Pacific International Journal, Vol. 5(4), 188-199; 2022
ISSN (Print) 2663-8991, ISSN (Online) 2616-4825
DOI: 10.55014/pij.v5i4.279
https://rclss.com/index.php/pij

An Empirical Research on Data Science Talent Training and Enterprise Demand

Dai Jinhua[1, 2*], Su Kun[3]

[1] Jining Normal University, Jining, Inner Mongolia Autonomous Region, China

[2] Center for International Education, Philippine Christian University, Manila, Philippines

[3] Baotou Medical college, Baotou, Inner Mongolia Autonomous Region, China

Abstract: The arrival of the information age has spawned massive amounts of data. How to use data well and give full play to the value of data has become crucial. In this paper, the recruitment information for data science positions from the well-known online recruitment platform BOSS was crawled to deeply analyze the current situation and characteristics of data science talent demand. The results indicate that the demand for data science talents involves various industries that are mainly distributed in coastal cities such as Beijing, Shanghai, and Shenzhen. It also involves various stages of product development, implementation, operation and maintenance, analysis, and testing, indicating that data science talents have broad employment prospects. The professional skills data science talents must acquire are no longer simple Excel, but higher practical experience, such as Python, SQL, Deep Learning, Java, Data Analysis, and Data Mining, etc... This paper suggests that in addition to the basic courses in mathematics and computer science, the training units should also offer Python, Data Analysis, Deep Learning, Data Mining, and other courses, and set up enough experimental courses according to the big data platform tutorial. At the same time, special courses such as financial big data, traffic big data, geographic big data, and bioengineering big data should be set up to realize the construction of multidisciplinary fields. undergraduates are encouraged to practice in enterprises with data application needs as soon as they finish their courses, which could make them establish an intuitive feeling of data and algorithm models to solve various industry problems with real projects and data.

Keywords: data science, talent training, enterprise demand, empirical research

Introduction

With the continuous development of information and communication technologies and industries such as the Internet, mobile Internet, Internet of Things, and 5G, everyone can generate copious data and so do the countless systems and devices which could even automatically generate data every minute. Nowadays global data is showing explosive growth [1]. The scale of the global big data market has increased from $28 billion in 2016 to $57.2 billion in 2020, with a compound annual growth rate of about 19.55%. It is expected to reach $71.8 billion in 2022; the size of China's big data market this year will be CNY84.9 billion, a year-on-year increase of 19.1%, and is expected to reach CNY104.9 billion in 2023 [2-3]. With the continuous expansion of the scale of the big data industry and the great enrichment of application scenarios, the demand for data engineers, data analysts, and data scientists is also increasing. It is estimated that the demand for data science talents in China will reach 2.5 million in 2025 [2-3]. Early in 2015, China issued the "Action Outline for Promoting the Development of Big Data", encouraging colleges and universities to set up majors related to data science and data engineering and to carry out interdisciplinary big data comprehensive talent training. By the year 2021, 674 colleges and universities have been approved

[Received 15 Dec 2022; Accepted 17 Dec 2022; Published (online) 31 Dec 2022]

to open this undergraduate major. Most domestic colleges and universities are actively establishing "data science majors" based on their respective subject expertise and foundations. However, due to the lack of clear definitions and connotation discussions on "data science" majors from the overall respective, relatively consistent training objectives and courses have not yet been formed. There is a common phenomenon that the system and training methods of data science are similar to big data technology applications, statistics, computer science, and other professional construction. China's data science professional education system is still in the stage of exploration and development [4-6]. The various unknown factors and training difficulties of this profession are in great contrast to the strong demand of the data science industry.

Up to date, there is still a lack of accurate answers to some questions, for example, what qualities and skills data science talents should have, how to carry out their work, and how to cultivate such talents, and so on. The current research on data science is mainly focused on two aspects: one is to take data science talents as the research object and propose the qualities and basic skills that data science talents should have for specific fields, such as civil engineering [7], medical and health [8-9], marketing [10], engineering education [11] and other fields. The other one is to study the knowledge and skills that data science talents should acquire from the perspective of talent training. For example, Sun Chiyao and Liu Ji'an et al. discuss the merits of the education system for data science undergraduates in top universities in the United States [12]. Wu Juan and Chen Qingli et al. used the College of Geophysics and Petroleum Resources of Changjiang University as an example to illustrate the data science training program from the supply side [13]. Zhi-Feng Dai and Chun-Yan Li et al. pointed out the necessity of cultivating complex data science talents with multidisciplinary domain knowledge [14]. Dan He and Yanli Wang et al. introduced the construction of data science majors from the perspective of new engineering [15-16].

