DP-100 Q&As

Designing and Implementing a Data Science Solution on Azure

Pass Microsoft DP-100 Exam with 100% Guarantee

Free Download Real Questions & Answers **PDF** and **VCE** file from:

https://www.pass4itsure.com/dp-100.html

100% Passing Guarantee
100% Money Back Assurance

Following Questions and Answers are all new published by Microsoft Official Exam Center

- **Instant Download** After Purchase
- **100% Money Back** Guarantee
- **365 Days** Free Update
- **800,000+** Satisfied Customers

QUESTION 1

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:

1.

Bikes

2.

Cars

3.

Vans

4.

Boats

You are building a regression model using the scikit-learn Python package.

You need to transform the text data to be compatible with the scikit-learn Python package.

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

NOTE: Each correct selection is worth one point.

Hot Area:

```
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:

```
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)
```

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/

QUESTION 2

HOTSPOT

You train a model by using Azure Machine Learning. You use Azure Blob Storage to store production data.

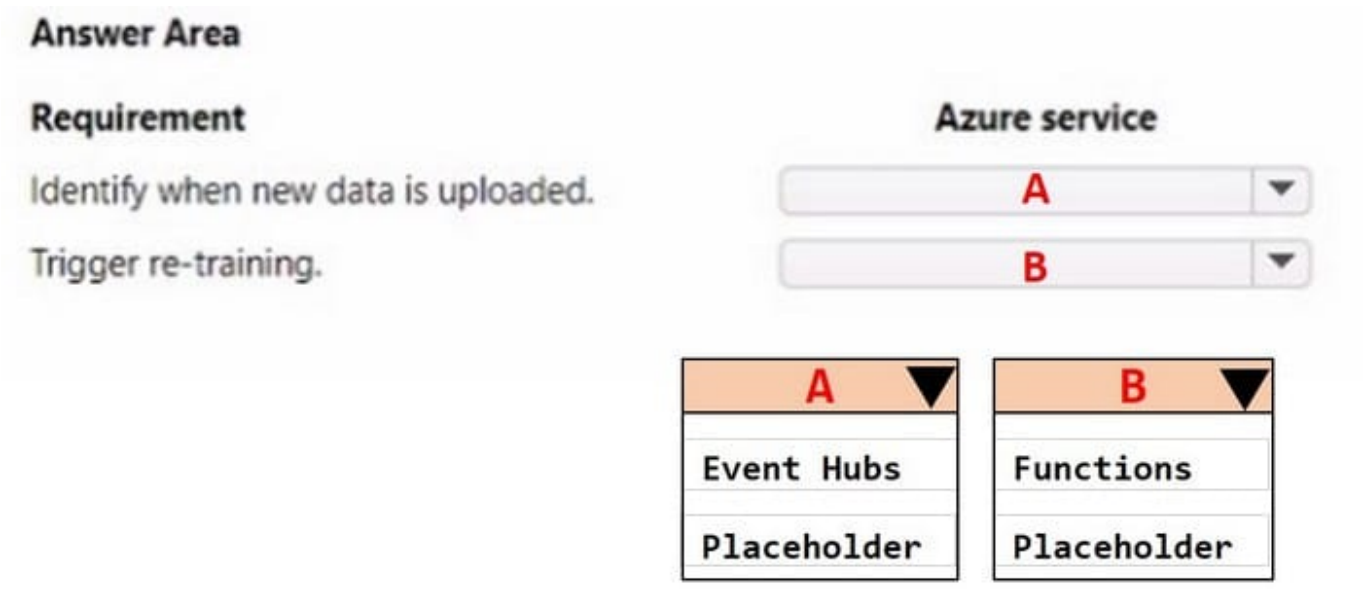
The model must be re-trained when new data is uploaded to Azure Blob Storage. You need to minimize development and coding.

You need to configure Azure services to develop a re-training solution.

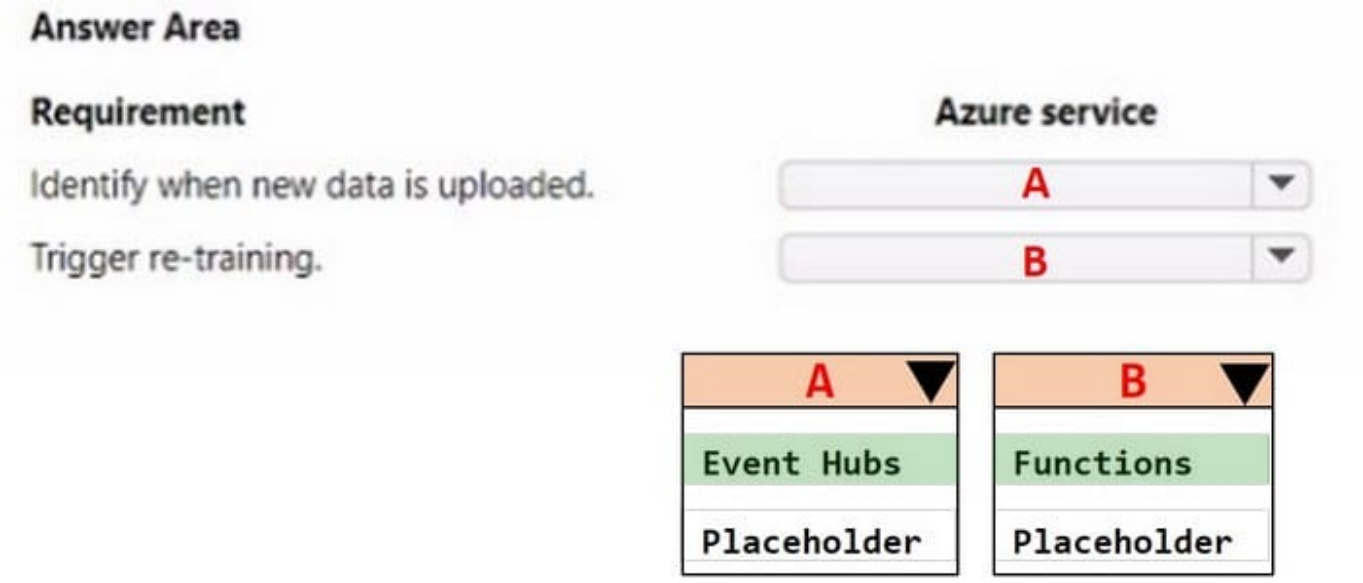
Which Azure services should you use? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Hot Area:



Correct Answer:



QUESTION 3

You create an Azure Machine Learning pipeline named pipeline 1 with two steps that contain Python scripts. Data processed by the first step is passed to the second step.

You must update the content of the downstream data source of pipeline 1 and run the pipeline again.

You need to ensure the new run of pipeline 1 fully processes the updated content.

Solution: Change the value of the compute.target parameter of the PythonScriptStep object in the two steps.

Does the solution meet the goal?

A. Yes

B. No

Correct Answer: B

QUESTION 4

DRAG DROP

You create an Azure Machine Learning workspace.

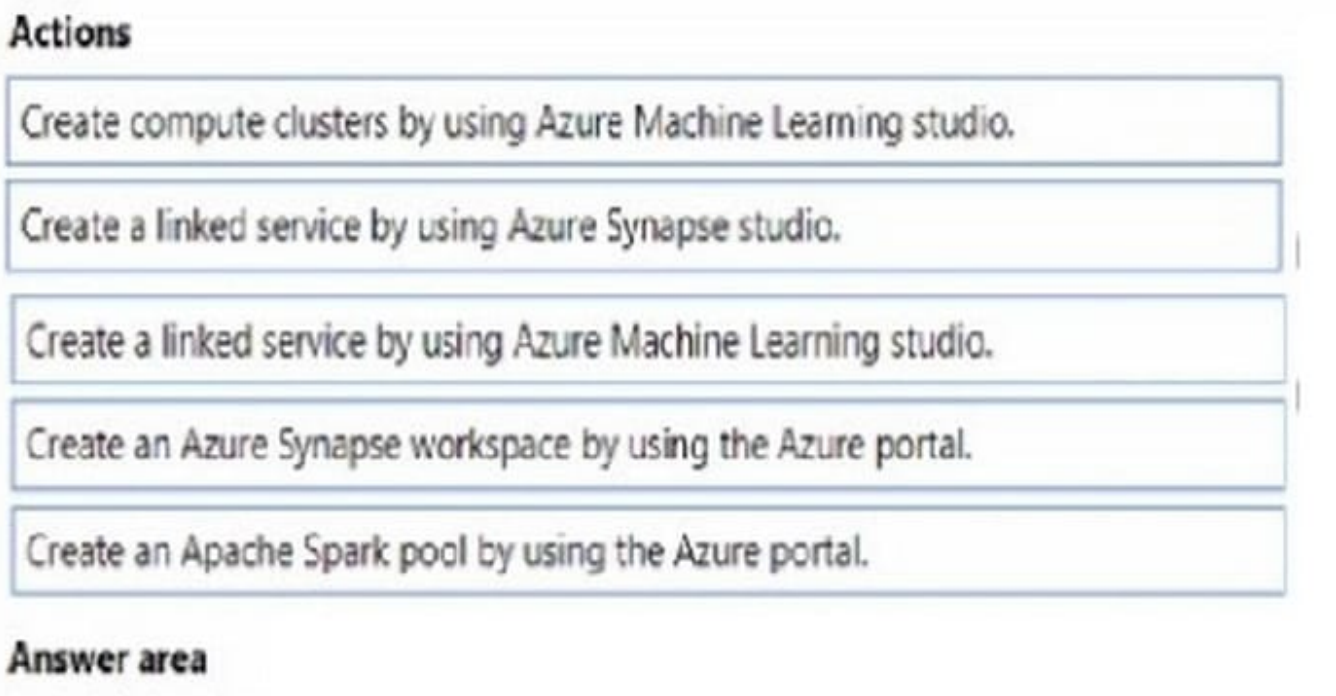
You must implement dedicated compute for model training in the workspace by using Azure Synapse compute resources. The solution must attach the dedicated compute and start an Azure Synapse session.

You need to implement the compute resources.

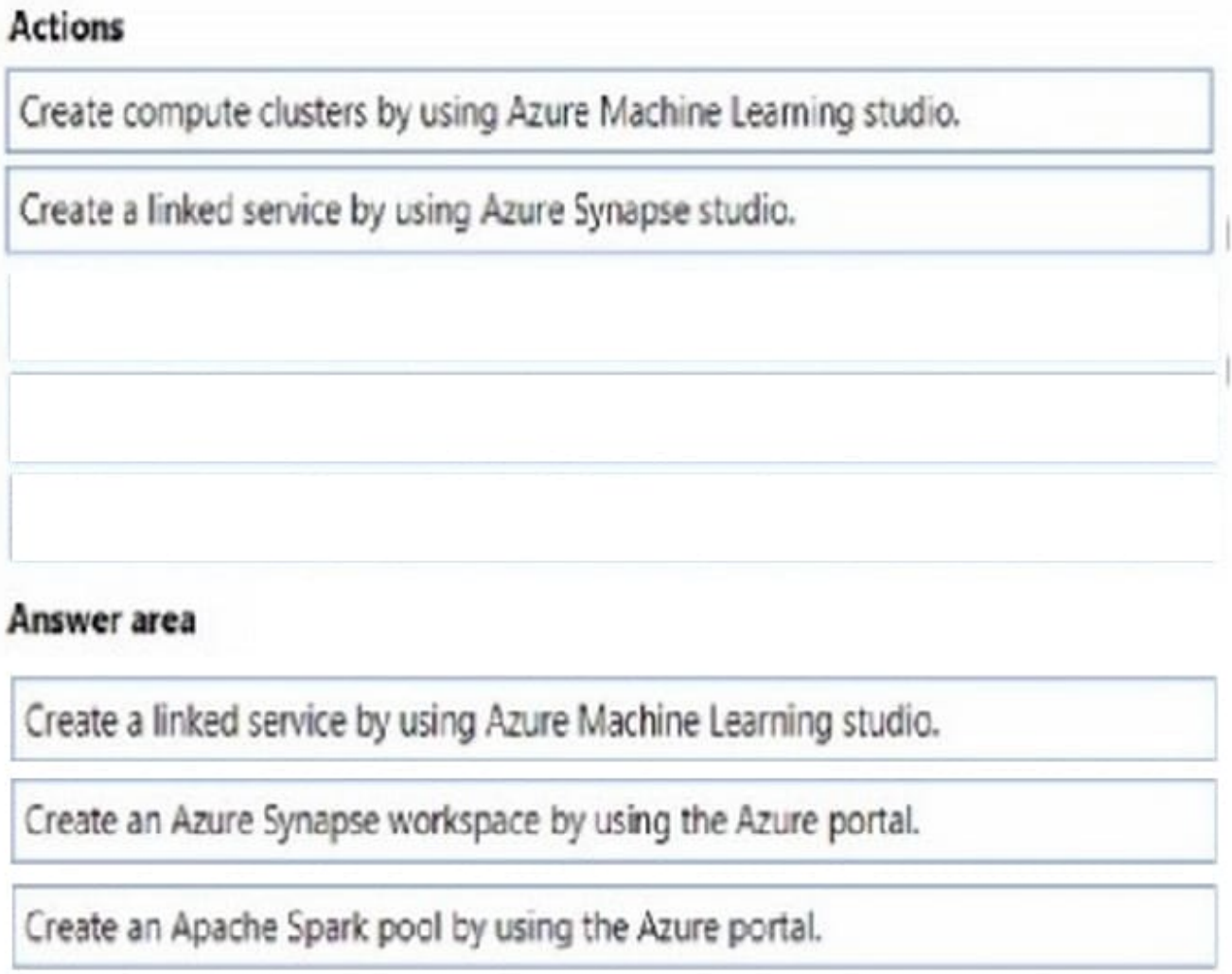
Which three actions should you perform in sequence? To answer, move the appropriate actions from the list of actions

to the answer area and arrange them in the correct order.

Select and Place:



Correct Answer:



QUESTION 5

You need to consider the underlined segment to establish whether it is accurate.

To transform a categorical feature into a binary indicator, you should make use of the Clean Missing Data module.

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

A. No adjustment required.

B. Convert to Indicator Values

C. Apply SQL Transformation

D. Group Categorical Values

Correct Answer: B

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

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

QUESTION 6

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

Correct Answer: B

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:

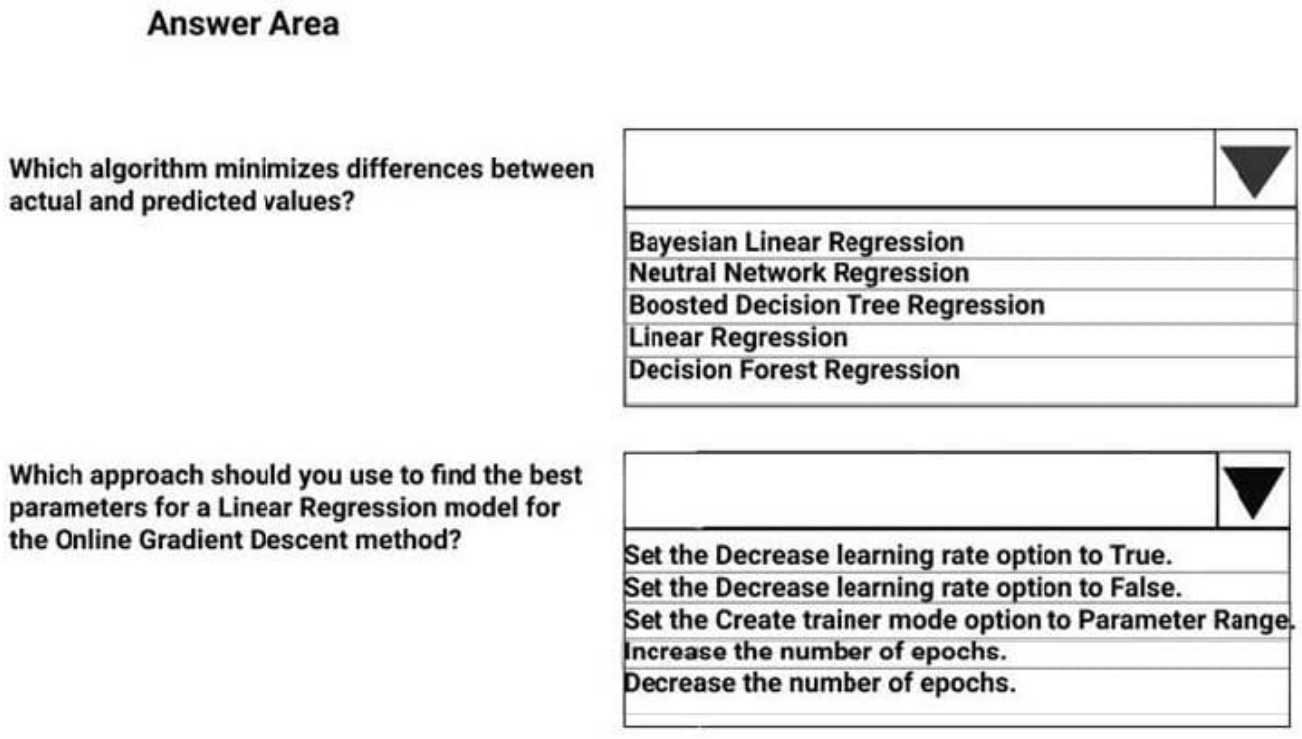
