Assessor Guide

BSBXBD403

Analyse big data

Assessment 2 of 4

Short Answer Questions

Version 1

Assessment Instructions

Task overview

This assessment task contains 10 short answer questions. Read each question carefully before typing your response in the space provided.

Assessment Information

Submission

You are entitled to three (3) attempts to complete this assessment satisfactorily. Incomplete assessments will not be marked and will count as one of your three attempts.

All questions must be responded to correctly to be assessed as satisfactory for this assessment.

Answers must be typed into the space provided and submitted electronically via the LMS. Hand-written assessments will not be accepted unless previously arranged with your assessor.

Reasonable adjustment

Students may request a reasonable adjustment for assessment tasks.

Reasonable adjustment usually involves varying:

- the processes for conducting the assessment (e.g. allowing additional time)
- the evidence gathering techniques (e.g. oral rather than written questioning, use of a scribe, modifications to equipment)

However, the evidence collected must allow the student to demonstrate all requirements of the unit.

Refer to the Student Handbook or contact your Trainer for further information.

Please consider the environment before printing this assessment.

Question 1

Identify three (3) Australian Privacy Principles (APPs) that relate to analysing big data and outline the obligations of each.

(Word count: 25 - 45 words for each APP)

Assessor instructions: Students must list three APPs in the answer table. The descriptions provided under the column 'Obligations as it relates to analysing big data' are likely to include different wording than the sample answer provided. However, the acceptable responses must:

- be within the specified word limit
- reflect the characteristics described in the exemplar answer.

Refer to the 'Guide to Data Analytics and the Australian Privacy Principles' (Long URL: https://www.oaic.gov.au/privacy/guidance-and-advice?a=3086) for more information on APPs. This resource is provided to the student as part of the learning related to this unit.

A sample answer is provided below.

Table 1 - Answer table for Question 1

#	Australian Privacy Principle	Obligations as it relates to analysing big data *(25– 45 words)*
1	APP 3 – Collection of personal information	Organisations should be mindful to only: - collect information by lawful and fair means - collect sensitive information with the individual's consent (unless an exception applies).
2	APP5 – Notification	Organisations should be careful not to use personal information for a purpose other than the primary purpose it was collected for unless an exception applies.
3	APP 11 – Security of personal information	Organisations should actively consider whether they are permitted to retain personal information. When personal information is no longer needed it should be destroyed or de-identified.
	Other answers include: APP10 – Quality of personal information	Organisations should take responsible steps to ensure that personal information: - collected is accurate, up-to-date and complete - used or disclosed is having regard to the purpose of the use or disclosure is accurate, up-to-date, complete and relevant.
	APP6 – Using and disclosing personal information	Organisations should carefully consider whether the uses and disclosures of personal information for data analytics activities are compatible with the original purpose of collection (particularly when the information is collected directly from a third party).
	APP7 – Direct marketing	Organisations should have a good understanding of how they use data analytics for direct marketing, and if this includes facilitating other organisations' direct marketing, they need to comply with additional obligations including the Spam Act 2003 or the Do Not Call Register Act 2006.
	APP8 – Quality of personal information	Organisations should take rigorous steps to ensure the personal information collected via creation is accurate, complete and up-to-date by checking that third parties, from which personal information is collected, have implemented appropriate practices, procedures and systems to ensure the quality of personal information.

Question 2

Outline the relationship between 'raw data' and 'dataset', as it relates to big data analysis.

(Word count: 25 - 35 words)

Assessor instructions: Student responses are likely to include different wording than the sample answer provided. However, the acceptable responses must:

- be within the specified word limit
- reflect the characteristics described in the exemplar answer.

A sample answer is provided below.

- A dataset is derived from raw data and can be a subset or a representative sample of raw data.
- A dataset is created by performing transformations, cleaning and organising raw data.

Other answers include:

- Raw data after being transformed and cleaned (i.e. free of missing information, duplicates etc) is used to create a dataset, which is further processed and analysed to gain insights.

Question 3

Outline the role of cloud technology when performing big data analytics. Provide three [3] examples of cloud technology platforms commonly used for analysing big data in your answer.

