

# Microsoft

## Exam Questions DP-100

Designing and Implementing a Data Science Solution on Azure



**NEW QUESTION 1**

- (Exam Topic 3)

You have an Azure Machine Learning workspace that contains a training cluster and an inference cluster. You plan to create a classification model by using the Azure Machine Learning designer.

You need to ensure that client applications can submit data as HTTP requests and receive predictions as responses.

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

Actions

Create a real-time inference pipeline and run the pipeline on the compute cluster.

Create a batch inference pipeline and run the pipeline on the compute cluster.

Deploy a service to the compute cluster.

Create a pipeline that trains a classification model and run the pipeline on the compute cluster.

Deploy a service to the inference cluster.

Answer area

- A. Mastered
- B. Not Mastered

**Answer:** A

**Explanation:**

Actions

Create a real-time inference pipeline and run the pipeline on the compute cluster.

Create a batch inference pipeline and run the pipeline on the compute cluster.

Deploy a service to the compute cluster.

Create a pipeline that trains a classification model and run the pipeline on the compute cluster.

Deploy a service to the inference cluster.

Answer area

Create a pipeline that trains a classification model and run the pipeline on the compute cluster.

Create a batch inference pipeline and run the pipeline on the compute cluster.

Deploy a service to the inference cluster.

**NEW QUESTION 2**

- (Exam Topic 3)

HOTSPOT

You create a script for training a machine learning model in Azure Machine Learning service. You create an estimator by running the following code:

```
from azureml.core import Workspace, Datastore
from azureml.core.compute import ComputeTarget
from azureml.train.estimator import Estimator
work_space = Workspace.from_config()
data_source = work_space.get_default_datastore()
train_cluster = ComputeTarget(workspace=work_space, name= 'train-cluster')
estimator = Estimator(source_directory =
    'training-experiment',
    script_params = { ' --data-folder' : data_source.as_mount(), ' --regularization':0.8},
    compute_target = train_cluster,
    entry_script = 'train.py',
    conda_packages = ['scikit-learn'])
```

For each of the following statements, select Yes if the statement is true. Otherwise, select No.

NOTE: Each correct selection is worth one point.

	Yes	No
The estimator will look for the files it needs to run an experiment in the training-experiment directory of the local compute environment.	<input type="radio"/>	<input type="radio"/>
The estimator will mount the local data-folder folder and make it available to the script through a parameter.	<input type="radio"/>	<input type="radio"/>
The train.py script file will be created if it does not exist.	<input type="radio"/>	<input type="radio"/>
The estimator can run Scikit-learn experiments.	<input type="radio"/>	<input type="radio"/>

- A. Mastered  
B. Not Mastered

Answer: A

Explanation:

Box 1: Yes  
Parameter source\_directory is a local directory containing experiment configuration and code files needed for a training job.  
Box 2: Yes  
script\_params is a dictionary of command-line arguments to pass to the training script specified in entry\_script.  
Box 3: No  
Box 4: Yes  
The conda\_packages parameter is a list of strings representing conda packages to be added to the Python environment for the experiment.

NEW QUESTION 3

- (Exam Topic 3)  
You are using C-Support Vector classification to do a multi-class classification with an unbalanced training dataset. The C-Support Vector classification using Python code shown below:

```
from sklearn.svm import svc
import numpy as np
svc = SVC(kernel= 'linear', class_weight= 'balanced', C=1.0, random_state=0)
modell = svc.fit(X_train, y)
```

You need to evaluate the C-Support Vector classification code.  
Which evaluation statement should you use? To answer, select the appropriate options in the answer area. NOTE: Each correct selection is worth one point.

Code Segment	Evaluation Statement
class_weight=balanced	<div>Automatically select the performance metrics for the classification. Automatically adjust weights directly proportional to class frequencies in the input data. Automatically adjust weights inversely proportional to class frequencies in the input data.</div>
C parameter	<div>Penalty parameter Degree of polynomial kernel function Size of the kernel cache</div>

- A. Mastered  
B. Not Mastered

Answer: A

Explanation:

Box 1: Automatically adjust weights inversely proportional to class frequencies in the input data  
The “balanced” mode uses the values of y to automatically adjust weights inversely proportional to class frequencies in the input data as n\_samples / (n\_classes \* np.bincount(y)).  
Box 2: Penalty parameter  
Parameter: C : float, optional (default=1.0) Penalty parameter C of the error term. References:  
<https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVC.html>

NEW QUESTION 4

- (Exam Topic 3)



## HOTSPOT

You register the following versions of a model.

Model name	Model version	Tags	Properties
healthcare_model	3	'Training context': 'CPU Compute'	value:87.43
healthcare_model	2	'Training context': 'CPU Compute'	value:54.98
healthcare_model	1	'Training context': 'CPU Compute'	value:23.56

You use the Azure ML Python SDK to run a training experiment. You use a variable named run to reference the experiment run. After the run has been submitted and completed, you run the following code:

```
run.register_model(model_path='outputs/model.pkl',  
                  model_name='healthcare_model',  
                  tags={'Training context': 'CPU Compute'})
```

For each of the following statements, select Yes if the statement is true. Otherwise, select No.

NOTE: Each correct selection is worth one point.

	Yes	No
The code will cause a previous version of the saved model to be overwritten.	<input type="radio"/>	<input type="radio"/>
The version number will now be 4.	<input type="radio"/>	<input type="radio"/>
The latest version of the stored model will have a property of value: 87.43.	<input type="radio"/>	<input type="radio"/>

- A. Mastered  
B. Not Mastered

**Answer:** A

### Explanation:

Reference:

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

## NEW QUESTION 5

- (Exam Topic 3)

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

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

An IT department creates the following Azure resource groups and resources:

Resource group	Resources
ml_resources	<ul style="list-style-type: none"><li>an Azure Machine Learning workspace named amlworkspace</li><li>an Azure Storage account named amlworkspace12345</li><li>an Application Insights instance named amlworkspace54321</li><li>an Azure Key Vault named amlworkspace67890</li><li>an Azure Container Registry named amlworkspace09876</li></ul>
general_compute	<p>A virtual machine named mlvm with the following configuration:</p> <ul style="list-style-type: none"><li>Operating system: Ubuntu Linux</li><li>Software installed: Python 3.6 and Jupyter Notebooks</li><li>Size: NC6 (6 vCPUs, 1 vGPU, 56 Gb RAM)</li></ul>

The IT department creates an Azure Kubernetes Service (AKS)-based inference compute target named aks-cluster in the Azure Machine Learning workspace. You have a Microsoft Surface Book computer with a GPU. Python 3.6 and Visual Studio Code are installed. You need to run a script that trains a deep neural network (DNN) model and logs the loss and accuracy metrics.

Solution: Install the Azure ML SDK on the Surface Book. Run Python code to connect to the workspace. Run the training script as an experiment on the aks-cluster compute target.

Does the solution meet the goal?

- A. Yes  
B. No

**Answer:** B

### Explanation:

Need to attach the mlvm virtual machine as a compute target in the Azure Machine Learning workspace.

Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/concept-compute-target>

## NEW QUESTION 6

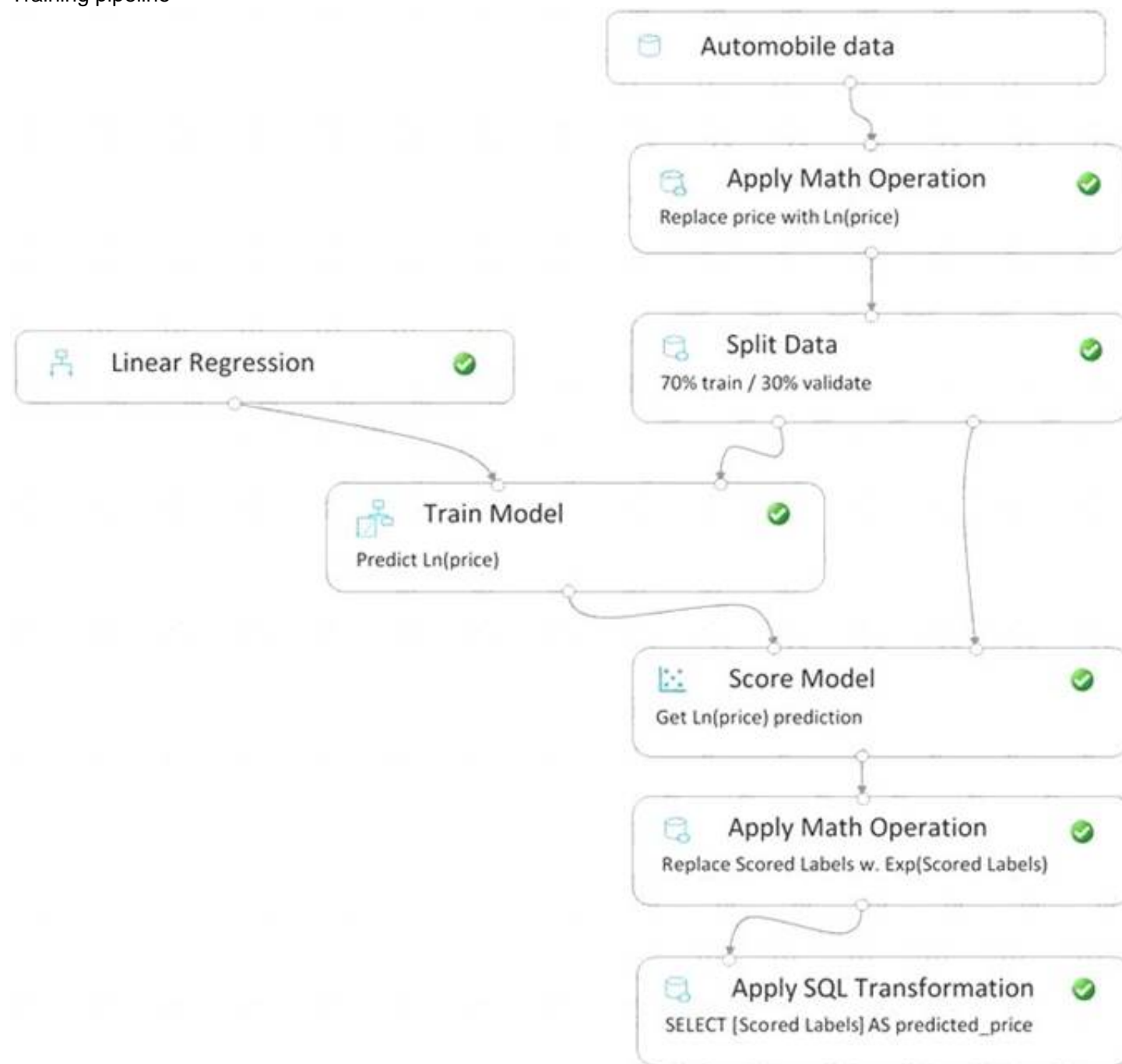
- (Exam Topic 3)

You create a pipeline in designer to train a model that predicts automobile prices.

Because of non-linear relationships in the data, the pipeline calculates the natural log (Ln) of the prices in the training data, trains a model to predict this natural log of price value, and then calculates the exponential of the scored label to get the predicted price.

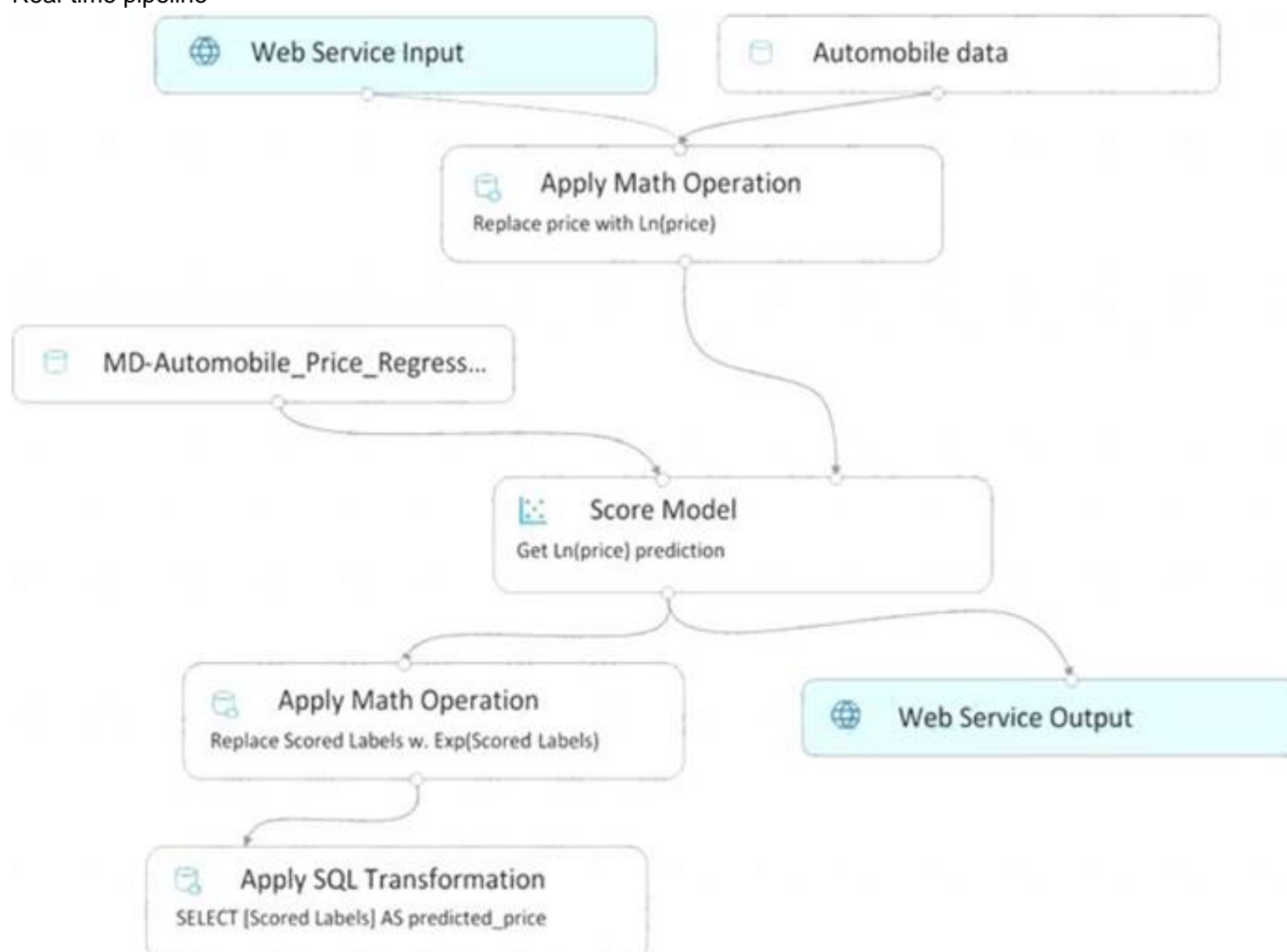
The training pipeline is shown in the exhibit. (Click the Training pipeline tab.)

Training pipeline



You create a real-time inference pipeline from the training pipeline, as shown in the exhibit. (Click the Real-time pipeline tab.)

Real-time pipeline



You need to modify the inference pipeline to ensure that the web service returns the exponential of the scored label as the predicted automobile price and that client applications are not required to include a price value in the input values.

Which three modifications must you make to the inference pipeline? Each correct answer presents part of the solution.

NOTE: Each correct selection is worth one point.

- A. Connect the output of the Apply SQL Transformation to the Web Service Output module.
- B. Replace the Web Service Input module with a data input that does not include the price column.
- C. Add a Select Columns module before the Score Model module to select all columns other than price.
- D. Replace the training dataset module with a data input that does not include the price column.
- E. Remove the Apply Math Operation module that replaces price with its natural log from the data flow.
- F. Remove the Apply SQL Transformation module from the data flow.

**Answer:** ACE

#### NEW QUESTION 7

- (Exam Topic 3)

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

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

You are creating a new experiment in Azure 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 Stratified split for the 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 underrepresented cases in a dataset used for machine learning. SMOTE is a better way of increasing the number of rare cases than simply duplicating existing cases.

References:

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

#### NEW QUESTION 8

- (Exam Topic 3)

You create a training pipeline using the Azure Machine Learning designer. You upload a CSV file that contains the data from which you want to train your model.

