



**Microsoft**

**Exam Questions DP-100**

Designing and Implementing a Data Science Solution on Azure

**NEW QUESTION 1**

- (Exam Topic 3)

You are determining if two sets of data are significantly different from one another by using Azure Machine Learning Studio.

Estimated values in one set of data may be more than or less than reference values in the other set of data. You must produce a distribution that has a constant Type I error as a function of the correlation.

You need to produce the distribution.

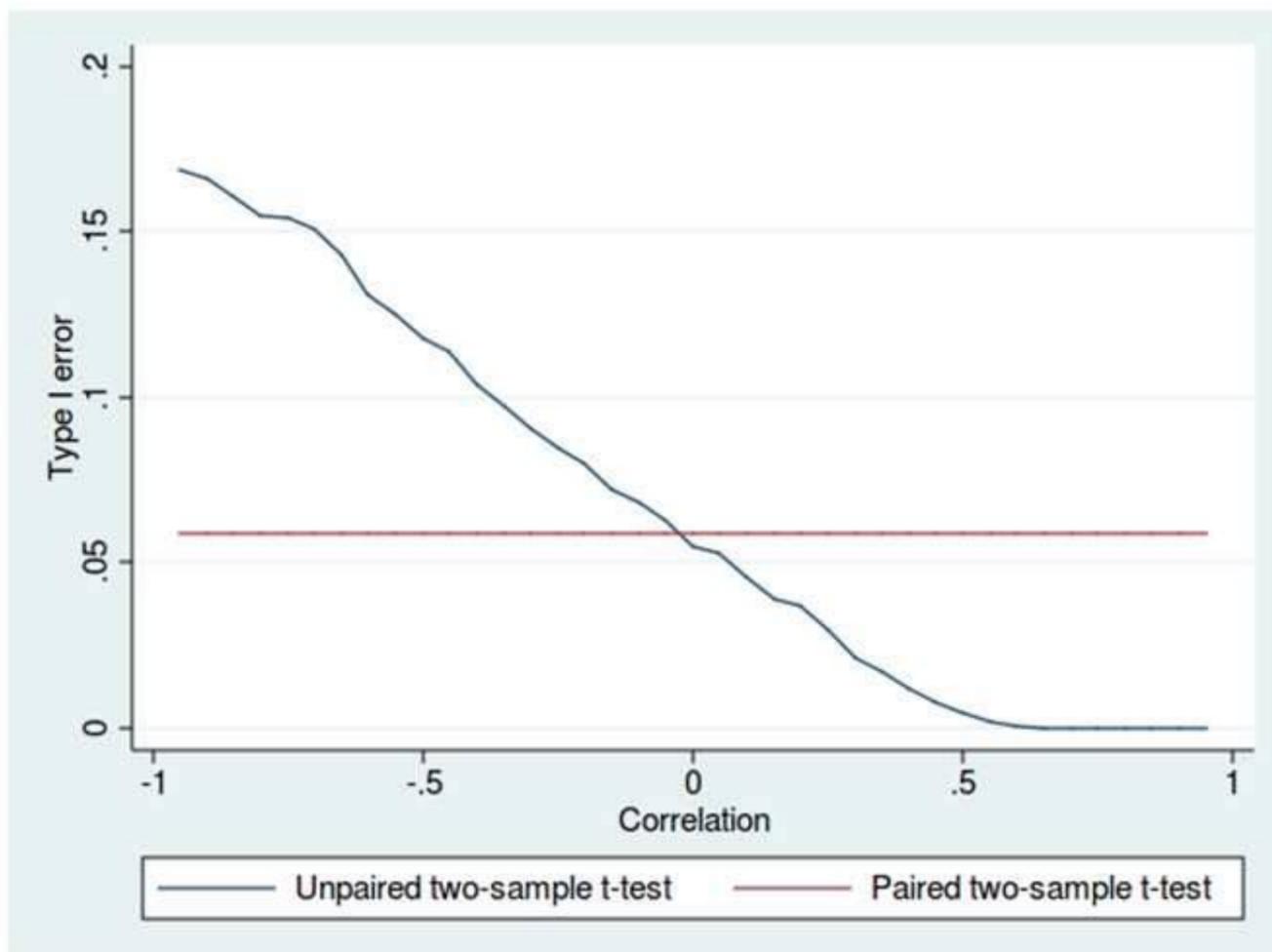
Which type of distribution should you produce?

- A. Paired t-test with a two-tail option
- B. Unpaired t-test with a two tail option
- C. Paired t-test with a one-tail option
- D. Unpaired t-test with a one-tail option

**Answer:** A

**Explanation:**

Choose a one-tail or two-tail test. The default is a two-tailed test. This is the most common type of test, in which the expected distribution is symmetric around zero. Example: Type I error of unpaired and paired two-sample t-tests as a function of the correlation. The simulated random numbers originate from a bivariate normal distribution with a variance of 1.



Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/test-hypothesis-using-t-test> [https://en.wikipedia.org/wiki/Student%27s\\_t-test](https://en.wikipedia.org/wiki/Student%27s_t-test)

**NEW QUESTION 2**

- (Exam Topic 3)

You are creating an experiment by using Azure Machine Learning Studio.

You must divide the data into four subsets for evaluation. There is a high degree of missing values in the data. You must prepare the data for analysis.

You need to select appropriate methods for producing the experiment.

Which three modules should you run in sequence? To answer, move the appropriate actions from the list of actions to the answer area and arrange them in the correct order.

NOTE: More than one order of answer choices is correct. You will receive credit for any of the correct orders you select.

### Actions

- Build Counting Transform
- Missing Values Scrubber
- Feature Hashing
- Clean Missing Data
- Replace Discrete Values
- Import Data
- Latent Dirichlet Transformation
- Partition and Sample

### Answer Area



- A. Mastered
- B. Not Mastered

**Answer:** A

**Explanation:**

The Clean Missing Data module in Azure Machine Learning Studio, to remove, replace, or infer missing values.

**NEW QUESTION 3**

- (Exam Topic 3)

Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.

After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.

You are a data scientist using Azure Machine Learning Studio.

You need to normalize values to produce an output column into bins to predict a target column. Solution: Apply a Quantiles normalization with a QuantileIndex normalization.

Does the solution meet the GOAL?

- A. Yes
- B. No

**Answer:** B

**Explanation:**

Use the Entropy MDL binning mode which has a target column. References:

<https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/group-data-into-bins>

**NEW QUESTION 4**

- (Exam Topic 1)

You need to implement a model development strategy to determine a user's tendency to respond to an ad. Which technique should you use?

- A. Use a Relative Expression Split module to partition the data based on centroid distance.
- B. Use a Relative Expression Split module to partition the data based on distance travelled to the event.
- C. Use a Split Rows module to partition the data based on distance travelled to the event.
- D. Use a Split Rows module to partition the data based on centroid distance.

**Answer:** A

**Explanation:**

Split Data partitions the rows of a dataset into two distinct sets.

The Relative Expression Split option in the Split Data module of Azure Machine Learning Studio is helpful when you need to divide a dataset into training and testing datasets using a numerical expression.

Relative Expression Split: Use this option whenever you want to apply a condition to a number column. The number could be a date/time field, a column containing age or dollar amounts, or even a percentage. For example, you might want to divide your data set depending on the cost of the items, group people by age ranges, or separate data by a calendar date.

Scenario:

Local market segmentation models will be applied before determining a user's propensity to respond to an advertisement.

The distribution of features across training and production data are not consistent References:

<https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/split-data>

**NEW QUESTION 5**

- (Exam Topic 1)

You need to define an evaluation strategy for the crowd sentiment models.

Which three actions should you perform in sequence? To answer, move the appropriate actions from the list of actions to the answer area and arrange them in the

correct order.

Actions	Answer Area
Define a cross-entropy function activation.	
Add cost functions for each target state.	
Evaluate the classification error metric.	⬆
Evaluate the distance error metric.	⬇
Add cost functions for each component metric.	
Define a sigmoid loss function activation.	

- A. Mastered
- B. Not Mastered

Answer: A

**Explanation:**

Step 1: Define a cross-entropy function activation

When using a neural network to perform classification and prediction, it is usually better to use cross-entropy error than classification error, and somewhat better to use cross-entropy error than mean squared error to evaluate the quality of the neural network.

Step 2: Add cost functions for each target state. Step 3: Evaluated the distance error metric. References:

<https://www.analyticsvidhya.com/blog/2018/04/fundamentals-deep-learning-regularization-techniques/>

**NEW QUESTION 6**

- (Exam Topic 1)