(Word count: 55 – 75 words)

Assessor instructions: Student responses are likely to include different wording than the sample answer provided. However, the acceptable responses must:

- be within the specified word limit
- reflect the characteristics described in the exemplar answer
- list three examples of cloud technology platforms.

A sample answer is provided below.

Cloud technology allows large amounts of data to be stored using cloud storage platforms and have made datasets more accessible and easier to connect to. These datasets can then be processed using the built-in intelligence tools and powerful features of the cloud analytic platforms to:

- further analyse these datasets
- generate reports, and dashboards
- share results from the analysis with stakeholders in an organisation as required.

Examples:

- AWS
- Google cloud
- Microsoft Azure

<u>Other answers and examples include:</u>
Cloud technology provides great advantages for businesses in terms of the underlying infrastructure, technology and equipment and tools required for analysing big data. Such as:

- Cost-effectiveness (pay for only what you use)
- Availability of a variety of cloud services to perform analytical functions (e.g. Machine Learning etc)
- Accessible from anywhere, any device etc.
- Increased opportunities to monetise data
- Separation of compute from the storage means that you can leverage (cheap) storage and can rapidly scale resources for querying and other analytic workloads
- Ability to store large amounts of data
- Elasticity and scalability

Examples:

- Snowflake
- Cloudera
- IBM Cloud

Question 4

Outline the role of automation tools when performing big data analytics. Provide two (2) examples of commonly used automation tools used for analysing big data in your answer.

(Word count: 55 – 75 words)

Assessor instructions: Student responses are likely to include different wording than the sample answer provided. However, the acceptable responses must:

- be within the specified word limit
- reflect the characteristics described in the exemplar answer
- list two examples of automation tools.

A sample answer is provided below.

Automation tools allow businesses to easily perform analysis of their big data; especially those time-consuming analytical tasks such as:

- detecting anomalies in time-series data
- data categorisation
- data preparation and modelling
- providing real-time insights.

The use of automation tools can help to:

- fast-track the big data analysis process
- get faster results
- handle more data
- save costs
- increase productivity.

Examples of some automation tools include:

- SAS Visual Forecasting
- R scripts (to automate repeatable steps)

Other examples include:

- Python (to automate repeatable steps)
- Oracle analytic cloud – includes automation capabilities
- HEVO, Fivetran – for data pipeline automation
- dotData
- Alteryx
- BI Automation tools – Wherescape (automates common code written, built-in meta data repository, ability to track back and forward)
- Microsoft Purview (data lineage, data categorization, manage personal identifiable information)

Question 5

Discuss the following data classification categories of analytics by outlining:

- what each classification category represents with three (3) examples (Word count: 25 - 35 words for each category)

- the purpose of analysing each category of data for a business (Word count: 15 - 30 words for each category).

Assessor instructions: Student responses are likely to include different wording than the sample answer provided. However, the acceptable responses must:

- be within the specified word limit
- reflect the characteristics described in the exemplar answer.

A sample answer is provided below.

Table 2 - Answer table for Question 5

Data classification categories	**What does it represent? *(Include 3 examples)*** *(25 - 35 words)*	**Purpose for businesses** *(15 – 30 words)*
Text	Data in the form of words. This may include sequences of words and non-word characters (numeric, alphanumeric) organised by sections and sub-sections. Examples include: • online reviews • email body text • customer survey data Other examples include: • text from documents and records • transcripts	Used to analyse sentiment, urgency, emotion, or topical categories To understand the context and human language.
Audio/video	This represents data in the form of sound and/or motion pictures. Some examples include: • Call-centre recordings (audio), • Data from surveillance cameras (video) • Data from live streams and broadcasts (audio/video]. Other examples include: • Dashcams in vehicles • Satellite data • Uploaded audio/video on streaming sites (e.g. YouTube etc)	For surveillance of traffic, and suspicious activity. Researchers use this information to analyse user behaviour or actions in specific scenarios.
Web	Data that is available in the World Wide Web and captured within semi-structured formats. Examples include: • HTML web pages, • Really Simple Syndication (RSS) feeds, • JavaScript Object Notation (JSON) files. Other examples include: • XML • CSS (Cascading Style Sheets)	Data from the web classification can be used for market research, product/customer surveys and competitor analysis
Network	Data obtained from information technology, telecommunication systems and devices that are connected to the internet. This may include: - data from IoT sensors - wearable devices - computer system data	To understand user activity on the network, and frequency of access to certain systems, services, and websites. To conduct digital forensics in the event of any

Data classification categories	What does it represent? *(Include 3 examples)* *(25 - 35 words)*	Purpose for businesses *(15 – 30 words)*
	Other examples may include: - device logs - network activity logs - location tracking data	misconduct, or suspicious activity.