You need to use the designer to create a pipeline that includes steps to perform the following tasks:

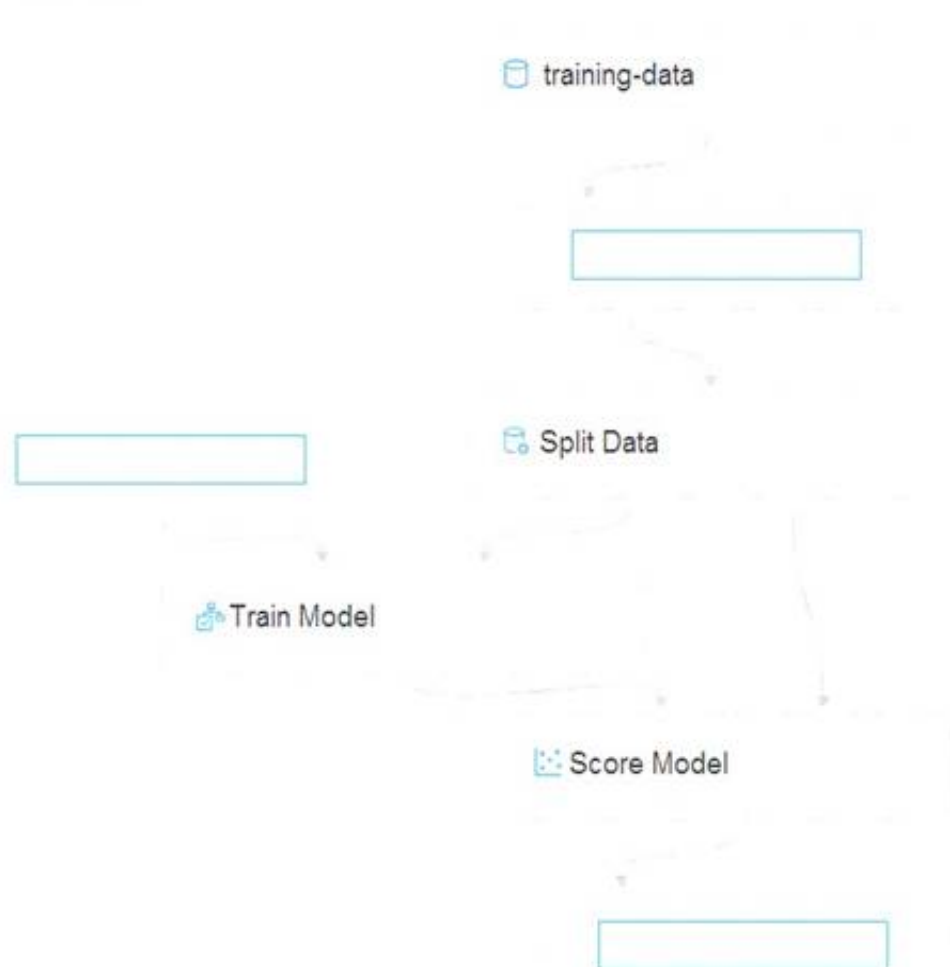
- > Select the training features using the pandas filter method.
- > Train a model based on the naive\_bayes.GaussianNB algorithm.
- > Return only the Scored Labels column by using the query `SELECT [Scored Labels] FROM t1`; Which modules should you use? To answer, drag the appropriate modules to the appropriate locations. Each module name 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.

#### Modules

#### Answer Area

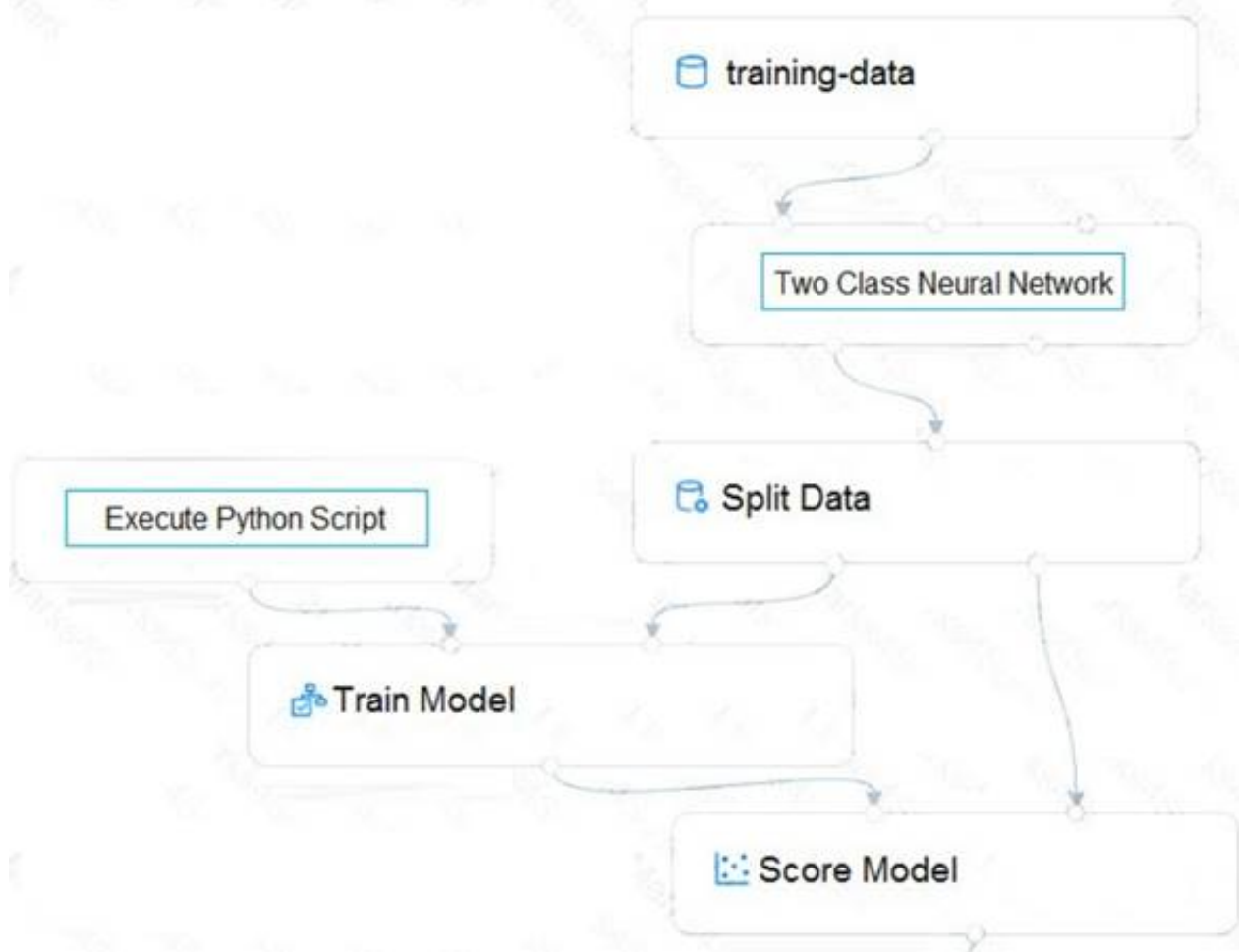
- Create Python Model
- Train Model
- Two Class Neural Network
- Execute Python Script
- Apply SQL Transformation
- Select Columns in Dataset



- A. Mastered
- B. Not Mastered

**Answer:** A

Explanation:



**NEW QUESTION 9**

- (Exam Topic 3)

Your Azure Machine Learning workspace has a dataset named `real_estate_data`. A sample of the data in the dataset follows.

postal_code	num_bedrooms	sq_feet	garage	price
12345	3	1300	0	23,9000
54321	1	950	0	11,0000
12346	2	1200	1	15,0000

You want to use automated machine learning to find the best regression model for predicting the price column. You need to configure an automated machine learning experiment using the Azure Machine Learning SDK. How should you complete the code? To answer, select the appropriate options in the answer area.  
 NOTE: Each correct selection is worth one point.

**Answer Area**

```

from azureml.core import Workspace
from azureml.core.compute import ComputeTarget
from azureml.core.runconfig import RunConfiguration
from azureml.train.automl import AutoMLConfig

ws = Workspace.from_config()
training_cluster = ComputeTarget(workspace=ws, name= 'aml-cluster1')
real_estate_ds = ws.datasets.get('real_estate_data')
split1_ds, split2_ds = real_estate_ds.random_split(percentage=0.7, seed=123)
automl_run_config = RunConfiguration(framework= "python")
automl_config = AutoMLConfig(
    task= 'regression',
    compute_target= training_cluster,
    run_configuration=automl_run_config,
    primary_metric='r2_score',
    

▼ =split1_ds,
      X
      Y
      X_valid
      Y_valid
      training_data



▼ =split2_ds
      X
      Y
      X_valid
      Y_valid
      validation_data
      training_data



▼ ='price')
      y
      y_valid
      y_max
      label_column_name
      exclude_nan_labels


```



- A. Mastered
- B. Not Mastered

**Answer:** A

**Explanation:**

Box 1: training\_data

The training data to be used within the experiment. It should contain both training features and a label column (optionally a sample weights column). If training\_data is specified, then the label\_column\_name parameter must also be specified.

Box 2: validation\_data

Provide validation data: In this case, you can either start with a single data file and split it into training and validation sets or you can provide a separate data file for the validation set. Either way, the validation\_data parameter in your AutoMLConfig object assigns which data to use as your validation set.

Example, the following code example explicitly defines which portion of the provided data in dataset to use for training and validation.

```
dataset = Dataset.Tabular.from_delimited_files(data)
training_data, validation_data = dataset.random_split(percentage=0.8, seed=1) automl_config = AutoMLConfig(compute_target = aml_remote_compute, task = 'classification',
primary_metric = 'AUC_weighted', training_data = training_data,
validation_data = validation_data, label_column_name = 'Class'
)
```

Box 3: label\_column\_name label\_column\_name:

The name of the label column. If the input data is from a pandas.DataFrame which doesn't have column names, column indices can be used instead, expressed as integers.

This parameter is applicable to training\_data and validation\_data parameters. Reference:

<https://docs.microsoft.com/en-us/python/api/azureml-train-automl-client/azureml.train.automl.automlconfig.auto>

**NEW QUESTION 10**

- (Exam Topic 3)

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

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

You have a Python script named train.py in a local folder named scripts. The script trains a regression model by using scikit-learn. The script includes code to load a training data file which is also located in the scripts folder.

You must run the script as an Azure ML experiment on a compute cluster named aml-compute.

You need to configure the run to ensure that the environment includes the required packages for model training. You have instantiated a variable named aml-compute that references the target compute cluster.

Solution: Run the following code:

```
from azureml.train.dnn import TensorFlow
sk_est = TensorFlow(source_directory='./scripts',
compute_target=aml-compute,
entry_script='train.py')
```

Does the solution meet the goal?

- A. Yes
- B. No

**Answer:** B

**Explanation:**

The scikit-learn estimator provides a simple way of launching a scikit-learn training job on a compute target. It is implemented through the SKLearn class, which can be used to support single-node CPU training.

Example:

```
from azureml.train.sklearn import SKLearn
}
estimator = SKLearn(source_directory=project_folder, compute_target=compute_target, entry_script='train_iris.py')
)
```

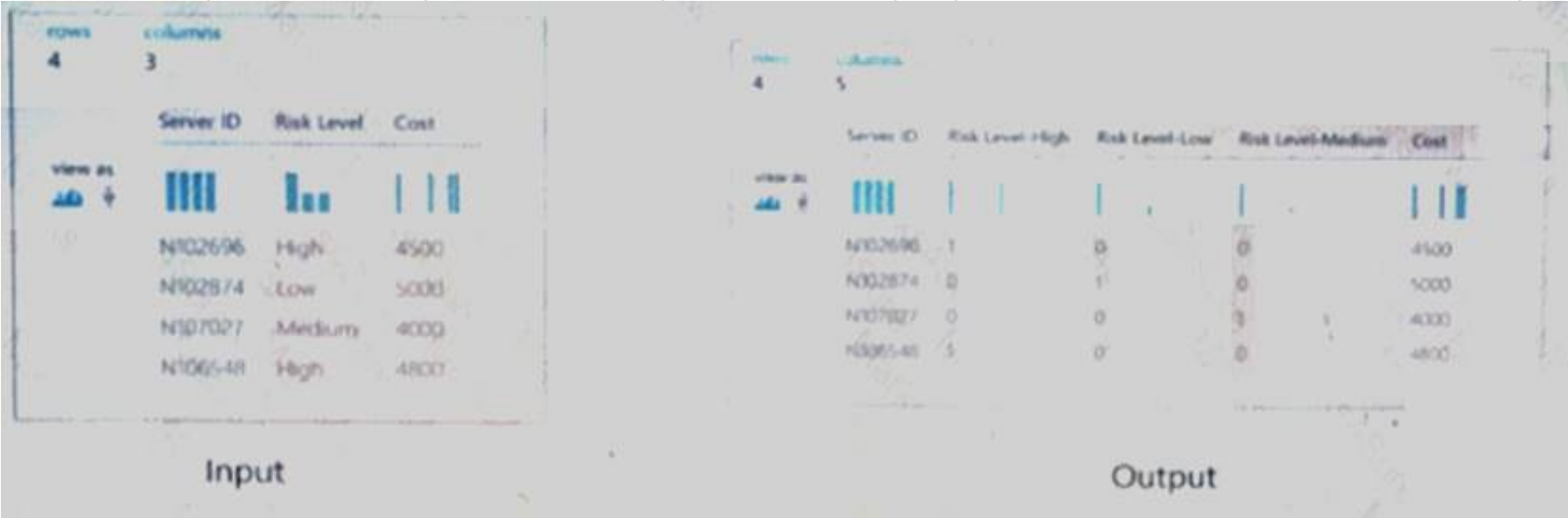
Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/how-to-train-scikit-learn>

**NEW QUESTION 10**

- (Exam Topic 3)

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





Use the drop-down menus to select the answer choice that answers each question based on the information presented in the graphic.  
 NOTE: Each correct selection is worth one point.

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

What is the expected input column type for this transformation?

Apply Filter
Build Counting Transform
Convert to Indicator Values

Categorical
Numerical
String

- A. Mastered
- B. Not Mastered

**Answer:** A

**Explanation:**

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

What is the expected input column type for this transformation?

Apply Filter
Build Counting Transform
Convert to Indicator Values

Categorical
Numerical
String

#### NEW QUESTION 14

- (Exam Topic 3)

You use the Two-Class Neural Network module in Azure Machine Learning Studio to build a binary classification model. You use the Tune Model Hyperparameters module to tune accuracy for the model.

You need to select the hyperparameters that should be tuned using the Tune Model Hyperparameters module. Which two hyperparameters should you use? Each correct answer presents part of the solution.

NOTE: Each correct selection is worth one point.

- A. Number of hidden nodes
- B. Learning Rate
- C. The type of the normalizer
- D. Number of learning iterations
- E. Hidden layer specification

**Answer:** DE

**Explanation:**

D: For Number of learning iterations, specify the maximum number of times the algorithm should process the training cases.

E: For Hidden layer specification, select the type of network architecture to create.

Between the input and output layers you can insert multiple hidden layers. Most predictive tasks can be accomplished easily with only one or a few hidden layers.

References:

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

#### NEW QUESTION 19

- (Exam Topic 3)

You use the Azure Machine Learning designer to create and run a training pipeline.

The pipeline must be run every night to inference predictions from a large volume of files. The folder where the files will be stored is defined as a dataset.

You need to publish the pipeline as a REST service that can be used for the nightly inferencing run. What should you do?

- A. Create a batch inference pipeline
- B. Set the compute target for the pipeline to an inference cluster
- C. Create a real-time inference pipeline
- D. Clone the pipeline

**Answer:** A

**Explanation:**

Azure Machine Learning Batch Inference targets large inference jobs that are not time-sensitive. Batch Inference provides cost-effective inference compute scaling, with unparalleled throughput for asynchronous applications. It is optimized for high-throughput, fire-and-forget inference over large collections of data.

You can submit a batch inference job by pipeline\_run, or through REST calls with a published pipeline. Reference:

<https://github.com/Azure/MachineLearningNotebooks/blob/master/how-to-use-azureml/machine-learning-pipeli>

#### NEW QUESTION 24

- (Exam Topic 3)

You are training machine learning models in Azure Machine Learning. You use Hyperdrive to tune the hyperparameters. In previous model training and tuning runs, many models showed similar performance. You need to select an early termination policy that meets the following requirements:

- accounts for the performance of all previous runs when evaluating the current run
- avoids comparing the current run with only the best performing run to date

Which two early termination policies should you use? Each correct answer presents part of the solution. NOTE: Each correct selection is worth one point.

- A. Bandit

- B. Median stopping
- C. Default
- D. Truncation selection

**Answer:** BC

**Explanation:**

The Median Stopping policy computes running averages across all runs and cancels runs whose best performance is worse than the median of the running averages.

If no policy is specified, the hyperparameter tuning service will let all training runs execute to completion. Reference:

<https://docs.microsoft.com/en-us/python/api/azureml-train-core/azureml.train.hyperdrive.medianstoppingpolicy> <https://docs.microsoft.com/en-us/python/api/azureml-train-core/azureml.train.hyperdrive.truncationselectionpoli> <https://docs.microsoft.com/en-us/python/api/azureml-train-core/azureml.train.hyperdrive.banditpolicy>

**NEW QUESTION 25**

- (Exam Topic 3)

You plan to use a Deep Learning Virtual Machine (DLVM) to train deep learning models using Compute Unified Device Architecture (CUDA) computations. You need to configure the DLVM to support CUDA. What should you implement?

- A. Intel Software Guard Extensions (Intel SGX) technology
- B. Solid State Drives (SSD)
- C. Graphic Processing Unit (GPU)
- D. Computer Processing Unit (CPU) speed increase by using overcloking
- E. High Random Access Memory (RAM) configuration

**Answer:** C

**Explanation:**

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

References:

<https://azuremarketplace.microsoft.com/en-au/marketplace/apps/microsoft-ads.dsvm-deep-learning>

**NEW QUESTION 30**

- (Exam Topic 3)

You are solving a classification task.

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

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

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

**Answer:** C

**Explanation:**

Leave One Out (LOO) cross-validation

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

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

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

**NEW QUESTION 31**

- (Exam Topic 3)

You create an Azure Machine Learning workspace.

You must create a custom role named DataScientist that meets the following requirements:

- Role members must not be able to delete the workspace.
- Role members must not be able to create, update, or delete compute resource in the workspace.
- Role members must not be able to add new users to the workspace.

You need to create a JSON file for the DataScientist role in the Azure Machine Learning workspace. The custom role must enforce the restrictions specified by the IT Operations team.

Which JSON code segment should you use?

A)

```
{
  "Name": "DataScientist",
  "IsCustom": true,
  "Description": "Project Data Scientist role",
  "Actions": ["*"],
  "NotActions": [
    "Microsoft.MachineLearningServices/workspaces/*/delete",
    "Microsoft.MachineLearningServices/workspaces/computes/*/write",
    "Microsoft.MachineLearningServices/workspaces/computes/*/delete",
    "Microsoft.Authorization/*/write"
  ],
  "AssignableScopes": [
    "/subscriptions/<id>/resourceGroups/ml-rg/providers/Microsoft/MachineLearningServices/workspaces/ml-ws"
  ]
}
```

B)

```
{
  "Name": "DataScientist",
  "IsCustom": true,
  "Description": "Project Data Scientist role",
  "Actions": ["*"],
  "NotActions": [],
  "AssignableScopes": [
    "/subscriptions/<id>/resourceGroups/ml-rg/providers/Microsoft.MachineLearningServices/workspaces/ml-ws"
  ]
}
```

C)

```
{
  "Name": "DataScientist",
  "IsCustom": true,
  "Description": "Project Data Scientist role",
  "Actions": [
    "Microsoft.MachineLearningServices/workspaces/*/delete",
    "Microsoft.MachineLearningServices/workspaces/computes/*/write",
    "Microsoft.MachineLearningServices/workspaces/computes/*/delete",
    "Microsoft.Authorization/*/write"
  ],
  "NotActions": [],
  "AssignableScopes": [
    "/subscriptions/<id>/resourceGroups/ml-rg/providers/Microsoft.MachineLearningServices/workspaces/ml-ws"
  ]
}
```

D)

```
{
  "Name": "DataScientist",
  "IsCustom": true,
  "Description": "Project Data Scientist role",
  "Actions": [],
  "NotActions": ["*"],
  "AssignableScopes": [
    "/subscriptions/<id>/resourceGroups/ml-rg/providers/Microsoft.MachineLearningServices/workspaces/ml-ws"
  ]
}
```

- A. Option A
- B. Option B
- C. Option C
- D. Option D

**Answer:** A

#### Explanation:

The following custom role can do everything in the workspace except for the following actions:

- It can't create or update a compute resource.
- It can't delete a compute resource.
- It can't add, delete, or alter role assignments.
- It can't delete the workspace.

To create a custom role, first construct a role definition JSON file that specifies the permission and scope for the role. The following example defines a custom role named "Data Scientist Custom" scoped at a specific workspace level:

data\_scientist\_custom\_role.json :

```
{
  "Name": "Data Scientist Custom", "IsCustom": true,
  "Description": "Can run experiment but can't create or delete compute.", "Actions": ["*"],
  "NotActions": [
    "Microsoft.MachineLearningServices/workspaces/*/delete", "Microsoft.MachineLearningServices/workspaces/write",
    "Microsoft.MachineLearningServices/workspaces/computes/*/write", "Microsoft.MachineLearningServices/workspaces/computes/*/delete",
    "Microsoft.Authorization/*/write"
  ],
  "AssignableScopes": [ "/subscriptions/<subscription_id>/resourceGroups/<resource_group_name>/providers/Microsoft.MachineLearnin"
]
```

Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/how-to-assign-roles>

#### NEW QUESTION 34

- (Exam Topic 3)

You are a lead data scientist for a project that tracks the health and migration of birds. You create a multi-image classification deep learning model that uses a set of labeled bird photos collected by experts. You plan to use the model to develop a cross-platform mobile app that predicts the species of bird captured by app users.

You must test and deploy the trained model as a web service. The deployed model must meet the following requirements:

- An authenticated connection must not be required for testing.
- The deployed model must perform with low latency during inferencing.



> The REST endpoints must be scalable and should have a capacity to handle large number of requests when multiple end users are using the mobile application.

You need to verify that the web service returns predictions in the expected JSON format when a valid REST request is submitted.

Which compute resources should you use? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

## Context

## Resource

Test

	▼
ds-workstation notebook VM	
aks-compute cluster	
cpu-compute cluster	
gpu-compute cluster	

Production

	▼
ds-workstation notebook VM	
aks-compute cluster	
cpu-compute cluster	
gpu-compute cluster	

- A. Mastered
- B. Not Mastered

**Answer:** A

### Explanation:

Box 1: ds-workstation notebook VM

An authenticated connection must not be required for testing.

On a Microsoft Azure virtual machine (VM), including a Data Science Virtual Machine (DSVM), you create local user accounts while provisioning the VM. Users then authenticate to the VM by using these credentials.

Box 2: gpu-compute cluster

Image classification is well suited for GPU compute clusters

Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/data-science-virtual-machine/dsvm-common-identity> <https://docs.microsoft.com/en-us/azure/architecture/reference-architectures/ai/training-deep-learning>

### NEW QUESTION 39

- (Exam Topic 3)

You are creating a machine learning model. You need to identify outliers in the data.

Which two visualizations can you use? Each correct answer presents a complete solution.

NOTE: Each correct selection is worth one point. NOTE: Each correct selection is worth one point.

- A. box plot
- B. scatter
- C. random forest diagram
- D. Venn diagram
- E. ROC curve

**Answer:** AB

### Explanation:

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

One other way to quickly identify Outliers visually is to create scatter plots. References:

<https://blogs.msdn.microsoft.com/azuredev/2017/05/27/data-cleansing-tools-in-azure-machine-learning/>

### NEW QUESTION 41

- (Exam Topic 3)

You are using the Hyperdrive feature in Azure Machine Learning to train a model. You configure the Hyperdrive experiment by running the following code:

```
from azureml.train.hyperdrive import RandomParameterSampling
param_sampling = RandomParameterSampling( {
    "learning_rate": normal(10, 3),
    "keep_probability": uniform(0.05, 0.1),
    "batch_size": choice(16, 32, 64, 128)
    "number_of_hidden_layers": choice(range(3,5))
})
```

For each of the following statements, select Yes if the statement is true. Otherwise, select No.

NOTE: Each correct selection is worth one point.

	Yes	No
By defining sampling in this manner, every possible combination of the parameters will be tested.	<input type="radio"/>	<input type="radio"/>
Random values of the learning_rate parameter will be selected from a normal distribution with a mean of 10 and a standard deviation of 3.	<input type="radio"/>	<input type="radio"/>
The keep_probability parameter value will always be either <b>0.05</b> or <b>0.1</b> .	<input type="radio"/>	<input type="radio"/>
Random values for the number_of_hidden_layers parameter will be selected from a normal distribution with a mean of 3 and a standard deviation of 5.	<input type="radio"/>	<input type="radio"/>

- A. Mastered  
 B. Not Mastered

**Answer:** A

**Explanation:**

Box 1: Yes

In random sampling, hyperparameter values are randomly selected from the defined search space. Random sampling allows the search space to include both discrete and continuous hyperparameters.

Box 2: Yes

learning\_rate has a normal distribution with mean value 10 and a standard deviation of 3.

Box 3: No

keep\_probability has a uniform distribution with a minimum value of 0.05 and a maximum value of 0.1.

Box 4: No

number\_of\_hidden\_layers takes on one of the values [3, 4, 5].

Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/how-to-tune-hyperparameters>

**NEW QUESTION 44**

- (Exam Topic 3)

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

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

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

Which parameter sweep mode should you use?

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

**Answer:** D

**Explanation:**

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

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

**NEW QUESTION 49**

- (Exam Topic 3)

You create a multi-class image classification deep learning model that uses a set of labeled images. You create a script file named train.py that uses the PyTorch 1.3 framework to train the model.

You must run the script by using an estimator. The code must not require any additional Python libraries to be installed in the environment for the estimator. The time required for model training must be minimized.

You need to define the estimator that will be used to run the script. Which estimator type should you use?

- A. TensorFlow  
 B. PyTorch  
 C. SKLearn  
 D. Estimator

**Answer:** B

**Explanation:**

For PyTorch, TensorFlow and Chainer tasks, Azure Machine Learning provides respective PyTorch, TensorFlow, and Chainer estimators to simplify using these frameworks.

Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/how-to-train-ml-models>

### NEW QUESTION 50

- (Exam Topic 3)

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

```
from azureml.pipeline.core import Pipeline
```

```
from azureml.core.experiment import Experiment
```

```
pipeline = Pipeline(workspace=ws, steps=[parallelrun_step]) pipeline_run = Experiment(ws, 'batch_pipeline').submit(pipeline)
```

You need to monitor the progress of the pipeline execution.

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

NOTE: Each correct selection is worth one point.

A. Run the following code in a notebook:

```
from azureml.contrib.interpret.explanation.explanation_client import ExplanationClient
client = ExplanationClient.from_run(pipeline_run)
explanation = client.download_model_explanation()
explanation = client.download_model_explanation(top_k=4)
global_importance_values = explanation.get_ranked_global_values()
global_importance_names = explanation.get_ranked_global_names()
print('global importance values: {}'.format(global_importance_values))
print('global importance names: {}'.format(global_importance_names))
```

B. Use the Inference Clusters tab in Machine Learning Studio.

C. Use the Activity log in the Azure portal for the Machine Learning workspace.

D. Run the following code in a notebook:

```
from azureml.widgets import RunDetails
RunDetails(pipeline_run).show()
```

E. Run the following code and monitor the console output from the PipelineRun object:

```
pipeline_run.wait_for_completion(show_output=True)
```

A. Option A

B. Option B

C. Option C

D. Option D

E. Option E

**Answer:** DE

#### Explanation:

A batch inference job can take a long time to finish. This example monitors progress by using a Jupyter widget. You can also manage the job's progress by using:

➤ Azure Machine Learning Studio.

➤ Console output from the PipelineRun object. `from azureml.widgets import RunDetails RunDetails(pipeline_run).show()`

`pipeline_run.wait_for_completion(show_output=True)` Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/how-to-use-parallel-run-step#monitor-the-parallel-run>

### NEW QUESTION 54

- (Exam Topic 3)

You create an Azure Machine Learning workspace and a new Azure DevOps organization. You register a



model in the workspace and deploy the model to the target environment.

All new versions of the model registered in the workspace must automatically be deployed to the target environment.

You need to configure Azure Pipelines to deploy the model.

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.

Actions	Answer Area
Create a service connection	
Create a release pipeline	
Create a build pipeline	
Create an Azure DevOps project	
Install the Machine Learning extension for Azure Pipelines	

A. Mastered

B. Not Mastered

**Answer:** A

#### Explanation:

Graphical user interface, text, application, email Description automatically generated

Step 1: Create an Azure DevOps project

Step 2: Create a release pipeline



- > Sign in to your Azure DevOps organization and navigate to your project.
- > Go to Pipelines, and then select New pipeline.

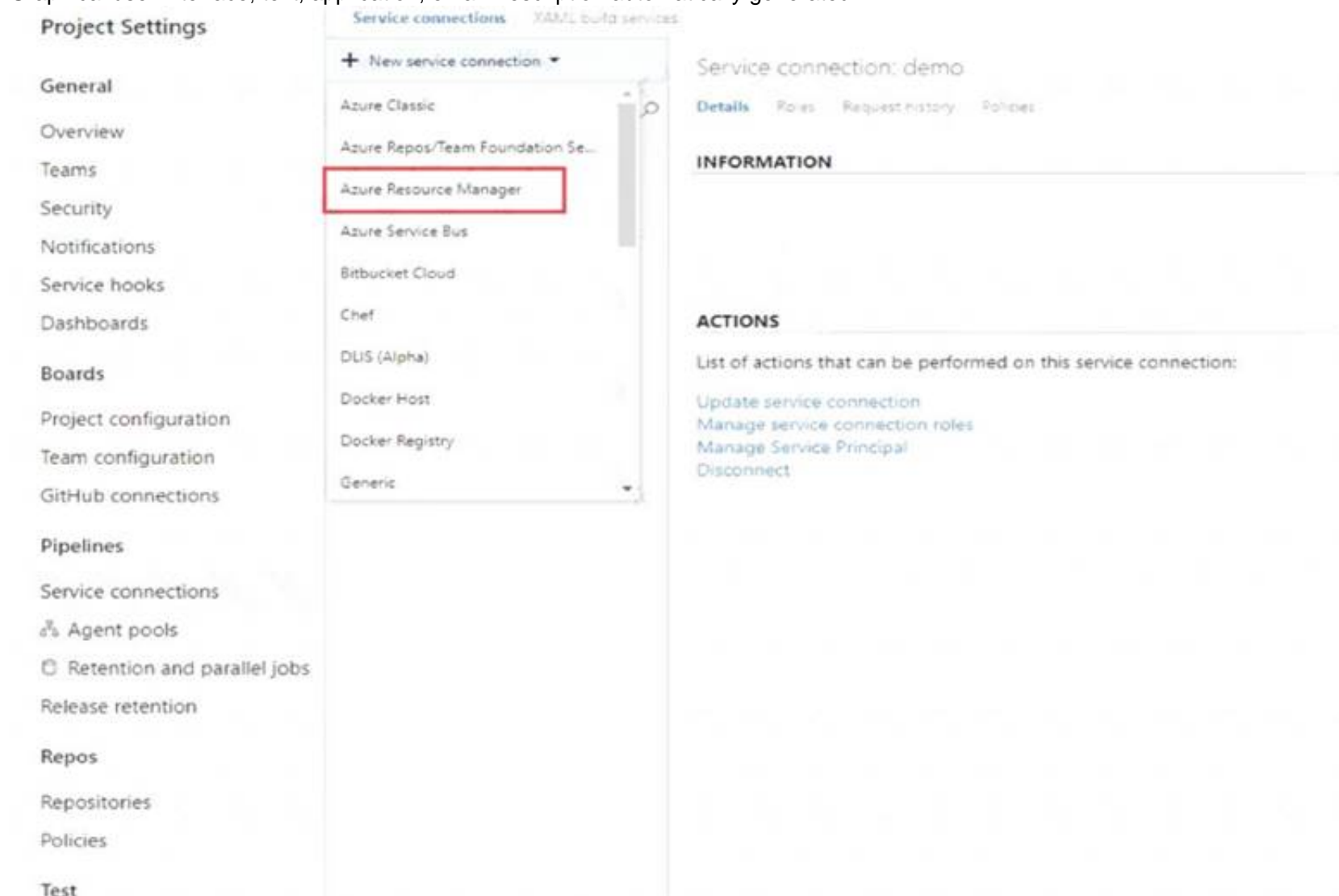
Step 3: Install the Machine Learning extension for Azure Pipelines

You must install and configure the Azure CLI and ML extension.

Step 4: Create a service connection

How to set up your service connection

Graphical user interface, text, application, email Description automatically generated



Select AzureMLWorkspace for the scope level, then fill in the following subsequent parameters. Graphical user interface, text, application Description automatically generated

Note: How to enable model triggering in a release pipeline

- > Go to your release pipeline and add a new artifact. Click on AzureML Model artifact then select the appropriate AzureML service connection and select from the available models in your workspace.
- > Enable the deployment trigger on your model artifact as shown here. Every time a new version of that model is registered, a release pipeline will be triggered.

Reference:

<https://marketplace.visualstudio.com/items?itemName=ms-air-aiagility.vss-services-azureml> <https://docs.microsoft.com/en-us/azure/devops/pipelines/targets/azure-machine-learning>

## NEW QUESTION 57

- (Exam Topic 3)

You create a datastore named training\_data that references a blob container in an Azure Storage account. The blob container contains a folder named csv\_files in which multiple comma-separated values (CSV) files are stored.

You have a script named train.py in a local folder named ./script that you plan to run as an experiment using an estimator. The script includes the following code to read data from the csv\_files folder:

```
import os
import argparse
import pandas as pd

from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from azureml.core import Run

run = Run.get_context()
parser = argparse.ArgumentParser()
parser.add_argument('--data-folder', type=str, dest='data_folder', help='data reference')
args = parser.parse_args()

data_folder = args.data_folder
csv_files = os.listdir(data_folder)
training_data = pd.concat((pd.read_csv(os.path.join(data_folder, csv_file)) for csv_file in csv_files))

# Code goes on to split the training data and train a logistic regression model
```

You have the following script.

```
from azureml.core import Workspace, Datastore, Experiment
from azureml.train.sklearn import SKLearn

ws = Workspace.from_config()
exp = Experiment(workspace=ws, name='csv_training')
ds = Datastore.get(ws, datastore_name='training_data')
data_ref = ds.path('csv_files')

# Code to define estimator goes here

run = exp.submit(config=estimator)
run.wait_for_completion(show_output=True)
```

You need to configure the estimator for the experiment so that the script can read the data from a data reference named data\_ref that references the csv\_files folder in the training\_data datastore.

Which code should you use to configure the estimator?

- A. 

```
estimator = SKLearn(source_directory='./script',
    inputs=[data_ref.as_named_input('data-folder').to_pandas_dataframe()],
    compute_target='local',
    entry_script='train.py')
```
- B. 

```
script_params = {
    '--data-folder': data_ref.as_mount()
}
estimator = SKLearn(source_directory='./script',
    script_params=script_params,
    compute_target='local',
    entry_script='train.py')
```
- C. 

```
estimator = SKLearn(source_directory='./script',
    inputs=[data_ref.as_named_input('data-folder').as_mount()],
    compute_target='local',
    entry_script='train.py')
```
- D. 

```
script_params = {
    '--data-folder': data_ref.as_download(path_on_compute='csv_files')
}
estimator = SKLearn(source_directory='./script',
    script_params=script_params,
    compute_target='local',
    entry_script='train.py')
```
- E. 

```
estimator = SKLearn(source_directory='./script',
    inputs=[data_ref.as_named_input('data-folder').as_download(path_on_compute='csv_files')],
    compute_target='local',
    entry_script='train.py')
```

- A. Option A
- B. Option B
- C. Option C
- D. Option D
- E. Option E

**Answer:** B

#### Explanation:

Besides passing the dataset through the inputs parameter in the estimator, you can also pass the dataset through script\_params and get the data path (mounting point) in your training script via arguments. This way, you can keep your training script independent of azureml-sdk. In other words, you will be able use the same training script for local debugging and remote training on any cloud platform.

Example:

```
from azureml.train.sklearn import SKLearn
script_params = {
    # mount the dataset on the remote compute and pass the mounted path as an argument to the training script '--data-folder':
    mnist_ds.as_named_input('mnist').as_mount(),
    '--regularization': 0.5
}
est = SKLearn(source_directory=script_folder, script_params=script_params, compute_target=compute_target, environment_definition=env,
    entry_script='train_mnist.py')
# Run the experiment
run = experiment.submit(est)
run.wait_for_completion(show_output=True)
```

Reference:  
<https://docs.microsoft.com/es-es/azure/machine-learning/how-to-train-with-datasets>

#### NEW QUESTION 60

- (Exam Topic 3)


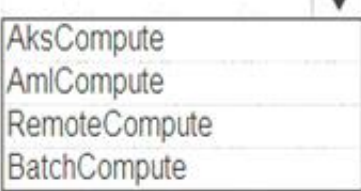



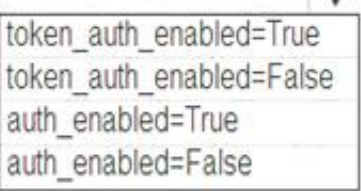
You use Azure Machine Learning to train and register a model.

You must deploy the model into production as a real-time web service to an inference cluster named service-compute that the IT department has created in the Azure Machine Learning workspace.

Client applications consuming the deployed web service must be authenticated based on their Azure Active Directory service principal.

You need to write a script that uses the Azure Machine Learning SDK to deploy the model. The necessary modules have been imported.

How should you complete the code? To answer, select the appropriate options in the answer area.  
 NOTE: Each correct selection is worth one point.

```
# Assume the necessary modules have been imported
deploy_target =  (ws, "service-compute")

deployment_config =  .deploy_configuration(cpu_cores=1, memory_gb=1,

 )

service = Model.deploy(ws, "ml-service",
    [model], inference_config, deployment_config, deploy_target)
service.wait_for_deployment(show_output = True)
```

- A. Mastered
- B. Not Mastered

**Answer: A**

**Explanation:**

Box 1: AksCompute Example:

aks\_target = AksCompute(ws,"myaks")

# If deploying to a cluster configured for dev/test, ensure that it was created with enough  
 # cores and memory to handle this deployment configuration. Note that memory is also used by  
 # things such as dependencies and AML components.

deployment\_config = AksWebservice.deploy\_configuration(cpu\_cores = 1, memory\_gb = 1)

service = Model.deploy(ws, "myservice", [model], inference\_config, deployment\_config, aks\_target) Box 2: AksWebservice

Box 3: token\_auth\_enabled=Yes

Whether or not token auth is enabled for the Webservice.

Note: A Service principal defined in Azure Active Directory (Azure AD) can act as a principal on which authentication and authorization policies can be enforced in Azure Databricks.

The Azure Active Directory Authentication Library (ADAL) can be used to programmatically get an Azure AD access token for a user.

Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/how-to-deploy-azure-kubernetes-service> <https://docs.microsoft.com/en-us/azure/databricks/dev-tools/api/latest/aad/service-prin-aad-token>

**NEW QUESTION 63**

- (Exam Topic 3)

You are building a machine learning model for translating English language textual content into French language textual content.

You need to build and train the machine learning model to learn the sequence of the textual content. Which type of neural network should you use?

- A. Multilayer Perceptions (MLPs)
- B. Convolutional Neural Networks (CNNs)
- C. Recurrent Neural Networks (RNNs)
- D. Generative Adversarial Networks (GANs)

**Answer: C**

**Explanation:**

To translate a corpus of English text to French, we need to build a recurrent neural network (RNN).

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.

References:

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

**NEW QUESTION 66**

- (Exam Topic 3)

You use the following code to run a script as an experiment in Azure Machine Learning:



```
from azureml.core import Workspace, Experiment, Run
from azureml.core import RunConfig, ScriptRunConfig
ws = Workspace.from_config()
run_config = RunConfiguration()
run_config.target='local'
script_config = ScriptRunConfig(source_directory='./script', script='experiment.py', run_config=run_config)
experiment = Experiment(workspace=ws, name='script experiment')
run = experiment.submit(config=script_config)
run.wait_for_completion()
```

You must identify the output files that are generated by the experiment run. You need to add code to retrieve the output file names. Which code segment should you add to the script?

- A. files = run.get\_properties()
- B. files= run.get\_file\_names()
- C. files = run.get\_details\_with\_logs()
- D. files = run.get\_metrics()
- E. files = run.get\_details()

**Answer: B**

**Explanation:**

You can list all of the files that are associated with this run record by called run.get\_file\_names() Reference:  
<https://docs.microsoft.com/en-us/azure/machine-learning/how-to-track-experiments>

**NEW QUESTION 67**

- (Exam Topic 3)

You are the owner of an Azure Machine Learning workspace.

You must prevent the creation or deletion of compute resources by using a custom role. You must allow all other operations inside the workspace.

You need to configure the custom role.

How should you complete the configuration? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

**Answer Area**

```
{
  "Name": "Data Scientist Custom",
  "IsCustom": true
  "Description": "Description"
  "Actions": [
    Microsoft.MachineLearningServices/workspaces/*/read
    Microsoft.MachineLearningServices/workspaces/computes/*/write
    Microsoft.MachineLearningServices/workspaces/delete
    Microsoft.MachineLearningServices/workspaces/*/write
    Microsoft.MachineLearningServices/workspaces/computes/*/write
    Microsoft.MachineLearningServices/workspaces/delete
  ],
  "NotActions": [
    Microsoft.MachineLearningServices/workspaces/*/read
    Microsoft.MachineLearningServices/workspaces/*/write
    Microsoft.MachineLearningServices/workspaces/computes/*/delete
    Microsoft.MachineLearningServices/workspaces/*/read
    Microsoft.MachineLearningServices/workspaces/*/write
    Microsoft.MachineLearningServices/workspaces/computes/*/write
  ],
  "AssignableScopes": [
    "/subscriptions/<subscription_id>"
  ]
}
```

- A. Mastered
- B. Not Mastered

**Answer: A**

**Explanation:**

Graphical user interface, application Description automatically generated

Graphical user interface, application Description automatically generated

Box 1: Microsoft.MachineLearningServices/workspaces/\*/read

Reader role: Read-only actions in the workspace. Readers can list and view assets, including datastore credentials, in a workspace. Readers can't create or update these assets.

Box 2: Microsoft.MachineLearningServices/workspaces/\*/write

If the roles include Actions that have a wildcard (\*), the effective permissions are computed by subtracting the NotActions from the allowed Actions.

Box 3: Box 2: Microsoft.MachineLearningServices/workspaces/computes/\*/delete

Box 4: Microsoft.MachineLearningServices/workspaces/computes/\*/write Reference:  
<https://docs.microsoft.com/en-us/azure/role-based-access-control/overview#how-azure-rbac-determines-if-a-use>

**NEW QUESTION 69**  
- (Exam Topic 3)

You create a multi-class image classification deep learning experiment by using the PyTorch framework. You plan to run the experiment on an Azure Compute cluster that has nodes with GPU's.

You need to define an Azure Machine Learning service pipeline to perform the monthly retraining of the image classification model. The pipeline must run with minimal cost and minimize the time required to train the model.

Which three pipeline steps 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.

Actions

Configure a DataTransferStep() to fetch new image data from public web portal, running on the cpu-compute compute target.

Configure an EstimatorStep() to run an estimator that runs the bird\_classifier\_train.py model training script on the gpu\_compute compute target.

Configure a PythonScriptStep() to run both image\_fetcher.py and image\_resize.py on the cpu-compute compute target.

Configure an EstimatorStep() to run an estimator that runs the bird\_classifier\_train.py model training script on the cpu\_compute compute target.

Configure a PythonScriptStep() to run image\_fetcher.py on the cpu-compute compute target.

Configure a PythonScriptStep() to run image\_resize.py on the cpu-compute compute target.

Configure a PythonScriptStep() to run bird\_classifier\_train.py on the cpu-compute compute target.

Configure a PythonScriptStep() to run bird\_classifier\_train.py on the gpu-compute compute target.

Answer Area

- A. Mastered
- B. Not Mastered

Answer: A

**Explanation:**

Step 1: Configure a DataTransferStep() to fetch new image data...

Step 2: Configure a PythonScriptStep() to run image\_resize.y on the cpu-compute compute target. Step 3: Configure the EstimatorStep() to run training script on the gpu\_compute computer target.

The PyTorch estimator provides a simple way of launching a PyTorch training job on a compute target. Reference:  
<https://docs.microsoft.com/en-us/azure/machine-learning/how-to-train-pytorch>

**NEW QUESTION 73**  
- (Exam Topic 3)

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

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

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

One class has a much smaller number of observations than tin- 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 Principal Components Analysis (PCA) 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 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.

**NEW QUESTION 77**  
- (Exam Topic 3)

You create an Azure Machine Learning workspace named ML-workspace. You also create an Azure Databricks workspace named DB-workspace. DB-workspace contains a cluster named DB-cluster.

You must use DB-cluster to run experiments from notebooks that you import into DB-workspace.

You need to use ML-workspace to track MLflow metrics and artifacts generated by experiments running on DB-cluster. The solution must minimize the need for custom code.

What should you do?

- A. From DB-cluster, configure the Advanced Logging option.
- B. From DB-workspac
- C. configure the Link Azure ML workspace option.
- D. From ML-workspac
- E. create an attached compute.
- F. From ML-workspac
- G. create a compute cluster.

**Answer:** B

**Explanation:**

Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/how-to-use-mlflow-azure-databricks>

**NEW QUESTION 78**

- (Exam Topic 3)

You are using a Git repository to track work in an Azure Machine Learning workspace. You need to authenticate a Git account by using SSH.

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

Actions

Generate a public/private key pair

Add the private key to the Git account

Clone the Git repository by using an SSH repository URL

Add the public key to the Git account

Create a new Azure Key Vault resource

Answer Area

- A. Mastered
- B. Not Mastered

**Answer:** A

**Explanation:**

Graphical user interface, text, application, chat or text message Description automatically generated

Authenticate your Git Account with SSH: Step 1: Generating a public/private key pair Generate a new SSH key

\* 1. Open the terminal window in the Azure Machine Learning Notebook Tab.

\* 2. Paste the text below, substituting in your email address. `ssh-keygen -t rsa -b 4096 -C "your_email@example.com"`

This creates a new ssh key, using the provided email as a label.

> Generating public/private rsa key pair.

Step 2: Add the public key to the Git Account

In your terminal window, copy the contents of your public key file. Step 3: Clone the Git repository by using an SSH repository URL

\* 1. Copy the SSH Git clone URL from the Git repo.

\* 2. Paste the url into the git clone command below, to use your SSH Git repo URL. This will look something like:

`git clone git@example.com:GitUser/azureml-example.git` Cloning into 'azureml-example'.

Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/concept-train-model-git-integration>

**NEW QUESTION 81**

- (Exam Topic 3)

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

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

What should you recommend?

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

**Answer:** B

**Explanation:**

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



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

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

**NEW QUESTION 85**

- (Exam Topic 3)

You plan to deliver a hands-on workshop to several students. The workshop will focus on creating data visualizations using Python. Each student will use a device that has internet access.

Student devices are not configured for Python development. Students do not have administrator access to install software on their devices. Azure subscriptions are not available for students.

You need to ensure that students can run Python-based data visualization code. Which Azure tool should you use?

- A. Anaconda Data Science Platform
- B. Azure BatchAI
- C. Azure Notebooks
- D. Azure Machine Learning Service

**Answer:** C

**Explanation:**  
References:  
<https://notebooks.azure.com/>

**NEW QUESTION 87**

- (Exam Topic 3)

You are using a decision tree algorithm. You have trained a model that generalizes well at a tree depth equal to 10.

You need to select the bias and variance properties of the model with varying tree depth values.

Which properties should you select for each tree depth? To answer, select the appropriate options in the answer area.

**Answer Area**

Tree Depth	Bias	Variance
5	<div><div></div><div>High</div><div>Low</div><div>Identical</div></div>	<div><div></div><div>High</div><div>Low</div><div>Identical</div></div>
15	<div><div></div><div>High</div><div>Low</div><div>Identical</div></div>	<div><div></div><div>High</div><div>Low</div><div>Identical</div></div>

- A. Mastered
- B. Not Mastered

**Answer:** A

**Explanation:**  
In decision trees, the depth of the tree determines the variance. A complicated decision tree (e.g. deep) has low bias and high variance.

Note: In statistics and machine learning, the bias–variance tradeoff is the property of a set of predictive models whereby models with a lower bias in parameter estimation have a higher variance of the parameter estimates across samples, and vice versa. Increasing the bias will decrease the variance. Increasing the variance will decrease the bias.

References:  
<https://machinelearningmastery.com/gentle-introduction-to-the-bias-variance-trade-off-in-machine-learning/>

**NEW QUESTION 90**

- (Exam Topic 3)

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

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

You need to select appropriate methods for producing the experiment.

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

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

## Actions

## Answer Area

Build Counting Transform

Missing Values Scrubber

Feature Hashing

Clean Missing Data

Replace Discrete Values

Import Data

Latent Dirichlet Transformation

Partition and Sample



- A. Mastered
- B. Not Mastered

**Answer:** A

### Explanation:

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

### NEW QUESTION 95

- (Exam Topic 3)

You register a file dataset named csvfolder that references a folder. The folder includes multiple comma-separated values (CSV) files in an Azure storage blob container. You plan to use the following code to run a script that loads data from the file dataset. You create and instantiate the following variables:

Variable	Description
remote_cluster	References the Azure Machine Learning compute cluster
ws	References the Azure Machine Learning workspace

You have the following code:

```
from azureml.train.estimator import Estimator
file_dataset = ws.datasets.get('csv_folder')
estimator = Estimator(source_directory=script_folder,

                        compute_target = remote_cluster,
                        entry_script='script.py')
run = experiment.submit(config=estimator)
run.wait_for_completion(show_output=True)
```

You need to pass the dataset to ensure that the script can read the files it references. Which code segment should you insert to replace the code comment?

A)

```
inputs=[file_dataset.as_named_input('training_files').to_pandas_dataframe()],
```

B)

```
inputs=[file_dataset.as_named_input('training_files').as_mount()],
```

C)

```
script_params={'--training_files': file_dataset},
```

D)

```
inputs=[file_dataset.as_named_input('training_files')],
```

- A. Option A
- B. Option B
- C. Option C
- D. Option D

**Answer:** D

### Explanation:

Example:

```
from azureml.train.estimator import Estimator script_params = {
```

```
# to mount files referenced by mnist dataset
'--data-folder': mnist_file_dataset.as_named_input('mnist_opendataset').as_mount(),
'--regularization': 0.5
}
est = Estimator(source_directory=script_folder, script_params=script_params, compute_target=compute_target, environment_definition=env,
entry_script='train.py')
Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/tutorial-train-models-with-aml
```

**NEW QUESTION 96**

- (Exam Topic 3)  
You create a multi-class image classification deep learning model.  
The model must be retrained monthly with the new image data fetched from a public web portal. You create an Azure Machine Learning pipeline to fetch new data, standardize the size of images, and retrain the model.  
You need to use the Azure Machine Learning SDK to configure the schedule for the pipeline.  
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.

Actions

Publish the pipeline.

Retrieve the pipeline ID.

Create a ScheduleRecurrence(frequency= 'Month', interval=1, start\_time='2019-01-01T00:00:00') object.

Define a pipeline parameter named **RunDate**.

Define a new Azure Machine Learning pipeline StepRun object with the step ID of the first step in the pipeline.

Define an Azure Machine Learning pipeline schedule using the schedule.create method with the defined recurrence specification.

Answer Area

<

>

↑

↓

- A. Mastered
- B. Not Mastered

**Answer:** A

**Explanation:**  
Step 1: Publish the pipeline.  
To schedule a pipeline, you'll need a reference to your workspace, the identifier of your published pipeline, and the name of the experiment in which you wish to create the schedule.  
Step 2: Retrieve the pipeline ID. Needed for the schedule.  
Step 3: Create a ScheduleRecurrence..  
To run a pipeline on a recurring basis, you'll create a schedule. A Schedule associates a pipeline, an experiment, and a trigger.  
First create a schedule. Example: Create a Schedule that begins a run every 15 minutes: recurrence = ScheduleRecurrence(frequency="Minute", interval=15)  
Step 4: Define an Azure Machine Learning pipeline schedule.. Example, continued:  
recurring\_schedule = Schedule.create(ws, name="MyRecurringSchedule", description="Based on time", pipeline\_id=pipeline\_id, experiment\_name=experiment\_name, recurrence=recurrence)  
Reference:  
<https://docs.microsoft.com/en-us/azure/machine-learning/how-to-schedule-pipelines>

**NEW QUESTION 99**

- (Exam Topic 3)  
You plan to use Hyperdrive to optimize the hyperparameters selected when training a model. You create the following code to define options for the hyperparameter experiment



```
import azureml.train.hyperdrive.parameter_expressions as pe
from azureml.train.hyperdrive import GridParameterSampling, HyperDriveConfig

param_sampling = GridParameterSampling({
    "max_depth" : pe.choice(6, 7, 8, 9),
    "learning_rate" : pe.choice(0.05, 0.1, 0.15)
})

hyperdrive_run_config = HyperDriveConfig(
    estimator = estimator,
    hyperparameter_sampling = param_sampling,
    policy = None,
    primary_metric_name = "auc",
    primary_metric_goal = PrimaryMetricGoal.MAXIMIZE,

    estimator = estimator,
    hyperparameter_sampling = param_sampling,
    policy = None,
    primary_metric_name = "auc",
    primary_metric_goal = PrimaryMetricGoal.MAXIMIZE,
    max_total_runs = 50,
    max_concurrent_runs = 4)
```

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

#### Answer Area

	Yes	No
There will be 50 runs for this hyperparameter tuning experiment.	<input type="radio"/>	<input type="radio"/>
You can use the policy parameter in the HyperDriveConfig class to specify a security policy.	<input type="radio"/>	<input type="radio"/>
The experiment will create a run for every possible value for the learning rate parameter between 0.05 and 0.15.	<input type="radio"/>	<input type="radio"/>

- A. Mastered  
B. Not Mastered

Answer: A

#### Explanation:

Box 1: No

max\_total\_runs (50 here)

The maximum total number of runs to create. This is the upper bound; there may be fewer runs when the sample space is smaller than this value.

