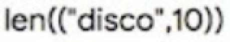
✓ What is the result of the following? 1/1



```
len(("disco",10))
```

○ 5

○ 6

● 2 ✓

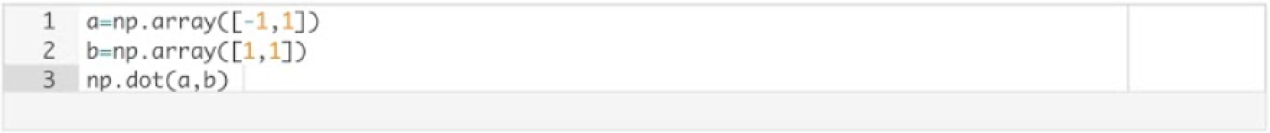
✓ What is the syntax to clone the list A and assign the result to list B 1/1

○ B = A

● B = A[:] ✓

○ B = copy(A)

✓ What is the result of the following lines of code: 1/1

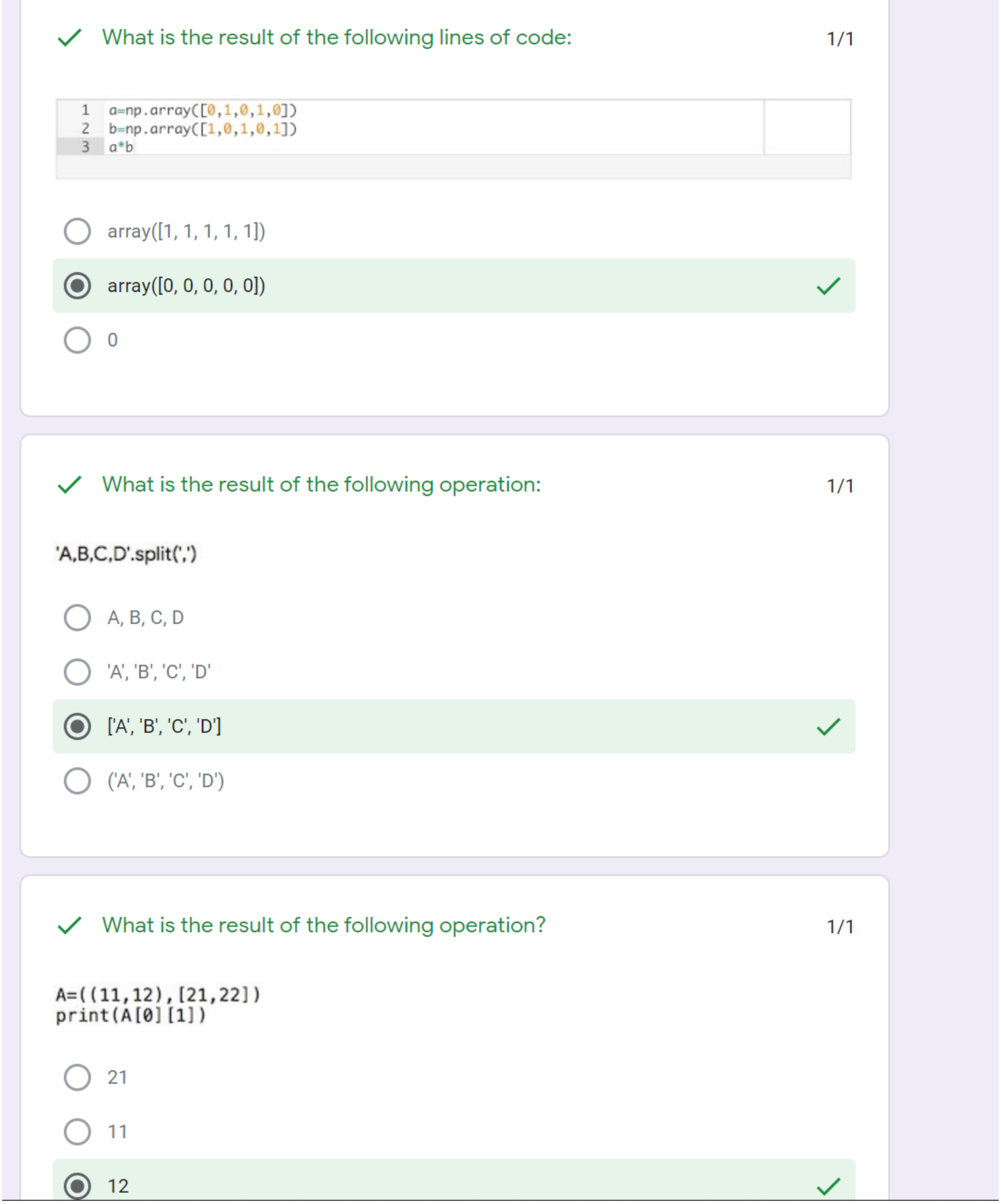


```
a=np.array([-1,1])
b=np.array([1,1])
np.dot(a,b)
```

● 0 ✓

○ 1

○ array([0, 2])



✓ What is the result of the following lines of code: 1/1

```
a=np.array([0,1,0,1,0])
b=np.array([1,0,1,0,1])
a*b
```

○ array([1, 1, 1, 1, 1])

● array([0, 0, 0, 0, 0]) ✓

○ 0

✓ What is the result of the following operation: 1/1

'A,B,C,D'.split(',')

○ A, B, C, D

○ 'A', 'B', 'C', 'D'

● ['A', 'B', 'C', 'D'] ✓

○ ('A', 'B', 'C', 'D')

✓ What is the result of the following operation? 1/1

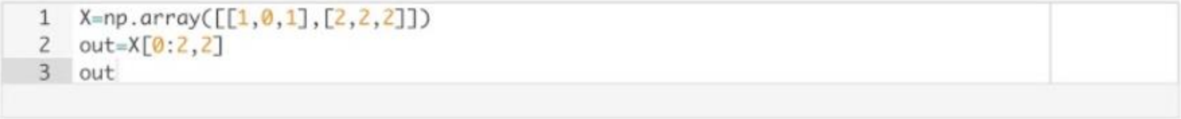
```
A=((11,12),[21,22])
print(A[0][1])
```

○ 21

○ 11

● 12 ✓

✓ What values does the variable out take if the following lines of code are run: 1/1



```
X=np.array([[1,0,1],[2,2,2]])
out=X[0:2,2]
out
```

● array([1, 2]) ✓

○ array([1, 1])

○ array([1, 0])

