Saint Petersburg State University
Graduate School of Management

Master of Business Analytics and Big Data

Paid search analysis: Case of Big Data Analytics for French Cosmetic Company

Master's Thesis by the 2nd year students
Concentration — BM.5783.2019
Master in Business Analytics and Big Data
Li Xiran
Li Yijun

Research Advisor:
Zhukova Sofia
Professor, Information Technologies in Management Department

Saint Petersburg 2021

ЗАЯВЛЕНИЕ О САМОСТОЯТЕЛЬНОМ ХАРАКТЕРЕ ВЫПОЛНЕНИЯ ВЫПУСКНОЙ КВАЛИФИКАЦИОННОЙ РАБОТЫ

Мы, Ли Сижань, студент 2 курса магистратуры ВМ.5783.2019 «Бизнес-аналитика и большие данные (Master in Business Analytics and Big Data – MiBA)», и Ли Ицзюнь, студент 2 курса магистратуры ВМ.5783.2019 «Бизнес-аналитика и большие данные (Master in Business Analytics and Big Data – MiBA)», подтверждаем, что в нашей магистерской диссертации на тему «Платный поисковый анализ на примере анализа больших данных для французской косметической компании», представленной в службу обеспечения программ магистратуры для последующей передачи в государственную аттестационную комиссию для публичной защиты, не содержится элементов плагиата.

Все прямые заимствования из печатных и электронных источников, а также из защищенных ранее курсовых и выпускных квалификационных работ, кандидатских и докторских диссертаций имеют соответствующие ссылки.

Нам известно содержание п. 9.7.1 Правил обучения по основным образовательным программам высшего и среднего профессионального образования в СПбГУ о том, что «ВКР выполняется индивидуально каждым студентом под руководством назначенного ему научного руководителя», и п. 51 Устава федерального государственного бюджетного образовательного учреждения высшего профессионального образования «Санкт-Петербургский государственный университет» о том, что «студент подлежит отчислению из Санкт-Петербургского университета за представление курсовой или выпускной квалификационной работы, выполненной другим лицом (лицами)».

_____ Ли Сижань

_____ Ли Ицзюнь

STATEMENT ABOUT THE INDEPENDENT CHARACTER OF THE MASTER THESIS

We, Li Xiran, 2nd year master student of the BM.5783.2019 master program “Master in Business Analytics and Big Data - MiBA”, and Li Yijun, 2nd year master student of the BM.5783.2019 master program “Business Analytics and Big Data (Master in Business Analytics and Big Data - MiBA) ”, state that our master thesis on the topic “Paid search analysis: Case of Big Data Analytics for French Cosmetic Company” which is presented to the Master Office to be submitted to the Official Defense Committee for the public defense, does not contain any elements of plagiarism.

All direct borrowings from printed and electronic sources, as well as from master theses, PhD and doctorate theses which were defended earlier, have appropriate references.

We are aware that according to paragraph 9.7.1. of Guidelines for instruction in major curriculum programs of higher and secondary professional education at St. Petersburg University «A master thesis must be completed by each of the degree candidates individually under the supervision of his or her advisor», and according to paragraph 51 of Charter of the Federal State Institution of Higher Professional Education Saint-Petersburg State University «a student can be expelled from St. Petersburg University for submitting of the course or graduation qualification work developed by other person (persons)».

_____ Li Xiran

_____ Li Yijun

АННОТАЦИЯ

Авторы	Ли Сижань Ли Ицзюнь
Название магистерской диссертации	Платный поисковый анализ на примере анализа больших данных для французской косметической компании
Факультет	Высшая школа менеджмента
Направление подготовки	Бизнес-аналитика и большие данные
Год	2021
Научный руководитель	Жукова София
Описание цели, задач и основных результатов	Цель этого исследования-выбрать метрику для измерения эффективности платного поиска и помочь команде электронной коммерции Lancôme лучше понять пользователей платного поиска. Задачами исследования являются: 1. Выбрать метрику для измерения эффективности платной поисковой кампании. 2. Найти самые эффективные ключевые слова в платном поисковом канале. 3. Проанализировать основные функции пользователей платного поиска, поведение пользователей и обоснование. Основные результаты включают измерение эффективности платного поиска (на практике было протестировано 8 эффективных ключевых слов платного поиска), отчет о понимании пользователями платного поиска и стратегические рекомендации по бизнесу платного поиска.
Ключевые слова	Эффективность Маркетинговой Кампании, Платный Поиск, Поведение Пользователей

ABSTRACT

Master Student Names	Li Xiran Li Yijun
Master Thesis Title	Paid search analysis: Case of Big Data Analytics for French Cosmetic Company
Title Faculty	Graduate School of Management
Main field of study	Business Analytics and Big Data
Year	2021
Academic Advisor's Name	Zhukova Sofia
Description of the goals, tasks, and main results	The goal of this study is to choose a metric to measure paid search effectiveness and help Lancôme e-commerce team gain a better understanding of paid search users. The research tasks are: 1. Choose metric to measure paid search campaign effectiveness. 2. Find top efficient keywords in paid search channel. 3. Analyze core paid search user features, user behaviors, and the rationale behind. The main results include paid search efficiency measurement (with 8 efficient paid search keywords were tested in practice), paid search user understanding report, and strategic recommendations on paid search business.
Keywords	Marketing Campaign Effectiveness, Paid Search, User Behavior

Content

Content 4
1. Introduction 6
1.1. Background 6
1.2. Research motivation and goal 7
1.3. Research questions 8
2. Literature review on theoretical background of paid search 9
2.1. Digital economy 9
2.1.1 The concept of digital economy 9
2.1.2 Characteristics of digital economy 10
2.1.3 The change of marketing strategy during Digital Economy Era 11
2.1.4 The significance of marketing strategy transformation 11
2.2 Search engine marketing 12
2.2.1 The concept of search engine marketing 12
2.2.2 The mode of search engine marketing 13
2.2.3 Paid search advertising 14
Summary 15
3. Paid search business 17
3.1. Russian cosmetic market overview 17
3.1.1 Market sales 17
3.1.2 Category segmentation 18
3.1.3 Market distribution 18
3.2. Five forces analysis on Russian cosmetic market 19
3.2.1 Buyer power 19
3.2.2 Supplier power 19
3.2.3 New entrants 20
3.2.4 Threat of substitutes 21
3.2.5 Degree of rivalry 22
3.2.6 Conclusion 23
3.3 Lancôme paid search business background 24
3.3.1 Business Profile 24
3.3.2 Current paid search working process from e-commerce perspective 25
3.3.3 Managerial problems 26
3.4 Research methodology and research framework 27
Summary 30
4. Paid search analysis 31

4.1 Part one - paid search keyword efficiency 32
4.1.1 KPI on marketing campaign 32
4.1.2 Regression analysis 35
4.2 Part two - paid search user behaviour understanding 43
4.2.1 Data introduction 44
4.2.2 Methodology 46
4.2.3 Result 47
Summary 60
5. Discussions 62
5.1 Keywords test on real business 62
5.2 Impact of the Covid-19 epidemic 63
5.2.1 On Russian cosmetics market 63
5.2.2 On Lancôme paid search 64
5.3 Paid search in practice 65
Implications 67
Conclusion 68
References 69
Academic Literature in English 69
Academic Literature in Other Languages 72
Appendix 73

1. Introduction

1.1. Background

With the continuous development and popularization of the Internet, people's lifestyle has changed. At the same time, due to the continuous improvement of people's revenue level, people's consumption power and purchasing power are also gradually improving. Therefore, market products are becoming more and more diversified and personalized, which can not only meet the different needs of different consumers, but also promote the rapid economic development. However, product sales and marketing are inseparable. Only by building a scientific and reasonable marketing strategy can companies strengthen people's understanding of products and gradually increase product sales. Nowadays, in the era of material oversupply, the choice of consumers has become diversified. One effective way for enterprises to develop is to continuously promote the personalization of marketing strategies and product upgrading.

In the era of digital economy driven by Internet technology, a series of significant changes have taken place in the behaviour mode of consumers. It has become an important part of the behaviour mode of consumers in the Internet era to accurately search their own products through search engines. This means that the demand driven by consumers is generated, and the search engine is just in line with the marketing concept of the digital economy era. At the same time, search engine marketing has advantages in cost control and brand promotion. In this context, search engine marketing has become an important way and channel of enterprise product marketing.

Internet advertising has been the fastest growing marketing channel in recent years. In the United States alone, it accounted for about $60 billion in 2015. The rise of digital advertising is remarkable - it has doubled in the past five years alone. Paid searches that place ads next to search results account for the largest share of online advertising spending.[1]

When search engine first appeared, scholars thought that it would become an important marketing tool. (Yu, 2019) In 1994, the emergence of classified directory search engines such as Yahoo marked the birth of search engine marketing. In 2000,

[1] 于朝晖. CNNIC 发布第 44 次《中国互联网络发展状况统计报告》[J]. 网信军民融合, 2019(09):30-31.

the generation of pay per click mode led to the development of search engine marketing. With the development of Internet and the growth of Internet users, the search engine market expanded rapidly. According to relevant data, nearly 80% of Internet users choose search engine to obtain information, and a large part of Internet users have a strong sense of dependence on search engine.[2] Now, search engine advertising has become the best promotion channel for small and medium-sized enterprises.

1.2. Research motivation and goal

Paid search has become an important way of search engine marketing, and the measurement of its efficiency is also a topic concerned by scholars and enterprise managers.

Motivation. Previous research in the field of paid search basically only focused on the optimization of paid search algorithms. But, although data-driven is the concept of scientific decision-making, data-driven is not only about data theory. Behind the decision, we also should pay attention to the paralleled results of user experience and the big data.

Therefore, in addition to algorithm optimization, we should also gain insights into the behavioural habits of target consumer groups and dive into consumer needs. Currently, there is a lack of research on this part in the paid search field.

Goal. The goal of this study is to choose a metric to measure paid search effectiveness and help Lancôme e-commerce team gain a better understanding on paid search users.

Lancôme brand entered Russia in 1990. In recent years, its sales in Eastern Europe are gradually increasing, especially in Russia. Last year, all countries and markets were affected by the pandemic and lockdown measures. Online consumption continued to grow, which brought us to deep dive into paid search performance. Therefore, the research on Lancôme's paid search efficiency could help us have a better understanding of:

a) Working process in a big cosmetics company
b) The current circumstances of paid search in practice, including players, stakeholders, decision-making process, implications, and future trends.

[2] 李凯,邓智文,严建援.搜索引擎营销研究综述及展望[J].外国经济与管理,2014,36(10):13-21.

c) Measuring the efficiency of keywords in paid search channel of Lancôme company.
d) Current Lancôme paid search user behaviour and the rationale behind it.

1.3. Research questions

Part one - Paid search keyword efficiency measurement

a) What is the common way to measure paid search efficiency in business?
b) Which metric is suitable for Lancôme?
c) What is the correlation between efficiency and other indicators?

Part two - Paid search user analysis

a) How decision-making process works in paid search channel in Lancôme?
b) What is the current Lancôme paid search user behaviour and the rationale behind?

2. Literature review on theoretical background of paid search

2.1. Digital economy

2.1.1 The concept of digital economy

The term "digital economy" was first proposed as an academic concept in 1996 by Don Tapscott, the father of the digital economy, in his book Rethinking Promise and Peril in the Age of Networked Intelligence[3]. Since then, scholars have mainly understood the digital economy from the perspectives of resources, technology, and activities (Jianzhuang Zheng, 2020).

From the perspective of resources, many scholars understand the digital economy as an economy that uses numbers, data or digital information as key resources or elements. From a technical point of view, some scholars define the digital economy as an economic activity driven by digital technology. From the perspective of activity, some scholars regard it as a series of economic activities that use information technology to improve efficiency. At present, scholars studying the digital economy generally believe that resources, technology, and activities are the three core elements of the digital economy. In addition, knowledge in the form of data is the main production factor.[4] Although data is very important in digital economic activities, data cannot be directly understood as the production factors of the digital economy. The reason is that fragmented data has the nature of public goods and is not scarce. Only processed data are private. Only processed data can be directly put into production. Therefore, the ability to process data, which present as the "knowledge" behind the data, is the production factor.

In summary, the digital economy is an economy system of digital industry with knowledge in the form of data as its main production factor and a new generation of information technology as its core driving force.

---

[3] Tapscott D, *Rethinking Promise and Peril in the Age of Networked Intelligenc*e, McGraw-Hill Education, 2015

[4] Xianrong Yi, Yinyin Chen, Yushuang Wei, 《数字经济中的几个重大理论问题研究———基于现代经济学的一般性分析》, 2019.

2.1.2 Characteristics of digital economy

- **Break constraints of time and space**

Digital marketing in the Internet age can break the time and space constraints and ensure the convenience and integrity of services, thereby enabling enterprises and consumers to establish mutually beneficial behavior patterns. Consumers can learn about product-related information from all aspects of the Internet and choose products that meet their needs. At the same time, digital marketing also breaks geographical restrictions, and consumers can buy products on their actual demand regardless of the geographical nature of the product. This breaks regional protection to a certain extent and provides consumers with more diversified products and services to form the market healthy competition, which is conducive to enterprises' continuous development and progress.

- **Multimedia and Interactivity**

Digital marketing can not only occupy the network position but also use multimedia to help consumers understand product information in all aspects, thereby changing the traditional marketing strategy of single-channel and effectively promoting consumers' desire to buy. At the same time, companies can strengthen the analysis of consumers and show consumers products of different types and different positioning. This is conducive to product diversification and can also increase consumer attention rate. In addition, consumers can interact with companies online and can form specific cognitions of the company's service attitudes and service concepts. By communicating with consumers, companies can also strengthen their understanding of consumers and introduce them to the products which they need to achieve a win-win situation between the company and consumers.

- **Digital marketing has economic advantages**

Traditional marketing methods usually carry out marketing activities through stores or traditional media. This marketing model is not conducive to exaggerating the influence of enterprises. Therefore, the productivity and sales ability of enterprises are also limited. Digital marketing can diversify traditional marketing and expand marketing influence and attract consumers from different regions to buy products. Digital marketing is also conducive to the self-regulation and adjustment of enterprises. Moreover, although there are many ways of digital marketing and the scale is large,

because enterprise carries out marketing campaigns according to the habits of consumers, it can also reduce the investment of traditional marketing for enterprises to reduce the marketing pressure of enterprises.

2.1.3 The change of marketing strategy during Digital Economy Era

- **Using network to develop market**

Enterprises in the network marketing, not only through the network to develop the market, but also to promote market research's rationalization, provide users with more accurate services. Enterprises can use big data to push product links and marketing videos to users in need to improve market share. In addition, enterprises can increase the analysis of consumers, which can help enterprises gradually understand the purchase of consumers to develop different personalized services for different consumers.

- **Create demand for consumers**

With the improvement of people's economic level, people's consumption demand is becoming more and more diversified. Enterprises should change the traditional marketing strategy of meeting demand into the marketing strategy of creating demand. This can gradually enrich consumers' lives and create a personalized service system for consumers, effectively enhance the influence of products, and create a better brand image for enterprises.

- **Diversified marketing channels**

In the Internet age, all kinds of multimedia software and network platforms influence and transform people's lives. People gradually recognize this kind of entertainment. Therefore, enterprises should fully strengthen the cooperation of network software and app, which can promote the diversification of marketing means and enable different types of user groups to strengthen their understanding of the enterprise.

2.1.4 The significance of marketing strategy transformation

- **Promote the diversification of business transaction mode**

In the era of network economy, people's choices are gradually diversified. Therefore, enterprises should gradually diversify the marketing model and transaction

mode to meet the needs of consumers and provide a more convenient service experience for consumers. At the same time, many consumers are used to shopping on the e-commerce platform. This purchase habit has hit the traditional store transaction. Enterprises should also strengthen the adjustment of marketing strategies and promote the diversification of enterprise transaction mode to promote the continuous development of enterprises and make network marketing more reasonable.

- **Promote the change of traditional consumption and production mode**

Traditional consumption mode has time limit, but network marketing can effectively break this limit. Consumers can choose the products they need all day and have a more intuitive understanding of product information. At the same time, consumers can understand the market evaluation before buying products, which can also help consumers choose more economical and reasonable related products. Therefore, enterprises must pay more attention to product quality, which will also form a positive interaction, making the productivity and sales of enterprises gradually improve. Moreover, by understanding the types and times of consumers' browsing, we can also carry out fixed-point marketing for consumers to stimulate consumers' desire to buy and constantly expand the market space of enterprises.

2.2 Search engine marketing

2.2.1 The concept of search engine marketing

The concept of "search engine marketing" was first proposed by GoTo company (Edelman, 2006). From the user's point of view, search engine marketing is to transfer marketing information to target users as much as possible according to the way users use search engines and the opportunity of users to retrieve information (Sen, 2005). From the perspective of enterprises, search engine marketing refers to the marketing means that the enterprise adopts to enhance the natural ranking, launch paid search advertising, and other search engine-related behaviors through the enterprise website to make the enterprise website significantly listed on the search engine. Its purpose is to attract the target audience to visit the enterprise website (Telang, 2004). Scholars who study marketing practice generally regard search engine marketing as an important part of network marketing, through page optimization and advertising to enhance the

probability of enterprise website being concerned in the search engine results page. To sum up, search engine marketing is a marketing activity that enterprise websites that use search engines to promote products and services by changing their position on the search results page.

2.2.2 The mode of search engine marketing

Search engine is not only an important method of website promotion, but also the most common way for users to find new websites (Telang, 2004). Nowadays, there are two mainstream search engine marketing modes: paid search advertising and search engine optimization. In the paid search advertising mode, enterprises as advertisers buy keywords from search engine providers according to the characteristics of their products and services. If the keywords entered by the user in the search match the keywords purchased by the advertiser, the advertiser's webpage link will appear in the promotion link area of the search result page. According to the charging mechanism, paid search advertising can be divided into pricing ranking and bidding ranking. The pricing ranking is fixed according to the period. The first three locations have higher fixed prices, and the enterprise links on the back locations will be displayed in turn. The pricing and ranking mechanism has been basically replaced by the bidding and ranking mechanism. Bidding ranking refers to advertisers competing for the position of their web links in the promotion link area through keyword auction. After advertisers bid for a certain keyword, search engine providers mainly determine the location of advertisers' web links based on bidding and keyword relevance. The cost paid by advertisers is determined by the actual number of hits and the price, that is, to adopt the pay-per-click advertising mode.

Search engine optimization is to make the enterprise's website easier to be included by the search engine and get a good position on the search results when users search through the search engine to achieve the purpose of website promotion (Li and Lin, 2013). Search engine optimization can ensure the effective connection between the enterprise website and the search engine and improve the ranking of the website in the search results page under the selected keywords. Search engine optimization is subdivided into website content optimization, keyword optimization, search engine login, etc. Through these methods, the enterprise does not need to pay to the search engine provider so that the search engine can include the enterprise's website links in

the front of the organic search area in the search results page (Yalom and Kose, 2010). Generally, search engine optimization is known as the most fundamental and effective way to improve website ranking. It should be noted that the impact of search engine optimization on search user satisfaction is fuzzy. On the one hand, search engine optimization technology can help enterprises capture and index potential users more easily through search engines; on the other hand, search engine optimization technology can also use the loopholes of search algorithm to manipulate natural ranking, thus disturbing natural search. It is generally believed that search engine optimization brings new opportunities to the Internet market and brings great challenges to the search engine ranking of the market and web pages (Li and Lin, 2013).

Paid search advertising and search engine optimization are the main search engine marketing mode, and both have their own advantages and disadvantages. On the one hand, for search engine optimization, advertisers spend less and mostly one-time, while for paid search advertising, advertisers need to pay for each click of search users, which costs more. At the same time, paid search advertising also faces the risk of invalid click. But on the other hand, although the cost of different keywords is different, paid search ads tend to show higher stability in the search results after paying the cost, and the location retention time is longer. Search engine optimization is greatly affected by the search engine algorithm. To ensure that each keyword appears in the top position, there are many uncertain factors. When the search engine adopts the new reference algorithm, the existing optimization results will become worthless. In addition, in the case of competitors also take optimization strategy, the ranking time is short (Chen, 2011).

2.2.3 Paid search advertising

- **Keywords price forecast and bidding strategy**

Keywords are the core elements of paid search advertising. Advertisers want to get the best position with the lowest price, which involves understanding the search engine auction mechanism and the bidding competition with other advertisers. In terms of price forecasting, most existing researchers establish their own price forecasting models or functions based on different theories, such as regression models, genetic algorithms, etc. (Yang, 2006). Some researchers divide keywords into brand keywords and general keywords and find that users' behaviors are different in the face of different

types of keywords from the Perspective of users. In terms of bidding strategy, the optimal bidding strategy of advertisers is often closely related to bidding rules. Many researchers choose the game theory model for general discussion. But the bidding among enterprises is a dynamic process, and the information that enterprises grasp is not enough.

- **Keyword click through rate and conversion rate**

In general, the possibility of users clicking on ads decreases rapidly with the decline of ad location. Only when the ads with high click rate are placed in the front position, can the advertisers' revenue be improved and the users' satisfaction with the ads be improved. If users judge that the brand information in the search results fits their own needs according to their own knowledge and experience, then the user's click-through rate and conversion rate are relatively high. In recent years, through different analysis methods, many achievements have been made in this field. Some scholars find that although advertisers will buy a group of keywords, most of them cannot bring benefits that means the conversion rate is 0. By studying the influence of the location of paid search ads on the profit generated by ads, scholars found that the top ads have the highest click-through rate, but the profit is not the highest. Sometimes, the cost to get a higher location is greater than the profit.

- **Auction mechanism**

The auction mechanism determines the final ranking of advertisers. A good auction mechanism should ensure that the high bid ads get more hits to improve the advertisers' willingness to pay and ultimately increase the revenue of search engine providers. Compared with the mechanism based on revenue ranking, the competitive ranking could increase search engine providers' profits.

Summary

In this chapter, the analysis of the theoretical background related to paid search starts from the general background and gradually understands it.

According to many scholars, today's era has become the era of digital economy. In the era of digital economy, digitized knowledge is a significant factor of production. Data has also become an important resource in business activities. The efficiency of digitized knowledge and data utilization can determine whether a company can catch

up with the times. For the Lancôme brand, sales and marketing through online channels have become an important part of its business.

Now, online marketing has become an important part of brand advertising. Internet marketing has the characteristics of not being restricted by time and space, multimedia, interactivity, and economy. Companies need to use these characteristics of network marketing to open markets and create demand for consumers. At the same time, it diversifies its own marketing methods and adopts targeted marketing methods for different consumers.

After that, we learned about search engine marketing and paid search advertising. Paid search advertising has several characteristics: keywords price forecast and bidding strategy, keyword click-through rate, and auction mechanism.

Through the research on these related theories and combining the data in this research, we determined the focus of the research to be the evaluation and improvement of the efficiency of paid search for Lancôme brand.

3.1. Russian cosmetic market overview

The cosmetic market mainly consists of sales of skincare, makeup, fragrance, and body care.

From 2015 to 2020, the Russian cosmetic market has achieved strong growth. In 2020, the total revenue of the Russian cosmetic market reached 605 million RUB. The compound annual growth rate (CAGR) from 2015 to 2020 is 3.9%.

Facial skincare sales accounted for the highest value in the Russian cosmetic market in 2020, equivalent to 19% of the market's overall value. In comparison, perfume sales reached an equating to 18% of the total market value, which drops slightly.

3.1.1 Market sales

The Russian cosmetic market grew by 4% in 2020. The compound annual growth rate of the market in the period 2015–2020 was 3.9%.

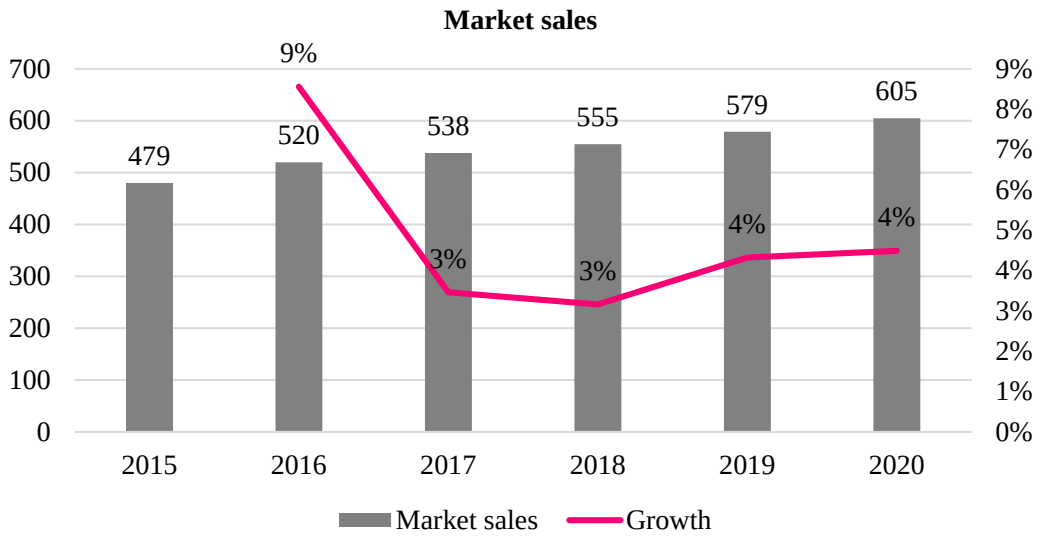


Figure 1 Total sales of Russian cosmetic market. Source: Росстат

3.1.2 Category segmentation

Facial skincare is the largest segment of the cosmetic market in Russia, which has increased in recent years. The perfume segment accounted for a further 18% of the market in 2018, followed by the makeup segment accounts for 15% of the market in 2018.

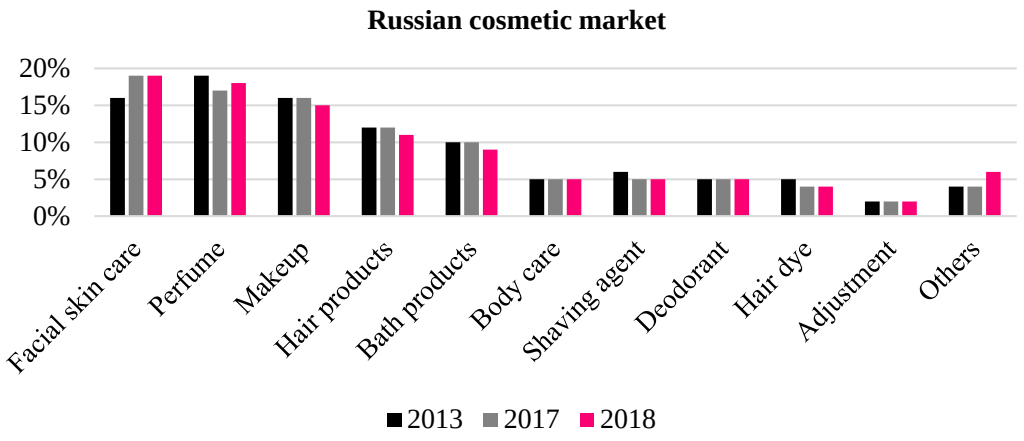


Figure 2 Total sales of different cosmetic category. Source: Poccmam

3.1.3 Market distribution

Direct Sellers form the leading distribution channel in the Russian cosmetic market, accounting for a 27.8% share of the total market's value. Health and Beauty Stores accounted for 27% of the market.

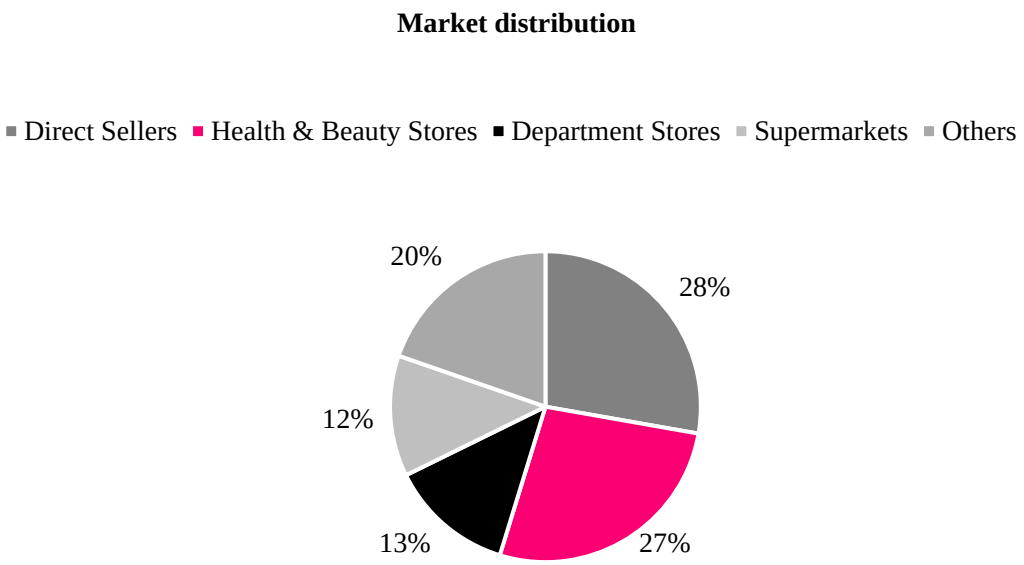


Figure 3 Market distribution of cosmetic. Source: Marketline

3.2. Five forces analysis on Russian cosmetic market

The cosmetic market will be analyzed, taking manufacturers of cosmetic products as players. The key buyers are retailers, chemical companies, and mineral producers as the key suppliers.

3.2.1 Buyer power

Direct sellers are the leading distribution channel in the Russian cosmetic market, accounting for 27.8% of the overall value in 2019.

Advertising is the key to the cosmetic marketing. Large companies have a large scale of marketing budget. This allows these companies to build brand loyalty through many marketing means, including expensive celebrity endorsements.

Retailers are positioned at the end of the value chain, which means they have an obligation to provide buyers with what they want in a market affected by unpredictable changes in fashion. This reduces buyers' purchasing power, as most retailers have to reserve popular brands to maintain their sales.

With brand loyalty exists, consumers prefer brands with better design to retailer brands, although some designer brands also have their own retail business.

Overall, buyer power in the Russian cosmetic market is moderate.

3.2.2 Supplier power

Cosmetic products are usually made from various chemical and mineral products, such as essential oils, which are widely available from many chemical companies.

Packaging is another important input, for example, plastic for Eyeshadow containers and glass for nail polish bottles.

The quality of raw materials is important, and the chemicals used in these products must meet the appropriate standards for the manufacture of consumer goods. In addition, they must be tested to ensure that they do not cause any adverse reactions

to human health. Any ingredients used must be proven safe, and, in many countries, regulatory approval must be obtained. This may weaken the strength of suppliers, as they have to prove that they can meet these conditions, although suppliers with good track records may increase their bargaining power.

For cosmetic manufacturers, supplier switching costs are negligible because inputs are usually not different (if they met regulatory standards). Products can be manufactured with a range of alternative raw materials. This reduces the capability of the supplier.

Suppliers tend to be smaller than large cosmetic manufacturers, so their strength has declined.

Overall, supplier power in the cosmetic market is moderate.

3.2.3 New entrants

Russian cosmetic market is composed of few brands (such as L'Oréal, Coty) some of which are widely recognized and have a strong market position. Large companies benefit from economies of scale, enabling them to compete more effectively on price. As a result, new companies entering the market will find it difficult to compete on their own.

Retaliation by existing players, such as launching a price war, is possible, especially when new entrants enter more concentrated markets. The financial strength of big companies such as L'Oréal and Avon make it difficult for smaller newcomers to compete on price. In addition, this financial strength means that these players have a larger marketing budget, so they can use celebrity services to enhance the attractiveness of their brands. Smaller entrants will find it difficult to match this form of capital-intensive marketing.

Starting a business in this market usually requires many kinds of funding resources, including production, distribution, and advertising investment (which is important to the success of this market). However, due to the high sales volume of cosmetics products and low degree of product differentiation, it is not uncommon for new companies to enter the low-end market of cosmetics.

New entrants may be able to start on a small scale, operating within a particular niche, such as make-up with anti-aging components. However, product testing and research would be both time-consuming and costly.

In addition, new entrants will need to persuade stores to stock their products, and major retailers, aware of their importance in the distribution chain, may be unwilling to risk displacing existing brands for the sake of new ones.

The regulation of cosmetics in Russia is very strict. Government legislation covers all aspects of the market, from labeling to testing methods and safety.

Due to the high brand strength of leading cosmetics manufacturers, it is difficult for new companies to develop their own brands and compete at the international level. The brand strength of major manufacturers is likely to offset most of the impact of low switching costs.

Despite this, strong market growth during recent years may attract new entrants.

Overall, the threat of new entrants with respect to the cosmetic market is moderate.

3.2.4 Threat of substitutes

There are few substitutes for cosmetics.

Manufacturers may face indirect competition from traditional cosmetics, although these cosmetics are unlikely to be a serious threat on a large scale in major markets. In some underdeveloped markets, the use of homemade products may be significant. Domestic hair care products can prevent users from contacting some potential allergic ingredients.

Mineral cosmetics are another alternative. They are more natural, light, and non-allergenic because there are no chemicals, spices or preservatives in their contents. Mineral cosmetics include two kinds of foundation and loose powder. These products claim to eliminate wrinkles that consumers get during the aging process, acne, acne scars or any other skin stains.

Counterfeited cosmetics constitute a serious substitutive threat for retail outlets and manufacturers, offering the benefit of a much lower price, though at the potential cost of lower quality.

Overall, the threat of substitutes is assessed as weak.

3.2.5 Degree of rivalry

The cosmetics market in Russia is moving towards concentration. In 2019, the top four companies accounted for 48.4% of the total market in terms of market value.

As many large enterprises have their own production facilities, fixed costs are relatively high, which helps to enhance competition.

Retailers may be reluctant to switch between market participants because their customers may seek leading brands. The diversified product range provided by some major players, including skincare products and hair care products, has reduced their dependence on cosmetics, thus alleviating competition to a certain extent. This diversification protects the company's business from competitive pressures in any market.

Because most companies have large production facilities, they need to divest these assets when they exit the market, which constitutes a high exit barrier, so it is the driving force of competition. However, most of the leading companies are geographically diversified, which weakens competition to a certain extent.

The strong market growth in the review period tends to ease competition to a certain extent. Overall, the competition in the Russian cosmetic market is moderate.

3.2.6 Conclusion

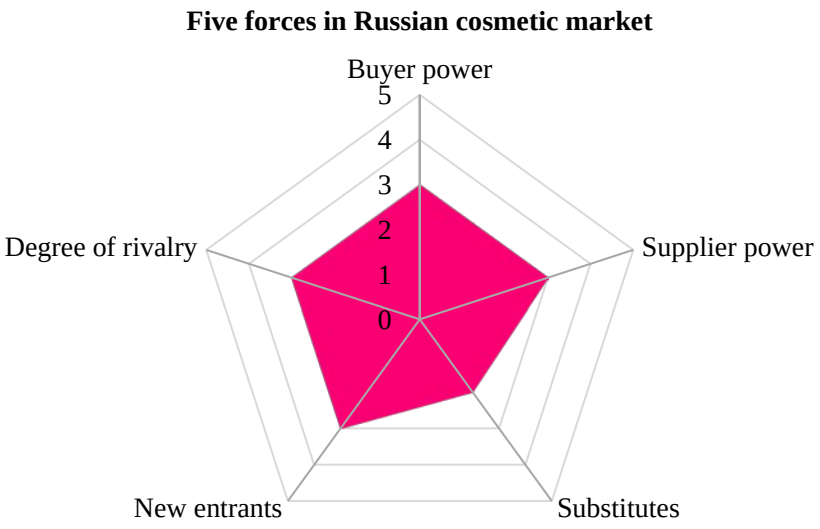


Figure 4 Five forces analysis of Russian cosmetic market

To sum up, the rivalry in the market is assessed as moderate. Some big companies (such as L'Oréal and Coty) have strong competitive brands and have the characteristics of regional diversification. In addition, relatively high fixed costs exacerbate competition as they increase the cost of exiting the market.

