Dataset used: `shampoo.csv`

```python
import pandas as pd
import matplotlib.pyplot as plt
from pandas import to_datetime

def parser(x):
    return pd.to_datetime("190"+x, format='%Y-%m')

dataset = pd.read_csv("../data/shampoo.csv", parse_dates=["Month"],
date_format="%Y-%m")
dataset["Month"] = dataset["Month"].apply(parser)
dataset
```

```
        Month  Sales
0   1901-01-01  266.0
1   1901-02-01  145.9
2   1901-03-01  183.1
3   1901-04-01  119.3
4   1901-05-01  180.3
5   1901-06-01  168.5
6   1901-07-01  231.8
7   1901-08-01  224.5
8   1901-09-01  192.8
9   1901-10-01  122.9
10  1901-11-01  336.5
11  1901-12-01  185.9
12  1902-01-01  194.3
13  1902-02-01  149.5
14  1902-03-01  210.1
15  1902-04-01  273.3
16  1902-05-01  191.4
17  1902-06-01  287.0
18  1902-07-01  226.0
19  1902-08-01  303.6
20  1902-09-01  289.9
21  1902-10-01  421.6
22  1902-11-01  264.5
23  1902-12-01  342.3
24  1903-01-01  339.7
25  1903-02-01  440.4
26  1903-03-01  315.9
27  1903-04-01  439.3
28  1903-05-01  401.3
29  1903-06-01  437.4
30  1903-07-01  575.5
31  1903-08-01  407.6
32  1903-09-01  682.0
```

```
33 1903-10-01  475.3
34 1903-11-01  581.3
35 1903-12-01  646.9
```

Manual plotting without setting index column

```python
plt.plot(dataset["Month"], dataset["Sales"])
plt.xlabel("Month")
plt.ylabel("Sales")
plt.xticks(dataset["Month"][::5], rotation=45)
plt.show()
```
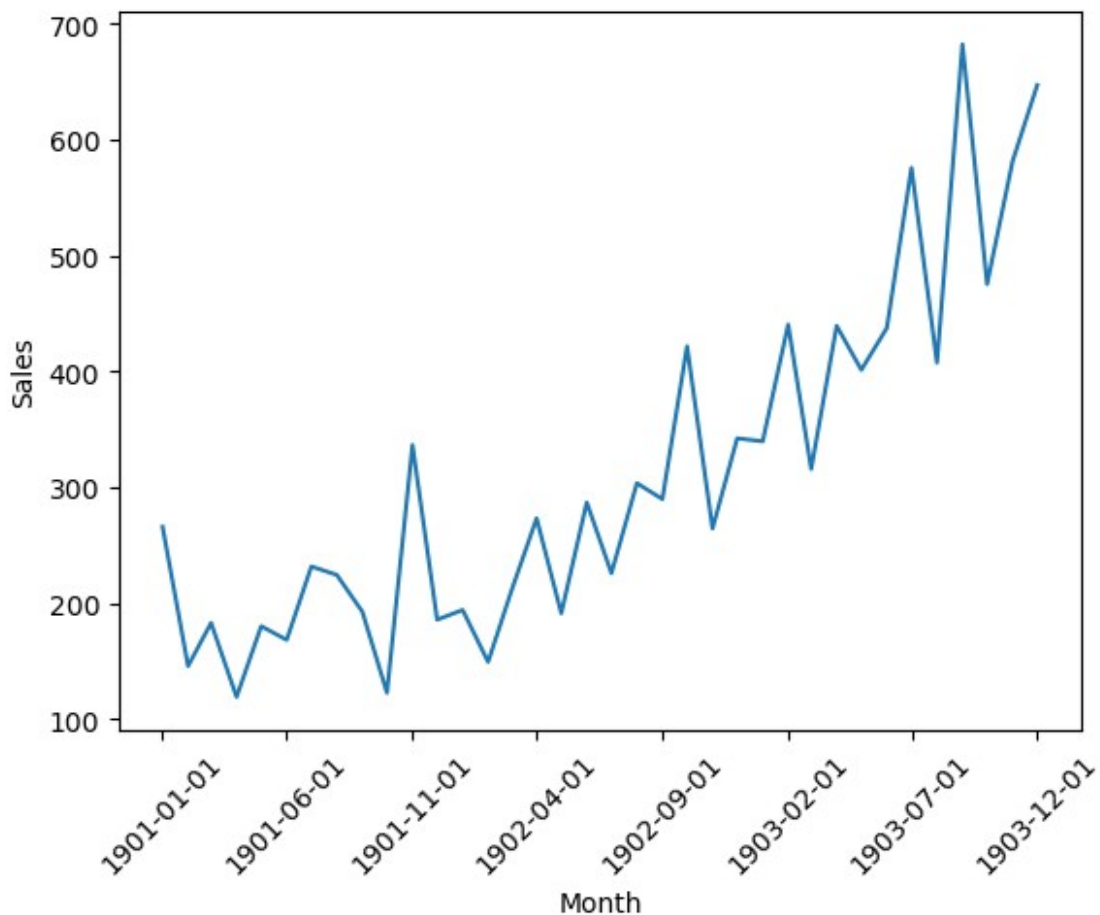

