

[Print friendly format](#)


Question 1

Correct

Mark 1.00 out
of 1.00

What is polynomial regression?

Select one:

- ☐ a. Fitting many lines to data.
- ☐ b. Many numbers regression.
- ☒ c. Fitting a curve defined by a polynomial function to data.  Correct. Polynomial regression means fitting a curve to data.
- ☐ d. Fitting a curve to a line.

Your answer is correct.

The correct answer is: Fitting a curve defined by a polynomial function to data.

Question 2

Correct

Mark 1.00 out
of 1.00


What can you deduce from the following data?

Age, Loan

20, No

21, Yes

Select one:

- ☐ a. Older people don't get loans.
- ☒ b. This 20 year old didn't get a loan.  Correct. You can deduce this 20 year old didn't get a loan.
- ☐ c. Younger people get loans.
- ☐ d. Loan approvals are random.

Your answer is correct.

Your answer is correct.

The correct answer is: This 20 year old didn't get a loan.

Question 3

Incorrect

Mark 0.00 out of 1.00

What is the meaning of the term "The Unreasonable Effectiveness of Data"?

Select one:

- ☐ a. The data is unreasonable.
- ☒ b. Good algorithms are better. ✖ Incorrect. The Unreasonable Effectiveness of Data doesn't mean good algorithms are better.
Review [Introduction to Data Analysis](#) in [Data Analysis Process](#) and then try again.
- ☐ c. More data is usually better.
- ☐ d. There's too much data.

Your answer is incorrect.

The correct answer is: More data is usually better.

Question 4

Correct

Mark 1.00 out of 1.00

Why was the Google Flu Study criticised?

Select one:

- ☒ a. All options are correct. ✔ Correct. The Google Flu Study was criticised due to a number of serious systematic errors.
- ☐ b. Because of the proprietary nature of Google's search engine's internals, scientific reproducibility is very difficult.
- ☐ c. Google apparently did not use standard time series models.
- ☐ d. Google's search engine and its query logs are constantly changing so it makes the sources unclear.

Your answer is correct.

The correct answer is: All options are correct.

Question 5

Incorrect

Mark 0.00 out of 1.00

What is the purpose of linear regression?

Select one:

- ☒ a. To fit data to a model. ✖ Incorrect. Review [Activity: Regression in iPython](#) in [Data Analysis Process](#) and then try again.
- ☐ b. To draw a line through the middle of data.
- ☐ c. To fit a model to data.
- ☐ d. To draw a line through data.

Your answer is incorrect.

The correct answer is: To fit a model to data.

Question 6

Correct

Mark 1.00 out of 1.00

Complete this sentence. A 'good fit' linear regression line can be achieved by:

Select one:

- ☐ a. Drawing a line backwards.
- ☒ b. Drawing a line through the middle of data to minimise the distance from the line to all the points.
✔ Correct. Linear regression means to draw a line through the middle of data to minimise the distance from the line to all the points.
- ☐ c. Drawing a line through data.
- ☐ d. Drawing a line through the middle of data.

Your answer is correct.

The correct answer is: Drawing a line through the middle of data to minimise the distance from the line to all the points.

Question 7

Correct

Mark 1.00 out of 1.00

Complete this sentence. $y = ax + b$ is a function describing a line:

Select one:

- ☐ a. Where **a** is the **intercept** and **b** is the **incline**.
- ☒ b. Where **b** is the **intercept** and **a** is the **incline**. ✔ Correct. $y = ax + b$ is a function describing a line where **b** is the **intercept** and **a** is the **incline**.
- ☐ c. Can't say without **a** and **b**.

Your answer is correct.

The correct answer is: Where **b** is the **intercept** and **a** is the **incline**.

Question 8

Correct

Mark 1.00 out of 1.00

Complete this sentence. In a discussion forum, off-topic posts and spam are regarded as an example of:

Select one:

- ☐ a. Correlation versus causation.
- ☒ b. Signal versus noise. ✔ Correct. Off-topic posts and spam are regarded as an example of signal versus noise.
- ☐ c. No-free Lunch Theorem.
- ☐ d. Ensembles

Your answer is correct.

The correct answer is: Signal versus noise.

Question 9

Correct

Mark 1.00 out of 1.00

Complete this sentence. The 'ML' in BigML means:

Select one:

- ☐ a. Markup Language.
- ☒ b. Machine Learning. ✔ Correct. The 'ML' in BigML means machine learning.
- ☐ c. A new clothes size, not quite 'M', not quite 'L'.
- ☐ d. Modelling Language.

Your answer is correct.

The correct answer is: Machine Learning.

Mark 1.00 out of 1.00

Select one:

- ☐ a. Markup Language.
- ☒ b. Machine Learning. ✔ Correct. The 'ML' in BigML means machine learning.
- ☐ c. A new clothes size, not quite 'M', not quite 'L'.
- ☐ d. Modelling Language.

Your answer is correct.

The correct answer is: Machine Learning.

The correct answer is: Machine Learning.

Question 10

Correct

Mark 1.00 out of 1.00

What rule can you deduce from the following data?

Age, Loan

20, No

21, Yes

22, No

Select one:

- ☐ a. Not much, more rows of data may help.
- ☐ b. Younger people get loans.
- ☒ c. Not much, more columns of data may help. ✓ Correct. More data may help. e.g. Salary
- ☐ d. Older people don't get loans.

Your answer is correct.

The correct answer is: Not much, more columns of data may help.

[Full memory format](#)

Question 1

Incorrect

Mark 0.00 out of 1.00

Complete this sentence. Chris Bain:

Select one:

- ☐ a. talks about levels of hospital data protection, both in policy, in access control, and in technical capability that prevent inappropriate use of data.
- ☒ b. is an advocate of open data, everyone in the hospital has access to all data. ✗ Unlikely. Review **Interview on Data Management** and then try again.
- ☐ c. is an advocate of closed hospital data, nobody has access because of privacy concerns.
- ☐ d. doesn't even work in a hospital.

Your answer is incorrect.

The correct answer is: talks about levels of hospital data protection, both in policy, in access control, and in technical capability that prevent inappropriate use of data.

Question 2

Correct

Mark 1.00 out of 1.00

Complete this sentence. Data is a non-fungible asset means:

Select one:

- ☒ a. it's irreplaceable. ✓ Correct. A non-fungible asset cannot be replaced.
- ☐ b. it doesn't decay.
- ☐ c. it's not refundable.
- ☐ d. it's not possible to have fun with data.

Question 3

Correct

Mark 1.00 out of 1.00

Complete this sentence. Data Governance:

Select one:

- ☐ a. is a made up word, there's a lot of that.
- ☐ b. means the same as Data Management.
- ☒ c. means the authority, control and shared decision making over the management of data assets.
✔ Correct. Data Governance means means the authority, control and shared decision making over the management of data assets.
- ☐ d. means who owns the data.

Your answer is correct.

The correct answer is: means the authority, control and shared decision making over the management of data assets.

Question 4

Correct

Mark 1.00 out of 1.00

Complete this sentence. Privacy and confidentiality in Data Science:

Select one:

- ☒ a. have different meanings, one is about oneself, the other is about your data.
✔ Correct. Privacy and confidentiality in Data Science have different meanings.
- ☐ b. are the same.
- ☐ c. is about having control over how one shares oneself with others.
- ☐ d. is about how information about an individual is treated.

Your answer is correct.

The correct answer is: have different meanings, one is about oneself, the other is about your data.

Question 5

Correct

Mark 1.00 out of 1.00

Choose the correct option to complete this statement.

The DCC lifecycle model.

Select one:

- ☐ a. Every aspect of the model is 'action'.
- ☐ b. Has eight types of action, from Conceptualise to Transform.
- ☒ c. Has three types of action: Full lifecycle, Sequential and Occasional.
✔ Correct. The DCC lifecycle model has three types of actions.
- ☐ d. Doesn't have any specific 'actions'.

Your answer is correct.

Question 6


Correct

Mark 1.00 out of 1.00

Choose the correct option to complete this statement.

There are **four** political camps in data science.

Select one:

- ☐ a. Actually there are **five** because one of them can be split into two groups.
- ☐ b. Actually there are only two, liberal and conservative.
- ☒ c. Is just an opinion.  Correct. It's just an opinion that there are four political camps in data science.
- ☐ d. Is one of the Laws of Data Science.

Your answer is correct.

The correct answer is: Is just an opinion.


Question 7

Correct

Mark 1.00 out of 1.00

Complete this sentence. Data Management:

Select one:

- ☐ a. is the same as data governance.
- ☒ b. is about the development, execution and supervision of plans, policies, programs and practices that control, protect, deliver and enhance the value of data and information assets.  Correct. Data management is about the development, execution and supervision of plans, policies, programs and practices that control, protect, deliver and enhance the value of data and information assets.
- ☐ c. is the same as data analysis.
- ☐ d. is the same as data curation.

Your answer is correct.

The correct answer is: is about the development, execution and supervision of plans, policies, programs and practices that control, protect, deliver and enhance the value of data and information assets.


Question 8

Correct

Mark 1.00 out of 1.00

Complete this sentence. Data management project frameworks in Medicine and Health are best based on:

Select one:

- ☒ a. are their own unique area.  Correct. Data management project frameworks are their own unique area.
- ☐ b. the Government framework.
- ☐ c. a combination of Business/Organisational and Science frameworks.
- ☐ d. a combination of Government and Science frameworks.

Your answer is correct.

The correct answer is: are their own unique area.


Question 9

Correct

Mark 1.00 out of 1.00

Complete this sentence. Liking a curly fry page on Facebook being an indicator of high intelligence is an example of:

Select one:

- ☐ a. Facebook marketing.
- ☐ b. curly fry advertising.
- ☐ c. causality.
- ☒ d. homophily.  Correct. Liking a curly fry page on Facebook being an indicator of high intelligence is an example of homophily.

Your answer is correct.

The correct answer is: homophily.

Question 10


Correct

Mark 1.00 out of 1.00

Bob posts on Facebook that he's "going for lunch with Jim". His boss, a canny Data Scientist, discovers there's no Jim in his online friends or followers and deduces that this is a reference to Jim's Tavern (and surprises him having a beer during work hours).

What type of data did Bob make available from his post to Facebook?

Select one:

- ☐ a. Implicit data.
- ☐ b. Implicit then explicit data.
- ☒ c. Explicit then implicit data.  Correct. Bob made explicit and then implicit data available from his post to Facebook.
- ☐ d. Explicit data.

Your answer is correct.

The correct answer is: Explicit then implicit data.

Question 1

Correct

Mark 1.00 out of 1.00

Which of the following would not be regarded as a step in the data value chain?

Select one:

- ☐ a. Transform and prepare.
- ☒ b. Security. ✓ Correct. This is but a small part of the management or governance step.
- ☐ c. Presentation.
- ☐ d. Engineering.

Your answer is correct.

The correct answer is: Security.

Question 2

Correct

Mark 1.00 out of 1.00

What is The Big Data Landscape?

Select one:

- ☐ a. A value chain for big data.
- ☐ b. A famous photograph by Ansel Adams.
- ☒ c. A way of organising different big data businesses and projects hierarchically. ✓ Correct. The Big Data Landscape is a way of organising different big data businesses and projects hierarchically.
- ☐ d. A website by Matt Turck.

Question 3

Correct

Mark 1.00 out

What is a dataframe?

Select one:

https://lms.monash.edu/mod/quiz/review.php?attempt=14244012&cmid=6454739&print=1

Page 1 of 1

Check your understanding of data models in organisations: Attempt review

8/3/20, 2:31

of 1.00

- ☒ a. A structure that stores tabular data. ✓ Correct. A dataframe is a structure that stores tabular data.
- ☐ b. A list.
- ☐ c. An array.
- ☐ d. A theory about data.

Your answer is correct.

The correct answer is: A structure that stores tabular data.

Question 4

Correct

Mark 1.00 out of 1.00

When working with SAS Visual Analytics, what is the meaning of the term "drill down"?

Select one:

- ☐ a. To put your tools down and have a break.
- ☐ b. To build the database.
- ☒ c. To explore lower levels of data in a hierarchy. ✓ Correct. The analogy is drilling for gold or oil.
- ☐ d. To enter data.

Your answer is correct.

The correct answer is: To explore lower levels of data in a hierarchy.

Question 5

Correct

Mark 1.00 out of 1.00

Complete this sentence. An Influence Diagram:

Select one:

- ☐ a. is an alternative to decision tree.
- ☐ b. is a model giving possible situations or outcomes.
- ☒ c. consists of nodes and arcs and is an alternative to decision tree. ✓ Correct. An Influence Diagram consists of nodes and arcs and is an alternative to decision tree.
- ☐ d. consists of nodes and arcs.

Your answer is correct.

The correct answer is: consists of nodes and arcs and is an alternative to decision tree.

Question 6

Correct

Mark 1.00 out of 1.00

What is .ipynb?

Select one:

- ☐ a. An illegal file extension, 3 is the maximum length.
- ☐ b. nb is Nota bene.
- ☒ c. Interactive Python Notebook. ✓ Correct. .ipynb is an Interactive Python Notebook.
- ☐ d. A terrible spelling of 'pineapple'.

Question 7
Incorrect
Mark 0.00 out of 1.00

Pivotal's data value chain takes a broader view of the use of data, from capture and storage through to changing business practices. How does their notion of a data scientist differ from that given by our Standard Value Chain?

Select one:

- ☐ a. Their DS doesn't do the major data extraction task.
- ☐ b. Their DS doesn't propose changes to applications.
- ☐ c. Their DS doesn't do visualisation.
- ☒ d. Their DS doesn't identify candidate business tasks to apply data science too. ✖ Incorrect. Well, we cannot be sure of this one.
Review Miller and Mork's **Data to Decisions: A Value Chain for Big Data in Data and Decision Models** and then try again.

Your answer is incorrect.

The correct answer is: Their DS doesn't do the major data extraction task.

Question 8
Correct
Mark 1.00 out of 1.00

Which of the following is an optimization or prescriptive analytics task (as opposed to a predictive analytics task)?

Select one:

- ☐ a. Recommend a credit card transaction be checked for fraud.
- ☐ b. Recommending a web advertisement to show a website visitor.
- ☒ c. Recommending a traffic route based on prior data for the time of data and incident reports. ✔ Correct. Not only must predictions be made, but an optimisation done to get the best route.
- ☐ d. Recommending that a doctor check the patient's nursing notes for evidence of fungal infection.

Your answer is correct.

The correct answer is: Recommending a traffic route based on prior data for the time of data and incident reports.

Question 9
Correct
Mark 1.00 out of 1.00

If a value y is determined by a function $y = ax + b$ where a is -27 and b is 1548 , then a and b are:

Select one:

- ☐ a. Constants.
- ☒ b. Slope and intercept respectively. ✔ Correct. For example, if age is 32 then salary is on the line: $y = -27 \cdot 32 + 1548 = 648$
- ☐ c. Can't say because we don't know x .
- ☐ d. Unknown.

Your answer is correct.

The correct answer is: Slope and intercept respectively.

Question 10
Correct
Mark 1.00 out of 1.00

NAB uses a customer relationship management approach to target customers:

Select one:

- ☒ a. Using all the information about a customer they have. ✔ Correct. They try to be as targeted as possible.
- ☐ b. Not using their behavioural data due to privacy concerns.
- ☐ c. Using a one-size-fits all approach because it is more robust.

Question 1
Incorrect
Mark 0.00 out
of 1.00

What software tool does Dr Rami Mukhtar mention in his interview about software and tools?

Select one:

- ☐ a. Paypal Rabbit
- ☐ b. Vowpal Wabbit
- ☐ c. Focal Rabbit
- ☒ d. Power-Grabbit ✖ Incorrect. Review **Interview on Software and Tools** in **Data Resources, Processes, Standards and Tools** and try again.

Your answer is incorrect.

The correct answer is: Vowpal Wabbit

Question 2
Correct
Mark 1.00 out
of 1.00

What does data wrangling mean?

Select one:

- ☐ a. To capture data from sensors.
- ☒ b. To tidy up or clean data. ✔ Correct. Data wrangling means to tidy up or clean data.
- ☐ c. To brand or own data.
- ☐ d. To steal data.

Your answer is correct.

The correct answer is: To tidy up or clean data.


Question 3

Correct

Mark 1.00 out of 1.00

What were the main types of data problem found in the patients data file?

Select one:

- ☐ a. Metadata.
- ☒ b. Missing values, invalid values, duplicate values.  Correct. The main types of data problem found in the patients data file were missing values, invalid values, duplicate values.
- ☐ c. Can't possibly say because it's source is unknown.
- ☐ d. Length, not enough data.

Your answer is correct.

The correct answer is: Missing values, invalid values, duplicate values.


Question 4

Correct

Mark 1.00 out of 1.00

What is the meaning of pharmacovigilance?

Select one:

- ☐ a. It is a made up word.
- ☐ b. Drug companies protecting their patents.
- ☐ c. Drug companies worrying about their income.
- ☒ d. Means the practice of monitoring the effects of drugs.  Correct. Pharmacovigilance means the practice of monitoring the effects of drugs.

Your answer is correct.

The correct answer is: Means the practice of monitoring the effects of drugs.


Question 5

Correct

Mark 1.00 out of 1.00

Which of the following are metadata?

Select one:

- ☒ a. All options are correct.  Correct. They are all data about data, name, type and range.
- ☐ b. Heart rate
- ☐ c. Numeric
- ☐ d. 40 to 100

Your answer is correct.

The correct answer is: All options are correct.


Question 6

Correct

Mark 1.00 out of 1.00

Which of the following are metadata?

Select one:

- ☒ a. All options are correct.  Correct. They are all data about data, name, type and range.

- ☐ b. 40 to 100
- ☐ c. Heart rate
- ☐ d. Numeric

Your answer is correct.

The correct answer is: All options are correct.

Question 7
Correct
Mark 1.00 out of 1.00

Why is the exact phrase "Apache Cassandra" is difficult to research in Google trends?

Select one:

- ☐ a. Because "Cassandra" is a girls name and a DB name.
- ☐ b. Because the Cassandra DB isn't famous.
- ☒ c. The search excludes instances of "Cassandra". ✓ Correct. Solve the ambiguity problem, create another.
- ☐ d. Because Apache is a tribe of American Indians too.

Your answer is correct.

The correct answer is: The search excludes instances of "Cassandra".

Question 8
Correct
Mark 1.00 out of 1.00

What type of data format is used in the patients data file?

Select one:

- ☒ a. All options are correct. ✓ Correct. The patients data file uses text, fixed width and flat file formatting.
- ☐ b. Fixed width
- ☐ c. Text
- ☐ d. Flat file

Your answer is correct.

The correct answer is: All options are correct.

Question 9
Correct
Mark 1.00 out of 1.00

A vector of ages data was saved to file in the following format:

```
{"Age":{"0":39,"1":28,"2":44,"3":25,"4":32,"5":33,"6":31,"7":26,"8":22,"9":25,"10":28}}
```

What format is this?

Select one:

- ☐ a. CSV
- ☒ b. JSON ✓ Correct. The vector of ages data was saved in JSON.
- ☐ c. XML
- ☐ d. RDF

Your answer is correct.

The correct answer is: JSON

Question 10

Correct

Mark 1.00 out of 1.00

What would be the output if the following patient data was imported, cleaned up and then saved to .csv?

001M11/11/1998 88140 80 10

Select one:

- ☒ a. 001,M,11/11/1998,88,140,80,1,0 ✓ Correct. Remember, there should be **no** spaces between each comma separated value.
- ☐ b. 001,M,11/11/1998, 88,140, 80, 1,0
- ☐ c. 001M11/11/1998 88140 80 10

Your answer is correct.

The correct answer is: 001,M,11/11/1998,88,140,80,1,0

Question 1

Correct

Mark 1.00 out of 1.00

How many dimensions of data can Motion Charts display? You need to select more than one.

Select one:

- ☐ a. 2 dimensions (e.g. x and y)
- ☒ b. 4+ dimensions ✓ Correct. In fact data should be in the format **name/category, date/time, 4 columns of data** as mentioned in **Activity: Motion Charts**.
- ☐ c. 1 dimension
- ☐ d. 3 dimensions

Your answer is correct.

The correct answers are: 3 dimensions, 4+ dimensions

Question 2

Correct

Mark 1.00 out of 1.00

Which of the following is **not** a characteristic of a data scientist according to Mike Loukides of O'Reilly?

Select one:

- ☐ a. Be versatile.
- ☒ b. Be able to read a statistical journal article. ✓ Correct. Some data scientists might need to read a book on machine learning, and all need to be able to use some standard statistical or machine learning software.
- ☐ c. Be interdisciplinary.
- ☐ d. Be able to program computation on data.
- ☐ e. Be entrepreneurial.


Question 3

Correct

Mark 1.00 out of 1.00

Which of the following has **not** been proposed as a definition of Data Science?

Select one:

- ☐ a. Extraction of actionable knowledge from data through the complete lifecycle process.
- ☐ b. Almost everything concerned with data, especially applications.
- ☐ c. Organisation, storage, modelling, reporting on data and the delivery of value.
- ☐ d. The use of learning on big data in a distributed computing environment.
- ☒ e. Data-driven processes in science or business.  Correct. This seems somewhat reasonable, but it does not include Government. That means the definition is incomplete. Remember, although Business Intelligence is data-driven it isn't always Data Science.

Your answer is correct.

The correct answer is: Data-driven processes in science or business.


Question 4

Correct

Mark 1.00 out of 1.00

Which of the following is **not** a historical precursor to Data Science?

Select one:

- ☐ a. Statistical platforms such as R, SAS or Weka.
- ☒ b. IBM's cognitive computing.  Correct. IBM's cognitive computing is an application of sophisticated natural language processing. It's related, but not a historical precursor to Data Science.
- ☐ c. Cheap computing hardware.
- ☐ d. Google and Amazon's data-driven operations.

Your answer is correct.

The correct answer is: IBM's cognitive computing.


Question 5

Correct

Mark 1.00 out of 1.00

When did Data Science begin?

Select one:

- ☐ a. 1974, when Peter Naur used the term 'data science' freely in 'Concise Survey of Computer Methods'.
- ☐ b. 1962, with the article "The Future of Data Analysis".
- ☐ c. 1996, IFCS, "Data science, classification, and related methods".
- ☒ d. Data Science has been happening for a long time, it just called something else.  We have seen several views of the history of Data Science, most differ, some go back to the dawn of civilisation (e.g. Wolfram). You can trace the history of the term "Data Science" but that's not the same.

Question 6

Correct

Mark 1.00 out of 1.00

Complete this sentence. The Scientific Method can be impacted by Data Science because:

Select one:

- ☐ a. Non-scientists can contribute via data science competitions.
- ☒ b. The technology to adopt data-driven science exists, and it's importance is more widely understood. ✓ Correct. The Scientific Method can be impacted by Data Science because the technology to adopt data-driven science exists, and it's importance is more widely understood.
- ☐ c. It offers new statistical theories to use.
- ☐ d. It shows correlation is adequate for providing deep understanding.

