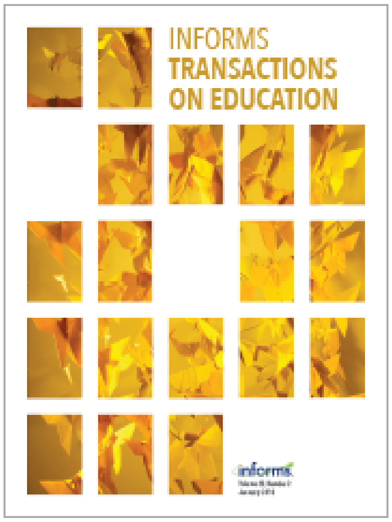


INFORMS Transactions on Education

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Data Analytics Research-Informed Teaching in a Digital Technologies Curriculum

Jing Lu

INFORMS TRANSACTIONS ON EDUCATION
Vol. 20, No. 2, January 2020, pp. 57–72
ISSN 1532-0545 (online)
http://pubsonline.informs.org/journal/ited

Data Analytics Research-Informed Teaching in a Digital Technologies Curriculum

Jing Lu[a]

[a] Department of Digital Futures, University of Winchester Business School, Winchester SO22 5HT, United Kingdom
Contact: jing.lu@winchester.ac.uk, http://orcid.org/0000-0001-9163-8795 (JL)

Received: December 10, 2018
Revised: February 17, 2019; March 23, 2019; April 23, 2019
Accepted: April 24, 2019
Published Online in Articles in Advance: January 2, 2020

https://doi.org/10.1287/ited.2019.0215

Abstract. In the business environment, the goal of data analytics can be characterized as improving decision making and its links to big data and other data-driven technologies. In UK higher education, degree apprenticeships are business-led and government-supported nationally recognized qualifications, where delivery is tailored to partner employer requirements. This paper focuses on the development of the data analytics specialism of the BSc Digital and Technology Solutions degree apprenticeship at the University of Winchester Business School informed by current research and practice. A data-driven analytical framework is first proposed to provide an overarching methodology for extracting knowledge and insights from (big) data. It covers key components of the analytics lifecycle from data management, data preprocessing, and integration through data modeling and business intelligence to insight management. Software tools related to collecting, cleansing, processing, analyzing, and visualizing data have been systematically discussed to provide the technological dimension. The methodology is then applied to the development of the specialist modules in data analytics, which represent the core thematic structure of the degree apprenticeship pathway. The culmination of the paper is an evaluation of the educational innovation of this digital technologies curriculum, highlighting teaching, learning, and assessment from the perspective of data-analytic thinking.

Keywords: business environment • information systems • data analytics • degree apprenticeship • digital and technology solutions • data-driven analytical framework • curriculum development • undergraduate specialist modules • research-informed teaching • analytical tools • data mining • business intelligence • insight management • decision support • big data

1. Introduction

1.1. Academic and Professional Context

The digital economy has facilitated an explosion in the data available to the world. This has affected businesses, jobs, education, and healthcare. The broad availability of data has led to a growing interest in technologies for extracting useful information and knowledge from data—the realm of data analytics (Provost and Fawcett 2013). Organizations are increasingly driven by data analytics, which is crucial to business strategy and improving decision making through its links to big data.

Within the University of Winchester (UoW) Business School, an Honours Degree Apprenticeship Programme started delivery in 2015 to encompass studying for an undergraduate degree in tandem with work-based professional training and experience. The BSc Digital and Technology Solutions (D&TS) aims to develop creative and responsible managers and leaders of the future able to meet the challenges of managing and leading digital projects and businesses in complex and changing global markets. The program equips degree apprentices with the skills that they need to work in graduate-level technology and digital roles, such as business analyst, cyber security analyst, data analyst, software engineer, and information technology (IT) consultant. It covers the priority areas of business transformation through cloud services, cyber security and risk management, and data analytics and insight management as well as other high-level aspects of digital technology for business.

Degree apprenticeships in the United Kingdom are business-led and government-supported nationally recognized qualifications developed in close collaboration with partner employers. Programs are tailored to ensure that apprentices develop the knowledge and skills that they require to further their careers and become digital specialists who are in tune with the future needs of their industry. The UoW currently works in partnership with a number of employers to support them in attracting and retaining their top digital talent, including CGI, Fujitsu, OceanWise, Premium Credit Limited, Transactor Global Solutions, and Walnut Unlimited.

The core D&TS degree covers the key concepts, skills, and competencies in digital and technology management alongside professional and project-based studies. D&TS also includes specific pathways, such as business management, cyber security analyst, and data analytics. This paper focuses on the development and delivery of the BSc Digital and Technology Solutions (Data Analytics) specialism informed by data analytics research and practice.

Following the literature overview presented below, Section 2 starts with a research methodology for data analytics—adapted from a data-driven framework used in the context of big data—which subsequently underpins the curriculum. Software tools related to collecting, cleansing, processing, analyzing, and visualizing data are discussed in Section 3—widely available data sources for digital technology projects have also been recommended. The methodology is then applied to the development of specialist modules for the data analytics pathway. Section 4 covers detailed discussion on the taught modules across three levels, including key themes and learning outcomes. Section 5 then highlights the innovative teaching and learning approach motivated by data analytic thinking—the mapping of modules to software tools, the structure of learning resources, and the design of assessment are all described before reflecting on outcomes and feedback so far during the first two years' delivery of the specialism. The paper draws to a close with some concluding remarks on the status of progress with the ongoing final year and a pointer to some future work in relation to the proposed BSc Data Science Programme at the UoW.

1.2. Literature Overview

Data science and analytics are postgraduate courses at many universities. Nonetheless, there is an interesting review of a business analytics curriculum for *undergraduate* majors from Wilder and Ozgur (2015). The appropriate skills level and breadth of knowledge required for business school graduates to be successful have been defined and discussed. Their proposed Business Analytics Programme is designed around five knowledge areas: project lifecycle, data management, analytical techniques, deployment, and a functional area. In addition, a capstone course is included to provide students with challenging project work. It is emphasized that business analytics offers a multidisciplinary education, where individual courses involve a combination of business and quantitative disciplines. Among the guidelines for successful program implementation, universities are advised not to stray from the important objective of preparing the *data-savvy managers* of the future.

Data have become some of the most vital resources of this century, and they have been used across virtually every major function in business, government, healthcare, education, and other nonprofit organizations. Data support a variety of commercial purposes and also, provide key inputs to decision models. Many organizations are overwhelmed by today's big data, which refer to data sets so large and complex that it would be impossible to analyze them using traditional methods. Big data have been defined by Gartner in 2001 as "high-volume, high-velocity and/or high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making" (https://www.gartner.com/it-glossary/big-data). *Volume* is the amount of data that is created. *Velocity* refers to the speed at which new data are generated as well as the speed that they move around—this can help to appreciate the difference between large data sets and big data. *Variety* is the number of types of data and places that are creating it.

The three Vs can be used to set up a common ground and also, point out where big data challenges and opportunities arise. With many forms of big data, quality and accuracy are less controllable. IBM data scientists break big data down into four dimensions by adding a fourth V as *veracity*, referring to the trustworthiness of the data. However, all of this volume of fast-moving data of different variety and veracity has to be turned into *value*, which leads to the fifth V for big data (Marr 2017). Big data challenges are often related to the characteristics of the data itself, with Sivarajah et al. (2017) summarizing key aspects as volume, velocity, variety, variability, veracity, visualization, and value. This leads to significant data processing opportunities through data acquisition and warehousing, data mining and cleansing, data aggregation and integration, analysis, and modeling as well as data interpretation.

