

## Unit 5: Big Data & Visualisation

**Unit code** J/618/6251

**Unit level** 4

**Credit value** 15

---

### Introduction

Exploring and analysing big data translates information into insight. The purposeful, systematic exploitation of big data, coupled with analytics, reveals opportunities for improved decision making and better business outcomes. All this data is useful when processed but requires visualisation to bring to life. Data visualisation makes big data easier for the human brain to understand and detect patterns, trends and meaning in complicated data sets. With such rapid advancement in this area, there have been considerable challenges for data specialists to develop the skills, experience and growth required to maintain innovation in the sector. Similarly, the public and private sectors have struggled to keep up with progress, meaning that the introduction of legislation and community norms have been retrospective and, at times, reactive. As data continues to be the fuel for the digital economy, this area remains a constant topic of conversation for organisations and governments, and the public who share an interest in its growing commercial use, manipulation and presentation.

This unit introduces students to the concepts of big data and visualisation and how this is used for decision making. Students will explore the industry software solutions available for investigating and presenting data, before assessing the role and responsibility of data specialists in the current environment. Students will examine topics including data-driven decision making, manipulating data and automation, and building ethics into a data-driven culture. Students will demonstrate their use of tools and software to manipulate and prepare a visual presentation for a given data set. They will also assess how data specialists are responsible for adhering to legislation and ensuring data compliance.

On successful completion of this unit, students will be able to investigate the value of data for decision making to both end-users and organisations; compare how different industry leading tools and software solutions are used to analyse and visualise data; carry out queries to summarise and group a given data set, and analyse the challenges faced when building ethics into a data-driven culture. Students will have the opportunity to progress to a range of roles in the digital sector and will develop industry-led skills, analysis, and interpretation, which are crucial for developing practical experiences with big data and for gaining employment.

## Learning Outcomes

By the end of this unit, students will be able to:

- LO1 Examine big data and visualisation for decision making
- LO2 Investigate statistical and graphical techniques, tools and industry software solutions for big data and visualisation
- LO3 Demonstrate the use of industry software to manipulate data and prepare visual presentations for a given data set
- LO4 Assess the role, responsibilities and challenges for data specialists.

## Essential Content

### LO1 Examine big data and visualisation for decision making

#### *Big data:*

Explore common fundamental concepts, e.g. Doug Laney's Three Vs of Big Data (volume, velocity and variety), an extension of Vs (variability, veracity, visualisation and value).

Big data lifecycle to include: purpose, capturing data, searching and filtering, retrieving data for processing, validation and cleansing, visualisation, analysis and querying, utilisation and storage, obsolete and deleted data.

Organisation-led big data strategies such as data lake vs data warehouse.

#### *Visualisation:*

Understanding the target audience needs, e.g. context, reporting, dissemination, accessibility, breadth of data, depth of analysis.

Phases of data visualisation design process to include: formulating the brief, working with data, establishing editorial thinking and developing design solution.

Applying principles of good design to data visualisation, e.g. Dieter Rams' Ten Principles for Good Design, Gestalt principles of visual perception, use of storytelling for presenting data.

Effective visual elements, e.g. charts, graphs, plots, tables, points, lines, bars, area, maps, narratives, and aesthetics, e.g. position, size, shape, colour and transparency.

#### *Data for decision making:*

Process of data-driven decision making (DDDM) and steps to include: define an objective, establish a hypothesis, identify data need, build data process, sampling methods, collect data, analyse data, interpret results and make a decision.

Advantages of data-driven decision-making, e.g. continuous improvement and planning, collaborative decisions, reduce costs, real-time insights and new opportunities, digital literacy and data-driven cultures and challenges, e.g. inconsistent and unstandardised data, aligning decision making with business strategy, bias and discrimination, descriptive vs predictive trends and probabilities.

## **LO2 Investigate statistical and graphical techniques, tools and industry software solutions for big data and visualisation**

*Statistical and graphical techniques for big data analysis and visualisation:*

Popular big data techniques and analysis to include: business intelligence, cluster analysis, data mining, A/B testing, textual analysis, sentiment analysis, predictive, optimisation and prescriptive modelling.

Organising semi-structured and unstructured data variety, e.g. word-cloud visuals to display frequency, data catalogue, taxonomies and ontologies.

Forecasting estimates of future values, e.g. applied forecasting and decision tree algorithms.

*Industry leading tools and software solutions to analyse data:*

e.g. SQL, MySQL, Node XL, Oracle Analytics, Qlik Analytics Platform, Google Fusion Tables, Open Refine, Python, R Programming, Apache Spark/Hadoop, SAS Sentiment Analysis, Microsoft Azure, AWS, MATLAB.

*Industry-leading tools and software solutions to visualise data:*

e.g. Microsoft Excel, Power BI and Azure, AWS, Oracle Visual Analyzer, Qlikview, Google Chart, Canvas, Tableau, SAS Visual Analytics.

## **LO3 Demonstrate the use of industry software to manipulate data and prepare visual presentations for a given data set**

*Manipulating data:*

General activities when using industry software to manipulate data to include: importing datasets, data cleansing, data frame manipulation, testing and training a model, summarising analysis process and steps taken.

Query basics, e.g. prepare reports using commands including unions and subqueries, calculate aggregate statistics during a query, create queries against multiple data sources by using join commands, transform the output format of queries using built-in functions, perform queries across a group of rows using windowing functions, summarising and grouping data.

Explore advanced data manipulation and automation concepts, e.g. generalised linear models and regression, multilevel modelling and techniques, data pipelines, machine learning, 4-bit deep learning and deep reinforcement learning (DRL).

#### *Prepare visual presentations:*

General activities when using industry software to prepare visual presentations to include: selecting a visual element and aesthetic design, use insight analysis to understand data in context, find and filter content in dashboards, view and export data from dashboards to create a report, presentation and/or infographic.

Big data visualisation, e.g. onemilliontweetmap.com, the internet in real time (betfy.co.uk/internet-realtime), earth wind map (earth.nullschool.net), and data breaches (informationisbeautiful.net).

#### *Data set requirements:*

The data and its topic/context; summary of data collection, sampling procedures and data type; stakeholder requirements, interests and needs.

### **LO4 Assess the role, responsibilities and challenges for data specialists**

#### *Roles and responsibilities:*

Roles in a data-driven industry, e.g. data analyst, data scientist, data engineer, visualisation specialist, data administrator, business analyst, middle managers and senior management teams.

Responsibilities of a data specialist to include: preparing, analysing, modelling, managing and visualising data, storage and access rights. A chief data specialist will be expected to establish and enforce data policies and standards.

#### *Strategies to ensure data compliance:*

Organisational data architecture, policies, standards and rules, e.g. how data is stored, managed, used and disseminated.

Data protection, informed consent and privacy issues for compliance to include: personally identifiable information, protected health information, General Data Protection Regulation (GDPR) rights and obligations, enforcement and regulatory legal penalties.

Choice of industry-leading compliance management software and tools, e.g. Microsoft Compliance Manager, Amazon Web Services (AWS) Compliance, IBM DataOps.

*Challenges for data specialists:*

Data governance framework to ensure the value of outcomes, accountability, trust, collaboration, transparency, risks and security, and role of the data steward.

Guard against poor practice, e.g. cherry picking, disclosure of assumptions, conflict of interest, bias from a single view and/or choice of technique.

Building ethics into a data-driven culture and joining community of good practice, e.g. Data for Good Exchange (D4GX); Fairness, Accountability and Transparency in Machine Learning group (FAT/ML), Data Ethics Framework (gov.uk).

## Learning Outcomes and Assessment Criteria

Pass		Merit	Distinction
<b>LO1</b> Examine big data and visualisation for decision making			<b>D1</b> Evaluate the potential impact of data on both users and organisations when using data for decision making.
<b>P1</b> Explain the fundamental concepts of big data. <b>P2</b> Investigate the value of data for decision making to both end users and organisations.	<b>M1</b> Analyse the advantages and challenges of data-driven decision making to an organisation.		
<b>LO2</b> Investigate statistical and graphical techniques, tools and industry software solutions for big data and visualisation			<b>LO2 and LO3</b> <b>D2</b> Evaluate own data preparation and manipulation, justifying your choice of statistical techniques, to show how this meets the needs of stakeholders for a given data set.
<b>P3</b> Describe statistical and graphical techniques for big data and visualisation used in industry. <b>P4</b> Review different industry-leading tools and software solutions available for analysing and visualising data.	<b>M2</b> Compare how different industry-leading tools and software solutions are used to analyse and visualise data, with examples.		
<b>LO3</b> Demonstrate the use of industry software to manipulate data and prepare visual presentations for a given data set			
<b>P5</b> Select an industry-leading tool and software solution to manipulate data for a given data set. <b>P6</b> Demonstrate the use of queries to summarise and group data for a given data set.	<b>M3</b> Prepare a visual presentation to summarise data for a given data set.		
<b>LO4</b> Assess the role, responsibilities and challenges for data specialists			<b>D3</b> Analyse the role, responsibilities and challenges faced by data specialists when building ethics into a data-driven culture.
<b>P7</b> Explain the different roles, responsibilities and challenges faced by data specialists.	<b>M4</b> Review the different strategies used by data specialists to ensure data compliance.		

## Recommended Resources

### Textbooks

Dietel, P. (2020) *Intro to Python for Computer Science and Data Science: Learning to Program with AI, Big Data and The Cloud*. London: Pearson.

Franks, B. (2020) *97 Things About Ethics Everyone in Data Science Should Know*. USA: O'Reilly Media.

Graesser, L. and Keng, W. L. (2020) *Foundations of Deep Reinforcement Learning: Theory and Practice in Python*. London: Addison-Wesley Professional.

Kirk, A. (2019) *Data Visualisation: A Handbook for Data Driven Design*. London: Sage Publications.

Knaflic, C. N. (2015) *Storytelling with Data: A Data Visualization Guide for Business Professionals*. USA: John Wiley & Sons.

Loukides, M., Mason, H. and Patil, D. J. (2018) *Ethics in Health Data Science*. USA: O'Reilly Media.

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

McCormick, K. and Salcedo, J. (2017) *SPSS Statistics for Data Analysis and Visualization*. USA: John Wiley & Sons.

Freeman, M., Ross, J. (2019) *Data Science Foundations Tools and Techniques: Core Skills for Quantitative Analysis with R and Git*. London: Addison-Wesley Professional.

Viescas, J. L. (2018) *SQL Queries for Mere Mortals: A Hands-On Guide to Data Manipulation in SQL*. 4th Edition. London: Addison-Wesley Professional.

Wilke, C. O. (2019) *Fundamentals of Data Visualization: A Primer on Making Informative and Compelling Figures*. USA: O'Reilly Media.

### Journals

*Big Data & Society*, SAGE Journals

*International Journal of Computer Applications (IJCA)*

*Journal of Big Data*, SpringerOpen

*Journal of Data Science, Statistics and Visualisation*, International Association for Statistical Computing (IASC)



## Websites

gov.uk	UK government – Data Ethics Framework
ukdataservice.ac.uk	UK Data Service
	(General reference)

## Links

This unit links to the following related units:

*Unit 12: Data Analytics*

*Unit 19: Business Intelligence*

*Unit 29: Applied Analytical Models*