CAMBRIDGE ADVANCED NATIONALS

OCR
Oxford Cambridge and RSA

**Sample question paper and mark scheme**  **DRAFT**

**LEVEL 3 CAMBRIDGE ADVANCED NATIONAL (AAQ) IN**

# IT: DATA ANALYTICS

**Extended Certificate H119**

For first teaching in 2025

## F201: Big data and machine learning

# Introduction

**This is Sample Assessment Material (SAM). It is an example exam paper that we publish alongside a new specification to help illustrate the intended style and structure of our question papers.**

During the lifetime of the qualification, updates to the question paper template may happen. We always recommend you look at the most recent set of past papers where available.

We also produce two further specific resources to support you with using this SAM:

- An assessment story. We explain the research we have undertaken during the development of the qualification and how consultation with teachers, students and schools have helped shape our assessment approach.

- Annotated SAMs. We take you through the key points of the assessment and highlight the different types of questions your students will experience in the exam.

## Tell us what you think

Your feedback plays an important role in how we develop, market, support and resource qualifications now and into the future. Here at OCR, we want teachers and students to enjoy and get the best out of our qualifications and resources, but to do that we need honest opinions to tell us whether we're on the right track or not. That's where you come in.

You can email your thoughts to **ProductDevelopment@OCR.org.uk** or visit the **OCR feedback page** to learn more about how you can help us improve our qualifications.

**Designed and tested with teachers and students**

**Helping young people develop an ethical view of the world**

**Equality, diversity, inclusion and belonging (EDIB) are part of everything we do**

# Summary of updates

| Date | Version | Page number | Summary of change |
|---|---|---|---|
| July 2023 | 1 DRAFT | All | Creation of document |

# Teacher support

## We have a range of support services to help you at every stage, from preparation to delivery.

Our teacher support is designed to make teaching our qualifications straightforward, whether you are an experienced teacher, new to teaching, new to OCR, or not a subject specialist of the qualification you are teaching.

We offer free access to services such as Access to Scripts, ExamBuilder and ActiveResults, and you'll find comprehensive teaching resources and a range of professional development courses on our teacher website, Teach Cambridge.

Our OCR subject advisors provide support and information to centres, including:

- Specification and non-exam assessment advice

- Updates on resource developments and training opportunities

- Information on our subject networks giving an opportunity to share ideas and expertise.

## Further help and support

Visit our subject pages to find out more about the assessment and request trial access to **Teach Cambridge**.

Not a registered or approved OCR centre? Discover the benefits of becoming one on the **OCR website**.

Have more questions about teaching OCR qualifications? Explore our **Online Support Centre** or contact our **Customer Support team**.

Contact details are available on the **final page** of the SAM.

## Work with us

As one of the major UK exam boards, OCR is responsible for creating and marking exams taken by over a million students each year. We work with over 10,000 professionals to create, check, mark and moderate our assessments.

If you would like to find out more about becoming an OCR examiner, please visit **our website** for more information.

# People and planet

**OCR is part of Cambridge University Press & Assessment, which has clear commitments to champion sustainability, diversity, trust and respect for our people and planet.**

We are committed to supporting a curriculum that helps young people develop an ethical view of the world. This enables them to take social responsibility, understand environmental issues and prepare them for the green jobs of the future.

Our equality, diversity, inclusion and belonging principles are that we:

- are respectful and considerate

- celebrate differences and promote positive attitudes to belonging

- include perspectives that reflect the diverse cultural and lifestyle backgrounds of our society

- challenge prejudicial views and unconscious biases

- promote a safe and supportive approach to learning

- are accessible and fair, creating positive experiences for all

- provide opportunities for everyone to perform at their best

- are contemporary, relevant and equip everyone to live and thrive in a global, diverse world

- create a shared sense of identity in a modern mixed society with one humanity.

**To learn more, including our work on accessibility in our assessment materials, visit our People and planet page.**

---

If you prefer to use a printed copy of the SAMs, consider printing a selection of pages. The following are the pages which you might find useful to print:

**Question paper** pages 5-20

**Mark scheme** pages 21-32

# Level 3 Cambridge Advanced National (AAQ) in IT: Data Analytics (Extended Certificate)

F201: Big data and machine learning

## Sample Assessment Material (SAM)

**Time allowed: 1 hour 30 minutes**

No extra materials are needed.

Please write clearly in black ink. **Do not write in the barcodes.**

Centre number ☐☐☐☐☐   Candidate number ☐☐☐☐

First name(s) _____

Last name _____

Date of birth ☐☐ ☐☐ ☐☐☐☐

## INSTRUCTIONS

- Use black ink.
- Write your answer to each question in the space provided. You can use extra paper if you need to, but you must clearly show your candidate number, the centre number and the question numbers.
- In the live exam there might be lined pages at the end of the question paper for you to use if you need extra space. Remember, you must clearly show the question numbers.
- Answer **all** the questions.

## INFORMATION

- The total mark for this paper is **60**.
- The marks for each question are shown in brackets **[ ]**.
- This document has **16** pages.

## ADVICE

- Read each question carefully before you start your answer.

OCR is an exempt Charity

**Turn over**

A large retailer has an online store and 250 physical stores across the UK. There are currently over 12 million members of its loyalty scheme. Customers use an online form on the retailer's website to register to join the loyalty scheme. Each customer can only join once. After registration, customers are emailed a membership number. Customers can use the membership number when shopping online or in store.

**1 (a)** State the **characteristic** of big data that relates to the completion of an online form.

..................................................................................................................................... **[1]**

**(b)** State **two** ways the volume of data can characterise data as big data.

1...............................................................................................................................

..................................................................................................................................

2...............................................................................................................................

..................................................................................................................................
**[2]**

Here are some of the customer loyalty scheme records.

| Membership number | Given name | Family name | Email address | Postcode |
|---|---|---|---|---|
| 124XC78 | Mia | Ling | ML.asl.com | L90 67BD |
| 987YT99 | Jane | Taylor | Jane@BBt.com | |
| 390BB41 | Kareem | Riley | 247@ihl.com | NP01 79V |
| 622MB83 | Jamal | Amit | Ja27@btm.com | LX99 |
| 124XC78 | Mia | Ling | ML.asl.com | L90 67BD |

Data cleaning and data wrangling will be carried out on the customer loyalty scheme records.

**(c)** Identify **three data cleaning** techniques that need to be carried out on the customer loyalty scheme records above.

1.................................................................................................................................

2.................................................................................................................................

3.................................................................................................................................

**[3]**

**Turn over**

**(d)** Complete the sentences to explain what **data wrangling** is.

Use words from the list.

You can use each word once, more than once or not at all.

| analysed | formatted | processed | raw | reliable | verified |
|----------|-----------|-----------|-----|----------|----------|

Data wrangling transforms and prepares …………………..... data from one or more sources

into a format that can be …………….…... or used by a computer system or application. It

can be time-consuming but essential for obtaining …………….……..…… results from

data analysis.

**[3]**

**2** When a customer registers for the loyalty scheme, they enter personal data. The retailer must comply with UK General Data Protection Regulations (UK GDPR).

**(a)** Describe the purpose of UK GDPR

...................................................................................................................................

.............................................................................................................................. **[1]**

**(b)** State **two** principles of UK GDPR

1...............................................................................................................................

...................................................................................................................................

2...............................................................................................................................

...................................................................................................................................

**[2]**

**(c)** Analyse the possible impacts on the retailer of **non-compliance** with UK GDPR.

.......................................................................................................................................................

.......................................................................................................................................................

.......................................................................................................................................................

.......................................................................................................................................................

.......................................................................................................................................................

.......................................................................................................................................................

.......................................................................................................................................................

.......................................................................................................................................................

.......................................................................................................................................................

.......................................................................................................................................................

.......................................................................................................................................................

.......................................................................................................................................................

**[6]**

**Turn over**

**3** Customers use their membership number when they shop online or in store. Transactional data is captured by the retailer and saved against the customer record.

**(a)** Why does the retailer capture the transactional data of each customer?

.......................................................................................................................................................

.................................................................................................................................................... **[1]**

**(b)** Explain **one** benefit to the **retailer** of using transactional data.

.......................................................................................................................................................

.......................................................................................................................................................

.......................................................................................................................................................

.................................................................................................................................................... **[2]**

**(c)** Transactional data is structured data.

State **two other** examples of **structured data.**

1 ....................................................................................................................................................

2 ....................................................................................................................................................

**[2]**

**4** The transactional data is stored in a private cloud storage area.

**(a)** State **two** characteristics of a **private cloud**.

1................................................................................................................................................

................................................................................................................................................

2................................................................................................................................................

................................................................................................................................................

**[2]**

**(b)** Explain **one** limitation to the **retailer** of using a private cloud to store the transactional data.

................................................................................................................................................

................................................................................................................................................

................................................................................................................................................

................................................................................................................................................ **[2]**

**(c)** State **one other** storage area the retailer could use to store the transactional data.

................................................................................................................................................ **[1]**

**5** When customers use their membership number to shop online, the website shows product recommendations.

Explain how **Artificial Intelligence (AI)** can be used to provide the recommendations.

.......................................................................................................................................................

.......................................................................................................................................................

.......................................................................................................................................................

................................................................................................................................................... **[2]**

**6** The retailer wants to add new products to their stock. Data analytics will be used to make informed decisions about the new products to be added.

**(a)** Describe the purpose of data analytics.

...................................................................................................................................................

.......................................................................................................................................... **[1]**

**(b)** Explain how data analytics software could help the retailer to make informed decisions about the new products.

...................................................................................................................................................

...................................................................................................................................................

...................................................................................................................................................

.......................................................................................................................................... **[2]**

**Turn over**

Customer records, including the transactional records, are used as part of the process of making decisions about new products.

The retailer wants to use the data analytic technique of **cluster analysis** to make decisions about the new products.

**(c)** Discuss whether the retailer should use **cluster analysis** to make decisions about new products.

In your answer you **must** write about:

- Any benefits to the retailer
- Any limitations to the retailer
- Whether you would recommend that the retailer use cluster analysis to make decisions about new products **and** your reasons.

**[9]**

........................................................................................................................................

........................................................................................................................................

........................................................................................................................................

........................................................................................................................................

........................................................................................................................................

........................................................................................................................................

........................................................................................................................................

........................................................................................................................................

........................................................................................................................................

........................................................................................................................................

........................................................................................................................................

........................................................................................................................................

.......................................................................................................................................

.......................................................................................................................................

.......................................................................................................................................

.......................................................................................................................................

.......................................................................................................................................

**(d)** State **two** data analytic techniques, **other than** cluster analysis.

1.......................................................................................................................................

2.......................................................................................................................................

**[2]**

The results of the data analytics will be presented at a meeting.

**(e)** Explain **one** benefit and **one** limitation to the **retailer** of using data visualisation software to present these findings.

Benefit.............................................................................................................................

.......................................................................................................................................

.......................................................................................................................................

Limitation.........................................................................................................................

.......................................................................................................................................

.......................................................................................................................................

.......................................................................................................................................

**[4]**

**7** The new products have been introduced to the retailer's stock. The retailer is sending money-off vouchers to customers who have previously bought similar products.

**Automated decision making** will be used to select which customers will receive the vouchers.

**(a)** Explain how automated decision making could **increase** the risk of bias.

.................................................................................................................................

.................................................................................................................................

.................................................................................................................................

........................................................................................................................ **[2]**

**(b)** A questionnaire will be sent to some customers asking for their views on the new products. Customers will be selected using the transactional records.

   **(i)** State **two inclusive data principles** the retailer should consider when selecting customers to receive the questionnaire.

   1 ..........................................................................................................................

   .............................................................................................................................

   2 ..........................................................................................................................

   ......................................................................................................................... **[2]**

**(ii)** Explain why **each** principle in **7(b)(i)** would need to be considered by the retailer.

1 ...............................................................................................................................

.................................................................................................................................

.................................................................................................................................

.................................................................................................................................

2 ...............................................................................................................................

.................................................................................................................................

.................................................................................................................................

.................................................................................................................................

**[4]**

**8** The retailer stores all their products in a warehouse. Staff pick, pack and despatch online orders from this warehouse.

The retailer wants the warehouse to be as energy efficient as possible.

Explain **one** benefit and **one** limitation to the **retailer** of using big data to increase the energy efficiency of the warehouse.

Benefit..................................................................................................................................

...............................................................................................................................................

...............................................................................................................................................

