# Preprocessing data

## Question 1)

Which preprocessing technique is most appropriate to use when dealing with a dataset that has numerical features with varying units of measurement and ranges that need to be compared and analyzed in a machine learning model?

A) Normalization
B) Standardization
C) Imputation
D) Discretization

## Question 2)

Which of the following statements is true about the Encoding technique in scikit-learn preprocessing?

A) Encoding is used to handle missing values in a dataset.
B) Encoding is used to transform categorical features into numerical features.
C) Encoding is used to standardize the range of features in a dataset.
D) Encoding is used to discretize continuous features in a dataset.

## Question 3)

Which preprocessing technique would be the best choice to apply in the following example?

You have a dataset with categorical features that have missing values. You want to fill in the missing values to prepare the data for a machine learning model.

A) Normalization
B) Standardization
C) Imputation
D) Discretization

## Question 4)

Which of the following statements is true about the Scaling technique in scikit-learn?

A) Scaling is a way to rescale the range of features in a dataset, typically to a standard range of 0 to 1, making it easier to compare and analyze features with different units of measurement and ranges.

B) Min-max scaling scales the features to have a mean of zero and a standard deviation of one, whereas z-score scaling scales the features between 0 and 1.

C) Scaling is a technique to convert categorical data into numerical data by assigning unique integer values to each category.

D) Scaling is a technique to fill missing values in a dataset by assigning a value based on the mean or median of the feature.

E) Scaling is only useful for datasets with discrete features, and not for datasets with continuous features.

## Question 5)

Which of the following statements is true about the Imputation technique for handling missing values?

A) It replaces the missing values with the mean of the non-missing values in the column.
B) It replaces the missing values with the median of the non-missing values in the column.
C) It replaces the missing values with the mode of the non-missing values in the column.
D) It can only be applied to categorical features.
E) It replaces the missing values with the minimum value of the non-missing values in the column.

## Answers

1) B
2) B
3) C
4) A
5) A