Babylonian Journal of Internet of Things
Vol.2026, **pp**. 202–213
DOI: https://doi.org/10.58496/BJIoT/2026/014; ISSN: 3006-1083
https://mesopotamian.press/journals/index.php/BJIoT

Research Article

From Ancient Records to Agentic Intelligence: A Comprehensive Historical Review of Data Science and Its Role in Intelligent Internet of Things Analytics

Aamo Iorliam[1,*], Iveren Blessing Iorliam[2], Yahaya Isah Shehu[3]

[1] *Department of Mathematics/Computer Science, Rev. Fr. Moses Orshio Adasu University, Makurdi (Formerly Benue State University, Makurdi), Nigeria.*

[2] *Science Laboratory Technology Department, Nigerian Army College of Environmental Science and Technology (NACEST), Makurdi, Nigeria.*

[3] *Department of Computer Science, Shehu Shagari University of Education, Sokoto, Nigeria.*

ARTICLE INFO

Article History
Received 11 Jul 2026
Revised 19 Aug 2026
Accepted 16 Sep 2026
Published 18 Oct 2026

Keywords
Data Science,
Historical Evolution,
Foundation Models,
Agentic AI,
Internet of Things (IoT).

ABSTRACT

Since the era of the Mesopotamia between 2900 BCE - 2350 BCE, the goals of "data science-like" approaches have always revolved around collecting fragments of information and trying to make insights out of such information. Data Science was not clearly termed as Data Science from the onset. Moreover, there exists several questions about Data Science just been a polished name for Statistics. Therefore, there is a need to look carefully at how Data Science has evolved over the years and where it is currently. The study defines Data Science as a multidisciplinary field that integrates statistics, mathematics, computer science, domain expertise, and ethics. This is important because, it will give Data Scientists and the Data Science enthusiasts a clear understanding of this discipline. In addition, the paper demonstrates how Data Science transforms massive Internet of Things (IoT) data streams into actionable insights through predictive analytics, anomaly detection, explainable artificial intelligence, cybersecurity analytics, digital twins, and intelligent decision support. The findings reveal that the evolution of Data Science reflects a transition from descriptive statistical analysis to autonomous, explainable, and trustworthy intelligent systems capable of supporting real-time decision-making across diverse domains. This work provides a unified historical perspective, highlights emerging research directions, and offers a valuable reference for researchers, educators, and practitioners seeking to understand the past, present, and future direction of Data Science.

1. INTRODUCTION

The increasing availability of data today has necessitated the concentration on using data as a means to gain a competitive edge in practically every sector including Internet of Things (IoTs). The huge amount and diversity of data have greatly gone beyond manual analysis and, in certain instances, the capabilities of traditional databases [1], [2]. The combination of these factors has led to the growing importance of the role and use of data science in making insights from the huge data generated all over the world especially on IoT devices.

Historically, data science can be traced as far back as 2900 BCE - 2350 BCE, when the ancient Mesopotamia's provided the foundation for record keeping and administration [3]. Since then, there have been several evolutions of "data science-like" key events that have subsequently led to the term Data Science.

Data science at the moment attempts to integrate existing methods to convert the huge data that is generated almost every second into insights/meanings for people, companies, and society as a whole. Data scientists employ a range of techniques, methods and tools, such as artificial intelligence, machine learning, and data visualization, to examine and understand big datasets [4], [47]. There is also a huge demand for the development of new Data Science techniques that can make insights from any given dataset [5].

It is very important to explore the key events that led to the creation of Data Science and also the events that have happened till 2026. Understanding the history of data science is crucial as it provides context, helping practitioners appreciate the discipline's evolution and relevance. It offers valuable lessons from past successes and failures, aiding in the avoidance of repeating mistakes. Again, it unveils technological advancements, empowering data science practitioners to leverage on cutting-edge tools effectively. Furthermore, historical data insights will inspire innovation by showcasing breakthroughs and

*Corresponding author. Email: aamoiorliam@gmail.com

innovative ideas in the Data Science domain. Additionally, it will inform future directions and trends, guiding Data Science research and innovation. Moreover, it will yield cross-disciplinary insights that enrich problem-solving approaches. This will assist in also deducing the application areas of Data Science as well as the benefits this discipline provides especially in the area of IoT.

The rest of the paper is structured as follows: Section 2 gives the History of Data Science, Section 3 explains how Statistics, Mathematics, Computing and Historical Data are rudiments of Data Science, Section 4 presents deductions from the history of data science with a focus Internet of things (IoT), and Conclusion and Future Work is presented in Section 5.

2. HISTORY OF DATA SCIENCE

The history of Data Science is tracked in phases and presented into the Early Foundations of Data Science and the Emergence of Modern Data Science.

2.1. Early Foundations of Data Science

2900 BCE - 2350 BCE: Early Practices in Mesopotamian, Egyptian, and Greek Civilizations

Mesopotamians, Egyptians, and Greeks utilized rudimentary forms of data collection and analysis for administrative, economic, and scientific purposes. The Mesopotamians, for instance, kept meticulous records of agricultural production and transactions on clay tablets, laying the groundwork for systematic data management [6] and this approach looks "data science-like".

1550 BCE - 1070 BCE: Papyrus, Precision, and the Roots of Record Keeping

The ancient Egyptian scribes documented census data, tax records, and astronomical observations on papyrus scrolls, demonstrating an early recognition of the value of empirical evidence [7].

600-300 BCE: The Philosophical and Mathematical Roots of Data Science in Ancient Greece

The development of mathematics in ancient Greece further contributed to the foundation of data science. Scholars like Pythagoras, Euclid, and Archimedes laid the groundwork for quantitative reasoning and geometric analysis, essential skills for modern data scientists [8]. Moreover, the philosophical tradition of empiricism championed by thinkers like Aristotle emphasized the importance of observation and experimentation in acquiring knowledge about the natural world, a principle that underpins contemporary data-driven approaches [9].

