practical3

April 18, 2024

1 Perform the following operations on any open source dataset (e.g., data.csv)

[3]:
```
import pandas as pd
import statistics as stat
from sklearn import datasets
iris= datasets.load_iris()
```

[5]:
```
df = pd.read_csv('IRIS.csv')
```

2 2. Write a Python program to display some basic statistical details like percentile, mean, standard deviation etc. of the species of 'Iris-setosa', 'Iris-versicolor' and 'Iris-versicolor' of iris.csv dataset.

[13]:
```
df.head()
```

[13]:

	sepal_length	sepal_width	petal_length	petal_width	species
0	5.1	3.5	1.4	0.2	Iris-setosa
1	4.9	3.0	1.4	0.2	Iris-setosa
2	4.7	3.2	1.3	0.2	Iris-setosa
3	4.6	3.1	1.5	0.2	Iris-setosa
4	5.0	3.6	1.4	0.2	Iris-setosa

[14]:
```
df.describe()
```

[14]:

	sepal_length	sepal_width	petal_length	petal_width
count	150.000000	150.000000	150.000000	150.000000
mean	5.843333	3.054000	3.758667	1.198667
std	0.828066	0.433594	1.764420	0.763161
min	4.300000	2.000000	1.000000	0.100000
25%	5.100000	2.800000	1.600000	0.300000
50%	5.800000	3.000000	4.350000	1.300000
75%	6.400000	3.300000	5.100000	1.800000
max	7.900000	4.400000	6.900000	2.500000

[15]:
```
df.
↪groupby("species")[['sepal_width','sepal_length','petal_length','petal_width']].
↪mean()
```

[15]:

species	sepal_width	sepal_length	petal_length	petal_width
Iris-setosa	3.418	5.006	1.464	0.244
Iris-versicolor	2.770	5.936	4.260	1.326
Iris-virginica	2.974	6.588	5.552	2.026

[16]:
```
df.
↪groupby("species")[['sepal_width','sepal_length','petal_length','petal_width']].
↪median()
```

[16]:

species	sepal_width	sepal_length	petal_length	petal_width
Iris-setosa	3.4	5.0	1.50	0.2
Iris-versicolor	2.8	5.9	4.35	1.3
Iris-virginica	3.0	6.5	5.55	2.0

[17]:
```
df.
↪groupby("species")[['sepal_width','sepal_length','petal_length','petal_width']].
↪min()
```

[17]:

species	sepal_width	sepal_length	petal_length	petal_width
Iris-setosa	2.3	4.3	1.0	0.1
Iris-versicolor	2.0	4.9	3.0	1.0
Iris-virginica	2.2	4.9	4.5	1.4

[18]:
```
df.
↪groupby("species")[['sepal_width','sepal_length','petal_length','petal_width']].
↪max()
```

[18]:

species	sepal_width	sepal_length	petal_length	petal_width
Iris-setosa	4.4	5.8	1.9	0.6
Iris-versicolor	3.4	7.0	5.1	1.8
Iris-virginica	3.8	7.9	6.9	2.5

[19]:
```
df.
↪groupby("species")[['sepal_width','sepal_length','petal_length','petal_width']].
↪quantile()
```

[19]:

species	sepal_width	sepal_length	petal_length	petal_width
Iris-setosa	3.4	5.0	1.50	0.2

Iris-versicolor	2.8	5.9	4.35	1.3
Iris-virginica	3.0	6.5	5.55	2.0

3 1 Provide summary statistics (mean, median, minimum, maximum, standard deviation) for a dataset (age, income etc.) with numeric variables grouped by one of the qualitative (categorical) variable. For example, if your categorical variable is age groups and quantitative variable is income, then provide summary statistics of income grouped by the age groups. Create a list that contains a numeric value for each response to the categorical variable.

[ ]: