1.Question 1

What type of file saves data in a text-based tabular format?

XLSX

PDF

HTML

**CSV**

Correct

Correct! A CSV saves data in a text-based tabular format.

**1 / 1 point**

**2.** Question 2

Which Python libraries are primarily *Algorithmic Libraries*?

Matplotlib, Seaborn

Jupyter, Regression

Pandas, Numpy, SciPy

Scikit-learn, Statsmodels

**Correct**

Correct! These Python libraries are primarily *Algorithmic Libraries.*

**1 / 1 point**

**3.** Question 3

What is a file path?

Describes the format of the data

Indicates the way data is encoded

Describes the schema of the data

Describes where the data is stored

**Correct**

Correct! A file path describes where the data is stored.

**1 / 1 point**

**4.** Question 4

For a Pandas data frame, what does the attribute “**dtypes**” return?

It returns the last five rows of the data frame

It returns the first five rows of the data frame

It returns the data type of the object

It returns the data types of each column

**Correct**

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

**1 / 1 point**

**5.** Question 5

The Pandas library allows us to read what?

File types

Only headers

Only rows

Data sets into a data frame

**Correct**

Correct! The Pandas library allows us to read data sets into a data frame.

**1 / 1 point**

**6.** Question 6

The Matplotlib library is mostly used for what?

Data analysis

Statistical modeling

Data visualization

Machine learning algorithms

**Correct**

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

**1 / 1 point**

**7.** Question 7

What is the output of the following code segment of the data frame **df**?

**df.head(5)**

It returns the header of the data frame

It returns all of the rows of the data frame

It returns the first five rows of the data frame

It returns the last five rows of the data frame

**Correct**

Correct! The code **df.head(5)** returns the first five rows of data frame **df.**

**1 / 1 point**

**8.** Question 8

What is the **dropna()** method used for?

Replacing missing values

Dropping specified values

Identifying missing values

Dropping missing values

**Correct**

Correct! The **dropna()** method is used to drop missing values.

**1 / 1 point**

**9.** Question 9

Which type of plot is binning best suited to graph?

Box plot

Scatter plot

Histogram

Line plot

**Correct**

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

**1 / 1 point**

**10.** Question 10

Which of the following describes how to standardize a set of values?

Subtracts the mean from each value in the “length” column and divides them by the standard deviation.

Finds the mean of the values in the “length” column and divides them by the standard deviation.

Subtract the standard deviation from each value in the “length” column and divide that by the mean.

Finds the standard deviation of the values in the “length” column and divides it by the mean.

**Correct**

Correct! This describes how to standardize a set of values.

**1 / 1 point**

**11.** Question 11

What is it called when you subtract the mean from the values in a data set and divide by the standard deviation?

Data standardization

Binning

Min-max method

One-hot encoding

**Correct**

Correct! Data standardization is calculated by subtracting the mean from the values in a data set and divide by the standard deviation.

**1 / 1 point**

**12.** Question 12

What segment of code calculates the mean of the column ‘peak-rpm’?

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

**mean = df.['peak-rpm']**

**mean( df['peak-rpm'])**

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

**Correct**

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

**1 / 1 point**

**13.** Question 13

What does a positive linear relationship between an input variable and an output variable imply?

The output does not adequately explain the input.

That as the input increases, the output increases at an ever-increasing rate.

That as the input increases, the output increases at about the same rate.

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.

**1 / 1 point**

**14.** Question 14

If you have data for the price of 400 cars, how many price values will be in each quartile?

1600

400

50

100

**Correct**

Correct! Quartiles split the data into four equal groups.

**1 / 1 point**

**15.** Question 15

If the predicted function is:

y^=b0+b1x*y*^​=*b*0​+*b*1​*x*y, with, hat, on top, equals, b, start subscript, 0, end subscript, plus, b, start subscript, 1, end subscript, x

The method is:

Polynomial Regression

Multiple Linear Regression

Linear regression

Exponential Regression

**Correct**

Correct! A linear regression has 2 parameters and no exponent on the input variable.

**1 / 1 point**

**16.** Question 16

Which of the following statements is true about a model estimator?

Has multiple input variables.

Has slope and intercept parameters.

It relates one or more independent variables to an output variable.

Has an order greater than one.

**Correct**

Correct! A model estimator relates one or more independent variables to an output variable.

**1 / 1 point**

**17.** Question 17

What does a residual plot help you determine?

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 coefficients of determination are accurate.

It’s a visual way to see if a linear plot is appropriate.

It’s a visual way to see if your model needs more input variables.

**Correct**

Correct! A residual plot is a visual way to see if a linear plot is appropriate.

**1 / 1 point**

**18.** Question 18

What is true about noise in the data?

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.

It is random and cannot be predicted.

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.

**1 / 1 point**

**19.** Question 19

Say you have several differently ordered polynomial models. Which of the following statistics is used to control the magnitude of the polynomial coefficients?

Alpha

Coefficient of determination

Correlation coefficient

Mean-squared error

**Correct**

Correct! Ridge regression helps you determine if your model is overfit or underfit.

**1 / 1 point**

**20.** Question 20

What does the **GridSearchCV()** method do?

It gives you R2 values for different orders of polynomial models.

It iterates over hyperparameters using cross-validation.

It selects the appropriate hyperparameters for your model.

It’s another way to cross-validate your data set.

**Correct**

Correct! The **GridSearchCV()** iterates over hyperparameters using cross-validation.