1046 BCE–256 BCE: Yin-Yang and Five Elements Theory Shaped Early Scientific Inquiry in China

This provided a comprehensive framework for understanding natural phenomena, human health, and cosmological processes. These philosophies shaped how early Chinese thinkers observed, categorized, and explained the world around them, blending empirical observation with metaphysical concepts. Chinese scholars meticulously recorded data on agriculture, astronomy, and medicine, laying the groundwork for empirical inquiry and evidence-based decision-making [10].

Mid-7th Century - Mid-13th Century: Foundations of Data Science in Classical and Medieval Mathematics

Scholars like Al-Kindi, Al-Khwarizmi, and Ibn Al-Haytham made significant contributions to mathematics, including the development of algebra and the scientific method [11]. Their work laid the foundation for statistical analysis and hypothesis testing, essential components of modern data science. Furthermore, medieval European scholars like Leonardo Fibonacci and Nicole Oresme made pioneering contributions to mathematics and statistics, advancing techniques for data analysis and probability theory between the 12th and 13th century [12].

1614: From Logarithms to Data Science and the Impact of Mathematical Innovation

The invention of logarithms by John Napier in 1614 revolutionized numerical computations, simplifying calculations involving multiplication and division, and providing a foundational tool for subsequent advancements in data analysis [13].

17th Century – 19th Century: The Evolution of Probability, Statistics, and Data Science Based on Foundational Theories to Modern Analytical Practices

Pioneers like Blaise Pascal and Pierre de Fermat laid the foundations of probability theory [14], paving the way for statistical inference and decision-making under uncertainty. Their correspondence on problems of gambling and chance led to the development of combinatorial analysis and the concept of expected value. In the 18th century, mathematicians like Carl Friedrich Gauss and Pierre-Simon Laplace made groundbreaking contributions to statistics and probability theory. Gauss's method of least squares laid the foundation for regression analysis and hypothesis testing, while Laplace's work on probability distributions and Bayesian inference [15] provided powerful tools for modeling uncertainty and making predictions. Furthermore, the emergence of mathematical statistics in the late 19th century, spearheaded by figures like Francis Galton and Karl Pearson, formalized the principles of data analysis and experimental design, laying the groundwork for modern statistical methods.

2.2. Emergence of Modern Data Science

1962: Data Science Predicted

An American Mathematician John Tukey in the paper titled "The Future of Data Analysis" predicted and formulated the real concept of Data Science. Turkey considered "Data Analysis" as an empirical science [16]. In essence, Turkey was looking at how statistics and computers could merge in solving statistical and mathematical problems. Technically speaking, the transition from data analysis to data science can be traced not only to statistics, but also mathematics [16], [17].

1974: Data Science Term Repeated

Peter Naur in his important book titled "concise survey of computer methods" clearly defined data science as "the science of dealing with data, once they have been established, while the relation of the data to what they represent is delegated to other fields and science" [18].

1977: Formation of the International Association for Statistical Computing (IASC)

The IASC's first mission statement was more formed on how data can be converted into information and knowledge [19]. It clearly stated: "It is the mission of the IASC to have traditional statistical methodology, modern computer technology and the knowledge of domain experts in order to convert data into information and knowledge". Currently IASC publishes a high impact journal titled Computational Statistics and Data Analysis and organizes top conferences such as IASC world conferences, ARS conferences on statistical computing in Asia and COMPSTAT in Europe. The IASC has played a major role in shaping modern statistical computing, bridging the gap between statistics and computer science, and supporting innovation in data analysis and statistical software development since 1977.

1989: Knowledge Discovery in Databases (KDD) Workshop

The first knowledge discovery in database workshop was organized and chaired by Gregory Piatetsiry-Shapiro [20]. The focus of the workshop was to bring to the attention of scholars the emerging area of "knowledge discovery in databases". Later in 1995, this workshop was renamed ACM SIGKDD conference on knowledge Discovery and Data Mining. The mission of SIGKDD is partly focused on mining data from any available source. The ACM SIGKDD focused on data mining and knowledge discovery, which would become central to data science.

1994: Database Marketing Published

A cover story was published by the business week titled "Database Marketing" [21]. This cover story emphasized on the importance of database-marketing frontier. It stated that businesses had begun collecting vast amounts of personal data with the intention of launching new marketing campaigns. Many business managers felt overwhelmed and confused by the huge amount of different information, not knowing how to make sense of it or what to do with it.

1996: International Federation of Classification Societies (IFCS) Biennial Conference

The conference clearly induced "data science" in their title: "Data Science, classification, and related methods". The IFCS currently supports the journal of classification (JoC), the journal of data Analysis and Classification (ADAC), and organizes a biennial conference of International Federation of Classification Societies [22].

1996: From Data Mining to knowledge Discovery in Databases

Usama Fayyad, Gregory Piatetsiey-Shapiro, and Padhraic Smyth published a fundamental paper in AI magazine volume 17 number 3, 1996. The article focused on the historical idea of extracting meaningful patterns from data, often known as "data mining," "knowledge extraction," "information discovery," and "data archaeology" [23].

1997: Professor C. F. Jeff Wu, Called for Statistics to be Renamed

Professor C. F. Jeff Wu gave a public lecture at the University of Michigan titled: "Statistics = Data Science?". He argued that the field of statistics should rebrand itself as "data science" [17], [24].

1999: Statistics VS Data Science by Jacob Zahavi

Jacob Zahavi in Mining Data for Nuggets of Knowledge, published in 1999, started highlighting certain issues with statistics and describes such issues that appeared as data science, noting that while modern databases contain millions of rows and dozens of columns of data, conventional statistical approaches perform well for small data sets. Zahavi identified scalability as a major problem in data mining. Additionally, creating models that are more adept at data analysis, nonlinear relationship detection, and element interaction are technical difficulties. It might be necessary to create specialized data mining techniques to handle website decisions [25].