As can be seen from the brief reference review above, the existing literature has mainly studied the competency of existing data science talents and the cultivation of future data science talents from two perspectives and lacks systematic research on the competency requirements and job responsibilities of data science talents from the perspective of employers. In this paper, we study the competency requirements and job responsibilities of data science talents from the demand-side perspective, deeply analyze the data science job recruitment information of enterprises, summarize the demand pattern of data science talents at various levels, explore the construction of data science talents training program oriented by market demand, and provide reference basis for talent training units. Technically, this empirical research will be carried out as follows: Firstly, we use Web Scraper, a self-coded crawler, to collect job information from the professional job information website of Boss Direct, and pre-process and analyze the original data content by using Python language and third-party extension libraries such as Numpy and Pandas. In the second part, the pre-processed data samples are analyzed visually and statistically using tools such as Matplotlib, Pyecharts, and Word cloud, and the job information is fully mined to summarize the law of data science job demand. The third part will be the discussion and conclusion. In this part, some suggestions for data science talent training in colleges and universities will be put forward to provide a reference basis for talent training units to develop training programs.

Data crawling and pre-processing

Due to the characteristics of the information age, a large amount of talent demand information has been released through the Internet, and employers use some professional recruitment websites as intermediaries to release recruitment information. This paper collects recruitment data for data science-related talent demand from a large domestic recruitment website (BOSS Zhipin, https://www.zhipin.com/). The original data is pre-processed by data cleaning, and each piece of information mainly contains three aspects: basic information about the job, basic information about the enterprise, and the requirements of the job for talents.

Crawling of original data

A web scraper is a tool for extracting data from websites, typically in the form of a Chrome browser plugin. In this paper, we use Web Scraper to crawl the data on the demand for data science talents in November 2022 on a large professional job site (BOSS Zhipin) in China and save it in CSV format, with a total of 2025 data samples.

The advent of the digital intelligence era has given rise to massive amounts of data, and data-centric careers and positions such as data engineers, data analysts, and data scientists have emerged. According to the job content of data science-related positions, this paper collects recruitment data from three employment directions: data research and development, data management, and data mining algorithms, respectively. Each recruitment record includes three aspects of information: basic information about the position, basic information about the company, and the requirements of the position for talents. The

typical positions included in the three employment directions are shown in Table 1.

R&D positions	Management positions	Data mining algorithm positions
Big Data Development Engineer	Big Data Product Manager	Data analyst
Big Data Architect		Data mining engineer
Big data operation and maintenance engineer		Deep learning algorithm engineer
Data Engineer		Data Scientist

Table 1: Types of positions

Pre-processing of raw data

Since the structure of the recruitment information data is complex and the data is input by each company, there are bound to be some invalid records or data with irregularities and inconsistent formats in the crawled data. To ensure the accuracy of data analysis, this paper uses Python to remove and standardize the missing values, abnormal values, and duplicate data of the original data and finally obtains a total of 1322 data samples after data cleaning.

To enable better descriptive statistical analysis of the data, the two fields for the maximum and minimum salary are averaged and commonly logarithmized to reduce the impact of outliers. After basic data cleaning, the contents of the structured data fields include the following aspects:

Education requirements: Divided into 6 levels such as unlimited education, junior college/technology, college, bachelor, master, doctor, etc.

Work experience: Divided into 7 levels: unlimited experience, school/recent, intern, within 1 year, 1-3 years, 3-5 years, 5-10 years, etc.

Salary: The monthly salary is divided into 6 levels: RMB3000 or less, RMB3000-5000, RMB5000-10000, RMB 10000-20000, RMB20000-50000, RMB50000 or more, etc. The maximum and minimum values of the salary are used to calculate the average, which is taken as the common logarithm.

Job title: Divided into 10 types of big data operation and maintenance engineers, big data development engineers, data engineers, big data architects, big data product managers, data analysts, data mining engineers, deep learning algorithm engineers, machine learning algorithm engineers, and data scientists.

Company size: Divided into 6 levels: 0-20 people, 20-99 people, 100-499 people, 500-999 people, 1000-9999 people, and 10000 people or more.

Company financing: Divided into 8 levels such as Angel Round, Round A, Round B, Round C, Round D, and Above, Listed, No Financing Required, and No Financing.

This section introduces the data collection and data pre-processing for the requirements for data science talent. By processing the raw data with missing values, outliers, and duplicate data, we realize the cleaning and organizing of the data to prepare for the later descriptive statistical analysis.

Employment data analysis of data science talent

Echarts is an open-source data visualization JavaScript library developed by Baidu, and Pyecharts is a class library that uses Python to call Echarts to generate data visualization. This section analyzes the pre-processed recruitment data from the perspectives of job quantity, popular cities, company industry, company size, salary, education requirements, skill literacy,

etc., and visualizes the data using Pyecharts to visually display the current situation of the job demand for data talents through statistical charts such as a bar chart, pie chart, word cloud chart, map, etc.

The job demand

The recruitment quantity of data science-related positions is shown in Figure 1. It can be seen from the figure, Data Analyst demand is 271, ranking first; Deep Learning Algorithm Engineer demand is 195, ranking second; Big Data Development Engineer Big Data Development Engine demand is 174, ranking third; Data Mining Engineer and Machine Learning Algorithm Engineer The demand for Data Mining Engineer and Machine Learning Algorithm Engineer is close to the average value of 132; the demand for Data Scientist is only 64 and the demand for Big Data Operations Engineer is less, respectively 64 and 46.

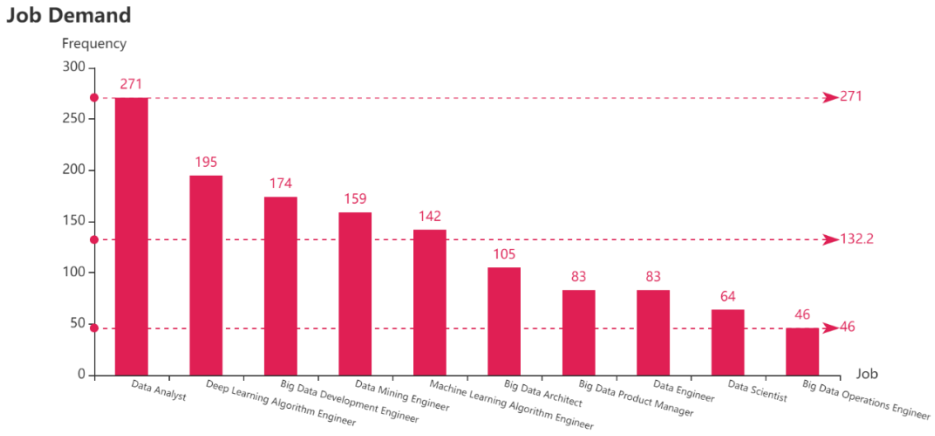

Figure 1 Data science-related job Demand

In terms of copious position types for Data science talent, we will show them by drawing a word cloud chart, of which the font sizes reflect the number of data science talents demanded in this position. From figure 2, we can see that Data Analyst, Deep Learning Algorithm Engineer, Big Data Development Engine, Data Mining Engineer, and other data mining algorithm jobs have more demand, accounting for 60% of the total demand, the future domestic data mining algorithm talent gap is large, especially Data Analyst. Data Scientists, Big Data Operations engineers are in less demand. Data science positions involve development, implementation, operation and maintenance, analysis, testing, and other stages of product development, employment direction, and good prospects.

Figure 2 The job demand word cloud

Company Industry Distribution

Data science should study data collection, cleaning, management, analysis, and mining and explore how to use data science to solve various practical problems and provide ideas and methods for the natural and social sciences. Therefore, in this paper, variable statistics on company industries are conducted based on the recruitment data of data science, and the ten industries with the highest frequency are extracted and plotted into a bar chart, as shown in Figure 3. Figure 3 shows that the job demand

is mostly distributed in the Internet, computer software, e-commerce, finance, and data services industries. Among them, the demand for Internet and computer software is above average, especially the Internet industry talent demand of as much as 417, accounting for 31.54% of the total sample. In addition to medical health and intelligent hardware data science, talent demand is also more prominent.

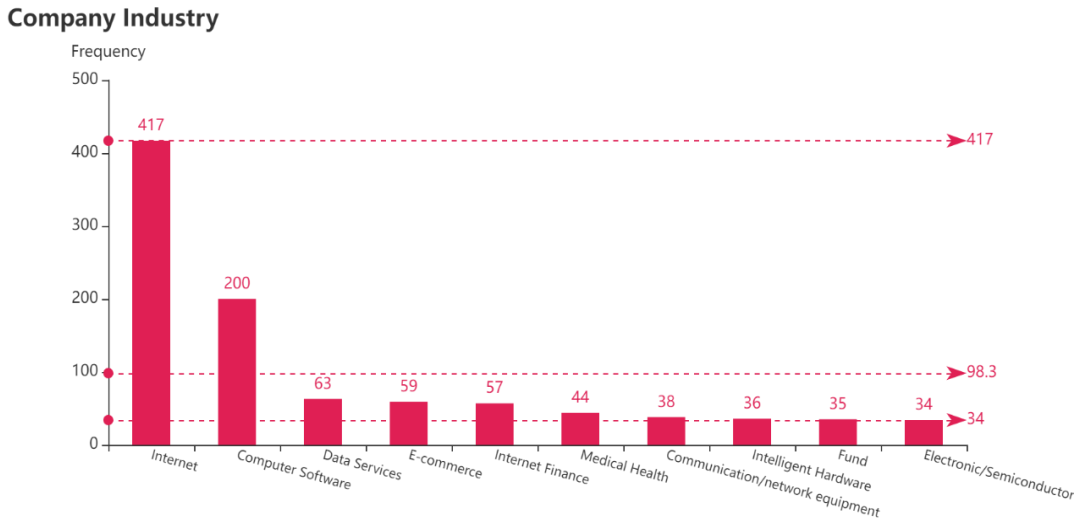

Figure 3 Company industry distribution map

In addition to the above ten industries with a large demand for data science talents, all industries have a demand for data science talents. In the word cloud diagram below, the size of the font in the word cloud diagram reflects the amount of data science talent demanded by this industry. As seen in Figure 4, there is a shortage of data science talents in many industries, such as automotive, gaming, information security, new energy, etc. In addition, traditional industries such as wholesale and retail, trade, import and export, online education, clothing, textiles, leather, etc. have also set up data science positions, indicating that data science has penetrated traditional industries. Both traditional industries and emerging sunrise industries have a strong demand for data science talents, and data science majors will see greater development shortly.

Therefore, on the one hand, data science jobs involved in the expanding industry sectors, will further drive the data collection, cleaning, management, analysis, mining, and other research deepening; on the other hand, the great abundance of data in different fields will further drive numerous applications, and penetrate a variety of scenarios, giving rise to a huge industry; Wherever there is data, there will be a need for people to manage and leverage that data well, which creates a massive data science talent gap.

Data science talents should not only master the format, attributes, and surface meaning of data but also understand the essential meaning behind the data. Future data science talents need to learn industry background knowledge, use data to understand and explore the world, solve various practical problems and create social value. At the same time, in practice, they need to continuously innovate the methods of data collection, management, analysis, mining, and presentation to deepen the connotation of data science.

Figure 4 Word cloud of company industry distribution

City Location Distribution

The data science job information obtained from job sites in this paper covers different cities throughout China. To analyze the popular cities in demand for big data jobs, this paper draws the distribution of the number of hires in the major cities in demand for data science talents into a ring diagram, as shown in Figure 5.

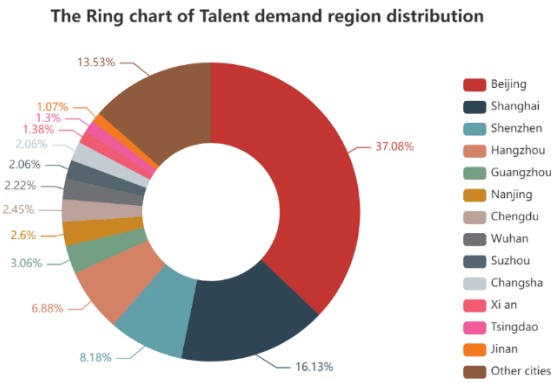

Figure 5 The ring chart of talent demand regional distribution

As can be seen from Figure 5, the top ten cities in demand for data science talent are Beijing, Shanghai, Shenzhen, Hangzhou, Guangzhou, Nanjing, Chengdu, Wuhan, Suzhou, and Changsha in that order. Beijing's demand accounted for 37.08%, topping the country; Shanghai accounted for 16.13%, ranking second in the country in terms of demand; followed by Shenzhen and Hangzhou, accounting for 8.18% and 6.88%, respectively. These cities with developed digital economies are still popular cities for data science jobs, with Beijing, Shanghai, Shenzhen, Guangzhou, and Hangzhou accounting for 71.33% of the national demand. Nanjing, Chengdu, Wuhan, Xi'an, and Jinan—these provincial capitals have a high demand for data science talents, which coincides with the economic development momentum of provincial capitals in recent years and the importance of talent introduction and technology development.

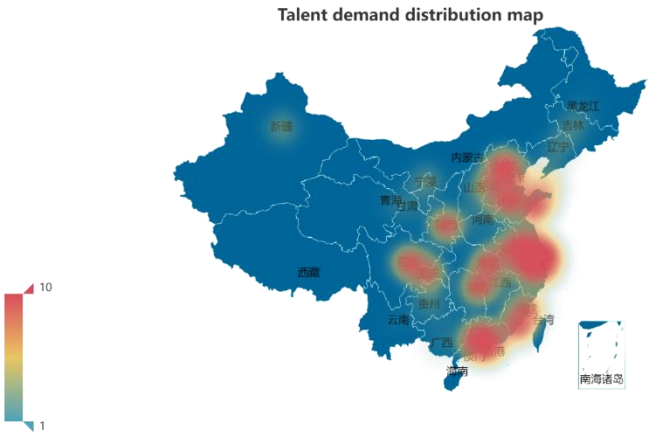

Figure 6 The talent demand distribution map

The demand distribution map6 shows that Beijing is the capital of China and the political, cultural, and technological innovation center of China, and the demand for data science talents also ranks first in the country; Shanghai, the economic and financial center of China, follows closely; while Hangzhou, with its many startups, ranks fourth in the demand for data science talents, just after Guangdong, the first economic province of China. As seen in Figure 6, the overall trend shows that the northern cities are weak and the southern cities are strong, and the coastal cities provide more employment opportunities that are in the national leading position leading the development of the national data science industry. Wuhan, Chengdu, Changsha, and other central regions also have broader development prospects, and the pace of data industry development is accelerating. Xinjiang, Jilin, Inner Mongolia, and Liaoning also have a small number of data science talents in demand, and these relatively less developed regions are beginning to attach importance to the value of data, and other provinces have less demand for talent. From the changes in the ranking, we can see that the level of development of urban data science is positively correlated with the level of economic development of cities, and the difference and differentiation of the development level of the data science industry in different regions are significant.

Salary distribution and related factors

According to the highest salary and lowest salary offered by each position, we calculate the average salary of each job and divide the average salary into six grades: RMB3000 or less, RMB3000-5000, RMB5000-10000, RMB10000-20000, RMB20000-50000, and above RMB50000 according to the salary filtering conditions of the recruitment website.

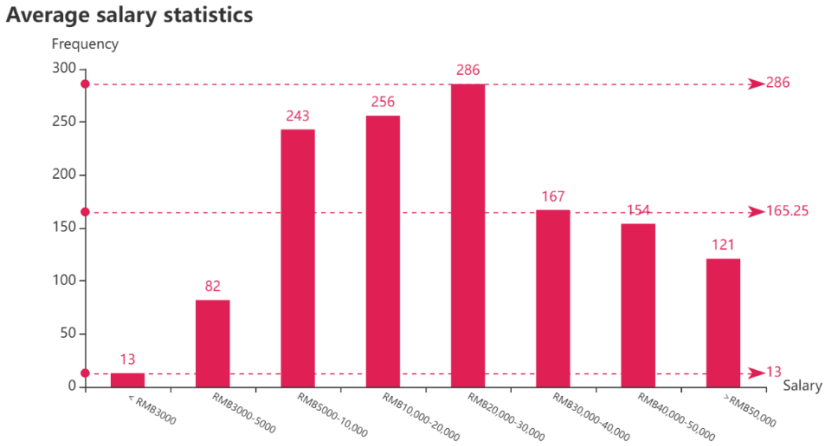

Figure 7 Statistical chart of average salary

The average salary distribution is shown in Figure 7, with only 13 positions with a salary less than RMB 3,000, accounting for only 0.098%; the most positions with a salary from RMB20,000 to 30,000, reaching 21.63%; followed by positions with

a salary from RMB10,000 to 20,000, accounting for 19.36%. The number of jobs with salaries between 10,000 and 40,000 RMB exceeds more than half of the total, reaching 53.63%, indicating that the salaries of data science talents are higher, mainly concentrated between 5,000 and 40,000, with a large gap in demand and a high job threshold.

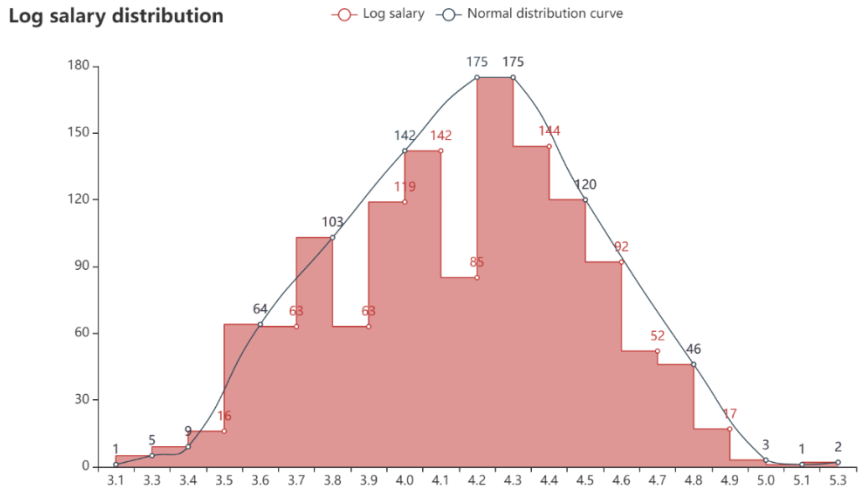

Figure 8 Logarithmic Salary Distribution Chart

Salary package and related factors

In this paper, the factors affecting the log-average salary in terms of both work experience and educational requirements are analyzed. For many majors, the salary is directly proportional to the degree of the applicant, and the same is true for data science majors. The higher the degree of the applicant, the better the salary.

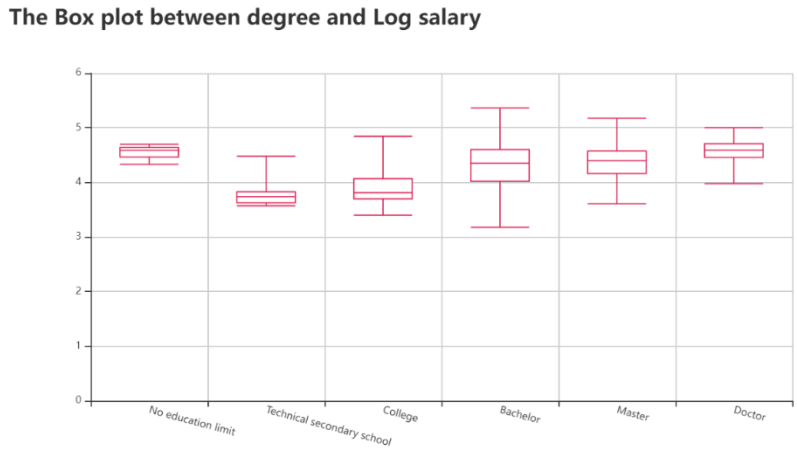

Figure 9 The box plot between degree and log salary

Figure 9 shows that the salary in employers' demand for data science talents increases significantly with education, with bachelor's and master's degrees paying significantly more than college degrees. At the moment, jobs for data science majors are primarily concentrated in the middle and senior degrees of bachelor's and master's degrees, while secondary school and below secondary technical recruitment is limited. Salary ranges for bachelor's and master's degrees are close, with Ph. D.s able to earn higher salaries, though the median position for Ph. D.s is higher.

Salary and work experience is another critical factor. As shown in Figure 10, with the increase in work experience, the salary given by the employer increases significantly; the difference between the salary level within one year of work is not significant, but the salary above one year of work will increase significantly; the difference between the salary level within one year is not significant; the salary of data science talents above 3 years will increase significantly, and the salary will not increase significantly when the work experience reaches above 5 years. Therefore, work experience is crucial for data science, and work experience is accumulated to a certain extent to have better data science thinking. However, since data science is a relatively new profession, companies do not require much work experience—less than three years—and they can even provide internship positions and higher-paid positions for students and fresh graduates. It is easy to conclude that there is a large gap between data science talents and the demand of enterprises, which leads to a broad employment prospect for big data talents.

Figure 10 The box plot between experience and log salary

Knowledge and Skills

As can be seen from Figure 10, the professional skills required for the data discipline include the common programming languages "Python", "C/C++" and "Java", and the common database "SQL". Data analysis mainly includes "Deep Learning", "Data Analysis", "Data Mining "Data Mining" and so on. In addition, some data science-related positions require mastery of big data distributed processing related to Hadoop and Hive software. It can be seen that SQL, Python, Deep Learning, and Java are popular and cost-effective professional skills nowadays. Therefore, the professional skills of data science talents are no longer simply Excel, and after clarifying the skills and mastery of the depth of the situation, it also determines the salary level and promotion space of the future positions in data science.

Figure 11 Word cloud map of knowledge skills

Results and Discussion

In an era when data is so important, there is still a lack of accurate answers to some questions, such as how to use it well and make use of its value becomes crucial, what qualities and skills data science talents should possess, how to carry out their work, and how to cultivate such talents, and so on. In this paper, we address these questions by examining the current employment status, industry distribution, job distribution, salary distribution, and professional skills of mathematical science talents in the Chinese labor market. Here the research findings are summarized to give some suggestions for the cultivation of data science talents in universities.

Job distribution

Data science positions involve all stages of product development, such as development, implementation, operation and maintenance, analysis, testing, etc. 60% of the data science talents are in the data mining algorithm category, and there are large gaps in four data science positions: Data Analyst, Deep Learning Algorithm Engineer, Big Data Development Engine, Data Mining Engineer, and there are lags and shortages in the training of data science talents. There is a lag and shortage in the training of data science talents.

Regional Distribution

The demand for data science jobs is highly consistent with the level of economic development of cities, and the development level of the data science industry in different regions differs and differentiates significantly. The top ten cities in terms of the demand for data science talent are Beijing, Shanghai, Shenzhen, Hangzhou, Guangzhou, Nanjing, Chengdu, Wuhan, Suzhou, and Changsha, and the northwest and northeast regions lag behind other regions of the country in terms of data science talent pool. The first-tier cities still have a greater advantage in terms of the attractiveness of data science talents.

Industry Distribution.

The demand for data science talent is mostly distributed in the Internet, computer software, e-commerce, finance, and data service industries, among which the Internet industry has the most demand for talent, accounting for 31.54% of the total sample. In addition, traditional industries such as wholesale/retail, trade/import/export, online education, and clothing/textile/leather also set up data science positions, indicating that data science has penetrated traditional industries.

Salary distribution.

The positions recruited by data science majors are mainly concentrated in the middle and senior degrees of bachelor and master students, and the salary increases significantly with the increase in education, and a doctoral degree can get a higher salary. With the increase in work experience, the salary offered by employers increases significantly, the work experience requirement is mainly concentrated within three years, and internship positions are offered. Higher-paying positions are also offered for current students and recent graduates. This further indicates that there is a large gap in data science talents.

Professional Skills.

The emphasis on data science talents is not on programming skills and EXCEL, but more on big data processing technologies such as SQL, Python, Deep Learning, Java, and Spark.

In short, data science talents need to be proficient in using big data platforms, proficient in various tools of data management, data mining, machine learning, and data visualization models, and also can learn industry knowledge quickly, understand industry problems, understand the deep meaning reflected behind the data, solve various industry problems and create social value. Therefore, to cultivate excellent data science talents, we should first have a full understanding and knowledge of the current demand for data science talents, and establish an effective training mechanism for data science talents based on industry, function, and skill characteristics, and oriented by market demand.

Suggestions for the cultivation of data science talents in universities

In an era when data is so important, it has become crucial to cultivate a group of human resources who work around data, to create a series of data-centered practitioners who can shoulder the mission given by the times, and to use data well and bring the value of data into play. There is still a lack of accurate answers and clear ideas on how higher education institutions can use and give full play to their strengths to cultivate such talents. Through the study of data science talents' needs, this paper

provides some suggestions with a reference value for data science talent training units from three aspects: training program, curriculum, and comprehensive quality.

Cultivation program. Data science needs to be cultivated with real projects and data. Higher education institutions have the advantages of teaching and researching theoretical data science techniques, but students are too short of practical opportunities and lack business understanding and real-world experience, which makes it difficult to meet the market demand. First, improve the ability of students to practice in internships. Undergraduates should be encouraged to find an internship after the course is completed, for example, they can go to many companies in Beijing, Shanghai, Guangzhou, Hangzhou, and other areas offering internship positions to get in touch with the most cutting-edge professional practical knowledge, to communicate and learn more from customers and industry experts, to handle data more hands-on, and furtherly to build an intuitive feeling of model characteristics. Secondly, by introducing numerous project topics between the school and enterprises and the government, the school encourages students to participate in practical joint projects to become familiar with the operation of big data platforms and gradually build up an in-depth understanding of data. Once again, schools can also establish counterpart internship bases in enterprises to achieve collaborative training of universities and business bodies. Finally, the government can establish an open business park to build an open cooperation platform for schools and enterprises to achieve the sharing of information resources, understanding of business, and practical experience, it is difficult to meet the needs of the market.

