The Multi-Dimensional Sales Analysis of Chinese Culture Products on Cross-Border E-Commerce Platforms Based on Web Texts

Tianya Guo[1] , Hui Du[2*]

[1]*School of Economics and Management, Beijing Jiaotong University, Weihai, China*
[2] *School of Economics and Management, Beijing Jiaotong University, Beijing, China*
[*]*Corresponding author. Email: guotianya737@outlook.com*

ABSTRACT

In views of lacking web texts based multi-dimensional sales analysis of Chinese culture products on cross-border e-commerce platforms, this paper firstly put forward the analysing dimensions and indicators depending on literature reviews, then collected and cleaned data about products, sellers, and comments from popular cross-border e-commerce platforms, and finally employed Tableau to make multi-dimensional sales analysis and provides sellers with corresponding suggestions. This paper is divided into introduction section, data preparation section, data analysis section and the conclusion.

Keywords: Web text, cross-border e-commerce, Chinese cultural goods, sales analysis, multidimensional analysis

1. INTRODUCTION

1.1. Background

With the rapid development of China, Chinese culture is becoming more internationally influential. Since 2002, the cultural products exports have been in a steady growth trend, showing great potential. At the same time, with the in-depth implementation of "Internet +" and "The Belt and Road Initiatives", cross-border e-commerce has become an essential channel for the export of cultural products [1]. So, it is not hard to figure out that e-commerce is a perfect opportunity to promote the growth the Chinese culture spreading.

In addition, compared with traditional off-line stores, e-commerce earns huge amounts of data resources in its web texts. So, in nowadays data-driven age, it is significant to employ such natural born advantage of e-commerce to help the healthy and steady growth of cross-board Chinese culture market by applying data based multi-dimensional analysis and data mining technologies on it.

1.2. Literature Reviews

Now the researches about web text based multi-dimensional sales analysis have existed: Jin Zhenjie [2] employed Python to collect book products data from e-commerce platforms, trying to provide purchasing guidelines for buyers. Lyu Boqing [3] designed a web crawler based on Selenium and collect clothes products data to make analysis on area, price, and category dimensions to support sellers' decision. Tu Jingwen [4] helped sellers choose the right products based on the clustering analysis of multi-platforms products data collected by Python. However, on the one hand, multi-dimensional analysis researches targeting at Chinese products have not existed, and on the other hand, the analysing dimensions are not comprehensive enough. Concurrently, researches about Chinese culture products has published: Zhang Yue [5] used Python to collect Chinese products data from several cross-border e-commerce platforms, and provide sellers with suggestions on choosing products by applying sentiment analysis; but the analysis is limited to the comments data with a single perspective. Given all the above, researches about web texts based multi-dimensional sales analysis of Chinese culture products on cross-border e-commerce platforms are still in deficiency.

1.3. Research Design and Paper Structure

To give a relatively comprehensive view of e-commerce Chinese culture products market, this research firstly tried to determine the analysing dimensions and Indicators, then used web crawler scripts to collect data from e-commerce platforms and cleaned them based on the determined indicators, and finally used visualization tools to visualize the data and analyse them from various dimensions.

This paper is organized as follows: in section 2, the analysing dimensions and indicators are introduced and the products data are collected and cleaned. In section 3, analysing results on time, buyers' location, price range and categories dimensions are put forward, followed by the

suggestions to sellers. In section 4, conclusions and the next work is provided.

2. DATA PREPARATION

2.1. Dimensions and Indicators

According to the summary of e-commerce multi-dimensional analysis researches studied by Zhang Yumei [6], Yang Linlin [7] and Wang Tao [8], this paper determined the analysing dimensions: time, reviewer locations, price section and category. In a further step, fundamental analysing indicators (can be obtained directly from the web texts) are determined: revenue, sales quantity, item numbers, price, products rating, seller quantity, seller rating comments quantity, and reviewer quantity.

2.2. Date Gathering

In this paper, four popular cross-border e-commerce platforms: AliExpress, Amazon, DHgate and eBay, are determined to be the sources of Chinese cultural products data. The preliminary investigation found that the Chinese cultural products are mainly in four categories: clothing, home, jewellery, and collectibles. So, the web crawler program runs based on the retrieval keywords "Chinese clothing", "Chinese home", "Chinese jewellery" and "Chinese collectibles" to collect the sample Chinese cultural goods data. In the end, the Python based web crawler gathered 73348 products information (including sellers' information) and around 200 thousand comments information in total. Selenium, IP proxy and automatic GUI are used to deal with anti-crawling problems.

2.3. Data Cleaning

Data Cleaning has five steps in all:
1. Denoising. Depending on whether the key words are included, 7268 Chinese culture unrelated products are got rid of. The relevance ratio is about 88.4% based on sampling inspection. For collected data, some punctuations are presented in ASCII, then use Pandas to replace the ASCII to regular punctuations.
2. Standardization. Use Pandas to standardize the price, time, location, and comments data from various platforms.
3. Absent data processing. For numeric absent data, use "NULL" to replace it. For absent characters, use "Unbranded" or "Unknown" to replace it.
4. Duplicated data processing: Use function "drop_duplicate ()" from Pandas to reject repeated data.
5. Data integration. Integrate data from various platforms into one final MySQL database. The database connection is based on PyMySQL under Python.

After cleaning, finally 66080 products information, 9900 sellers Information and 206018 comments data are remained for the following analysis.

3. DATA ANALYSIS

Based on the cleaned data, the data analysis part employs Tableau to analysis the Chinese culture e-commerce market situation from time, buyers' location, price section and category dimensions.

3.1. Time Dimension Analysis

As shown in Figure 1, the number of new sellers of Chinese cultural products on cross-border e-commerce platforms increased year by year, especially between 2016 and 2018. Although the growth rate seems slackened in 2019, the growing trend remains.

Usually, a product with high reviewers and comments also means relatively more revenue and buyers. So as seen in Figure 2, from 2010 to 2019, the revenue and buyer quantity of Chinese culture products can be considered in a rapid growth along with its reviewer number and comments number.

To sum up, the Chinese cultural commodity market is still in the stage of development. Sellers still have considerable potency to grow once they target the right market and products.
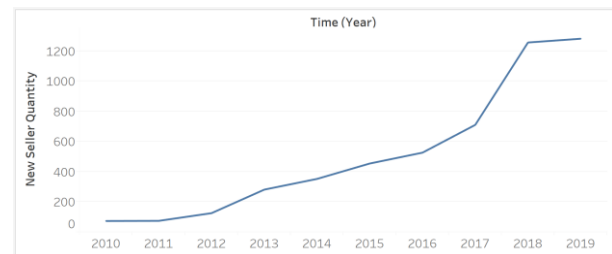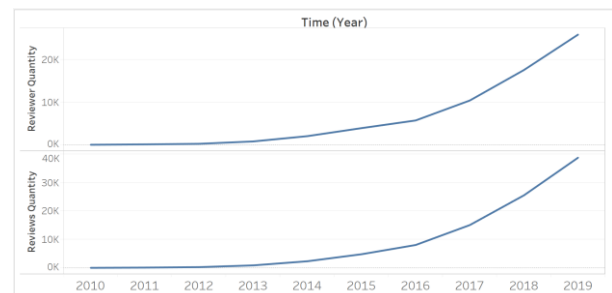

Figure 1 Annually New Sellers Quantity Line Chart

Figure 2 Annually Comments and Reviewer Quantity Line Chart

3.2. Buyers' Location Dimension Analysis

The Figure 3 shows that Russia and the United States rank closely top 2 in reviewer quantity ranking, and the number of reviewers from these two countries is significantly more than that from any of other countries.

The Figure 4 tells that Russia and the United States rank top 2 in comments quantity ranking but Russia has comments obviously much more than the United States. In other words, the number of products bought per capita in Russia is much more that that from other countries.

To sum up, sellers should mainly focus on Russia and the United States market. For Russia, in addition to increasing publicity, it is also necessary to enhance buyers' experience, such as adding Russian interface, or specially provide service in Russian.
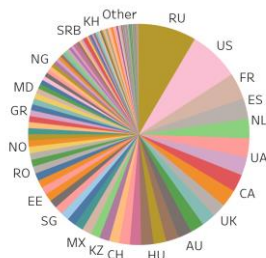

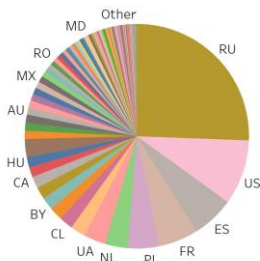**Figure 3** Reviewer Quantity Divided by Countries

Figure 4 Comments Quantity Divided by Countries

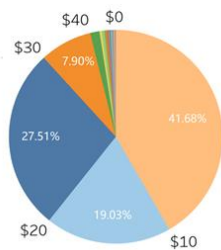### 3.3. Price Section Dimension Analysis

Figure 5 Products Quantity Divided by Price Range

As can been seen in Figure 5, products marks $10 or less occupies more than 40% of the total sales. While the products marks $10 - $20 (including $20) and the products marks $20 - $30 (including $30) have little difference in sales, the decline is significant compared with products less than $10. Similarly, products marks $30 - $40 fell significantly again while products mark higher than $40 only account for 3.88% of the total sale. It is apparent that with the rise of the price section, the proportion of the products sales shows a clear downward trend.

The competition index refers to the ratio of total merchandise number to the total sales volume, and is a derived indicator based on the fundamental indexes, and the smaller the index, the smaller the competition between sellers, vice versa [4]. As shown in Figure 6, the competition index of commodities within the price range of $40 is generally small and close, while the index for markets above $40 is generally much higher and more oscillated.

In summary, it is recommended that the sellers give priority to selling products under $40. For other products higher than $40, sellers should try to set the price within a price range with a lower competition index.
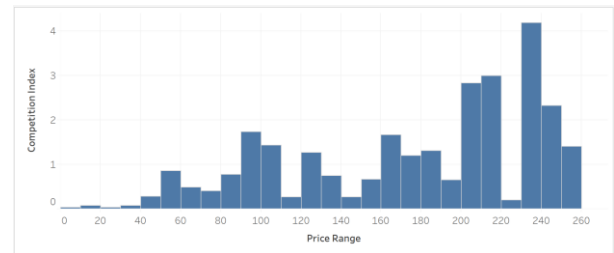

Figure 6 Price Ranged Competition Index Histogram

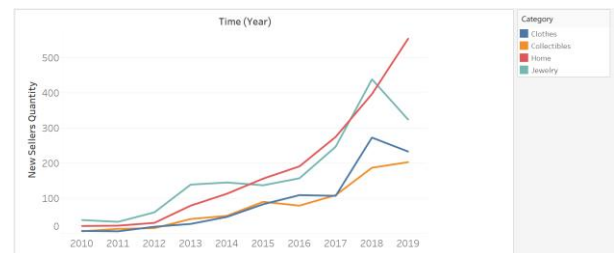### 3.4. Category Dimension Analysis

Figure 7 Line Chart of Annually New Sellers Quantity in Each Category Market

As can be seen from Figure 7, the number of new sellers in all categories markets is on an overall rapid growth before 2019. During 2019, the new sellers' number in home and collectibles market continue to rise, while jewellery and clothing declined significantly.
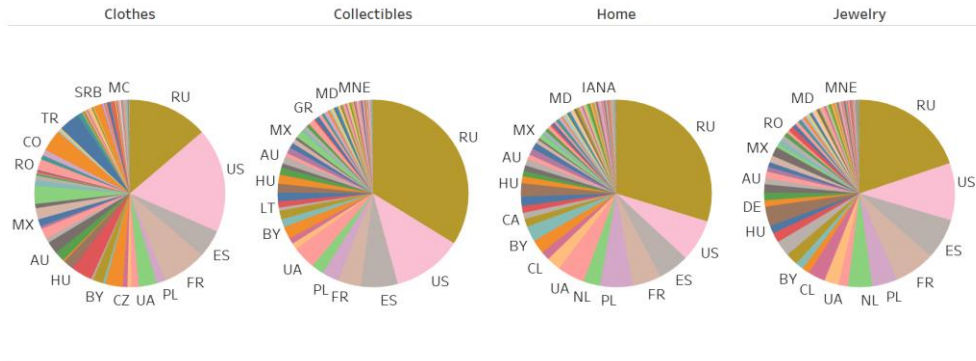
Figure 8 Pie Chart of Comments Quantity in Each Category Market Divided by Countries

Figure 8 shows that the United States accounts for the largest share of clothing sales, and is close to Russia the second one, and the both countries have tremendous precedence on other countries. Russia occupies the largest share of other three categories markets, and is much more than any other countries.

In conclusion, sellers can mainly focus on home and collectibles markets, and should pay attention to Russia market while keep an eye on the United States market

4. CONCLUSION

In views of lacking web texts based multi-dimensional sales analysis of Chinese culture products on cross-border e-commerce platforms, this paper made a research about this field based on fresh data collected from popular e-commerce platforms, and derive a structured view on the developing situation of the e-commerce Chinese culture products market.

In practice, based on the derived information and visualized charts, this paper provides some executable suggestions to Chinese culture products sellers, hoping to be beneficial to their development.

This paper also has limitations: the sources of data can be expanded to make the result more precise, and more analysis dimensions can be added to view the developing situation more comprehensively.

The next urgent and valuable work is to establish a Tableau-based automatic analysing platform for comprehensive dimensions and indicators.

REFERENCES

[1] Lei Xiaoqing, The expansion of international market of foreign trade manufacturing through cross-border electronic commerce under "mobile internet plus", the 4th International Conference on Economics, Management, Law and Education, (2018).

[2] Jin Zhenjie et al., Research on data crawler of electric business books based on python, Journal of Beijing Institute of Graphic Communication, 26 (2018) 39-42.

[3] Lyu Boqing, Analysis of e-commerce information based on crawler and data mining, Lanzhou University, (2018).

[4] Tu Jingwen, Research on seller-side product categories strategies in the cross-border e-commerce platform—take AliExpress as an example, South China University of Technology, (2019).

[5] Zhang Yue, Review sentiment analysis and sales forecast based item selection of foreign trade e-commerce, Beijing Jiaotong University, (2019).

[6] Zhang Yumei, Jin Yibo, Sci-Tech & Development of Enterprise, 9 (2019) 161-163+165.

[7] Yang Linlin, Statistics and Management, 10 (2016) 140-147.

[8] Wang Tao et al., Analysis and enlightenment of sales characteristics of fresh litchi on e-commerce platform, Guangdong Agricultural Science, 45 (2018) 140-147.