ISSN: 2473-4901
v12 n6945

Longitudinal Change in Data Science and Business Intelligence Analyst Competencies: Latent Dirichlet Allocation and Topic Prevalence in O*NET Data

Loreen Powell
lpowell@marywood.edu
Marywood University
Scranton, PA, 18509, USA

Zachary Brunner
zjbrunner@maryu.marywood.edu
Marywood University
Scranton, PA, 18509, USA

Carl M. Rebman, Jr.
carlr@sandiego.edu
University of San Diego
San Diego, CA, 92110-2492, USA

Richard Clarke
richardclarke@sandiego.edu
University of San Diego
San Diego, CA, 92110-2492, USA

Steven LevKoff
slevkoff@sandiego.edu
University of San Diego
San Diego, CA, 92110-2492, USA

Hayden Wimmer
hwimmer@georgiasouthern.edu
Georgia Southern University
Savannah, GA 31419, USA

Theodocia Zayac
tlzayac@maryu.marywood.edu
Marywood University
Scranton, PA, 18509, USA

Abstract

The rapid growth of artificial intelligence (AI), cloud computing, and digital transformation has increased the importance of understanding how analytics occupations evolve. This study conducted a longitudinal text-mining analysis of Occupational Information Network (O*NET) data published from 2016 through 2026 to examine documented competency requirements for Business Intelligence (BI) Analysts and Data Scientists. Occupational descriptions, tasks, detailed work activities, software requirements, and competency records were analyzed using frequency comparisons and Latent Dirichlet Allocation (LDA) topic modeling. The final corpus contained 4,186 records: 3,645 for BI Analysts across 11 annual releases and 541 for Data Scientists across six annual releases. BI Analysts retained 17 substantively consistent core tasks while documented technology or software requirements increased from 146 to 203 records, an increase of 39.0%. Data Scientists retained 16 core tasks, while documented technology or software requirements increased modestly from 84 records in 2023 to 87 in 2026. LDA models produced six interpretable themes for BI Analysts (Cv = 0.621) and three for Data Scientists (Cv = 0.610). Overall, the published O*NET profiles indicate technological modernization and expansion of supporting tools rather than fundamental occupational transformation. The findings inform analytics curriculum development, workforce planning, and future research on occupational change. Net growth in BI Analyst technologies reflected turnover rather than pure accumulation, with 98 distinct titles added and 41 removed. Importance ratings changed only at two discrete survey updates (the 2021 and 2025 releases) rather than drifting continuously, and the BI Analyst profile contained no artificial-intelligence or generative-AI software records in any release through 2026.

Keywords: Data Science, Business Intelligence Analysts, O*NET, topic modeling, occupational competencies, technological modernization

 ISSN: 2473-4901
v12 n6945

Longitudinal Change in Data Science and Business Intelligence Analyst Competencies: Latent Dirichlet Allocation and Topic Prevalence in O*NET Data

Loreen Powell, Carl M. Reman Jr., Steven Levkoff, Theodocia Zayac,
Zachary Brunner, Richard Clarke, and Hayden Wimmer

1. INTRODUCTION

Digital technologies generate data across nearly every area of organizational activity, including accounting, finance, management, marketing, sales, supply chain management, information technology, healthcare administration, and sports management (Organisation for Economic Co-operation and Development [OECD], 2022). Organizations collect data through enterprise systems, digital platforms, connected devices, customer interactions, and online transactions. When effectively managed and analyzed, these data can support evidence-based decision making, operational efficiency, innovation, customer experience, and competitive advantage (Lem, 2024; Syed et al., 2025). Consequently, the capacity to transform raw data into actionable business intelligence has become an important organizational capability rather than a narrowly specialized technical function (Lem, 2024).

The strategic importance of organizational data has been accompanied by substantial investment in analytics technologies and the workforce needed to use them. As organizations expand their use of business intelligence, predictive analytics, cloud computing, and AI, demand has grown for professionals who can acquire, prepare, analyze, visualize, and communicate data across organizational settings (Hartzel & Ozturk, 2022; Levendis & Indika, 2025). Davenport and Patil (2012) famously characterized data scientist as "the sexiest job of the 21st century," reflecting the occupation's growing strategic visibility.

Employer requirements for analytics professionals continue to evolve alongside technological and organizational change. The COVID-19 pandemic accelerated organizational reliance on cloud computing, automation, remote collaboration, and digital business processes (OECD, 2023; World Economic Forum [WEF], 2025). More recently, generative AI tools have been incorporated into coding, data exploration, visualization, and decision support (Brynjolfsson et al., 2025; OECD, 2024). These developments suggest that analytics professionals increasingly need to combine technical capability with critical thinking, ethical judgment, business understanding, communication, and the ability to evaluate AI-assisted outputs (Microsoft, 2025; OECD, 2024; Peslak et al., 2025).

Prior studies have identified analytics competencies through cross-sectional analyses of job advertisements and position descriptions (Booker et al., 2024; Hartzel & Ozturk, 2022; Levendis & Indika, 2025). These studies provide useful snapshots of labor-market expectations but offer limited evidence about how published occupational profiles change over time. A longitudinal analysis of O*NET releases can therefore help distinguish enduring occupational responsibilities from changes in documented technologies and competencies. Such evidence is useful to higher education institutions seeking to align data science and business intelligence curricula with contemporary workforce requirements.

This study examines longitudinal change in the published competency profiles for Data Scientists and BI Analysts in O*NET, the U.S. Department of Labor's occupational database, across a period encompassing the COVID-19 pandemic and the emergence of widely accessible generative AI. The study does not attempt to establish that these events caused observed changes. Instead, it evaluates which occupational requirements remained stable, which expanded or declined, and whether latent competency themes changed across successive O*NET releases. The remainder of the paper presents the literature review, research questions, methods, results, discussion, limitations, and conclusion.

ISSN: 2473-4901
v12 n6945

2. LITERATURE REVIEW

2.1 Data Science and Business Intelligence Analyst Professions

The modern analytics profession emerged through the convergence of mathematics, statistics, computer science, information technology, and organizational decision making (Davenport & Harris, 2007; Provost & Fawcett, 2013). This evolution is reflected in O*NET, the U.S. Department of Labor's occupational database. Earlier O*NET taxonomy reports documented the emergence of interdisciplinary data-intensive occupations, including bioinformatics roles, and the decline or reclassification of narrower mathematical-technical occupations as information technology and database systems expanded (National Center for O*NET Development, 2006, 2009).

Davenport and Patil (2012) highlighted the growing strategic importance of professionals capable of extracting value from large and complex datasets. At that time, Data Scientist had not yet been fully established as a distinct occupation within U.S. occupational classification systems (Wimmer & Powell, 2016). Data Scientists are now among the fastest-growing U.S. occupations; the U.S. Bureau of Labor Statistics (2025) projects employment growth of 34% from 2024 through 2034.

2.2 Knowledge and Skills

Early analytics roles often emphasized spreadsheet analysis, database querying, recurring reports, and descriptive summaries. The expansion of business intelligence platforms, cloud-based data environments, visualization software, predictive modeling, and automated analytics has broadened the technologies used to perform this work (Booker et al., 2024; Paul & MacDonald, 2020). Contemporary position descriptions frequently identify SQL, Microsoft Excel, Python, Tableau, Microsoft Power BI, databases, statistical analysis, and data visualization, together with communication, teamwork, critical thinking, and business problem solving (Booker et al., 2024; Hartzel & Ozturk, 2022).

