



Міністерство освіти і науки України
Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»
Факультет інформатики та обчислювальної техніки
Кафедра інформаційних систем та технологій

Лабораторна робота №1
з дисципліни
Аналіз даних з використанням мови Python

Виконав:

студент(ка) групи ІА-24:
Криворучек В.С.

Перевірила:

ст. викладач
Тимофєєва Ю.С.

Тема: Базове знайомство з бібліотеками NumPy та Pandas

Мета роботи: Ознайомитись з основними можливостями роботи з масивами бібліотеки NumPy, структурами даних Pandas, визначити основні статистичні характеристики.

Хід роботи

Завдання 1. Створюємо програму, яка:

- а) генерує випадкові і не випадкові масиви різними способами, зазначеними в теоретичних відомостях.
- б) демонструє звернення до елементів масиву за допомогою індексів, в тому числі від'ємних; виділення підмасивів як одновимірних, так і багатовимірних масивів.
- в) демонструє основні арифметичні операції над масивами, а також роботу методів reduce, accumulate, outer.

Код програми:

```
import numpy as np
import pandas as pd
import os

# Завдання 1
print("\n--- Завдання 1 ---")

# а) Створення різних типів масивів

print("\na) Генерація масивів:")

arr1 = np.arange(1, 16, 2)
print("Масив із кроком 2 у діапазоні [1, 16):\n", arr1)

arr2 = np.ones(9, dtype=int)
print("Масив із одиниць:\n", arr2)

arr3 = np.zeros((2, 4), dtype=float)
print("Двовимірний масив нулів:\n", arr3)

arr4 = np.linspace(0, 1, 5)
print("Рівномірний розподіл значень від 0 до 1:\n", arr4)

arr5 = np.random.rand(4, 3)
print("Масив випадкових чисел (4x3):\n", arr5)

arr6 = np.random.randint(0, 100, size=(4, 4))
print("Масив випадкових цілих чисел (4x4):\n", arr6)

arr7 = np.empty(5)
print("Порожній масив із 5 елементів:\n", arr7)
```

```

# б) Доступ до елементів масиву

print("\nб) Індексція та зрізи:")
print("Третій рядок у масиві arr6:\n", arr6[2])
print("Третій елемент з кінця:\n", arr6[-3])
print("Елемент першого рядка та останнього стовпця:\n", arr6[0, -1])

print("\nВиділення підмасивів:")
print("Перші два рядки:\n", arr6[:2])
print("Другий та третій стовпці:\n", arr6[:, 1:3])

# в) Арифметичні операції

print("\nv) Операції з масивами:")
arr8 = np.array([1, 2, 3])
arr9 = np.array([4, 5, 6])

print("Сума масивів:\n", arr8 + arr9)
print("Добуток масивів:\n", arr8 * arr9)
print("Функція reduce (сума елементів arr8):\n", np.add.reduce(arr8))
print("Функція accumulate (накопичена сума arr8):\n",
np.add.accumulate(arr8))
print("Функція outer (перемноження кожного з кожним):\n",
np.multiply.outer(arr8, arr9))

```

Результат виконання:

```

--- Завдання 1 ---

а) Генерація масивів:
Масив із кроком 2 у діапазоні [1, 16):
[ 1  3  5  7  9 11 13 15]
Масив із одиниць:
[1 1 1 1 1 1 1 1]
Двовимірний масив нулів:
[[0. 0. 0. 0.]
 [0. 0. 0. 0.]]
Рівномірний розподіл значень від 0 до 1:
[0.   0.25 0.5  0.75 1.  ]
Масив випадкових чисел (4x3):
[[0.33899579 0.72964243 0.42635429]
 [0.78185251 0.04645263 0.94407529]
 [0.43022123 0.29121402 0.51835588]
 [0.84028072 0.56597859 0.3928728 ]]
Масив випадкових цілих чисел (4x4):
[[51 53 66  7]
 [42 26 49 49]
 [60 26 61 58]
 [69 64 98  3]]
Порожній масив із 5 елементів:
[0.   0.25 0.5  0.75 1.  ]

б) Індексція та зрізи:
Третій рядок у масиві arr6:
[60 26 61 58]
Третій елемент з кінця:
[42 26 49 49]
Елемент першого рядка та останнього стовпця:
7

```

```

Виділення підмасивів:
Перші два рядки:
[[51 53 66 7]
 [42 26 49 49]]
Другий та третій стовпці:
[[53 66]
 [26 49]
 [26 61]
 [64 98]]

в) Операції з масивами:
Сума масивів:
[5 7 9]
Добуток масивів:
[ 4 10 18]
Функція reduce (сума елементів arr8):
6
Функція accumulate (накопичена сума arr8):
[1 3 6]
Функція outer (перемноження кожного з кожним):
[[ 4  5  6]
 [ 8 10 12]
 [12 15 18]]

```

Завдання 2. Варіант 10.

- а) Вивести основні статистичні характеристики кількісних ознак;
- б) Порахувати кількість та вивести інформацію щодо товарів з низьким пріоритетом та вагою до 2кг, що будуть відправлені кораблем;
- в) Додати новий стовпець, який містить перерахунок ціни за кілограм.

Код програми:

```

# Завдання 2
print("\nЗавдання 2:")

file_path = 'CommerceShipping.csv'

if os.path.exists(file_path):
    # Завантаження даних
    data = pd.read_csv(file_path)

    print("\nВиведення перших рядків таблиці:")
    print(data.head().to_string(index=False))

    # Вивід статистичних характеристик
    print("\nОсновні статистичні характеристики кількісних ознак:")
    print(data.describe().to_string())

    # Фільтрація товарів
    low_prior_under_2kg = data[(data['Product_importance'] == 'low') &
                               (data['Mode_of_Shipment'] == 'Ship') &
                               (data['Weight_in_gms'] < 2000)]

    # Виведення загальної кількості знайдених товарів
    print(

```

```

f"\nКількість товарів з низьким пріоритетом, вагою до 2кг, що
будуть відправлені кораблем: {len(low_prior_under_2kg)}")

# Виведення лише перших 5 рядків таблиці
if not low_prior_under_2kg.empty():
    print("\nПерелік таких товарів:")
    print(low_prior_under_2kg.to_string(index=False))
else:
    print("\nТоварів, що відповідають цим критеріям, немає.")

# Додаємо новий стовпець з ціною за кг
data['Price_per_kg'] = data['Cost_of_the_Product'] /
(data['Weight_in_gms'] / 1000)

print("\nТаблиця з новим стовпцем:")
# Перевіряємо результат (виведемо перші 5 рядків)
print(data.head().to_string(index=False))

else:
    print("Файл не знайдено!")

```

Результат:

Виведення перших рядків таблиці:

ID	Warehouse_block	Mode_of_Shipment	Customer_care_calls	Customer_rating	Cost_of_the_Product	Prior_purchases	Product_id
1	D	Flight	4	2	177	3	1
2	F	Flight	4	5	216	2	2
3	A	Flight	2	2	183	4	3
4	B	Flight	3	3	176	4	4
5	C	Flight	2	2	184	3	5

Основні статистичні характеристики кількісних ознак:

	ID	Customer_care_calls	Customer_rating	Cost_of_the_Product	Prior_purchases	Discount_offered	Weight_in_gms
count	1457.000000	1457.000000	1457.000000	1457.000000	1457.000000	1457.000000	1457.000000
mean	729.000000	3.732327	2.986960	194.671929	3.277282	32.793411	23.000000
std	420.743984	1.018917	1.404162	45.940277	1.424370	18.896276	9.000000
min	1.000000	2.000000	1.000000	96.000000	2.000000	1.000000	10.000000
25%	365.000000	3.000000	2.000000	157.000000	2.000000	16.000000	15.000000
50%	729.000000	4.000000	3.000000	190.000000	3.000000	33.000000	22.000000
75%	1093.000000	4.000000	4.000000	234.000000	4.000000	49.000000	30.000000
max	1457.000000	7.000000	5.000000	294.000000	10.000000	65.000000	78.000000

Кількість товарів з низьким пріоритетом, вагою до 2кг, що будуть відправлені кораблем: 232

Перелік таких товарів:

ID	Warehouse_block	Mode_of_Shipment	Customer_care_calls	Customer_rating	Cost_of_the_Product	Prior_purchases
23	C	Ship	2	5	156	1
44	F	Ship	3	1	174	1
45	A	Ship	3	1	151	1
48	F	Ship	5	1	160	1
52	B	Ship	4	4	141	1
53	C	Ship	4	3	165	1
54	F	Ship	2	1	170	1
68	F	Ship	3	3	158	1
78	F	Ship	3	3	113	1
82	B	Ship	2	5	132	1
104	F	Ship	5	1	144	1
156	F	Ship	5	1	114	1
160	B	Ship	3	3	149	1
165	A	Ship	3	5	141	1
168	F	Ship	4	5	152	1

Таблиця з новим стовпцем:

ID	Warehouse_block	Mode_of_Shipment	Customer_care_calls	Cus
1		D	Flight	4
2		F	Flight	4
3		A	Flight	2
4		B	Flight	3
5		C	Flight	2

ht_in_gms	Reached.on.Time_Y.N	Price_per_kg
1233	1	143.552311
3088	1	69.948187
3374	1	54.238293
1177	1	149.532710
2484	1	74.074074

Висновок: У ході виконання даної лабораторної роботи я зміг ознайомитись з основними можливостями роботи з масивами бібліотеки NumPy, структурами даних Pandas, визначити основні статистичні характеристики.