Associate - Data Science and Big Data Analytics V2 Exam

1. In the Map Reduce framework, what is the purpose of the Reduce function?

A. It writes the output of the Map function to storage

B. It aggregates the results of the Map function and generates processed output

C. It distributes the input to multiple nodes for processing

D. It breaks the input into smaller components and distributes to other nodes in the cluster

Answer(s): B

---

2. When creating a presentation for a technical audience, what is the main objective?

A. Show that you met the project goals

B. Show if the model will meet the SLA

C. Show the technique to be used in the production environment

D. Show how you met the project goals

Answer(s): D

---

3. Consider a database with 4 transactions:

A. {bread, milk} => {cheese}

B. {juice} => {soda}

C. {bread} => {milk}

D. {bread} => {cheese}

Answer(s): A

---

4. In linear regression modeling, which action can be taken to improve the linearity of the relationship between the dependent and independent variables?

A. Use a different statistical package

B. Calculate the R-Squared value

C. Apply a transformation to a variable

D. Change the units of measurement on the independent variable

Answer(s): C

---

5. Consider these itemsets:

A. 40%

B. 66%

C. 50%

D. 60%

Answer(s): B

---

6. Which word or phrase completes the statement; "A theater actor is to 'artistic and expressive' as a data scientist is to."?

A. Communicative and collaborative

B. Introverted and technical

C. Independent and intelligent

D. Logical and steadfast

Answer(s): A

---

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

A. Matrix

B. Data frame

C. Array

D. List

Answer(s): B

---

8. Which word or phrase completes the statement? Business Intelligence is to ad-hoc reporting and dashboards as Data Science is to __________.

A. Sales and profit reporting

B. Structured Data and Data Sources

C. Optimization and Predictive Modeling

D. Alerts and Queries

Answer(s): C

---

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

A. SAS

B. Excel

C. PostgreSQL

D. R

Answer(s): C

---

10. In association rules, given X -> Y, what is confidence?

A. How many times more often X and Y occur together than expected if they were statistically independent, expressed as a ratio

B. Percentage of transactions that contain the itemset

C. Difference in the probability of X and Y appearing together compared with expectations if they were statistically independent

D. Percentage of transactions with X that also contain Y

Answer(s): D

---

11. Data has been collected on visitors' viewing habits at a bank's website. Which technique is used to identify pages commonly viewed during the same visit to the website?

A. Regression

B. Classification

C. Clustering

D. Association Rules

Answer(s): D

---

12. Consider the example of an analysis for fraud detection on credit card usage. You will need to ensure higher-risk transactions that may indicate fraudulent credit card activity are retained in your data for analysis, and not dropped as outliers during pre-processing.

A. ELT

B. OLTP

C. EDW

D. ETL

Answer(s): A

---

13. You received 100,000 home loan records and want to quickly determine if there is any correlation between mortgage age and mortgage amount before conducting advanced analysis.

A. Box and Whisker plot

B. Scatter plot

C. Stacked Bar chart

D. Histogram

Answer(s): B

---

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

A. Sets nv to TRUE or FALSE depending on whether all elements of vector v are less than 1000

B. Selects the values in vector v that are less than 1000 and assigns them to the vector nv

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(s): B

---

15. During a study to understand the population growth of a certain bacterial culture, you plot the data and identify a quadratic growth trend over time. Which transformation should you apply to linearize the data?

A. Square root

B. Add a constant

C. Square

D. Cube

Answer(s): A

---

16. What is an appropriate data visualization to use in a presentation for an analyst audience?

A. Stacked bar chart

B. Pie chart

C. ROC curve

D. Area chart

Answer(s): C

---

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

A. Ratio

B. Nominal

C. Real

D. Ordinal

Answer(s): D

---

18. What are considerations in a data science and Big Data analytics project?

A. Analysis flexibility and decision making

B. Building data silos and bypassing data privacy rules

C. Applying the latest technologies to demonstrate technical skills

D. Ignoring executive stakeholders and business users

Answer(s): A

---

19. A business colleague who is new to Hadoop approaches you with a question. The colleague wants to know the best approach to access their dat a. The colleague has previously worked extensively with SQL and databases.

A. Pig

B. HBase

C. Hive

D. Howl

Answer(s): C

---

20. Refer to the exhibit, which shows pairwise counts for items purchased together.

A. Eggs -> Milk

B. Milk -> Eggs

C. Milk -> Bread

D. Bread -> Milk

Answer(s): A