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

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

You need to write code to access the datastore from a notebook. How should you complete the code segment? To answer, select the appropriate options in the answer area. NOTE : Each correct selection is worth one point.

☐ CHECK BELOW THE RIGHT ANSWER

Answer Area

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

Box 1 options:
- Workspace
- Datastore
- Experiment
- Run

Box 2 options:
- ws
- run
- experiment
- log

Box 3 options:
- demo_datastore
- demo_datacontainer
- demo_account
- Datastore

CORRECT ANSWER:
KEEP OPEN

EXPLANATION:

Explanation: Box 1: DataStore To get a specific datastore registered in the current workspace, use the get() static method on the Datastore class: # Get a named datastore from the current workspace datastore = Datastore.get(ws, datastore_name='your datastore name') Box 2: ws Box 3: demo_datastore Reference: https://docs.microsoft.com/en-us/azure/machine-learning/how-to-access-data

HOTSPOT 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.

☐ CHECK BELOW THE RIGHT ANSWER

Answer Area

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
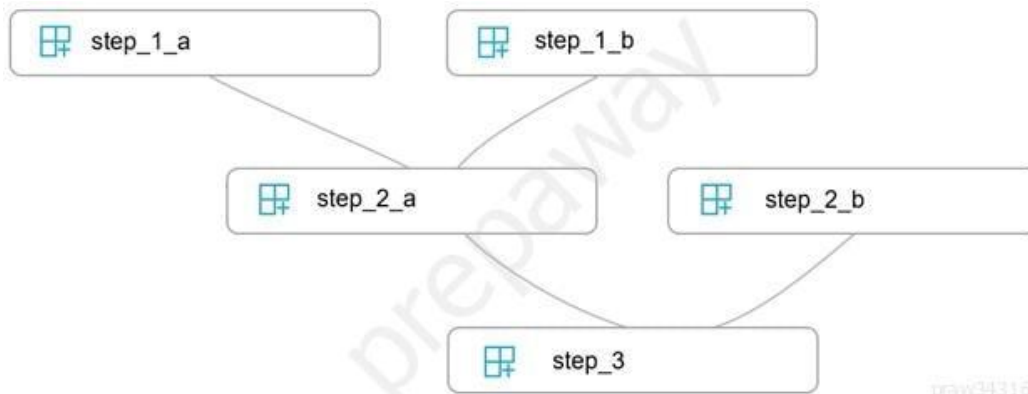Azure Machine Learning compute clusters

CORRECT ANSWER:
KEEP OPEN

EXPLANATION:

Explanation: 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

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

```
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])
```

☐ C

```
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)
```

□ D

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

CORRECT ANSWER: A
KEEP OPEN

EXPLANATION:

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

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

CORRECT ANSWER: B
KEEP OPEN

EXPLANATION:

Explanation: Use the Entropy MDL binning mode which has a target column. Reference: https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/group-data-into-bins

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').as_mount()],
  compute_target='local',
  entry_script='train.py')
```

☐ B

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

☐ C

```
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'
```

☐ D

```
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')
```

☐ E

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

CORRECT ANSWER: E
KEEP OPEN

EXPLANATION:

Explanation: Besides passing the dataset through the input parameters 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) Incorrect Answers: A: Pandas DataFrame not used. Reference: https://docs.microsoft.com/es-es/azure/machine-learning/how-to-train-with-datasets

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

Load Trained Model

☐ B

Tune Model-Hyperparameters

☐ C

Partition and Sample

☐ D

Assign Data to Clusters

CORRECT ANSWER: C
KEEP OPEN

EXPLANATION:

Explanation: Partition and Sample with the Stratified split option outputs multiple datasets, partitioned using the rules you specified. Reference: https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/partition-and-sample

Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution. After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen. You are a data scientist using Azure Machine Learning Studio. You need to normalize values to produce an output column into bins to predict a target column. Solution: Apply a Quantiles normalization with a QuantileIndex normalization. Does the solution meet the goal?

☐ A

No

☐ B

Yes

CORRECT ANSWER: A
KEEP OPEN

EXPLANATION:

Explanation: Use the Entropy MDL binning mode which has a target column. Reference: https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/group-data-into-bins

Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution. After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen. You are analyzing a numerical dataset which contains missing values in several columns. You must clean the missing values using an appropriate operation without affecting the dimensionality of the feature set. You need to analyze a full dataset to include all values. Solution: Remove the entire column that contains the missing data point. Does the solution meet the goal?

☐ A
Yes

☐ B
No

CORRECT ANSWER: B
KEEP OPEN

EXPLANATION:

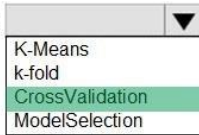Explanation: Use the Multiple Imputation by Chained Equations (MICE) method. Reference: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3074241/ https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/clean-missing-data
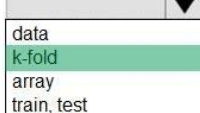
HOTSPOT You are evaluating a Python NumPy array that contains six data points defined as follows: data = [10, 20, 30, 40, 50, 60] You must generate the following output by using the k-fold algorithm implantation in the Python Scikit-learn machine learning library: train: [10 40 50 60], test: [20 30] 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.

☐ CHECK BELOW THE RIGHT ANSWER

Answer Area

```
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-        ▼    ,shuffle - True, random_state-1)
                        ┌──────────┐
                        │ 1        │
                        │ 2        │
                        │ 3        │
                        │ 6        │
                        └──────────┘
for train, test in kFold, split(        ▼  ):
                                ┌──────────────┐
                                │ data         │
                                │ k-fold       │
                                │ array        │
                                │ train, test  │
                                └──────────────┘
print('train: %s, test: %5' % (data[train], data[test])
```

CORRECT ANSWER:
KEEP OPEN

EXPLANATION:

Explanation: Box 1: k-fold Box 2: 3 K-Folds 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] Reference: https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.KFold.html

You are with a time series dataset in Azure Machine Learning Studio. You need to split your dataset into training and testing subsets by using the Split Data module. Which splitting mode should you use?

☐ A

Relative Expression Split

☐ B

Split Rows with the Randomized split parameter set to true

☐ C

Recommender Split

☐ D

Regular Expression Split

CORRECT ANSWER: B
KEEP OPEN

EXPLANATION:

Explanation: Split Rows: Use this option if you just want to divide the data into two parts. You can specify the percentage of data to put in each split, but by default, the data is divided 50-50. Incorrect Answers: B: Regular Expression Split: Choose this option when you want to divide your dataset by testing a single column for a value. C: Relative Expression Split: Use this option whenever you want to apply a condition to a number column. Reference: https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/split-data

You use Azure Machine Learning to train a model based on a dataset named dataset1. You define a dataset monitor and create a dataset named dataset2 that contains new data. You need to compare dataset1 and dataset2 by using the Azure Machine Learning SDK for Python. Which method of the DataDriftDetector class should you use?

☐   A

update

☐   B

run

☐   C

backfill

☐   D

get

CORRECT ANSWER: C

KEEP OPEN

EXPLANATION:

Explanation: A backfill run is used to see how data changes over time. Reference: https://docs.microsoft.com/en-us/python/api/azureml-datadrift/azureml.datadrift.datadriftdetector.datadriftdetector

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

☐ A
Computer Linear Correlation

☐ B
Execute Python Script

☐ C
Summarize Data

☐ D
Export Count Table

☐ E
Convert to Indicator Values

CORRECT ANSWERS: C,D
KEEP OPEN

EXPLANATION:

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

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_details_with_logs()

☐ C

files= run.get_file_names()

☐ D

files = run.get_details()

☐ E

files = run.get_metrics()

CORRECT ANSWER: C
KEEP OPEN

EXPLANATION:

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

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

CORRECT ANSWER: A
KEEP OPEN

EXPLANATION:

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

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. You need to select a data cleaning method. Which method should you use?

☐ A

Replace using Probabilistic PCA

☐ B

Synthetic Minority Oversampling Technique (SMOTE)

☐ C

Normalization

☐ D

Replace using MICE

CORRECT ANSWER: A
KEEP OPEN

EXPLANATION:

Explanation: Replace using Probabilistic PCA: Compared to other options, such as Multiple Imputation using Chained Equations (MICE), this option has the advantage of not requiring the application of predictors for each column. Instead, it approximates the covariance for the full dataset. Therefore, it might offer better performance for datasets that have missing values in many columns. Reference: https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/clean-missing-data

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

☐ CHECK BELOW THE RIGHT ANSWER

Answer Area

datastore = Datastore.[register_azure_blob_container / **register_azure_file_share** / register_azure_data_lake / register_azure_sql_database] (workspace = ws,

```
datastore_name = 'finance_datastore',
container_name = 'finance-data',
account_name = 'fintrainingdatastorage',
account_key = 'FWUYORRv3XoyNe...',
```
[create_if_not_exists = True / create_if_not_exists = False / **overwrite = True** / overwrite = False]

CORRECT ANSWER:
KEEP OPEN

EXPLANATION:

Explanation: Box 1: register_azure_blob_container Register an Azure Blob Container to the datastore. Box 2: create_if_not_exists = False Create the file share if it does not exist, defaults to False. Reference: https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.core.datastore.datastore

HOTSPOT You create a new Azure Databricks workspace. You configure a new cluster for long-running tasks with mixed loads on the compute cluster as shown in the image below.

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

□ CHECK BELOW THE RIGHT ANSWER

Answer Area

Code for each user runs as a separate process   ▼
Yes
No

The number of workers is fixed for the entire duration of the job   ▼
Yes
No

CORRECT ANSWER:
KEEP OPEN

EXPLANATION:

Explanation: Box 1: No Running user code in separate processes is not possible in Scala. Box 2: No Autoscaling is enabled. Minimum 2 workers, Maximum 8 workers. Reference: https://docs.databricks.com/clusters/configure.html

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

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

- [ ] A
resource_group
- [ ] B
login
- [ ] C
key
- [ ] D
region
- [ ] E
subscription_id

CORRECT ANSWERS: A,E
KEEP OPEN

EXPLANATION:

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

HOTSPOT 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.

☐ CHECK BELOW THE RIGHT ANSWER

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>"
 ]
}
```