✓ What is an important difference between lists and tuples? 1/1

○ Tuples can only have integers

○ List and tuples are the same

○ List can't contain strings

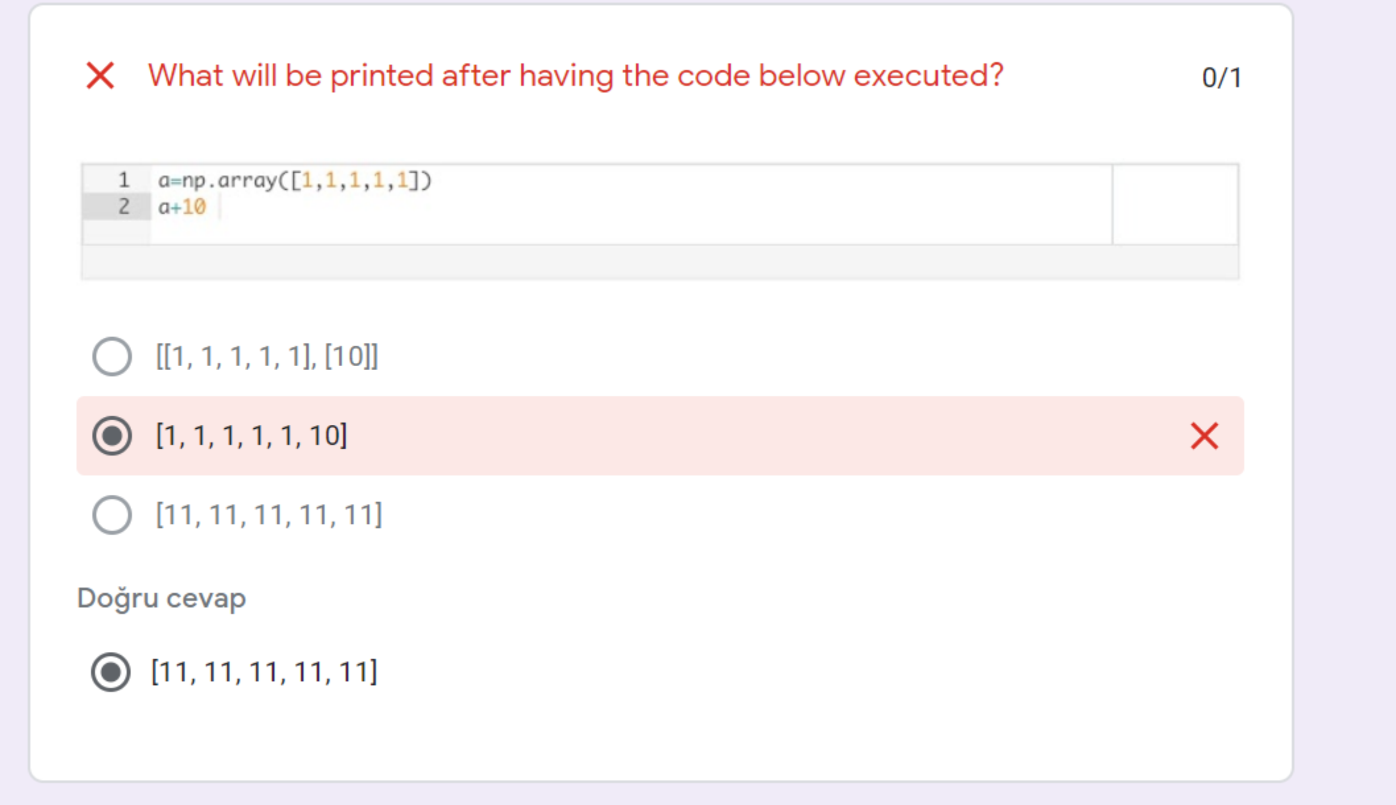
● Lists are mutable tuples are not ✓

✓ How would you select the columns with the headers: "artist, length and genre" from the dataframe df and assign them to the variable y 1/1

● y = df[['artist', 'length', 'genre']] ✓

○ y = df[['artist'], ['length'], ['genre']]

○ y = df['artist', 'length', 'genre']



✗ What will be printed after having the code below executed? 0/1

```
a=np.array([1,1,1,1,1])
a+10
```

○ [[1, 1, 1, 1, 1], [10]]

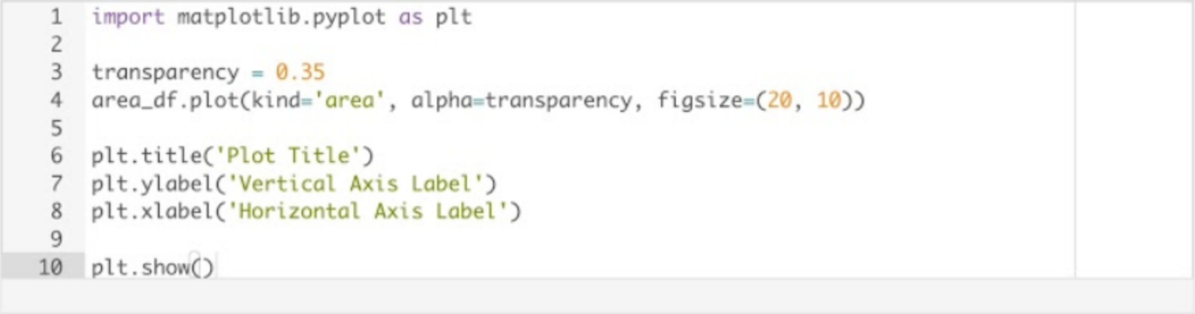
● [1, 1, 1, 1, 1, 10] ✗

○ [11, 11, 11, 11, 11]

Doğru cevap

● [11, 11, 11, 11, 11]

✗ The following code will create an unstacked area plot of the data in the pandas dataframe, area_df, with a transparency value of 0.35? 0/1



```
import matplotlib.pyplot as plt

transparency = 0.35
area_df.plot(kind='area', alpha=transparency, figsize=(20, 10))

plt.title('Plot Title')
plt.ylabel('Vertical Axis Label')
plt.xlabel('Horizontal Axis Label')

plt.show()
```

○ False

● True ✗

Doğru cevap

● False

✓ The easiest way to create a waffle chart in Python is using the Python package, PyWaffle. 1/1

○ False

● True ✓

✓ Stamen Terrain is the right tile style of Folium maps for visualizing and exploring river meanders and coastal zones of a given geographical area. 1/1

○ True

● False ✓

✗ Which of the following are examples of Matplotlib magic functions? Choose all that apply. 0/1

☑ %matplotlib inline ✓

□ %matplotlib notebook

☑ #matplotlib inline ✗

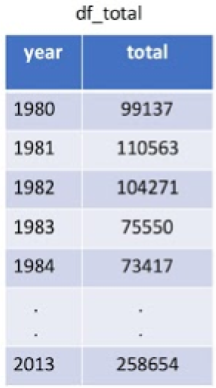
□ $matplotlib outline

Doğru cevap

☑ %matplotlib notebook

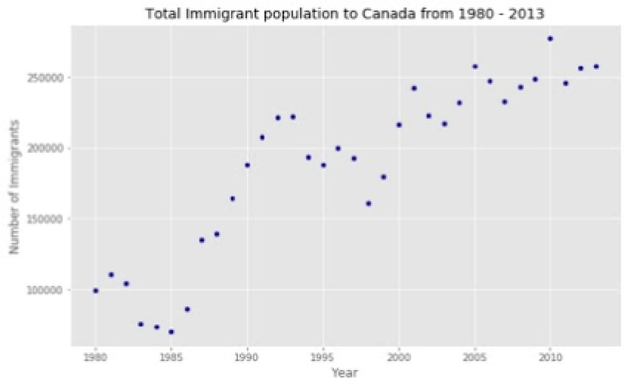
☑ %matplotlib inline

✗ Which of the code below will create the following scatter plot, given the pandas dataframe, df_total? (Note: Forget about the title and labels. Focus on the plot function only.) 0/1



df_total

year	total
1980	99137
1981	110563
1982	104271
1983	75550
1984	73417
.	.
.	.
2013	258654



○ df_total.plot(type='scatter', x='year', y='total')

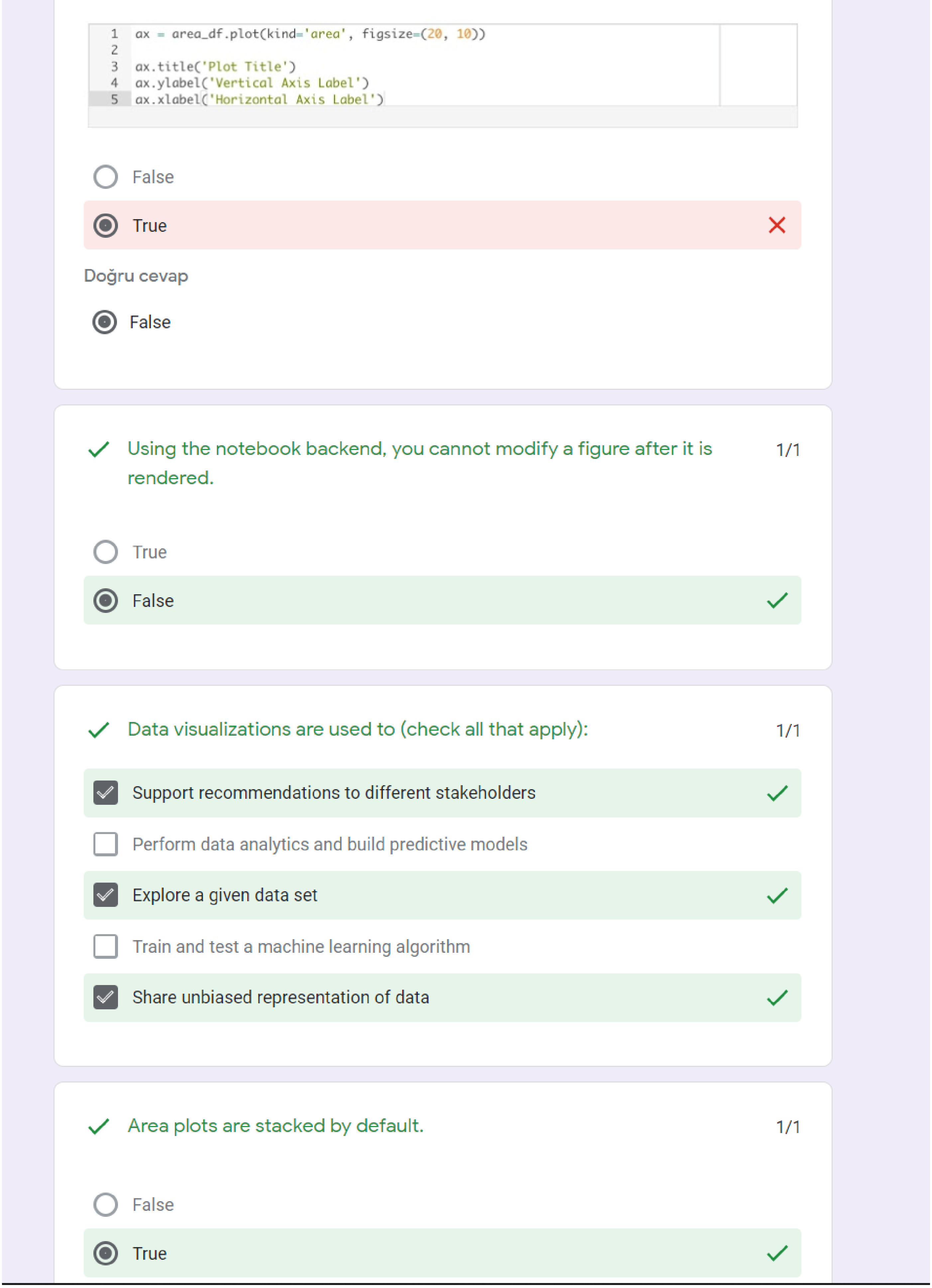
● plot(kind='scatter', x='year', y='total', data=df_total) ✗

○ plot(type='scatter', x='year', y='total', data=df_total)

○ df_total.plot(kind='scatter', x='year', y='total')

Doğru cevap

● df_total.plot(kind='scatter', x='year', y='total')



```
ax = area_df.plot(kind='area', figsize=(20, 10))

ax.title('Plot Title')
ax.ylabel('Vertical Axis Label')
ax.xlabel('Horizontal Axis Label')
```

○ False

● True ✗

Doğru cevap

● False

✓ Using the notebook backend, you cannot modify a figure after it is rendered. 1/1

○ True

● False ✓

✓ Data visualizations are used to (check all that apply): 1/1

☑ Support recommendations to different stakeholders ✓

□ Perform data analytics and build predictive models

☑ Explore a given data set ✓

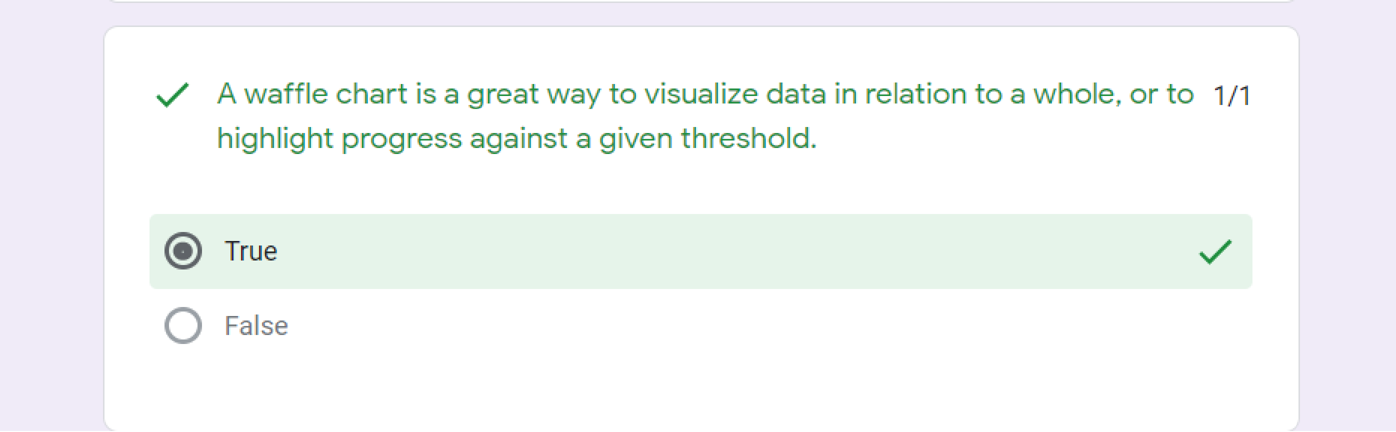
□ Train and test a machine learning algorithm

☑ Share unbiased representation of data ✓

✓ Area plots are stacked by default. 1/1

○ False

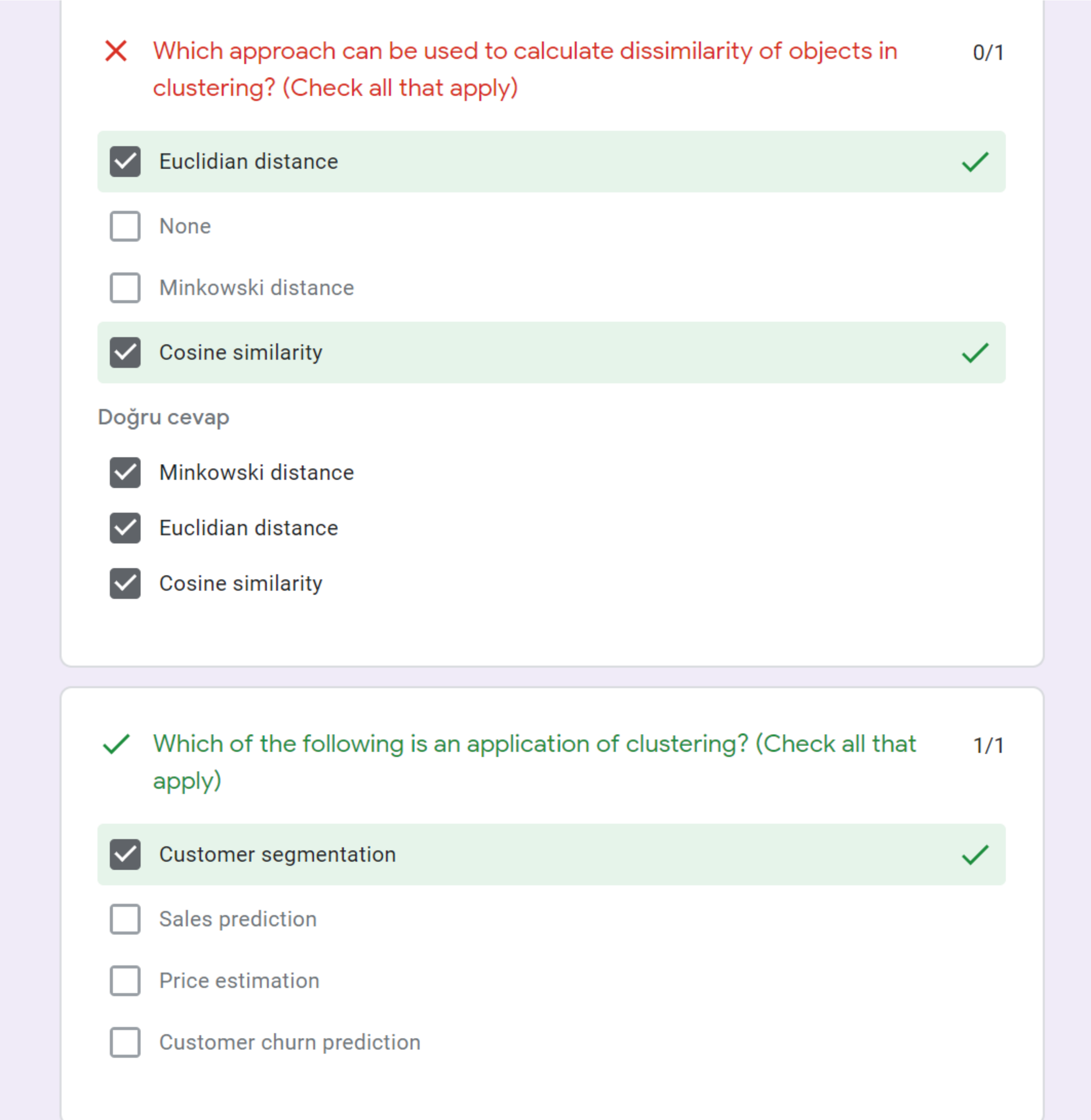
● True ✓



✓ A waffle chart is a great way to visualize data in relation to a whole, or to highlight progress against a given threshold. 1/1

● True ✓

○ False



✗ Which approach can be used to calculate dissimilarity of objects in clustering? (Check all that apply) 0/1

☑ Euclidian distance ✓

□ None

□ Minkowski distance

☑ Cosine similarity ✓

Doğru cevap

☑ Minkowski distance

☑ Euclidian distance

☑ Cosine similarity

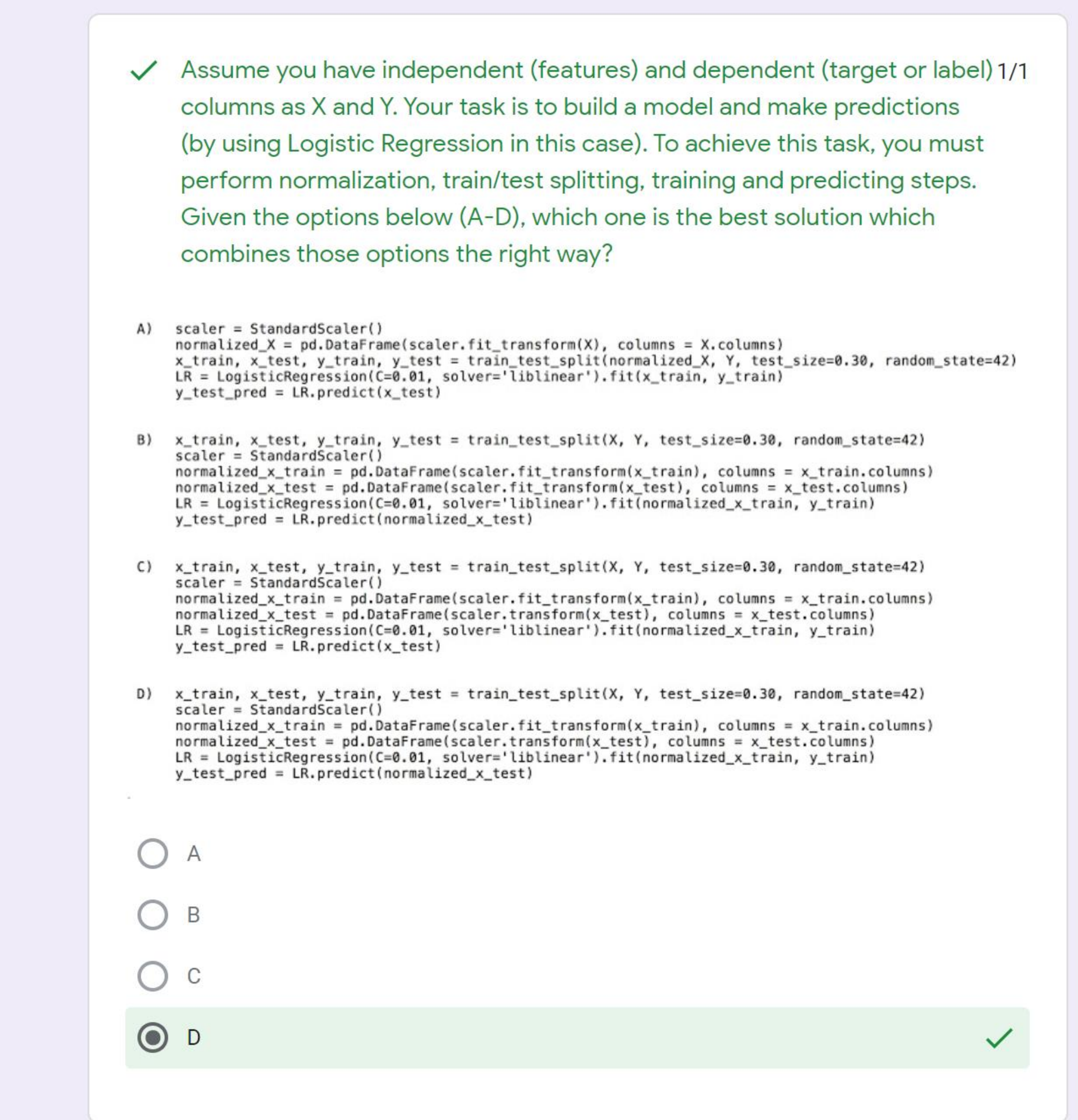
✓ Which of the following is an application of clustering? (Check all that apply) 1/1

