

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 testowanie hipotez w praktyce” Wariant 10	Anna Więzik Informatyka II stopień, niestacjonarne, 1 semestr, gr.1b
--	---

1. Polecenie:

10. Przeanalizuj dane o dziennej temperaturze w wybranych miastach:

- Pobierz dane z Daily Temperature of Major Cities.
- Oblicz średnie ruchome dla okien 7 i 30 dni.
- Przeprowadź analizę ACF i PACF.
- Zdekomponuj dane na składniki trendu, sezonowości i reszt.

Link do repozytorium: https://github.com/AnaShiro/NoD1_2024

2. Opis programu opracowanego

- Pobierz dane z Daily Temperature of Major Cities.
- Oblicz średnie ruchome dla okien 7 i 30 dni.

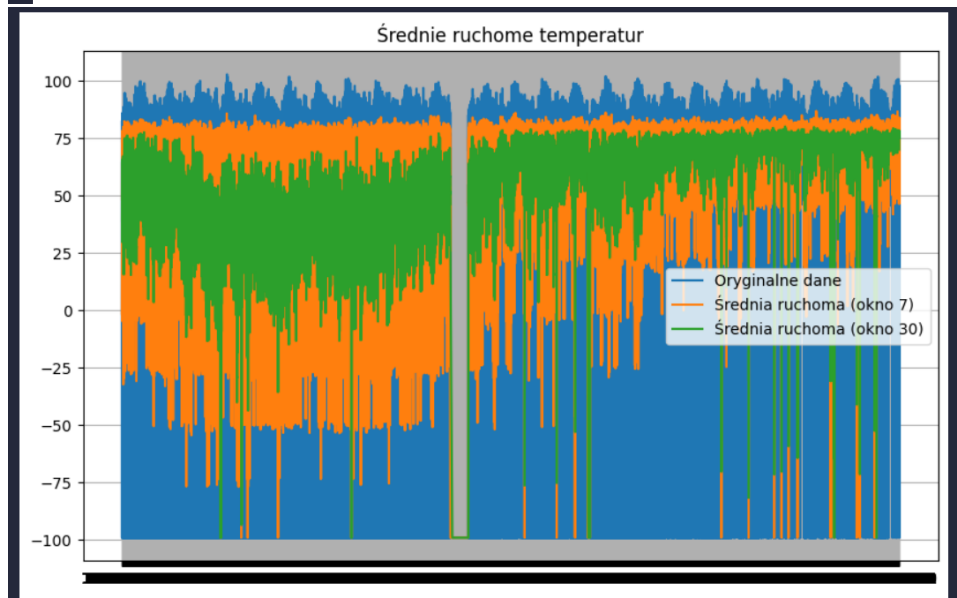
```
import matplotlib.pyplot as plt    Import "matplotlib.pyplot" could not be resolved from source
import pandas as pd               Import "pandas" could not be resolved from source

data = pd.read_csv('city_temperature.csv', parse_dates=['Date': ['Year', 'Month', 'Day']])

# Sortowanie danych po dacie
data = data.sort_values(by='Date')
data.set_index('Date', inplace=True)

# Obliczanie średnich ruchomych
data['MA_7'] = data['AvgTemperature'].rolling(window=7).mean()
data['MA_30'] = data['AvgTemperature'].rolling(window=30).mean()

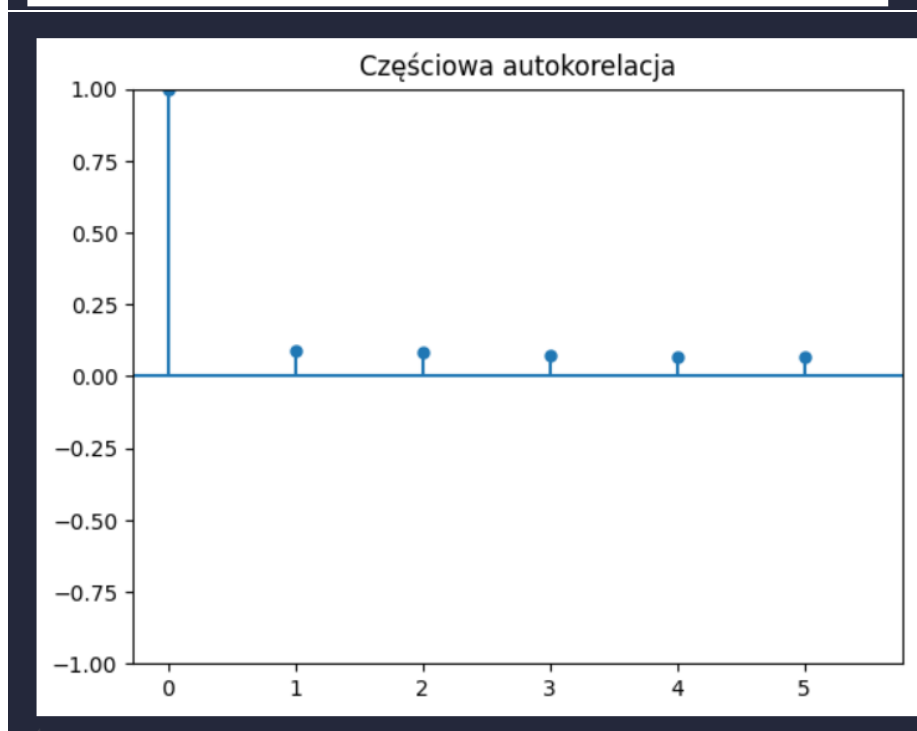
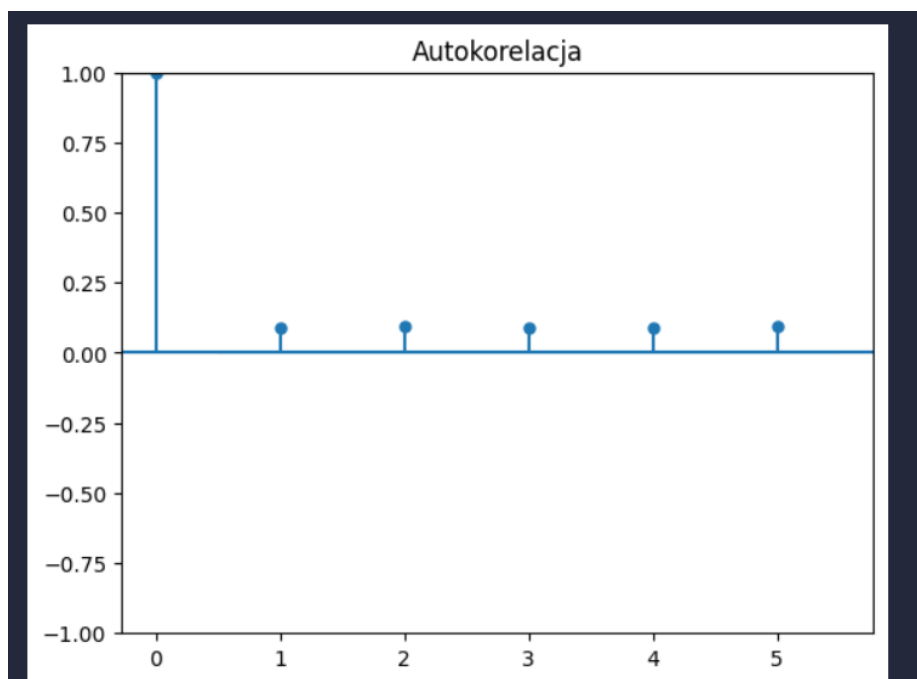
# Wizualizacja
plt.figure(figsize=(10, 6))
plt.plot(data['AvgTemperature'], Label="Oryginalne dane")
plt.plot(data['MA_7'], Label="Średnia ruchoma (okno 7)")
plt.plot(data['MA_30'], Label="Średnia ruchoma (okno 30)")
plt.title("Średnie ruchome temperatur")
plt.legend()
plt.grid()
plt.show()
✓ 2m 6.7s
```



- Przeprowadź analizę ACF i PACF.

```
from statsmodels.graphics.tsaplots import plot_acf, plot_pacf    Import "statsmodels.graphics

# ACF i PACF
plot_acf(data['AvgTemperature'], lags=5, title="Autokorelacja")    ValueError: could not conv
plot_pacf(data['AvgTemperature'], lags=5, title="Częściowa autokorelacja", method='ywmm')
plt.show()
✓ 49.2s
```



- Zdekomponuj dane na składniki trendu, sezonowości i reszt.



3. Wnioski

Analiza szeregów czasowych pozwala na dekompozycję danych na składniki takie jak trend, sezonowość i reszty, co umożliwia lepsze zrozumienie ich struktury oraz identyfikację wzorców. Wykorzystanie funkcji autokorelacji (ACF) jest kluczowe do wykrywania sezonowości i trendów, podczas gdy funkcja PACF pomaga określić istotne opóźnienia przydatne w modelach autoregresyjnych. Średnie ruchome są niezastąpionym narzędziem do wygładzania danych, eliminacji szumów oraz identyfikacji trendów i krótkoterminowego prognozowania. Projektowanie eksperymentów i testowanie hipotez statystycznych, w tym poprawne formułowanie hipotez zerowej i alternatywnej oraz wybór odpowiednich testów, takich jak test t-Studenta czy chi-kwadrat, pozwala na uzyskanie wiarygodnych wniosków z danych. Narzędzia takie jak Python, z bibliotekami takimi jak statsmodels i pandas, oraz środowisko KNIME, wspierają efektywną analizę i wizualizację danych. Te techniki znajdują szerokie zastosowanie w praktyce, między innymi w analizie sprzedaży, danych meteorologicznych, finansowych czy w wykrywaniu anomalii. Kluczowe jest także świadome dobranie parametrów analizy, takich jak długość okna w średnich ruchomych czy zakres opóźnień w ACF i PACF, co znacząco wpływa na jakość prognoz i wyciąganych wniosków.