Collectively, this literature indicates that analytics occupations require an interdisciplinary competency profile combining technical knowledge, analytical reasoning, business understanding, communication, and problem-solving ability (Booker et al., 2024; Hartzel & Ozturk, 2022; Levendis & Indika, 2025). This profile distinguishes analytics professionals from occupations focused primarily on software development, database administration, or highly specialized statistical modeling.

2.3 COVID-19 Accelerated Digital Transformation

The COVID-19 pandemic significantly accelerated organizational digital transformation by forcing firms to adopt technologies that could support remote operations, digital customer interactions, organizational continuity, and rapid decision making. Amankwah-Amoah et al. (2021) characterized the pandemic as a "great accelerator" of digitalization because it compressed technology-adoption timelines that had previously unfolded gradually.

As organizations transitioned to cloud-based operations, e-commerce platforms, virtual collaboration environments, and digitally enabled business processes, the volume and diversity of organizational data increased substantially. Digital interactions that had traditionally occurred through face-to-face meetings, paper-based transactions, or manual workflows were increasingly captured through enterprise information systems, collaboration platforms, customer relationship management systems, online commerce, and cloud-based applications, creating unprecedented quantities of digital data available for organizational analysis (Kudyba, 2020; Reuschl et al., 2022). Consequently, organizations became increasingly dependent on data to monitor rapidly changing business conditions, evaluate organizational performance, forecast demand, optimize operations, and support strategic decision-making during periods of significant uncertainty (Abidi et al., 2023).

The rapid expansion of organizational data also elevated the strategic importance of data analytics. In their examination of firms responding to COVID-19, Abidi et al. (2023) found that organizations with stronger digital capabilities demonstrated greater resilience during the pandemic, suggesting that the ability to collect, integrate, and analyze digital information became an important organizational

ISSN: 2473-4901
v12 n6945

capability rather than merely a technical function. Similarly, Reuschl et al. (2022) reported that organizations were required to move beyond simply implementing new technologies and instead develop sustainable digital capabilities that incorporated data-driven decision making into organizational processes. Collectively, these studies suggest that the pandemic shifted organizational priorities from merely adopting digital technologies toward leveraging the data generated by those technologies to improve organizational agility, operational performance, and managerial decision making.

As reliance on data-driven decision making increased, organizations also required employees capable of transforming rapidly expanding data resources into actionable business intelligence. Data Analysts became increasingly responsible for acquiring, preparing, integrating, analyzing, visualizing, and communicating organizational data generated across diverse digital systems, enabling managers to respond more effectively to evolving business conditions (Hartzel & Ozturk, 2022; Levendis & Indika, 2025). The expansion of digital data environments therefore provides an important theoretical foundation for anticipating changes in employer expectations for Data Analyst positions following the COVID-19 pandemic, particularly with respect to technical competencies related to data management, analytics, visualization, cloud technologies, and decision support.

2.4 Generative AI and Workforce Transformation

The emergence of generative AI has extended the technological change that accelerated during the pandemic. Microsoft's 2025 Work Trend Index describes increased organizational interest in AI-enabled work and digital labor. These developments provide context for examining whether AI-related technologies have entered O*NET profiles for analytics occupations.

The increasing integration of AI has further expanded the boundaries of Data Science and analytics-related employment. In an analysis of 435 AI-related job descriptions, Peslak et al. (2025) found that employers sought technical competencies involving Python, natural language processing, cloud platforms, and generative AI, together with communication, teamwork, interpersonal, and problem-solving skills. Although AI-focused positions are not equivalent to Data Analyst positions, these findings illustrate the increasing convergence of analytics and AI competencies within the labor market. As AI-enabled technologies become integrated into data preparation, coding, visualization, forecasting, and decision support, Data Analysts may increasingly be expected to work with intelligent tools while evaluating the accuracy, relevance, limitations, and ethical implications of AI-assisted outputs (Microsoft, 2025; OECD, 2024; Peslak et al., 2025).

Competency requirements should therefore be examined as potentially dynamic rather than assumed to be static. Cross-sectional job-advertisement studies identify what employers request at a particular time, but longitudinal occupational data are needed to determine whether documented responsibilities, competency ratings, software requirements, and latent themes remain stable or change across successive releases. O*NET provides a standardized archival source for conducting that comparison, although changes in its taxonomy and data structure must be distinguished from genuine occupational change.

3. RESEARCH GOAL AND QUESTIONS

The goal of this study is to examine longitudinal change in the published O*NET competency profiles for Data Scientists and Business Intelligence Analysts from 2016 through 2026. Three research questions guide the analysis:

1. How did the documented competency profiles of Business Intelligence Analysts and Data Scientists change across the selected O*NET releases?

2. Which occupational responsibilities, competency importance ratings, and technology requirements remained stable, increased, or decreased over the study period?

3. What latent competency themes emerged from the O*NET records, and how did their prevalence change over time?

ISSN: 2473-4901
v12 n6945

4. METHODS

This study employed a quantitative longitudinal text-mining design to examine published O*NET competency profiles for Data Scientists and Business Intelligence Analysts. O*NET is sponsored by the U.S. Department of Labor, Employment and Training Administration (USDOL/ETA) and developed by the National Center for O*NET Development (O*NET, 2026). One representative annual release was selected for each year from 2016 through 2026 to avoid overrepresenting years containing multiple, nearly identical production releases. Table 1 provides the selected releases versions used and retrieval location. The annual extraction contained 11 BI Analyst profiles and six Data Scientist profiles.

Table 1.
O*NET Dataset Used

Data Set Used (Version and Date)	URL
20.3, 21.3, 22.3, 23.3, 24.3, 25.3, 26.3, 27.3, 28.3, 29.3, 30.3	https://www.onetcenter.org/db_releases.html

Note. Data were obtained from O*NET Version 30.3 and archived database releases produced by USDOL/ETA. The data are used under the Creative Commons Attribution 4.0 license. O*NET® is a trademark of USDOL/ETA.

O*NET was selected because it provides standardized occupational information that is updated over time and is publicly available for research. The database includes occupational descriptions, tasks, detailed work activities, knowledge, skills, abilities, and technology or software requirements. Because O*NET update procedures and file structures change over time, the analysis treats the records as published occupational profiles rather than direct annual measures of employer behavior.

The analysis was restricted to the O*NET occupations Business Intelligence Analysts and Data Scientists because both are directly relevant to business analytics and information systems education. For each selected annual release, records associated with the two occupations were extracted and combined into a longitudinal corpus. The final dataset contained 4,186 records: 3,645 BI Analyst records from 2016 through 2026 and 541 Data Scientist records from 2021 through 2026. Occupational descriptions, tasks, detailed work activities, technology or software requirements, and competency records with available importance ratings were retained.

Text preprocessing was conducted in Python within the Anaconda data-science environment. Each O*NET record served as a textual observation. Text was converted to lowercase, and punctuation, numbers, special characters, and common stop words were removed. Terms were tokenized and lemmatized with the Natural Language Toolkit (NLTK) to reduce grammatical variation while preserving meaning. The cleaned text was then represented in a document-term matrix. Because many O*NET records are carried forward across releases, repeated text was retained as part of the published annual profile; this design choice means that the analysis evaluates stability and change in O*NET documentation rather than independent annual samples of employer requirements.

