CertyIQ


# CertyIQ

## Premium exam material

Get certification quickly with the CertyIQ Premium exam material.
Everything you need to prepare, learn & pass your certification exam easily. Lifetime free updates
First attempt guaranteed success.

https://www.CertyIQ.com

# About CertyIQ

We here at CertyIQ eventually got enough of the industry's greedy exam paid for. Our team of IT professionals comes with years of experience in the IT industry Prior to training CertIQ we worked in test areas where we observed the horrors of the paywall exam preparation system.

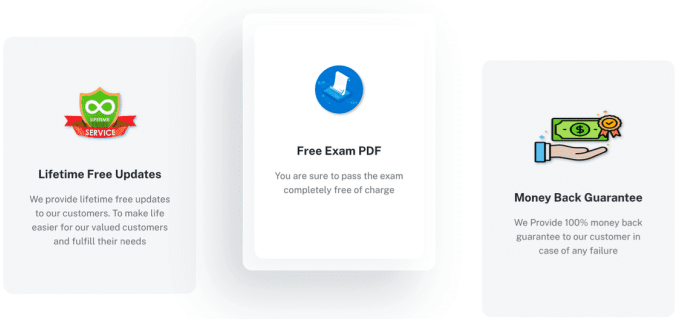
The misuse of the preparation system has left our team disillusioned. And for that reason, we decided it was time to make a difference. We had to make In this way, CertyIQ was created to provide quality materials without stealing from everyday people who are trying to make a living.

# Doubt Support

We have developed a very scalable solution using which we are able to solve 400+ doubts every single day with an average rating of 4.8 out of 5.

https://www.certyiq.com

Mail us on - certyiqofficial@gmail.com




**John**

October 19, 2022

★★★★★

Thanks you so much for your help. I scored 972 in my exam today. More than 90% were from your PDFs!

**Dana**

September 04, 2022

★★★★★

Thanks a lot for this updated AZ-900 Q&A. I just passed my exam and got 974, I followed both of your Az-900 videos and the 6 PDF, the PDFs are very much valid, all answers are correct. Could you please create a similar video/PDF for DP900, your content/PDF's is really awesome. The team did a really good job. Thank You 😊.

**Ahamed Shibly**

2 months ago

★★★★★

Customer support is realy fast and helpful, I just finished my exam and this video along with the 6 PDF helped me pass! Definitely recommend getting the PDFs. Thank you!

October 22, 2022

★★★★★

Passed my exam today with 891 marks. Out of 52 questions, 51 were from certyiq PDFs including Contoso case study.
Thank You certyiq team!

**Henry Rome**

2 months ago

★★★★★

These questions are real and 100 % valid. Thank you so much for your efforts, also your 4 PDFs are awesome, I passed the DP900 exam on 1 Sept. With 968 marks. Thanks a lot, buddy!

**Esmaria**

2 months ago

★★★★★

Simple easy to understand explanations. To anyone out there wanting to write AZ900, I highly recommend 6 PDF's.Thank you so much, appreciate all your hard work in having such great content. Passed my exam Today - 3 September with 942 score.

# Microsoft

(DP-100)

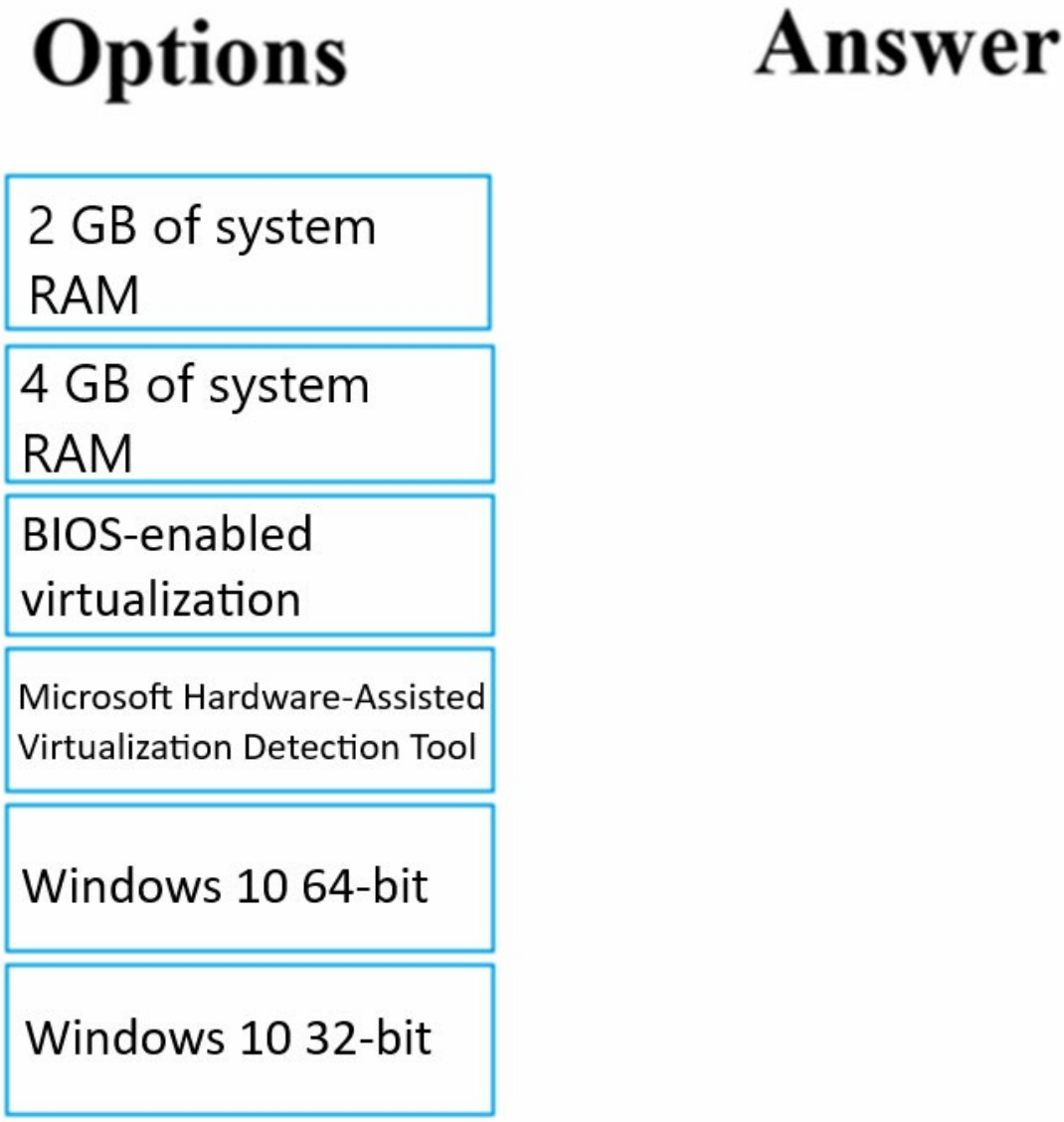
Designing and Implementing a Data Science Solution on Azure (beta)

Total: **484 Questions**

Link: https://certyiq.com/papers?provider=microsoft&exam=dp-100

**Question: 1** CertyIQ

DRAG DROP -
You are planning to host practical training to acquaint staff with Docker for Windows.
Staff devices must support the installation of Docker.
Which of the following are requirements for this installation? Answer by dragging the correct options from the list to the answer area.
Select and Place:




**Answer:**

| Options | Answer |
| --- | --- |
| 2 GB of system RAM | 4 GB of system RAM |
| | BIOS-enabled virtualization |
| | Windows 10 64-bit |
| Microsoft Hardware-Assisted Virtualization Detection Tool | |
| | |
| Windows 10 32-bit | |

**Explanation:**

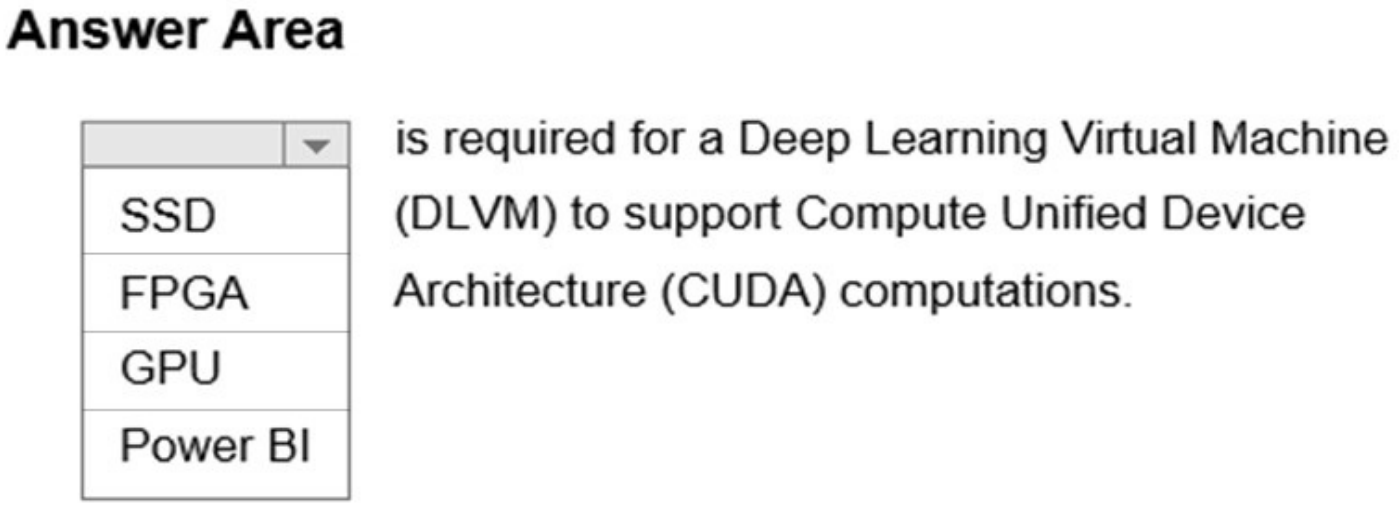
Reference:
https://docs.docker.com/toolbox/toolbox_install_windows/
https://blogs.technet.microsoft.com/canitpro/2015/09/08/step-by-step-enabling-hyper-v-for-use-on-windows-10/ https://docs.docker.com/docker-for-windows/install/
" target="_blank" style="word-break: break-all;">

## WSL 2 backend

- Windows 10 64-bit: Home or Pro 2004 (build 19041) or higher, or Enterprise or Education 1909 (build 18363) or higher.
- Enable the WSL 2 feature on Windows. For detailed instructions, refer to the Microsoft documentation.
- The following hardware prerequisites are required to successfully run WSL 2 on Windows 10:
  - 64-bit processor with Second Level Address Translation (SLAT)
  - 4GB system RAM
  - BIOS-level hardware virtualization support must be enabled in the BIOS settings. For more information, see Virtualization.
- Download and install the Linux kernel update package.

**Question: 2** CertyIQ

HOTSPOT -
Complete the sentence by selecting the correct option in the answer area.
Hot Area:




**Answer:**


**Explanation:**
A Deep Learning Virtual Machine is a pre-configured environment for deep learning using GPU instances.

**Question: 3** CertyIQ

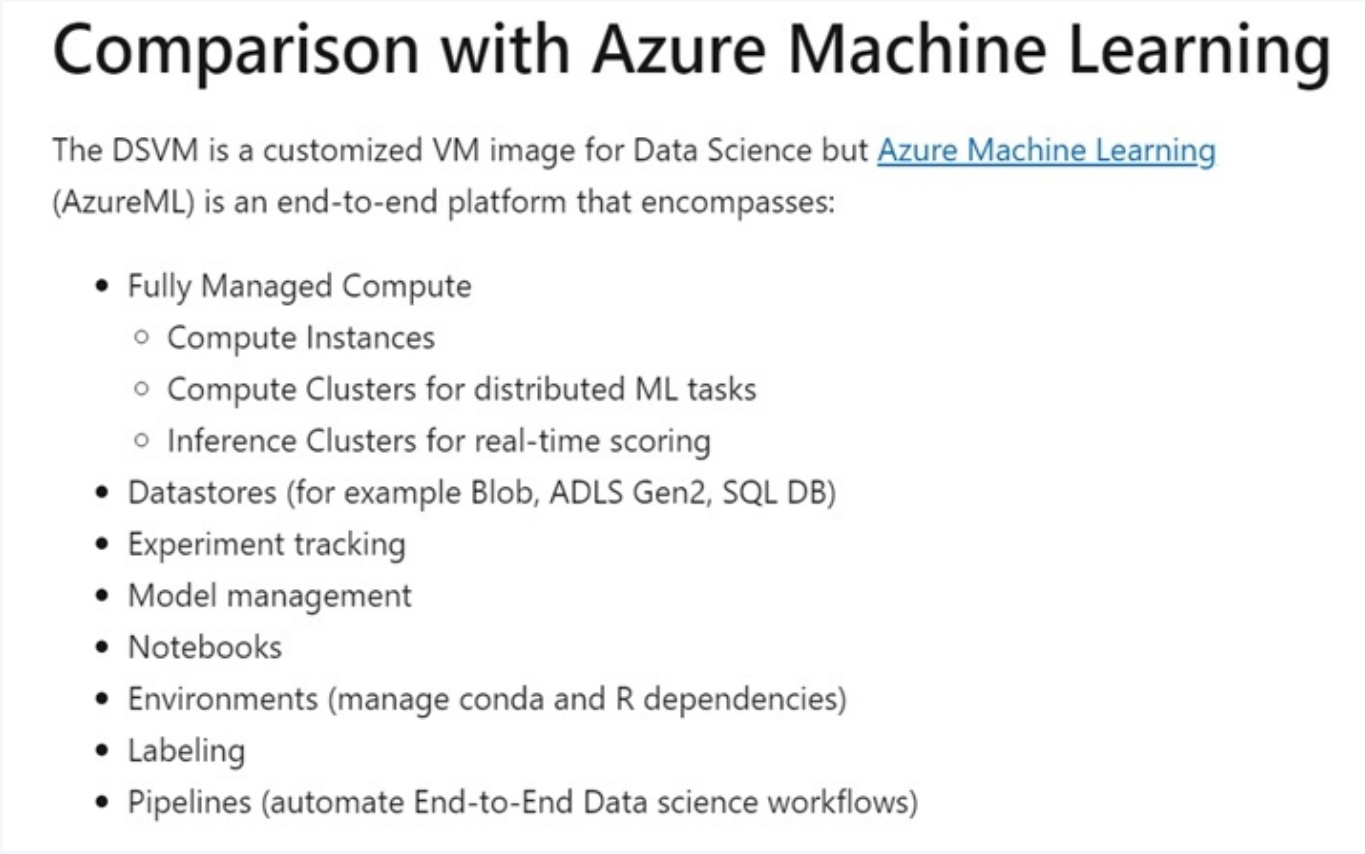
You need to implement a Data Science Virtual Machine (DSVM) that supports the Caffe2 deep learning framework.
Which of the following DSVM should you create?

A. Windows Server 2012 DSVM
B. Windows Server 2016 DSVM
C. Ubuntu 16.04 DSVM
D. CentOS 7.4 DSVM

**Answer: C**

**Explanation:**
Caffe2 is supported by Data Science Virtual Machine for Linux.
Microsoft offers Linux editions of the DSVM on Ubuntu 16.04 LTS and CentOS 7.4.
However, only the DSVM on Ubuntu is preconfigured for Caffe2.

Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/data-science-virtual-machine/overview
" target="_blank" style="word-break: break-all;">




**Question: 4** CertyIQ

This question is included in a number of questions that depicts the identical set-up. However, every question has a distinctive result. Establish if the recommendation satisfies the requirements.
You have been tasked with employing a machine learning model, which makes use of a PostgreSQL database and needs GPU processing, to forecast prices.
You are preparing to create a virtual machine that has the necessary tools built into it.
You need to make use of the correct virtual machine type.
Recommendation: You make use of a Geo AI Data Science Virtual Machine (Geo-DSVM) Windows edition.
Will the requirements be satisfied?

A. Yes
B. No

**Answer: B**

**Explanation:**
The Azure Geo AI Data Science VM (Geo-DSVM) delivers geospatial analytics capabilities from Microsoft's Data Science VM. Specifically, this VM extends the AI and data science toolkits in the Data Science VM by adding ESRI's market-leading ArcGIS Pro Geographic Information System.

Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/data-science-virtual-machine/overview

**Question: 5** CertyIQ

This question is included in a number of questions that depicts the identical set-up. However, every question has a

distinctive result. Establish if the recommendation satisfies the requirements.
You have been tasked with employing a machine learning model, which makes use of a PostgreSQL database and needs GPU processing, to forecast prices.
You are preparing to create a virtual machine that has the necessary tools built into it.
You need to make use of the correct virtual machine type.
Recommendation: You make use of a Deep Learning Virtual Machine (DLVM) Windows edition.
Will the requirements be satisfied?

A. Yes
B. No

**Answer: A**

**Explanation:**
1. I will also stand with a Yes.A Deep Learning Virtual Machine (DLVM) is a pre-configured virtual machine image on Microsoft Azure that is optimized for training deep learning models and includes popular tools such as TensorFlow, PyTorch, and Caffe2. The DLVM Windows edition includes support for GPU processing, making it suitable for the task of running a machine learning model that requires GPU processing to forecast prices. Additionally, the DLVM can be configured to use PostgreSQL as the database, satisfying the requirement for a PostgreSQL database. Therefore, the recommendation to use a DLVM Windows edition will satisfy the requirements.

**Question: 6** CertyIQ

This question is included in a number of questions that depicts the identical set-up. However, every question has a distinctive result. Establish if the recommendation satisfies the requirements.
You have been tasked with employing a machine learning model, which makes use of a PostgreSQL database and needs GPU processing, to forecast prices.
You are preparing to create a virtual machine that has the necessary tools built into it.
You need to make use of the correct virtual machine type.
Recommendation: You make use of a Data Science Virtual Machine (DSVM) Windows edition.
Will the requirements be satisfied?

A. Yes
B. No

**Answer: A**

**Explanation:**

In the DSVM, your training models can use deep learning algorithms on hardware that's based on graphics processing units (GPUs).

PostgreSQL is available for the following operating systems: Linux (all recent distributions), 64-bit installers available for macOS (OS X) version 10.6 and newer "

Windows (with installers available for 64-bit version; tested on latest versions and back to Windows 2012 R2.

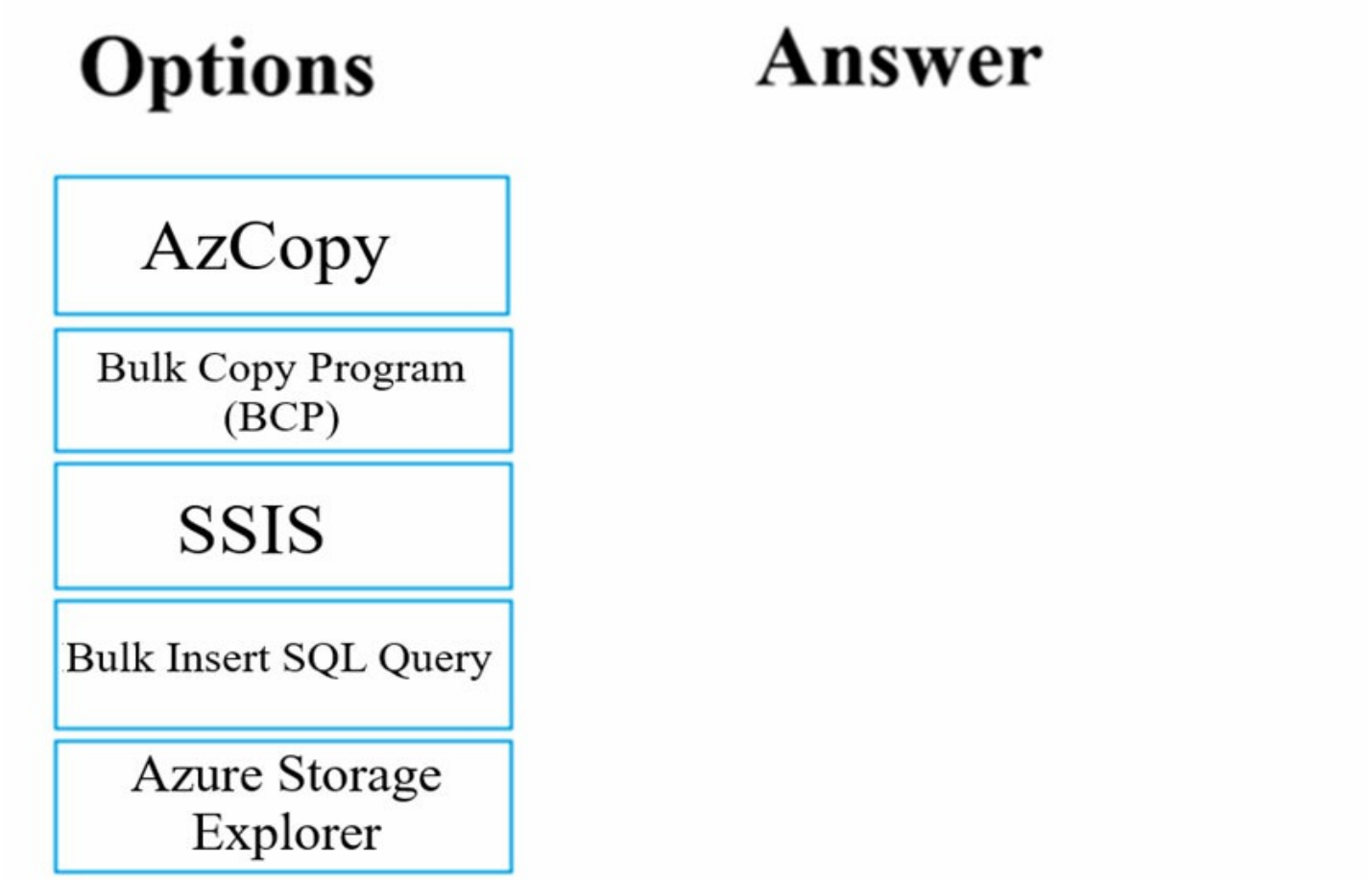
References:

https://docs.microsoft.com/en-us/azure/machine-learning/data-science-virtual-machine/overview

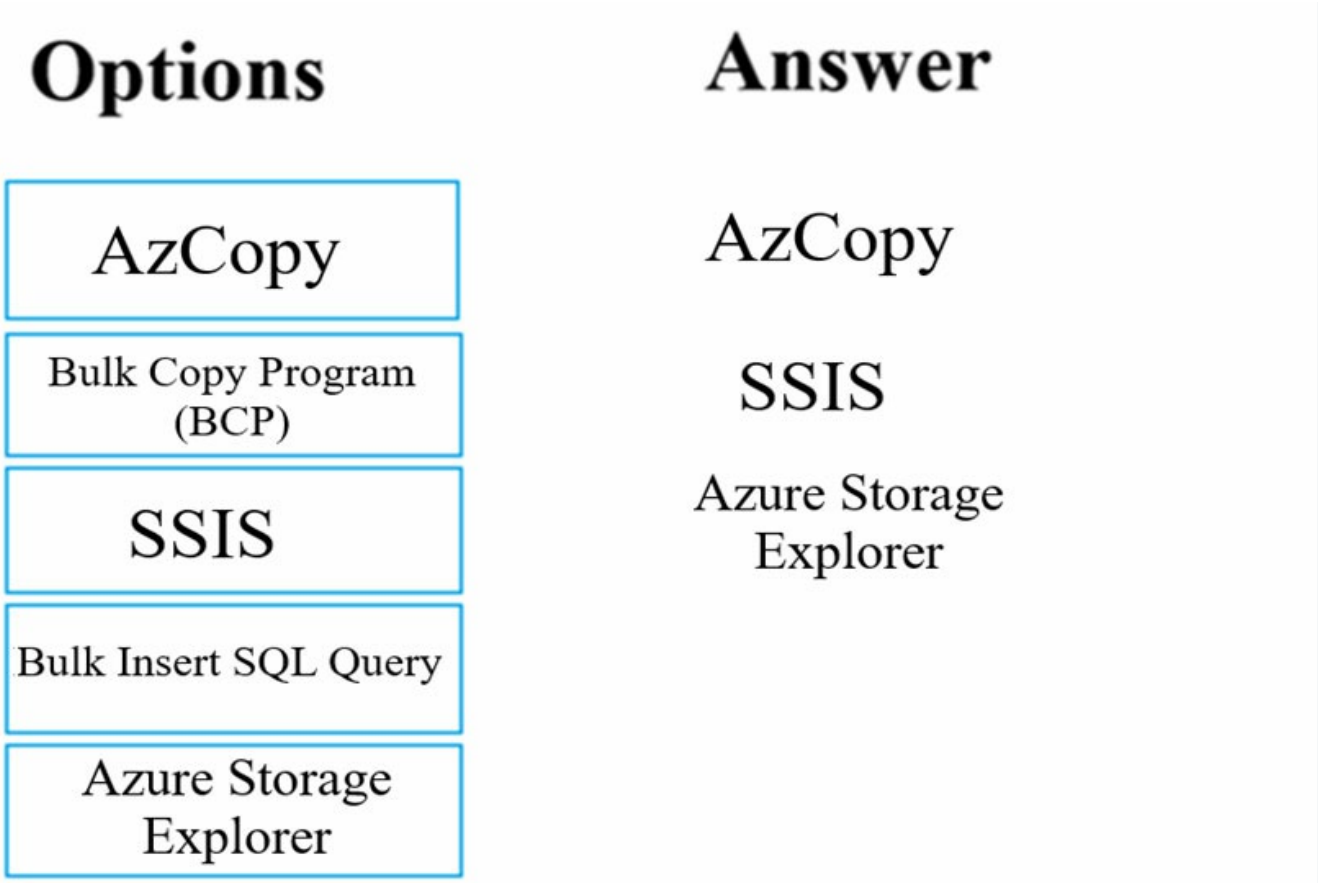
**Question: 7** CertyIQ

DRAG DROP -
You have been tasked with moving data into Azure Blob Storage for the purpose of supporting Azure Machine

Learning.
Which of the following can be used to complete your task? Answer by dragging the correct options from the list to the answer area.
Select and Place:

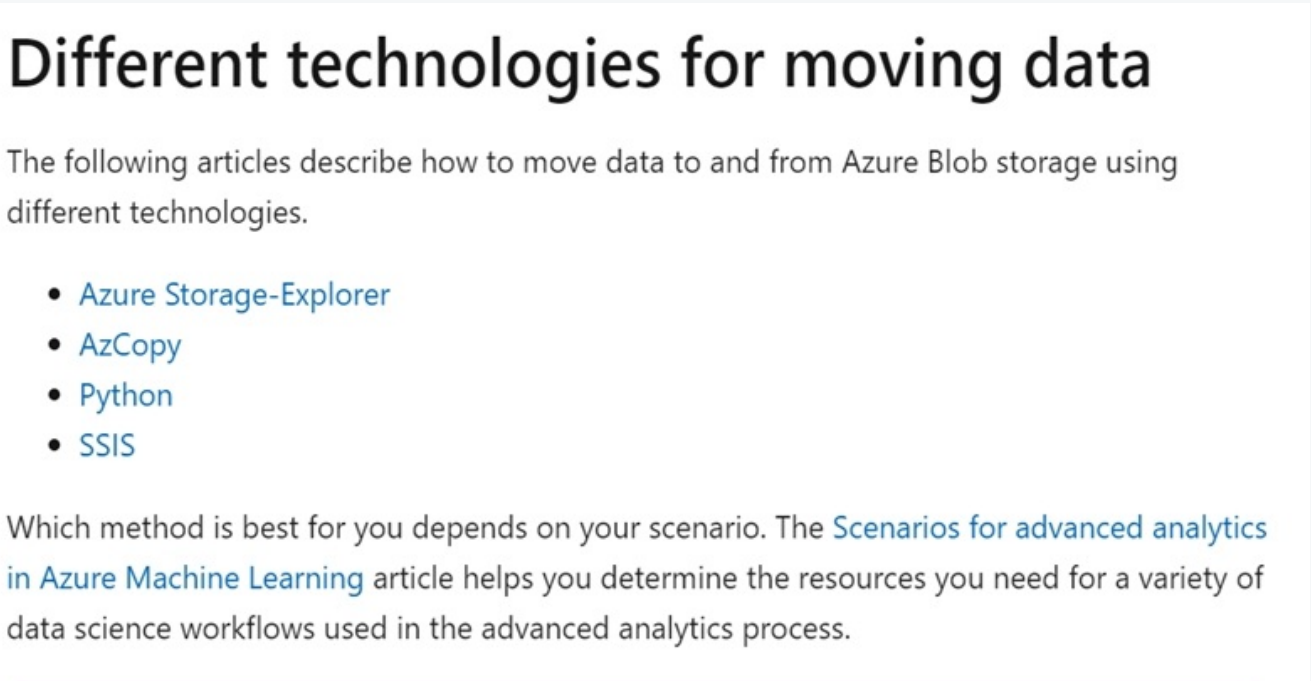



**Answer:**



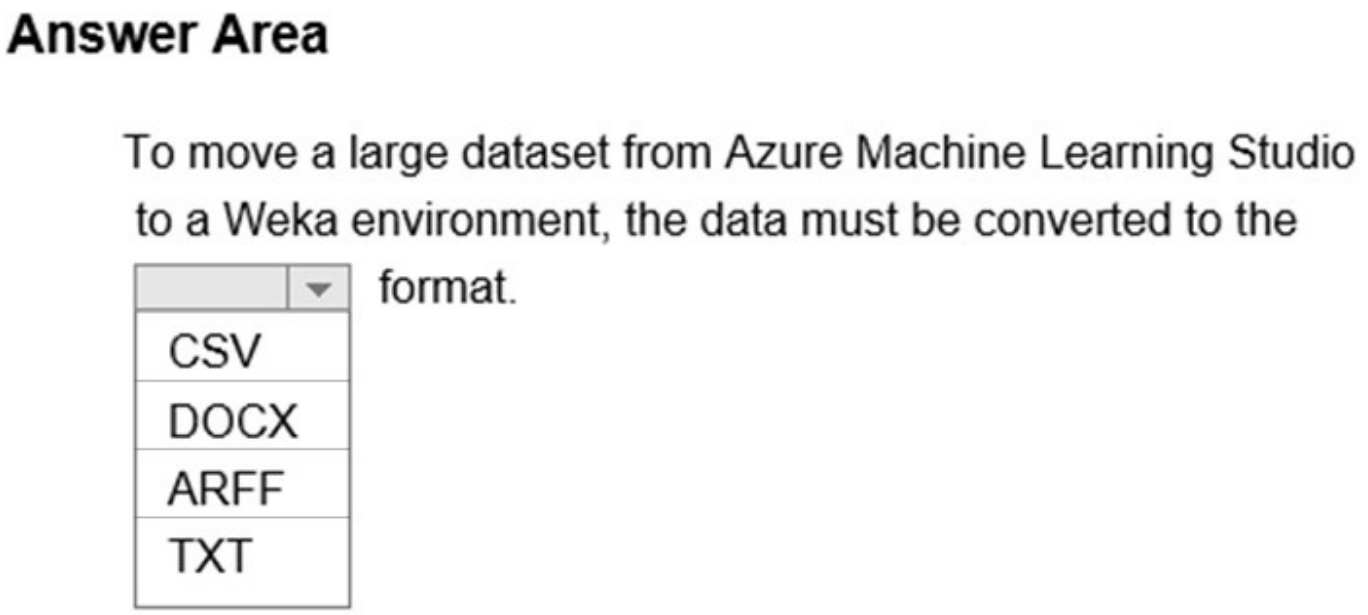

**Explanation:**
You can move data to and from Azure Blob storage using different technologies:
- Azure Storage-Explorer
- AzCopy
- Python
- SSIS

Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/team-data-science-process/move-azure-blob" target="_blank" style="word-break: break-all;">



**Question: 8** CertyIQ

HOTSPOT -
Complete the sentence by selecting the correct option in the answer area.
Hot Area:




**Answer:**


**Explanation:**
Use the Convert to ARFF module in Azure Machine Learning Studio, to convert datasets and results in Azure Machine Learning to the attribute-relation file format used by the Weka toolset. This format is known as ARFF.
The ARFF data specification for Weka supports multiple machine learning tasks, including data preprocessing, classification, and feature selection. In this format, data is organized by entities and their attributes, and is contained in a single text file.

Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/convert-to-arff
" target="_blank" style="word-break: break-all;">

# Module overview

This article describes how to use the Convert to ARFF module in Machine Learning Studio (classic), to convert datasets and results the attribute-relation file format used by the Weka toolset. This format is known as ARFF.

The ARFF data specification for Weka supports multiple machine learning tasks, including data preprocessing, classification, and feature selection. In this format, data is organized by entites and their attributes, and is contained in a single text file. You can find details of the Weka file format in the Technical Notes section.

In general, conversion to the Weka file format is required only if you want to use both Machine Learning and Weka, and intend to move your training data back and forth between them.

For more information about the Weka toolset, see this Wikipedia article: Weka (machine learning)

**Question: 9** CertyIQ

You have been tasked with designing a deep learning model, which accommodates the most recent edition of Python, to recognize language.
You have to include a suitable deep learning framework in the Data Science Virtual Machine (DSVM).
Which of the following actions should you take?

A. You should consider including Rattle.
B. You should consider including TensorFlow.
C. You should consider including Theano.
D. You should consider including Chainer.

**Answer: B**

**Explanation:**

Reference:
https://www.infoworld.com/article/3278008/what-is-tensorflow-the-machine-learning-library-explained.html" target="_blank" style="word-break: break-all;">

Machine learning is a complex discipline. But implementing machine learning models is far less daunting and difficult than it used to be, thanks to machine learning frameworks—such as Google's TensorFlow—that ease the process of acquiring data, training models, serving predictions, and refining future results.

Created by the Google Brain team, TensorFlow is an open source library for numerical computation and large-scale machine learning. TensorFlow bundles together a slew of machine learning and deep learning (aka neural networking) models and algorithms and makes them useful by way of a common metaphor. It uses Python to provide a convenient front-end API for building applications with the framework, while executing those applications in high-performance C++.

**Question: 10** CertyIQ

This question is included in a number of questions that depicts the identical set-up. However, every question has a distinctive result. Establish if the recommendation satisfies the requirements.
You have been tasked with evaluating your model on a partial data sample via k-fold cross-validation.
You have already configured a k parameter as the number of splits. You now have to configure the k parameter for the cross-validation with the usual value choice.
Recommendation: You configure the use of the value k=3.
Will the requirements be satisfied?

A. Yes
B. No

**Answer: B**

**Explanation:**

Usual choice is key word here and usual choice is K=5 or 10. So answer is B. You can use as many splits as you want. It all depends on the data. Train/test/validate is basically 3 splits that are just swapped around. 3 is perfectly fine.

Reference:

https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/cross-validate-model

**Question: 11** CertyIQ

This question is included in a number of questions that depicts the identical set-up. However, every question has a distinctive result. Establish if the recommendation satisfies the requirements.
You have been tasked with evaluating your model on a partial data sample via k-fold cross-validation.
You have already configured a k parameter as the number of splits. You now have to configure the k parameter for the cross-validation with the usual value choice.
Recommendation: You configure the use of the value k=10.

Will the requirements be satisfied?

A. Yes
B. No

**Answer: A**

**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.

**Question: 12** CertyIQ

You construct a machine learning experiment via Azure Machine Learning Studio.
You would like to split data into two separate datasets.
Which of the following actions should you take?

A. You should make use of the Split Data module.
B. You should make use of the Group Categorical Values module.
C. You should make use of the Clip Values module.
D. You should make use of the Group Data into Bins module.

**Answer: A**

**Explanation:**

The ans selected as B is selected based on old Machine Learning Studio but now azure machine learning is used to perform any task and according to that A is right.

Reference:

https://learn.microsoft.com/en-us/azure/machine-learning/component-reference/split-data

https://learn.microsoft.com/en-us/previous-versions/azure/machine-learning/studio-module-reference/cross-validate-model

**Question: 13** CertyIQ

You have been tasked with creating a new Azure pipeline via the Machine Learning designer.
You have to makes sure that the pipeline trains a model using data in a comma-separated values (CSV) file that is published on a website. A dataset for the file for this file does not exist.
Data from the CSV file must be ingested into the designer pipeline with the least amount of administrative effort as possible.
Which of the following actions should you take?

A. You should make use of the Convert to TXT module.

B. You should add the Copy Data object to the pipeline.

C. You should add the Import Data object to the pipeline.

D. You should add the Dataset object to the pipeline.

**Answer: C**

**Explanation:**

Both dataset and import data modules can do the job, but dataset module requires one to register the dataset first before using it in designer, so violating the least admin efforts. Import data component is used to import data from data sources such as web URLs with minimum effort. Reference:

https://docs.microsoft.com/en-us/azure/machine-learning/component-reference/import-data

**Question: 14** CertyIQ

This question is included in a number of questions that depicts the identical set-up. However, every question has a distinctive result. Establish if the recommendation satisfies the requirements.
You are in the process of creating a machine learning model. Your dataset includes rows with null and missing values.
You plan to make use of the Clean Missing Data module in Azure Machine Learning Studio to detect and fix the null and missing values in the dataset.
Recommendation: You make use of the Replace with median option.
Will the requirements be satisfied?

A. Yes

B. No

**Answer: A**

**Explanation:**

This is an incomplete question. We don't know what type of data it is. Continuous or categorical. If it's continuous then it's A else its B, Clean Missing Data module has the option to replace with median

**Question: 15** CertyIQ

This question is included in a number of questions that depicts the identical set-up. However, every question has a distinctive result. Establish if the recommendation satisfies the requirements.
You are in the process of creating a machine learning model. Your dataset includes rows with null and missing values.
You plan to make use of the Clean Missing Data module in Azure Machine Learning Studio to detect and fix the null and missing values in the dataset.
Recommendation: You make use of the Custom substitution value option.
Will the requirements be satisfied?

A. Yes

B. No

**Answer: A**

**Explanation:**

Clean Missing Data module also provides "Custom substitution value" cleaning modebecause we can use '0'

for numeric and 'na' for text column.

**Question: 16** CertyIQ

This question is included in a number of questions that depicts the identical set-up. However, every question has a distinctive result. Establish if the recommendation satisfies the requirements.
You are in the process of creating a machine learning model. Your dataset includes rows with null and missing values.
You plan to make use of the Clean Missing Data module in Azure Machine Learning Studio to detect and fix the null and missing values in the dataset.
Recommendation: You make use of the Remove entire row option.
Will the requirements be satisfied?

A. Yes
B. No

**Answer: A**

**Explanation:**
Remove entire row: Completely removes any row in the dataset that has one or more missing values. This is useful if the missing value can be considered randomly missing.

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

**Question: 17** CertyIQ

You need to consider the underlined segment to establish whether it is accurate.
To transform a categorical feature into a binary indicator, you should make use of the Clean Missing Data module.
Select `No adjustment required` if the underlined segment is accurate. If the underlined segment is inaccurate, select the accurate option.

A. No adjustment required.
B. Convert to Indicator Values
C. Apply SQL Transformation
D. Group Categorical Values

**Answer: B**

**Explanation:**
Use the Convert to Indicator Values module in Azure Machine Learning Studio. The purpose of this module is to convert columns that contain categorical values into a series of binary indicator columns that can more easily be used as features in a machine learning model.

Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/convert-to-indicator-values

**Question: 18** CertyIQ

You need to consider the underlined segment to establish whether it is accurate.
To improve the amount of low incidence cases in a dataset, you should make use of the SMOTE module.

Select `No adjustment required` if the underlined segment is accurate. If the underlined segment is inaccurate, select the accurate option.

A. No adjustment required.
B. Remove Duplicate Rows
C. Join Data
D. Edit Metadata

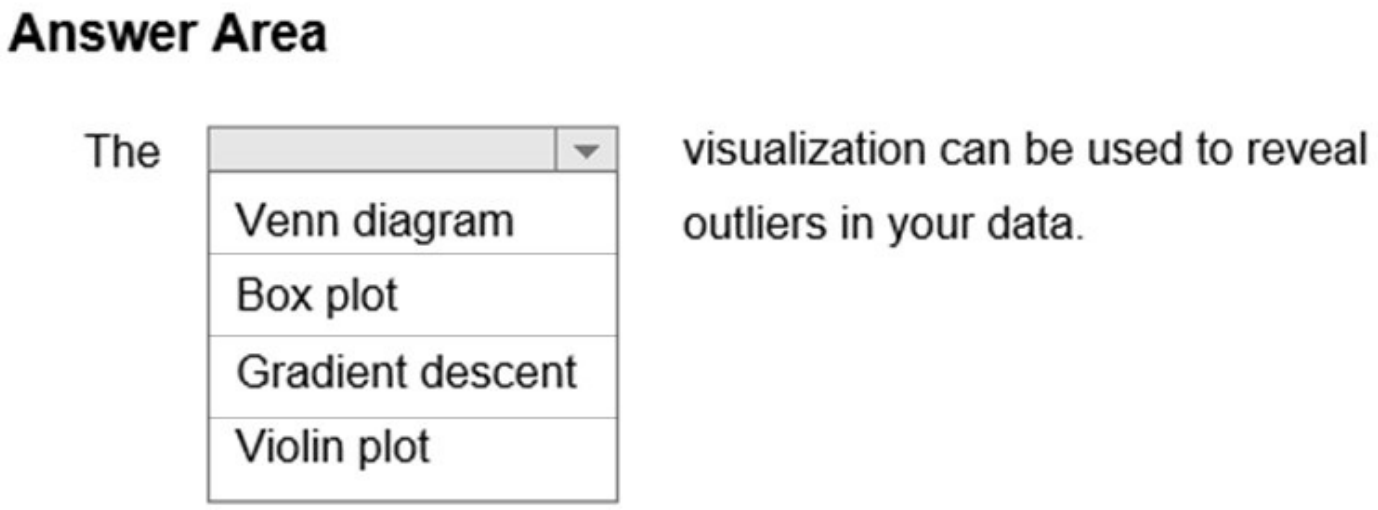
**Answer: A**

**Explanation:**
Use the SMOTE module in Azure Machine Learning Studio 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.

