DASCA Online Practice-Test / Test 1

Senior Data Scientist (SDS™)

Top of Form

**Q 1. Which of the following does the Big Data Business Model Maturity Index provide the roadmap for? Choose from the answers provided:**

A. How organizations can integrate data into their business models

B. How organizations can integrate analytics into their business models

C. Both A and B

D. How organizations can integrate Big Data Analytics tools into their business models

E. How organizations can integrate strategy into their business models

**Q 2. There is a particular phase that the Big Data Business Model Maturity Index does NOT focus on , with regard to external market opportunities. Identify the phase from the options below:**

A. Business Metamorphosis

B. Data Monetization

C. Business Optimization

D. All of the above

E. None of the above

**Q 3. Identify the option, from the following, which does NOT provide a process of a use case:**

A. Understand your organization's key business initiatives or business challenge

B. Identify your key business stakeholders

C. Capture the decisions that the key business stakeholders need to make in order to support the organization's key business initiatives

D. Brainstorm the outcomes that the key stakeholders need to answer to facilitate making the decisions

E. Brainstorm the questions that the key stakeholders need to answer to facilitate making the decisions

**Q 4. A sports team uses an analytical method to determine the relative value of a player vis-à-vis their success on the overall team. Which analysis does it use?**

A. Trend Analysis

B. Prescriptive Analysis

C. Descriptive Analysis

D. Predictive Analysis

E. Cohorts Analysis

**Q 5. One of the following options does NOT exemplify a graphical model. Identify which one:**

A. Road maps

B. Electrical circuits

C. Computer networks

D. Geographical networks

E. Flow charts

**Q 6. Which of the following does the ElementTree sub library provide direct access to?**

A. Parse tree of the XML

B. Delete tree of the XML

C. Copy tree of the XML

D. Insert tree of the XML

E. None of the above

**Q 7. Exploratory analytics algorithms is best exemplified by which of the following?**

A. Trend Analysis

B. Geographical Analytics

C. Box Plots

D. Pair plots

E. All

**Q 8. Market Basket Analysis is fundamentally based on:**

A. Trend Analysis

B. Box plot Analysis

C. Association Analysis

D. Scatter Plot Analysis

E. None of the above

**Q 9. The concept of clustering is very frequently applied to one of the areas mentioned below. Identify which one:**

A. Food and hygiene

B. Sound engineering

C. Image compression

D. Memory utilization

**Q 10. What is a tuple?**

A. A list that cannot be modified

B. A list that can be modified

C. Cannot say

D. Programmer dependent entity

DASCA Online Practice-Test / Test 2

Senior Data Scientist (SDS™)

Top of Form

**Q 1. "A supermarket store analyzes the CCTV footage of its customers movement to Improve "in store" marketing". What type of Big Data are we talking about?**

A. Structured Data

B. Unstructured Data

C. Semi-Structured Data

D. None of these

**Q 2. Which among the following analytics techniques has highest business value:**

A. Discovery

B. Predictive

C. Exploratory

D. Prescriptive

**Q 3. Which of the following actions does the microservices architecture require?**

A. Breaking a product or project into independent services

B. Breaking a product or project into dependent services

C. Both A and B

D. None of the above

**Q 4. Which of the following is true about Discovery Analytics:**

A. Uses slicing and Dicing to enable easy analysis

B. Scorecards and Dashboards are used in Discovery analytics to measure KPIs

C. Advance filtering, Pivots and graphics helps to profile customers and study patterns

D. In Discovery analytics samples are scrutinized to identify trends and samples

**Q 5. What is maximum likelihood estimation? Identify from the options below:**

A. A method of framing a large class of problems in Data Science

B. A method of framing a small class of problem in the Data Science

C. A method of framing a large class of problems in HDFS

D. A method of framing a small class of problem in the HDFS

E. Both A and C

**Q 6. One of the following is a Python library used largely for fitting Bayesian Networks to real data. Identify which one:**

A. SciLib

B. PyMC

C. MyLib

D. MCMC

E. SCIMC

**Q 7. Clustering, Segmentation, Decision trees are part of which analytics techniques:**

A. Exploratory

B. Descriptive

C. Prescriptive

D. Predictive

**Q 8. What is the job of Amazon EMR?**

A. Amzon EMR is a managed cluster for that simplifies running big data frameworks such as Apache Hadoop, Apache Spark on AWS

B. Handles the details of capacity provisioning, load balancing, scaling, and application health monitoring

C. Both A and B

D. None of the above

**Q 9. Which of the following is possible uses of behavioral analytics:**

A. Improving overall customer satisfaction

B. Improving conversion of an audience to customer

C. Driving innovation in business strategy

D. All of the above

**Q 10. Structured Data that contains tags or markup elements which help in organizing and analyzing data to certain extent are called \_\_\_\_\_\_\_\_\_\_\_**

A. Structured Data

B. Unstructured Data

C. Semi-Structured Data

D. None of these

DASCA Online Practice-Test / Test 3

Senior Data Scientist (SDS™)

Top of Form

**Q 1. Why do many companies have trouble governing their data lakes effectively?**

A. Data governance often isn't a priority with data lakes because they usually don't support mission-critical transaction processing applications.

B. Once data is put into a data lake, it's impossible to make any changes to it, rendering governance efforts ineffective.

C. Data lakes typically contain broad sets of data stored in different formats and multiple data repositories, which complicates governance efforts.

D. All of the above.

**Q 2. Which is true about the comparison between Data Lake and Data Warehouse?**

A. Data Lake stores structured, unstructured, semi-structured as well as raw data whereas Data Warehouse stores only structured data.

B. Data Warehouse stores structured, unstructured, semi-structured as well as raw data whereas Data Lake stores only structured data.

C. Both Data Lake and Data Warehouse store only structured data.

D. Both Data Lake and Data Warehouse store structured, unstructured, semi-structured, and raw data.

**Q 3. What is the difference between Business Intelligence and Data Science?**

A. Data Science focuses on descriptive analytics whereas Business Intelligence focuses on predictive and presriptive analytics.

B. Business Intelligence focuses on descriptive analytics whereas Data Science focuses on predictive and presriptive analytics.

C. Business Intelligence focuses on predictive analytics whereas Data Science focuses on presriptive analytics.

D. Business Intelligence focuses on predictive analytics whereas Data Science focuses on descriptive analytics.

E. Business Intelligence focuses on prescriptive analytics whereas Data Science focuses on predictive analytics.

**Q 4. You are the head of a major social media website. Which type of database would you choose to store data that changes structure frequently, doesn't require updates too often but requires quick read times.**

A. Relational Database

B. Wide Column Store

C. Key-Value Store

D. Document Store

**Q 5. What is the probability that a 5 digit number, the sum of whose individual digits is 41, is divisible by 11?**

A. "2/15"

B. "13/36"

C. "11/36"

D. "6/35"

E. "3/35"

**Q 6. What is the output of the following python program: def get\_values(): a = 7 b = a a = a + 1 print b, a x = [4, 5] y = x x.append(6) print x, y get\_values():**

A. 8 7[4, 5, 6] [4, 5, 6]

B. 7 8[4, 5, 6] [4, 5]

C. 8 8[4, 5, 6] [4, 5, 6]

D. 7 8[4, 5, 6] [4, 5, 6]

E. 8 8[4, 5, 6] [4, 5]

**Q 7. Following cases fall under which type of machine learning: Case 1: Data of blood samples of patients and results if they ever had or later developed cancer. Using this to data to assess cancer risk for other patients. Case 2: Drawing inferences by clustering to unlock hidden patterns**

A. Case 1: Unsupervised learning Case 2: Supervised learning

B. Case 1: Unsupervised learning Case 2: Unsupervised learning

C. Case 1: Supervised learning Case 2: Supervised learning

D. Case 1: Supervised learning Case 2: Unsupervised learning

**Q 8. Which of the following is an assumption about the data in Time Series Analysis?**

A. Data may or may not have systamatic patterns

B. Data does not have noise/ errors

C. Data consists of systamatic patterns and noise (errors)

D. Data consists of systamatic patterns but no noise

**Q 9. Which of the following is correct:**

A. Calling split() on a string generates a list.

B. Calling split() on a string generates a tuple.

C. Calling split() on a string generates a dictionary.

D. Calling split() on a string generates smaller strings.

**Q 10. If 2 squares are randomly chosen on a chessboard, what is the probability that they have a side in common?**

A. "1/18"

B. "1/9"

C. "1/24"

D. "63/64"

DASCA Online Practice-Test / Test 4

Senior Data Scientist (SDS™)

**Q 1. Which is true about the comparison between Data Lake and Data Warehouse?**

A. Data Lake stores structured, unstructured, semi-structured as well as raw data whereas Data Warehouse stores only structured data.

B. Data Warehouse stores structured, unstructured, semi-structured as well as raw data whereas Data Lake stores only structured data.

C. Both Data Lake and Data Warehouse store only structured data.

D. Both Data Lake and Data Warehouse store structured, unstructured, semi-structured, and raw data.

**Q 2. In least squares regression, which of the following is not a required assumption about the error term e?**

A. The expected value of the error term is one

B. The variance of the error term is the same for all values of x.

C. The values of the error term are independent.

D. The error term is normally distributed.

**Q 3. Where would dimensionality reduction make most sense?**

A. When processing audio, video or image files.

B. When processing heights of 1 billion people.

C. When processing grades of 1000 students in 4 subjects.

D. Both B and C.

**Q 4. The sentence "The quick brown fox jumped over the lazy dog" can be broken into how many 3-gram words?**

A. 9

B. 8

C. 7

D. 6

**Q 5. A coefficient of correlation is computed to be -0.95. This means:**

A. The relation between the variables is weak

B. The relation between the variables is strong and positive

C. The relation between the variables is strong and negative

D. The relation between the variables is weak and negative

E. Correlation coefficient cannot have this value

**Q 6. Which of the following is correct:**

A. Calling split() on a string generates a list.

B. Calling split() on a string generates a tuple.

C. Calling split() on a string generates a dictionary.

D. Calling split() on a string generates smaller strings.

**Q 7. What is Text Mining?**

A. Process of deriving usable information from text files.

B. Process of Creating structured data from unstructured data.

C. Process of creating semi-structured data from unstructured data.

D. Both A and B.

E. All of the above.

**Q 8. A regression analysis is inappropriate when**

A. you have two variables that are measured on an interval or ratio scale.

B. you want to make predictions for one variable based on information about another variable.

C. the pattern of data points form a reasonably straight line.

D. there is heteroscedasticity in the scatter plot.

**Q 9. In which case should you use NLP?**

A. Extract information from data that has a regular structure.

B. There is a lack of training data.

C. To check if two documents have similar content.

D. Both A and B.

**Q 10. In k fold cross validation there are**

A. k trained classifiers and k performance metrics

B. 1 trianed classifier and k performance metrics

C. k trianed classifiers and k + 1 performance metrics

D. k + 1 trianed classifiers and k performance metrics

E. 1 trianed classifier and 1 performance metric

DASCA Online Practice-Test / Test 5

Senior Data Scientist (SDS™)

**Q 1. Identify the option, from the following, which does NOT provide a process of a use case:**

A. Understand your organization's key business initiatives or business challenge

B. Identify your key business stakeholders

C. Capture the decisions that the key business stakeholders need to make in order to support the organization's key business initiatives

D. Brainstorm the outcomes that the key stakeholders need to answer to facilitate making the decisions

E. Brainstorm the questions that the key stakeholders need to answer to facilitate making the decisions

**Q 2. Which of the following is an assumption about the data in Time Series Analysis?**

A. Data may or may not have systamatic patterns

B. Data does not have noise/ errors

C. Data consists of systamatic patterns and noise (errors)

D. Data consists of systamatic patterns but no noise

**Q 3. Which of the following does the Big Data Business Model Maturity Index provide the roadmap for? Choose from the answers provided:**

A. How organizations can integrate data into their business models

B. How organizations can integrate analytics into their business models

C. Both A and B

D. How organizations can integrate Big Data Analytics tools into their business models

E. How organizations can integrate strategy into their business models

**Q 4. Exploratory analytics algorithms is best exemplified by which of the following?**

A. Trend Analysis

B. Geographical Analytics

C. Box Plots

D. Pair plots

E. All

**Q 5. Which of the following actions does the microservices architecture require?**

A. Breaking a product or project into independent services

B. Breaking a product or project into dependent services

C. Both A and B

D. None of the above

**Q 6. There is a particular phase that the Big Data Business Model Maturity Index does NOT focus on , with regard to external market opportunities. Identify the phase from the options below:**

A. Business Metamorphosis

B. Data Monetization

C. Business Optimization

D. All of the above

E. None of the above

**Q 7. Which of the following is an example of “amortized performance”?**

A. Hadoop dictionaries

B. Python dictionaries

C. HDFS dictionaries

D. MapReduce dictionaries

E. All of the above

**Q 8. What does the “entropy” method measure?**

A. “how random” a random variable is, and it comes from the field of information theory

B. “how random” a continuous variable is, and it comes from the field of information theory

C. “how random” a random or continuous variable is as it comes from the field of information theory

D. Both A and B

E. All of the above

**Q 9. Market Basket Analysis is fundamentally based on:**

A. Trend Analysis

B. Box plot Analysis

C. Association Analysis

D. Scatter Plot Analysis

E. None of the above

**Q 10. One of the following is a Python library used largely for fitting Bayesian Networks to real data. Identify which one:**

A. SciLib

B. PyMC

C. MyLib

D. MCMC

E. SCIMC

Bottom of Form