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.leads4pass.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

What would be considered "Big Data"?

A. An OLAP Cube containing customer demographic information about 100, 000, 000 customers

B. Daily Log files from a web server that receives 100, 000 hits per minute

C. Aggregated statistical data stored in a relational database table

D. Spreadsheets containing monthly sales data for a Global 100 corporation

Correct Answer: B

QUESTION 2

If R factors are categorical variables, which data classification level are they most closely related?

A. Nominal

B. Ordinal

C. Interval

D. Ratio

Correct Answer: A

QUESTION 3

What describes a true property of Logistic Regression method?

A. It is robust with redundant variables and correlated variables.

B. It handles missing values well.

C. It works well with discrete variables that have many distinct values.

D. It works well with variables that affect the outcome in a discontinuous way.

Correct Answer: A

QUESTION 4

In which lifecycle stage are appropriate analytical techniques determined?

A. Model planning

B. Model building

C. Data preparation

D. Discovery

Correct Answer: A

QUESTION 5

A Data Scientist is assigned to build a model from a reporting data warehouse. The warehouse contains data collected from many sources and transformed through a complex, multi-stage ETL process. What is a concern the data scientist should have about the data?

A. It is too processed

B. It is not structured

C. It is not normalized

D. It is too centralized

Correct Answer: A

QUESTION 6

On analyzing your time series data you suspect that the data represented as y1, y2, y3, ... , yn-1, yn 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. (y3-y2) ?(y2-y1) = ...= (yn-yn-1)-(yn-1-yn-2)

B. (y2-y1) = (y3-y2) = ... = (yn-yn-1)

C. ((y2-y1) /y1) * 100% = ...((yn-yn-1)/yn-1) * 100%

D. (y4-y2) ?(y3-y1) = ...= (yn-yn-2)-(yn-1-yn-3)

Correct Answer: A

QUESTION 7

You have completed your model and are handing it off to be deployed in production. What should you deliver to the production team, along with your commented code?

A. The production team needs to understand how your model will interact with the processes they already support. Give them documentation on expected model inputs and outputs, and guidance on error-handling.

B. The production team are technical, and they need to understand how the processes that they support work, so give them the same presentation that you prepared for the analysts.

C. The production team supports the processes that run the organization, and they need context to understand how

your model interacts with the processes they already support. Give them the same presentation that you prepared for the project sponsor.

D. The production team supports the processes that run the organization, and they need context to understand how your model interacts with the processes they already support. Give them the executive summary.

Correct Answer: A

QUESTION 8

You are provided four different datasets. Initial analysis on these datasets show that they have identical mean, variance and correlation values. What should your next step in the analysis be?

A. Visualize the data to further explore the characteristics of each data set

B. Select one of the four datasets and begin planning and building a model

C. Combine the data from all four of the datasets and begin planning and bulding a model

D. Recalculate the descriptive statistics since they are unlikely to be identical for each dataset

Correct Answer: A

QUESTION 9

What describes a true property of a Logistic Regression method?

A. Robust with redundant variables and correlated variables

B. Handles missing values well

C. Works well with discrete variables that have many distinct values

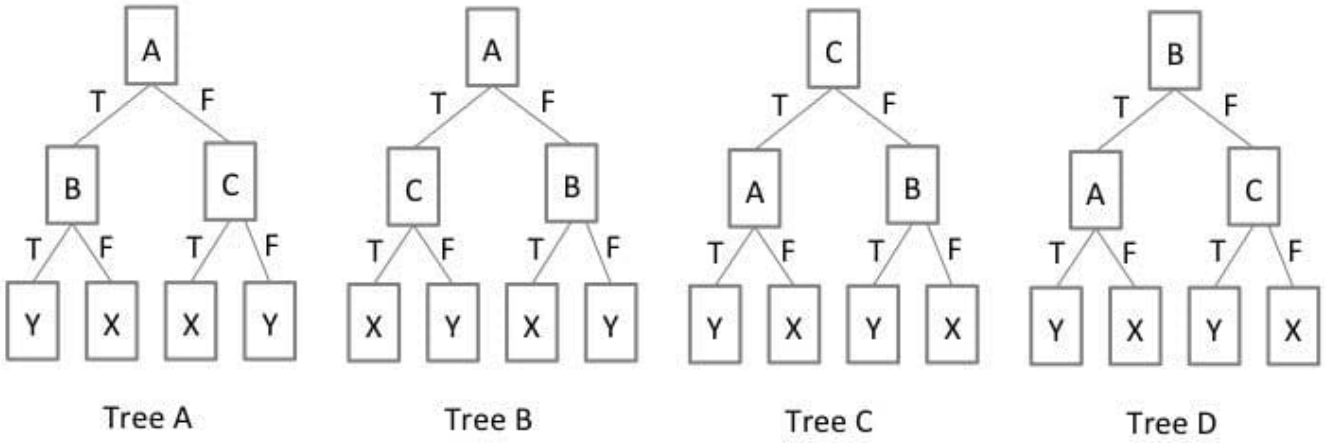
D. Works well with variables that affect the outcome in a discontinuous way

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

Consider the following itemsets: (hat, scarf, coat)

(hat, scarf, coat, gloves)

(hat, scarf, gloves)

(hat, gloves)

(scarf, coat, gloves)

What is the confidence of the rule (hat, scarf) => gloves?