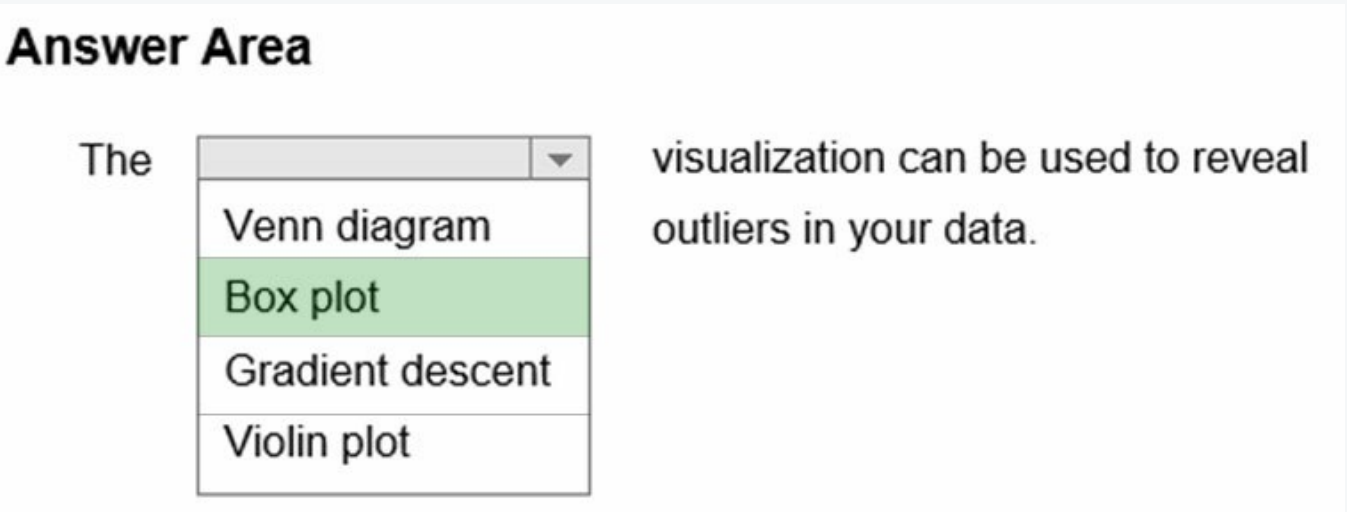
Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/smote

**Question: 19** CertyIQ

HOTSPOT -
You need to consider the underlined segment to establish whether it is accurate.
Hot Area:




**Answer:**



**Explanation:**
The box-plot algorithm can be used to display outliers.

Reference:
https://medium.com/analytics-vidhya/what-is-an-outliers-how-to-detect-and-remove-them-which-algorithm-are-sensitive-towards-outliers-2d501993d59

**Question: 20** CertyIQ

You are planning to host practical training to acquaint learners with data visualization creation using Python.
Learner devices are able to connect to the internet.
Learner devices are currently NOT configured for Python development. Also, learners are unable to install software on their devices as they lack administrator permissions. Furthermore, they are unable to access Azure subscriptions.
It is imperative that learners are able to execute Python-based data visualization code.
Which of the following actions should you take?

A. You should consider configuring the use of Azure Container Instance.
B. You should consider configuring the use of Azure BatchAI.
C. You should consider configuring the use of Azure Notebooks.
D. You should consider configuring the use of Azure Kubernetes Service.

**Answer: C**

**Explanation:**

Azure Notebooks are accessible by any device that can connect to the internet. The other options require installation of additional software, which is not permitted in this scenario.

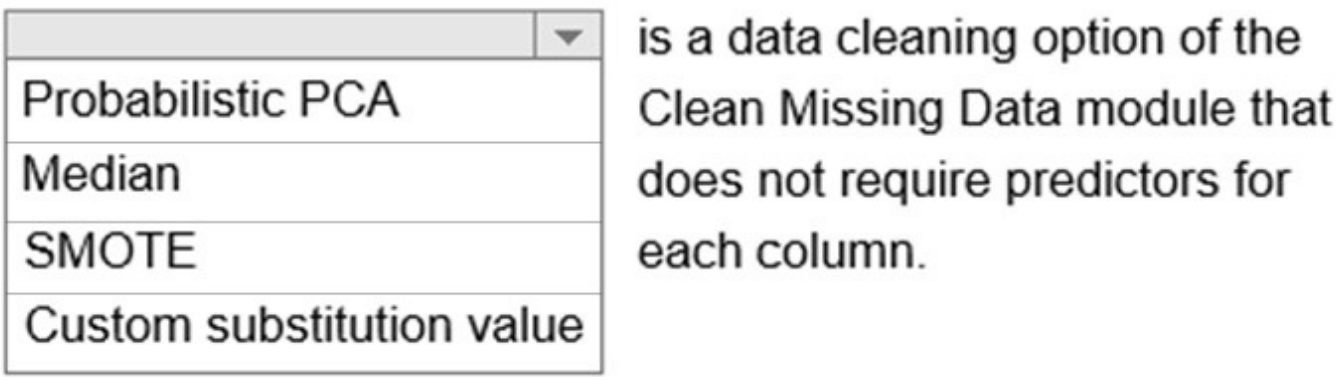
Reference:

https://notebooks.azure.com/

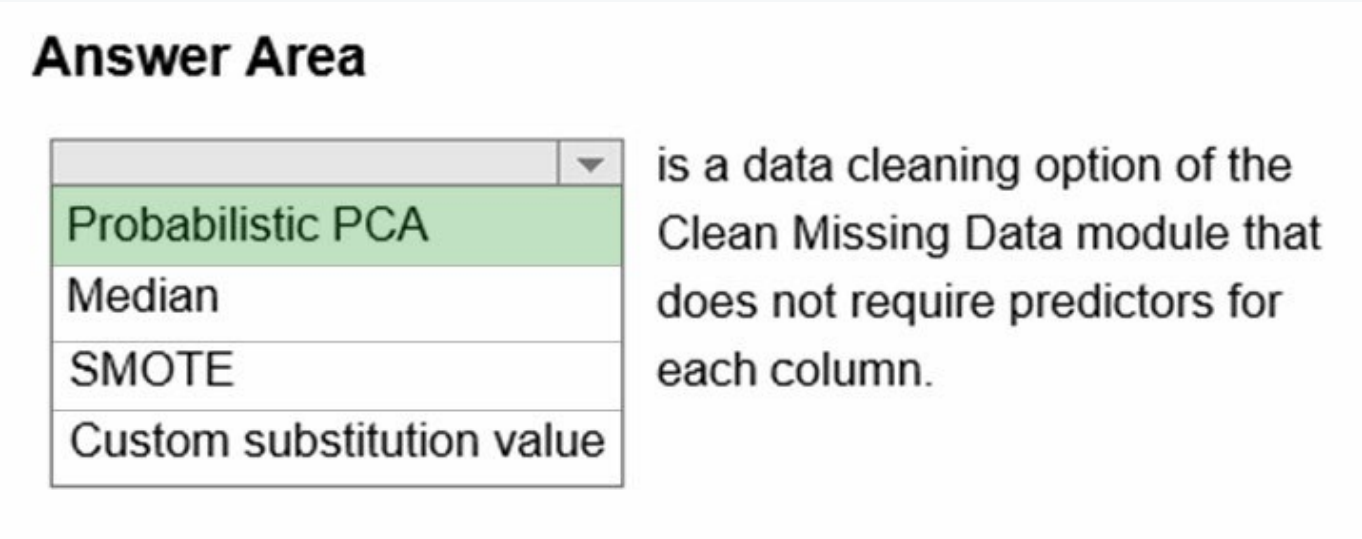
**Question: 21** CertyIQ

HOTSPOT -
Complete the sentence by selecting the correct option in the answer area.
Hot Area:

**Answer Area**



**Answer:**




**Explanation:**
Replace using Probabilistic PCA: Compared to other options, such as Multiple Imputation using Chained Equations (MICE), this option has the advantage of not requiring the application of predictors for each column. Instead, it approximates the covariance for the full dataset. Therefore, it might offer better performance for datasets that have missing values in many columns.

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

**Question: 22** CertyIQ

You have recently concluded the construction of a binary classification machine learning model.
You are currently assessing the model. You want to make use of a visualization that allows for precision to be used as the measurement for the assessment.
Which of the following actions should you take?

A. You should consider using Venn diagram visualization.
B. You should consider using Receiver Operating Characteristic (ROC) curve visualization.
C. You should consider using Box plot visualization.
D. You should consider using the Binary classification confusion matrix visualization.

**Answer: D**

**Explanation:**

Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-understand-automated-ml#confusion-matrix

**Question: 23** CertyIQ

This question is included in a number of questions that depicts the identical set-up. However, every question has a distinctive result. Establish if the recommendation satisfies the requirements.
You have been tasked with evaluating your model on a partial data sample via k-fold cross-validation.
You have already configured a k parameter as the number of splits. You now have to configure the k parameter for the cross-validation with the usual value choice.
Recommendation: You configure the use of the value k=1.

Will the requirements be satisfied?

A. Yes
B. No

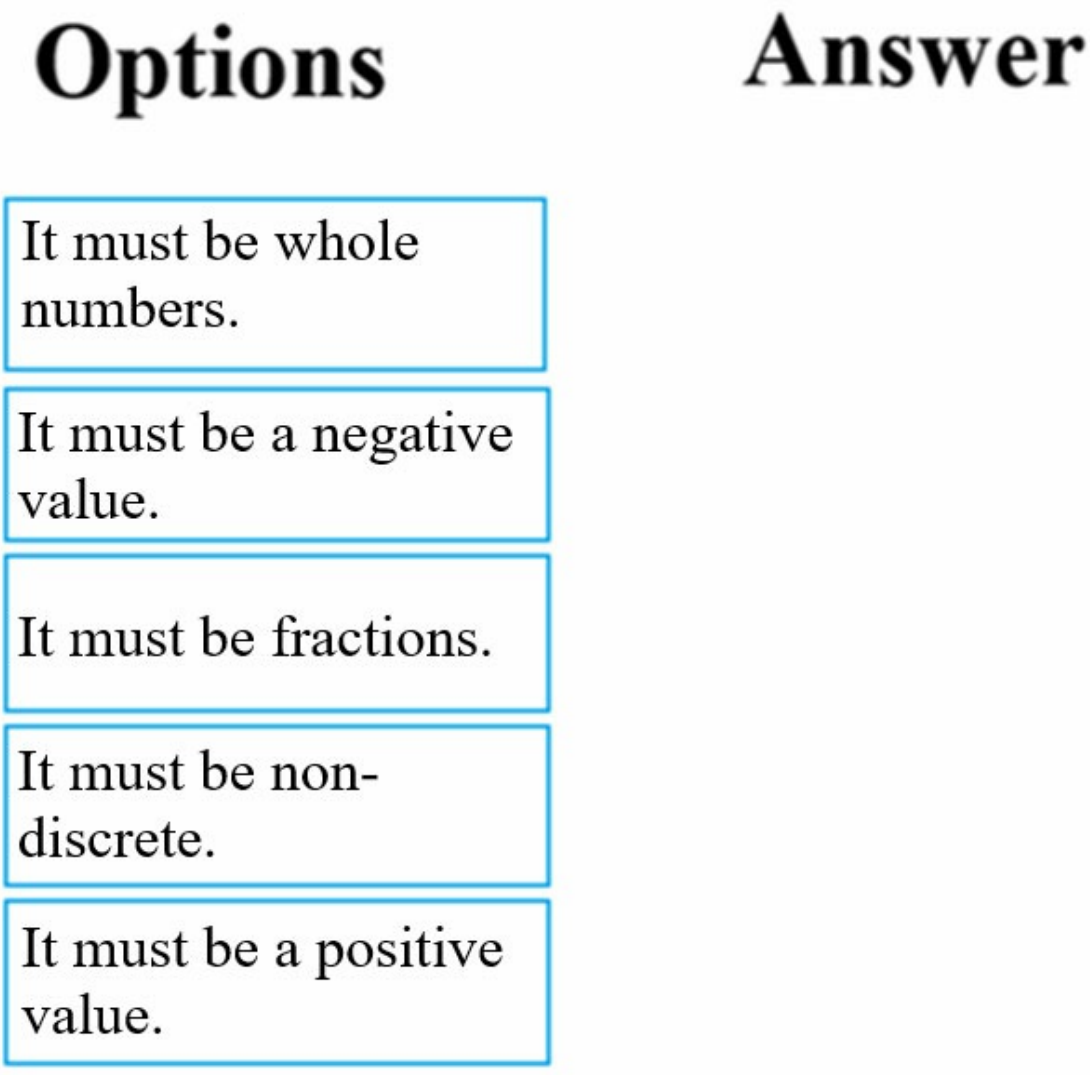
**Answer: B**

**Explanation:**

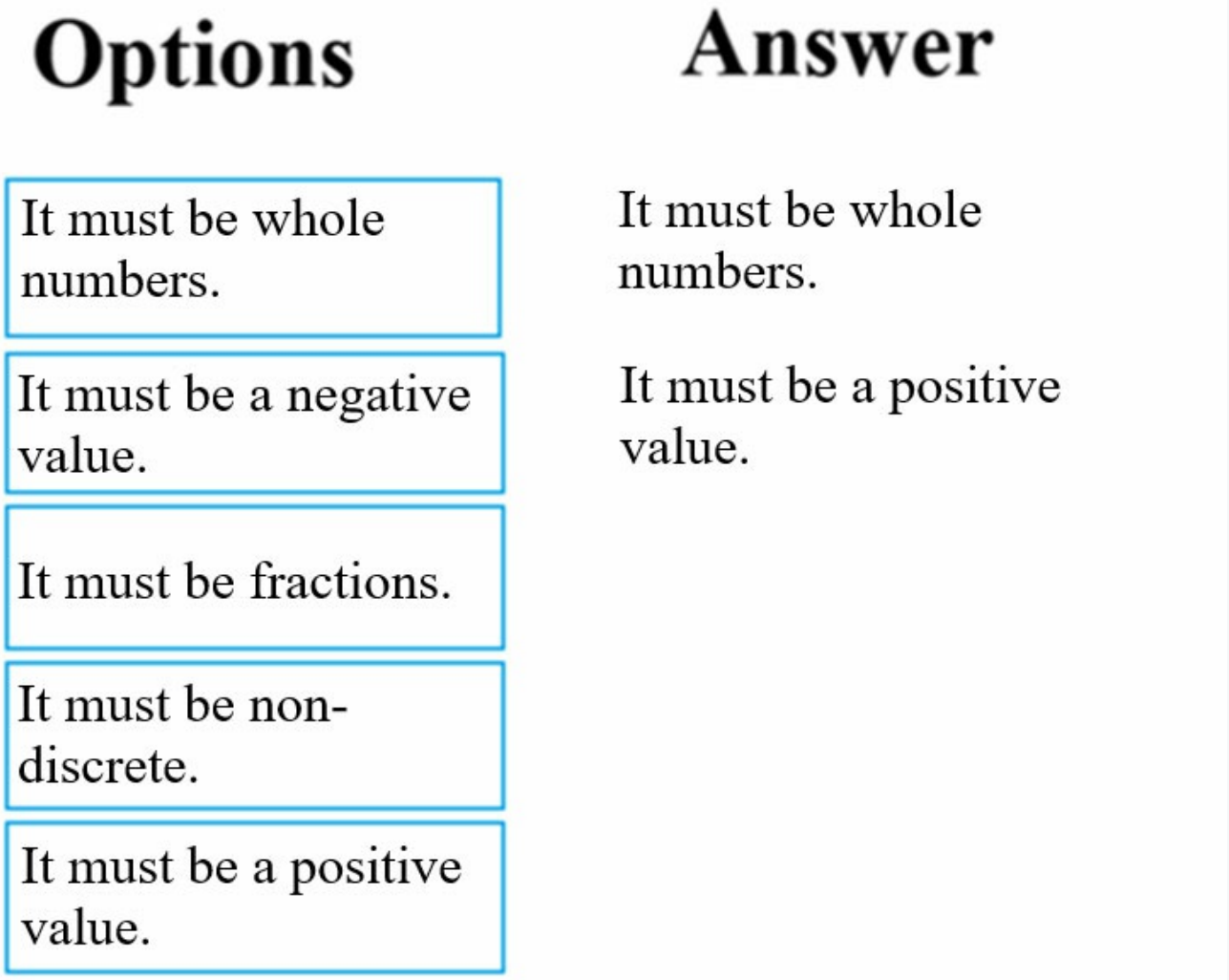
k can be 5 or 10More specifically, k MUST be within: 1 < k <= N

**Question: 24** CertyIQ

DRAG DROP -
You are in the process of constructing a regression model.
You would like to make it a Poisson regression model. To achieve your goal, the feature values need to meet certain conditions.
Which of the following are relevant conditions with regards to the label data? Answer by dragging the correct options from the list to the answer area.
Select and Place:




**Answer:**




**Explanation:**
Poisson regression is intended for use in regression models that are used to predict numeric values, typically counts. Therefore, you should use this module to create your regression model only if the values you are trying to predict fit the following conditions:
✑ The response variable has a Poisson distribution.
✑ Counts cannot be negative. The method will fail outright if you attempt to use it with negative labels.
✑ A Poisson distribution is a discrete distribution; therefore, it is not meaningful to use this method with non-whole numbers.

Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/poisson-regression

**Question: 25** CertyIQ

This question is included in a number of questions that depicts the identical set-up. However, every question has a distinctive result. Establish if the recommendation satisfies the requirements.
You are in the process of carrying out feature engineering on a dataset.
You want to add a feature to the dataset and fill the column value.
Recommendation: You must make use of the Group Categorical Values Azure Machine Learning Studio module.
Will the requirements be satisfied?

A. Yes
B. No

**Answer: B**

**Explanation:**

The typical use for grouping categorical values is to merge multiple string values into a single new level.

**Question: 26** CertyIQ

This question is included in a number of questions that depicts the identical set-up. However, every question has a distinctive result. Establish if the recommendation satisfies the requirements.
You are in the process of carrying out feature engineering on a dataset.
You want to add a feature to the dataset and fill the column value.
Recommendation: You must make use of the Join Data Azure Machine Learning Studio module.
Will the requirements be satisfied?

A. Yes
B. No

**Answer: B**

**Explanation:**

Add Columns need to be used. Join data is needed only for database style joins join is used for datasets not for columns b is correct

Reference:

https://docs.microsoft.com/bs-cyrl-ba/azure/machine-learning/component-reference/add-columns

https://docs.microsoft.com/bs-cyrl-ba/azure/machine-learning/component-reference/join-data

**Question: 27** CertyIQ

This question is included in a number of questions that depicts the identical set-up. However, every question has a distinctive result. Establish if the recommendation satisfies the requirements.
You are in the process of carrying out feature engineering on a dataset.
You want to add a feature to the dataset and fill the column value.
Recommendation: You must make use of the Edit Metadata Azure Machine Learning Studio module.
Will the requirements be satisfied?

A. Yes
B. No

**Answer: B**

**Explanation:**

Edit meta data cannot add a new column, it can change properties of existing column

**Question: 28** CertyIQ

You have been tasked with ascertaining if two sets of data differ considerably. You will make use of Azure Machine Learning Studio to complete your task.

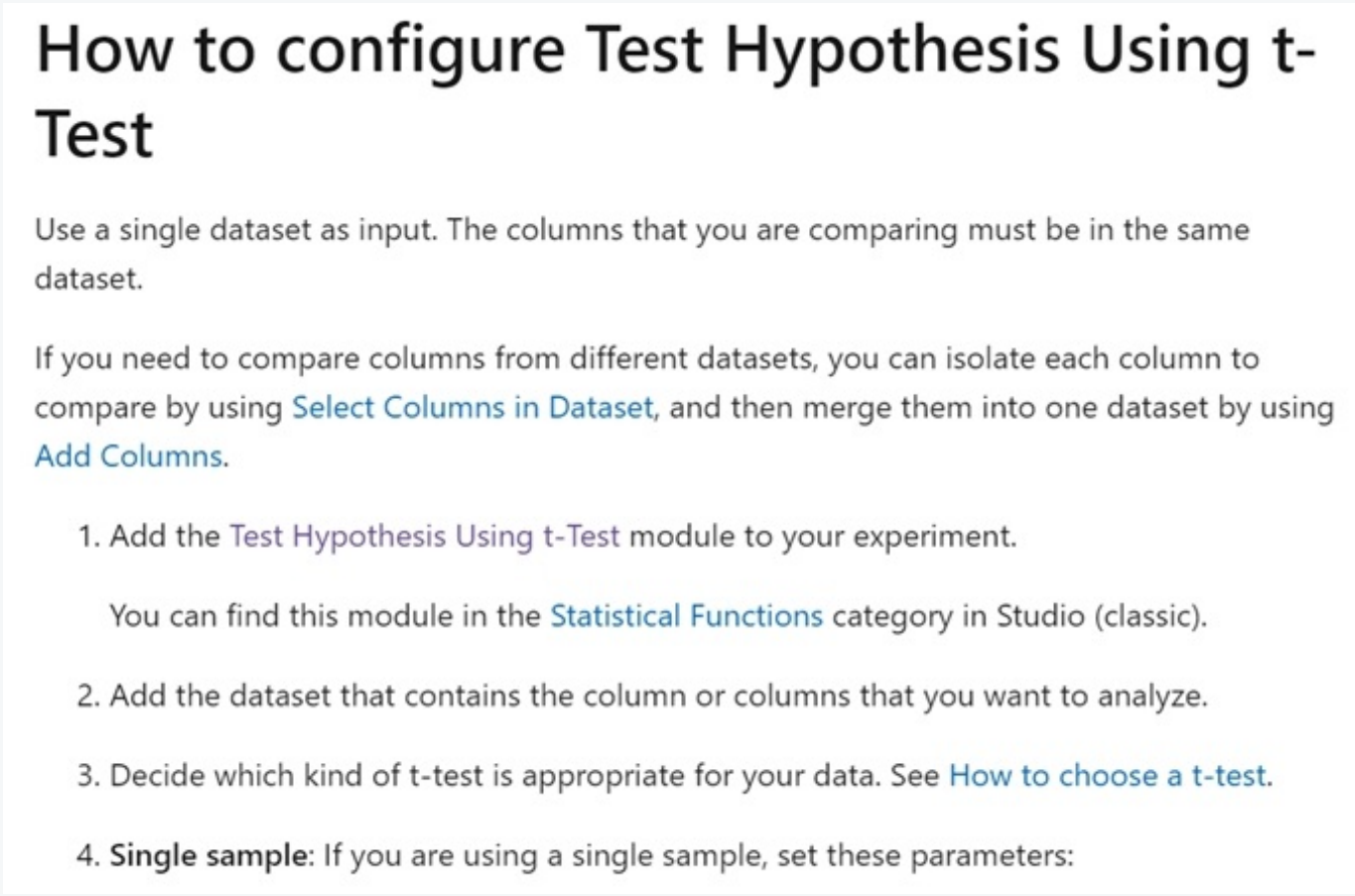
You plan to perform a paired t-test.
Which of the following are conditions that must apply to use a paired t-test? (Choose all that apply.)

A. All scores are independent from each other.
B. You have a matched pairs of scores.
C. The sampling distribution of d is normal.
D. The sampling distribution of x1- x2 is normal.

**Answer: BC**

**Explanation:**

Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/test-hypothesis-using-t-test
" target="_blank" style="word-break: break-all;">



# How to configure Test Hypothesis Using t-Test

Use a single dataset as input. The columns that you are comparing must be in the same dataset.

If you need to compare columns from different datasets, you can isolate each column to compare by using Select Columns in Dataset, and then merge them into one dataset by using Add Columns.

1. Add the Test Hypothesis Using t-Test module to your experiment.

   You can find this module in the Statistical Functions category in Studio (classic).

2. Add the dataset that contains the column or columns that you want to analyze.

3. Decide which kind of t-test is appropriate for your data. See How to choose a t-test.

4. **Single sample**: If you are using a single sample, set these parameters:

**Question: 29** CertyIQ

You want to train a classification model using data located in a comma-separated values (CSV) file.
The classification model will be trained via the Automated Machine Learning interface using the Classification task type.
You have been informed that only linear models need to be assessed by the Automated Machine Learning.
Which of the following actions should you take?

A. You should disable deep learning.
B. You should enable automatic featurization.
C. You should disable automatic featurization.
D. You should set the task type to Forecasting.

**Answer: A**

**Explanation:**

Disabling automatic featurization does not cause the models evaluated to be linear only. Automatic featurization has the following steps listed how-to-configure-auto-features#automatic-featurization The only way to force linear algorithms to be evaluated is to use the blocked algorithms. list.

**Question: 30** CertyIQ

You are preparing to train a regression model via automated machine learning. The data available to you has features with missing values, as well as categorical features with little discrete values.
You want to make sure that automated machine learning is configured as follows:
- missing values must be automatically imputed.
- categorical features must be encoded as part of the training task.

Which of the following actions should you take?

A. You should make use of the featurization parameter with the 'auto' value pair.
B. You should make use of the featurization parameter with the 'off' value pair.
C. You should make use of the featurization parameter with the 'on' value pair.
D. You should make use of the featurization parameter with the 'FeaturizationConfig' value pair.

**Answer: A**

**Explanation:**

Featurization str or FeaturizationConfig

Values: 'auto' / 'off' / FeaturizationConfig

Indicator for whether featurization step should be done automatically or not, or whether customized featurization should be used.

Column type is automatically detected. Based on the detected column type preprocessing/featurization is done as follows:

Categorical: Target encoding, one hot encoding, drop high cardinality categories, impute missing values.

Numeric: Impute missing values, cluster distance, weight of evidence.

DateTime: Several features such as day, seconds, minutes, hours etc.

Text: Bag of words, pre-trained Word embedding, text target encoding.

# Constructor

Python Copy

```
AutoMLConfig(task: str, path: typing.Union[str, NoneType] = None, iterations:
typing.Union[int, NoneType] = None, primary_metric: typing.Union[str, NoneType] =
None, positive_label: typing.Union[typing.Any, NoneType] = None, compute_target:
typing.Union[typing.Any, NoneType] = None, spark_context:
typing.Union[typing.Any, NoneType] = None, X: typing.Union[typing.Any, NoneType]
= None, y: typing.Union[typing.Any, NoneType] = None, sample_weight:
typing.Union[typing.Any, NoneType] = None, X_valid: typing.Union[typing.Any,
NoneType] = None, y_valid: typing.Union[typing.Any, NoneType] = None,
sample_weight_valid: typing.Union[typing.Any, NoneType] = None,
cv_splits_indices: typing.Union[typing.List[typing.List[typing.Any]], NoneType] =
None, validation_size: typing.Union[float, NoneType] = None, n_cross_validations:
typing.Union[int, NoneType] = None, y_min: typing.Union[float, NoneType] = None,
y_max: typing.Union[float, NoneType] = None, num_classes: typing.Union[int,
NoneType] = None, featurization: typing.Union[str,
azureml.automl.core.featurization.featurizationconfig.FeaturizationConfig] =
'auto', max_cores_per_iteration: int = 1, max_concurrent_iterations: int = 1,
iteration_timeout_minutes: typing.Union[int, NoneType] = None, mem_in_mb:
typing.Union[int, NoneType] = None, enforce_time_on_windows: bool = True
```

Reference:

https://docs.microsoft.com/en-us/python/api/azureml-train-automl-

**Question: 31** CertyIQ

You make use of Azure Machine Learning Studio to develop a linear regression model. You perform an experiment to assess various algorithms.
Which of the following is an algorithm that reduces the variances between actual and predicted values?

A. Fast Forest Quantile Regression
B. Poisson Regression
C. Boosted Decision Tree Regression
D. Linear Regression

**Answer: D**

**Explanation:**

Linear regression minimizes the sum of squares, i.e. it minimizes S=Sum[(y_i-f_i)^2, i,1,n ], where y_i is the actual value and f_i the predicted value (we can see it as the average).

Since the variance of y_i is S/n, minimizing S is equivalent to minimizing the variance.

**Question: 32** CertyIQ

This question is included in a number of questions that depicts the identical set-up. However, every question has a distinctive result. Establish if the recommendation satisfies the requirements.

You have been tasked with constructing a machine learning model that translates language text into a different language text.
The machine learning model must be constructed and trained to learn the sequence of the.
Recommendation: You make use of Convolutional Neural Networks (CNNs).
Will the requirements be satisfied?

A. Yes
B. No

**Answer: B**

**Explanation:**

Use Reccurent Neural Network (RNN) for translations.

**Question: 33** CertyIQ

This question is included in a number of questions that depicts the identical set-up. However, every question has a distinctive result. Establish if the recommendation satisfies the requirements.
You have been tasked with constructing a machine learning model that translates language text into a different language text.
The machine learning model must be constructed and trained to learn the sequence of the.
Recommendation: You make use of Generative Adversarial Networks (GANs).
Will the requirements be satisfied?

A. Yes
B. No

**Answer: B**

**Explanation:**

GANs used for image translation

**Question: 34** CertyIQ

This question is included in a number of questions that depicts the identical set-up. However, every question has a distinctive result. Establish if the recommendation satisfies the requirements.
You have been tasked with constructing a machine learning model that translates language text into a different language text.
The machine learning model must be constructed and trained to learn the sequence of the.
Recommendation: You make use of Recurrent Neural Networks (RNNs).
Will the requirements be satisfied?

A. Yes
B. No

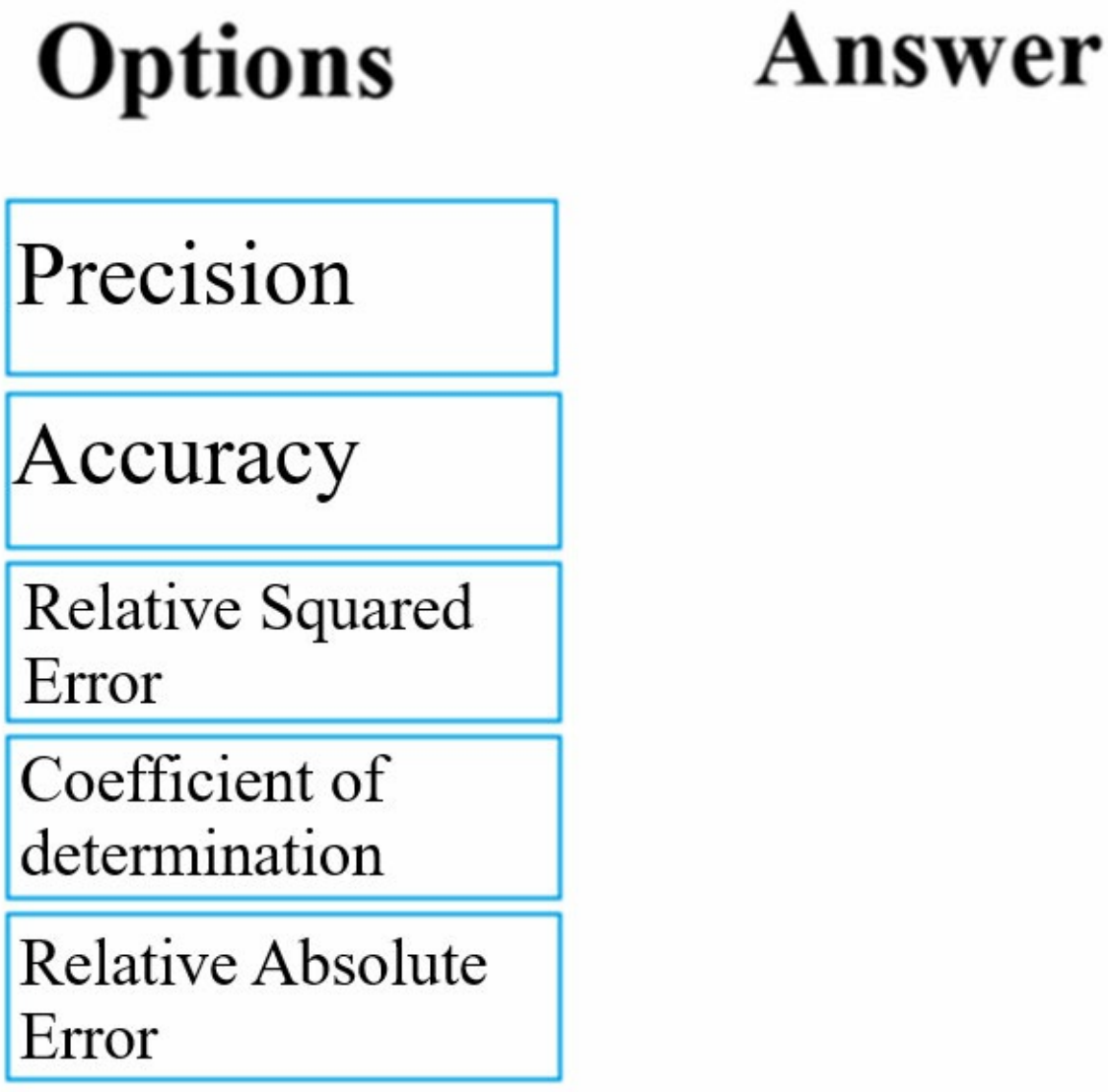
**Answer: A**

**Explanation:**
Note: RNNs are designed to take sequences of text as inputs or return sequences of text as outputs, or both.
They're called recurrent because the network's hidden layers have a loop in which the output and cell state from each time step become inputs at the next time step. This recurrence serves as a form of memory.
It allows contextual information to flow through the network so that relevant outputs from previous time steps can be applied to network operations at the current time step.

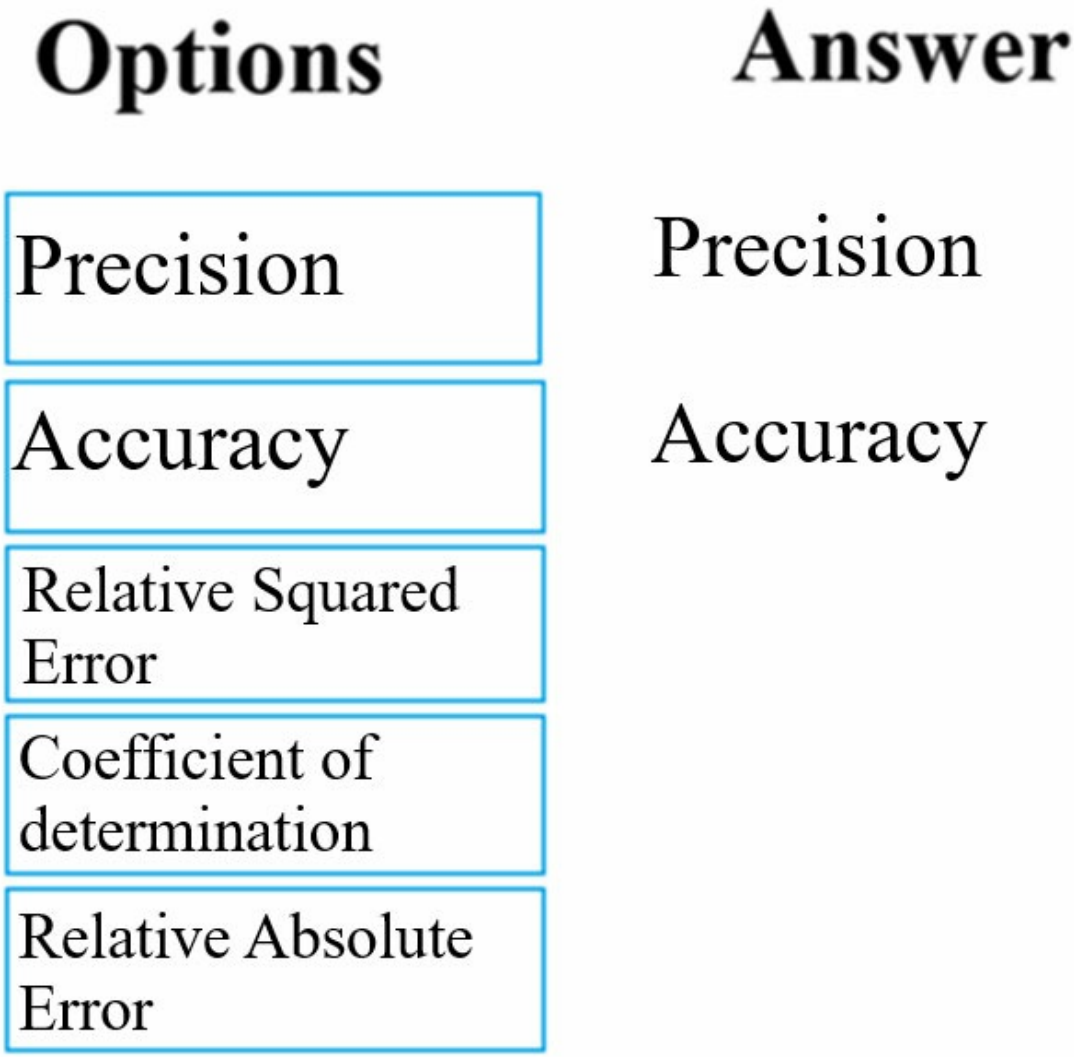
Reference:
https://towardsdatascience.com/language-translation-with-rnns-d84d43b40571

**Question: 35** CertyIQ

DRAG DROP -
You have been tasked with evaluating the performance of a binary classification model that you created.
You need to choose evaluation metrics to achieve your goal.
Which of the following are the metrics you would choose? Answer by dragging the correct options from the list to the answer area.
Select and Place:




**Answer:**




**Explanation:**
The evaluation metrics available for binary classification models are: Accuracy, Precision, Recall, F1 Score, and AUC.
Note: A very natural question is: 'Out of the individuals whom the model, how many were classified correctly (TP)?'
This question can be answered by looking at the Precision of the model, which is the proportion of positives that are classified correctly.

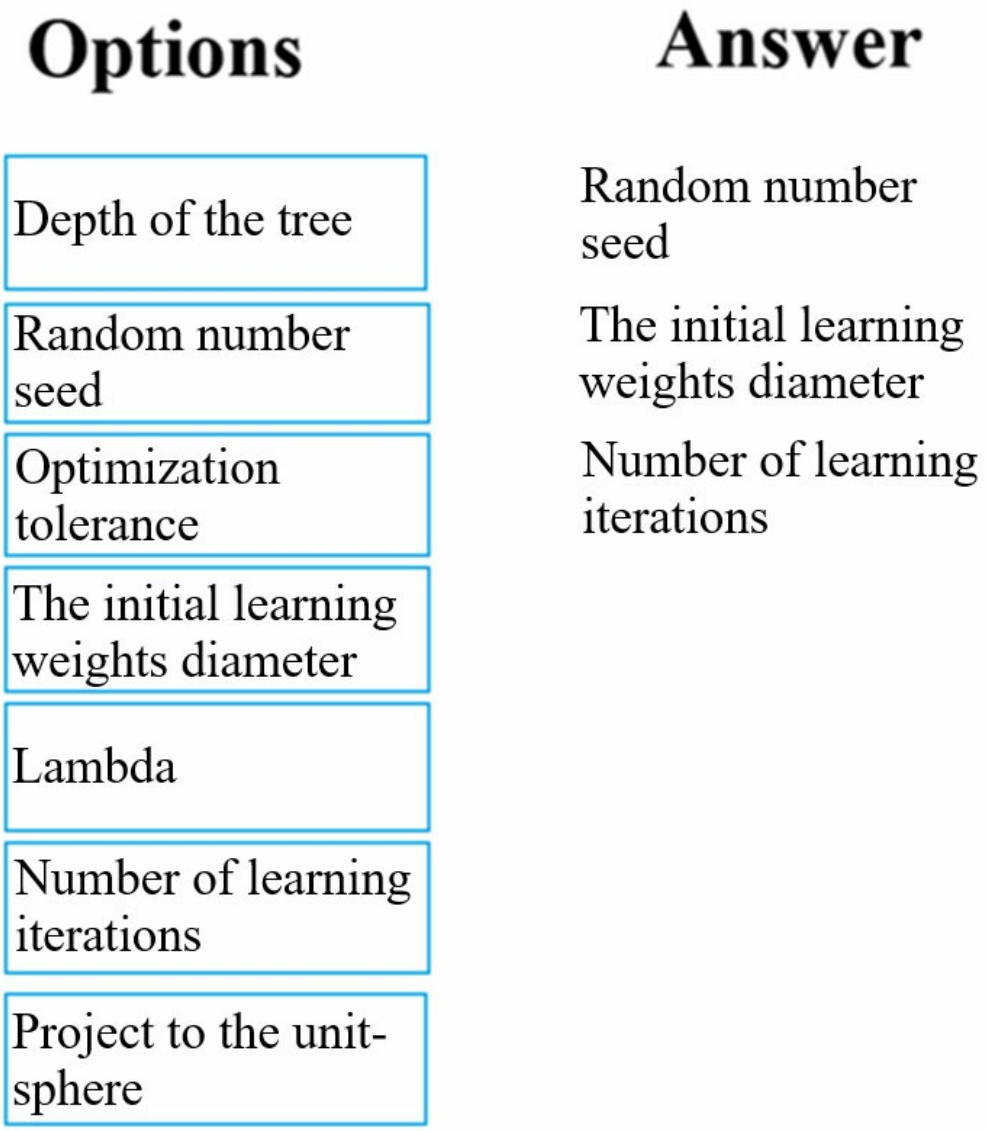
Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/studio/evaluate-model-performance

**Question: 36** CertyIQ

DRAG DROP -
You build a binary classification model using the Azure Machine Learning Studio Two-Class Neural Network module.
You are preparing to configure the Tune Model Hyperparameters module for the purpose of tuning accuracy for the model.
Which of the following are valid parameters for the Two-Class Neural Network module? Answer by dragging the correct options from the list to the answer area.
Select and Place:

| Options | Answer |
| --- | --- |
| Depth of the tree | |
| Random number seed | |
| Optimization tolerance | |
| The initial learning weights diameter | |
| Lambda | |
| Number of learning iterations | |
| Project to the unit-sphere | |

**Answer:**




**Explanation:**

Random seed is a parameter for binary classification, but I do not understand "Hyperparameters"

Reference

https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/two-class-neural-network

**Question: 37** CertyIQ

You make use of Azure Machine Learning Studio to create a binary classification model.
You are preparing to carry out a parameter sweep of the model to tune hyperparameters. You have to make sure that the sweep allows for every possible combination of hyperparameters to be iterated. Also, the computing resources needed to carry out the sweep must be reduced.
Which of the following actions should you take?