...............................................................................................................................................

Limitation.............................................................................................................................

...............................................................................................................................................

...............................................................................................................................................

...............................................................................................................................................

**[4]**

**END OF QUESTION PAPER**

**BLANK PAGE**

This is sample assessment material for our specification. It is to help show how the live assessment materials will look. During the lifetime of the qualification you might see small adjustments to the assessment materials. This is part of continuous improvement, designed to help you and your students. We recommend you look at the most recent set of past papers where available.

**OCR**
Oxford Cambridge and RSA

**XXXX/YYMM**

# Level 3 Cambridge Advanced National (AAQ) in IT: Data Analytics (Extended Certificate)

## F201: Big data and machine learning

**SAMPLE ASSESSMENT MATERIAL**

MARK SCHEME

This document has **12** pages.

# MARKING INSTRUCTIONS

## Crossed-out answers

If a student has crossed out an answer and written a clear alternative, do **not** mark the crossed-out answer.

If a student has crossed out an answer and **not** written a clear alternative, give the student the benefit of the doubt and mark the crossed-out answer if it's readable.

## Multiple choice question answers

When a multiple choice question has only one correct answer and a student has written two or more answers (even if one of these answers is correct), you should **not** award a mark.

## When a student writes more than one answer

### 1. Questions that ask for a set number (including 1) of short answers or points

If a question asks for a set number of short answers or points (e.g. **two** reasons for something), mark only the **first set number** of answers/points.

**First** mark the answers/points against any printed numbers on the answer lines, marking the **first** answer/point written against each printed number. **Then**, if students have not followed the printed numbers, mark the answers/points from left to right on each line and **then** line by line until the set number of answers/points have been marked. Do **not** mark the remaining answers/points.

### 2. Questions that ask for a single developed answer

If a student has written two or more answers to a question that only requires a single (developed) answer, and has **not** crossed out unintended answers, mark only the first answer.

### 3. Contradictory answers in points-based questions

When a student has written contradictory answers, do **not** award any marks, even if one of the answers is correct.

## Levels of Response marking

**1. To determine the level** start at the highest level and work down until you reach the level that best describes the answer

**2. To determine the mark within the level**, consider the following:

| Quality of the answer | Award mark |
|---|---|
| Consistently meets the criteria for this level | At the top of the level (6 and 9 mark questions) |
| Meets the criteria but with some inconsistency | At the middle of the level (9 mark questions) |
| On the borderline of this level and the one below | At the bottom of the level (6 and 9 mark questions) |

**ANNOTATIONS**

| Annotation | Meaning |
|---|---|
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

**MARK SCHEME**

| 1 (a) | |
|---|---|
| Max mark | 1 |
| Answer | • Variety of data types (1) |
| Guidance | Do **not** accept:<br>• Variety |

| 1 (b) | |
|---|---|
| Max mark | 2 |
| Answer | Any **two** from:<br>• The amount of data that exists (1)<br>• The initial size of data collected (1)<br>• The amount of data is growing (1)<br>**Credit any other appropriate response.** |
| Guidance | 1 mark for each correct answer.<br><br>The focus of the question is generic. The answer does **not** have to be applied to the context but accept answers which are applied to the context. |

| 1 (c) | |
|---|---|
| Max mark | 3 |
| Answer | • Removing duplicates (1)<br>• Fix missing data (1)<br>• Fix structural errors (1) |
| Guidance | 1 mark for each correct answer |

| 1 (d) | |
|---|---|
| Max mark | 3 |
| Answer | • Raw (1)<br>• Analysed (1)<br>• Reliable (1) |
| Guidance | Correct answers only |

| 2 (a) | |
|---|---|
| Max mark | 1 |
| Answer | Any **one** from:<br>• To protect an individual's personal data (1)<br>• To set out the key principles / rights / obligations for the processing of personal data. (1)<br>**Credit any other appropriate response.** |
| Guidance | |

| 2 (b) | |
|---|---|
| Max mark | 2 |
| Answer | Any **two** from:<br><br>• Processed lawfully, fairly and in a transparent way (in relation to individuals) / Lawfulness, fairness and transparency (1)<br><br>• Collected for specified, explicit and legitimate purposes and not further processed in a way that is incompatible with those purposes / Purpose limitation (1)<br><br>• Adequate, relevant and limited to what is necessary in relation to the purposes (for which they are processed) / Data minimisation (1) |

