[Add your name here] assignment #2 Iris Species

August 31, 2023

[ ]:
```
# add comment
import pandas as pd

from sklearn.datasets import load_iris
iris_dataset = load_iris()

# nothing will be printed to output but you must run this cell
```

[ ]:
```
# add comment
print("Keys of iris_dataset: \n{}".format(iris_dataset.keys()))
```

[ ]:
```
# comment
print(iris_dataset['DESCR'])
```

[ ]:
```
# comment
print("Target name: {}".format(iris_dataset['target_names']))
```

[ ]:
```
# comment explain features
print("Feature names: \n{}".format(iris_dataset['feature_names']))
```

[ ]:
```
# comment explain "type"
print("Type of data: {}".format(type(iris_dataset['data'])))
```

[ ]:
```
# comment explain "shape"
print("Shape of data: {}".format(iris_dataset['data'].shape))
```

[ ]:
```
# comment
print("First 5 columns of data: \n{}".format(iris_dataset['data'][:5]))
```

[ ]:
```
# create dataframe to store information about the features
iris_df = pd.DataFrame(iris_dataset.data)

# automatically prints out the first 5 rows of data
iris_df.head()
```

[ ]:
```
# add comment that explain why this was done
iris_df.columns = ['sepal_length', 'sepal_width', 'petal_length', 'petal_width']
```

```
iris_df.head()
```

[ ]:
```
# the target frame is only one column but when combined with both dataframes␣
 ↪show the
# features and which flower it belongs to.
target_df = pd.DataFrame(iris_dataset.target)
target_df = target_df.rename(columns = {0: 'target'})
target_df.head()
```

[ ]:
```
# combined the two dataframes to see the features and target (labels) together
df = pd.concat([iris_df,target_df], axis=1)
df
```

[ ]:
```
# numbers given by the iris_datset array: 0 setosa, 1 versicolor and 2 virginica
print("Target: \n{}".format(iris_dataset['target']))
```

[ ]:
```
# add comment
import matplotlib.pyplot as plt

# indices of features to plot
X = 0
y = 1

# formatter labels the color bar with the target names
formatter = plt.FuncFormatter(lambda i, *args: iris_dataset.
 ↪target_names[int(i)])

# nothing will be printed out
```

[ ]:
```
# Two feature scatter plot of the iris dataset using matplotlib

plt.figure(figsize=(5,4))
plt.scatter(iris_dataset.data[:, X], iris_dataset.data[:,y], c=iris_dataset.
 ↪target)
plt.colorbar(ticks=[0,1,2], format=formatter)
plt.xlabel(iris_dataset.feature_names[X])
plt.ylabel(iris_dataset.feature_names[y])


class_labels = iris_dataset.target_names
legend_aliases = [plt.Line2D([0,0],[0,0], color=['purple','green','yellow'][i],␣
 ↪marker='o', linestyle='', label=label)
    for i, label in enumerate(class_labels)]


plt.legend(handles=legend_aliases)
```

```
plt.tight_layout()
plt.show()
```

[ ]: