Question 1
Which of the following methods should you use to replace a missing value of an attribute with continuous values? 1 / 1 point
Use the average of the other values in the column
Use an educated guess
Use the mean square error of the other data in the column
Use the difference between the minimum and maximum values of the other data in the column

Correct
Correct! The average is often a good choice to fill in a missing value for an attribute with continuous values.

Question 2
Which of the following helps you decide on bin values when pre-processing data? 1 / 1 point

Visualize the distribution using a histogram
Use the interquartile range
Divide the average by the standard deviation
Convert objects to ints

Correct
Correct! Creating a histogram of values can help you decide how to group your data.

Question 3
Which of the following data types should numbers with decimals be if you want to use them as input for training a statistical model?

666, 1.1, 232, 23.12 1 / 1 point

int
data frame
float
object

Correct
Correct! Statistical models mostly take numerical values as inputs, and since these values contain decimals, float is the best type to use.

Question 4
Which of the following is the primary purpose of simple feature scaling? 0 / 1 point

So all the variables have a similar influence on the models you build
To make comparing and analyzing values easier.
It brings data into a common standard of expression
To get rid of “not a number” or NaN values

Question 5
Which of the following is the primary purpose of the get_dummies() method? 1 / 1 point

Converts categorical values into numerical ones
To help you group your data into bins
Converts numerical values into categorical ones
Converts the data’s data type

Correct
Correct! It creates a separate column with names as the entries of the variable’s categorical values. It assigns numerical values to each column based on the value of the actual attribute.