1. What type of file saves data in a text-based tabular format? 1 point

- ○ HTML
- ◉ CSV
- ○ PDF
- ○ XLSX

2. What is the Python library *Scikit-learn* primarily designed for? 1 point

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

3. What tells us the way the data is encoded? 1 point

- ○ The data path
- ○ The file type
- ◉ Encoding scheme
- ○ The file path

4. What attribute or function returns the data types of each column of a data frame? 1 point

- ○ `datatypes`
- ○ `head()`
- ○ `tail()`
- ◉ `dtypes`

5. What is a header? 1 point

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

6. The Pandas library is mostly used for what? 1 point

- ◉ Data analysis
- ○ Data visualization
- ○ Machine learning algorithms
- ○ Statistical modeling

7. What is the output of the following code segment of the data frame **df**? 1 point

```
df.tail(10)
```

- (●) It returns the last 10 rows of the data frame
- ( ) It returns the header of the data frame
- ( ) It returns all of the rows of the data frame
- ( ) It returns the first 10 rows of the data frame

8. What task does the following code segment do to a data frame **df**? 1 point

```
mean = df["price"].mean() df["price"].replace(np.nan, mean)
```

- ( ) It drops rows that contain missing values
- (●) It replaces the missing values in the column "price" with the mean values of that column
- ( ) It replaces the data in the column "price" with normalized values
- ( ) It calculates the mean of the data in the column "price"

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

- (●) Histogram
- ( ) Scatter plot
- ( ) Line plot
- ( ) Box plot

10. Consider the column "length" in the data frame **df**. What does this line of code do? 1 point

```
df["length"] = (df["length"]-df["length"].mean())/df["length"].std()
```

- ○ It finds the mean of the values in the length column and divides by the standard deviation
- ○ It finds the number of standard deviations that each value is from the mean of the values in the "length" column
- ○ It calculates the standard deviation of the values in the length column
- ◉ It standardizes the values in the "length" column

11. What is the primary purpose of *one-hot encoding*? 1 point

- ○ To convert numeric variables into categorical ones
- ◉ To convert categorical variables into numeric ones
- ○ To convert numeric data types into object data types
- ○ To convert object data types into numeric data types

12. What task does the following line of code perform in the data frame **df**? 1 point

```
df['peak-rpm'].replace(np.nan, 5,inplace=True)
```

- ○ Replaces the values equal to 5 in the column `'peak-rpm'` with the value `'nan'`
- ○ Adds 5 to the values in the column `'peak-rpm'`
- ◉ Replaces the *not a number* values with 5 in the column `'peak-rpm'`
- ○ Renames the column `'peak-rpm'` to 5

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

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

14. If you have data for the price of 400 cars, how many price values will be in each quartile? 1 point

- ○ 50
- ◉ 100
- ○ 400
- ○ 1600

15. If the predicted function is: 1 point

$$\hat{y} = b_0 + b_1 x$$

The method is:

- ○ Multiple Linear Regression
- ○ Exponential Regression
- ○ Polynomial Regression
- ◉ Linear regression

16. Which of the following statements is true about a model estimator? 1 point

- ○ Has an order greater than one.
- ○ Has multiple input variables.
- ○ Has slope and intercept parameters.
- ◉ It relates one or more independent variables to an output variable.

17. How are residuals calculated? 1 point

- ○ $x - \hat{x}$
- ○ $\hat{y} - \bar{y}$
- ○ $b_0 + b_1 x$
- ◉ $y - \hat{y}$

18. What is true about noise in the data? 1 point

- ○ Your model accounts for it with a parameter.
- ○ If your testing data fits your function well, you will not see noise in your predicted values.
- ○ If your training data fits your function well, you will not see noise in your predicted values.
- ◉ It is random and cannot be predicted.

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

- ○ If your model needs to be a higher order or lower order function.
- ◉ The lower the alpha value, the better the fit.
- ○ The accuracy of your R^2 value.
- ○ The bigger the alpha value, the better the fit.

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

- ○ A Python dictionary where the key is the column header of your data set, and the values are the different values in that column.
- ◉ 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 R^2 values.
- ○ The normalized input values from your data set.