



Vendor: EMC

Exam Code: E20-007

Exam Name: Data Science for Big Data Analytics

Version: DEMO

QUESTION 1

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

Answer: A

QUESTION 2

You are analyzing a time series and want to determine its stationarity. You also want to determine the order of autoregressive models.

How are the autocorrelation functions used?

- A. ACF as an indication of stationarity, and PACF for the correlation between X_t and X_{t-k} not explained by their mutual correlation with X_1 through X_{k-1} .
- B. PACF as an indication of stationarity, and ACF for the correlation between X_t and X_{t-k} not explained by their mutual correlation with X_1 through X_{k-1} .
- C. ACF as an indication of stationarity, and PACF to determine the correlation of X_1 through X_{k-1} .
- D. PACF as an indication of stationarity, and ACF to determine the correlation of X_1 through X_{k-1} .

Answer: A

QUESTION 3

Which word or phrase completes the statement? A spreadsheet is to a data island as a centralized database for reporting is to a _____?

- A. Data Warehouse
- B. Data Repository
- C. Analytic Sandbox
- D. Data Mart

Answer: A

QUESTION 4

What is one modeling or descriptive statistical function in MADlib that is typically not provided in a standard relational database?

- A. Linear regression
- B. Expected value
- C. Variance
- D. Quantiles

Answer: A

QUESTION 5

In which phase of the data analytics lifecycle do Data Scientists spend the most time in a project?

- A. Discovery
- B. Data Preparation
- C. Model Building
- D. Communicate Results

Answer: B

QUESTION 6

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.

Answer: A

QUESTION 7

Data visualization is used in the final presentation of an analytics project. For what else is this technique commonly used?

- A. Data exploration
- B. Descriptive statistics
- C. ETLT
- D. Model selection

Answer: A

QUESTION 8

Which functionality do regular expressions provide?

- A. text pattern matching
- B. underflow prevention
- C. increased numerical precision
- D. decreased processing complexity

Answer: A

QUESTION 9

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

Answer: A

QUESTION 10

Which word or phrase completes the statement? Business Intelligence is to monitoring trends as Data Science is to _____ trends.

- A. Predicting
- B. Discarding
- C. Driving
- D. Optimizing

Answer: A

QUESTION 11

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

Answer: A

QUESTION 12

Which key role for a successful analytic project can provide business domain expertise with a deep understanding of the data and key performance indicators?

- A. Business Intelligence Analyst
- B. Project Manager
- C. Project Sponsor
- D. Business User

Answer: A

QUESTION 13

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

Answer: A

QUESTION 14

What does R code `nv <- v[v < 1000]` do?

- A. Selects the values in vector `v` that are less than 1000 and assigns them to the vector `nv`
- B. Sets `nv` to TRUE or FALSE depending on whether all elements of vector `v` are less than 1000
- C. Removes elements of vector `v` less than 1000 and assigns the elements ≥ 1000 to `nv`
- D. Selects values of vector `v` less than 1000, modifies `v`, and makes a copy to `nv`

Answer: A

QUESTION 15

For which class of problem is MapReduce most suitable?

- A. Embarrassingly parallel
- B. Minimal result data
- C. Simple marginalization tasks
- D. Non-overlapping queries

Answer: A

QUESTION 16

In data visualization, which type of chart is recommended to represent frequency data?

- A. Line chart
- B. Histogram
- C. Q-Q chart
- D. Scatterplot

Answer: B

QUESTION 17

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

Answer: A

Thank You for Trying Our Product

Lead2pass Certification Exam Features:

- ★ More than **99,900** Satisfied Customers Worldwide.
- ★ Average **99.9%** Success Rate.
- ★ **Free Update** to match latest and real exam scenarios.
- ★ **Instant Download** Access! No Setup required.
- ★ Questions & Answers are downloadable in **PDF** format and **VCE** test engine format.
- ★ Multi-Platform capabilities - **Windows, Laptop, Mac, Android, iPhone, iPod, iPad.**
- ★ **100%** Guaranteed Success or **100%** Money Back Guarantee.
- ★ **Fast**, helpful support **24x7**.



View list of all certification exams: <http://www.lead2pass.com/all-products.html>



Microsoft



ORACLE



CITRIX



JUNIPER
NETWORKS



EMC²
where information lives®

10% Discount Coupon Code: ASTR14