Which XML statement(s) given below is/are well-formed? (Check all that 2/2
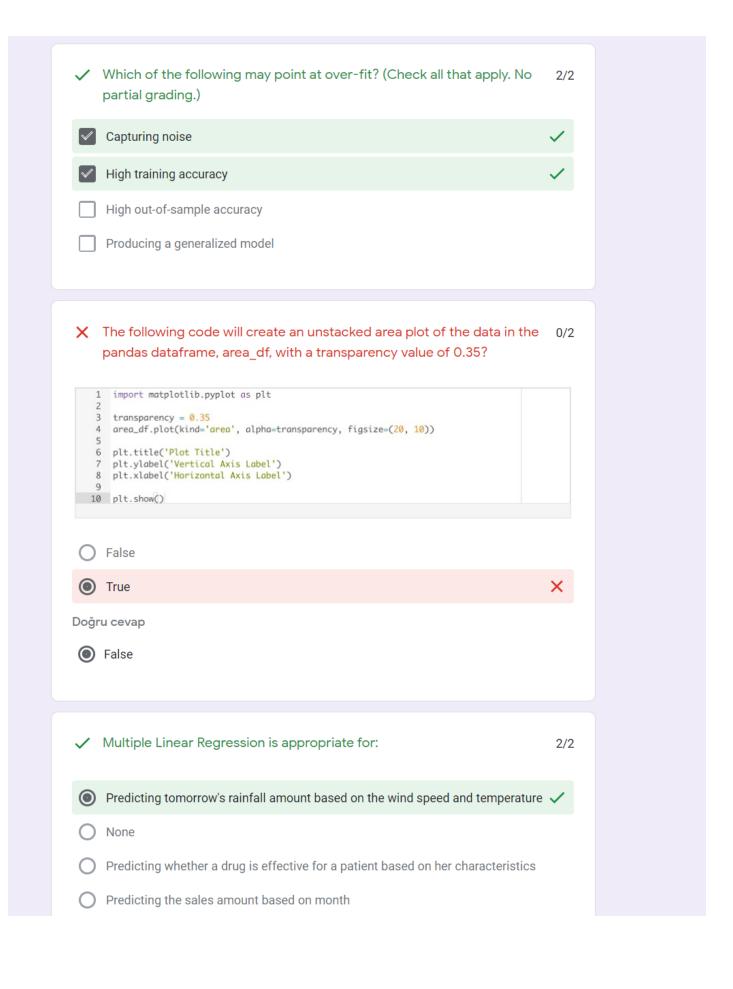apply. No partial grading.)

```
a) <!-- This is a comment -- quite useful one -->
b) <myElement myAttribute="value1 <!-anotherAttribute="value2" --> />
c) <!-- This is a comment: <xml:comment > -->
d) <!-- This is a comment <--! This is my old comment--> -->
```

☐ a

☐ b

☑ c ✓

☐ d

✓ Which of the following is an application of clustering? (Check all that 2/2
apply. No partial grading.)

☐ Customer churn prediction

☐ Price estimation

☐ Sales prediction

☑ Customer segmentation ✓

✓ Stamen Terrain is the right tile style of Folium maps for visualizing and 2/2
exploring river meanders and coastal zones of a given geographical area.

○ True

◉ False ✓

✓ The following code uses the artist layer to create a stacked area plot of 2/2
the data in the pandas dataframe, area_df.

```
1  ax = area_df.plot(kind='area', figsize=(20, 10))
2
3  ax.title('Plot Title')
4  ax.ylabel('Vertical Axis Label')
5  ax.xlabel('Horizontal Axis Label')
```

○ True

◉ False   ✓

✓ Which XML statement(s) given below is/are well-formed? (Check all that 2/2
apply. No partial grading.)

```
a) <myElement myAttribute="value1 > value2"/>
b) <myElement myAttribute="value1'value2"/>
c) <myElement myAttribute="value1 <= value2"/>
d) <myElement myAttribute="value1 & value2"/>
```

☑ a   ✓

☑ b   ✓

☐ c

☐ d

✓ Which approach can be used to calculate dissimilarity of objects in 2/2
clustering? (Check all that apply. No partial grading.)

☐ None

☑ Euclidian distance   ✓

☑ Cosine similarity   ✓

☑ Minkowski distance   ✓

✓ Which of the following may point at over-fit? (Check all that apply. No partial grading.) 2/2

- ☑ Capturing noise ✓
- ☑ High training accuracy ✓
- ☐ High out-of-sample accuracy
- ☐ Producing a generalized model

✗ 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? 0/2

```
1   import matplotlib.pyplot as plt
2
3   transparency = 0.35
4   area_df.plot(kind='area', alpha=transparency, figsize=(20, 10))
5
6   plt.title('Plot Title')
7   plt.ylabel('Vertical Axis Label')
8   plt.xlabel('Horizontal Axis Label')
9
10  plt.show()
```

- ○ False
- ◉ True ✗

Doğru cevap

- ◉ False

✓ Multiple Linear Regression is appropriate for: 2/2

- ◉ Predicting tomorrow's rainfall amount based on the wind speed and temperature ✓
- ○ None
- ○ Predicting whether a drug is effective for a patient based on her characteristics
- ○ Predicting the sales amount based on month

Final Exam Key

Part II

```
df_canada = df_original.drop(columns=['Type', 'Coverage', 'AREA', 'REG', 'DEV'])
df_canada.rename(columns={'OdName':'Country', 'AreaName':'Continent', 'RegName':'Region'}, inplace=True)
df_canada.set_index('Country', inplace=True)
df_canada['Total'] = df_canada.sum(axis=1)
```

Part III

Virtual partitioning requires full replication.

Coexisting schemas: 1) To serve different query classes 2) To serve ad-hoc queries

Your landlord as friend: 1) Common friends of friends 2) Maybe he is on your phone book

```
rdd.reduce(lambda x, y : x + y)
```

FoFoF scenario. Because, huge joins are required.

Part IV

```
//Chair/Professor/Last_Name
//Course[@Enrollment > 500]/Title
//Department[Course/Prerequisites/Prereq = "CS106B"]/Title⇔
```

exam