The large cosmetic companies own a variety of recognized brands and operate in various segments of the market. Fashion is a major influence on the cosmetic market, with consumers differentiating themselves quite strongly through the styles and brands of cosmetic products offered.

The pull of consumers, the range of products consumers need, and the scale of large cosmetics manufacturers such as L'Oréal and Coty prevent the buyer from becoming extremely powerful.

Due to the availability of raw materials of end products and low conversion cost, the supplier's ability is also moderate.

Access to this market depends largely on growth prospects and the size of existing players. Many markets are dominated by large multinational companies with a large marketing budget. This has deterred potential new entrants. Cosmetics have few

alternatives, but manufacturers may face indirect competition from traditional cosmetics. However, they are unlikely to pose a serious threat to the main markets.

3.3 Lancôme paid search business background

3.3.1 Business Profile

In L'Oréal Group, Helena and Giorgio Armani are in the same brand category as Lancôme. They belong to the luxury brand with high-end products. Each brand has its own development strategy and business goals each year.

Within Lancôme, paid search is a cross-department business, with two departments using paid search channels at the same time to help themselves achieve business growth independently. They are marketing department and e-commerce department. The two departments have completely different purposes and KPIs but use the same service platforms in the Russian market (Google and Yandex). The marketing department focuses on the conversion and awareness before and after marketing activities, while the e-commerce department focuses on the final purchase behavior of users. In the end, how many people finally complete the purchase by clicking on the advertising link is the focus of the e-commerce team. Therefore, although the two departments use the same channel, the keywords setting, and the traffic and budget setting of each keyword are different.

In Lancôme's paid search case, there are mainly 5 participants involved as the figure 4 shows. The first participant is the brand itself. The brand director adjusts its strategic goals according to the development requirements of the group every year. The second participant is the e-commerce department, responsible for the operation and execution of e-commerce sales and part of the paid search. The third participant is the marketing department, which is responsible for all marketing business and promotion things. Paid search is an important channel for the marketing team to track people's awareness of the brand and products, especially after marketing campaigns. The fourth participant is the paid search user group. The user group is the receiver of paid search services. People type words in search engines such as Google and Yandex search information they are interested in, some of them will click on the advertising link shown

on the top of the result page. After they look through the information on the website, they close pages or complete the purchase. The last participant is the paid search service provider. Yandex and Google provide search engines and recommendation algorithms to help brands use paid search services.

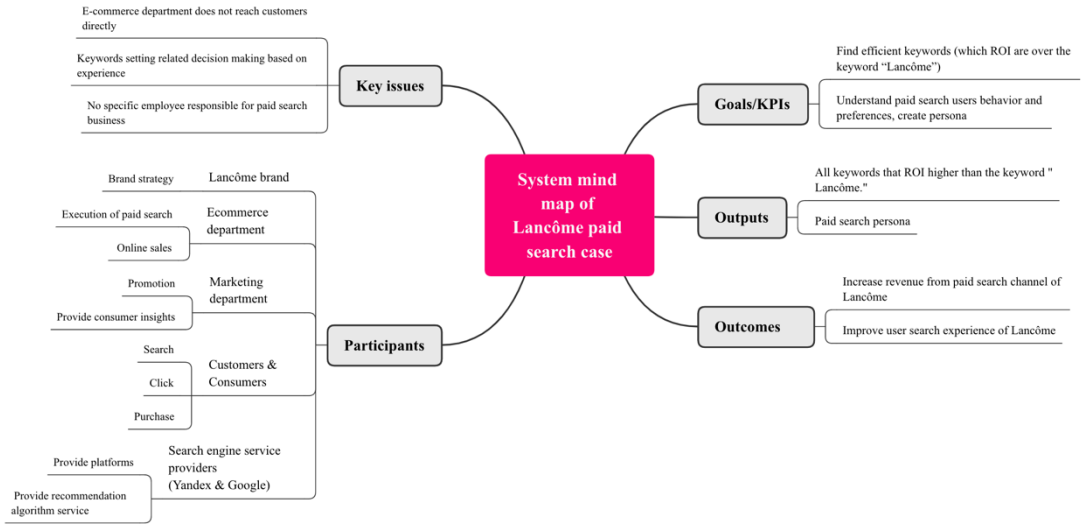


Figure 5 System mind map of the research case of Lancôme

3.3.2 Current paid search working process from e-commerce perspective

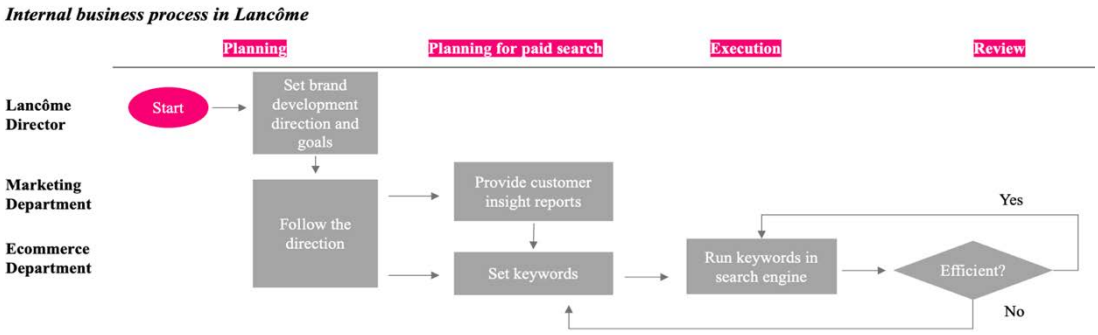


Figure 6 Current paid search process in Lancôme

The current paid search working process from the e-commerce department perspective mainly consists of four stages, involving three participants.

The first stage is the general planning stage. The Lancôme director is responsible for setting all the development strategies and goals for the brand. The marketing department and e-commerce department set their own department's work goals based on the overall macro strategies.

Next is the paid search planning stage. The e-commerce team sets the keywords in paid search channel according to their business needs. The keyword setting is based on previous work experience on the one hand and consumer insight data on the other. The marketing department and the e-commerce department share the same research reports on consumer insights. These consumer insights report mainly come from the marketing department. Their insights and analysis are mainly derived from external market research, marketing activities effect analysis, and internal data analysis from the Lancôme CRM system (Customer Relationship Management system).

The third stage is the paid search execution stage. The e-commerce team sets keywords and the traffic and budget (CPC) corresponding to each keyword in the backstage of the search engines.

The last stage is the review. Before 2020, the e-commerce team reviewed the paid search performance twice a year. The main task of the review is to check the efficiency of keywords subjectively. They defined efficiency as keywords that bring positive income. They replaced inefficient keywords with new keywords, and the remained efficient keywords would continue to be used in the next cycle. Affected by the epidemic in 2020, paid search traffic has increased, so the e-commerce team has increased the frequency of paid search performance review, updated keywords once a quarter.

3.3.3 Managerial problems

In this case study, we completed the research work with the e-commerce department. Through interviews, we found three main pain points that the Lancôme e-commerce team faces:

1. In paid search business, they do not directly contact users, so they know very little about the paid search users.

2. Data support departments, such as Research & Development department or Data science department, are only available in the L'Oréal group scale, but these departments do not serve the internal brand. Therefore, in paid search business, although the team has log data from Google Analytics and Yandex Metrica, no one

responsible for the data analysis work. So basically, all the decisions about the keywords setting in paid search are based on previous work experience, not based on data support.

3. The penetration of paid search business in Russia has increased due to the spread of internet access. The development of paid search channel accelerated under the influence of the 2020 Covid-19 epidemic. However, there is no dedicated team or employee responsible for this business within the e-commerce team. Even they got data in magnitude in 2020, but they still cannot get a better understanding of what the characteristics of their core users are.

Therefore, we mainly focused on the paid search user group that finally completes the purchase in our study. On the one hand, we aim to find the most efficient keywords through data analysis. On the other hand, we aim to have a deep understanding of the groups who complete the purchase to support the business growth of the Lancôme e-commerce team.

3.4 Research methodology and research framework

To reach our goal, which is to build a metric to measure paid search effectiveness and gain a comprehensive understanding of paid search users, the research involved two tasks: keywords efficiency measurement and paid search user analysis. The sub-research questions and methodologies of the tasks are in table 1.

	Task 1	Task 2
	Paid search keyword efficiency measurement	Paid search user analysis
Research questions	1.1 What is the common way to measure paid search efficiency in business? 1.2 Which metric is suitable for Lancôme? 1.3 What is the correlation between efficiency and other indicators?	2.1 How decision-making process works in paid search channel in Lancôme? 2.2 What is the current Lancôme paid search user behavior and rationale behind?
Data source	Paid search channel traffic - Lancôme keywords	Lancôme log data in paid search channel from Google Analytics and Yandex Metrica.
Methodology	Quantitative analysis by Excel Logit regression by python	In-depth interview with Lancôme e-commerce manager. Quantitative analysis by Excel Features extraction by python.

Table 1 The sub-research questions and methodologies of the tasks

Our research started from December 2020, ended in May 2021. The detailed works of each task shown in research timetable below:

		Month					
		Dec	**Jan**	**Feb**	**Mar**	**Apr**	**May**
	Kick-off meetings and research preparations						
	Interview with Lancôme e-commerce manager to get familiar with the whole business context, including paid search working process, decision-making process, etc.						
Task 1: Paid search keyword efficiency measurement	Choose suitable metric for keyword efficiency measurement						
	Data preprocessing: data cleansing and data recombination and efficiency measurement						
	Find the correlation between keyword efficiency and other indicators in the dataset by using logit regression						
	Top efficient keywords test in real business						
Task 2: Paid search user analysis	Data understanding and cleansing						
	Quantitative analysis on Paid search user profile, user preference, and user behavior						
	Choose suitable machine learning algorithm to understand the correlation between search words and purchase behavior.						

Table 2 Research timetable with detailed works

Summary

In this chapter, we first researched the Russian cosmetic market to better understand Lancôme's paid search business context and apply the five forces analysis method to make an overall analysis of the Russian cosmetic market.

The Russian cosmetic market has maintained a compound annual growth rate of 3.9% in the last five years, and different types of cosmetic products have also seen general sales growth. From the Five Forces analysis perspective, buyer power, degree of rivalry, supplier power, and new entrants in the Russian cosmetics market are moderate, while substitutes are relatively low.

Based on our analysis, we think Lancôme should follow the market trend - focus on customer-oriented service. In the era of the digital economy, "consumer-centric" has become the purpose of many companies' marketing campaigns. The marketing method of paid search is the embodiment of consumer-centric thinking. Our research on paid search efficiency can improve Lancôme's marketing efficiency in the cosmetic market.

Then, we organized and recorded the entire paid search business and decision-making process in Lancome through interviews with the Lancome e-commerce manager to have a complete background understanding of our study case.

Finally, we planned task details chose the methodologies for further research work.

4. Paid search analysis

As shown in the Figure 7, in Lancôme marketing funnel, paid search is in the third stage. First, brand advertising and marketing approaches raise users' awareness of Lancôme. Then, the brand promotion activities make users think about Lancôme products and prompt them to search online to get further information about the brand and products.

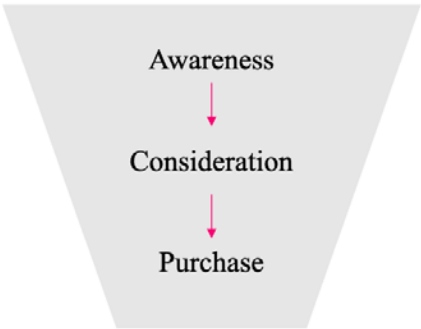


Figure 7 Marketing funnel

For Lancôme, paid search is a form where search engines (Google and Yandex) show Lancôme ads on their search engine results pages. Some paid search users will complete their purchases after searching online.

The effectiveness of paid search advertising depends largely on the keywords. The more the keyword setting is closer to the users' search content and the users' interest preferences, the more users will click on the link, which means that there is a greater possibility that users will purchase through the advertising link, the more effective paid search advertising is.

Therefore, we are mainly focused on two parts in this research:

1. Choose suitable metric for keywords effectiveness measurement.

2. Analyze the group's profile and behavior that finally completes the purchase.

We first defined keywords' effectiveness, established measurement indicators, identified keywords with outstanding performance, and explored the common characteristics of such words. Next, we used the logistic regression algorithm to find the most relevant indicators to paid search efficiency. Finally, we mainly carry out

statistical analysis around important indicators to provide more meaningful insights into the target group's profile and behavior who completed the purchase.

4.1 Part one - paid search keyword efficiency

4.1.1 KPI on marketing campaign

It is a complex problem faced by all companies in digital business to measure the returns brought by marketing campaigns to companies. Companies no longer only regard marketing as a price they must pay for business activities but as an investment. The metric has become a measurable indicator to improve the quality level, quantity, or financial categories. (Kerzner, 2011) The choice of metric should correspond to the business practices and strategy of different companies.

Frantisek Milichovsky and Iveta Simberova have divided marketing metrics into four main parts: single financial output measure (comparing the financial output of marketing campaigns and cost to conclude, such as return on ad spend and return on marketing investment), non-financial measures (outputs through non-financial or qualitative metrics, such as market share, customer satisfaction, or brand equity), input measures (analyzing their resource absorption such marketing budget, or the marketing units' behavior), multiple measures (integrating multiple metrics to analyze marketing campaigns at a macro level).[5]

The choice of KPI can affect the final marketing effect. In this research, in terms of the work practice of Lancôme's e-commerce department, they mainly focus on the financial income brought by the paid search channel. In addition, the information provided by our data is mainly income and expenditure. So in this research, we should choose financial output measure. According to the research of Frantisek Milichovsky and Iveta Simberova, metrics that can measure marketing campaigns, including both revenue and expenditure, are return on ad spend and return on marketing investment. (Frantisek Milichovsky and Iveta Simberova, 2015) This research will mainly focus on and analyze these two metrics and select the appropriate one as the measurement of efficiency analysis.

[5] Milichovský F, Šimberová I. Marketing effectiveness: Metrics for effective strategic marketing[J]. Engineering economics, 2015, 26(2): 211-219.

Return on ad spend (ROAS)

ROAS is a measure used by many retail companies. ROAS is calculated by dividing Gross Revenue by Cost of Ad Campaign. From the most basic point of view, ROAS can intuitively measure the income after marketing activities, which is helpful for the operation of enterprises in terms of capital. However, ROAS emphasizes the income in marketing activities but ignores other marketing investment expenditures. In this case, ROAS ignores the attribution of marketing campaigns. Sometimes the income of a company is not only brought by marketing campaigns, and ROAS cannot deal with this problem. In this way, it may be misleading. At the same time, ROAS has another problem: it cannot measure some "non-ad" marketing campaigns. The increase in other costs caused by marketing campaigns, such as product storage costs, transportation costs, and gift costs, cannot be reflected in ROAS. Because of this problem, companies have found that they need one that can evaluate the efficiency of all marketing activities, not just advertising activities.

Return on marketing investment (ROMI)

Many companies currently use ROMI to measure marketing effects. ROMI is calculated by dividing the revenue generated by marketing activities after deducting marketing costs and dividing by the costs related to the marketing campaigns. ROMI does not pay attention to the expenses related to the business but only pays attention to the expenses related to marketing work. ROMI can be considered as the costs of all marketing activities, not just advertising expenditures. It may include content creation, merchandise discounts, etc.

Selection of KPI

The biggest problem with ROAS is that it lets the ad network take credit for all sales revenue while ignoring any inconvenient costs and calculations that tell a different story. ROMI is a more realistic indicator that can tell the manager the true effect of marketing activities and marketing return on investment.

In the actual work of Lancôme, ROMI is generally used as an indicator to measure the effect of marketing activities. In this article, the marketing effect of Lancôme’s paid search should also be combined with Lancôme’s work practice.

Besides, when ROMI measures the effect of keywords in paid search, using ROMI as a measurement indicator can intuitively understand the marketing effect of each keyword. Based on these two reasons, we finally chose ROMI as our metric to measure the efficiency of keywords. And, using this as a benchmark, we define keywords with a ROMI value greater than 0 (that can bring pure profits) as "efficient" keywords, and analyze their characteristics.

The specific steps for calculating ROMI are as follows:

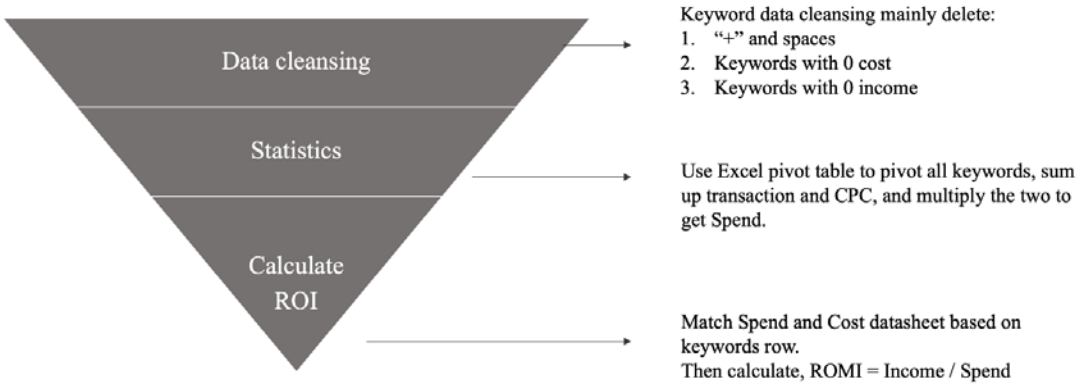


Figure 8 Steps for calculating ROMI of keywords from paid search channel

Besides, comparing ROMI of keywords in 2019 and 2020, and ROMI of keywords in Yandex and Google, we came to the following observations:

- In 2019, there are 146248 records and 1449 validate keywords in total, of which 71 were efficient keywords; in 2020, there are 117252 records and 2128 validate keywords in total, of which 185 were efficient keywords.

	Spend		Income			Total ROMI
The number of	Records	Keywords	Records	Keywords	Keywords (brought income)	
2019	74986	1810	146248	1449	71	0.61
2020	54039	2090	117252	2128	185	3.57

Table 3 Result of data preview

4.1.2 Regression analysis

Regression analysis refers to a statistical analysis method to find the quantitative relationship between two or more variables. The use of regression models can reveal the correlation between dependent variables and independent variables and can show the degree of influence of multiple independent variables on a dependent variable. Regression analysis also allows us to compare the impact of variables measured on different scales, such as the impact of price changes and the number of promotional activities. The advantage of this is that it can help market researchers, data analysts, etc., evaluate and select the best set of variables for building predictive models.

In this case, by analyzing the characteristics of the keywords in the paid search channel, such as the number of clicks, income, transaction rate, etc., to find out the main characteristics of the effective keywords, and then use this as a basis for Lancôme's paid search The keyword setting strategy was improved.

Regression analysis approaches

There are four main regression techniques that can be used to make predictions. These regression models are mainly related to three indicators: the number of

independent variables, the types of independent variables, and the types of regression lines. Generally, there are several common regression models:

- Linear regression

Linear regression is a regression analysis that uses the least square function called linear regression equation to model the relationship between one or more independent variables and dependent variables. This function is a linear combination of one or more model parameters called regression coefficients. The situation with only one independent variable is called simple regression, and the situation with more than one independent variable is called multiple regression.

In linear regression, the data uses a linear prediction function to build a regression model, and unknown model parameters are also estimated through the data. These models are called linear models. Linear regression is the first type of regression analysis that is widely used in practical applications. This is because a model that linearly depends on its unknown parameters is easier to fit than a model that non-linearly depends on its unknown parameters, and the statistical properties of the resulting estimates are easier to determine.

In linear regression analysis, independent variables and dependent variables must satisfy a linear relationship. There may be multicollinearity, autocorrelation, and heteroscedasticity in multiple regression analysis. Linear regression is very sensitive to outliers, which will seriously affect the final prediction results.

The following is an example of a linear regression model:

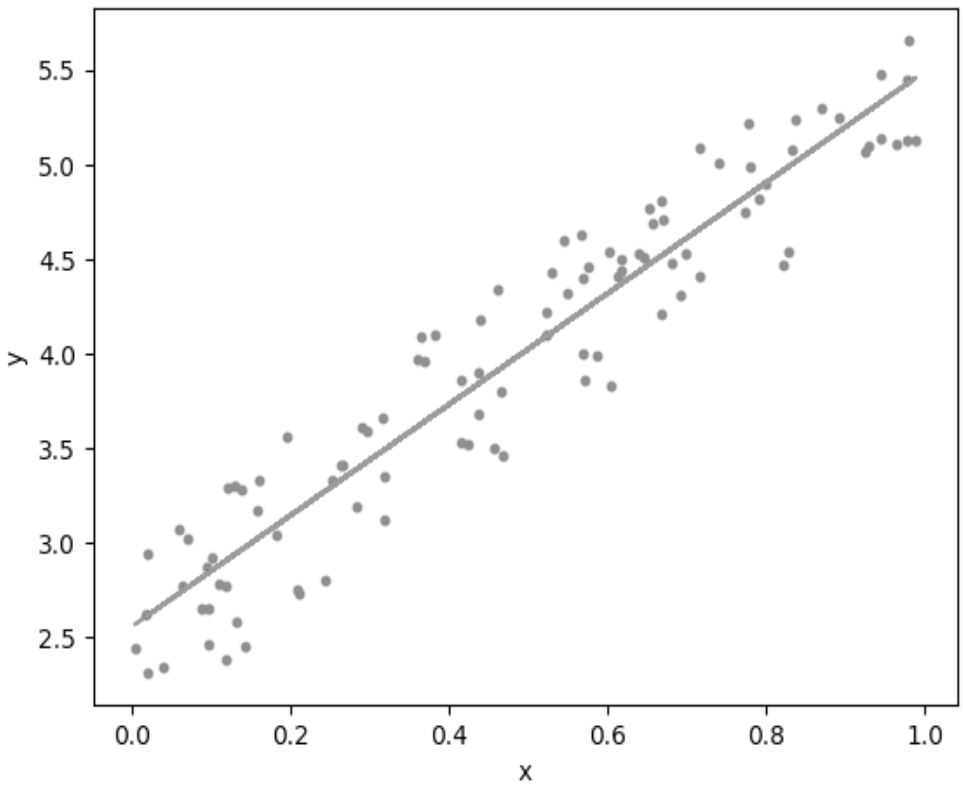


Figure 9 Example of linear regression model. Source: https://towardsdatascience.com/linear-regression-using-python-b136c91bf0a2

- Logistic regression

Logistic regression is a basic classification technique and one of the common methods of data mining and machine learning. It is a supervised learning method. It analyzes and predicts generalized linear regression of discrete dependent variables based on one or more continuous or discrete independent variables. Binary logistic regression is one of the most commonly used forms, and its dependent variable contains only two categorical values.

Logistic regression and multiple linear regression analysis have many similarities; the difference lies in their different dependent variables. Logistic regression is used to calculate the probability of success or failure of an event. When the dependent variable is binary (0/1, True/False, Yes/No), logistic regression should be used. The result variable value obtained by logistic regression fitting is a probability value (0-100%). According to the size of the probability value and the classification threshold determined according to the business scenario, it is mapped to the

classification value of the final predicted target variable, such as: the probability value is greater than or equal to 50%, the classification value of the mapped target variable is 1; the probability value is less than 50%, the classification value of the mapped target variable is 0. The value range of the dependent variable is [0,1]. Because of the value range of the variable, it is necessary to choose a suitable activation function to map the output to [0,1], and the Logit function meets the requirements. In addition, logistic regression does not require a linear relationship between the dependent variable and the independent variable. It can handle multiple types of relationships because it performs a nonlinear log transformation on the predicted output.

The following is an example of logistic regression model:

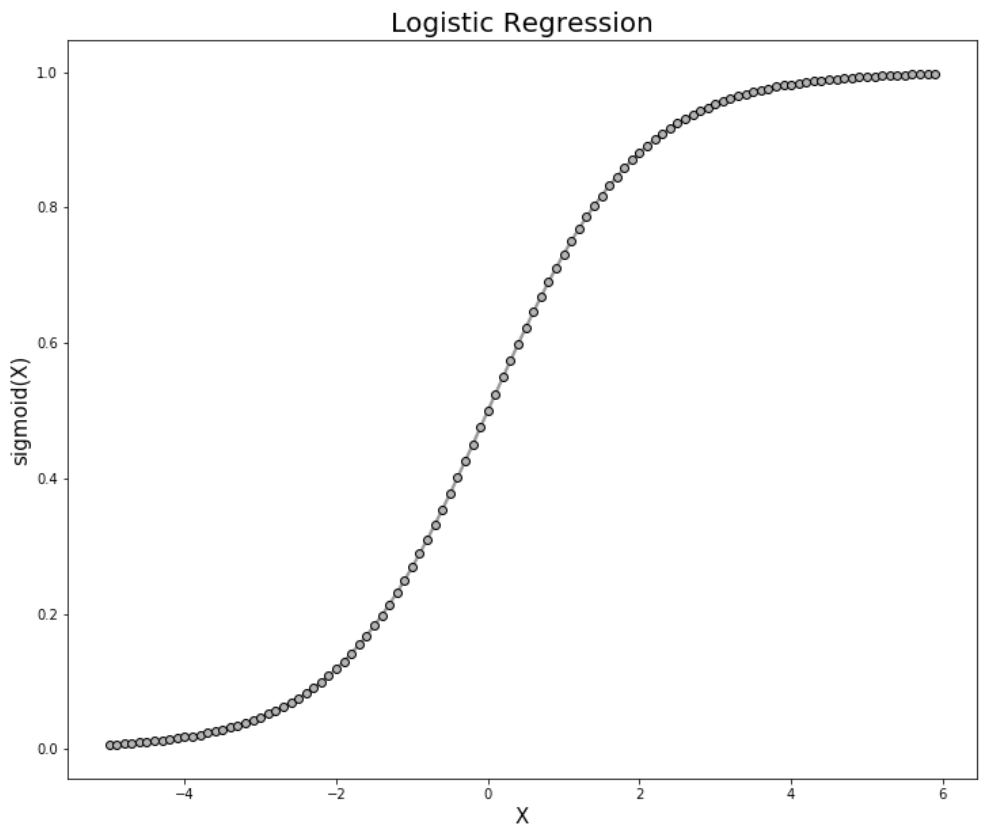


Figure 10 Example of logistic regression model. Source: https://realpython.com/logistic-regression-python/#logistic-regression-in-python-with-scikit-learn-example-1

- Polynomial regression

Polynomial regression is a form of regression analysis in which the relationship between the independent variable x and the dependent variable y is modeled as a polynomial of degree n with respect to x.

The polynomial regression model is a type of linear regression model, where the regression function is linear with respect to the regression coefficients. Since any function can be approximated by a polynomial, polynomial regression has a wide range of applications. Linear regression studies the regression problem between a dependent variable and an independent variable, but in actual situations, there are often more than one independent variable that affect the dependent variable. Therefore, it is necessary to perform a regression analysis between a dependent variable and multiple independent variables. Namely multiple regression analysis. The study of polynomial regression analysis method between a dependent variable and one or more independent variables is called Polynomial Regression. If there is only one independent variable, it is called univariate polynomial regression; if there are multiple independent variables, it is called multiple polynomial regression. In univariate regression analysis, if the relationship between the dependent variable y and the independent variable x is non-linear, but no appropriate function curve can be found to fit, then univariate polynomial regression can be used.

The following is an example of a polynomial regression model:

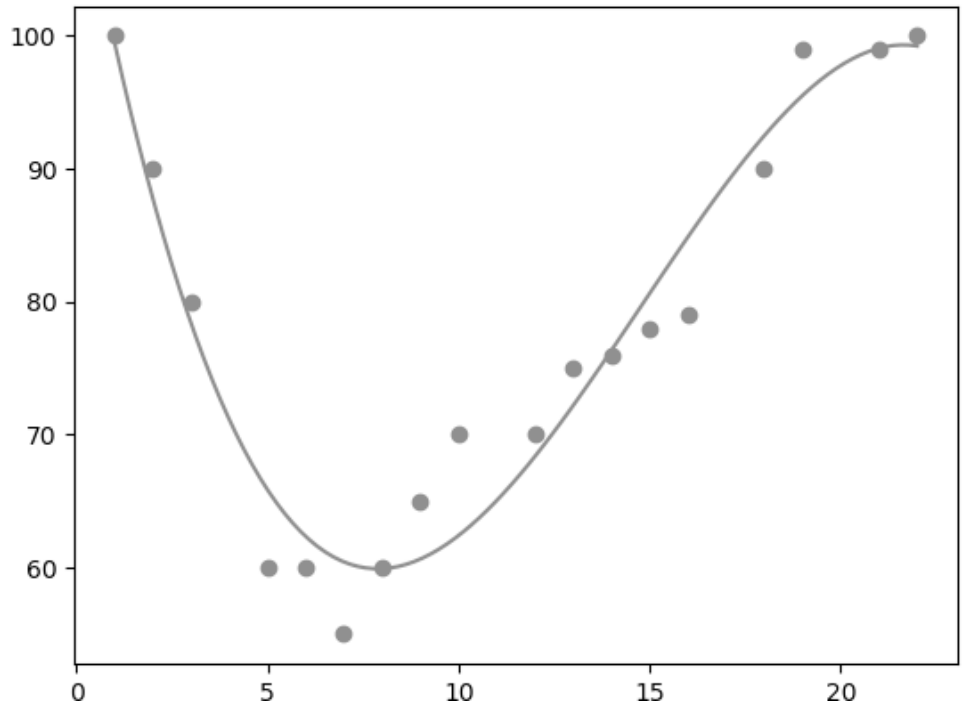


Figure 11 Example of polynomial regression model. Source: https://www.w3schools.com/python/python_ml_polynomial_regression.asp

- Stepwise regression

The basic idea of stepwise regression analysis method is to automatically select the most important variable from a large number of variables to establish the prediction or interpretation model of regression analysis. When dealing with multiple independent variables, stepwise regression is a better method. In this technique, the selection of independent variables is accomplished by means of automatic process, which does not involve manual intervention. Stepwise regression is to observe statistical values, such as R-square, t-stats, AIC indicators to identify important variables. Firstly, the explained variables are used to make simple regression for each explanatory variable under consideration, and then based on the regression equation corresponding to the explanatory variable that contributes the most to the explained variable, the remaining explanatory variables are gradually introduced. After stepwise regression, the explanatory variables remained in the model are not only important, but also have no serious multicollinearity.

The purpose of this modeling technique is to obtain the maximum predictive ability by using the least independent variables. It is also one of the methods to deal with high-dimensional data sets.

The following is a process of Stepwise regression model:

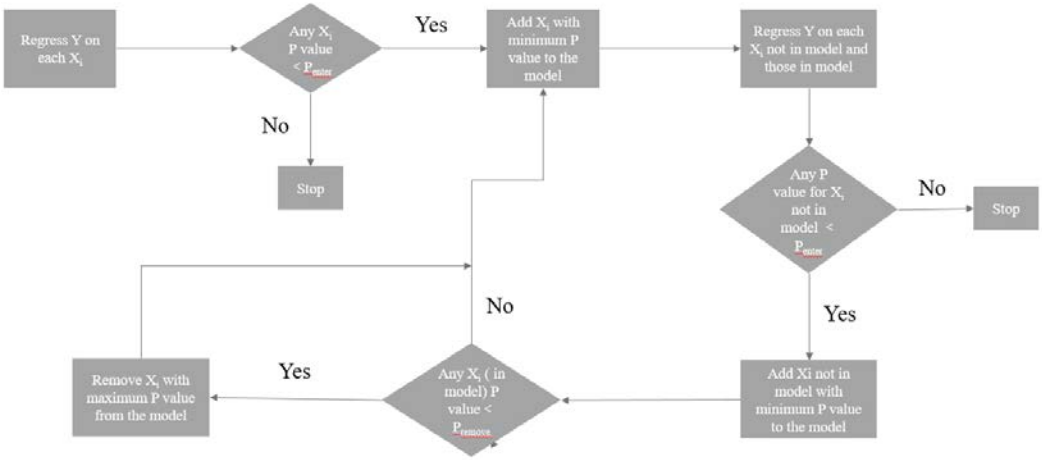


Figure 12 Process of stepwise regression model. Source: https://www.spcforexcel.com/knowledge/root-cause-analysis/stepwise-regression#regression_review

The choice of regression model

In the case of this article, the efficiency of a keyword in paid search is defined as whether its ROMI is greater than 0. Keywords with ROMI greater than 0 are efficient, while keywords with ROMI less than 0 are inefficient. The effectiveness as a dependent variable is a binary dependent variable, which meets the requirements of the logistic regression model and is suitable for logistic regression. Therefore, we will use logistic regression model to analyze the efficiency of keywords in paid search.

Logistic regression analysis

- Data preprocessing

In the data, the independent variables related to the keywords in Lancôme's paid search channel include the number of users, the number of new users, sessions, bounce rate of each click, transaction number, click number and include. Our data includes the search and purchase records of Lancôme users in Google and Yandex from 2019 to 2020.

First, we delete the invalid keywords, including the simple digital ID record. Since then, we have integrated the same keywords and counted all the data of the same keyword in 2019 and 2020. Since then, we have merged the obvious typo, such as "Lankam", "Lonkome", etc. Finally, we screened out 179 keywords in the paid search. In addition, we calculate the bounce rate and transaction rate of keywords in the whole time period as independent variables.

- Model results and analysis

The coefficient of each independent variable is as follows:

Users Number	-2.735e-02
New Users	1.424e-02
Sessions	-2.362e-02
Bounce Rate	0.337e-02
Transaction	0.026e-02
Transaction Rate	0.166e-02
Income	0.114e-02
Click Number	0.114e02

Table 4 The coefficient in logistic regression model

The prediction result of the logistic regression model：

Actual 0s	**42**	**3**
Actual 1s	**9**	**125**
	Predicted 0s	**Predicted 1s**

Table 5 The prediction result of the logistic regression model

The quality of the logistic regression model：

	Precision	**Recall**	**F1-score**	**Support**
0.0	0.82	0.93	0.87	45
1.0	0.98	0.93	0.95	134
Accuracy			0.93	179
Macro avg	0.90	0.93	0.91	179
Weighted avg	0.94	0.93	0.93	179

Table 6 The quality of the logistic regression model

The prediction accuracy of the model has reached 90%, which is a more reliable logistic regression model.

The coefficient of each independent variable is analyzed. The independent variables positively related to efficiency include new users, bounce rate, transaction, transaction rate, income and click number.

The independent variables negatively related to efficiency include user numbers, sessions.

Among the positively correlated independent variables, new users, bounce rate and transaction rate are more significant. We think that customers who search Lancôme related product keywords for the first time have higher purchase intention and can bring higher transaction rate and revenue. On this basis, Lancôme's paid search keyword settings and keyword budget can be more inclined to new users who search for the first time.

4.2 Part two - paid search user behaviour understanding

The depth of behavior analysis has represented the refinement of data analysis needs of some fast-growing companies in the past years. Simultaneously, it has also spawned a new generation of data analysis tools such as Google Analytics and Yandex Metrica. Common user analysis indicators include page view, unique visitor, bounce rate, session duration, etc. In this research, these indicators are all statistical indicators, reflecting the overall situation of the paid search performance. At the same time, we noticed that the value of data reflects the status quo of paid search and, more importantly, reflects user usage scenarios and guides business growth based on user behavior and the time and frequency of the behavior.