Question 6

Outline the statistical concepts relating to big data analytics for each listed criterion.

(Word count: 35 - 45 words for each criterion)

Assessor instructions: Student responses are likely to include different wording than the sample answer provided. However, the acceptable responses must:

- be within the specified word limit
- reflect the characteristics described in the exemplar answer.

A sample answer is provided below.

Table 3 - Answer table for Question 6

Criterion	Answer *(25 - 45 words)*
Measures of location	These provide estimates of a single value that in some way represent where the data is centered. Several statistical measures that characterise this include: • Mean • Median • Mode • Midrange • Outliers
Measures of dispersion	These provide estimations that relate to the degree of variation in the data. Several statistical measures that characterise dispersion include: • Range • Interquartile Range (IQR) • Variance • Standard deviation
Measures of shape	These help to describe the distribution or pattern of the data within a dataset either as symmetrical or asymmetrical. Several statistical measures that characterise the shape of data include: • Skewness – may be positively or negatively skewed. • Kurtosis – indicates whether the distribution is taller or thinner.

Answer **questions 7 to 10** based on given scenarios.

Scenario:

'AUS Retail' started off as a single retail store based in Sydney NSW. They now have retail store locations across several other states and territories in Australia and continue to grow with the goal of eventually setting up stores across all states in Australia.

Question 7

Outline the purpose and benefits to the organisation (AUS Retail) of big data analysis.

(Word count: 55 - 75 words)

Assessor instructions: Student responses are likely to include different wording than the sample answer provided. However, the acceptable responses must:

- be within the specified word limit
- reflect the characteristics described in the exemplar answer.

A sample answer is provided below.

The purpose of analysing big data is to gain insights into a business and to make use of 'data' considering it as an asset of a business.
Analysing big data helps the organisation to:

- gain competitive advantage
- drive supply chain efficiencies
- enable data-driven decision making
- implement better strategies to meet business goals
- minimise losses to the organisation and increase its revenue

Other answers include:

- improve what they know about customer's wants and needs
- optimise business processes by knowing the inefficiencies and opportunities for improvement in current business practices.

Scenario continued…

As the business is growing rapidly, AUS Retail's management demands a more accurate and efficient way to gain insights into their business operations.

AUS Retail currently has 10 retail stores across five different states and territories in Australia.

- 3 stores in NSW (Sydney, New Castle and Dubbo)
- 1 store in ACT (Canberra)
- 2 stores in QLD (Brisbane and Cairns)
- 2 stores in VIC (Melbourne and Bendigo)
- 2 stores in WA (Perth, Bunbury)

There are around 1857 different products from 315 vendors that are sold at each of the retail stores. However, recent reports from several stores indicated that:

- some products don't sell very often in some store locations
- the gross profit margins were below the target, which is 30% per annum.

The production department had indicated that most vendors have increased the product prices, therefore:

- they have to decrease the number of products purchased from the vendors for distribution across the stores due to resource constraints.
- It is becoming more and more difficult to identify which products to give priority to, as the production department has no visibility of the demand for products in each store.

AUS Retail's Marketing department also recently noticed that the organisation's social media site had collected a lot of product review data from customers, which is a combination of both negative and positive feedback.

Question 8

According to the scenario, outline four (4) organisational requirements for big data analysis.

(Word count: 55 - 75 words)

Assessor instructions: Student responses are likely to include different wording than the sample answer provided. However, the acceptable responses must:

- be within the specified word limit
- list four (4) organisational requirements
- reflect the characteristics described in the exemplar answer.

A sample answer is provided below.

