TC1 LAB 01

Name - Sarvesh Karanjkar

PRN - 20210812002

In [ ]:

```
'''
1. What is CSV file?
-->
2) What are categorical variables and types?
--> 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.

3) What are numerical variables and types?
-->
A numeric variable (also called quantitative variable) is a quantifiable
characteristic whose values are numbers (except numbers which are codes standing up for
categories). Numeric variables may be either continuous or discrete. There are
two types of numerical variables, namely; discrete and continuous.

'''
```

Creating DataFrame

In [31]:

```
import pandas as pd
import numpy as np

columns = ["Id", "SepalLengthCm","PetalLengthCm","PetalWidthCm","Species"]

df  = pd.read_csv("Data Set - copy.csv")
df.describe()
```

Out[31]:

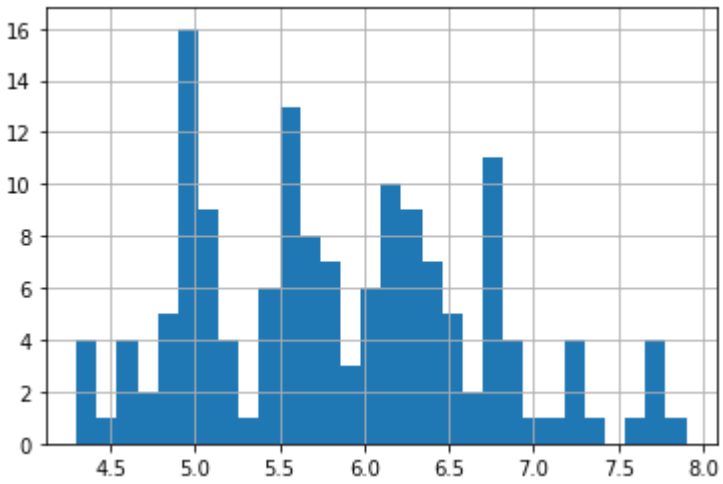
	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm
count	150.000000	150.000000	150.000000	150.000000	150.000000
mean	75.500000	5.843333	3.054000	3.758667	1.198667
std	43.445368	0.828066	0.433594	1.764420	0.763161
min	1.000000	4.300000	2.000000	1.000000	0.100000
25%	38.250000	5.100000	2.800000	1.600000	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

In [68]:

```
df['SepalLengthCm'].hist(bins = 30)
```

Out[68]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x1cb1322a8b0>
```

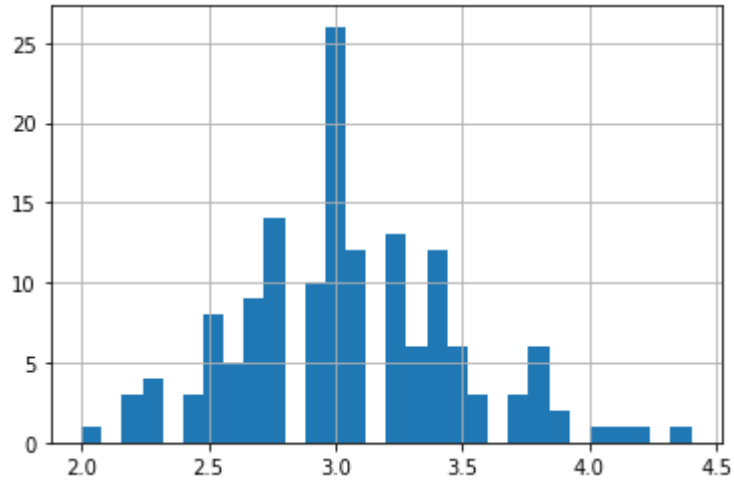


In [66]:

```
df['SepalWidthCm'].hist(bins=30)
```

Out[66]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x1cb131bf820>
```

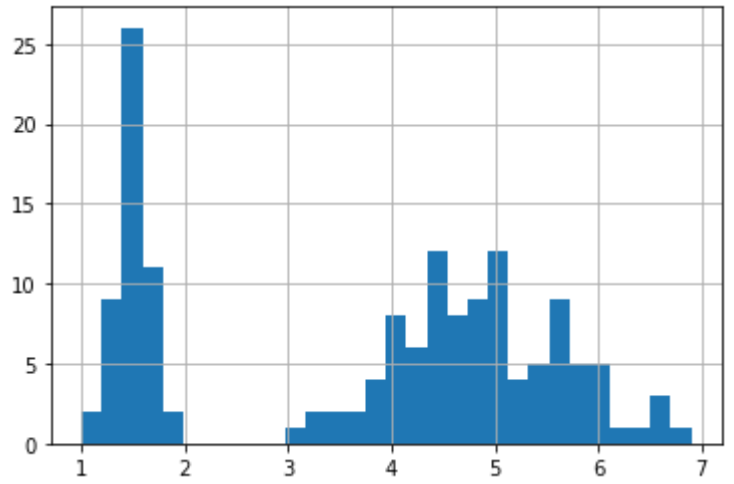


In [67]:

```
df['PetalLengthCm'].hist(bins = 30)
```

Out[67]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x1cb13279d00>
```



Classification OF Numerical And Categorical Variables

In [59]:

```
for i in columns:
    if df[i].dtype != object:
        print(f"{i} = Numerial Variable")
    else:
        print(f"{i} = Categorical Variable")
```

```
Id = Numerial Variable
SepalLengthCm = Numerial Variable
PetalLengthCm = Numerial Variable
PetalWidthCm = Numerial Variable
Species = Categorical Variable
```

Contigecy Table

In [60]:

```
cont_table = pd.crosstab(df['SepalLengthCm'] , df['PetalWidthCm'] )
print(cont_table)
```

```
PetalWidthCm   0.1  0.2  0.3  0.4  0.5  0.6  1.0  1.1  1.2  1.3  ...  1.6  \
SepalLengthCm                                                     ...
4.3              1    0    0    0    0    0    0    0    0    0  ...    0
4.4              0    3    0    0    0    0    0    0    0    0  ...    0
4.5              0    0    1    0    0    0    0    0    0    0  ...    0
4.6              0    3    1    0    0    0    0    0    0    0  ...    0
4.7              0    2    0    0    0    0    0    0    0    0  ...    0
4.8              1    3    1    0    0    0    0    0    0    0  ...    0
4.9              3    1    0    0    0    0    1    0    0    0  ...    0
5.0              0    5    1    1    0    1    2    0    0    0  ...    0
5.1              0    3    2    2    1    0    0    1    0    0  ...    0
5.2              1    2    0    0    0    0    0    0    0    0  ...    0
5.3              0    1    0    0    0    0    0    0    0    0  ...    0
5.4              0    2    0    3    0    0    0    0    0    0  ...    0
5.5              0    2    0    0    0    0    1    1    1    2  ...    0
5.6              0    0    0    0    0    0    0    1    0    3  ...    0
5.7              0    0    1    1    0    0    1    0    1    3  ...    0
5.8              0    1    0    0    0    0    1    0    2    0  ...    0
5.9              0    0    0    0    0    0    0    0    0    0  ...    0
6.0              0    0    0    0    0    0    1    0    0    0  ...    2
6.1              0    0    0    0    0    0    0    0    1    1  ...    0
6.2              0    0    0    0    0    0    0    0    0    1  ...    0
6.3              0    0    0    0    0    0    0    0    0    1  ...    1
6.4              0    0    0    0    0    0    0    0    0    1  ...    0
6.5              0    0    0    0    0    0    0    0    0    0  ...    0
6.6              0    0    0    0    0    0    0    0    0    1  ...    0
6.7              0    0    0    0    0    0    0    0    0    0  ...    0
6.8              0    0    0    0    0    0    0    0    0    0  ...    0
6.9              0    0    0    0    0    0    0    0    0    0  ...    0
7.0              0    0    0    0    0    0    0    0    0    0  ...    0
7.1              0    0    0    0    0    0    0    0    0    0  ...    0
7.2              0    0    0    0    0    0    0    0    0    0  ...    1
7.3              0    0    0    0    0    0    0    0    0    0  ...    0
7.4              0    0    0    0    0    0    0    0    0    0  ...    0
7.6              0    0    0    0    0    0    0    0    0    0  ...    0
7.7              0    0    0    0    0    0    0    0    0    0  ...    0
7.9              0    0    0    0    0    0    0    0    0    0  ...    0

