



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.passapply.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

You are testing two new weight-gain formulas for puppies. The test gives the results: Control group: 1% weight gain
Formula A. 3% weight gain

Formula B. 4% weight gain A one-way ANOVA returns a p-value = 0.027 What can you conclude?

- A. Either Formula A or Formula B is effective at promoting weight gain.
- B. Formula B is more effective at promoting weight gain than Formula A.
- C. Formula A and Formula B are both effective at promoting weight gain.
- D. Formula A and Formula B are about equally effective at promoting weight gain.

Correct Answer: A

QUESTION 2

Which word or phrase completes the statement? Mahout is to Hadoop as MADlib is to _____ .

- A. PostgreSQL
- B. R
- C. Excel
- D. SAS

Correct Answer: A

QUESTION 3

What describes a true limitation of a Logistic Regression method?

- A. Does not handle missing values well
- B. Does not handle redundant variables well
- C. Does not handle correlated variables well
- D. Does not have explanatory values

Correct Answer: A

QUESTION 4

What are the characteristics of Big Data?



- A. Data volume, processing complexity, and data structure variety.
- B. Data volume, business importance, and data structure variety.
- C. Data type, processing complexity, and data structure variety.
- D. Data volume, processing complexity, and business importance.

Correct Answer: A

QUESTION 5

In a Student's t-test, what is the meaning of the p-value?

- A. Area under the appropriate tails of the student's distribution
- B. Power of the Student's t-test
- C. Mean of the distribution for the null hypothesis
- D. Mean of the distribution for the alternate hypothesis

Correct Answer: A

QUESTION 6

In the MapReduce framework, what is the purpose of the Reduce function?

- A. It aggregates the results of the Map function and generates processed output
- B. It distributes the input to multiple nodes for processing
- C. It writes the output of the Map function to storage
- D. It breaks the input into smaller components and distributes to other nodes in the cluster

Correct Answer: A

QUESTION 7

Which SQL OLAP extension provides all possible grouping combinations?

- A. CUBE
- B. ROLLUP
- C. UNION ALL
- D. CROSS JOIN

Correct Answer: A



QUESTION 8

Your company has 3 different sales teams. Each team's sales manager has developed incentive offers to increase the size of each sales transaction. Any sales manager whose incentive program can be shown to increase the size of the average sales transaction will receive a bonus.

Data are available for the number and average sale amount for transactions offering one of the incentives as well as transactions offering no incentive.

The VP of Sales has asked you to determine analytically if any of the incentive programs has resulted in a demonstrable increase in the average sale amount. Which analytical technique would be appropriate in this situation?

- A. One-way ANOVA
- B. Multi-way ANOVA
- C. Student's t-test
- D. Wilcoxon Rank Sum Test

Correct Answer: A

QUESTION 9

Which analytical method is considered unsupervised?

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

Correct Answer: A

QUESTION 10

Which word or phrase completes the statement; "A data scientist would consider a RDBMS is to a table as R is to a _____."?

- A. Data frame
- B. List
- C. Matrix
- D. Array

Correct Answer: A



QUESTION 11

What is a core deliverable at the end of the analytic project?

- A. An implemented database design
- B. A whitepaper describing the project and the implementation
- C. A presentation for project sponsors
- D. The training materials

Correct Answer: C

QUESTION 12

The Marketing department of your company wishes to track opinion on a new product that was recently introduced. Marketing would like to know how many positive and negative reviews are appearing over a given period and potentially retrieve each review for more in-depth insight. They have identified several popular product review blogs that historically have published thousands of user reviews of your company's products.

You have been asked to provide the desired analysis. You examine the RSS feeds for each blog and determine which fields are relevant. You then craft a regular expression to match your new product's name and extract the relevant text from each matching review.

What is the next step you should take?

- A. Convert the extracted text into a suitable document representation and index into a review corpus
- B. Use the extracted text and your regular expression to perform a sentiment analysis based on mentions of the new product
- C. Read the extracted text for each review and manually tabulate the results
- D. Group the reviews using Naïve Bayesian classification

Correct Answer: A

QUESTION 13

What is holdout data?

- A. a subset of the provided data set selected at random and used to validate the model
- B. a subset of the provided data set selected at random and used to initially construct the model
- C. a subset of the provided data set that is removed by the data scientist because it contains data errors
- D. a subset of the provided data set that is removed by the data scientist because it contains outliers

Correct Answer: A



QUESTION 14

In which lifecycle stage are appropriate analytical techniques determined?

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

Correct Answer: A

QUESTION 15

If your intention is to show trends over time, which chart type is the most appropriate way to depict the data?

- A. Line chart
- B. Bar chart
- C. Stacked bar chart
- D. Histogram

Correct Answer: A

[E20-007 VCE Dumps](#)

[E20-007 Practice Test](#)

[E20-007 Braindumps](#)