
Exam : SDS

Title : Senior Data Scientist

Version : DEMO

1. Which of the following is used to summarize a dataset by showing the median, quantiles, and min/max values for each of the variables?

- A. Box Plots
- B. Pie Charts
- C. Histogram
- D. Scatter Chart
- E. Bar Charts

Answer: A

Explanation:

A Box Plot (also called Whisker Plot) is a visualization tool used to summarize data distribution using five-number summary:

Minimum, First quartile (Q1), Median (Q2), Third quartile (Q3), Maximum.

It also highlights outliers explicitly.

Option A (Box Plots): Correct.

Option B (Pie Charts): Show proportions, not distribution.

Option C (Histogram): Shows frequency distribution but not quartiles/median.

Option D (Scatter Chart): Used for relationships between two variables, not summary statistics.

Option E (Bar Charts): Compare categories, not statistical spread.

Thus, the correct answer is Option A (Box Plots).

Reference: DASCA Data Scientist Knowledge Framework (DSKF) – Data Visualization Tools: Box Plots and Statistical Summaries.

2. Image files can be broken down into two broad categories:

- i. Rasterized
- ii. Vectorized
- iii. Sectorized

- A. i, ii
- B. ii, iii
- C. i, iii
- D. None of the above

Answer: A

Explanation:

Images are broadly categorized based on how they store visual information:

Rasterized images (Option i):

Composed of a grid of pixels (bitmap).

Each pixel has color information.

Examples: JPEG, PNG, BMP.

Best for photos or complex visuals.

Vectorized images (Option ii):

Composed of paths defined by mathematical formulas.

Scalable without quality loss.

Examples: SVG, EPS, AI.

Best for logos, icons, and illustrations.

Sectorized images (Option iii):

Not a standard category in computer graphics.

Thus, image files are categorized into Rasterized and Vectorized, making Option A (i, ii) correct.

Reference: DASCA Data Scientist Knowledge Framework (DSKF) – Data Types & Multimedia Data Management.

3. Self-driving car is an example of:

- A. Supervised learning
- B. Unsupervised learning
- C. Reinforcement learning
- D. All of the above

Answer: C

Explanation:

Self-driving cars (autonomous vehicles) are an application of Reinforcement Learning (RL) in machine learning:

In RL, an agent (car) interacts with an environment (roads, obstacles, traffic) and learns to maximize rewards (e.g., safe driving, efficient navigation).

The system improves performance through trial-and-error learning, guided by reward signals such as staying in a lane or avoiding collisions.

Supervised learning (A): Used in some supporting tasks like image recognition (e.g., identifying stop signs), but not the core paradigm for self-driving.

Unsupervised learning (B): Useful for clustering sensor data, but again not the main paradigm.

Reinforcement learning (C): Correct, since self-driving fundamentally depends on RL decision-making. Thus, the correct answer is Option C (Reinforcement Learning).

Reference: DASCA Data Scientist Knowledge Framework (DSKF) – Machine Learning Paradigms: Reinforcement Learning and Autonomous Systems.

4. Exploratory analytic algorithms help the Data Science team to better:

- A. Understand the data content
- B. Gain a high-level understanding of relationships
- C. Understand patterns in the data
- D. Both A and B
- E. All of the above

Answer: E

Explanation:

Exploratory analytics (often referred to as Exploratory Data Analysis – EDA) is a fundamental step in data science, enabling practitioners to discover initial insights, detect anomalies, and understand the structure of datasets before applying predictive or prescriptive modeling.

Option A (Understand the data content): Correct. EDA techniques (descriptive statistics, summary tables, profiling) reveal missing values, data types, and distributions.

Option B (Gain a high-level understanding of relationships): Correct. Correlation analysis, scatter plots, and cross-tabulations help identify dependencies between variables.

Option C (Understand patterns in the data): Correct. Visualization and clustering methods help discover hidden structures, seasonalities, and outliers.

Since exploratory algorithms contribute to all of these objectives, the correct answer is Option E (All of

the above).

Reference: DASCA Data Scientist Knowledge Framework (DSKF) – Analytics and Machine Learning: Exploratory Analytics & EDA.

5. Which of the following is TRUE for Tensor?

- A. Tensor is an array of floating-point numbers
- B. In Tensor, there can be arbitrarily many dimensions to the array
- C. Tensor is used to describe multidimensional arrays of numbers on which we perform linear operations
- D. Both B and C
- E. All of the above

Answer: E

Explanation:

A Tensor is a fundamental data structure in modern machine learning frameworks (e.g., TensorFlow, PyTorch). It is best described as a generalization of vectors and matrices to potentially higher dimensions.

Option A: Correct. Tensors typically store numeric values (commonly floating-point numbers) in structured formats.

Option B: Correct. A tensor can have any number of dimensions (rank).

For example:

A scalar is a 0-D tensor.

A vector is a 1-D tensor.

A matrix is a 2-D tensor.

Higher-rank tensors can represent images, videos, or multidimensional datasets.

Option C: Correct. Tensors are explicitly designed to allow linear algebra operations, which are the foundation of deep learning computations (matrix multiplications, dot products, etc.).

Therefore, since all three statements are true, the correct answer is Option E (All of the above).

Reference: DASCA Data Scientist Knowledge Framework (DSKF) – Analytics and Machine Learning, Deep Learning Concepts; Official DASCA Study Guide.