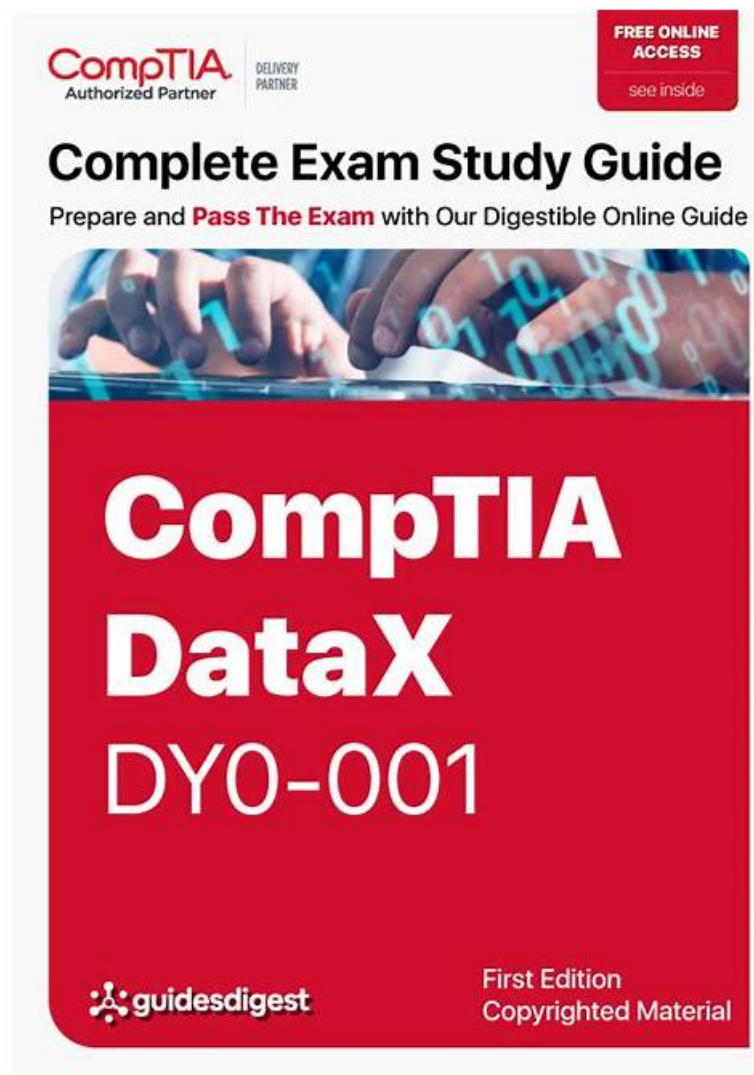


DY0-001덤프자료 - DY0-001시험대비덤프공부자료



BONUS!!! KoreaDumps DY0-001 시험 문제집 전체 버전을 무료로 다운로드하세요: <https://drive.google.com/open?id=18fkNUNUvY0rubjODRZoWN1TCYW2Ctutm>

CompTIA DY0-001 시험을 어떻게 통과할수 있을가 고민중이신 분들은KoreaDumps를 선택해 주세요. KoreaDumps는 많은 분들이 IT인증시험을 응시하여 성공하도록 도와주는 사이트입니다. 최고급 품질의CompTIA DY0-001시험대 비 덤프는CompTIA DY0-001시험을 간단하게 패스하도록 힘이 되어드립니다. KoreaDumps 의 덤프는 모두 엘리트 한 전문가들이 만들어낸 만큼 시험문제의 적중률은 아주 높습니다.

CompTIA DY0-001 시험요강:

주제	소개
주제 1	<ul style="list-style-type: none">Mathematics and Statistics: This section of the exam measures skills of a Data Scientist and covers the application of various statistical techniques used in data science, such as hypothesis testing, regression metrics, and probability functions. It also evaluates understanding of statistical distributions, types of data missingness, and probability models. Candidates are expected to understand essential linear algebra and calculus concepts relevant to data manipulation and analysis, as well as compare time-based models like ARIMA and longitudinal studies used for forecasting and causal inference.

주제 2	<ul style="list-style-type: none"> Modeling, Analysis, and Outcomes: This section of the exam measures skills of a Data Science Consultant and focuses on exploratory data analysis, feature identification, and visualization techniques to interpret object behavior and relationships. It explores data quality issues, data enrichment practices like feature engineering and transformation, and model design processes including iterations and performance assessments. Candidates are also evaluated on their ability to justify model selections through experiment outcomes and communicate insights effectively to diverse business audiences using appropriate visualization tools.
주제 3	<ul style="list-style-type: none"> Operations and Processes: This section of the exam measures skills of an AI ML Operations Specialist and evaluates understanding of data ingestion methods, pipeline orchestration, data cleaning, and version control in the data science workflow. Candidates are expected to understand infrastructure needs for various data types and formats, manage clean code practices, and follow documentation standards. The section also explores DevOps and MLOps concepts, including continuous deployment, model performance monitoring, and deployment across environments like cloud, containers, and edge systems.
주제 4	<ul style="list-style-type: none"> Machine Learning: This section of the exam measures skills of a Machine Learning Engineer and covers foundational ML concepts such as overfitting, feature selection, and ensemble models. It includes supervised learning algorithms, tree-based methods, and regression techniques. The domain introduces deep learning frameworks and architectures like CNNs, RNNs, and transformers, along with optimization methods. It also addresses unsupervised learning, dimensionality reduction, and clustering models, helping candidates understand the wide range of ML applications and techniques used in modern analytics.
주제 5	<ul style="list-style-type: none"> Specialized Applications of Data Science: This section of the exam measures skills of a Senior Data Analyst and introduces advanced topics like constrained optimization, reinforcement learning, and edge computing. It covers natural language processing fundamentals such as text tokenization, embeddings, sentiment analysis, and LLMs. Candidates also explore computer vision tasks like object detection and segmentation, and are assessed on their understanding of graph theory, anomaly detection, heuristics, and multimodal machine learning, showing how data science extends across multiple domains and applications.

>> DY0-001덤프자료 <<

DY0-001시험대비 덤프공부자료, DY0-001시험기출문제

CompTIA 인증DY0-001인증 시험공부자료는KoreaDumps에서 제공해드리는CompTIA 인증DY0-001덤프가 가장 좋은 선택입니다. KoreaDumps에서는 시험문제가 업데이트되면 덤프도 업데이트 진행하도록 최선을 다하여 업데이트 서비스를 제공해드려 고객님의소유하신 덤프가 시장에서 가장 최신버전덤프로 되도록 보장하여 시험을 맞이할 수 있게 도와드립니다.

최신 CompTIA Data+ DY0-001 무료샘플문제 (Q36-Q41):

질문 # 36

Which of the following is a key difference between KNN and k-means machine-learning techniques?

- A. KNN is used for classification, while k-means is used for clustering.
- B. KNN operates exclusively on continuous data, while k-means can work with both continuous and categorical data.
- C. KNN performs better with longitudinal data sets, while k-means performs better with survey data sets.
- D. KNN is used for finding centroids, while k-means is used for finding nearest neighbors.

정답: A

설명:

K-Nearest Neighbors (KNN) is a supervised machine learning algorithm used primarily for classification and regression. It labels a new instance by majority vote (or averaging, in regression) of its k-nearest labeled neighbors.

k-Means is an unsupervised learning algorithm used for clustering. It partitions unlabeled data into k groups based on feature similarity, using centroids.

Thus, the key difference is in their purpose:

* KNN # Classification (Supervised)

* K-Means # Clustering (Unsupervised)

Why the other options are incorrect:

* A: Both can technically operate on continuous or categorical data (with preprocessing).

* B: This is not a meaningful or standardized distinction.

* C: This reverses the actual roles. k-means finds centroids; KNN finds nearest neighbors.

Official References:

* CompTIA DataX (DY0-001) Official Study Guide - Section 4.1 (Classification vs. Clustering): "KNN is a supervised learning algorithm for classification tasks. K-means is an unsupervised clustering technique that groups data by proximity to centroids."

* Data Science Handbook, Chapter 5: "One key distinction: KNN uses labeled data to classify or regress; k-means uses unlabeled data to identify groupings."

-

질문 # 37

A data scientist is building a model to predict customer credit scores based on information collected from reporting agencies. The model needs to automatically adjust its parameters to adapt to recent changes in the information collected. Which of the following is the best model to use?

- A. Linear discriminant analysis
- B. Random forest
- C. XGBoost
- D. Decision tree

정답: C

설명:

XGBoost (Extreme Gradient Boosting) is a high-performance, scalable ensemble algorithm that builds decision trees in sequence and adjusts to errors iteratively. It also supports incremental training, making it adaptive to changing data patterns - ideal for dynamically updated credit information.

Why the other options are incorrect:

* A: Decision trees are static once trained and don't adapt unless retrained.

* B: Random forest is an ensemble of trees but lacks the adaptive boosting component.

* C: LDA is a linear classification technique - not suited for adapting to changing data distributions.

Official References:

* CompTIA DataX (DY0-001) Official Study Guide - Section 4.3: "XGBoost is highly efficient and supports iterative learning making it well-suited for data environments that evolve over time."

* Applied Machine Learning Guide, Chapter 8: "XGBoost adapts to changes by refining errors across iterations, providing robustness in dynamic systems."

-

질문 # 38

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. Perform analysis on all of the data and create a summary report on the results relevant to chemical operations.
- B. Ingest data from the hard drive containing the most data and present sample results on the 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 the data from all of the hard drives and perform exploratory data analysis to identify which business is responsible for chemical operations.

정답: C

설명:

The most efficient and practical approach is to consult the business stakeholders to understand which sites or data partitions relate to chemical operations. This avoids unnecessary processing of irrelevant data and aligns with the data science best practice of combining domain knowledge with technical methods.

Why the other options are incorrect:

* A: Ingesting all data without guidance is time- and resource-intensive.

- * B: Analyzing all data indiscriminately can dilute the focus on chemical business specifics.
- * D: Using the largest data set arbitrarily may not reflect chemical operations and lacks targeted relevance.

Official References:

- * CompTIA DataX (DY0-001) Official Study Guide - Section 5.1: "Collaboration with domain experts and stakeholders ensures the data scientist focuses on relevant sources and minimizes inefficiency in data preparation."
- * CRISP-DM Model - Business Understanding Phase: "Clarifying project objectives with business input is key to aligning data selection with analytical goals."

-

질문 # 39

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

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

정답: A

설명:

Principal Component Analysis (PCA) is an unsupervised technique used to reduce the dimensionality of large datasets by transforming correlated features into a smaller set of uncorrelated components (principal components) while retaining the most variance.

Why the other options are incorrect:

- * B: Classification is a predictive modeling task; PCA is not inherently predictive.
- * C: Regression models numerical relationships; PCA does not predict outcomes.
- * D: Recommendation systems use collaborative or content filtering, not PCA directly.

Official References:

- * CompTIA DataX (DY0-001) Study Guide - Section 3.3: "PCA is primarily used for reducing the number of variables while preserving data structure and minimizing information loss."
- * Pattern Recognition and Machine Learning, Chapter 12: "PCA identifies principal axes of variation and is widely used in preprocessing for dimensionality reduction."

-

질문 # 40

A data scientist needs to:

Build a predictive model that gives the likelihood that a car will get a flat tire.

Provide a data set of cars that had flat tires and cars that did not.

All the cars in the data set had sensors taking weekly measurements of tire pressure similar to the sensors that will be installed in the cars consumers drive.

Which of the following is the most immediate data concern?

- A. Insufficient domain expertise
- B. Multivariate outliers
- C. Granularity misalignment
- D. Lagged observations

정답: C

설명:

Granularity misalignment refers to a mismatch between the level of detail in the predictor variables and the event being predicted. In this case, flat tires are likely discrete, infrequent events, while tire pressure is measured weekly. If the prediction model is trying to link a specific tire pressure value to a binary outcome (flat tire: yes/no), and the timing doesn't align precisely, the predictor variable (pressure) may not be granular enough to accurately associate with the event.

Why the other options are incorrect:

- * B: While outliers can exist, they are not the most immediate concern given the time-series nature of the data.
- * C: While domain expertise is helpful, it doesn't directly address the data structure issue.
- * D: Lagged observations can be engineered in modeling but aren't the primary problem here.

Official References:

