

Aim: Write a python program to open Comma Separated Value (CSV) and perform given statistical operations.

Theory:

a) CSV: A CSV(Comma Separated Values) file is a delimited text file that uses a comma to separate values. Each line of the file is a data record.

b) Categorical variables: Categorical variables represent groupings of some kind. They are sometimes recorded as numbers, but the numbers represent categories rather than actual amounts of things.

There are three types of categorical variables: binary, nominal, and ordinal variables.

1. Binary variables: They represent data either YES or NO like Heads/tails in a coin flip.

2. Nominal variables: Data has group or order between them like Species names.

3. Ordinal variables: Data are ranked in specific order like Rating scale responses in a survey.

c) Numerical variables: The data variable that takes on any value within a finite or infinite interval. They are also called Continuous variables because it exhibits the features of Continuous Data.

Operations:

1. Identification of categorical and numerical variables

```
1 import pandas as pd
2 dataSet = pd.read_csv("./Data Set.csv")
3 print(dataSet.dtypes)
4
```

```
C:\WINDOWS\system32\cmd.exe
dataSet[set(column)].tolist()
Id int64
SepalLengthCm float64
SepalWidthCm float64
PetalLengthCm float64
PetalWidthCm float64
Species object
dtype: object
dataSet, "SepalLengthCm")
dataSet, "SepalWidthCm")
dataSet, "PetalLengthCm")
```

2. Mean

```
SepalLen Mean: 5.843333333333335 sepalWid mean: 3.0540000000000007
petalLen mean: 3.7586666666666693 petalWid mean: 1.1986666666666672
(5.843333333333335, 3.0540000000000007, 3.7586666666666693, 1.1986666666666672)
```