

## Data Visualization

In this part, after cleaning dataset then we go to the next step of processing data and this step is (**Data Visualization**) lets talk about it.

**Data visualization:** is the practice of translating information into a visual context, such as a map or graph, to make data easier for the human brain to understand and pull insights from. The main goal of data visualization is to make it easier to identify patterns, trends and outliers in large data sets. The term is often used interchangeably with others, including information graphics, information visualization and statistical graphics.

Visualization depends on Seaborn.

**Seaborn:** is a Python data visualization library based on matplotlib. It provides a high-level interface for drawing attractive and informative statistical graphics.

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

After cleaning data we used clean dataset .

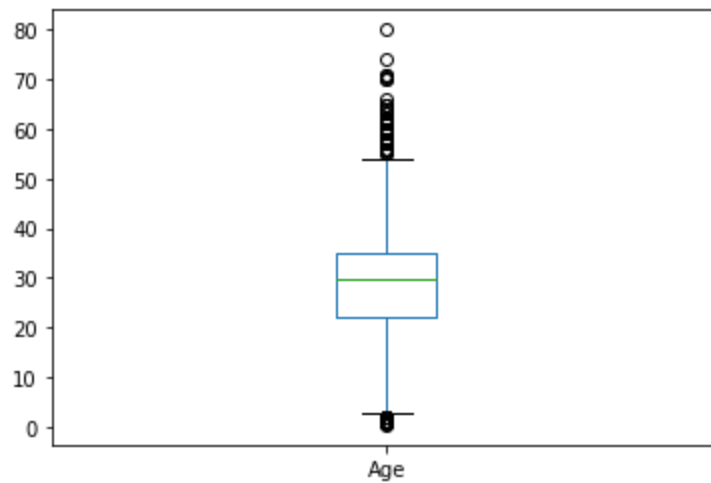
First thing we do in Visualization step is import seaborn library and matplotlib library

**Box plots:**

**Age column:**

```
In [38]: cleaned_data.Age.plot(kind = "box")
```

```
Out[38]: <matplotlib.axes._subplots.AxesSubplot at 0x17d4e0fe580>
```



---

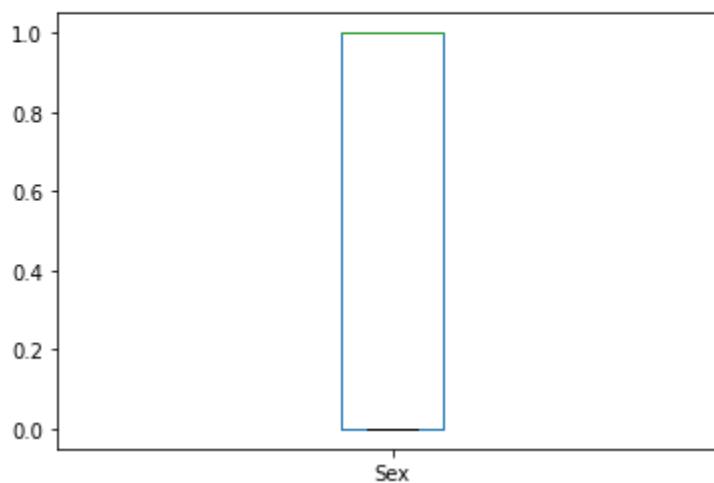
### Conclusion:

the age column has an outliers but we shouldn't remove it because its normal to have very old people and children. the median age is about thirty and the range from zero to eighty.

### Sex column:

```
In [39]: cleaned_data.Sex.plot(kind = "box")
```

```
Out[39]: <matplotlib.axes._subplots.AxesSubplot at 0x17d4e1c7eb0>
```



---

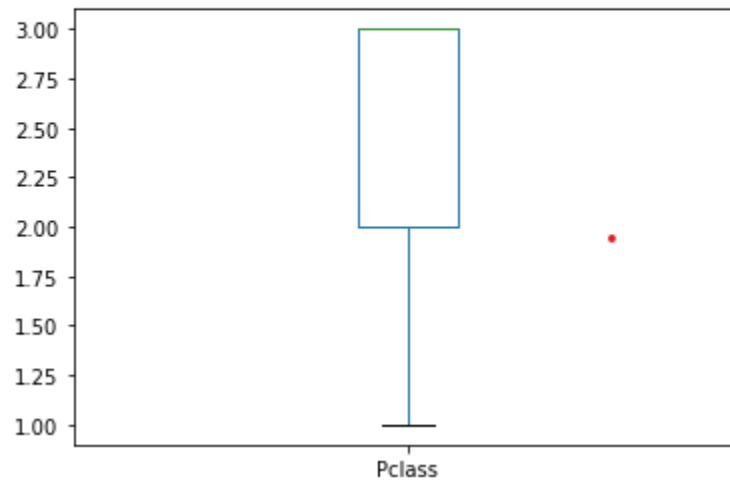
### Conclusion:

Sex column is has only two values : one ,which represent men and zero which represent women.

### Pclass Column:

```
In [40]: cleaned_data.Pclass.plot(kind = "box")
```

```
Out[40]: <matplotlib.axes._subplots.AxesSubplot at 0x17d4e219d60>
```



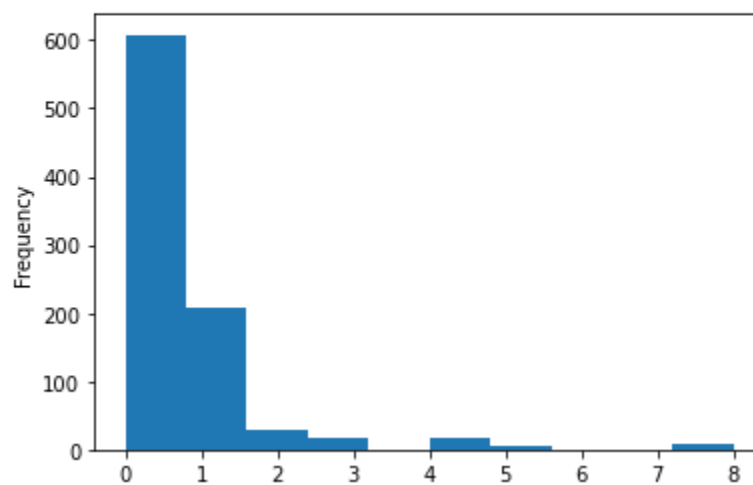
### Conclusion:

passenger class minimum value equal one , the maximum value equal three and the mean equal (and median) equal 2.

### Number od Sublings Column:

```
In [41]: cleaned_data["Number of sublings"].plot(kind = "hist")
```

```
Out[41]: <matplotlib.axes._subplots.AxesSubplot at 0x17d4e280730>
```



### Conclusion:

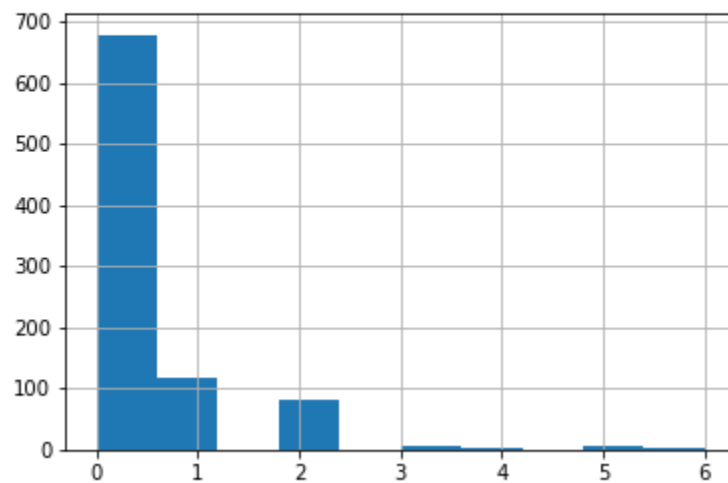
most of the passengers have no sublings on the ship so the mode of Number of sublings column equal 0 and the range is from 0 to 8.

### Histogram:

#### Number of parents and Children:

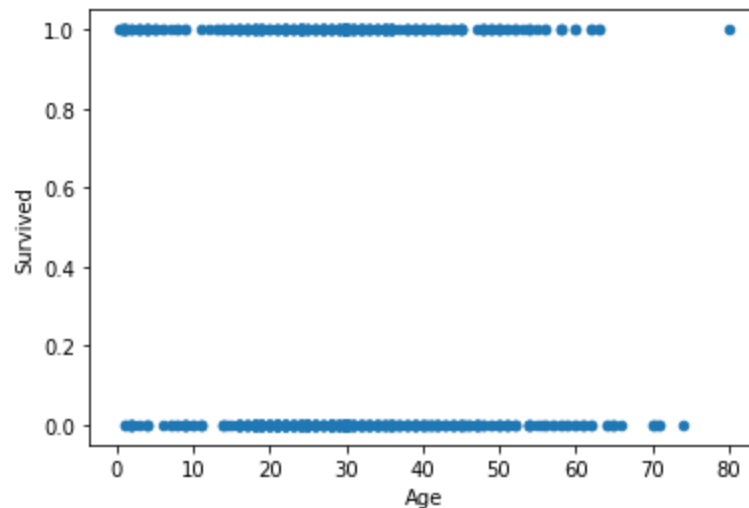
```
In [42]: cleaned_data["Number of paremnts and children"].hist()
```

```
Out[42]: <matplotlib.axes._subplots.AxesSubplot at 0x17d4e31cee0>
```



```
In [43]: cleaned_data.plot(x = "Age",y = "Survived",kind = "scatter")
```

```
Out[43]: <matplotlib.axes._subplots.AxesSubplot at 0x17d4e398c10>
```

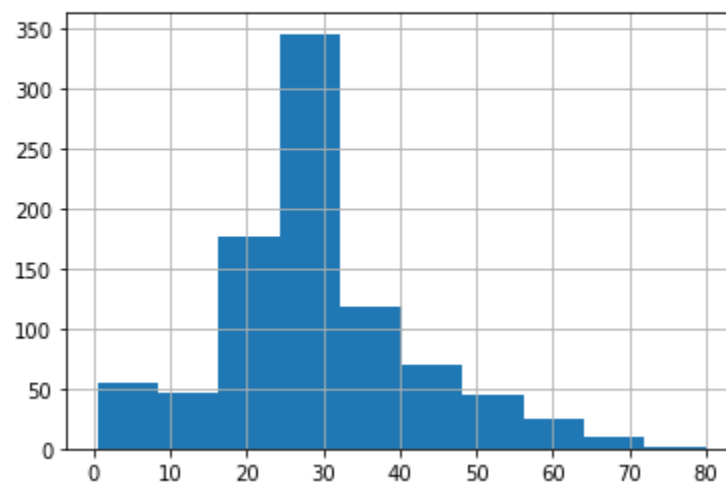


In this step we had plotted Age column and Survived Column to compare the Number of Survived People and their Ages ,then we do a histogram to each column in this dataset

### 1. Age

```
In [42]: cleaned_data.Age.hist()
```

```
Out[42]: <AxesSubplot:>
```

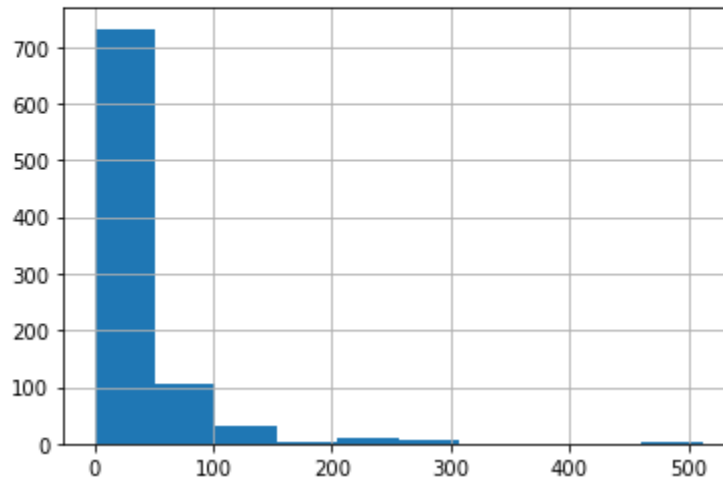


the passengers ages follow right skewed normal distribution which have mean and mode thirty

## 2.Fare

```
In [44]: cleaned_data.Fare.hist()
```

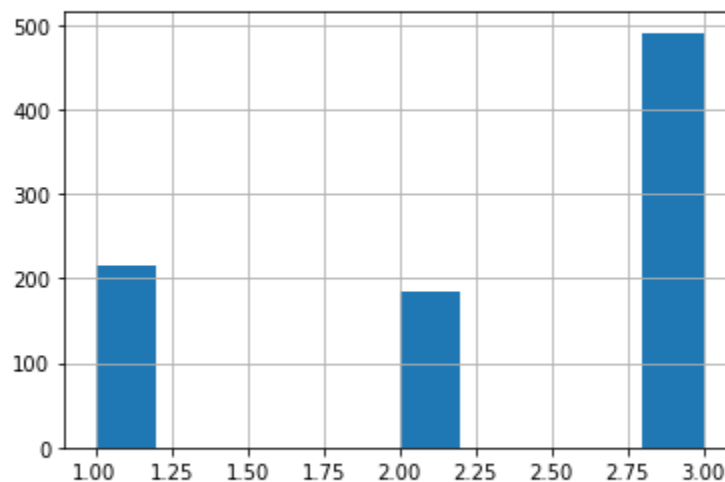
```
Out[44]: <matplotlib.axes._subplots.AxesSubplot at 0x17d4e3fb640>
```



## 3.Pclass

```
In [45]: cleaned_data.Pclass.hist()
```

```
Out[45]: <matplotlib.axes._subplots.AxesSubplot at 0x17d4e466fd0>
```

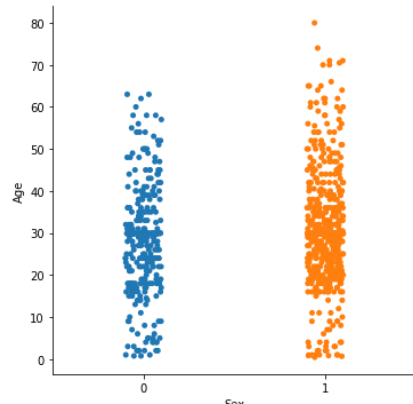


the most frequent passenger class is 3.

```
In [45]: seaborn.catplot("Sex", "Age", data = cleaned_data)
```

```
C:\Users\ALYOSER\anaconda3\lib\site-packages\seaborn\_decorators.py:36: FutureWarning: Pass the following variables as keyword args: x, y. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit keyword will result in an error or misinterpretation.
  warnings.warn(
```

```
Out[45]: <seaborn.axisgrid.FacetGrid at 0x17941f78790>
```

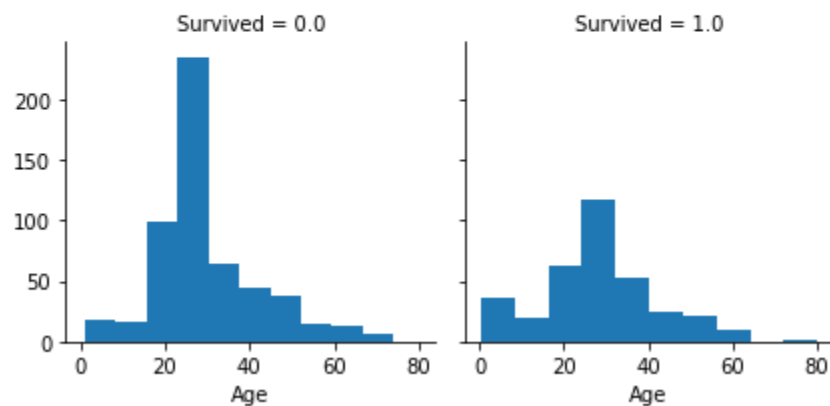


## Conclusion:

the survivor people have a range greater than the dead ones

```
In [46]: Survive = seaborn.FacetGrid(cleaned_data, col = "Survived")
Survive.map(plt.hist, "Age")
```

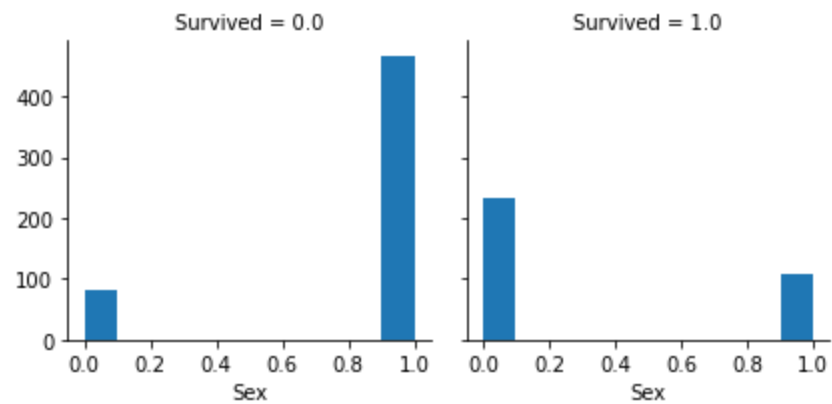
```
Out[46]: <seaborn.axisgrid.FacetGrid at 0x17d4e4dec10>
```



from the charts above we can conclude that youth from 20 to 40 have a more chance to die and childrens and old people have a more chance to survive

```
In [47]: Survive = seaborn.FacetGrid(cleaned_data,col = "Survived")  
Survive.map(plt.hist,"Sex")
```

```
Out[47]: <seaborn.axisgrid.FacetGrid at 0x17d4e1ad250>
```



from the visualization we can conclude that the women(which take zero values) has more chance to survive than men.