Some analysis must be applied to the data, because the value is not in raw bits and bytes but rather, the insights gathered from them. Available data analytics technology today is stretching from simple statistical tools to more sophisticated machine learning approaches, with deep learning among the latest trends. This includes traditional data mining and advanced analytics techniques, such as classification, clustering, regression, and association rules, along with machine learning and artificial intelligence techniques, such as neural networks, decision trees, and pattern-based analytics. Moreover, time series analysis can be used for analyzing sequences of data points that represent values over successive periods. Text analysis, social network analysis, and sentiment analysis can also be applied if the data are in the appropriate form (Elgendy and Elragal 2016). Additionally, graph analyses can be considered when representing complex networks, path analyses can describe directed dependencies among variables, and clickstream analyses can be used for web data.

Big data technologies, with their potential to ascertain valued insights for an enhanced decision-making process, have attracted substantial interest from both academics and practitioners. Data analytics is increasingly becoming a trending practice that many organizations are adopting with the purpose of determining valuable information. There is a range of analytical techniques used to extract value and insight from data, including descriptive analytics, inquisitive analytics, predictive analytics, prescriptive analytics, and preemptive analytics (Sivarajah et al. 2017). People understand the power and importance of big data, but many fail to grasp the actionable steps and resources required to utilize it effectively. Marr (2016a) aims to fill the knowledge gap by showing how major companies are using big data every day from an up-close and on-the-ground perspective.

Despite a breakthrough in technological development over recent years, there is limited understanding of how organizations can translate big data into actual social and economic value. Günther et al. (2017) have conducted a systematic review of the information systems literature on the topic and concluded that two sociotechnical features of big data that influence value realization are portability and interconnectivity. They argue that, in practice, organizations need to continuously realign work practices, organizational models, and stakeholder interests in order to reap the benefits from big data.

Although the constantly growing body of academic research on big data analytics is mostly technology oriented, a better understanding of the strategic implications of big data is urgently needed. Wang et al. (2018) have identified five different strategies reflecting how big data analytics can be effectively used to create business value through analytical capability for patterns, unstructured data-analytical capability, decision support capability, predictive capability, and traceability.

In terms of methodology, the Cross-Industry Standard Process for Data Mining (CRISP-DM) has been an industry standard for the last two decades. It breaks up the overall task of finding patterns from data into a set of well-defined subtasks: business understanding, data understanding, data preparation, modeling, evaluation, and deployment (Shearer 2000). In order to convert the promise of big data into real-world business results, a five-step SMART model has been proposed by Marr (2015): Start with Strategy, Measure Metrics and Data, Apply Analytics, Report Results, and Transform Business. Other frameworks include, for example, the data analytics lifecycle of EMC Education Services (2015) and the Knowledge Discovery in Databases (KDD) process model (Fayyad et al. 1996). Additionally, based on contemporary research and practice in data analytics (Lu 2018), a data-driven framework for business analytics has been adapted for this paper to provide an overarching methodology.

2. Methodology

2.1. Data-Driven Analytical Framework

A data-driven analytical framework is proposed (see Figure 1), which follows the underlying structure introduced by Lu (2018). It covers key aspects of the analytics lifecycle from data management, data preprocessing, and integration through data modeling and business intelligence (BI) to insight management—indicated down the right-hand side of the framework in Figure 1. The five Vs of big data are all referenced within the framework and labeled down the left-hand side of Figure 1. An important feature across the main stages of the framework is the sixth V representing *visualization*.

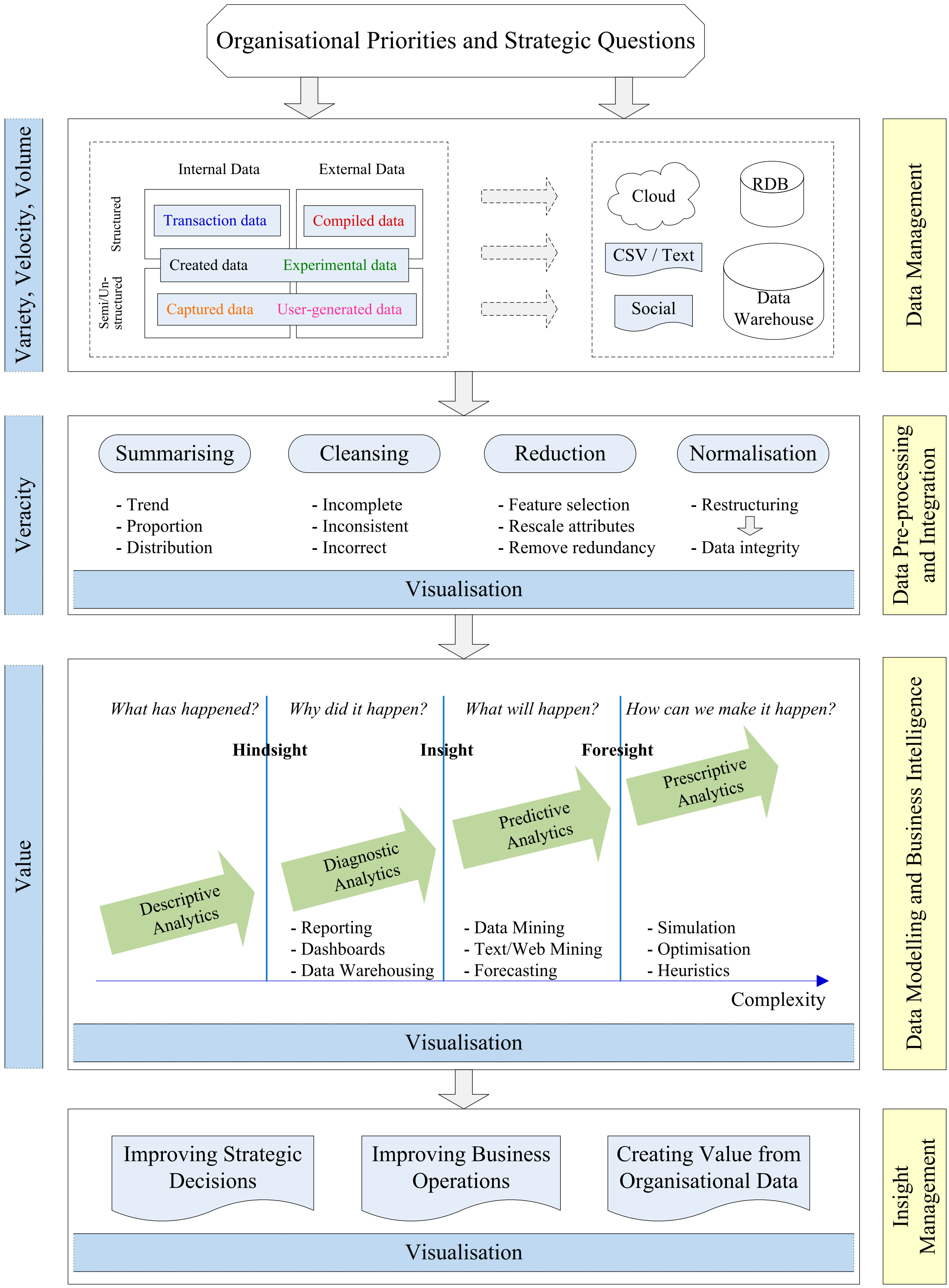
Instead of starting with the data itself, it is vital that analytics for the business environment starts with strategy. A good data strategy is not determined by what data are readily or potentially available—it is more about what the business wants to achieve and how data can help to get there (Marr 2017). Having a clear data strategy informs organizational priorities and can reveal unanswered questions. This will help to identify what data are needed—often a combination of existing “small” data and newer faster-moving “big” data.

2.1.1. Data Management. Data management is the data layer of the framework, where data exist in a variety of formats and types: *structured*—data or information that have a predefined data model or are organized in a predetermined way (e.g., information stored in relational databases, where each record in a table follows the same structure); *semistructured*—data may have a certain structure, but not all of the information follows the same structure (e.g., XML, email, and Facebook posts); and *unstructured*—either does not have a predefined data model or is not organized in a specified way (e.g., photos, graphic images, videos, and text files).

Data can be collected and/or generated through various methods (Marr 2015): for example, *created data* are collected by market research or focus groups, *transaction data* are generated every time that a customer buys something (online and off), *captured data* are related to information gathered passively from an individual’s behavior (e.g., search terms or location data), and *user-generated data* are consciously or knowingly generated by individuals and companies (e.g., blogs, tweets, and YouTube videos).

2.1.2. Data Preprocessing and Integration. Informed decision making is highly dependent on good-quality data that have been captured correctly, that have been

Figure 1. A Data-driven Analytical Framework in the Context of Big Data



Note. CSV, comma-separated values; RDB, relational database.

carefully defined, and for which data storage, retrieval, and reporting have been optimally designed to meet the needs of the business. Data preprocessing is used to prepare raw data sets from various sources before integration for use in the next stage. It typically consists of the following: *data summarizing*—not only describing data but also presenting them visually, providing a simple overview of data while helping to determine quality before analysis; *data cleansing*—detecting, correcting, or removing corrupt or inaccurate values and then replacing, modifying, or deleting the dirty data as required; *data reduction*—creating transforms of data through rescaled attribute values and attributes, identifying and removing irrelevant and redundant features; and *data normalization*—establishing subsets of data that are logical to compare and contrast, improving the performance of analysis and further interpretation.

2.1.3. Data Modeling and Business Intelligence. Data modeling offers analytical techniques of varying complexity that can address the questions highlighted in Figure 1. The quality of decision making through the hindsight, insight, and foresight achieved is the remit of BI support in organizations. BI is a technology-driven process for analyzing data and presenting actionable information to help executives, managers, and other corporate end users make informed business decisions. Data analytics can be divided into the following four key categories in order to facilitate the fifth V of big data, value: *descriptive* analytics manipulates data to provide insight into what is happening in the business, *diagnostic* analytics might use the same metrics but helps to understand why an event is happening, *predictive* analytics estimates the likelihood of a future outcome (e.g., understanding end-of-year sales or predicting the items that customers will purchase together), and *prescriptive* analytics not only predicts what will happen but also why it will happen, providing recommendations on actions using simulation and optimization algorithms.

2.1.4. Insight Management. Data are the most valuable asset of many organizations today but only if interpretation and impact deliver competitive advantage. Insight represents information from which value is derived and subsequently, that makes a difference to the business. Insight management is concerned with understanding organizational priorities and requirements and then, managing the information flow to achieve positive outcomes. There are three core areas where data really matter to organizations (Marr 2017): they enable businesses to collect better market and customer intelligence in order to improve decision making, help companies gain efficiencies and improve their operations, and provide an opportunity for all organizations to build big data into their product/service offering, ascribing real value to the data itself.

2.2. Data Analytics Methods/Techniques

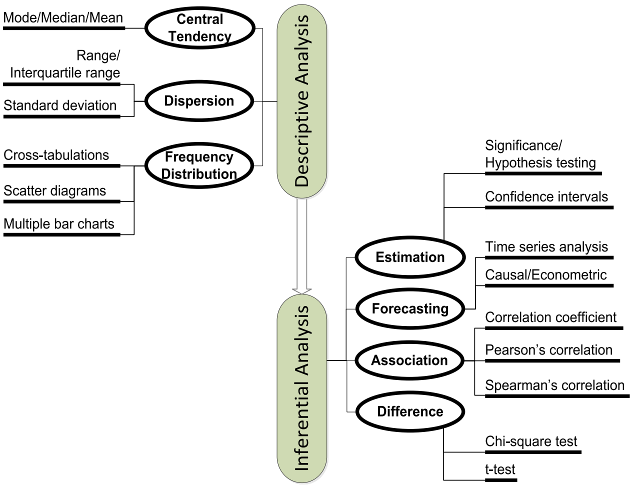
Real-world data-analytical applications occur in different domains, using not just one set of computational tools but a merger of (for example) applied mathematics, numerical analysis, informatics, and statistics. Diverse applications and case studies often share a similar core mathematical or statistical problem, and therefore, they can be addressed using common techniques across data management, computational analytics, data modeling, and business intelligence. There is a range of different analytical methods that can be brought to bear in the areas of statistical analysis and simulation, machine learning and data mining, graphical visualization, information modeling, and decision support.

2.2.1. Statistical Methods, Simulation, and Optimization. Statistics is intimately tied to scientific understanding of the world, and therefore, it is natural to apply statistical concepts to the business environment to help make good business decisions and improve performance. Figure 2 represents the process of data exploration by indicating the tasks involved in descriptive analysis as well as the process of inferential analysis using statistical techniques.

Simulation can be used as a methodology to analyze the behavior of business activity as well as gain an understanding of how a present system operates. Although optimization was originally in the domain of operations management, it is now used in all areas of business to determine the most effective solutions. Optimization models are prescriptive when their outcomes are recommendations for decision makers on what to do (Evans 2013).

2.2.2. Machine Learning and Data Mining. Machine learning and data mining are fundamentally involved with the extraction of information from data sets—consequently, there is often some overlap with statistical inference. Four common analytics techniques are classification, clustering, regression, and dimension reduction. Corresponding problems can be broadly segregated by whether the output is continuous or categorical (classes) and whether it is supervised (includes desired outputs) or unsupervised.

Figure 2. Data Exploration and Statistical Analysis



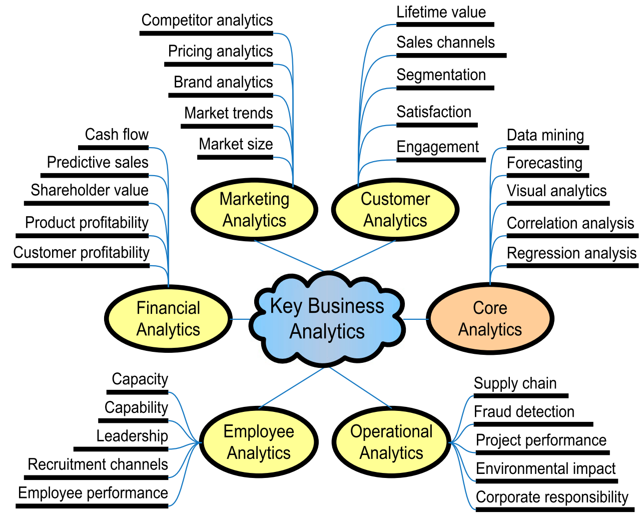
Classification aims to accurately allocate objects to a set of predetermined classes according to the *differences* among data elements. *Clustering* divides data elements into different groups based on the *similarity* between elements within a single group. *Regression* is similar to classification, but the output variable is continuous instead of categorical (a set of classes), the aim being to accurately and precisely estimate or predict the response. *Dimension reduction* is a common tool in big data analytics, where the performance of algorithms is often improved with a limited number of variables when large numbers of records can be collected into homogeneous groups (Shmueli et al. 2016).

2.2.3. Visual Analytics. Visualization focuses on techniques for presenting data in visual ways to support understanding of the underlying patterns whether performed for research, science, or decision-making purposes. It plays a key role in business intelligence, where data, knowledge, and users are tightly related. Visual analytics is an integrated approach that combines data analysis with data visualization and human interaction. It normally commences with a predetermined task and then goes through an iterative process to get the data, choose an appropriate visual structure, view the data, form insights, and then, act. This process involves users moving across different steps as new data insights and questions are revealed.

Visual analytics connects the strengths of visualization and analytical reasoning to help with decision making. This can support understanding the results of experiments and simulations and show the content of large-scale data sets in meaningful ways. Visual analytics emerged as a response to the *information overload* problem and the way that decision makers are bombarded with irrelevant information or inappropriate data. It allows exploration of data through both visual and theoretical models to potentially understand underlying patterns and generate new knowledge. The result often produces a graphical interface illustrating the trends and changes in data over time, which is effectively used to communicate the message to nonexpert users.

2.2.4. Domain-Based Analytics. The generation of the right information and insights for decision makers is a major challenge for many organizations. The challenge is in coping with a burgeoning amount of multifarious data, analyzing data, and ensuring that it reaches decision makers in a timely and meaningful manner. Key analytics in business can help organizations apply tools to turn data into valuable insights that enable them to better understand their customers, optimize their internal processes, and identify cost savings and growth opportunities (Marr 2016b). Figure 3 shows some representative domain-based approaches within the business categories of customer analytics, marketing analytics, financial analytics, employee analytics, and operational analytics alongside core analytics.

Figure 3. Domain-based Business Analytics



Health informatics is another example of domain-based analytics. Lu and Keech (2015) reviewed how emerging database technologies can be used to provide additional value from available health information. They also cover the integration of data from various sources and how to share and connect the resources. Subsequent research (Lu et al. 2017) presented a methodology for health data analytics through a case study for modeling cancer patient records. Modern data-mining software and visual analytics tools permit new insights into temporally structured clinical data. The challenges and outcomes of the application of such software-based systems to this complex data environment were also reported. A range of approaches was tested on the resulting data set, including multidimensional modeling, sequential patterns mining, and classification.

3. Digital Technologies

3.1. Software Tools

Technologies related to collecting, cleansing, storing, processing, analyzing, and visualizing big data are evolving at a fast pace. The choice of software can depend on a range of factors. Several analytical tools deployed in the program are described in Table 1, expanding on the treatment in Lu (2018). This derives from the data-driven analytical framework in the previous section and outlines specific techniques in preprocessing, data modeling, and visualization.

Microsoft Excel is still one of the most widely used tools for business analysis. It provides standard spreadsheet functionality, including generating graphical formats, which makes it useful for data manipulation and presentation. In addition, XLMiner is the

Table 1. Analytical Tools and Techniques

Tools and techniques	Preprocessing	Data modeling	Visualization
Excel: electronic spreadsheet program	Show missing data in pivot table	Analysis ToolPak, XLCubed, PowerPivot	Graphs/charts, PivotCharts
Alteryx: data preparation, blending, and analysis	Drag and drop to eliminate SQL coding/formulas	Prepackaged procedures for predictive analytics	Workflow for self-service data analytics
SPSS: statistical analysis software package	Validate data, unusual cases, and optimal binning	Descriptive statistics, inferential analysis, prediction	Chart builder, Graphboard Template Chooser
R: statistical computing and graphics environment	Raw → correct → consistent data	Statistics, time series, classification, clustering	Base, grid, lattice, and ggplot2
iNZight: data exploration and insight generation	Quick explore → missing values	Relationships, estimation, time series	Visual Inferential Tools
Weka: machine learning and data-mining software	Discretize, normalize, attribute selection	Classification, clustering, association rules, and sequential patterns mining	Plot, receiver operating characteristic, tree/graph/boundary visualizer
Tableau: data visualization and analytics	Joins, unions, splits, and pivots	Segmentation and cohort analysis, predictive analysis	Interactive and visual analytics
Analytic Solver: predictive and prescriptive analytics in Excel	Data cleansing via dimensionality reduction	Simulation, optimization, and data mining	View detailed charts of data using Chart Wizard
Python: flexible language for general purpose programming	Standardize, cleanse, transform, normalize	Statistics, machine learning, network analysis	Matplotlib, Seaborn, ggplot, Bokeh, pygal, Plotly, etc.

Note. SQL, Structured Query Language.

comprehensive data-mining plug in for Excel, now superseded by Analytic Solver.

Alteryx is a tool primarily designed to extract, transform, and load data (https://www.alteryx.com). The online self-paced training materials provide hands-on experience with solving practical business problems (e.g., connecting to and cleansing data from multiple sources, improving data quality with profiling, and offering repeatable workflow design to assist data integrity during the process).

The IBM SPSS software platform used to be well known just for descriptive and statistical analyses. However, within the context of big data applications, it is now more integrated with data mining and machine learning algorithms (IBM 2018). SPSS Modeler can be used for data preparation and discovery, predictive analytics, model management, and deployment, whereas plug ins are also available for R and Python.

R has become one of the most popular software environments for data analysis and manipulation (https://www.r-project.org). It is commonly used for big data management and analysis, and it provides a wide variety of statistical and graphical techniques. Developed using R, iNZight software and its online materials provide interfaces for data exploration and statistical understanding that are intuitive, covering multivariable graphics and time series analysis (https://www.stat.auckland.ac.nz/~wild/iNZight/).

Weka is a long-standing open source workbench for knowledge discovery and data mining (Hall et al. 2009). It contains a large number of algorithms for data preprocessing, feature selection, classification, clustering, sequential patterns mining, and finding association rules. The main Knowledge Explorer interface gives access to the workbench and a variety of panels, including one for visualization.

As illustrated in Figure 1, visualization is a key aspect of analytics. Tableau Software provides interactive visualization facilities, which focus on delivering insights from data for business intelligence (https://www.tableau.com). Features include connecting, integrating, blending, cleaning, and visualizing data from multiple sources as well as segmentation and cohort analysis, what-if analysis, and predictive analytics.

The Analytic Solver platform is a versatile tool for predictive and prescriptive analytics in Excel that provides functionalities in the areas of risk analysis, simulation, optimization, forecasting, and data mining (https://analyticsolver.com/). It includes capabilities for automatically determining optimal decisions using stochastic programming and robust optimization methods and new ways to apply data mining and visualization to simulation results.

Last but not least, the general purpose Python programming language has become even more popular and powerful with the advent of data science (https://www.python.org). Data analysis, information visualization, and machine learning techniques can be applied through Python software libraries and toolkits, such as Pandas, Scikit-Learn, StatsModels, and Matplotlib, to gain additional insight from data.

3.2. Data Sources

Data scientists are expected to be proficient in searching, extracting, understanding, and presenting information from structured and unstructured data sources. Keeping up to date with the latest trends in technological development is important for effective analytics.

Moreover, degree apprentices find it useful to be able to access resources for digital technology projects. Table 2 is a list of data sources used by the program in the areas of data cleansing, data visualization, machine learning, and data mining—adapted from material on https://www.dataquest.io/ (hyperlinks are provided to individual repositories).

The data.world source is a versatile place to search for, copy, analyze, and download data sets. In addition, users can upload their data to data.world to collaborate with others. One key differentiator of data.world is that there are tools built to facilitate working with data (e.g., writing SQL queries within their interface to explore data and join multiple data sets). They also have software developer kits for R and Python to make it easier to acquire and work with data in the tool of choice. The website offers the ability to automatically match and enhance data: for example, identifying when data contain a matchable column and offering to pull in a clean version based on standard industry classifications or suggesting additional related data to join into analysis, enhancing options to *slice and dice* data.

Tableau Public is a free platform that can connect to a spreadsheet or file and create interactive data visualizations. Its resources provide sample data sets to help initiate a student analytics project in the categories of business, government, science, technology, health, education, and sports.

When people are working on a machine learning project, they sometimes want to be able to predict a variable from the other variables in a data set. In order to be able to do this, it is necessary to make sure that data have been cleaned, that there is an interesting target column to make predictions for, and that the other variables have some explanatory significance for the target column.

There are several online repositories of data sets that are specifically for data mining or machine learning. The UCI Machine Learning Repository is one of the oldest sources of data sets on the web. Although the data sets are user contributed and thus, have varying levels of structure and documentation, the vast majority are clean and ready for machine learning to be applied. UCI is the first stop when looking for data sets in science and engineering, but it can also be used for business.

Kaggle is a data science community that hosts machine learning competitions, and there is a variety of interesting data sets on the site. Kaggle has both live and historical competitions where people can download data by signing up—each competition has its own associated data set. Externally contributed data sets can be retrieved from the new Kaggle Datasets offering.