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-train-scikit-learn

QUESTION 7

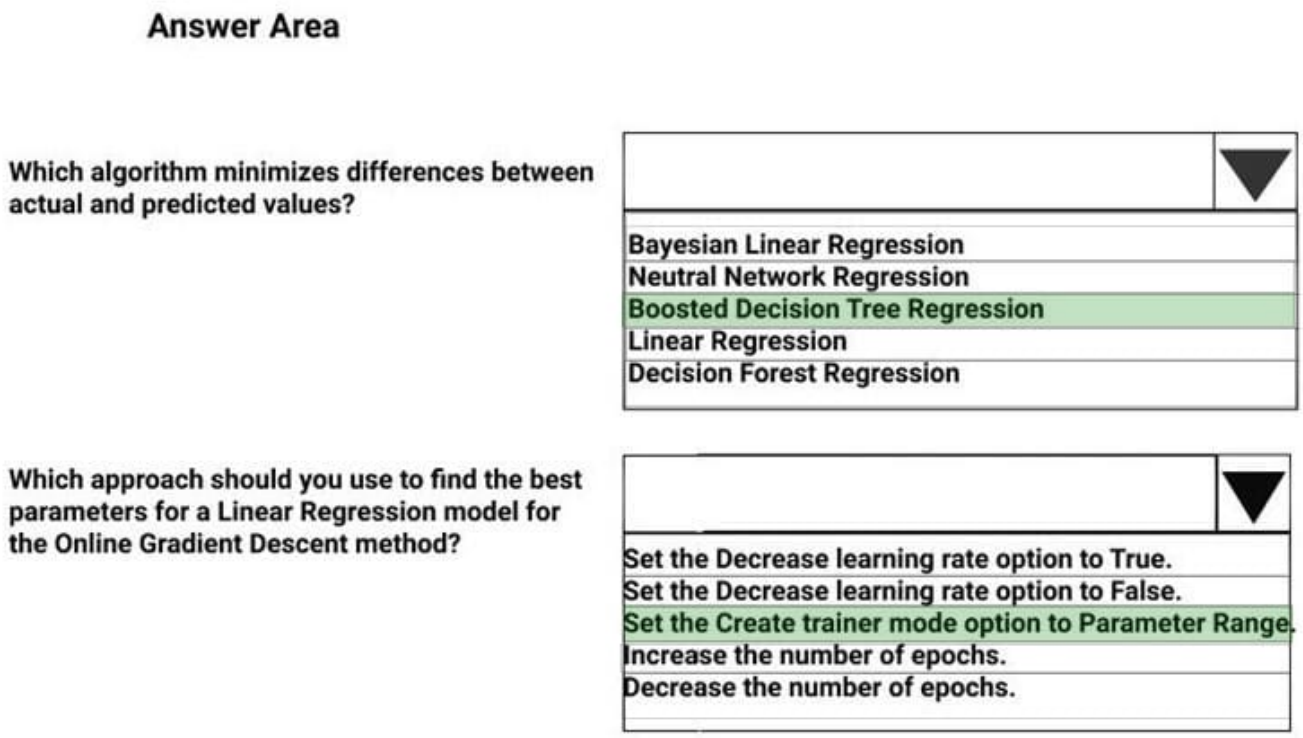
Use the drop-down menus to select the answer choice that answers each question based on the information presented in the image.

NOTE: Each correct selection is worth one point.

Hot Area:



Correct Answer:



Box 1: Boosted Decision Tree Regression

Mean absolute error (MAE) measures how close the predictions are to the actual outcomes; thus, a lower score is better.

Box 2:

Online Gradient Descent: If you want the algorithm to find the best parameters for you, set Create trainer mode option to Parameter Range. You can then specify multiple values for the algorithm to try.

References:

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

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

QUESTION 8

DRAG DROP

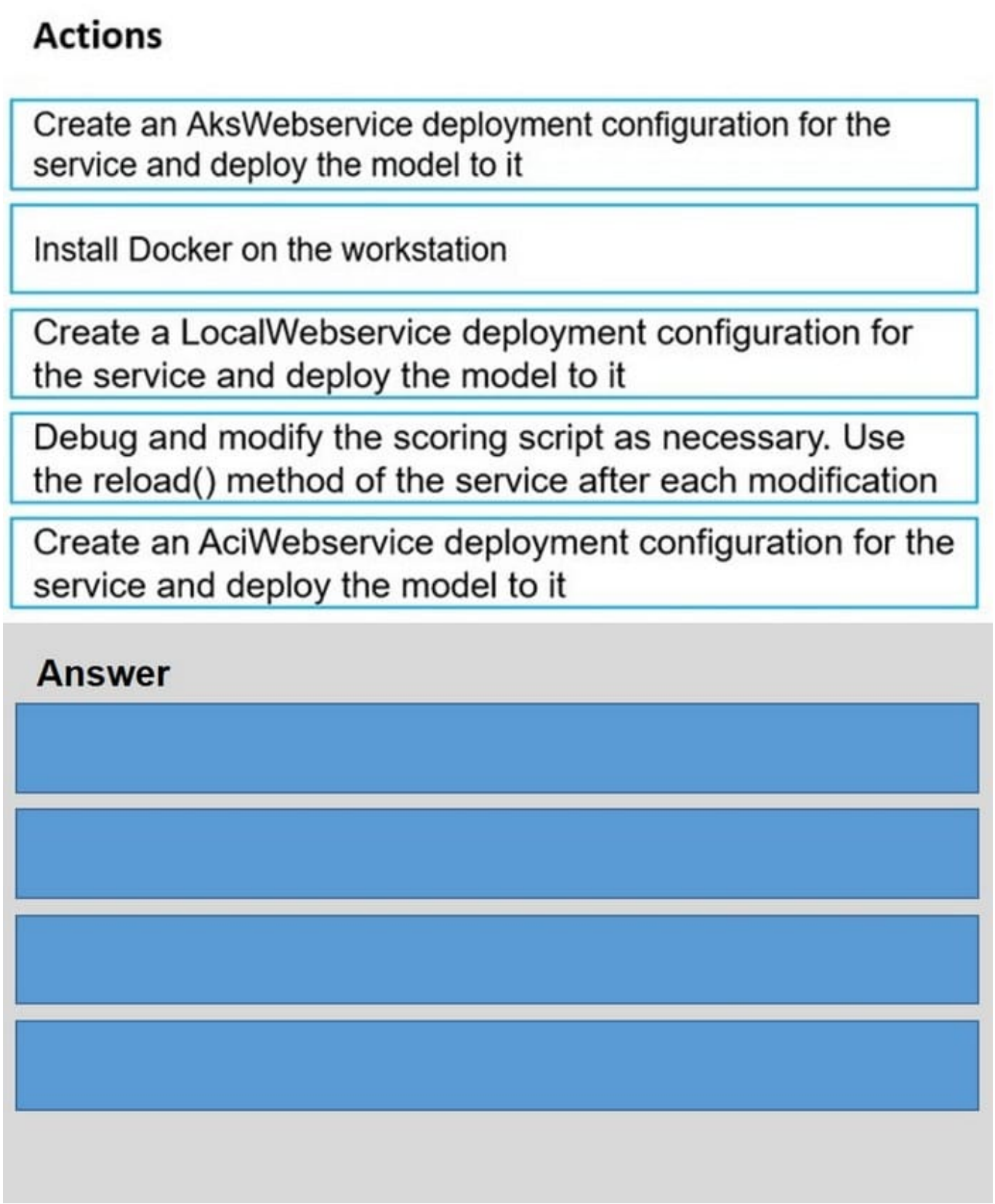
You train and register a model by using the Azure Machine Learning SDK on a local workstation. Python 3.6 and Visual Studio Code are installed on the workstation.

When you try to deploy the model into production as an Azure Kubernetes Service (AKS)-based web service, you experience an error in the scoring script that causes deployment to fail.

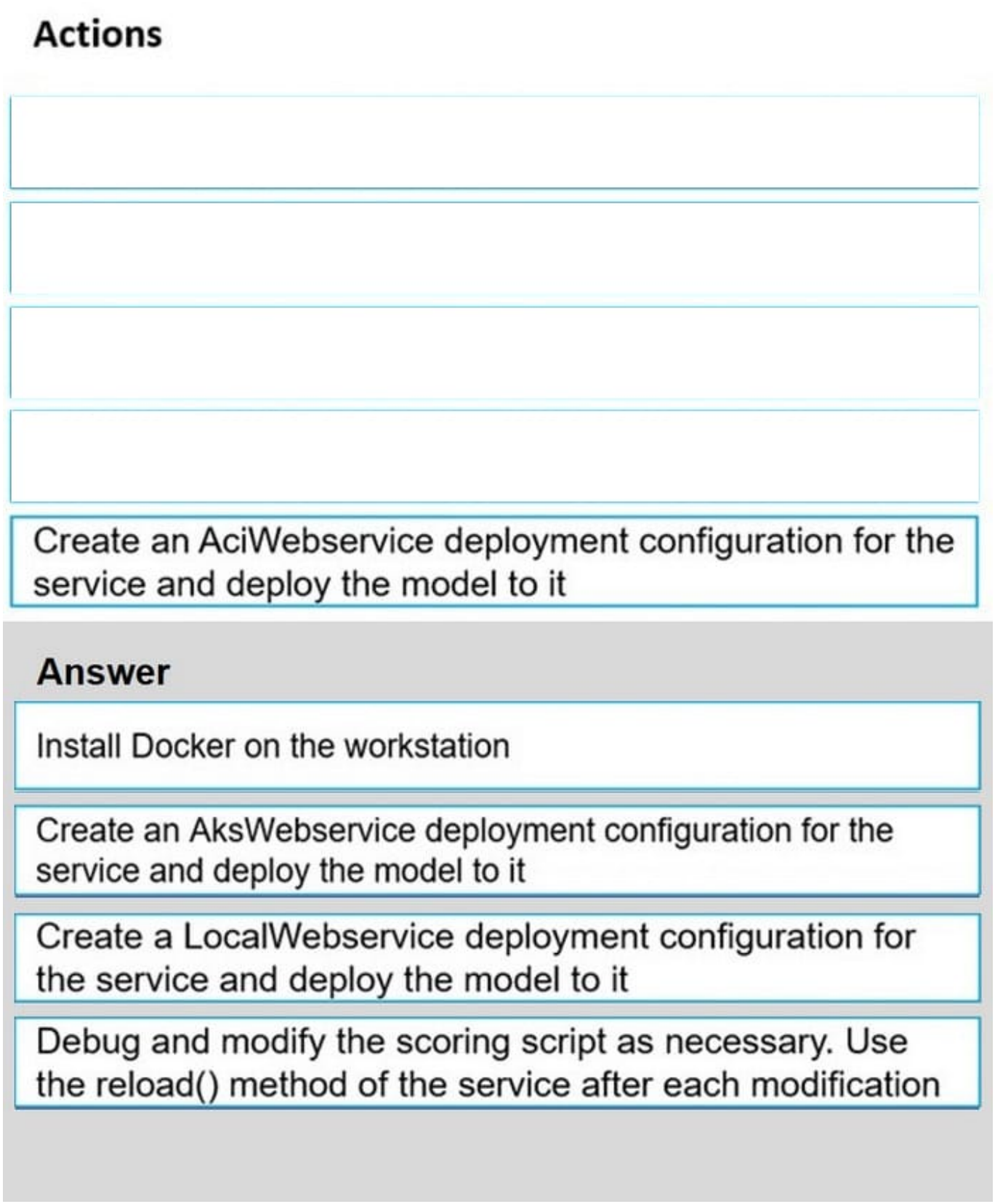
You need to debug the service on the local workstation before deploying the service to production.

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

Select and Place:



Correct Answer:



Step 1: Install Docker on the workstation

Prerequisites include having a working Docker installation on your local system.

Build or download the dockerfile to the compute node.

Step 2: Create an AksWebservice deployment configuration and deploy the model to it

To deploy a model to Azure Kubernetes Service, create a deployment configuration that describes the compute resources needed.

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)

service.wait_for_deployment(show_output = True)

print(service.state)

print(service.get_logs())

Step 3: Create a LocalWebservice deployment configuration for the service and deploy the model to it

To deploy locally, modify your code to use LocalWebservice.deploy_configuration() to create a deployment configuration. Then use Model.deploy() to deploy the service.

Step 4: Debug and modify the scoring script as necessary. Use the reload() method of the service after each modification.

During local testing, you may need to update the score.py file to add logging or attempt to resolve any problems that you\\'ve discovered. To reload changes to the score.py file, use reload(). For example, the following code reloads the script for

the service, and then sends data to it.

Incorrect Answers:

AciWebservice: The types of web services that can be deployed are LocalWebservice, which will deploy a model locally, and AciWebservice and AksWebservice, which will deploy a model to Azure Container Instances (ACI) and Azure

Kubernetes Service (AKS), respectively.

Reference:

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

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

QUESTION 9

HOTSPOT

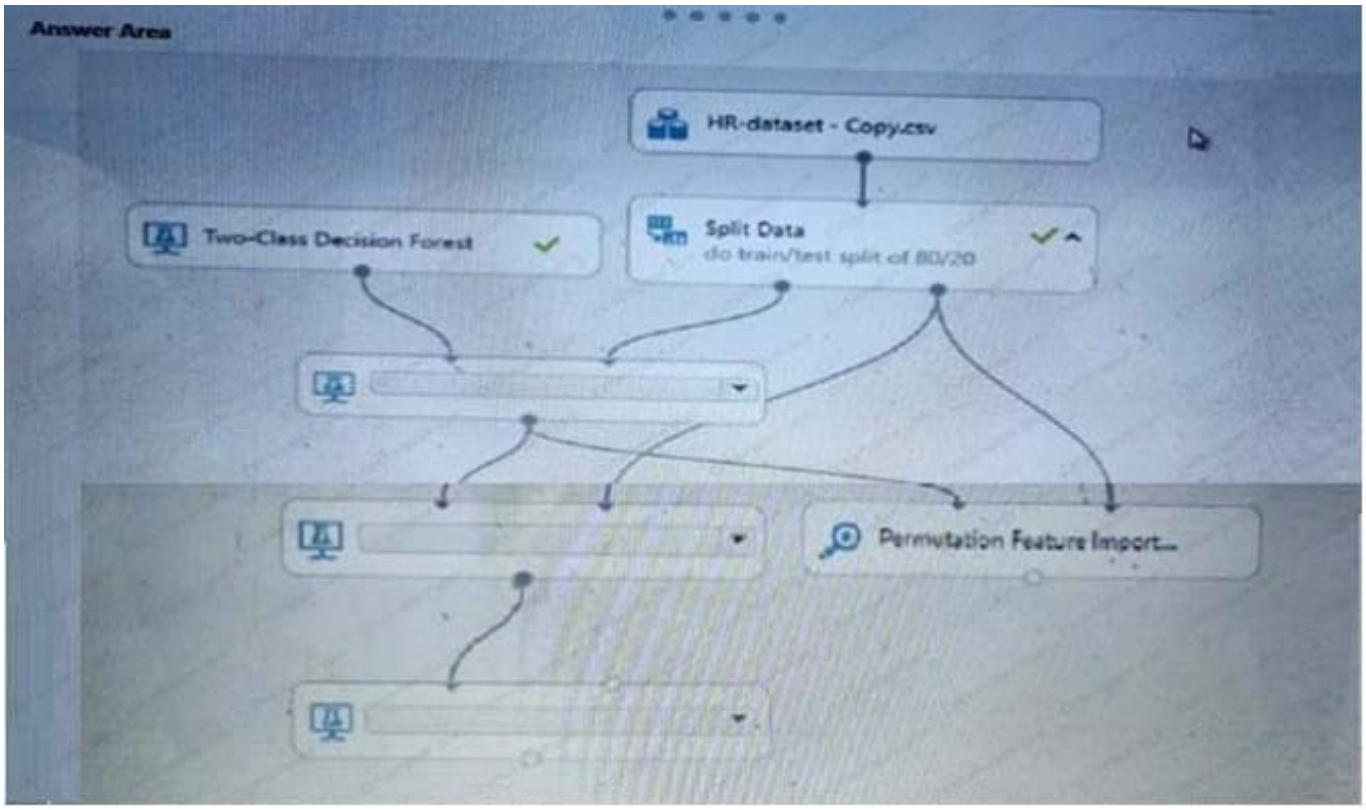
You create a binary classification model using Azure Machine Learning Studio.

You must use a Receiver Operating Characteristic (RO C) curve and an F1 score to evaluate the model.

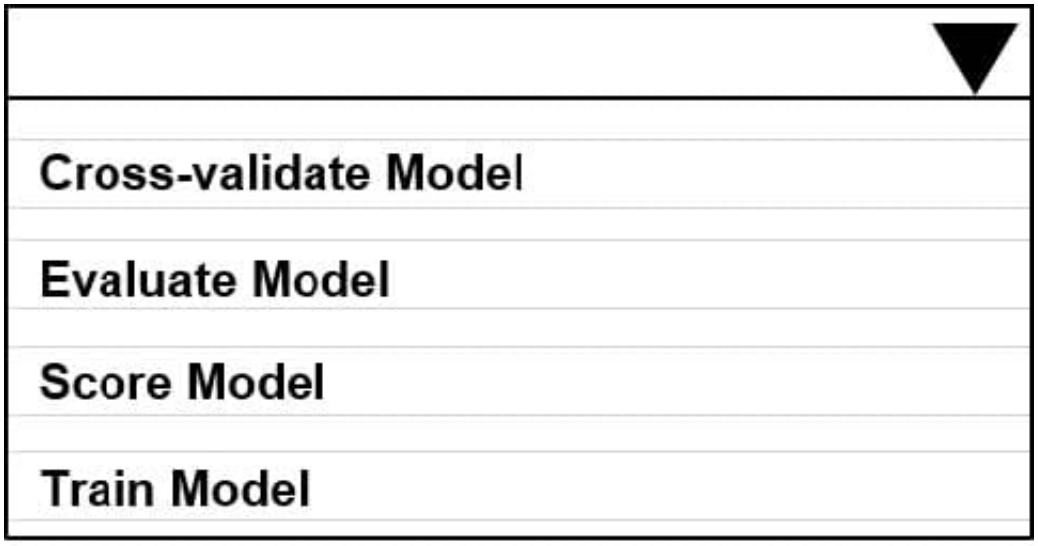
You need to create the required business metrics.

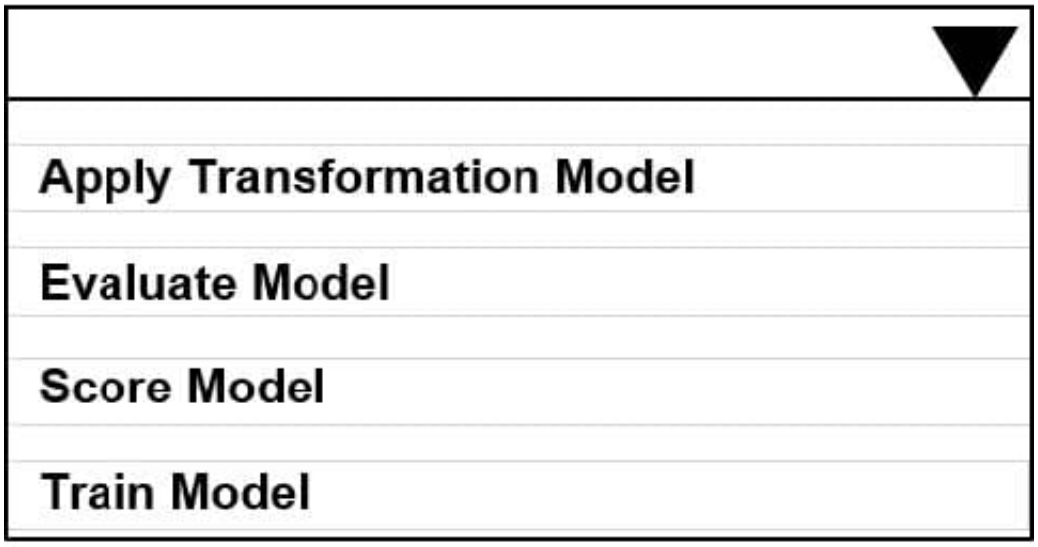
How should you complete the experiment? To answer, select the appropriate options in the dialog box in the answer area;

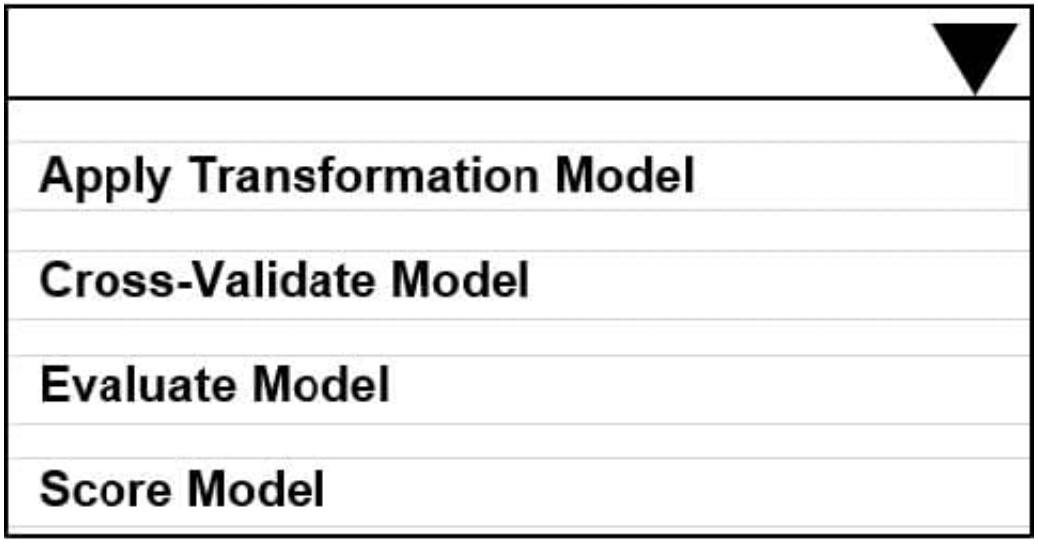
NOTE: Each correct selection is worth one point.



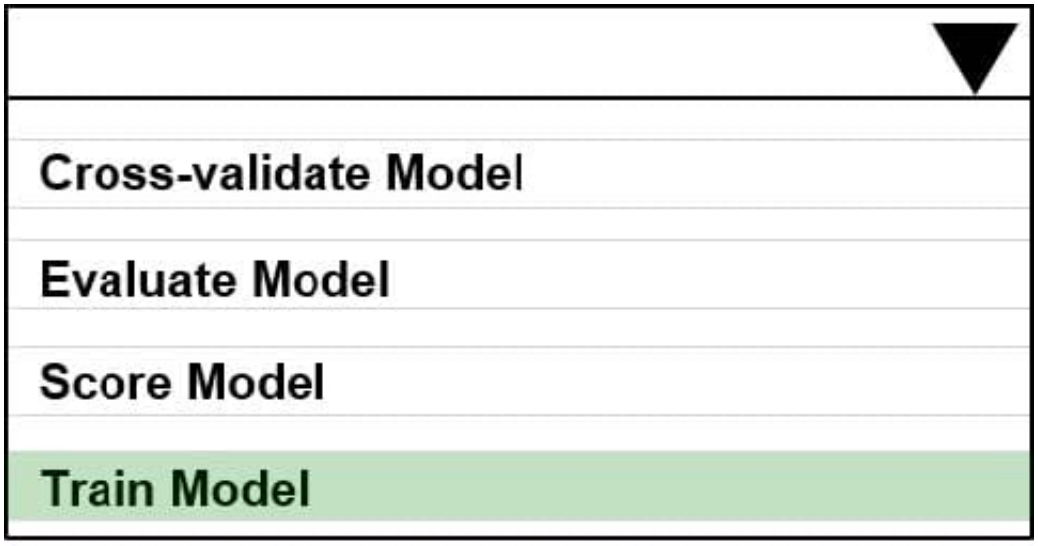
Hot Area:

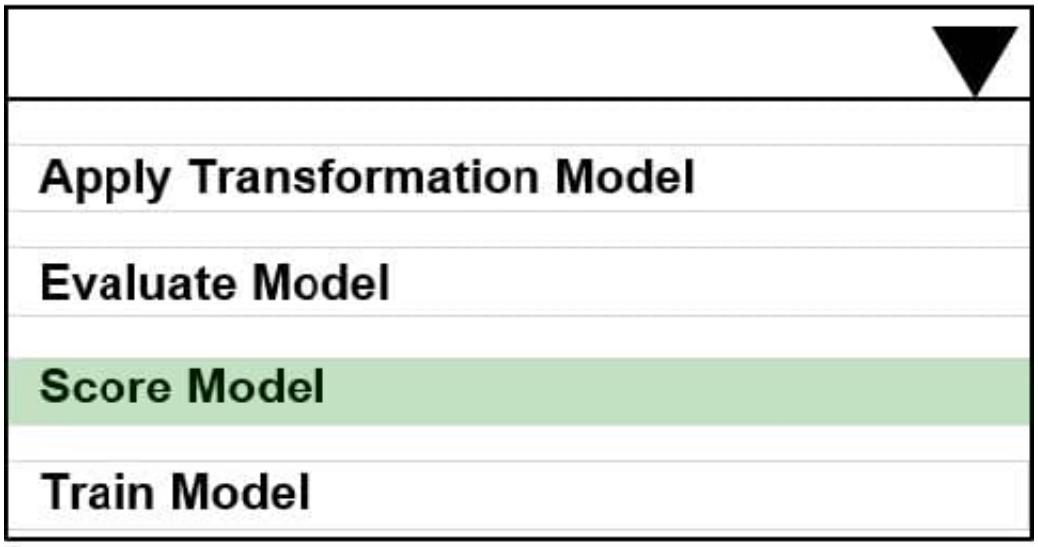

Cross-validate Model
Evaluate Model
Score Model
Train Model

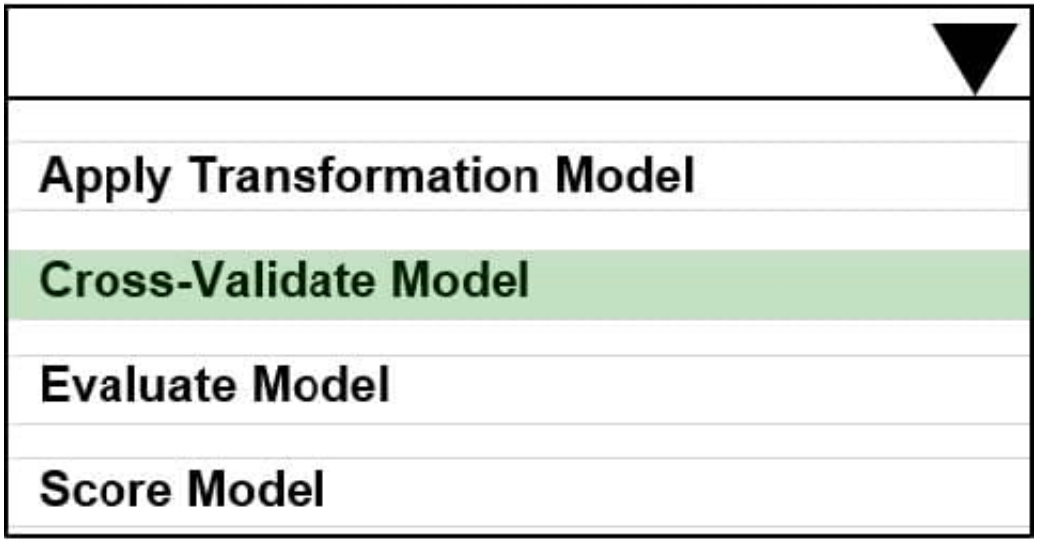

Apply Transformation Model
Evaluate Model
Score Model
Train Model


Apply Transformation Model
Cross-Validate Model
Evaluate Model
Score Model

Correct Answer:


Cross-validate Model
Evaluate Model
Score Model
Train Model


Apply Transformation Model
Evaluate Model
Score Model
Train Model


Apply Transformation Model
Cross-Validate Model
Evaluate Model
Score Model

QUESTION 10

You use the Azure Machine Learning service to create a tabular dataset named training_data. You plan to use this dataset in a training script.

You create a variable that references the dataset using the following code:

training_ds = workspace.datasets.get("training_data")

You define an estimator to run the script.

You need to set the correct property of the estimator to ensure that your script can access the training_data dataset.

Which property should you set?

A. environment_definition = {"training_data":training_ds}

B. inputs = [training_ds.as_named_input(\\'training_ds\\')]

C. script_params = {"--training_ds":training_ds}

D. source_directory = training_ds

Correct Answer: B

Example:

Get the training dataset

diabetes_ds = ws.datasets.get("Diabetes Dataset")

Create an estimator that uses the remote compute

hyper_estimator = SKLearn(source_directory=experiment_folder, inputs=[diabetes_ds.as_named_input(\\'diabetes\\')], # Pass the dataset as an input compute_target = cpu_cluster,

conda_packages=[\\'pandas\\',\\'ipykernel\\',\\'matplotlib\\'],

pip_packages=[\\'azureml-sdk\\',\\'argparse\\',\\'pyarrow\\'],

entry_script=\\'diabetes_training.py\\')

Reference:

https://notebooks.azure.com/GraemeMalcolm/projects/azureml-primers/html/04%20-%20Optimizing%20Model%20Training.ipynb

QUESTION 11

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.

Hot Area:

Answer Area

	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.	○	○
The estimator will mount the local data-folder folder and make it available to the script through a parameter.	○	○
The train.py script file will be created if it does not exist.	○	○
The estimator can run Scikit-learn experiments.	○	○

Correct Answer:

Answer Area

	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.	○	○
The estimator will mount the local data-folder folder and make it available to the script through a parameter.	○	○
The train.py script file will be created if it does not exist.	○	○
The estimator can run Scikit-learn experiments.	○	○

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.

QUESTION 12

DRAG DROP

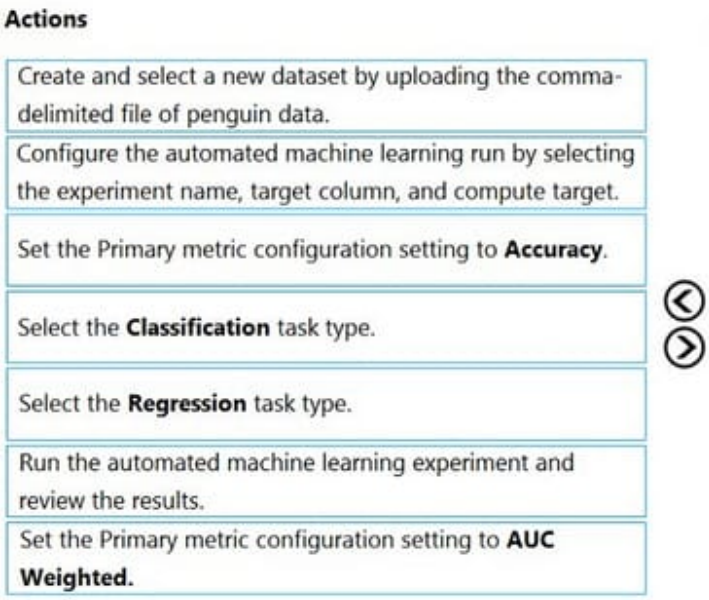
You are creating a machine learning model that can predict the species of a penguin from its measurements. You have a file that contains measurements for three species of penguin in comma-delimited format.

The model must be optimized for area under the received operating characteristic curve performance metric, averaged for each class.

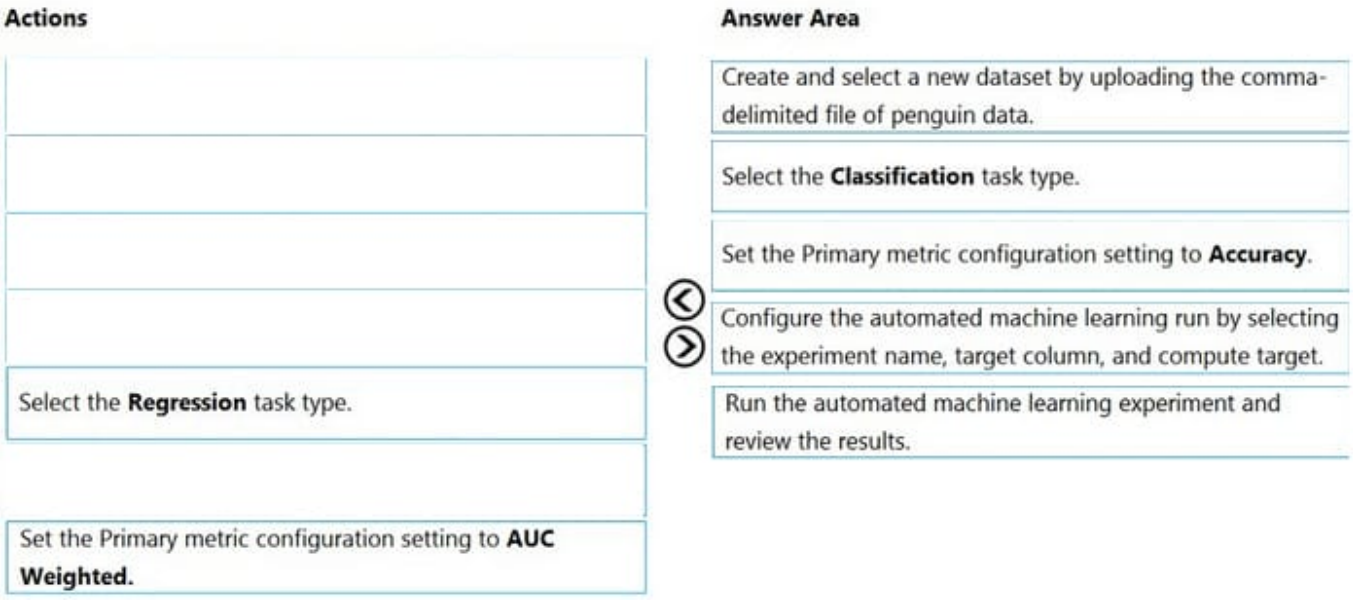
You need to use the Automated Machine Learning user interface in Azure Machine Learning studio to run an experiment and find the best performing model.

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