2001: William S. Cleveland Published a Fundamental Paper titled – "Data Science: An Action Plan for Expanding the Technical Areas of the Field of Statistics" and Leo Breiman Published a Top Paper titled – "Statistical Modeling: The Two Cultures"

Cleveland outlines a strategic framework to evolve statistics into a more comprehensive discipline and the author termed it "data science". This key paper laid the groundwork for the development of data science as an interdisciplinary field [26]. Furthermore, Leo Breiman explained the two predominant approaches in statistical modeling: data modeling and algorithmic modeling. The author noted that focusing solely on data models limits the ability to tackle a wide array of contemporary problems effectively. This paper influenced the development of modern machine learning techniques and prompted a reevaluation of the goals and methodologies in statistical modeling [27].

2002 and 2003: Launch of Data Science Journal, and Journal of Data Science in 2002 and 2003 respectively.
The Data Science Journal was launched by "Committee on Data for Science and Technology (CODATA)" in 2002 [25]. The journal's url is "https://datascience.codata.org/". Furthermore, the "Journal of Data Science" which is an "an open-access, peer-reviewed academic journal" was established in 2003 and the journal's url is: https://jds-online.org/journal/JDS.
2005: The National Science Board Report
In 2005, the National Science Board released "Long-lived Digital Data Collections: Enabling Research and Education in the 21st Century". The report addressed the transformative role of digital data in research and education [28].
2006: Davenport Thomas Published a Fundamental Paper: "Competing on Analytics"
The paper "Competing on Analytics" was published in 2006, by Davenport Thomas and discussed how organizations can leverage data and analytics as a strategic asset to outperform competitors [29]. This paper underscores the transformative power of data and analytics in modern business. By developing the right capabilities, aligning analytics with strategic goals, and fostering a data-driven culture, organizations can position themselves to lead in their respective industries [29].
2007: The Research Center for Dataology and Data Science Established at Fudan University, Shanghai, China.
This center focuses on Data Science and Dataology and holds annual symposiums. The center aims to " study on the theories, methods and technologies of Dataology and Data Science" [30].
2008: Joint Information Systems Committee (JISC) Publishes a Report Mentioning Data Scientist
The report stated that the role of JISC is to "examine and make recommendations on the role and career development of data scientists and the associated supply of specialist data curation skills to the research community". As such the name "data scientists" and "data curation" which are key terms in data science today are clearly mentioned in this report [31].
2009: Harnessing the Power of Digital Data for Science and Society is Published and Troy Sadkowsky Creates Data Scientists Group
The Interagency Working Group on Digital Data's report to the Committee on Science of the National Science and Technology Council stated that "The nation needs to identify and promote the emergence of new disciplines and specialists expert in addressing the complex and dynamic challenges of digital preservation, sustained access, reuse and repurposing of data. Many disciplines are seeing the emergence of a new type of data science and management expert, accomplished in the computer, information, and data sciences arenas and in another domain science. These individuals are key to the current and future success of the scientific enterprise. However, these individuals often receive little recognition for their contributions and have limited career paths" [25], [32].
Also in 2009, Troy Sadkowsky established the data scientists group on LinkedIn in June 2009 to complement his website datascientists.com [25],[32].
2010: Data, Data Everywhere: Bringing all the Data Together for Continuous School Improvement was Published
The Economist Special Report "Data, Data Everywhere," published in February, 2010, quotes Kenneth Cukier as saying that "a new kind of professional has emerged, the data scientist, who combines the skills of software programmer, statistician, and storyteller/artist to extract the nuggets of gold hidden under mountains of data" [33].
In "What is Data Science?" from June 2010, Mike Loukides explains: "Data scientists mix entrepreneurship with patience, a willingness to construct data products slowly, the capacity to research, and the capacity to iterate over a solution [34].
Again, In 2010, Mason, and Wiggins (2010) in "A Taxonomy of Data Science" stated that the job a data scientist does is usually to "Obtain, Scrub, Explore, Model, and iNterpret" [35].
Also in 2010, Drew Conway proposed "The Data Science Venn Diagram" [36]. Even though, Taylor [36] explained more about Data Science Venn diagrams and noted a few flaws in the Drew Conways's proposed data science venn diagram.
2011: Defining Data Science as an Emergence of a Distinct Discipline Beyond Traditional Boundaries
Pate Waden stated: "There is no widely accepted boundary for what's inside and outside of data science's scope. I believe that the recent abundance of data has sparked something new in the world, and when I look around I see people with shared characteristics who don't fit into traditional categories. These people tend to work beyond the narrow specialties that dominate the corporate and institutional world, handling everything from finding the data, processing it at scale, visualizing it and writing it up as a story. They also seem to start by looking at what the data can tell them, and then picking interesting threads to follow, rather than the traditional scientist's approach of choosing the problem first and then finding data to shed light on it." [37].
Also, in 2011, David Smith wrote in "Data Science': What's in a name?" that "the terms 'Data Science' and 'Data Scientist' have only been in common usage for a little over a year, but they've really taken off since then: many companies are now hiring for 'data scientists', and entire conferences are run under the name of 'data science'. But despite the widespread adoption, some have resisted the change from the more traditional terms like 'statistician' or 'quant' or 'data analyst'…. I think 'Data Science' better describes what we actually do: a combination of computer hacking, data analysis, and problem solving." [38], [39].
Matthew J. Graham presents on "The Art of Data Science" in June 2011

The author stated that "We need to develop new talents in order to thrive in the new data-intensive environment of 21st-century research..." We must comprehend the principles that [data] abides by, the symbols and language used to represent it, as well as how it relates to actual space and time. The author pointed out that the focus should be more on data and not necessarily number [40].

According to Harlan Harris' in the article "Data Science, Moore's Law, and Moneyball" in 2011: "What Data Scientists' perform is referred to as Data Science". The work of data scientists has been extensively explored, and it includes everything from data collection and cleaning, and data mining, to using statistics, machine learning, and other related approaches, as well as interpreting, communicating, and visualizing the results.

Also in 2011, Patil wrote in "Building Data Science Teams" that "meeting was the start of data science as a distinct professional specialization.... we realized that as our organizations grew, we both had to figure out what to call the people on our teams. 'Business analyst' seemed too limiting. 'Data analyst' was a contender; many of the people on our teams had deep engineering expertise. 'Research scientist' was a reasonable job title used by companies like Sun, HP, Xerox, Yahoo, and IBM. However, The term that seemed to fit best was data scientist: those who use both data and science to create something new" [38].