- Determine which products have the most demand at each store location.
- Determine the quantities sold of the products that are most in demand at each store location.
- Determine the gross profit margins at each store to identify which stores have less than the target level of 30%.
- Determine which vendors have increased the product prices and the frequency of product price increases.

Other answers include:

- Determine the quality of products sold by considering the customer feedback to determine which products have received the most negative feedback or positive feedback.

Scenario continued…

AUS Retail wants to conduct an analysis of the following types of data that only relates to the business operations of the stores in NSW. The type of data collected for the analysis includes:

1. Customer details and location (who purchased the products, from which location}
2. Data on population growth in NSW from the Australian Bureau of Statistics official website.
3. Store sales details (transactions from the 3 individual stores within NSW)
4. Customer reviews on store services from the AUS Retail's social media site which relates to the stores in NSW,
5. Product details (which products from which categories were sold and their quantities within the stores in NSW.)
6. Product quality survey data from third-party service provider collected from customers in NSW.

Refer to the ***AUS Retail_Data source identification policy.pdf*** document for more details on the organisational policies, procedures and legislative requirements that apply when identifying internal and external sources of big data to be analysed.

Question 9

Identify the internal and external sources of each type of data collected for the analysis according to the given scenario, organisational policies and procedures.

You must complete the answer table by,

a) listing the six (6) types of data collected for the analysis according to the scenario

b) indicating the data source as either internal or external according to the organisational policies and procedures
c) indicating the source system details (e.g. name of the source system, website, platform or provider).

Assessor instructions: Student responses are likely to include different wording than the sample answer provided for 'Type of data collected' and 'Source system details' columns. However, the acceptable responses must:

- list six (6) types of data collected
- identify the type of source either as 'Internal source' or 'External source'
- mention the name of the source system
- reflect the characteristics described in the exemplar answer.

A sample answer is provided below.

Table 4 - Answer table for Question 9

#	Type of data collected:	Type of source *(Indicate as Internal or External)*	Source system details *(Name of the system, website, platform, provider etc.)*
1.	Customer details	☒ Internal source ☐ External source	From AUS Retail's internal Customer Relationship Management System (CRMS).
2.	Population growth data for NSW.	☐ Internal source ☒ External source	From the Australian Bureau of Statistics official website.
3.	Store sales details	☒ Internal source ☐ External source	From AUS Retail's internal Transaction processing system.
4.	Customer reviews on store services	☐ Internal source ☒ External source	From social media posts.
5.	Product details	☒ Internal source ☐ External source	From AUS Retail's internal Product management system.
6.	Product quality data	☐ Internal source ☒ External source	From a third-party service provider

Question 10

Outline the legislative requirements that apply when identifying internal and external sources of big data for analysis according to the given scenario, organisational policies and procedures.

Use the table given below to record your answers for each criterion:

a) Legislative requirements for:
   a. Identifying internal sources of big data for analysis
   b. Identifying external sources of big data for analysis
   (15 – 25 words for each)

b) Two (2) examples of legislation that apply for each requirement.

Assessor instructions: Student responses are likely to include different wording than the sample answer provided. However, the acceptable responses must:

- be within the specified word limit
- reflect the characteristics described in the exemplar answer
- list two (2) examples of legislation for each requirement.

A sample answer is provided below.

Table 5 - Answer table for Question 10

Criterion:	Identifying internal sources of big data for analysis	Identifying external sources of big data for analysis
Legislative requirement: *(15 – 35 words for each requirement)*	All data that contain customer information need to be de-identified before using it in the analysis.	Check that the third-party service providers have good privacy practices in place to provide accurate information De-identify any personal information in the third-party data before using it in the analysis.
Examples of applicable legislation: *(Include two examples for each requirement)*	1. Data Availability and Transparency Act 2022 2. Private and Personal Information Protection Act 1998 Other examples include: • Privacy Act 1988 • Privacy Regulation 2013 • Australian Privacy Principles	1. Privacy Act 1988 2. Australian Privacy Principles Other examples include: • Data Availability and Transparency Act 2022 • Data-matching program Act 1990 • Privacy Regulation 2013

Assessment checklist:

Students must have completed all questions within this assessment before submitting. This includes:

1	10 short answer questions to be completed in the spaces provided.	☐

Congratulations you have reached the end of Assessment [2]!