CORRECT ANSWER:
KEEP OPEN

EXPLANATION:

Explanation: 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

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)
```

You plan to use this configuration to run a script that trains a random forest model and then tests it with validation data. The label values for the validation data are stored in a 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:

```
import numpy as np
from sklearn.metrics import roc_auc_score
# code to train model omitted
auc = roc_auc_score(y_test, y_predicted)
print(np.float(auc))
```

Does the solution meet the goal?

☐ A
Yes

☐ B
No

CORRECT ANSWER: B
KEEP OPEN

EXPLANATION:

Explanation Use a solution with logging.info(message) instead. Note: Python printing/logging example: logging.info(message) Destination: Driver logs, Azure

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

```
/sales
  /01-2019
    /sales.csv
  /02-2019
    /sales.csv
  /03-2019
    /sales.csv
...aw343167
```

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

☐ A

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

☐ B

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

☐ C

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

☐ D

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

EXPLANATION:

Explanation: Specify the path. Example: The following code gets the workspace existing workspace and the desired datastore by name. And then passes the datastore and file locations to the path parameter to create a new TabularDataset, weather_ds. from azureml.core import Workspace, Datastore, Dataset datastore_name = 'your datastore name' # get existing workspace workspace = Workspace.from_config() # retrieve an existing datastore in the workspace by name datastore = Datastore.get(workspace, datastore_name) # create a TabularDataset from 3 file paths in datastore datastore_paths = [(datastore, 'weather/2018/11.csv'), (datastore, 'weather/2018/12.csv'), (datastore, 'weather/2019/*.csv')] weather_ds = Dataset.Tabular.from_delimited_files(path=datastore_paths)

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

☐ CHECK BELOW THE RIGHT ANSWER

Answer Area

Virtual machine size

🔍 Search by name...

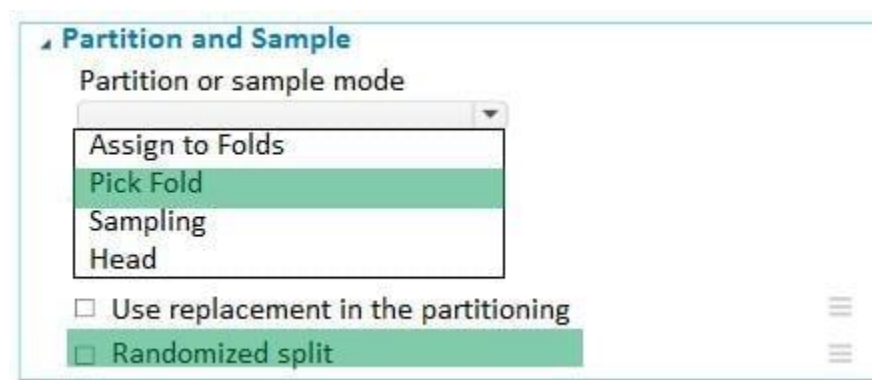| Name ↑	vCPUs	GPUs	RAM	Resource disk
BASIC_A0	1		0.75 GB	20 GB
STANDARD_D3_V2	4		14 GB	200 GB
STANDARD_E64_V3	64		432 GB	1,600 GB
STANDARD_M64LS	64		512 GB	2,000 GB
STANDARD_NC12	12	2	112 GB	680 GB
STANDARD_NC24	24	4	224 GB	1,440 GB

CORRECT ANSWER:
KEEP OPEN

EXPLANATION:

Explanation: CUDA is a parallel computing platform and programming model developed by Nvidia for general computing on its own GPUs (graphics processing units). CUDA enables developers to speed up compute-intensive applications by harnessing the power of GPUs for the parallelizable part of the computation. Reference: https://www.infoworld.com/article/3299703/what-is-cuda-parallel-programming-for-gpus.html

HOTSPOT You have a dataset that contains 2,000 rows. You are building a machine learning classification model by using Azure Learning Studio. You add a Partition and Sample module to the experiment. You need to configure the module. You must meet the following requirements: Divide the data into subsets Assign the rows into folds using a round-robin method Allow rows in the dataset to be reused How should you configure the module? To answer, select the appropriate options in the dialog box in the answer area. NOTE: Each correct selection is worth one point.

☐ CHECK BELOW THE RIGHT ANSWER
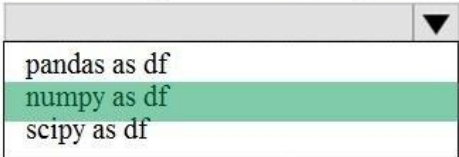

CORRECT ANSWER:
KEEP OPEN

EXPLANATION:

Explanation: Use the Split data into partitions option when you want to divide the dataset into subsets of the data. This option is also useful when you want to create a custom number of folds for cross-validation, or to split rows into several groups. Add the Partition and Sample module to your experiment in Studio (classic), and connect the dataset. For Partition or sample mode, select Assign to Folds. Use replacement in the partitioning: Select this option if you want the sampled row to be put back into the pool of rows for potential reuse. As a result, the same row might be assigned to several folds. If you do not use replacement (the default option), the sampled row is not put back into the pool of rows for potential reuse. As a result, each row can be assigned to only one fold. Randomized split: Select this option if you want rows to be randomly assigned to folds. If you do not select this option, rows are assigned to folds using the round-robin method. Reference: https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/partition-and-sample
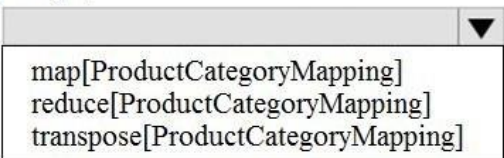
HOTSPOT You are creating a machine learning model in Python. The provided dataset contains several numerical columns and one text column. The text column represents a product's category. The product category will always be one of the following: Bikes Cars Vans Boats You are building a regression model using the scikit-learn Python package. You need to transform the text data to be compatible with the scikit-learn Python package. How should you complete the code segment? To answer, select the appropriate options in the answer area. NOTE: Each correct selection is worth one point.

☐ CHECK BELOW THE RIGHT ANSWER

Answer Area

```
from sklearn import linear_model
import ▼
    pandas as df
    numpy as df
    scipy as df

dataset = df.read_csv("data\\ProductSales.csv")
ProductCategoryMapping = {"Bikes":1, "Cars":2, "Boats": 3,
"Vans": 4}
dataset['ProductCategoryMapping'] =
dataset['ProductCategory']. ▼
    map[ProductCategoryMapping]
    reduce[ProductCategoryMapping]
    transpose[ProductCategoryMapping]

regr = linear_model.LinearRegression()
X_train = dataset[['ProductCategoryMapping', 'ProductSize',
'ProductCost']]
y_train = dataset[['Sales']]
regr.fit(X_train, y_train)
```

CORRECT ANSWER:
KEEP OPEN

EXPLANATION:

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/

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)
```

You plan to use this configuration to run a script that trains a random forest model and then tests it with validation data. The label values for the validation data are stored in a 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:

```
import json, os
from sklearn.metrics import roc_auc_score
# code to train model omitted
auc = roc_auc_score(y_test, y_predicted)
os.makedirs("outputs", exist_ok = True)
with open("outputs/AUC.txt", "w") as file_cur:
    file_cur.write(auc)
```

Does the solution meet the goal?

☐ A
No

☐ B
Yes

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 and register a dataset by using TabularDataset class that references the Azure blob storage containing bird photographs.

☐ B

Register the Azure blob storage containing the bird photographs as a datastore in Azure Machine Learning service.

☐ C

Create an Azure Cosmos DB database and attach the Azure Blob containing bird photographs storage to the database.

☐ D

Copy the bird photographs to the blob datastore that was created with your Azure Machine Learning service workspace.

☐ E

Create an Azure Data Lake store and move the bird photographs to the store.

CORRECT ANSWER: B
KEEP OPEN

EXPLANATION:

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

HOTSPOT You are preparing to build a deep learning convolutional neural network model for image classification. You create a script to train the model using CUDA devices. You must submit an experiment that runs this script in the Azure Machine Learning workspace. The following compute resources are available: a Microsoft Surface device on which Microsoft Office has been installed. Corporate IT policies prevent the installation of additional software a Compute Instance named ds-workstation in the workspace with 2 CPUs and 8 GB of memory an Azure Machine Learning compute target named cpu-cluster with eight CPU-based nodes an Azure Machine Learning compute target named gpu-cluster with four CPU and GPU-based nodes You need to specify the compute resources to be used for running the code to submit the experiment, and for running the script in order to minimize model training time. Which resources should the data scientist use? To answer, select the appropriate options in the answer area. NOTE: Each correct selection is worth one point.

☐ CHECK BELOW THE RIGHT ANSWER

Answer Area

Resource type	Option
Run code to submit the experiment	the Microsoft Surface device the ds-workstation compute instance **the cpu-cluster compute target** the gpu-cluster compute target
Run the training script	the ds-workstation compute instance the cpu-cluster compute target the gpu-cluster compute target **the Microsoft Surface device**

CORRECT ANSWER:
KEEP OPEN

EXPLANATION:

Explanation: Box 1: the ds-workstation compute instance A workstation notebook instance is good enough to run experiments. Box 2: the gpu-cluster compute target Just as GPUs revolutionized deep learning through unprecedented training and inferencing performance, RAPIDS enables traditional machine learning practitioners to unlock game-changing performance with GPUs. With RAPIDS on Azure Machine Learning service, users can accelerate the entire machine learning pipeline, including data processing, training and inferencing, with GPUs from the NC_v3, NC_v2, ND or ND_v2 families. Users can unlock performance gains of more than 20X (with 4 GPUs), slashing training times from hours to minutes and dramatically reducing time-to-insight. Reference: https://azure.microsoft.com/sv-se/blog/azure-machine-learning-service-now-supports-nvidia-s-rapids/

You create an Azure Machine Learning compute resource to train models. The compute resource is configured as follows: Minimum nodes: 2 Maximum nodes: 4 You must decrease the minimum number of nodes and increase the maximum number of nodes to the following values: Minimum nodes: 0 Maximum nodes: 8 You need to reconfigure the compute resource. What are three possible ways to achieve this goal? Each correct answer presents a complete solution. NOTE: Each correct selection is worth one point.

☐ A

Use the Azure Machine Learning studio.

☐ B

Run the update method of the AmlCompute class in the Python SDK.

☐ C

Use the Azure Machine Learning designer.

☐ D

Run the refresh_state() method of the BatchCompute class in the Python SDK.

☐ E

Use the Azure portal.

CORRECT ANSWERS: A,B,E
KEEP OPEN

EXPLANATION:

Explanation: A: You can manage assets and resources in the Azure Machine Learning studio. B: The update(min_nodes=None, max_nodes=None, idle_seconds_before_scaledown=None) of the AmlCompute class updates the ScaleSettings for this AmlCompute target. C: To change the nodes in the cluster, use the UI for your cluster in the Azure portal. Reference: https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.core.compute.amlcompute(class)

Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution. After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen. You are analyzing a numerical dataset which contains missing values in several columns. You must clean the missing values using an appropriate operation without affecting the dimensionality of the feature set. You need to analyze a full dataset to include all values. Solution: Replace each missing value using the Multiple Imputation by Chained Equations (MICE) method. Does the solution meet the goal?

☐ A
Yes

☐ B
No

CORRECT ANSWER: A
KEEP OPEN

EXPLANATION:

Explanation: Replace using MICE: For each missing value, this option assigns a new value, which is calculated by using a method described in the statistical literature as "Multivariate Imputation using Chained Equations" or "Multiple Imputation by Chained Equations". With a multiple imputation method, each variable with missing data is modeled conditionally using the other variables in the data before filling in the missing values. Note: Multivariate imputation by chained equations (MICE), sometimes called "fully conditional specification" or "sequential regression multiple imputation" has emerged in the statistical literature as one principled method of addressing missing data. Creating multiple imputations, as opposed to single imputations, accounts for the statistical uncertainty in the imputations. In addition, the chained equations approach is very flexible and can handle variables of varying types (e.g., continuous or binary) as well as complexities such as bounds or survey skip patterns.   Reference: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3074241/ https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/clean-missing-data

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

☐ A

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

☐ B

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

☐ C

Create an Azure Kubernetes Service (AKS) inference cluster.

☐ D

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

☐ E

Create a tabular dataset that supports versioning.

CORRECT ANSWERS: A,C,E
KEEP OPEN

EXPLANATION:

Incorrect Answers: C, E: The UI is included the Enterprise edition only. Reference: https://azure.microsoft.com/en-us/pricing/details/machine-learning/

You use an Azure Machine Learning workspace. You have a trained model that must be deployed as a web service. Users must authenticate by using Azure Active Directory. What should you do?

☐ A

Deploy the model to Azure Container Instances. During deployment, set the token_auth_enabled parameter of the target configuration object to true

☐ B

Deploy the model to Azure Kubernetes Service (AKS). During deployment, set the auth.enabled parameter of the target configuration object to true

☐ C

Deploy the model to Azure Kubernetes Service (AKS). During deployment, set the token_auth_enabled parameter of the target configuration object to true

☐ D

Deploy the model to Azure Container Instances. During deployment, set the auth_enabled parameter of the target configuration object to true

CORRECT ANSWER: C
KEEP OPEN

EXPLANATION:

Explanation: To control token authentication, use the token_auth_enabled parameter when you create or update a deployment Token authentication is disabled by default when you deploy to Azure Kubernetes Service. Note: The model deployments created by Azure Machine Learning can be configured to use one of two authentication methods: key-based: A static key is used to authenticate to the web service. token-based: A temporary token must be obtained from the Azure Machine Learning workspace (using Azure Active Directory) and used to authenticate to the web service. Incorrect Answers: C: Token authentication isn't supported when you deploy to Azure Container Instances. Reference: https://docs.microsoft.com/en-us/azure/machine-learning/how-to-authenticate-web-service

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

☐ A

Assign Data to Clusters

☐ B

Split Data

☐ C

Group Data into Bins

☐ D

Load Trained Model

CORRECT ANSWER: C
KEEP OPEN

EXPLANATION:

Explanation: The Group Data into Bins module supports multiple options for binning data. You can customize how the bin edges are set and how values are apportioned into the bins. Reference: https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/group-data-into-bins

Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution. After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen. You are creating a new experiment in Azure Machine Learning Studio. One class has a much smaller number of observations than the other classes in the training set. You need to select an appropriate data sampling strategy to compensate for the class imbalance. Solution: You use the Scale and Reduce sampling mode. Does the solution meet the goal?

☐ A
No

☐ B
Yes

CORRECT ANSWER: A
KEEP OPEN

EXPLANATION:

Explanation: Instead use the Synthetic Minority Oversampling Technique (SMOTE) sampling mode. Note: SMOTE is used to increase the number of underepresented cases in a dataset used for machine learning. SMOTE is a better way of increasing the number of rare cases than simply duplicating existing cases. Incorrect Answers: Common data tasks for the Scale and Reduce sampling mode include clipping, binning, and normalizing numerical values. Reference: https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/smote https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/data-transformation-scale-and-reduce

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

☐ A

Remove entire row

☐ B

Hot Deck

☐ C

Replace with mode

☐ D

Custom substitution value

☐ E

Replace with mean

☐ F

Remove entire column

CORRECT ANSWER: A
KEEP OPEN

EXPLANATION:

Explanation: Remove entire row: Completely removes any row in the dataset that has one or more missing values. This is useful if the missing value can be considered randomly missing. Reference: https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/clean-missing-data

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

CORRECT ANSWER: B
KEEP OPEN

EXPLANATION:

Explanation: Use the Entropy MDL binning mode which has a target column. Reference: https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/group-data-into-bins

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

Azure BatchAI

☐ B

Azure Notebooks

☐ C

Azure Machine Learning Service

☐ D

Anaconda Data Science Platform

CORRECT ANSWER: B
KEEP OPEN

EXPLANATION:

Reference: https://notebooks.azure.com/

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

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

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

☐ CHECK BELOW THE RIGHT ANSWER

Answer Area

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

CORRECT ANSWER:
KEEP OPEN

EXPLANATION:

Explanation: Box 1: No If a compute cluster already exists it will be used. Box 2: Yes The wait_for_completion method waits for the current provisioning operation to finish on the cluster. Box 3: Yes Low Priority VMs use Azure's excess capacity and are thus cheaper but risk your run being pre-empted. Box 4: No Need to use training_compute.delete() to deprovision and delete the AmlCompute target. Reference: https://notebooks.azure.com/azureml/projects/azureml-getting-started/html/how-to-use-azureml/training/train-on-amlcompute/train-on-amlcompute.ipynb https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.core.compute.computetarget

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)
```

You plan to use this configuration to run a script that trains a random forest model and then tests it with validation data. The label values for the validation data are stored in a 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

CORRECT ANSWER: A
KEEP OPEN

Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution. After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen. You are analyzing a numerical dataset which contains missing values in several columns. You must clean the missing values using an appropriate operation without affecting the dimensionality of the feature set. You need to analyze a full dataset to include all values. Solution: Use the Last Observation Carried Forward (LOCF) method to impute the missing data points. Does the solution meet the goal?

☐ A
No
☐ B
Yes

CORRECT ANSWER: A
KEEP OPEN

EXPLANATION:

Explanation: Instead use the Multiple Imputation by Chained Equations (MICE) method. Replace using MICE: For each missing value, this option assigns a new value, which is calculated by using a method described in the statistical literature as "Multivariate Imputation using Chained Equations" or "Multiple Imputation by Chained Equations". With a multiple imputation method, each variable with missing data is modeled conditionally using the other variables in the data before filling in the missing values. Note: Last observation carried forward (LOCF) is a method of imputing missing data in longitudinal studies. If a person drops out of a study before it ends, then his or her last observed score on the dependent variable is used for all subsequent (i.e., missing) observation points. LOCF is used to maintain the sample size and to reduce the bias caused by the attrition of participants in a study. Reference: https://methods.sagepub.com/reference/encyc-of-research-design/n211.xml https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3074241/

HOTSPOT You have an Azure Machine Learning workspace named workspace1 that is accessible from a public endpoint. The workspace contains an Azure Blob storage datastore named store1 that represents a blob container in an Azure storage account named account1. You configure workspace1 and account1 to be accessible by using private endpoints in the same virtual network. You must be able to access the contents of store1 by using the Azure Machine Learning SDK for Python. You must be able to preview the contents of store1 by using Azure Machine Learning studio. You need to configure store1. What should you do? To answer, select the appropriate options in the answer area. NOTE: Each correct selection is worth one point.

☐ CHECK BELOW THE RIGHT ANSWER

Answer Area

Requirement	Action
Access the contents of store1 by using the Azure Machine Learning SDK for Python.	⌄
	Set store1 as the default datastore.
	Disable data validation for store1.
	Update authentication for store1.
	Regenerate the keys of account1.
Preview the contents of store1 by using Azure Machine Learning studio.	⌄
	Set store1 as the default datastore.
	Disable data validation for store1.
	Update authentication for store1.
	Regenerate the keys of account1.

CORRECT ANSWER:
KEEP OPEN

EXPLANATION:

Explanation: Box 1: Regenerate the keys of account1. Azure Blob Storage support authentication through Account key or SAS token. To authenticate your access to the underlying storage service, you can provide either your account key, shared access signatures (SAS) tokens, or service principal Box 2: Update the authentication for store1. For Azure Machine Learning studio users, several features rely on the ability to read data from a dataset; such as dataset previews, profiles and automated machine learning. For these features to work with storage behind virtual networks, use a workspace managed identity in the studio to allow Azure Machine Learning to access the storage account from outside the virtual network. Note: Some of the studio's features are disabled by default in a virtual network.

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	• an Azure Machine Learning workspace named amlworkspace • an Azure Storage account named amlworkspace12345 • an Application Insights instance named amlworkspace54321 • an Azure Key Vault named amlworkspace67890 • an Azure Container Registry named amlworkspace09876
general_compute	A virtual machine named mlvm with the following configuration: • Operating system: Ubuntu Linux • Software installed: Python 3.6 and Jupyter Notebooks • Size: NC6 (6 vCPUs, 1 vGPU, 56 Gb RAM)

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. Does the solution meet the goal?

☐ A
No
☐ B
Yes

CORRECT ANSWER: B

KEEP OPEN

EXPLANATION:

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

You create a new Azure subscription. No resources are provisioned in the subscription. You need to create an Azure Machine Learning workspace. What are three possible ways to achieve this goal? Each correct answer presents a complete solution. NOTE: Each correct selection is worth one point.

☐ A

Run Python code that uses the Azure ML SDK library and calls the Workspace.create method with name, subscription_id, and resource_group parameters.

☐ B

Use the Azure Command Line Interface (CLI) with the Azure Machine Learning extension to call the az group create function with --name and --location parameters, and then the az ml workspace create function, specifying –w and –g parameters for the workspace name and resource group.

☐ C

Navigate to Azure Machine Learning studio and create a workspace.

☐ D

Navigate to Azure Machine Learning studio and create a workspace.

☐ E

Run Python code that uses the Azure ML SDK library and calls the Workspace.get method with name, subscription_id, and resource_group parameters.

CORRECT ANSWERS: B,C,D
KEEP OPEN

EXPLANATION:

Explanation: B: You can create a workspace in the Azure Machine Learning studio C: You can create a workspace for Azure Machine Learning with Azure CLI Install the machine learning extension. Create a resource group: az group create --name <resource-group-name> --location <location> To create a new workspace where the services are automatically created, use the following command: az ml workspace create -w <workspace-name> -g <resource-group-name> D: You can create and manage Azure Machine Learning workspaces in the Azure portal. Sign in to the Azure portal by using the credentials for your Azure subscription. In the upper-left corner of Azure portal, select + Create a resource. Use the search bar to find Machine Learning. Select Machine Learning. In the Machine Learning pane, select Create to begin.

Machine Learning 🖶

Create a machine learning workspace

Basics   Networking   Advanced   Tags   Review + create

Project details

Select the subscription to manage deployed resources and costs. Use resource groups like folders to organize and manage all your resources.

Subscription * ⓘ	Documentation-team ⌄
└─ Resource group * ⓘ	docs-ws ⌄
	Create new

Workspace details

Specify the name, region, and edition for the workspace.

Workspace name * ⓘ	docs-mlw ✓
Region * ⓘ	West Central US ⌄
Workspace edition * ⓘ	Basic ⌃
	Basic
	Enterprise

ⓘ For your convenience, these resources a
Application Insights, Azure Key Vault

Reference: https://docs.microsoft.com/en-us/azure/machine-learning/how-to-create-workspace-template https://docs.microsoft.com/en-us/azure/machine-learning/how-to-manage-workspace-cli https://docs.microsoft.com/en-us/azure/machine-learning/how-to-manage-workspace

HOTSPOT 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.

☐ CHECK BELOW THE RIGHT ANSWER

Answer Area

Statements	Yes	No
The user can perform all actions in the workspace	○	●
The user can delete a compute resource in the workspace	○	●
The user can write metrics to the workspace	●	○

CORRECT ANSWER:
KEEP OPEN

EXPLANATION:

Explanation: 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-user-has-access-to-a-resource

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

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

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

☐ CHECK BELOW THE RIGHT ANSWER

Answer Area

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

CORRECT ANSWER:
KEEP OPEN

EXPLANATION:

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

Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution. After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen. You are analyzing a numerical dataset which contains missing values in several columns. You must clean the missing values using an appropriate operation without affecting the dimensionality of the feature set. You need to analyze a full dataset to include all values. Solution: Calculate the column median value and use the median value as the replacement for any missing value in the column. Does the solution meet the goal?

☐ A

No

☐ B

Yes

CORRECT ANSWER: A
KEEP OPEN

EXPLANATION:

Explanation: Use the Multiple Imputation by Chained Equations (MICE) method. Reference: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3074241/ https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/clean-missing-data

HOTSPOT 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.

☐ CHECK BELOW THE RIGHT ANSWER

Answer Area

Parameter	Value
Estimator	▼

Estimator
SKLearn
PyTorch
Tensorflow
Chainer

| Training compute | ▼ |

12 vCPU, 48 GB memory, 96 GB SSD
12 vCPU, 112 GB memory, 680 GB SSD, 2 GPU, 24 GB GPU memory
16 vCPU, 128 GB memory, 160 GB HDD, 80 GB NVME disk (4000 MBps)
44 vCPU, 352 GB memory, 3.4 GHz CPU frequency all cores

CORRECT ANSWER:
KEEP OPEN

EXPLANATION:

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

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

☐ CHECK BELOW THE RIGHT ANSWER

Answer Area

▲ 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

CORRECT ANSWER:
KEEP OPEN

EXPLANATION:

Explanation: Box 1: Sampling Create a sample of data This option supports simple random sampling or stratified random sampling. This is useful if you want to create a smaller representative sample dataset for testing. 1. Add the

Partition and Sample module to your experiment in Studio, and connect the dataset. 2. Partition or sample mode: Set this to Sampling. 3. Rate of sampling. See box 2 below. Box 2: 0 3. Rate of sampling. Random seed for sampling: Optionally, type an integer to use as a seed value. This option is important if you want the rows to be divided the same way every time. The default value is 0, meaning that a starting seed is generated based on the system clock. This can lead to slightly different results each time you run the experiment. Reference: https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/partition-and-sample

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 resources 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":["*"],
"AssignableScopes": [
    "/subscriptions/<id>/resourceGroups/ml-rg/providers/Microsoft.MachineLearningServices/workspaces/ml-ws"
]
}
```

☐ B

```
{
 "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"
  ]
}
```

☐ 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"
]
}
```

praw343167

CORRECT ANSWER: B
KEEP OPEN

EXPLANATION:

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.MachineLearningServices/workspaces/<workspace_name>"] } Reference: https://docs.microsoft.com/en-us/azure/machine-learning/how-to-assign-roles

You are analyzing a dataset containing historical data from a local taxi company. You are developing a regression model. You must predict the fare of a taxi trip. You need to select performance metrics to correctly evaluate the regression model. Which two metrics can you use? Each correct answer presents a complete solution? NOTE: Each correct selection is worth one point.

☐ A

an R-Squared value close to 0

☐ B

an F1 score that is high

☐ C

an R-Squared value close to 1

☐ D

an F1 score that is low

☐ E

a Root Mean Square Error value that is low

☐ F

a Root Mean Square Error value that is high

CORRECT ANSWERS: C,E
KEEP OPEN

EXPLANATION:

Explanation: RMSE and R2 are both metrics for regression models. A: Root mean squared error (RMSE) creates a single value that summarizes the error in the model. By squaring the difference, the metric disregards the difference between over-prediction and under-prediction. D: Coefficient of determination, often referred to as R2, represents the predictive power of the model as a value between 0 and 1. Zero means the model is random (explains nothing); 1 means there is a perfect fit. However, caution should be used in interpreting R2 values, as low values can be entirely normal and high values can be suspect. Incorrect Answers: C, E: F-score is used for classification models, not for regression models. Reference: https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/evaluate-model

HOTSPOT You are performing a classification task in Azure Machine Learning Studio. You must prepare balanced testing and training samples based on a provided data set. You need to split the data with a 0.75:0.25 ratio. Which value should you use for each parameter? To answer, select the appropriate options in the answer area. NOTE: Each correct selection is worth one point.

□ CHECK BELOW THE RIGHT ANSWER

Answer Area

Parameter	Value
Splitting mode	▼
	Split rows
	Recommender Split
	Regular Expression Split
	Relative Expression Split
Fraction of rows in the first output dataset	▼
	0.75
	0.25
	0.5
	1
Randomized split	▼
	True
	False
Stratified split	▼
	True
	False

CORRECT ANSWER:
KEEP OPEN

EXPLANATION:

Explanation: Box 1: Split rows Use the Split Rows option if you just want to divide the data into two parts. You can specify the percentage of data to put in each split, but by default, the data is divided 50-50. You can also randomize the selection of rows in each group, and use stratified sampling. In stratified sampling, you must select a single column of data for which you want values to be apportioned equally among the two result datasets. Box 2: 0.75 If you specify

a number as a percentage, or if you use a string that contains the "%" character, the value is interpreted as a percentage. All percentage values must be within the range (0, 100), not including the values 0 and 100. Box 3: Yes To ensure splits are balanced. Box 4: No If you use the option for a stratified split, the output datasets can be further divided by subgroups, by selecting a strata column. Reference: https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/split-data