Select and Place:



Correct Answer:



Step 1:Create and select a new dataset by uploading he command-delimited file of penguin data.

Step 2: Select the Classification task type

Step 3: Set the Primary metric configuration setting to Accuracy.

The available metrics you can select is determined by the task type you choose.

Primary metrics for classification scenarios:

Post thresholded metrics, like accuracy, average_precision_score_weighted, norm_macro_recall, and precision_score_weighted may not optimize as well for datasets which are very small, have very large class skew (class imbalance), or

when the expected metric value is very close to 0.0 or 1.0. In those cases, AUC_weighted can be a better choice for the primary metric.

Step 4: Configure the automated machine learning run by selecting the experiment name, target column, and compute target

Step 5: Run the automated machine learning experiment and review the results.

Reference:

https://docs.microsoft.com/en-us/azure/machine-learning/how-to-configure-auto-train

QUESTION 13

HOTSPOT

You write code to retrieve an experiment that is run from your Azure Machine Learning workspace.

The run used the model interpretation support in Azure Machine Learning to generate and upload a model explanation.

Business managers in your organization want to see the importance of the features in the model.

You need to print out the model features and their relative importance in an output that looks similar to the following.

Feature	Importance
0	1.5627435610083558
2	0.6077689312583112
4	0.5574002432900718
3	0.42858759955671777
1	0.3501361539771977

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

NOTE: Each correct selection is worth one point.

Hot Area:

Answer Area