After setting index column as "Month"

```python
dataset.set_index('Month', inplace=True)
dataset
```

```
                Sales
Month
1901-01-01  266.0
1901-02-01  145.9
```

```
1901-03-01   183.1
1901-04-01   119.3
1901-05-01   180.3
1901-06-01   168.5
1901-07-01   231.8
1901-08-01   224.5
1901-09-01   192.8
1901-10-01   122.9
1901-11-01   336.5
1901-12-01   185.9
1902-01-01   194.3
1902-02-01   149.5
1902-03-01   210.1
1902-04-01   273.3
1902-05-01   191.4
1902-06-01   287.0
1902-07-01   226.0
1902-08-01   303.6
1902-09-01   289.9
1902-10-01   421.6
1902-11-01   264.5
1902-12-01   342.3
1903-01-01   339.7
1903-02-01   440.4
1903-03-01   315.9
1903-04-01   439.3
1903-05-01   401.3
1903-06-01   437.4
1903-07-01   575.5
1903-08-01   407.6
1903-09-01   682.0
1903-10-01   475.3
1903-11-01   581.3
1903-12-01   646.9
```

```
dataset.plot()
```

```
<Axes: xlabel='Month'>
```
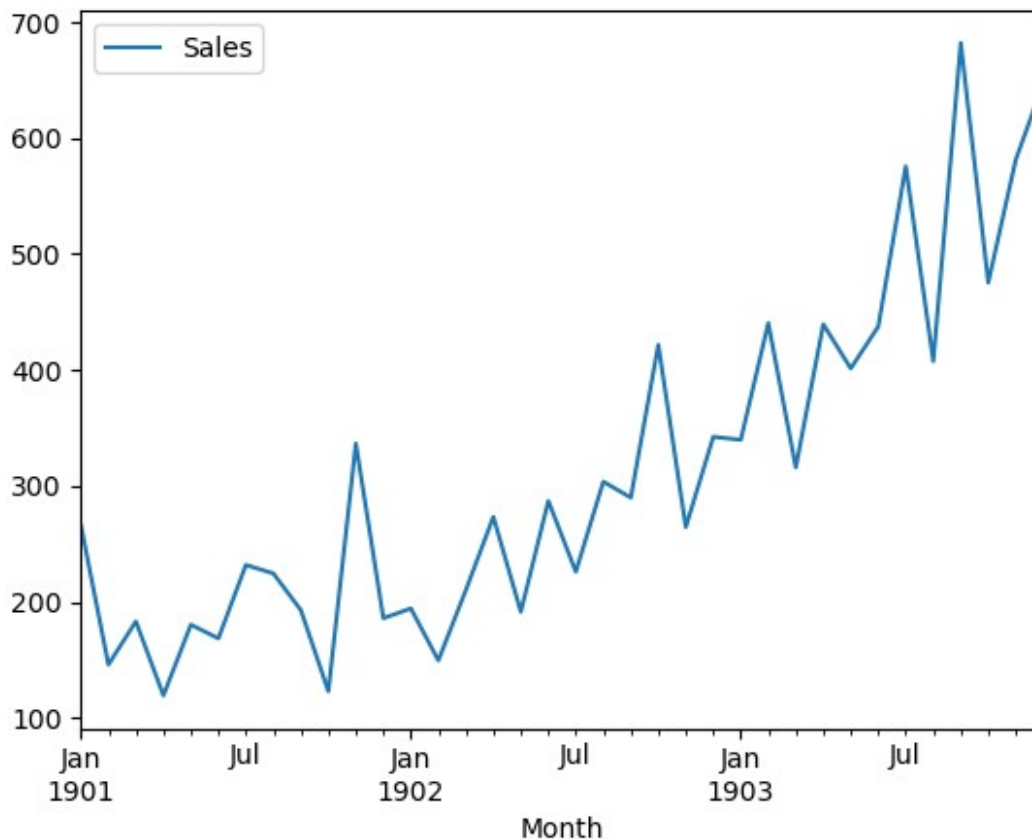
Upsampling