2012: The Emergence and Impact of Data Science as a Transformative Force in Modern Business

Davenport and Patil wrote in their paper "Data Scientist: The Sexiest Job of the 21st Century" that "The title has been around for only a few years…But thousands of data scientists are already working at both start-ups and well-established companies." [25], [38].

2013: Data Science in Aviation Workshop Holds in Madrid

The data science in aviation workshop took place in Madrid in October, 2013. This workshop targeted areas in aviation safety and other innovative areas linked to aviation [39].

2015: Artificial intelligence (AI), Deep Learning, and Machine Learning as Fields of Data Science

The development of self-driving cars, personalized entertainment, and personalized purchasing have all been fuelled by these technologies during the past years, along with all the insights necessary to successfully implement these practical AI applications in our daily lives [41].

2016: Explainable AI (XAI) Introduced

Defense Advanced Research Projects Agency (DARPA) launched the XAI program, which was a formal and influential turning point in artificial intelligence and data science domain. This program aimed to create a suite of machine learning techniques that produce more explainable models while maintaining a high level of learning performance [42].

2018 and 2019: Major Breakthroughs in Artificial Intelligence and Machine Learning

Key advances in data science occurred such as development of Bidirectional Encoder Representations from Transformers (BERT), Automated machine learning and Neural Architecture Search, AI for Protein Folding (AlphaFold's Early Version), GPT-2 by OpenAI, and Federated Machine Learning all developed as tools for data science purposes [43].

2020 to 2021: Global Rise of Python for AI and Computational Applications

Python became the most used programming language in the world around this era for building data models, predictions and simulations.

Additional advancements in AI and machine learning took place in this period, and there was an increasing need for experienced Big Data specialists [43].

2022: Rapid Rise in Generative AI models

In 2022, one of the major stories in data science was the rapid rise and impact of generative AI models, especially following the launch and widespread adoption of large language models (LLMs) like OpenAI's GPT-3. OpenAI produced ChatGPT combined with Natural Language Processing for smart thinking [48].

2023: From GPT-4 to Global AI Ethics

GPT-4 was introduced in 2023 [44] and the increasing focus on ethical AI and data governance increased. This was sparked by growing awareness around biases in AI models, privacy concerns (like those related to General Data Protection Regulation (GDPR) and data regulations worldwide), and the push for more transparent, explainable AI systems.

2024: Rapid Advances that Included Major Data Science Conferences and Summits

Conferences and summits totally focused on data science were held in 2024 to include DeepFest 2024 that held in Riyadh, Saudi Arabia that attracted over 48,000 attendees, and Databricks Data + AI Summit that held in San Francisco, USA and featured over 500 sessions amongst several other data science conferences.

2025: AI-Powered Revolution in Weather Forecasting

The Aardvark's innovative use of AI and machine learning represents a significant leap forward in weather forecasting, offering a faster, more efficient, and accessible alternative to traditional supercomputing methods that has been into existence [45].

2025: More Conferences and Summits Targeted at Data Science Held in 2025
Some of these conferences and summits include Data Innovation Summit 2025 (which held in Stockholm, and attracted over 3,000 delegates and featured more than 250 sessions), Data + AI Summit 2025 (Databricks) (which had 20,000 attendees for over 700 sessions), GITEX Asia 2025 (which held in Singapore, and attracted over 5,000 exhibitors and 50,000 attendees), and AI Action Summit 2025 (which was co-chaired by French President Emmanuel Macron and Indian Prime Minister Narendra Modi, and it had over 1,000 participants from more than 100 countries).
2025: Rise of Autonomous Analytics
Autonomous artificial intelligence (AI) agents which assist or independently perform key stages of the analytics lifecycle, including data collection, preprocessing, feature engineering, model selection, interpretation, and decision support became more pronounced [46], [49]. Agentic systems leverage large language models, reasoning capabilities, and tool orchestration to adapt dynamically to changing data and objectives. Agentic analytics enhances productivity, accelerates insight generation, and improves decision-making across domains such as healthcare, finance, cybersecurity, and scientific research while maintaining human oversight for critical decisions [49].
2025: Foundation Models Become Core Data Science Tools
Foundation models have rapidly evolved into essential tools for modern data science by providing powerful capabilities for data understanding, feature extraction, prediction, and natural language interaction [52]. Built on large-scale pretraining, models such as transformer-based large language models (LLMs) enable data scientists to automate coding, exploratory data analysis, feature engineering, model development, and result interpretation. Recent studies highlight that foundation models improve productivity, reduce development time, and support multimodal analytics by integrating text, images, audio, and structured data within a unified framework. Despite these advantages, researchers emphasize challenges related to model bias, hallucinations, computational costs, privacy, and explainability, necessitating responsible deployment and effective human oversight.
2025: Generative AI Integrated into Data Science
Generative Artificial Intelligence (GenAI) became increasingly integrated into data science by automating tasks across the entire analytical lifecycle, from data preparation and feature engineering to model development, visualization, and report generation. Powered by large language models and multimodal foundation models, GenAI enables natural language interaction with complex datasets, accelerates exploratory data analysis, and assists in generating high-quality code and documentation [48],[53].
2026: Real-Time Agentic Data Streaming
Real-time agentic data streaming recently emerged as an advancement in data science that combines autonomous AI agents with continuous data processing to enable intelligent, low-latency analytics. Agentic architectures dynamically monitor, interpret, and respond to streaming data with minimal human intervention. Recent research demonstrates their effectiveness in applications such as Internet of Things (IoT), cybersecurity, financial fraud detection, healthcare monitoring, and smart cities, where rapid decision-making is essential. By integrating event-driven architectures, foundation models, and autonomous reasoning, agentic streaming systems improve scalability, adaptability, and operational efficiency [54].
2026: Decision Governance for AI Agents
As AI agents increasingly perform complex tasks with limited human intervention, governance frameworks integrate policy constraints, human oversight, explainability, risk assessment, and continuous monitoring to align agent behavior with organizational objectives and regulatory requirements. In 2026, researchers are using guardrails, audit trails, reinforcement learning with human feedback (RLHF), and formal verification techniques to enhance trustworthiness [50].
2026: XAI techniques for Zero-day Threats Detection
Daniel-Beston and Iorliam [51] proposed an explainable framework that integrates Isolation Forest, Bidirectional LSTM Autoencoder, SHAP, and LIME, demonstrating how data science can deliver accurate, interpretable, and trustworthy IoT anomaly detection for cybersecurity applications. The proposed model identified anomalous behavioural patterns without requiring labelled attack data, making them suitable for detecting zero-day threats.
2026: XAI techniques for Mango Fruits Shelf Life Prediction
Iorliam and Iorliam [52] utilised mango fruit data generated from experimental data and proposed an interpretable stacking ensemble machine learning framework for accurate shelf-life prediction of mango fruit using explainable artificial intelligence (XAI).

