

Code for Q1 related age income qn

In []: *# full code in 1 cell*

```
import pandas as pd

df = pd.read_csv('data.csv')
df.head()

grouped_stats1=df.groupby('Gender')[['Age']].agg(['mean' , 'median' , 'min' , 'max' , 'std'])
grouped_stats2=df.groupby('Gender')[['Income']].agg(['mean' , 'median' , 'min' , 'max' , 'std'])

print(grouped_stats1)
print(grouped_stats2)

grouped_dict = {group: df[df['Gender'] == group]['Income'].tolist() for group in df['Gender'].unique()}
print(grouped_dict)
```

In []: *#code for Q2*

In []: *# full code in 1 cell*

```
import pandas as pd

df = pd.read_csv('iris.csv')

setosa_stats = df[df['species'] == 'Iris-setosa'].describe()
versicolor_stats = df[df['species'] == 'Iris-versicolor'].describe()
virginica_stats = df[df['species'] == 'Iris-virginica'].describe()

print("Iris-setosa Stats:\n", setosa_stats)
print("Iris-versicolor Stats:\n", versicolor_stats)
print("Iris-virginica Stats:\n", virginica_stats)
```

In [4]: `import pandas as pd`

In [6]: `df=pd.read_csv('data.csv')`

In [7]: `df.head()`

Out[7]:

	Name	Gender	Age	Income
0	Alice	Female	23	45000
1	Bob	Male	25	52000
2	Charlie	Male	30	61000
3	Diana	Female	28	58000
4	Ethan	Male	22	48000

In [14]: `grouped_stats1=df.groupby('Gender')[['Age']].agg(['mean' , 'median' , 'min' , 'max' , 'std'])`
`grouped_stats2=df.groupby('Gender')[['Income']].agg(['mean' , 'median' , 'min' , 'max' , 'std'])`
`print(grouped_stats1)`
`print(grouped_stats2)`

```

      Age
      mean median min max      std
Gender
Female 25.25   25.0  23  28  2.217356
Male   26.00   26.0  22  30  3.366502
      Income
      mean   median   min   max      std
Gender
Female 50000.0  48500.0  45000  58000  5715.476066
Male   53750.0  53000.0  48000  61000  5439.056291

```

In [19]: `grouped_dict={group: df[df['Gender']==group]['Income'].tolist() for group in df['Gender'].unique()}`
`print(grouped_dict)`

```
{'Female': [45000, 58000, 50000, 47000], 'Male': [52000, 61000, 48000, 54000]}
```

Code for second QN

In [20]: `import pandas as pd`

```
In [21]: df=pd.read_csv('iris (3).csv')
```

```
In [26]: setosa=df[df['species']=='Iris-setosa'].describe()
versicolor=df[df['species']=='Iris-versicolor'].describe()
virginica=df[df['species']=='Iris-virginica'].describe()
```

```
In [35]: print("Iris setosa stats:\n",setosa)
```

```
Iris setosa stats:
      sepal_length  sepal_width  petal_length  petal_width
count      4.000000      4.000000        4.000      4.000
mean       4.925000      3.325000        1.375      0.200
std        0.170783      0.275379        0.050      0.000
min        4.700000      3.000000        1.300      0.200
25%        4.850000      3.150000        1.375      0.200
50%        4.950000      3.350000        1.400      0.200
75%        5.025000      3.525000        1.400      0.200
max        5.100000      3.600000        1.400      0.200
```

```
In [37]: print("Iris versicolor stats:\n",versicolor)
```

```
Iris versicolor stats:
      sepal_length  sepal_width  petal_length  petal_width
count      3.000000      3.000000        3.000000      3.000000
mean       5.866667      2.433333        4.233333      1.166667
std        0.321455      0.321455        0.404145      0.152753
min        5.500000      2.200000        4.000000      1.000000
25%        5.750000      2.250000        4.000000      1.100000
50%        6.000000      2.300000        4.000000      1.200000
75%        6.050000      2.550000        4.350000      1.250000
max        6.100000      2.800000        4.700000      1.300000
```

```
In [38]: print("Iris virginica stats:\n",virginica)
```

Iris virginica stats:

	sepal_length	sepal_width	petal_length	petal_width
count	3.000000	3.000000	3.000000	3.000000
mean	6.133333	3.233333	5.500000	2.200000
std	0.208167	0.208167	0.458258	0.360555
min	5.900000	3.000000	5.100000	1.800000
25%	6.050000	3.150000	5.250000	2.050000
50%	6.200000	3.300000	5.400000	2.300000
75%	6.250000	3.350000	5.700000	2.400000
max	6.300000	3.400000	6.000000	2.500000

In []: