

STL - KRZYSZTOF KOWALSKI

0. Wczytanie danych

```
In [1]: import pandas as pd
import matplotlib.pyplot as plt
from statsmodels.tsa.seasonal import STL
import matplotlib.pyplot as plt
import numpy as np

In [2]: dane = pd.read_csv('CO2_dataset.csv', index_col = 'Decimal Date')

In [3]: dane

Out[1]:
```

	Year	Month	Carbon Dioxide (ppm)	Seasonally Adjusted CO2 (ppm)	Carbon Dioxide Fit (ppm)	Seasonally Adjusted CO2 Fit (ppm)
Decimal Date						
1958.0411	1958	1	NaN	NaN	NaN	NaN
1958.1280	1958	2	NaN	NaN	NaN	NaN
1958.2027	1958	3	315.69	314.42	316.18	314.89
1958.2877	1958	4	317.45	315.15	317.30	314.98
1958.3699	1958	5	317.50	314.73	317.83	315.06
...
2017.8219	2017	8	NaN	NaN	NaN	NaN
2017.7068	2017	9	NaN	NaN	NaN	NaN
2017.7890	2017	10	NaN	NaN	NaN	NaN
2017.8740	2017	11	NaN	NaN	NaN	NaN
2017.9562	2017	12	NaN	NaN	NaN	NaN

720 rows x 6 columns

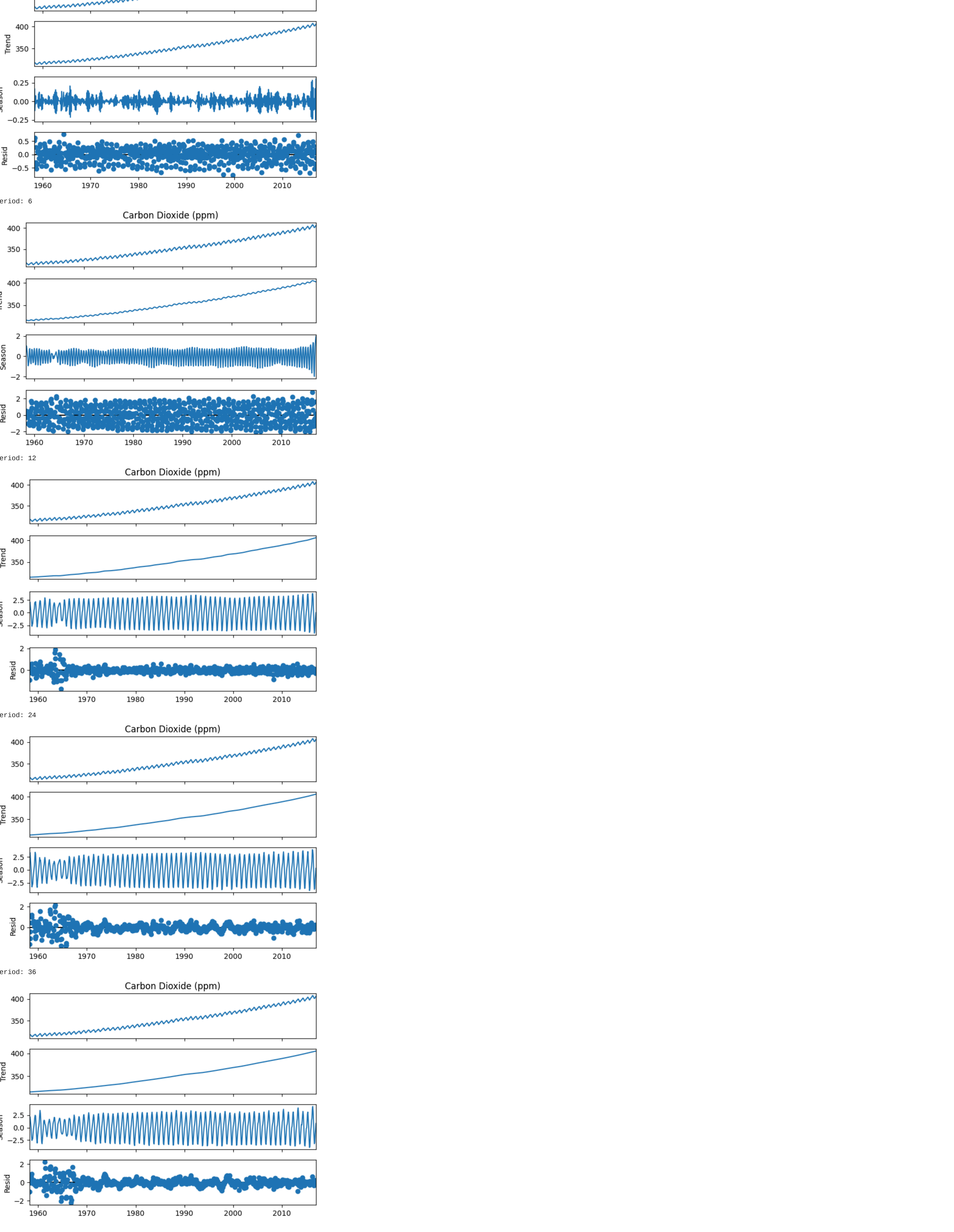
1. Przeprowadź dekompozycję szeregu czasowego za pomocą metody STL, korzystając z modułu dostępnego w bibliotece statsmodels.

```
In [4]: dane = dane.dropna()

stl = STL(dane['Carbon Dioxide (ppm)'], period = 12)
res = stl.fit()
fig = res.plot()
plt.show()
```