PetalWidthCm   1.7  1.8  1.9  2.0  2.1  2.2  2.3  2.4  2.5
SepalLengthCm
4.3              0    0    0    0    0    0    0    0    0
4.4              0    0    0    0    0    0    0    0    0
4.5              0    0    0    0    0    0    0    0    0
4.6              0    0    0    0    0    0    0    0    0
4.7              0    0    0    0    0    0    0    0    0
4.8              0    0    0    0    0    0    0    0    0
4.9              1    0    0    0    0    0    0    0    0
5.0              0    0    0    0    0    0    0    0    0
5.1              0    0    0    0    0    0    0    0    0
5.2              0    0    0    0    0    0    0    0    0
5.3              0    0    0    0    0    0    0    0    0
5.4              0    0    0    0    0    0    0    0    0
5.5              0    0    0    0    0    0    0    0    0
5.6              0    0    0    1    0    0    0    0    0
5.7              0    0    0    1    0    0    0    0    0
5.8              0    0    2    0    0    0    0    1    0
```

5.9	0	2	0	0	0	0	0	0	0
6.0	0	1	0	0	0	0	0	0	0
6.1	0	1	0	0	0	0	0	0	0
6.2	0	1	0	0	0	0	1	0	0
6.3	0	2	1	0	0	0	0	1	1
6.4	0	1	1	0	1	1	1	0	0
6.5	0	1	0	2	0	1	0	0	0
6.6	0	0	0	0	0	0	0	0	0
6.7	1	1	0	0	1	0	1	1	1
6.8	0	0	0	0	1	0	1	0	0
6.9	0	0	0	0	1	0	2	0	0
7.0	0	0	0	0	0	0	0	0	0
7.1	0	0	0	0	1	0	0	0	0
7.2	0	1	0	0	0	0	0	0	1
7.3	0	1	0	0	0	0	0	0	0
7.4	0	0	1	0	0	0	0	0	0
7.6	0	0	0	0	1	0	0	0	0
7.7	0	0	0	1	0	1	2	0	0
7.9	0	0	0	1	0	0	0	0	0

[35 rows x 22 columns]

calculate mean for all columns

In [25]:

```
for i in columns:
    if df[i].dtype != object:
        mean = sum(df[i]) / len(df[i])
        print(f' Mean for Column {i} = {mean}')
```

```
 Mean for Column Id = 75.5
 Mean for Column SepalLengthCm = 5.843333333333335
 Mean for Column PetalLengthCm = 3.7586666666666693
 Mean for Column PetalWidthCm = 1.1986666666666672
```

Calulate Median For All Columns

In [30]:

```
for i in columns:
    if df[i].dtype != object:
        if len(df[i]) %2 == 0:
            ind = len(df[i])/2
            print(f'Median for column {i} = {df[i][ind]}')
        else:
            ind = len(df[i])//2
            print(f'Median for column {i} = {df[i][ind]}')
```

```
Median for column Id = 76
Median for column SepalLengthCm = 6.6
Median for column PetalLengthCm = 4.4
Median for column PetalWidthCm = 1.4
```

Calculate Mode for Column

In [55]:

```
temp=  {}
cnt = 0
for i in df["SepalLengthCm"].unique():
    temp[i] = 0
for i in df["SepalLengthCm"]:
    if i in temp:
        temp[i] += 1
print("Mode for Column = " ,end = "")
print(max(temp.values()))
```

```
Mode for Column = 10
```

Calculate Variance

In [57]:

```
for i in columns:
    if df[i].dtype != object:
        variance = df[i].var()
        print(f"Variance For column {i} = {variance}")
```

```
Variance For column Id = 1887.5
Variance For column SepalLengthCm = 0.6856935123042505
Variance For column PetalLengthCm = 3.1131794183445156
Variance For column PetalWidthCm = 0.5824143176733784
```

Calculate Standard Deviation

In [58]:

```
for i in columns:
    if df[i].dtype != object:
        sd = df[i].std()
        print(f"Variance For column {i} = {sd}")
```

```
Variance For column Id = 43.445367992456916
Variance For column SepalLengthCm = 0.8280661279778629
Variance For column PetalLengthCm = 1.7644204199522617
Variance For column PetalWidthCm = 0.7631607417008414
```

Show categorical Data(Binary, Ordinal , Nominal)

In [63]:

```
print(f"In given data set {columns[-1]} is nominal Data and there is no any Binary Or Ordin
```

```
In given data set Species is nominal Data and there is no any Binary Or Ordi
nal Data
```

Conclusion - performed various statistical operations on given dataset.