```
from google.colab import files
upload=files.upload()
```

```
Choose Files IRIS.csv
• IRIS.csv(text/csv) - 4617 bytes, last modified: 5/3/2023 - 100% done
Saving IRIS.csv to IRIS (4).csv
```

```
import pandas as pd
df=pd.read_csv("IRIS.csv")
```

```
df.head()
```

	sepal_length	sepal_width	petal_length	petal_width	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

```
df.shape
```

```
(150, 5)
```

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 150 entries, 0 to 149
Data columns (total 5 columns):
 #   Column        Non-Null Count  Dtype  
---  ------        --------------  -----  
 0   sepal_length  150 non-null    float64
 1   sepal_width   150 non-null    float64
 2   petal_length  150 non-null    float64
 3   petal_width   150 non-null    float64
 4   species       150 non-null    object 
dtypes: float64(4), object(1)
memory usage: 6.0+ KB
```

```
df["species"]
```

```
0         Iris-setosa
1         Iris-setosa
2         Iris-setosa
3         Iris-setosa
4         Iris-setosa
            ...
145    Iris-virginica
146    Iris-virginica
147    Iris-virginica
148    Iris-virginica
149    Iris-virginica
Name: species, Length: 150, dtype: object
```

Double-click (or enter) to edit

checking the categories in the data set

```
df.species.unique()
```

```
array(['Iris-setosa', 'Iris-versicolor', 'Iris-virginica'], dtype=object)
```

there are 4 categories in the species column

summary of the table

```
df.describe()
```

	sepal_length	sepal_width	petal_length	petal_width
count	150.000000	150.000000	150.000000	150.000000
mean	5.843333	3.054000	3.758667	1.198667
std	0.828066	0.433594	1.764420	0.763161
min	4.300000	2.000000	1.000000	0.100000
25%	5.100000	2.800000	1.600000	0.300000
50%	5.800000	3.000000	4.350000	1.300000
75%	6.400000	3.300000	5.100000	1.800000
max	7.900000	4.400000	6.900000	2.500000

checkng the corelation

```
df.corr()
```

	sepal_length	sepal_width	petal_length	petal_width
sepal_length	1.000000	-0.109369	0.871754	0.817954
sepal_width	-0.109369	1.000000	-0.420516	-0.356544
petal_length	0.871754	-0.420516	1.000000	0.962757
petal_width	0.817954	-0.356544	0.962757	1.000000

from above corelation table we can detrmine that the correlation between petal width and petal_length is high

checking is there null values or not

```
df.isnull()
```

	sepal_length	sepal_width	petal_length	petal_width	species
0	False	False	False	False	False
1	False	False	False	False	False
2	False	False	False	False	False
3	False	False	False	False	False
4	False	False	False	False	False
...	...	...	...	...	...
145	False	False	False	False	False
146	False	False	False	False	False
147	False	False	False	False	False
148	False	False	False	False	False
149	False	False	False	False	False

150 rows × 5 columns

checking how many categories present in the dataset

```
df.nunique()
```

```
sepal_length    35
sepal_width     23
petal_length    43
petal_width     22
species          3
dtype: int64
```

```
df.duplicated()
```

```
0      False
1      False
2      False
3      False
4      False
       ...
145    False
146    False
147    False
148    False
149    False
Length: 150, dtype: bool
```

```
df.sepal_length.unique()
```

```
array([5.1, 4.9, 4.7, 4.6, 5. , 5.4, 4.4, 4.8, 4.3, 5.8, 5.7, 5.2, 5.5,
       4.5, 5.3, 7. , 6.4, 6.9, 6.5, 6.3, 6.6, 5.9, 6. , 6.1, 5.6, 6.7,
       6.2, 6.8, 7.1, 7.6, 7.3, 7.2, 7.7, 7.4, 7.9])
```

```
df.isna()
```

	sepal_length	sepal_width	petal_length	petal_width	species
0	False	False	False	False	False
1	False	False	False	False	False
2	False	False	False	False	False
3	False	False	False	False	False
4	False	False	False	False	False
...	...	...	...	...	...
145	False	False	False	False	False
146	False	False	False	False	False
147	False	False	False	False	False
148	False	False	False	False	False
149	False	False	False	False	False

150 rows × 5 columns

```
df.isna().sum()
```

```
sepal_length    0
sepal_width     0
petal_length    0
petal_width     0
species         0
dtype: int64
```

```
df.isnull().sum()
```

```
sepal_length    0
sepal_width     0
petal_length    0
petal_width     0
species         0
dtype: int64
```

separating the categories

```
df[df["species"]=="Iris-setosa"].shape
```

```
(50, 5)
```

```
df[df["species"]=="Iris-versicolor"].shape
```

```
(50, 5)
```

```
df[df["species"]=="Iris-virginica"].shape
```

```
(50, 5)
```

counnting the how many feature are there below their miin/max value baesd on the descirbe()

```
df[df["sepal_length"]<=4.300000].count()
```

```
sepal_length    1
sepal_width     1
petal_length    1
petal_width     1
species         1
dtype: int64
```

```
df[df["sepal_width"]<=2].count()
```

```
sepal_length    1
sepal_width     1
petal_length    1
petal_width     1
species         1
dtype: int64
```

```
df[df["petal_length"]<=1].count()
```

```
sepal_length    1
sepal_width     1
petal_length    1
petal_width     1
species         1
dtype: int64
```

```
df[df["petal_width"]<=0.100000].count()
```

```
sepal_length    6
sepal_width     6
petal_length    6
petal_width     6
species         6
dtype: int64
```

seperating X and Y variable

```
X=df.iloc[:,0:4]
Y=df["species"]
```

```
df.hist(None)
```

```
array([[<Axes: title={'center': 'sepal_length'}>,
        <Axes: title={'center': 'sepal_width'}>],
       [<Axes: title={'center': 'petal_length'}>,
        <Axes: title={'center': 'petal_width'}>]], dtype=object)
```

```
df.hist(None)
```

```
array([[<Axes: title={'center': 'sepal_length'}>,
        <Axes: title={'center': 'sepal_width'}>],
       [<Axes: title={'center': 'petal_length'}>,
        <Axes: title={'center': 'petal_width'}>]], dtype=object)
```

```
import seaborn as sns
import matplotlib.pyplot as plt


sns.boxplot(x="petal_length", y="petal_width", data=df )
plt.show()


bargraph = df. plot. bar(x = 'sepal_length', y = "petal_length", fontsize='9')
```

Double-click (or enter) to edit

Double-click (or enter) to edit

✓ 1s completed at 11:42 AM