

## 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 11	Szymon Nycz Informatyka II stopień, niestacjonarne, 1 semestr, gr.1b
--	---

### 1. Polecenie:

11. Wykonaj analize danych dotyczących dziennych danych klimatycznych w Delhi:

- Pobierz dane z Daily Climate Time Series Data.
- Oblicz średnie ruchome dla okien 7 i 30 dni.
- Przeprowadź analize ACF i PACF.
- Zdekomponuj dane na składniki trendu, sezonowości i reszt.

Link do repozytorium:

## 2. Opis programu opracowanego

- Pobierz dane z Daily Time Series Data
- Oblicz średnie ruchome dla okien 7 i 30 dni

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

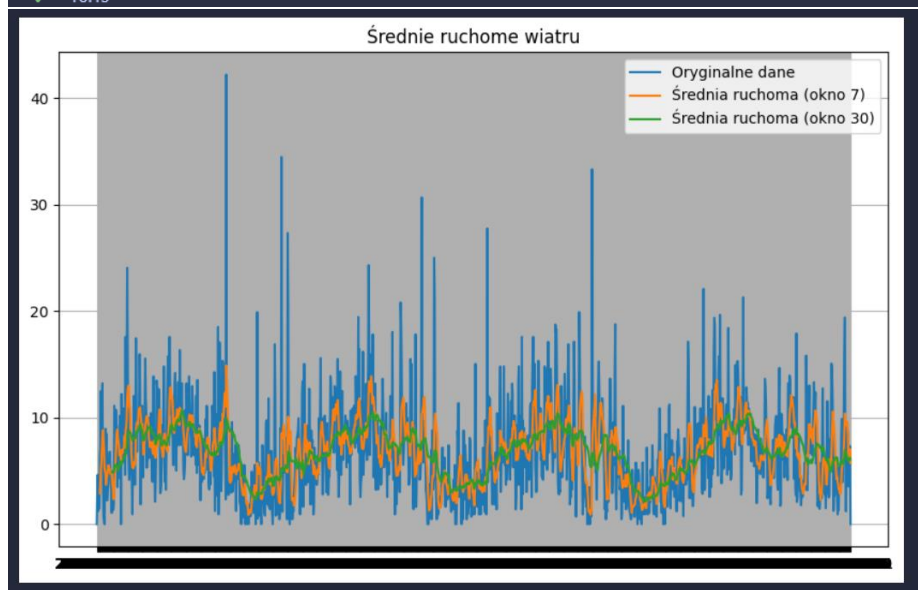
data = pd.read_csv('DailyDelhiClimateTrain.csv')

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

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

# Wizualizacja
plt.figure(figsize=(10, 6))
plt.plot(data['wind_speed'], 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 wiatru")
plt.legend()
plt.grid()
plt.show()

✓ 16.1s
```

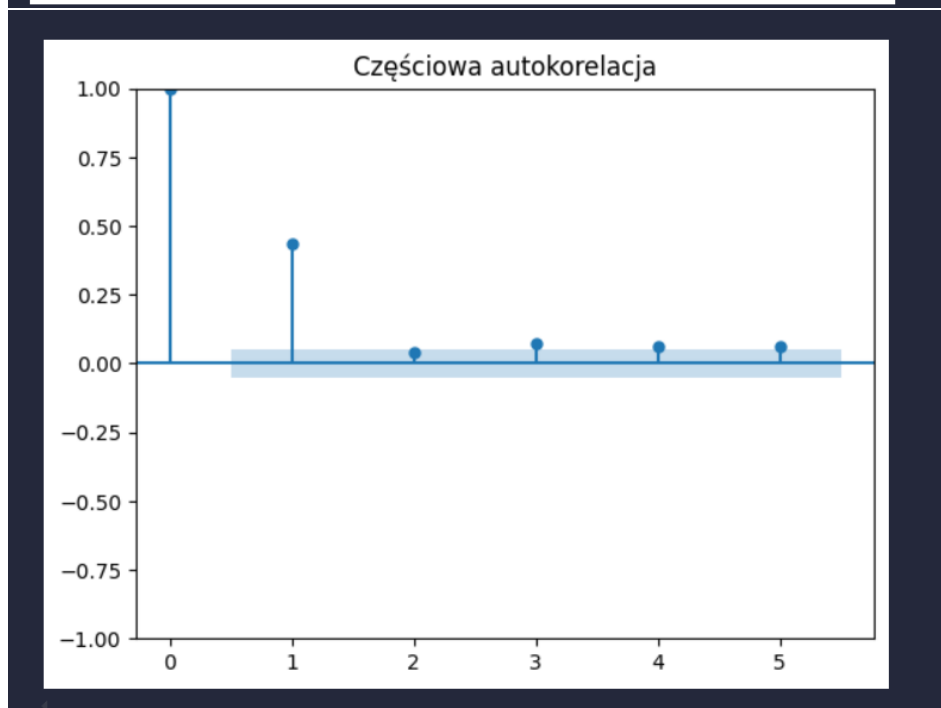
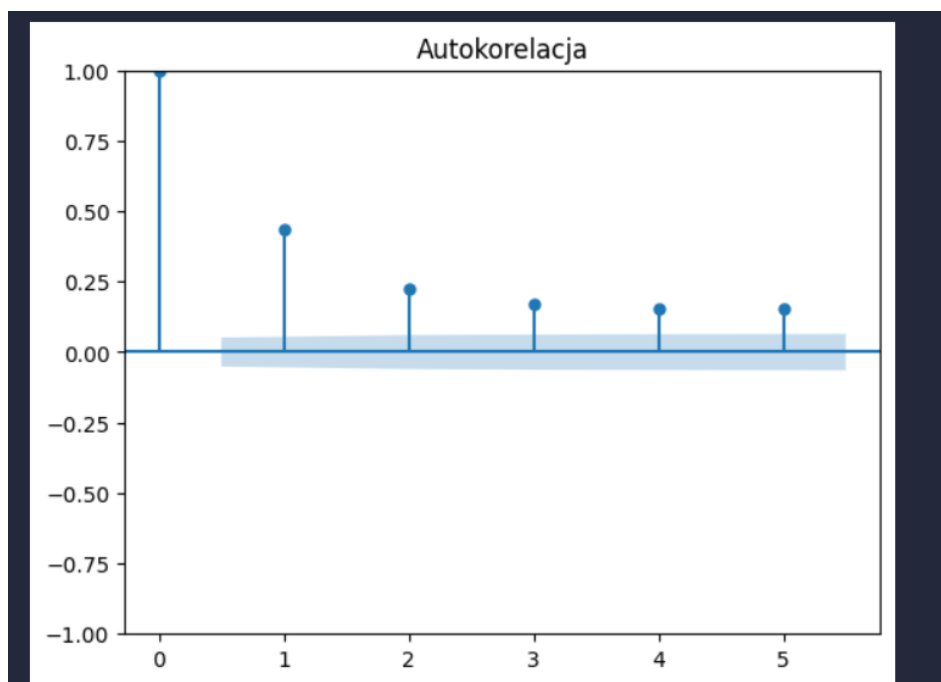


- Przeprowadź analizę ACF i PACF

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

# ACF i PACF
plot_acf(data['wind_speed'], lags=5, title="Autokorelacja")    ValueError: could not co
plot_pacf(data['wind_speed'], lags=5, title="Częściowa autokorelacja", method='ywm')
plt.show()

✓ 7.7s
```



- Zdekomponuj dane na składniki trendu, sezonowości i reszt



### 3. Wnioski

Analiza szeregów czasowych umożliwia rozłożenie danych na elementy takie jak trend, sezonowość i składniki losowe, co pozwala lepiej zrozumieć ich strukturę i odkryć występujące wzorce. Funkcja autokorelacji (ACF) jest szczególnie przydatna do identyfikowania trendów oraz sezonowości, natomiast funkcja PACF pomaga wskazać kluczowe opóźnienia istotne w modelowaniu autoregresyjnym. Wykorzystanie średnich ruchomych pozwala na wygładzanie danych, eliminację zakłóceń oraz lepsze dostrzeganie długoterminowych trendów, co ułatwia krótkoterminowe prognozowanie. Z kolei projektowanie eksperymentów i testowanie hipotez, obejmujące odpowiednie sformułowanie hipotez zerowej i alternatywnej oraz dobór właściwych testów statystycznych, takich jak test t-Studenta czy test chi-kwadrat, pozwala na rzetelną interpretację wyników. Do realizacji takich analiz doskonale sprawdzają się narzędzia takie jak Python, z jego bibliotekami `statsmodels` i `pandas`, oraz platforma KNIME. Techniki te mają szerokie zastosowanie praktyczne, między innymi w badaniach nad sprzedażą, danymi meteorologicznymi czy finansowymi, a także w identyfikowaniu anomalii. Ważne jest również precyzyjne dopasowanie parametrów analizy, na przykład długości okien w średnich ruchomych czy zakresu opóźnień w ACF i PACF, co ma znaczący wpływ na trafność prognoz i wiarygodność wyciąganych wniosków.