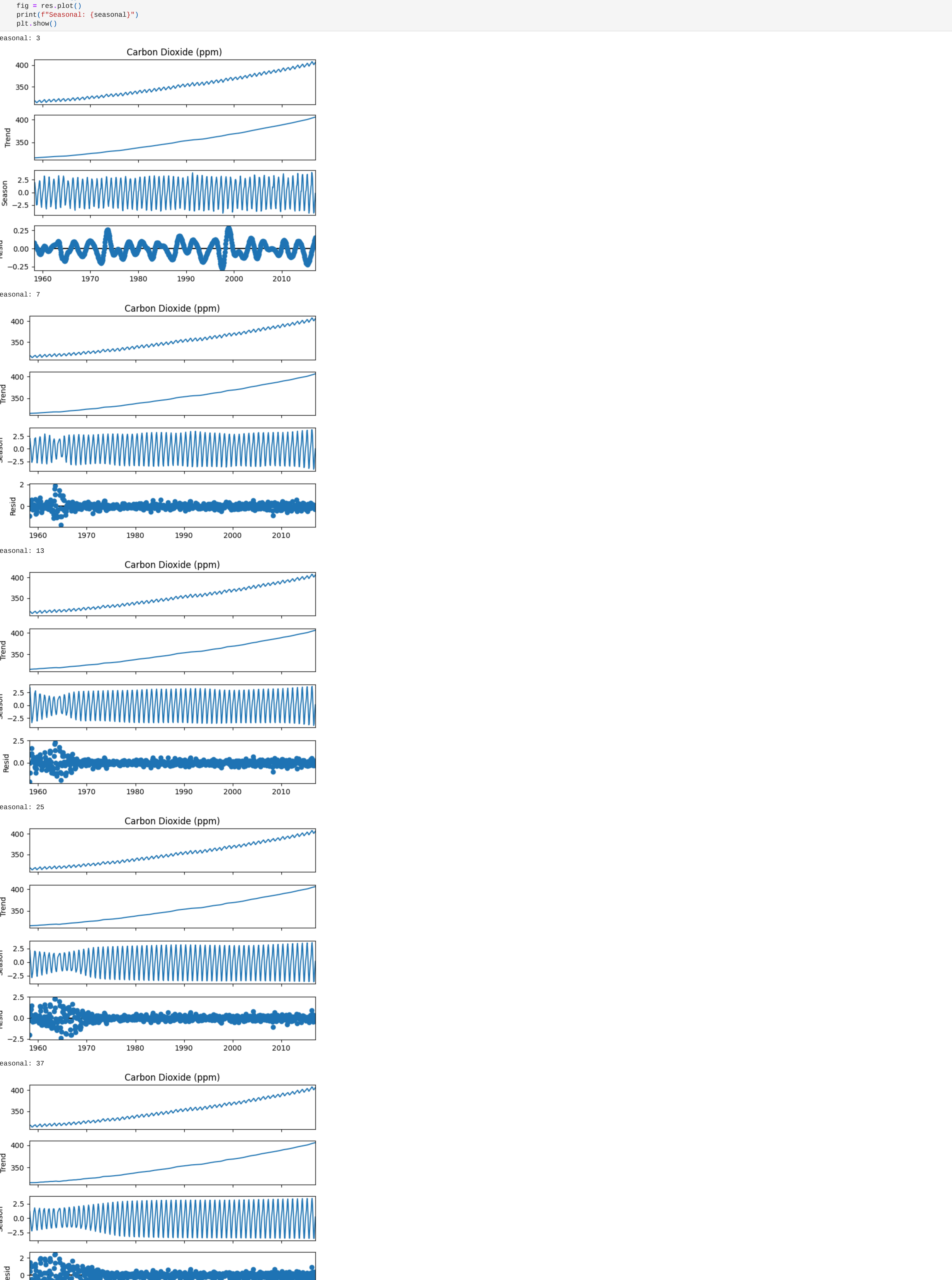
2. Przeprowadź testy dla różnych wartości parametrów period i seasonal, a następnie spis swoje obserwacje z przeprowadzonych badań.

```
In [5]: periods = [2, 6, 12, 24, 36]
for period in periods:
    stl = STL(dane['Carbon Dioxide (ppm)'], period = period)
    res = stl.fit()
    fig = res.plot()
    print(f"Period: {period}")
    plt.show()
```



