

Name : Shivam Indrabhan Borse

Roll No : 21119

Subject: Software Laboratory III (DATA SCIENCE)

Assignment No : 03

Problem statement:

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) variables. 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.

CODE :

```
----- Iris Flowers Classification -----
+ Code + Markdown

import numpy as np
import pandas as pd
import matplotlib.pyplot as plt

iris = pd.read_csv("Iris.csv")

print(iris.head(10))
```

| | | Id | SepallengthCm | SepalwidthCm | PetallengthCm | PetalwidthCm | Species |
|---|----|----|---------------|--------------|---------------|--------------|-------------|
| 0 | 1 | | 5.1 | 3.5 | 1.4 | 0.2 | Iris-setosa |
| 1 | 2 | | 4.9 | 3.0 | 1.4 | 0.2 | Iris-setosa |
| 2 | 3 | | 4.7 | 3.2 | 1.3 | 0.2 | Iris-setosa |
| 3 | 4 | | 4.6 | 3.1 | 1.5 | 0.2 | Iris-setosa |
| 4 | 5 | | 5.0 | 3.6 | 1.4 | 0.2 | Iris-setosa |
| 5 | 6 | | 5.4 | 3.9 | 1.7 | 0.4 | Iris-setosa |
| 6 | 7 | | 4.6 | 3.4 | 1.4 | 0.3 | Iris-setosa |
| 7 | 8 | | 5.0 | 3.4 | 1.5 | 0.2 | Iris-setosa |
| 8 | 9 | | 4.4 | 2.9 | 1.4 | 0.2 | Iris-setosa |
| 9 | 10 | | 4.9 | 3.1 | 1.5 | 0.1 | Iris-setosa |

```
print(iris.describe())
```

[7] ✓ 0.0s

```
...
      count      Id  SepalLengthCm  SepalWidthCm  PetalLengthCm  PetalWidthCm
mean    75.500000   5.843333      3.054000      3.758667      1.198667
std     43.445368   0.828066      0.433594      1.764420      0.763161
min      1.000000   4.300000      2.000000      1.000000      0.100000
25%     38.250000   5.100000      2.800000      1.600000      0.300000
50%     75.500000   5.800000      3.000000      4.350000      1.300000
75%    112.750000   6.400000      3.300000      5.100000      1.800000
max    150.000000   7.900000      4.400000      6.900000      2.500000
```

```
print("Target Labels", iris["Species"].unique())
```

[8] ✓ 0.0s

```
... Target Labels ['Iris-setosa' 'Iris-versicolor' 'Iris-virginica']
```

```
import plotly.express as px
fig = px.scatter(iris, x="SepalLengthCm", y="SepalWidthCm", color="Species")
fig.show()
```

