***Introduction to Ensemble Learning and Bagging***

***Quiz:***

1. **Check null values**

**How many null values are present in the dataset?**

***There could be more than just one correct answer.***

MinTemp and MaxTemp are the only two variables without missing values

All columns are missing at least some values

WindGustSpeed and Humidity9am are all missing more than 38% of their values

2. **Transform RainTomorrow**

Choice the correct code to transform **RainTomorrow** to a numeric variable using map.

data['RainTomorrow'].map({'Yes':1,'No':4})

data['RainTomorrow'] = data['RainTomorrow'].map({'Yes':1,'No':0})

data['RainTomorrow'] = data['RainTomorrow'].map({1:'Yes',0:'No'})

3. **Split the dataset**

Select the correct way to split the dataset into 30% test and 70% train.

*There could be more than just one correct answer.*

X\_train, X\_val, y\_train, y\_val = train\_test\_split(X, Y, test\_size=0.30, random\_state=0)

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, Y, test\_size=0.70, random\_state=0)

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, Y, test\_size=0.20, random\_state=0)

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, Y, test\_size=0.30, random\_state=0)

4. **Evaluate the accuracy**

Evaluate the accuracy of the training and testing dataset

The accuracy of the testing is 0.75

The accuracy of the testing is 0.99

The accuracy of the testing is 0.85

The accuracy of the training is 0.99

5. **Decision boundaries**

After visualizing the decision boundaries. Answer the following question.

*Do you have to normalize the data?*

No

Yes