Your answer is correct.

The correct answer is: The technology to adopt data-driven science exists, and it's importance is more widely understood.

Question 7

Correct

Mark 1.00 out of 1.00

Which of the following are examples of correlation?

Select one:

- ☒ a. All options are correct. ✓ Correct. These are all correlations i.e. events that seem to be or are related. If a supermarket had strong evidence that shoppers are buying beer and chips they may like to put them side by side and see if sales increase. Review 'market basket analysis or 'affinity analysis' in the [Affinity analysis](#) section of Wikipedia to learn more about correlation.
- ☐ b. There is a spill on the floor and a broken bottle.
- ☐ c. In a supermarket, people who bought beer also bought chips.
- ☐ d. In a supermarket, people who bought chips also bought beer.

Your answer is correct.

The correct answer is: All options are correct.

Question 8

Correct

Mark 1.00 out of 1.00

Which of the following are examples of causality?

Select one:

- ☐ a. In a supermarket, shoppers who bought beer also bought chips.
- ☐ b. All options are correct.
- ☐ c. There is a spill on the floor, a broken bottle of beer and only one shopper near by, they must have caused the mess.
- ☒ d. A shopper dropped a beer bottle on the floor and it smashed. ✓ Correct. There is a cause and effect here: oops, gravity, mess, sad. Review **Activity: Motion Charts** to learn more about causality and then try again.

Your answer is correct.

The correct answer is: A shopper dropped a beer bottle on the floor and it smashed.


Question 9

Correct

Mark 1.00 out of 1.00

Motion Charts is given data from 1900 to 2000, but only intermediary years are displayed. What process is Motion Charts using to display this?

Select one:

- ☐ a. Data munging
- ☐ b. Extrapolation
- ☐ c. Hyperpolation
- ☒ d. Interpolation  Correct. Motion Charts is 'filling in the gaps'. Interpolation is 'between the points' of a sequence of values.

Your answer is correct.

The correct answer is: Interpolation


Question 10

Incorrect

Mark 0.00 out of 1.00

Complete this sentence. According to Cukier and Mayer-Schoenberger:

Select one:

- ☐ a. Observing and making use of correlations from big data is a move away from trying to develop deep causal models, traditionally done in science.
- ☒ b. Datafication is the digitization of formerly analogue content.  Incorrect. It is a lot more. It is the digitization of many aspects of life.
Review **Overview of Data Science** and try again.
- ☐ c. Google's clever use of high-quality translations allowed them to develop a better automated translation system.
- ☐ d. A key technology for using big data is the use of statistical sampling, being able to infer information from small, carefully curated samples.

Your answer is incorrect.

The correct answer is: Observing and making use of correlations from big data is a move away from trying to develop deep causal models, traditionally done in science.

[Link memory format](#)


Question 1

Correct

Mark 1.00 out
of 1.00

Complete this sentence. The two most prevalent types of analysis on the Enron data set include:

Select one:

- ☐ a. Social network analysis and visualisation.
- ☐ b. Thematic structure analysis.
- ☒ c. Social network analysis and visualisation, and thematic structure analysis.  Correct. Social network analysis and visualisation, and thematic structure analysis are the two most prevalent types of analysis.
- ☐ d. Thematic structure analysis and topic modelling.

Your answer is correct.

The correct answer is: Social network analysis and visualisation, and thematic structure analysis.


Question 2

Correct

Mark 1.00 out
of 1.00

What is the correct order, in **decreasing** size of the following units of digital information?

Select one:

- ☐ a. Mega, giga, tera, peta.
- ☐ b. Exa, zetta, peta, tera, giga, mega.
- ☐ c. Exa, peta, zetta, tera, giga, mega.
- ☒ d. Exa, peta, tera, giga, mega.  Correct. There was a time when 'giga' was a mystery, now 'tera' is in common usage.

