```
1 ## iris flowers classification
```

```
1 import pandas as pd
2 import numpy as np
3 import seaborn as sns
4 %matplotlib inline
5 import matplotlib.pyplot as plt
6 sns.set(style="white", color_codes=True)
7 import os
```

```
1 df = pd.read_csv('/content/IRIS NEW FILE.csv')
```
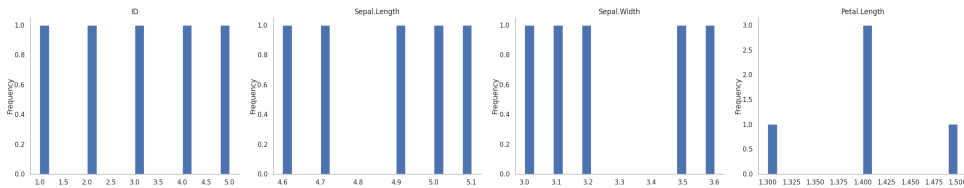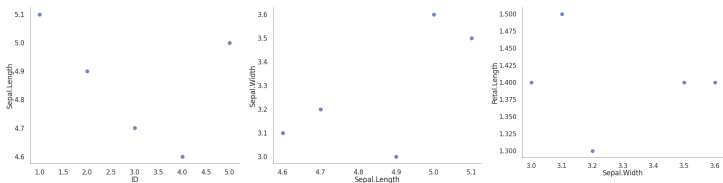
```
1 df.head()
```

	ID	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
0	1	5.1	3.5	1.4	0.2	setosa
1	2	4.9	3.0	1.4	0.2	setosa
2	3	4.7	3.2	1.3	0.2	setosa
3	4	4.6	3.1	1.5	0.2	setosa
4	5	5.0	3.6	1.4	0.2	setosa

Distributions

2-d distributions

Time series
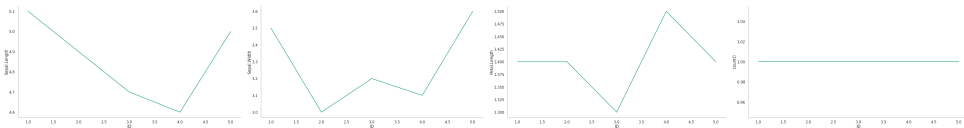
Values
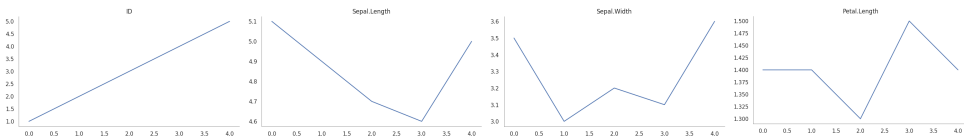

```
1 df.tail()
```

	ID	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
145	146	6.7	3.0	5.2	2.3	virginica
146	147	6.3	2.5	5.0	1.9	virginica
147	148	6.5	3.0	5.2	2.0	virginica
148	149	6.2	3.4	5.4	2.3	virginica
149	150	5.9	3.0	5.1	1.8	virginica

Distributions
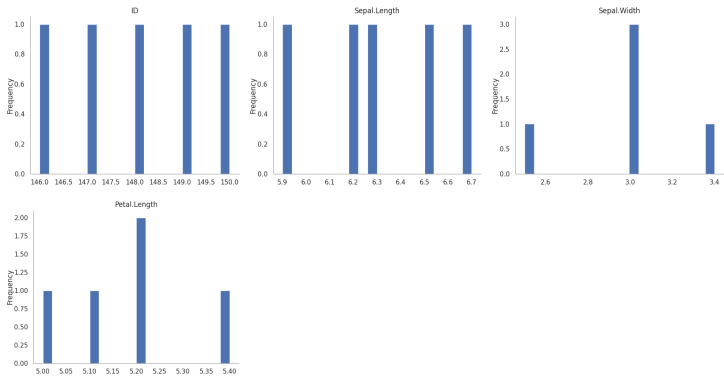
2-d distributions


```
1 df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 150 entries, 0 to 149
Data columns (total 6 columns):
 #   Column        Non-Null Count  Dtype
---  ------        --------------  -----
 0   ID            150 non-null    int64
 1   Sepal.Length  150 non-null    float64
 2   Sepal.Width   150 non-null    float64
 3   Petal.Length  150 non-null    float64
 4   Petal.Width   150 non-null    float64
 5   Species       150 non-null    object
dtypes: float64(4), int64(1), object(1)
memory usage: 7.2+ KB
```

```
1 df.corr()
```

```
<ipython-input-210-2f6f6606aa2c>:1: FutureWarning: The default value of numeric_only in
  df.corr()
```

	ID	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width
ID	1.000000	0.716676	-0.402301	0.882637	0.900027
Sepal.Length	0.716676	1.000000	-0.117570	0.871754	0.817941
Sepal.Width	-0.402301	-0.117570	1.000000	-0.428440	-0.366126
Petal.Length	0.882637	0.871754	-0.428440	1.000000	0.962865

```
1 df.describe()
```

	ID	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width
count	150.000000	150.000000	150.000000	150.000000	150.000000
mean	75.500000	5.843333	3.057333	3.758000	1.199333
std	43.445368	0.828066	0.435866	1.765298	0.762238
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

Distributions

2-d distributions
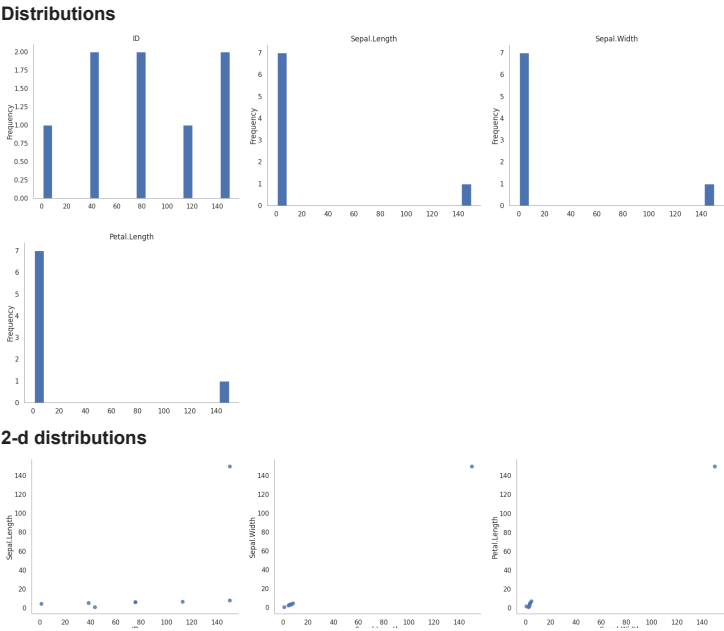

```
1 df.dropna(inplace=True)
```

```
1 df.columns
```

```
Index(['ID', 'Sepal.Length', 'Sepal.Width', 'Petal.Length', 'Petal.Width',
       'Species'],
      dtype='object')
```

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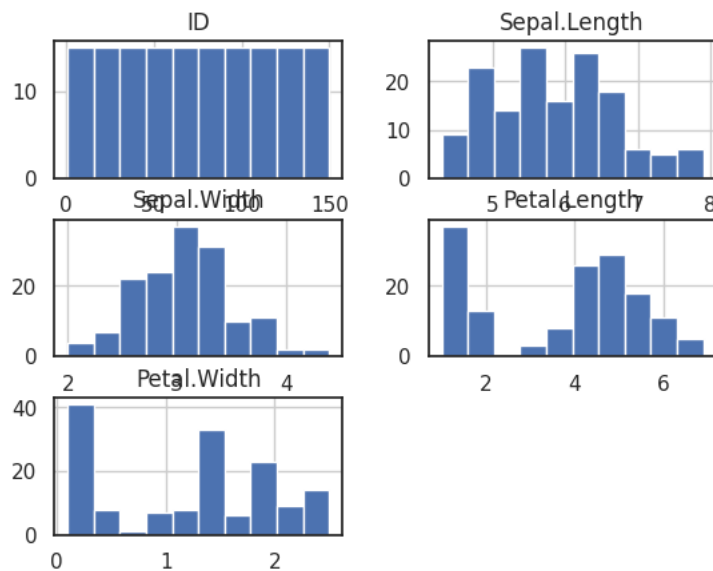
```
0      False
1      False
2      False
3      False
4      False
       ...
145    False
146    False
147    False
148    False
149    False
Length: 150, dtype: bool
```

```
1 df.nunique()
```

```
ID              150
Sepal.Length     35
Sepal.Width      23
Petal.Length     43
Petal.Width      22
Species           3
dtype: int64
```

```
1 df.hist()
```
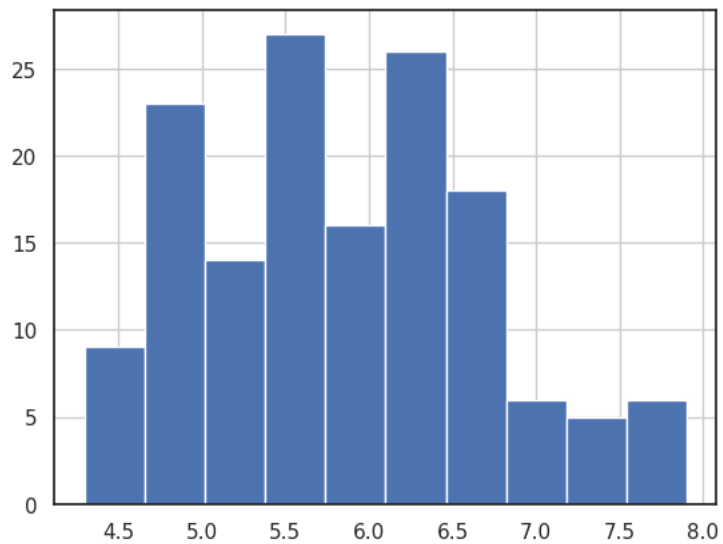
```
array([[<Axes: title={'center': 'ID'}>,
        <Axes: title={'center': 'Sepal.Length'}>],
       [<Axes: title={'center': 'Sepal.Width'}>,
        <Axes: title={'center': 'Petal.Length'}>],
       [<Axes: title={'center': 'Petal.Width'}>, <Axes: >]], dtype=object)
```



```
1 df.boxplot()
```

```
<Axes: >
```



```
1 df["Species"].value_counts()
```

```
setosa        50
versicolor    50
virginica     50
Name: Species, dtype: int64
```
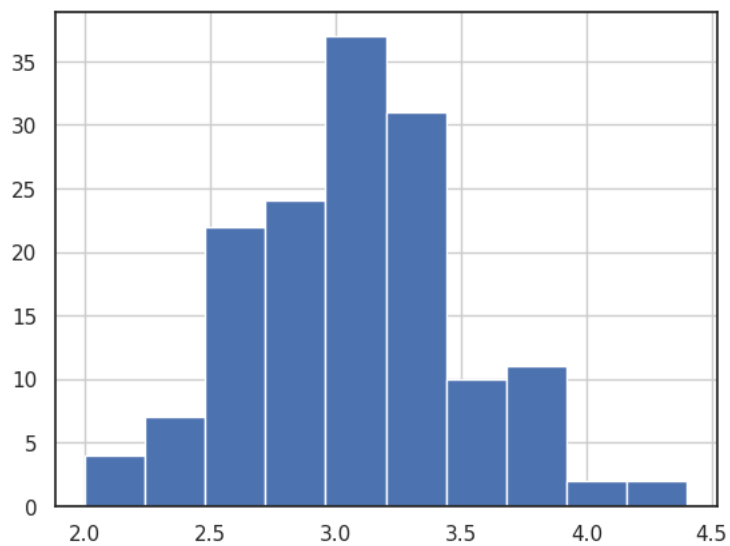
```
1 df['Sepal.Length'].hist()
```