3. DATA SCIENCE AS AN INTERDISCIPLINARY FIELD

Research has attempted to replace similar data-related disciplines of statistics, mathematics and computing with data science [24], [25], [34]. Data science is a highly interdisciplinary field that draws on knowledge and techniques from multiple domains.

The major interdisciplinary fields that are related to data science are Statistics, Computer Science, Mathematics, Machine Learning and Artificial Intelligence (AI), Data Engineering, Visualization and Communication, Operations Research, Business and Management, Ethics and Social Sciences, and Domain Expertise.
Table 1 summarises these interdisciplinary areas and how they contribute to the field of Data Science.

TABLE I: INTERDISCIPLINARY FIELDS AND DATA SCIENCE

S/No.	Field	Contribution to Data Science
1.	Statistics	Data analysis, inference, uncertainty estimation
2.	Computer Science	Programming, algorithm design, system scalability
3.	Mathematics	Foundations of models and algorithms
4.	Machine Learning/AI	Pattern recognition and prediction
5.	Data Engineering	Infrastructure for data handling
6.	Visualization	Communicating insights effectively
7.	Operations Research	Optimization and decision making strategies
8.	Business and Management	Strategic alignment and value creation
9.	Ethics and Social Sciences	Addressing fairness, privacy, and societal impact
10.	Domain Expertise	Contextual understanding of problems and data

Data science, at its core, revolves around extracting insights from data. One of the most critical sources of such data is historical data which is past records and observations that have been collected over time. These datasets act as the rudiments or building blocks upon which modern data science practices are founded. Historical data is the foundation of data science, enabling predictive modeling, decision-making, and strategic planning.

In reality, data science is an interdisciplinary field, and No data = No Data Science.

4. DATA SCIENCE WITH A FOCUS ON INTERNET OF THINGS (IOT)

Data Science serves as the analytical intelligence layer of the Internet of Things, transforming massive volumes of heterogeneous sensor data into actionable knowledge that supports informed decision-making [57], [61], [67] . Through data acquisition, preprocessing, exploratory analysis, predictive modelling, anomaly detection, classification, clustering, time-series forecasting, explainable AI, edge intelligence, optimization, cybersecurity analytics, digital twins, and sustainability applications, Data Science uncovers patterns that would otherwise remain hidden within continuous IoT data streams [51]. Emerging paradigms including foundation models, generative AI, federated learning, multimodal analytics, and agentic AI are further enhancing the scalability, autonomy, and interpretability of IoT analytics [67]. As billions of connected devices continue to generate increasingly complex data, the integration of advanced Data Science methodologies will remain essential for converting raw observations into reliable insights that improve operational efficiency, resilience, security, sustainability, and evidence-based decision-making across diverse application domains. Table 2 summarises how Data Science can provide solutions to IoT.

TABLE II: DATA SCIENCE SOLUTIONS TO IOT

S/No.	Key Data Science Theme	Analytical Deductions from IoT Data	Knowledge/Insights Derived	Data Science Techniques and Solutions
1.	Understanding IoT Data Characteristics [53]	Identify the nature, quality, structure, and behaviour of IoT data streams.	High volume, velocity, variety, temporal dependency, noise, and high dimensionality from IoT datasets.	Exploratory Data Analysis (EDA), descriptive statistics, visualization, dimensionality reduction, and data profiling.
2.	Data Acquisition and Integration [54]	Combine heterogeneous sensor data into a unified analytical dataset for IoT datasets.	Relationships among multiple sensors, synchronized events, and comprehensive system behavior of IoT systems.	Data integration, sensor fusion, timestamp synchronization, data warehousing, and feature engineering.
3.	Data Preprocessing and Quality Improvement [55]	Detect and correct data quality issues before analysis on IoT datasets.	Missing values, noisy signals, sensor drift, duplicate records, and outliers in IoT datasets.	Data cleaning, normalization, filtering, imputation, outlier detection, and feature scaling.
4.	Exploratory Data Analysis and Pattern Discovery [56]	Discover hidden relationships and trends within IoT datasets.	Seasonal trends, correlations, behavioural patterns, and operational variability on IoT datasets.	Statistical analysis, visualization, clustering, correlation analysis, and pattern mining.
5.	Predictive Analytics [57]	Forecast future events using historical IoT data.	Equipment failure prediction, disease forecasting, traffic prediction, and energy demand	Machine learning and deep learning models such as Random Forest, XGBoost, LSTM, and Transformers.

S/No.	Key Data Science Theme	Analytical Deductions from IoT Data	Knowledge/Insights Derived	Data Science Techniques and Solutions
			estimation using IoT generated IoT datasets.	
6.	Anomaly Detection [58]	Identify abnormal behaviour that deviates from normal operations in IoT environments.	Cyberattacks, equipment malfunction, fraud, medical emergencies, and unusual sensor activity in IoT systems.	Isolation Forest, Autoencoders, One-Class SVM, Local Outlier Factor, and unsupervised learning techniques.
7.	Classification and Decision Support [57]	Categorize IoT observations into predefined classes for automated decisions.	Normal vs abnormal behaviour, healthy vs diseased systems, benign vs malicious traffic in IoT datasets	Decision Trees, Random Forest, Logistic Regression, Support Vector Machines, and Deep Neural Networks.
8.	Clustering and Behaviour Discovery [59]	Group similar observations without labelled IoT data.	User segmentation, operational modes, traffic patterns, and behavioural profiling of IoT systems.	K-Means, DBSCAN, Hierarchical Clustering, Gaussian Mixture Models, and Spectral Clustering.
9.	Time-Series Analytics [60]	Analyse sequential IoT sensor observations over time.	Trends, seasonality, forecasting, temporal anomalies, and change-point detection in IoT datasets.	ARIMA, Prophet, LSTM, GRU, Temporal Convolutional Networks, and Transformer-based forecasting.
10.	Explainable Artificial Intelligence (XAI) [61]	Explain how machine learning models reach their predictions on IoT datasets	Feature importance, prediction transparency, model interpretability, and decision confidence in IoT datasets.	SHAP, LIME, Integrated Gradients, attention mechanisms, and counterfactual explanations.
11	Real-Time Analytics and Edge Intelligence [62]	Process streaming IoT data with minimal latency.	Immediate event detection, rapid response, reduced communication delay, and local intelligence in IoT systems.	Edge computing, streaming analytics, online machine learning, and edge AI deployment.
12.	Optimization and Prescriptive Analytics [63]	Recommend optimal actions based on predictive insights on IoT datasets.	Resource optimization, maintenance scheduling, energy efficiency, and operational improvement in IoT systems.	Reinforcement Learning, Linear Programming, Genetic Algorithms, Swarm Intelligence, and optimization models.
13.	Cybersecurity Intelligence [64]	Detect and predict cyber threats targeting IoT infrastructures.	Intrusion detection, malware identification, DDoS attacks, behavioural anomalies, and threat intelligence in IoT datasets.	AI-driven intrusion detection systems, behavioural analytics, anomaly detection, and threat prediction models.
14.	Digital Twins and Simulation [65]	Create virtual models of physical IoT assets for monitoring and prediction.	Equipment performance, predictive maintenance, system simulation, and operational optimization in IoT systems.	Digital Twin technology, simulation modelling, predictive analytics, and real-time synchronization.
15.	Sustainability and Environmental Intelligence [66]	Support sustainable resource management using IoT data analytics.	Water conservation, energy optimization, pollution monitoring, precision agriculture, and climate resilience in IoT systems.	AI-driven environmental monitoring, predictive resource management, smart agriculture, and intelligent energy systems.
16.	Emerging Intelligent IoT Analytics [67]	Enable autonomous, privacy-preserving, and adaptive IoT intelligence.	Autonomous decision-making, multimodal learning, privacy protection, continual learning, and scalable analytics in IoT systems.	Foundation Models, Generative AI, Agentic AI, Federated Learning, Graph Neural Networks, Self-Supervised Learning, and Privacy-Preserving AI.

5. CONCLUSION and FUTURE WORK

This paper presented a comprehensive historical and conceptual review of the evolution of Data Science, tracing its origins from early record-keeping practices in ancient civilizations through the development of mathematics, statistics, computing, and artificial intelligence to the emergence of modern agentic and foundation-model-driven data science in 2026. The review demonstrates that Data Science is not a newly created discipline but rather the culmination of centuries of scientific progress, integrating advances in quantitative reasoning, computational methods, statistical inference, and intelligent decision-making. The study further established that contemporary Data Science has evolved into a highly interdisciplinary field that combines statistics, computer science, mathematics, machine learning, artificial intelligence, data engineering, visualization, operations research, ethics, and domain expertise to transform raw data into actionable knowledge.

A major contribution of this paper is the synthesis of historical milestones with emerging technological developments to illustrate how Data Science continues to redefine modern analytics. The analysis further highlighted the pivotal role of Data Science in Internet of Things (IoT) environments, where advanced analytical techniques enable predictive modelling, anomaly detection, explainable artificial intelligence, cybersecurity analytics, digital twins, and autonomous decision

support. Recent developments such as foundation models, generative AI, agentic analytics, real-time streaming intelligence, and governance frameworks for AI agents demonstrate that the discipline is transitioning from descriptive and predictive analytics toward increasingly autonomous, trustworthy, and explainable intelligent systems.

This paper presents a timeline of data science developments, but this timeline is particularly valuable in providing a clear historical perspective, highlighting major milestones, and demonstrating how successive technological and methodological advances have shaped the evolution of present-day data science.

Moreover, this paper suggests that the transition from data analysis to data science represents a significant development in the field, highlighting the growing integration of computational methods, machine learning, artificial intelligence, and advanced analytical techniques. This transition provides important context for understanding how data science has evolved into a multidisciplinary field with broad applications.

Again, this paper presents a timeline of advancements but does not focus solely on historical developments. Instead, it uses these milestones to contextualize the evolution of data science, demonstrate the progression of methodologies and technologies, and establish a foundation for understanding contemporary practices, emerging trends, interdisciplinary applications, and future directions within the field of data science.

Future research should extend this historical synthesis by establishing standardized benchmarks, ethical governance frameworks, reproducible evaluation protocols, and sustainability metrics for ensuring that future Data Science systems remain transparent, robust, human-centred, and capable of delivering reliable evidence-based decision-making in increasingly data-intensive environments.

Conflicts of Interest

The authors declare no conflict of interest.

Funding

This research received no external funding.

Acknowledgment

None.

References

[1] F. Provost and T. Fawcett, “Data science and its relationship to big data and data-driven decision making,” *Big Data*, vol. 1, no. 1, pp. 51–59, 2013.

[2] S. Ayyalasomayajula, “The dawn of big data: Origins and early promise of big data,” in *AI and the Revival of Big Data*, IGI Global Scientific Publishing, pp. 1–22, 2025.

[3] C. J. Lucas, “The scribal tablet-house in ancient Mesopotamia,” *History of Education Quarterly*, vol. 19, no. 3, pp. 305–332, 1979.

[4] G. Arora, S. Kalra, and G. Kaur, “Introduction to data science and analytics,” in *Revolutionizing Data Science and Analytics for Industry Transformation*, IGI Global Scientific Publishing, pp. 1–28, 2025.

[5] A. Iorliam and Ö. Asilkan, “Can Benford’s law serve as a data science tool?,” 2024.

[6] D. Garbutt, “The significance of ancient Mesopotamia in accounting history,” *Accounting Historians Journal*, vol. 11, no. 1, pp. 83–101, 1984.

[7] D. T. Whiteside, “And John Napier created logarithms...,” *BSHM Bulletin: Journal of the British Society for the History of Mathematics*, vol. 29, no. 3, pp. 154–166, 2014.

[8] D. J. Hand, “Statistics and computing: the genesis of data science,” *Statistics and Computing*, vol. 25, no. 4, pp. 705–711, 2015.

[9] B. Allen, “Experience, experiments, and the history of empiricism,” *Revista Portuguesa de Filosofia*, vol. 80, no. 3, pp. 805–812, 2024.

[10] K. M. Law, “Organisational sustainability from a systematic approach—a five-elements systems thinking (FEST) model,” *Enterprise Information Systems*, p. 2510349, 2025.

[11] B. Karagözoğlu and B. Karagözoğlu, “Contribution of Muslim scholars to science and technology,” *Science and Technology from Global and Historical Perspectives*, pp. 137–184, 2017.

[12] E. Neuenschwander, *Sourcebook in the Mathematics of Medieval Europe and North Africa*, 2018.

[13] B. Rice, E. A. González-Velasco, and A. Corrigan, *The Life and Works of John Napier*. Cham: Springer, 2017.

[14] T. Johnson and T. Johnson, "Practical mathematics: The development of probability theory," *Ethics in Quantitative Finance: A Pragmatic Financial Market Theory*, pp. 149–169, 2017.

[15] G. E. Box and G. C. Tiao, *Bayesian Inference in Statistical Analysis*. John Wiley & Sons, 2011.

[16] J. W. Tukey, "The future of data analysis," *The Annals of Mathematical Statistics*, vol. 33, no. 1, pp. 1–67, 1962.

[17] D. Donoho, "50 years of data science," *Journal of Computational and Graphical Statistics*, vol. 26, no. 4, pp. 745–766, 2017, doi: 10.1080/10618600.2017.1384734.

[18] P. Naur, *Concise Survey of Computer Methods*, 1974.

[19] S. Computing, "Formation of the international association for statistical computing," *Mathematical Programming*, vol. 15, pp. 119–120, 1978.

[20] G. Piatetsky-Shapiro, "Knowledge discovery in databases vs. personal privacy (draft)," *IEEE Expert*, 1995.

[21] J. Berry, "Database marketing," *Bloomberg*, Sep. 4, 1994. [Online]. Available: https://www.bloomberg.com/news/articles/1994-09-04/database-marketing. [Accessed: Jun. 15, 2025].

[22] C. Hayashi, K. Yajima, H. H. Bock, N. Ohsumi, Y. Tanaka, and Y. Baba, Eds., *Data Science, Classification, and Related Methods*. Springer, 2013.

[23] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, "From data mining to knowledge discovery in databases," *AI Magazine*, vol. 17, no. 3, pp. 37–37, 1996.

[24] C. F. J. Wu, "Statistics = Data science?," Inaugural lecture for the H. C. Carver Professorship, Univ. of Michigan, 1997. [Online]. Available: https://www2.isye.gatech.edu/~jeffwu/presentations/datascience.pdf.

[25] G. Jifa and Z. Lingling, "Data, DIKW, big data and data science," *Procedia Computer Science*, vol. 31, pp. 814–821, 2014.

[26] 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.

[27] L. Breiman, "Statistical modeling: The two cultures," *Statistical Science*, vol. 16, no. 3, pp. 199–231, 2001.

[28] National Science Board, *Long-Lived Digital Data Collections: Enabling Research and Education in the 21st Century*, Tech. Rep. NSB-05-40, 2005.

[29] T. H. Davenport, "Competing on analytics," *Harvard Business Review*, vol. 84, no. 1, pp. 98, 2006.

[30] J. Zhang, A. Fu, H. Wang, and S. Yin, "The development of data science education in China from the LIS perspective," *International Journal of Librarianship*, vol. 2, no. 2, pp. 3–17, 2017.

[31] A. Swan and S. Brown, *The Skills, Role and Career Structure of Data Scientists and Curators: An Assessment of Current Practice and Future Needs*. Programme/Project deposit, 2008. [Online]. Available: http://repository.jisc.ac.uk/245/

[32] I. O. Muraina, "Exploration of data science techniques in predicting students' relevant courses of study on getting to higher institutions," 2022.

[33] K. Cukier, "Special report: data, data everywhere," *The Economist*, 2010.

[34] M. Loukides, "What is data science?," 2010.

[35] H. Mason and C. Wiggins, "A taxonomy of data science," *dataists.com*, 2010.

[36] D. Taylor, "Battle of the data science Venn diagrams," *KDNuggets News*, 2016.

[37] P. Warden, "Why the term 'data science' is flawed but useful," 2011.

[38] G. Press, "A very short history of data science," *Forbes*, May 28, 2013. [Online]. Available: https://www.forbes.com/sites/gilpress/2013/05/28/a-very-short-history-of-data-science/.

[39] H. Ureta, "Coining data science," 2017. [Online]. Available: https://datascience.aero/coining-data-science/. [Accessed: Jun. 18, 2025].

[40] M. J. Graham, "The art of data science," in *Astrostatistics and Data Mining*, New York, NY, USA: Springer, pp. 47–59, 2011.

[41] M. Yıldız, "What is data science? A beginner's guide to data science," 2021. [Online]. Available: https://clarusway.com/what-is-data-science/. [Accessed: Jun. 13, 2025].

