ASSIGNMENT - 3

In [18]:
```
import pandas as pd
import numpy as np
```

In [2]:
```
df = pd.read_csv('iris.csv')
```

In [3]:
```
df
```

Out[3]:

	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
...	...	...	...	...	...
145	6.7	3.0	5.2	2.3	Iris-virginica
146	6.3	2.5	5.0	1.9	Iris-virginica
147	6.5	3.0	5.2	2.0	Iris-virginica
148	6.2	3.4	5.4	2.3	Iris-virginica
149	5.9	3.0	5.1	1.8	Iris-virginica

150 rows × 5 columns

In [5]:
```
df.head()
```

Out[5]:

	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

In [6]:
```
df.tail()
```

Out[6]:

	sepal_length	sepal_width	petal_length	petal_width	species
145	6.7	3.0	5.2	2.3	Iris-virginica
146	6.3	2.5	5.0	1.9	Iris-virginica
147	6.5	3.0	5.2	2.0	Iris-virginica
148	6.2	3.4	5.4	2.3	Iris-virginica
149	5.9	3.0	5.1	1.8	Iris-virginica

In [13]:
```
#Group data by a categorical variale and calculate summary statistics
summary_stats = df.groupby('species')["petal_length"].describe()
```

In [14]:
```
summary_stats
```

Out[14]:

species	count	mean	std	min	25%	50%	75%	max
Iris-setosa	50.0	1.464	0.173511	1.0	1.4	1.50	1.575	1.9
Iris-versicolor	50.0	4.260	0.469911	3.0	4.0	4.35	4.600	5.1
Iris-virginica	50.0	5.552	0.551895	4.5	5.1	5.55	5.875	6.9

In [15]:
```
#Convert grouped statistics to a list
numeric_summary = df.groupby('species')["petal_length"].apply(list)
```

In [16]:
```
numeric_summary
```

Out[16]:
```
species
Iris-setosa        [1.4, 1.4, 1.3, 1.5, 1.4, 1.7, 1.4, 1.5, 1.4, ...
Iris-versicolor    [4.7, 4.5, 4.9, 4.0, 4.6, 4.5, 4.7, 3.3, 4.6, ...
Iris-virginica     [6.0, 5.1, 5.9, 5.6, 5.8, 6.6, 4.5, 6.3, 5.8, ...
Name: petal_length, dtype: object
```

In [20]:
```
# Calculate 25th, 50th, and 75th percentiles
percentiles = df["sepal_length"].quantile([0.25, 0.5, 0.75])
```

In [21]:
```
percentiles
```

Out[21]:
```
0.25    5.1
0.50    5.8
0.75    6.4
Name: sepal_length, dtype: float64
```

In [23]:
```
iris = pd.read_csv('iris.csv')
```

In [24]:
```
iris.head(5)
```

Out[24]:

	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

In [26]:
```
# Filtering species
species_groups = iris.groupby('species')
```

In [27]:
```
# Calculate statistics for each species
for species, group in species_groups:
 print(f"Statistics for {species}:")
```

```
Statistics for Iris-setosa:
Statistics for Iris-versicolor:
Statistics for Iris-virginica:
```

In [28]:
```
group.describe()
```

Out[28]:

	sepal_length	sepal_width	petal_length	petal_width
count	50.00000	50.000000	50.000000	50.00000
mean	6.58800	2.974000	5.552000	2.02600
std	0.63588	0.322497	0.551895	0.27465
min	4.90000	2.200000	4.500000	1.40000
25%	6.22500	2.800000	5.100000	1.80000
50%	6.50000	3.000000	5.550000	2.00000
75%	6.90000	3.175000	5.875000	2.30000
max	7.90000	3.800000	6.900000	2.50000

In [ ]:
```

```