Name: Pranav Mehendale

Roll No.: TCOD34

Batch: T11

Assignment 3 Group A

Descriptive Statistics - Measures of Central Tendency and variability

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

Provide the codes with outputs and explain everything that you do in this step.

```
import pandas as pd
import numpy as np
from sklearn.datasets import load_iris

iris = load_iris()
iris.feature_names
```

```
['sepal length (cm)',
 'sepal width (cm)',
 'petal length (cm)',
 'petal width (cm)']
```

```
df = pd.DataFrame(iris['data'])
df.head(3)
```

```
     0    1    2    3
0  5.1  3.5  1.4  0.2
1  4.9  3.0  1.4  0.2
2  4.7  3.2  1.3  0.2
```

```
df[4] = iris['target']
```

```
df.head(3)
```

```
     0    1    2    3  4
0  5.1  3.5  1.4  0.2  0
1  4.9  3.0  1.4  0.2  0
2  4.7  3.2  1.3  0.2  0
```

```
df.rename(columns={0:'SepalLengthCm', 1:'SepalWidthCm',
2:'PetalLengthCm', 3:'PetalWidthCm', 4:'Species'}, inplace = 'True')
df.head(3)
```

```
   SepalLengthCm  SepalWidthCm  PetalLengthCm  PetalWidthCm  Species
0            5.1           3.5            1.4           0.2        0
1            4.9           3.0            1.4           0.2        0
2            4.7           3.2            1.3           0.2        0
```

```
df['Species'].value_counts()
```

```
Species
0    50
1    50
2    50
Name: count, dtype: int64
```

```
df.describe()
```

```
       SepalLengthCm  SepalWidthCm  PetalLengthCm  PetalWidthCm
Species
count     150.000000    150.000000     150.000000    150.000000
150.000000
mean        5.843333      3.057333       3.758000      1.199333
1.000000
std         0.828066      0.435866       1.765298      0.762238
0.819232
min         4.300000      2.000000       1.000000      0.100000
0.000000
25%         5.100000      2.800000       1.600000      0.300000
0.000000
50%         5.800000      3.000000       4.350000      1.300000
1.000000
75%         6.400000      3.300000       5.100000      1.800000
2.000000
max         7.900000      4.400000       6.900000      2.500000
2.000000
```

Mean, Median & Mode

```
df.mean()
```

```
SepalLengthCm    5.843333
SepalWidthCm     3.057333
```

```
PetalLengthCm    3.758000
PetalWidthCm     1.199333
Species          1.000000
dtype: float64
```

```
df.median()
```

```
SepalLengthCm    5.80
SepalWidthCm     3.00
PetalLengthCm    4.35
PetalWidthCm     1.30
Species          1.00
dtype: float64
```

```
df.mode()
```

	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
0	5.0	3.0	1.4	0.2	0
1	NaN	NaN	1.5	NaN	1
2	NaN	NaN	NaN	NaN	2

Summary statistics grouped by Species

```
df.groupby(['Species']).mean()
```

Species	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm
0	5.006	3.428	1.462	0.246
1	5.936	2.770	4.260	1.326
2	6.588	2.974	5.552	2.026

```
df.groupby(['Species']).median()
```

Species	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm
0	5.0	3.4	1.50	0.2
1	5.9	2.8	4.35	1.3
2	6.5	3.0	5.55	2.0

```
df.groupby(['Species']).std()
```

Species	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm
0	0.352490	0.379064	0.173664	0.105386
1	0.516171	0.313798	0.469911	0.197753
2	0.635880	0.322497	0.551895	0.274650