

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 8	Imię Nazwisko Hubert Mentel Informatyka II stopień, niestacjonarne, 1 semestr, gr.1a
--	---

1. Zadanie:

Celem ćwiczenia jest zapoznanie się z podstawowymi technikami projektowania eksperymentów oraz metodami testowania hipotez statystycznych. Studenci nauczą się:

- Projektować eksperymenty w celu zbierania danych odpowiednich do analizy statystycznej,
- Formułować hipotezy statystyczne,
- Wykonywać testy statystyczne z wykorzystaniem języka Python oraz środowiska KNIME,
- Interpretować wyniki testów i wyciągać wnioski.

Wariant 8:

Przeanalizuj dane o dziennych danych finansowych:

- Pobierz dane z Financial Time Series Datasets.
- Oblicz Średnie ruchome dla okien 7 i 30 dni.
- Przeprowadź analizę ACF i PACF.
- Zakomponuj dane na składniki trendu, sezonowości i reszt.

Pliki dostępne są na GitHubie pod linkiem:

<https://github.com/HubiPX/NOD/tree/master/Zadanie%209>

2. Opis programu opracowanego (kody Źródłowe, zrzuty ekranu)

```
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 z pliku CSV
file_path = 'AMZN.csv'
data = pd.read_csv(file_path)

# Konwersja kolumny 'Date' na datetime
data['Date'] = pd.to_datetime(data['Date'])

# Ustawienie daty jako indeks
data.set_index('Date', inplace=True)
```

```
# Wyświetlenie pierwszych kilku wierszy
```

```
print(data.head())
```

```
# Obliczanie Średnich ruchomych (7 i 30 dni)
```

```
data['MA_7'] = data['Adj Close'].rolling(window=7).mean()
```

```
data['MA_30'] = data['Adj Close'].rolling(window=30).mean()
```

```
# Wizualizacja wykresu cen i Średnich ruchomych
```

```
plt.figure(figsize=(10, 6))
```

```
plt.plot(data['Adj Close'], label='Cena skorygowana')
```

```
plt.plot(data['MA_7'], label='Średnia ruchoma 7 dni')
```

```
plt.plot(data['MA_30'], label='Średnia ruchoma 30 dni')
```

```
plt.legend()
```

```
plt.title('Średnie ruchome dla AMZN')
```

```
plt.grid(True)
```

```
plt.show()
```

```
# Rysowanie wykresów ACF i PACF
```

```
plt.figure(figsize=(12, 6))
```

```
plot_acf(data['Adj Close'].dropna(), lags=30, ax=plt.gca())
```

```
plt.title("Autokorelacja (ACF)")
```

```
plt.show()
```

```
plt.figure(figsize=(12, 6))
```

```
plot_pacf(data['Adj Close'].dropna(), lags=30, ax=plt.gca(), method='ywm')
```

```
plt.title("Częściowa autokorelacja (PACF)")
```

```
plt.show()
```

```
# Dekompozycja szeregów czasowych
```