A. You should consider making use of the Selective grid sweep mode.
B. You should consider making use of the Measured grid sweep mode.
C. You should consider making use of the Entire grid sweep mode.
D. You should consider making use of the Random grid sweep mode.

**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.
C: Entire grid: When you select this option, the module loops over a grid predefined by the system, to try different combinations and identify the best learner. This option is useful for cases where you don't know what the best parameter settings might be and want to try all possible combination of values.

Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/tune-model-hyperparameters

**Question: 38** CertyIQ

You are in the process of constructing a deep convolutional neural network (CNN). The CNN will be used for image classification.
You notice that the CNN model you constructed displays hints of overfitting.
You want to make sure that overfitting is minimized, and that the model is converged to an optimal fit.
Which of the following is TRUE with regards to achieving your goal?

A. You have to add an additional dense layer with 512 input units, and reduce the amount of training data.
B. You have to add L1/L2 regularization, and reduce the amount of training data.
C. You have to reduce the amount of training data and make use of training data augmentation.
D. You have to add L1/L2 regularization, and make use of training data augmentation.
E. You have to add an additional dense layer with 512 input units, and add L1/L2 regularization.

**Answer: D**

**Explanation:**

Discarded because that is increasing the complexity of the architecture. B, C are suggesting reducing the amount of data. D will generate more data for the CNN to be able to generalize more.

**Question: 39** CertyIQ

This question is included in a number of questions that depicts the identical set-up. However, every question has a distinctive result. Establish if the recommendation satisfies the requirements.
You are planning to make use of Azure Machine Learning designer to train models.
You need choose a suitable compute type.
Recommendation: You choose Attached compute.
Will the requirements be satisfied?

A. Yes
B. No

**Answer: A**

**Explanation:**

Because we can use databricks or vm as attached compute for training purposes

Reference:

https://docs.microsoft.com/en-us/azure/machine-learning/how-to-create-attach-compute-studio

**Question: 40** CertyIQ

This question is included in a number of questions that depicts the identical set-up. However, every question has a distinctive result. Establish if the recommendation satisfies the requirements.
You are planning to make use of Azure Machine Learning designer to train models.
You need choose a suitable compute type.
Recommendation: You choose Inference cluster.
Will the requirements be satisfied?

A. Yes
B. No

**Answer: B**

**Explanation:**

Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-create-attach-compute-studio

**Question: 41** CertyIQ

This question is included in a number of questions that depicts the identical set-up. However, every question has a distinctive result. Establish if the recommendation satisfies the requirements.
You are planning to make use of Azure Machine Learning designer to train models.
You need choose a suitable compute type.
Recommendation: You choose Compute cluster.
Will the requirements be satisfied?

A. Yes
B. No

**Answer: A**

**Explanation:**

Says even Compute Instance is possible Answer

Reference:

https://docs.microsoft.com/en-us/azure/machine-learning/how-to-create-attach-compute-studio

**Question: 42** CertyIQ

You are making use of the Azure Machine Learning to designer construct an experiment.
After dividing a dataset into training and testing sets, you configure the algorithm to be Two-Class Boosted Decision Tree.

You are preparing to ascertain the Area Under the Curve (AUC).
Which of the following is a sequential combination of the models required to achieve your goal?

A. Train, Score, Evaluate.
B. Score, Evaluate, Train.
C. Evaluate, Export Data, Train.
D. Train, Score, Export Data.

**Answer: A**

**Explanation:**

Train,score,evaluate

**Question: 43** CertyIQ

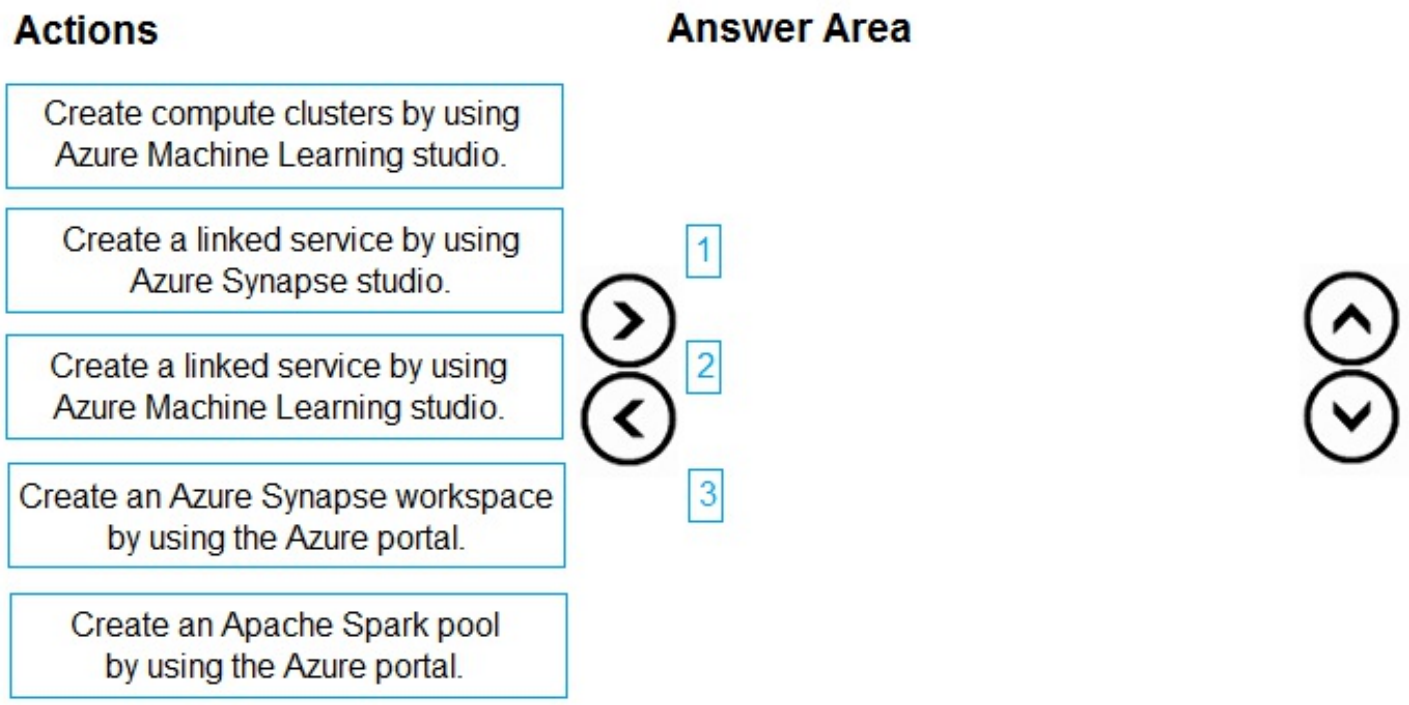
DRAG DROP
-

You create an Azure Machine Learning workspace.

You must implement dedicated compute for model training in the workspace by using Azure Synapse compute resources. The solution must attach the dedicated compute and start an Azure Synapse session.

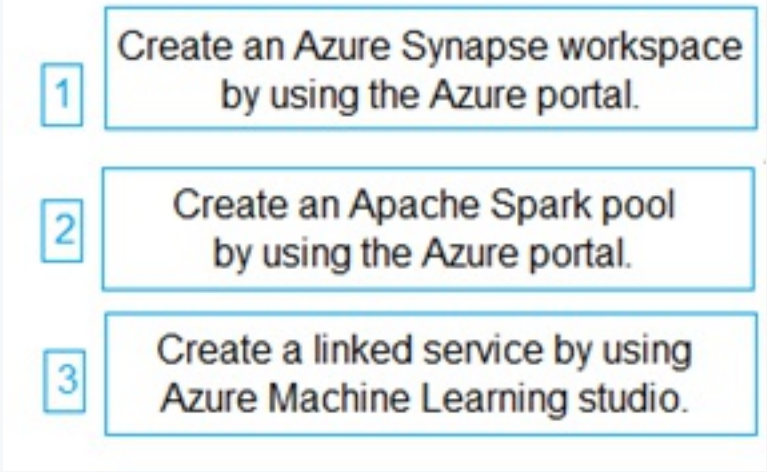
You need to implement the computer resources.

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.




**Answer:**

**Answer Area**




**Question: 44** CertyIQ

You create an Azure Machine Learning workspace. You are preparing a local Python environment on a laptop computer. You want to use the laptop to connect to the workspace and run experiments.
You create the following config.json file.

```
{
  "workspace_name" : "ml-workspace"
}
```

You must use the Azure Machine Learning SDK to interact with data and experiments in the workspace.
You need to configure the config.json file to connect to the workspace from the Python environment.
Which two additional parameters must you add to the config.json file in order to connect to the workspace? Each correct answer presents part of the solution.
NOTE: Each correct selection is worth one point.

A. login
B. resource_group
C. subscription_id
D. key
E. region

**Answer: BC**

**Explanation:**
To use the same workspace in multiple environments, create a JSON configuration file. The configuration file saves your subscription (subscription_id), resource
(resource_group), and workspace name so that it can be easily loaded.
The following sample shows how to create a workspace.
from azureml.core import Workspace
ws = Workspace.create(name='myworkspace',
subscription_id='<azure-subscription-id>',
resource_group='myresourcegroup',
create_resource_group=True,
location='eastus2'
)

Reference:
https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.core.workspace.workspace

**Question: 45** CertyIQ

You create a batch inference pipeline by using the Azure ML SDK. You configure the pipeline parameters by executing the following code:

```
from azureml.contrib.pipeline.steps import ParallelRunConfig
parallel_run_config = ParallelRunConfig(
      source_directory=scripts_folder,
      entry_script= "batch_pipeline.py",
      mini_batch_size= "5",
      error_threshold=10,
      output_action= "append_row",
      environment=batch_env,
      compute_target=compute_target,
      logging_level= "DEBUG",
      node_count=4)
```

You need to obtain the output from the pipeline execution.
Where will you find the output?

A. the digit_identification.py script
B. the debug log
C. the Activity Log in the Azure portal for the Machine Learning workspace
D. the Inference Clusters tab in Machine Learning studio
E. a file named parallel_run_step.txt located in the output folder

**Answer: E**

**Explanation:**
output_action (str): How the output is to be organized. Currently supported values are 'append_row' and 'summary_only'.
✑ 'append_row' " All values output by run() method invocations will be aggregated into one unique file named parallel_run_step.txt that is created in the output location.
'summary_only'
▪

Reference:
https://docs.microsoft.com/en-us/python/api/azureml-contrib-pipeline-steps/azureml.contrib.pipeline.steps.parallelrunconfig

**Question: 46** CertyIQ

You write a Python script that processes data in a comma-separated values (CSV) file.
You plan to run this script as an Azure Machine Learning experiment.
The script loads the data and determines the number of rows it contains using the following code:

```
from azureml.core import Run
import pandas as pd

run = Run.get_context()
data = pd.read_csv('./data.csv')
rows = (len(data))
# record row_count metric here
...
```

You need to record the row count as a metric named row_count that can be returned using the get_metrics method of the Run object after the experiment run completes.
Which code should you use?

A. run.upload_file('row_count', './data.csv')
B. run.log('row_count', rows)
C. run.tag('row_count', rows)
D. run.log_table('row_count', rows)
E. run.log_row('row_count', rows)

**Answer: B**

**Explanation:**
Log a numerical or string value to the run with the given name using log(name, value, description=''). Logging a metric to a run causes that metric to be stored in the run record in the experiment. You can log the same metric multiple times within a run, the result being considered a vector of that metric.
Example: run.log("accuracy", 0.95)
Incorrect Answers:
E: Using log_row(name, description=None, **kwargs) creates a metric with multiple columns as described in kwargs. Each named parameter generates a column with the value specified. log_row can be called once to log an arbitrary tuple, or multiple times in a loop to generate a complete table.
Example: run.log_row("Y over X", x=1, y=0.4)

Reference:
https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.core.run

**Question: 47** CertyIQ

You are developing a hands-on workshop to introduce Docker for Windows to attendees.
You need to ensure that workshop attendees can install Docker on their devices.
Which two prerequisite components should attendees install on the devices? Each correct answer presents part of the solution.
NOTE: Each correct selection is worth one point.

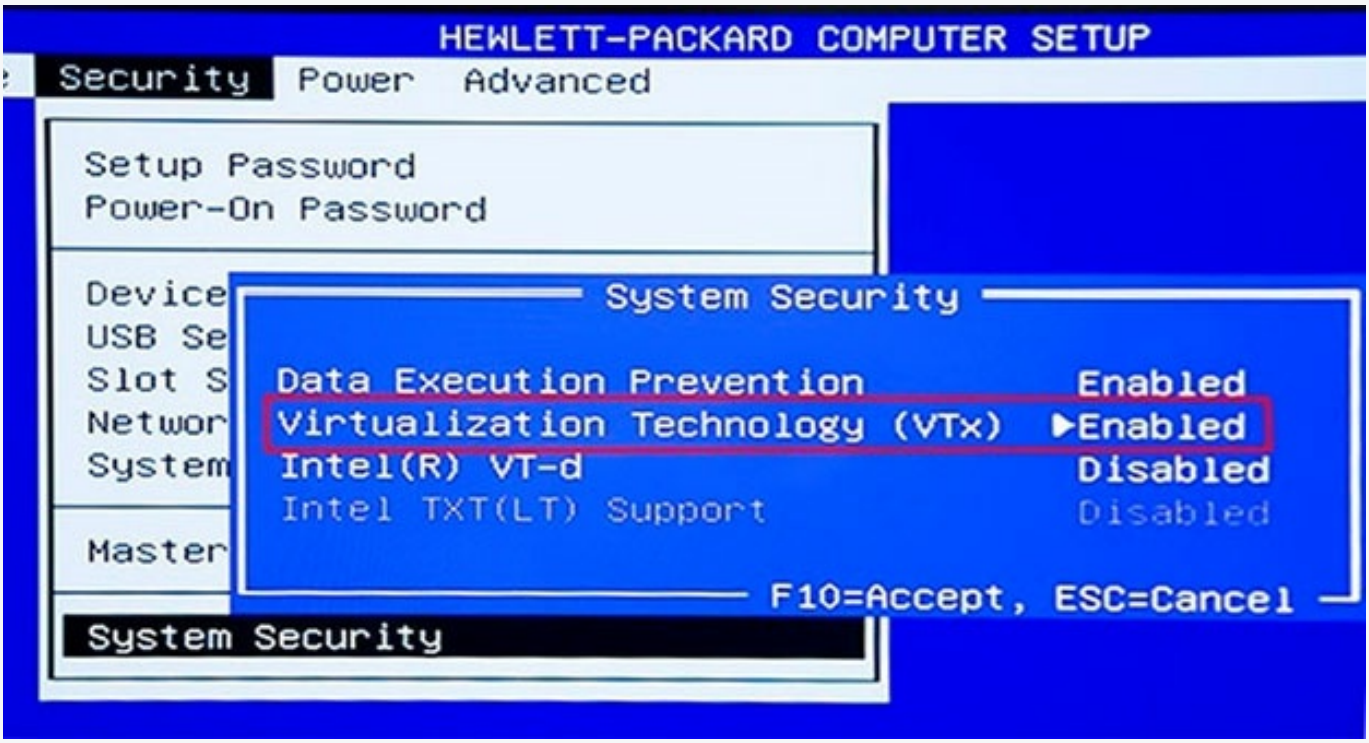
A. Microsoft Hardware-Assisted Virtualization Detection Tool
B. Kitematic
C. BIOS-enabled virtualization
D. VirtualBox
E. Windows 10 64-bit Professional

**Answer: CE**

**Explanation:**

C: Make sure your Windows system supports Hardware Virtualization Technology and that virtualization is enabled.

Ensure that hardware virtualization support is turned on in the BIOS settings. For example:




E: To run Docker, your machine must have a 64-bit operating system running Windows 7 or higher.

Reference:
https://docs.docker.com/toolbox/toolbox_install_windows/
https://blogs.technet.microsoft.com/canitpro/2015/09/08/step-by-step-enabling-hyper-v-for-use-on-windows-10/

**Question: 48** CertyIQ

Your team is building a data engineering and data science development environment.
The environment must support the following requirements:
- support Python and Scala
- compose data storage, movement, and processing services into automated data pipelines
- the same tool should be used for the orchestration of both data engineering and data science
- support workload isolation and interactive workloads
- enable scaling across a cluster of machines

You need to create the environment.
What should you do?

A. Build the environment in Apache Hive for HDInsight and use Azure Data Factory for orchestration.
B. Build the environment in Azure Databricks and use Azure Data Factory for orchestration.
C. Build the environment in Apache Spark for HDInsight and use Azure Container Instances for orchestration.
D. Build the environment in Azure Databricks and use Azure Container Instances for orchestration.

**Answer: B**

**Explanation:**

In Azure Databricks, we can create two different types of clusters.
- Standard, these are the default clusters and can be used with Python, R, Scala and SQL
- High-concurrency

Azure Databricks is fully integrated with Azure Data Factory.

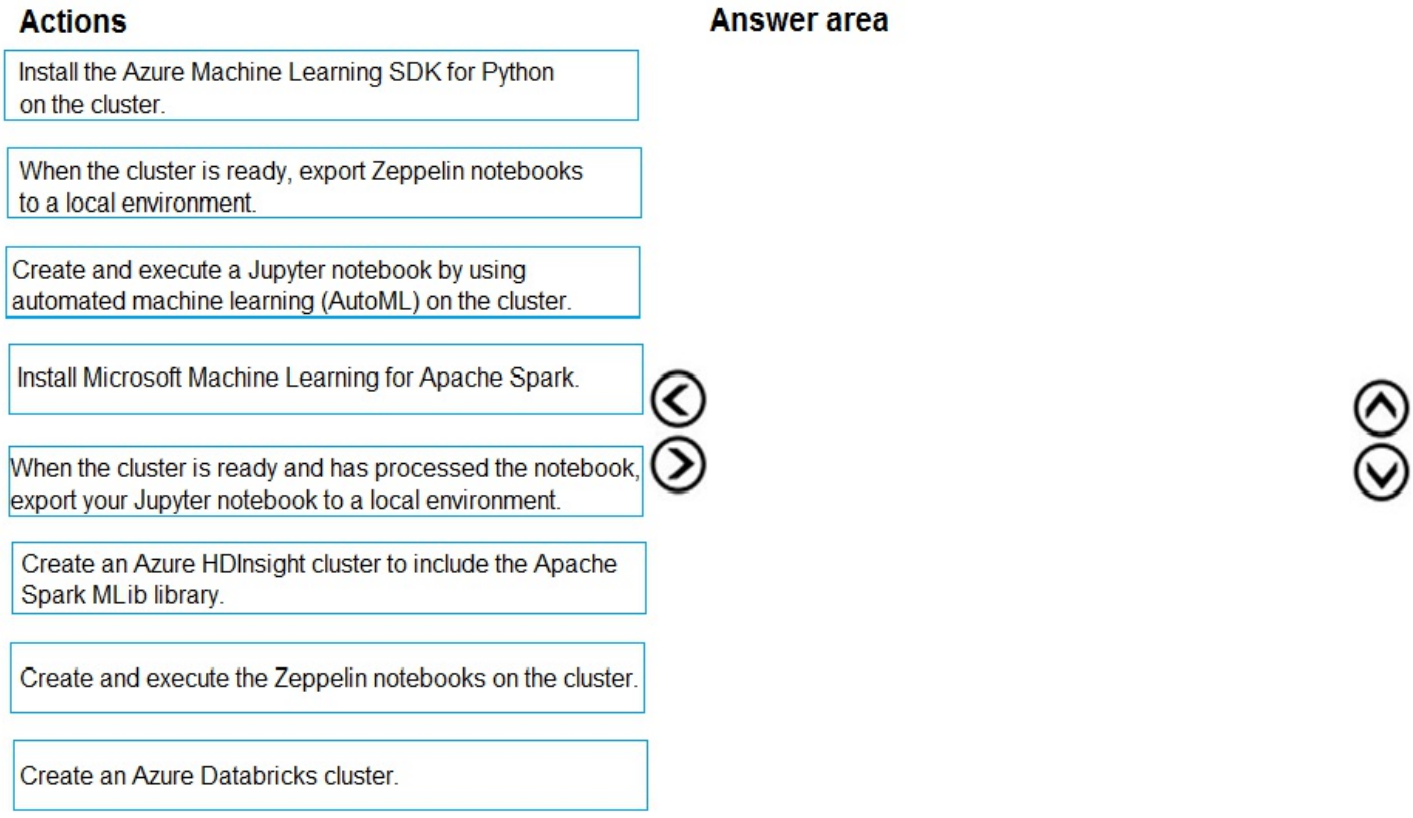
Incorrect Answers:
D: Azure Container Instances is good for development or testing. Not suitable for production workloads.

Reference:
https://docs.microsoft.com/en-us/azure/architecture/data-guide/technology-choices/data-science-and-machine-learning

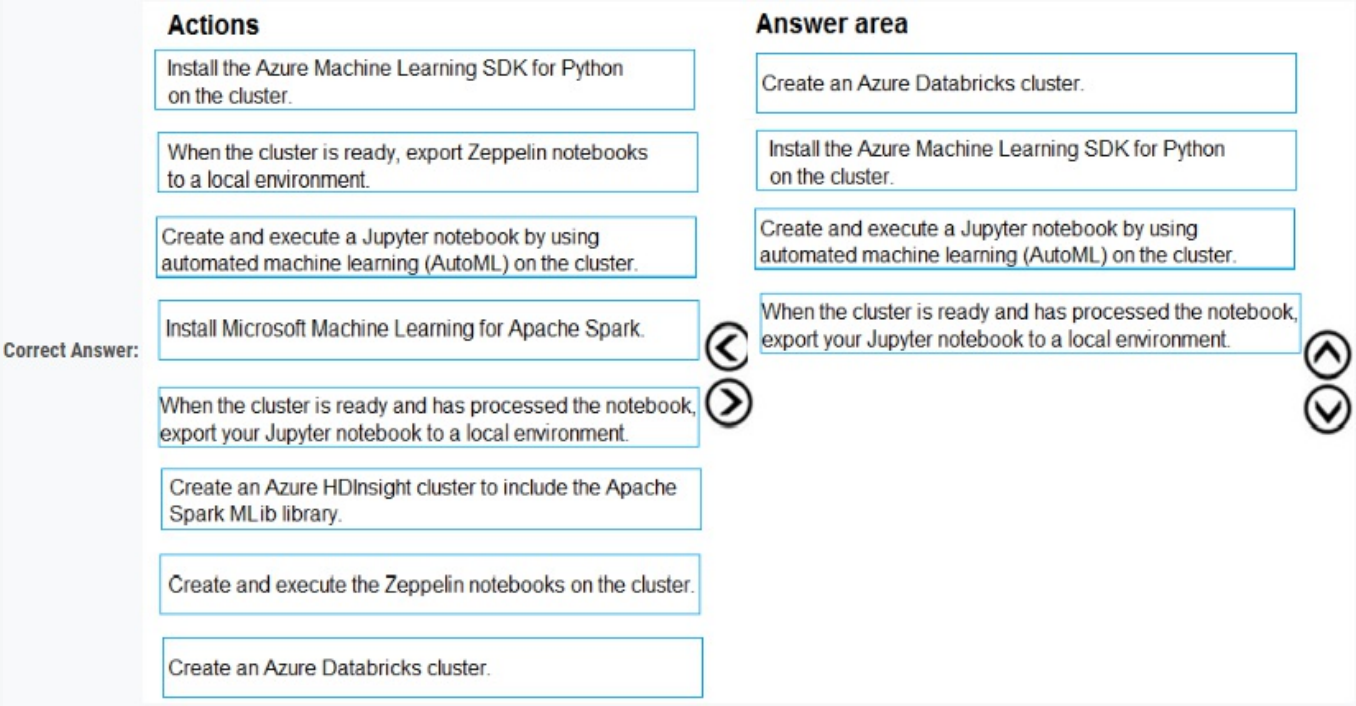
**Question: 49** CertyIQ

DRAG DROP -
You are building an intelligent solution using machine learning models.
The environment must support the following requirements:
- Data scientists must build notebooks in a cloud environment
- Data scientists must use automatic feature engineering and model building in machine learning pipelines.
- Notebooks must be deployed to retrain using Spark instances with dynamic worker allocation.
- Notebooks must be exportable to be version controlled locally.

You need to create the environment.
Which four 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.
Select and Place:




**Answer:**




**Explanation:**

Create Azure Databricks cluster -> Install Azure ML SDK for Python -> Create and exec Jupyter notebook using AutoML -> Export Jupyter to local env. That because you need auto feature engineering provided by autoML

**Question: 50** CertyIQ

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.

Incorrect Answers:

B: Use Machine Learning Studio when you want to experiment with machine learning models quickly and easily, and the built-in machine learning algorithms are sufficient for your solutions.

Reference:

https://docs.microsoft.com/en-us/azure/machine-learning/data-science-virtual-machine/overview

**Question: 51** CertyIQ

You are implementing a machine learning model to predict stock prices.
The model uses a PostgreSQL database and requires GPU processing.
You need to create a virtual machine that is pre-configured with the required tools.
What should you do?

A. Create a Data Science Virtual Machine (DSVM) Windows edition.
B. Create a Geo Al Data Science Virtual Machine (Geo-DSVM) Windows edition.
C. Create a Deep Learning Virtual Machine (DLVM) Linux edition.
D. Create a Deep Learning Virtual Machine (DLVM) Windows edition.

**Answer: C**

**Explanation:**
1. Only Windows and Linux Ubuntu two options when talking about machine / deep learning VMs; unless very specific MS / Windows products, e.g., SQL Server, otherwise always select Linux Ubuntu as your answer

**Question: 52** CertyIQ

You are developing deep learning models to analyze semi-structured, unstructured, and structured data types.
You have the following data available for model building:
- Video recordings of sporting events
- Transcripts of radio commentary about events
- Logs from related social media feeds captured during sporting events

You need to select an environment for creating the model.
Which environment should you use?

A. Azure Cognitive Services
B. Azure Data Lake Analytics
C. Azure HDInsight with Spark MLib
D. Azure Machine Learning Studio

**Answer: A**

**Explanation:**
Azure Cognitive Services expand on Microsoft's evolving portfolio of machine learning APIs and enable developers to easily add cognitive features " such as emotion and video detection; facial, speech, and vision recognition; and speech and language understanding " into their applications. The goal of Azure Cognitive Services is to help developers create applications that can see, hear, speak, understand, and even begin to reason. The catalog of services within Azure Cognitive
Services can be categorized into five main pillars - Vision, Speech, Language, Search, and Knowledge.

Reference:
https://docs.microsoft.com/en-us/azure/cognitive-services/welcome

**Question: 53** CertyIQ

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

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

## Question: 54

CertyIQ

You are moving a large dataset from Azure Machine Learning Studio to a Weka environment.
You need to format the data for the Weka environment.
Which module should you use?

A. Convert to CSV
B. Convert to Dataset
C. Convert to ARFF
D. Convert to SVMLight

**Answer: C**

**Explanation:**
Use the Convert to ARFF module in Azure Machine Learning Studio, to convert datasets and results in Azure Machine Learning to the attribute-relation file format used by the Weka toolset. This format is known as ARFF.
The ARFF data specification for Weka supports multiple machine learning tasks, including data preprocessing, classification, and feature selection. In this format, data is organized by entites and their attributes, and is contained in a single text file.

Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/convert-to-arff

## Question: 55

CertyIQ

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. Rattle
B. TensorFlow
C. Weka
D. Scikit-learn

**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.
Incorrect Answers:
A: Rattle is the R analytical tool that gets you started with data analytics and machine learning.
C: Weka is used for visual data mining and machine learning software in Java.
D: Scikit-learn is one of the most useful libraries for machine learning in Python. It is on NumPy, SciPy and matplotlib, this library contains a lot of efficient tools for machine learning and statistical modeling including classification, regression, clustering and dimensionality reduction.

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

**Question: 57** CertyIQ

You plan to use a Data Science Virtual Machine (DSVM) with the open source deep learning frameworks Caffe2 and PyTorch.
You need to select a pre-configured DSVM to support the frameworks.
What should you create?

A.Data Science Virtual Machine for Windows 2012
B.Data Science Virtual Machine for Linux (CentOS)
C.Geo AI Data Science Virtual Machine with ArcGIS
D.Data Science Virtual Machine for Windows 2016
E.Data Science Virtual Machine for Linux (Ubuntu)

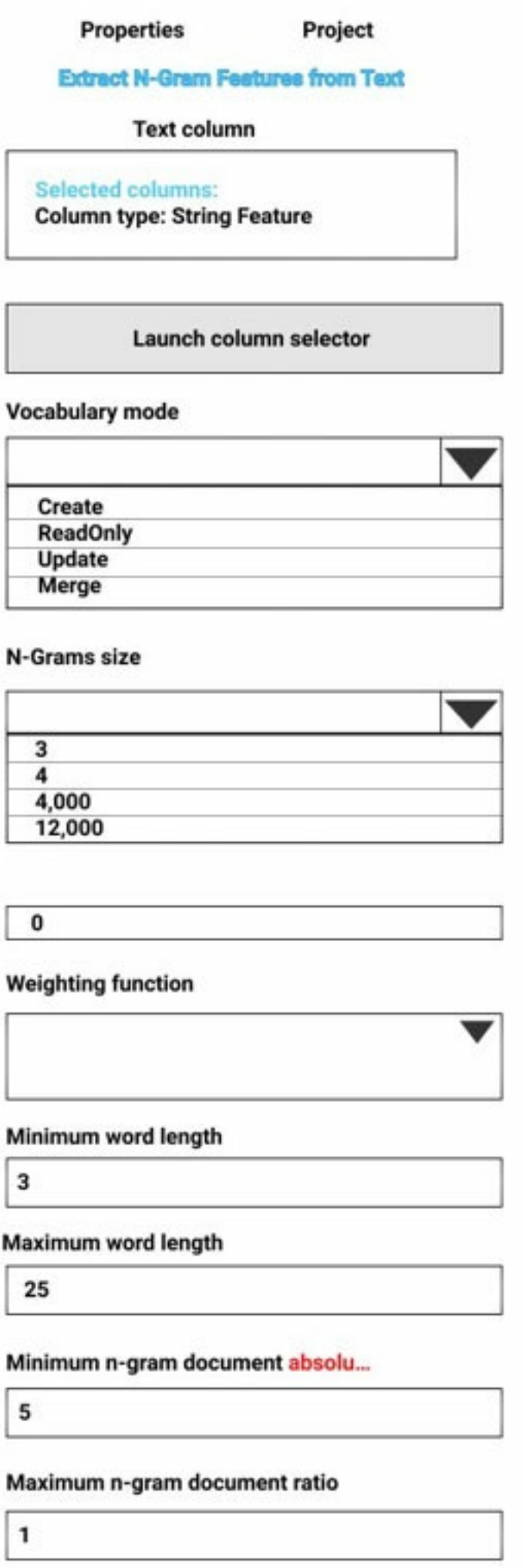
**Answer: E**

**Explanation:**
Caffe2 and PyTorch is supported by Data Science Virtual Machine for Linux.
Microsoft offers Linux editions of the DSVM on Ubuntu 16.04 LTS and CentOS 7.4.
Only the DSVM on Ubuntu is preconfigured for Caffe2 and PyTorch.
Incorrect Answers:
D: Caffe2 and PytOCH are only supported in the Data Science Virtual Machine for Linux.

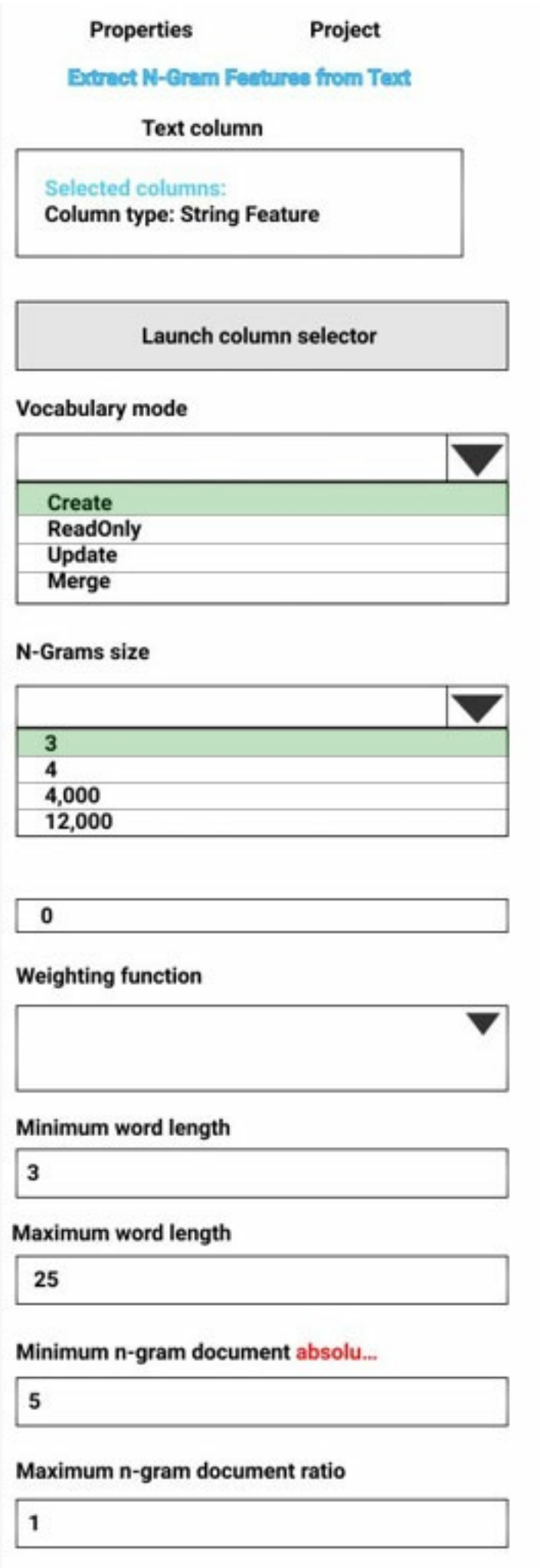
Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/data-science-virtual-machine/overview

**Question: 58** CertyIQ

HOTSPOT -
You are performing sentiment analysis using a CSV file that includes 12,000 customer reviews written in a short sentence format. You add the CSV file to Azure
Machine Learning Studio and configure it as the starting point dataset of an experiment. You add the Extract N-Gram Features from Text module to the experiment to extract key phrases from the customer review column in the dataset.
You must create a new n-gram dictionary from the customer review text and set the maximum n-gram size to trigrams.
What should you select? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.
Hot Area:


Properties
Project
Extract N-Gram Features from Text
Text column
Selected columns:
Column type: String Feature
Launch column selector
Vocabulary mode
Create
ReadOnly
Update
Merge
N-Grams size
3
4
4,000
12,000
0
Weighting function
Minimum word length
3
Maximum word length
25
Minimum n-gram document absolu...
5
Maximum n-gram document ratio
1

**Answer:**




**Explanation:**

Vocabulary mode: Create -

For Vocabulary mode, select Create to indicate that you are creating a new list of n-gram features.

N-Grams size: 3 -
For N-Grams size, type a number that indicates the maximum size of the n-grams to extract and store. For example, if you type 3, unigrams, bigrams, and trigrams will be created.

Weighting function: Leave blank -
The option, Weighting function, is required only if you merge or update vocabularies. It specifies how terms in the two vocabularies and their scores should be weighted against each other.

Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/extract-n-gram-features-from-text

**Question: 59** CertyIQ

You are developing a data science workspace that uses an Azure Machine Learning service.
You need to select a compute target to deploy the workspace.
What should you use?

A. Azure Data Lake Analytics
B. Azure Databricks
C. Azure Container Service
D. Apache Spark for HDInsight

**Answer: C**

**Explanation:**
Azure Container Instances can be used as compute target for testing or development. Use for low-scale CPU-based workloads that require less than 48 GB of
RAM.

Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/service/how-to-deploy-and-where

**Question: 60** CertyIQ

You are solving a classification task.
The dataset is imbalanced.
You need to select an Azure Machine Learning Studio module to improve the classification accuracy.
Which module should you use?

A. Permutation Feature Importance
B. Filter Based Feature Selection
C. Fisher Linear Discriminant Analysis
D. Synthetic Minority Oversampling Technique (SMOTE)

**Answer: D**

**Explanation:**
Use the SMOTE module in Azure Machine Learning Studio (classic) to increase the number of underrepresented cases in a dataset used for machine learning.

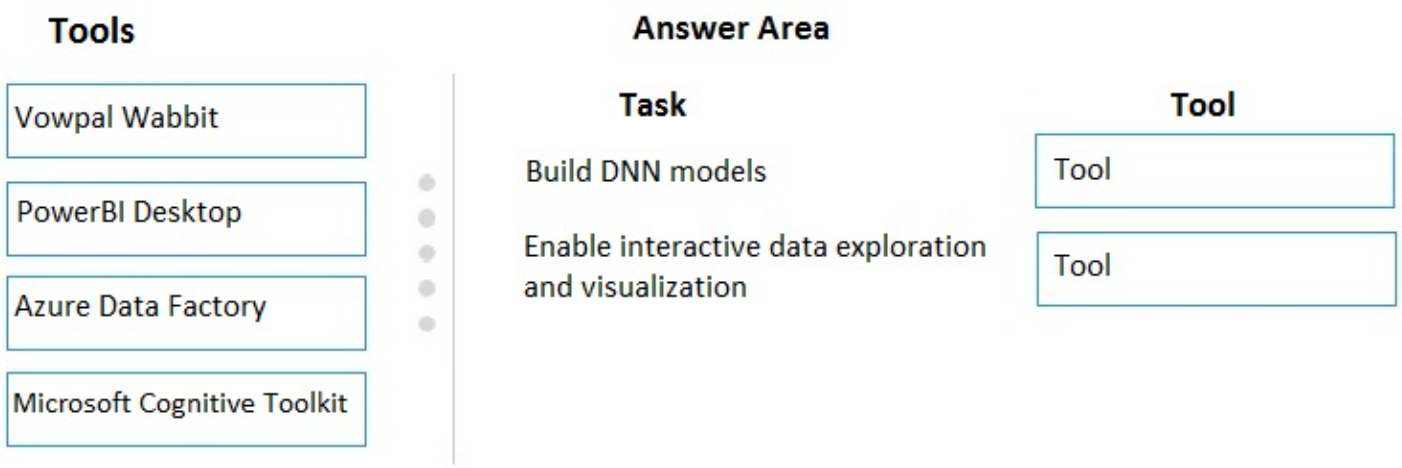
SMOTE is a better way of increasing the number of rare cases than simply duplicating existing cases.
You connect the SMOTE module to a dataset that is imbalanced. There are many reasons why a dataset might be imbalanced: the category you are targeting might be very rare in the population, or the data might simply be difficult to collect. Typically, you use SMOTE when the class you want to analyze is under- represented.

Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/smote

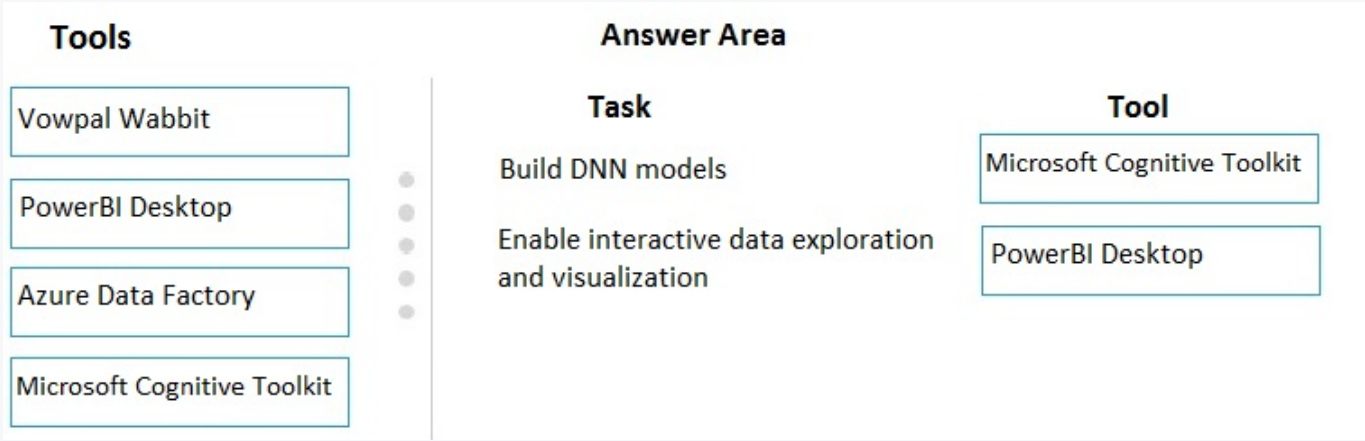
**Question: 61** CertyIQ

DRAG DROP -
You configure a Deep Learning Virtual Machine for Windows.
You need to recommend tools and frameworks to perform the following:
- Build deep neural network (DNN) models
- Perform interactive data exploration and visualization

Which tools and frameworks should you recommend? To answer, drag the appropriate tools to the correct tasks. Each tool may be used once, more than once, or not at all. You may need to drag the split bar between panes or scroll to view content.
NOTE: Each correct selection is worth one point.
Select and Place:




**Answer:**




**Explanation:**

Be Microsoft Cognitive Kit. Vowpal Wabbit is for Machine Learning. I will go with microsoft cognitive toolkit for DNN:

PowerBI Desktop -Power BI Desktop is a powerful visual data exploration and interactive reporting toolBI is a name given to a modern approach to business decision making in which users are empowered to find, explore, and share insights from data across the enterprise.

Reference:

https://docs.microsoft.com/en-us/cognitive-toolkit/examples#c-examples

**Question: 62** CertyIQ

You use Azure Machine Learning Studio to build a machine learning experiment.
You need to divide data into two distinct datasets.
Which module should you use?

A. Assign Data to Clusters
B. Load Trained Model
C. Partition and Sample
D. Tune Model-Hyperparameters

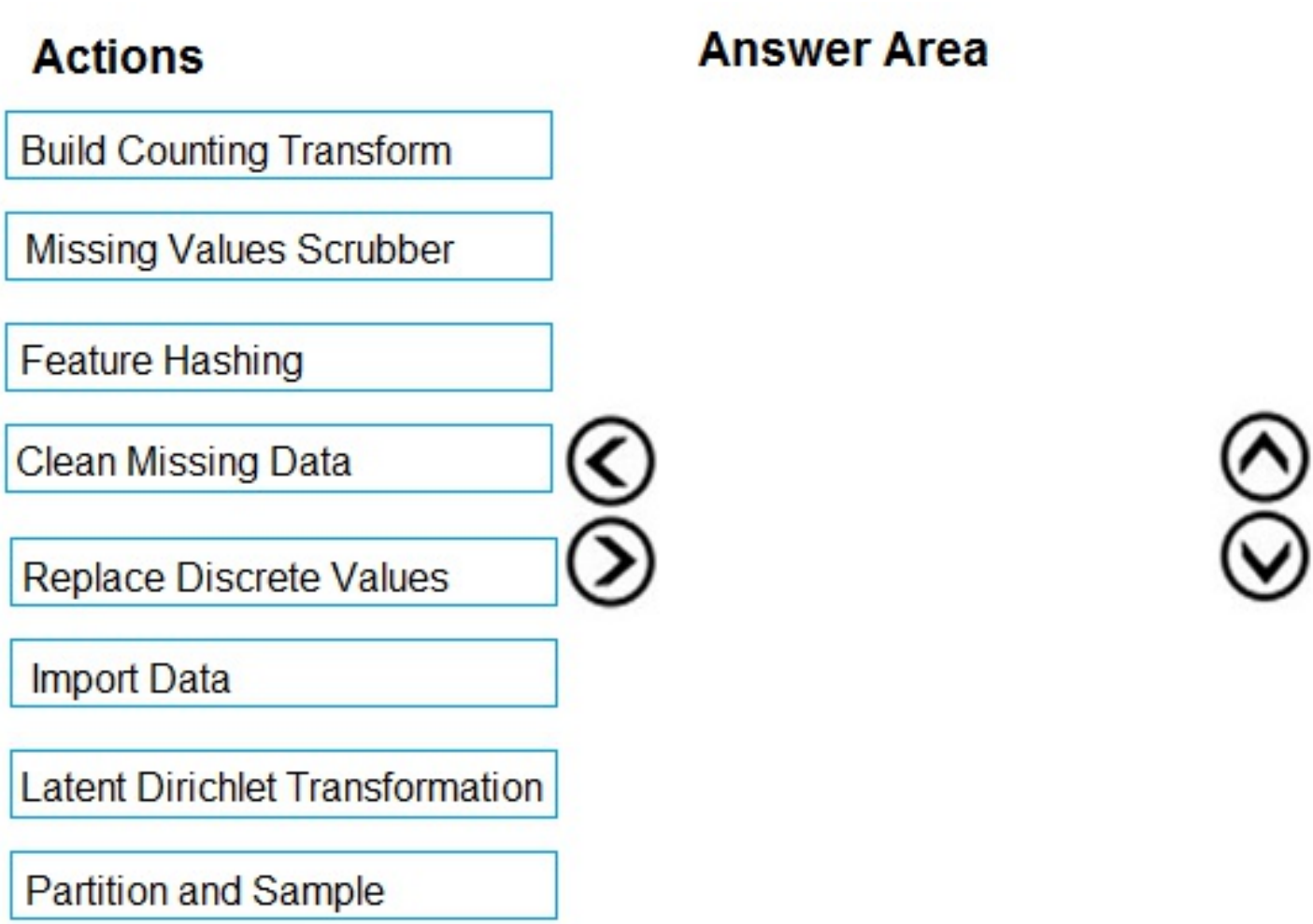
**Answer: C**

**Explanation:**
Partition and Sample with the Stratified split option outputs multiple datasets, partitioned using the rules you specified.

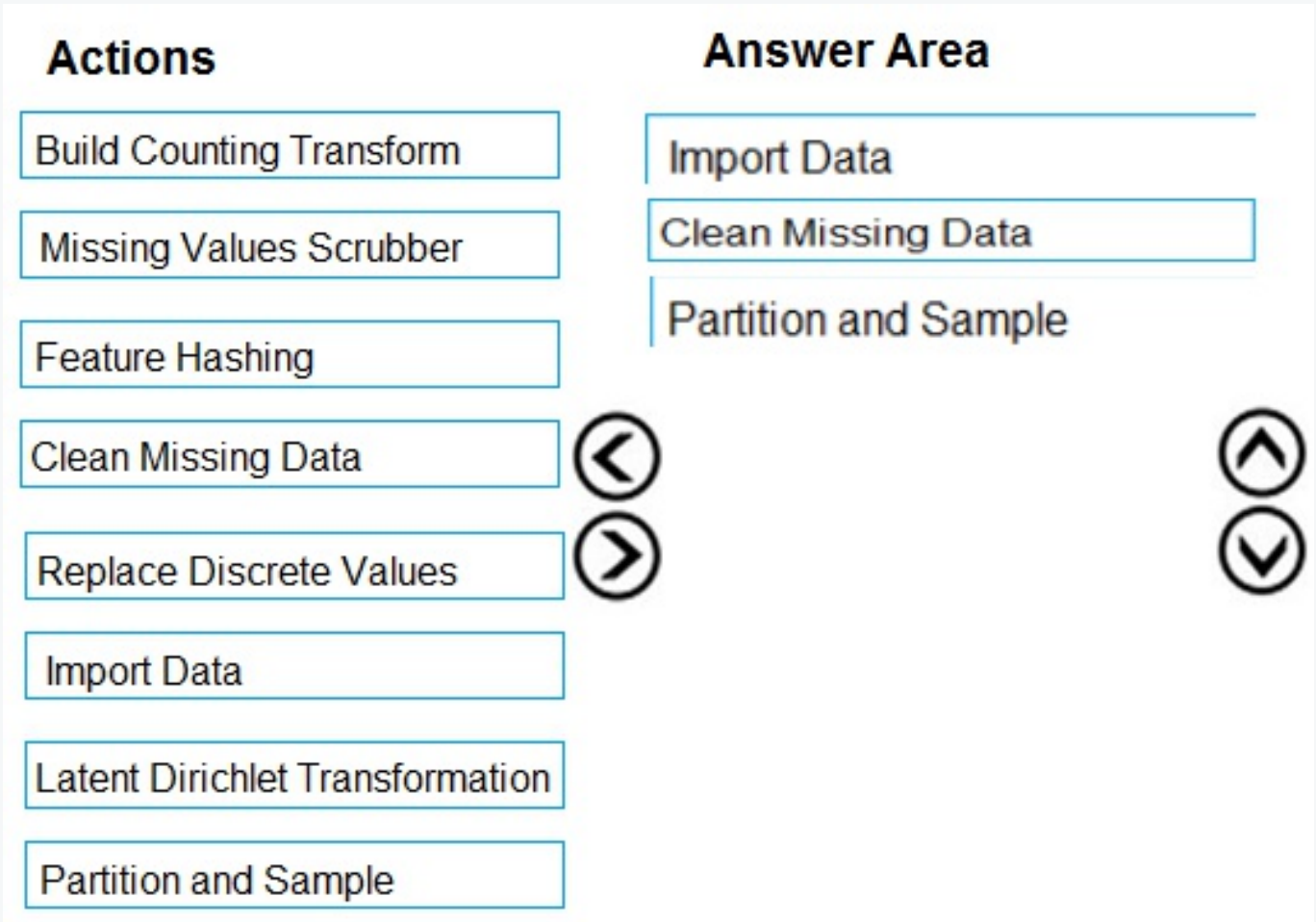
Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/partition-and-sample

**Question: 63** CertyIQ

DRAG DROP -
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.
Select and Place:




**Answer:**




**Explanation:**

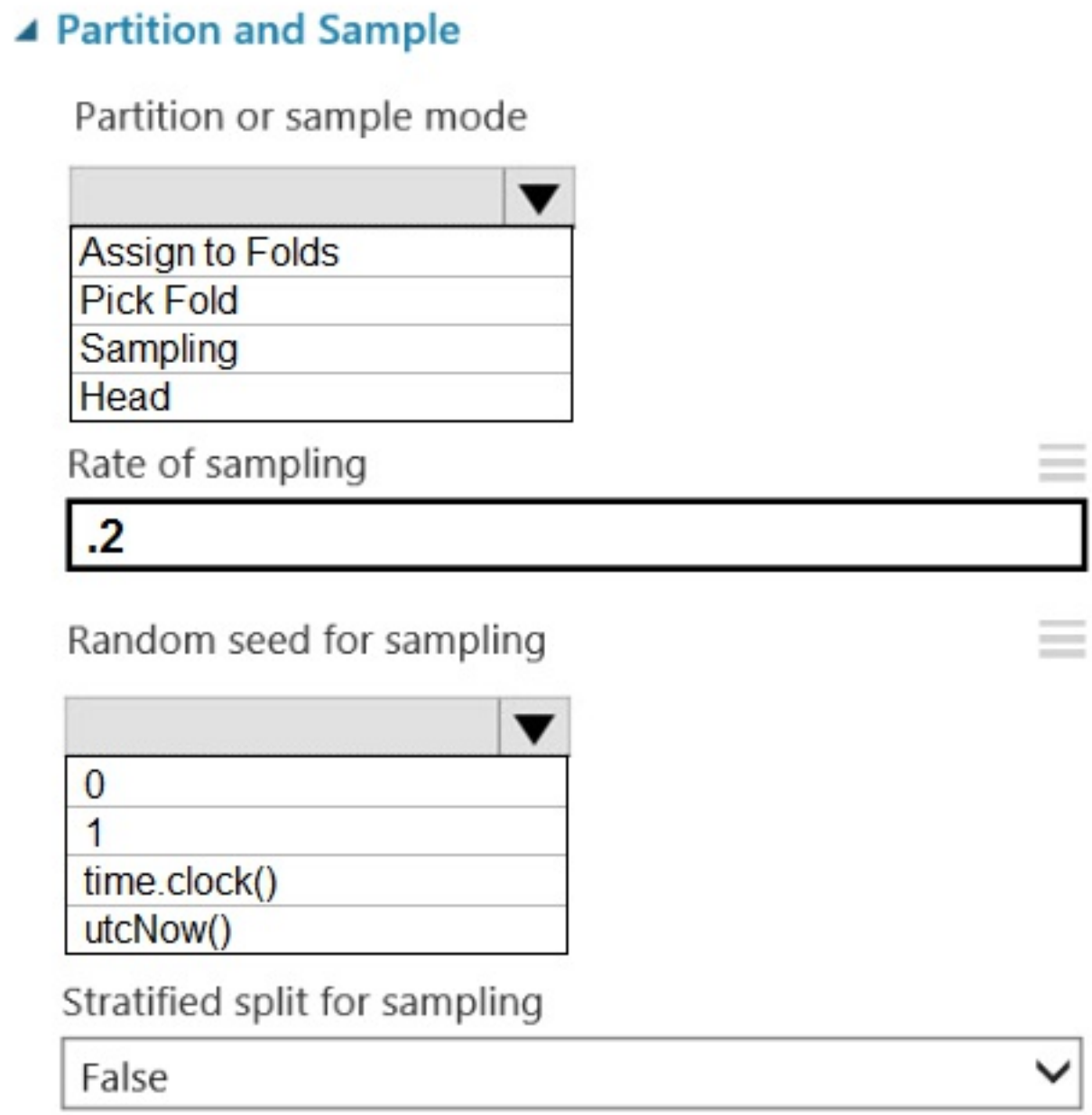
The Clean Missing Data module in Azure Machine Learning Studio, to remove, replace, or infer missing values.
Incorrect Answers:
✑ Latent Direchlet Transformation: Latent Dirichlet Allocation module in Azure Machine Learning Studio, to group otherwise unclassified text into a number of categories. Latent Dirichlet Allocation (LDA) is often used in natural language processing (NLP) to find texts that are similar. Another common term is topic modeling.
✑ Build Counting Transform: Build Counting Transform module in Azure Machine Learning Studio, to analyze training data. From this data, the module builds a count table as well as a set of count-based features that can be used in a predictive model.
Missing Value Scrubber: The Missing Values Scrubber module is deprecated.
✑ Feature hashing: Feature hashing is used for linguistics, and works by converting unique tokens into integers.
✑ Replace discrete values: the Replace Discrete Values module in Azure Machine Learning Studio is used to generate a probability score that can be used to represent a discrete value. This score can be useful for understanding the information value of the discrete values.

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

**Question: 64** CertyIQ

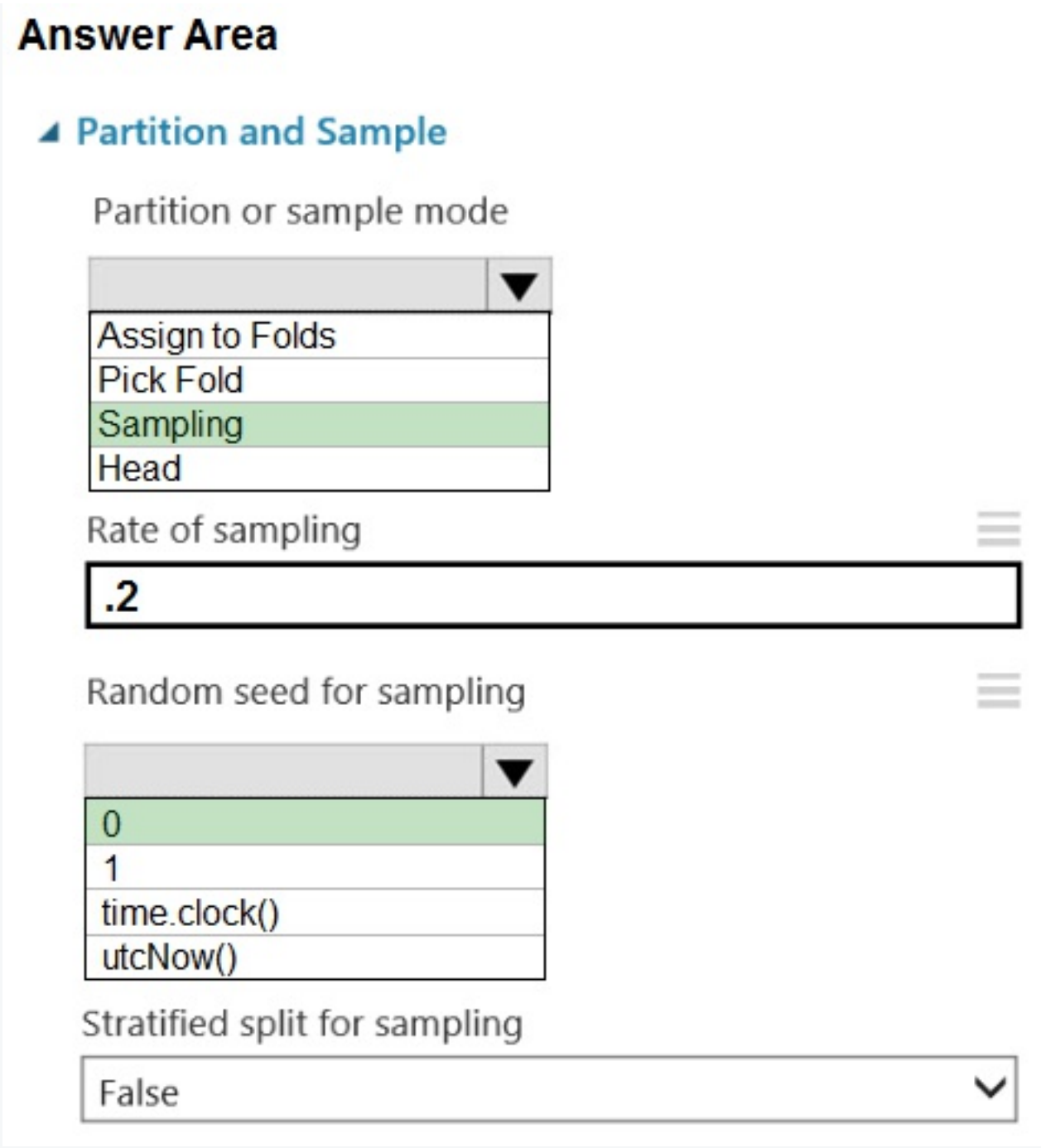
HOTSPOT -
You are retrieving data from a large datastore by using Azure Machine Learning Studio.
You must create a subset of the data for testing purposes using a random sampling seed based on the system clock.
You add the Partition and Sample module to your experiment.
You need to select the properties for the module.
Which values should you select? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.
Hot Area:

## Answer Area




**Answer:**

## Answer Area




**Explanation:**
Box 1: Sampling -

Create a sample of data -
This option supports simple random sampling or stratified random sampling. This is useful if you want to create a smaller representative sample dataset for testing.
1. Add the Partition and Sample module to your experiment in Studio, and connect the dataset.
2. Partition or sample mode: Set this to Sampling.
3. Rate of sampling. See box 2 below.

Box 2: 0 -
3. Rate of sampling. Random seed for sampling: Optionally, type an integer to use as a seed value.
This option is important if you want the rows to be divided the same way every time. The default value is 0, meaning that a starting seed is generated based on the system clock. This can lead to slightly different results each time you run the experiment.

Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/partition-and-sample

**Question: 65** CertyIQ

You are creating a machine learning model. You have a dataset that contains null rows.
You need to use the Clean Missing Data module in Azure Machine Learning Studio to identify and resolve the null and missing data in the dataset.
Which parameter should you use?

A. Replace with mean
B. Remove entire column
C. Remove entire row
D. Hot Deck
E. Custom substitution value
F. Replace with mode

**Answer: C**

**Explanation:**
Remove entire row: Completely removes any row in the dataset that has one or more missing values. This is useful if the missing value can be considered randomly missing.

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

**Question: 66** CertyIQ

HOTSPOT -
The finance team asks you to train a model using data in an Azure Storage blob container named finance-data.
You need to register the container as a datastore in an Azure Machine Learning workspace and ensure that an error will be raised if the container does not exist.
How should you complete the code? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.
Hot Area:

**Answer Area**

```
datastore = Datastore.[ register_azure_blob_container | register_azure_file_share | register_azure_data_lake | register_azure_sql_database ](workspace = ws,
    datastore_name = 'finance_datastore',
    container_name = 'finance-data',
    account_name = 'fintrainingdatastorage',
    account_key = 'FWUYORRv3XoyNe...',
    [ create_if_not_exists = True | create_if_not_exists = False | overwrite = True | overwrite = False ])
```

**Answer:**

**Answer Area**

```
datastore = Datastore.[register_azure_blob_container | register_azure_file_share | register_azure_data_lake | register_azure_sql_database] (workspace = ws,
    datastore_name = 'finance_datastore',
    container_name = 'finance-data',
    account_name = 'fintrainingdatastorage',
    account_key = 'FWUYORRv3XoyNe...',
    [create_if_not_exists = True | create_if_not_exists = False | overwrite = True | overwrite = False])
```

**Explanation:**

Box 1: register_azure_blob_container

Register an Azure Blob Container to the datastore.

Box 2: create_if_not_exists = False

Create the file share if it does not exist, defaults to False.

