Congratulations! You passed!
TO PASS 80% or higher

Keep Learning

GRADE
100%

Graded Quiz: Test Your Project Understanding

LATEST SUBMISSION GRADE
100%

1. For a pandas data frame **df,** what function is used to return the first five rows? **1 / 1 point**

- (•) `df.head()`
- ( ) `df.info()`
- ( ) `df.tail()`

Correct
Correct! The pandas.DataFrame.head( ) method is used to return the first five rows of the data frame. This function returns the first *n* rows for the object based on position when the *n* argument is specified. Eg:

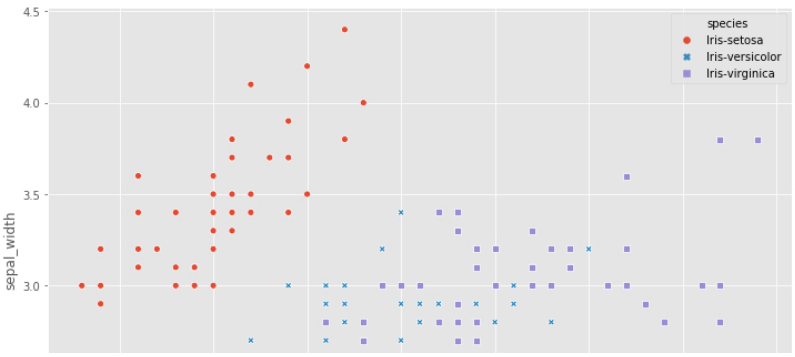
```
df.head(n=10)
```

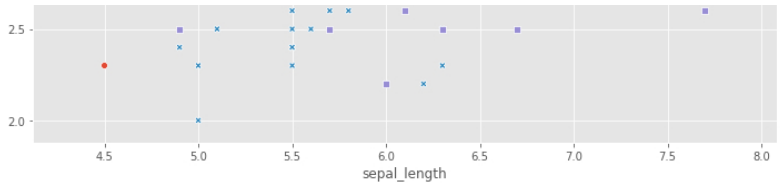
2. How would you drop rows in a dataframe **df** if all the values are missing? **1 / 1 point**

- ( ) `df.dropnull(how='all', inplace=True)`
- ( ) `df.dropna(how='all', inplace=False)`
- (•) `df.dropna(how='all', inplace=True)`

Correct
Correct! This removes rows that contain all missing values.

3. In the hands-on component of this course, you created the below figure. It is a scatter plot of the sepal width against the sepal length of three types of iris plant. What argument did you use to color the data points by the type of plant? **1 / 1 point**





- style
- x
- y
- hue

Correct

Correct! 'hue' is a grouping variable that will produce points with different colors.

```
sns.scatterplot(x = iris.sepal_length, y = iris.sepal_width,
                hue = iris.species, style = iris.species);
```

4. It is important to standardize the values in the data set before applying PCA. **1 / 1 point**

- True
- False

Correct

Correct! With great power to reduce dimensionality, comes great responsibility. One must take care to preprocess the input data appropriately. One must zero-out the mean from each feature (subtract the mean of each feature from the training set), and normalize the values if your features have differing units. PCA is also best used when the data is linear, because it is projecting it onto a linear subspace spanned by the eigenvectors. This is an important step in many machine learning algorithms, and especially so in the case of PCA. We want the PCA algorithm to give equal weight to each of the features while making the projection. If one or more features are in a different scale than the rest, those non-standardized features will dominate the eigenvalues and give you an incorrect result.

This is a direct consequence of how PCA works. It is going to project our data into directions that maximize the variance along the axes.

5. Say you have a standardized data set stored in a numpy array **X**. How would you use NumPy to calculate the covariance matrix Σ ? **1 / 1 point**

```
covariance_matrix = np.linalg.cov(X.T)
```

```
covariance_matrix = np.cov(X.T)
```

```
covariance_matrix = np.corr(X.T)
```

Correct

Correct!

6. Now that you have calculated the covariance matrix, how do you use NumPy to compute the eigenvectors and eigenvalues associated with the covariance matrix? **1 / 1 point**

```
eigenvalues, eigenvectors = np.linalg.eig(covariance_matrix)
```

```
eigenvalues, eigenvectors = np.eig(covariance_matrix)
```

Correct

Correct! To compute eigenvalues and eigenvectors, you need to the NumPy linear algebra functions.