4. Data Analytics Pathway Development

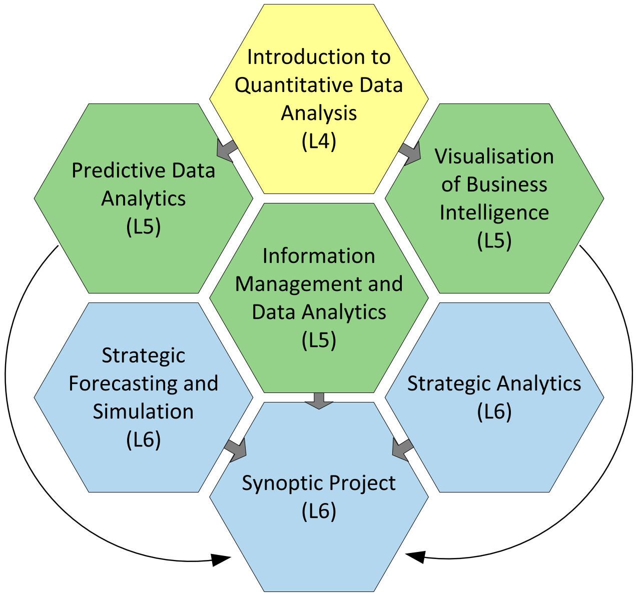
4.1. Overview

Students undertaking a BSc Digital and Technology Solutions (Data Analytics) degree apprenticeship at the UoW learn how to use data to develop significant business insights, helping companies to improve both their strategic decisions and their performance. Although Figure 1 provides the framework and motivation here, Figure 4 illustrates the schematic structure of the seven specialist modules for the data analytics pathway—from the one at Level 4 (first year) to the three modules at each of Levels 5 and 6. Arrows indicate pathway prerequisites in the modular flow.

Table 2. Data Sources for Digital Technology Projects

Data sources	Description
Data cleansing	
data.world	A social-based data source that allows users to share/clean/improve data collectively—can write SQL within the interface to explore data and join multiple data sets
The World Bank	The platform provides several tools, like Open Data Catalog, world development indices, education indices
Reddit	A community discussion site that has a section devoted to sharing interesting data sets
Data visualization	
FiveThirtyEight	Interactive news and sports site with data-driven articles—each data set includes the data, a dictionary, and a link to the story
FlowingData	Catalog of data sources described in detail with examples—it explores how statisticians, data scientists, and others use analysis and visualization
Tableau Public	Sample data for visual analytics in the categories of education, public, government, science, technology, health, business, sports, and entertainment
Machine learning	
UCI/ML Repository	Established source of data sets online—vast majority are clean and ready for machine learning
Kaggle	A data science community that hosts machine learning competitions—contains externally contributed data sets
Quandl	Financial data sets—useful for building models to predict economic indicators/stock prices

Figure 4. Data Analytics Specialist Modules



The Synoptic Project represents a substantial piece of work designed to showcase the application of specialist skills, knowledge, and behaviors using the occupational domain for the real-world problem or business scenario. For example, a data analyst would be expected to demonstrate a range of analytical techniques, such as data mining, time series forecasting, and information modeling, to identify and predict trends and patterns in data. The taught specialist modules are described in more detail in the following subsections, including the topics covered and learning outcomes.

4.2. Level 4 Specialist Module

The aim of the first-year undergraduate program is to build a strong foundation in digital technologies and information systems. Students are introduced to academic study and the different core areas required of data analysts, including business analysis and quantitative data analysis. The learning outcomes at this program level include the ability to do the following:

- State and explain relevant models, principles, and practices applicable to the study of information systems and quantitative analysis.
- Collect, manage, and present information, ideas, and concepts and interpret data using suitable techniques.

There is just the one specialist module at Level 4.

4.2.1. Introduction to Quantitative Data Analysis (IQDA).

This module introduces the apprentice student to quantitative analytics concepts, procedures, and software tools (Excel and SPSS) for specific data analysis tasks, which are fundamental to the work of the professional data analyst. It takes a real-world approach to understanding the value and meaning of data as well as collecting and preparing data for processing. This provides students with a practical knowledge of quantitative analysis, which also underpins more advanced study found later in the pathway. By the conclusion of this module, a student will be expected to be able to do the following:

- Identify appropriate analytical tools and techniques for use in data analysis.
- Describe ways in which data are collected for business decision making and new insight.
- Apply industry standard tools and methods used in interpretation and analysis of quantitative data.
- Explain how conclusions gained from analyzing business data can be reported.

Introduction to quantitative data analysis (IQDA) topics include introduction to quantitative research; summarizing and presenting data; descriptive analysis; analysis of relationships, differences, and trends; and moving from data to insight.

4.3. Level 5 Specialist Modules

In the second year, students explore data analytics in more depth and gain experience with different software tools and environments. They pursue topics in information management, predictive modeling, and visualization. At this program level, the learning outcomes include the ability to do the following:

- Use business intelligence and knowledge in the design and testing of computer systems for information management.
- Effectively use a variety of techniques, such as data mining, predictive analytics, and visualization, for complex data analysis through the whole data lifecycle.

There are three specialist modules at Level 5—all of the topics are organized around themes (also called blocks) of study that reflect the main areas of interest.

4.3.1. Information Management and Data Analytics (IM&DA).

This is a core module across the whole Digital and Technology Solutions Degree Apprenticeship Programme. The information management and data analytics (IM&DA) module encapsulates the challenges faced in deriving insights from data to underpin fact-based decisions. It first considers information management, examining the different types of systems and data that can flow into organizations while evaluating the characteristics and value of diverse data sources. With an awareness of these issues, data may be creatively integrated and analyzed to deliver insights. Data analytics is then explored from the perspective of decision makers. The three themes for the IM&DA module are in chronological order.

4.3.1.1. Theme 1—Database Management and SQL.

The relational model remains the most widely used

Database Management System (DBMS) model in industry for many reasons, not least meeting organizational needs. It is well understood, and there are many robust and mature products based on the relational model. This theme includes topics such as the database development process, relational databases and Relational Database Management Systems (RDBMS), entity relationship diagrams, structured query language, database design and normalization, and database integrity.

4.3.1.2. Theme 2—Data Warehousing and Information Modeling. As operational databases grow, management comes to realize that they represent an increasingly useful information resource for the business. The operational data stored in databases and a next generation of tools and methods can lead to new and better opportunities for the analysis of business data. The business intelligence and data warehousing terms are often used in the context of this advance. This theme not only includes the topics of BI and data warehousing but also, includes Extraction Transformation Loading (ETL), measures versus dimensions, Online Transaction Processing (OLTP) versus Online Analytical Processing (OLAP), Star schema, Snowflake schema, and OLAP cubes.

4.3.1.3. Theme 3—Data Mining and Knowledge Discovery. Organizations are becoming increasingly aware that there may be undiscovered knowledge in operational databases that has the potential to benefit the business. This theme introduces a cross-industry standard process model for data-mining projects, which offers a structured approach. An informative case study is presented that is relevant to business management and sufficient to demonstrate the data-mining lifecycle, prospective opportunities, and several of the key issues. This theme includes topics such as data mining, knowledge discovery in databases, CRISP-DM, market basket analysis, and association rules mining.

4.3.2. Predictive Data Analytics (PDA). This module provides experience of predictive modeling and analytics across a range of domains, promoting relevant higher-level practical skills to create data visualizations and carry out analyses. It aims to help students understand and unlock the power of large data sets—topics are organized around two themes.

4.3.2.1. Theme 1—Statistical Computing and Graphics. Statistical techniques offer methodical approaches to predictive analytics and reduce the danger of reaching erroneous conclusions owing to bias and uncertainty. Measuring and predicting human behavior in big data are the main case study for this theme, which includes topics such as predicting stock market movement and measuring and predicting spread of disease. R is the software environment used here, and it is demonstrated through the essentials of R programming, exploratory data analysis, data manipulation, and predictive modeling.