☑ Customer segmentation ✓

□ Sales prediction

□ Price estimation

□ Customer churn prediction



✓ Assume you have independent (features) and dependent (target or label) columns as X and Y. Your task is to build a model and make predictions (by using Logistic Regression in this case). To achieve this task, you must perform normalization, train/test splitting, training and predicting steps. Given the options below (A-D), which one is the best solution which combines those options the right way? 1/1

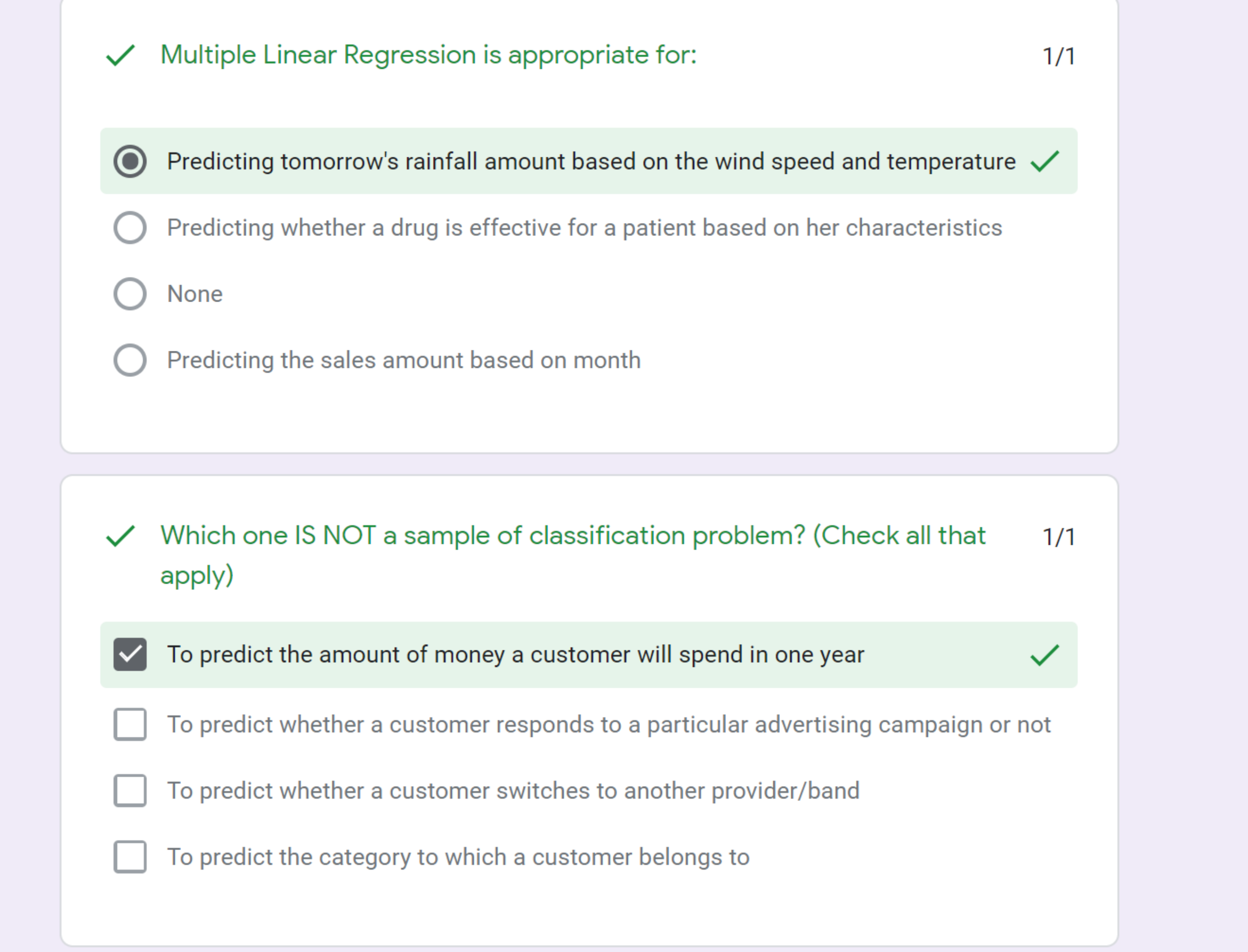
```
A)   scaler = StandardScaler()
     normalized_X = pd.DataFrame(scaler.fit_transform(X), columns = X.columns)
     x_train, x_test, y_train, y_test = train_test_split(normalized_X, Y, test_size=0.30, random_state=42)
     LR = LogisticRegression(C=0.01, solver='liblinear').fit(x_train, y_train)
     y_test_pred = LR.predict(x_test)

B)   x_train, x_test, y_train, y_test = train_test_split(X, Y, test_size=0.30, random_state=42)
     scaler = StandardScaler()
     normalized_x_train = pd.DataFrame(scaler.fit_transform(x_train), columns = x_train.columns)
     normalized_x_test = pd.DataFrame(scaler.fit_transform(x_test), columns = x_test.columns)
     LR = LogisticRegression(C=0.01, solver='liblinear').fit(normalized_x_train, y_train)
     y_test_pred = LR.predict(normalized_x_test)

C)   x_train, x_test, y_train, y_test = train_test_split(X, Y, test_size=0.30, random_state=42)
     scaler = StandardScaler()
     normalized_x_train = pd.DataFrame(scaler.fit_transform(x_train), columns = x_train.columns)
     normalized_x_test = pd.DataFrame(scaler.transform(x_test), columns = x_test.columns)
     LR = LogisticRegression(C=0.01, solver='liblinear').fit(normalized_x_train, y_train)
     y_test_pred = LR.predict(x_test)

D)   x_train, x_test, y_train, y_test = train_test_split(X, Y, test_size=0.30, random_state=42)
     scaler = StandardScaler()
     normalized_x_train = pd.DataFrame(scaler.fit_transform(x_train), columns = x_train.columns)
     normalized_x_test = pd.DataFrame(scaler.transform(x_test), columns = x_test.columns)
     LR = LogisticRegression(C=0.01, solver='liblinear').fit(normalized_x_train, y_train)
     y_test_pred = LR.predict(normalized_x_test)
```

○ A

○ B

○ C

● D ✓



✓ Multiple Linear Regression is appropriate for: 1/1

● Predicting tomorrow's rainfall amount based on the wind speed and temperature ✓

○ Predicting whether a drug is effective for a patient based on her characteristics

○ None

○ Predicting the sales amount based on month

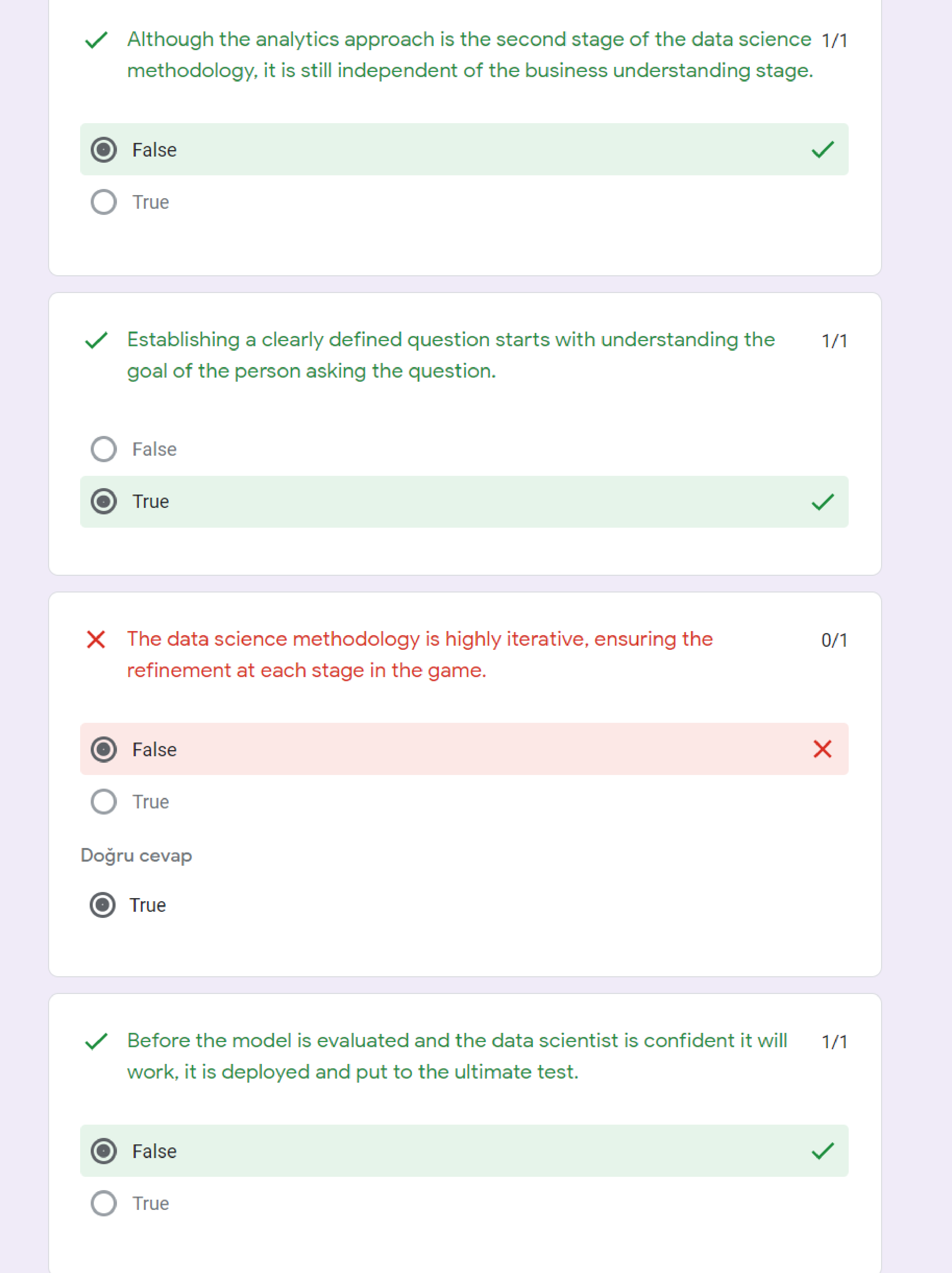
✓ Which one IS NOT a sample of classification problem? (Check all that apply) 1/1

☑ To predict the amount of money a customer will spend in one year ✓

□ To predict whether a customer responds to a particular advertising campaign or not

□ To predict whether a customer switches to another provider/band

□ To predict the category to which a customer belongs to



✓ Although the analytics approach is the second stage of the data science methodology, it is still independent of the business understanding stage. 1/1

● False ✓

○ True

✓ Establishing a clearly defined question starts with understanding the goal of the person asking the question. 1/1

○ False

● True ✓

✗ The data science methodology is highly iterative, ensuring the refinement at each stage in the game. 0/1

● False ✗

○ True

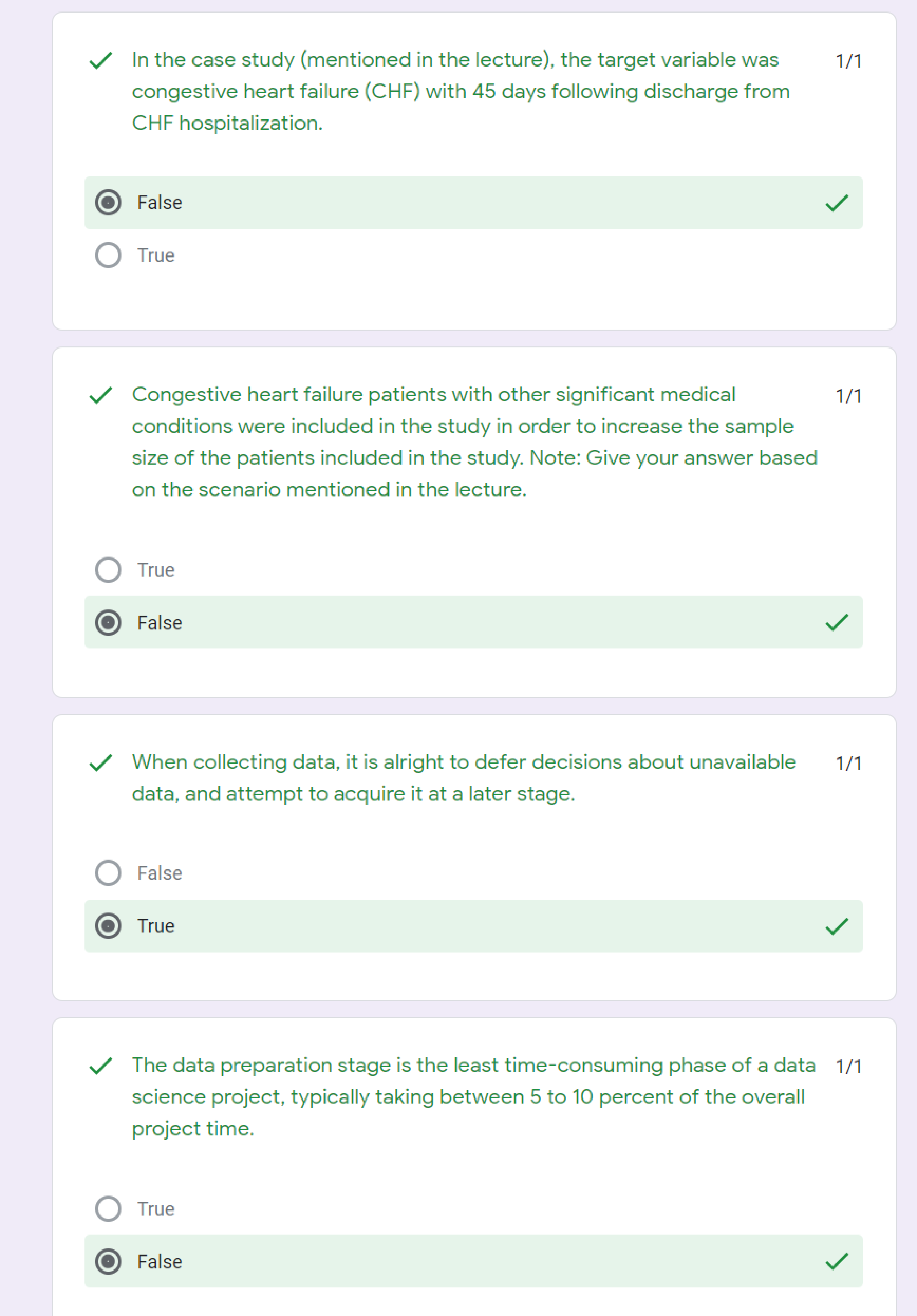
Doğru cevap

● True

✓ Before the model is evaluated and the data scientist is confident it will work, it is deployed and put to the ultimate test. 1/1

