Quantitative Analysis of Salary Data in the Big Data Era

Biqi Li*
School of Information and Business Management, Dalian Neusoft University of Information, Dalian, China

*Corresponding author e-mail: libiqi@neusoft.edu.cn

Abstract. The advent of the big data era has brought new opportunities and challenges to every industry, as well as the new insights for the development of human resource management. Compensation management is regarded as an important part of human resource management. Applying big data principles and methods into this field to improve business efficiency and make wiser decisions, which has become a frequent topic of discussion among scholars. This article firstly focuses on the quantitative analysis of compensation data in the era of big data, and then clarifies the main idea of big data human resource management and quantitative analysis of salary data, identifies the problems on the traditional quantitative analysis of salary data, and finally presents the path of quantitative analysis of salary data in the era of big data.

Keywords: Data Analysis, Compensation Management, Big Data

1. Introduction
With the rapid change in Internet and technology, there is a booming increase on the amount of global data, which means that people have entered the age of "big data". The advent of the era of big data has posed a big challenge to traditional human resource management. Applying big data technology into human resource management (HRM) to drive better outcomes for business and employees has become the top priority [1]. As an important part of HRM, compensation management is not only related to the internal fairness of the enterprise, but also has a huge impact on the competitiveness of the company in the industry. The prevalence and popularity of big data calls for a transform and update in compensation field.

2. Main Definition
2.1. Big Data Human Resource Management
The definition of big data in human resource management varies. Big data in human resource management develops from the electronic human resource management (e-HRM), becoming the upgraded path of e-HRM [2]. There is not a broad consensus on the definition of big data in human resource management among authors. This article integrates the opinions of domestic and foreign scholars and believes that, different from e-HRM, big data in human resource management strengthens the character of "data", and emphasizes its application in the business operation through the "scale" of data to achieve business intelligence [2,3]. This means that it can explain and predict employees’
behavior by collecting, processing, analyzing and mining data (including personnel information data, employee behavior data, employee social network data, etc.) beyond the traditional scale, which promotes to offer insights for decision-making on attracting, acquiring, developing, and retaining employees [4,5,6].

2.2. Quantitative Analysis of Salary Data
In traditional compensation management practice, HR usually do quantitative analysis on job evaluation, salary survey, pay structure and salary budget, mainly using Excel and SPSS. In the era of big data, HR have easier access to explosive data, therefore quantitative analysis of salary data need to be transformed to improve the reliability and validity of data, thereby improving the accuracy of decision-making [1]. In this paper, the quantitative analysis of salary data in the age of big data is reflected on the use of more effective data analysis tools (such as SQL, Python and other tools) to collect, process, and analyze massive amounts of data, which ultimately drives productive business decisions.

3. Problems on Traditional Quantitative Analysis of Salary Data

3.1. The Inaccuracy of Salary Data
Many companies implement the salary confidentiality policy, hoping that salary confidentiality can reduce mutual comparisons between employees and public discussion, which has a negative effect on employees' job satisfaction [7]. However, with the popularization of the big data technology, many recruitment websites will disclose salary information, and many consulting agencies will publish industry salary reports every year. In the era of big data, employees have various channels to understand salary composition and check whether the salary level is reasonable, so using the scale of big data and data-based analysis can provide enterprises with latest and accurate salary data, thereby improving the level of talent management [1,6].

3.2. Limited Usage of Traditional Salary Data Analysis Tools
HR normally use Excel for descriptive analysis of data. However, compared with Python, Excel has its limitations. The differences between these tools are mainly reflected on the following aspects:

3.2.1. The Scale of Processing Data. Excel can literally handle hundreds of and thousands of data, but in actual operation, when the amount of data reaches a certain level and the types of data are complex, the speed and efficiency of Excel will be slow, therefore it will be difficult to respond the business needs quickly. On the contrast, Python is far more efficient than Excel in processing big data, and it operates productively in complex calculations, data crawling, automated processing, machine learning, etc. This can fully cope with the massive scale of data and meet different business needs without any delay.

3.2.2. Functions. Excel can realize small-scale data statistical analysis, and combine with VBA for data visualization analysis. Python is a glue language that can not only implement basic descriptive statistics, but also integrate with many programming languages and applications [8]. Python can connect to various databases and extract, write, and modify database data by writing SQL statements. At the same time, Python can link to application APIs to write automated scripts to operate applications. In the process of data exploration, python also provides Pandas for data analysis and a large number of visualization libraries. Furthermore, Python has a rich machine learning and deep learning library. HR can use Python to create machine learning models easily and dig deeper into the value of data to predict employee behavior.

3.2.3. The Complexity of the Operation. When using Excel for quantitative data analysis, every step requires a click with mouse from data cleaning to data analysis, which is very complicated and easy to
report error. In contrast, when using Python to process data, it can be implemented simply with several lines of code. Besides, the analysis code can be saved as a script, and if there are similar tasks in the future, just change the work path, this code can be used directly. Therefore, using python for data analysis is conducive to continuous optimization for code version, and helps the process efficient and reproducible.

3.3. Lack the Integrity of Data-based Thinking on Salary

The coming of the big data era has affected HRM to a certain extent. However, in HRM practice, especially in the field of salary management, some HR workers still get used to addressing problems with traditional thoughts. They fail to recognize the impact of the external environment on the salary field and ignore the background of the big data era. Salary and benefit field usually contain a large amount of data, which applies big data mind frequently, deeply, and widely. However, many HR only focus on certain types of data for analysis in practice. To be specific, they only pay attention to total salary and benefits budgets, ignore other factors’ effect, such as the number of employees of the company, the company's past experience, and the external comparisons in the same industry, and pursue the meaningless number of total compensation blindly. In the era of big data, the significance of data analysis on total amounts is to compare the absolute value of internal and external standards, find differences, and provide insights [6]. For example, when considering the connection between total compensation and employee turnover rate, HR do cross-comparison on several sets of data to find out the potential issues. If HR do not develop a complete data-based mindset or form a cycle on the quantitative analysis of decision-making on compensation management, it will be difficult to face the impact of big data.

4. The Path of Quantitative Analysis of Salary Data in Big Data Era

Quantitative analysis is intertwined with the development of "big data". Through the quantitative analysis of salary data, companies can clarify the main problems in the salary field and use data to drive better decisions. Generally speaking, quantitative analysis of salary data is mainly divided into five steps.

4.1. Find the Highest Priority Business Issues in the Compensation Field

When performing quantitative analysis, it is very necessary to understand the business. The challenges in salary practice will vary depending on the industry, region, and organization scale, so the focus of actual operation in each company is different. Large companies may be more concerned about the external competitiveness, and small enterprises will pay more attention to salary budget and labor costs. That is to say, when doing quantitative analysis, HR should start with the best way to increase the company's value and find the best practice, which can bring greater value to the company [8].

4.2. Collect Salary Data

Collecting massive amount of data is the basis for the subsequent quantitative analysis. The data is large in scale and complex in data types in the salary field. In terms of data scope, it is mainly divided into internal salary data and external salary data [3]. From the data level, it can be divided into individual, team and organization [8]. Internal data involves in organization's salary structure, salary system, employee benefits system, bonus distribution system, stock option plan, implementation and adjustment of salary plan in recent years, organizational performance target (organization level), team incentive plan (team level), basic salary of employees, performance salary, year-end bonus, personalized benefits (individual level), which are mainly obtained from the company by internal human resource system (e-HR), internal reports and business activities, interviews, questionnaires, etc. The main source of external data are Internet search engines, recruitment websites, human resource consulting agencies, outsourcing service platforms, etc [9]. The data, targeting at the organizational level and individual level, can gained by crawler software (such as Houyi collector, Octopus collector, etc.).
4.3. Process Salary Data
Due to different data sources, many salary data will present incomplete, missing, and duplicate records. Therefore, after collecting internal and external salary data, the data must be reviewed and preprocessed. Generally speaking, during data cleaning, it is necessary to check the validity of the data, delete duplicate data, check data labels, calculate missing values, check abnormal values, and define effective data output. Through these six steps, the completeness, accuracy and reliability of the data will be guaranteed [3,4,8].

4.4. Analyse Data

4.4.1. Types of Quantitative Analysis of Salary Data. Quantitative analysis of salary data is mainly divided into descriptive analysis, predictive analysis and guidance analysis. Descriptive analysis mainly refers "what is happening now" and relates to average, median, frequency, frequency, etc., which are achieved through Excel; predictive analysis refers "what will happen" and "why this happened", realizing through Python by using machine learning tools (such as KNN, Naive Bayes, and decision trees, etc.); guidance analysis is the highest level and mainly aimed at "what should we do" through the decision model, providing customized solutions for future decisions and strategies [4,10].

Through analyzing the data, it is presented in the form of a data dashboard. The dashboard includes histogram, bar chart, pie chart, radar chart, scatter chart, map, rose chart, word cloud, etc. The data visualization helps senior managers (department managers, corporate executives) understand the company operation, summarize the past experience and formulate strategic goals for the future. At present, when most companies in China conduct quantitative analysis of salary data, they generally stay at the level of descriptive analysis. Few companies achieve predictive analysis and guidance analysis.

4.4.2. Tools of Quantitative Analysis of Salary Data. Both Excel and Python are excellent tools; HR can select them according to different needs. Small data scale, quick obtained results and simple logical relationship are suitable to be achieved by Excel and VBA, whereas when meeting large scale data and complicated calculation, Python's integration and powerful database can be easily solved.

4.4.3. Content of Quantitative Analysis of Salary Data. In the salary field, quantitative analysis can be carried out from six aspects: labor cost analysis, internal fairness analysis, external competitiveness analysis, personal reward, salary structure analysis and salary satisfaction analysis [11]. In terms of labor cost analysis, analyze the company's revenue, profit, salary budget cost and its change range; when conducting internal fairness analysis, recognize the relationship between employees' education background, skills, department, position, rank, position and internal equity; analyzing external competitiveness need to understand the salary level of external data in the industry based on position, and calculate the frequency, central tendency and dispersion of the salary level; when analyzing the individual reward, consider comprehensively employees’ years of service, attendance, skills, positions, ranks, performance and promotion; as for analyzing salary satisfaction, identify the key factors from position, department and organization perspectives [3,11].

4.5. Offer Insights to Drive Decision-making
According to the results of data analysis, HR should provide optimized suggestions for the company’s salary management. For example, companies need to keep the balance between employee compensation and labor costs. This means that based on the external salary survey, HR can determine the company’s compensation strategy to make it more competitive to attract talents. Besides, it also ensures the relative fairness of the company's internal equity considering internal job evaluations. At the same time, it is necessary to check the labor cost within the reasonable range, reduce the remuneration expenditure as much as possible, and make wise decision, leading to realizing the legitimacy, rationality, relative fairness and effectiveness of compensation management [10,12].
5. Conclusion
The arrival of the age of big data has brought new opportunities and challenges to human resource management. With the support of big data technology, quantitative analysis of salary data can effectively help companies solve problems, thereby retaining core talents, improving employee satisfaction and meeting employee personal interests. Through cooperating with other modules in HRM, quantitative analysis of salary data can promote the human resource management functions, improve the quality and efficiency of decision-making, and finally achieve the company's strategic goals.

Acknowledgments
2019 Liaoning Provincial Social Science Planning Fund Project, Research on the Coordinated Development of Scientific and Technological Innovation Talents and Regional Economy Under the Reform of the Labor Supply Side, Project Number L19BRK002.
2020 Major (Key) Issues of Dalian Federation of Social Sciences, Research on the Cultivation and Incentive Mechanism of Technological Innovation Talents in Dalian, Project Number 2020dlskzd322.

References
[1] James Manyika, et al, “Big Data: The Next Frontier for Innovation, Competition, and Productivity. 2011,” McKinsey Global Institute (2011).
[2] Nan X, Yuming L, Jianfeng P, Haigang M, From e-HRM to HRM: The Evaluation of Human Resource Data Management, Technology of HRD, 05 (2017) 79-88.
[3] Kai Y, Hongyi G, Big Data in Human Resource Management: Transformation and Challenge, Fudan Journal (Social Sciences), 03 (2018) 146-155.
[4] Gerard George, Martine R. Haas, and Alex Pentland, “Big Data and Management,” Academy of Management Journal 57. 2 (2014): 321 -326.
[5] David Angrave, et al, “HR and Analytics: Why HR Is Set to Fail the Big Data Challenge,” Human Resource Management Journal 26. 1 (2016): 1 -11.
[6] Jean Paul Isson, and Jesse S. Harriott, People Analytics in the Era Of Big Data: Changing the Way You Attract, Acquire, Develop, And Retain Talent ( John Wiley & Sons, 2016).
[7] Day, N.E. "Pay Equity as a Mediator of the Relationships Among Attitudes and Communication About Pay Level Determination and Pay Secrecy", Journal of Leadership & Organizational Studies, 19 (2012) 462-476.
[8] Erik Van Vulpen, the Basic Principles of People Analysis, Publishing House of Electroinic Industry, Beijing, 2020.
[9] Hsinchun Chen, Roger HL Chiang, and Veda C. Storey, “Business Intelligence and Analytics: From Big Data to Big Impact,” MIS quarterly 36. 4 (2012).
[10] Van den Heuvel, Sjoerd, and Tanya Bondarouk, “The Rise (and Fall?) of HR Analytics: A Study into the Future Application, Value, Structure, and System Support,” Journal of Organizational Effectiveness: People and Performance 4. 2 (2017).
[11] Xin L, Compensation Management, China Remin University Press, Beijing, 2020.
[12] Laney, D., LeHong, H., & Lapkin, A. (2013). What big data means for business. Financial Times. (Accessed November 15, 2013) http://www.ft.com/intl/cms/s/0/b1dec7f4-b686-11e2-93ba-.html#axzz2gBZKTk4m