Question 1
When looking at a CSV file, what character separates each value? 1 / 1 point

A tab
An apostrophe
A comma
An equal sign

Correct
Correct! A comma separates each value in a CSV file.

Question 2
What is the Python library Scikit-learn primarily designed for? 1 / 1 point

Exploratory data analysis
Statistical modeling, including regression and classification
Operations on matrices
Fast array processing

Correct
Correct! The Python library Scikit-learn is primarily designed for statistical modeling, including regression and classification.

Question 3
In order to read data using the Python Pandas package, what are the two most important factors? 1 / 1 point

Format and file path
File types and encoding scheme
File types and format
Encoding scheme and file path

Correct
Correct! In order to read data using the Python Pandas package, the two most important factors are format and file path.

Question 4
For a Pandas data frame, what does the attribute “dtypes” return? 1 / 1 point

It returns the data type of the object
It returns the data types of each column
It returns the last five rows of the data frame
It returns the first five rows of the data frame

Correct
Correct! For a Pandas data frame, the attribute “dtypes” returns the data types of each column.

Question 5
What is a header? 1 / 1 point

The name of the rows
The name of the columns
The first value in a column
The first value in a row

Correct
Correct! The header refers to the names of the columns.

Question 6
The Matplotlib library is mostly used for what? 1 / 1 point

Data analysis
Statistical modeling
Data visualization
Machine learning algorithms

Correct
Correct! The Matplotlib library is mostly used for data visualization.

Question 7
What code returns the last 10 rows of the data frame df? 1 / 1 point

df.last(10)
df.last()
df.tail()
df.tail(10)

Correct
Correct! The code df.tail(10) returns the last 10 rows of the data frame df.

Question 8
What function should you use to remove rows and columns with null or NaN values? 1 / 1 point

findna()
dropna()

replacena()
removena()

Correct
Correct! The dropna() method removes rows and columns with null or NaN values.

Question 9
Which type of plot is binning best suited to graph? 1 / 1 point

Histogram
Line plot
Box plot
Scatter plot

Correct
Correct! The Histogram plot is best suited for binning to graph.

10.
Question 10
Consider the column "length" in the data frame df. What does this line of code do?
df["length"] = (df["length"]-df["length"].mean())/df["length"].std() 1 / 1 point

It calculates the standard deviation of the values in the length column
It standardizes the values in the “length” column
It finds the number of standard deviations that each value is from the mean of the values in the “length” column
It finds the mean of the values in the length column and divides by the standard deviation

Correct
Correct! This line of code standardizes the values in the “length” column.

Question 11
Consider the following image. What is the name of the operation that transformed the column “fuel” into quantitative variables?

Rows where fuel=gas have gas=1 and diesel=0. Rows where fuel=diesel have gas=0 and diesel=1. 1 / 1 point

Data standardization
Data pre-processing

One-hot encoding
Data numeration

Correct
Correct! The operation shown in the diagram is referred to as one-hot encoding.

Question 12
What segment of code calculates the mean of the column ‘peak-rpm’?
1 / 1 point

df.mean(['peak-rpm'])
mean = df.['peak-rpm']
df['peak-rpm'].mean()
mean(df['peak-rpm'])

Correct
Correct! This segment of code calculates the mean of the column ‘peak-rpm’.

Question 13
What does a positive linear relationship between an input variable and an output variable imply? 1 / 1 point

That as the input increases, the output increases at about the same rate.
That as the input increases, the output increases at an ever-increasing rate.
The output does not adequately explain the input.
That as the input increases, the output decreases at about the same rate.

Correct
Correct! A positive linear relationship between an input variable and an output variable implies that as the input increases, the output increases at about the same rate.

14.
Question 14
What is the interquartile range of a data set? 1 / 1 point

The difference in the range of values in the uppermost quartile with the range of values in the lower-most quartile

The middle of the data
The range of the data, split into four equal-sized groups
The data between the upper and lower quartiles represents the interquartile range.

Correct
Correct! The data between the upper and lower quartiles represents the interquartile range.

15.
Question 15
If the predicted function is:

=b
0

x
3
+b
1

x
2
+b
2

x+b

3

The method is:

1 / 1 point

Multiple Linear Regression
Exponential Regression
Linear regression
Polynomial Regression

Correct
Correct! This method always has an exponent on one of the input values.

16.
Question 16
What is a model estimator? 1 / 1 point

The mean, mode, median, and standard deviation of a data set
The descriptive statistics of a data set
The estimate of the output value of a data set given an input value
A mathematical equation that can be used to predict values not in the data set

Correct
Correct! A model estimator is a mathematical equation that can be used to predict values not in the data set.

17.
Question 17
What does a residual plot help you determine? 0 / 1 point

It's a visual way to see if there is a pattern in your data set.
It's a visual way to see if your model needs more input variables.
It's a visual way to see if a linear plot is appropriate.
It's a visual way to see if your coefficients of determination are accurate.

18.
Question 18
What is true about noise in the data?

1 / 1 point

It is random and cannot be predicted.

If your testing data fits your function well, you will not see noise in your predicted values.
Your model accounts for it with a parameter.
If your training data fits your function well, you will not see noise in your predicted values.

Correct
Correct! Noise in the data is random and cannot be predicted.

19.
Question 19
What can the hyperparameter, alpha, help you decide? 1 / 1 point

The accuracy of your R2 value.
If your model needs to be a higher order or lower order function.
The lower the alpha value, the better the fit.
The bigger the alpha value, the better the fit.

Correct
Correct! Alpha values indicate overfitting or underfitting, thus, it helps you to determine the order of your model if you have several models that appear to be a good fit.

20.
Question 20
What is one of the arguments of the GridSearchCV() method? 1 / 1 point

The normalized input values from your data set.
A Python dictionary with the key-value pair of the hyperparameter and a list of its possible values.
A data frame containing different possible models and their R2 values.
A Python dictionary where the key is the column header of your data set, and the values are the different values in that column.

Correct
Correct! One of the arguments of the GridSearchCV() method is a Python dictionary with the key-value pair of the hyperparameter and a list of its possible values.