Upsampling to daily sales information

```
upsampled = dataset.resample('D').mean()
upsampled

            Sales
Month
1901-01-01  266.0
1901-01-02    NaN
1901-01-03    NaN
1901-01-04    NaN
1901-01-05    NaN
...           ...
1903-11-27    NaN
1903-11-28    NaN
1903-11-29    NaN
1903-11-30    NaN
1903-12-01  646.9

[1065 rows x 1 columns]
```
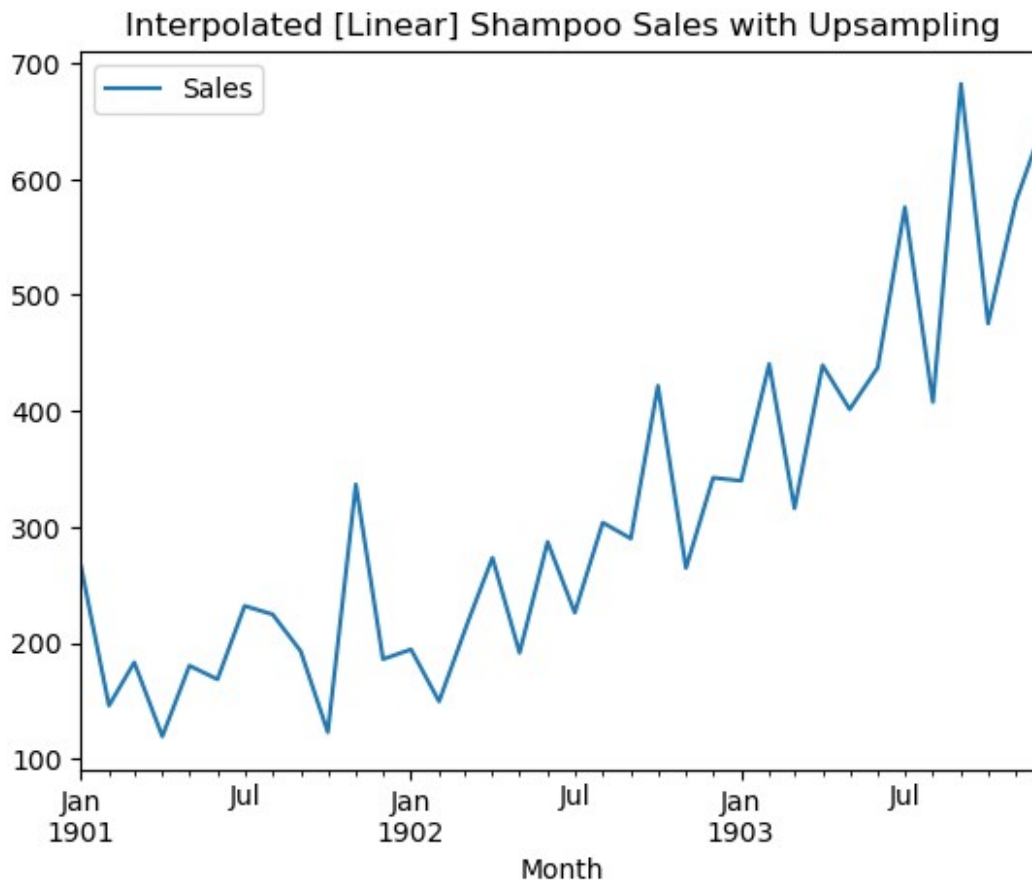
Interpolation

Linear Interploation

```
interpolated_linear = upsampled.interpolate(method="linear")
interpolated_linear

                  Sales
Month
1901-01-01   266.000000
1901-01-02   262.125806
1901-01-03   258.251613
1901-01-04   254.377419
1901-01-05   250.503226
...                 ...
1903-11-27   638.153333
1903-11-28   640.340000
1903-11-29   642.526667
1903-11-30   644.713333
1903-12-01   646.900000

[1065 rows x 1 columns]

interpolated_linear.plot()
# plt.figure(figsize=(20, 20))
plt.title("Interpolated [Linear] Shampoo Sales with Upsampling")
plt.show()
```

Interpolated [Linear] Shampoo Sales with Upsampling

Spline Interpolation

```
interpolated_spline = upsampled.interpolate(method="spline", order=2)
interpolated_spline
```

```
                Sales
Month
1901-01-01   266.000000
1901-01-02   258.630160
1901-01-03   251.560886
1901-01-04   244.720748
1901-01-05   238.109746
...                 ...
1903-11-27   645.060131
1903-11-28   645.754167
1903-11-29   646.316013
1903-11-30   646.745668
1903-12-01   646.900000

[1065 rows x 1 columns]
```

```
interpolated_spline.plot()
plt.title("Interpolated [Spline(2)] Shampoo Sales with Upsampling")
plt.show()
```
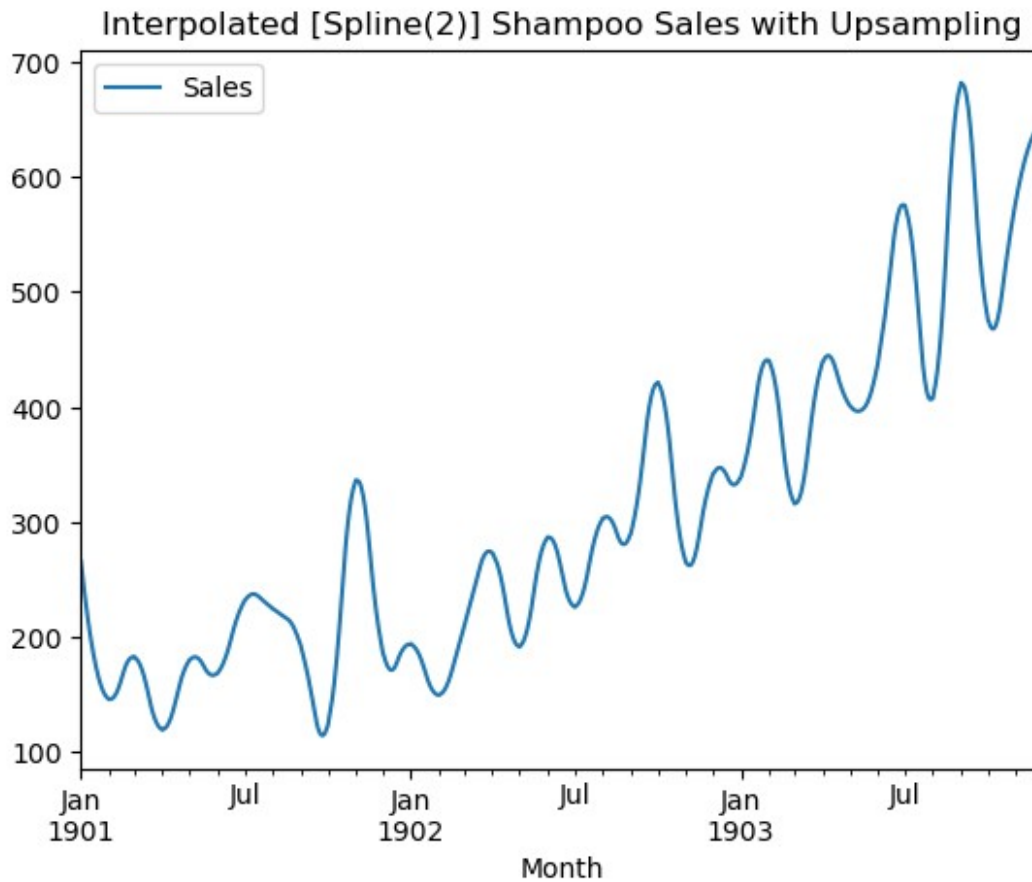

Downsampling

Downsampling to quarterly sales information

```
downsampled = dataset.resample('QE').mean()
downsampled
```

```
                Sales
Month
1901-03-31   198.333333
1901-06-30   156.033333
1901-09-30   216.366667
1901-12-31   215.100000
1902-03-31   184.633333
1902-06-30   250.566667
1902-09-30   273.166667
```
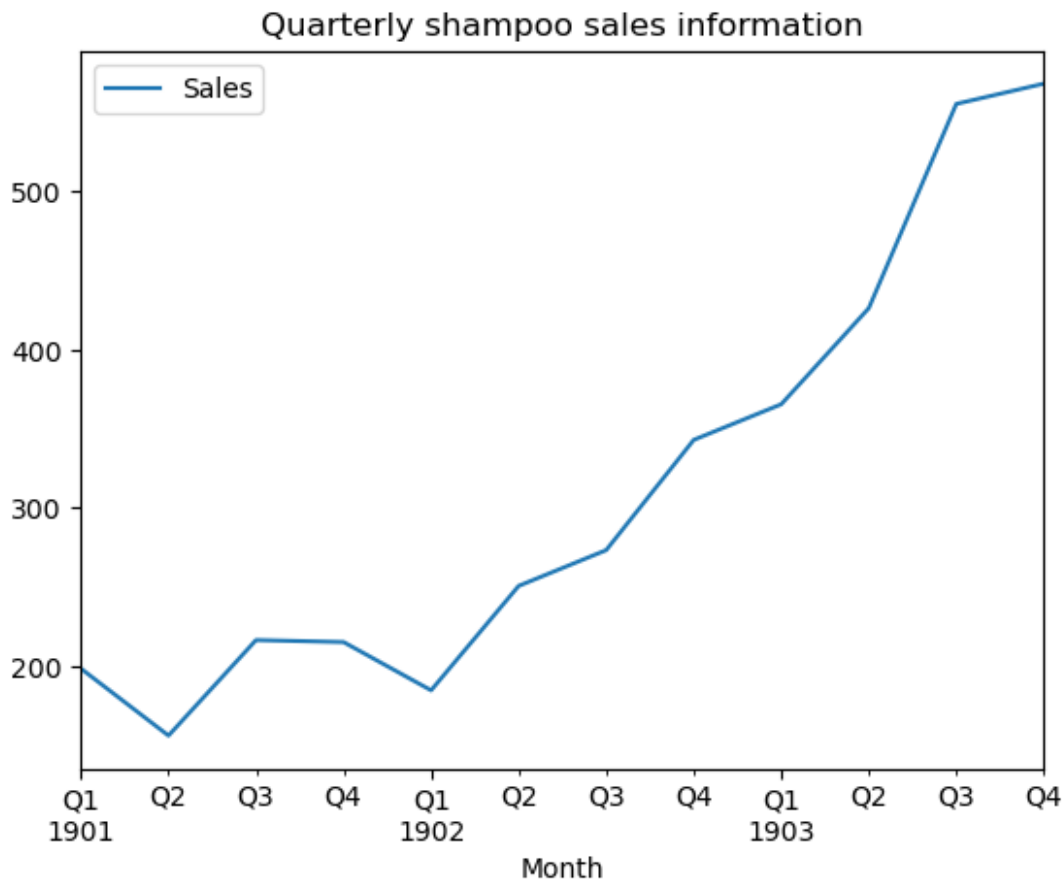
```
1902-12-31  342.800000
1903-03-31  365.333333
1903-06-30  426.000000
1903-09-30  555.033333
1903-12-31  567.833333

downsampled.plot()
plt.title("Quarterly shampoo sales information")
plt.show()
```


Quarterly shampoo sales information

Downsampling to yearly sales information

```
downsampled_2 = dataset.resample('YE').mean()
downsampled_2
```

```
                Sales
Month
1901-12-31  196.458333
1902-12-31  262.791667
1903-12-31  478.550000
```

```
downsampled_2.plot()
plt.title("Yearly shampoo sales information")
plt.show()
```


Yearly shampoo sales information