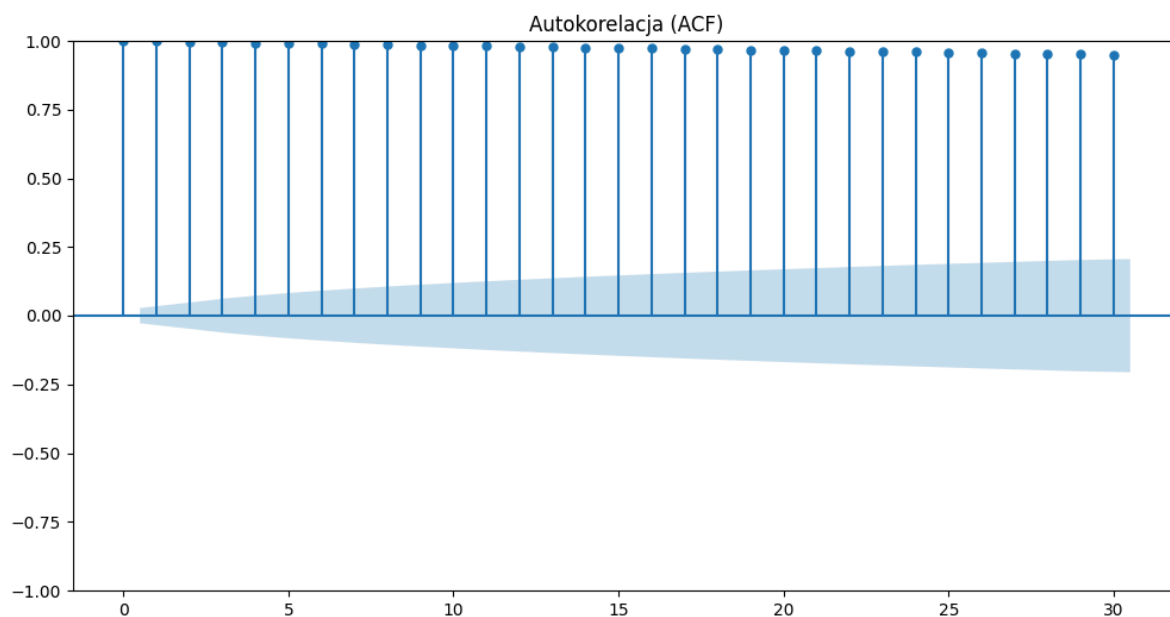
```
result = seasonal_decompose(data['Adj Close'].dropna(), model='additive',  
period=7) # Okres 7 dni
```

```
result.plot()
```

