Unit 10: Big Data and Business Analytics

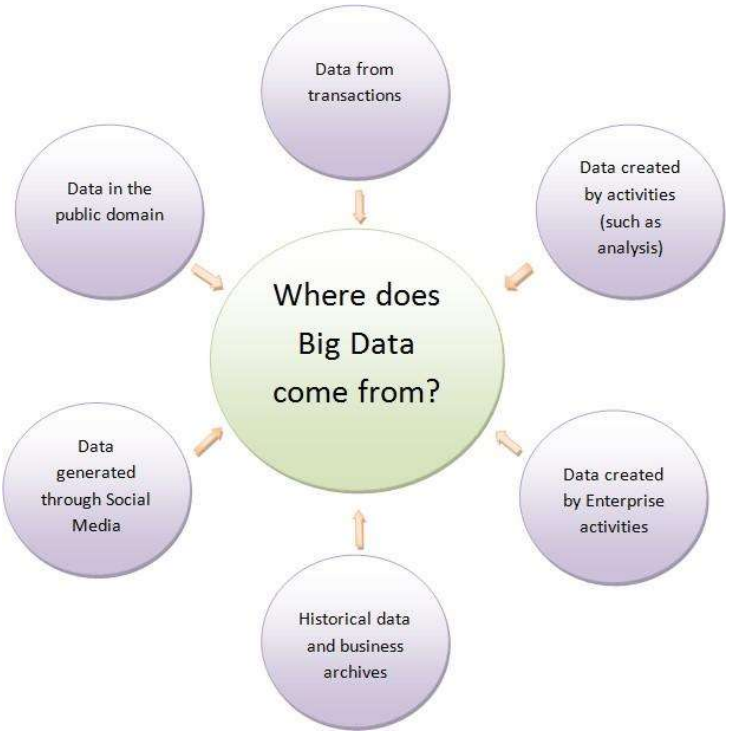
Delivery guidance

This internally assessed unit introduces learners to one of the most important concepts that have emerged in recent years. This is because data, big data and business analytics will only continue to present challenges (with more and more data added to systems every day) to organisations.

Although not required, it would be beneficial if learners had already studied Unit 2 (*Creating Systems to Manage Information*) before beginning this unit. This means they will already be familiar with terminology around data and data structures (as stored in databases etc.). They should also have a good grasp of basic maths as they will be required to undertake statistical analysis. While you could argue that software exists that will do the calculations, learners must still understand what standard deviation, a central tendency or an interquartile range are.

Not only is more and more data being stored, the amount of data that needs to be gathered and manipulated is growing. Databases (and data stores in general) are getting bigger and the ways in which they are being used are becoming increasingly diverse.

Data comes in all shapes and sizes; it can be unstructured, structured or semi- structured, and it is now common to have data in volumes of multiple terabytes and more. Data can be drawn from an increasing range of sources:



In delivering this unit the learners will explore the concept of data, including legislation related to its use, how data is stored, analysed (including how and why), and how data should be selected for analysis and presented to data users.

Learners will develop analytical skills, draw on other units for business contexts and ideas for analysis (e.g. to improve performance, in planning, to support decision making) and will learn a range of statistical techniques that support business activities.

As there are many different job roles that are reliant on good quality data, there is an opportunity here to involve a range of data users to talk to learners from an employer perspective (such as senior managers who set strategy, production and finance, and marketing managers who use data to make operational decisions).

Approaching the unit

In order for analytical activities to have some meaning, you will need to acquire and prepare different data sets and provide access to industry software. Links to websites that offer data for analysis have been provided in the resources section of this delivery guide.

Available data includes:

- Business and economic data
- Crime and justice
- Defence
- Education
- Environment
- Government
- Government spending
- Health
- Mapping
- Society
- Towns and cities
- Transport
- FBI Crime Data (USA)
- CDC Cause of Death data (USA)

Tutors should also explore local sources of data

The approach to this unit should be:

As practical as possible to give learners the best chance to apply a good range of analytical techniques.

Delivering the learning aims

Learning aim A

This aim sets the context for the unit by exploring from where data is gathered, how it is used and why it is important to different types of organisations.

Learners begin by exploring different business decision-making processes and how data

contributes to this activity both strategically and operationally. Examples could include data used to explore customer behaviour (sales and marketing), data used to manage the structure of the business (finance and operations), data used to cost products or services (labour and materials) and data used for benchmarking to help organisations measure their success in relation to their competitors and their own previous performance.

The challenges of data such as the costs of gathering, storing and analysing it will be considered. Learners will be introduced to a range of skills and techniques that they will need to master if they wish to work as analysts in this sector, and it is imperative that they understand the security and compliance requirements as applied to data.

Types of data are explored and data is categorised as internal or external with examples so that learners appreciate the diverse range of data sources. Whether internal or external, how data needs to be gathered, cleansed, stored until needed and how this may happen (both in terms of the technology and the way the data are structured) is covered.

Understanding the role of data warehousing and the role of the more limited data mart will help learners to build a knowledge of likely data sources or places where data can be stored in the real world.

The aim completes with learners considering big data analysis including OLAP (Online Analytical Processing) technology and the importance of checking any data used for validity, accuracy, sufficiency etc. They will investigate the four stages (or levels) of business analytics (descriptive, diagnostic, predictive and prescriptive analytics) and should be able to define these terms and the specific techniques that are characteristics of each stage.

Learning aim B

This aim focuses heavily on statistical techniques and the types of software that are available to analyse data sets. The software is not prescribed, which will allow you to choose a suitable or available product. A link has been provided in the resources that will enable you to choose free software from 50 choices that can be used at your centre and also by learners at home. Learners will have the challenge of using and navigating the software and some packages are much more difficult to master than others.

Tutorialspoint (see the link in the resources section) provides a series of tutorials for some of the more common analytical tools. In addition to proprietary software, learners should explore the analysis capabilities of Microsoft Excel (or a similar spreadsheet software).

You will need to introduce learners to a range of statistics from the simpler measure of central tendency (mean, median and mode) to the more complex operations including standard deviation, range and dispersion.

Probability is a key decision-making tool and is a learning aim in its own right. You should ensure that learners can carry out routine and non-routine operations accurately. To ensure learners are able to do these calculations they should not only provide the answers but should also be able to confidently show their working out and explain the processes.

They will explore the relationship between both independent and dependent variables using different techniques and should be able to work things out both with and without the use of software (so, e.g. using a calculator).

Learners will need to use at least one proprietary industry relevant software package in addition to Microsoft Excel to ensure they acquire a range of skills.

Learning aim C

This final learning aim will begin to teach learners the skills they will also need in the project. At this level, learners should be able to not only select the right tools and techniques for particular situations, but also to justify their choices and actions, being able to answer questions such as why did they choose one particular data set or another. At Level 4, learners will build on the learning from this unit and study how to integrate different data sets to provide a new set. Although there is no requirement to learn about this in this unit, learners may well question what they would do if a data set was incomplete. Data integration would provide one possible solution to the problem.

Wherever the data set comes from, learners should be able to show that they can select data and clean it (removing extraneous, inconsistent, inaccurate and incomplete data), format it consistently and, using appropriate software, produce outcomes from analysis that are relevant, valid and accurate.

Assessment model

Learning aim	Key content areas	Recommended assessment approach
A Investigate the role of big data and business analytics to improve performance, for benchmarking and/or to trigger innovation in organisations	**A1** Business information **A2** Types and storage of data **A3** Analysing big data **A4** Types of business analytics	A presentation focusing on how business analytics can be used by organisations to inform decisions that improve performance, for benchmarking and how this may result in innovation
B Explore the statistical software tools and techniques used to analyse data in organisations	**B1** Statistical techniques **B2** Probability distributions **B3** Mathematical modelling of data to find a goodness of fit	An informal report containing the results of learners' calculations and analysis of measured and supplied data, using appropriate software
C Carry out analysis of statistical data to meet the needs of an organisation	**C1** Selecting data for analysis **C2** Evaluating a data set and presenting the outcomes	A formal report using software tools to analyse a data set for a given organisation and to present the outcome (visualisation)

Assessment guidance

The assessment for this internally assessed unit would benefit from being divided into three assignments as shown above.

Assignment 1 (Learning aim A)

A presentation that focuses on how analytics is used to inform business decisions. The presentation should explain the benefits and challenges that analytics brings to any business. You should provide a realistic scenario for learners to focus on their presentation, particularly as it will give them an opportunity to research and include some specific analytics that would be useful to such a client.

The scenario could be based on a business that has been asked to provide analytical services to a client and the learner has been asked to persuade a less than enthusiastic middle management team about how the information gained from the process could help them in their roles and in their everyday activity.

Learners should ensure that the presentation is suitable for the audience as defined in the scenario they are given.

The tutor could prepare two or three different data sets/scenarios that would make watching the presentations more interesting and give learners an exposure to a wider range of situations.

You may choose to link this assessment to assignment 3, in which case the scenario should

be drawn from the data set. The set will need to have at least 500 records to comply with the requirements for assignment 3.

Assignment 2 (Learning aim B)

This assignment suggests an informal report that requires learners to show their ability in carrying out a range of statistical calculations and applying analytical techniques. This includes carrying out t-tests on two sets of data. Executing operations and producing results that should be communicated appropriately, the learners should be able to evidence that they can logically structure an analysis, the results and demonstrate the use of the correct methods.

The scenario for this assignment will be drawn from the data set.

As all learners will essentially be producing the same answers, you may include an optional verbal assessment to assess individual learner skills. This does not need to be extensive and cover all of the grading criteria/assignment content, but you might find this useful to support your decisions.

Assignment 3 (Learning aim C)

You have the option for this assignment to link the activity to the scenario (or scenarios) used in assignment 1.

The data should require cleaning and reformatting. You may therefore need to change formatting, make it inconsistent and add data that would need to be cleaned.

You should instruct learners to begin their formal report by restating the purpose of the analysis and the question that will be addressed.

For this assignment, learners will not need to produce their calculations, but should be able to explain (if asked) how they arrived at particular answers.

Getting started

This provides you with a starting place for one way of delivering the unit, based around the recommended assessment approach in the specification.

Unit 10: Big Data and Business Analytics

Introduction

Ask learners to share their own experiences of using data/information to make decisions.

Introduce the unit with the documentary listed in the resources section of this guide. It is a really interesting documentary that asks some fascinating questions that will make learners think about data in a number of ways.

Learning aim A – Investigate the role of big data and business analytics to improve performance, for benchmarking and/or to trigger innovation in organisations

A1: Business information

- Guest speaker presents to learners about how they use the results of data analysis in their everyday activity. This does not mean that you should invite a data analyst, and the speaker should be a data user. For example, a manufacturing manager who uses projected sales data to plan resource needs for the short and medium terms (raw materials, staff etc.), or a sales or a marketing manager who needs to use demographic information to target a sales or a marketing campaign.
- Present the reasons why organisations analyse data (using the wider range of examples listed in the specification).
- Support a discussion on the general challenges of analysing data:
  - Why can gathering data be an expensive process?
  - What are the issues with data storage?
  - What are the main skills needed by staff who analyse data?
  - How important are maths skills? Learners may well say that as software exists which will do all this, maths is not so important – but you should counter that they at least understand what a calculation means (in other words, the software may calculate the median – but what is a median? What is it for? What does it tell the analyst?).
  - What is the analyst's role in the security of the data that they use and interrogate?
  - What is the legal implication? In particular, you cite the Data Protection Act and the clause that says that data should only be stored for the purpose for which they are intended.
- Learners carry out group work (or paired work) to investigate legislative, ethical and security issues (in relation to both commercial and personal data). Split the class into three and then into pairs or small groups. Each group investigates one area. Learners create a slide presentation carousel to be used by a small business advisory service on its website to inform its members about the regulations (learners can add a narrative).

A2: Types and storage of data

Types of data:

- Explain the difference between qualitative and quantitative data, and internal and external data. You should provide a series of examples that learners should categorise, for example:
  - Wages data are quantitative internal data.
  - Feedback from customer through a website about levels of service experienced will be external data, and if the responses are to open questions rather than answers with numeric values it will be qualitative.
  - Data from sensors in a greenhouse that is climatically controlled are internal and quantitative.
  - HR data about employees' sickness and holiday leaves which contribute to a calculation about the loss of production days which is internal and quantitative.

Storing data:

- When it comes to storing data there are a number of considerations. Learners work in pairs to find out what each of the following terms means:
  - structured data versus unstructured data
  - volume, velocity and variety
  - methods of ensuring data security (such as passwords and encryption)
  - data warehouse versus data mart (including variants of data marts such as confirmed, stand-alone).

  Learners should create either a crossword puzzle or dominoes game using their research (see links in the resources section). The other options (such as wordsearch maker) are not relevant for this level. The completed game should be tried/used by other learners in the class.

Accessing data:

- To complete this topic, present to learners how data is shared across organisations, cite examples of software that is used by organisations to analyse the data and explain how data is managed to ensure only the right people have access to some or all of the data.

A3: Analysing big data

- Learners research and write a short report on the challenges of analysing big data (skills needed, software and technology available to analyse data, and the quality of the data themselves).

A4: Types of business analytics

Four stages (or levels) of business analytics:

Learners work in pairs to create teaching materials that would explain the four stages (descriptive, diagnostic, predictive and prescriptive) to a group of learners. This gives learners an opportunity to explore the ways that they would like to be taught and to create resources that they would find interesting. Choice of resources created is at the discretion of learners.

Learning aim B – Explore the statistical software tools and techniques used to analyse data in organisations

B1: Statistical techniques

- You should ensure that you have acquired or prepared a series of worksheets that will allow learners to develop and practice their statistical skills across the range of routine and non-routine operations listed in the specification. There are two useful URLs in the resources section in this delivery guide that provide some prepared materials.
- Learners will need to spend time using a range of statistical techniques, practicing the calculations and checking their answers against prepared answers.

B2: Probability distributions

- Prepare (or acquire) a series of worksheets that will allow learners to develop and practice their statistical skills in probability as listed in the specification. There are two useful URLs in the resources section in this delivery guide that provide some prepared materials.
- Learners use software to carry out some of the calculations around distribution as shown in the unit content.
- Learners will need to spend time practicing probability distributions and checking their answers against prepared answers.

B3: Mathematical modelling of data to find a goodness of fit

- Prepare (or acquire) a series of worksheets that will allow learners to develop and practice their statistical skills in modelling and regression as listed in the specification. There are two useful URLs in the resources section in this delivery guide that provide some prepared materials. Learners should use spreadsheet functionality to support this activity.
- Learners will need to spend time practicing mathematical modelling and checking their answers against prepared answers.

Learning aim C – Carry out analysis of statistical data to meet the needs of an organization

C1: Selecting data for analysis

- Presentation that focuses on how data for analysis should be selected. This is best achieved by walking through a practical example (maybe using one of the data sets provided by the links in the resources section of this guide).
- Learners should practice data selection against defined criteria. For each activity learners should be asked to explain their reasons for the approach taken.

C2: Evaluating a dataset and presenting the outcomes

- In anticipation of the assignment, learners should cleanse and format a data set to prepare it for analysis. You may find it useful to manipulate a downloaded data set introducing formatting errors or adding errors (such as changing headings in a series of columns so that one obviously repeats (e.g. 2012, 2013, 2013, 2015 – where the second instance of 2013 should be 2014 – this error could be transitioned into the legends in a graph or a chart).
- Challenge learners by demonstrating poor examples of data presentation (such as pie

charts with no legends or titles, graphs where the series have produced excessively small bars and lots of white space and spreadsheets with columns not suitably formatted to display all of the numbers in a cell). Ask learners what would need to be changed, added etc. to improve the presentation.

- Learners analyse the data set cleansed in the first activity against defined criteria and prepare the reports that will communicate the outcomes to different audiences.
- In order to demonstrate that learners are able to adapt the presentation of outcomes to suit different audiences, you could ask learners to present the outcomes for two or three different audience types rather than repeating the process a number of times simply so that an understanding of this can be shown. The data being presented should be appropriately supported by text, images and narrative as appropriate to the context.
- Learners create a short presentation that explains how the outcomes of the analysis were valid, accurate and relevant.

Details of links to other BTEC units and qualifications, and to other relevant units/qualifications

Depending on the choice of project, the following units will provide useful underpinning knowledge and skills that can be drawn on to create the deliverables:

- Unit 1: Information Technology Systems – Strategy, Management and Infrastructure
- Unit 2: Creating Systems to Manage Information
- Unit 11: Cyber Security and Incident Management
- Unit 14: Customising and Integrating Applications
- Unit 18: The Internet of Things
- Unit 19: Enterprise in IT

Resources

Journals

Search "*Springer Open Journal of Big Data*"– This is a collection of published papers on a variety of data-related subjects.

Search "*Elsevier Big Data Research*"– Big data research journal – some samples and articles can be viewed.

The KDNuggets website - offers a series of journals, magazines in analytics, big data, data mining and knowledge discovery.

Videos

YouTube:

Choosing which statistical test to use – statistics help (9.5 minutes)

Data Marts (1.5 minutes – no dialogue)

Data warehousing – An overview (9 minutes)

Big Data (53 minutes) – a range of videos introducing Big Data

Websites

Visit the Maths Answers website – A series of printable notes and worksheets on a range of statistics (Edexcel A-level syllabus), but useful in this context.

Go to the PAT Research website search "Top 48 free Statistical software"– This page provides details about, and access to, the top 50 free statistical software products.

Visit the Resourceaholic site – A large resource bank for supporting the teaching of statistics. This includes teaching notes, questions and answer sheets.

Go to the Springboard Blog and search "19 Free Public Data Sets for Your Data Science Project"– Additional data sets including FBI Crime Data, CDC Cause of Death data, store data for Walmart and emails released after the collapse of Enron.

Go to the Tools for Educators website - Educational games maker. This website is an online tool for creating a crossword puzzle. Learners carry out research and use the research to provide the answers, writing clues for each word included.

Go to the Tools for Educators website - – Domino maker – this is a free online game maker for paper/card based dominoes. Learners supply the content (term and definition) and complete a series of dominoes that can be printed and played as an educational game.

Search "*Tutorials point Tutorials Library*" and choose from the selection of big data tutorials – This link has a series of tutorials for big data analytical tools.

Pearson is not responsible for the content of any external internet sites. It is essential for tutors to preview each website before using it in class so as to ensure that the URL is still accurate, relevant and appropriate. We suggest that tutors bookmark useful websites and consider enabling learners to access them through the school/college intranet