4.3.2.2. Theme 2—Machine Learning for Data Mining. Machine learning methods allow the analysis of data to obtain data-driven insights, providing for more evidence-based decision making. The classification approach is the focus for this theme, which includes topics such as building a classifier and evaluating a classifier's performance. The Weka software environment is recommended for machine learning and data mining because of its large collection of algorithms and visualization tools for data analysis and predictive modeling.

4.3.3. Visualization of Business Intelligence (VBI). This module focuses on techniques for data extraction and preparation while analyzing data in visual ways to generate insight for business intelligence and decision making. The business need is often for data to be presented in perceptible, comprehensible, relevant, and usable visual forms to communicate complex ideas that support decision making. The choice and selection of a range of visual formats are considered and practiced using examples related to real-world applications. Students have the opportunity to develop skills with visualization tools used in commerce and industry.

4.3.3.1. Theme 1—Data Blending and Integration. This theme introduces the activities of preparing data for presentation through cleansing and validation processes. Alteryx is used as a platform for self-service data analytics management, which includes workflow on input data, prepare and blend, enrich, analyze, share, and output data. An Alteryx tutorial checklist is created to help students complete practice activities from beginner to intermediate levels and then, more advanced levels.

4.3.3.2. Theme 2—BI and Visual Analytics. This theme introduces students to presentational techniques for visualization as a form of recording, understanding, and communicating transformed data. Visualization techniques are important, because they can present large and overwhelming amounts of multisource and multiformat data in a meaningful manner. The theme covers the analysis of data using industry-standard techniques, such as Tableau, highlighting the advantages and limitations of a wide range of visualization approaches from the customary statistical charts through to the more complex formats.

4.4. Level 6 Specialist Modules

In the final year, students pursue advanced topics in strategic analytics (SA), forecasting, and simulation.

They also undertake a synoptic project to showcase the analytical skills that they have developed. At this program level, the learning outcomes include the ability to do the following:

- Apply the concepts and principles in key areas of information systems, including data and information management, computer simulation, decision support, and data mining.
- Create formal models for specific organizational tasks and processes, using them to discover hidden patterns and propose optimizations, develop new models and methods if necessary, and recommend and influence organizational improvement based on strategic data analysis.

Apart from the synoptic project, there are two specialist modules at Level 6—note that both the strategic forecasting and simulation (SF&S) and the SA modules are new and being delivered during the 2018–2019 academic year.

4.4.1. Strategic Forecasting and Simulation (SF&S).

This module covers the business prediction topics of forecasting and simulation to develop advanced models and solutions to real-world problems. Forecasting is the analysis of trends in data, and it is a subject related to probability, risk, and uncertainty. Simulation, in a business context, is about constructing data-driven models to emulate real-world systems with sufficient fidelity and validity so that the possible impacts caused by changes to component variables can be explored experimentally. Students have the opportunity to develop advanced spreadsheet modeling and problem structuring methods.

SF&S topics include fundamental considerations in business forecasting; exploration and visualization for time series data; performance evaluation and reporting; and time series analysis, smoothing methods, and regression-based models. Different tasks are designed for each topic, and in terms of the implementation, students can either choose using R—which has been developed at Level 5 in the PDA module—or the new Analytic Solver software.

4.4.2. Strategic Analytics (SA).

This module provides students with a deeper understanding of how data are used by strategic decision makers, covering the approaches to data analysis as well as strategic data analytics challenges. Students gain advanced knowledge through studying the concepts of big data and data storage. The current strategic issues of concern to the data analyst are also considered. The module concludes with a data analysis case study, where the student is required to work through the lifecycle of data analytics using appropriate techniques and methods, reporting on the findings and making critical recommendations to a given stakeholder.

Strategic analytics topics follow the SMART model mentioned previously in Section 1.2 (i.e., start with strategy, measure metrics and data, apply analytics, report results, and transform business). Python is used as the main programming language here and demonstrated through the basics of coding; filter, clean, and visualize data; combine and transform data; and advanced analytics techniques.

5. Teaching and Learning Innovation

5.1. Data-Analytic Thinking

One of the primary goals of teaching data analytics is to help degree apprentices view business problems from a data perspective and understand strategic approaches to extracting useful knowledge from data. A data perspective will provide students with a framework to systematically analyze real-world business problems. As students get better at data-analytic thinking, they will develop intuition as to how and where to apply creativity and domain knowledge.

At a high level, data analytics can follow a set of fundamental principles that guides the extraction of knowledge from data (Provost and Fawcett 2013)—four of the principles have been adopted to underpin educational innovation for the Digital and Technology Solutions (Data Analytics) pathway at Winchester.

Principle 1. *Data and the capacity to extract useful knowledge from data should be regarded as key strategic assets.*

Typically for business school students, the popular method to collect data is through primary research (e.g., interviews and questionnaire surveys). On the one hand, it is time consuming; on the other hand, the data can be limited and not representative—performing useful analysis can often be a challenge. Through online resources deployed on the Canvas virtual learning environment at the UoW, a rich list of secondary data sources has been published for students to search, review, and download: for instance, in the categories of marketing, sales, accounting, finance, and the wider business context—Table 2 illustrates some sample data sources for student projects.

Principle 2. *Extracting useful knowledge from data to solve business problems can be treated systematically by following a process with reasonably well-defined stages.*

Several established models have been introduced during teaching (e.g., KDD, SMART, CRISP-DM, and the Problem Plan Data Analysis Conclusion model). Students are required to either (1) pick one of these process models or (2) adapt their own data-driven framework/model for analytics—in order to support their research approach and guide the analytical process for their assessments. This practice has been successful through the formative and summative assessments to date, culminating in (for example) quantitative analysis, predictive analytics, and business intelligence reports.

Principle 3. *From a large mass of data, information technology can be used to find informative descriptive attributes of entities of interest.*

All of the tools and techniques listed previously in Table 1 have been used in teaching for the data analytics pathway—see Table 3 for the mapping with specialist modules and themes across three levels. Among the list, some software licenses are provided by the UoW (e.g., Excel and SPSS), and some software is open source and free to use (e.g., SQL, R, iNZight, Weka, and Python). Other commercial software companies provide the educational licenses for university students (e.g., "Alteryx for Good" is a program designed specifically for nonprofits, students, and educators; "Tableau for Teaching" provides one-year licenses for instructors, university laboratories, and students; and finally, "Analytic Solver Platform for Education" is a special version of Frontline Systems' software for Microsoft Excel, including the former RSPE and XLMiner capabilities).

Principle 4. *Formulating data-analytical solutions and evaluating the results involve thinking carefully about the context in which they will be used.*

Case studies have been used across the different levels/modules—see Table 3 again for illustrations. Analyzing case studies improves students' ability to approach problems *data analytically*. Promoting such a perspective is another of the primary goals in the teaching. When faced with a business problem, students should be able to assess whether and how data can be used to address issues and improve performance. It is important to realize that business strategy and decision making are increasingly driven by data analytics. Degree apprentices will move on to apply the analytical skills that they have developed to their own workplaces, formulating domain-based solutions in a relevant context.

5.2. Teaching, Learning, and Assessment

Weekly teaching of specialist modules each semester is clearly scheduled through the "Pre-Reading," "Lecture Notes," and "Seminar Practice Activities." Teaching materials cover both a theoretical understanding and a practical understanding of the key methods for data analytics—in a corresponding business decision-making context. Real-world cases have been used within modules to demonstrate the application and interpretation of data analytics methods, where data are either searched for/collected by students according to their interest or provided by the tutor. This enhances learning and provides students with the opportunity to understand the strength of data analytics and challenges that arise in the process.