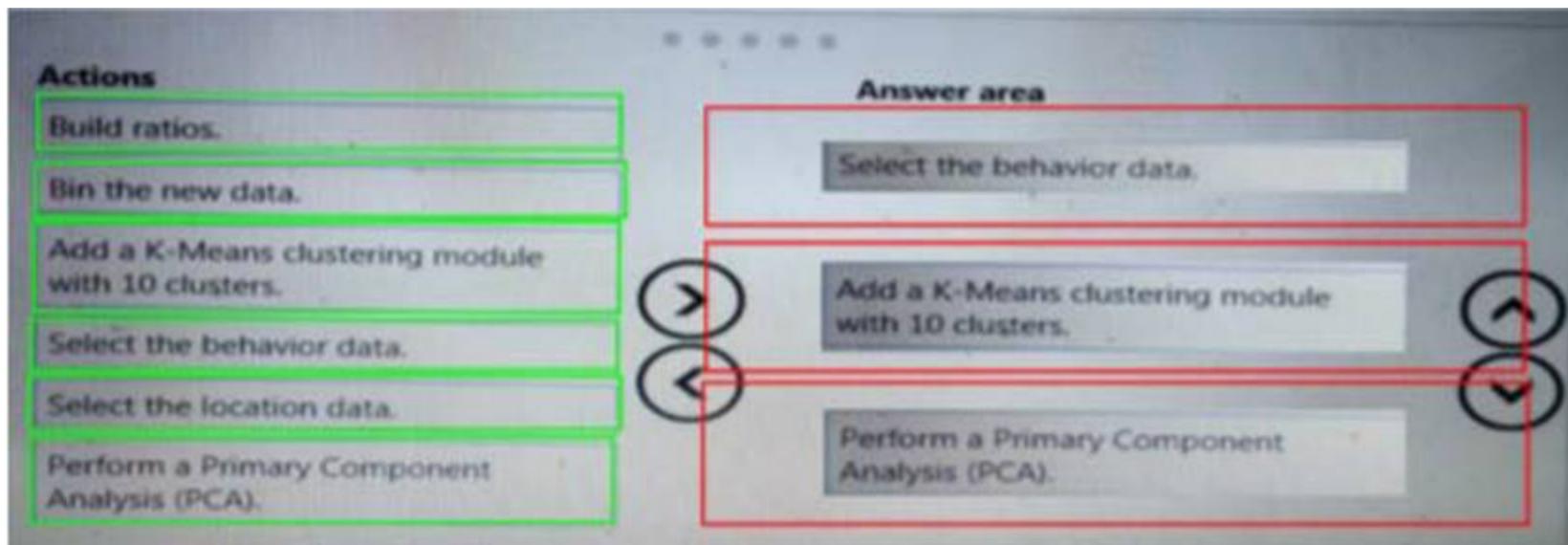
You need to modify the inputs for the global penalty event model to address the bias and variance issue.

Which three actions should you perform in sequence? To answer, move the appropriate actions from the list of actions to the answer area and arrange them in the correct order.

- A. Mastered
- B. Not Mastered

Answer: A

**Explanation:**



**NEW QUESTION 7**

- (Exam Topic 3)

Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.

After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.

You are creating a new experiment in Azure Learning learning Studio.

One class has a much smaller number of observations than the other classes in the training

You need to select an appropriate data sampling strategy to compensate for the class imbalance. Solution: You use the Synthetic Minority Oversampling Technique (SMOTE) sampling mode. Does the solution meet the goal?

- A. Yes
- B. No

**Answer:** A

**Explanation:**

SMOTE is used to increase the number of underrepresented cases in a dataset used for machine learning. SMOTE is a better way of increasing the number of rare cases than simply duplicating existing cases.

References:

<https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/sMOTE>

**NEW QUESTION 8**

- (Exam Topic 3)

You must store data in Azure Blob Storage to support Azure Machine Learning. You need to transfer the data into Azure Blob Storage.

What are three possible ways to achieve the goal? Each correct answer presents a complete solution.

NOTE: Each correct selection is worth one point.

- A. Bulk Insert SQL Query
- B. AzCopy
- C. Python script
- D. Azure Storage Explorer
- E. Bulk Copy Program (BCP)

**Answer:** BCD

**Explanation:**

You can move data to and from Azure Blob storage using different technologies: Azure Storage-Explorer

AzCopy Python SSIS

References:

<https://docs.microsoft.com/en-us/azure/machine-learning/team-data-science-process/move-azure-blob>

**NEW QUESTION 9**

- (Exam Topic 3)

You plan to create a speech recognition deep learning model. The model must support the latest version of Python.

You need to recommend a deep learning framework for speech recognition to include in the Data Science Virtual Machine (DSVM).

What should you recommend?

- A. Apache Drill
- B. Tensorflow
- C. Rattle
- D. Weka

**Answer:** B

**Explanation:**

TensorFlow is an open source library for numerical computation and large-scale machine learning. It uses Python to provide a convenient front-end API for building applications with the framework

TensorFlow can train and run deep neural networks for handwritten digit classification, image recognition, word embeddings, recurrent neural networks, sequence-to-sequence models for machine translation, natural language processing, and PDE (partial differential equation) based simulations.

References:

<https://www.infoworld.com/article/3278008/what-is-tensorflow-the-machine-learning-library-explained.html>

**NEW QUESTION 10**

- (Exam Topic 3)

You create a classification model with a dataset that contains 100 samples with Class A and 10,000 samples with Class B. The variation of Class B is very high. You need to resolve imbalances. Which method should you use?

