In [1]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

In [4]:
```
sales = pd.read_csv('C:/Users/prajw/Desktop/Indexs/DSBDA print/DSBDA3/data.csv')
```

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

Out[5]:

	Name	AgeGroup	Income
0	John	20-30	30000
1	Alice	30-40	45000
2	Mike	20-30	28000
3	Sarah	40-50	52000
4	Rahul	30-40	46000

In [7]:
```
sales.tail()   #Show me the last 5 rows from the table named sales
```

Out[7]:

	Name	AgeGroup	Income
5	Priya	40-50	51000
6	Ravi	20-30	31000
7	Sneha	30-40	47000
8	Amit	20-30	29000
9	Neha	40-50	50000

In [9]:
```
sales.shape  #Show me number of rows and columm
```

Out[9]:
```
(10, 3)
```

In [11]:
```
sales.isnull().sum()  #It tells you how many missing values (nulls) are in each column of your sales table.
```

Out[11]:
```
Name        0
AgeGroup    0
Income      0
dtype: int64
```

In [13]:
```
sales.count()   #tells you how many non-missing (non-null) values are present in each column
```

Out[13]:
```
Name        10
AgeGroup    10
Income      10
dtype: int64
```

In [14]:
```
sales.sum()
```

Out[14]:
```
Name            JohnAliceMikeSarahRahulPriyaRaviSnehaAmitNeha
AgeGroup    20-3030-4020-3040-5030-4040-5020-3030-4020-304...
Income                                                 409000
dtype: object
```

In [17]:
```
print(sales.columns)  #This will show you the actual column names
```

```
Index(['Name', 'AgeGroup', 'Income'], dtype='object')
```

In [20]:
```
grouped = sales.groupby('AgeGroup')['Income']
```

In [23]:
```
summary_stats = grouped.agg(['mean', 'median', 'min', 'max', 'std'])
print("Summary Statistics grouped by AgeGroup:")
print(summary_stats)                                                    # Calculate summary statistics
```

```
Summary Statistics grouped by AgeGroup:
             mean   median    min    max          std
AgeGroup
20-30     29500.0  29500.0  28000  31000  1290.994449
30-40     46000.0  46000.0  45000  47000  1000.000000
40-50     51000.0  51000.0  50000  52000  1000.000000
```

In [22]:
```
income_groups = grouped.apply(list).to_dict()
print("\nIncome values grouped by AgeGroup:")
for group, values in income_groups.items():
    print(f"{group}: {values}")
```

```
Income values grouped by AgeGroup:
20-30: [30000, 28000, 31000, 29000]
30-40: [45000, 46000, 47000]
40-50: [52000, 51000, 50000]
```

In [ ]:
```
# Iris dataset
```

In [1]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
```

In [2]:
```
iris = pd.read_csv('C:/Users/prajw/Desktop/Indexs/DSBDA print/DSBDA3/iris.csv')
```

In [3]:
```
iris.head()
```

Out[3]:

	sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)	species
0	5.1	3.5	1.4	0.2	setosa
1	4.9	3.0	1.4	0.2	setosa
2	4.7	3.2	1.3	0.2	setosa
3	4.6	3.1	1.5	0.2	setosa
4	5.0	3.6	1.4	0.2	setosa

In [4]:
```
iris.tail()
```

Out[4]:

	sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)	species
145	6.7	3.0	5.2	2.3	virginica
146	6.3	2.5	5.0	1.9	virginica
147	6.5	3.0	5.2	2.0	virginica
148	6.2	3.4	5.4	2.3	virginica
149	5.9	3.0	5.1	1.8	virginica

In [5]:
```
iris.shape
```

Out[5]:
```
(150, 5)
```

In [6]:
```
print(iris.columns)
```

```
Index(['sepal length (cm)', 'sepal width (cm)', 'petal length (cm)',
       'petal width (cm)', 'species'],
      dtype='object')
```

In [10]:
```
iris.describe()
```

Out[10]:

	sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)
count	150.000000	150.000000	150.000000	150.000000
mean	5.843333	3.057333	3.758000	1.199333
std	0.828066	0.435866	1.765298	0.762238
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

In [14]:
```
print(iris['species'].unique())
```

```
['setosa' 'versicolor' 'virginica']
```

In [15]:
```
# Check what species are available
print("Available species in dataset:", iris['species'].unique())
```

```
Available species in dataset: ['setosa' 'versicolor' 'virginica']
```

In [16]:
```
# Now use the correct species names (based on the above output)
species_list = iris['species'].unique()

for species in species_list:
    print(f"\nStatistical summary for {species}:")
    subset = iris[iris['species'] == species]
    print(subset.describe(percentiles=[.25, .5, .75]))
```

```
Statistical summary for setosa:
       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

Statistical summary for versicolor:
       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

Statistical summary for virginica:
       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)
count          50.00000
mean            2.02600
std             0.27465
min             1.40000
25%             1.80000
50%             2.00000
75%             2.30000
max             2.50000
```

In [17]:
```
iris['species'] = iris['species'].str.strip()
```

In [18]:
```
iris[iris['species'] == 'Iris-setosa']
```

Out[18]:

	sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)	species

In [ ]: