

# SPRAWOZDANIE

Zajęcia: Nauka o danych I

Prowadzący: prof. dr hab. Vasyl Martsenyuk

Laboratorium Nr 9 Data 25.01.2025 Temat: Wprowadzenie do analizy czasowych szeregów danych, projektowanie eksperymentów i test hipotez. Wariant drugi (2)	Bartosz Bieniek Informatyka II stopień, stacjonarne, 1 semestr, gr.A
--	---

## 1. Polecenie: Wariant drugi zadania

2. Przeanalizuj dane o miesięcznej produkcji energii elektrycznej w USA:

- Pobierz dane z [Time Series Datasets](#).
- Oblicz średnie ruchome dla okien 6 i 12 miesięcy.
- Przeprowadź analizę ACF i PACF.
- Zdekomponuj dane na składniki trendu, sezonowości i reszt.

## 2. Opis programu opracowanego [[Kod Źródłowy github.com/mindgoner](https://github.com/mindgoner)]

```
import pandas as pd
import matplotlib.pyplot as plt
from statsmodels.graphics.tsaplots import plot_acf, plot_pacf
from statsmodels.tsa.seasonal import seasonal_decompose

# Wczytanie danych
dane = pd.read_csv('dane.csv', parse_dates=['DATE'], index_col='DATE')
dane.columns = ['wartosci']
```

[2] ✓ 0.0s

Rys. 1. Wczytanie wymaganych bibliotek oraz załadowanie danych.

Standardowym krokiem w wykonywaniu programów do analizy danych jest przygotowanie wymaganych bibliotek oraz załadowanie danych. Dane zostały załadowane do zmiennej “dane”.

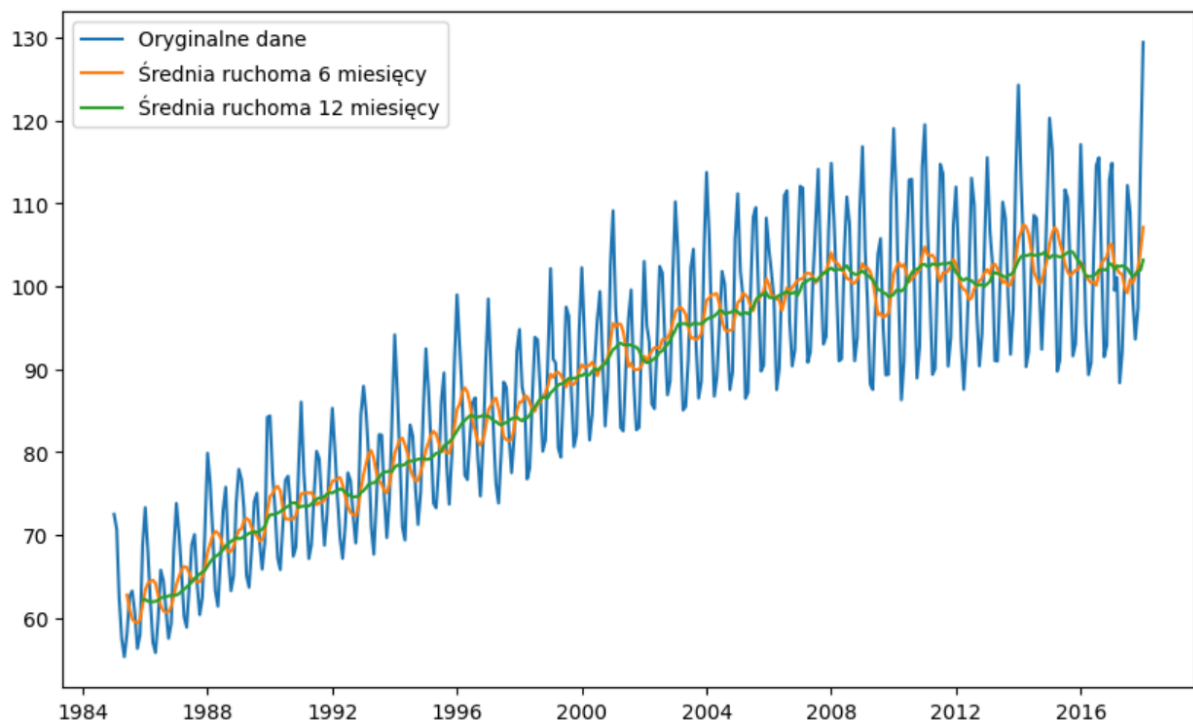
```
# 1. Średnie ruchome
dane['srednia_6'] = dane['wartosci'].rolling(window=6).mean()
dane['srednia_12'] = dane['wartosci'].rolling(window=12).mean()

plt.figure(figsize=(10, 6))
plt.plot(dane['wartosci'], label='Oryginalne dane')
plt.plot(dane['srednia_6'], label='Średnia ruchoma 6 miesięcy')
plt.plot(dane['srednia_12'], label='Średnia ruchoma 12 miesięcy')
plt.legend()
plt.show()
```