Figure 5 presents an overview structure of learning resources for the data analytics specialist modules. Resources on the Canvas VLE have been designed in such a way that allows the learning outcomes to be achieved within a flexible "Prepare–Do–Improve" iterative cycle. This independent and activity-based learning is at the heart of the student experience and comprises (1) *prepare*—students are advised to undertake guided prereading to prepare for the lecture/seminar; (2) *do*—a series of activities is included during class to help students further understand the analytical concepts, methods, and theories; and (3) *improve*—formative feedback is provided on a regular basis so that students have opportunities to improve their practice and be ready for their summative assessments. In addition, Canvas has a "course analytics" function that helps tutors to monitor students' learning activity online to ensure sufficient engagement.

The left-hand column of Figure 5 shows that staff research practice in analytics has informed the teaching through various perspectives. For instance, the data-driven framework and methods highlighted in Section 2 demonstrate the methodology and process model for the analytics lifecycle. In addition, the research cited on health informatics represents a domain-based collaborative project with a university hospital. Emerging technologies were applied to electronic patient records through data warehousing, OLAP, visual analytics, data mining, and machine learning techniques.

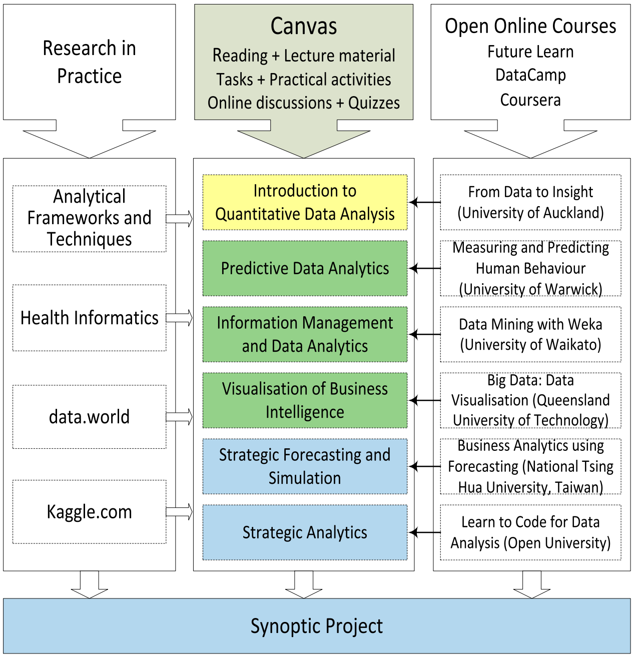
As illustrated in the right-hand column of Figure 5, some open online courses drawn especially from FutureLearn (2018) have been adapted into key learning resources. The individual courses across the categories of statistical analysis, big data analytics, and visualization have been mapped to corresponding specialist modules. A range of applications includes business, health, public transport, and government. The online courses from FutureLearn share common features and structure (e.g., online discussion that allows students to share their experience and learn from each other, quizzes that help students to test and consolidate their knowledge and understanding, and activities that provide hands-on practical opportunities using modern analytical software).

The student learning all comes together for the synoptic project through analyzing real-world data while devising and deploying solutions in their occupational domain: in particular, by applying digital technologies to perform analytics in relation to a significant data-driven business scenario.

Table 3. Modules, Themes, Tools, and Data Sets

		Software tools										
Specialist modules	Themes/topics	SQL	Excel	Alteryx	SPSS	R	iNZight	Weka	Tableau	Analytic Solver	Python	Case studies/data sets
Level 4												
Introduction to quantitative data analysis	Data exploration Descriptive analysis From data to insight		√		√		√					National Health and Nutrition Examination Survey
Level 5												
Information management and data analytics	RDBMS and SQL Data warehousing and information modeling Data mining and knowledge discovery: association rules	√	√					√	√			MyOrder—a medium-sized retail company
Predictive data analytics	Statistical computing and graphics Machine learning for data mining (e.g., classification)				√	√		√				Measuring and predicting human behavior; direct marketing in banking
Visualization of business intelligence	Data integration, extraction, and preparation Self-service and visual analytics		√	√					√			Global superstore orders and returns
Level 6												
Strategic forecasting and simulation	Exploring time series data Performance evaluation Process and methods for forecasting—smoothing; regression models		√			√	√			√		Air travel; department store sales; shipments of household appliances; company working hours
Strategic analytics	Start with strategy Measure metrics Coding for data analysis Reporting results Decision making and transforming business	√					√	√			√	World Health Organization; life expectancy project
Synoptic project	Analytical techniques: data mining, machine learning, and information modeling Identify and predict trends and patterns in data-driven business scenario	√	√	√	√	√	√	√	√	√	√	Workplace related

Figure 5. Overview Structure of Learning Resources



Aside from the various practical activities already described, key components of assessment also include portfolios and written reports across the analytical modules. For example, the summative assessment for the first-year IQDA module contains two deliverables in the form of an ePortfolio: Part A—students are required to apply industry standard tools (Excel and SPSS) to a range of business data, interpreting the results and highlighting any limitations of the quantitative analysis; Part B—a case study is provided to and/or agreed on with students, who are required to (1) identify the types of data available to the business and the ways in which data may be collected for decision making and new insight and (2) produce a professional report on the quantitative analysis of the business data.

Moving onto second-year assessment, the three specialist modules are thematic in their approach. For IM&DA, students are given a scenario where the IT analyst for a company is asked to contribute to a project to evaluate enterprise information systems technologies, methods, and tools that may benefit the business. The primary areas of interest are database management, data warehousing, information modeling, data mining, and knowledge discovery. Fundamental to informing the resulting evaluation report is that students consider the application of analytical tools and techniques in the context of the business and the module themes.

For the PDA module, the summative assessment contains two deliverables: Part A—students are required to select and apply appropriate predictive data analysis techniques to meet given business information requirements for a specified case study; Part B—students are advised to write a predictive analytics report, which (1) applies appropriate data analysis techniques to support decision-making processes in relation to a real-world issue/problem, (2) evaluates the results of predictive data analysis, and (3) justifies conclusions while suggesting recommendations drawn from the analysis which indicate a way forward.

Additionally, for the visualization of business intelligence module, students are required to use multisource data and produce an ePortfolio that consists of (1) data extraction and preparation tasks—identification, transformation, and integration of data required to answer business intelligence questions; (2) visualization—creation of a range of appropriate visualizations that will address the decision-making needs of stakeholders; (3) a business intelligence report, which documents and presents the visualizations in a coherent manner, explaining how the visual analytics can provide intelligence to the business.

In the final year, the two specialist-taught modules are both strategic in their outlook. For SF&S, students will be required to complete exercises using forecasting and simulation techniques based on real-world data while making critical recommendations. Summative assessment is designed to contain several deliverables in the form of (1) online discussion demonstrating learning engagement with the process of business forecasting; (2) quizzes on forecasting process and goals, data set exploration, and naive forecasts; (3) problems and solutions on the visualization of time series data, smoothing methods, and regression models; and (4) a case study with strategic assessment of relevant sources of data, preprocessing and forecasting approach, and evidence for the analysis and evaluation of results.

For the SA module, the summative assessment is designed around a strategic analysis report, where students will be required to investigate the analytics lifecycle for a chosen case study, demonstrating the ability to (1) integrate and document appropriate methods and techniques associated with the data analysis; (2) critically evaluate the practical strategic application of the storage, extraction, processing, and analysis of big data held in enterprise and central repositories; (3) apply methods for data mining and analysis to determine relationships in the data that provide insight; and (4) report and comment critically on findings and recommendations made to the case study stakeholder as a result of the analysis.

5.3. Outcomes—Reflection and Feedback

There have been two full years of delivery of the data analytics pathway to date (through the end of 2018). Tutors complete module evaluation and reflection forms each semester based on their own experiences as well as ongoing feedback from students. Below is a summarized reflection on the delivery thus far.

