NIKHIL MALVI B 50 EXP 3

```
from sklearn.datasets import load_iris
import pandas as pd

# Load the built-in Iris dataset from scikit-learn
iris = load_iris()

# Convert to pandas DataFrame and add species names as a new column
df = pd.DataFrame(data=iris.data, columns=iris.feature_names)
df['Species'] = pd.Categorical.from_codes(iris.target, iris.target_names)

# Get the list of unique species in the dataset
species_list = df['Species'].unique().tolist()
print("\nList of Unique Species:\n", species_list)

# Generate descriptive statistics for each species individually
setosa_stats = df[df['Species'] == 'setosa'].describe()
versicolor_stats = df[df['Species'] == 'versicolor'].describe()
virginica_stats = df[df['Species'] == 'virginica'].describe()

# Print the statistics for each species
print("\nSetosa Statistics:\n", setosa_stats)
print("\nVersicolor Statistics:\n", versicolor_stats)
print("\nVirginica Statistics:\n", virginica_stats)

```

```
List of Unique Species:
 ['setosa', 'versicolor', 'virginica']

Setosa Statistics:
        sepal length (cm)  sepal width (cm)  petal length (cm)  \
count           50.00000         50.000000          50.000000
mean             5.00600          3.428000           1.462000
std              0.35249          0.379064           0.173664
min              4.30000          2.300000           1.000000
25%              4.80000          3.200000           1.400000
50%              5.00000          3.400000           1.500000
75%              5.20000          3.675000           1.575000
max              5.80000          4.400000           1.900000

       petal width (cm)
count         50.000000
mean           0.246000
std            0.105386
min            0.100000
25%            0.200000
50%            0.200000
75%            0.300000
max            0.600000

Versicolor Statistics:
        sepal length (cm)  sepal width (cm)  petal length (cm)  \
count           50.000000         50.000000          50.000000
mean             5.936000          2.770000           4.260000
std              0.516171          0.313798           0.469911
min              4.900000          2.000000           3.000000
25%              5.600000          2.525000           4.000000
50%              5.900000          2.800000           4.350000
75%              6.300000          3.000000           4.600000
max              7.000000          3.400000           5.100000

       petal width (cm)
count         50.000000
mean           1.326000
std            0.197753
min            1.000000
25%            1.200000
50%            1.300000
75%            1.500000
max            1.800000

Virginica Statistics:
        sepal length (cm)  sepal width (cm)  petal length (cm)  \
count           50.00000         50.000000          50.000000
mean             6.58800          2.974000           5.552000
std              0.63588          0.322497           0.551895
min              4.90000          2.200000           4.500000
```

```
25%             6.22500          2.800000          5.100000
50%             6.50000          3.000000          5.550000
75%             6.90000          3.175000          5.875000
max             7.90000          3.800000          6.900000

       petal width (cm)
```

```
from sklearn.datasets import load_iris
import pandas as pd

# Load the Iris dataset
iris = load_iris()

# Convert to pandas DataFrame for easy viewing
df = pd.DataFrame(data=iris.data, columns=iris.feature_names)
df['target'] = iris.target

# Display the headers and first 5 rows
print("Headers:", df.columns.tolist())
print("First 5 rows:")
print(df.head())

```

```
Headers: ['sepal length (cm)', 'sepal width (cm)', 'petal length (cm)', 'petal width (cm)', 'target']
First 5 rows:
   sepal length (cm)  sepal width (cm)  petal length (cm)  petal width (cm)  \
0                5.1               3.5                1.4               0.2
1                4.9               3.0                1.4               0.2
2                4.7               3.2                1.3               0.2
3                4.6               3.1                1.5               0.2
4                5.0               3.6                1.4               0.2

   target
0       0
1       0
2       0
3       0
4       0
```

```
from sklearn.datasets import load_iris
import pandas as pd

# Load the Iris dataset
iris = load_iris()

# Create DataFrame
df = pd.DataFrame(data=iris.data, columns=iris.feature_names)
df['Species'] = pd.Categorical.from_codes(iris.target, iris.target_names)

# Group by Species and calculate statistics
grouped_stats = df.groupby('Species').agg(['mean', 'median', 'min', 'max', 'std'])

# Display
print("Statistics Grouped by Species:\n", grouped_stats)