When we start this research, we found that the Lancôme e-commerce team did not directly reach customers. Their insights on target consumers basically came from the research report of the marketing department and their agency. Most of the time, the decision-making on paid search keywords setting in the e-commerce team is based on former experience. For example, if the keyword "Lancôme Hypnôse" brings outstanding revenue this quarter, they would continue to set this keyword in the next quarter. The problem is that there are thousands of keywords in paid search settings, Lancôme e-commerce team cannot efficiently evaluate all the keywords monthly.

Describing users in highly refined user characteristics can help Lancôme's e-commerce team understand users more easily, reduce the complexity of keyword settings, optimize advertising content, accurately target users, and increase advertising ROI. At the same time, it is also convenient for managers to set key business KPIs more flexibly. Therefore, we are going to build a detailed and complete user portrait by supplementing user behavior data.

4.2.1 Data introduction

- Data source

Lancôme internal log data. Indicators including members, new members, sessions, bounce rate, page/session, average session duration, transactions, income, and transaction rate. Data recorded by gender, age, location, devices, browsers, and product performance in Google Analytics and Yandex Direct from 2019 to 2020 and presented by 6 Excel files. Each Excel file (Lancome_age.xlsx and Lancome_gender.xlsx as examples) looks like Figure 13 and Figure 14.

```
In [12]: df_age = pd.read_excel('Lancome_age.xlsx', 'GA2019-2020', index_col=None, na_values=['NA'])

In [13]: df_age.head(n=5)

Out[13]:
```

	Date	Source or channel	Age	Member	New member	Sessions	Bounce rate	Bounce	Pages / session	Average session duration	Transactions	Income	AOI	Transactions/Members	Trar
0	20200110.0	yandex / paid_search	55-64	56	54	58	0.775862	45.0	1.431034	40.879310	0	0.0	0.0	0.0	
1	20200110.0	yandex / paid_search	35-44	49	47	50	0.760000	38.0	1.400000	25.500000	0	0.0	0.0	0.0	
2	20200110.0	yandex / paid_search	25-34	34	34	36	0.694444	25.0	2.500000	57.777778	0	0.0	0.0	0.0	
3	20200110.0	yandex / paid_search	65+	31	31	33	0.757576	25.0	1.515152	18.969697	0	0.0	0.0	0.0	
4	20200110.0	yandex / paid_search	45-54	25	23	26	0.923077	24.0	1.115385	11.846154	0	0.0	0.0	0.0	

```
In [14]: df_age.shape

Out[14]: (5759, 16)

In [15]: df_age.columns

Out[15]: Index(['Date', 'Source or channel', 'Age', 'Member', 'New member', 'Sessions',
         'Bounce rate', 'Bounce', 'Pages / session', 'Average session duration',
         'Transactions', 'Income', 'AOI', 'Transactions/Members',
         'Transaction rate', 'Conversion rate goal'],
        dtype='object')
```

Figure 13 Dataset example of "Lancome_age"

```
In [2]: # reading the data
        df_gender = pd.read_excel('Lancome_gender.xlsx', sheet_name='GA2019-2020', header=0, index_col=0, na_values=['NA'])

In [3]: df_gender.head(n=5)
Out[3]:
```

Date	Source or channel	Gender	Member	New member	Sessions	Bounce rate	Pages/session	Average session duration	Transactions	Income	AOI	Transaction/Members	Transa
20190101	google / cpc	female	310	216	383	0.503916	2.590078	92.467363	0	0.0	0.0	0.000000	0.00
20190101	google / cpc	male	54	36	69	0.695652	2.579710	64.724638	0	0.0	0.0	0.000000	0.00
20190101	yandex / paid_search	female	280	198	328	0.664634	2.192073	146.006098	0	0.0	0.0	0.000000	0.00
20190101	yandex / paid_search	male	49	33	58	0.637931	2.655172	163.327586	0	0.0	0.0	0.000000	0.00
20190102	google / cpc	female	286	187	344	0.488372	2.784884	127.377907	1	4786.0	4786.0	0.003497	0.00

```
In [5]: df_gender.shape
Out[5]: (2176, 13)

In [4]: df_gender.columns
Out[4]: Index(['Source or channel', 'Gender', 'Member', 'New member', 'Sessions',
               'Bounce rate', 'Pages/session', 'Average session duration',
               'Transactions', 'Income', 'AOI', 'Transaction/Members',
               'Transaction rate'],
              dtype='object')
```

Figure 14 Dataset example of "Lancome gender"

- Data introduction

GA: Google Analytics, a website traffic statistics and analytics tool provided by Google.

Google: Google search engine.

Yandex: Yandex search engine.

Users: People who have initiated at least one session in a day.

Sessions: A period of time a user spend on Lancôme website.

Bounce rate: Bounce rate is single-page sessions divided by all sessions.

Page/session: Average number of pages viewed in a session.

Average session duration: Total duration of sessions in seconds divided by the number of sessions (counted by day by different dimensions such as age, gender, location, etc.). This indicator is unimportant in this paid search advertising research since no evidence shows the time period people spend on a website is related to the final purchase behavior.

Transactions: Number of purchases on Lancôme website in a day.

Transaction rate: Total number of transactions divided by total number of sessions in a day.

Missing data: The third-quarter data in Google Analytics in 2020 is unavailable, and the fourth-quarter data is incomplete. Therefore, in the user behavior analysis part, the data of 2019 is often used.

4.2.2 Methodology

In this paid search user behaviour understanding part, we mainly use statistical analysis methods and use Tableau as a data visualization tool to show data analysis content. According to the characteristics of the datasets, the data analysis concludes with 3 dimensions.

- User profile: the demographical analysis contains basic information about gender, age, location.
- User preference: mobile device model, browser, operating system, search preference, product purchase preference.
- User behavior: data analysis based on session duration, number of clicks, bounce rate, conversion, and transaction.

When studying search preference, we used FreqDist in Natural Language Toolkit (NLTK) to analyze the frequency of keywords. FreqDist is a function that can calculate word frequency and show the frequency distribution. After FreqDist accepts the parameter words, it will count the frequency of each word in words and return a dictionary, where value is the frequency of word appearing in words.

When analyzing product purchase preference, we used TfidfVectorizer in Scikit-learn to vectorize the text set. Then we used the CatBoost classifier to explore the feature importance for search words. CatBoost is an open-sourced machine learning library provided by Russian search engine company Yandex in 2017. CatBoost, XGBoost, and LightGBM are collectively known as the three mainstream algorithms of Gradient boosting, which are all improved implementations under the framework of the gradient boosting algorithm.

4.2.3 Result

- **User profile:**

Because of the higher cost of advertising cost on Google than Yandex, Lancôme set more traffic on Yandex. So, the sample size of Yandex is larger than Google, which is almost doubled. Lancôme paid search user distribution by age in Figure 15 shows. 25-44 are the main paid search users, and Google's main user group is younger than Yandex.

The age distribution not only depends on the characteristics of the product lines of Lancôme but also depends on the loyalty of different age stages. In cosmetic industry, young people are more likely influenced by social media influencers and new coming fancy products. They prefer to try new things rather than continue to buy luxury products of high quality from the same brand.

Now for Lancôme brand, they want to reach more younger generations. Considering that the loyalty of young users to brands and products is not as high as that of middle-aged users, brands could consider launching diversified products suitable for young people.

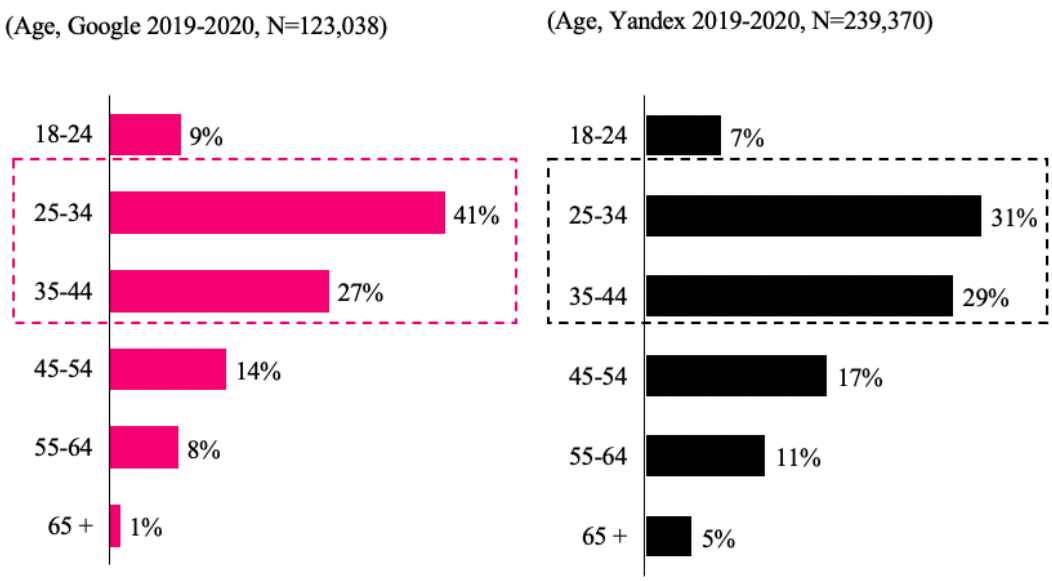


Figure 15 Data statistics of the age of Lancôme customers

The gender distribution of Yandex and Google is not very different, so we did the statistics according to the overall data. Women are the main paid search group, accounting for 84% of the total paid search users.

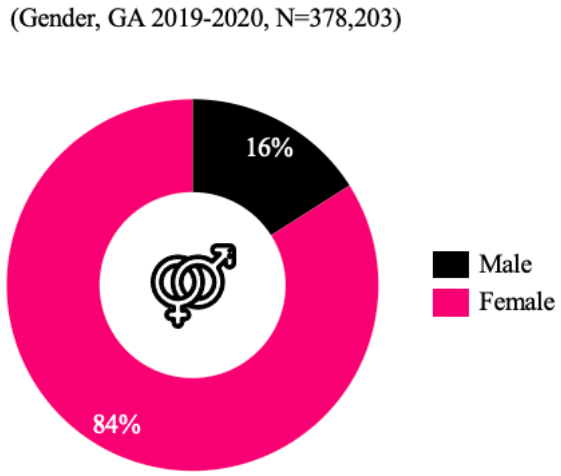


Figure 16 Data statistics of the gender of Lancôme customers

Regarding regional distribution, the top five regions in terms of the number of paid search users are the Moscow region (including Moscow and Moscow Oblast), St. Petersburg, Krasnodar Territory, Sverdlovsk Oblast, and the Republic of Bashkortostan. The total population of these five regions accounted for more than 50% of the overall sample, and each remaining region accounted for no more than 2.5%.

It is worth mentioning that the number of Yandex users in the Moscow area is as high as 37%, greater than Google by 7%, which means that Lancôme advertising in Yandex can focus more on the Moscow area.

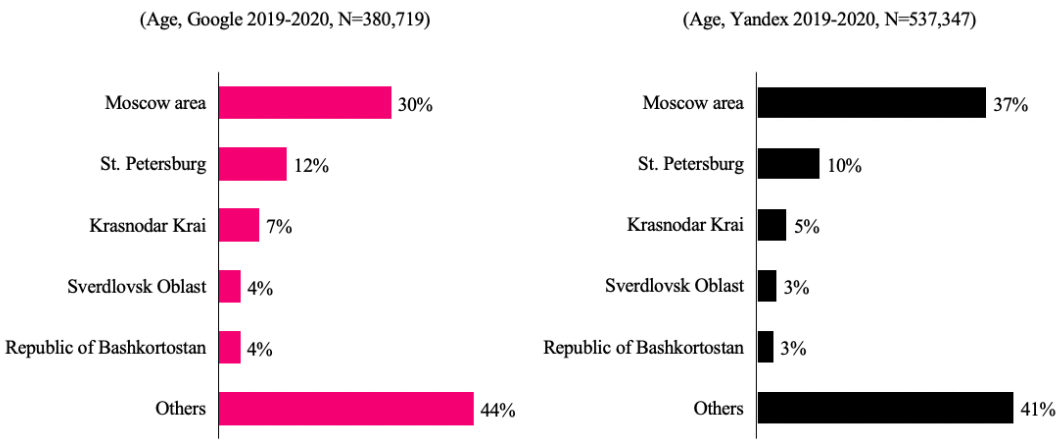


Figure 17 Data statistics of the gender of Lancôme customers

- **User preference:**

Mobile devices. According to mobile devices tracking data, 90% paid search users are using Apple, Samsung, Huawei, and Xiaomi. In descending order of income, top10 brands consist of 50% total income, they are LG, Samsung, Meizu, Asus, Xiaomi, Nokia, Lenovo, Apple, ZTE, Huawei. Top15 brand consist of ~97% total users. Four main brands (Apple, Samsung, Huawei, Xiaomi) consist of 90% proportion of total users.

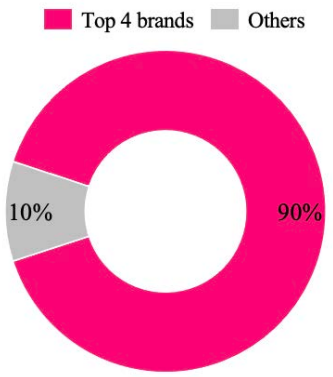


Figure 18 User share by mobile devices

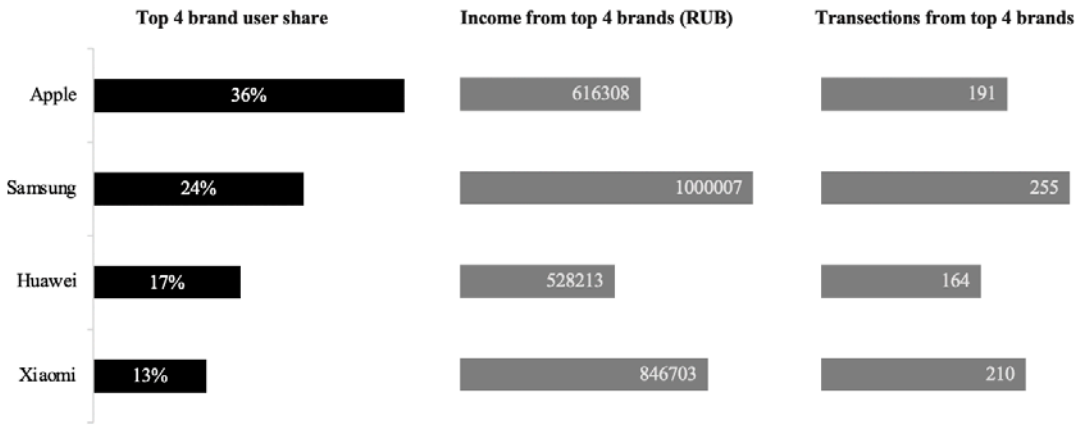


Figure 19 Key user distribution and performance by mobile devices

Browsers. Chrome is the most popular browser, accounting for 50% of the total paid search users, followed by Safari, Yandex Browser, 22%, and 14%, respectively. Comparing within the three browsers, the number of transactions by Chrome users has the highest with 3529 transaction records and the lowest average unit income with 3650 Rubles per transaction. Yandex Browser has the least transaction but the highest per customer price.

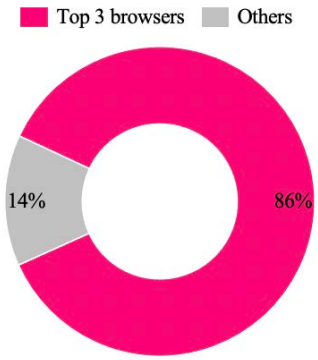


Figure 20 User share by browsers

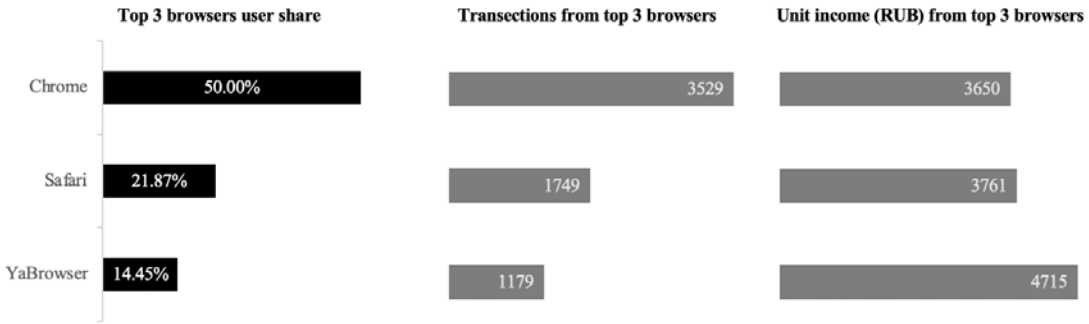


Figure 21 Key user distribution and performance by browsers

Operating systems. Android, Windows, and iOS are the three main operating systems, and users account for 97.61%. Among the three operating systems, Android has the highest average bounce rate (46.2%) and the lowest income (7.6 million RUB), versus Windows, has the lowest average bounce rate (35.09%), and the highest income (10.7 million RUB).

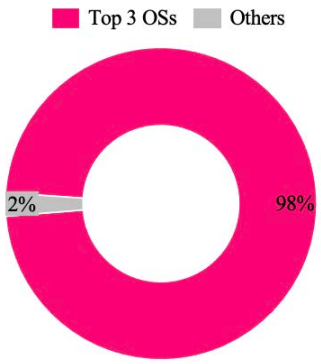


Figure 22 User share by operating systems

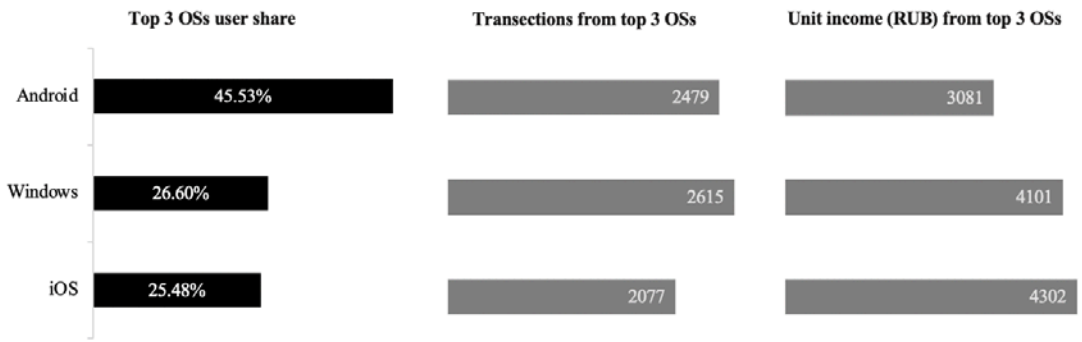


Figure 23 Key user distribution and performance by operating systems

Product consumptions. The dataset of product performance records daily product income, unique shopping, average price, and other information with product SKU as the dimension. We labeled them based on dates and product characteristics for four seasons and four major product categories (skincare, makeup, fragrance, and suits) to understand the relationship between seasons and product category sales. In terms of product sales, skincare products account for the largest share of Lancôme's total revenue. The additional sales characteristics of the product are connected in series. Autumn is the most prosperous season for product sales, and gift sets have the highest average transaction volume in autumn.

