Problem Chosen	2020 MCM/ICM Summary Sheet	Team Control Number
C		2013548

Build a Successful Product on the Online Shopping Platform

Information is one of the most valuable wealth now. In recent years, applied statistics has become more popular. Data science and business analysis have been more attractive. The key to turning information into wealth is not only to collect a large number of data, but also to extract useful information from these data.

First, this paper analyzes the basic numerical characteristics of the three data sets in the attachment. In this paper, NLP (natural language processing) is used to convert text-based reviews into numerical variables, to show the emotional characteristics of customers in reviews. The numerical score is used to indicate whether the emotional types carried by reviews are positive or negative. In this paper, NLP, PCA (Principal Component Analysis), regression analysis, spectrum analysis and other tools are used to establish several evaluation indicators based on star ratings and reviews to complete the measurement of product competitiveness, sales status, reputation and success potential.

Secondly, this paper considers the influence of each variable on the reliability of star ratings and reviews given by customers based on the weight function. NLP and PCA are used to reduce dimension and simplify data, and the most informative data measures are constructed. Through the analysis, the paper find it is important for the score of the product whether the customer who gives the evaluation is the Amazon verified vine user, and whether the customer is confirmed to have purchased the product on the Amazon platform when giving the evaluation.

Moreover, this paper establishes a time-based product reputation score via PCA. We can judge whether the reputation of a product is increasing or decreasing over a period. In this paper, the rate of change of sales volume is introduced, and a Boolean prediction function δ_{rate} is established to measure the success or failure potential of a product by regression fitting. When the value of δ_{rate} rate is 1, it can be judged that the product has the potential of success.

Finally, the correlation between star ratings and reviews is discussed by using the spectral analysis model. The analysis shows that specific star ratings will cause more comments. This paper believes that this is a kind of herding effect. After seeing some specific star ratings, people will express and vent their emotions through reviews. In addition, we use Fourier transform to transform star ratings and reviews emotion scores into time series. The paper finds that there is a positive correlation between high stars and positive reviews, low stars and negative reviews. Specifically, there will be "love", "happy" and other words containing satisfaction and pleasure in the reviews of high star, while "not", "dissatisfaction" and other words expressing disappointment in the reviews of low star.

Key words: NLP, PCA, Reviews emotion scores model, Spectral analysis, Regression model

Contents

1. Introduction

1.1 Restatement of the Problem

In the information age, data is one of the most valuable wealth. People can sum up successful experience by analyzing past data, compare differences with competitors, find development advantages, and formulate future strategies. Huge data is the most valuable wealth, but there are many kinds and forms of data, so it is not easy to get useful information from them. Amazon online shopping platform provides Sunshine Company with three kinds of product record data based on star ratings and reviews information. These data sets are relatively large and complex. We need to help Sunshine Complete the following tasks:

- Filter data set and reduce dimension and simplify data.
- Establish data measures and models by using data sets to gather information value.
- Use models and measures to solve the requirements put forward by Sunshine.
- Provide Sunshine with an attractive online sales strategy based on the results of the model and the reality.

1.2 Our Work

We analyze the data set provided in the attachment and get some numerical characteristics. We use NLP, PCA, regression fitting and other tools to build several basic models, and based on the models, we build a series of evaluation indexes about product competitiveness, product reputation and product success potential. We solve the requirements and problems put forward by the Marketing Director by improving the models. We pay more attention to the visualization of data and get the corresponding conclusion by analyzing the trend of data characteristics in each graph. We study the correlation between star ratings and reviews via spectrum analysis. Moreover, we analyze and discuss the experimental results based on the actual situation and data in order to provide a practical explanation for numerical indicators. Finally, we combine the characteristics of the Internet market, people's behavior psychology and relevant information in real life, and use our analysis results to provide some reasonable sales strategies for the Marketing Director of Sunshine Company.

2. Model Assumptions

1. Assume that customers' star ratings and review emotions are random variables.
2. Assume that the emotional characteristics of reviews can be converted into numerical measures. It is easy to judge the positivity and negativity of a review by numerical value.
3. Assume that the customers' label will have an impact on the credibility of star ratings.
4. Assume that the growth of sales of means that the products have the potential for success
5. Assume that the story of long reviews will attract more customers' attention.
6. Assume that the star ratings and reviews given by customers with multiple reviews are more guiding and valuable for other customers.

3. Notations

Symbol	Description
g_k	Article k comprehensive score of star ratings recorded
S_k	Article k star ratings recorded
δ_{vote}	Weight function determined by the number of votes
$\delta_{verified}$	Weight function determined by verified label
δ_{vine}	Weight function determined by vine label
$vote_k$	Article k the total number of votes received record
$help_k$	Article k the proportion of helpful votes received recorded
M_t	Comprehensive star ratings of a product
E_t	Comprehensive emotion score of a product
e_k	Article k reviews emotion score recorded
C_t	Comprehensive evaluation index of commodity competitiveness
P_k	Article k commodity reputation score recorded
X_m	Comprehensive reputation score of a product in year m
$Q(t)$	Time series of star ratings of a product
$R(t)$	Time series of reviews emotion scores of a product
$FQ(t)$	Wave obtained by Fourier transform of sequence $Q(t)$
$FR(t)$	Wave obtained by Fourier transform of sequence $R(t)$
δ_{var}	Weight function determined by star ratings variance
δ_l	Weight function determined by length of reviews
δ_c	Weight function determined by customers with multiple reviews
V	Comprehensive index of star ratings information
W	Information measurement vector based on star ratings and reviews
δ_{rate}	Measure function of whether a product has the potential for success
$rate$	Annual sales growth rate of products
R_1	Correlation coefficient between star rating and reviews
R_2	Correlation coefficients of spectrum $FQ(t)$ and $FQ(t)$

4. Data Analysis and Basic Model

4.1 Distribution Characteristics of Data Sets

The basic analysis of distribution characteristics is essential for a large data set. The distribution of star ratings, average and mode of product star ratings can directly reflect the performance of the products in the entire Amazon online market, such as overall satisfaction and competitiveness among different brands.

The distribution characteristics of the number of votes in statistical reviews can show the degree of attention that commodities get and the degree of enthusiasm that they participate in discussions. Comparing the distribution characteristics of different brands of the same product can find out which brand has more competitiveness and success potential.

The frequency histogram of star ratings of three commodities is as follows:

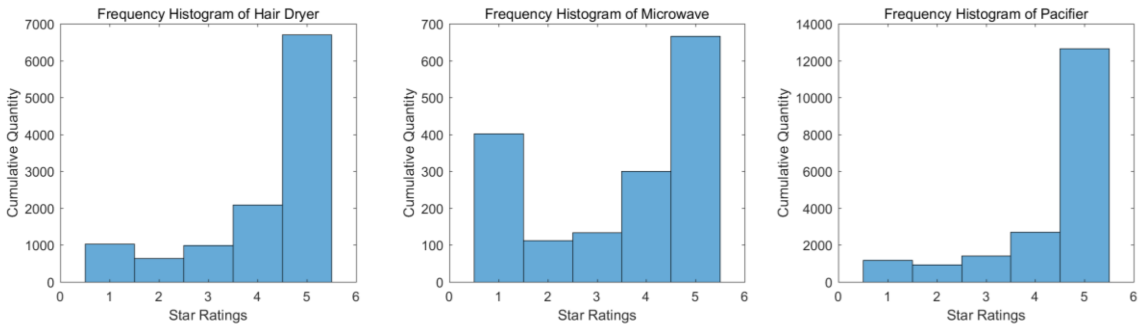

Figure 1：Frequency histogram of star ratings of three commodities

It can be seen from Figure 1 that the star ratings distribution characteristics of hair dryer and pacifier are roughly the same, and most of the customers give five star to hair dryer and pacifier. The star ratings of microwave oven show obvious polarization. While many people give five star of microwave oven, many people only give one star.