Curriculum. Firstly, while emphasizing statistics, mathematics, and computer as the three traditional basic courses of data science, higher education institutions should add courses on deep learning, machine learning, data mining, data analysis, and distributed processing of big data, so that students can be proficient in various data processing tools, proficient in using big data platforms and do some experiments concerning big data platform tutorials. *Secondly,* data science talents should not only master the format, attributes, and surface meaning of data, but also understand the deep meaning reflected behind the data to deeply understand the data, and future data science talents need to learn industry background knowledge. Therefore, colleges and universities can carry out collaborative cooperation among various departments within the school and help students build a more comprehensive and solid knowledge structure system and realize the construction of multidisciplinary intersection fields by offering special courses for students to choose independently, such as financial big data, transportation big data, geographic big data and biological big data engineering, etc. We can vigorously promote mutual learning and exchange cooperation among multiple disciplines to help students build a more comprehensive and solid knowledge structure system. *Finally,* a career planning guidance course can be set up in the basic general education course to guide students to fully understand the development trend and frontier dynamics of today's data science field, help them set up career plans that meet their characteristics and strive to achieve full and highly qualified employment.

Improving the overall quality. Talent training institutions should focus on cultivating students' comprehensive quality, establishing correct values, and continuously improving their cognitive ability, learning ability, innovation ability, and communication ability, taking the knowledge system as the core, and constantly refining and self-improving in the process of practice.

***Acknowledgments*:** We acknowledge the support of Dr. Danny Wong Barrenechea for his constructive suggestions. This research work was supported by the Research on the cultivation mode of school-local-enterprise tripartite cooperation mechanism of data science and big data technology (Grant No.: NGJGH-XH2019084). We also acknowledge the support of our various colleagues at Jining Normal University, for their grateful comments and insights in improving the paper.

About Author: Jinhua Dai (1980-), Chinese female lecturer at Jining Normal University (Inner Mongolia Autonomous Region, China), specializing in information technology and machine learning; Ph.D. candidate in business management with a major in information technology at Philippine Christian University.

Conflicts of Interest: The authors declare no conflict of interest.

REFERENCES

[1] Liu Hongying. (2022). Exploration of talent training mode of data science and big data technology. Talent Resource Development, 13:77-78.

[2] China Big Data Industry White Paper and Top 100 List. (2022). Beijing: Big Data Industry Research Center of Saidi Consulting Co., 2022:1-5.

[3] Hire. (2022). 2022 Future Talent Employment Trends Report. Shanghai: National Development and Reform Commission, Ministry of Science and Technology, etc. and Shanghai Municipal People's Government, 2022:1-5

[4] Chao LM. (2019). Data science theory and practice. 2nd edition. Beijing: Tsinghua University Press,

[5] Zhou Aoying. (2020). Zhou Xuan. Data professional talent training program and core curriculum system construction. China University Teaching, 2020(6).

[6] Chao Le Men, Zhang Chen, Sun Zhi Zhong. (2022). Advances in data science: core theory and typical practice. Chinese Journal of Library Science, 48(01):77-93.

[7] Grajdura Sarah,Niemeier Deb. (2023). State of Programming and Data Science Preparation in Civil Engineering Undergraduate Curricula. Journal of Civil Engineering Education, 149(2).

[8] Mainali Shraddha, Park Soojin. (2023). Artificial Intelligence and Big Data Science in Neurocritical Care. Critical care clinics, 39(1).

[9] Wang M, Wang F L, Zhang L X. (2022). Stress information literacy in clinicians and cultivate data scientists. Zhonghua nei ke za zhi, 61(3).

[10] AhlemeyerStubbe Andrea, Müller Agnes. (2022). Why domain knowledge is essential for data scientists in marketing. Applied Marketing Analytics, 7(4).

[11] Wu, Jingshan, Shi, Jincheng, Zhu, Ling. (2020). Data-enabled transformation of engineering education: an analysis based on five U.S. data science advisory reports. . Research on Higher Engineering Education, 4:41-47.

[12] Sun Chiyao, Liu Ji'an, Xu Yanru. (2022). Undergraduate data science education system of top universities in the United States: construction path, system characteristics, and development trend. Library and Information Work, 66(08):134-143.

[13] Wu Juan, Chen Qingli, Bai Min. (2021). Exploration of talent cultivation objectives of data science and big data technology majors--taking local petroleum universities as an example. Educational Teaching Forum, 20:5-8

[14] Dai ZF, Li Chunyan, Jin Hong. (2022). Teaching reform and practice of multidisciplinary cross-fertilization of data science and big data technology courses. Journal of Hubei University of Economics (Humanities and Social Sciences Edition), 19(07):149-152.

[15] Wang Yanli, Zhang Min, You Guoqiang. (2020). A preliminary study on the construction of data science and big data technology in the context of new engineering. Shaanxi Education (Higher Education), 2:44-45.

[16] He Dan, Niu Yi, Nie Huabei. (2022). Exploration of the construction of data science and big data technology in the context of new engineering. Computer Knowledge and Technology, 18(22):13-15.