Box 2: Yes

Policy EarlyTerminationPolicy

The early termination policy to use. If None - the default, no early termination policy will be used. Box 3: No

Discrete hyperparameters are specified as a choice among discrete values. choice can be: one or more comma-separated values

> a range object

> any arbitrary list object Reference:

<https://docs.microsoft.com/en-us/python/api/azureml-train-core/azureml.train.hyperdrive.hyperdriveconfig> <https://docs.microsoft.com/en-us/azure/machine-learning/how-to-tune-hyperparameters>

#### NEW QUESTION 103

- (Exam Topic 3)

You train and register a machine learning model. You create a batch inference pipeline that uses the model to generate predictions from multiple data files.

You must publish the batch inference pipeline as a service that can be scheduled to run every night. You need to select an appropriate compute target for the inference service.

Which compute target should you use?

- A. Azure Machine Learning compute instance  
B. Azure Machine Learning compute cluster  
C. Azure Kubernetes Service (AKS)-based inference cluster  
D. Azure Container Instance (ACI) compute target

Answer: B

#### Explanation:

Azure Machine Learning compute clusters is used for Batch inference. Run batch scoring on serverless compute. Supports normal and low-priority VMs. No support for real-time inference.

Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/concept-compute-target>

### NEW QUESTION 107

- (Exam Topic 3)

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

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

You create the following Python data frames:

Name	Description
X_train	training feature set
Y_train	training class labels
x_train	testing feature set
y_train	testing class labels

You need to apply the Principal Component Analysis (PCA) method to reduce the dimensionality of the feature set to 10 features in both training and testing sets.

How should you complete the code segment? To answer, select the appropriate options in the answer area. NOTE: Each correct selection is worth one point.

```
from sklearn.decomposition import PCA
pca = PCA()
PCA(n_components = 150)
PCA(n_components = 10)
PCA(n_components = 10000)
X_train = pca
model
sklearn.decomposition.fit_transform(X_train)
x_test = pca.x_test
X_train
fit(x_test)
transform(x_test)
```

- A. Mastered  
B. Not Mastered

**Answer:** A

#### Explanation:

Box 1: PCA(n\_components = 10)

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

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

Box 2: pca

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

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

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

### NEW QUESTION 109

- (Exam Topic 3)

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 AI 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.  
E. Create a Data Science Virtual Machine (DSVM) Linux edition.

**Answer:** E

### NEW QUESTION 112

- (Exam Topic 3)

You use the Azure Machine Learning SDK in a notebook to run an experiment using a script file in an experiment folder.

The experiment fails.

You need to troubleshoot the failed experiment.

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

- A. Use the get\_metrics() method of the run object to retrieve the experiment run logs.  
B. Use the get\_details\_with\_logs() method of the run object to display the experiment run logs.  
C. View the log files for the experiment run in the experiment folder.  
D. View the logs for the experiment run in Azure Machine Learning studio.  
E. Use the get\_output() method of the run object to retrieve the experiment run logs.

**Answer:** BD

#### Explanation:

Use `get_details_with_logs()` to fetch the run details and logs created by the run.

You can monitor Azure Machine Learning runs and view their logs with the Azure Machine Learning studio. Reference:

<https://docs.microsoft.com/en-us/python/api/azureml-pipeline-core/azureml.pipeline.core.steprun> <https://docs.microsoft.com/en-us/azure/machine-learning/how-to-monitor-view-training-logs>

#### NEW QUESTION 114

- (Exam Topic 3)

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

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

You create an Azure Machine Learning service datastore in a workspace. The datastore contains the following files:

- /data/2018/Q1 .csv
- /data/2018/Q2.csv
- /data/2018/Q3.csv
- /data/2018/Q4.csv
- /data/2019/Q1.csv

All files store data in the following format: id,M,f2,l

1,1,2,0

2,1,1,1

32,10

You run the following code:

```
data_store = Datastore.register_azure_blob_container(workspace=ws,
datastore_name='data_store',
container_name='quarterly_data',
account_name='companydata',
account_key='NRPxk8duxbM3...'
create_if_not_exists=False)
```

You need to create a dataset named `training_data` and load the data from all files into a single data frame by using the following code:

```
data_frame = training_data.to_pandas_dataframe()
```

Solution: Run the following code:

```
from azureml.core import Dataset
paths = (data_store, 'data/*/*.csv')
training_data = Dataset.Tabular.from_delimited_files(paths)
```

Does the solution meet the goal?

- A. Yes
- B. No

**Answer: A**

#### NEW QUESTION 116

- (Exam Topic 3)

You are building a recurrent neural network to perform a binary classification. You review the training loss, validation loss, training accuracy, and validation accuracy for each training epoch.

You need to analyze model performance.

Which observation indicates that the classification model is over fitted?

- A. The training loss stays constant and the validation loss stays on a constant value and close to the training loss value when training the model.
- B. The training loss increases while the validation loss decreases when training the model.
- C. The training loss decreases while the validation loss increases when training the model.
- D. The training loss stays constant and the validation loss decreases when training the model.

**Answer: B**

#### NEW QUESTION 121

- (Exam Topic 3)

You plan to explore demographic data for home ownership in various cities. The data is in a CSV file with the following format:

age,city,income,home\_owner 21,Chicago,50000,0 35,Seattle,120000,1 23,Seattle,65000,0 45,Seattle,130000,1 18,Chicago,48000,0

You need to run an experiment in your Azure Machine Learning workspace to explore the data and log the results. The experiment must log the following information:

- > the number of observations in the dataset
- > a box plot of income by home\_owner
- > a dictionary containing the city names and the average income for each city

You need to use the appropriate logging methods of the experiment's run object to log the required information.

How should you complete the code? To answer, drag the appropriate code segments to the correct locations. Each code segment 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.



## Code segments

log  
log\_list  
log\_row  
log\_table  
log\_image

## Answer Area

```
from azureml.core import Experiment, Run
import pandas as pd
import matplotlib.pyplot as plt
# Create an Azure ML experiment in workspace
experiment = Experiment(workspace = ws, name = "demo-experiment")
# Start logging data from the experiment
run = experiment.start_logging()
# load the dataset
data = pd.read_csv('research/demographics.csv')
# Log the number of observations
row_count = (len(data))
run. Segment ("observations", row_count)
# Log box plot for income by home_owner
fig = plt.figure(figsize=(9, 6))
ax = fig.gca()
data.boxplot(column = 'income', by = "home_owner", ax = ax)
ax.set_title('income by home_owner')
ax.set_ylabel('income')
run. Segment (name = 'income_by_home_owner', plot = fig)
# Create a dataframe of mean income per city
mean_inc_df = data.groupby('city')['income'].agg(np.mean).to_frame().reset_index()
# Convert to a dictionary
mean_inc_dict = mean_inc_df.to_dict('dict')
# Log city names and average income dictionary
run. Segment (name="mean_income_by_city", value= mean_inc_dict)
# Complete tracking and get link to details
run.complete()
```

- A. Mastered  
B. Not Mastered

**Answer:** A

### Explanation:

Box 1: log

The number of observations in the dataset. run.log(name, value, description="")

Scalar values: Log a numerical or string value to the run with the given name. 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)

Box 2: log\_image

A box plot of income by home\_owner.

log\_image Log an image to the run record. Use log\_image to log a .PNG image file or a matplotlib plot to the run. These images will be visible and comparable in the run record.

Example: run.log\_image("ROC", plot=plt) Box 3: log\_table

A dictionary containing the city names and the average income for each city. log\_table: Log a dictionary object to the run with the given name.

## NEW QUESTION 123

- (Exam Topic 3)

You deploy a model as an Azure Machine Learning real-time web service using the following code.

```
# ws, model, inference_config, and deployment_config defined previously
service = Model.deploy(ws, 'classification-service', [model], inference_config, deployment_config)
service.wait_for_deployment(True)
```

The deployment fails.

You need to troubleshoot the deployment failure by determining the actions that were performed during deployment and identifying the specific action that failed.

Which code segment should you run?

- A. service.get\_logs()  
B. service.state  
C. service.serialize()  
D. service.update\_deployment\_state()

**Answer:** A

### Explanation:

You can print out detailed Docker engine log messages from the service object. You can view the log for ACI, AKS, and Local deployments. The following example demonstrates how to print the logs.

# if you already have the service object handy print(service.get\_logs())

# if you only know the name of the service (note there might be multiple services with the same name but different version number)

print(ws.webservices['mysvc'].get\_logs()) Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/how-to-troubleshoot-deployment>

## NEW QUESTION 126

- (Exam Topic 3)

HOTSPOT

You have an Azure blob container that contains a set of TSV files. The Azure blob container is registered as a datastore for an Azure Machine Learning service workspace. Each TSV file uses the same data schema. You plan to aggregate data for all of the TSV files together and then register the aggregated data as a dataset in an Azure Machine Learning workspace by using the Azure Machine Learning SDK for Python. You run the following code.

```
from azureml.core.workspace import Workspace
from azureml.core.datastore import Datastore
from azureml.core.dataset import Dataset
import pandas as pd
datastore_paths = (datastore, './data/*.tsv')
myDataset_1 = Dataset.File.from_files(path=datastore_paths)
myDataset_2 = Dataset.Tabular.from_delimited_files(path=datastore_paths, separator='\t')
```

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

	Yes	No
The myDataset_1 dataset can be converted into a pandas dataframe by using the following method: using myDataset_1.to_pandas_dataframe()	<input type="radio"/>	<input type="radio"/>
The myDataset_1.to_path() method returns an array of file paths for all of the TSV files in the dataset.	<input type="radio"/>	<input type="radio"/>
The myDataset_2 dataset can be converted into a pandas dataframe by using the following method: myDataset_2.to_pandas_dataframe()	<input type="radio"/>	<input type="radio"/>

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Box 1: No  
FileDataset references single or multiple files in datastores or from public URLs. The TSV files need to be parsed.  
Box 2: Yes  
to\_path() gets a list of file paths for each file stream defined by the dataset. Box 3: Yes  
TabularDataset.to\_pandas\_dataframe loads all records from the dataset into a pandas DataFrame. TabularDataset represents data in a tabular format created by parsing the provided file or list of files.  
Note: TSV is a file extension for a tab-delimited file used with spreadsheet software. TSV stands for Tab Separated Values. TSV files are used for raw data and can be imported into and exported from spreadsheet software. TSV files are essentially text files, and the raw data can be viewed by text editors, though they are often used when moving raw data between spreadsheets.  
Reference:  
<https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.data.tabulardataset>

NEW QUESTION 128

- (Exam Topic 3)  
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. D

- 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:

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

NEW QUESTION 129

- (Exam Topic 3)  
You have a Python data frame named salesData in the following format:

	shop	2017	2018
0	Shop X	34	25
1	Shop Y	65	76
2	Shop Z	48	55

The data frame must be unpivoted to a long data format as follows:

	shop	year	value
0	Shop X	2017	34
1	Shop Y	2017	65
2	Shop Z	2017	48
3	Shop X	2018	25
4	Shop Y	2018	76
5	Shop Z	2018	55

You need to use the pandas.melt() function in Python to perform the transformation.

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.

**Answer Area**

```
import pandas as pd
salesData = pd.melt(
```

▼

dataFrame

pandas

salesData

year

, id\_vars='

▼

shop

year

value

Shop X, Shop Y, Shop Z

', value\_vars='

▼

'shop'

'year'

['year']

['2017', '2018']

) )

- A. Mastered
- B. Not Mastered

**Answer: A**

**Explanation:**

Box 1: dataFrame

Syntax: pandas.melt(frame, id\_vars=None, value\_vars=None, var\_name=None, value\_name='value', col\_level=None)[source]

Where frame is a DataFrame Box 2: shop

Parameter id\_vars id\_vars : tuple, list, or ndarray, optional Column(s) to use as identifier variables.

Box 3: ['2017','2018']

value\_vars : tuple, list, or ndarray, optional

Column(s) to unpivot. If not specified, uses all columns that are not set as id\_vars. Example:

df = pd.DataFrame({'A': {0: 'a', 1: 'b', 2: 'c'},

'B': {0: 1, 1: 3, 2: 5},

'C': {0: 2, 1: 4, 2: 6}})

pd.melt(df, id\_vars=['A'], value\_vars=['B', 'C']) A variable value

0 a B 1

1 b B 3

2 c B 5

3 a C 2

4 b C 4

5 c C 6

References:

<https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.melt.html>

**NEW QUESTION 132**

- (Exam Topic 3)

You are planning to register a trained model in an Azure Machine Learning workspace.

You must store additional metadata about the model in a key-value format. You must be able to add new metadata and modify or delete metadata after creation.

You need to register the model. Which parameter should you use?

- A. description
- B. model\_framework
- C. cags
- D. properties

**Answer: D**

**Explanation:**

azureml.core.Model.properties:

Dictionary of key value properties for the Model. These properties cannot be changed after registration, however new key value pairs can be added.

Reference:

<https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.core.model.model>

**NEW QUESTION 134**

- (Exam Topic 3)

Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the



stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution. After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen. You are using Azure Machine Learning to run an experiment that trains a classification model. You want to use Hyperdrive to find parameters that optimize the AUC metric for the model. You configure a HyperDriveConfig for the experiment by running the following code:

```
hyperdrive = HyperDriveConfig(estimator=your_estimator,
                              hyperparameter_sampling=your_params,
                              policy=policy,
                              primary_metric_name='AUC',
                              primary_metric_goal=PrimaryMetricGoal.MAXIMIZE,
                              max_total_runs=6,
                              max_concurrent_runs=4)
```

variable named y\_test variable, and the predicted probabilities from the model are stored in a variable named y\_predicted. You need to add logging to the script to allow Hyperdrive to optimize hyperparameters for the AUC metric. Solution: Run the following code:

```
from sklearn.metrics import roc_auc_score
import logging
# code to train model omitted
auc = roc_auc_score(y_test, y_predicted)
logging.info("AUC: " + str(auc))
```

Does the solution meet the goal?

- A. Yes
- B. No

**Answer: A**

**Explanation:**

Python printing/logging example: logging.info(message)

Destination: Driver logs, Azure Machine Learning designer

Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/how-to-debug-pipelines>

**NEW QUESTION 139**

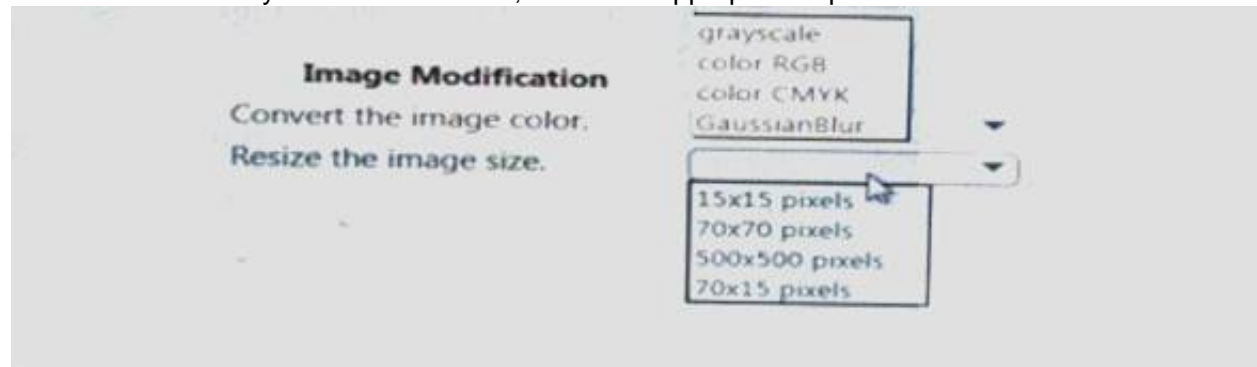
- (Exam Topic 3)

You are training a deep learning model to identify cats and dogs. You have 25,000 color images. You must meet the following requirements:

- Reduce the number of training epochs.
- Reduce the size of the neural network.
- Reduce over-fitting of the neural network.

You need to select the image modification values.

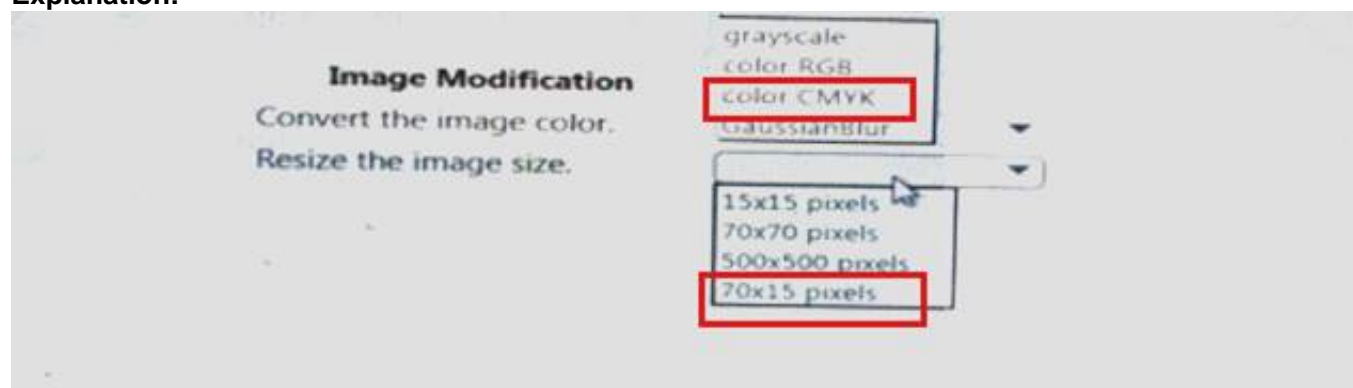
Which value should you use? To answer, select the appropriate Options in the answer area. NOTE: Each correct selection is worth one point.