Your answer is correct.

The correct answer is: Exa, peta, tera, giga, mega.

Question 3

Correct

Mark 1.00 out of 1.00

Which of the following are 'machine-readable data'?

Select one:

- ☐ a. RDF
- ☒ b. All options are correct. ✓ Correct. RDF, XML and JSON are all machine-readable data.
- ☐ c. JSON
- ☐ d. XML

Your answer is correct.

The correct answer is: All options are correct.

Question 4

Correct

Mark 1.00 out of 1.00

Which laws are predicated by (and derived from), Moore's Law?

Select one:

- ☒ a. All of them. ✓ Correct. Zimmermann's, Bell's, Koomey's laws are predicated by (and derived from) Moore's Law.
- ☐ b. Bell
- ☐ c. Koomey
- ☐ d. Zimmermann

Your answer is correct.

The correct answer is: All of them.

Question 5

Correct

Mark 1.00 out of 1.00

Complete this sentence. The fact that the Square Kilometer Array (SKA) is expected to generate 700 terabytes per day and there's a proposal to store data in 1 gram of DNA, also 700 terabytes, is:

Select one:

- ☐ a. good planning.
- ☐ b. good design.
- ☒ c. a coincidence. ✓ Correct. It's a coincidence that there's a proposal to store data in 1 gram of DNA.
- ☐ d. not possible, neither are true.

Your answer is correct.

The correct answer is: a coincidence.

Question 6

Correct

Mark 1.00 out

What are the 'walls' that the RDBMS/SQL paradigm has hit in business?

Select one:


<https://lms.monash.edu/mod/quiz/review.php?attempt=14244057&cmid=6454786#question-14574509-7>

Page 2

Check your understanding of data types and storage: Attempt review

8/3/20, 2:37

of 1.00

- ☐ a. Businesses require insights faster i.e. in real time.
- ☒ b. All options are correct.  Correct. The RDBMS/SQL paradigm has hit all of the 'walls' listed.
- ☐ c. Businesses require data driven decision-making, massive amounts of data.
- ☐ d. Businesses function in a continuously changing environment, therefore the DB schema must change.

Your answer is correct.

The correct answer is: All options are correct.


Question 7

Incorrect

Mark 0.00 out
of 1.00

What is MapReduce?

Select one:

- ☐ a. A multi stage process to break up then analyse data.
- ☐ b. Owned by Apache.
- ☒ c. No longer used, Google has found an alternative.  Incorrect. It's true that Google has alternatives but others use it.
Review **Big Data Processing** in **Data Types and Storage** and try again.
- ☐ d. A way to make maps smaller.

Your answer is incorrect.

The correct answer is: A multi stage process to break up then analyse data.


Question 8

Correct

Mark 1.00 out
of 1.00

When a scientist collects data from an explosion, which of the four Vs is paramount?

Select one:

- ☒ a. Velocity  Correct. The event window is milli or even microseconds, so the issue is to capture data 'coming at' you.
- ☐ b. Variety
- ☐ c. Volume
- ☐ d. Veracity

Your answer is correct.

The correct answer is: Velocity


Question 9

Correct

Mark 1.00 out
of 1.00

Which growth law says power consumption will decrease?

Select one:

- ☐ a. Zimmerman's Law.
- ☒ b. Koomey's Law.  Correct. Koomey's Law says power consumption will decrease.
- ☐ c. Moore's Law.

☐ d. Bell's Law.

Your answer is correct.

The correct answer is: Koomey's Law.

Question **10**

Correct

Mark 1.00 out
of 1.00

Complete this sentence. If 5GB is approximately one movie then 500TB is:

Select one:

- ☐ a. A Gigabyte of movies.
- ☒ b. 100,000 movies. ✓ Correct. $100 * 1,000$.
- ☐ c. 500 movies.
- ☐ d. 1,000 movies.

Your answer is correct.

The correct answer is: 100,000 movies.