Dataset: `Month_Value_1.csv`

```
import pandas as pd
import matplotlib.pyplot as plt

def parser(x):
    return pd.to_datetime(x, format='%d.%m.%Y')

dataset = pd.read_csv("../data/Month_Value_1.csv",
parse_dates=["Period"], date_format="%d.%m.%Y")
dataset["Period"] = dataset["Period"].apply(parser)
dataset.dropna(axis=0, inplace=True)
dataset.head(35)
```

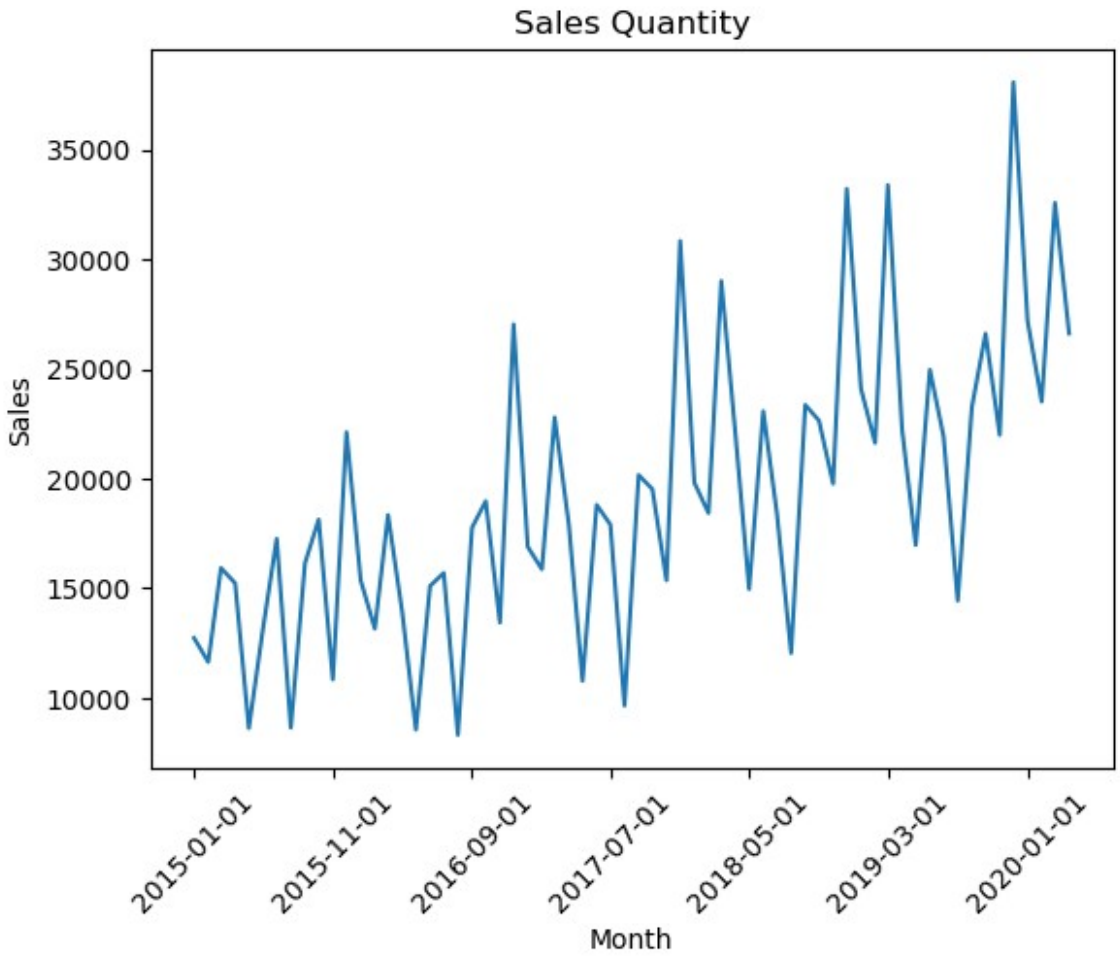
	Period	Revenue	Sales_quantity	Average_cost	\
0	2015-01-01	1.601007e+07	12729.0	1257.763541	
1	2015-02-01	1.580759e+07	11636.0	1358.507000	
2	2015-03-01	2.204715e+07	15922.0	1384.697024	
3	2015-04-01	1.881458e+07	15227.0	1235.606705	
4	2015-05-01	1.402148e+07	8620.0	1626.621765	
5	2015-06-01	1.678393e+07	13160.0	1275.374508	
6	2015-07-01	1.916189e+07	17254.0	1110.576805	
7	2015-08-01	1.520498e+07	8642.0	1759.428870	
8	2015-09-01	2.060394e+07	16144.0	1276.259909	
9	2015-10-01	2.099287e+07	18135.0	1157.588904	
10	2015-11-01	1.499337e+07	10841.0	1383.024597	
11	2015-12-01	2.779181e+07	22113.0	1256.808558	
12	2016-01-01	2.860159e+07	15365.0	1861.476505	
13	2016-02-01	2.236707e+07	13153.0	1700.530226	
14	2016-03-01	2.973861e+07	18339.0	1621.604699	
15	2016-04-01	2.835101e+07	13909.0	2038.321083	
16	2016-05-01	1.526460e+07	8553.0	1784.707557	
17	2016-06-01	2.438566e+07	15101.0	1614.837301	
18	2016-07-01	2.948652e+07	15695.0	1878.720425	
19	2016-08-01	1.527012e+07	8314.0	1836.675157	
20	2016-09-01	3.614103e+07	17764.0	2034.509545	
21	2016-10-01	2.791514e+07	18969.0	1471.619150	
22	2016-11-01	2.127205e+07	13433.0	1583.566541	
23	2016-12-01	4.201416e+07	27029.0	1554.410444	
24	2017-01-01	3.600738e+07	16889.0	2132.001934	
25	2017-02-01	3.039678e+07	15864.0	1916.085185	
26	2017-03-01	4.767813e+07	22786.0	2092.430910	
27	2017-04-01	2.701396e+07	17910.0	1508.317405	
28	2017-05-01	2.494884e+07	10777.0	2315.008323	
29	2017-06-01	3.110135e+07	18799.0	1654.414891	
30	2017-07-01	3.384882e+07	17899.0	1891.101303	
31	2017-08-01	1.645467e+07	9649.0	1705.323552	
32	2017-09-01	3.165009e+07	20159.0	1570.022950	

```
33 2017-10-01  3.157221e+07          19519.0   1617.511431
34 2017-11-01  2.244637e+07          15360.0   1461.352280
```

```
    The_average_annual_payroll_of_the_region
0                                 30024676.0
1                                 30024676.0
2                                 30024676.0
3                                 30024676.0
4                                 30024676.0
5                                 30024676.0
6                                 30024676.0
7                                 30024676.0
8                                 30024676.0
9                                 30024676.0
10                                30024676.0
11                                30024676.0
12                                27828571.0
13                                27828571.0
14                                27828571.0
15                                27828571.0
16                                27828571.0
17                                27828571.0
18                                27828571.0
19                                27828571.0
20                                27828571.0
21                                27828571.0
22                                27828571.0
23                                27828571.0
24                                27406473.0
25                                27406473.0
26                                27406473.0
27                                27406473.0
28                                27406473.0
29                                27406473.0
30                                27406473.0
31                                27406473.0
32                                27406473.0
33                                27406473.0
34                                27406473.0
```

Manual plotting without setting index column

```
plt.plot(dataset["Period"], dataset["Sales_quantity"])
plt.title("Sales Quantity")
plt.xlabel("Month")
plt.ylabel("Sales")
plt.xticks(dataset["Period"][::10], rotation=45)
plt.show()
```



After setting index column as "Month"

```
dataset.set_index('Period', inplace=True)
dataset.head(40)
```

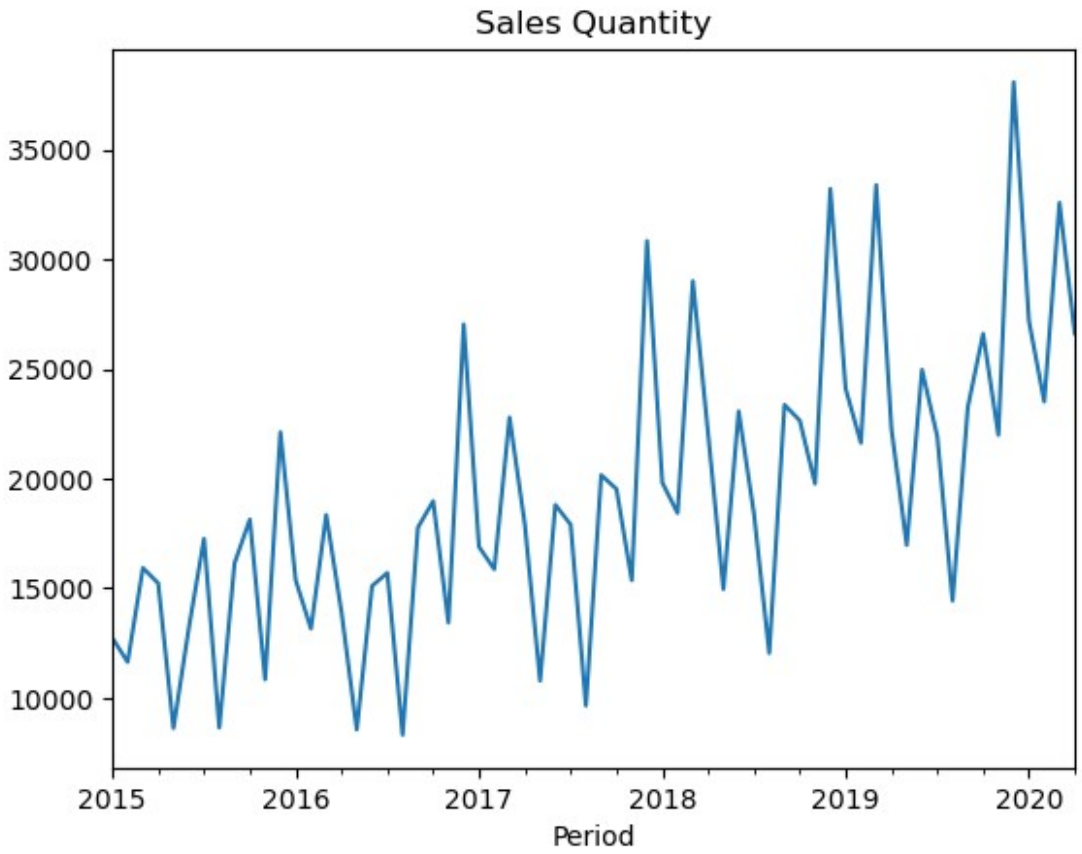
Period	Revenue	Sales_quantity	Average_cost	\
2015-01-01	1.601007e+07	12729.0	1257.763541	
2015-02-01	1.580759e+07	11636.0	1358.507000	
2015-03-01	2.204715e+07	15922.0	1384.697024	
2015-04-01	1.881458e+07	15227.0	1235.606705	
2015-05-01	1.402148e+07	8620.0	1626.621765	
2015-06-01	1.678393e+07	13160.0	1275.374508	
2015-07-01	1.916189e+07	17254.0	1110.576805	
2015-08-01	1.520498e+07	8642.0	1759.428870	
2015-09-01	2.060394e+07	16144.0	1276.259909	
2015-10-01	2.099287e+07	18135.0	1157.588904	
2015-11-01	1.499337e+07	10841.0	1383.024597	
2015-12-01	2.779181e+07	22113.0	1256.808558	
2016-01-01	2.860159e+07	15365.0	1861.476505	