Obviously, the consumption changes brought about by the epidemic in 2020. In 2020, people paid more attention to searching for information and shopping online and find that the overall revenue ranking of search channels increased by 90% in 2019. Behaviors such as isolation and wearing masks have affected consumers' demand for cosmetics. According to some people's attention to skincare products, the revenue of the cosmetics category decreased by 6% compared with 2019, but the skincare category increased by 6%.

Search preference. By using the keywords traffic dataset, we have conducted frequency statistics for all queries searched by users in Google from 2019 to 2020. The results are shown in Figure 24. The user's favorite search terms are mainly about the Lancôme brand itself (official website, online store, Lancôme) and several famous Lancôme products. The content searched by users has a strong positive relationship with the marketing efforts of certain products. Therefore, the more famous the product, the higher the frequency of online search for related information. In addition, the online search for Lancôme fragrance accounted for a large proportion of queries, indicating that users like to get information about fragrance through online searches.

```
freq_dist = FreqDist(df['text_proc'].values)
display(freq_dist)
plt.figure(figsize=(16, 4))
plt.title('50 most frequent search queries in Поисковый запрос column, Google')
freq_dist.plot(50, cumulative=False)
plt.show()
```

```
FreqDist({'ланком духи': 2384, 'lancome la vie est belle': 1622, 'la vie est belle': 1449, 'lancome': 1439, 'ланком':
1285, 'духи ланком': 1157, 'ланкоме': 1095, 'тушь ланком': 863, 'kfyrjv': 797, 'духи lancome': 758, ...})
```

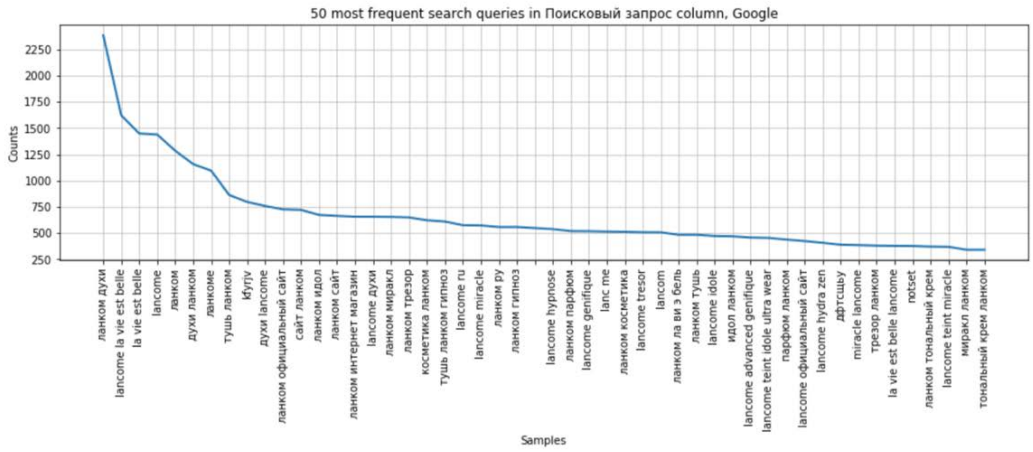


Figure 24 50 most frequent search queries relative to Lancome in Google

Relationship between search words and purchase. Furthermore, we extracted feature importance of keywords, found the features of keywords most relevant to income. As the result shows in Figure 25, when the keywords displayed are more result-oriented words such as "buy," "price," "store," etc., the easier it is for the target group to complete purchase behavior. In addition, users prefer to make purchases after searching for product-related reviews on the Internet. Products purchased online through paid search channels are mainly about fragrance and eye products.

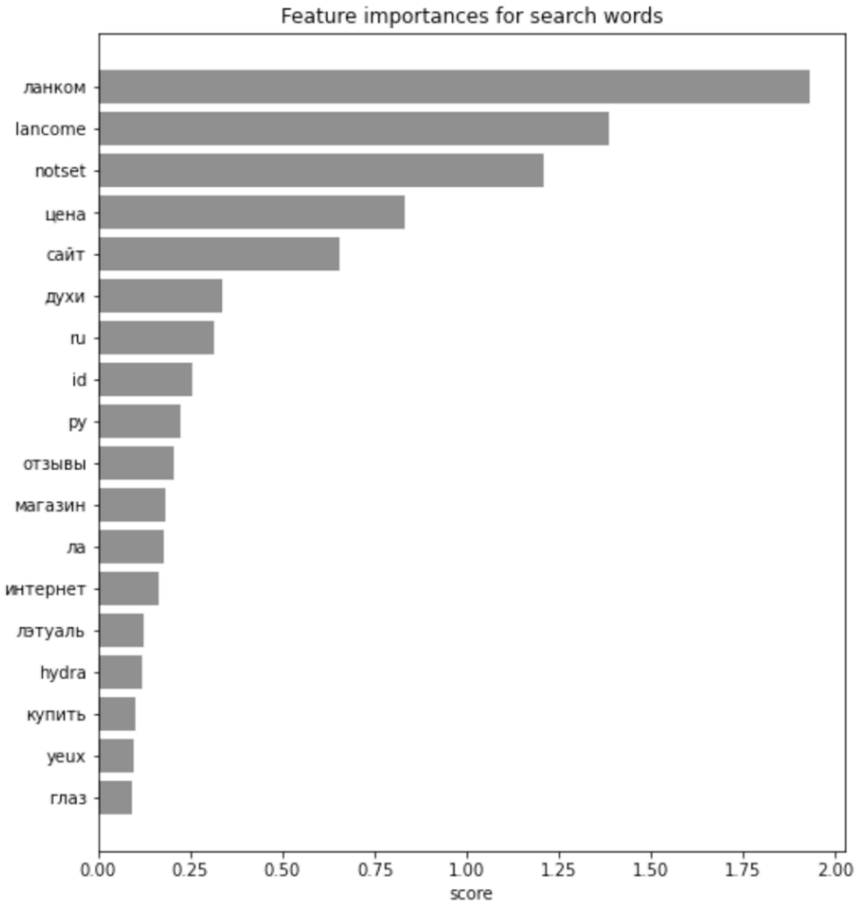


Figure 25 Analysis of feature importance for search words

- **User behavior**

Like the Lancôme marketing funnel, we analyze user behavior based on two dimensions (age and gender) with interactions, transactions, and final purchases.

1. **Age dimension:**

Interaction. Overall, 25-44 groups click on paid search ads the most. The activeness can be seen from the Interaction trends line chart that users are the most active in the third quarter, reaching the peak of search-click, followed by the fourth quarter and the first quarter, and the second quarter is the low season the year. In 2019, the changing trend of each age group was basically the same. Compared with 2020, the changing trends of all age groups have changed in the third and fourth quarters. The number of user interactions in the 18-24 age group has increased significantly in the fourth quarter of 2020. The number of user interactions in the 45-54 age group is lower than that of users in the fourth quarter's 18-24 age group. In addition, 65 above years

old age group interactions continuously grew in 2020, which shown that under the Covid-19 epidemic influence, elderly people began to buy cosmetic products online.

Insights 1:

In the fourth quarter, the Lancôme team could add keywords for products that 18-24 age group is interested in and increase the traffic and the budget.

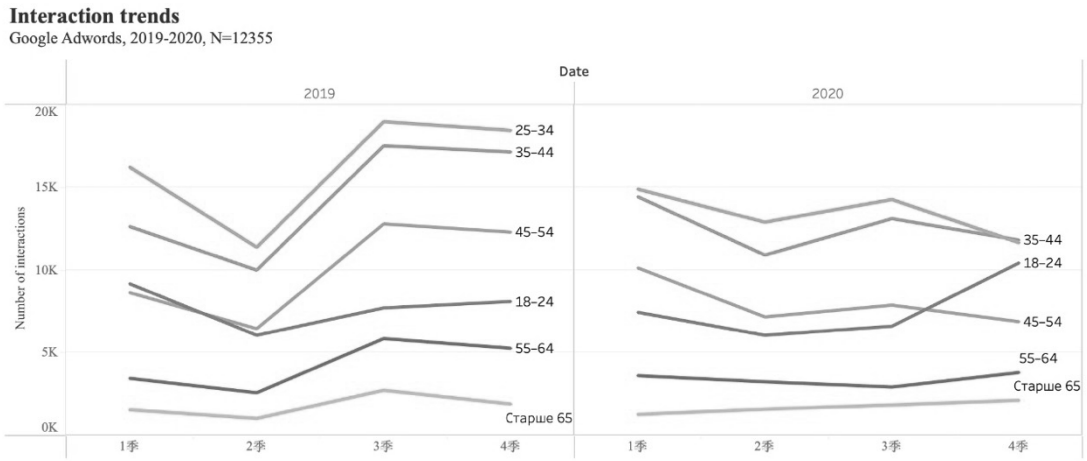


Figure 26 Interaction trends in Google by age

Transaction. Affected by brand strategy, holidays, and seasons, March, July, and November are the three consumption peaks. People are willing to buy products in promotion seasons and before holidays.

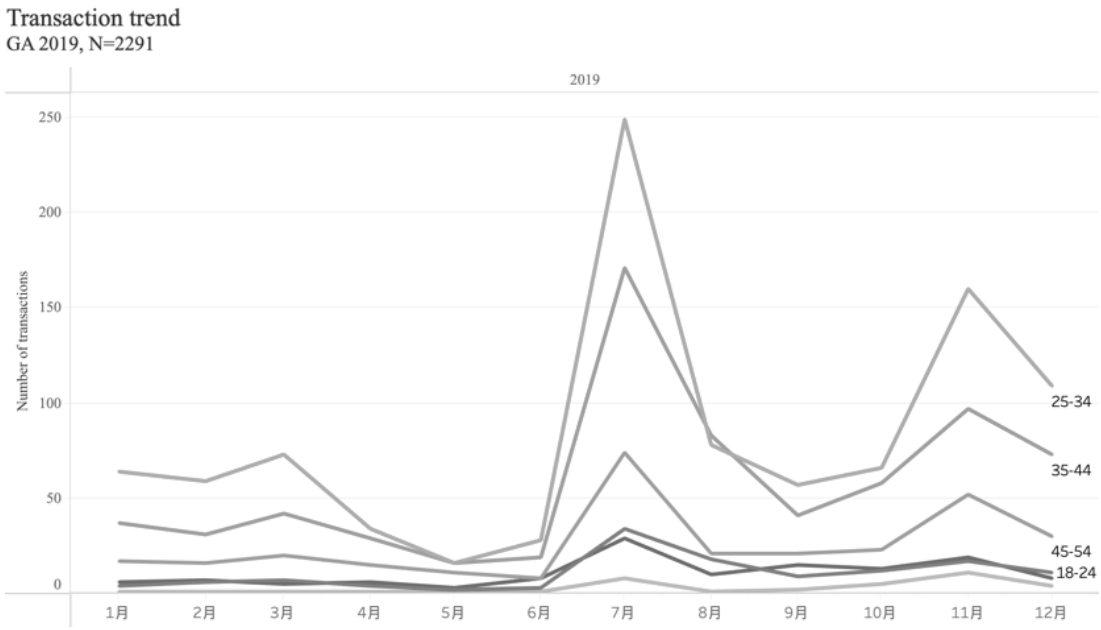


Figure 27 Transaction trend in Google by age

There was a significant increase in the transaction rate in 2020. The living behavior of young people and older people was affected the most by the lockdown policy during the Covid-19 pandemic. Among them, the 18-24 and 65+ age groups have the largest increase.

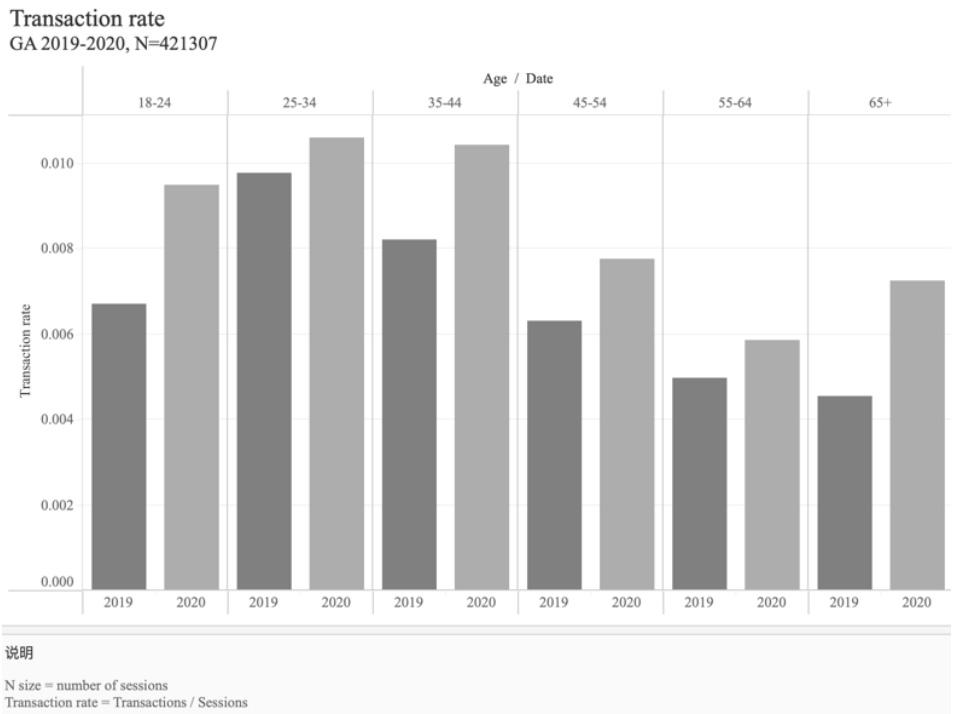


Figure 28 Transaction rate by age

Insights 2:

Changing keywords corresponding with the peak consumption in March, July, and November.

Insights 3:

Under the influence of the epidemic, people have become more purposeful in using paid search. Significant growth occurred in the 18-24 and 65 above age groups, who have shifted more from offline shopping to online. Although the size of these two groups is not large, these two groups should be paid attention to when conducting paid advertising in a particular season.

Income. In the first, second, and fourth quarters, the total income kept in decreasing order of 25-34, 35-44, 45-54, 55-64, 18-24, 65+. 35-44 age group had the highest consumption in the third quarter (Figure 29).

Although users in the 25-34 and 35-44 age groups have the highest total consumption amount, the average consumption per order is in the middle and lower stages from the perspective of average order income. The reasons behind this are that firstly the price of products suitable for the 25-44 age group, whether it is makeup or skincare products, is at a medium level; secondly, Lancôme provides limited product options for the 25-44 aged group.

In addition, the AOI level of the 18-24 age group varies significantly with the seasons. The AOI in the first quarter was the highest among the overall AOI and the lowest in the fourth quarter.

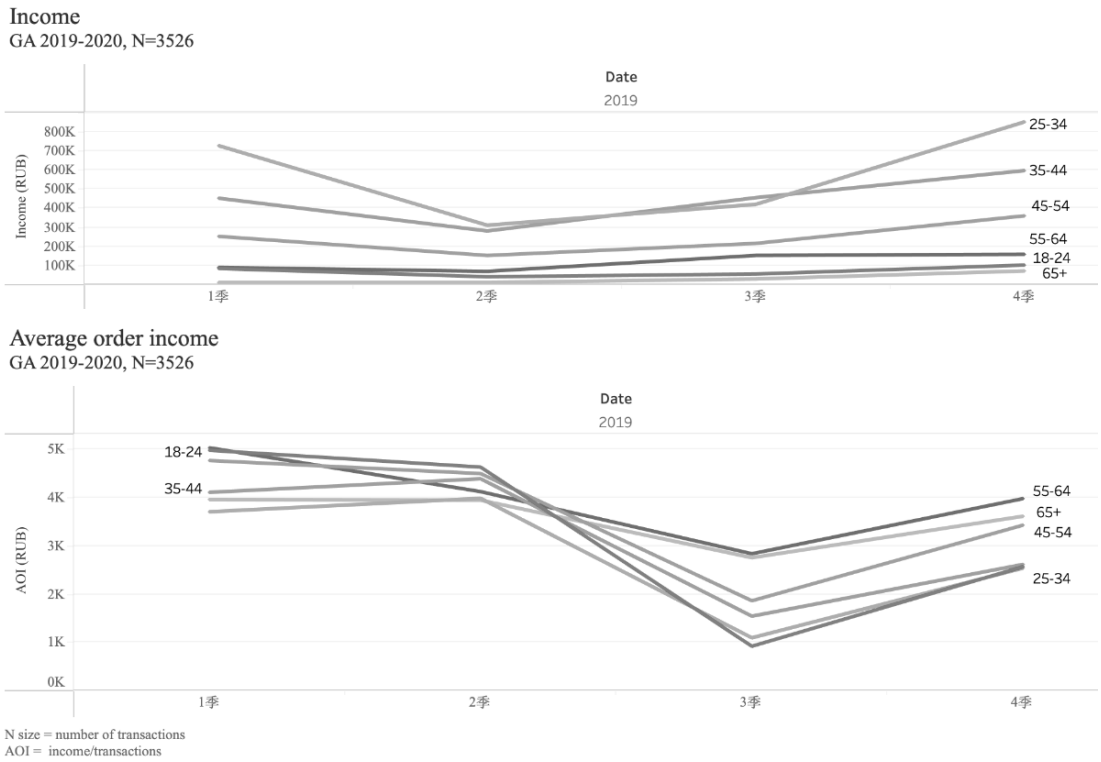


Figure 29 Income distribution and average order income by age

Average order income, the consumption level in 2019 increased with age (Figure 30). In 2020, the overall AOI was higher. The AOI of the 18-24 young group has increased significantly by 57% year-on-year, indicating that young people tend to use paid search under the influence of the epidemic to understand product information and complete purchases.

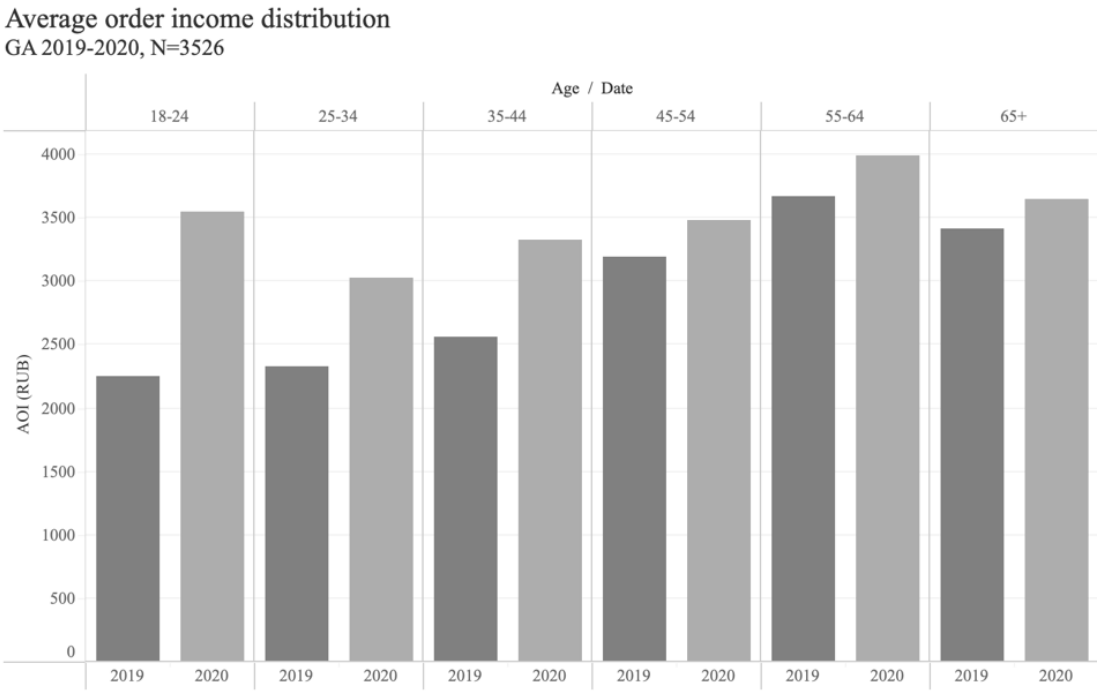


Figure 30 Average order income distribution by age

Insight 4:

In the third quarter, increase the keywords of products for the 35-44-year-old audience, and increase traffic and budget. In the third and fourth quarters, appropriately increase extra keywords of the products that senior people pay attention to and increase the traffic and budget at the same time.

Insight 5:

Although the number of interactions (Figure 26) was reduced, average order income is increased which might indicate that:

a) The Ad delivery was more accurate in 2020, and customers required no additional searching.
b) Users have a clear purpose than 2019.

2. Gender dimension:

Interaction. The overall number of female interactions is much higher than that of males. The trend of male and female interaction in 2019 is basically the same, and the third quarter is the peak of interaction. The overall third-quarter interaction in 2020 is lower than the first quarter. Men reached the peak of interaction in the fourth quarter.

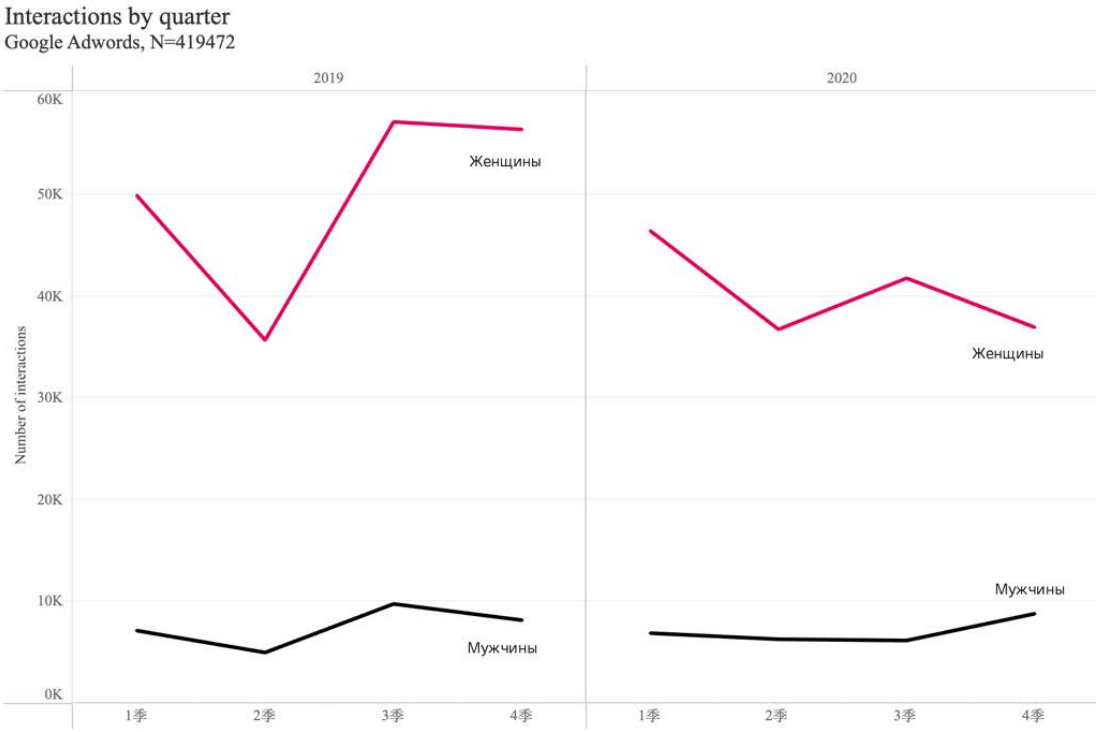


Figure 31 Interactions by quarter by gender

Transaction. The peak of male transactions was in the fourth quarter while the peak of female transactions was in the third quarter. The second quarter is an off-season for both men and women.

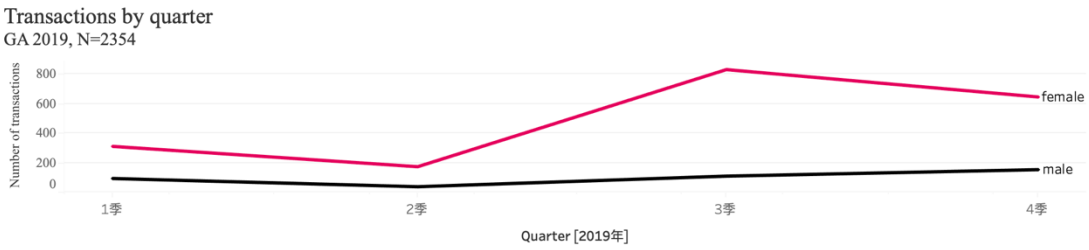


Figure 32 Transactions by quarter by gender

Although female transactions accounted for 81.42% of the total transaction volume (GA2019-2020, N female=3042, N male=694), from the transaction rate point of view, the transaction rate of men is greater, and in 2020, the transaction rate of men Reached 1%. It shows that men are more purposeful when searching for information on the Internet, and it is easier to complete the final purchase behavior.

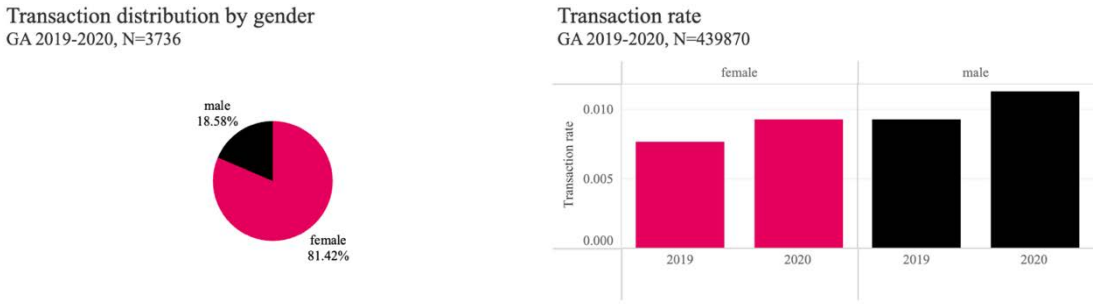


Figure 33 Transaction distribution and transaction rate by gender

Insights 1:

Keyword search settings need to focus on women in the third quarter and men in the fourth quarter.

Income. Overall, female consumers spend much more than men (69% higher in 2019 and 70% higher in 2020). From the perspective of Average order income, men are higher than women and reached the highest in the fourth quarter. The AOI of men in the fourth quarter of 2020 was 8221 RUB, compared with the AOI of men in the fourth quarter of 2019, which is 3830 RUB, an increase of 53%.

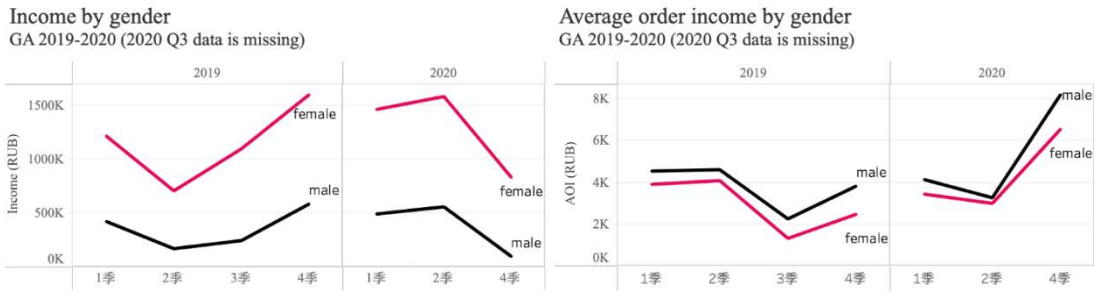


Figure 34 Income distribution and average order income by gender

Insight 2:

Considering the festivals and holidays in Russia in the fourth quarter. Male may need to buy gifts or choosing products for themselves and others. According to the seasons, brands can adjust their marketing strategies to cater to the audience's search preferences and search needs.

Summary

In this chapter, we have studied and discussed the actual data from Lancôme paid search business.

Firstly, we found efficient keywords set according to the keyword efficiency defined by the paid search application situation in Lancôme. In terms of the year comparison, the paid search performed better in 2020 than in 2019, with ROI in 2020 increased by 475% compared to 2019.

After that, we did a regression to explore the relationship between Efficiency and other paid search indicators. The most positively correlated independent variables with efficiency are new users, bounce rate, and transaction rate. We found that customers who search Lancôme related product keywords for the first time have higher purchase intention and can bring higher transaction rate and revenue.

Finally, we supplemented the portraits of paid search users to make up for Lancôme's lack of understanding of paid search users. We analyzed paid search users from three aspects: User profile, User preference, and User behavior, and got the following core findings:

a) Core Lancôme paid search users are around 25-44 years old. They are adults with purchasing power and decision-making power. They have higher requirements for cosmetics and know what they want.

b) Although the number of male users is much lower than that of females, their average purchasing power is higher than that of females.

c) Core Lancôme paid search users mainly live in fast-growing large cities, such as the Moscow area and St. Petersburg. With the penetration of e-commerce services in Russia, there are also a small number of transactions with high average order income in some remote cities.

d) The main mobile devices, operating systems, and browsers used by Lancôme paid search users are consistent with the mainstream in the country. The main users use Apple, Samsung, Huawei phones, Android and iOS operating systems, Chrome, Yandex, and Safari browsers.

e) Autumn is the peak consumption period. On the one hand, it is affected by brand sales strategies, marketing activities, and promotions mainly concentrated in this season. On the other hand, affected by holidays, people prefer to consume or buy gifts for their families and friends during the holidays.

f) The user's shopping preferences are greatly affected by the seasons. They buy cosmetics in spring, skincare products in summer and autumn, and perfumes in winter.

g) When people search for some more purposeful words, such as “online store,” “buy Lancôme,” “price,” etc., it means that they are more likely to be ready to buy through paid search ads link.

5. Discussions

5.1 Keywords test on real business

After calculating the ROMI of the keywords, based on the Lancôme e-commerce manager's interpretation of the Lancôme brand marketing trends in 2021, we tested our hypothesis by selecting some keywords (keywords marked in red in the following table 7) whose ROMI is higher than the keyword "ланком" in 2020.

The hypothesis is that with the increasing of the budget of cost-per-click and the traffic, the income brought by paid search channel will increase.

Thus, we set higher traffic and budget of these red-marked keywords to verify whether this approach will increase the clicks of paid search advertisements related to these keywords and bring the final purchase.

2020 Google Keywords	Total spend	Income	ROMI
ланком tresor midnight rose	22.32	3222.33	14437%
genifique yeux pearl	72.78	7180.01	9865%
visionnaire advanced multi correcting cream	64.89	4903.34	7556%
lancom.ru	430.64	30346.84	7047%
ланком россия	160.03	10074.41	6295%
lancome la vie est belle intense купить	60.19	3762.83	6252%
vie est belle купить духи	55.58	2275	4093%
lancome absolue yeux	114.77	4624.49	4029%
lancome magie noire edt	207.17	5651.26	2728%
lancome la nuit tresor купить	101.13	2462.5	2435%
hypnose men	190.97	4426.5	2318%
advanced genifique сыворотка активатор молодости	154.5	3527.51	2283%
тушь грандиоз	167.14	3203.66	1917%
ланком официальный сайт	19167.76	332829.85	1736%
lancome visionnaire advanced skin corrector	454.65	7564.67	1664%
трезор роуз	183.34	3042.6	1660%
lancome miracle forever	344.95	5697.74	1652%
lancome ru	19265.96	298356.66	1549%
blossom lancome miracle	552.52	7999	1448%
genifique serum	290.72	4092	1408%
ланком	**254154.29**	**3569280.67**	**1404%**

Table 7 Efficient keywords in 2020, Google

Due to time constraints, we only have two weeks to do the keyword test. The results and analysis of the two-week experiment are as follows:

a) By increasing budget and traffic, clicks on these keywords have increased by 50%, which is a good start as we expected.

b) Although the number of clicks has increased, we have not seen an increase in e-commerce income from paid search channel.

The reason may behind:

a) The test time is short. It is necessary to observe the long-term changes brought about by changing the budget and traffic for paid search.

b) According to the year's sales trends, the time of the test is in the off-season of the year. There are no major marketing campaigns in this quarter, and the number of holidays is also limited. Therefore, compared with other quarters of the year, the income from paid search channel in this quarter is reasonable correspondent smaller.

5.2 Impact of the Covid-19 epidemic

5.2.1 On Russian cosmetics market

In 2020, with the development of covid-19 epidemic in Russia, Russia began to gradually control residents' travel in March 2020. Due to the control measures, more users have reduced the activities of going out and gathering.

During this period, people began to choose online shopping more frequently. For the cosmetics market, offline channels received a lot of impact. Therefore, the improvement of online shopping channels has become an important work of cosmetics companies during this period. In addition, building a sound logistics channel is also the focus of cosmetics companies.

In addition to the impact of offline sales channels, the production and raw material sources of cosmetics companies have also been tested. After the interruption of many factories and transportation channels, whether the production of goods can be guaranteed is also the focus of cosmetics companies.

Similarly, due to the more balanced information consumers get when shopping online, the competition in the cosmetics industry is also fiercer, which may lead to more marketing activities to obtain consumers at low prices.

In general, due to the impact of the covid-19 epidemic, the challenges to the cosmetics industry are all-round, including the challenges of upstream supplier channels and the greater pressure of production and competition in the industry. For the cosmetics market, this is a challenge and an opportunity for progress.

5.2.2 On Lancôme paid search

Overall, the paid search performed better in 2020 than in 2019, with ROI in 2020 increased by 481% compared to 2019. Here are the reasons may behind it.

Macro reasons:

a) With the maturity of software and hardware technologies, people use the Internet more to search for information and purchase products.

b) Due to the Covid-19 impact on the economy, people use paid search in a more purpose-oriented way, reducing unnecessary page viewing and personal consumption expenditures.

Micro reasons:

a) The brand has made great efforts in marketing, so the brand influence has increased.

b) The working pace of the e-commerce team in paid search has changed. The frequency set by keywords has been increased from twice a year to once a quarter, which more effectively captures the changes in consumer preferences.

This year, the frequency of setting keywords has been increased to once a month, and we are optimistic about the paid search performance in 2021.

We believe that due to the impact of the coronavirus epidemic, online shopping scenarios are rapidly replacing traditional offline shopping scenarios and will continue for a period of time in the future.

According to a survey of American consumers by the research company Morning Consult, 24% of consumers currently say that they may feel uncomfortable when shopping in the mall in the next six months; 16% of consumers need three months' buffer time.[6] The same situation in Russia, a survey by Viber and Sbermarket showed online shopping is gradually replacing people with shopping, especially in 2020. During the COVID-19 pandemic, convenience, safety, and shopping frequency are more important for shoppers than ever before. Therefore, consumers are more motivated to choose online shopping than ever before.

This change in consumption mode will also provide more data support for e-commerce. At the same time, this stage is also an ideal experimental environment that supports fast-tracking and testing of related businesses to launch premium, a combination of experience and data-driven services to meet consumer requirements.

5.3 Paid search in practice

Advertising carries the role of promoting the development of brands, and advertising expenditure also accounts for a large part of enterprise expenses. Evaluating the effectiveness of advertising campaigns has always been a difficult task.

With the development of information technology, the advertising costs invested in the Internet and mobile Internet have increased dramatically. In terms of advertising cost, paid search advertising consists of an indispensable part.

In our research, Lancôme's paid search keywords mainly include two types: brand and brand-containing vocabulary, such as Lancôme, Lancôme perfume, Lancôme site, etc.; non-brand vocabulary, such as cream, mascara, skincare, etc. From the marketing perspective, the measure of effectiveness is the increase in user visits brought by paid search engine websites, and from the e-commerce perspective, it is the increase in revenue brought by paid search advertising.

[6] https://morningconsult.com/return-to-shopping/

The data research in this article shows that paid search advertising is better for new users or users who are not very active (such as men, remote regions, etc.), but most of the search promotion costs are occupied by active old users.[7]

Therefore, we must pay attention to this trap. As brands, they usually first choose the keywords that best reflect their own advantages and then select the target user group that best fits them. But for a globally well-known brand like Lancôme, the user group targeted by paid search is often already the old users of the brand, which leads to the fact that most of the clicks are contributed by old users.

Because the paid search service is paid by click, we should note that it is easier for old users to click on the content they are familiar with, but this part does not contribute to incremental revenue. For old users, even if they do not see the promotion information of the product in the search engine, they will find and purchase related products by directly visiting the website or other means. This is not the result that the brand side wants to see.

[7] Blake T, Nosko C, Tadelis S. Consumer heterogeneity and paid search effectiveness: A large-scale field experiment[J]. Econometrica, 2015, 83(1): 155-174.

Implications

Due to the change brought by Covid-19, the new consumption behaviour provided more data support for e-commerce. At the same time, this stage is also an ideal experimental environment that supports fast-tracking and testing related businesses to launch premium, a combination of experience, and data-driven services to meet consumer requirements.

Our recommendations are two parts: business recommendations and managerial recommendations.

Recommendations for paid search strategy:

a) Keywords should be added or deleted in time based on Lancôme brand advantages, marketing activities, and consumer behaviors to promptly capture paid search users' changing needs.

b) Lancôme, as a well-known brand should consider reducing paid search advertising budgets on the user-centralized area such as Moscow and St Petersburg. Instead, try to increase traffic and budget on users from remote areas, male users, or users who have not been active in recent time. This is the key to the success of search ads. And this recommendation will be validated by Lancôme team in the future.

c) When placing paid search ads, try to make consumers aware of things that are not in their knowledge. For example, well-known brands such as Lancôme should use paid search to promote products or services that are new for consumers.

Recommendations for managerial improvements:

a) Cross-department cooperation inside the company is good for knowledge transformation and innovation. It might bring a better ROMI of paid search as the marketing department has professional consumer insights, and the e-commerce department is the specialist in online sales strategy.

Conclusion

This research aims to understand the efficiency of the paid search business and the user characteristics of the paid search business of the Lancôme brand, and provide Lancôme with managerial suggestions that can improve the efficiency of paid search.

The data on paid search provided by Lancôme is presented in the form of tables with different dimensions. Dimensions include user age, user gender, user geographic distribution, keyword search, browsers, devices, and sales of different goods. And these data come from two different search engine channels, Google and Yandex. In the process of this research, we integrated data of different dimensions and analyzed them separately.

Based on the characteristics of the data collected by Lancôme in actual work, the study analyzed different indicators for evaluating marketing activities and finally concluded that it is most appropriate to use ROMI to evaluate the efficiency of paid search channels.

Taking ROMI as the efficiency evaluation standard, the study uses logistic regression to build a model for the keyword search data in Lancôme's paid search channels. Constructing the model is to analyze which keywords searched by users can effectively bring more profits to Lancôme and a higher input-output ratio. As a result, according to the finally constructed model, we believe that new users are groups that can bring more income to the paid search channel. Lancôme needs to increase its appeal to new users and acquire more new users.

Then, we analyzed paid search user profile, user preferences, and user behavior.

As a result, the research achieved the original research plan and business goals. The process of the research is to analyze the user-related data of the Lancôme brand in paid search channels, choose suitable efficiency evaluation method of the paid search channels, and complete paid search user analysis models to better measure and improve the performance of Lancôme e-commerce department in the paid search channel.

Finally, we provided managerial recommendations for Lancôme based on our study in order to help them to improve paid search activities.

References

Academic Literature in English

1. Al-Debi M M, El-Haddadeh R, Avison D. Defining the business model in the new world of digital business[J]. AMCIS 2008 proceedings, 2008: 300.
2. Berman R, Katona Z. The role of search engine optimization in search marketing[J]. Marketing Science, 2013, 32(4): 644-651.
3. Bharadwaj A, El Sawy O A, Pavlou P A, et al. Digital business strategy: toward a next generation of insights[J]. MIS quarterly, 2013: 471-482.
4. Blake T, Nosko C, Tadelis S. Consumer heterogeneity and paid search effectiveness: A large‐scale field experiment[J]. Econometrica, 2015, 83(1): 155-174.
5. Brooks N. The Atlas rank report: How search engine rank impacts traffic[J]. Insights, Atlas Institute Digital Marketing, 2004.
6. Chen C Y, Shih B Y, Chen Z S, et al. The exploration of internet marketing strategy by search engine optimization: A critical review and comparison[J]. African Journal of Business Management, 2011, 5(12): 4644-4649.
7. Dai D, Luca M. Effectiveness of paid search advertising: Experimental evidence[J]. Harvard Business School NOM Unit Working Paper, 2016 (17-025).
8. Drnevich P L, Croson D C. Information technology and business-level strategy: Toward an integrated theoretical perspective[J]. Mis Quarterly, 2013: 483-509.
9. Edelman B，Ostrovsky M and Schwarz M． Internet advertising and the generalized second-price auction: Selling billions of dollars worth of key- words［J］． American Economic Ｒeview，2006，97(1) : 242 －259．
10. Edelman B and Ostrovsky M． Strategic bidder behavior in sponsored search auctions［J］． Decision Support Systems，2007，43(1) : 192 －198
11. Falk T, Schepers J, Hammerschmidt M, et al. Identifying cross-channel dissynergies for multichannel service providers[J]. Journal of Service Research, 2007, 10(2): 143-160.
12. Gensler S, Leeflang P, Skiera B. Impact of online channel use on customer revenues and costs to serve: Considering product portfolios and self-selection[J]. International Journal of Research in Marketing, 2012, 29(2): 192-201.

13. Gluhovsky I. Forecasting click-through rates based on sponsored search advertiser bids and intermediate variable regression[J]. ACM Transactions on Internet Technology (TOIT), 2010, 10(3): 1-28.
14. Goldfarb A. What is different about online advertising?[J]. Review of Industrial Organization, 2014, 44(2): 115-129.
15. Kim A J, Jang S, Shin H S. How should retail advertisers manage multiple keywords in paid search advertising?[J]. Journal of Business Research, 2019.
16. Kritzinger W T, Weideman M. Search engine optimization and pay-per-click marketing strategies[J]. Journal of Organizational Computing and Electronic Commerce, 2013, 23(3): 273-286.
17. Jank W，Shmueli G and Wang S． Dynamic，real-time forecasting of online auctions via functional models［R］． Paper presented at the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining，ACM，2006．
18. Laffey D. Paid search: The innovation that changed the Web[J]. Business Horizons, 2007, 50(3): 211-218.
19. LeBaron M, Lee H. Bidding system for search engine marketing: U.S. Patent Application 12/006,009[P]. 2009-7-2.
20. Lewis R A, Reiley D H. Online ads and offline sales: measuring the effect of retail advertising via a controlled experiment on Yahoo![J]. Quantitative Marketing and Economics, 2014, 12(3): 235-266.
21. Li H, Kannan P K, Viswanathan S, et al. Attribution strategies and return on keyword investment in paid search advertising[J]. Marketing Science, 2016, 35(6): 831-848.
22. Li K, Lin M, Lin Z, et al. Running and Chasing--The Competition between Paid Search Marketing and Search Engine Optimization[C]//2014 47th Hawaii International Conference on System Sciences. IEEE, 2014: 3110-3119.
23. Mahdian M, Tomak K. Pay-per-action model for online advertising[C]//Proceedings of the 1st international workshop on Data mining and audience intelligence for advertising. 2007: 1-6.
24. Mesenbourg T L. Measuring the digital economy[J]. US Bureau of the Census, 2001 (1).

25. Milichovský F, Šimberová I. Marketing effectiveness: Metrics for effective strategic marketing[J]. Engineering economics, 2015, 26(2): 211-219.
26. Milgrom P R, Weber R J. A theory of auctions and competitive bidding[J]. Econometrica: Journal of the Econometric Society, 1982: 1089-1122.
27. Nambisan S, Wright M, Feldman M. The digital transformation of innovation and entrepreneurship: Progress, challenges and key themes[J]. Research Policy, 2019, 48(8): 103773.
28. Pagani M. Digital business strategy and value creation: Framing the dynamic cycle of control points[J]. Mis Quarterly, 2013: 617-632.
29. Richardson M, Dominowska E, Ragno R. Predicting clicks: estimating the click-through rate for new ads[C]//Proceedings of the 16th international conference on World Wide Web. 2007: 521-530.
30. Rutz O J, Bucklin R E. From generic to branded: A model of spillover in paid search advertising[J]. Journal of Marketing Research, 2011, 48(1): 87-102.
31. Rutz O J, Trusov M, Bucklin R E. Modeling indirect effects of paid search advertising: Which keywords lead to more future visits?[J]. Marketing Science, 2011, 30(4): 646-665.
32. Sen R. Optimal search engine marketing strategy[J]. International Journal of Electronic Commerce, 2005, 10(1): 9-25.
33. Tapscott D. The digital economy anniversary edition: Rethinking promise and peril in the age of networked intelligence[M]. McGraw Hill Professional, 2014.
34. Terrance A R, Shrivastava S, Kumari A. Importance of Search Engine Marketing in the Digital World[C]//ICITKM. 2017: 155-158.
35. Weltevreden J W J. Substitution or complementarity? How the Internet changes city centre shopping[J]. Journal of Retailing and consumer Services, 2007, 14(3): 192-207.
36. White A. Search engines: Left side quality versus right side profits[J]. International Journal of Industrial Organization, 2013, 31(6): 690-701.
37. Yang S, Ghose A. Analyzing the relationship between organic and sponsored search advertising: Positive, negative, or zero interdependence?[J]. Marketing science, 2010, 29(4): 602-623.
38. Yang Z, Shi Y, Wang B. Search engine marketing, financing ability and firm performance in E-commerce[J]. Procedia Computer Science, 2015, 55: 1106-1112.

Academic Literature in Other Languages

1. 李凯,邓智文,严建援.搜索引擎营销研究综述及展望[J].外国经济与管理,2014,36(10):13-21.
2. 陆芹. 网络经济环境下的市场营销转变思考[J]. 收藏, 2017, 24.
3. 于朝晖. CNNIC 发布第 44 次《 中国互联网络发展状况统计报告》[J]. 网信军民融合, 2019, 9.
4. 姚兴安,闫林楠.数字经济研究的现状分析及未来展望[J].技术经济与管理研究,2021,(02):3-8.
5. 易宪容,陈颖颖,位玉双.数字经济中的几个重大理论问题研究——基于现代经济学的一般性分析[J].经济学家,2019(07):23-31.
6. 郑健壮,李强.数字经济的基本内涵、度量范围与发展路径[J].浙江树人大学学报(人文社会科学),2020,20(06):33-39.

Appendix

Figure 35 Data preprocessing

```
In [2]: %%time
sheets = [
    'GA Гугл 2020',
    'GA Гугл 2019',
    'GA Яндекс 2020',
    'GA Яндекс 2019'
]
cols = [
    'Дата',
    'Ключевое слово',
    'Транзакции',
    'Доход'
]
df = pd.DataFrame()
for sheet in tqdm(sheets):
    df_temp = pd.read_excel(
        'LANCOME_Выгрузки/Lancome_Paid Search_Оплачиваемый поисковый трафик._KW.xlsx',
        sheet_name=sheet,
        index_col=None
    )
    print('read', df_temp.shape)
    df_temp = df_temp[cols]
    df_temp['source'] = 'google' if 'Гугл' in sheet else 'yandex'
    df = df.append(df_temp)
    print('done', df.shape)
df.head()

read (117251, 12)
done (117251, 5)
read (146247, 6)
done (263498, 5)
read (2679, 4)
done (266177, 5)
read (1615, 4)
done (267792, 5)
CPU times: user 3min 12s, sys: 1.58 s, total: 3min 13s
Wall time: 3min 15s
```

Out[2]:

	Дата	Ключевое слово	Транзакции	Доход	source
0	20201014	(not set)	0	0.00	google
1	20200812	ланком	102	86075.60	google
2	20201015	(not set)	1	2542.88	google
3	20201017	(not set)	0	0.00	google
4	20200320	ланком	56	174501.27	google

Figure 36 Frequency statistics

```
In [4]: df = pd.read_csv('search_words.csv', sep='\t')
        df.head()
```

Out[4]:

	Дата	Ключевое слово	Транзакции	Доход	source
0	20201014	(not set)	0	0.00	google
1	20200812	ланком	102	86075.60	google
2	20201015	(not set)	1	2542.88	google
3	20201017	(not set)	0	0.00	google
4	20200320	ланком	56	174501.27	google

```
In [5]: def preprocessing(text):
            for ch in ['\n', '\t', '\r']:
                text = text.replace(ch, ' ')
            result = re.sub('[^a-яА-Яа-zA-Z]+', ' ', text).strip().lower()
            result = re.sub('ё', 'е', result)
            return result

In [6]: df['Ключевое слово'] = df['Ключевое слово'].replace('(not set)', 'notset')
        df['text_proc'] = df['Ключевое слово'].apply(str).apply(preprocessing)

In [7]: text = df['text_proc'].values
        text = ' '.join(text)

In [8]: freq_dist = FreqDist(df['text_proc'].values)
        display(freq_dist)
        plt.figure(figsize=(16, 4))
        plt.title('50 most frequent search queries')
        freq_dist.plot(50, cumulative=False)
        plt.show()

FreqDist({'id': 53950, 'lancome': 14119, 'ланком': 13664, 'dynamic search ads': 11757, 'notset': 9603, 'духи ланком': 8416, 'lanc me': 6363, 'ланком цена': 3466, 'lancom
e духи': 3302, 'lankom': 3244, ...})
```

```
In [25]: show_importance(2, 20, title='Feature importances for search words')
```

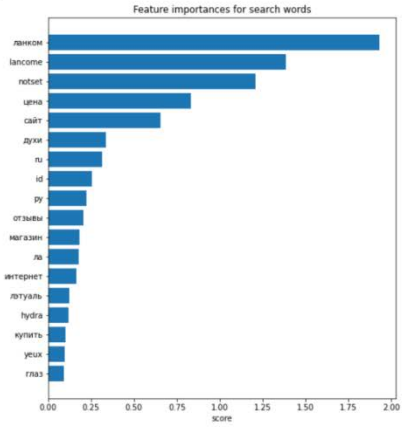
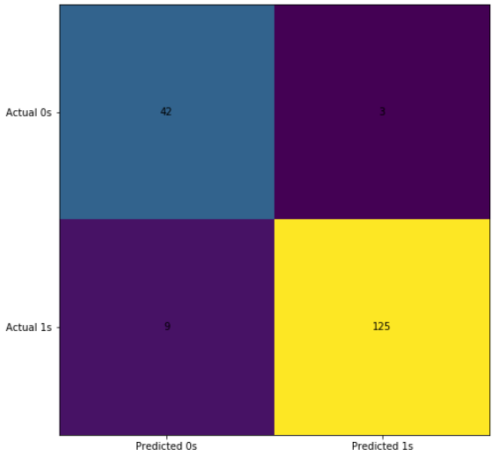


Figure 37 Logistic regression

```
In [1]: from openpyxl import load_workbook
        import matplotlib.pyplot as plt
        import numpy as np
        from sklearn.linear_model import LogisticRegression
        from sklearn.metrics import classification_report, confusion_matrix

        book = load_workbook(filename=r"data1.xlsx")

        sheet = book.get_sheet_by_name("Sheet1")

        data=np.zeros((179,10))

        for i in range(179):
            for j in range(9):
                data[i,j]=sheet.cell(row=i+1, column=j+1).value
            if(data[i,8]>1):
                data[i,9]=1

        X=data[:,0:7]
        y=data[:,9]

        model = LogisticRegression(solver='liblinear', random_state=0)

        model.fit(X , y)

        print('Users, NewUsers, Sessions, BounceRate, Transaction, TransactionRate, Income, Clicks')
        model.coef_

        Users, NewUsers, Sessions, BounceRate, Transaction, TransactionRate, Income, Clicks

        C:\Users\Yukino\Anaconda3\lib\site-packages\ipykernel_launcher.py:9: DeprecationWarning: Call to deprecated function get_sheet_by_name (Us
        e wb[sheetname]).
          if __name__ == '__main__':

Out[1]: array([[-0.02735292,  0.0142404 , -0.02362016,  0.00337189,  0.00026265,
                 0.00166717,  0.0011496 ]])
```

Figure 38 Forecast result of logistic regression

```
In [15]: cm = confusion_matrix(y, model.predict(X))

         fig, ax = plt.subplots(figsize=(8, 8))
         ax.imshow(cm)
         ax.grid(False)
         ax.xaxis.set(ticks=(0, 1), ticklabels=('Predicted 0s', 'Predicted 1s'))
         ax.yaxis.set(ticks=(0, 1), ticklabels=('Actual 0s', 'Actual 1s'))
         ax.set_ylim(1.5, -0.5)
         for i in range(2):
             for j in range(2):
                 ax.text(j, i, cm[i, j], ha='center', va='center', color='black')
         plt.show()
```