ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue III, March 2026

Data Science Talent Demand in China: A Large-Scale Job Posting Analysis and Implications for Curriculum Alignment

Yang Shiwei [1,2] and Ashardi Abas [3*]

[1]Faculty of Computing and Metal-Technology,Sultan Idris Education University

[2]Guizhou Education University

[3]Faculty of Computing and Metal-Technology,Sultan Idris Education University

DOI: https://doi.org/10.51583/IJLTEMAS.2026.150300052

Received: 25 March 2026; Accepted: 30 March 2026; Published: 10 April 2026

ABSTRACT

The rapid expansion of the digital economy has intensified demand for data science talent in China, yet higher education curricula often lag behind evolving industry requirements. While the skills gap is widely acknowledged, few studies offer large-scale empirical evidence linking labor-market signals to curriculum design in the Chinese context. This study analyzes 12,436 data science–related job postings from major Chinese recruitment platforms between 2022 and 2024 to map employer demands and their educational implications. Using text preprocessing, natural language processing, skill extraction, clustering, and regression techniques, we identify key patterns in geographic distribution, required competencies, and salary drivers. Results show that job demand is heavily concentrated in Tier-1 and Tier-2 cities. The most frequently required skills include Python, SQL, machine learning, big data tools (e.g., Spark and Hadoop), statistical analysis, and communication abilities. Salaries are most strongly influenced by city tier, company size, educational qualifications, and proficiency in specialized technical areas such as cloud platforms and deep learning. A notable mismatch persists between university training and market expectations—particularly in applied technical skills and interdisciplinary problem-solving. These findings provide an evidence-based foundation for curriculum redesign, stronger industry–academia collaboration, and more responsive educational planning. Future research should extend to longitudinal forecasting and cross-country comparisons.

Keywords: data science talent demand; job posting analytics; curriculum alignment; China

INTRODUCTION

The rapid expansion of the digital economy has elevated data science from a specialist technical field to a strategic capability that underpins innovation, productivity, and evidence-based decision-making across business, government, healthcare, manufacturing, and public services. As an interdisciplinary domain, data science integrates statistics, computing, machine learning, and domain knowledge to transform large-scale structured and unstructured data into actionable intelligence [1], [2]. In this context, the competitiveness of national economies increasingly depends not only on data infrastructure but also on the availability of highly skilled data professionals who can design models, manage data pipelines, and translate analytical outputs into organizational value [3].

China represents one of the most consequential settings for examining this transition. National initiatives such as the *Digital China* strategy, the *14th Five-Year Plan* for digital economy development, the *New Generation Artificial Intelligence Development Plan*, and the *Eastern Data, Western Computing* program have positioned data as a core factor of production and have accelerated demand for digital and data-intensive capabilities [3]–[5]. The policy effect is also visible in higher education: according to the article dataset and policy review, the number of universities approved to offer programs in data science and big data technology increased from only three in 2016 to more than 700 by the end of 2023 [5]. At the same time, local governments in Beijing, Shanghai, Shenzhen, Guizhou, Chongqing, Hainan, and other regions have launched complementary digital economy

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue III, March 2026

plans, suggesting that data talent demand in China is shaped not only by national policy but also by regional industrial strategies and spatially uneven innovation ecosystems.

Labor-market signals indicate that this policy momentum has translated into strong employer demand for data science talent. Based on 12,436 job postings collected from Zhaopin.com and 51job.com between January 2022 and December 2024, employers most frequently sought competencies in machine learning, Python, and big data tools such as Hadoop and Spark. They also showed increasing interest in hybrid profiles that combine technical expertise with project management and interdisciplinary collaboration. Salary differentials were associated with city tier, company size, and educational background, indicating that the value of data science talent is embedded within broader regional and organizational structures. Recent studies also confirm that online job advertisements have become an increasingly valuable source of labor-market intelligence for identifying skill trends, skill gaps, and changing recruitment patterns in digital occupations.

Problem Statement

Despite the rapid expansion of data science programs in China—from just 3 in 2016 to over 700 by 2023—many curricula remain overly theoretical and disconnected from industry practice. Employers increasingly seek 'T-shaped' professionals who combine deep technical skills (e.g., Python, Spark) with communication, collaboration, and business problem-solving abilities—competencies often underemphasized in current university training. This misalignment constitutes a core bottleneck in China′s data talent pipeline. Survey evidence cited in this study indicates that over 62% of employers express dissatisfaction with graduates′ practical readiness, particularly in applied data engineering and cross-functional teamwork. While prior research has documented skill gaps, few studies translate large-scale labor-market signals into actionable curriculum guidance tailored to China′s regional and industrial diversity.

Existing research has not fully resolved this challenge. Prior studies have applied NLP, topic modeling, and job-advertisement analysis to skill extraction, but many are limited by single-platform data, cross-sectional designs, a narrow occupational scope, or insufficient linkage between labor demand and educational response [6], [7]. Recent scholarship likewise recognizes the usefulness of online job postings for labor-market analytics, but also notes that many studies stop at describing demand patterns without converting those patterns into curriculum-relevant intelligence or institutionally actionable frameworks. In the Chinese context, this limitation is particularly significant because policy, geography, and industrial specialization jointly shape labor demand, making it inadequate to treat data science skills as static or nationally uniform categories.

The specific gap addressed by this article, therefore, lies at the intersection of three unresolved issues. First, there is limited large-scale empirical evidence on how demand for data science is structured across China’s labor market. Second, there is insufficient analysis of how skill clusters, salary premiums, and regional variation can be interpreted as signals for educational reform. Third, there is a lack of studies that move beyond descriptive labor-market analytics to develop a curriculum-alignment pespective grounded in employer demand. The present study addresses these gaps by combining large-scale job-posting analysis with an explicit curriculum-alignment lens, thereby linking labor-market intelligence to higher-education strategy in a more direct and operational manner.

Research Objectives and Research Questions

In response to the above problem, this study aims to analyze the structure of demand for data science talent in China and examine its implications for curriculum alignment in higher education. More specifically, This study pursues three objectives: first, to identify the scale, distribution, and competency structure of data science job demand in China using large-scale online recruitment data; second, to examine the key determinants of skill demand and salary variation across regions and organizational contexts; and third, to translate these labor-market signals into implications for curriculum alignment in data science education. Based on these objectives, the study addresses the following research questions: **RQ1:** What are the dominant patterns and structures of data science talent demand in China? **RQ2:** Which technical, professional, and contextual factors most strongly shape

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue III, March 2026

employer requirements and salary outcomes? **RQ3:** How can labor-market intelligence derived from job postings inform curriculum alignment in higher education? These objectives are consistent with the article's original analytical framework, which integrates labor-market signals, skill clusters, and curriculum mapping.

Significance of the Study

This study is significant both academically and practically. From an academic perspective, it contributes to the growing literature on labor-market intelligence, digital skill demand, and curriculum responsiveness by showing how online job postings can be used not only to identify market trends but also to generate education-relevant knowledge. Rather than treating labor demand as a descriptive endpoint, the study conceptualizes job postings as dynamic evidence for curriculum redesign, thereby connecting computational labor-market analysis with higher education research and workforce development. In this sense, this study extends prior work on skill extraction and recruitment analytics by introducing a curriculum-alignment perspective grounded in the Chinese context.

From a practical perspective, the study offers actionable value for multiple stakeholders. For universities, it provides an empirical basis for revising course structures, emphasizing applied toolchains, and strengthening interdisciplinary and industry-linked training. For policymakers, it offers evidence that can support more targeted digital talent strategies, regional planning, and education-industry coordination. For employers and industry partners, it clarifies the competency profiles associated with recruitment demand and salary premiums, potentially improving collaboration with universities on internship design, curriculum co-development, and graduate readiness. For students, it provides clearer signals regarding the skill combinations and market conditions associated with employability in the data science field. This study serves as a bridge between national digital strategies, labor-market realities, and educational reform—a role that remains central to its contribution.

Contribution of the Article

This article makes five contributions. First, it provides large-scale empirical evidence on demand for data science talent in China, using 12,436 job postings collected from two major national recruitment platforms over the 2022–2024 period. Second, it identifies the technical and hybrid skill structures most valued by employers, including the continued importance of machine learning, Python, and big data frameworks, as well as project management and interdisciplinary competencies. Third, it demonstrates that salary premiums are shaped by city tier, company size, and educational requirements, highlighting the contextualized value of data science skills rather than treating them as universally priced attributes. Fourth, it advances a scalable methodological approach for extracting labor-market intelligence through text mining, NLP, clustering, and regression analysis. Fifth, and most importantly, it translates these findings into implications for curriculum alignment, thereby offering a more practice-oriented framework for connecting labor-market demand with higher education design. These contributions are directly supported by the article's abstract, objectives, problem framing, and significance discussion.

Structure of the Article

The remainder of this article is organized as follows. Section 2 reviews the literature on data science talent demand, online job postings as a source of labor-market intelligence, and curriculum alignment in higher education. Section 3 explains the research design, dataset, preprocessing procedures, and analytical methods. Section 4 presents the empirical findings on demand distribution, skill clusters, and salary determinants. Section 5 discusses the implications of these findings for theory, policy, and curriculum alignment. Section 6 concludes the article by summarizing the key contributions, acknowledging limitations, and identifying directions for future research.

LITERATURE REVIEW

The purpose of this literature review is to establish the conceptual and empirical foundation for examining demand for data science talent in China through large-scale online job-posting analysis and to explain why this analysis is relevant to curriculum alignment in higher education. Rather than reviewing prior studies one by one,

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue III, March 2026

this section synthesizes the literature thematically around five connected issues: the rise of data science within digital economy transformation, the use of online job postings as labor-market intelligence, the evolution of skill demand in data science occupations, the responsiveness of higher education curricula, and the unresolved gaps linking market evidence to educational reform. This thematic structure is consistent with the article's own organization, which moves from the conceptualization of data science to labor-market demand, education structures, policy initiatives, and industry–academia coordination.

Data science has emerged as a core enabling capability of the digital economy because it supports the extraction of value from large, complex, and rapidly growing datasets. Data science is an interdisciplinary domain at the intersection of statistics, machine learning, computing, and domain knowledge, reflecting the now widely accepted view that data-intensive problem-solving requires more than isolated technical specialization. It further notes that, in the Chinese context, data science has moved from a supportive technical function to a strategic national capability under policy frameworks such as *Digital China*, the *14th Five-Year Plan*, and related digitalization agendas. This policy-led positioning matters because it implies that demand for talent in data science is shaped not only by firm-level recruitment needs but also by state-led digital restructuring, regional industrial priorities, and institutional pressures on universities to supply graduates with relevant competencies [1], [2].

The literature, therefore, converges on an important premise: data science talent demand should not be understood as a static list of technical requirements. It is embedded in broader transformations of production systems, industrial upgrading, digital infrastructure, and governance regimes. This point is reinforced by linking labor-market demand to both national strategic intent and local implementation—particularly through regional development policies, higher education expansion, and evolving employer expectations that favor hybrid profiles combining technical depth with communication, collaboration, and project execution. This broader framing is essential because it situates curriculum alignment not as an internal pedagogical concern alone but as part of a national talent-development challenge.

Online Job Postings as Labor-Market Intelligence Data

A second major theme in the literature concerns the use of online job postings as a source of labor-market intelligence. Traditional labor-market studies often rely on government statistics, employer surveys, or occupational classifications, which remain valuable but are often too aggregated, too infrequent, or too slow to reflect rapidly changing skill requirements in digital occupations. In contrast, online recruitment platforms provide high-volume, near-real-time information on job titles, required competencies, educational expectations, salary ranges, experience levels, and regional variation. For this reason, Online recruitment data are defined not merely as a hiring channel but as a source of structured and unstructured evidence for analyzing labor-market demand. It explicitly frames Zhaopin.com and 51job.com as the primary empirical basis for uncovering patterns in the Chinese data science labor market between 2022 and 2024 [3].

Recruitment data have increasingly been used with text mining, topic modeling, semantic clustering, and related NLP methods to infer employer preferences and occupational trends.Earlier studies used association-rule mining, recruitment information analysis, or data-mining approaches to identify skills and occupational patterns from online advertisements, while more recent work has applied more advanced language-processing techniques to large corpora of recruitment text. The methodological appeal of job postings lies in their granularity: they capture the vocabulary employers actually use to describe roles, making them particularly useful for emerging occupations such as data science, where formal occupational categories remain fluid and often lag behind labor-market realities [4]–[7].

However, the literature also reveals important limitations. First, recruitment data do not perfectly represent the entire labor market. They tend to overrepresent formal, platform-mediated, white-collar, and urban employment while underrepresenting internal hiring, informal recruitment, and niche sectors. Second, job postings describe employer demand signals rather than actual hiring outcomes, so they capture intent rather than completed matches. Third, the language of advertisements may vary across firms, regions, and sectors, which introduces challenges in standardizing skill labels and comparing positions.These limitations are acknowledged: reliance on online platforms may underrepresent niche or emerging sectors, and the observed relationships are associative

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue III, March 2026

rather than causal.These limitations do not invalidate job posting analysis, but they do mean that such data must be treated as a high-resolution proxy for market demand rather than a complete labor-market census.

Even with these constraints, the literature strongly supports the use of online job postings for skill-demand analysis in dynamic digital fields. For data science in particular, where new tools, frameworks, and hybrid role definitions evolve rapidly, recruitment text provides one of the most direct empirical windows into market expectations. This makes job postings highly suitable for a study that seeks to connect labor demand to curriculum redesign.

Skill Demand Analysis in Data Science and Related Occupations

A third theme in the literature concerns the structure of skill demand itself. Existing studies generally agree that data science occupations are characterized by a layered skill architecture rather than a single homogeneous competency set. At the technical level, recurring requirements include programming languages, database management, statistical modeling, machine learning, and big data frameworks. This consensus is reflected in the repeated emphasis by employers on competencies such as Python, SQL, and machine learning frameworks, a s well as platforms like Hadoop and Spark, along with cloud-related capabilities and data-processing proficiency. It also reports that more recent literature increasingly describes demand for “T-shaped” professionals, meaning candidates with deep technical knowledge alongside broader interdisciplinary, communicative, and organizational capabilities.

This shift toward hybrid competency models is one of the most important findings across the literature. Earlier studies of digital skill demand often centered on hard skills alone, particularly programming and analytics. More recent work, however, suggests that employers in data-intensive roles increasingly value soft skills and organizational competencies because data professionals are expected not only to build models, but also to communicate results, collaborate across functional units, manage projects, and translate technical outputs into operational decisions. This study adopts a broader interpretation by explicitly distinguishing technical competencies from cross-disciplinary skills and arguing that demand should be analyzed in clusters rather than as isolated terms.This is analytically important because the market rarely seeks skills in isolation; instead, job ads reflect bundles of competencies associated with specific organizational roles and sectoral needs.

At the same time, the literature reveals several unresolved issues. One is the instability of occupational labels. Roles such as data scientist, machine learning engineer, data analyst, big data engineer, and AI developer frequently overlap in practice, making it difficult to establish a universal occupational taxonomy.Another issue is that many studies treat the frequency of skill mentions as a sufficient indicator of value, even though demand intensity may vary according to location, industry, salary level, or organizational maturity. A third issue is the lack of integrated analysis across technical, professional, and contextual variables. Some studies identify skills; others model wages; others discuss education. Fewer bring these elements together in a unified framework that can support curriculum decisions.

The literature therefore supports the need for more sophisticated models of skill demand that do at least three things simultaneously: identify core skill clusters, explain how these clusters vary across regions and sectors, and connect them to labor-market rewards such as salary premiums. This is precisely the direction This study adopts an integrated approach combining skill extraction, clustering, and regression-based salary analysis.

Curriculum Alignment and Higher Education Responsiveness

A fourth theme in the literature concerns the responsiveness of higher education systems to labor-market change. Rapid expansion in data science education does not necessarily imply effective alignment with employer needs. It argues that China has seen substantial growth in data-related academic programs, but that many programs remain theoretically oriented, unevenly resourced, and insufficiently integrated with practical industry tools, applied problem-solving, or interdisciplinary collaboration. In this respect, the problem is not merely whether universities offer data science degrees, but whether those degrees cultivate the competency combinations actually demanded in the labor market.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue III, March 2026

Theoretical perspectives help explain this issue. Human capital theory suggests that education enhances productivity by developing relevant knowledge and competencies, while skill formation theory emphasizes the institutional processes through which skills are produced, transmitted, and updated. Labor-market matching models further suggest that inefficiencies arise when educational outputs do not align closely with employer demand. Together, these frameworks underscore the study's central concern: for universities to function effectively in a digital economy, curriculum design must be guided by real labor-market signals rather than static disciplinary traditions alone. The literature review synthesizes human capital theory, skill formation theory, and labor-market matching models into an integrative analytical framework.

The literature on curriculum alignment broadly agrees that data science education should be interdisciplinary, practice-oriented, and industry-responsive.Yet several structural obstacles complicate this ideal in China. These include curriculum overlap and misalignment, shortages of qualified faculty, limited access to real-world data and practical training environments, unequal regional resource distribution, and weak or inconsistent industry–academia collaboration. The absence of robust lifelong-learning pathways and the difficulty of aligning educational reform with regional policy priorities further compound these challenges. Together, these factors indicate that curriculum responsiveness cannot be achieved merely by adding a few new courses; it requires institutional redesign, stronger external partnerships, and more dynamic mechanisms for monitoring employer demand.

Another contribution of the literature concerns policy mediation. In the Chinese setting, curriculum alignment is shaped not only by university-level choices but also by state policy frameworks, local development initiatives, and sector-specific directives. This is illustrated by mapping policy drivers to programmatic actions and evaluative KPIs, showing that data science education sits at the intersection of national digital strategies and regional industrial needs.This makes China a particularly significant case for studying curriculum alignment because educational adaptation occurs within a coordinated, policy-intensive environment rather than a purely market-driven one.

Research Gaps in Existing Studies

The literature reviewed above reveals substantial progress, but also several important gaps. First, there remains a shortage of large-scale empirical studies that examine data science labor-market demand in China with sufficient temporal breadth, regional differentiation, and methodological depth. The lack of fine-grained, large-scale empirical evidence is identified as a core gap in prior research. Many earlier studies have been limited by small datasets, reliance on a single recruitment platform, narrow occupational categories, or cross-sectional snapshots that cannot adequately capture the evolving structure of demand.

Second, the literature is fragmented across domains. Some studies analyze digital skill demand from job postings; others examine higher education programs; others discuss policy or talent shortages. Fewer studies integrate these strands into a coherent framework that links employer demand, contextual labor-market variation, and curriculum implications. Earlier work often stops at describing labor demand or identifying skill gaps, without proposing empirically grounded curriculum frameworks or demonstrating how macro-level policies translate into micro-level hiring patterns. This fragmentation limits both theoretical accumulation and practical usefulness.

Third, there is insufficient attention to the clustered and contextual nature of data science skills. Existing studies often treat skills as independent items rather than interdependent bundles shaped by sector, city tier, firm size, or regional development strategy. As a result, they may overlook that market demand is not simply for "Python" or "machine learning" in abstract terms, but for combinations of tools, methods, and professional capabilities embedded in particular organizational and geographic settings.

Fourth, there is a weak connection between labor-market analytics and actionable curriculum alignment.This is identified as one of the most consequential shortcomings in the existing literature. Although prior research acknowledges skill gaps, relatively few studies extend labor-market evidence to structured educational implications, such as curriculum mapping, practice-based modules, industry partnership models, or evaluative

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue III, March 2026

indicators of responsiveness. Without that bridge, labor-market intelligence remains descriptive rather than transformative.

Conceptual Positioning of the Present Study

Against this background, the present study is conceptually positioned at the intersection of digital economy transformation, labor-market intelligence, and curriculum alignment. It treats online job postings as a high-resolution empirical interface between industry demand and educational supply. It also adopts the article's multidimensional analytical logic by integrating skill extraction, cluster analysis, salary modeling, and educational interpretation into a single framework. In doing so, the study departs from narrower descriptive approaches and instead conceptualizes labor-market data as a strategic input for curriculum redesign and talent development planning.

More specifically, the study contributes in three ways. First, it offers a large-scale evidence base for understanding demand for data science talent in China by analyzing 12,436 job postings from major recruitment platforms. Second, it analyzes employer demand as a structured phenomenon shaped by policy environment, region, industry, and organizational characteristics rather than as a simple aggregate of skill frequencies. Third, it extends the literature by explicitly translating labor-market patterns into implications for curriculum alignment in higher education. This positioning responds directly to the article's stated objective of integrating labor-market signals, skill clusters, and curriculum mapping into a single analytical framework.

In summary, the literature shows that demand for data science talent is a strategically important yet analytically complex phenomenon shaped by digital transformation, occupational fluidity, regional variation, and institutional responsiveness. Online job postings provide a powerful data source for capturing these dynamics, but prior studies have often remained either methodologically narrow or insufficiently connected to educational reform. The present study addresses these gaps by using large-scale recruitment data to analyze the structure of data science demand in China and by interpreting the resulting evidence through a curriculum alignment lens. On that basis, the next section explains the research design, data collection process, preprocessing strategy, and analytical methods used to operationalize this framework.

METHODOLOGY

This study adopted a **quantitative research design** grounded in computational social science and labor-market analytics. The central objective was to identify the structure, distribution, and determinants of demand for data science talent in China using large-scale online recruitment data. A quantitative design was appropriate because the study sought to examine observable market signals at scale, including the frequency of job postings, distribution across locations and firms, the co-occurrence of required skills, and the statistical relationship between job characteristics and salary outcomes. This study is explicitly defined as a labor-market analysis based on natural language processing and econometric modeling applied to online recruitment data, with the main analytical dimensions being demand structure, skill mapping, and salary dynamics.

More specifically, the study operationalized data science talent demand through the structured and unstructured content of job advertisements, including job titles, salary ranges, education requirements, work experience requirements, geographic location, company characteristics, and textual descriptions of skills and responsibilities. This design was suitable for addressing the research objectives because the paper aims to answer questions about market demand patterns, skill clusters, and salary determinants rather than to evaluate individual experiences or perceptions. In the broader article, these empirical findings were later interpreted for curriculum alignment, but the present article is anchored in the quantitative analytical core.

Data Sources and Sampling Strategy

The primary data source consisted of publicly available job advertisements collected from two major Chinese online recruitment platforms, **Zhaopin.com** and **51job.com**, covering the period from **January 2022 to December 2024**. The article identifies these platforms as leading recruitment channels for white-collar and technical roles and treats them as suitable proxies for labor-market demand in China's data-intensive

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue III, March 2026

occupations. The final dataset comprised **12,436 job postings**, providing the empirical basis for analyzing the demand structure, skill requirements, and salary patterns in the Chinese data science labor market.

The sampling strategy was **purposive and keyword-driven**. Job advertisements were included if their titles or descriptions matched data science-related professional categories such as data scientist, data analyst, machine learning engineer, big data engineer, and related roles. The article defines the thematic scope of the study in exactly these terms, noting that the research focused on professional and technical positions containing keywords linked to data science, data analysis, machine learning, and big data, while excluding broader information technology jobs that required only routine data processing or generic computing skills. This purposive sampling design was appropriate because the study aimed to isolate the market segment for data science talent rather than the broader digital labor market.

The geographical scope emphasized China's major urban labor markets and representative regional centers, especially first-tier and new first-tier cities such as Beijing, Shanghai, Shenzhen, and Guangzhou, while also accounting for regional variation in other provinces and municipalities. The dataset did not systematically include vacancies circulated through non-public channels such as headhunting, internal referrals, or offline job fairs. Accordingly, the sample should be interpreted as a structured representation of formal, platform-mediated labor demand rather than a full census of all hiring activity. This boundary is clearly acknowledged in the article and remains important for interpreting the findings.

Data Collection and Preprocessing

Data collection focused on extracting both **structured variables** and **unstructured text** from the selected job advertisements. Structured fields included job title, salary, city, work experience requirement, education requirement, company size, and industry category. Unstructured fields primarily consisted of the textual job description and skill requirements, which formed the basis for subsequent text mining and clustering analysis. The article specifies that online recruitment in this study refers to the collection of publicly available postings from Zhaopin.com and 51job.com and that these records include both structured and unstructured information relevant to labor-market analysis.

The preprocessing pipeline was designed to improve consistency, reduce noise, and enhance the quality of downstream NLP analysis. Numerical noise, punctuation artifacts, and non-relevant symbols were removed. Text was normalized through lowercase conversion for English text and standardized Chinese preprocessing for Chinese-language content. Tokenization was performed using language-appropriate tools, specifically **Jieba** for Chinese text and **NLTK** for English text. Stop words were removed, and lemmatization or stemming was used where appropriate to reduce lexical variation. High-frequency but semantically weak phrases, such as generic recruitment boilerplate, were filtered out using a predefined stop-word list. These steps ensured that variations of the same term were grouped into more analytically meaningful forms before model construction.

In addition to linguistic preprocessing, the study performed data filtering based on thematic relevance, retaining only postings associated with the target occupational domain. This stage ensured that the corpus reflected genuine data science-related demand and not loosely adjacent IT or administrative roles. The preprocessing stage was therefore not merely a technical exercise; it was an essential validity check that aligned the final analytical corpus with the study's research questions.

Variables and Analytical Framework

The analytical framework combined descriptive labor-market analysis, unsupervised machine learning, and econometric modeling. The key dependent variable in the salary analysis was **log-transformed salary**, used to reduce skewness in the raw salary distribution. The article explicitly states that salaries were transformed using **log(salary)** before regression analysis. This transformation is standard in wage modeling because salary data are typically right-skewed and heteroskedastic in raw form.

The independent variables included both structural controls and skill-related indicators. Structural controls comprised job type, city type, industry, work experience requirement, and education requirement. In addition,

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue III, March 2026

the study included skill-related feature groups capturing academic, computing, economic-management, scientific, leadership, database, visualization, big-data tool, deep-learning, coding, communication, attitude, and thinking-related attributes. The article also makes clear that talent demand was operationalized through the explicit and implicit requirements stated in the job postings, including technical skills, soft skills, educational qualifications, years of experience, and job volume across regions and industries.

The framework was explicitly designed to link three analytical layers: **demand structure**, **skill clusters**, and **salary dynamics**. Demand structure addressed the scale and distribution of postings across labor-market contexts; skill clusters captured the co-occurrence and latent grouping of employer requirements; salary dynamics estimated the extent to which geographic, organizational, and competency variables were associated with wage premiums. This multi-dimensional design directly reflects the article objectives and allowed the study to move beyond simple frequency counts toward a more integrated interpretation of labor-market demand.

Text Mining and Skill Extraction Procedures

To identify the competency structure embedded in job advertisements, the study used a multi-stage text-mining pipeline that combined keyword extraction, semantic expansion, and embedding-based representations. Skill extraction began with the cleaned job descriptions and requirement fields, in which recurring terms and phrases were identified, standardized, and grouped. The article defines talent demand operationally as the explicit and implicit skill requirements embedded in job postings and treats the extraction of these requirements as central to the study's analytical design.

To model semantic relationships among skill terms, the study trained **Word2Vec** embeddings using both **skip-gram** and **CBOW** configurations as robustness checks. The final specification used a **vector size of 200**, **window size of 5**, **minimum count of 3**, **negative sampling of 10**, **15 epochs**, **subsampling of 1e-4**, and a fixed random seed. Posting-level representations were then generated by averaging word embeddings across relevant terms. These dense vectors supported both skill expansion and downstream clustering. The article further notes that low-dimensional visualization of embeddings was performed using **t-SNE** and **UMAP** to reveal latent relationships among skill sets and occupational signals.

This procedure was particularly suitable for the present study because recruitment language is often variable, overlapping, and semantically unstable. Terms such as "machine learning," "AI," "modeling," "analytics," and platform-specific technical tools may appear in different combinations across industries and regions. Embedding-based modeling, therefore, improved the ability to detect related competencies even when employers used different wording. The resulting skill representation supported a more nuanced analysis of employer expectations than simple word-frequency counting alone.

Clustering and Topic Modeling of Skill Demand

The study next applied unsupervised learning methods to identify latent skill structures within the recruitment corpus. Three methods were tested in parallel: **K-means clustering**, **Non-negative Matrix Factorization (NMF)**, and **Latent Dirichlet Allocation (LDA)**. According to the article, the number of clusters or topics was selected using a combination of the **elbow method** and **grid-search-based coherence diagnostics**, including **C_v** and **UMass** coherence scores. This comparative strategy reduced dependence on a single unsupervised method and strengthened the interpretability of the final skill groupings.

For K-means clustering, the parameters included **k-means++ initialization**, **n_init = 20**, **max_iter = 500**, and fixed random seeds to ensure reproducibility. NMF was estimated using the coordinate descent solver with **β-divergence** as the cost function. LDA was implemented with **symmetric priors** of **$\alpha = 0.1$** and **$\eta = 0.01$**, and the model ran for **1,000 Gibbs passes**. The purpose of these techniques was to uncover recurring bundles of competencies and role profiles rather than isolated technical keywords. The article indicates that clusters such as "ML + Python," "Big-data (Hadoop/Spark)," "Cloud + DevOps," and "Cross-disciplinary project skills" were among the patterns revealed in the corpus.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue III, March 2026

This clustering stage was methodologically important because employer demand in data science is not expressed as a set of independent skills. Rather, it emerges as combinations of tools, knowledge domains, and professional capabilities attached to particular labor-market roles. By modeling the corpus through clustering and topic analysis, the study revealed the internal structure of demand and provided a more realistic basis for educational interpretation and curriculum alignment.

Regression Analysis of Salary Determinants

To analyze the wage structure associated with demand for data science jobs, the study employed multiple regression-based approaches. The primary modeling strategy employed linear regression with log-transformed salary as the outcome variable, supplemented by backward stepwise and **LASSO regression** for feature selection. The article states that backward stepwise regression used the **Akaike Information Criterion (AIC)** as the stopping criterion, while LASSO was tuned using **10-fold cross-validation**, with the optimal penalty parameter selected at the minimum cross-validation error.

The regression models incorporated structural controls such as job type, city type, industry, experience, and education, together with detailed skill indicators extracted from the text corpus. This enabled the study to estimate the associations between location, organizational context, human capital requirements, and technical or soft skills and salary premiums. The article positions this stage as a means of moving from descriptive labor-demand mapping to a more nuanced micro-level explanation of skill value formation in the labor market.

Model diagnostics were also reported. Multicollinearity was examined using **Variance Inflation Factors (VIF)**. Residual distributions and heteroskedasticity were checked, and **HC3 robust standard errors** were used. Predictive validity was evaluated through **10-fold cross-validation** and a stricter **time-based hold-out split**. These procedures increase confidence that the estimated relationships are not purely artifacts of overfitting or unstable model specification. In Q1 journal terms, this is an important strength of the empirical design because it balances interpretability with statistical rigor.

Ethical and Legal Considerations

The study relied on **publicly available online job postings** rather than direct human-subject experimentation. As a result, no sensitive personal identifiers from applicants were involved in the main labor-market dataset. The article defines online recruitment data as publicly accessible records from major recruitment platforms, including structured and unstructured posting content. The research, therefore, focused on organizational hiring signals rather than individual personal data.

At the same time, the study remained attentive to ethical and legal boundaries. First, analysis was conducted at the aggregate level, with the goal of identifying labor-market patterns rather than evaluating specific firms or individuals. Second, the data were used for scholarly analysis of skill demand, salary structure, and curriculum implications, which aligns with the article's stated academic purpose.

Third, the article recognizes broader concerns about data governance, lawful use of data, and institutional compliance in the Chinese educational and digital policy environment, reinforcing the importance of responsible handling of platform-derived data. Where the broader article included qualitative interviews, those components would require informed consent and confidentiality procedures; however, the present article centers on the job-posting analytics component.

Reliability, Validity, and Study Limitations

The study incorporated several procedures to strengthen reliability and validity. Reliability was supported through a reproducible analytical workflow, explicit preprocessing rules, fixed random seeds for clustering and embedding models, and clearly specified hyperparameters for major algorithms. Internal analytical consistency was enhanced by using multiple unsupervised methods—K-means, NMF, and LDA—rather than relying on a single clustering model. Predictive robustness in salary analysis was assessed through 10-fold cross-validation and time-based hold-out validation, while regression diagnostics addressed multicollinearity and heteroskedasticity. Together, these steps improved consistency, reproducibility, and interpretability.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue III, March 2026

Construct validity was addressed by closely aligning the operational definitions of the main variables with the study objectives. In the article, "talent demand" is explicitly operationalized through job-posting frequency, skill requirements, educational qualifications, work experience, and salary information, while "curriculum design" is operationalized as the mapping of empirical skill findings to course recommendations and learning objectives. This alignment ensures that the empirical indicators used in the study meaningfully represent the concepts under investigation.

Several limitations should nevertheless be acknowledged. As stated in the article, the dataset was drawn mainly from major online job portals and may therefore exclude niche recruitment channels, internal referrals, or less formal labor-market mechanisms.

The time frame was limited to 2022–2024, which provides a recent but still bounded snapshot of a rapidly evolving technological labor market. In addition, the broader institutional analysis covered benchmark universities rather than the full diversity of China's higher education system.

These limitations do not negate the study's value, but they do mean that the findings should be interpreted as a high-resolution analysis of formal online market demand rather than a complete representation of all talent dynamics. In summary, the methodology combined large-scale online recruitment data, rigorous text preprocessing, unsupervised skill clustering, embedding-based semantic modeling, and regression analysis of salary determinants to build an evidence-based picture of demand for data science talent in China.

This design was appropriate for identifying both the structural composition of the labor market and the specific competencies associated with demand and wage premiums. The next section presents the empirical findings, beginning with the overall distribution of data science job demand and then moving to skill clusters, salary determinants, and their implications for curriculum alignment.

RESULTS

This section presents the empirical findings in line with the study's objectives: to map the structure of data science talent demand in China, examine its geographic and sectoral distribution, identify the core technical and soft-skill requirements valued by employers, detect latent competency clusters, and estimate the salary premiums associated with structural and skill-related factors.

The findings are organized thematically. Section 4.1 establishes the overall market profile, Section 4.2 examines regional and sectoral concentration, Sections 4.3 and 4.4 analyze the competency architecture of demand, Section 4.5 reports the regression results on salary determinants, and Section 4.6 translates the labor-market evidence into implications for curriculum alignment. This structure mirrors the article's analytical logic and research questions.

Overview of Data Science Talent Demand in China

The results confirm that demand for data science talent in China is both substantial and structurally differentiated. The empirical dataset comprised 12,436 de-duplicated job postings collected from Zhaopin.com and 51job.com over the 2022–2024 period, capturing formal online demand for data science-related roles such as data scientist, data analyst, machine learning engineer, and big data engineer.

The study's operational definition of talent demand was based on the frequency and composition of skill requirements, educational expectations, experience levels, salary ranges, and job volume across locations and industries. The overall market profile suggests that data science has become a mainstream occupational category rather than a niche technical specialization. Salary data further indicate that these positions remain comparatively well compensated.

The article reports that most advertised salaries cluster between CNY 8,000 and CNY 30,000 per month, and that the log-transformed average salary is approximately normally distributed, suggesting a sufficiently stable wage structure for regression analysis. This finding is methodologically important because it supports the use of

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue III, March 2026

log-salary as the dependent variable in later models and substantively important because it indicates that data science talent continues to command meaningful labor-market value across multiple sectors.

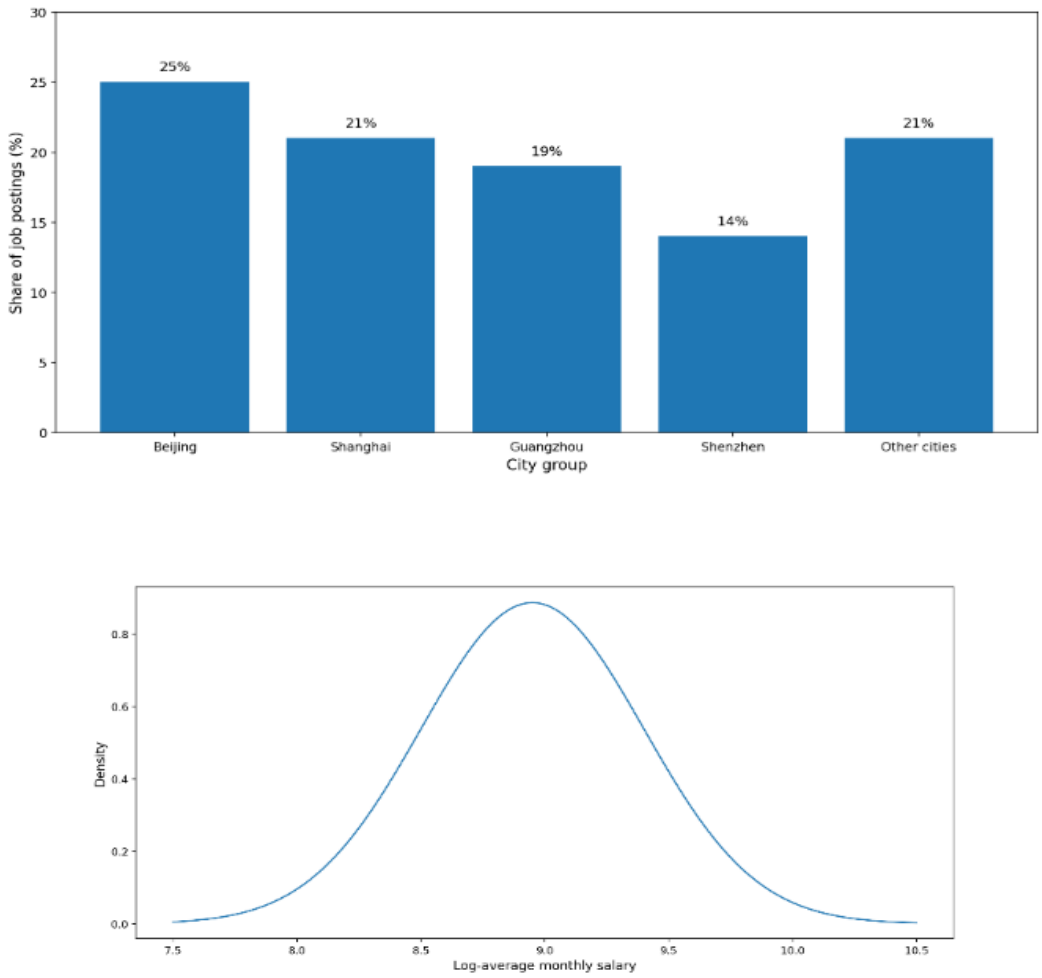

Figure 1. Overall distribution of data science job postings and log-average salary distribution.

Figure 1 shows that monthly salary rises clearly with work experience during the first five years, with the most pronounced increase occurring between three and five years; after five years, wage growth slows, although the probability of earning very high salaries continues to increase.

Table 1. Summary statistics of the 12,436 job-posting dataset (platform, city tier, sector, salary, education, experience).

Table 1. Summary statistics of the 12,436 job-posting dataset

Variable	**Category / Description**	**Summary statistic**
Dataset size	Unique job postings after cleaning and filtering	**12,436**
Observation period	January 2022 – December 2024	**36 months**
Recruitment platforms	Zhaopin.com and 51job.com	**2 major platforms**
Data type	Structured + unstructured attributes	Included

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue III, March 2026

Structured fields captured	Job title, salary range, location, company size	Included
	Industry sector, education requirement, work experience	Included
City coverage	Tier-1 cities (e.g., Beijing, Shanghai, Shenzhen, Guangzhou)	Included
	Selected Tier-2 hubs (e.g., Chengdu, Hangzhou, Wuhan)	Included
	Representative western regions (e.g., Guizhou, Chongqing)	Included
City-tier distribution	Tier-1 cities (Beijing, Shanghai, Guangzhou, Shenzhen)	**79% of postings**
	Beijing	**25%**
	Shanghai	**21%**
	Guangzhou	**19%**
	Shenzhen	**14%**
	Other cities combined	**21%**
Sectoral concentration	Internet / e-commerce	**1,872 postings (24.31%)**
	Finance / investment / securities	High concentration reported
	Computer software	High concentration reported
	Traditional sectors (retail, automobiles, FMCG, education, services)	Also represented
Salary distribution	Main monthly salary concentration	**CNY 8,000–30,000**
	Transformed salary form used for modeling	**Log-average salary**
Education field	Education requirement captured in structured data	Included
	Master's degree or above	Positive salary predictor reported
Experience field	Work experience requirement captured in structured data	Included
	Experience-linked salary progression	Increases strongly in first 5 years; slower thereafter
Company size field	Company size captured in structured data	Included
	Large company size	Positive salary predictor reported

Table 1 summarizes the core characteristics of the 12,436-job-posting dataset used in this study. The dataset was compiled from Zhaopin.com and 51job.com from January 2022 to December 2024 and includes both structured and unstructured recruitment information, such as salary range, location, industry, education, work experience,

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue III, March 2026

and textual skill requirements. The descriptive profile shows that the labor market for data science talent in China is highly concentrated in Tier-1 cities, which account for 79% of observed postings, and is particularly strong in the internet and e-commerce sector, which alone contributes 24.31% of vacancies. Salary information is concentrated mainly between CNY 8,000 and CNY 30,000 per month, while education level, company size, city tier, and work experience are all retained as important explanatory dimensions in subsequent analysis.

Geographic and Sectoral Distribution of Job Demand

A key result is the strong geographic concentration of demand for data science.Beijing, Shanghai, Guangzhou, and Shenzhen together account for 79% of the observed national demand, indicating that China's data science labor market is highly centralized in major first-tier innovation and commercial hubs.Among these cities, Beijing accounts for 25% of total demand, followed by Shanghai at 21%, Guangzhou at 19%, and Shenzhen at 14%. The findings, therefore, reveal a clear metropolitan concentration pattern consistent with agglomeration effects in finance, internet commerce, research, and digital services. Hangzhou also emerges as an important second-tier growth hub, while Guiyang is notable for its specialized role as a "big data city," indicating that regional policy and industrial positioning can shape local labor-market demand beyond the largest metropolitan centers.

Sectoral distribution exhibits a similarly uneven but interpretable structure. Demand is concentrated in the internet or e-commerce, financial investment and securities, and computer software. The article reports that the internet and e-commerce sector alone accounts for 1,872 postings, representing 24.31% of the total sample. At the same time, the results also show that demand is no longer confined to digital-native industries. Traditional sectors such as wholesale and retail, automobiles and spare parts, fast-moving consumer goods, professional services, and education and training also display significant hiring activity. This broadening of sectoral demand suggests that data science is increasingly embedded in both core digital industries and conventional sectors undergoing digital transformation.

Taken together, the geographic and sectoral findings indicate that demand for data science in China is shaped by both economic concentration and industrial diversification. First-tier cities remain dominant because they combine capital, infrastructure, advanced services, and technology ecosystems. However, the spread of demand into second-tier hubs and traditional sectors suggests that data science capabilities are diffusing across the broader economy. This pattern supports the view that talent development strategies should be differentiated by regional and industrial context rather than treated as nationally uniform.

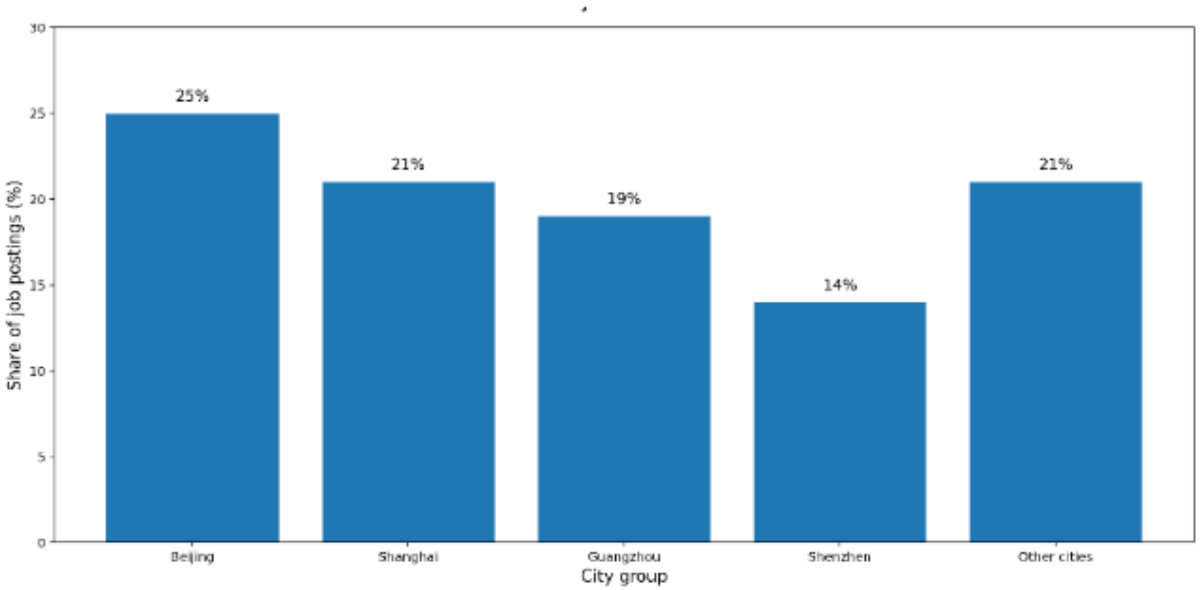

Figure 2. Regional distribution of demand for data science talent across major Chinese cities.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue III, March 2026

Figure 2 shows that demand for data science talent in China is strongly concentrated in major first-tier cities, with Beijing (25%), Shanghai (21%), Guangzhou (19%), and Shenzhen (14%) together accounting for 79% of all observed job postings, indicating a highly centralized urban labor-market structure.

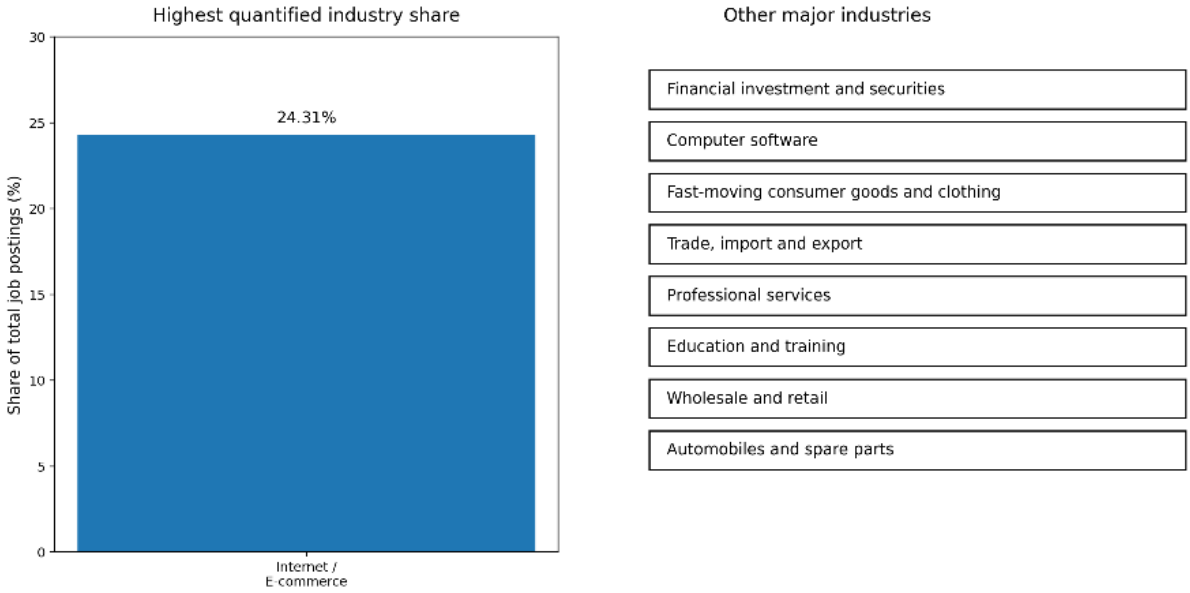

Figure 3. Distribution of demand for data science talent across major industries.

Figure 3 shows that demand for data science talent in China is led by the internet and e-commerce sector, which accounts for 1,872 postings or 24.31% of the total sample, while strong additional demand is also observed in finance, computer software, and several traditional industries undergoing digital transformation.

Table 2. Top cities and top sectors by share of data science job postings.

(a) Top cities by share of job postings

Rank	**City**	**Share of job postings (%)**	**Notes**
1	Beijing	25	Highest observed city share
2	Shanghai	21	Major national demand hub
3	Guangzhou	19	Strong concentration of demand
4	Shenzhen	14	Major technology and innovation hub
5	Other cities combined	21	Remaining national demand outside top four
	Top four cities total	**79**	Combined share of Beijing, Shanghai, Guangzhou, and Shenzhen

(b) Top sectors by share of job postings

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue III, March 2026

Rank	Sector	Share/volume reported in article	Notes
1	Internet/e-commerce	**1,872 postings (24.31%)**	Highest quantified sector share
2	Financial investment/securities	Major sector reported	High demand concentration reported
3	Computer software	Major sector reported	High demand concentration reported
4	Trade/import / export	Major sector reported	Significant demand reported
5	Professional services	Major sector reported	Significant demand reported
6	Education/training	Major sector reported	Significant demand reported
7	Wholesale / retail	Major sector reported	Traditional sector with rising demand
8	Automobiles / spare parts	Major sector reported	Traditional sector with rising demand

Table 2 shows that demand for data science jobs in China is highly concentrated both geographically and sectorally. At the city level, Beijing, Shanghai, Guangzhou, and Shenzhen together account for 79% of all observed postings, confirming a strong metropolitan concentration of demand.

At the sectoral level, the internet and e-commerce industry is the largest source of demand, accounting for 1,872 postings, or 24.31% of the total sample, while finance, computer software, and several traditional industries also show substantial hiring activity. Overall, the table indicates that demand for data science talent is shaped by both the concentration of urban innovation and the diffusion of digital transformation across multiple sectors.

Core Technical and Soft Skill Requirements

The results reveal that employer demand for data science talent is built on a hybrid competency profile that combines core technical expertise with soft skills and cross-functional capabilities. The article operationalizes talent demand by extracting explicit and implicit requirements from job postings, including technical skills such as Python, machine learning, TensorFlow, Hadoop, Spark, and SQL, as well as soft skills such as communication and teamwork. At the descriptive level, the findings show that core technical requirements center on programming, machine learning, big data tools, database management, and visualization technologies.

The salary modeling results also help identify which skills are especially valued. In the LASSO model, machine learning, Python, Hadoop/Spark, database management, deep learning, communication skills, problem-solving, visualization tools, and attitude all remain as non-zero predictors. Machine learning ($\beta = +0.143$), Python ($\beta = +0.131$), Hadoop/Spark ($\beta = +0.117$), database management ($\beta = +0.096$), and deep learning ($\beta = +0.087$) represent the most economically rewarded technical skill indicators, while communication skills ($\beta = +0.079$), problem-solving ($\beta = +0.068$), and attitude ($\beta = +0.047$) remain positively associated with salary. These results indicate that labor-market demand is not narrowly technical; instead, employers reward both technical depth and professional effectiveness.

This competency structure is consistent with the article's broader argument that data science roles increasingly require "T-shaped" professionals rather than narrowly defined technical specialists. Employers appear to value candidates who can combine algorithmic and programming competence with communication, project coordination, and applied problem-solving. The results, therefore, support a broader interpretation of employability in data science: the labor market rewards the integration of analytical capability, tool proficiency, and organizational readiness.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue III, March 2026

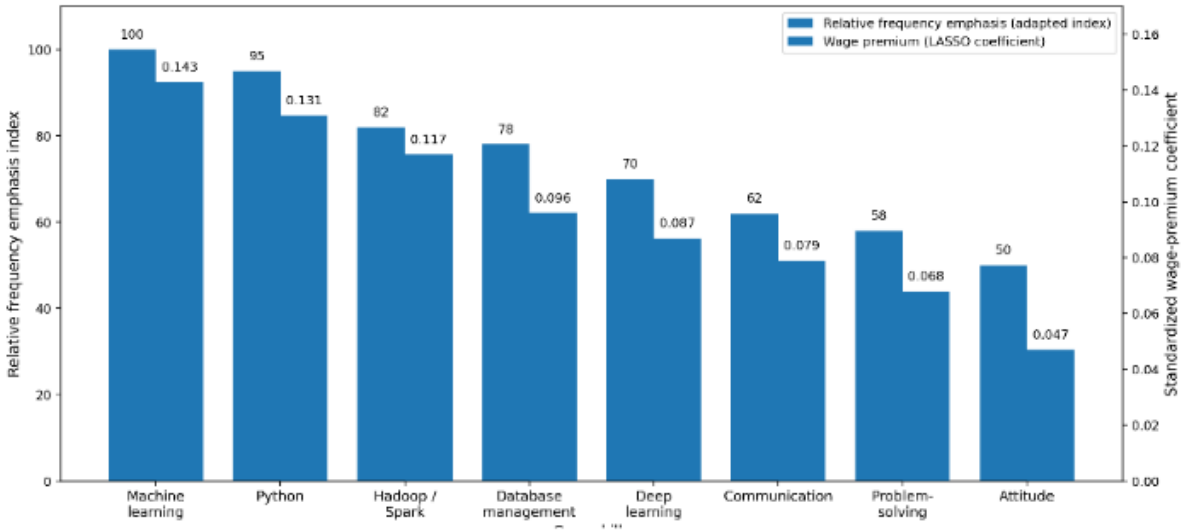

Figure 4. Frequency and wage premium of core technical and soft skills in data science postings.

Figure 4 shows that the highest-value skills in China's data science labor market combine core technical competencies and selected soft skills: machine learning, Python, Hadoop/Spark, database management, and deep learning carry the strongest wage premiums, while communication, problem-solving, and professional attitude also remain positively associated with salary, confirming that employers reward hybrid competency profiles rather than purely technical specialization.

Table 3. Core technical and professional competencies identified from job-posting analysis.

Competency category	**Competency**	**Role in job-posting analysis**	**Evidence from salary model/interpretation**
Technical	Machine learning	Core analytical and modeling requirements	Strongest positive wage premium (β = 0.143)
Technical	Python	Core programming and implementation skill	Strong positive wage premium (β = 0.131)
Technical	Hadoop / Spark	Big-data processing and distributed computing	Strong positive wage premium (β = 0.117)
Technical	Database management	Data storage, querying, and management capability	Positive wage premium (β = 0.096)
Technical	Deep learning	Advanced AI and predictive modeling capability	Positive wage premium (β = 0.087)
Professional	Communication	Ability to explain analytical results and collaborate	Positive wage premium (β = 0.079)
Professional	Problem-solving	Applied analytical reasoning in practical contexts	Positive wage premium (β = 0.068)
Professional	Professional attitude	Work readiness, reliability, and professional conduct	Positive wage premium (β = 0.047)

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue III, March 2026

Table 3 shows that employer demand for data science talent in China is built around a hybrid competency structure that combines advanced technical capabilities with professional workplace skills. Among the technical competencies, machine learning, Python, Hadoop/Spark, database management, and deep learning emerge as the most important market-valued capabilities. At the same time, communication, problem-solving, and professional attitude are also retained as positive predictors of salary, indicating that employers reward not only technical expertise but also the ability to operate effectively in organizational and collaborative environments. Overall, the table confirms that the Chinese data science labor market favors integrated competency profiles rather than narrowly specialized technical skill sets.

Skill Clusters and Emerging Competency Patterns

Beyond individual skills, the results show that employer demand is structured around recurring clusters of competencies. The article explicitly states that unsupervised learning methods were used to identify latent skill topics and cluster job requirements into meaningful patterns. These clusters include combinations such as "ML + Python," "Big-data (Hadoop/Spark)," "Cloud + DevOps," and "Cross-disciplinary project skills," indicating that employers usually recruit for bundled capability sets rather than isolated tools or keywords.

The identified patterns suggest at least four broad competency archetypes in the Chinese data science labor market. The first is a **modeling-oriented cluster** centered on machine learning, deep learning, and Python, likely associated with advanced analytics and AI roles. The second is a **big-data engineering cluster**, characterized by Hadoop, Spark, database systems, and platform-level data infrastructure. The third is an **application and integration cluster**, where visualization, data communication, and domain-linked analytical skills are emphasized. The fourth is a **cross-disciplinary collaboration cluster**, which links technical knowledge with teamwork, agile methods, project management, and communication. Evidence for this last profile is reinforced in the article's curriculum mapping discussion, which explicitly identifies "Soft Skills & Collaboration" as involving teamwork, agile methods, project management, and communication, with associated capstone and portfolio-based learning responses.

These cluster results are important because they show that the labor market organizes demand around role-specific capability packages. A curriculum aligned only to standalone programming or statistics courses would therefore be insufficient. The more relevant educational implication is that universities must help students develop integrated competency profiles that mirror actual hiring bundles. This finding strengthens the argument that data science training should be modular, applied, and interdisciplinary.

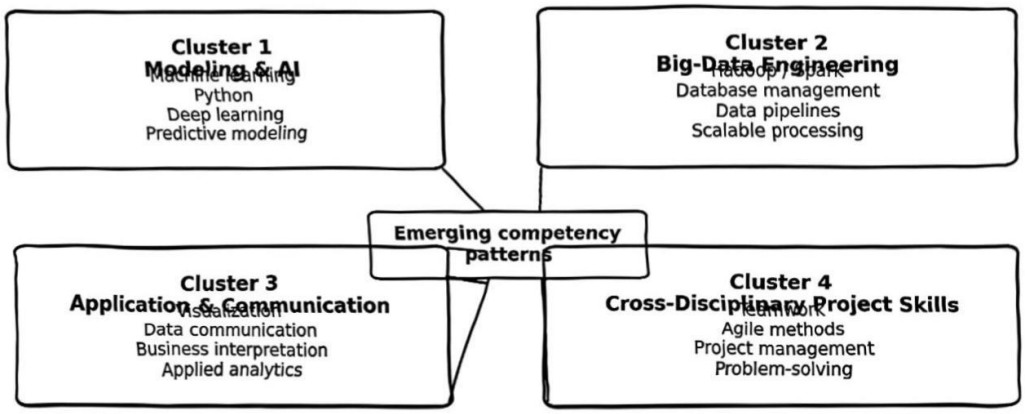

Figure 5. Cluster map of emerging competency patterns in data science job postings.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue III, March 2026

Figure 5 shows that demand for data science talent in China is organized around recurring competency bundles rather than isolated skills, with the strongest emerging patterns centered on modeling and AI, big data engineering, application and communication, and cross-disciplinary project capabilities.

Table 4. Interpreted skill clusters and associated role orientations.

Cluster	Core competency pattern	Main skills included	Interpreted role orientation	Curriculum implication
Cluster 1	**Modeling and AI**	Machine learning, Python, deep learning, predictive modeling	Data scientist/machine learning specialist / AI analyst	Strengthen applied machine learning, model building, Python programming, and AI project labs
Cluster 2	**Big-data engineering**	Hadoop, Spark, database management, data pipelines, scalable processing	Big-data engineer/data engineer/analytics infrastructure specialist	Add distributed computing, database systems, data engineering, and large-scale processing modules
Cluster 3	**Application and communication**	Visualization, data communication, business interpretation, applied analytics	Data analyst / business intelligence analyst / decision-support analyst	Embed visualization, storytelling with data, decision-making cases, and applied industry analytics
Cluster 4	**Cross-disciplinary project skills**	Teamwork, agile methods, project management, problem-solving, and communication	Project-oriented data professional/analytics consultant / interdisciplinary team member	Use team capstones, agile project work, portfolio assessment, and collaborative problem-based learning

Table 4 shows that employer demand for data science talent in China is structured around four interpretable competency clusters rather than isolated standalone skills. The first two clusters emphasize technical depth in modeling, AI, and big-data engineering, while the latter two reflect application-facing and cross-disciplinary professional roles that require communication, teamwork, and project execution. This pattern indicates that the labor market values role-based combinations of competencies, suggesting that universities should design curricula around integrated learning pathways rather than disconnected subject silos.

Salary Determinants by Region, Education, and Firm Characteristics

The regression analysis shows that structural factors are the strongest determinants of salary, although skill factors also generate meaningful wage premiums. In the LASSO model, the largest standardized coefficients are associated with city tier ($\beta = +0.284$), managerial job type ($\beta = +0.226$), and large company size ($\beta = +0.194$). This indicates that salary outcomes are strongly conditioned by the jobis location, the level of responsibility it entails, and the employer's scale. These findings align with the broader descriptive results, which show concentration in first-tier cities and major corporate sectors.

Educational attainment also matters. A master's degree or above is retained as a positive predictor ($\beta = +0.110$), indicating that advanced academic qualifications continue to yield salary advantages in the data science labor market. At the same time, several technical competencies carry independent premiums even after controlling for structural variables. Machine learning, Python, Hadoop/Spark, database management, and deep learning remain positively associated with higher salaries, confirming that specialized analytical and engineering capabilities have market value beyond general job characteristics. The out-of-sample performance of the LASSO model is

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue III, March 2026

also reasonably strong, with $R^2 = 0.612$ and MAE = 0.184 in log-salary units, suggesting that the model captures substantial variance in wage outcomes while maintaining an acceptable level of predictive error.

An especially noteworthy result is that soft skills remain salary-relevant rather than merely decorative requirements in recruitment language. Communication skills, problem-solving, and professional attitude persist as non-zero predictors even after regularization. This finding implies that employers do not treat these attributes as optional complements to technical skill; instead, they attach measurable economic value to them. The salary models therefore reinforce the broader interpretation that China's data science labor market rewards a multidimensional competency structure combining location, organizational context, qualifications, technical specialization, and professional skills.

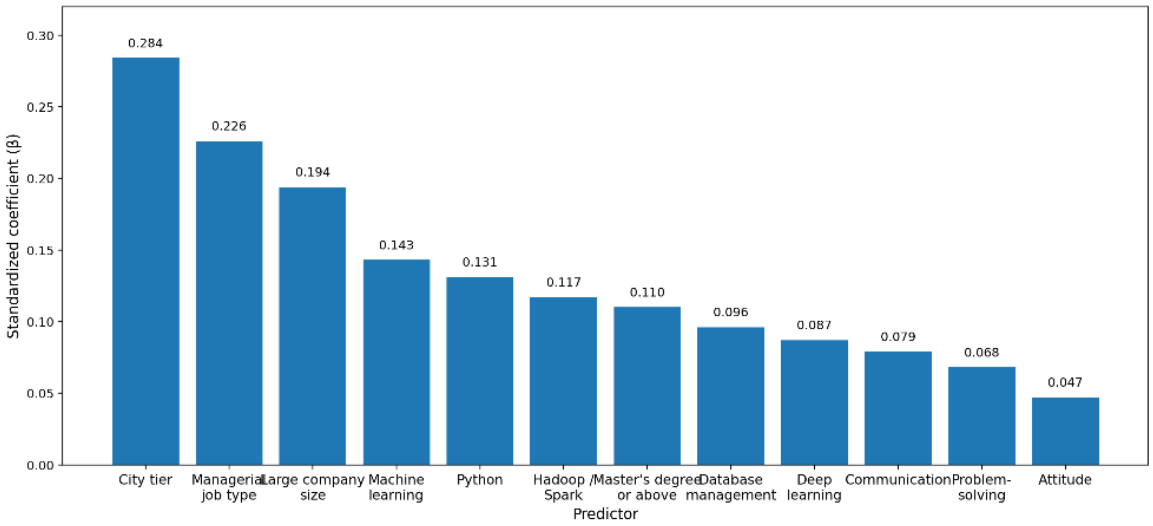

Figure 6. Standardized coefficients of non-zero predictors from the LASSO salary model.

Figure 6 shows that the strongest predictors of salary in China's data science labor market are structural factors—especially city tier, managerial job type, and large company size—followed by high-value technical skills such as machine learning, Python, and Hadoop/Spark, while education level and selected soft skills also retain positive wage effects.

Table 5. Salary determinants from regression analysis (city tier, job type, company size, education, and skills).

Predictor category	**Predictor**	**Standardized coefficient (β)**	**Direction of effect**	**Interpretation**
Location	City tier	**0.284**	Positive	Jobs in higher-tier cities are associated with higher salaries
Job structure	Managerial job type	**0.226**	Positive	Managerial or higher-responsibility roles receive higher pay
Firm characteristic	Large company size	**0.194**	Positive	Larger firms tend to offer stronger salary premiums
Technical skill	Machine learning	**0.143**	Positive	Advanced modeling capability is strongly rewarded
Technical skill	Python	**0.131**	Positive	Programming proficiency is a major salary-enhancing skill

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue III, March 2026

Technical skill	Hadoop / Spark	**0.117**	Positive	Big-data platform skills carry substantial market value
Education	Master's degree or above	**0.110**	Positive	Advanced educational qualification increases salary potential
Technical skill	Database management	**0.096**	Positive	Data storage and query capabilities improve wage outcomes
Technical skill	Deep learning	**0.087**	Positive	Specialized AI capability contributes to higher salaries
Professional skill	Communication	**0.079**	Positive	Communication ability has measurable economic value
Professional skill	Problem-solving	**0.068**	Positive	Applied analytical reasoning supports higher wages
Professional skill	Attitude	**0.047**	Positive	Professional attitude remains a retained salary predictor

Model performance

Model	Outcome variable	R^2	MAE (log-salary units)
LASSO regression	Log-average salary	**0.612**	**0.184**

Table 5 shows that salary outcomes in China's data science labor market are driven primarily by structural factors, especially city tier, managerial job type, and company size, which have the largest standardized coefficients in the regression model. At the same time, several technical skills, particularly machine learning, Python, Hadoop/Spark, database management, and deep learning, remain significant positive predictors of salary, while education level and selected professional skills, such as communication, problem-solving, and attitude, also contribute positively. Overall, the table indicates that wages in this labor market are shaped by a combination of regional context, organizational characteristics, formal qualifications, and hybrid technical-professional competencies.

Implications of Labor-Market Findings for Curriculum Alignment

The final set of results concerns translating labor-market evidence into curriculum implications. The article explicitly defines curriculum design in operational terms as the mapping of empirical findings on skills onto course recommendations, practical modules, and learning objectives for data science programs in higher education. It further indicates that the education-side analysis compared official university syllabi and program outcomes against industry requirements and developed a quantitative skill-matching index to identify curricular strengths and gaps. Although the full curriculum comparison is part of the broader article, the labor-market results alone already yield several clear signals of alignment.

First, the prominence of machine learning, Python, database systems, and Hadoop/Spark indicates that core curricula should move beyond introductory statistics and general computing into applied analytical toolchains and scalable data infrastructure. Second, the persistence of communication, problem-solving, teamwork, and project management in both descriptive and salary analyses suggests that soft skills should not be peripheral; they should be embedded into studio-based courses, capstone projects, agile collaboration exercises, and portfolio assessment. Third, the regional and sectoral concentration of demand implies that curriculum models may need localized specialization tracks, for example, in finance-oriented analytics, internet-platform data science, or manufacturing-linked intelligent systems, depending on institutional and regional context. Evidence for this policy-to-programmatic logic is evident in the article's framework, which links national directives to

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue III, March 2026

institutional actions and measurable KPIs, including AI/cloud-related course offerings, internship-to-hire conversion rates, and employer satisfaction indicators.

Overall, the results indicate that curriculum alignment should be guided by empirically observed demand bundles rather than static disciplinary assumptions. A market-responsive curriculum in data science should therefore integrate advanced technical modules, interdisciplinary applications, collaborative industry-oriented learning, and regionally sensitive program design. These implications provide the bridge from labor-market analysis to the discussion of how universities and policymakers can respond more strategically to China's evolving data science talent needs.

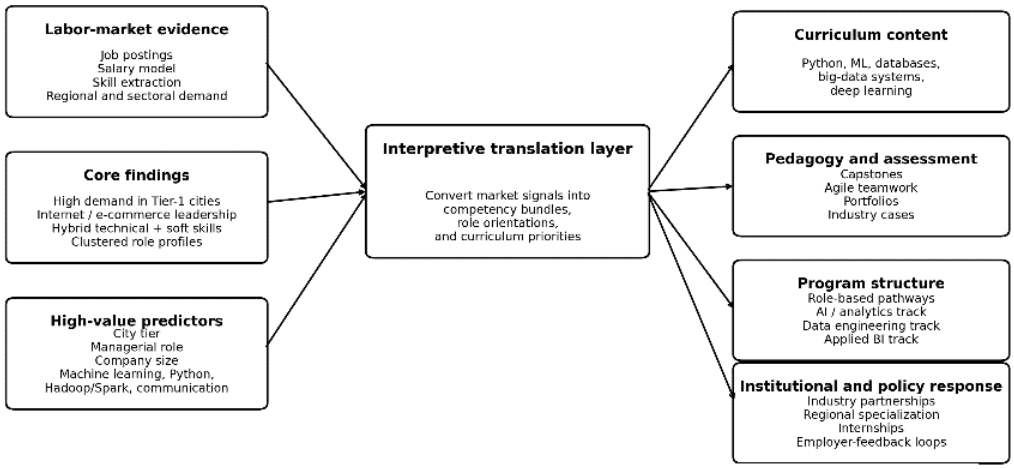

Figure 7. Evidence-to-curriculum alignment framework derived from labor-market findings.

Figure 7 illustrates how labor-market evidence—derived from job postings, salary determinants, skill clusters, and regional-sectoral demand can be translated into curriculum content, pedagogy, program pathways, and institutional responses, thereby providing an evidence-based framework for aligning data science education with evolving market needs.

Table 6. Labor-market findings and corresponding curriculum implications.

Labor-market finding	**Evidence from article**	**Curriculum implication**	**Suggested educational response**
Strong demand concentration in Tier-1 cities	Beijing, Shanghai, Guangzhou, and Shenzhen account for 79% of postings	Curriculum should reflect advanced metropolitan labor-market expectations	Strengthen industry-linked modules, internships, and case-based learning tied to major digital-economy hubs
Internet/e-commerce is the leading demand sector	Internet/e-commerce contributes 1,872 postings (24.31%)	Programs should prepare students for platform-based, data-intensive industries	Include e-commerce analytics, digital platform data, recommender systems, and user-behavior analysis
Data science demand is spreading into traditional industries	Demand also appears in finance, software, retail, automobiles, education, and services	Curriculum should not be designed only for tech firms	Develop sector-specific electives such as finance analytics, retail analytics, education data science, and intelligent industry applications

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue III, March 2026

Technical competencies dominate employer requirements	High-value technical skills include machine learning, Python, Hadoop/Spark, database management, and deep learning	Core curriculum must move beyond basic programming and descriptive statistics	Add applied modules in machine learning, Python programming, distributed systems, database analytics, and AI implementation
Soft skills are positively associated with salary	Communication, problem-solving, and attitude remain retained positive predictors in the LASSO model	Soft skills should be integrated into the core curriculum rather than treated as optional	Embed communication, teamwork, agile collaboration, project management, and reflective professional practice into assessment design
Employer demand is structured around competency bundles	Clusters include modeling/AI, big-data engineering, application/communication, and cross-disciplinary project skills	Curriculum should be organized around role-oriented pathways instead of isolated courses	Introduce specialization tracks such as AI and machine learning, data engineering, business analytics, and applied interdisciplinary analytics
Structural factors strongly shape wage outcomes	City tier, managerial role, and company size have the largest salary coefficients	Curriculum should prepare graduates not just technically, but also for high-responsibility organizational contexts	Include leadership exposure, project coordination, workplace simulation, and strategic decision-support scenarios
Advanced qualifications still yield wage advantages	Master's degree or above is a positive salary predictor	Undergraduate curriculum should build progression pathways toward advanced study and specialization	Create research pathways, advanced electives, honors tracks, and postgraduate articulation opportunities
Labor-market value depends on practical usability of skills	Wage premiums remain attached to applied technical skills	Curriculum should emphasize authentic, practice-oriented learning	Use capstone projects, real datasets, coding labs, portfolio-based assessment, and employer-validated tasks
Market-responsive curriculum requires institutional adaptation	Article links labor evidence to institutional and policy actions	Curriculum reform should be continuous and data-informed	Establish employer advisory panels, periodic curriculum review using labor-market analytics, and feedback loops with industry

Table 6 shows that the major labor-market signals identified in the study can be translated directly into curriculum design priorities for data science education. The findings indicate that curriculum reform should address not only high-value technical skills such as machine learning, Python, database systems, and big data tools, but also soft skills, role-based competency bundles, and sectoral variations that shape actual employability. Overall, the table demonstrates that effective curriculum alignment requires an integrated response involving course content, pedagogy, specialization pathways, and institutional collaboration with industry.

DISCUSSION

This section interprets the findings in relation to the study's three objectives: to identify the structure of data science talent demand in China, to explain the role of technical and contextual factors in shaping employer

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue III, March 2026

requirements and salary outcomes, and to derive implications f1or curriculum alignment in higher education. The discussion is organized around six interrelated themes. First, it interprets the overall structure of demand. Second, it compares the findings with prior literature. Third, it explains the regional, sectoral, and organizational variation observed in the results. Fourth, it examines how labor-market signals can be translated into curriculum alignment. Fifth, it outlines practical and policy implications. Finally, it acknowledges the study's limitations. This structure preserves the article's logic while sharpening the contribution for journal publication.

Interpreting the Structure of Data Science Talent Demand in China

A central finding of this study is that demand for data science talent in China is not organized around isolated technical skills but around layered, bundled competency structures. The evidence from 12,436 job postings shows that employers consistently seek combinations of programming, machine learning, big data infrastructure, database capabilities, and applied professional skills rather than narrowly defined single-skill profiles. This finding supports the study's interpretation that data science has matured into a hybrid occupational field in which analytical capability, computational implementation, and workplace effectiveness must coexist.

This pattern is consistent with the interdisciplinary definition of data science advanced in the article and in foundational literature, where the field is conceptualized as the integration of statistics, computing, and domain knowledge rather than as a branch of programming alone [1], [2]. It also supports the article's argument that the labor market increasingly values "T-shaped" professionals who combine technical depth with cross-functional adaptability. In theoretical terms, the finding aligns with human capital and labor-market matching perspectives: the market rewards not just possession of abstract knowledge, but possession of skill bundles that can be productively matched to organizational tasks and sectoral needs.

Another important interpretation is that demand in China has moved beyond a purely exploratory digital-economy phase into a more institutionalized talent market. The scale of demand, the stability of wage modeling, and the identifiable clustering of competencies suggest that data science roles are now embedded across multiple sectors rather than confined to experimental or elite technology teams. This broadening supports the article's broader claim that data science should be understood as an enabling capability of economic transformation rather than as a niche technical occupation.

Comparison with Prior Literature

The present findings are broadly consistent with earlier studies that identify programming, machine learning, databases, and big data technologies as the backbone of demand for data science [7]–[10]. The strong market signals for Python, machine learning, Hadoop/Spark, database management, and deep learning reinforce prior work arguing that technical competence in scalable data systems and predictive analytics remains central to employability in data-intensive occupations. The current results extend this literature by showing not only that such skills are frequently requested, but that several of them also carry independent salary premiums after controlling for structural labor-market variables.

At the same time, the findings also support more recent literature emphasizing hybridization in digital labor markets. The fact that communication skills, problem-solving, and attitude remained non-zero predictors in the regularized salary model suggests that soft skills are not merely symbolic additions in job advertisements. Rather, they have measurable economic relevance. This is important because some earlier labor-market studies focused heavily on technical keyword extraction and underemphasized the organizational and interpersonal dimensions of digital work. The present study, therefore, strengthens the argument that employers increasingly demand integrative professional capability rather than isolated technical proficiency.

The findings also differ from a narrower strand of prior research that treats demand for digital skills as relatively uniform across occupations or locations. In this study, demand is clearly stratified by city tier, company size, managerial level, and industry. This suggests that the value of data science skills is context-dependent rather than universally priced. In comparison with previous studies based on smaller or cross-sectional samples, the present analysis provides stronger evidence that regional development level and organizational characteristics

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue III, March 2026

substantially mediate the wage returns to data science competencies. In this sense, the study offers a more contextualized interpretation of digital skill demand than research relying solely on aggregate frequency analysis.

Explaining Regional, Sectoral, and Organizational Variations

One of the most important results is the strong concentration of demand in first-tier cities, especially Beijing, Shanghai, Guangzhou, and Shenzhen, which together account for 79% of the observed postings. This concentration is not surprising, but it is analytically significant. It indicates that the Chinese data science labor market is heavily shaped by metropolitan agglomeration effects, including digital infrastructure, venture capital concentration, high-value service sectors, research ecosystems, and dense pools of complementary firms. The finding, therefore, supports the article's argument that labor demand must be interpreted within the broader geography of digital economic transformation rather than as a nationally uniform phenomenon.

At the sectoral level, the dominance of internet and e-commerce, financial services, and software is also expected, but the spread of demand into retail, automobiles, consumer goods, professional services, and education is especially noteworthy. This suggests that data science capability is diffusing into conventional industries as part of digital transformation, automation, and platform-based decision-making. In other words, data science is not simply expanding within the digital sector; it is increasingly becoming a transversal capability embedded across sectoral boundaries. This helps explain why the labor market rewards both technical depth and business-facing soft skills: many data science roles are now positioned at the interface between analytical systems and operational decision environments.

Organizationally, the salary models indicate that large firms and managerial positions offer the strongest wage premiums, alongside city tier and postgraduate qualifications. This suggests that organizational maturity and scale condition the market value of data science talent. Larger firms may possess the infrastructure, data assets, and implementation capacity to extract greater value from advanced analytics, which in turn enables them to pay more for talent. Managerial or leadership-linked roles likely carry higher wages because they require not only technical execution but also coordination, strategy, and cross-functional influence. These findings are theoretically consistent with the view that wages reflect both human capital attributes and the complexity of the organizational environments in which those attributes are deployed.

An interesting result is that a master's degree remains a positive predictor of salary, but it does not dominate the model as strongly as structural variables such as city tier or company size. This may indicate that while advanced qualifications matter, labor-market rewards in data science are shaped just as much by where and how skills are applied as by formal educational credentials alone. That interpretation is important for higher education because it suggests that employability depends not only on degree level but also on the practical, contextual usability of graduates' capabilities.

From Job-Market Signals to Curriculum Alignment

The results have direct implications for curriculum alignment. First, the prominence of Python, machine learning, databases, Hadoop/Spark, and deep learning indicates that contemporary data science curricula must go beyond introductory programming and descriptive statistics. Universities need to design curricula that reflect the actual architecture of market demand, including scalable data systems, applied modeling, cloud-compatible workflows, and real-world data engineering tasks. The article's curriculum mapping logic strongly supports this interpretation by defining curriculum design as the translation of labor-market evidence into course content, skill outcomes, and learning pathways.

Second, the finding that communication, problem-solving, teamwork, and attitude remain economically relevant means that soft skills should not be treated as supplementary or generic graduate attributes. Instead, they should be embedded structurally into data science programs through capstone projects, interdisciplinary teamwork, industry-linked case studies, agile project environments, and portfolio assessment. The article explicitly links "Soft Skills & Collaboration" to teamwork, agile methods, project management, and communication, with corresponding educational responses such as team capstones and portfolio-based evaluation. This indicates that

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue III, March 2026

the labor market is rewarding forms of professional integration that curriculum designers must intentionally cultivate.

Third, the cluster-based results imply that curriculum design should be modular and pathway-oriented. Because employers recruit for competency bundles rather than standalone skills, universities may need to offer differentiated tracks such as machine learning and AI analytics, big-data engineering, business analytics and visualization, or applied interdisciplinary data science. A curriculum built only around disconnected courses will not adequately mirror market demand. A stronger design would integrate technical modules, applied labs, and industry problems into coherent role-oriented learning sequences.

Policy and Institutional Implications

The study also has implications beyond curriculum design. For universities, the findings support stronger industry-academia coordination in the design of syllabi, internships, microcredentials, and applied laboratories. Institutions should not revise programs solely through internal disciplinary debate; they should build continuous labor-market monitoring mechanisms that track evolving skill bundles, regional demand shifts, and salary signals. This would allow curriculum revision to become evidence-based and iterative rather than reactive and episodic. The article's broader framework, which links national policy, institutional response, and measurable KPIs, is highly relevant here.

For policymakers, the regional concentration of demand suggests a need for differentiated talent strategies. First-tier cities may require advanced specialization pipelines and industry-facing postgraduate development, while emerging regions may benefit more from capacity-building models that emphasize applied analytics, local industrial use cases, and distributed digital talent development. Policy should therefore consider not only the expansion of degree programs but also the regional fit of program design, talent-retention mechanisms, and the digital infrastructure needed to absorb graduates productively. The article's treatment of national and local digital policies supports this interpretation by showing that labor demand is mediated through both macro strategy and regional implementation.

For employers and professional bodies, the study suggests that recruitment data can serve as a strategic feedback mechanism for collaboration with higher education. Firms can contribute to curriculum co-design, problem-based learning, internship structuring, and competency validation. This is particularly important in a field like data science, where the market evolves more quickly than formal curricular approval cycles. For researchers, the study demonstrates the value of combining job-posting analytics, clustering, and wage modeling in a single framework. Future work can build on this by comparing countries, tracking longitudinal change, or linking recruitment signals to actual graduate outcomes.

Limitations of the Study

Several limitations should be acknowledged. First, the dataset was derived primarily from major public job portals and therefore reflects formal, platform-mediated recruitment rather than the entire labor market. Second, the study covers 2022–2024, providing a recent yet bounded snapshot of a rapidly evolving field. Third, job advertisements represent employer demand signals rather than confirmed hiring outcomes, so the results should be interpreted as market expectations rather than completed matches. Fourth, although the study derives meaningful curriculum implications from labor-market evidence, the deeper institutional curriculum comparison is broader than what is fully foregrounded here. These limitations do not weaken the core contribution, but they define the scope of inference and point to the need for future longitudinal, multi-source, and comparative studies.

In summary, the discussion shows that demand for data science talent in China is large-scale, geographically concentrated, sectorally diffuse, and structured around hybrid competency bundles that combine technical depth with professional effectiveness. The study extends prior literature by showing that these skill bundles are context-dependent and economically differentiated, and by demonstrating how labor-market evidence can be translated into curriculum alignment. These findings set up the final section, which consolidates the article's theoretical contribution, practical significance, and recommendations for future research.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue III, March 2026

CONCLUSION

Summary of Key Findings

This study addressed a central challenge in China's digital transformation agenda: the growing mismatch between rapidly evolving labor-market demand for data science talent and the ability of higher education curricula to respond in a timely, evidence-based manner. To address this problem, the study examined the structure of demand for data science talent in China through a large-scale analysis of online job postings and interpreted the results in relation to curriculum alignment. Specifically, it aimed to identify the dominant demand patterns in the labor market, determine the main technical and contextual factors associated with employer requirements and salary outcomes, and derive implications for curriculum reform in higher education. These objectives were pursued using a dataset of 12,436 job postings collected from major Chinese recruitment platforms between 2022 and 2024 and analyzed through text mining, clustering, and regression-based modeling.

The findings show that demand for data science talent in China is substantial, geographically concentrated, and increasingly structured around hybrid competency bundles rather than isolated technical skills. Demand is heavily concentrated in major first-tier cities such as Beijing, Shanghai, Guangzhou, and Shenzhen, while sectorally it is led by internet and e-commerce, finance, and software, with growing diffusion into more traditional industries. The study further demonstrates that employer demand centers on a combination of programming, machine learning, database capabilities, big data platforms, and soft skills such as communication and problem-solving. Salary outcomes are strongly influenced by city tier, company size, managerial role, educational level, and specific technical competencies, indicating that the economic value of data science talent is shaped by both skill possession and contextual labor-market conditions. These findings collectively suggest that curriculum alignment must be based on empirically observed demand structures rather than static academic assumptions.

Theoretical Contribution

This study contributes to the literature in several important ways. First, it fills a clear research gap by providing large-scale empirical evidence on demand for data science talent in China, using job-posting analytics over a recent multi-year period. Prior studies have often discussed data science education, skill shortages, or digital labor demand in conceptual terms or through relatively limited empirical samples. By contrast, this study integrates labor-market intelligence, skill extraction, competency clustering, and salary modeling into a single analytical framework. In doing so, it moves beyond descriptive accounts of talent shortages and offers a more systematic understanding of how demand for data science is structured across regions, sectors, and organizational contexts.

Second, the study contributes theoretically by reinforcing and extending the interdisciplinary view of data science as a field shaped by the interaction of technical, organizational, and institutional forces. The findings support the relevance of human capital and labor-market matching perspectives by showing that wages and employability are not determined by technical skills alone, but by the fit between bundled competencies and the environments in which they are applied. The results also strengthen the idea that digital occupations should be analyzed as hybrid roles situated within broader economic transformation rather than as isolated technical categories. Most importantly, the study advances a curriculum-alignment perspective that treats job postings not simply as descriptive labor-market data but as a dynamic source of evidence for educational redesign. This creates a stronger conceptual bridge between computational labor-market analysis and higher education reform.

Practical Contribution for Universities and Policymakers

From a practical perspective, the study provides clear implications for universities, policymakers, and industry stakeholders. For universities, the results indicate that data science curricula should be redesigned around role-relevant competency bundles rather than isolated subject silos. Programs should strengthen training in Python, machine learning, database systems, big-data tools, and scalable computing environments, while simultaneously embedding communication, teamwork, problem-solving, and project management into the core curriculum. This means that curriculum reform should not be limited to adding technical electives; it should include capstone

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue III, March 2026

projects, interdisciplinary labs, industry-linked assignments, and authentic problem-based learning environments that reflect the structure of actual labor-market demand.

For policymakers, the findings suggest that digital talent development strategies should be regionally differentiated and closely linked to industrial priorities. First-tier cities may require advanced specialist pipelines and stronger integration with postgraduate research, whereas emerging regions may benefit more from applied analytics tracks, localized industry partnerships, and capacity-building initiatives. Policy frameworks should therefore move beyond broad expansion of data science programs and focus instead on curriculum quality, practical readiness, employer engagement, and graduate absorption into regional economies. For industry stakeholders, the study highlights the value of sustained collaboration with higher education institutions in curriculum co-design, internship development, project supervision, and competency validation. In practical terms, job-market intelligence can serve as a recurring feedback mechanism to support a more adaptive and evidence-driven education ecosystem.

Study Limitations

Several limitations should be acknowledged. First, the dataset was drawn primarily from major public recruitment platforms, which means the study captures formal, platform-mediated demand rather than the full spectrum of hiring activity, including internal recruitment, headhunting, and informal channels. Second, the analysis covers the 2022–2024 period and therefore provides a recent but still bounded view of a rapidly evolving labor market. Third, job advertisements reflect employer demand signals rather than confirmed hiring outcomes or actual on-the-job performance, so the findings should be interpreted as indicators of expected competencies rather than direct evidence of realized employment matches. Finally, although strong curriculum implications are derived from labor-market evidence, the broader institutional and educational dimensions are not fully foregrounded in this paper. These limitations primarily affect generalizability rather than the internal logic of the findings.

Recommendations for Future Research

Future research should extend this work in at least four directions. First, longitudinal studies should track how skill demand evolves over longer periods, especially as generative AI, automation, and cloud-native analytics reshape the occupational structure of data science roles. Second, comparative studies across countries or regions would help determine whether the demand patterns observed in China are nationally specific or part of wider global convergence in data science labor markets. Third, future research should integrate job-posting analytics with graduate outcome data, employer interviews, and curriculum audits to test more directly whether curriculum alignment improves employability and wage outcomes. Fourth, more granular analysis of sector-specific pathways, such as finance analytics, manufacturing intelligence, educational data science, or healthcare AI, would help universities design more specialized and context-sensitive training models.

Closing Statement

In conclusion, this study demonstrates that demand for data science talent in China is both a labor-market phenomenon and an educational challenge. The value of the study lies not only in identifying which skills employers seek, but in showing how those signals can be translated into a more responsive, evidence-based curriculum framework for higher education. As digital transformation continues to reshape economic activity, the ability of universities and policymakers to interpret labor-market intelligence and act on it will become increasingly important. A data-driven approach to curriculum alignment is therefore not simply an academic exercise; it is a strategic requirement for building a future-ready talent ecosystem in China.

REFERENCES

1. W. S. Cleveland, “Data science: An action plan for expanding the technical areas of the field of statistics,” *International Statistical Review*, vol. 69, no. 1, pp. 21–26, 2001.
2. V. Dhar, “Data science and prediction,” *Communications of the ACM*, vol. 56, no. 12, pp. 64–73, 2013.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue III, March 2026

3. W. van der Aalst, "Data science in action," in *Process Mining*, Berlin, Germany: Springer, 2016, pp. 3–23.
4. L. Cao, "Data science: A comprehensive overview," *ACM Computing Surveys*, vol. 50, no. 3, pp. 1–42, 2017.
5. I. Y. Song and Y. Zhu, "Big data and data science: What should we teach?," *Expert Systems*, vol. 33, no. 4, pp. 364–373, 2016.
6. D. W. Xia and Z. L. Zhang, "On training mode of big-data talents in the age of data technology," *Journal of Southwest China Normal University (Natural Science Edition)*, vol. 41, no. 9, pp. 191–196, 2016.
7. Y. M. Chen, "On teaching environment and teaching mode based on 'Internet +'," *Journal of Southwest China Normal University (Natural Science Edition)*, vol. 41, no. 3, pp. 228–232, 2016.
8. Y. Y. Zhu and Y. Xiong, "Training data scientists in the era of big data," *Big Data Research*, vol. 2, no. 3, pp. 106–112, 2016.
9. L. B. Wu, "Cultivating big data talent by combining various disciplines and utilizing multiple resources," *Big Data Research*, vol. 2, no. 5, pp. 89–94, 2016.
10. K. C. C. Chan and T. T. He, "Data science: The demand and development of talents," *Big Data Research*, vol. 2, no. 5, pp. 95–106, 2016.
11. J. Chen, "Research and application of association rules algorithm in online recruitment system," M.S. article, Xi'an University of Science and Technology, Xi'an, China, 2009.
12. T. Cao, "Research on the application of data mining in employee online recruitment," M.S. article, Jinan University, Guangzhou, China, 2009.
13. J. Wu, "Analysis of occupation technique and ability in network engineering using recruitment information," in *Proc. 9th Int. Conf. Fuzzy Systems and Knowledge Discovery (FSKD)*, 2012, pp. 1396–1400.
14. Y. Zhang, W. Zhao, H. Bao, Y. Li, and K. Zhou, "Data mining of online recruitment information based on K-means and correlation analysis," *Software Engineering*, vol. 20, no. 5, pp. 10–14, 2017.
15. C. Liu, "Research on recruitment demand information for data jobs," M.S. article, Lanzhou University of Finance and Economics, Lanzhou, China, 2019.
16. H. Tang, "Research on talent demand and talent training approach of educational technology subject based on text mining," M.S. article, Beijing University of Posts and Telecommunications, Beijing, China, 2019.
17. C. Priyadarshini, S. Sreejesh, and M. R. Anusree, "Effect of information quality of employment website on attitude toward the website," *International Journal of Manpower*, vol. 38, no. 1, pp. 104–124, 2017.
18. R. S. Baker and P. S. Inventado, "Educational data mining and learning analytics," in *Learning Analytics*, New York, NY, USA: Springer, 2014, pp. 61–75.
19. A. P. Ayala, "Educational data mining: A survey and a data mining-based analysis of recent works," *Expert Systems with Applications*, vol. 41, no. 4, pp. 1432–1462, 2014.
20. D. J. Hand and N. M. Adams, "Data mining," *Wiley StatsRef: Statistics Reference Online*, pp. 1–7, 2014.
21. H. Qing, N. Li, W. Luo, and Z. Shi, "Overview of machine learning algorithms under big data," *Pattern Recognition and Artificial Intelligence*, vol. 27, no. 4, pp. 327–336, 2014.
22. S. Zhou, Z. Xu, and X. Tang, "Method for determining the optimal number of clusters in K-means algorithm," *Computer Applications*, vol. 30, no. 8, pp. 1995–1998, 2010.
23. Y. Wu, "Overview of clustering algorithms," *Computer Science*, vol. 42, no. 6A, pp. 491–499, 2015.
24. F. Å. Nielsen, *Data Mining Using Python*. 2017.
25. V. G. Nair, *Getting Started with Beautiful Soup*. Birmingham, U.K.: Packt Publishing, 2014.
26. G. Zaccone, *Python Parallel Programming Cookbook*. Birmingham, U.K.: Packt Publishing, 2015.
27. The State Council of the People's Republic of China, *New Generation Artificial Intelligence Development Plan*. Beijing, China, 2017.
28. The State Council of the People's Republic of China, *The 14th Five-Year Plan for Digital Economy Development*. Beijing, China, 2021.
29. National Development and Reform Commission of the People's Republic of China, *Implementation Plan for the Eastern Data and Western Computing Project*. Beijing, China, 2022.
30. Ministry of Education of the People's Republic of China, *Statistical Bulletin on National Education Development 2023*. Beijing, China, 2023.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue III, March 2026

31. Ministry of Education of the People's Republic of China, *Guidelines on Big Data and Artificial Intelligence Teaching in Universities*. Beijing, China, 2023.
32. China Big Data Industry Alliance, *China Big Data Industry Development Report 2023*. Beijing, China, 2023.