2016-02-01	2.236707e+07	13153.0	1700.530226
2016-03-01	2.973861e+07	18339.0	1621.604699
2016-04-01	2.835101e+07	13909.0	2038.321083
2016-05-01	1.526460e+07	8553.0	1784.707557
2016-06-01	2.438566e+07	15101.0	1614.837301
2016-07-01	2.948652e+07	15695.0	1878.720425
2016-08-01	1.527012e+07	8314.0	1836.675157
2016-09-01	3.614103e+07	17764.0	2034.509545
2016-10-01	2.791514e+07	18969.0	1471.619150
2016-11-01	2.127205e+07	13433.0	1583.566541
2016-12-01	4.201416e+07	27029.0	1554.410444
2017-01-01	3.600738e+07	16889.0	2132.001934
2017-02-01	3.039678e+07	15864.0	1916.085185
2017-03-01	4.767813e+07	22786.0	2092.430910
2017-04-01	2.701396e+07	17910.0	1508.317405
2017-05-01	2.494884e+07	10777.0	2315.008323
2017-06-01	3.110135e+07	18799.0	1654.414891
2017-07-01	3.384882e+07	17899.0	1891.101303
2017-08-01	1.645467e+07	9649.0	1705.323552
2017-09-01	3.165009e+07	20159.0	1570.022950
2017-10-01	3.157221e+07	19519.0	1617.511431
2017-11-01	2.244637e+07	15360.0	1461.352280
2017-12-01	4.496613e+07	30833.0	1458.376602
2018-01-01	4.406752e+07	19812.0	2224.284315
2018-02-01	3.602029e+07	18424.0	1955.074205
2018-03-01	4.699599e+07	29004.0	1620.327900
2018-04-01	3.553649e+07	22033.0	1612.875581

Period	The_average_annual_payroll_of_the_region
2015-01-01	30024676.0
2015-02-01	30024676.0
2015-03-01	30024676.0
2015-04-01	30024676.0
2015-05-01	30024676.0
2015-06-01	30024676.0
2015-07-01	30024676.0
2015-08-01	30024676.0
2015-09-01	30024676.0
2015-10-01	30024676.0
2015-11-01	30024676.0
2015-12-01	30024676.0
2016-01-01	27828571.0
2016-02-01	27828571.0
2016-03-01	27828571.0
2016-04-01	27828571.0
2016-05-01	27828571.0
2016-06-01	27828571.0
2016-07-01	27828571.0

```
2016-08-01                                    27828571.0
2016-09-01                                    27828571.0
2016-10-01                                    27828571.0
2016-11-01                                    27828571.0
2016-12-01                                    27828571.0
2017-01-01                                    27406473.0
2017-02-01                                    27406473.0
2017-03-01                                    27406473.0
2017-04-01                                    27406473.0
2017-05-01                                    27406473.0
2017-06-01                                    27406473.0
2017-07-01                                    27406473.0
2017-08-01                                    27406473.0
2017-09-01                                    27406473.0
2017-10-01                                    27406473.0
2017-11-01                                    27406473.0
2017-12-01                                    27406473.0
2018-01-01                                    28197847.0
2018-02-01                                    28197847.0
2018-03-01                                    28197847.0
2018-04-01                                    28197847.0
```

```
dataset["Sales_quantity"].plot()
plt.title("Sales Quantity")
plt.show()
```



Upsampling

Upsampling to daily sales information

```
upsampled = dataset.resample('D').mean()
upsampled.head(35)
```

Period	Revenue	Sales_quantity	Average_cost
2015-01-01	1.601007e+07	12729.0	1257.763541
2015-01-02	NaN	NaN	NaN
2015-01-03	NaN	NaN	NaN
2015-01-04	NaN	NaN	NaN
2015-01-05	NaN	NaN	NaN
2015-01-06	NaN	NaN	NaN
2015-01-07	NaN	NaN	NaN
2015-01-08	NaN	NaN	NaN
2015-01-09	NaN	NaN	NaN
2015-01-10	NaN	NaN	NaN
2015-01-11	NaN	NaN	NaN
2015-01-12	NaN	NaN	NaN

2015-01-13	NaN	NaN	NaN
2015-01-14	NaN	NaN	NaN
2015-01-15	NaN	NaN	NaN
2015-01-16	NaN	NaN	NaN
2015-01-17	NaN	NaN	NaN
2015-01-18	NaN	NaN	NaN
2015-01-19	NaN	NaN	NaN
2015-01-20	NaN	NaN	NaN
2015-01-21	NaN	NaN	NaN
2015-01-22	NaN	NaN	NaN
2015-01-23	NaN	NaN	NaN
2015-01-24	NaN	NaN	NaN
2015-01-25	NaN	NaN	NaN
2015-01-26	NaN	NaN	NaN
2015-01-27	NaN	NaN	NaN
2015-01-28	NaN	NaN	NaN
2015-01-29	NaN	NaN	NaN
2015-01-30	NaN	NaN	NaN
2015-01-31	NaN	NaN	NaN
2015-02-01	1.580759e+07	11636.0	1358.507000
2015-02-02	NaN	NaN	NaN
2015-02-03	NaN	NaN	NaN
2015-02-04	NaN	NaN	NaN

Period	The_average_annual_payroll_of_the_region
2015-01-01	30024676.0
2015-01-02	NaN
2015-01-03	NaN
2015-01-04	NaN
2015-01-05	NaN
2015-01-06	NaN
2015-01-07	NaN
2015-01-08	NaN
2015-01-09	NaN
2015-01-10	NaN
2015-01-11	NaN
2015-01-12	NaN
2015-01-13	NaN
2015-01-14	NaN
2015-01-15	NaN
2015-01-16	NaN
2015-01-17	NaN
2015-01-18	NaN
2015-01-19	NaN
2015-01-20	NaN
2015-01-21	NaN
2015-01-22	NaN
2015-01-23	NaN

```
2015-01-24                                        NaN
2015-01-25                                        NaN
2015-01-26                                        NaN
2015-01-27                                        NaN
2015-01-28                                        NaN
2015-01-29                                        NaN
2015-01-30                                        NaN
2015-01-31                                        NaN
2015-02-01                                 30024676.0
2015-02-02                                        NaN
2015-02-03                                        NaN
2015-02-04                                        NaN
```

Interpolation

Linear Interploation

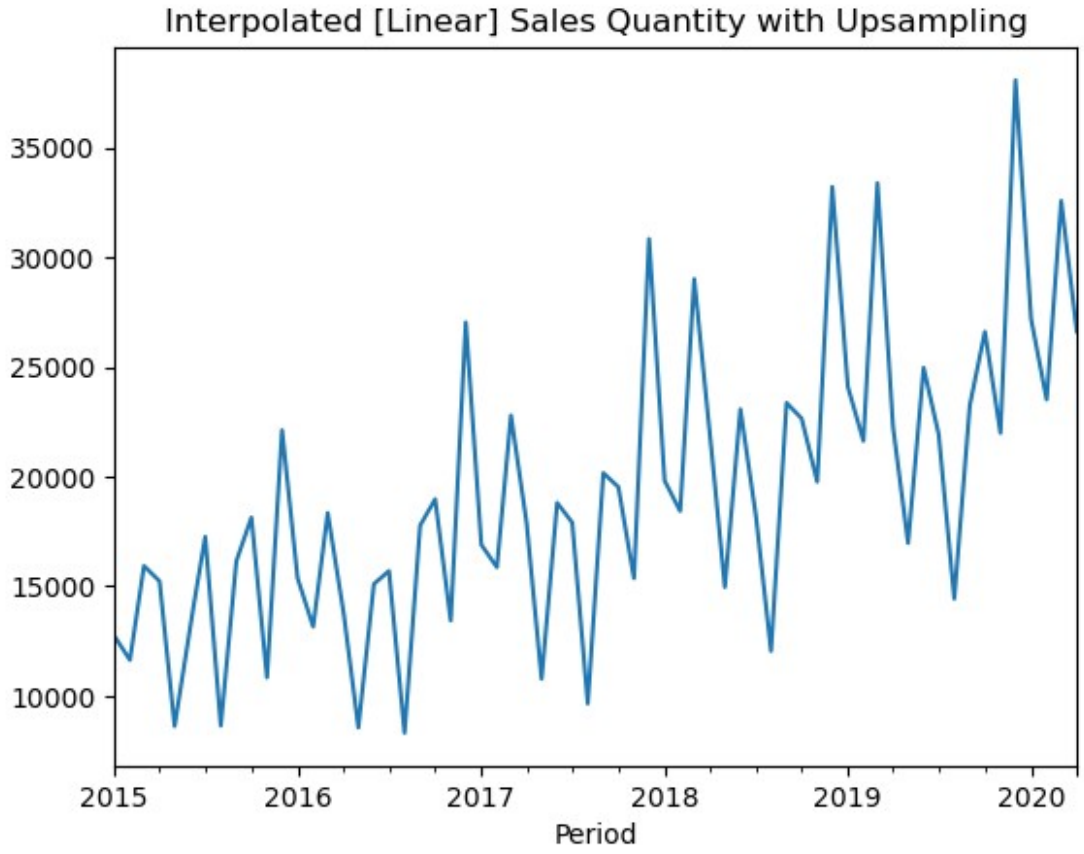
```
interpolated_linear = upsampled.interpolate(method="linear")
interpolated_linear
```

```
                 Revenue  Sales_quantity  Average_cost  \
Period
2015-01-01  1.601007e+07    12729.000000   1257.763541
2015-01-02  1.600354e+07    12693.741935   1261.013330
2015-01-03  1.599701e+07    12658.483871   1264.263119
2015-01-04  1.599048e+07    12623.225806   1267.512908
2015-01-05  1.598395e+07    12587.967742   1270.762697
...                  ...             ...           ...
2020-03-28  5.202411e+07    27383.258065   1910.357107
2020-03-29  5.209825e+07    27191.193548   1924.226528
2020-03-30  5.217240e+07    26999.129032   1938.095949
2020-03-31  5.224655e+07    26807.064516   1951.965370
2020-04-01  5.232069e+07    26615.000000   1965.834790

            The_average_annual_payroll_of_the_region
Period
2015-01-01                                30024676.0
2015-01-02                                30024676.0
2015-01-03                                30024676.0
2015-01-04                                30024676.0
2015-01-05                                30024676.0
...                                              ...
2020-03-28                                29044998.0
2020-03-29                                29044998.0
2020-03-30                                29044998.0
2020-03-31                                29044998.0
2020-04-01                                29044998.0
```

```
[1918 rows x 4 columns]
```

```
interpolated_linear["Sales_quantity"].plot()
plt.title("Interpolated [Linear] Sales Quantity with Upsampling")
plt.show()
```



Spline Interpolation

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

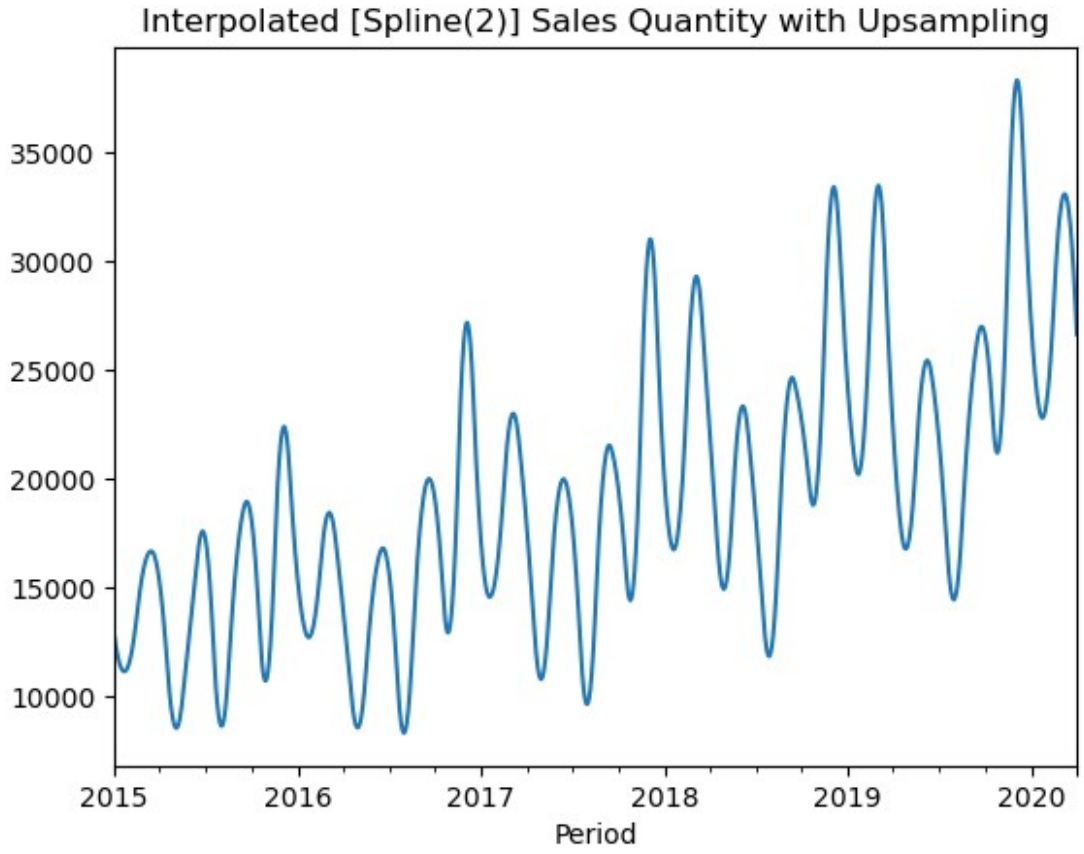
```
                 Revenue  Sales_quantity  Average_cost  \
Period
2015-01-01  1.601007e+07    12729.000000   1257.763541
2015-01-02  1.583837e+07    12570.038903   1261.461420
2015-01-03  1.567768e+07    12419.371290   1265.164313
2015-01-04  1.552800e+07    12276.947648   1268.835473
2015-01-05  1.538933e+07    12142.767976   1272.474901
...                  ...             ...           ...
2020-03-28  5.289016e+07    28565.761592   1872.398093
2020-03-29  5.277185e+07    28110.951415   1894.704359
2020-03-30  5.263750e+07    27634.244231   1917.713292
```

```
2020-03-31  5.248712e+07    27135.640041   1941.424890
2020-04-01  5.232069e+07    26615.000000   1965.834790