- Structured and organized weekly materials on Canvas linking to a range of resources.
- Hands-on opportunities provided for students to practice analytical tools—subsequently used in their summative assessment to demonstrate the capability of more advanced data analytics.
- Students were encouraged to adapt/create a data-driven framework as a guide for their analytical process; in addition, some students selected relevant business topics/issues from their own workplace to underpin their assessments.
- Recommendations on data sources for analytical projects have been welcomed by students, because they can select data sets according to their own interest, knowledge, and experience.

In respect to formal feedback, the school conducts independent module evaluation questionnaires with all students at the end of each semester. There are 12 questions using a rating scale as well as 2 open questions (i.e., "What have you enjoyed about this module?" and "How could this module be improved?"). The following is a representative summary from Level 4 and Level 5 cohorts during the 2016–2018 academic years.

5.3.1. Teaching and Learning.

- Enjoyed the variety of subjects/topics explored in the module—useful and relevant to the degree.
- Interactive seminar style helped learning process, including demonstrating the analytical software.
- A smaller class size would have made it easier for effective learning/study.
- More help needed with challenging parts for non-technical stream people.
- Group tasks and practical activities useful.
- Good amount of resources on Canvas.
- Learning new analytical skills.

5.3.2. Tutor Feedback.

- Individual feedback in class: good approach, enough support, and plenty of opportunities.
- Regular and good feedback throughout the module—really beneficial in helping to improve work.

5.3.3. Assessment.

- Students used the following phrases to describe the support for their assessment: focused, regular attention, and commitment to answer questions.
- Students appreciated the recommended structure provided for the summative assessment.

Information management and data analytics is the core second-year module and also the largest group in terms of student numbers (25–35). There was a clear relationship between the formative assessment submissions and summative outcomes. Some students had taken this module from the business management pathway and required additional support with certain technical aspects.

6. Conclusion and Future Work

A data-driven analytical framework has been proposed in the context of big data (see Figure 1). It provides the methodology and architecture for data analytics curriculum development and underpins the delivery of specialist modules. Digital technologies have also been explored and recommended through a range of software tools and data sources appropriate to the program.

The 2018–2019 cohort will be the first cohort at the UoW taking the Level 6 data analytics specialist modules—strategic forecasting & simulation and strategic analytics—as well as the synoptic project. The preferred teaching and learning practice already in place is being carried over to this level. In addition, it is planned to involve this group of final-year students in undergraduate research activities. The author's experience from research and knowledge exchange has helped to inform teaching through supervision of external projects and dissertations. The previous undergraduate project outcomes in health informatics had been selected among the excellent examples of students engaging with applied research projects. This practice will be transferred to a business setting going forward, potentially linking with the degree apprentice's own workplace.

The data analytics pathway development has had a positive impact on a diverse range of existing courses within Winchester Business School. Moreover, data science is an emerging field that requires multidisciplinary principles to guide the extraction of knowledge from data. Consequently, a new full-time BSc Data Science Programme is being developed in recognition of the growing importance of knowledge as a commodity to organizations. The curriculum will adopt a distinctive structure and pedagogy, building on the degree apprenticeship program as well as the brand-new computer science suite. This is articulated particularly through specialist modules, where technology and business-oriented activities are designed to focus on the data science context and concepts.

The new Data Science Programme is following the expectations of relevant subject benchmark statements in the United Kingdom. It will integrate key themes in information management, data mining, machine learning, and business intelligence. Understanding the employment market while defining specific skills sets associated with potential graduates are always important for courses in higher education. The Skills Framework for the Information Age (SFIA 2018) is being adopted to facilitate the mapping of corresponding employability skills related to data science. These could also be linked to an appropriate process model as well as the specialist modules across academic levels.

In terms of the next stage for development at Winchester, the EDISON Data Science Framework (EDISON 2018) will be highly influential. Their Data Science Model Curriculum in particular will be evaluated for its relevance and suitability within a business school environment. It will also be illuminating to investigate and discover the relative contrast between these comprehensive European efforts aimed at *building the data science profession* and the degree apprenticeship approach described in this paper.

References

EDISON (2018) Building the data science profession. Accessed November 26, 2018, http://edison-project.eu.

Elgendy N, Elragal A (2016) Big data analytics in support of the decision making process. *Procedia Comput. Sci.* 100:1071–1084.

EMC Education Services (2015) *Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data* (Wiley, Hoboken, NJ).

Evans JR (2013) *Statistics, Data Analysis, and Decision Modeling*, 5th ed. (Pearson Central Publishing, Harlow, UK).

Fayyad UM, Smyth P, Uthurusamy R, eds. (1996) *Advances in Knowledge Discovery and Data Mining*, vol. 21 (AAAI Press, Menlo Park, CA).

FutureLearn (2018) Digital education platform. Accessed November 26, 2018, https://www.futurelearn.com/.

Günther WA, Mehrizi MHR, Huysman M, Feldberg F (2017) Debating big data: A literature review on realizing value from big data. *J. Strategic Inform. Systems* 26(3):191–209.

Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH (2009) The WEKA data mining software: An update. *SIGKDD Explorations* 11(1):10–18.

IBM (2018) *IBM SPSS Statistics for Windows, Version 25.0* (IBM Corporation, Armonk, NY).

Lu J (2018) A data-driven framework for business analytics in the context of big data. Benczúr A, Thalheim B, Horváth T, Chiusano S, Cerquitelli T, Sidló C, Revesz PZ, eds. *New Trends in Databases and Information Systems. ADBIS 2018*, Communications in Computer and Information Science, vol. 909 (Springer, Cham, Switzerland), 339–351.

Lu J, Keech M (2015) Emerging technologies for health data analytics research: A conceptual architecture. Spies M, Wagner RR, Tjoa AM, eds. *Proc. DEXA 2015, 2nd Internat. Workshop NoSQL Databases* (IEEE, Piscataway, NJ), 225–229.

Lu J, Hales A, Rew D (2017) Modelling of cancer patient records: A structured approach to data mining and visual analytics. Bursa M, Holzinger A, Renda ME, Khuri S, eds. *ITBAM 2017. LNCS*, vol. 10443 (Springer International Publishing AG, Cham, Switzerland), 30–51.

Marr B (2015) *Big Data: Using Smart Big Data, Analytics and Metrics to Make Better Decisions and Improve Performance* (Wiley, Chichester, UK).

Marr B (2016a) *Big Data in Practice: How 45 Successful Companies Used Big Data Analytics to Deliver Extraordinary Results* (John Wiley & Sons, Oxford, UK).

Marr B (2016b) *Key Business Analytics: The 60+ Business Analysis Tools Every Manager Needs to Know* (Pearson, Harlow, UK).

Marr B (2017) *Data Strategy: How to Profit from a World of Big Data, Analytics and the Internet of Things* (Kogan Page, London).

Provost F, Fawcett T (2013) *Data Science for Business: What You Need to Know About Data Mining and Data-Analytic Thinking* (O'Reilly Media, Sebastopol, CA).

SFIA (2018) The skills framework for the information age. Accessed November 26, 2018, https://www.sfia-online.org/en.

Shearer C (2000) The CRISP-DM model: The new blueprint for data mining. *J. Data Warehousing* 5(4):13–22.

Shmueli G, Bruce PC, Patel NR (2016) *Data Mining for Business Analytics: Concepts, Techniques, and Applications with XLMiner*, 3rd ed. (Wiley-Blackwell, Hoboken, NJ).

Sivarajah U, Kamal MM, Irani Z, Weerakkody V (2017) Critical analysis of big data challenges and analytical methods. *J. Bus. Res.* 70:263–286.

Wang YC, Kung L, Byrd T (2018) Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations. *Tech. Forecasting Soc. Change* 126:3–13.

Wilder CR, Ozgur CO (2015) Business analytics curriculum for undergraduate majors. *INFORMS Trans. Ed.* 15(2):180–187.