Topic models were estimated in Python using the Gensim library. Tokenized texts were augmented with automatically detected two-word phrases (bigrams) prior to dictionary construction, and candidate models containing three through eight topics were estimated separately for each occupation. Topic prevalence was computed as the mean document-topic proportion across all records within each annual release, so that changes in prevalence reflect shifts in the composition of the published profile. An independent re-estimation using an alternative preprocessing configuration (no phrase detection; vocabulary restricted to terms appearing in at least five records and no more than 50% of records; 10 corpus passes; 200 iterations per document chunk; learned asymmetric priors; random seed 42) produced lower absolute coherence but the same qualitative pattern of stable BI Analyst themes and a Data Scientist shift concentrated at the 2023 addition of technology records; this robustness check is documented in the replication materials.

Descriptive frequency comparisons were used to identify changes in record counts, tasks, detailed work activities, technologies, and available competency importance ratings. LDA topic models were then estimated separately for BI Analysts and Data Scientists using the Gensim library. Candidate models containing three through eight topics were compared using the Cv coherence metric, which evaluates semantic consistency among terms within a topic (Mimno et al., 2011; Röder et al., 2015). The selected

 ISSN: 2473-4901
v12 n6945

models balanced coherence and interpretability. Topic prevalence was compared across annual releases to identify themes that remained stable, increased, decreased, or appeared in later profiles. Because the number and organization of technology-related records changed over time, topic-prevalence results are interpreted together with category-specific frequency comparisons rather than as independent evidence of occupational transformation. Figure 1 summarizes the overall data-extraction and analysis pipeline.

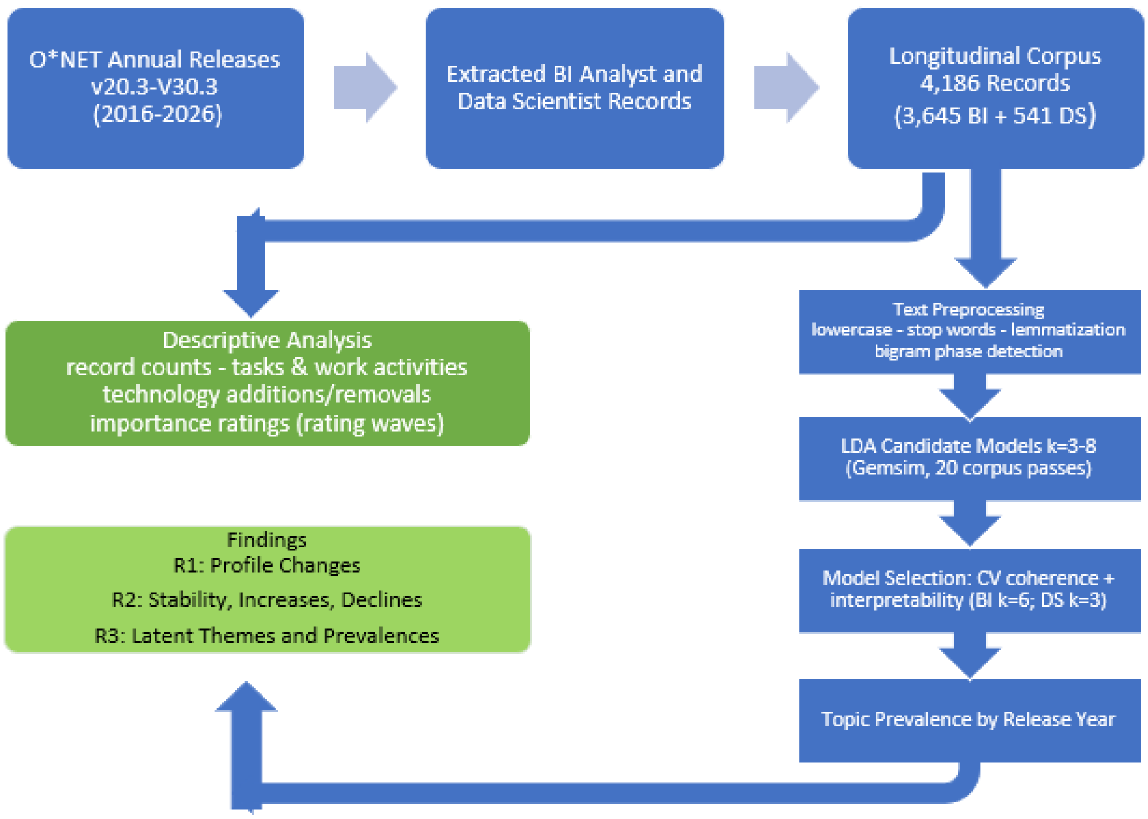

Figure 1. Flowchart overview of the data-extraction and analysis pipeline.
Source: authors' flowchart of O*NET releases 20.3–30.3 extraction and analysis

5. RESULTS

O*NET issued 42 production releases between 2016 and 2026, ranging from Version 20.2 through Version 30.3 and including a taxonomy change in Version 25.1. To avoid giving disproportionate weight to years with multiple similar releases, the analysis used one representative release per year, as shown in Table 1.

The BI Analyst occupational description remained substantively unchanged between the earliest and latest releases examined. O*NET continued to describe BI Analysts as professionals who produce financial and market intelligence by querying data repositories, generating reports, and identifying patterns and trends in available information. The occupation retained 17 substantively consistent core tasks; one task contained only a minor editorial wording change. The occupational code changed when O*NET implemented the 2019 SOC taxonomy, but this administrative change should not be interpreted as a change in the work itself.

Data Scientists were introduced in O*NET Version 25.1, released in November 2020, and first appear in the annual dataset used in this study for 2021. The occupation was coded as 15-2051.00 and described as developing and applying analytical techniques, programming, visualization, data mining, data

ISSN: 2473-4901
v12 n6945

modeling, natural language processing, and machine learning to transform structured and unstructured data into meaningful information. This occupational description remained unchanged across the annual releases examined.

The 2021 Data Scientist profile contained 33 records, reflecting a newly introduced and initially sparse profile, compared with more than 340 BI Analyst records in the same year. Across all selected releases, the extracted dataset contained 4,186 O*NET records: 3,645 for BI Analysts and 541 for Data Scientists.

Several apparent changes between the 2025 and 2026 releases resulted from changes in O*NET file organization rather than occupational change. The Technology Skills file was renamed Software Skills, while the former Skills file was reorganized into Essential Skills and Transferable Skills. Apparent disappearance of records across those categories therefore represents reclassification rather than elimination. Longitudinal interpretation consequently emphasizes comparable task descriptions, detailed work activities, software technologies, available importance ratings, and themes across the complete corpus.

