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.

[1]:
```
import pandas as pd
import numpy as np
```

[3]:
```
df = pd.read_csv(r"C:\\Users\\UNKNOWN_CODER\\DSDBA\\Assign3\\data-income-age-group.csv")
```

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.

[1]:
```
import pandas as pd
import numpy as np
```

[3]:
```
df = pd.read_csv(r"C:\\Users\\UNKNOWN_CODER\\DSDBA\\Assign3\\data-income-age-group.csv")
```

[7]:
```
df
```

[7]:

	Name	Age	Income	Age Group
0	Rob	27	70000	20-40
1	Michael	29	90000	20-40
2	Mohan	29	61000	20-40
3	Ismail	28	60000	20-40
4	Kory	42	150000	40-60
5	Gautam	39	155000	20-40
6	David	41	160000	40-60
7	Andrea	38	162000	20-40
8	Brad	36	156000	20-40
9	Angelina	35	130000	20-40
10	Donald	37	137000	20-40
11	Tom	26	45000	20-40
12	Arnold	27	48000	20-40
13	Jared	28	51000	20-40
14	Stark	29	49500	20-40
15	Ranbir	32	53000	20-40
16	Dipika	40	65000	40-60
17	Priyanka	41	63000	40-60
18	Nick	43	64000	40-60
19	Alia	39	80000	20-40
20	Sid	41	82000	40-60
21	Abdul	39	58000	20-40

[8]:
```
print("Statistical Summary: \n",)
df.describe()
```

```
Statistical Summary:
```

[8]:

	Age	Income
count	22.000000	22.000000
mean	34.818182	90431.818182
std	5.901060	43505.964412
min	26.000000	45000.000000
25%	29.000000	58500.000000
50%	36.500000	67500.000000
75%	39.750000	135250.000000
max	43.000000	162000.000000

[9]:
```
# Define age groups
bins = [0, 20, 40, 60, 100]  # Age groups: <20, 20-40, 40-60, >60
labels = ['<20', '20-40', '40-60', '>60']
df['Age Group'] = pd.cut(df['Age'], bins=bins, labels=labels, right=False)
grouped = df.groupby('Age')['Income'].describe()
grouped
```

[9]:

Age	count	mean	std	min	25%	50%	75%	max
26	1.0	45000.000000	NaN	45000.0	45000.0	45000.0	45000.0	45000.0
27	2.0	59000.000000	15556.349186	48000.0	53500.0	59000.0	64500.0	70000.0
28	2.0	55500.000000	6363.961031	51000.0	53250.0	55500.0	57750.0	60000.0
29	3.0	66833.333333	20870.633276	49500.0	55250.0	61000.0	75500.0	90000.0
32	1.0	53000.000000	NaN	53000.0	53000.0	53000.0	53000.0	53000.0
35	1.0	130000.000000	NaN	130000.0	130000.0	130000.0	130000.0	130000.0
36	1.0	156000.000000	NaN	156000.0	156000.0	156000.0	156000.0	156000.0
37	1.0	137000.000000	NaN	137000.0	137000.0	137000.0	137000.0	137000.0
38	1.0	162000.000000	NaN	162000.0	162000.0	162000.0	162000.0	162000.0
39	3.0	97666.666667	50856.005873	58000.0	69000.0	80000.0	117500.0	155000.0
40	1.0	65000.000000	NaN	65000.0	65000.0	65000.0	65000.0	65000.0
41	3.0	101666.666667	51403.631519	63000.0	72500.0	82000.0	121000.0	160000.0
42	1.0	150000.000000	NaN	150000.0	150000.0	150000.0	150000.0	150000.0
43	1.0	64000.000000	NaN	64000.0	64000.0	64000.0	64000.0	64000.0

Type Markdown and LaTeX: α^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

[12]:
```
df = pd.read_csv(r"C:\\Users\\UNKNOWN_CODER\\DSDBA\\Assign3\\iris.csv")
```