```

```
Statistics Grouped by Species:
             sepal length (cm)                          sepal width (cm)  \
                          mean median  min  max       std             mean
Species
setosa                  5.006    5.0  4.3  5.8  0.352490            3.428
versicolor              5.936    5.9  4.9  7.0  0.516171            2.770
virginica               6.588    6.5  4.9  7.9  0.635880            2.974

                                      petal length (cm)                  \
            median  min  max       std              mean median  min  max
Species
setosa         3.4  2.3  4.4  0.379064             1.462   1.50  1.0  1.9
versicolor     2.8  2.0  3.4  0.313798             4.260   4.35  3.0  5.1
virginica      3.0  2.2  3.8  0.322497             5.552   5.55  4.5  6.9

                     petal width (cm)
                 std             mean median  min  max       std
Species
setosa      0.173664            0.246    0.2  0.1  0.6  0.105386
versicolor  0.469911            1.326    1.3  1.0  1.8  0.197753
virginica   0.551895            2.026    2.0  1.4  2.5  0.274650
/tmp/ipykernel_31/1605368126.py:12: FutureWarning: The default of observed=False is deprecated and will be changed to True in a future v
  grouped_stats = df.groupby('Species').agg(['mean', 'median', 'min', 'max', 'std'])
```

```
from sklearn.datasets import load_iris
import pandas as pd

# Load the dataset
iris = load_iris()
df = pd.DataFrame(data=iris.data, columns=iris.feature_names)
df['Species'] = pd.Categorical.from_codes(iris.target, iris.target_names)

# List unique species
species_list = df['Species'].unique().tolist()
print("\nList of Unique Species:\n", species_list)

# Individual statistics
setosa_stats = df[df['Species'] == 'setosa'].describe()
versicolor_stats = df[df['Species'] == 'versicolor'].describe()
virginica_stats = df[df['Species'] == 'virginica'].describe()

# Display stats
print("\nSetosa Statistics:\n", setosa_stats)
print("\nVersicolor Statistics:\n", versicolor_stats)
print("\nVirginica Statistics:\n", virginica_stats)

```

```
List of Unique Species:
 ['setosa', 'versicolor', 'virginica']

Setosa Statistics:
        sepal length (cm)  sepal width (cm)  petal length (cm)  \
count           50.00000         50.000000          50.000000
mean             5.00600          3.428000           1.462000
std              0.35249          0.379064           0.173664
min              4.30000          2.300000           1.000000
25%              4.80000          3.200000           1.400000
50%              5.00000          3.400000           1.500000
75%              5.20000          3.675000           1.575000
max              5.80000          4.400000           1.900000

       petal width (cm)
count         50.000000
mean           0.246000
std            0.105386
min            0.100000
25%            0.200000
50%            0.200000
75%            0.300000
max            0.600000

Versicolor Statistics:
        sepal length (cm)  sepal width (cm)  petal length (cm)  \
count           50.000000         50.000000          50.000000
mean             5.936000          2.770000           4.260000
std              0.516171          0.313798           0.469911
min              4.900000          2.000000           3.000000
25%              5.600000          2.525000           4.000000
50%              5.900000          2.800000           4.350000
75%              6.300000          3.000000           4.600000
max              7.000000          3.400000           5.100000

       petal width (cm)
count         50.000000
mean           1.326000
std            0.197753
min            1.000000
25%            1.200000
50%            1.300000
75%            1.500000
max            1.800000

Virginica Statistics:
        sepal length (cm)  sepal width (cm)  petal length (cm)  \
count           50.00000         50.000000          50.000000
mean             6.58800          2.974000           5.552000
std              0.63588          0.322497           0.551895
min              4.90000          2.200000           4.500000
25%              6.22500          2.800000           5.100000
50%              6.50000          3.000000           5.550000
75%              6.90000          3.175000           5.875000
max              7.90000          3.800000           6.900000

       petal width (cm)
```