

Laboratorium Pandas III

Zadanie 1: Obliczanie miar położenia

Utwórz ramkę danych zawierającą informacje o wynikach egzaminów studentów i oblicz średnią arytmetyczną, medianę i modę dla każdej kolumny.

```
import pandas as pd

# Utwórz ramkę danych
dane_studentow = pd.DataFrame({
    'Student': ['Anna', 'Bartek', 'Celina', 'Dawid', 'Ewa', 'Filip',
                'Gosia', 'Hubert'],
    'Matematyka': [85, 90, 78, 92, 88, 76, 94, 85],
    'Fizyka': [76, 88, 90, 75, 82, 84, 91, 76],
    'Informatyka': [92, 85, 88, 95, 80, 75, 89, 92]
})

# Oblicz średnią, medianę i modę dla kolumn z ocenami
# Uzupełnij kod
```

Zadanie 2: Porównanie miar położenia

Stwórz dwie serie danych o różnych rozkładach i porównaj dla nich średnią, medianę i dominantę.

```
import pandas as pd

# Utwórz dwie serie danych
dane_symetryczne = pd.Series([10, 20, 30, 40, 50, 60, 70, 80, 90])
dane_skosne = pd.Series([10, 15, 18, 20, 22, 25, 30, 80, 90])

# Oblicz i porównaj miary położenia
# Uzupełnij kod
```

Zadanie 3: Analiza kwartyli

Stwórz ramkę danych z informacjami o cenach produktów i oblicz kwartyle dla tych cen.

```
import pandas as pd

# Utwórz ramkę danych z cenami produktów
ceny_produkow = pd.DataFrame({
    'Produkt': ['A', 'B', 'C', 'D', 'E', 'F', 'G', 'H', 'I', 'J'],
    'Cena': [25, 35, 40, 45, 50, 55, 60, 75, 90, 120]
})
```

```
})  
  
# Oblicz kwartyle cen  
# Uzupełnij kod
```

Zadanie 4: Miary zmienności

Utwórz ramkę danych z wynikami pomiarów i oblicz różne miary zmienności: rozstęp, odchylenie standardowe i wariancję.

```
import pandas as pd  
  
# Utwórz ramkę danych z wynikami pomiarów  
pomiary = pd.DataFrame({  
    'Próbka': ['P1', 'P2', 'P3', 'P4', 'P5', 'P6', 'P7', 'P8'],  
    'Wynik1': [45, 47, 49, 51, 46, 48, 50, 52],  
    'Wynik2': [120, 150, 110, 190, 200, 115, 140, 180]  
})  
  
# Oblicz miary zmienności dla każdej kolumny z wynikami  
# Uzupełnij kod
```

Zadanie 5: Analiza asymetrii i koncentracji

Stwórz trzy różne serie danych (o asymetrii lewostronnej, symetrycznej i o asymetrii prawostronnej) i oblicz dla nich współczynniki asymetrii i kurtozy.

```
import pandas as pd  
import numpy as np  
  
# Utwórz trzy serie danych o różnych rozkładach  
dane_lewostronne = pd.Series([90, 85, 82, 81, 80, 79, 75, 70, 65, 50])  
dane_symetryczne = pd.Series([50, 60, 65, 70, 75, 80, 85, 90, 95, 100])  
dane_prawostronne = pd.Series([50, 55, 60, 65, 70, 75, 76, 78, 85, 95])  
  
# Oblicz współczynniki asymetrii i kurtozy  
# Uzupełnij kod
```

Zadanie 6: Zastosowanie funkcji describe()

Stwórz ramkę danych z danymi sprzedażowymi i zastosuj funkcję describe() do wygenerowania statystyk opisowych.

```
import pandas as pd  
  
# Utwórz ramkę danych z danymi sprzedażowymi  
dane_sprzedazowe = pd.DataFrame({
```

```

    'Produkt': ['A', 'B', 'C', 'A', 'B', 'C', 'A', 'B', 'C', 'A'],
    'Ilość': [120, 85, 90, 110, 95, 105, 130, 75, 85, 140],
    'Cena_jednostkowa': [10, 15, 12, 10, 15, 12, 10, 15, 12, 10],
    'Wartość': [1200, 1275, 1080, 1100, 1425, 1260, 1300, 1125, 1020,
1400]
})

# Wygeneruj statystyki opisowe
# Uzupełnij kod

```

Zadanie 7: Grupowanie i analiza statystyczna

Utwórz ramkę danych z informacjami o pracownikach z różnych działów i oblicz statystyki dla wynagrodzeń pogrupowanych według działów.

```

import pandas as pd

# Utwórz ramkę danych z informacjami o pracownikach
pracownicy = pd.DataFrame({
    'Imię': ['Anna', 'Bartek', 'Celina', 'Dawid', 'Ewa', 'Filip',
    'Gosia', 'Hubert', 'Iza', 'Jan'],
    'Dział': ['IT', 'Marketing', 'HR', 'IT', 'Marketing', 'HR', 'IT',
    'Marketing', 'HR', 'IT'],
    'Wynagrodzenie': [7500, 6800, 5900, 8200, 7000, 6100, 7800, 6500,
6000, 8500],
    'Staż_lat': [3, 5, 2, 7, 4, 1, 6, 3, 2, 8]
})

# Wykonaj grupowanie według działów i oblicz statystyki dla wynagrodzeń
# Uzupełnij kod

```

Zadanie 8: Funkcja agg() i różne miary

Utwórz ramkę danych z wynikami sprzedaży i zastosuj funkcję agg() do obliczenia różnych miar dla wybranych kolumn.

```

import pandas as pd

# Utwórz ramkę danych z wynikami sprzedaży
sprzedaz = pd.DataFrame({
    'Produkt': ['A', 'B', 'C', 'A', 'B', 'C', 'A', 'B'],
    'Region': ['Wschód', 'Wschód', 'Wschód', 'Zachód', 'Zachód',
    'Zachód', 'Północ', 'Północ'],
    'Ilość': [120, 85, 90, 110, 95, 105, 130, 75],
    'Wartość': [1200, 1275, 1080, 1100, 1425, 1260, 1300, 1125]
})

```

```
# Zastosuj funkcję agg() do obliczenia różnych miar
# Uzupełnij kod
```

Zadanie 9: Wykrywanie wartości odstających

Stwórz ramkę danych z wynikami pomiarów i wykryj wartości odstające za pomocą reguły opartej na rozstępie międzykwartylowym.

```
import pandas as pd
import numpy as np

# Utwórz ramkę danych z wynikami pomiarów, zawierającą wartości
odstające
pomiar = pd.DataFrame({
    'Pomiar': ['P1', 'P2', 'P3', 'P4', 'P5', 'P6', 'P7', 'P8', 'P9',
              'P10'],
    'Wartość': [120, 125, 118, 122, 130, 200, 119, 124, 50, 127]
})

# Zidentyfikuj wartości odstające używając reguły IQR (1.5 * rozstęp
międzykwartylowy)
# Uzupełnij kod
```

Zadanie 10: Kompletna analiza danych

Utwórz ramkę danych z kompleksowymi informacjami o produktach i przeprowadź pełną analizę statystyczną, obejmującą miary położenia, zmienności, asymetrii i koncentracji.

```
import pandas as pd

# Utwórz ramkę danych z informacjami o produktach
produkty = pd.DataFrame({
    'ID': range(1, 11),
    'Kategoria': ['A', 'B', 'A', 'C', 'B', 'A', 'C', 'B', 'A', 'C'],
    'Cena': [120, 85, 110, 150, 95, 130, 200, 90, 125, 175],
    'Ilość_sprzedana': [450, 320, 380, 290, 410, 360, 240, 350, 400,
                       270],
    'Ocena_klientów': [4.5, 3.8, 4.2, 4.7, 3.9, 4.3, 4.8, 4.0, 4.4, 4.6]
})

# Przeprowadź kompletną analizę statystyczną
# Uzupełnij kod
```