● False ✓

○ True



✓ In the case study (mentioned in the lecture), the target variable was congestive heart failure (CHF) with 45 days following discharge from CHF hospitalization. 1/1

● False ✓

○ True

✓ Congestive heart failure patients with other significant medical conditions were included in the study in order to increase the sample size of the patients included in the study. Note: Give your answer based on the scenario mentioned in the lecture. 1/1

○ True

● False ✓

✓ When collecting data, it is alright to defer decisions about unavailable data, and attempt to acquire it at a later stage. 1/1

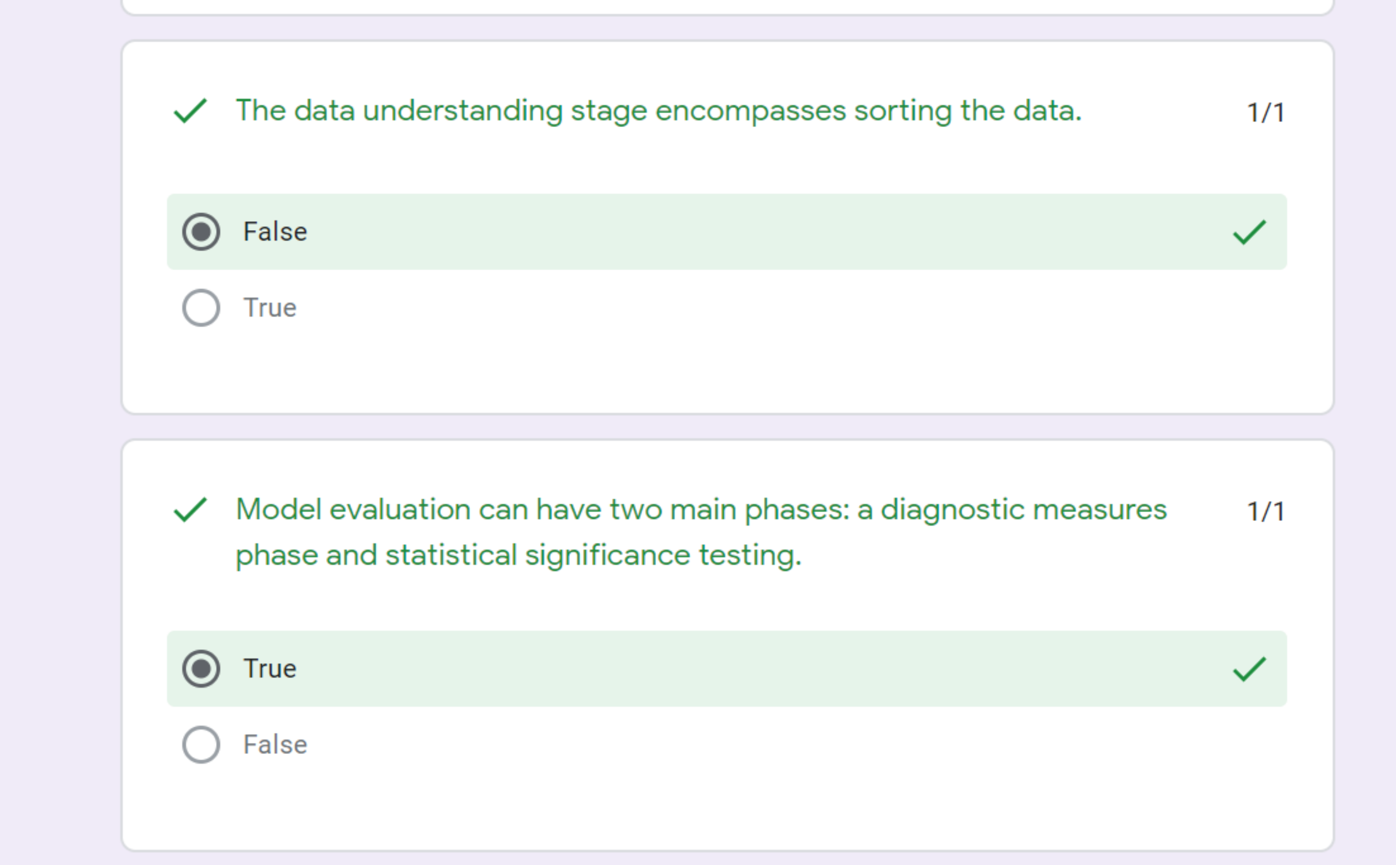
○ False

● True ✓

✓ The data preparation stage is the least time-consuming phase of a data science project, typically taking between 5 to 10 percent of the overall project time. 1/1

