Problem Chosen	2020	Team Control
C	**MCM/ICM**	**Number**
	Summary Sheet	**2002739**

With the rapid development of e-commerce, Amazon, as the premier online sales platform, uses ratings and reviews to make a relatively objective description of the product itself, thereby assisting other customers' purchasing decisions. Consumers use 1 (low rating, low satisfaction Degree) to 5 (high rating, high satisfaction) to indicate their satisfaction with the product, and customers can also provide richer opinions and information through text. The unique data combination helps the company to better understand the product market. Potentially successful product characteristics to develop a more favorable marketing strategy.

Sunshine plans to launch and sell three new products in the online market: microwave ovens, baby pacifiers, and hair dryers. Given that they are new to this particular data combination model, my team will act as a consultant, analyze and mine data, and serve as a sales director The proposed problem is to find a solution and applicable model. Based on the three data files provided by Sunshine's data center for this project: hair_dryer.tsv, microwave.tsv, and pacifier.tsv, perform the following operations.

First, 30,000 data are preprocessed by the spss tool, and secondly, a series of operations such as analytic hierarchy process, tf-idf algorithm are used to classify and quantify comments, and linear weighting is used for further data mining. With sufficient preparatory work , Explore the key data to measure the internal relationship and the fitting relationship between external variables such as time. Finally, after a series of big data calculations and comprehensive considerations, Sunshine Company gave a fair marketing strategy and recommended product design characteristics. To achieve accurate predictions of the potential success values of the three products issued by Sunshine.

Keywords: rating, product characteristics, marketing strategy

Contents

1. Research problem description

(I) Problem background

With the rapid development of e-commerce, more and more online merchants qualitatively evaluate products through ratings and reviews, thereby assisting other customers in their purchase decisions.At the same time, this specific combination helps the company to better understand the product market. Data metrics are tracked in time to develop more favorable marketing strategies.

(II)Problem Description

In consideration of this new combination mode, Sunshine Company commissioned me and my team to use data mining and algorithm analysis to provide three data files for the project from the data center: hair_dryer.tsv, microwave.tsv and pacer.tsv Draw out the potential success characteristics of the product and make the right marketing strategy for its plan to launch and sell three new products in the online market: microwave ovens, baby pacifiers and hair dryers.

(III) simple problem description

This paper uses 3 sets of data from Sunshine's data center to pre-process through SPSS tools, uses a series of operations such as analytic hierarchy process, TF-IDF algorithm, linear weighting, etc., uses matlab to fit the internal combination of data and external relations of time, and comprehensive analysis get conclusion.

2.Data set and feature selection

(I) Data source

This sample uses three datasets provided by Sunshine Company: hair_dryer.tsv, microwave.tsv, and pacifier.tsv. These data represent the ratings and reviews of microwave ovens, baby pacifiers, and hair dryers sold on the Amazon market during a specific time period , And the accompanying dataset definition.

(II)   Basic data analysis

Through observation, there are three data set tables, with a total of 32,022 samples and 15 features. Of the 10,755 product votes, a total of 49,760 valid votes were received. In the data "star_rating", buyers can use 1 (low rating, low satisfaction) to 5 (high rating, high satisfaction) to indicate their satisfaction with the product. Among the reviewers, they are subdivided into "vine" and "authentic_purchase" to explore the interior of the data portfolio Correlation between variables and time dimension.

(III)   Data processing methods and pre-processing processes

 The analysis of this sample data has two difficulties. The first is the quantification of reviews. For the quantification of reviews, we focus on the analysis of customer sentiment. We consider the Dependency-Sentiment-LDA model for sentiment binary classification to process text information, and then Use Python-TextBlob to score each review title and the entire review body. The score

range is [-1,1] (-1 is negative emotion, 1 is positive emotion). The second is for special "vine" and other special Parameter how we consider and deal with. We divide customers into "vine and buy", "vine and not buy",

"not vine but buy", "not vine and not buy", we ignore "vine and buy" based on realistic factors, where the NN customer group refers to non-members (people who comment objectively and truthfully) and have not been certified by the platform to complete the purchase There may be various problems such as large discounts, unrealistic purchases, and logistics delays, and their credibility is low. Therefore, our analysis mainly focuses on the weight ratio distribution of "vine and not buy" and "not vine but buy" people.

3.Solution

(I)Data measurement based on rating and comment weight ratio

（1）Comment quantification

When quantifying customer reviews, the topic focuses on the customer's emotions-satisfaction and dissatisfaction.We consider the Dependency-Sentiment-LDA model for emotion binary classification to process text information: documents and words are projected onto a set of topics , Trying to find the potential relationship between documents and words, between documents and documents, and words and words by topic; because LDA is an unsupervised algorithm, each topic does not require specifying conditions, but after clustering, it is calculated by statistics The probability distribution of the words on each topic, and those words with a high probability on the topic can describe the meaning of the topic very well. We use Python-TextBlob to score each review title and the entire review body, and the score range is [-1,1) (-1 is negative emotion, 1 is positive emotion).

（2）Examination review quantification

Considering that the customer's comments and ratings are the customer's subjective reflection and subjective score, respectively, we believe that all subjective reflections and subjective scores have a positive correlation, so we make a correlation analysis of reviews and ratings to ensure the accuracy of comment quantification. Data samples of three products in 2015 were included for all customers, without considering the validity of reviews.

Hair dryer products

Correlation		star	review
star	Pearson correlation	1	.495 **
	Sig. (Two-tailed)		.000
	Number of cases	11470	11470

		satr	revie w
rev iew	Pearson correlation	.495 **	1
	Sig. (Two-tailed)	.000	
	Number of cases	1147 0	1147 1

**. At 0.01 level (two-tailed), the correlation is significant.

Pacifier products

Correlation

		satr	revie w
satr	Pearson correlation	1	.448* *
	Sig. (Two-tailed)		.000
	Number of cases	5768	5768
rev iew	Pearson correlation	.448 **	1
	Sig. (Two-tailed)	.000	
	Number of cases	5768	5769

**. At 0.01 level (two-tailed), the correlation is significant.

Microwave products

Correlation

		revie w	star
rev iew	Pearson correlation	1	.500 **
	Sig. (Two-tailed)		.000
	Number of cases	549	549
star	Pearson correlation	.500 **	1
	Sig. (Two-tailed)	.000	

Number of cases	549	549

**. At 0.01 level (two-tailed), the correlation is significant.

Through the correlation analysis, the ratings of the three products are significantly correlated with the reviews, and are moderately related. Since it is not a strong correlation, we further analyze whether the review quantification is not accurate enough. Therefore, we performed anomalous case analysis on the sample data, and the default customer's rating Customer reviews are more reliable. Based on ratings, the number of anomalies in quantified reviews accounts for about 1% of the sample data. We roughly looked at the anomaly data in three products and found that

Although customers rated 3-5 stars when rating, they only pointed out the product shortcomings or used "although" statements in the reviews, which resulted in a large error between review scores and star ratings; in addition, customers also commented at will. The ratings are highly inconsistent. Considering the above reasons, we believe that our quantitative results are reliable and will be used in subsequent problem studies.

（3） Indicator weight

After reading part of the literature on reviews and ratings, we decided to use a judgment matrix for weighting.

The judgment matrix is   as follows

	Normal rating	Member rating	General comment	Member comments
Normal rating	1	1/3	1/4	1/7
Member rating	3	1	1/2	1/5
General comment	4	2	1	1/4
Member comments	7	5	4	1

After calculation, the consistency index $ci = 0.0347$ and the consistency ratio $cr = 0.0390$. Because $cr < 0.10$, the consistency of the judgment matrix a is acceptable.

The weights of the three methods are as follows:

Arithmetic mean	Geometric averaging	Eigenvalue method
0.0600	0.0580	0.0584
0.1331	0.1299	0.1288
0.2094	0.2087	0.2070
0.5975	0.6035	0.6058

In the end, we choose the intermediate amount, that is, the importance of ordinary ratings is 0.0580, the importance of member ratings is 0.1299, the importance of ordinary reviews is 0.2087, and the importance of member reviews is 0.6035.

Calculation of the average score of a product over a period of time

（4）Variable assumptions:

When calculating the average score of a product, we need to consider the validity of the review and the impact of external factors such as discounts and logistics on product sales.

The effective vote of a review is an objective reflection of other people's satisfaction with the product, so we consider the number of valid votes as the number of people who hold the review, and increase the influence of the review on the product. For external factors such as discounts, logistics, etc. People who comment on products are classified as non-ordinary non-members, and they are represented as vine-N & verified_purchase-N in the data sample.Except for specific cases, the ratings and comments of this group of people are not included in the total product score.

The quantitative indicators and ratings are unified indicators, and the indicator intervals are all [1,5].

Give the formula：review rating*(help vote+1)/(help vote+1)

(II)some exploration based on text and ratings

Through the analysis and quantification of comments, we give a review with a score ranging from -1 to 1, where 1 indicates infinite positive approach, and -1 indicates infinite negative approach. For these reviews, extract keywords for quantitative analysis. And for each comment that has been liked, then multiply his value by the number of likes, and finally calculate the total rating value.Then we average the star status of all members and non-members to get three Each product from 2005 to 2015, all members and ordinary users have their respective quantified values   and star rating results, and then we take the year as the unit, (the sources of the four weights are roughly mentioned) The data of each year is weighted to obtain the scores corresponding to the three products each year.The corresponding years and branches are fitted to obtain three corresponding curves.The fitting degrees of these three curves are different. There are exponential functions, quadratic functions, and cubic functions.

(a) explore the correlation between time and product reputation

（1）Pacifier

For the pacifier, according to the year and score data obtained, x represents the year time, and y represents the score score.

x=[2005,2006,2007,2008,2009,2010,2011,2012,2013,2014,2015];

y=[27.94,22.64,21.06,28.89,21.60,53.92,125.54,112.69,311.58,475.72,452.76];

first he performed an exponential fit y = exp (ax + b); at this time the goodness of fit is $R^2 = 0.8666$, the curve is as follows:
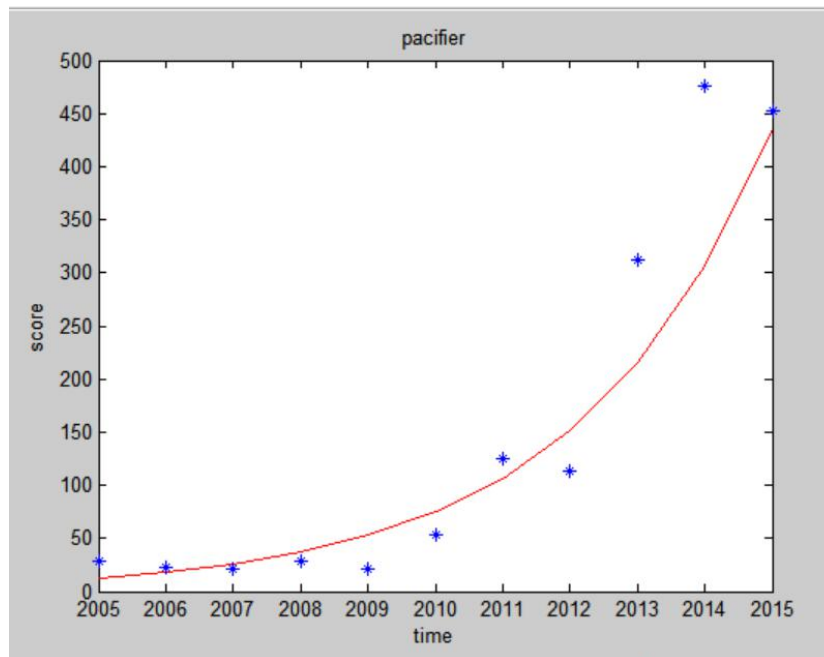
In addition, a second-order fit was performed on the function, that is, y = a * x. ^ 2 + b * x + c; at this time, the degree of fitting is higher R ^ 2 = 0.9371.

So suppose the result is:

  a = 8.3393; b = -3.3478e + 04; c = 3.3599e + 07

According to the function, it is known that the maximum value is x and takes the value of 449.70 in 2015.Observing the actual results shows that when x takes 2014, the maximum score is 475.72.
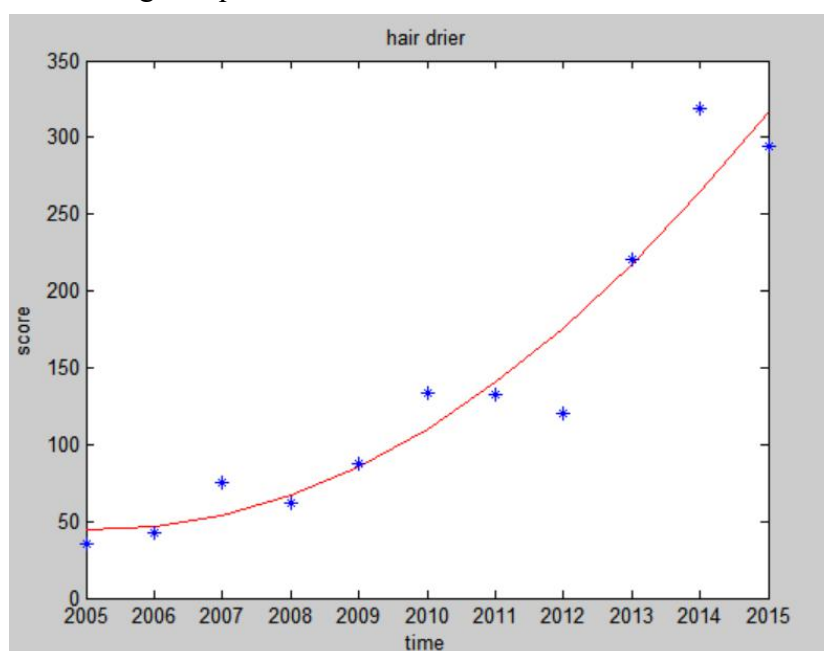
（2） Hair dryer

For a hair drier, the value obtained in this example is:

x=[2005,2006,2007,2008,2009,2010,2011,2012,2013,2014,2015];

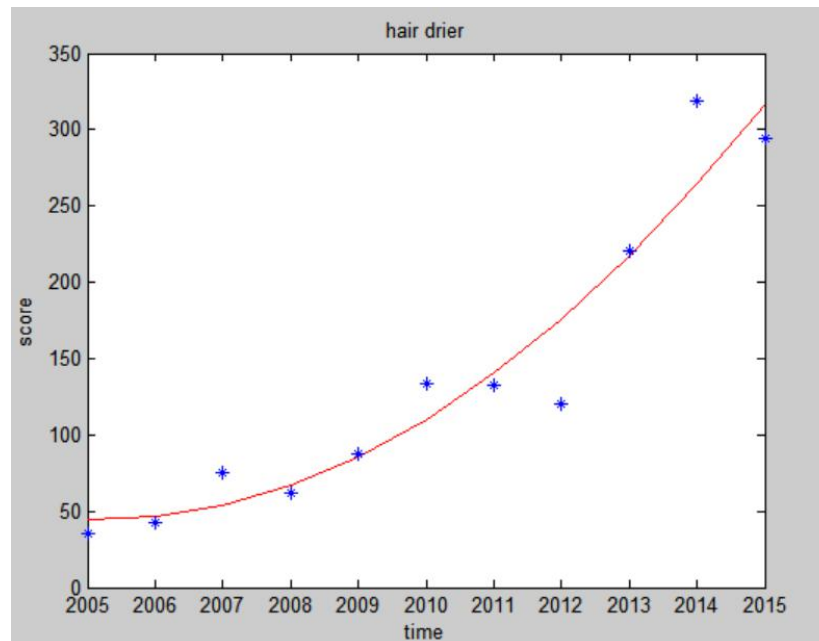y=[35.18,42.03,75.42,61.96,87.97,133.31,132.37,119.95,220.36,318.45,294.43];

The results of using the quadratic fit first are:

y=a*x.^2+b*x+c; a = 2.8014; b = -1.1235e+04; c = 1.1263e+07; R2 = 0.9191
The result of exponential fitting is:

R2 = 0.9194 The exponential function is better for this problem; a = 0.2152, b = -427.8121

y=exp(a*x+b);

According to the fitted function y is an increasing function of x, x = 2015 is, max (y) = 320.5170; the observation table shows that x = 2014 is max (y) = 318.45;

（3） Microwave oven

The data obtained in the microwave (microwave) is
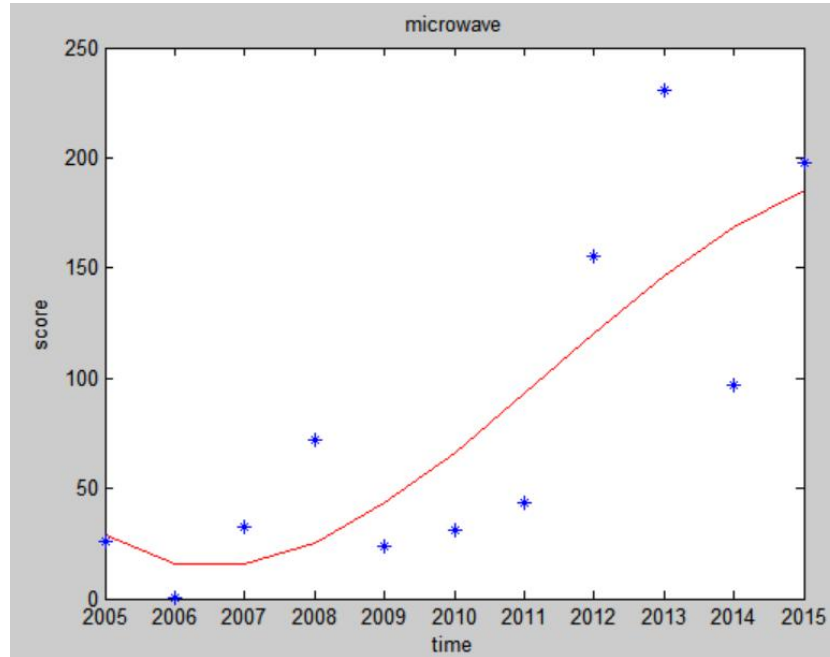
x=[2005,2006,2007,2008,2009,2010,2011,2012,2013,2014,2015];

y = [26.084, 0.06, 32.81, 71.76, 23.90, 31.27, 43.756, 155.026, 230.50, 96.89, 197.37]; first, the analysis result of the second fitting is
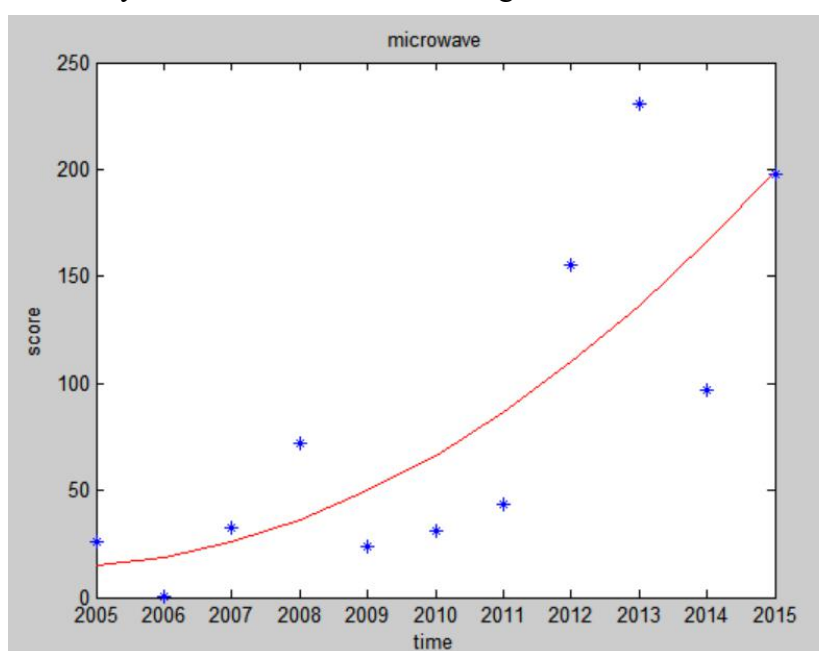
R2 = 0.6502 low degree of fit a = 1.6176b = -6.4843e + 03c = 6.4983e + 06
y=a*x.^2+b*x+c；

After making a fit about the cubic equation:

R2 = 0.6653, which improves the goodness of fitting compared to the quadratic fit; a = -0.3855, b = 2.3263e + 03

c = -4.6792e+06，d = 3.1372e+09

y1=a*x.^3+b*x.^2+c*x+d;

The maximum value observed was 230,50 in 2013, and the maximum value according to the fit was 2014, 185.0238.

Through the quantification of the review keywords, the assignment of rating and review weights is cleverly implemented, and the time is fitted to achieve changes in time through analysis and prediction of changes in product reputation. The three curves as a whole are extended to the upper right, that is, The reputation of the three products has changed over time.In addition to the temporary decline in the reputation of the microwave oven, the images of hair dryers and pacifiers have shown that their increase has increased over time.

(b) Analysis of product potential success characteristics based on text and ratings

Our team performed quantitative analysis and weighting on the score and text data combination for 05-15 years. Through the above steps, we obtained their fitted images with time and found the respective inflection points and peaks. The analysis can be obtained, and according to the function, we can know The optimal combination of reviews and ratings for the pacifier is the value of 449.70 in 2015 (this value is weighted). Observe that the actual optimal combination is 475.72 in 2014; the optimal combination of the function of the hair dryer and the optimal combination of actual observation are 2015. The year's 320.5170 and 2014's 318.45; the function optimal combination and the actual observation optimal combination of microwave ovens are 230.50 in 2013 and 185.0238 in 2014.

(III) Targeted impact of specific ratings on reviews

We explored whether the customer's emotions would move with the emotions of others by analyzing whether the ratings before the time node and the comments after the time node are relevant. When we analyzed the three products, we used the samples from 2015 Data selection data sample, the sample data volume is large and more reliable.

Due to study whether customers have a herd mentality and are affected by the emotions of others, we include all the people in our analysis.

Based on this problem we make the following assumptions:

1.There is a correlation between reviews and ratings within a time period.

2.Not all ratings in a time period will be correlated with comments and the smaller the time period, the greater the correlation, so it is assumed that there is a maximum range of time periods.

3. Only when the number of rated people is sufficient (> 10), the customer will be mobilized according to the emotions of others. Therefore, the time period can be week, half month and one month, depending on the sample size.

4.Reviews are related to the ratings of any previous customers. Therefore, no type of customer is excluded.

We first select the time period and time node, and then calculate the average score of the rating before a certain time node as the independent variable x.The comments after weighting the processing time node are regarded as the dependent variable y to get a set of (x, y), this operation is repeated, and finally multiple sets of data are obtained for correlation analysis.

We analyze the   the product of microwave and the time range of the rating and review is from January 1, 2015 to August 31, 2015. There are many days in the sample data of the microwave without relevant ratings and reviews,of which the sample data is small.So, we put 0 o'clock on the 1st of each month as a time node to perform data correlation analysis, the results are as follows:

Correlation

		y	x
y	Pearson correlation	1	-.685
	Sig. (Two-tailed)		.089
	Number of cases	7	7
x	Pearson correlation	-.685	1
	Sig. (Two-tailed)	.089	
	Number of cases	7	7

According to the results of the correlation analysis, the correlation between the

reviews of a certain month of microwave and the star rating of the previous month is not significant.It can be considered that there is no correlation.

Since there is a maximum range of hypothetical time periods, we shorten the time period to half a month under the guarantee of other assumptions. Select the 0 point on the 1st and the 0 point on the 15th as the time nodes for correlation. The results are as follows:

Correlation

		y	x
y	Pearson correlation	1	-.621*
	Sig. (Two-tailed)		.013
	Number of cases	15	15
x	Pearson correlation	-.621*	1
	Sig. (Two-tailed)	.013	
	Number of cases	15	15

*. At 0.05 level (two-tailed), the correlation is significant.

According to the results of the correlation analysis, the half-month reviews of microwave have a significant and negative correlation with the half-month star rating. It can be considered as a strong correlation owing to r at the range of 0.6-0.8. We can think that there is a high probability that customers will be motivated by the rating half a month ago and that emotion is the opposite of the rating emotion. We can think that the customer has both a herd mentality and an inverse mentality.

We analyze the product of pacifiers, and the time range of ratings and reviews is from January 3, 2015 to August 31, 2015. Because the sample data of the pacifiers has a lot of ratings every day , we put 7 days as a time range.The correlation analysis of the data is performed at 0 o'clock on the first day of every 7 days as the time node.The results are as follows:

Correlation

		review	star
review	Pearson correlation	1	-.344
	Sig. (Two-tailed)		.073
	Number of cases	28	28

tar	s Pearson correlation	-.344	1
	Sig. (Two-tailed)	.073	
	Number of cases	28	28

According to the results of the correlation analysis, the reviews of the pacifier for a week were not significantly related to the ratings of the previous week, and could be considered to be non-relevant.

Since there is a maximum range of hypothetical time periods, we guarantee data from August 4, 2015 to August 31, 2015 as sample data, while guaranteeing other assumptions, to shorten the time period to one day and exclude abnormal cases. Select 0 at each day as the time node for correlation analysis, the results are as follows:

Correlation

		y	x
y	Pearson correlation	1	-.330
	Sig. (Two-tailed)		.092
	Number of cases	27	27
x	Pearson correlation	-.330	1
	Sig. (Two-tailed)	.092	
	Number of cases	27	27

According to the results of the correlation analysis, the pacifier's comment on a certain day is not significantly related to the previous day's rating, so it can be considered as not related. Therefore, we can conclude that the customer of the pacifier product will not be rated by other customers Comments on the impact of this product, customers of this product do not have herd mentality when commenting, the comment is more sane and reliable.

We analyzed the product of hair dryer, and the time range of ratings and reviews was from January 3, 2015 to August 31, 2015. Because the sample data of the microwave oven has a lot of ratings every day, there are many sample data, we put 7 days as a time range. The correlation analysis of the data is performed at 0 o'clock on the first day of every 7 days as the time node.The results are as follows:

Correlation

			review	starweek
revie w	revie w	Pearson correlation	1	.376*
		Sig. (Two-tailed)		.049
		Number of cases	28	28
starw eek	starw eek	Pearson correlation	.376*	1
		Sig. (Two-tailed)	.049	
		Number of cases	28	28

*.   At   0.05   level   (two-tailed),   the correlation is significant.

According to the results of the correlation analysis, the weekly reviews of hair dryer are significantly and positively correlated with the ratings of the previous week. Since r is within 0.2-0.4, it can be considered a weak correlation. From this we can draw a conclusion : The customer's comments are positively related to the rating of the previous week. Customers of this product have a certain herd mentality when reviewing. When selling online, they must deal with bad reactions in time to ensure high scores and promote product sales.

(IV) The relevance of specific quality descriptions and ratings

In this topic, we used the LDA word frequency statistics method to count the frequency of the six keywords good like great bad awful worst in different star ratings of the three products, and plotted the following line chart.

Hair drier
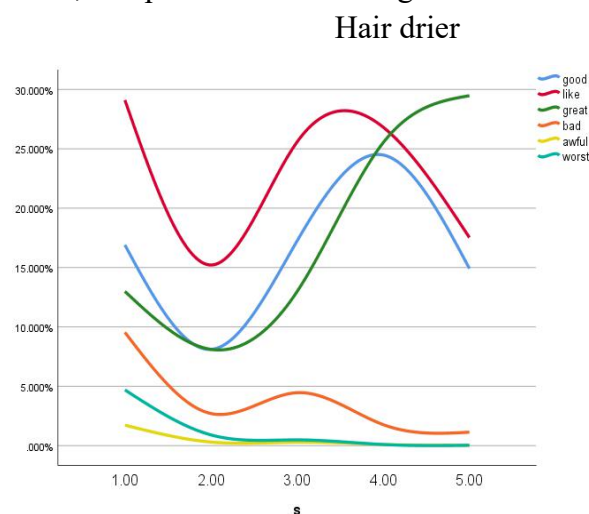
In the product evaluation of hair dryers, bad appears most frequently in one-star

reviews, bad appears more frequently in two-star reviews and Samsung reviews, and great appears most frequently in five-star reviews. The frequency is the highest. Therefore, it can be considered that great like has a higher correlation with positive evaluation, and Bad Worst awful has a higher correlation with negative evaluation.
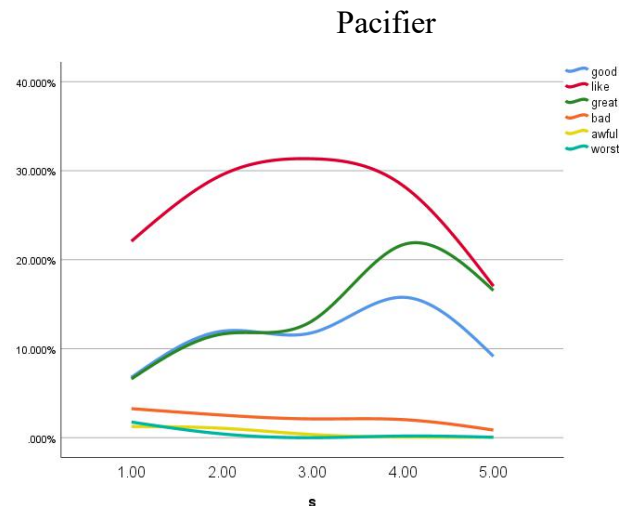
Pacifier

In the pacifier's review analysis, good great appears most frequently in four-star reviews, and the most frequent appears in five-star reviews. Like, Bad Worst awful appears less frequently with increasing star ratings, and appears in one-star reviews. Has the highest frequency and the lowest frequency in five-star reviews. Therefore, it can be considered that great like has a higher correlation with positive evaluation, and Bad Worst awful has a higher correlation with negative evaluation.
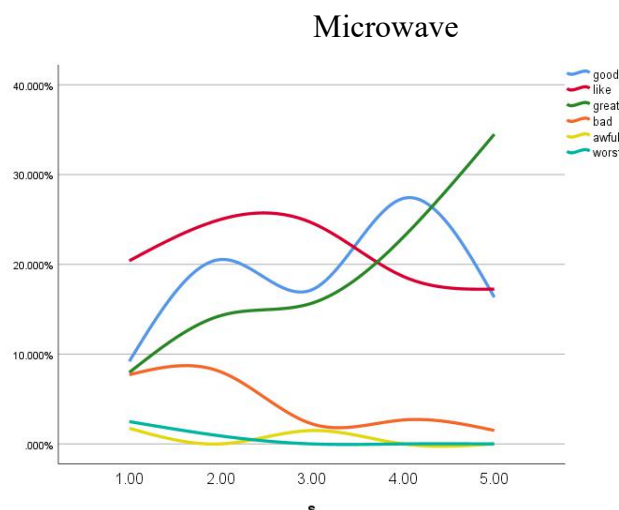
Microwave

In the evaluation of microwave ovens, five-star reviews have a greater proportion of great and good, and four-star reviews have a greater proportion of great, and like appears most frequently in two-star reviews. Bad Worst awful appears in one-star bad reviews It can be considered that great like has a higher correlation with positive evaluation, and Bad Worst awful has a higher correlation with negative evaluation.

4.Model optimization and reflection

After systematic consideration, we divided the indicators into several categories and adopted the analytic hierarchy process to obtain reasonable results. In determining the goals and measures, we will conduct a detailed analysis, and we show our result.
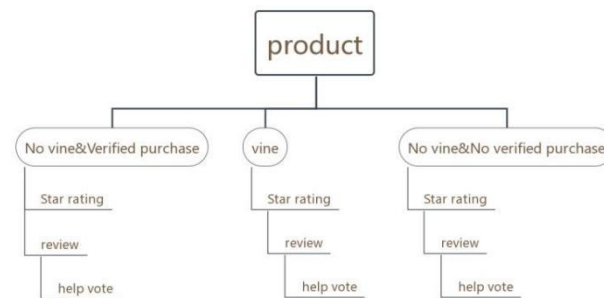
The disadvantage is that in the process of handling actual problems, we did not make full use of all the data, and the determination of weights was subjective.

When processing the annual data of the product, because the sales volume is also considered within the range of reputation, the quantified score of the product review is the total score, and no average operation is performed on it, but the corresponding user rating We performed an average operation from one star to five stars as an indicator of the star rating of the product.The weighted average of these four sets of data was later used to calculate the total situation of the product score of the year.

Dear director:

I am very honored to be a consultant for your company and engage in data

analysis. With the rapid development of e-commerce, more and more e-commerce platforms choose to use rating and review methods for product evaluation and information feedback. Simple operation can enable customers to get relatively intuitive and objective phenomena at the first sight of the product.Comments can further influence the customer's purchase decision. By voting on reviews, hire objective and fair users to become members and provide free Various measures such as user experience can not only optimize the customer's purchase experience, but also provide the company with more accurate and effective information feedback, which is conducive to the company's response and improvement. In recent years, the rise of big data cloud computing has enabled data mining not only Only as a reflection and summary of the past, and a basis and basis for predicting the future, I am very happy to see your company as a pioneer who dare to try this new type of data combination.My team and I will spare no effort to serve you.

From previous contacts, I learned that your company plans to launch and sell three new products in the online market: microwave ovens, baby pacifiers, and hair dryers. I got three of the projects provided by Sunshine's data center from the sales director. Data files: hair_dryer.tsv, microwave.tsv and pacifier.tsv.

In order to complete the director's requirements, we first analyzed and processed the data. First, 32022 data were pre-processed with SPSS tools, and there were 15 parameters in each sample, of which "time", "vine" and other special parameters Variables are worth paying attention to. Later we noticed that how to quantify comments is one of the most difficult points, so we adopted a series of operations such as TF-IDF and weighted analysis for further data mining. With sufficient preparatory work, we explored Key data measures internal relationships and fits to external variables such as time.