Next, we will specifically count the numerical characteristics of each product star ratings. The box line diagram of three products is as follows:

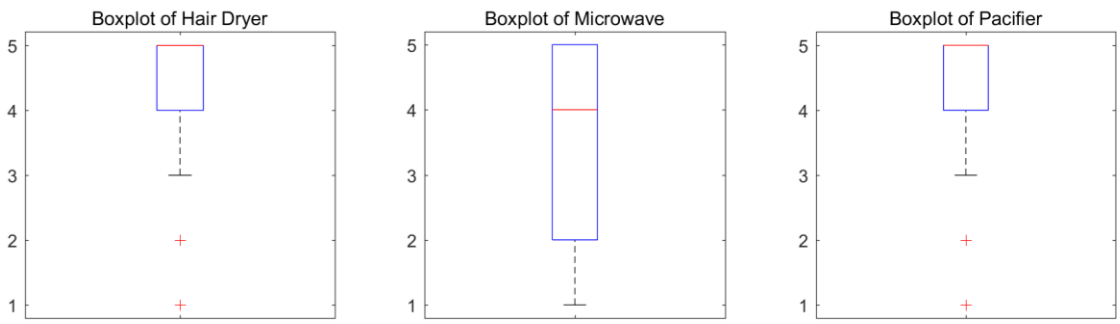

Figure 2：Star ratings characteristic boxplot of three commodities

The specific numerical characteristics of star ratings distribution are shown in the table:

Table 1：Numerical characteristics of star ratings of three commodities

Commodity	Average	Median	Mode Number	Variance
Hair Dryer	4.116	5	5	1.6909
Microwave	3.445	4	5	2.7068
Pacifier	4.305	5	5	1.4171

In addition to the comparison of star ratings distribution characteristics among the three commodities, the comparison of distribution characteristics among different brands of the same commodity is also quite important, which can help us preliminarily judge the competitiveness of different brands and find out the potential successful or failed products.

In this part, three different brands with product_parent ID of 983445543, 963066492 and 919751065 are selected as examples to compare their numerical distribution characteristics, to illustrate the basis and importance of this statistical analysis. The star ratings frequency histogram of three different brands of hair dryer is as follows：

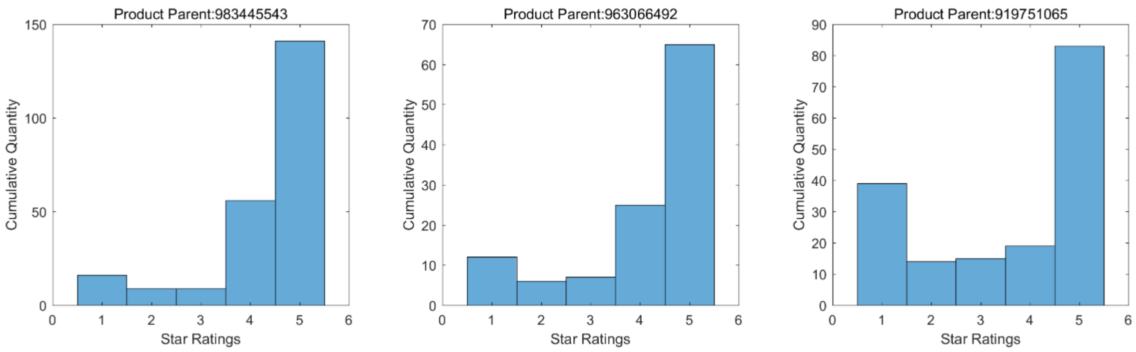

Figure 3：Frequency histogram of star ratings of three different brands

Draw the annual sales volume of three different brands of hair dryers over time as shown in the figure below：

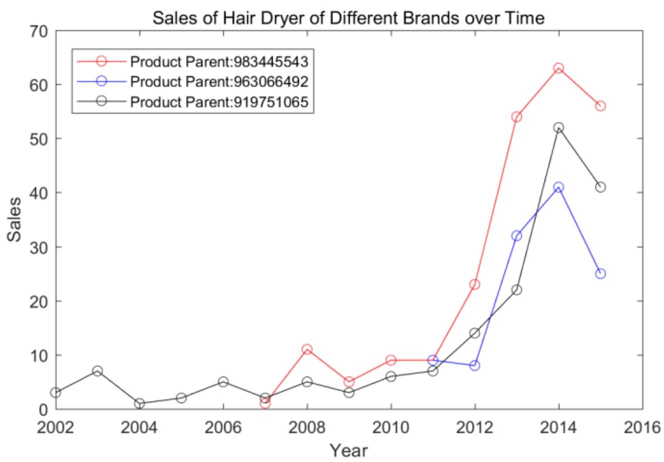

Figure 4：Changes of sales of hair dryer of three different brands over time

More numerical characteristics of the three brands are shown in the table below：

Table 2: Numerical characteristics of three hair dryer brands

Product Parent	983445543	963066492	919751065
Average Star Ratings	4.286	4.087	3.547
Annual Average of Comment Votes	28.2222	58.8000	29.9286
Annual Average Sales	25.6667	23.0000	12.1429

By analyzing the above statistical results, we can easily conclude that the product whose product_parent ID is 983445543 has the most potential for success. It has obvious advantages in annual sales volume and average star ratings. Nearly 90% of customers give four to five star evaluation, with the same reputation.

4.2 NLP Classifies the Emotions of Reviews

4.2.1 NLP Recognition Emotion

Star ratings and reviews are the two most important variables in each record. Star ratings is a numerical variable, which can show the customer's preference for the product intuitively (the closer the star is to 5, the more satisfied the customer is). And reviews is a string variable, which needs to be interpreted to understand whether the user experience of a product is pleasant or bad. Usually, most of the components of a sentence are neutral structured words. However, some special words, such as "love", "like", "happy", "good", "great", "satisfied", "generally", "not happy", "dissatisfied", "disappointed", etc., and some special symbols, such as "question mark" and "exclamation mark", these special words can reflect whether customers' attitudes towards products are positive, negative or neutral.

Using NLP (Natural Language Processing), we can divide a review into several emotional gradients: excellent, satisfied, general, dissatisfied and disappointed. The following three examples illustrate NLP's calculation results when recognizing reviews emotions：

Table 3: Examples of NLP recognition

<table>
<tr><td rowspan="2">Example1</td><td colspan="6">Great hair dryer! Love having the hair dryer ready to use but out of the way mounted to the wall. We have one in each bathroom!</td></tr>
<tr><td>Negative:</td><td>0.000</td><td>Neutral:</td><td>0.754</td><td>Positive:</td><td>0.246</td></tr>
<tr><td rowspan="2">Example2</td><td colspan="6">So handy, works great, frees up counter top space in my small bathroom.</td></tr>
<tr><td>Negative:</td><td>0.000</td><td>Neutral:</td><td>0.545</td><td>Positive:</td><td>0.455</td></tr>
<tr><td rowspan="2">Example3</td><td colspan="6">I suspect build up of lint on the blades is what eventually caused the other dryers to burn out, and with the other dryers it is not possible to access the blades (at least not easily). Thus, I expect this dryer to last longer.</td></tr>
<tr><td>Negative:</td><td>0.031</td><td>Neutral:</td><td>0.878</td><td>Positive:</td><td>0.090</td></tr>
</table>

We can roughly analyze the semantics of reviews in the three examples in the above table. Some special words and statements make them obtain different scores in negative, neutral and positive. We can see that in example 1, review scores higher in neutral, which can be classified as satisfactory emotion. In example 2, neutral and positive scores are relatively balanced, which can be classified as excellent emotion or satisfactory emotion.

4.2.2 Establishing Emotion Score Based on PCA

In the previous part, we proposed using NLP to convert the string variable reviews to three scores of negative, neutral and positive. Based on these scores, we can analyze and get some qualitative results about reviews emotion, but it is still lack of convenience and explanation to use them as numerical calculation. For each evaluation record, we can get its three scores through NLP. Now we use PCA to reduce the dimension of data, select principal components to build a comprehensive evaluation index, and establish the total score of each review emotion. The level of the score can explain the positivity or negativity of the review emotion

We take the product_parent ID 983445543 in the hair dryer as an example, PCA is used to establish a comprehensive evaluation index, and the emotional score of reviews is obtained. To a certain extent, star rating can reflect customer satisfaction. We take star ratings as a check and draw the change of star ratings and emotion scores with time as shown in the figure below:

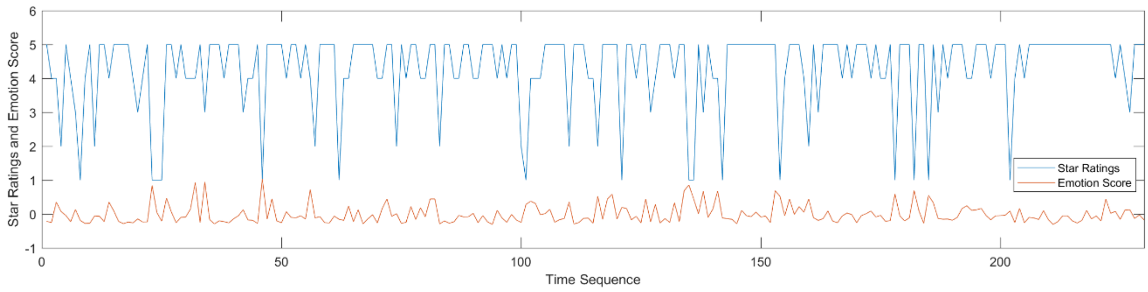

Figure 5: Time based changes in star ratings and emotional scores (1)

The comprehensive index of emotional score was obtained, but the index was lack of interpretation. Observe the change characteristics of stars and scores, we find that the emotion scores of low stars are high, and the emotion scores of high stars are low. The growth of the two curves has certain symmetry。In order to confirm the above conjecture, similar to star ratings, the emotional score is normalized to $1-5$ points through linear transformation, and the change of star ratings and emotion scores with time is plotted as shown in the figure below:

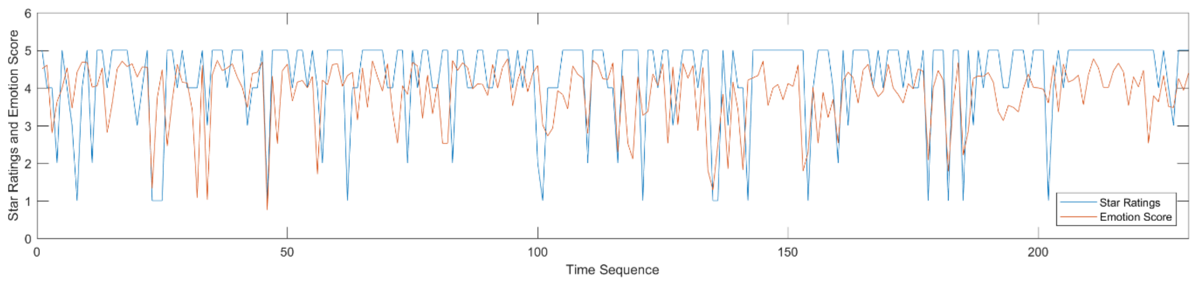

Figure 6: Time based changes in star ratings and emotional scores (2)

From Figure 6, we can see that the change characteristics of star ratings and emotion scores are similar, and the above conjecture is reasonable. Therefore, we can get the actual explanation of the normalized comprehensive emotion score of reviews after PCA:

The higher the score is, the more positive the emotions of reviews are. The lower the score is, the more negative the emotions of reviews are.

4.3 Comprehensive Evaluation Index of Each Commodity

Each star ratings and view has a strong subjectivity, and whether it helps to evaluate, whether the identity of the customer who gives the review is vine, and whether the customer who gives the review buys this product on Amazon platform can objectively reflect the credibility of the comment.

Therefore, the comprehensive evaluation index is constructed based on star ratings using that objective information, which can more truly show the competitiveness and performance of commodities on Amazon platform. At the same time, the dimension of data is reduced to make the comprehensive evaluation index contain more information. The structure of comprehensive evaluation index is as follows.

Suppose that there are n records for the product whose product_parent ID is t. We use the following formula to define the score of record k:

$$g_k = (S_k - 3) \times \delta_{vine}^k \times \delta_{verified}^k \times \delta_{vote}^k, \quad k = 1, 2, \cdots, n \tag{4.3.1}$$

In which, S_k represents the star ratings recorded in article k, and $S_k - 3$ classifies the star ratings from $[1\sim5]$ to $[-2\sim2]$. The positive and negative can more clearly indicate the nature of star level evaluation.

The definitions of $\delta_{vine}^k, \delta_{veri}^k, \delta_{vote}^k$ are as follows。

$$\delta_{vine}^k = \begin{cases} 1 & \text{Vine is Y} \\ \alpha & \text{Vine is N} \end{cases}, \quad 0 < \alpha < 1 \tag{4.3.2}$$

δ_{vine}^k indicates the support degree of customer identity to star ratings. If the customer identity is vine, it indicates that the evaluation is objective and fair, otherwise, the reliability of evaluation should be reduced appropriately. Parameter α is the regulatory parameter of reliability.

$$\delta_{veri}^k = \begin{cases} 1 & \text{Verified is Y} \\ \beta & \text{Verified is N} \end{cases}, \quad 0 < \beta < 1 \tag{4.3.3}$$

Same as the definition of δ_{vine}^k, δ_{veri}^k indicates whether the customer has purchased the product to support the star rating. If the customer does not purchase the product on Amazon platform for rating, the credibility of the rating should be appropriately reduced.

δ_{vote}^k is a segmented function related to the number of votes. Its definition is based on the numerical characteristics of the number of votes. The definition of δ_{vote}^k is as follows:

$$\delta_{vote}^{k}=\begin{cases}1.00 & (vote_k \geq 5)\wedge(help_k \geq 85\%)\\ 0.95 & (vote_k \geq 5)\wedge(55\% \leq help_k < 85\%)\\ 0.90 & (vote_k \geq 5)\wedge(45\% \leq help_k < 55\%)\\ 0.80 & (vote_k \geq 5)\wedge(30\% \leq help_k < 45\%)\\ 0.70 & (vote_k \geq 5)\wedge(help_k < 30\%)\\ 0.90 & (vote_k \leq 5)\end{cases} \tag{4.3.4}$$

δ_{vote}^{k} represents the objective performance of the number of votes evaluated on the reliability of stars, $vote_k$ represents the total number of votes recorded in article k, and $help_k$ represents the proportion of helpful votes to the total number of votes. Combined with the actual situation and the average number of votes of commodities, the number of votes is less, the statistics of votes contain more subjectivity of customer evaluation, and the voting results cannot objectively and fairly reflect the real situation of rating. When the number of votes is large, its subjectivity is reduced, and the objectivity embodied in the number of votes is enhanced. And in the case of many votes, the larger the proportion of $help_k$ that is helpful to vote, it shows that this comment is recognized by more people, and its star ratings reliability is also high. On the contrary, if the proportion of helpful votes $help_k$ is very small, it means that this comment violates the feelings and opinions of most people. At this time, we should question the credibility of stars, to reduce its proportion in the scoring indicators. After getting the score g_k of a single evaluation record, the comprehensive star rating M_t of product t is defined by the following formula：

$$M_t=\frac{1}{n}\sum_{k=1}^{n}\left(0.90+\frac{10\times vote_k}{vote_t}\right)g_k \tag{4.3.5}$$

where $vote_t$ is the total number of comments received with number t：

$$vote_t=\sum_{k=1}^{n} vote_k \tag{4.3.6}$$

In formula (4.3.5), each evaluation record of the same commodity is weighted and summed, and the weight is based on the proportion of current votes. The larger the proportion is, the more attention and discussion the evaluation causes, and the more objective the score is. It can be seen from the data set that most of the votes recorded in the evaluation are 0, so a constant of 0.90 is added before the weight to avoid a large number of invalid summations with a weight of 0 due to the proportion of votes.

Based on the comprehensive evaluation index of stars, we can also use the emotion score of reviews to construct a more perfect evaluation model. In Section 4.2, combining NLP and PCA, we can get the review emotion score of each record. We use E_t to evaluate the emotion score of products with number t：

$$E_t=\frac{1}{n}\sum_{k=1}^{n} E_k \tag{4.3.7}$$

Next, we construct a comprehensive evaluation index C_t using the star ratings score M_t and emotion score E_t as follows:

$$C_t = \varepsilon \cdot M_t + (1-\varepsilon) \cdot E_t,\ \ 0 < \varepsilon < 1 \tag{4.3.8}$$

In the above formula, the comprehensive evaluation index C_t is the weighted sum of the star ratings score M_t and the emotion score E_t, and the weight ε is the adjustable parameter, which indicates that the decision-maker is more inclined to believe in the star evaluation of or more subjective reviews. The emotional score E_t comes from the string variable reviews, which contains more subjective information of customers. Usually these subjective text reviews can drive or inhibit more customers' consumption than simple stars. Therefore, we set the adjustable weight parameter ε to make the index C_t more explanatory.

4.4 PCA Analysis of Time-based Performance of Commodities

Besides star ratings and reviews, the market performance of commodities is also one of the important indicators of commodity sales. Based on the star score g_k, emotion score e_k and annual sales $sale_k$ volume of each evaluation record obtained in Section 4.3, we use PCA to reduce the dimension again, and select the main components to construct the comprehensive score P_k of commodity market performance. By arranging P_k in chronological order and drawing changes chart with time, we can observe the change trend of the market performance of commodities with time. On the one hand, these upward or downward trends can indicate whether the reputation of commodities is increasing or decreasing. On the other hand, we can observe the potential successful or failed products from these trends.

Furthermore, we can calculate the average annual performance X_m of commodities by using P_k:

$$X_m = \frac{1}{d_m} \sum_{k}^{d_m} P_k \tag{4.4.1}$$

where m represents the year, d_m represents the total number of sales records of the year, and X_m change chart about time m can be drawn to observe the change trend of commodity expressiveness on a larger scale.

4.5 Spectrum Analysis of Star Ratings and Reviews Based on Time

There is a certain relationship between star ratings and reviews emotion. Usually, high star ratings mean that customers are more satisfied with the product, and a pleasant shopping experience will make customers' comments more positive. Low star ratings mean that the shopping experience is bad for customers, and customers will express their negative emotions in the comments.

We regard the emotion carried by star ratings and reviews as random variables and arrange them in time order. By analyzing the characteristics of the two time series changes, we can judge the strength of the correlation between them. In addition, by analyzing the influence of time sequence changes, we can observe the effect of special stars on subsequent reviews.

Let the star ratings time series of a commodity be $Q(t)$ and the time series of reviews emotion scores be $R(t)$, because both are discrete random variables. In order to enhance the characteristics of the wave, Fourier transform is carried out on the two series, and the sine signal waves $(FQ)(t)$ and $(FR)(t)$ with two changes are obtained as follows:

$$(FQ)(t) = \mathrm{FFT}(Q(t)),\ (FR)(t) = \mathrm{FFT}(R(t)) \tag{4.5.1}$$

Take the product whose product_parent ID is 983445543 in the hair dryer as an example. The change of the transformed signal wave is shown in the figure below:

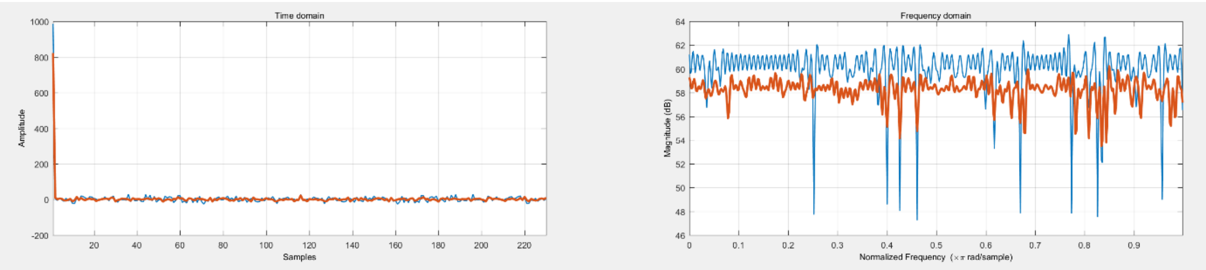

Figure 7: Time series signal wave of star ratings and emotion scores

Analyzing the same frequency of signal wave changes in Figure 7 and calculating the relevant numerical characteristics (such as correlation coefficient), we can see the strength of correlation between stars and comment types, which will be discussed in section 5.5.

In addition, we build another model chart for spectrum analysis, which will represent the effect of specific star ratings on reviews types. From the verified label, we found that the customers marked as “N” did not prove that they had purchased the product on Amazon, so the reliability of the reviews published by the customers labeled as “N” was low, and these comments and star ratings had a certain "induced".

In the timing chart of star rating $Q(t)$ and reviews emotion score $R(t)$, we use the hollow circle to indicate the record with the verified label content of “N”. The solid dot indicates the record with the verified label content of “Y”. By analyzing the characteristics of the graph, we focus on the spectral characteristics of the hollow circle and observe whether the changes of the comment are "induced".

Still take the product_parent ID 983445543 in the hair dryer as an example to draw the timing chart of star $Q(t)$ with verified label and reviews emotion score $R(t)$ as follows:

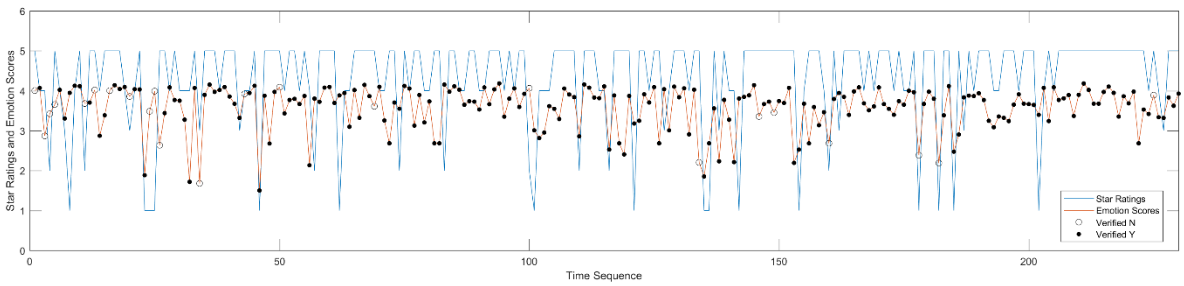

Figure 8: Time series of star ratings and emotion scores with verified labels

5. Model Solving and Question Answering

5.1 Data Measures Based on Ratings and Reviews (Requirement a.)

For Requirement a, we need to get the most informative data measure from star ratings and reviews. First, there are nine specific data items in the three data sets that have been given by statistical analysis, which belong to ratings and reviews. Seven of them are focused on, including star_ratings, helpful_votes, total_votes, vine, verified_purchase, review_body, review_date, etc.

Star ratings is intuitive and largely shows how satisfied consumers are. According to the analysis of the information given in the topic, the specific values are affected by the helpful_votes, total_votes, vine and verified_purchase. Consider changing star ratings into a data form that is easy for companies to track. For the convenience of the following other data impact directly on star ratings to produce an intuitive effect. Using the results in Section 4.3:

$$g_k = \left(S_k - 3\right) \times \delta_{vine}^k \times \delta_{verified}^k \times \delta_{vote}^k, \quad k = 1, 2, \cdots, n \tag{5.1.1}$$

This indicator can directly show the impact of other data on star rating, and it is representative to some extent. It will be the main indicator to be used in the following further problem solving.

By analyzing the sales data of a product, the annual average star ratings can be obtained, and then the variance of the annual average star ratings of any user after purchasing the product can be calculated. Variance describes the deviation degree of a single value from the average of all data, so when the star ratings variance of a user is large, the reliability of the star ratings is low. We set the function δ_{var} to describe the influence of star variance on star ratings:

$$\delta_{\mathrm{var}}^k = \begin{cases} 1 & \mathrm{var}_k > d \\ \lambda & \mathrm{var}_k \le d \end{cases}, \quad 0 < \lambda < 1, \quad d > 0 \tag{5.1.2}$$

where $d$and λ are fixed values and adjustable parameters.

Research shows that consumers tend to believe in the evaluation of a product by acquaintances. If there is no opinion reference of acquaintances, when an evaluation is long and has a story, which is helpful for consumers to understand the evaluators, the credibility of the evaluation will also be improved. Based on this theory, set the function δ_l to describe the effect of long comment on Star value:

$$\delta_l^k = \begin{cases} 1 & words_k > N \\ \gamma & words_k \le N \end{cases}, \quad 0 < \gamma < 1, \quad N > 0 \tag{5.1.3}$$

Where N and γ are fixed values and adjustable parameters.

Research shows that consumers tend to believe in the evaluation of products given by users with more evaluation times. In solving this problem, users with more than m evaluation times can be identified according to the customer_ID and recorded in a list. When analyzing the reliability of an evaluation, if the evaluation is evaluated by users in the list, the reliability of the evaluation will be improved. Based on this theory, a function δ_c is set to describe the influence of users with more reviews on star ratings：

$$\delta_c^k = \begin{cases} 1 & \text{customerID is in the list} \\ \mu & \text{customerID is not in the list} \end{cases}, \quad 0 < \mu < 1 \tag{5.1.4}$$

where μ is fixed value and adjustable parameter.

The annual average star ratings V after standardization shall be：

$$V = S \times \delta_{vine} \times \delta_{verified} \times \delta_{vote} \times \delta_{\text{var}} \times \delta_l \times \delta_c \tag{5.1.5}$$

Reviews data includes review_header and review_body as text information, which need to be processed to extract useful information. According to the analysis, the necessary information is mainly evaluation length and evaluation emotion. In the specific processing, we choose to use NLP in the field of artificial intelligence to process data.

According to the model in Section 4.2, combined with NLP and PCA, we can get the comprehensive emotion score E of reviews. Combining the data of star ratings and reviews, we can get the expression of evaluating product competitiveness F:

$$F = q \cdot V + (1-q) \cdot E \tag{5.1.6}$$

the company's managers can decide the size of q according to their will tendency.

when q is large, the company pays more attention to star ratings evaluation on product evaluation; when q is small, the company pays more attention to text evaluation on product evaluation. In addition, considering the requirements, in order to make the measurement contain more information, we can directly choose to use the measurement vector w to investigate the competitiveness of products in the market:

$$W = \begin{pmatrix} V & E \end{pmatrix} \tag{5.1.7}$$

5.2 Time-based Reputation Measures (Requirement b.)

By analyzing the data categories in the data set, data that can show the product's reputation in the online market includes star_ratings, helpful_votes, total_votes, verified_purchase, vine, review_body, review_dates, sales volume. In general, the reviewer is vine, reviewer's verified purchase label is ‘Y’, and the larger the sales volume of a product, the higher the star rating, the more active and helpful the reviews are, the higher the reputation level of the product in the online market; on the contrary, the lower the reputation level.

Using the results of Sections 4.3 and 5.1, we only need to analyze the metrics and characteristics of star ratings and reviews over time. Due to the correlation between star rating and review to a certain extent, in order to analyze the problem comprehensively and systematically, the model will analyze the star rating and review level of the product based on the time node by PCA , and get the comprehensive evaluation index of each product to show the reputation level of the product.

Using the model in section 4.4, we select two commodities with large sales volume from each category of products for analysis. The annual average score is obtained via formula (4.4.1), and then standardize the score to the range $[-2,2]$, the specific results are as follows:

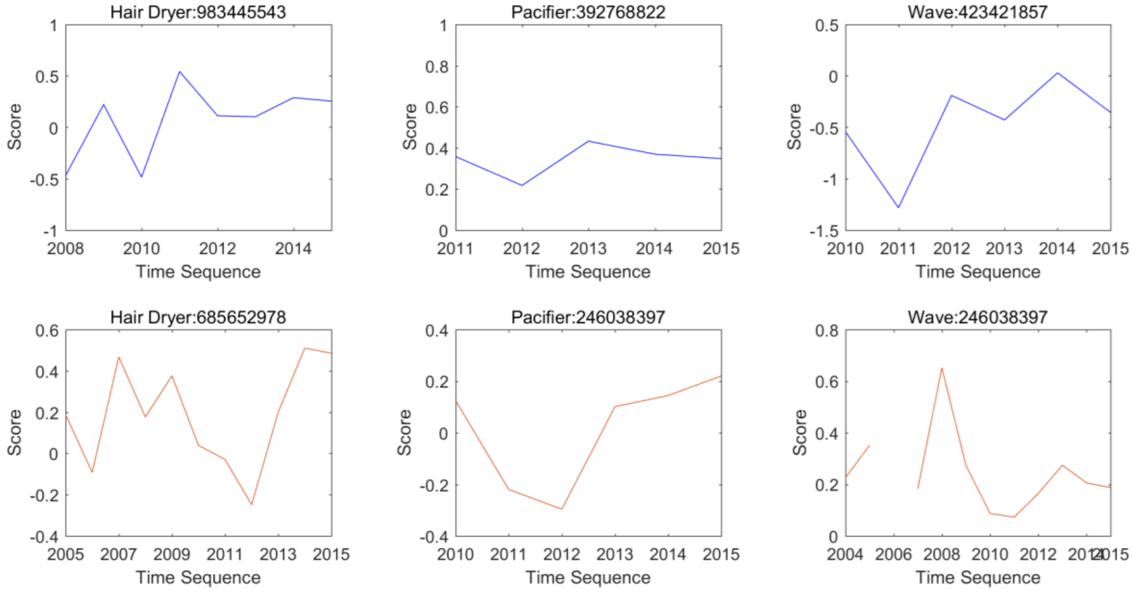

Figure 9：Time-based reputation changes of six products

From Figure 9, the evaluation indexes of all products have a fluctuating process over time. Next, we compare the reputation changes of the same kind of products:

Figure 10：Comparison of reputation changes of hair dryer

From Figure 10, we can see that the reputation of the two brands of hair dryers fluctuates with time. Specifically, the product_parent ID 983445543 of hair dryer reached the peak of reputation in 2011, and in the next few years, it showed a decline in reputation trend. The reputation of another hair dryer has been declining from 2009 to 2012 and reached its lowest level in 2012 and in the following years, the reputation began to show an upward trend.

Comparing the two products, the product_parent ID 685652978 has a better reputation than the other in terms of the latest year.

Here's another example of a microwave reputation comparison

Figure 11: Comparison of reputation changes of microwave

From Figure 11, it can be concluded that the reputation changes of the two brands are significantly different. The reputation of product_parent ID 246038387 of microwave has been declining in the following years since it reached its reputation peak in 2008. As for another product, its reputation has risen directly from 2011.

5.3 Measures of Success Based on Regression Model (Requirement c.)

The topic requires combinations of text-based measure(s) and ratings-based measures that best indicate a potentially successful or failing product. Therefore, we first define the meaning of a successful product. According to the marketing rules, the products whose annual average sales volume does not decrease could be considered as successful products, with more potential for long-term development and sustainable profitability. Thus, we calculate the annual average sales growth rate of products $rate$, and suggest that products whose $rate \geq 0$ are successful. Set δ_{rate} as a variable to describe whether a product is a successful product or not, then:

$$\delta_{rate} = \begin{cases} 1 & rate > 0 \\ 0 & rate \leq 0 \end{cases} \tag{5.3.1}$$

In Section 5.1, the measure vector W with the largest amount of information is given:

$$W = \begin{pmatrix} V & E \end{pmatrix} \tag{5.3.2}$$

V contains the rating-based information based on star ratings considering the influence of four data items: vine, verified_purchase, helpful_votes and total_votes; E contains the text-based information based on the scores which we get from review_body after using NLP and PCA processed. In section 5.1, a combination of information F that can be set by the company's officials is given:

$$F = q \cdot V + (1-q) \cdot E \tag{5.3.3}$$

But F influenced by q is more subjective. And the wanted combination should be more objective to predict the success of products in the future. So, we decided to do the regression analysis of δ_{rate} with vector W. Considering the value of the δ_{rate} is Boolean, we should do the logistic regression analysis with sigmoid function. And the definition of regression function is as follows:

$$\delta_{rate}^{i+1} = \frac{1}{1+e^{-\left(A^T W_i + b\right)}}, \quad i = 0,1,2,\cdots \tag{5.3.4}$$

According to the above equation, when we know the vector W_i of the year i, the equation will offer us the δ_{rate} of the year $i+1$, then we can predict whether the product is a potential successful product by seeing whether δ_{rate} is non-zero. When the future $\delta_{rate} = 1$, it would be safe to say that the sales of the product will increase, and it is a product with the potential of success.

By randomly selecting training set and test set from three data sets, the accuracy of regression fitting of δ_{rate} can reach 100% at most and 67% at least. In this way, we can say this regression model can predict whether a product has the potential for success with higher than 67% of the accuracy.

5.4 Specific Star Ratings Influence Reviews (Question d.)

There is a lot of mathematical evidence that star ratings influence reviews in the data set. Here we select one of the products as an example for corresponding analysis. Select the product with the product_parent ID 544821753 in the microwave, and use the spectral analysis method in section 4.5 to draw the time series diagram of the star rating $Q(t)$ and the review emotion scores $R(t)$ with the verified label as follows:

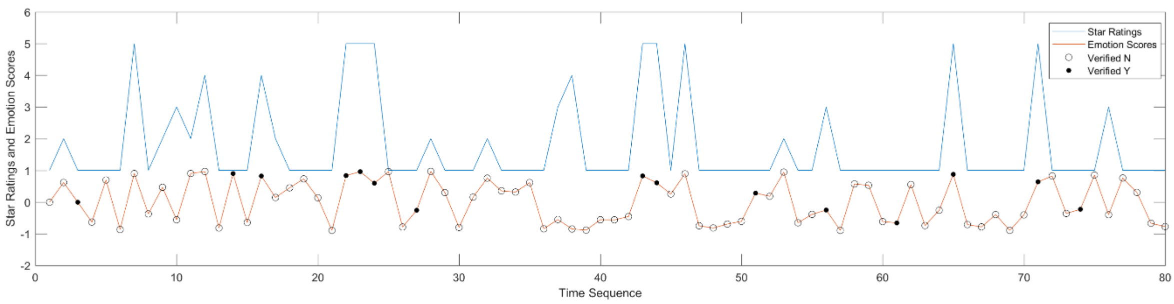

Figure 12: Timing chart of $Q(t)$ and $R(t)$ with verified label

Observing the above time series, there are two regions that show obvious characteristics of specific star ratings affecting reviews. Individually examine a sequence of time between 20 and 50, as shown below:

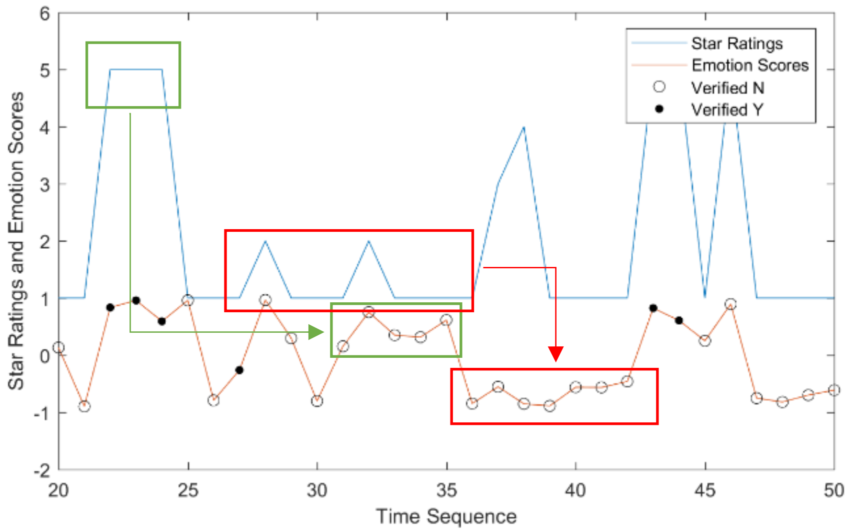

Figure 13: Part feature of Figure 12 (1)

From the figure above, it can be clearly observed that certain star reviews will have some kind of "inductiveness" on the evaluation of subsequent customers. The strong evidence to support this view is that the verified tags of customers who give positive or negative evaluation are all n (points marked by the hollow circle in the figure), that is, customers who have the same type of evaluation have not purchased this product on the Amazon platform.

We should not only doubt the reliability of these evaluations, can also suspect that they gave these similar evaluations because of seeing that the previous evaluations had some resonance, so that they could express their emotions through evaluations without purchasing. Another region that show characteristics of specific star ratings affecting reviews is as follows:

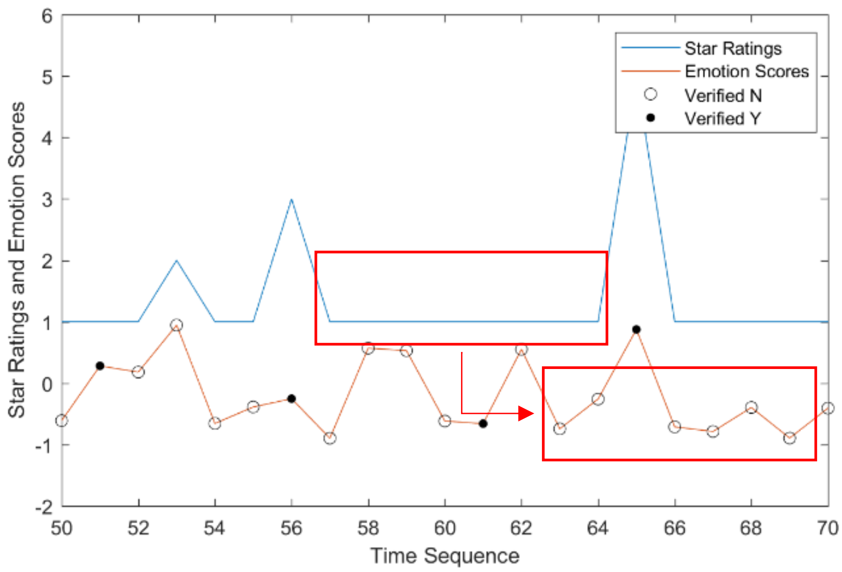

Figure 14: Part feature of Figure 12 (2)

Taking the above analysis as an example, by analyzing the data of other products in the data set, we get the following conclusions:

(1) A specific star rating will indeed bring more reviews, and most of the customers of these reviews have a verified label of N, that is, they have not been confirmed that they have purchased these products on the Amazon platform.

(2) A higher star rating and a lower probability of star rating will lead to more positive and negative reviews respectively, and they will express their sense of approval of these star ratings through the reviews. Low star ratings are more likely to trigger customers to express their bad, disappointed, and dissatisfied emotions than satisfaction and pleasure.

(3) The above phenomenon is a vent of customer's identification on the online platform. It's just that negative emotions are more likely to attract people's attention than positive emotions. In addition to online shopping platforms, online books, movies and other evaluation platforms will also have a similar phenomenon. This herd effect can make some customers known as "followers" of special opinions or reviews, and they will collectively express their favorite or disliked emotions, even if they do not fully understand the thing.

5.5 Correlation between Star Ratings and Reviews (Question e.)

In Section 4.2, we use NLP and PCA to convert the string variable reviews into quantifiable emotion scores, and compared with the star ratings, we determine the actual explanation of using the principal component as the score indicator. At the same time, in section 4.5, we use two discrete random sequences of star ratings and reviews emotion score, which are transformed into spectrum by Fourier transform. After the discussion and analysis of spectrum, we can preliminarily get the correlation between star rating and reviews.

Select the product_parent ID 983445543 in the hair dryer as an example, and use the model in section 4.5, the change of the transformed signal wave $(FQ)(t)$ and $(FR)(t)$ is shown in the figure below

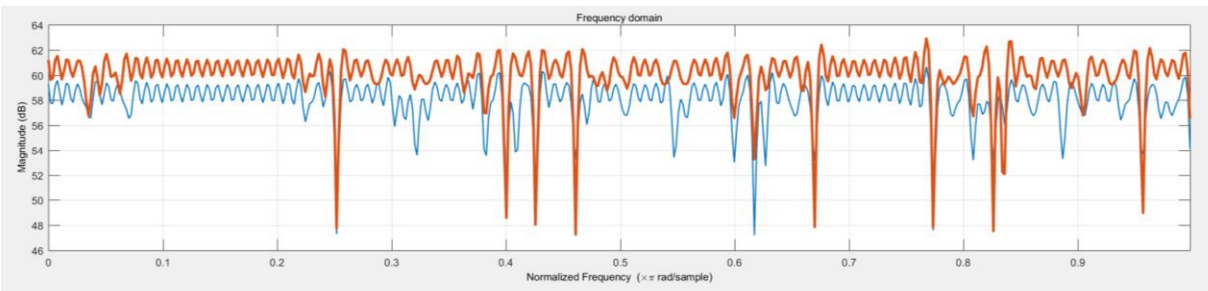

Figure 15: Time series signal wave of $(FQ)(t)$ and $(FR)(t)$

It is easy to get from the analysis of the above figure. Although the two waves are not identical, on the whole, the transformation of the two waves $(FQ)(t)$ and $(FR)(t)$ has strong same frequency vibration. We can use correlation coefficient to quantify the linear correlation between stars and emotional scores, use R_1 to represent the correlation coefficient of stars and emotional scores of random series, and R_2 to represent the correlation coefficient of two waves $(FQ)(t)$ and $(FQ)(t)$. R_1 and R_2 are defined as follows

$$R_1 = \frac{\mathrm{cov}(S,E)}{\sqrt{D(S)D(E)}} \tag{5.5.1}$$

$$R_2 = \frac{\mathrm{cov}(FQ(t),FR(t))}{\sqrt{D(FQ(t))D(FQ(t))}} \tag{5.5.2}$$

where $cov(FQ(t), FR(t))$ represents the covariance of two columns of wave $(FQ)(t)$ and $(FR)(t)$, defined as:

$$\mathrm{cov}(FQ(t),FR(t)) = E\left(FQ(t)\cdot FR(t)\right) - E\left(FQ(t)\right)\cdot E\left(FR(t)\right) \tag{5.5.3}$$

For the above example, the numerical calculation results:

$$\Re\left(R_2\right) \approx 0.9777\,,\quad R_1 \approx 0.6874 \tag{5.5.4}$$

Because the value of R_2 approaches to 1, the same frequency vibration of two waves $(FQ)(t)$ and $(FR)(t)$ is verified, which shows that the changes of two waves are closely related. However, $R_1 \approx 0.6874$ indicates that there is a strong positive correlation between stars and reviews.

Calculate the correlation coefficient of all products in the three data sets to get the average value. The average value of the correlation coefficient is as follows:

$$\overline{\Re\left(R_2\right)} \approx 0.8732\,,\quad \overline{R_1} \approx 0.6483 \tag{5.5.5}$$

Based on the above analysis, we can draw the following conclusions:

(1) The change of star ratings and the change of customers' reviews with emotion are very synchronous.

(2) There is a strong positive correlation between the star ratings and types of reviews. Specifically, high star ratings often correspond to the more positive reviews of emotions. According to the results of NLP, these reviews often contain "love", "like", "satisfied", "good" and other words that contain positive information and express pleasant emotions. On the contrary, the low star ratings often correspond to the more negative emotional comments, which often contain negative information such as "no", "unhappy", "bad", "dissatisfied" and words expressing disappointment.

(3) There are some limitations in the above-mentioned correlation between star ratings and text reviews. Sometimes, although customers give low star or high star comments, their reviews do not carry words to express their feelings, and they prefer to publish objective fact analysis in the comments. Therefore, the mean value of the correlation coefficient R_1 is not too large, which can reflect the positive correlation between high star ratings and positive reviews, low star ratings and negative reviews.

6. Model Evaluation and Analysis

6.1 The Advantages

(1) For basic data analysis in the early stages of model building, this paper considers star ratings, helpful_votes, total_votes, vine, verified_purchase, review_body, review_dates and sales volume and other data items, which can represent the performance of the product in the online market in more detail.
(2) For text-based user review_body, in this paper, NLP is used to solve the problem that natural language itself is complex and difficult to classify and quantify manually. It can quickly and effectively transform the user's review_body into a mathematical language that can be analyzed and classified quantitatively.
(3) In this paper, a combination of NLP and PCA processing method is used to deal with reviews. After NLP is used to analyze reviews' emotions, three numerical terms that can represent review' emotions will be obtained and will use PCA to carry out dimensionality reduction processing, which not only ensures the comprehensiveness of data, but also gives comprehensive indicators that can represent reviews emotions scores.
(4) For the use of time-based data metrics to express the reputation of the product, this article uses the PCA processing method, which integrates star_ratings, helpful_votes, total_votes, the vine, the verified_purchase, the review_body, the review_dates, sales volume and other data items ensure the comprehensiveness of the data, while giving a comprehensive indicator that can measure the reputation of the product.

6.2 Errors and Weakness

(1) We ignore the analysis of product_title, the attraction of attractive name to consumers, and the influence on product sales and reputation level. We should consider the influence of an excellent product name on the sales volume and reputation level.
(2) Due to the limitation of data set, we ignore the influence of market characteristics in different regions on product sales and reputation level. The conclusion of this paper can only be used in the sales strategy analysis of US region. The market characteristics and sales data of different regions should be considered for targeted analysis.
(3) We ignore the products category and lacks the analysis of the applicable population of products. We should consider the preferences of people who use the product and their views on the product, so as to improve the quality of the product
(4) We ignore the analysis of the reasons for the change of product sales and reputation level. We should not only analyze the trend of product sales and reputation level, but also consider the reasons behind it.
(5) We ignore the subjective impact of review_title, and only consider the emotion scores of reviews on review_body. In fact, review_title can bring more direct attraction and influence to other customers.

7. Letter

Dear Marketing Director:

We are the professional team to study market rules and customize scientific sales strategies based on big data, which you hired, and we are writing to introduce our latest research results to you with proud.

For the three products your company will sell online: microwave oven, baby pacifier and hair dryer, our team has made a detailed analysis based on the past data of other competing products, making full use of the data.

Key data measures have been successfully identified by us which will be most informative, easy to track and can make accurate predictions of product reputation and sales. We do hold the firm belief that with the help of our team's research results, your company will be able to easily grasp the market trend of online products in the future, and then formulate the best sales plans in time to dominate the market. Here is a brief summary of our team's work:

1. First, our team comprehensively analyzed possible effects of each data item on predicting product sales and reputation with economic expertise, aiming at obtaining the maximum amount of information;
2. Next, the data items retained after the above analysis were quantified and standardized. With the processed data items, we built time-based measures and patterns scientifically which will be able to facilitate your company's monitoring market and provide the largest amount of information.

   For example: After NLP, the principal component analysis method was used to finally convert the reviews into scores which could measure the feelings of the text;
3. Then, we used the previous time-based measures and patterns to construct specific models to predict the product sales and reputation; and after testing the models on existing data sets, we find that our model possesses high prediction precision and excellent forecasting effect.

   For example: The model to indicate the success of a product, which was built with combinations of text-based measures and ratings-based measures, had the accuracy higher than 90%;
4. Finally, based on the existing data, we used the measures and models identified in the previous research to analyze the present market trend and discover more market rules;

   For example: Based on the method of spectrum analysis, it was found that specific star ratings will incite more reviews on the same point of views even without purchasing products. And this kind of reviews are often not reliable, but they will have a huge impact on sales;

The above research results show that there is no doubt that our team has really grasped some market laws, so we can confidently recommend some practical sales strategies to your company:

Premise: during our research based on time variables, we discovered that sales and reputation are more easily affected by evaluations of the purchased users at the early stage of products' selling online, so managing the ratings and reviews of your products attaches great importance.

1. While selling products online, some preferential rules about making proper comments should be issued: after purchasing products, users offer higher star ratings and reviews with more words can enjoy a certain percentage of the price rebate;

   Reasons: when analyzing the effects of different data items on product reputation, the results of Principal Component Analysis show that: higher star ratings play a great role in improving product reputation. And when the number of words in a positive review is large, more consumers will be attracted to read and believe this evaluation;

2. In the early stage of product sales, it is suggested actively contacting some Amazon Vine members and offering them some free copies of products. Then warmly welcome them to write detailed reviews of products online and listen carefully to their advice about the products;

   Reason: specific research shows that the products used by Amazon Vine members and given positive evaluations will attract more users to purchase;

3. If a series of low star ratings are given, contact and negotiate with the customers who offered these comments in time to improve the evaluation;

   Reason: Based on the spectrum analysis, the comment and purchase behavior of consumers show obvious Herd Effect, which means after seeing a series of low star ratings customers are more likely to give the product poor evaluation even without purchasing one. And according to the previous Logistic Regression analysis, a large number of negative evaluations will directly lead to the reduction of product reputation and then sales.

4. Use our research results to timely track the reputation of your products and predict the future sales, so that your company can the sales strategies always in time and generate more profits;

In short, our research is based on real and effective data sets, and the research results also worked out excellently in the test: time-based data measures can truly reflect the information your company wants, and specific models can smoothly predict the market situations of your products. We do hope your company can adopt our research results after investigation, and we wish your company's products will sell like hot cakes in the future!

Sincerely
Team # 2013548

8. References and Appendix

8.1 References

[1] Xing Lao. Research on the influence of online reviews on commodity sales [D]. 2017.
[2] A first course in mathematical modeling: Frank R. Giordano and Maurice D. Weir. Monterey, CA: Brooks/Cole Publishing Company, 1985. 382 pp. $00.00 ISBN 0-53-403367-9[J]. 8(2):250-0.
[3] Gong Shiyang, Liu Xia, Zhao Ping. How does online consumer reviews affect product sales? -- An Empirical Study Based on online book reviews [J]. Soft science in China (6):176-188.
[4] Xia Huosong, Zhen Huachun, Zhang Yingye, et al. Research on the construction of professional domain knowledge model for the effectiveness classification of online commodity Reviews [J]. Journal of Information Science, 2016(35):954.

