

E20-007^{Q&As}

Data Science and Big Data Analytics

Pass EC-COUNCIL E20-007 Exam with 100% Guarantee

Free Download Real Questions & Answers **PDF** and **VCE** file from:

<https://www.pass2lead.com/e20-007.html>

100% Passing Guarantee
100% Money Back Assurance

Following Questions and Answers are all new published by EC-COUNCIL Official Exam Center

- ⚙️ **Instant Download** After Purchase
- ⚙️ **100% Money Back** Guarantee
- ⚙️ **365 Days** Free Update
- ⚙️ **800,000+** Satisfied Customers



QUESTION 1

On analyzing your time series data you suspect that the data represented as $y_1, y_2, y_3, \dots, y_{n-1}, y_n$ may have a trend component that is quadratic in nature. Which pattern of data will indicate that the trend in

the time series data is quadratic in nature?

- A. $(y_3 - y_2) - (y_2 - y_1) = \dots = (y_n - y_{n-1}) - (y_{n-1} - y_{n-2})$
- B. $(y_2 - y_1) = (y_3 - y_2) = \dots = (y_n - y_{n-1})$
- C. $((y_2 - y_1) / y_1) * 100\% = \dots = ((y_n - y_{n-1}) / y_{n-1}) * 100\%$
- D. $(y_4 - y_2) - (y_3 - y_1) = \dots = (y_n - y_{n-2}) - (y_{n-1} - y_{n-3})$

Correct Answer: A

QUESTION 2

You have two tables of customers in your database. Customers in `cust_table_1` were sent an e-mail promotion last year, and customers in `cust_table_2` received a newsletter last year. Customers can only be entered in once per table. You want to create a table that includes all customers, and any of the communications they received last year. Which type of join would you use for this table?

- A. Full outer join
- B. Inner join
- C. Left outer join
- D. Cross join

Correct Answer: A

QUESTION 3

Which analytical method is considered unsupervised?

- A. K-means clustering
- B. Naïve Bayesian classifier
- C. Decision tree
- D. Linear regression

Correct Answer: A

QUESTION 4

A business colleague who is new to Hadoop approaches you with a question. The colleague wants to know the best approach to access their data. The colleague has previously worked extensively with SQL and databases.

Which query interface should be recommended?

- A. Hive
- B. Pig
- C. Howl
- D. HBase

Correct Answer: A

QUESTION 5

What is required in a presentation for business analysts?

- A. Budgetary considerations and requests
- B. Operational process changes
- C. Detailed statistical explanation of the applicable modeling theory
- D. The presentation author's credentials

Correct Answer: B

QUESTION 6

You have used k-means clustering to classify behavior of 100,000 customers for a retail store. You decide to use household income, age, gender and yearly purchase amount as measures. You have chosen to use 8 clusters and notice that 2 clusters only have 3 customers assigned. What should you do?

- A. Decrease the number of clusters
- B. Increase the number of clusters
- C. Decrease the number of measures used
- D. Identify additional measures to add to the analysis

Correct Answer: A

QUESTION 7

Which word or phrase completes the statement? Unix is to bash as Hadoop is to:

- A. Pig

- B. HDFS
- C. Sqoop
- D. NameNode

Correct Answer: A

QUESTION 8

Which data asset is an example of quasi-structured data?

- A. Webserver log
- B. XML data file
- C. Database table
- D. News article

Correct Answer: A

QUESTION 9

Which word or phrase completes the statement? Emphasis color is to standard color as _____ .

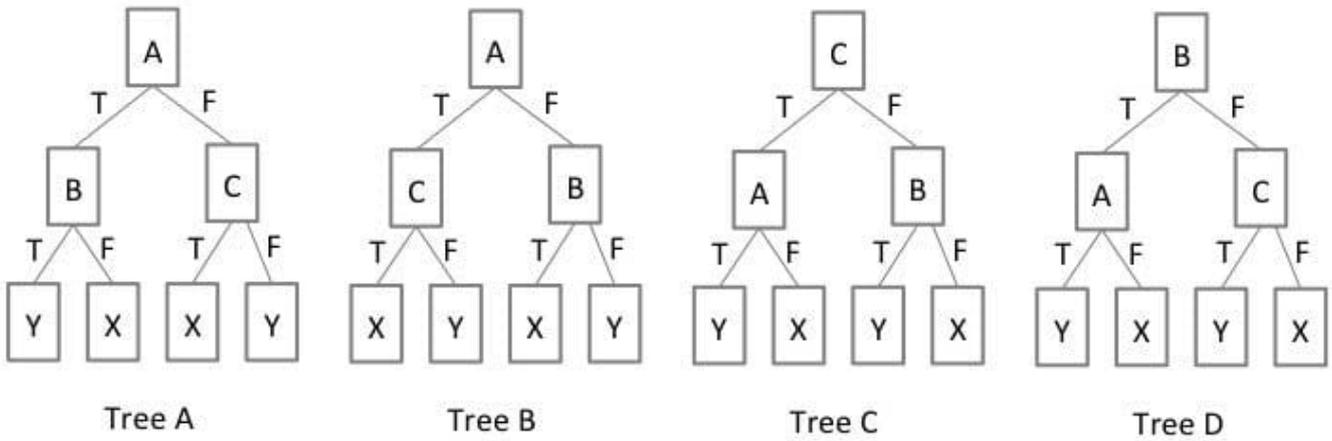
- A. Main message is to context
- B. Main message is to key findings
- C. Frequent item set is to item
- D. Pie chart is to proportions

Correct Answer: A

QUESTION 10

Refer to the Exhibit.

A	B	C	CLASS
T	T	T	X
T	T	F	Y
T	F	T	X
F	F	F	Y
F	T	T	X
F	F	T	Y



In the Exhibit, the table shows the values for the input Boolean attributes "A", "B", and "C". It also shows the values for the output attribute "class". Which decision tree is valid for the data?

- A. Tree B
- B. Tree A
- C. Tree C
- D. Tree D

Correct Answer: A

QUESTION 11

In linear regression, what indicates that an estimated coefficient is significantly different than zero?

- A. A small p-value
- B. R-squared near 1
- C. R-squared near 0

D. The estimated coefficient is greater than 3

Correct Answer: A

QUESTION 12

The average purchase size from your online sales site is \$17, 200. The customer experience team believes a certain adjustment of the website will increase sales. A pilot study on a few hundred customers showed an increase in average purchase size of \$1.47, with a significance level of $p=0.1$.

The team runs a larger study, of a few thousand customers. The second study shows an increased average purchase size of \$0.74, with a significance level of 0.03. What is your assessment of this study?

- A. The change in purchase size is not practically important, and the good p-value of the second study is probably a result of the large study size.
- B. The change in purchase size is small, but may aggregate up to a large increase in profits over the entire customer base.
- C. The difference in the change in purchase size between the two studies is troubling; The team should run another, larger study.
- D. The p-value of the second study shows a statistically significant change in purchase size. The new website is an improvement.

Correct Answer: A

QUESTION 13

In which lifecycle stage are appropriate analytical techniques determined?

- A. Model planning
- B. Model building
- C. Data preparation
- D. Discovery

Correct Answer: A

QUESTION 14

You do a Student's t-test to compare the average test scores of sample groups from populations A and B. Group A averaged 10 points higher than group B. You find that this difference is significant, with a p-value of 0.03. What does that mean?

- A. There is a 3% chance that you have identified a difference between the populations when in reality there is none.
- B. The difference in scores between a sample from population A and a sample from population B will tend to be within 3% of 10 points.

C. There is a 3% chance that a sample group from population A will score 10 points higher than a sample group from population B.

D. There is a 97% chance that a sample group from population A will score 10 points higher than a sample group from population B.

Correct Answer: A

QUESTION 15

Refer to the exhibit.

		<u>True Class</u>	
		p	n
<u>Prediction</u>	P	262	15
	N	26	347

You have scored your Naive Bayesian classifier model on a hold out test data for cross validation and determined the way the samples scored and tabulated them as shown in the exhibit.

What are the False Positive Rate (FPR) and the False Negative Rate (FNR) of the model?

A. $FPR = 15/262$ $FNR = 26/288$

B. $FPR = 26/288$ $FNR = 15/262$

C. FPR = 262/15 FNR = 288/26

D. FPR = 288/26 FNR = 262/15

Correct Answer: A

[Latest E20-007 Dumps](#)

[E20-007 Practice Test](#)

[E20-007 Study Guide](#)