| | • Accurate and, where necessary, kept up to date / Accuracy (1) |
|---|---|
| | • Kept in a form which allows identification of data subjects for no longer than is necessary (for the purposes for which the personal data are processed) / Storage limitation (1) |
| | • Processed in a way that ensures appropriate security of the personal data (including protection against unauthorised or unlawful processing and against accidental loss, destruction or damage) using appropriate technical or organisational measures / Integrity and confidentiality (1) |
| **Guidance** | 1 mark for each correct answer.<br>Accept alternative wording |

| 2( c) | |
|---|---|
| **Max mark** | 6 |
| **Levels of Response** | **Level 3 (high) 5-6 marks**<br><br>A **thorough** analysis, which includes:<br>• identification of a **range** of characteristics or elements<br>• **detailed** knowledge and understanding in the context of the question<br>• **clear** explanation<br>• **consistent** use of appropriate subject terminology.<br><br>**Level 2 (mid) 3-4 marks**<br><br>An **adequate** analysis, which includes:<br>• identification of **some** characteristics or elements<br>• **sound** knowledge and understanding in the context of the question<br>• **adequate** explanation<br>• **some** use of appropriate subject terminology.<br><br>**Level 1 (low) 1-2 marks**<br><br>A **basic** analysis, which includes:<br>• identification of **at least one** characteristic or element<br>• **limited** knowledge and understanding in the context of the question<br>• **basic** explanation<br>• use of appropriate subject terminology is **limited**.<br><br>**0 marks**<br>Answer is **not** worthy of credit. |

| Indicative content | Answers can include some of the following: |
|---|---|
| | Impacts:<br>• Financial Penalties<br>     o Non-compliance with data protection laws can result in fines and penalties, which can be up to 4% of the retailer's annual revenue or €20 million (whichever is greater).<br>• Increased costs<br>     o Staff training will need to be increased to ensure all staff are aware of their responsibilities when handing data and increased levels of security will need to be implemented to ensure compliance and reduce the risks of a data breach. Given the size of the retailer and the number of branches it has, there will be a large number of staff to train which will be very costly.<br>• Reputational Damage<br>     o Data breaches and non-compliance can result in negative publicity and damage to reputation. This can lead to a loss of customer trust, decreased brand value, and reduced business opportunities. Given the size and scale of the retailer, this is likely to be more widely reported in the media and therefore potentially more significant.<br>• Legal Action<br>     o Non-compliance can result in legal action, such as civil lawsuits or criminal charges. This can result in significant legal costs and may require the retailer to pay compensation to affected individuals.<br>• Business Disruption<br>     o A data breach or other non-compliance issues can cause significant disruption to business operations, as well as damage to IT systems and infrastructure. This can result in additional costs to repair the damage and restore operations. With 250 stores and an online operation, this disruption has the potential to be severe.<br>• Regulatory Investigations<br>     o Non-compliance with data protection laws can result in regulatory investigations, which can be time-consuming and costly. Investigations can also result in further penalties or sanctions if non-compliance is found.<br><br>**Credit other relevant analysis, points and examples.** |

| 3 (a) | |
|---|---|
| **Max mark** | 1 |
| **Answer** | Any **one** from:<br>• To keep a record of items sold when / what / if online or in store (1)<br>• To record the purchases made by the customer (1)<br>**Credit any other appropriate response.** |
| **Guidance** | |

| 3 (b) | |
|---|---|
| **Max mark** | 2 |
| **Answer** | Up to **two** marks for **one** benefit. **One** mark for identifying benefit. **One** mark for application to the retailer, e.g.:<br>• Stock levels of products can be monitored (1) so the retailer can order more stock of popular products (1)<br>• Poor selling products can be identified (1) so these can be reduced in price / put on special offer (1)<br>• Trends can be identified (1) so stock can be increased to meet customer's needs (1)<br>• Analysis can be carried out (1) so the retailer can identify the most / less profitable products (1) |

| | Credit any other appropriate response. |
|---|---|
| **Guidance** | Up to **two** marks for **one** benefit.<br>Examples / expansion (application to the retailer) **must** relate to the retailer in terms of on line **or** in store. |

| 3 (c) | |
|---|---|
| **Max mark** | 2 |
| **Answer** | Any **two** from:<br>• Continuous data (1)<br>• Discrete data (1)<br>• Relational database (1)<br>• Spreadsheet file data (1) |
| **Guidance** | **One** mark for each correct answer. Correct answers only<br>Do **not** accept:<br>• Transactional data |

| 4 (a) | |
|---|---|
| **Max mark** | 2 |
| **Answer** | Any **two** from:<br>• The infrastructure is only accessible to one organisation (1)<br>• It is accessed through an intranet (1)<br>• Accessed through permission levels remotely through the internet (1)<br>**Credit any other appropriate response** |
| **Guidance** | **One** mark for each correct answer. |

| 4 (b) | |
|---|---|
| **Max mark** | 2 |
| **Answer** | Up to **two** marks for **one** limitation. **One** mark for identifying limitation. **One** mark for application to the retailer, e.g.:<br>• Higher staff costs (1) the retailer will need specialist staff to maintain the cloud area to maintain access to the data (1)<br>• Unreliability of access (1) if systems / internet access is not available the retailer will not be able to access the data (1)<br>• Maintenance costs / time (1) as no SLA is in place with an external company time to solve issues may increase (1)<br>**Credit any other appropriate response** |
| **Guidance** | Up to **two** marks.<br>Maximum **one** limitation.<br><br>The focus of the question is on the retailer. However, this question tests PO1 (Show knowledge and understanding) since limitations identified will apply to many contexts, not just this context. Accept both answers which are specifically applied to the context and those that are more generic. |

| 4 (c) | |
|---|---|
| **Max mark** | 1 |
| **Answer** | Any **one** from:<br><br>• Public Cloud (1)<br>• Data Lake (1)<br>• Data warehouse (1)<br>• Solid State Drives (1) |
| **Guidance** | Correct answer only |

| 5 | |
|---|---|
| **Max mark** | 2 |
| **Answer** | Up to two marks for an explanation of how AI is used to provide recommendations. **One** mark for identifying the feature. **One** mark for saying how it can be used, e.g.:<br>• Algorithms are used to consider customers' transaction history (1) using this to select products based on the previous products bought (1)<br>• AI can provide recommendations based on search criteria (1) input by the customer based on product history (1)<br>**Credit any other appropriate response** |
| **Guidance** | Up to **two** marks.<br>Maximum **one** explanation. |

| 6 (a) | |
|---|---|
| **Max mark** | 1 |
| **Answer** | **One** mark for a description of data analytics, e.g.:<br>• To help the retailer make sense of the collected data (1)<br>• To analyse the raw data to identify trends and patterns / help in decision making (1)<br>**Credit any other appropriate response** |
| **Guidance** | Maximum **one** description. |

| 6 (b) | |
|---|---|
| **Max mark** | 2 |
| **Answer** | Up to **two** marks for an explanation. **One** mark for identifying an action. **One** mark for how it can be used by the retailer, e.g.:<br>• The software will identify trends / patterns (1) so the retailer will know what products are popular with customers (1)<br>• The price of the products can be analysed against purchases (1) so the retailer can determine prices / discounts of products (1)<br>**Credit any other appropriate response** |
| **Guidance** | Up to **two** marks.<br>Maximum **one** explanation.<br>The expansion (how it can be used by the retailer) **must** relate to the retailer / product buying |

| 6 (c) | |
|---|---|
| **Max mark** | 9 |
| **Levels of Response** | **Level 3 (high) 7 - 9 marks**<br><br>A **thorough** discussion which shows **detailed** evaluation, which includes:<br><br>• a **range** of examples of the use of cluster analysis<br>• the benefits **and** limitations of cluster analysis for the retailer<br>• a **detailed** analysis in the context of the question<br>• a **clear** conclusion(s) with **detailed** reasons/justifications<br>• **consistent** use of appropriate subject terminology.<br><br>**Level 2 (mid) 4 - 6 marks**<br><br>An **adequate** discussion which shows **sound** evaluation, which includes:<br><br>• **some** examples of the use of cluster analysis<br>• **some** points relating to benefits or limitations of cluster analysis for the retailer<br>• **some** analysis in the context of the question<br>• an **adequate** conclusion(s) with **relevant** reasons/justifications |