[42] D. Gunning and D. Aha, "DARPA's explainable artificial intelligence (XAI) program," *AI Magazine*, vol. 40, no. 2, pp. 44–58, 2019.

[43] D. Zha, Z. P. Bhat, K. H. Lai, F. Yang, Z. Jiang, S. Zhong, and X. Hu, "Data-centric artificial intelligence: A survey," *ACM Computing Surveys*, vol. 57, no. 5, pp. 1–42, 2025.

[44] O. Caelen and M. A. Blete, *Developing Apps with GPT-4 and ChatGPT: Build Intelligent Chatbots, Content Generators, and More*. O'Reilly Media, 2024.

[45] A. Allen, S. Markou, W. Tebbutt, J. Requeima, W. P. Bruinsma, T. R. Andersson *et al.*, "End-to-end data-driven weather prediction," *Nature*, pp. 1–3, 2025.

[46] Haque, A. K. M., Ridoy, A. A. I., Rayhan, M., & Porres, I. (2026). Emergence of Agentic AI: A Review on Evolution, Background, Working Principles, Applications, Adoption Factors, and Future Research Directions. *Computers, Materials & Continua*, *88*(2), 1.

[47] Kutz, J. N. (2026). *Data-driven modeling & scientific computation: methods for complex systems & big data*. Oxford University Press.

[48] Rajkumar, N., Viji, C., Nachiappan, B., Mohanraj, A., & Kovilpillai, J. A. (2026). Generative AI as an Accelerator for Data Engineering. *Generative AI-Powered Data Architectures: From Governance to Autonomous Analytics*, 57-82.

[49] Patel, A. K. (2026). Agentforce+ Data Cloud: The New Era of Real-Time, Adaptive, and Predictive Lightning Applications. *International Journal of AI, BigData, Computational and Management Studies*, *7*(1), 1-14.

[50] Farag, M. I. H. (2026). Artificial Intelligence and Strategic Decision-Making for Sustainable Governance. *Knowledge and Decision Systems with Applications*, *2*, 447-468.

[51] A. K. Daniel-Beston and A. Iorliam, *An Explainable Hybrid Deep Learning Framework for Unsupervised Network Intrusion Detection Using Isolation Forest and Bidirectional Long Short-Term Memory Autoencoder*. *Babylonian Journal of Internet of Things*, 2026, pp. 46–56.

[52] A. Iorliam and I. B. Iorliam, *Interpretable Stacking Ensemble Models for Accurate Shelf-Life Prediction of Mango Fruit Using Explainable Artificial Intelligence*. Unpublished, 2026.

[53] Evkaya, O., & de Carvalho, M. (2026). Using ChatGPT for Data Science Analyses. *Harvard Data Science Review*.

[54] Hossain, M. Z. (2026). Emerging trends in forensic accounting: Data analytics, cyber forensic accounting, cryptocurrencies, and blockchain technology for fraud investigation and prevention. *Journal of Artificial Intelligence and Technological Development*, 2(2), 183–208.

[55] Deora, R., Chilakapati, P., Selvarajan, G. P., & Chowdhury, A. G. (2026). Real-World Applications of Data Analytics and Machine Learning in Continuous Improvement for Operational Excellence. *Innovative Computing and Communications: Proceedings of ICICC 2025, Volume 8*, 81–88.

[56] Xyntarakis, M., & Antoniou, C. (2026). Data science and data visualization. *Mobility Patterns, Big Data and Transport Analytics*, 117–149.

[57] Jahani, H., Jain, R., & Ivanov, D. (2026). Data science and big data analytics: A systematic review of methodologies used in the supply chain and logistics research. *Annals of Operations Research*, 359(2), 1297–1354.

[58] Onyekwere, C. K., Nwankwo, C. K., & Apameh, D. G. (2026). A hybrid machine learning framework for multi-objective performance optimization and anomaly detection in maritime operations. *Innovation in Computer and Data Sciences*, 2(1), 1–10.

[59] Akintola, K. G., Ifedayo-Ojo, B. A., Olatunde, O. V., Muritala, O. E., & Lawal, S. O. (2026). Roles of Big Data Analytics on Customer Experience in eCommerce and Digital Financial Platforms. *Annals of Data Science*, 1–13.

[60] UmaDevi, A., Singh, J., Raha, S., Shaik, N., Turukmane, A. V., & Singh, I. (2026). Advances in Time-Series Analysis: Techniques and Applications for Predictive Forecasting. *Data-Driven Modeling*, 121–142.

[61] Kaya, O., Shah, A. S., Karabulut, M. A., Karahan, S. N., Osmanca, M. S., & Acır, N. (2026). Explainable artificial intelligence (XAI): Concepts, applications, challenges, and future perspectives. *IEEE Access*.

[62] Patil, R. S., & Moantri, S. (2026). Challenges and Opportunities in Real-Time Data Processing: Advancements and Limitations in Real-Time Data Analytics. *Artificial Intelligence and Machine Learning in Neurology*, 2, 647–683.

[63] Chen, L., Sim, M., Zhang, X., Zhao, L., & Zhou, M. (2026). Robust actionable prescriptive analytics. *Operations Research*, 74(1), 550–571.

[64] Çakir, E., & Tolga, A. Ç. (2026). A Review of Artificial Intelligence's Impact on Cybersecurity in the Big Data Era. *International Conference on Computational Science and Its Applications*, 182–192.

[65] Kabbaj, N., & El-Amin, M. F. (2026). An introduction to digital twins and big data analytics. *Mathematical Modeling for Big Data Analytics*, 91–105.

[66] Dhar, S., Das, A., Banerjee, J. S., Bhattacharyya, S., Platoš, J., Zelinka, I., & Mondal, N. K. (2026). Data-driven environmental intelligence: Unlocking the power of data for a sustainable planet. *Data-Driven Environmental Intelligence*, 1–27.

[67] Mamodiya, U., & Upadhyay, H. K. (2026). A Critical Review of Proactive Maintenance Strategies for Smart Factories in Industry 5.0. *Industry 5.0*, 111–125.