Vendor: CompTIA

Exam Code: DY0-001

Exam Name: CompTIA DataAI Certification Exam

Version: DEMO

QUESTION 1
A statistician notices gaps in data associated with age-related illnesses and wants to further aggregate these observations. Which of the following is the best technique to achieve this goal?

A. Label encoding
B. Linearization
C. Binning
D. Imputing

Answer: C
Explanation:
Binning groups continuous age values into discrete intervals (e.g., age ranges), filling gaps by aggregating observations into broader categories. This directly addresses uneven or sparse age data by creating consistent age groups.

QUESTION 2
A data scientist needs to analyze a company's chemical businesses and is using the master database of the conglomerate company. Nothing in the data differentiates the data observations for the different businesses. Which of the following is the most efficient way to identify the chemical businesses' observations?

A. Ingest the data from all of the hard drives and perform exploratory data analysis to identify which business is responsible for chemical operations.
B. Perform analysis on all of the data and create a summary report on the results relevant to chemical operations.
C. Consult with the business team to identify which sites are responsible for chemical operations and ingest only the relevant data for analysis.
D. Ingest data from the hard drive containing the most data and present sample results on the chemical operations.

Answer: C
Explanation:
Engaging the business team leverages domain expertise to pinpoint which records pertain to chemical operations, allowing you to extract and analyze just the relevant subset. This avoids the time and resource waste of ingesting and sifting through unrelated data.

QUESTION 3
A data scientist built several models that perform about the same but vary in the number of features. Which of the following models should the data scientist recommend for production according to Occam's razor?

A. The model with the fewest features and highest performance
B. The model with the fewest features and the lowest performance
C. The model with the most features and the lowest performance
D. The model with the most features and the highest performance

Answer: A
Explanation:
According to Occam's razor, when models perform equivalently, you choose the simplest one - in this case, the model that achieves the needed performance with the fewest features.

QUESTION 4
A data analyst wants to use compression on an analyzed data set and send it to a new destination for further processing. Which of the following issues will most likely occur?

A. Library dependency will be missing.
B. Server CPU usage will be too high.
C. Operating system support will be missing.
D. Server memory usage will be too high.

Answer: B
Explanation:
Compression and decompression are CPU-intensive operations; on large data sets, the extra processing load can significantly spike CPU utilization. Memory, OS support, or library dependencies are far less likely to be the primary bottleneck in a standard compression workflow.

QUESTION 5
The most likely concern with a one-feature, machine-learning model is high error due to:

A. bias
B. dimensionality.
C. variance.
D. probability.

Answer: A
Explanation:
A model with only one feature is unlikely to capture the true complexity of the data's underlying relationships, leading to systematic underfitting - i.e., high bias.

QUESTION 6
Given matrix

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \\ 3 & 2 & 1 \end{bmatrix}$$

Which of the following is A^T?

A. $\begin{bmatrix} 3 & 2 & 1 \\ 2 & 1 & 3 \\ 1 & 2 & 3 \end{bmatrix}$

B. $\begin{bmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \\ 3 & 2 & 1 \end{bmatrix}$

C. $\begin{bmatrix} 1 & 2 & 3 \\ 2 & 1 & 2 \\ 3 & 3 & 1 \end{bmatrix}$

D. $\begin{bmatrix} 3 & 3 & 1 \\ 2 & 1 & 2 \\ 1 & 2 & 3 \end{bmatrix}$

Answer: C
Explanation:
Transposing swaps rows and columns, so the (i, j) entry becomes the (j, i) entry.

QUESTION 7
A data scientist is clustering a data set but does not want to specify the number of clusters present. Which of the following algorithms should the data scientist use?

A. DBSCAN
B. k-nearest neighbors
C. k-means
D. Logistic regression

Answer: A
Explanation:
DBSCAN discovers clusters based on density without requiring you to predefine the number of clusters, automatically finding arbitrarily shaped groups and identifying noise points.

QUESTION 8
A data analyst wants to find the latitude and longitude of a mailing address. Which of the following is the best method to use?

A. One-hot encoding
B. Binning
C. Geocoding
D. Imputing

Answer: C
Explanation:
Geocoding is the process of converting a postal address into geographic coordinates (latitude and longitude), making it the appropriate method.

QUESTION 9
Which of the following describes the appropriate use case for PCA?

A. Dimensionality reduction
B. Classification
C. Regression
D. Recommendation

Answer: A
Explanation:
Principal Component Analysis transforms correlated features into a smaller set of uncorrelated components that capture most of the variance, making it ideal for reducing dimensionality before modeling or visualization.

QUESTION 10
A data scientist observes findings that indicate that as electrical grids in a country become more and more connected over time, the frequency of brownouts and blackouts in total decrease, and the frequency of major brownouts and blackouts increase. Which of the following distribution

metrics could best be identified?

A. Scale axis magnitudes
B. Kurtosis
C. Skewness
D. Normality

Answer: B
Explanation:
Kurtosis quantifies how heavy or light the tails of a distribution are. In this case, fewer overall events but more extreme (major) brownouts/blackouts indicates heavier tails over time. This is exactly what an increasing kurtosis would reveal.

QUESTION 11
A data scientist is merging two tables. Table 1 contains employee IDs and roles. Table 2 contains employee IDs and team assignments. Which of the following is the best technique to combine these data sets?

A. inner join between Table 1 and Table 2
B. left join on Table 1 with Table 2
C. right join on Table 1 with Table 2
D. outer join between Table 1 and Table 2

Answer: A
Explanation:
An INNER JOIN merges records only where the employee ID exists in both tables, yielding a single combined table of each employee's role paired with their team assignment.

QUESTION 12
Which of the following is a classic example of a constrained optimization problem?

A. The cold start problem
B. The traveling salesman
C. Calculating local maximum
D. Calculating gradient descent

Answer: B
Explanation:
The traveling-salesman problem seeks the shortest possible route that visits each city exactly once and returns to the start, making it a textbook example of optimization under explicit constraints.

QUESTION 13
A data scientist wants to digitize historical hard copies of documents. Which of the following is the best method for this task?

A. Word2vec
B. Optical character recognition
C. Latent semantic analysis
D. Semantic segmentation

Answer: B
Explanation:
OCR converts scanned images of text into machine-readable characters, making it the appropriate tool for digitizing printed or handwritten historical documents.

QUESTION 14
A data scientist trained a model for departments to share. The departments must access the model using HTTP requests. Which of the following approaches is appropriate?

A. Utilize distributed computing.
B. Deploy containers.
C. Create an endpoint.
D. Use the File Transfer Protocol.

Answer: C
Explanation:
Exposing the model behind an HTTP endpoint (for example, a REST API) allows other departments to send requests and receive predictions directly over HTTP. The other options don't inherently provide a request瑩esponse interface for sharing a model.

QUESTION 15
Given the following:

$$X_t = \delta + \phi_1 X_{t-1} + \omega_t \text{where} \quad \omega_t \sim N(0, \sigma_\omega^2)$$

Which of the following time series models **best** represents this process?

A. ARIMA(1,1,1)
B. ARMA(1,1)
C. SARIMA(1, 1, 1) x (1, 1, 1)1
D. AR(1)

Answer: D
Explanation:
The model has a single autoregressive term and only white-noise errors, matching the definition of an AR(1) process.

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

10% Discount Coupon Code: **ASTR14**