[3] ✓ 0.1s

Rys. 2. Średnie ruchome

W pierwszym kroku przetwarzania danych, przy pomocy metody rolling z argumentem window=6 i window=12, uzyskano średnie ruchome dla danych z danego okresu. Wartości oznaczono oraz wrysowano biblioteką pyplot.



Rys. 3. Średnie ruchome - wizualizacja

W kolejnym kroku dokonano analizy ACF i PACF, odrzucając wartości puste, które mogłyby powodować błędy podczas plotowania.

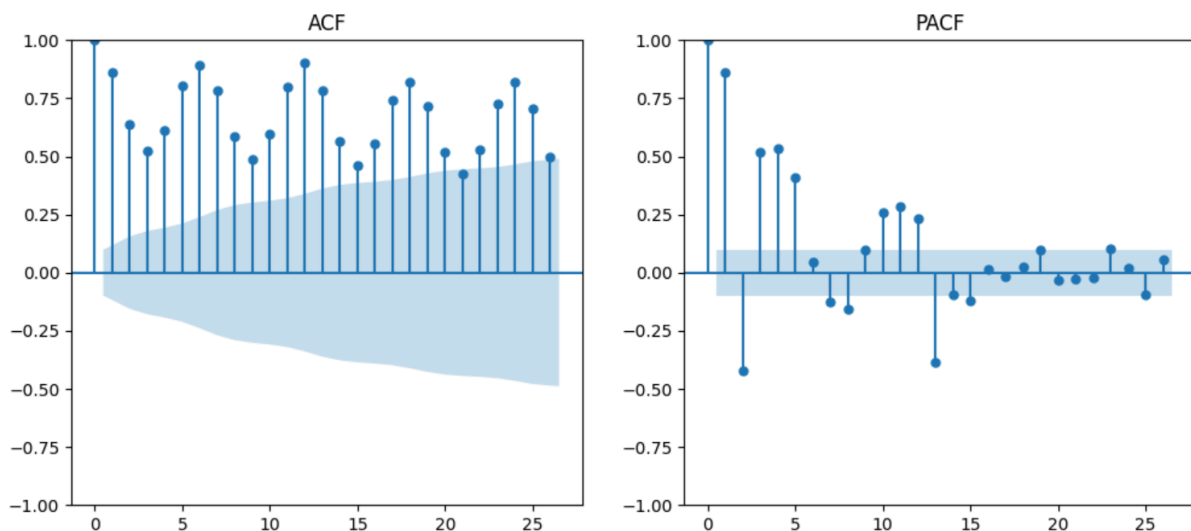
```
# 2. Analiza ACF i PACF
plt.figure(figsize=(12, 5))
plt.subplot(1, 2, 1)
plot_acf(dane['wartosci'].dropna(), ax=plt.gca())
plt.title('ACF')

plt.subplot(1, 2, 2)
plot_pacf(dane['wartosci'].dropna(), ax=plt.gca(), method='ywmm')
plt.title('PACF')
plt.show()
```

[4] ✓ 0.1s

Rys. 4. Analiza ACF i PACF

Wynik wykonania powyższego kodu przedstawia rysunek piąty. Z jego analizy możemy wywnioskować, że wykres ACF pokazuje silne korelacje dla wielu opóźnień. To sugeruje, że dane mają silny składnik autoregresyjny (AR). Potwierdzeniem istnienia silnego składnika AR jest brak silnego zanikania (korelacje nie maleją szybko do zera).



Rys. 5. Wizualizacja ACF i PACF.

