



# E20-007<sup>Q&As</sup>

Data Science and Big Data Analytics

## Pass EMC E20-007 Exam with 100% Guarantee

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

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

100% Passing Guarantee  
100% Money Back Assurance

Following Questions and Answers are all new published by EMC  
Official Exam Center

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





### QUESTION 1

A data scientist is given an R data frame (i.e., empdata) with the following columns: Age Salary Occupation Education Gender. The scientist wants to examine only the Salary and Occupation columns for ages greater than 40. Which

command extracts the appropriate rows and columns from the data frame?

- A. `empdata[empdata$Age > 40, c("Salary","Occupation")]`
- B. `empdata[c("Salary","Occupation"), empdata$Age > 40]`
- C. `empdata[Age > 40, ("Salary","Occupation")]`
- D. `empdata[, c("Salary","Occupation")]$Age > 40`

Correct Answer: A

---

### QUESTION 2

You have fit a decision tree classifier using 12 input variables. The resulting tree used 7 of the 12 variables, and is 5 levels deep. Some of the nodes contain only 3 data points. The AUC of the model is

0.85. What is your evaluation of this model?

- A. The tree is probably overfit. Try fitting shallower trees and using an ensemble method.
- B. The AUC is high, and the small nodes are all very pure. This is an accurate model.
- C. The tree did not split on all the input variables. You need a larger data set to get a more accurate model.
- D. The AUC is high, so the overall model is accurate. It is not well-calibrated, because the small nodes will give poor estimates of probability.

Correct Answer: A

---

### QUESTION 3

How does Pig's use of a schema differ from that of a traditional RDBMS?

- A. Pig's schema is optional
- B. Pig's schema requires that the data is physically present when the schema is defined
- C. Pig's schema is required for ETL
- D. Pig's schema supports a single data type

Correct Answer: A

---



#### QUESTION 4

What is the reason for using LOESS?

- A. Fits a smoothed curve to scatterplot data; providing a general idea of the data's behavior
- B. Significance test for the correlation between two variables
- C. Plots a continuous variable versus a discrete variable; comparing distributions across classes
- D. Runs after a one-way ANOVA; determining which population has the highest mean value

Correct Answer: A

---

#### QUESTION 5

Which word or phrase completes the statement? Structured data is to OLAP data as quasi- structured data is to\_\_\_\_\_

- A. Clickstream data
- B. XML data
- C. Text documents
- D. Image files

Correct Answer: A

[Latest E20-007 Dumps](#)

[E20-007 PDF Dumps](#)

[E20-007 Exam Questions](#)