A. 40%

B. 50%

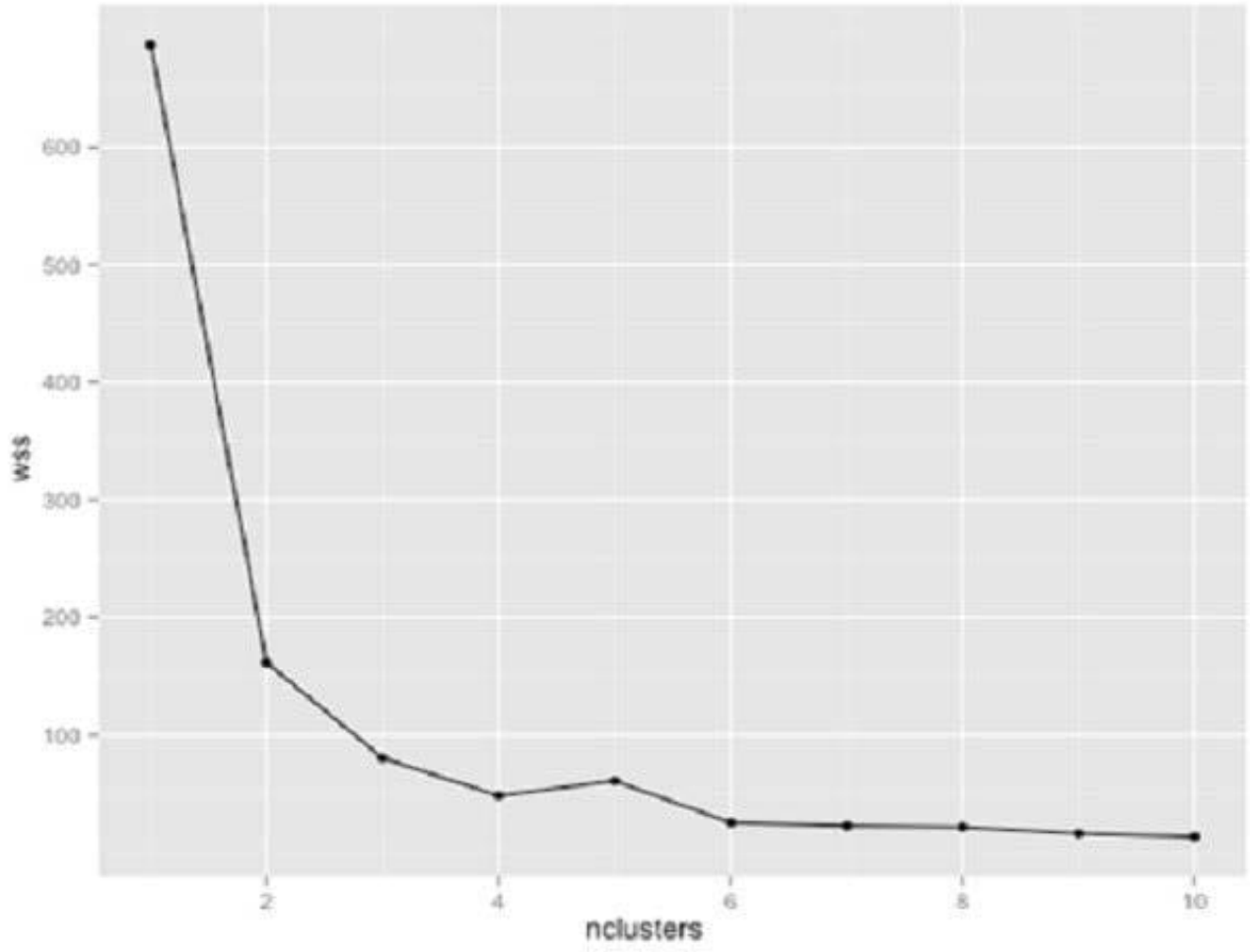
C. 60%

D. 66%

Correct Answer: D

QUESTION 12

Refer to the exhibit.



You are using k-means clustering to discover groupings within a data set. You plot within- sum-of-squares (wss) of multiple cluster sizes. Based on the exhibit, how many clusters should you use in your analysis?

A. 4

B. 2

C. 8

D. 10

Correct Answer: A

QUESTION 13

A data scientist is given an R data frame, "empdata", with the columns Age, Salary, Occupation, Education, and Gender. The data scientist would like 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 14

Which SQL OLAP extension provides all possible grouping combinations?

A. CUBE

B. ROLLUP

C. UNION ALL

D. CROSS JOIN

Correct Answer: A

QUESTION 15

What is LOESS used for?

A. It fits a smoothed curve to scatterplot data, to give a general sense of the data\\'s behavior.

B. It is a significance test for the correlation between two variables.

C. It plots a continuous variable versus a discrete variable, to compare distributions across classes.

D. It is run after a one-way ANOVA, to determine which population has the highest mean value.

Correct Answer: A

E20-007 PDF Dumps E20-007 VCE Dumps E20-007 Study Guide