c) the size does not fit for every baby
   d) hard to clean
☆☆:
   a) no size choice
   b) the color is limit
   c) easy to damage
☆☆☆:
   a) the weight is not heavy enough
   b) inconvenient to take
☆☆☆☆:
   a) good surface design
   b) good seal
   c) which is not easy to be damaged when dropping on the floor
☆☆☆☆☆:
   a) various size choice
   b) kids can breathe well
   c) no smell
   d) easy to clean and take
   e) looks cute

II.   For vine consumers
☆:0
☆☆:
a) to be broken few months later
b) looks cute
c) the size is too big
☆☆☆:
a) the size is not fit
b) the quality problem
c) not worth the price
☆☆☆☆:
a) cute
☆☆☆☆☆:
a) baby can easy to hold it
b) easy to clean and safe

III.   About the association between the specific description and star rating:
a) for one and two stars with positive reviews
   8.109% (136/1677)
b) for four and five stars with negative reviews
   0.891% (119/13351)

4) Summary:
a) The reviews direction for one and two stars are negative, three stars are sitting on the fence and most of four and five stars are positive. Therefore, most of buyers will have a coincident star rating and reviews.
b) According to the tables, the five stars rating of each good are all over fifty percent of all sales. Therefore, those three kinds of goods in each field are very popular among the consumers.

5. Model weaknesses and limiting Assumptions
- Since the model is based on only 3 products on Amazon, there still are some other products or other platform will give more accurate results
- The number of views of each product is incomplete. The data sets need more detailed information of the products' brand, so that the more completed analysis can be made.
- The databases can be updated to give a more reasonable conclusion.
- We focus on the market analysis and data trend summary instead of using mathematical formulae or equations.

6. Letter for Sunshine Company

Dear Sunshine Company:

Thank you for inviting our team to predict your company sales, we are pleasure of analyzing the data to give you some advice about the sale strategy.

According to the analysis of the competitive company data, some conclusions and suggestions are derived.

Here are some suggestions about the goods and marketing.

I. Goods suggestions:
a) Hair dryer:
- Quality problems:
according to the analysis of data, there main problem about the hair dryer was the quality problem. Therefore, before releasing the products, please check the quality of components especially for the buttons and the whole electric machine. At the same time, the using-life time should be increased.
-Technical problem:
the area restriction problem appeared when using the hair dryer, so the advice for Sunshine Company is solving the region limit using problem. (e.g. can alter voltage or provide plug adaptor). Meanwhile, if you cannot do that above, the other suggestion is label the area that can be operated on the goods sale's homepage.

b) Microwave Oven:
-Quality problems:
according to the feedback from consumers, they prefer the product is easy to use, therefore, your company should design a simple functional microwave oven. Also, the component part is also the main problem need to be resolved. Lots of buyers comment that the buttons and working time is disappointed. So, your company should focus on these quality problems.

c)Baby pacifier:
-Design and manufacture:
firstly, due to the market data, a cute outlook design will more abstract the baby than the regular one, therefore, your company can choose some adorable pattern for sale. Secondly, produce more alternative size and various color, because most buyers think the size did not fit their babies' month and limited color choice. Thirdly, because of the use of babies, the material of pacifier should be healthy which is non-toxic, non-toxic smell and easy to clean. Finally, the design style can be high quantity, most people will put eyes on the quantity.
-Distinguish:
the baby pacifier may have different using for different age-period babies. Therefore, the function they use is different. Because of that your company should label the

corresponding age's baby.

II. Marketing suggestions:

a) Price：

the price made should be reasonable due to the product quality, the data shows the consumers prefer to buy the high cost-performance ration product rather than the cheaper or more expensive one.

b) Make sure the consumption groups:

one of the common problems appeared in these three products is this problem. The unclear mark has misunderstood the consumers which gave rise to negative comments for the product. In case, the suggestion for your company is label all functions and recipients.

-Delivery company choice:

most consumers will have a not bad emotion to the product due to the fast and kind attitude delivery, therefore, a good delivery company choice will leave a good memory to the buyers.

-Details:

before dispatching, your company should double check the details and products of consumers, because there were many cases without a careful check which leads to the wrong way to the consumers or with wrong products, these details your company should be noticed.

-About verified purchaser:

according to the data, there were lots of non-verified consumers to write some malice reviews under the comments, which influence buyers to distinguish the quality of the goods. Because of that your company should take more care about those significant negative comments after publishing the products to protect your reputation. (some suggested actions: 1.report this behavior to the platform in order to let the manager to delete the malice comments. 2. Develop some technical skills to identify those comments.)

We hope our suggestions and analysis report can be helpful to your company.
And wish your success in your business!

Sincerely,
Team #2022666.