Immediately afterwards, through the quantification of evaluations and the allocation of rating weights, we defined a new value as the data metric that best fits Sunshine Company tracking after the three products are sold on the online market. On this basis, we will The data of this metric is fitted with the time of the external variables, and the three curves are extended to the upper right as a whole, that is, the reputation of the three products changes with time, except that the reputation of the microwave oven has a short-term downward trend. It shows that their increase increases with time. In addition, we get their fitted images with time and find the respective inflection points and peaks. Analysis can be obtained, according to the function, the optimal combination of reviews and ratings of the pacifier is the value of 2015 449.70 (the value is weighted), the actual optimal combination of observation is 475.72 in 2014; the optimal combination of the function of the hair dryer and the optimal combination of actual observation are 320.5170 in 2015 and 318.45 in 2014; The best combination and the best combination of actual observations are 230.50 in 2013 and 185.0238 in 2014. At the same time, we also pay close attention to the internal relationship of reviews and ratingsThere is a positive correlation between reviews and star ratings.A series of high-scoring ratings tend to be recognized reviews, while disappointed and angry reviews often have less than ideal grades.

For data mining and product analysis, combined with market trends and actual

operations, we make judgments and recommendations on the potential success characteristics of each product.

Microwave ovens do a good job in terms of balance of safety, size, and materials, and can maintain a reasonable cost. So the price is his competitiveness. Most users of microwave ovens are looking for convenience, fast cooking methods, and fast heating performance. The few, easy-to-understand operations that add recipes and instructions will definitely become the highlights and advantages of our products. Of course, for this appliance, safety guarantee and excellent after-sales are first.

The use process and use effect are indicators of the hair dryer. For most users, especially women, the feeling of using the hair and the effect on the hair quality after blowing is very important. Of course, the price is also an important factor in product sales.

Pacifier is a baby product, especially a product that is in direct contact with the oral cavity. Safety, hygiene and softness are the most important. For babies, they are relatively fragile and their safety comes first. And for the baby who cares for the baby For the nurses, it is also important whether the product is convenient.For example, compared to cheap pacifiers, water will leak and water stains will nourish the growth of black mildew.They prefer slightly more expensive but more hygienic products, so the price becomes slightly Minor factor.

In summary, according to the different characteristics of the product and the audience, we should specify a special marketing strategy. First of all, for the product itself, I think the positioning of the microwave oven is "affordable, good quality, easy to operate". According to data analysis, it is found that people The inherent impression is fast and simple. Cheap and easy to operate products are far more popular than expensive multifunctional products. Therefore, our design cost is mainly in appearance, product quality improvement, and after-sales service. Compared with products in the market, we It also has a price advantage. I think that hair dryers can be launched in two series, one of which is mainly "quality and cheap", while the other one takes the high-end sales route in terms of design, performance, and materials; corresponding to the general public and more for hairdressing. For the high-end needs of beauty people, firmly grasp the market customers to meet a variety of requirements. For baby's rubber pacifiers, I recommend the high-end route. First, we must choose safe materials, excellent technology, strict production, and produce certain Cost, the second is that this price is in line with the expected price range of most households with medium consumption levels, and the third is that good quality will attract many repeat customers to form products.Effect, but also to avoid a lot of product quality disputes.

Thanks again to Sunshine for the trust in me and my team.I   hope that my analysis and suggestions can be used for you.

Yours
Friends

References

[1]郝媛媛,叶强,李一军.基于影评数据的在线评论有用性影响因素研究[J].管理科学学报,2010,13(08):78-88+96.

[2]张丽. 在线评论的客户参与动机与评论有效性研究[D].南开大学,2011.

[3]马松岳,许鑫.基于评论情感分析的用户在线评价研究——以豆瓣网电影为例[J].图书情报工作,2016,60(10):95-102.

Appendix

Emotional analysis:

```
#-*- coding: UTF-8 -*-
from textblob import TextBlob
import time
f_path=r'C:\Users\DOCTOR\Desktop\temp.txt'
ww=0
data=open('C://Users//DOCTOR//Desktop//out.txt','w+')
with open(f_path,encoding='gb18030',errors='ignore') as f:
    for line in f:
        text = line
        blob = TextBlob(text)
        if(blob.sentiment[0]==0):
            ww=ww+1
            print(blob.sentiment[0],file=data)
        else:
            print(blob.sentiment[0], file=data)
print(ww)
```

matlab:

```
x=[2005,2006,2007,2008,2009,2010,2011,2012,2013,2014,2015];
y=[26.084,0.06,32.81,71.76,23.90,31.27,43.756,155.026,230.50,96.89,197.37];
p=polyfit(x,y,3);
a=p(1);b=p(2);c=p(3);d=p(4);
y1=a*x.^3+b*x.^2+c*x+d;
plot(x,y1,'r',x,y,'*');
xlabel('time');ylabel('score');title('microwave');
A=sum((y1-y).^2);S=sum((y-mean(y)).^2);
R2=1-A/S;
max(y)
```