Your grade: 100%

Your latest: **100%** • Your highest: **100%** • To pass you need at least 70%. We keep your highest score.

Next item →

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

- [ ] An equal sign
- [ ] A tab
- [x] A comma
- [ ] An apostrophe

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

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

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

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

3. What is a file path? 1 / 1 point

- [ ] Indicates the way data is encoded
- [ ] Describes the schema of the data
- [x] Describes where the data is stored
- [ ] Describes the format of the data

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

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

- [x] `dtypes`
- [ ] `head()`
- [ ] `tail()`
- [ ] `datatypes`

Correct
Correct! The `dtypes` attribute returns the data types of each column.

5. In a data set, what term refers to the column name? 1 / 1 point

- [ ] Type
- [ ] Title
- [x] Header
- [ ] Row

Correct
Correct! In a data set, the term *header* refers to the column name.

6. What library is primarily used for data analysis? 1 / 1 point

- [x] Pandas
- [ ] Scikit-learn
- [ ] Matplotlib
- [ ] Sklearn

Correct
Correct! The Pandas library is primarily used for data analysis.

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

- [ ] `df.last()`
- [ ] `df.last(10)`
- [ ] `df.tail()`
- [x] `df.tail(10)`

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

8. What is the `dropna()` method used for? 1 / 1 point

- [x] Dropping missing values
- [ ] Replacing missing values
- [ ] Dropping specified values
- [ ] Identifying missing values

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

9. What is the primary purpose of binning values of a data set? 1 / 1 point

- [x] To gain a better understanding of the data distribution
- [ ] To convert numeric values to categorical ones
- [ ] To train a data set
- [ ] To convert categorical values to numeric ones

Correct
Correct! Binning is primarily used to gain a better understanding of the data distribution.

10. What is the primary purpose of standardizing a set of values? 1 / 1 point

- [x] It places different variables on the same scale, allowing you to compare them more easily.
- [ ] So you can see the spread of the data set and identify outliers.
- [ ] To find how well a data set fits a model.
- [ ] To see how many standard deviations each value is from the mean.

Correct
Correct! Standardizing values serves to place different variables on the same scale, allowing you to compare them more easily.

11. Since most statistical models cannot take objects or strings as inputs, what action needs to be performed? 1 / 1 point

- [ ] Convert numerical values into categorical variables
- [ ] Convert numeric data types into object data types
- [x] Convert categorical variables into numerical values
- [ ] Convert object data types into numeric data types

Correct
Correct! Most statistical models cannot take objects or strings as inputs so it is helpful to convert categorical variables into numerical values.

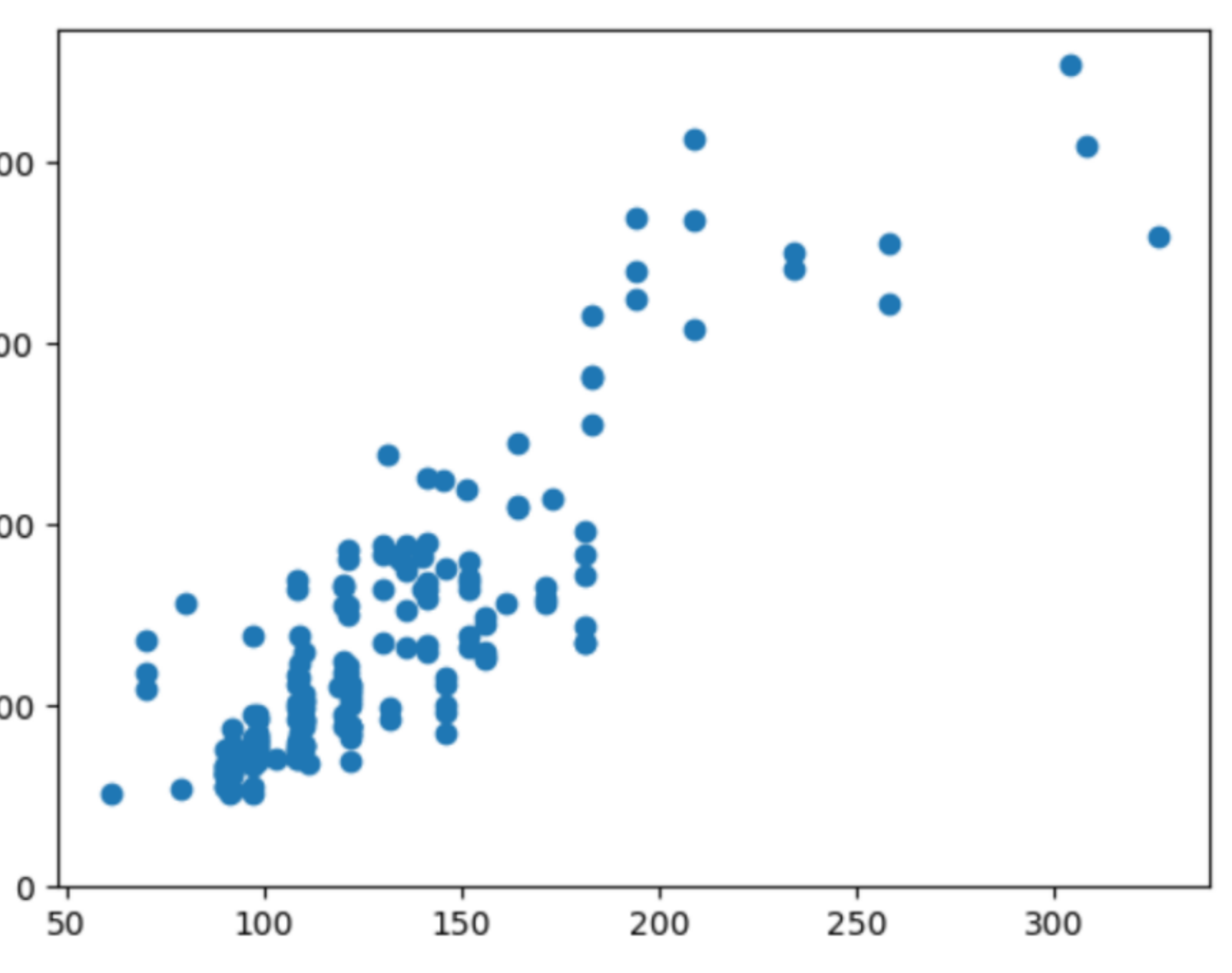
12. What task does the following line of code perform in the data frame **df**? 1 / 1 point

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

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

Correct
Correct! This segment of code replaces the *not a number* values with 5 in the column `'peak-rpm'`.

13. Consider the following scatter plot; what kind of relationship do the two variables have? 1 / 1 point



- [ ] No clear relationship
- [ ] Negative linear relationship
- [x] Positive linear relationship
- [ ] Exponential relationship

Correct
Correct! Since the price tends to increase per unit of increase in the engine size, this plot has a positive linear relationship.

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

- [ ] The middle of the data
- [ ] The difference in the range of values in the uppermost quartile with the range of values in the lower-most quartile
- [ ] The range of the data, split into four equal-sized groups
- [x] 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. If the predicted function is: 1 / 1 point

$$\hat{y} = b_0x^3 + b_1x^2 + b_2x + b_3$$

The method is:

- [ ] Multiple Linear Regression
- [ ] Exponential Regression
- [x] Polynomial Regression
- [ ] Linear regression

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

16. Say you are trying to predict the price of a car based on its gas mileage, and you find an equation in terms of and x to $\hat{y}$ predict these values. What is this equation called? 1 / 1 point

- [x] Model estimator
- [ ] Mean square error
- [ ] Multiple linear regression
- [ ] Coefficient of determination

Correct
Correct! A mathematical equation with an input variable and an output variable used to predict values is called a model estimator.

17. Why might you want to use a histogram in conjunction with your residuals? 1 / 1 point

- [ ] To see if there is curvature in your predicted values.
- [ ] To calculate the accuracy of your model.
- [x] To look at the distribution of your residuals in a multiple linear regression.
- [ ] To standardize your output values.

Correct
Correct! These plots are extremely useful for visualizing models with more than one independent variable or feature.

18. Say you have multiple polynomials that seem to fit your data. What type of graph can you use to help determine which order polynomial is the best fit for your data? 1 / 1 point

- [ ] The input values of your testing data on the x-axis and their MSE values on the y-axis.
- [ ] The order of the polynomials on the x-axis, and the MSE on the y-axis with your training data set.
- [ ] The input values of your training data on the x-axis and their MSE values on the y-axis.
- [x] The order of the polynomials on the x-axis, and the MSE on the y-axis with your testing data set.

Correct
Correct! A graph of MSE vs. order of the testing data set will help you determine which order polynomial is the best fit for your data.

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

- [x] If your model needs to be a higher order or lower order function.
- [ ] The bigger the alpha value, the better the fit.
- [ ] The accuracy of your R^2 value.
- [ ] The lower 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. What does the `GridSearchCV()` method do? 1 / 1 point

- [ ] It gives you R2 values for different orders of polynomial models.
- [ ] It selects the appropriate hyperparameters for your model.
- [x] It iterates over hyperparameters using cross-validation.
- [ ] It's another way to cross-validate your data set.

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