- A. Partition and Sample
- B. Cluster Centroids
- C. Tomek links
- D. Synthetic Minority Oversampling Technique (SMOTE)

**Answer: D**

**NEW QUESTION 11**

- (Exam Topic 3)

You plan to build a team data science environment. Data for training models in machine learning pipelines will be over 20 GB in size. You have the following requirements:

- > Models must be built using Caffe2 or Chainer frameworks.
- > Data scientists must be able to use a data science environment to build the machine learning pipelines and train models on their personal devices in both connected and disconnected network environments.
- > Personal devices must support updating machine learning pipelines when connected to a network. You need to select a data science environment.

Which environment should you use?

- A. Azure Machine Learning Service
- B. Azure Machine Learning Studio
- C. Azure Databricks
- D. Azure Kubernetes Service (AKS)

**Answer: A**

**Explanation:**

The Data Science Virtual Machine (DSVM) is a customized VM image on Microsoft's Azure cloud built specifically for doing data science. Caffe2 and Chainer are supported by DSVM. DSVM integrates with Azure Machine Learning.

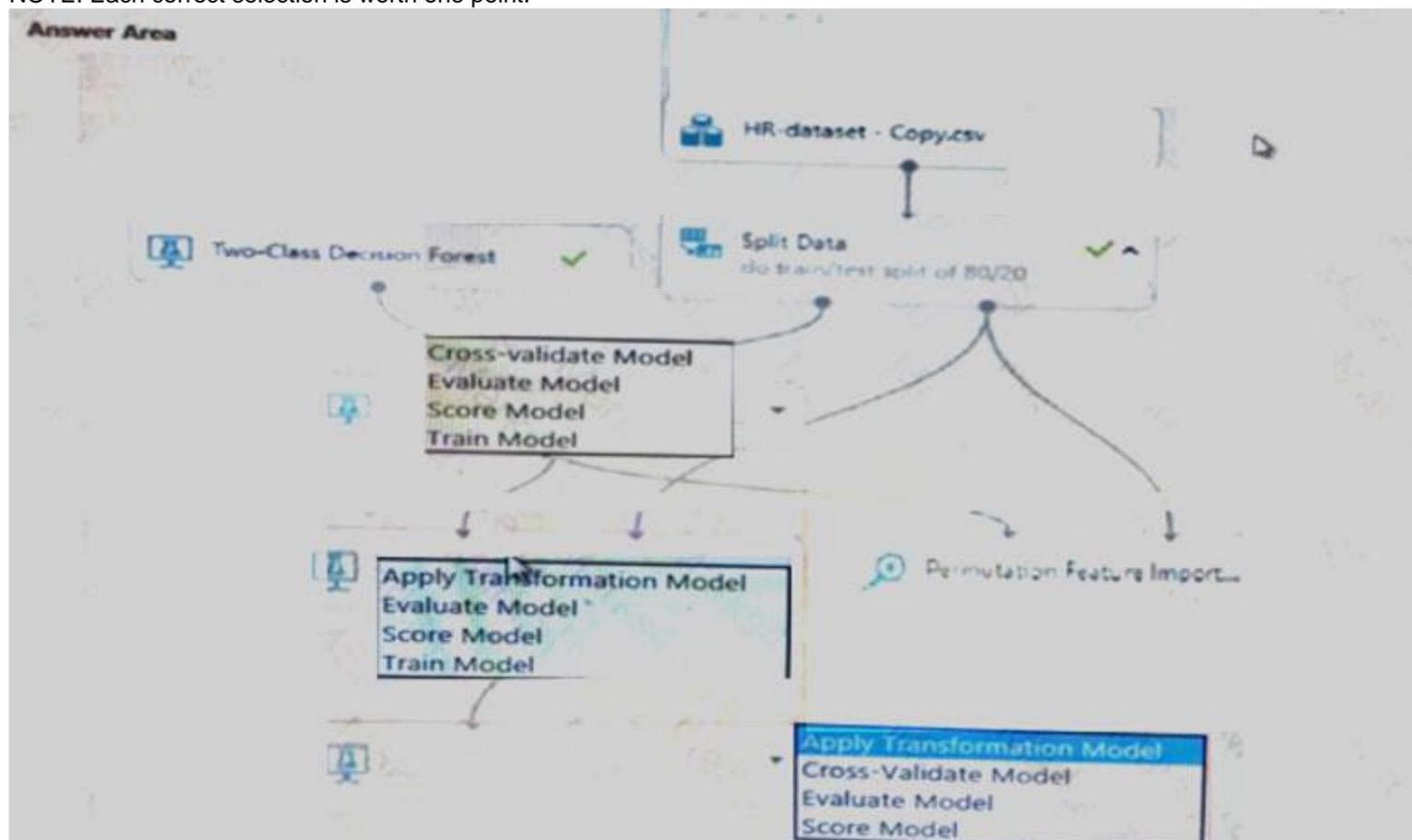
**NEW QUESTION 12**

- (Exam Topic 3)

You create a binary classification model using Azure Machine Learning Studio.

You must use a Receiver Operating Characteristic (ROC) curve and an F1 score to evaluate the model. You need to create the required business metrics. How should you complete the experiment? To answer, select the appropriate options in the dialog box in the answer area.

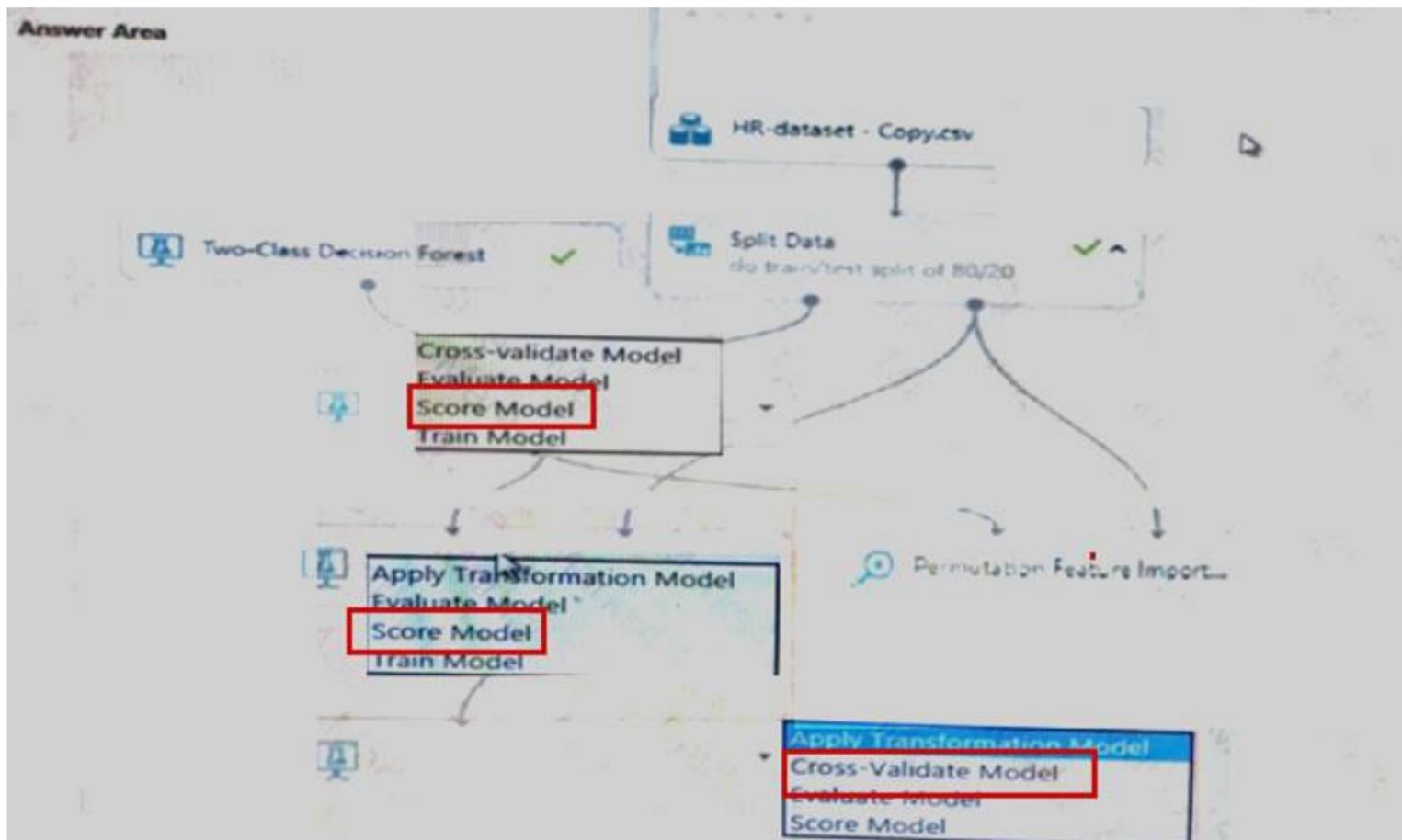
NOTE: Each correct selection is worth one point.



- A. Mastered
- B. Not Mastered

**Answer: A**

**Explanation:**



**NEW QUESTION 13**

- (Exam Topic 3)

Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.

After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.

You are creating a model to predict the price of a student's artwork depending on the following variables: the student's length of education, degree type, and art form.

You start by creating a linear regression model. You need to evaluate the linear regression model.

Solution: Use the following metrics: Accuracy, Precision, Recall, F1 score and AUC. Does the solution meet the goal?

- A. Yes
- B. No

**Answer:** B

**Explanation:**

Those are metrics for evaluating classification models, instead use: Mean Absolute Error, Root Mean Absolute Error, Relative Absolute Error, Relative Squared Error, and the Coefficient of Determination.

References:

<https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/evaluate-model>

**NEW QUESTION 14**

- (Exam Topic 3)

You are developing a machine learning, experiment by using Azure. The following images show the input and output of a machine learning experiment:

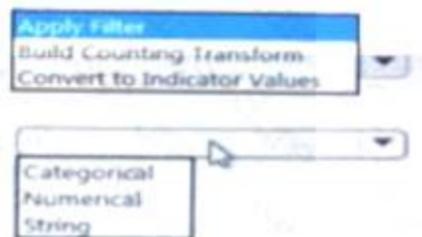


Use the drop-down menus to select the answer choice that answers each question based on the information presented in the graphic.

NOTE: Each correct selection is worth one point.

You need to perform the data transformation applied to the Risk Level column. Which module should you use?

What is the expected input column type for this transformation?



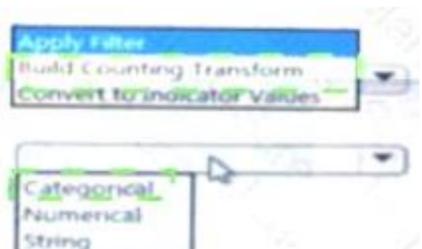
- A. Mastered
- B. Not Mastered

**Answer:** A

**Explanation:**

You need to perform the data transformation applied to the Risk Level column. Which module should you use?

What is the expected input column type for this transformation?



### NEW QUESTION 15

- (Exam Topic 3)

Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.

After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.

You are analyzing a numerical dataset which contains missing values in several columns.

You must clean the missing values using an appropriate operation without affecting the dimensionality of the feature set.

You need to analyze a full dataset to include all values.

Solution: Replace each missing value using the Multiple Imputation by Chained Equations (MICE) method. Does the solution meet the goal?

- A. Yes
- B. NO

**Answer:** A

**Explanation:**

Replace using MICE: For each missing value, this option assigns a new value, which is calculated by using a method described in the statistical literature as "Multivariate Imputation using Chained Equations" or "Multiple Imputation by Chained Equations". With a multiple imputation method, each variable with missing data is modeled conditionally using the other variables in the data before filling in the missing values.

Note: Multivariate imputation by chained equations (MICE), sometimes called "fully conditional specification" or "sequential regression multiple imputation" has emerged in the statistical literature as one principled method of addressing missing data. Creating multiple imputations, as opposed to single imputations, accounts for the statistical uncertainty in the imputations. In addition, the chained equations approach is very flexible and can handle variables of varying types (e.g., continuous or binary) as well as complexities such as bounds or survey skip patterns.

References: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3074241/>

<https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/clean-missing-data>

### NEW QUESTION 16

- (Exam Topic 3)

You are solving a classification task.

You must evaluate your model on a limited data sample by using k-fold cross validation. You start by configuring a k parameter as the number of splits.

You need to configure the k parameter for the cross-validation. Which value should you use?

- A. k=0.5
- B. k=0
- C. k=5
- D. k=1

**Answer:** C

**Explanation:**

Leave One Out (LOO) cross-validation

Setting  $K = n$  (the number of observations) yields n-fold and is called leave-one out cross-validation (LOO), a special case of the K-fold approach.

LOO CV is sometimes useful but typically doesn't shake up the data enough. The estimates from each fold are highly correlated and hence their average can have high variance.

This is why the usual choice is  $K=5$  or  $10$ . It provides a good compromise for the bias-variance tradeoff.

### NEW QUESTION 17

- (Exam Topic 3)

You need to select a feature extraction method. Which method should you use?

- A. Spearman correlation
- B. Mutual information
- C. Mann-Whitney test

D. Pearson's correlation

**Answer:** D

**NEW QUESTION 18**

- (Exam Topic 3)

You create a binary classification model by using Azure Machine Learning Studio.

You must tune hyperparameters by performing a parameter sweep of the model. The parameter sweep must meet the following requirements:

- > iterate all possible combinations of hyperparameters
- > minimize computing resources required to perform the sweep
- > You need to perform a parameter sweep of the model.

Which parameter sweep mode should you use?

- A. Random sweep
- B. Sweep clustering
- C. Entire grid
- D. Random grid
- E. Random seed

**Answer:** D

**Explanation:**

Maximum number of runs on random grid: This option also controls the number of iterations over a random sampling of parameter values, but the values are not generated randomly from the specified range; instead, a matrix is created of all possible combinations of parameter values and a random sampling is taken over the matrix. This method is more efficient and less prone to regional oversampling or undersampling.

If you are training a model that supports an integrated parameter sweep, you can also set a range of seed values to use and iterate over the random seeds as well. This is optional, but can be useful for avoiding bias introduced by seed selection.

**NEW QUESTION 19**

- (Exam Topic 3)

You have a dataset created for multiclass classification tasks that contains a normalized numerical feature set with 10,000 data points and 150 features.

You use 75 percent of the data points for training and 25 percent for testing. You are using the scikit-learn machine learning library in Python. You use X to denote the feature set and Y to denote class labels.

You create the following Python data frames:

You need to apply the Principal Component Analysis (PCA) method to reduce the dimensionality of the feature set to 10 features in both training and testing sets. How should you complete the code segment? To answer, select the appropriate options in the answer area. NOTE: Each correct selection is worth one point.

```

from sklearn.decomposition import PCA
pca =
PCA()
PCA(n_components = 150)
PCA(n_components = 10)
PCA(n_components = 10000)

X_train =
pca
model
sklearn.decomposition

x_test = pca.
x_test
X_train
fit(x_test)
transform(x_test)
    
```

- A. Mastered
- B. Not Mastered

**Answer:** A

**Explanation:**

Box 1: PCA(n\_components = 10)

Need to reduce the dimensionality of the feature set to 10 features in both training and testing sets. Example:

from sklearn.decomposition import PCA pca = PCA(n\_components=2) ;2 dimensions principalComponents = pca.fit\_transform(x)

Box 2: pca

fit\_transform(X[, y])fits the model with X and apply the dimensionality reduction on X. Box 3: transform(x\_test)

transform(X) applies dimensionality reduction to X. References:

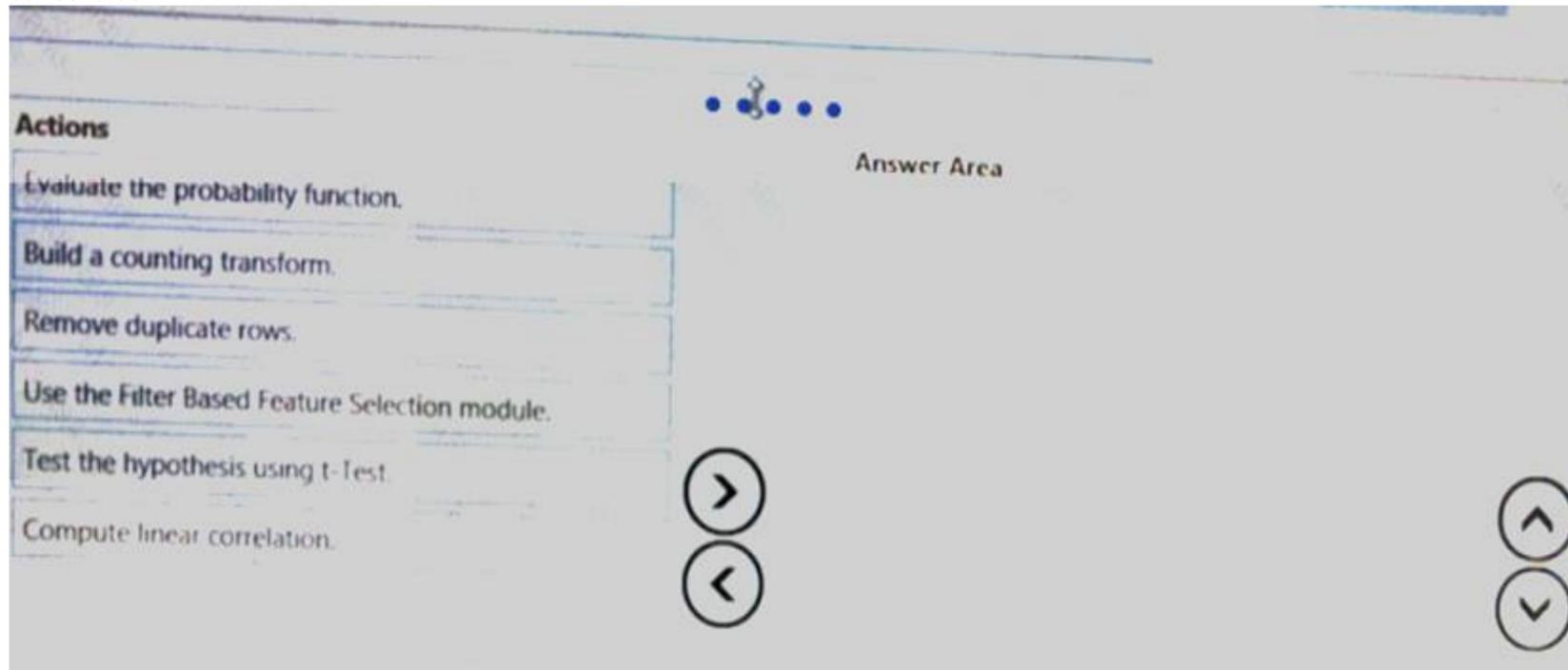
<https://scikit-learn.org/stable/modules/generated/sklearn.decomposition.PCA.html>

**NEW QUESTION 20**

- (Exam Topic 3)

You are producing a multiple linear regression model in Azure Machine learning Studio. Several independent variables are highly correlated. You need to select appropriate methods for conducting elective feature engineering on all the data.

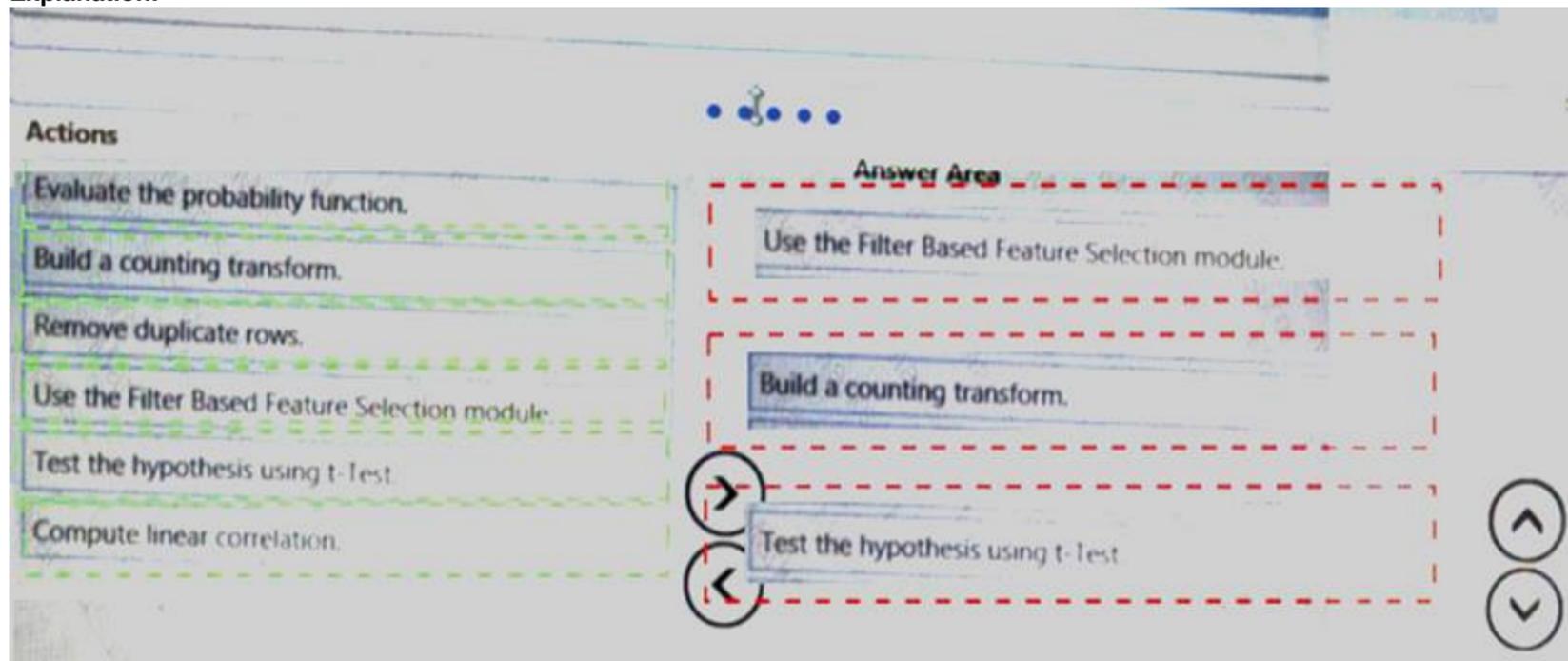
Which three actions should you perform in sequence? To answer, move the appropriate Actions from the list of actions to the answer area and arrange them in the correct order.



- A. Mastered
- B. Not Mastered

Answer: A

Explanation:



NEW QUESTION 21

.....

## Thank You for Trying Our Product

### We offer two products:

1st - We have Practice Tests Software with Actual Exam Questions

2nd - Questions and Answers in PDF Format

### DP-100 Practice Exam Features:

- \* DP-100 Questions and Answers Updated Frequently
- \* DP-100 Practice Questions Verified by Expert Senior Certified Staff
- \* DP-100 Most Realistic Questions that Guarantee you a Pass on Your FirstTry
- \* DP-100 Practice Test Questions in Multiple Choice Formats and Updatesfor 1 Year

**100% Actual & Verified — Instant Download, Please Click**  
[Order The DP-100 Practice Test Here](#)