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

[13]:

	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	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]:
```
grouped_species = df.groupby('Species')
```

[15]:
```
print("\nBasic Statistical Details of each Species:")
for species, group in grouped_species:
    print(f"\nStatistics for {species}:")
    numeric_group = group.select_dtypes(include=[np.number])
    # Calculate and display basic statistics: mean, standard deviation, min, and max
    print("Mean:")
    print(numeric_group.mean())

    print("\nStandard Deviation:")
    print(numeric_group.std())

    print("\nMinimum Values:")
    print(numeric_group.min())

    print("\nMaximum Values:")
    print(numeric_group.max())

    print("\nPercentiles:")
    percentiles = np.percentile(numeric_group, [25, 50, 75], axis=0)
    print(f"25th Percentile: {percentiles[0]}")
    print(f"50th Percentile (Median): {percentiles[1]}")
    print(f"75th Percentile: {percentiles[2]}")
    print("="*40)
```

```
Basic Statistical Details of each Species:

Statistics for Iris-setosa:
Mean:
SepalLengthCm    5.006
SepalWidthCm     3.418
PetalLengthCm    1.464
PetalWidthCm     0.244
dtype: float64

Standard Deviation:
SepalLengthCm    0.352490
SepalWidthCm     0.381024
PetalLengthCm    0.173511
PetalWidthCm     0.107210
dtype: float64

Minimum Values:
SepalLengthCm    4.3
SepalWidthCm     2.3
PetalLengthCm    1.0
PetalWidthCm     0.1
dtype: float64

Maximum Values:
SepalLengthCm    5.8
SepalWidthCm     4.4
PetalLengthCm    1.9
PetalWidthCm     0.6
dtype: float64

Percentiles:
25th Percentile: [4.8   3.125 1.4   0.2  ]
50th Percentile (Median): [5.  3.4 1.5 0.2]
75th Percentile: [5.2   3.675 1.575 0.3  ]
========================================

Statistics for Iris-versicolor:
Mean:
SepalLengthCm    5.936
SepalWidthCm     2.770
PetalLengthCm    4.260
PetalWidthCm     1.326
dtype: float64

Standard Deviation:
SepalLengthCm    0.516171
SepalWidthCm     0.313798
PetalLengthCm    0.469911
PetalWidthCm     0.197753
dtype: float64

Minimum Values:
SepalLengthCm    4.9
SepalWidthCm     2.0
PetalLengthCm    3.0
PetalWidthCm     1.0
dtype: float64
```

```
Maximum Values:
SepalLengthCm    7.0
SepalWidthCm     3.4
PetalLengthCm    5.1
PetalWidthCm     1.8
dtype: float64

Percentiles:
25th Percentile: [5.6   2.525 4.    1.2  ]
50th Percentile (Median): [5.9  2.8  4.35 1.3 ]
75th Percentile: [6.3 3.  4.6 1.5]
=======================================

Statistics for Iris-virginica:
Mean:
SepalLengthCm    6.588
SepalWidthCm     2.974
PetalLengthCm    5.552
PetalWidthCm     2.026
dtype: float64

Standard Deviation:
SepalLengthCm    0.635880
SepalWidthCm     0.322497
PetalLengthCm    0.551895
PetalWidthCm     0.274650
dtype: float64

Minimum Values:
SepalLengthCm    4.9
SepalWidthCm     2.2
PetalLengthCm    4.5
PetalWidthCm     1.4
dtype: float64

Maximum Values:
SepalLengthCm    7.9
SepalWidthCm     3.8
PetalLengthCm    6.9
PetalWidthCm     2.5
dtype: float64

Percentiles:
25th Percentile: [6.225 2.8   5.1   1.8  ]
50th Percentile (Median): [6.5  3.   5.55 2.  ]
75th Percentile: [6.9   3.175 5.875 2.3  ]
=======================================
```