|   |   |
|---|---|
|   | • **some** use of appropriate subject terminology.<br><br>**Level 1 (low) 1-3 marks**<br><br>A **basic** discussion which shows **limited** evaluation, which includes:<br><br>• a **few** examples of the use of cluster analysis<br>• a **few** points relating to benefits or limitations of cluster analysis for the retailer<br>• a **limited** analysis in the context of the question<br>• a **brief** conclusion(s) with **limited** reasons/justifications<br>• use of appropriate subject terminology is **limited**.<br><br>**0 marks**<br><br>Answer is **not** worthy of credit. |
| Indicative content | Answers can include some of the following:<br><br>**Purpose:**<br>• To sort different objects or data points into groups to identify the degree of association<br><br>**Benefits:**<br>• Clear identification of links between data, such as the association between groups.<br>• Different groups can be analysed, such as age / location/ in store / online. The buyer behaviour of these groups can be analysed, e.g.how often each group makes purchases. This will help the retailer make informed decisions.<br>• To Identify "bad" transactions which could indicate fraudulent activity to the retailer<br>• Grouping data could reduce the volume of data for the retailer, by filtering out irrelevant data.<br><br>**Limitations**:<br>• Incorrect selection of data can lead to incorrect results<br>• Some associations can be false if the groups are not selected correctly<br>• Incorrect analysis can lead to incorrect decision making<br><br>**Examples:**<br>• Links between products bought and age range.<br>• Links between products bought and geographical location.<br>• Links between range of products bought online / in store.<br><br>**Credit other relevant conclusions, points, and examples.** |

| 6 (d) | |
|---|---|
| **Max mark** | 2 |
| **Answer** | Any **two** from:<br>• Regression analysis (1)<br>• Monte Carlo simulation (1)<br>• Factor analysis (1)<br>• Cohort analysis (1)<br>• Time series analysis (1) |
| **Guidance** | 1 mark for each correct answer.<br>Do **not** accept:<br>• Cluster analysis |

| 6 (e) | |
|---|---|

| Max mark | 4 |
|---|---|
| **Answer** | Up to **two** marks for a benefit and up to **two** marks for a limitation of using data visualisation software: **One** mark for identifying the benefit / limitation. **One** mark for saying how it will affect the retailer, e.g.:<br><br>**Benefit:**<br>• Charts / graphs can be created (1) so the retailer can visualise the data rather than looking at raw data (1)<br>• The graphs / charts can be imported into other software (1) so the results can be shared across different departments (warehouse and stores) in the retailer (1)<br>• Data can be processed more accurately (1) meaning the retailer can rely on the results of the processing to make informed decisions (1)<br>• Data visualisation software enables real-time data monitoring (1) so results will be presented instantly, providing up-to-the-minute information the retailer can use (1)<br><br>**Limitation:**<br>• Using the incorrect data can lead to incorrect processing (1) meaning that incorrect decisions will be made by the retailer (1)<br>• The output of the software may be cluttered / not fit for the target audience (1) meaning the retailer may find it difficult to extrapolate the correct / relevant results (1)<br>• If the incorrect output type is selected (1) it may be misleading meaning incorrect decisions are made (1)<br>• The quality of the data selected and used is important (1) as the data visualisation software will only process and display the data input, i.e. if the input is low quality, the output will not meet the retailer's requirements (1)<br><br>**Credit any other appropriate response** |
| **Guidance** | Up to **two** marks each for **one** valid benefit and **one** valid limitation.<br>Maximum **one** benefit.<br>Maximum **one** limitation. |

| 7 (a) | |
|---|---|
| **Max mark** | 2 |
| **Answer** | Up to **two** marks for an explanation of how automated decision making could increase bias. **One** mark for identifying the cause. **One** mark for the implication, e.g.:<br>• Initial lack of training data being collected for certain groups of customers (1) so the algorithm has not learnt to pick up patterns (1)<br>• The human creating the algorithm has unconscious bias (1) meaning this bias will be translated into the final algorithm (1)<br>• If historical data is used to train the algorithm this data may include bias (1) which will be correlated into the patterns learnt by the algorithm (1)<br>**Credit any other appropriate response** |
| **Guidance** | Up to **two** marks.<br>Maximum **one** explanation.<br>The focus of the question is on the retailer. However, this question tests PO1 (Show knowledge and understanding) since implications identified will apply to many contexts, not just this context. Accept both answers which are specifically applied to the context and those that are more generic. |

| 7 (b) (i) | |
|---|---|
| **Max mark** | 2 |
| **Answer** | Any **two** of:<br>• All population groups must be included in the data (1)<br>• The data must be up to date (1)<br>• The data must be transparent (1) |

| | |
|---|---|
| | Credit any other appropriate response |
| Guidance | |

| 7 (b) (ii) | |
|---|---|
| Max mark | 4 |
| Answer | Up to **two** marks for each principle. **One** mark for identifying the reason why. **One** mark for application to the retailer, e.g:<br><br>**Principle: All population groups must be included in the data**<br>• To ensure that the questionnaire is sent to a range of customers who represent the retailer's customer base (1) if not the views expressed may demonstrate bias for / discrimination against a population group (1)<br>• To get a representative set of views about the products (1) to ensure full and accurate views can be acted on by the retailer (1)<br><br>**Principle: The data must be up to date**<br>• The transactional records used must be up to date with data cleaning having been carried out (1)<br>    ○ this will ensure that customers sent the questionnaire are still members of the loyalty scheme. (1)<br>    ○ this will ensure that only customers who have bought the product will be sent the questionnaire (1)<br><br>**Principle: The data must be transparent**<br>• The selection of the customers must be fair and equitable (1) with the process used to select customers being available to anyone who requests to see it (1)<br>• Pre-defined process / algorithms should be used (1) to ensure a true representation of the customer base is sent the questionnaires / to enable a range of views to be elicited (1)<br><br>**Credit any other appropriate response** |
| Guidance | Each explanation **must** refer to each principle detailed in **(i)**.<br>Where **zero** marks are awarded for **(i) zero** marks are awarded for **(ii)** |

| 8 | |
|---|---|
| Max mark | 4 |
| Answer | Up to **two** marks for a benefit and up to **two** marks for a limitation of using big data to increase energy efficiency. **One** mark for identifying the benefit / limitation. **One** mark **for saying** how it will affect the retailer, e.g.:<br><br>**Benefit:**<br>• The analysis of the big data can show trends / patterns in when / where people use the warehouse (1) meaning energy can be managed in those areas / at those times (1)<br>• The data can be used to introduce green energy as an alternative source (1) meaning energy costs for the warehouse will decrease (1)<br>• Smart technology can be used to control energy use (1) meaning AI can be used to switch off / on energy at the required / appropriate time(s) (1)<br><br>**Limitation:**<br>• Trends / patterns can be identified on 'normal' working days which may change, e.g. Covid working (1) which may mean the algorithm is wrong leading to a waste of energy / increased costs (1)<br>• If the energy devices e.g. boilers / air conditioning units malfunction (1) then the algorithm will send instructions, but these may not be carried out and it may be difficult to identify the problem (1) |

| | | |
|---|---|---|
| | • The warehouse may need extra insulation / equipment to be installed which adds to initial costs (1) to maximise the efficient use of the big data when controlling energy efficiency (1)<br><br>**Credit any other appropriate response** | |
| **Guidance** | Up to **two** marks each for **one** valid benefit and **one** valid limitation.<br>Answers **must** focus on the retailer who operates the warehouse. | |

# Examine *with* us

- Build confidence supporting your students with assessment

- Enhance subject knowledge

- Great for professional development

Join our team: **ocr.org.uk/assessor**

**OCR**
Oxford Cambridge and RSA

**Contact the team at:**

📞 **01223 553998**
✉ **computerscience@ocr.org.uk**
◉ **ocr.org.uk**
⊙ **facebook.com/ocrexams**
🐦 **@OCR_ict**
⊙ **instagram.com/ocrexaminations**
in **linkedin.com/company/ocr**
▶ **youtube.com/ocrexams**

To stay up to date with all the relevant news about our qualifications, register for email updates at **ocr.org.uk/updates**

Visit our Online Support Centre at **support.ocr.org.uk**