5.1 Business Intelligence Analyst Findings

BI Analysts retained 17 substantively consistent core tasks across the 11 annual releases. The number of detailed work activities increased from 10 in 2016 to 12 in 2026 through the addition of two activities:

- Develop models of information or communications systems
- Report information to managers or other personnel

Documented BI Analyst technology or software requirements increased from 146 records in 2016 to 203 in 2026, an increase of approximately 39.0%. Newer requirements included Amazon Web Services, Amazon Redshift, Amazon DynamoDB, Alteryx, Airtable, Apache Cassandra, Apache Groovy, and other cloud-computing and database-query platforms. This net growth reflected turnover rather than pure accumulation: 98 distinct technology titles were added and 41 were removed between the 2016 and 2026 releases (102 and 45 when category reassignments of unchanged titles are also counted), and a portion of that churn consisted of vendor renaming (for example, "Adobe Systems Adobe Acrobat" became "Adobe Acrobat") rather than substantive change. The timing of additions also indicates that O*NET documentation lags industry adoption. Amazon Web Services first appeared in the 2020 release, Apache Spark in 2021, Alteryx in 2023, and Microsoft Power BI—one of the most widely used BI platforms—did not enter the BI Analyst profile until the 2024 release. Figure 2 presents the growth of documented technology records for both occupations, and Figure 3 presents year-over-year additions and removals.

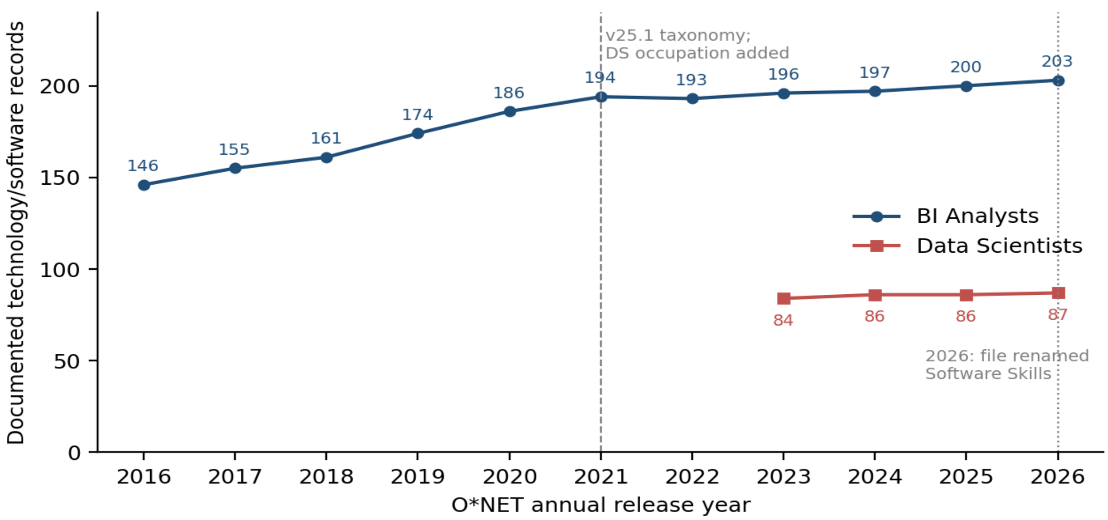

Figure 2. Documented technology/software records per annual O*NET 2016-2026. *Dashed lines mark Version 25.1 taxonomy change and the 2026 renaming of the technology skills to software skills. Source: authors' analysis of O*NET releases 20.3–30.3.*

ISSN: 2473-4901
v12 n6945

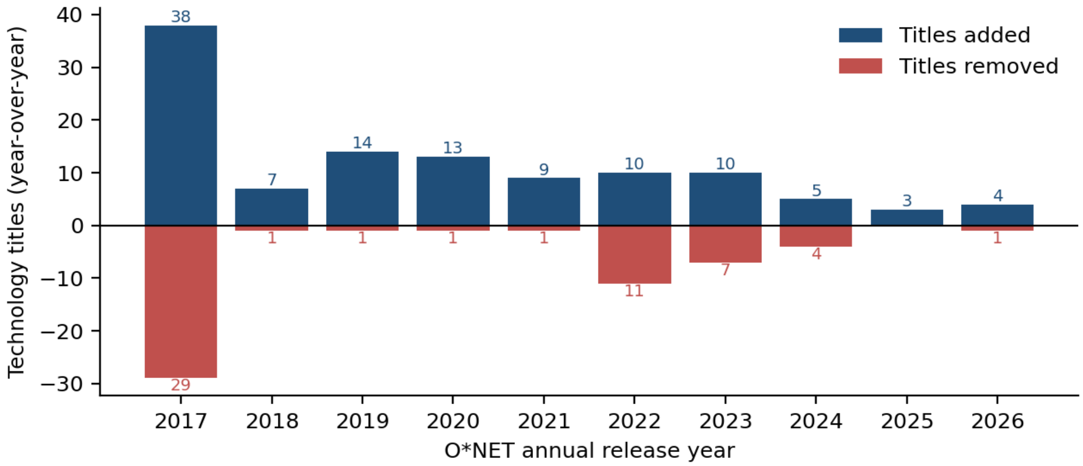

Figure 3. ***Year-over-year additions and removals of BI Analyst technology titles.*** *The large 2017 churn primarily reflects vendor renaming rather than substantive turnover.*
*Source: authors' analysis of O*NET releases 20.3–30.3.*

The expansion of documented technologies indicates that the BI Analyst profile became more technologically extensive, particularly in cloud computing, distributed databases, data integration, scalable analytics, and modern BI platforms. The stable task structure nevertheless suggests that these technologies expanded the tools used to perform the occupation rather than replacing its central organizational purpose. Figure 4 shows the change in documented BI Analyst technology records by O*NET software category between the 2016 and 2026 releases. Growth was concentrated in development environments, database user-interface and query software, operating systems, and business intelligence and data analysis software, with a single category (graphics and photo-imaging software) declining. The remaining 40 of the 54 categories were small and largely static, typically holding one to three records and changing by at most one record over the decade, and spanned areas such as accounting, project management, geographic information systems, and web content management. Complete year-by-year technology listings by category are provided in the study's replication materials. Notably, the BI Analyst profile contained no artificial-intelligence or generative-AI software records in any release through 2026, indicating that the published profile has not yet incorporated the AI tools emphasized in the recent workforce literature.

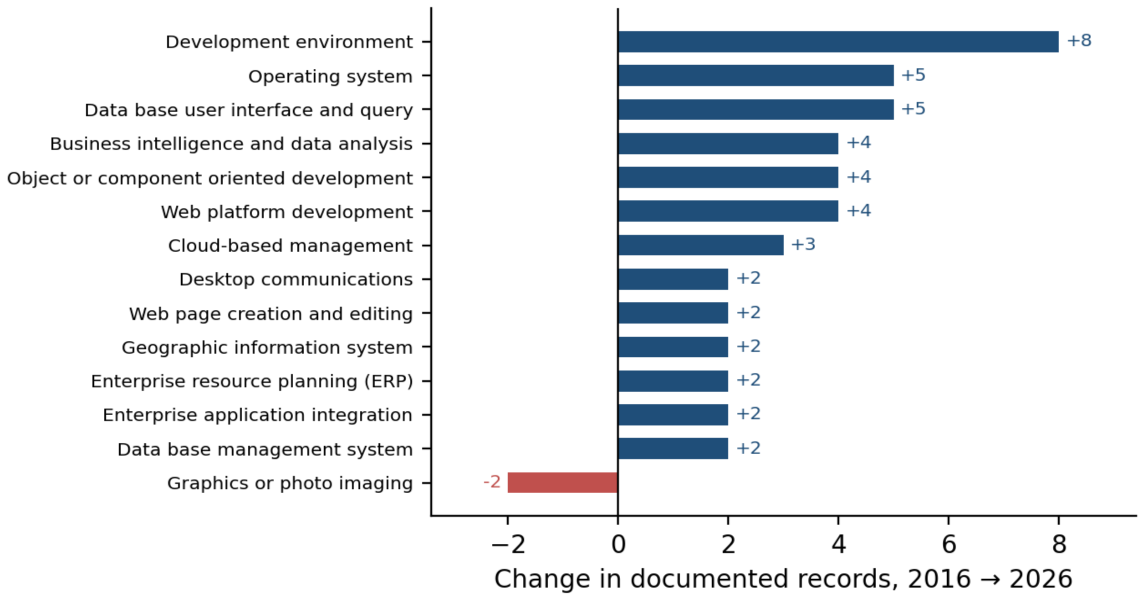

Figure 4. **Change in documented BI Analyst technology records by O*NET software category, 2016–2026**. Categories changing by two or more records are shown; the remaining 40 categories changed by at most one record.
*Source: authors' analysis of O*NET releases 20.3–30.3.*

Available importance ratings also show that BI Analyst work became more technically intensive. The largest increases occurred in Computers and Electronics (+1.00), Design (+0.93), Mathematics (+0.49), Written Expression (+0.38), and Customer and Personal Service (+0.31). The inclusion of written expression and customer service indicates that technical growth did not eliminate communication and stakeholder-facing requirements. Table 2 presents the competencies with the largest increases. These endpoint differences should be interpreted as cumulative changes across discrete rating updates rather than continuous annual drift. Importance ratings changed at only two points during the study period: the 2021 release (coinciding with the Version 25.1 update) and the 2025 release, which appears to reflect a new incumbent survey wave. The data therefore contain three rating regimes (2016-2020, 2021-2024, and 2025-2026), and some paths were non-monotonic; for example, Computers and Electronics declined from 3.29 to 3.05 in 2021 before rising to 4.29 in 2025 (Figure 5). Tables 2 and 3 exclude physical and psychomotor abilities. These descriptors appear in the profile because O*NET's Content Model rates every occupation on a standardized descriptor set—including 52 abilities, many of them physical or psychomotor—through incumbent surveys, so analytical occupations receive ratings on physically oriented abilities by design. Several such abilities showed rating movements as large as the tabled competencies (for example, Trunk Strength +0.63, Wrist-Finger Speed +0.62, and Finger Dexterity -0.75), but these movements lack face validity for analytical work and most plausibly reflect measurement noise in incumbent ratings; their magnitude reinforces the need for caution when interpreting importance changes of similar size.

 ISSN: 2473-4901
v12 n6945

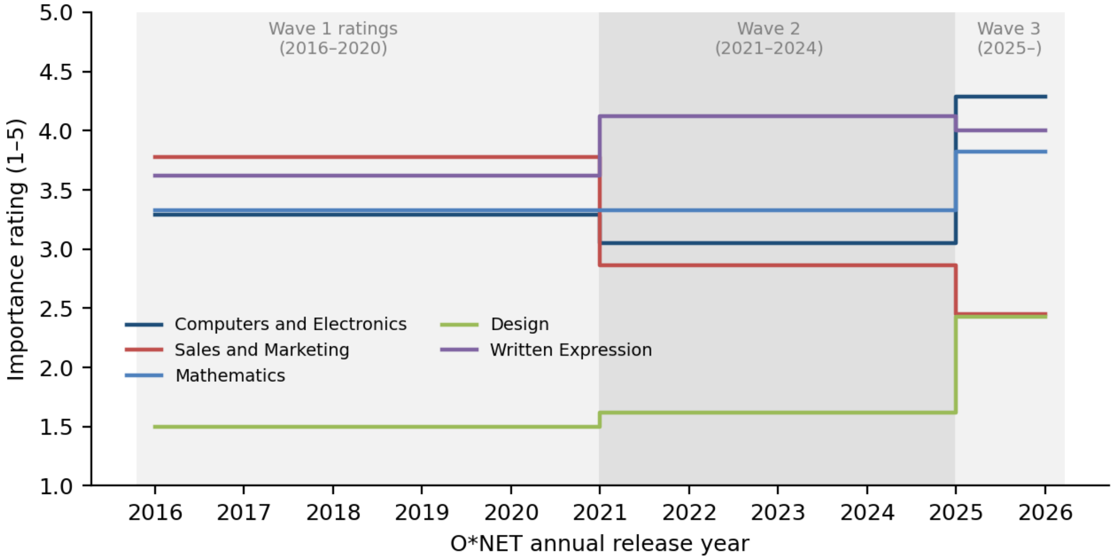

Figure 5. *Selected BI Analyst importance ratings by release year. Ratings are constant within three survey waves (2016-2020, 2021-2024, 2025-2026) and change only at the 2021 and 2025 releases.*
*Source: authors' analysis of O*NET releases 20.3–30.3.*

Table 2.
Competencies Showing the Largest Increases in Importance

Competency	Domain	Earlier Importance	Later Importance	Change
Computers and Electronics	Knowledge	3.29	4.29	**+1.00**
Design	Knowledge	1.50	2.43	+0.93
Mathematics	Knowledge	3.33	3.82	+0.49
Written Expression	Ability	3.62	4.00	+0.38
Customer and Personal Service	Knowledge	2.79	3.10	+0.31

Note. Earlier = 2016 release (Version 20.3); Later = 2026 release (Version 30.3). Ratings use O*NET's 1–5 importance scale; physical and psychomotor abilities are excluded (see text).

Several traditional business, resource-management, and interpersonal competencies declined in documented importance. Sales and Marketing showed the largest decrease (-1.33), followed by Telecommunications (-0.66), Management of Personnel Resources (-0.63), Technology Design (-0.63), and Foreign Language (-0.60). These changes suggest a profile increasingly centered on analytics technologies, mathematics, systems, and written communication rather than broad sales or resource-management knowledge. Table 3 presents the largest declines.

ISSN: 2473-4901
v12 n6945

Table 3.
Competencies Showing the Largest Decreases in Importance

Competency	Domain	Earlier Importance	Later Importance	Change
Sales and Marketing	Knowledge	3.78	2.45	**−1.33**
Telecommunications	Knowledge	2.21	1.55	−0.66
Management of Personnel Resources	Transferable Skill	2.75	2.12	−0.63
Technology Design	Knowledge	2.25	1.62	−0.63
Foreign Language	Knowledge	1.96	1.36	−0.60
Management of Material Resources	Transferable Skill	2.25	1.75	−0.50
Management of Financial Resources	Transferable Skill	2.38	1.88	−0.50
Social Perceptiveness	Transferable Skill	3.38	2.88	−0.50

Note. Earlier = 2016 release (Version 20.3); Later = 2026 release (Version 30.3). Ratings use O*NET's 1–5 importance scale; physical and psychomotor abilities are excluded (see text).