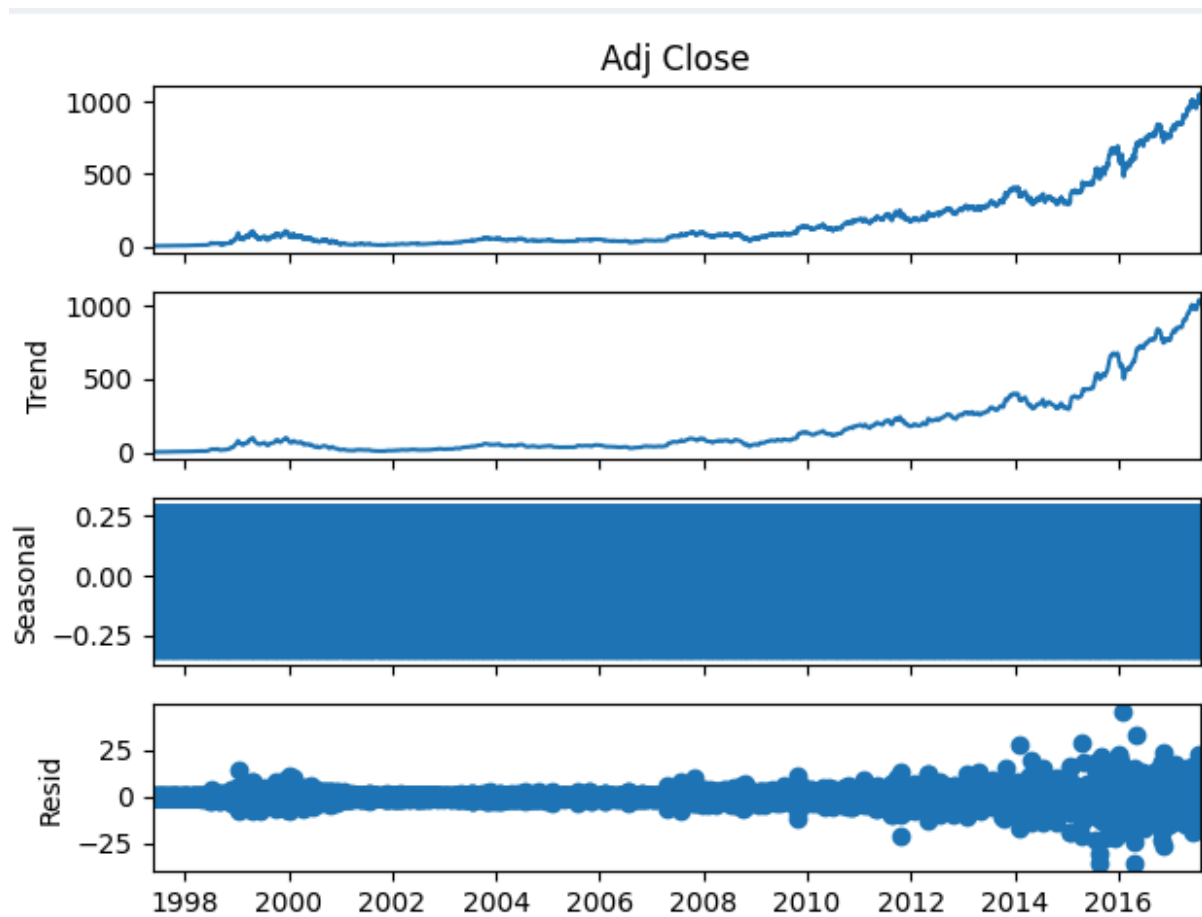
```
plt.show()
```



Wykres ten pokazuje, że cena akcji AMZN ma tendencję wzrostową, która jest wygładzona bardziej przez średnią ruchomą 30-dniową. Średnia ruchoma 7-dniowa wykazuje bardziej "szumowe" zmiany, ale również śledzi ogólny wzrost cen. Obie średnie ruchome podążają za wzrostem ceny, ale 30-dniowa robi to w sposób mniej wrażliwy na krótkoterminowe fluktuacje.



Wykres ACF pokazuje, że wartości szeregu czasowego (ceny akcji) są silnie skorelowane z wcześniejszymi obserwacjami, szczególnie na małych opóźnieniach (lag). Z czasem autokorelacja maleje, co sugeruje, że wpływ przeszłych danych na bieżące staje się mniejszy w miarę jak patrzymy na odleglejsze opóźnienia. Jest to typowe dla danych finansowych, gdzie krótkoterminowe zależności mogą być silne, ale długoterminowe zależności słabną.



PACF to wykres pokazujący, jak wartości szeregu czasowego są skorelowane z wcześniejszymi wartościami, ale z uwzględnieniem eliminacji wpływu pośrednich opóźnień. Innymi słowy, PACF mierzy zależność po usunięciu wpływu innych opóźnień.

3. Wnioski

W celu analizy dziennych danych finansowych dla akcji Amazon (AMZN), w programie zastosowałem kilka technik szeregów czasowych, takich jak obliczanie średnich ruchomych, analiza autokorelacji (ACF), częściowej autokorelacji (PACF) oraz dekompozycja szeregów czasowych. Na początek obliczyłem średnie ruchome dla okien 7-dniowych i 30-dniowych, które zostały użyte do wygładzenia danych i uzyskania lepszego obrazu ogólnych trendów cenowych. Wykres średnich ruchomych oraz cen skorygowanych ujawnia wzrostową tendencję cen akcji AMZN, z widocznymi fluktuacjami, które są wygładzane przez średnie ruchome o różnej długości okresu. Warto zauważyć, że średnia ruchoma 30-dniowa bardziej wygładza zmiany, co pozwala dostrzec długoterminowy trend wzrostu.

Kolejnymi ważnymi elementami analizy były wykresy autokorelacji (ACF) i częściowej autokorelacji (PACF), które umożliwiają ocenę zależności między bieżącymi i przeszłymi wartościami szeregu czasowego. Wykres ACF pokazał, że ceny akcji są silnie skorelowane z wartościami z przeszłości, szczególnie na mniejszych opóźnieniach, co wskazuje na istotne zależności czasowe w krótkim okresie. Natomiast wykres PACF ujawnił, które opóźnienia mają największy wpływ na dane, pomagając określić odpowiednią liczbę opóźnień, które należy uwzględnić w modelach prognozujących. Na koniec, dekompozycja szeregów czasowych rozbiła dane na składniki trendu, sezonowości i reszt, co pozwoliło na dokładniejsze zrozumienie struktury danych i identyfikację potencjalnych sezonowych fluktuacji.