8.2 Appendix

Appendix 1:

Pyhton package used for calculation in this paper:

```
import xlrd
import pandas as pd
import numpy as np
from numpy import    *
import nltk
from nltk.sentiment.vader import SentimentIntensityAnalyzer
from sklearn.decomposition import PCA
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score
```

Appendix 2:

Schedule 1: Several sample data used in the paper

star_rating	helpful_votes	total_votes	vine	verified_purchase	emotion
5	4	5	0	1	-0.21672
4	1	1	0	0	-0.24949
4	1	1	0	0	0.351672
2	1	1	0	0	0.074009
5	0	0	0	0	-0.04306
4	3	5	0	1	-0.22509
3	1	1	0	1	0.133212
1	5	7	0	1	-0.18837
4	1	1	0	1	-0.27706

5	16	17	0	1	-0.27164
2	2	2	0	0	-0.05457
5	0	0	0	1	-0.06644
5	0	0	0	0	-0.22306
4	1	1	0	1	0.349389
5	1	1	0	1	0.092323
5	1	1	0	0	-0.21587
5	2	2	0	1	-0.28394
5	39	40	0	1	-0.23639
4	32	34	0	1	-0.26103
3	0	0	0	0	-0.14348
4	3	3	0	1	-0.23359
5	0	0	0	1	-0.2308
1	1	1	0	1	0.842979
1	0	0	0	0	0.042841
1	2	6	0	0	-0.21011
5	0	0	0	0	0.468399
5	0	0	0	1	0.065864
4	8	8	0	1	-0.25671
5	0	0	0	1	-0.09834
4	7	7	0	1	-0.09002
4	0	0	0	1	0.148002
4	0	0	0	1	0.926935
5	0	0	0	1	-0.2488
3	0	0	0	0	0.945962
5	1	1	0	1	-0.16312
5	0	0	0	1	-0.29116
5	1	1	0	1	-0.20037
4	0	0	0	1	-0.22433
5	0	0	0	1	-0.26108
5	0	0	0	1	-0.14193
5	1	1	0	1	-0.05021
3	0	0	0	1	0.125549
4	0	0	0	0	-0.17547
4	6	6	0	1	-0.18624
5	3	4	0	1	-0.27638
1	2	2	0	1	1.034664
5	0	0	0	1	-0.15211
5	1	1	0	1	0.446568
5	0	0	0	1	-0.20011
5	1	1	0	0	-0.2581
4	0	0	0	1	0.069681

5	0	0	0	1	-0.09894
5	0	0	0	1	-0.11305
4	0	0	0	1	-0.04817
5	0	0	0	1	-0.14934
4	1	1	0	1	0.718566
2	0	0	0	1	-0.11628
5	0	0	0	1	-0.07584
5	1	1	0	1	-0.2535
5	0	0	0	1	-0.26164
5	1	1	0	1	-0.06142
1	1	2	0	1	-0.15632
4	1	1	0	1	-0.186
4	0	0	0	1	0.236741
5	0	0	0	1	-0.22355
5	2	2	0	1	0.124989
5	0	0	0	1	-0.28776
5	0	0	0	1	-0.14618
5	0	0	0	0	-0.01934
4	0	0	0	1	-0.26285
4	0	0	0	1	0.158092
5	0	0	0	1	0.443596
5	0	0	0	1	-0.06657
2	2	2	0	1	0.010695
5	0	0	0	1	-0.27413
4	0	0	0	1	-0.24207
5	6	6	0	1	0.222812
5	1	1	0	1	-0.16133
4	1	1	0	1	0.182398
4	1	1	0	1	-0.08135
5	0	0	0	1	0.443596
5	1	1	0	1	0.443596
2	2	2	0	1	-0.29129
5	1	1	0	1	-0.20265
5	0	0	0	1	-0.27055
4	1	1	0	1	-0.22765
4	0	0	0	1	-0.03905
5	0	0	0	1	-0.08381
5	0	0	0	1	-0.07779
4	0	0	0	1	0.020191
4	0	0	0	1	-0.25601
5	0	0	0	1	-0.04583
5	0	0	0	1	-0.23161

5	3	3	0	1	-0.3059
4	0	0	0	1	0.110727
5	0	0	0	1	-0.11507
4	0	0	0	1	-0.24389
5	0	0	0	1	-0.01166
5	0	0	0	1	-0.17576
2	1	1	0	0	-0.24594
1	0	0	0	1	0.281034
4	0	0	0	1	0.378468
4	0	0	0	1	0.308759
4	1	1	0	1	-0.02205
5	0	0	0	1	0.01138
5	3	3	0	1	0.139089
5	0	0	0	1	-0.24191
5	0	0	0	1	-0.16757
5	0	0	0	1	-0.13443
2	1	1	0	1	0.356372
5	0	0	0	1	-0.29265
5	2	3	0	1	-0.25346
5	1	1	0	1	-0.12894
4	1	2	0	1	-0.12149
4	0	0	0	1	-0.26929
2	0	1	0	1	0.521975
5	0	0	0	1	-0.15594
5	0	0	0	1	0.443596
5	0	0	0	1	0.582779
5	0	0	0	1	-0.14974
1	1	1	0	1	0.196061
5	0	0	0	1	0.161939
5	0	0	0	1	-0.17049
4	0	0	0	1	-0.06725
5	0	0	0	1	-0.26044
5	0	0	0	1	0.443596
3	2	2	0	1	-0.23399
4	0	0	0	1	0.281581
5	0	0	0	1	-0.26591
5	0	0	0	1	-0.13315
5	0	0	0	1	-0.24521
4	0	0	0	1	0.331809
5	0	0	0	1	-0.22851
5	0	0	0	0	0.683107
1	2	2	0	1	0.859205

1	0	0	0	1	0.443596
5	0	0	0	1	0.006653
3	0	0	0	1	0.669283
5	1	1	0	1	-0.10023
4	0	0	0	1	0.148022
4	0	0	0	1	0.679397
1	0	0	0	1	-0.11704
5	0	0	0	1	-0.13707
5	0	0	0	1	-0.15707
5	23	23	0	1	-0.28452
5	0	0	0	0	0.108779
5	1	1	0	1	-0.04744
5	0	0	0	1	-0.07793
5	1	1	0	0	0.057994
5	1	1	0	1	-0.0861
5	0	0	0	1	-0.06131
5	0	0	0	1	-0.25086
5	0	0	0	1	0.689031
1	0	0	0	1	0.522868
4	0	0	0	1	-0.05221
5	0	0	0	1	0.443596
5	1	1	0	1	-0.01028
5	0	0	0	1	0.217694
4	0	0	0	1	0.053971
2	1	3	0	0	0.443596
5	0	0	0	1	-0.11108
3	0	0	0	1	-0.18707
5	1	1	0	1	-0.13766
5	0	0	0	1	0.087778
5	0	0	0	1	-0.20561
5	0	0	0	1	-0.25737
4	0	0	0	1	-0.05688
5	1	1	0	1	0.030304
4	0	0	0	1	-0.01827
5	0	0	0	1	-0.25688
5	0	0	0	1	-0.04702
5	0	0	0	1	0.01138
4	0	0	0	1	0.086482
5	1	1	0	1	-0.0846
4	0	0	0	1	-0.03962
4	0	0	0	1	-0.21417
5	0	0	0	1	-0.19567

1	0	1	0	0	0.592318
5	0	0	0	1	-0.04919
5	0	0	0	1	-0.20232
5	0	0	0	1	-0.11396
1	0	0	0	0	0.691059
5	0	0	0	1	0.093263
5	0	0	0	1	-0.27246
1	0	1	0	1	0.549188
5	0	0	0	1	0.333859
3	0	1	0	1	-0.1318
5	0	0	0	1	-0.15353
4	0	0	0	1	-0.14761
5	0	0	0	1	-0.1834
5	0	0	0	1	-0.09835
5	0	0	0	1	0.161759
4	0	0	0	1	0.242416
4	0	0	0	1	0.107588
5	0	0	0	1	0.124665
5	0	0	0	1	0.165337
5	0	0	0	1	-0.04049
4	0	0	0	1	-0.16992
5	0	0	0	1	-0.05221
5	0	0	0	1	-0.04702
5	0	0	0	1	-0.03651
1	2	3	0	1	0.08712
4	0	0	0	1	-0.25062
5	0	0	0	1	0.165197
4	0	0	0	1	-0.25836
5	0	0	0	1	-0.09889
5	0	0	0	1	-0.11937
5	0	0	0	1	-0.16045
5	0	0	0	1	0.101958
5	0	0	0	1	-0.16297
5	0	0	0	1	-0.30586
5	0	0	0	1	-0.2244
5	0	0	0	1	-0.05221
5	0	0	0	1	-0.05221
5	0	0	0	1	-0.1994
5	0	0	0	1	-0.26629
5	0	0	0	1	-0.18812
5	0	0	0	1	0.108779