            The_average_annual_payroll_of_the_region
Period
2015-01-01                              3.002468e+07
2015-01-02                              3.002468e+07
2015-01-03                              3.002468e+07
2015-01-04                              3.002468e+07
2015-01-05                              3.002468e+07
...                                              ...
2020-03-28                              2.904311e+07
2020-03-29                              2.904353e+07
2020-03-30                              2.904399e+07
2020-03-31                              2.904447e+07
2020-04-01                              2.904500e+07

[1918 rows x 4 columns]
```

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



Downsampling

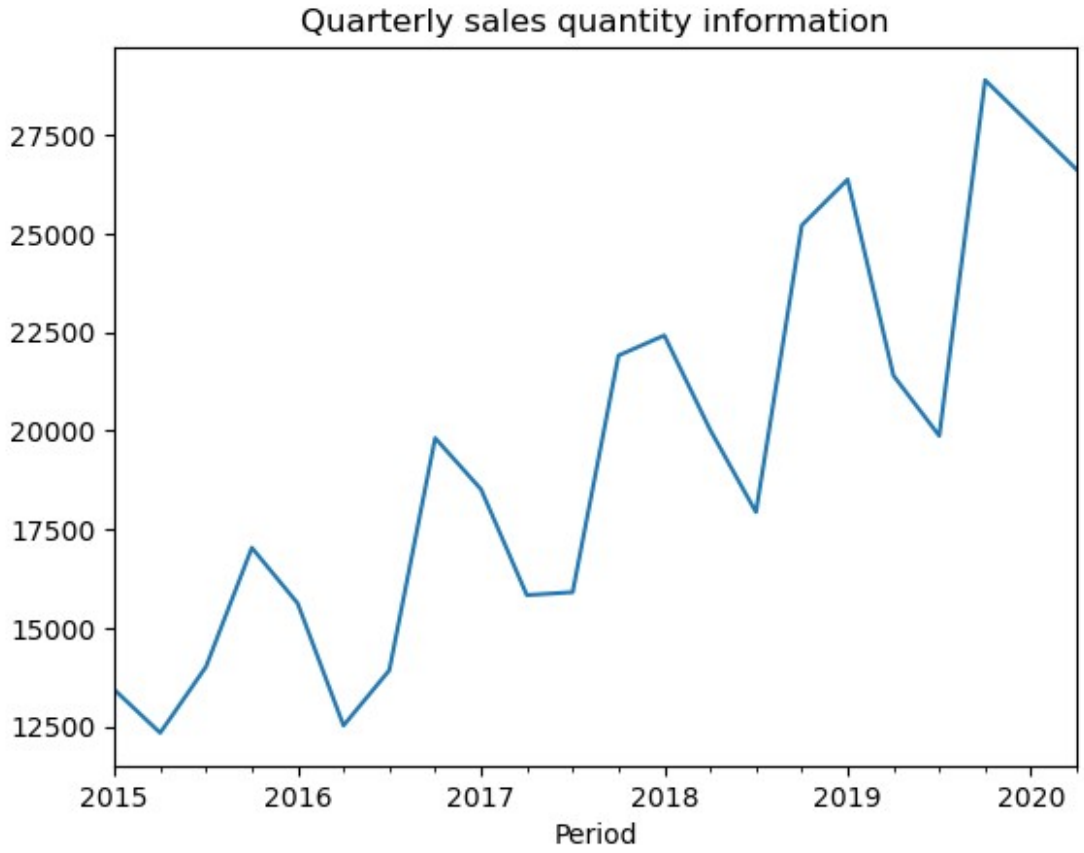
Downsampling to quarterly sales information

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

Period	Revenue	Sales_quantity	Average_cost	The_average_annual_payroll_of_the_region
2015-03-31	1.795494e+07	13429.000000	1333.655855	30024676.0
2015-06-30	1.654000e+07	12335.666667	1379.200992	30024676.0
2015-09-30	1.832361e+07	14013.333333	1382.088528	30024676.0
2015-12-31	2.125935e+07	17029.666667	1265.807353	30024676.0
2016-03-31	2.690242e+07	15619.000000	1727.870477	27828571.0
2016-06-30	2.266709e+07	12521.000000	1812.621980	27828571.0
2016-09-30	2.696589e+07	13924.333333	1916.635042	27828571.0
2016-12-31	3.040045e+07	19810.333333	1536.532045	27828571.0
2017-03-31	3.802743e+07	18513.000000	2046.839343	27406473.0
2017-06-30	2.768805e+07	15828.666667	1825.913540	27406473.0
2017-09-30	2.731786e+07	15902.333333	1722.149269	27406473.0
2017-12-31	3.299490e+07	21904.000000	1512.413438	27406473.0
2018-03-31	4.236127e+07	22413.333333	1933.228807	28197847.0
2018-06-30	3.283238e+07	20019.666667	1680.069495	28197847.0
2018-09-30	3.150623e+07	17933.333333	1798.589491	28197847.0
2018-12-31	4.001751e+07	25205.333333	1616.884383	
2019-03-31	4.240172e+07	26366.333333	1608.979630	
2019-06-30	3.766178e+07	21396.666667	1766.532304	
2019-09-30	4.381736e+07	19867.000000	2248.226241	
2019-12-31	4.643062e+07	28886.333333	1619.261325	
2020-03-31	4.884524e+07	27754.000000	1772.526926	
2020-06-30	5.232069e+07	26615.000000	1965.834790	

```
2018-12-31                                    28197847.0
2019-03-31                                    29878525.0
2019-06-30                                    29878525.0
2019-09-30                                    29878525.0
2019-12-31                                    29878525.0
2020-03-31                                    29044998.0
2020-06-30                                    29044998.0
```

```
downsampled["Sales_quantity"].plot()
plt.title("Quarterly sales quantity information")
plt.show()
```



Downsampling to yearly sales information

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

```
                 Revenue  Sales_quantity  Average_cost  \
Period
2015-12-31  1.851947e+07    14201.916667   1340.188182
2016-12-31  2.673396e+07    15468.666667   1748.414886
2017-12-31  3.150706e+07    18037.000000   1776.828897
2018-12-31  3.667935e+07    21392.916667   1757.193044
```

```
2019-12-31  4.257787e+07    24129.083333   1810.749875
2020-12-31  4.971410e+07    27469.250000   1820.853892

            The_average_annual_payroll_of_the_region
Period
2015-12-31                                   30024676.0
2016-12-31                                   27828571.0
2017-12-31                                   27406473.0
2018-12-31                                   28197847.0
2019-12-31                                   29878525.0
2020-12-31                                   29044998.0
```

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