5.2 Data Scientist Findings

The Data Scientist profile reflects expansion of a newly documented occupation rather than a long-established occupational profile. It contained 33 records in both 2021 and 2022, increased sharply to 117 records in 2023, and reached 120 records in 2026. The increase from 33 to 120 records is approximately 264%, but most of the 2023 expansion resulted from the addition of detailed technology-skill information. It should therefore be interpreted as expansion of O*NET documentation rather than evidence that the occupation itself changed by the same magnitude. Notably, the 2021 and 2022 Data Scientist profiles contained no technology-skill records at all; documented technology requirements first appeared in the 2023 release. The Data Scientist analysis is necessarily briefer than the BI Analyst analysis: the occupation entered O*NET only in 2021, providing six annual profiles rather than 11, no archived importance ratings, and sparse early records. This asymmetry is itself informative, illustrating how O*NET documents a newly recognized occupation.

The Data Scientist occupational description and 16 core tasks remained stable across the six annual releases, and the number of detailed work activities did not change. Documented technology or software requirements increased modestly from 84 records in 2023 to 87 in 2026, an increase of approximately 3.6%. Technologies added by 2026 included Microsoft Power BI, OpenAI ChatGPT, and Snowflake. OpenAI ChatGPT first appeared in the 2024 release, and Snowflake in 2026.

These additions reflect the incorporation of business intelligence visualization, generative AI, and cloud-based data platforms into the published profile. However, the modest numerical increase and stable tasks do not indicate extensive occupational restructuring. Longitudinal importance ratings for Data Scientist knowledge, skills, and abilities were not consistently available in the archived files; their absence is a data limitation and should not be interpreted as evidence of stability or decline.

5.3 LDA Topic-Modeling Results

LDA topic modeling was conducted separately for BI Analysts and Data Scientists to identify recurring competency themes. Candidate models containing three through eight topics were evaluated using Cv coherence and substantive interpretability. Higher coherence indicates that frequently co-occurring topic words form more semantically consistent groupings (Mimno et al., 2011; Röder et al., 2015). Because annual O*NET profiles contain repeated records and unequal growth across content categories, topic results are interpreted as patterns in the published corpus rather than as causal measures of changing employer demand. Table 4 reports Cv coherence for all candidate models. For both occupations, the selected model achieved the highest Cv coherence among the candidate solutions (0.621 for the six-

ISSN: 2473-4901
v12 n6945

topic BI Analyst model and 0.610 for the three-topic Data Scientist model), aligning statistical fit with substantive interpretability.

Table 4.
Cv Coherence for Candidate Topic Models

Number of Topics (k)	BI Analysts (Cv)	Data Scientists (Cv)
3	0.576	**0.610 (selected)**
4	0.573	0.537
5	0.609	0.551
6	**0.621 (selected)**	0.543
7	0.601	0.540
8	0.578	0.551

5.4 BI Analysts LDA Modeling Results

The selected BI Analyst model contained six themes:

1. Analytical systems and enterprise infrastructure
2. Enterprise resource planning and development environments
3. Web, platform, and application development
4. Database querying, monitoring, and market intelligence
5. Reporting, SQL, project, and information management
6. Financial analysis, web platforms, and enterprise coordination

Table 5.
BI Analyst Topics and Twelve Highest-Probability Terms

Topic	Top Terms
1. Analytical systems and enterprise infrastructure	analysis, server, management, systems, medical, communications, intelligence, red_hat, linux_operating, resources, adobe, mining
2. Enterprise resource planning and development environments	environment, development, erp, enterprise, resource_planning, microsoft, oracle, systems, operating, oriented, object, ibm
3. Web, platform, and application development	development, oriented, object_component, platform, web, oracle, information, management, microsoft, crm, analytical_scientific, server
4. Database querying, monitoring, and market intelligence	base, query, user_interface, monitoring, intelligence, oracle, reporting, hewlett_packard, analyze, market, information, language
5. Reporting, SQL, project, and information management	management, base, microsoft, reporting, information, server, enterprise, project, sql, reports, scientific, oracle
6. Financial analysis, web platforms, and enterprise coordination	analysis, web, development, platform, intelligence, enterprise, microsoft, financial, markup_language, dynamic, coordination, oracle

The six themes consistently emphasized enterprise information systems, databases, reporting, SQL, project coordination, financial analysis, and organizational decision support. Later annual profiles contained greater representation of cloud analytics, enterprise data platforms, and modern database technologies. This pattern is consistent with the 39.0% increase in documented technology or software records. The result is best interpreted as expansion and modernization of the technical environment supporting BI work while its core responsibilities remained stable. Aggregate topic proportions nevertheless remained nearly constant across releases (Figure 6), with no topic varying by more than

ISSN: 2473-4901
v12 n6945

about 0.02 across the eleven years. This indicates that newly documented technologies entered existing thematic clusters rather than forming new themes, and it quantifies the thematic stability that the task-level comparisons suggest. Table 5 lists the twelve highest probability terms for each theme.

The BI Analyst model produced a Cv coherence score of 0.621. The score supported an interpretable six-topic solution, although coherence should be treated primarily as a comparative model-selection measure rather than a universal threshold of model quality.

5.5 Data Scientists LDA Modeling Results

The selected Data Scientist model contained three themes:

1. Mathematical modeling, scientific analysis, and programming
2. Database management, querying, reporting, and information analysis
3. Statistical analysis, development environments, and analytical platforms

Table 6.
Data Scientist Topics and Twelve Highest-Probability Terms

Topic	Top Terms
1. Mathematical modeling, scientific analysis, and programming	development, object_component, oriented, scientific, analytical_scientific, microsoft, intelligence_analysis, server, mathematical, analyze, analytic_trends, read
2. Database management, querying, reporting, and information analysis	base, management, query, user_interface, apache, analyze, programming_languages, activities, reports, analytical, information, identify
3. Statistical analysis, development environments, and analytical platforms	development, environment, base, analysis, user_interface, query, results, management, apache, analytical_scientific, microsoft, statistical

The three themes emphasized mathematical and statistical modeling, programming, database management, querying, reporting, and analytical platforms. These themes were present throughout the available Data Scientist releases, indicating that the central competency structure was already well established when the occupation entered the annual study corpus. Table 6 lists the twelve highest-probability terms for each theme.

Later profiles incorporated modestly greater representation of cloud platforms, BI visualization, and emerging AI tools. This pattern is consistent with the increase from 84 technology-skill records in 2023 to 87 software-skill records in 2026, including Microsoft Power BI, OpenAI ChatGPT, and Snowflake. Topic prevalence reflects this documentation effect: when technology-skill records entered the profile in 2023, the mathematical-modeling and programming theme rose from 0.249 to 0.269 (reaching 0.285 by 2026) and the database-management theme declined from 0.334 to 0.303, while the statistical-analysis theme remained essentially unchanged (Figure 6). The modest size of these shifts indicates that the new technology records were absorbed into the existing thematic structure rather than reorganizing it.

The Data Scientist model produced a Cv coherence score of 0.610. The score supported an interpretable three-topic solution. Its small difference from the BI Analyst score should not be interpreted as evidence that one occupation is substantively more coherent than the other.

