



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.geekcert.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

You are building a regression model for estimating the number of calls during an event.

You need to determine whether the feature values achieve the conditions to build a Poisson regression model.

Which two conditions must the feature set contain? Each correct answer presents part of the solution.

NOTE: Each correct selection is worth one point.

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

Correct Answer: BD

Poisson regression is intended for use in regression models that are used to predict numeric values, typically counts. Therefore, you should use this module to create your regression model only if the values you are trying to predict fit the following conditions:

1.

The response variable has a Poisson distribution.

2.

Counts cannot be negative. The method will fail outright if you attempt to use it with negative labels.

3.

A Poisson distribution is a discrete distribution; therefore, it is not meaningful to use this method with non-whole numbers.

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

QUESTION 2

Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while

others might not have a correct solution.

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

You are creating a model to predict the price of a student's artwork depending on the following variables:



the student's length of education, degree type, and art form.

You start by creating a linear regression model.

You need to evaluate the linear regression model.

Solution: Use the following metrics: Mean Absolute Error, Root Mean Absolute Error, Relative Absolute Error, Relative Squared Error, and the Coefficient of Determination.

Does the solution meet the goal?

A. Yes

B. No

Correct Answer: A

The following metrics are reported for evaluating regression models. When you compare models, they are ranked by the metric you select for evaluation.

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

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.

Relative absolute error (RAE) is the relative absolute difference between expected and actual values; relative because the mean difference is divided by the arithmetic mean.

Relative squared error (RSE) similarly normalizes the total squared error of the predicted values by dividing by the total squared error of the actual values.

Mean Zero One Error (MZOE) indicates whether the prediction was correct or not. In other words:

$\text{ZeroOneLoss}(x,y) = 1$ when $x \neq y$; otherwise 0.

Coefficient of determination, often referred to as R^2 , 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 R^2 values, as low values can be entirely normal and high values can be suspect.

AUC.

Reference:

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

QUESTION 3

You run a script as an experiment in Azure Machine Learning.

You have a Run object named run that references the experiment run. You must review the log files that were generated during the experiment run.

You need to download the log files to a local folder for review.



Which two code segments can you run to achieve this goal? Each correct answer presents a complete solution.

NOTE: Each correct selection is worth one point.

- A. `run.get_details()`
- B. `run.get_file_names()`
- C. `run.get_metrics()`
- D. `run.download_files(output_directory=\\./runfiles\\)`
- E. `run.get_all_logs(destination=\\./runlogs\\)`

Correct Answer: AE

The run Class `get_all_logs` method downloads all logs for the run to a directory.

The run Class `get_details` gets the definition, status information, current log files, and other details of the run.

Incorrect Answers:

B: The run `get_file_names` list the files that are stored in association with the run.

Reference: [https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.core.run\(class\)](https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.core.run(class))

QUESTION 4

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

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

You create the following Python data frames:

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

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

NOTE: Each correct selection is worth one point.

Hot Area:



Answer Area

```
from sklearn.decomposition import PCA
pca = PCA()
X_train = pca.fit_transform(X_train)
x_test = pca.transform(x_test)
```

Correct Answer:

Answer Area

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



Box 1: `PCA(n_components = 10)`

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

Example:

```
from sklearn.decomposition import PCA
```

```
pca = PCA(n_components=2) ;2 dimensions
```

```
principalComponents = pca.fit_transform(x)
```

Box 2: `pca`

`fit_transform(X[, y])` fits the model with X and apply the dimensionality reduction on X.

Box 3: `transform(x_test)`

`transform(X)` applies dimensionality reduction to X.

References:

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

QUESTION 5

You use the Azure Machine learning SDK for Python to create a pipeline that includes the following step:

The output of the step run must be cached and reused on subsequent runs when the `source.directory` value has not changed.

You need to define the step.

What should you include in the step definition?

- A. `allow_reuse`
- B. `hash_path`
- C. `data_as_input(name=..)`
- D. `version`

Correct Answer: A

[Latest DP-100 Dumps](#)

[DP-100 VCE Dumps](#)

[DP-100 Study Guide](#)