Reference:

https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.core.datastore.datastore

## Question: 67

CertyIQ

You plan to provision an Azure Machine Learning Basic edition workspace for a data science project.
You need to identify the tasks you will be able to perform in the workspace.
Which three tasks will you be able to perform? Each correct answer presents a complete solution.
NOTE: Each correct selection is worth one point.

A. Create a Compute Instance and use it to run code in Jupyter notebooks.

B. Create an Azure Kubernetes Service (AKS) inference cluster.

C. Use the designer to train a model by dragging and dropping pre-defined modules.

D. Create a tabular dataset that supports versioning.

E. Use the Automated Machine Learning user interface to train a model.

**Answer: ABD**

**Explanation:**

Incorrect Answers:

C, E: The UI is included the Enterprise edition only.

Reference:

https://azure.microsoft.com/en-us/pricing/details/machine-learning/

## Question: 68

CertyIQ

HOTSPOT -
A coworker registers a datastore in a Machine Learning services workspace by using the following code:

```
Datastore.register_azure_blob_container(workspace=ws,
 datastore_name='demo_datastore',
 container_name='demo_datacontainer',
 account_name='demo_account',
 account_key='0A0A0A-0A0A00A-0A00A0A0A0A0A',
 create_if_not_exists=True)
```

You need to write code to access the datastore from a notebook.
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.
Hot Area:

**Answer Area**

```
import azureml.core
from azureml.core import Workspace, Datastore
ws = Workspace.from_config()
datastore = [ ▼ ].get( [ ▼ ] , ' [ ▼ ] ')
```

| | | |
|---|---|---|
| Workspace | ws | demo_datastore |
| Datastore | run | demo_datacontainer |
| Experiment | experiment | demo_account |
| Run | log | Datastore |

**Answer:**

**Answer Area**

```
import azureml.core
from azureml.core import Workspace, Datastore
ws = Workspace.from_config()
datastore = [ Datastore ▼ ].get( [ ws ▼ ] , ' [ demo_datastore ▼ ] ')
```

**Explanation:**
Box 1: DataStore -
To get a specific datastore registered in the current workspace, use the get() static method on the Datastore class:
# Get a named datastore from the current workspace
datastore = Datastore.get(ws, datastore_name='your datastore name')

Box 2: ws -

Box 3: demo_datastore -

Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-access-data

**Question: 69** CertyIQ

A set of CSV files contains sales records. All the CSV files have the same data schema.

Each CSV file contains the sales record for a particular month and has the filename sales.csv. Each file is stored in a folder that indicates the month and year when the data was recorded. The folders are in an Azure blob container for which a datastore has been defined in an Azure Machine Learning workspace. The folders are organized in a parent folder named sales to create the following hierarchical structure:

```
/sales
  /01-2019
    /sales.csv
  /02-2019
    /sales.csv
  /03-2019
    /sales.csv
  ...
```

At the end of each month, a new folder with that month's sales file is added to the sales folder.
You plan to use the sales data to train a machine learning model based on the following requirements:
- You must define a dataset that loads all of the sales data to date into a structure that can be easily converted to a dataframe.
- You must be able to create experiments that use only data that was created before a specific previous month, ignoring any data that was added after that month.
- You must register the minimum number of datasets possible.

You need to register the sales data as a dataset in Azure Machine Learning service workspace.
What should you do?

A. Create a tabular dataset that references the datastore and explicitly specifies each 'sales/mm-yyyy/sales.csv' file every month. Register the dataset with the name sales_dataset each month, replacing the existing dataset and specifying a tag named month indicating the month and year it was registered. Use this dataset for all experiments.

B. Create a tabular dataset that references the datastore and specifies the path 'sales/*/sales.csv', register the dataset with the name sales_dataset and a tag named month indicating the month and year it was registered, and use this dataset for all experiments.

C. Create a new tabular dataset that references the datastore and explicitly specifies each 'sales/mm-yyyy/sales.csv' file every month. Register the dataset with the name sales_dataset_MM-YYYY each month with appropriate MM and YYYY values for the month and year. Use the appropriate month-specific dataset for experiments.

D. Create a tabular dataset that references the datastore and explicitly specifies each 'sales/mm-yyyy/sales.csv' file. Register the dataset with the name sales_dataset each month as a new version and with a tag named month indicating the month and year it was registered. Use this dataset for all experiments, identifying the version to be used based on the month tag as necessary.

**Answer: D**

**Explanation:**

B does not allow you to get the data from before a specific month. With D you create only one dataset with multiple versions (1 version per month). Similar example in 'Versioning best practice': the solution. Because we dont want to add data before specific month. So the option B adding everything is not desirable.

Reference:

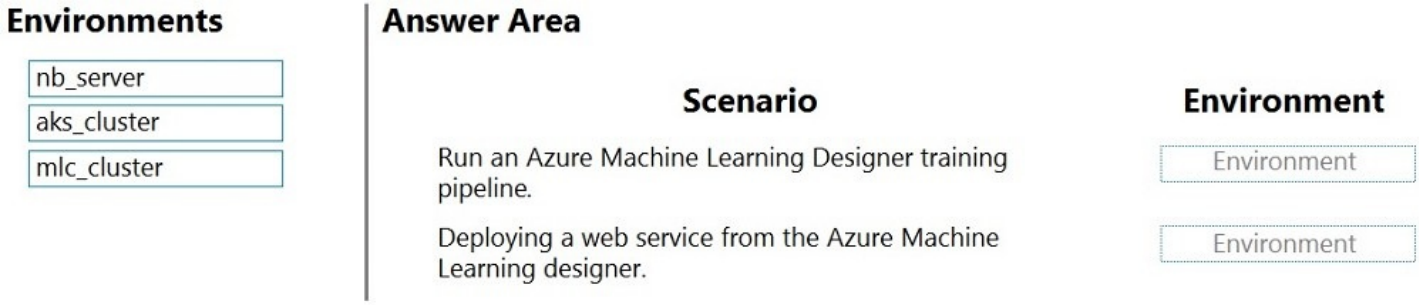
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-version-track-datasets

**Question: 70** CertyIQ

DRAG DROP -
An organization uses Azure Machine Learning service and wants to expand their use of machine learning.
You have the following compute environments. The organization does not want to create another compute environment.

| Environment name | Compute type |
|---|---|
| nb_server | Compute Instance |
| aks_cluster | Azure Kubernetes Service |
| mlc_cluster | Machine Learning Compute |

You need to determine which compute environment to use for the following scenarios.
Which compute types should you use? To answer, drag the appropriate compute environments to the correct scenarios. Each compute environment may be used once, more than once, or not at all. You may need to drag the split bar between panes or scroll to view content.
NOTE: Each correct selection is worth one point.
Select and Place:




**Answer:**


**Explanation:**

mlc_cluster, aks_cluster

**Question: 71** CertyIQ

HOTSPOT -
You create an Azure Machine Learning compute target named ComputeOne by using the STANDARD_D1 virtual machine image.
ComputeOne is currently idle and has zero active nodes.
You define a Python variable named ws that references the Azure Machine Learning workspace. You run the following Python code:

```
from azureml.core.compute import ComputeTarget, AmlCompute
from azureml.core.compute_target import ComputeTargetException
the_cluster_name = "ComputeOne"
try:
  the_cluster = ComputeTarget(workspace=ws, name=the_cluster_name)
  print('Step1')
except ComputeTargetException:
  config = AmlCompute.provisioning_configuration(vm_size='STANDARD_DS12_v2', max_nodes=4)
  the_cluster = ComputeTarget.create(ws, the_cluster_name, config)
  print('Step2')
```

For each of the following statements, select Yes if the statement is true. Otherwise, select No.
NOTE: Each correct selection is worth one point.
Hot Area:

**Answer Area**

| | Yes | No |
|---|---|---|
| A new machine learning compute resource is created with a virtual machine size of STANDARD_DS12_v2 and a maximum of four nodes. | ○ | ○ |
| Any experiments configured to use `the_cluster` will run on ComputeOne. | ○ | ○ |
| The text **Step1** will be printed to the screen. | ○ | ○ |

**Answer:**

**Answer Area**

| | Yes | No |
|---|---|---|
| A new machine learning compute resource is created with a virtual machine size of STANDARD_DS12_v2 and a maximum of four nodes. | ○ | ○ |
| Any experiments configured to use `the_cluster` will run on ComputeOne. | ○ | ○ |
| The text **Step1** will be printed to the screen. | ○ | ○ |

**Explanation:**

Box 1: no

Box 2: Yes -

Box 3: yes

Reference:

https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.core.compute.computetarget

**Question: 72** CertyIQ

HOTSPOT -
You are developing a deep learning model by using TensorFlow. You plan to run the model training workload on an Azure Machine Learning Compute Instance.
You must use CUDA-based model training.
You need to provision the Compute Instance.
Which two virtual machines sizes can you use? To answer, select the appropriate virtual machine sizes in the answer area.
NOTE: Each correct selection is worth one point.
Hot Area:

**Answer Area**

Virtual machine size

Search by name...

| Name ↑ | vCPUs | GPUs | RAM | Resource disk |
|---|---|---|---|---|
| BASIC_A0 | 1 | | 0.75 GB | 20 GB |
| STANDARD_D3_V2 | 4 | | 14 GB | 200 GB |
| STANDARD_E64_V3 | 64 | | 432 GB | 1,600 GB |
| STANDARD_M64LS | 64 | | 512 GB | 2,000 GB |
| STANDARD_NC12 | 12 | 2 | 112 GB | 680 GB |
| STANDARD_NC24 | 24 | 4 | 224 GB | 1,440 GB |

**Answer:**

**Answer Area**

Virtual machine size

Search by name...

| Name ↑ | vCPUs | GPUs | RAM | Resource disk |
|---|---|---|---|---|
| BASIC_A0 | 1 | | 0.75 GB | 20 GB |
| STANDARD_D3_V2 | 4 | | 14 GB | 200 GB |
| STANDARD_E64_V3 | 64 | | 432 GB | 1,600 GB |
| STANDARD_M64LS | 64 | | 512 GB | 2,000 GB |
| STANDARD_NC12 | 12 | 2 | 112 GB | 680 GB |
| STANDARD_NC24 | 24 | 4 | 224 GB | 1,440 GB |

**Explanation:**

CUDA is a parallel computing platform and programming model developed by Nvidia for general computing on its own GPUs (graphics processing units). CUDA enables developers to speed up compute-intensive applications by harnessing the power of GPUs for the parallelizable part of the computation.

Reference:
https://www.infoworld.com/article/3299703/what-is-cuda-parallel-programming-for-gpus.html

**Question: 73**

CertyIQ

DRAG DROP -
You are analyzing a raw dataset that requires cleaning.
You must perform transformations and manipulations by using Azure Machine Learning Studio.
You need to identify the correct modules to perform the transformations.
Which modules should you choose? To answer, drag the appropriate modules to the correct scenarios. Each module may be used once, more than once, or not at all.
You may need to drag the split bar between panes or scroll to view content.
NOTE: Each correct selection is worth one point.
Select and Place:

**Answer Area**

| Methods | Scenario | Module |
|---|---|---|
| Clean Missing Data | Replace missing values by removing rows and columns. | |
| SMOTE | Increase the number of low-incidence examples in the dataset. | |
| Convert to Indicator Values | Convert a categorical feature into a binary indicator. | |
| Remove Duplicate Rows | Remove potential duplicates from a dataset. | |
| Threshold Filter | | |

**Answer:**

**Answer Area**

| Methods | Scenario | Module |
|---|---|---|
| Clean Missing Data | Replace missing values by removing rows and columns. | Clean Missing Data |
| SMOTE | Increase the number of low-incidence examples in the dataset. | SMOTE |
| Convert to Indicator Values | Convert a categorical feature into a binary indicator. | Convert to Indicator Values |
| Remove Duplicate Rows | Remove potential duplicates from a dataset. | Remove Duplicate Rows |
| Threshold Filter | | |

**Explanation:**
Box 1: Clean Missing Data -

Box 2: SMOTE -
Use the SMOTE module in Azure Machine Learning Studio to increase the number of underepresented 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.
Box 3: Convert to Indicator Values
Use the Convert to Indicator Values module in Azure Machine Learning Studio. The purpose of this module is to convert columns that contain categorical values into a series of binary indicator columns that can more easily be used as features in a machine learning model.

Box 4: Remove Duplicate Rows -

Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/smote https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/convert-to-indicator-values

**Question: 74**

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 using Azure Machine Learning Studio to perform feature engineering on a dataset.
You need to normalize values to produce a feature column grouped into bins.
Solution: Apply an Entropy Minimum Description Length (MDL) binning mode.
Does the solution meet the goal?

A. Yes

B. No

**Answer: B**

**Explanation:**

Apply a Quantiles normalization with a QuantileIndex normalization. you can specify the following binning modes: - Entropy MDL - Quantiles - Equal Width - Custom Edges - Equal Width with Custom Start and Stop From all of these, the only binning mode which supports normalization is Quantiles. In particular, Entropy MDL does NOT support normalization.

Reference:

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

**Question: 75** CertyIQ

HOTSPOT -
You are preparing to use the Azure ML SDK to run an experiment and need to create compute. You run the following code:

```
from azureml.core.compute import ComputeTarget, AmlCompute
from azureml.core.compute_target import ComputeTargetException
ws = Workspace.from_config()
cluster_name = 'aml-cluster'
try:
  training_compute = ComputeTarget(workspace=ws, name=cluster_name)
except ComputeTargetException:
  compute_config = AmlCompute.provisioning_configuration(vm_size='STANDARD_D2_V2', vm_priority='lowpriority',
max_nodes=4)
  training_compute = ComputeTarget.create(ws, cluster_name, compute_config)
  training_compute.wait_for_completion(show_output=True)
```

For each of the following statements, select Yes if the statement is true. Otherwise, select No.
NOTE: Each correct selection is worth one point.
Hot Area:

**Answer Area**

| | Yes | No |
|---|---|---|
| If a compute cluster named aml-cluster already exists in the workspace, it will be deleted and replaced. | ○ | ○ |
| The `wait_for_completion()` method will not return until the aml-cluster compute has four active nodes. | ○ | ○ |
| If the code creates a new aml-cluster compute target, it may be preempted due to capacity constraints. | ○ | ○ |
| The aml-cluster compute target is deleted from the workspace after the training experiment completes. | ○ | ○ |

**Answer:**

**Answer Area**

| | Yes | No |
|---|---|---|
| If a compute cluster named aml-cluster already exists in the workspace, it will be deleted and replaced. | ○ | ○ |
| The `wait_for_completion()` method will not return until the aml-cluster compute has four active nodes. | ○ | ○ |
| If the code creates a new aml-cluster compute target, it may be preempted due to capacity constraints. | ○ | ○ |
| The aml-cluster compute target is deleted from the workspace after the training experiment completes. | ○ | ○ |

**Explanation:**

Box 1: No -

If a compute cluster already exists it will be used.

Box 2: No

Box 3: Yes -

Low Priority VMs use Azure's excess capacity and are thus cheaper but risk your run being pre-empted.

Box 4: No -

Need to use training_compute.delete() to deprovision and delete the AmlCompute target.

Reference:

https://notebooks.azure.com/azureml/projects/azureml-getting-started/html/how-to-use-azureml/training/train-on-amlcompute/train-on-amlcompute.ipynb https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.core.compute.computetarget

**Question: 76** CertyIQ

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: A**

**Explanation:**

If you select the Quantiles binning mode, use the Quantile normalization option to determine how values are normalized prior to sorting into quantiles. Note that normalizing values transforms the values, but does not affect the final number of bins. For an example, see Effects of Different Normalization Methods.

The following normalization types are supported:

Percent: Values are normalized within the range [0,100]

PQuantile: Values are normalized within the range [0,1]

QuantileIndex: Values are normalized within the range [1,number of bins]

**Question: 77** CertyIQ

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 Machine Learning Studio.
One class has a much smaller number of observations than the other classes in the training set.
You need to select an appropriate data sampling strategy to compensate for the class imbalance.
Solution: You use the Scale and Reduce sampling mode.
Does the solution meet the goal?

A. Yes
B. No

**Answer: B**

**Explanation:**
Instead use the Synthetic Minority Oversampling Technique (SMOTE) sampling mode.
Note: SMOTE is used to increase the number of underepresented 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.
Incorrect Answers:
Common data tasks for the Scale and Reduce sampling mode include clipping, binning, and normalizing numerical values.

Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/smote https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/data-transformation-scale-and-reduce

**Question: 78** CertyIQ

You are analyzing a dataset by using Azure Machine Learning Studio.
You need to generate a statistical summary that contains the p-value and the unique count for each feature column.
Which two modules can you use? Each correct answer presents a complete solution.
NOTE: Each correct selection is worth one point.

A. Computer Linear Correlation
B. Export Count Table
C. Execute Python Script
D. Convert to Indicator Values
E. Summarize Data

**Answer: BE**

**Explanation:**
The Export Count Table module is provided for backward compatibility with experiments that use the Build Count Table (deprecated) and Count Featurizer
(deprecated) modules.
E: Summarize Data statistics are useful when you want to understand the characteristics of the complete dataset. For example, you might need to know:
✑ How many missing values are there in each column?
✑ How many unique values are there in a feature column?
✑ What is the mean and standard deviation for each column?
✑ The module calculates the important scores for each column, and returns a row of summary statistics for each variable (data column) provided as input.
Incorrect Answers:
A: The Compute Linear Correlation module in Azure Machine Learning Studio is used to compute a set of Pearson correlation coefficients for each possible pair of variables in the input dataset.
C: With Python, you can perform tasks that aren't currently supported by existing Studio modules such as:
Visualizing data using matplotlib
Using Python libraries to enumerate datasets and models in your workspace
Reading, loading, and manipulating data from sources not supported by the Import Data module
D: The purpose of the Convert to Indicator Values module is to convert columns that contain categorical values into a series of binary indicator columns that can more easily be used as features in a machine learning model.

Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/export-count-table https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/summarize-data

**Question: 79** CertyIQ

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: Use the Last Observation Carried Forward (LOCF) method to impute the missing data points.
Does the solution meet the goal?

A. Yes

B. No

**Answer: A**

**Explanation:**

Absolutely possible to replace data with LOCF to keep the same number of rows and columns of the original dataset! That is the only requirements for the question!

**Question: 80** CertyIQ

HOTSPOT -
You are creating a machine learning model in Python. The provided dataset contains several numerical columns and one text column. The text column represents a product's category. The product category will always be one of the following:
- Bikes
- Cars
- Vans
- Boats

You are building a regression model using the scikit-learn Python package.
You need to transform the text data to be compatible with the scikit-learn Python package.
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.
Hot Area:

**Answer Area**

```
from sklearn import linear_model
import [ pandas as df | numpy as df | scipy as df ]
dataset = df.read_csv("data\\ProductSales.csv")
ProductCategoryMapping = {"Bikes":1, "Cars":2, "Boats": 3,
"Vans": 4}
dataset['ProductCategoryMapping'] =
dataset['ProductCategory'].[ map[ProductCategoryMapping] | reduce[ProductCategoryMapping] | transpose[ProductCategoryMapping] ]
regr = linear_model.LinearRegression()
X_train = dataset[['ProductCategoryMapping', 'ProductSize',
'ProductCost']]
y_train = dataset[['Sales']]
regr.fit(X_train, y_train)
```

**Answer:**

**Answer Area**

```
from sklearn import linear_model
import [pandas as df | numpy as df | scipy as df]
dataset = df.read_csv("data\\ProductSales.csv")
ProductCategoryMapping = {"Bikes":1, "Cars":2, "Boats": 3,
"Vans": 4}
dataset['ProductCategoryMapping'] =
dataset['ProductCategory'].[map[ProductCategoryMapping] | reduce[ProductCategoryMapping] | transpose[ProductCategoryMapping]]
regr = linear_model.LinearRegression()
X_train = dataset[['ProductCategoryMapping', 'ProductSize',
'ProductCost']]
y_train = dataset[['Sales']]
regr.fit(X_train, y_train)
```

**Explanation:**

Box 1: pandas as df -

Pandas takes data (like a CSV or TSV file, or a SQL database) and creates a Python object with rows and columns called data frame that looks very similar to table in a statistical software (think Excel or SPSS for example.

Box 2: transpose[ProductCategoryMapping]

Reshape the data from the pandas Series to columns.

Reference:

https://datascienceplus.com/linear-regression-in-python/

# Thank you

Thank you for being so interested in the premium exam material.
I'm glad to hear that you found it informative and helpful.

## But Wait

I wanted to let you know that there is more content available in the full version.
The full paper contains additional sections and information that you may find helpful, and I encourage you to download it to get a more comprehensive and detailed view of all the subject matter.

Download Full Version Now

**Future is Secured**

100% Pass Guarantee

**24/7 Customer Support**

Mail us - certyiqofficial@gmail.com

**Free Updates**

Lifetime Free Updates!

Total: **484 Questions**

Link: https://certyiq.com/papers?provider=microsoft&exam=dp-100