ISSN: 2473-4901
v12 n6945

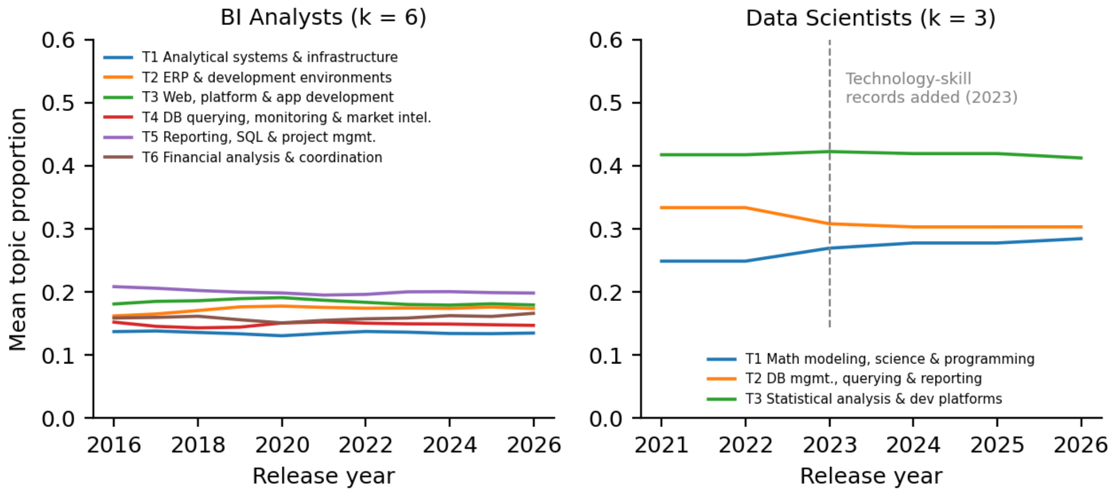

Figure 6. *Mean topic prevalence by annual release, computed from the fitted models. BI Analyst topic proportions are nearly constant across 2016-2026; Data Scientist proportions shift modestly when technology-skill records entered the profile in 2023.*
*Source: authors' analysis of O*NET releases 20.3–30.3.*

5.6 Cross-Occupation Interpretation

Across both occupations, the task structure and dominant themes were more stable than the documented technology environment. BI Analysts showed a larger expansion of technology requirements, whereas Data Scientists showed rapid early expansion of their O*NET profile followed by modest growth after 2023. The evidence therefore supports technological augmentation and profile maturation rather than occupational disruption. At the same time, part of the apparent increase in technology-related topic prevalence is mechanically related to the addition of technology or software records; the LDA results must therefore be interpreted alongside the category-specific counts. The prevalence estimates in Figure 6 make this point concrete: BI Analyst topic proportions varied by no more than about 0.02 despite a 39.0% increase in technology records, and the only visible Data Scientist movement was a modest shift concentrated at the 2023 addition of technology-skill records.

6. DISCUSSION

The findings answer the research questions by showing that the published O*NET profiles changed unevenly across occupational components. In answer to RQ1, core descriptions, tasks, and dominant competency themes remained stable for both occupations even as the documented technology environment expanded. Regarding stability, increases, and declines (RQ2), BI Analysts experienced a substantial expansion in documented software requirements and increased importance of Computers and Electronics, Design, Mathematics, and Written Expression. The Data Scientist profile expanded rapidly when O*NET added detailed technology records, but its core tasks and later technology counts were comparatively stable, while several business and resource-management competencies declined in documented importance. Regarding latent themes (RQ3), the topic models identified six stable BI Analyst themes centered on enterprise data infrastructure, reporting, and analytical development, and three stable Data Scientist themes centered on mathematical-statistical analysis, database management, and development platforms; topic prevalence was nearly constant for BI Analysts and shifted for Data Scientists only when technology records entered the profile in 2023. These patterns align with prior research describing analytics work as an interdisciplinary combination of technical, business, analytical, and communication competencies (Booker et al., 2024; Hartzel & Ozturk, 2022; Levendis & Indika, 2025). Three additional patterns refine this account. First, importance-rating changes occurred at only two discrete survey updates, so competency change in O*NET is registered in steps rather than trends, and single-year comparisons can miss or misdate change. Second, technology growth involved meaningful churn, with 98 title additions offset by 41 removals, including simple vendor

ISSN: 2473-4901
v12 n6945

renaming. Third, the published profiles lag industry practice: Microsoft Power BI did not enter the BI Analyst profile until 2024, and no AI or generative-AI tools appear in the BI Analyst profile in any release through 2026, even as ChatGPT entered the Data Scientist profile in 2024. For educators, this lag implies that O*NET is best treated as a conservative floor for curriculum content, to be supplemented with more current labor-market signals for rapidly emerging technologies.

For curriculum design, the results support continued emphasis on durable foundations such as SQL, databases, statistical and mathematical reasoning, reporting, visualization, and communication. BI curricula should also provide exposure to cloud platforms, distributed databases, data integration, and enterprise analytics environments. Data Science curricula should maintain strong foundations in statistical modeling, programming, machine learning, and data management while incorporating modern cloud platforms, BI visualization, and emerging AI tools. Broader recommendations concerning AI literacy, ethical reasoning, verification of AI-assisted outputs, and human-AI collaboration are supported by the surrounding workforce literature rather than by the O*NET frequency results alone (Microsoft, 2025; OECD, 2024; Peslak et al., 2025).

The study also clarifies an important distinction between occupational change and database-profile change. An increase in the number of documented technologies demonstrates that the published profile has expanded, but it does not by itself show that the occupation has been transformed. Similarly, the stability of repeated O*NET text may partly reflect records being carried forward across releases. The combined evidence is therefore most consistent with modernization of the tools and infrastructure used to perform established analytical functions.

7. LIMITATIONS AND FUTURE RESEARCH

This study is limited to published O*NET data for two occupations. O*NET updates may lag labor-market developments, and the available observation windows are unequal: 11 annual profiles for BI Analysts and six for Data Scientists. Taxonomy and file-structure changes, particularly the 2026 reorganization of Skills and Technology Skills, complicate direct comparisons. Data Scientist importance ratings were not consistently available. In addition, repeated occupational text across annual releases can increase apparent thematic stability, while unequal growth in technology-related records can mechanically increase technology-topic prevalence. LDA topic labels require researcher interpretation, and the analysis does not establish causal effects of COVID-19, generative AI, or other historical events. Because importance ratings changed at only two survey updates, endpoint comparisons summarize discrete rating revisions rather than continuous annual drift. Importance ratings derive from incumbent surveys with modest per-occupation samples, and the magnitude of rating movements observed for analytically irrelevant physical descriptors (approximately ±0.6–0.75) provides an internal yardstick for how much apparent change may reflect measurement noise. In addition, O*NET’s incumbent samples are drawn from sampled establishments and may not evenly represent all firm sizes, geographic regions, or industry segments, and ratings are refreshed on multi-year cycles rather than annually. Future research should assess topic stability across alternative random initializations and preprocessing choices, normalize topic prevalence within comparable content categories, and triangulate O*NET findings with longitudinal job advertisements, employer surveys, curricula, and additional occupations.