```
# Assume required modules are imported

ws = Workspace.from_config()
feature_importances = explanation.[ ▼ ]( workspace = ws,
                                         experiment_name='train_and_explain',
                                         run_id='train_and_explain_12345')
explanation = client.[ ▼ ]()
feature_importances = explanation.[ ▼ ]()

for key, value in feature_importances.items():
        print(key, "\t", value)
```

- [ ]: from_run, list_model_explanations, from_run_id, download_model_explanation
- [ ]: upload_model_explanation, list_model_explanations, run, download_model_explanation
- [ ]: explanation, explanation_client, get_feature_important_dict, download_model_explanation

Correct Answer:

Answer Area

```
# Assume required modules are imported

ws = Workspace.from_config()
feature_importances = explanation.[ from_run_id ]( workspace = ws,
                                                   experiment_name='train_and_explain',
                                                   run_id='train_and_explain_12345')
explanation = client.[ list_model_explanations ]()
feature_importances = explanation.[ explanation ]()

for key, value in feature_importances.items():
        print(key, "\t", value)
```

Box 1: from_run_id

from_run_id(workspace, experiment_name, run_id)

Create the client with factory method given a run ID.

Returns an instance of the ExplanationClient.

Parameters

Workspace Workspace An object that represents a workspace.

experiment_name str The name of an experiment.

run_id str A GUID that represents a run.

Box 2: list_model_explanations

list_model_explanations returns a dictionary of metadata for all model explanations available.

Returns

A dictionary of explanation metadata such as id, data type, explanation method, model type, and upload time, sorted by upload time

Box 3: explanation

Reference:

https://docs.microsoft.com/en-us/python/api/azureml-contrib-interpret/azureml.contrib.interpret.explanation.explanation_client.explanationclient?view=azure-ml-py

QUESTION 14

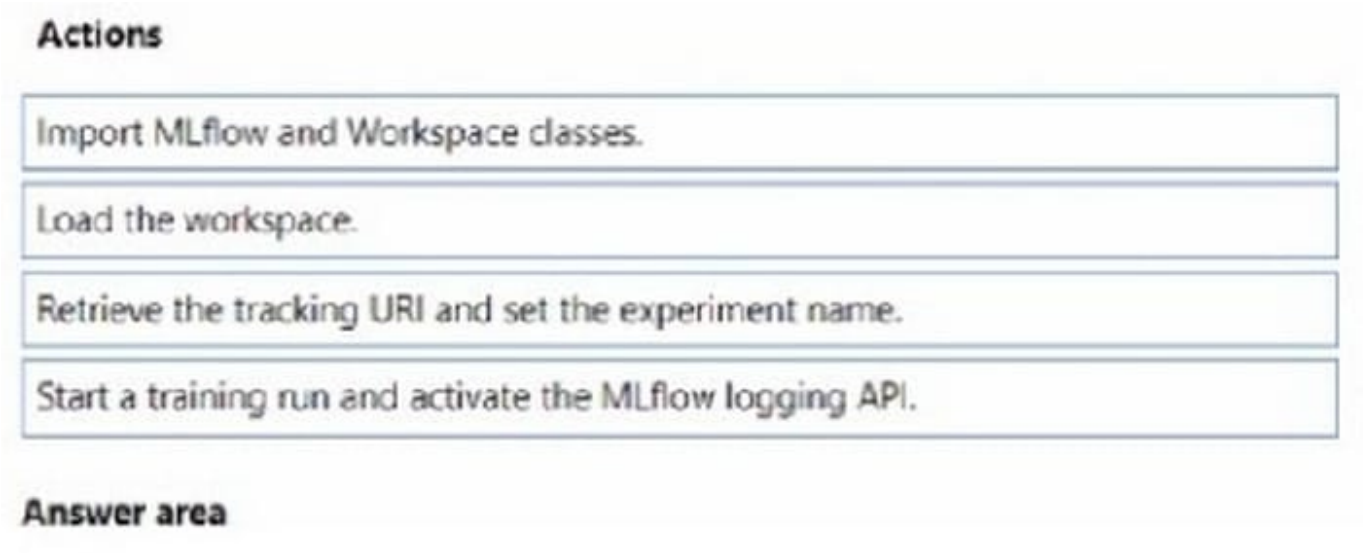
DRAG DROP

You have an Azure Machine Learning workspace. You are running an experiment on your local computer.

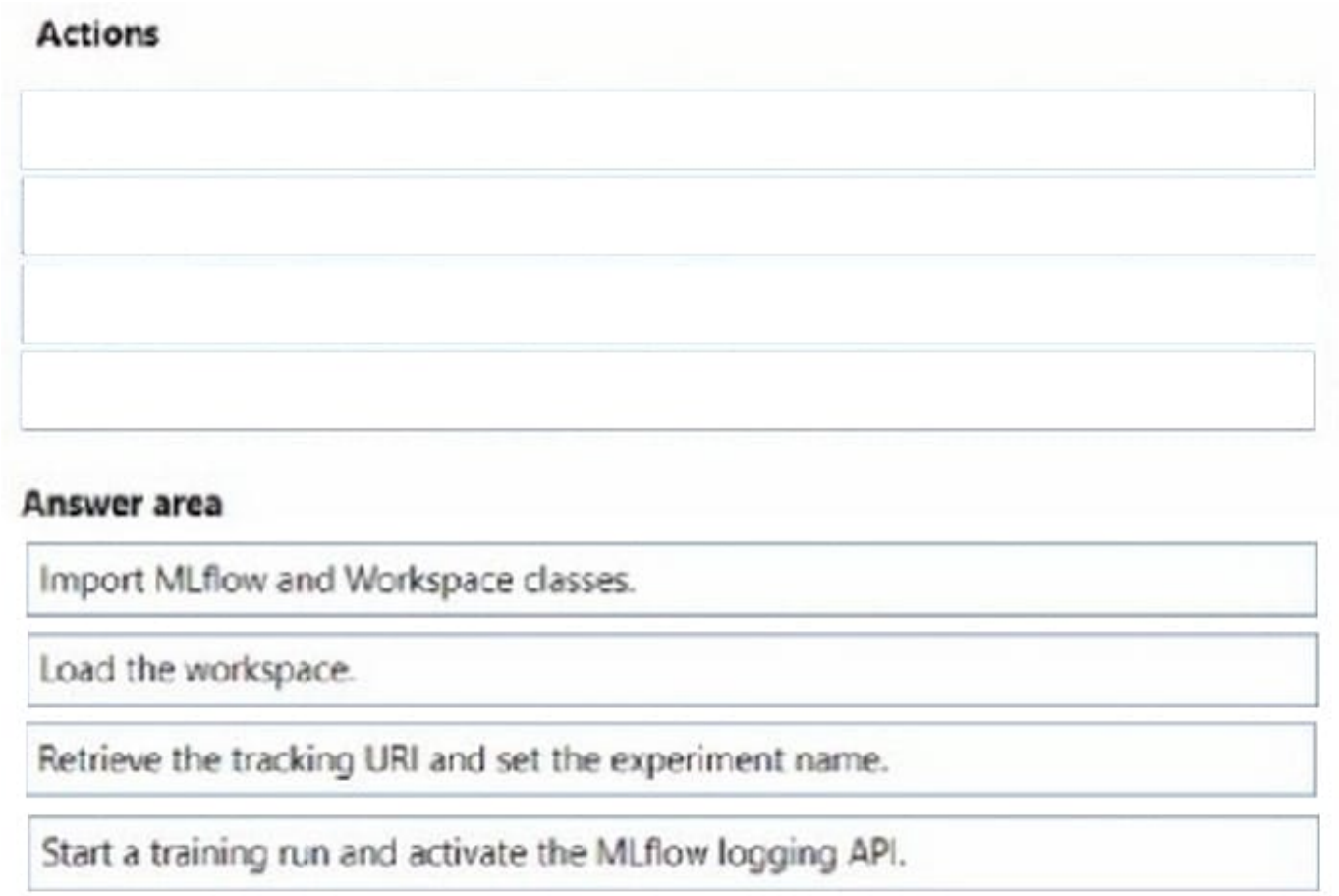
You need to use MLflow Tracking to store metrics and artifacts from your local experiment runs in the workspace.

In which order should you perform the actions? To answer, move all actions from the list of actions to the answer area and arrange them in the correct order.

Select and Place:



Correct Answer:



QUESTION 15

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_clasification	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')
best_iter = automl_ex.archived_time.find('11/11/2019 11:00:00 AM')
```

B.
```
from azureml.core import Workspace
from azureml.train.automl.run import AutoMLRun
automl_ex = ws.experiments.get('auto_ml_classification')
automl_run = AutoMLRun(automl_ex, 'AutoML_1234567890-123')
best_iter = automl_run.current_run
```

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
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]
```

E.
```
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 = automl_ex.get_runs('AutoML_1234567890-123')
```

A. Option A

B. Option B

C. Option C

D. Option D

E. Option E

Correct Answer: D

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-machine-learning/classification-with-deployment/auto-ml-classification-with-deployment.ipynb