- A. Mastered
- B. Not Mastered

**Answer: A**

**Explanation:**



**NEW QUESTION 144**

- (Exam Topic 3)

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

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

An IT department creates the following Azure resource groups and resources:

Resource group	Resources
ml_resources	<ul style="list-style-type: none"> <li>• an Azure Machine Learning workspace named amlworkspace</li> <li>• an Azure Storage account named amlworkspace12345</li> <li>• an Application Insights instance named amlworkspace54321</li> <li>• an Azure Key Vault named amlworkspace67890</li> <li>• an Azure Container Registry named amlworkspace09876</li> </ul>
general_compute	A virtual machine named mlvm with the following configuration: <ul style="list-style-type: none"> <li>• Operating system: Ubuntu Linux</li> <li>• Software installed: Python 3.6 and Jupyter Notebooks</li> <li>• Size: NC6 (6 vCPUs, 1 vGPU, 56 Gb RAM)</li> </ul>

The IT department creates an Azure Kubernetes Service (AKS)-based inference compute target named aks-cluster in the Azure Machine Learning workspace. You have a Microsoft Surface Book computer with a GPU. Python 3.6 and Visual Studio Code are installed. You need to run a script that trains a deep neural network (DNN) model and logs the loss and accuracy metrics.

Solution: Install the Azure ML SDK on the Surface Book. Run Python code to connect to the workspace and then run the training script as an experiment on local compute.

- A. Yes  
B. No

**Answer: B**

**Explanation:**

Need to attach the mlvm virtual machine as a compute target in the Azure Machine Learning workspace. Reference: <https://docs.microsoft.com/en-us/azure/machine-learning/concept-compute-target>

**NEW QUESTION 146**

- (Exam Topic 3)

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 value count for each feature column.

Which two modules can you users? Each correct answer presents a complete solution. NOTE: Each correct selection is worth one point.

- A. Execute Python Script  
B. Export Count Table  
C. Convert to Indicator Values  
D. Summarize Data  
E. Compute linear Correlation

**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.

References:

<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>

**NEW QUESTION 148**

- (Exam Topic 3)

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

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

An IT department creates the following Azure resource groups and resources:

Resource group	Resources
ml_resources	<ul style="list-style-type: none"> <li>• an Azure Machine Learning workspace named amlworkspace</li> <li>• an Azure Storage account named amlworkspace12345</li> <li>• an Application Insights instance named amlworkspace54321</li> <li>• an Azure Key Vault named amlworkspace67890</li> <li>• an Azure Container Registry named amlworkspace09876</li> </ul>
general_compute	A virtual machine named mlvm with the following configuration: <ul style="list-style-type: none"> <li>• Operating system: Ubuntu Linux</li> <li>• Software installed: Python 3.6 and Jupyter Notebooks</li> <li>• Size: NC6 (6 vCPUs, 1 vGPU, 56 Gb RAM)</li> </ul>

The IT department creates an Azure Kubernetes Service (AKS)-based inference compute target named aks-cluster in the Azure Machine Learning workspace.

You have a Microsoft Surface Book computer with a GPU. Python 3.6 and Visual Studio Code are installed. You need to run a script that trains a deep neural network (DNN) model and logs the loss and accuracy metrics.

Solution: Attach the mlvm virtual machine as a compute target in the Azure Machine Learning workspace. Install the Azure ML SDK on the Surface Book and run Python code to connect to the workspace. Run the training script as an experiment on the mlvm remote compute resource.

- A. Yes
- B. No

Answer: A

**Explanation:**

Use the VM as a compute target.  
Note: A compute target is a designated compute resource/environment where you run your training script or host your service deployment. This location may be your local machine or a cloud-based compute resource.  
Reference:  
<https://docs.microsoft.com/en-us/azure/machine-learning/concept-compute-target>

**NEW QUESTION 149**

- (Exam Topic 3)  
You must store data in Azure Blob Storage to support Azure Machine Learning. You need to transfer the data into Azure Blob Storage. What are three possible ways to achieve the goal? Each correct answer presents a complete solution.  
NOTE: Each correct selection is worth one point.

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

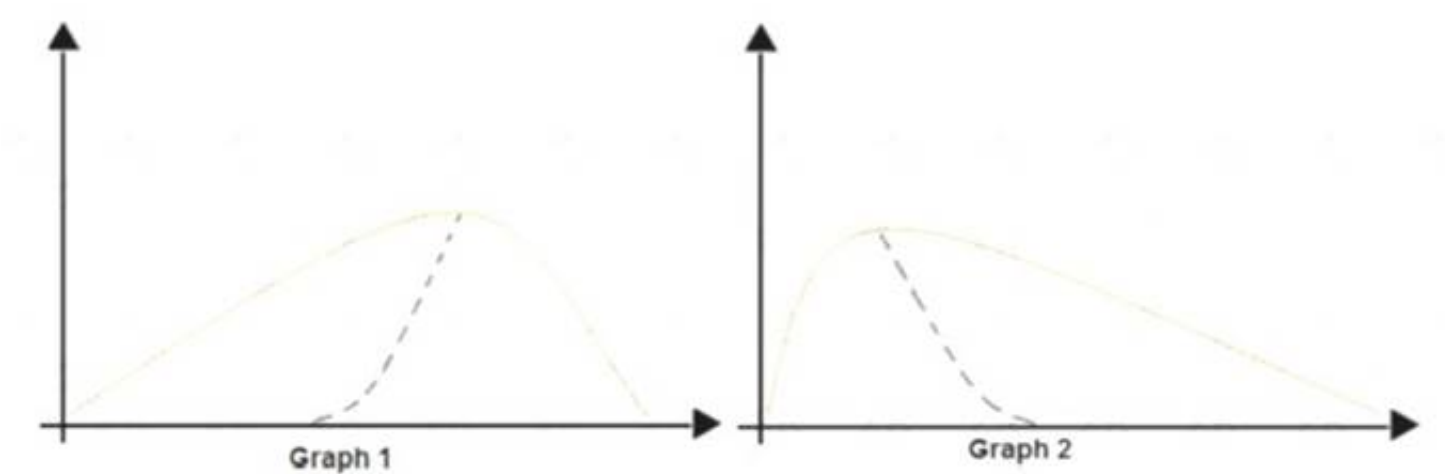
Answer: BCD

**Explanation:**

You can move data to and from Azure Blob storage using different technologies: Azure Storage-Explorer  
AzCopy Python SSIS  
References:  
<https://docs.microsoft.com/en-us/azure/machine-learning/team-data-science-process/move-azure-blob>

**NEW QUESTION 151**

- (Exam Topic 3)  
You are analyzing the asymmetry in a statistical distribution.  
The following image contains two density curves that show the probability distribution of two datasets.



Use the drop-down menus to select the answer choice that answers each question based on the information presented in the graphic.  
NOTE: Each correct selection is worth one point.

Question	Answer choice
Which type of distribution is shown for the dataset density curve of Graph 1?	<div><div></div><div>Negative skew</div><div>Positive skew</div><div>Normal distribution</div><div>Bimodal distribution</div></div>
Which type of distribution is shown for the dataset density curve of Graph 2?	<div><div></div><div>Negative skew</div><div>Positive skew</div><div>Normal distribution</div><div>Bimodal distribution</div></div>

- A. Mastered
- B. Not Mastered

Answer: A

**Explanation:**

Box 1: Positive skew  
Positive skew values means the distribution is skewed to the right. Box 2: Negative skew



Negative skewness values mean the distribution is skewed to the left. References:  
<https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/compute-elementary-statistic>

#### NEW QUESTION 156

- (Exam Topic 3)

You are creating a new experiment in Azure Machine Learning Studio. You have a small dataset that has missing values in many columns. The data does not require the application of predictors for each column. You plan to use the Clean Missing Data module to handle the missing data. You need to select a data cleaning method. Which method should you use?

- A. Synthetic Minority
- B. Replace using Probabilistic PAC
- C. Replace using MICE
- D. Normalization

**Answer: B**

#### NEW QUESTION 157

- (Exam Topic 3)

You are evaluating a Python NumPy array that contains six data points defined as follows: data = [10, 20, 30, 40, 50, 60]

You must generate the following output by using the k-fold algorithm implementation in the Python Scikit-learn machine learning library:

train: [10 40 50 60], test: [20 30]

train: [20 30 40 60], test: [10 50]

train: [10 20 30 50], test: [40 60]

You need to implement a cross-validation to generate the output.

How should you complete the code segment? To answer, select the appropriate code segment in the dialog box in the answer area.

NOTE: Each correct selection is worth one point.

```
from numpy import array
from sklearn.model_selection import K-Means
k-fold
CrossValidation
ModelSelection

data = array([10, 20, 30, 40, 50, 60])
kfold = Kfold(n_splits=1
2
3
6, shuffle = True, random_state=1)

for train, test in kFold, split(data
k-fold
array
train, test):

print('train: %s, test: %s' % (data[train], data[test]))
```

- A. Mastered
- B. Not Mastered

**Answer: A**

#### Explanation:

Box 1: k-fold

Box 2: 3

K-F olds cross-validator provides train/test indices to split data in train/test sets. Split dataset into k consecutive folds (without shuffling by default).

The parameter n\_splits ( int, default=3) is the number of folds. Must be at least 2. Box 3: data

Example: Example:

>>>

>>> from sklearn.model\_selection import KFold

>>> X = np.array([[1, 2], [3, 4], [1, 2], [3, 4]])

>>> y = np.array([1, 2, 3, 4])

>>> kf = KFold(n\_splits=2)

>>> kf.get\_n\_splits(X) 2

>>> print(kf)

KFold(n\_splits=2, random\_state=None, shuffle=False)

>>> for train\_index, test\_index in kf.split(X): print("TRAIN:", train\_index, "TEST:", test\_index) X\_train, X\_test = X[train\_index], X[test\_index] y\_train, y\_test =

y[train\_index], y[test\_index] TRAIN: [2 3] TEST: [0 1]

TRAIN: [0 1] TEST: [2 3]

References:

[https://scikit-learn.org/stable/modules/generated/sklearn.model\\_selection.KFold.html](https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.KFold.html)

#### NEW QUESTION 161

- (Exam Topic 3)

You train and register a model in your Azure Machine Learning workspace.

You must publish a pipeline that enables client applications to use the model for batch inferencing. You must use a pipeline with a single ParallelRunStep step that

runs a Python inferencing script to get predictions from the input data.  
You need to create the inferencing script for the ParallelRunStep pipeline step.  
Which two functions should you include? Each correct answer presents part of the solution.  
NOTE: Each correct selection is worth one point.

- A. run(mini\_batch) D
- B. main()
- C. batch()
- D. init()
- E. score(mini\_batch)

**Answer:** AD

**Explanation:**  
Reference:  
<https://github.com/Azure/MachineLearningNotebooks/tree/master/how-to-use-azureml/machine-learningpipelin>

**NEW QUESTION 166**

- (Exam Topic 3)  
HOTSPOT

You collect data from a nearby weather station. You have a pandas dataframe named weather\_df that includes the following data:

Temperature	Observation_time	Humidity	Pressure	Visibility	Days_since_last observation
74	2019/10/2 00:00	0.62	29.87	3	0.5
89	2019/10/2 12:00	0.70	28.88	10	0.5
72	2019/10/3 00:00	0.64	30.00	8	0.5
80	2019/10/3 12:00	0.66	29.75	7	0.5

The data is collected every 12 hours: noon and midnight.  
You plan to use automated machine learning to create a time-series model that predicts temperature over the next seven days. For the initial round of training, you want to train a maximum of 50 different models.  
You must use the Azure Machine Learning SDK to run an automated machine learning experiment to train these models.  
You need to configure the automated machine learning run.  
How should you complete the AutoMLConfig definition? To answer, select the appropriate options in the answer area.  
NOTE: Each correct selection is worth one point.

```

automl_config = AutoMLConfig(task="
                                regression
                                forecasting
                                classification
                                deep learning

                                training_data=weather_df,
                                label_column_name="
                                humidity
                                pressure
                                visibility
                                temperature
                                days_since_last
                                observation_time

                                time_column_name="
                                humidity
                                pressure
                                visibility
                                temperature
                                days_since_last
                                observation_time

                                max_horizon=
                                2
                                6
                                7
                                12
                                14
                                50

                                iterations=
                                2
                                6
                                7
                                12
                                14
                                50

                                iteration_timeout_minutes=5,
                                primary_metric="r2_score")

```

- A. Mastered  
 B. Not Mastered

**Answer:** A

**Explanation:**

Box 1: forecasting

Task: The type of task to run. Values can be 'classification', 'regression', or 'forecasting' depending on the type of automated ML problem to solve.

Box 2: temperature

The training data to be used within the experiment. It should contain both training features and a label column (optionally a sample weights column).

Box 3: observation\_time

time\_column\_name: The name of the time column. This parameter is required when forecasting to specify the datetime column in the input data used for building the time series and inferring its frequency. This setting is being deprecated. Please use forecasting\_parameters instead.

Box 4: 7

"predicts temperature over the next seven days"

max\_horizon: The desired maximum forecast horizon in units of time-series frequency. The default value is 1. Units are based on the time interval of your training data, e.g., monthly, weekly that the forecaster should predict out. When task type is forecasting, this parameter is required.

Box 5: 50

"For the initial round of training, you want to train a maximum of 50 different models."

Iterations: The total number of different algorithm and parameter combinations to test during an automated ML experiment. Reference:

<https://docs.microsoft.com/en-us/python/api/azureml-train-automl-client/azureml.train.automl.automlconfig.auto>

**NEW QUESTION 171**

- (Exam Topic 3)

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.

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		

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Box 1: Clean Missing Data Box 2: SMOTE

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.

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 References:  
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

NEW QUESTION 176

- (Exam Topic 3)

You are using Azure Machine Learning to train machine learning models. You need a compute target on which to remotely run the training script. You run the following Python code:

```
from azureml.core.compute import ComputeTarget, AmlCompute
from azureml.core.compute_target import ComputeTargetException
the_cluster_name = "NewCompute"
config = AmlCompute.provisioning_configuration(vm_size='STANDARD_D2', max_nodes=3)
the_cluster = ComputeTarget.create(ws, the_cluster_name, config)
```

Answer Area	
	Yes No
The compute is created in the same region as the Machine Learning service workspace.	<input checked="" type="checkbox"/> <input type="checkbox"/>
The compute resource created by the code is displayed as a compute cluster in Azure Machine Learning studio.	<input type="checkbox"/> <input type="checkbox"/>
The minimum number of nodes will be zero.	<input type="checkbox"/> <input type="checkbox"/>

- A. Mastered
- B. Not Mastered

Answer: A

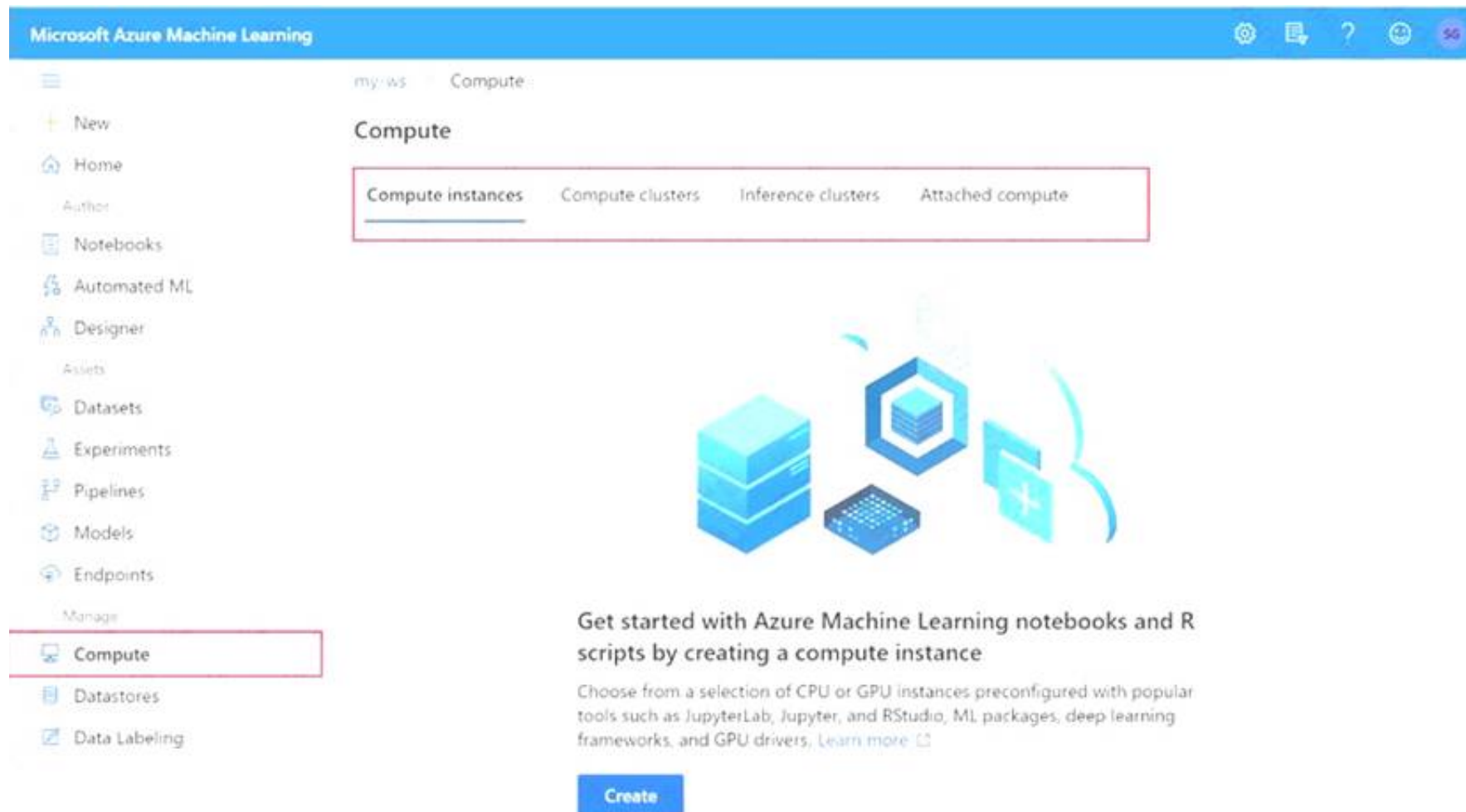
Explanation:

Box 1: Yes

The compute is created within your workspace region as a resource that can be shared with other users. Box 2: Yes

It is displayed as a compute cluster. View compute targets

- \* 1. To see all compute targets for your workspace, use the following steps:
- \* 2. Navigate to Azure Machine Learning studio.
- \* 3. Under Manage, select Compute.
- \* 4. Select tabs at the top to show each type of compute target.



Box 3: Yes

min\_nodes is not specified, so it defaults to 0. Reference:

<https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.core.compute.amlcompute.amlcomputeprovider> <https://docs.microsoft.com/en-us/azure/machine-learning/how-to-create-attach-compute-studio>

#### NEW QUESTION 178

- (Exam Topic 3)

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

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

You plan to use a Python script to run an Azure Machine Learning experiment. The script creates a reference to the experiment run context, loads data from a file, identifies the set of unique values for the label column, and completes the experiment run:

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

run = Run.get_context()
data = pd.read_csv('data.csv')
label_vals = data['label'].unique()
# Add code to record metrics here
run.complete()
```

The experiment must record the unique labels in the data as metrics for the run that can be reviewed later.

You must add code to the script to record the unique label values as run metrics at the point indicated by the comment.

Solution: Replace the comment with the following code: for label\_val in label\_vals:

run.log('Label Values', label\_val) Does the solution meet the goal?

A. Yes

B. No

**Answer: A**

#### Explanation:

The run\_log function is used to log the contents in label\_vals: for label\_val in label\_vals:

run.log('Label Values', label\_val) Reference:

<https://www.element61.be/en/resource/azure-machine-learning-services-complete-toolbox-ai>

#### NEW QUESTION 182

- (Exam Topic 3)

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

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

You plan to use a Python script to run an Azure Machine Learning experiment. The script creates a reference to the experiment run context, loads data from a file, identifies the set of unique values for the label column, and completes the experiment run:

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

run = Run.get_context()
data = pd.read_csv('data.csv')
label_vals = data['label'].unique()
# Add code to record metrics here
run.complete()
```

The experiment must record the unique labels in the data as metrics for the run that can be reviewed later.

You must add code to the script to record the unique label values as run metrics at the point indicated by the comment.

Solution: Replace the comment with the following code:

run.upload\_file('outputs/labels.csv', './data.csv') Does the solution meet the goal?

A. Yes

B. No

**Answer: B**

**Explanation:**

label\_vals has the unique labels (from the statement label\_vals = data['label'].unique()), and it has to be logged.

Note:

Instead use the run\_log function to log the contents in label\_vals: for label\_val in label\_vals:

run.log('Label Values', label\_val) Reference:

<https://www.element61.be/en/resource/azure-machine-learning-services-complete-toolbox-ai>

**NEW QUESTION 186**

- (Exam Topic 3)

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

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

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

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

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

A. Yes

B. No

**Answer: B**

**Explanation:**

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

References:

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

**NEW QUESTION 189**

- (Exam Topic 3)

You publish a batch inferencing pipeline that will be used by a business application.

The application developers need to know which information should be submitted to and returned by the REST interface for the published pipeline.

You need to identify the information required in the REST request and returned as a response from the published pipeline.

Which values should you use in the REST request and to expect in the response? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

**Answer Area**

REST Request	Value
Request Header	<div>JSON containing the run ID JSON containing the pipeline ID JSON containing the experiment name JSON containing an OAuth bearer token</div>
Response	<div>JSON containing the run ID JSON containing the pipeline ID JSON containing the experiment name JSON containing an OAuth bearer token</div>
Response	<div>JSON containing the run ID JSON containing a list of predictions JSON containing the experiment name JSON containing a path to the parallel_run_step.txt output file</div>

A. Mastered

B. Not Mastered

**Answer: A**

**Explanation:**

Box 1: JSON containing an OAuth bearer token Specify your authentication header in the request.

To run the pipeline from the REST endpoint, you need an OAuth2 Bearer-type authentication header. Box 2: JSON containing the experiment name

Add a JSON payload object that has the experiment name. Example:

```
rest_endpoint = published_pipeline.endpoint response = requests.post(rest_endpoint, headers=auth_header, json={"ExperimentName": "batch_scoring", "ParameterAssignments": {"process_count_per_node": 6}}) run_id = response.json()["Id"]
```

Box 3: JSON containing the run ID

Make the request to trigger the run. Include code to access the Id key from the response dictionary to get the value of the run ID.

Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/tutorial-pipeline-batch-scoring-classification>

**NEW QUESTION 192**

- (Exam Topic 3)

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.

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

- A. Mastered
- B. Not Mastered

**Answer:** A

**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/>

**NEW QUESTION 194**

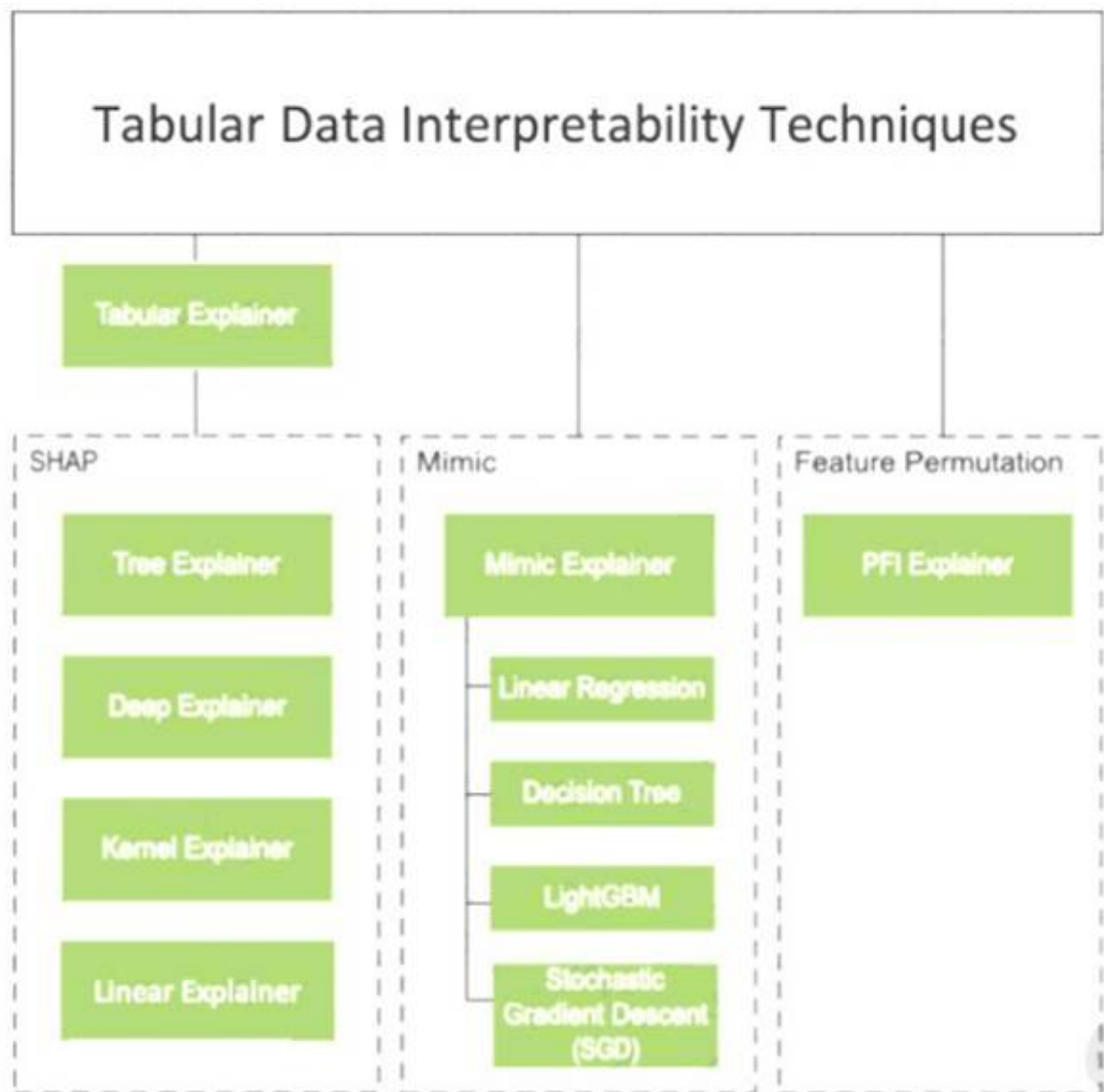
- (Exam Topic 3)  
 Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.  
 After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.  
 You train a classification model by using a logistic regression algorithm.  
 You must be able to explain the model's predictions by calculating the importance of each feature, both as an overall global relative importance value and as a measure of local importance for a specific set of predictions.  
 You need to create an explainer that you can use to retrieve the required global and local feature importance values.  
 Solution: Create a TabularExplainer. Does the solution meet the goal?

- A. Yes
- B. No

**Answer:** B

**Explanation:**

Instead use Permutation Feature Importance Explainer (PFI). Note 1:



Note 2: Permutation Feature Importance Explainer (PFI): Permutation Feature Importance is a technique used to explain classification and regression models. At a high level, the way it works is by randomly shuffling data one feature at a time for the entire dataset and calculating how much the performance metric of interest changes. The larger the change, the more important that feature is. PFI can explain the overall behavior of any underlying model but does not explain individual predictions.

Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/how-to-machine-learning-interpretability>

#### NEW QUESTION 199

- (Exam Topic 3)

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

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

You are a data scientist using Azure Machine Learning Studio.

You need to normalize values to produce an output column into bins to predict a target column. Solution: Apply an Equal Width with Custom Start and Stop binning mode.

Does the solution meet the goal?

- A. Yes
- B. No

**Answer: B**

#### Explanation:

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

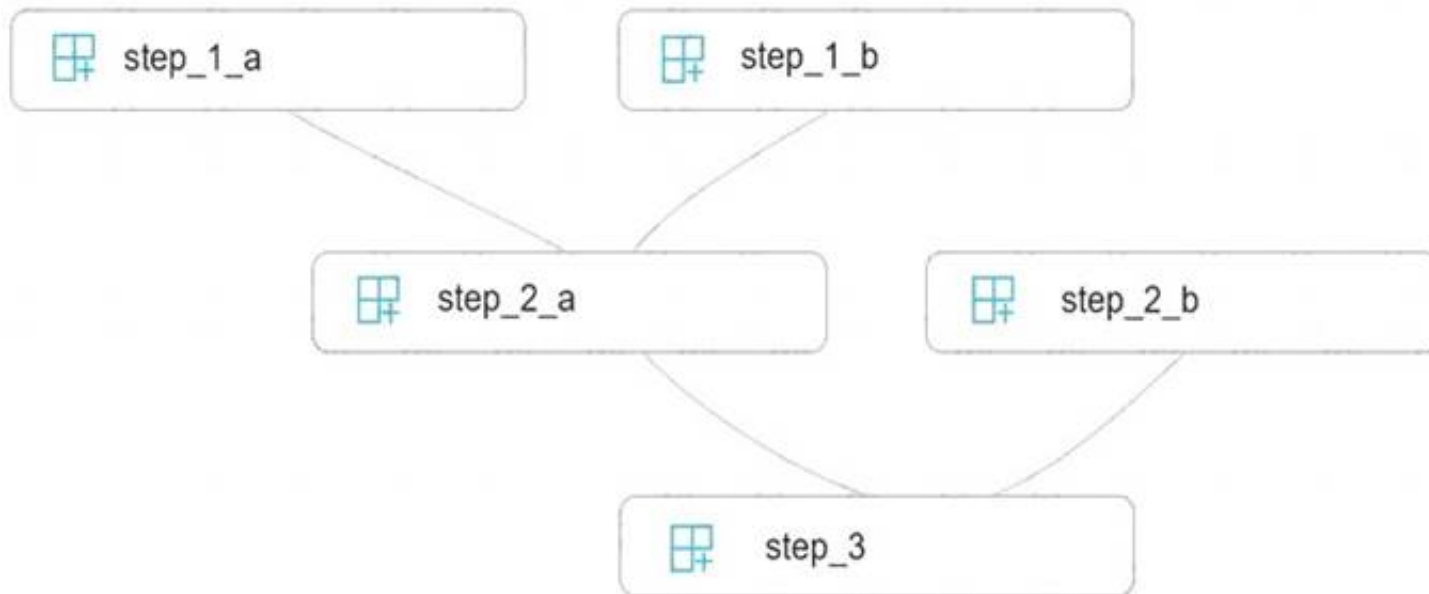
<https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/group-data-into-bins>

#### NEW QUESTION 204

- (Exam Topic 3)

You write five Python scripts that must be processed in the order specified in Exhibit A – which allows the same modules to run in parallel, but will wait for modules with dependencies.

You must create an Azure Machine Learning pipeline using the Python SDK, because you want to script to create the pipeline to be tracked in your version control system. You have created five PythonScriptSteps and have named the variables to match the module names.



You need to create the pipeline shown. Assume all relevant imports have been done. Which Python code segment should you use?

- A. `p = Pipeline(ws, steps=[[[[step_1_a, step_1_b], step_2_a], step_2_b], step_3])`
- B. 

```

pipeline_steps = {
    "Pipeline": {
        "run": step_3,
        "run_after": [{
            {"run": step_2_a,
             "run_after": [
                 {"run": step_1_a},
                 {"run": step_1_b}
             ]},
            {"run": step_2_b}
        ]}
    ]
}
p = Pipeline(ws, steps=pipeline_steps)

```
- C. `step_2_a.run_after(step_1_b)`  
`step_2_a.run_after(step_1_a)`  
`step_3.run_after(step_2_b)`  
`step_3.run_after(step_2_a)`  
`p = Pipeline(ws, steps=[step_3])`
- D. `p = Pipeline(ws, steps=[step_1_a, step_1_b, step_2_a, step_2_b, step_3])`

- A. Option A
- B. Option B
- C. Option C
- D. Option D

**Answer:** A

**Explanation:**

The steps parameter is an array of steps. To build pipelines that have multiple steps, place the steps in order in this array.

Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/how-to-use-parallel-run-step>

**NEW QUESTION 207**

- (Exam Topic 3)

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. Partition and Sample
- B. Assign Data to Clusters
- C. Group Data into Bins
- D. Test Hypothesis Using t-Test

**Answer:** A

**Explanation:**

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

References:

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

**NEW QUESTION 210**

- (Exam Topic 3)

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

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

You train and register a machine learning model.

You plan to deploy the model as a real-time web service. Applications must use key-based authentication to use the model.

You need to deploy the web service. Solution:



Create an AciWebService instance.  
Set the value of the ssl\_enabled property to True.  
Deploy the model to the service. Does the solution meet the goal?

- A. Yes
- B. No

**Answer:** B

**Explanation:**

Instead use only auth\_enabled = TRUE Note: Key-based authentication.  
Web services deployed on AKS have key-based auth enabled by default. ACI-deployed services have key-based auth disabled by default, but you can enable it by setting auth\_enabled = TRUE when creating the ACI web service. The following is an example of creating an ACI deployment configuration with key-based auth enabled.  
deployment\_config <- aci\_webservice\_deployment\_config(cpu\_cores = 1, memory\_gb = 1, auth\_enabled = TRUE) Reference:  
<https://azure.github.io/azureml-sdk-for-r/articles/deploying-models.html>

**NEW QUESTION 215**

- (Exam Topic 3)

You are performing feature engineering on a dataset.  
You must add a feature named CityName and populate the column value with the text London.  
You need to add the new feature to the dataset.  
Which Azure Machine Learning Studio module should you use?

- A. Edit Metadata
- B. Preprocess Text
- C. Execute Python Script
- D. Latent Dirichlet Allocation

**Answer:** A

**Explanation:**

Typical metadata changes might include marking columns as features. References:  
<https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/edit-metadata>

**NEW QUESTION 217**

- (Exam Topic 3)

You plan to preprocess text from CSV files. You load the Azure Machine Learning Studio default stop words list.  
You need to configure the Preprocess Text module to meet the following requirements:

- Ensure that multiple related words from a single canonical form.
- Remove pipe characters from text.
- Remove words to optimize information retrieval.

Which three options should you select? To answer, select the appropriate options in the answer area. NOTE: Each correct selection is worth one point.

#### Preprocess Text

Language  

English

Remove by part of speech  

False

Text column to clean  

**Selected columns:**  
Column names: **String, Feature**

Launch column selector

☐

 Remove stop words

☐

 Lemmatization

☐

 Detect sentences

☐

 Normalize case to lowercase

☐

 Remove numbers

☐

 Remove special characters

☐

 Remove duplicate characters

☐

 Remove email addresses

☐

 Remove URLs

☐

 Expand verb contractions

☐

 Normalize backslashes to slashes

☐

 Split tokens on special characters

- A. Mastered  
B. Not Mastered

**Answer:** A

#### Explanation:

Box 1: Remove stop words

Remove words to optimize information retrieval.

Remove stop words: Select this option if you want to apply a predefined stopwords list to the text column. Stop word removal is performed before any other processes.

Box 2: Lemmatization

Ensure that multiple related words from a single canonical form. Lemmatization converts multiple related words to a single canonical form Box 3: Remove special characters

Remove special characters: Use this option to replace any non-alphanumeric special characters with the pipe | character.

References:

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

#### NEW QUESTION 221

- (Exam Topic 3)

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.

Partition and Sample

Partition or sample mode

▼

Assign to Folds

Pick Fold

Sampling

Head

Rate of sampling

.2

Random seed for sampling

▼

0

1

time.clock()

utcNow()

Stratified split for sampling

False

▼

- A. Mastered
- B. Not Mastered

Answer: A

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.

References:

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

NEW QUESTION 224

- (Exam Topic 3)

You create an Azure Machine Learning workspace and set up a development environment. You plan to train a deep neural network (DNN) by using the Tensorflow framework and by using estimators to submit training scripts.

You must optimize computation speed for training runs.

You need to choose the appropriate estimator to use as well as the appropriate training compute target configuration.

Which values should you use? To answer, select the appropriate options in the answer area. NOTE: Each correct selection is worth one point.

Answer Area

Parameter	Value
Estimator	<div><div>Estimator</div><div>SKLearn</div><div>PyTorch</div><div>Tensorflow</div><div>Chainer</div></div>
Training compute	<div><div>12 vCPU, 48 GB memory, 96 GB SSD</div><div>12 vCPU, 112 GB memory, 680 GB SSD, 2 GPU, 24 GB GPU memory</div><div>16 vCPU, 128 GB memory, 160 GB HDD, 80 GB NVME disk (4000 MBps)</div><div>44 vCPU, 352 GB memory, 3.4 GHz CPU frequency all cores</div></div> <div>Use the selections for training compute</div>

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Box 1: Tensorflow

TensorFlow represents an estimator for training in TensorFlow experiments. Box 2: 12 vCPU, 112 GB memory...,2 GPU,... Use GPUs for the deep neural network. Reference:



<https://docs.microsoft.com/en-us/python/api/azureml-train-core/azureml.train.dnn>

**NEW QUESTION 227**

- (Exam Topic 3)

You create an Azure Machine Learning workspace named workspace1. You assign a custom role to a user of workspace1. The custom role has the following JSON definition:

```
{
  "Name": "MyRole",
  "IsCustom": true,
  "Description": "New custom role description.",
  "Actions": ["*"],
  "NotActions": [
    "Microsoft.MachineLearningServices/workspaces/write",
    "Microsoft.MachineLearningServices/workspaces/computes/*/write",
    "Microsoft.MachineLearningServices/workspaces/computes/*/delete",
    "Microsoft.Authorization/*/write"
  ],
  "AssignableScopes": [
    "/subscriptions/<subscription_id>/resourceGroups/resourcegroup1/providers/
    Microsoft.MachineLearningServices/workspaces/workspace1"
  ]
}
```

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

Statements	Yes	No
The user can perform all actions in the workspace	<input type="radio"/>	<input type="radio"/>
The user can delete a compute resource in the workspace	<input type="radio"/>	<input type="radio"/>
The user can write metrics to the workspace	<input type="radio"/>	<input type="radio"/>

A.

Answer:

**Explanation:**

Graphical user interface, text, application, email Description automatically generated

Box 1: No

The actions listed in NotActions are prohibited.

If the roles include Actions that have a wildcard (\*), the effective permissions are computed by subtracting the NotActions from the allowed Actions.

Box 2: No

Deleting compute resources in the workspace is in the NotActions list. Box 3: Yes

Writing metrics is not listed in NotActions. Reference:

<https://docs.microsoft.com/en-us/azure/role-based-access-control/overview#how-azure-rbac-determines-if-a-use>

**NEW QUESTION 232**

- (Exam Topic 3)

You have a Jupyter Notebook that contains Python code that is used to train a model.

You must create a Python script for the production deployment. The solution must minimize code maintenance.

Which two actions should you perform? Each correct answer presents part of the solution.

NOTE: Each correct selection is worth one point.

- A. Refactor the Jupyter Notebook code into functions
- B. Save each function to a separate Python file
- C. Define a main() function in the Python script
- D. Remove all comments and functions from the Python script

Answer: AC

**Explanation:**

Reference:

<https://www.guru99.com/learn-python-main-function-with-examples-understand-main.html> <https://towardsdatascience.com/from-jupyter-notebook-to-deployment-a-straightforward-example-1838c203a43>

**NEW QUESTION 237**

- (Exam Topic 3)

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

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

You are 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:** A

**Explanation:**

Entropy MDL binning mode: This method requires that you select the column you want to predict and the column or columns that you want to group into bins. It then makes a pass over the data and attempts to determine the number of bins that minimizes the entropy. In other words, it chooses a number of bins that allows the data column to best predict the target column. It then returns the bin number associated with each row of your data in a column named <colname>quantized.

References:

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

**NEW QUESTION 242**

- (Exam Topic 3)

You use Azure Machine Learning designer to create a real-time service endpoint. You have a single Azure Machine Learning service compute resource. You train the model and prepare the real-time pipeline for deployment. You need to publish the inference pipeline as a web service. Which compute type should you use?

- A. HDInsight
- B. Azure Databricks
- C. Azure Kubernetes Services
- D. the existing Machine Learning Compute resource
- E. a new Machine Learning Compute resource

**Answer:** C

**Explanation:**

Azure Kubernetes Service (AKS) can be used for real-time inference. Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/concept-compute-target>

**NEW QUESTION 247**

- (Exam Topic 3)

You create a binary classification model. You need to evaluate the model performance.

Which two metrics can you use? Each correct answer presents a complete solution.

NOTE: Each correct selection is worth one point.

- A. relative absolute error
- B. precision
- C. accuracy
- D. mean absolute error
- E. coefficient of determination

**Answer:** BC

**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.

References:

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

**NEW QUESTION 251**

- (Exam Topic 3)

You are building a regression model to estimate the number of calls during an event.

You need to determine whether the feature values achieve the conditions to build a Poisson regression model. Which two conditions must the feature set contain?

Each correct answer presents part of the solution. NOTE:

Each correct selection is worth one point.

- A. The label data must be a negative value.
- B. The label data can be positive or negative.
- C. The label data must be a positive value.
- D. The label data must be non-discrete.
- E. The data must be whole numbers.

**Answer:** CE

**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.

References:

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

#### NEW QUESTION 254

- (Exam Topic 3)

You create an Azure Machine Learning compute target named ComputeOne by using the STANDARD\_D1 virtual machine image.

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.

	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.	<input type="radio"/>	<input type="radio"/>
Any experiments configured to use <code>the_cluster</code> will run on ComputeOne.	<input type="radio"/>	<input type="radio"/>
The text <b>Step1</b> will be printed to the screen.	<input type="radio"/>	<input type="radio"/>

- A. Mastered
- B. Not Mastered

**Answer:** A

#### Explanation:

Box 1: Yes

ComputeTargetException class: An exception related to failures when creating, interacting with, or configuring a compute target. This exception is commonly raised for failures attaching a compute target,

missing headers, and unsupported configuration values. Create(workspace, name, provisioning\_configuration)

Provision a Compute object by specifying a compute type and related configuration. This method creates a new compute target rather than attaching an existing one. Box 2: Yes

Box 3: No

The line before print('Step1') will fail. Reference:

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

#### NEW QUESTION 257

- (Exam Topic 3)

You are creating a new Azure Machine Learning pipeline using the designer.

The pipeline must train a model using data in a comma-separated values (CSV) file that is published on a website. You have not created a dataset for this file.

You need to ingest the data from the CSV file into the designer pipeline using the minimal administrative effort.

Which module should you add to the pipeline in Designer?

- A. Convert to CSV
- B. Enter Data Manually D
- C. Import Data
- D. Dataset

**Answer:** D

#### Explanation:

The preferred way to provide data to a pipeline is a Dataset object. The Dataset object points to data that lives in or is accessible from a datastore or at a Web URL. The Dataset class is abstract, so you will create an instance of either a FileDataset (referring to one or more files) or a TabularDataset that's created by from one or more files with delimited columns of data.

Example:

```
from azureml.core import Dataset
```

```
iris_tabular_dataset = Dataset.Tabular.from_delimited_files([(def_blob_store, 'train-dataset/iris.csv')])
```

 Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/how-to-create-your-first-pipeline>

#### NEW QUESTION 259

- (Exam Topic 3)

You run an automated machine learning experiment in an Azure Machine Learning workspace. Information about the run is listed in the table below:



Experiment	Run ID	Status	Created on	Duration
auto_ml_classification	AutoML_1234567890-123	Completed	11/11/2019 11:00:00 AM	00:27:11

You need to write a script that uses the Azure Machine Learning SDK to retrieve the best iteration of the experiment run. Which Python code segment should you use?

A)

```
from azureml.core import Workspace
from azureml.train.automl.run import AutoMLRun
ws = Workspace.from_config()
automl_ex = ws.experiments.get('auto_ml_classification')
automl_run = AutoMLRun(automl_ex, 'AutoML_1234567890-123')
best_iter = automl_run.get_output()[0]
```

B)

```
from azureml.core import Workspace
from azureml.train.automl.run import AutoMLRun
ws = Workspace.from_config()
automl_ex = ws.experiments.get('auto_ml_classification')
best_iter = list(automl_ex.get_runs())[0]
```

C)

```
from azureml.core import Workspace
from azureml.train.automl.run import AutoMLRun
ws = Workspace.from_config()
automl_ex = ws.experiments.get('auto_ml_classification')
best_iter = list(automl_ex.get_runs())[0]
```

D)

```
from azureml.core import Workspace
```

- A. Option A
- B. Option B
- C. Option C
- D. Option D

**Answer: A**

**Explanation:**

The get\_output method on automl\_classifier returns the best run and the fitted model for the last invocation. Overloads on get\_output allow you to retrieve the best run and fitted model for any logged metric or for a particular iteration.

In [ ]:

best\_run, fitted\_model = local\_run.get\_output() Reference:

<https://notebooks.azure.com/azureml/projects/azureml-getting-started/html/how-to-use-azureml/automated-mach>

**NEW QUESTION 262**

- (Exam Topic 3)

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

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

You are a data scientist using Azure Machine Learning Studio.

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

Does the solution meet the goal?

- A. Yes
- B. No

**Answer: B**

**Explanation:**

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

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

**NEW QUESTION 267**

- (Exam Topic 3)

You are using an Azure Machine Learning workspace. You set up an environment for model testing and an environment for production.

The compute target for testing must minimize cost and deployment efforts. The compute target for production must provide fast response time, autoscaling of the deployed service, and support real-time inferencing.

You need to configure compute targets for model testing and production.

Which compute targets should you use? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

## Environment

## Compute target

Testing

	▼
Local web service	
Azure Kubernetes Services (AKS)	
Azure Container Instances	
Azure Machine Learning compute clusters	

Production

	▼
Local web service	
Azure Kubernetes Services (AKS)	
Azure Container Instances	
Azure Machine Learning compute clusters	

- A. Mastered
- B. Not Mastered

**Answer:** A

### Explanation:

Text, application Description automatically generated

Box 1: Local web service

The Local web service compute target is used for testing/debugging. Use it for limited testing and troubleshooting. Hardware acceleration depends on use of libraries in the local system.

Box 2: Azure Kubernetes Service (AKS)

Azure Kubernetes Service (AKS) is used for Real-time inference. Recommended for production workloads.

Use it for high-scale production deployments. Provides fast response time and autoscaling of the deployed service

Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/concept-compute-target>

### NEW QUESTION 271

- (Exam Topic 3)

You are a lead data scientist for a project that tracks the health and migration of birds. You create a multi-class image classification deep learning model that uses a set of labeled bird photographs collected by experts.

You have 100,000 photographs of birds. All photographs use the JPG format and are stored in an Azure blob container in an Azure subscription.

You need to access the bird photograph files in the Azure blob container from the Azure Machine Learning service workspace that will be used for deep learning model training. You must minimize data movement.

What should you do?

- A. Create an Azure Data Lake store and move the bird photographs to the store.
- B. Create an Azure Cosmos DB database and attach the Azure Blob containing bird photographs storage to the database.
- C. Create and register a dataset by using TabularDataset class that references the Azure blob storage containing bird photographs.
- D. Register the Azure blob storage containing the bird photographs as a datastore in Azure Machine Learning service.
- E. Copy the bird photographs to the blob datastore that was created with your Azure Machine Learning service workspace.

**Answer:** D

### Explanation:

We recommend creating a datastore for an Azure Blob container. When you create a workspace, an Azure blob container and an Azure file share are automatically registered to the workspace.

Reference:

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

### NEW QUESTION 272

- (Exam Topic 3)

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. subscription\_id

- B. Key
- C. resource\_group
- D. region
- E. Login

**Answer:** AC

**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')
```

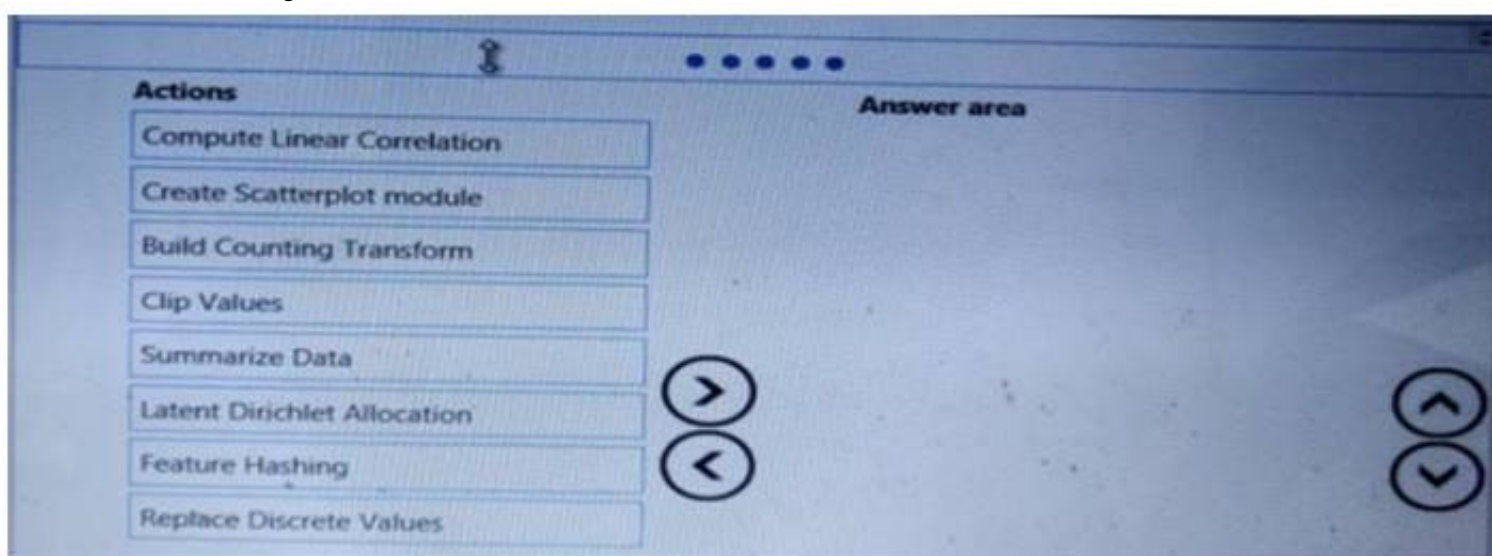
Reference:  
<https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.core.workspace.workspace>

**NEW QUESTION 275**

- (Exam Topic 2)

You need to visually identify whether outliers exist in the Age column and quantify the outliers before the outliers are removed.

Which three Azure Machine Learning Studio modules should you use in sequence? To answer, move the appropriate modules from the list of modules to the answer area and arrange them in the correct order.



- A. Mastered
- B. Not Mastered

**Answer:** A

**Explanation:**

Create Scatterplot Summarize Data Clip Values

You can use the Clip Values module in Azure Machine Learning Studio, to identify and optionally replace data values that are above or below a specified threshold. This is useful when you want to remove outliers or replace them with a mean, a constant, or other substitute value.

References:

<https://blogs.msdn.microsoft.com/azuredev/2017/05/27/data-cleansing-tools-in-azure-machine-learning/> <https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/clip-values>

**NEW QUESTION 276**

- (Exam Topic 2)

You need to replace the missing data in the AccessibilityToHighway columns.

How should you configure the Clean Missing Data module? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.



Properties Project

Clean Missing Data

Columns to be cleaned

Selected columns:

Column names: AccessibilityToHighway

Launch column selector

Minimum missing value ratio

0

Maximum missing value ratio

1

Cleaning mode

Replace using MICE  
Replace with Mean  
Replace with Median  
Replace with Mode

Cols with all missing values.

Propagate  
Remove

☒ Generate missing value indicator column

Number of iterations

5

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Box 1: Replace using MICE

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

Scenario: The AccessibilityToHighway column in both datasets contains missing values. The missing data must be replaced with new data so that it is modeled conditionally using the other variables in the data before filling in the missing values.

Box 2: Propagate

Cols with all missing values indicate if columns of all missing values should be preserved in the output. References:  
<https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/clean-missing-data>

NEW QUESTION 280

.....

## Thank You for Trying Our Product

### We offer two products:

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

2nd - Questions and Answers in PDF Format

### DP-100 Practice Exam Features:

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

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