# E20-026<sup>Q&As</sup>

Enterprise Storage Networking Specialist Exam

### Pass EMC E20-026 Exam with 100% Guarantee

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

https://www.leads4pass.com/e20-026.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

Leads4Pass

800,000+ Satisfied Customers



# Leads4Pass

#### **QUESTION 1**

You have run the association rules algorithm on your data set, and the two rules {banana, apple} => {grape} and {apple, orange}=> {grape} have been found to be relevant. What else must be true?

- A. {grape,apple,orange} must be a frequent itemset.
- B. {banana,apple,grape,orange} must be a frequent itemset.
- C. {grape} => {banana,apple} must be a relevant rule.
- D. {banana,apple} => {orange} must be a relevant rule.

Correct Answer: A

#### **QUESTION 2**

When creating a project sponsor presentation, what is the main objective?

- A. Show that you met the project goals
- B. Show how you met the project goals
- C. Show how well the model will meet the SLA (service level agreement)
- D. Clearly describe the methods and techniques used

Correct Answer: A

#### **QUESTION 3**

Consider these 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. 66%

- B. 40%
- C. 50%
- D. 60%

## Leads4Pass

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**

What is an example of a null hypothesis?

A. that a newly created model does not provide better predictions than the currently existing model

B. that a newly created model provides a prediction of a null sample mean

C. that a newly created model provides a prediction of a null population mean

D. that a newly created model provides a prediction that will be well fit to the null distribution

Correct Answer: A

#### **QUESTION 6**

Consider a scale that has five (5) values that range from "not important" to "very important". Which data classification best describes this data?

- A. Ordinal
- B. Nominal
- C. Real
- D. Ratio

Correct Answer: A

#### **QUESTION 7**

Which word or phrase completes the statement? Data-ink ratio is to data visualization as \_\_\_\_\_

- A. Confusion matrix is to classifier
- B. Data scientist is to big data
- C. Seasonality is to ARIMA
- D. K-means is to Naive Bayes

Correct Answer: A

#### **QUESTION 8**

You have been assigned to do a study of the daily revenue effect of a pricing model of online transactions. All the data currently available to you has been loaded into your analytics database; revenue data, pricing data, and online transaction data. You find that all the data comes in different levels of granularity. The transaction data has timestamps (day, hour, minutes, seconds), pricing is stored at the daily level, and revenue data is only reported monthly. What is your next step?

A. Report back to the business owner that the current data model does not support the business question.

B. Interpolate a daily model for revenue from the monthly revenue data.

C. Aggregate all data to the monthly level in order to create a monthly revenue model.

D. Disregard revenue as a driver in the pricing model, and create a daily model based on pricing and transactions only.

Correct Answer: A

#### **QUESTION 9**

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 10**

Refer to the exhibit.

Click on the calculator icon in the upper left corner. You are going into a meeting where you know your manager will have a question on your dataset -- specifically relating to customers that are classified as renters with good credit status.

In order to prepare for the meeting, you create a rule: RENTER => GOOD CREDIT. What is the

confidence of the rule?

	FREE HOUSING	HOME OWNER	RENTER	TOTAL
BAD CREDIT	54	476	270	800
GOOD CREDIT	75	1245	460	1780
TOTAL	129	1721	730	2580

A. 63%

B. 41%

C. 18%

D. 73%

Correct Answer: A

#### **QUESTION 11**

Refer to the Exhibit.

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?

С	В	А	CLASS
т	Т	Т	х
т	Т	F	Y
T	F	т	x
F	F	F	Y
F	Т	Т	x
F	F	т	Y



- A. Tree B
- B. Tree A
- C. Tree C
- D. Tree D
- Correct Answer: A

#### **QUESTION 12**

Which activity might be performed in the Operationalize phase of the Data Analytics Lifecycle?

- A. Run a pilot
- B. Try different analytical techniques
- C. Try different variables
- D. Transform existing variables

### Leads4Pass

Correct Answer: A

#### **QUESTION 13**

Your colleague, who is new to Hadoop, approaches you with a question. They want to know how best to access their data. This colleague has previously worked extensively with SQL and databases. Which query interface would you recommend?

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

Correct Answer: A

#### **QUESTION 14**

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 15**

Refer to the exhibit.

Click on the calculator icon in the upper left corner. You are given a list of pre-defined association rules:

	FREE HOUSING	HOME OWNER	RENTER	TOTAL
BAD CREDIT	54	476	270	800
GOOD CREDIT	75	1245	460	1780
TOTAL	129	1721	730	2580

For your next analysis, you must limit your dataset based on rules with confidence greater than 60%. Which of the rules will be kept in the analysis?

#### A. RENTER => BAD CREDIT

#### B. RENTER => GOOD CREDIT



- C. HOME OWNER => BAD CREDIT
- D. HOME OWNER => GOOD CREDIT
- E. FREE HOUSING => BAD CREDIT
- F. FREE HOUSING => GOOD CREDIT
- A. Rules B and D
- B. Rules A and F
- C. Rules C and E
- D. Rules D and E
- Correct Answer: A

Latest E20-026 Dumps

E20-026 PDF Dumps

E20-026 Exam Questions