○ True

● False ✓



✓ The data understanding stage encompasses sorting the data. 1/1

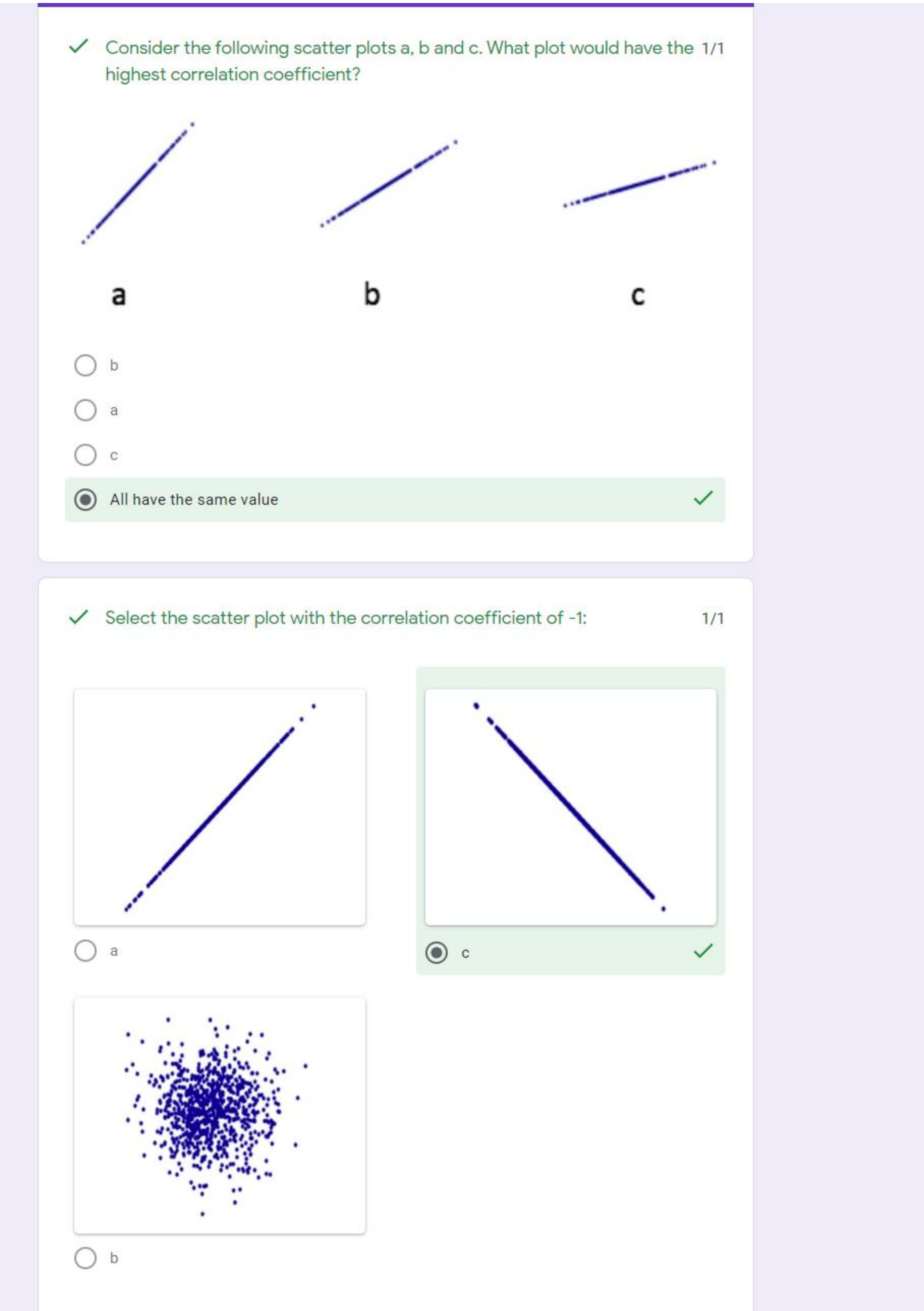
● False ✓

○ True

✓ Model evaluation can have two main phases: a diagnostic measures phase and statistical significance testing. 1/1

● True ✓

○ False



✓ Consider the following scatter plots a, b and c. What plot would have the highest correlation coefficient? 1/1

○ b

○ a

○ c

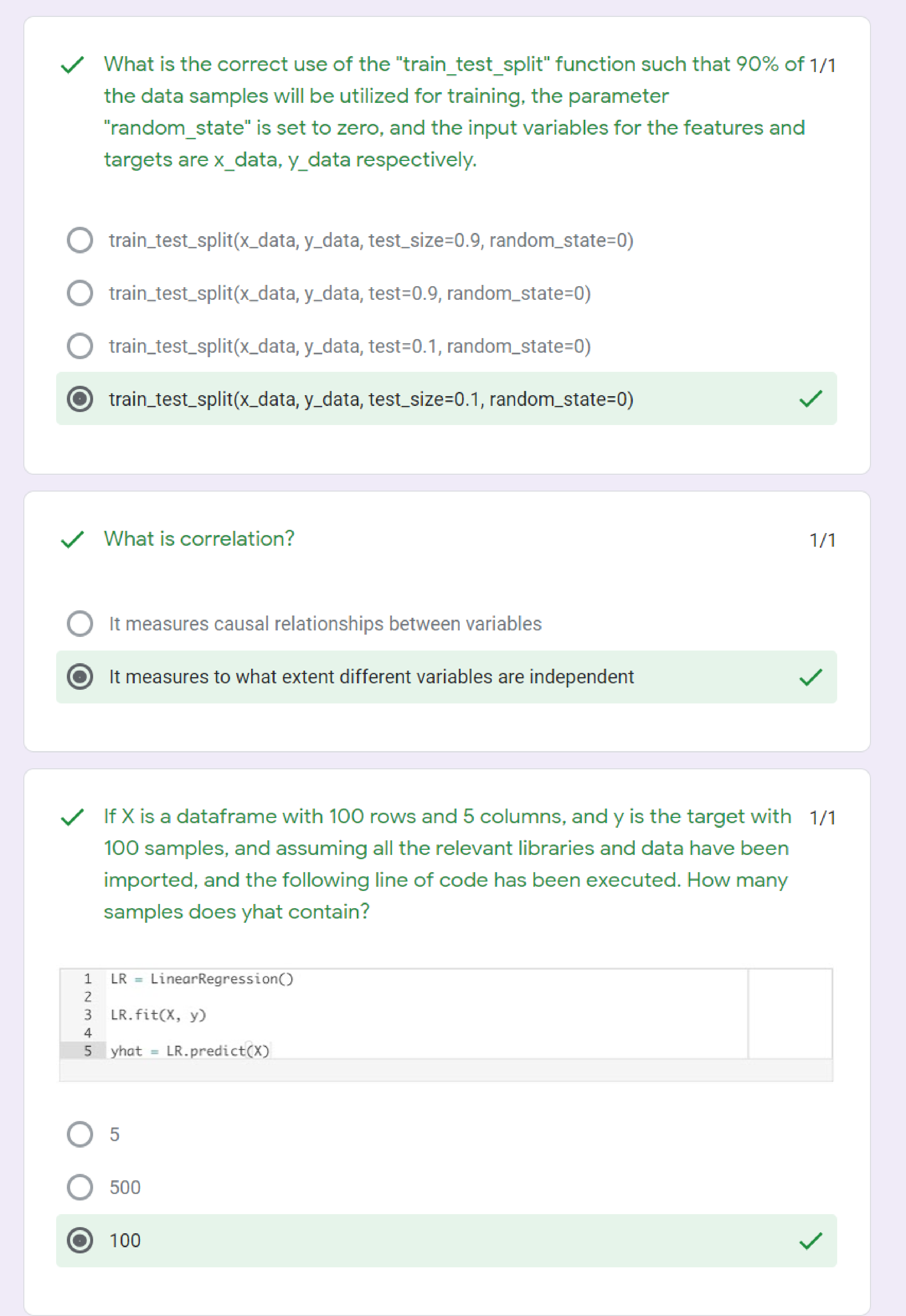
● All have the same value ✓

✓ Select the scatter plot with the correlation coefficient of -1: 1/1

○ a

● c ✓

○ b



✓ What is the correct use of the "train_test_split" function such that 90% of the data samples will be utilized for training, the parameter "random_state" is set to zero, and the input variables for the features and targets are x_data, y_data respectively. 1/1

○ train_test_split(x_data, y_data, test_size=0.9, random_state=0)

○ train_test_split(x_data, y_data, test=0.9, random_state=0)

○ train_test_split(x_data, y_data, test=0.1, random_state=0)

● train_test_split(x_data, y_data, test_size=0.1, random_state=0) ✓

✓ What is correlation? 1/1

○ It measures causal relationships between variables

● It measures to what extent different variables are independent ✓

✓ If X is a dataframe with 100 rows and 5 columns, and y is the target with 100 samples, and assuming all the relevant libraries and data have been imported, and the following line of code has been executed. How many samples does yhat contain? 1/1

```
LR = LinearRegression()

LR.fit(X, y)

yhat = LR.predict(X)
```

○ 5

○ 500

● 100 ✓