Z powyższego wykresu można zauważyć, że PACF szybko spada do zera, co sugeruje, że składnik średniej ruchomej “MA” jest niewielki. Ponadto najsilniejsza korelacja występuje w pierwszym opóźnieniu, wskazując na obecność wyżej wymienionego składnika “MA”.

```

# 3. Dekompozycja danych
dekompozycja = seasonal_decompose(dane['wartosci'], model='additive', period=12)

plt.figure(figsize=(10, 8))
dekompozycja.plot()
plt.show()

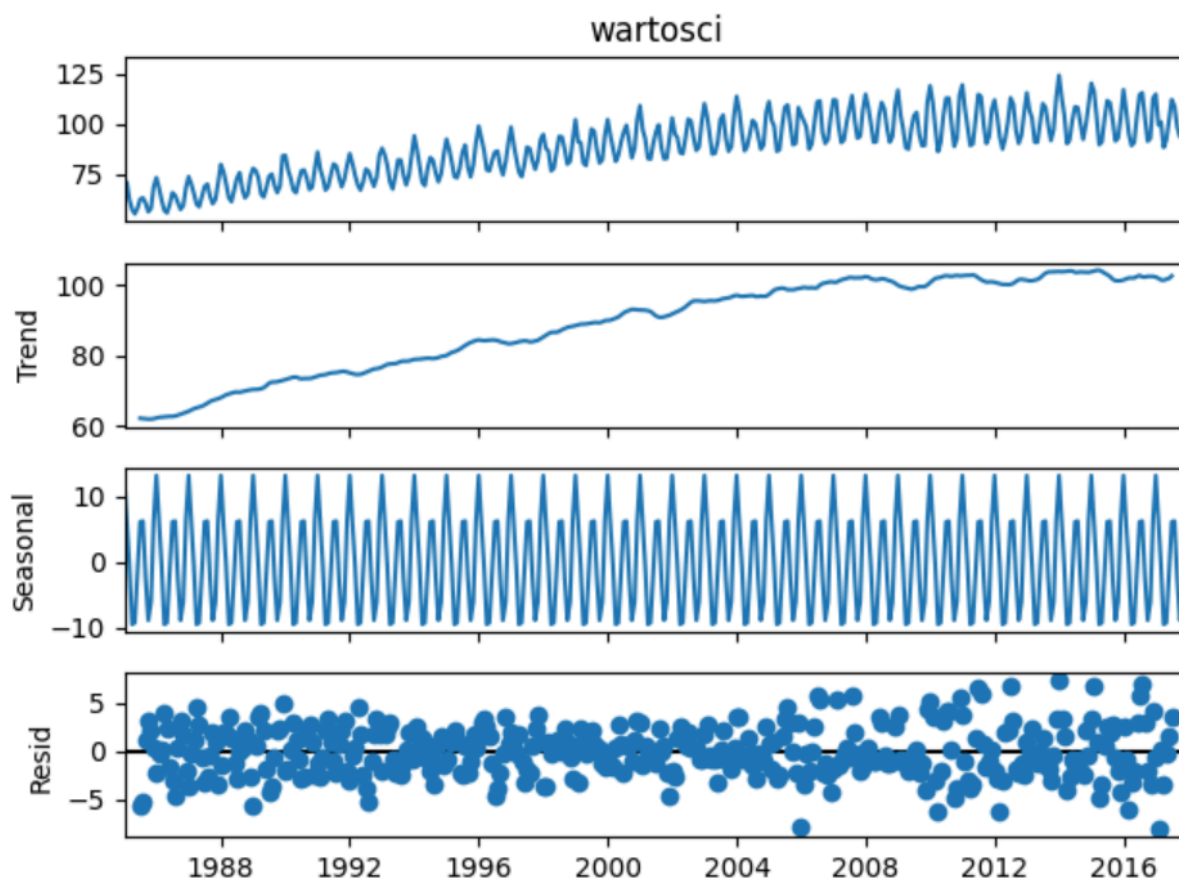
```

[5] ✓ 0.3s

Rys. 6. Dekompozycja danych.

Dekompozycja danych szeregów czasowych jest kluczową techniką prognozowania i analizy. Pomaga wykonać szereg czasowy na podstawowych elementach, ujawniając podstawowe wzorce i trendy.

Wynik dekompozycji przedstawia rysunek siódmy.



*Rys. 7. Dekompozycja danych - wizualizacja*

Z analizy powyższego rysunku można wywnioskować, że dekompozycja wykazuje wyraźną sezonowość z cyklem rocznym, co widać na wykresie "Seasonal". Zauważalny jest również trend wzrostowy "Trend", który pokrywa się z prawdą - produkcja energii przez USA wzrastała na przestrzeni lat.

Residua są rozproszone wokół zera, co sugeruje, że model uwzględnił większość zmienności danych. Pomimo wyraźnego rozrzucenia danych po roku 2000.

### 3. Wnioski

Analiza wykresów ACF i PACF sugeruje, że dane prawdopodobnie dobrze opisuje model autoregresyjny (AR) z potencjalnym niewielkim składnikiem średniej ruchomej (MA).

Dokonując dekompozycji danych otrzymujemy informacje o sezonowości, trendzie i reszt danych.