8. CONCLUSION

This study provides longitudinal evidence that the published O*NET profiles for BI Analysts and Data Scientists have been characterized more by technological modernization than by fundamental occupational transformation. BI Analysts retained stable core tasks while their documented software environment expanded substantially. Data Scientists entered O*NET with an already recognizable analytical and statistical core, and subsequent profile growth primarily added supporting technologies. These results support curricula that preserve durable analytical foundations while adapting to cloud platforms, enterprise data environments, visualization tools, and emerging AI technologies. The findings should be interpreted as changes in documented O*NET occupational profiles, not as direct causal measures of employer behavior. Real occupational change may therefore exceed what the profiles record: job-advertisement studies indicate that employer requirements, particularly for AI-related competencies, move faster than O*NET registers (Levendis & Indika, 2025; Peslak et al., 2025), and broader organizational and market shifts may reshape these occupations in ways the database captures

ISSN: 2473-4901
v12 n6945

only with delay. At the same time, the absence of AI-related records from the BI Analyst profile through 2026, and the late arrival of widely adopted tools such as Microsoft Power BI, indicate that O*NET documents occupational change conservatively and with a lag; curriculum designers should therefore pair O*NET evidence with more current labor-market sources when planning coverage of emerging technologies.

Acknowledgements
This paper includes information from the O*NET 30.3 Database and archived releases produced by the U.S. Department of Labor, Employment and Training Administration (USDOL/ETA). The information is used under the Creative Commons Attribution 4.0 license. O*NET® is a trademark of USDOL/ETA. The authors modified portions of the information; USDOL/ETA has not approved, endorsed, or tested those modifications.

9. REFERENCES

Abidi, N., El Herradi, M., & Sakha, S. (2023). Digitalization and resilience during the COVID-19 pandemic. *Telecommunications Policy, 47*(4), Article 102522. https://doi.org/10.1016/j.telpol.2023.102522

Amankwah-Amoah, J., Khan, Z., Wood, G., & Knight, G. (2021). COVID-19 and digitalization: The great acceleration. *Journal of Business Research, 136*, 602–611. https://doi.org/10.1016/j.jbusres.2021.08.011

Booker, Q. E., Rebman, C. M., Wimmer, H., Levkoff, S. B., Powell, L., & Breese, J. L. (2024). *Data analytics position description analysis: Skills review and implications for data analytics curricula*. *Information Systems Education Journal, 22*(3), 76–87. https://doi.org/10.62273/SSQL7979

Brynjolfsson, E., Li, D., & Raymond, L. R. (2025). Generative AI at work. The Quarterly Journal of Economics, 140(2), 889–942. https://doi.org/10.1093/qje/qjae044

Davenport, T. H., & Harris, J. G. (2007). Competing on analytics: The new science of winning. Harvard Business School Press.

Davenport, T. H., & Patil, D. J. (2012). Data scientist: The sexiest job of the 21st century. *Harvard Business Review, 90*(10), 70–76, 128. https://pubmed.ncbi.nlm.nih.gov/23074866/

Hartzel, K. S., & Ozturk, P. (2022). Analytics in the business school: Insights from the literature. *Journal of Information Systems Education, 33*(2), 169–181. https://jise.org/Volume33/n2/JISE2022v33n2pp169-181.pdf

Kudyba, S. (2020). *COVID-19 and the acceleration of digital transformation and the future of work*. *Information Systems Management, 37*(4), 284–287. https://doi.org/10.1080/10580530.2020.1818903

Lem, K. W. (2024). Data analytics strategy and internal information quality. *Contemporary Accounting Research, 41*(2), 1376–1410. https://doi.org/10.1111/1911-3846.12942

Levendis, J., & Indika, N. (2025). The business analytics skills employers are looking for. *Decision Sciences Journal of Innovative Education, 23*(2), e70004. https://doi.org/10.1111/dsji.7000

Microsoft. (2025). *2025 Work Trend Index annual report: The year the frontier firm is born*. Microsoft. https://www.microsoft.com/en-us/worklab/work-trend-index/the-year-the-frontier-firm-is-born

Mimno, D., Wallach, H. M., Talley, E., Leenders, M., & McCallum, A. (2011). *Optimizing semantic coherence in topic models*. In *Proceedings of the 2011 Conference on Empirical Methods in Natural*

ISSN: 2473-4901
v12 n6945

Language Processing (pp. 262–272). Association for Computational Linguistics. https://aclanthology.org/D11-1024/

National Center for O*NET Development. (2006). New and emerging occupations methodology development report. https://www.onetcenter.org/dl_files/NewEmerging.pdf

National Center for O*NET Development. (2009). New and emerging occupations of the 21st century: Updating the O*NET®-SOCtaxonomy. https://www.onetcenter.org/dl_files/UpdatingTaxonomy2009_Summary.pdf

Occupational Information Network [O*NET]. (2026). O*NET Resource Center. https://www.onetcenter.org/license_db.html

Organisation for Economic Co-operation and Development. (2022). *Going digital to advance data governance for growth and well-being*. OECD Publishing. https://doi.org/10.1787/e3d783b0-en

Organisation for Economic Co-operation and Development. (2023). OECD employment outlook 2023: Artificial intelligence and the labour market. OECD Publishing. https://doi.org/10.1787/08785bba-en

Organisation for Economic Co-operation and Development. (2024). *OECD employment outlook 2024: The net-zero and AI transition*. OECD Publishing. https://doi.org/10.1787/ac8b3538-e

Paul, J. A., & MacDonald, L. (2020). Analytics curriculum for undergraduate and graduate students. *Decision Sciences Journal of Innovative Education, 18*(1), 22–58. https://doi.org/10.1111/dsji.12196

Peslak, A., Ceccucci, W., Jones, K., & Gachau, M. (2025). *Examining artificial intelligence job descriptions: An analysis of current jobs*. In *Proceedings of the 2025 Information Systems & Computing Academic Professionals (ISCAP) Conference* (Vol. 11, Paper 6313). Information Systems & Computing Academic Professionals. https://iscap.us/proceedings/2025/pdf/6313.pdf

Provost, F., & Fawcett, T. (2013). Data science and its relationship to big data and data-driven decision making. *Big Data, 1*(1), 51–59. https://doi.org/10.1089/big.2013.150

Reuschl, A. J., Deist, M. K., & Maalaoui, A. (2022). Digital transformation during a pandemic: Stretching the organizational elasticity. *Journal of Business Research, 144*, 1320–1332. https://doi.org/10.1016/j.jbusres.2022.01.088

Röder, M., Both, A., & Hinneburg, A. (2015). Exploring the space of topic coherence measures. *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining (WSDM '15)*, 399–408. https://doi.org/10.1145/2684822.2685324

Syed, T. A., Bhatti, Z. A., Clifft, S., Siraj, S., Pahuja, A., & Nawaz, R. (2025). Strategic alignment of big data analytics: Leveraging operational and market capabilities for organizational performance. *British Journal of Management*. https://doi.org/10.1111/1467-8551.70014

U.S. Bureau of Labor Statistics. (2025). *Occupational outlook handbook: Data scientists*. https://www.bls.gov/ooh/math/data-scientists.htm

Wimmer, H., & Powell, L. M. (2016). A comparison of open source tools for data science. Journal of Information Systems Applied Research, 9(2), 4–12. http://jisar.org/2016-9/

ISSN: 2473-4901
v12 n6945

World Economic Forum. (2025). *The future of jobs report 2025*. World Economic Forum. https://www.weforum.org/publications/the-future-of-jobs-report-2025/