Wartości parametru seasonal również wpływa na wykresy rezydów i sezonowości. Ogromne zaburzenia są dla małych wartości seasonal, optymalna wydaje się być wartość 7.

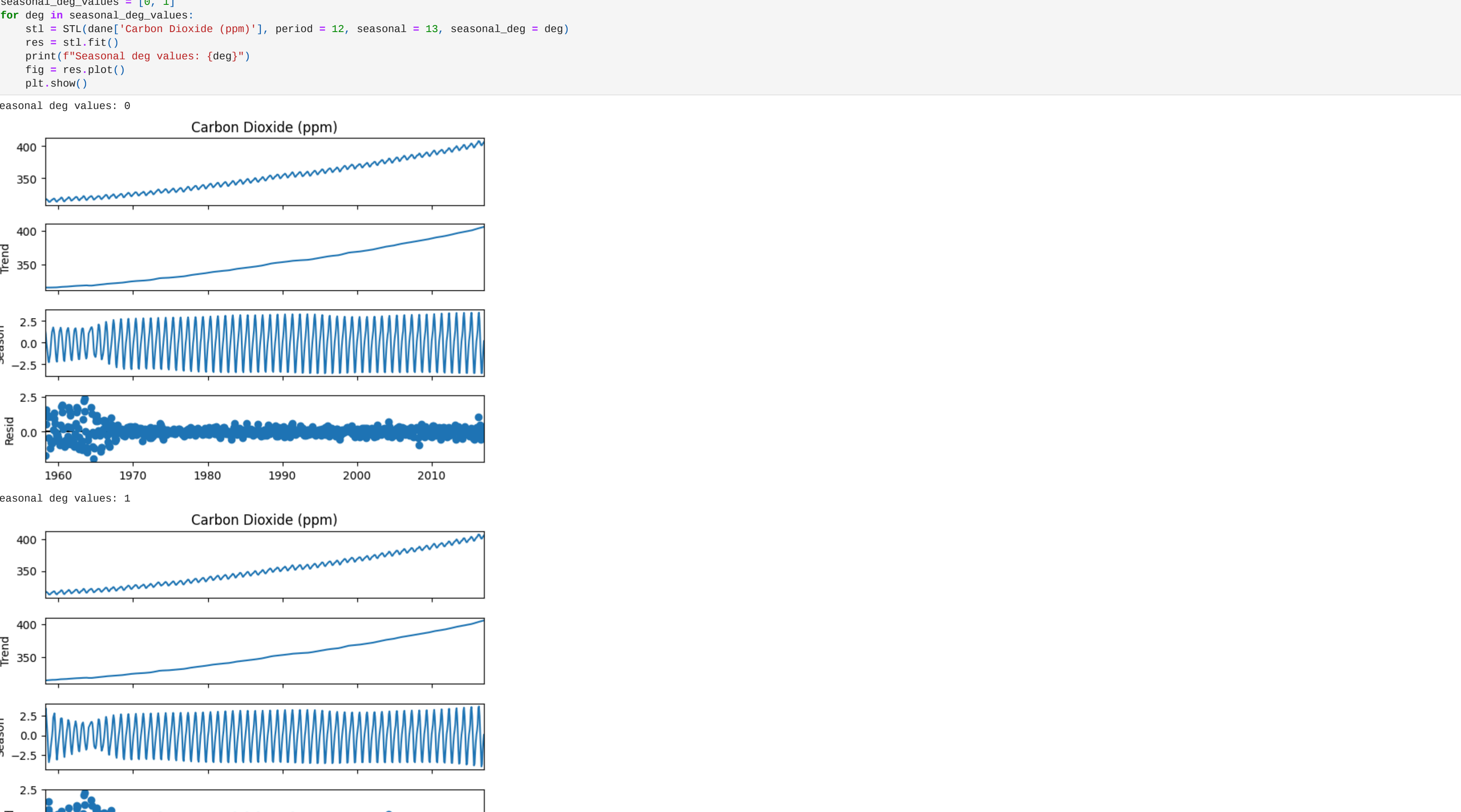
```
In [6]: seasonals = [3, 7, 13, 25, 37]
for seasonal in seasonals:
    stl = STL(dane['Carbon Dioxide (ppm)'], period = 12, seasonal = seasonal)
    res = stl.fit()
    fig = res.plot()
    plt.show()
```



3. Zbadaj różnice w magnitudzie sezonowości pomiędzy stałą a zmienną w czasie (parametr seasonal_deg równy 0 oznacza stałą magnitudę sezonową w czasie). Przedstaw opis wyników swoich badań.

Różnica dla stałej a zmiennej magnitudzie w czasie jest znikoma. Widoczna jest różnica dla wykresu sezonowości (szczególnie na jego początku).

```
In [7]: seasonal_deg_values = [0, 1]
for deg in seasonal_deg_values:
    stl = STL(dane['Carbon Dioxide (ppm)'], period = 12, seasonal = 13, seasonal_deg = deg)
    res = stl.fit()
    print(f"Seasonal deg values: {deg}")
    fig = res.plot()
    plt.show()
```



4. Na podstawie wykonanych ćwiczeń z dekompozycji tradycyjnej oraz punktów 1-3 dzisiejszego ćwiczenia, zaleć, którą dekompozycję i dlaczego warto wykorzystać do generowania cech do modelu uczenia maszynowego.

Lepiej wydaje się skorzystanie z stl, zamiast tradycyjnej metody dekompozycji. Pozwala ona lepiej radzić sobie z wartościami odstającymi, co wpływa na sezonowość oraz rezydów. Ustawienie parametru odpowiadającego za zmienną magnitudę w czasie pozwala na lepsze dopasowanie się do danych. W przypadku występowania wartości brakujących lepszym rozwiązaniem będzie skorzystanie z tradycyjnych metod dekompozycji.

5. Dokonaj interpretacji trendu, sezonowości oraz reszt dla dekompozycji, która Twoim zdaniem przynosi najbardziej optymalne wyniki.

Basując na powyższych uzasadnieniach postanowiłem na wybór parametrów widocznych poniżej. Uważam, że najlepiej dopasowują się do analizowanego zbioru.

```
In [8]: stl = STL(dane['Carbon Dioxide (ppm)'], period = 12, seasonal = 7, seasonal_deg = 1)
res = stl.fit()
fig = res.plot()
plt.show()
```