<Axes: >


```
1 df['Sepal.Width'].hist()
```
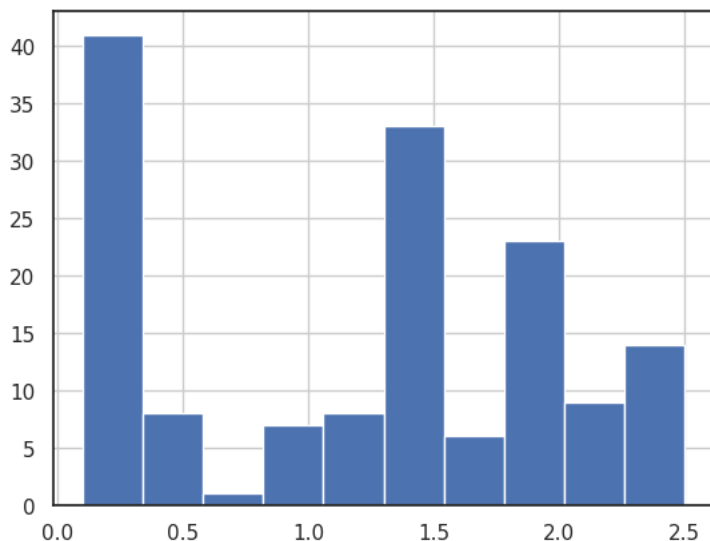
<Axes: >


```
1
```

```
1 df['Petal.Length'].hist()
```

<Axes: >


```
1 df['Petal.Width'].hist()
```

<Axes: >


```
1 df.mean
```

```
<bound method NDFrame._add_numeric_operations.<locals>.mean of        ID  Sepal.Length  Sepal.Width  Petal.Length  Petal.Width
Species
0      1           5.1          3.5           1.4          0.2     setosa
1      2           4.9          3.0           1.4          0.2     setosa
2      3           4.7          3.2           1.3          0.2     setosa
3      4           4.6          3.1           1.5          0.2     setosa
4      5           5.0          3.6           1.4          0.2     setosa
..   ...           ...          ...           ...          ...        ...
145  146           6.7          3.0           5.2          2.3  virginica
146  147           6.3          2.5           5.0          1.9  virginica
147  148           6.5          3.0           5.2          2.0  virginica
148  149           6.2          3.4           5.4          2.3  virginica
149  150           5.9          3.0           5.1          1.8  virginica

[150 rows x 6 columns]>
```

```
1 df.median
```

```
<bound method NDFrame._add_numeric_operations.<locals>.median of        ID  Sepal.Length  Sepal.Width  Petal.Length  Petal.Width
Species
0      1           5.1          3.5           1.4          0.2     setosa
1      2           4.9          3.0           1.4          0.2     setosa
2      3           4.7          3.2           1.3          0.2     setosa
3      4           4.6          3.1           1.5          0.2     setosa
4      5           5.0          3.6           1.4          0.2     setosa
..   ...           ...          ...           ...          ...        ...
145  146           6.7          3.0           5.2          2.3  virginica
146  147           6.3          2.5           5.0          1.9  virginica
147  148           6.5          3.0           5.2          2.0  virginica
148  149           6.2          3.4           5.4          2.3  virginica
149  150           5.9          3.0           5.1          1.8  virginica

[150 rows x 6 columns]>
```