

01.04.04. Доступ до даних (індексація)

Пакет `pandas` у мові програмування Python забезпечує широкі можливості для довільного доступу до даних, що є ключовим фактором ефективного аналізу великих наборів інформації. Завдяки структурованим підходам до зберігання й обробки даних, `pandas` дозволяє швидко і легко отримувати доступ до будь-яких необхідних елементів чи груп елементів у масиві або таблиці даних.

Однією з фундаментальних особливостей `pandas` є використання спеціалізованих структур даних – `Series` та `DataFrame`. `Series` - це одновимірна структура з індексацією, що забезпечує швидкий доступ до даних за числовими чи мітковими індексами. `DataFrame` дозволяє отримувати доступ до даних у двовимірному вигляді за допомогою як числових, так і текстових міток стовпців та рядків. Це забезпечує зручність і точність під час роботи з великими й складними масивами даних.

`Pandas` реалізує кілька методів довільного доступу до даних. Зокрема, атрибути `.loc` та `.iloc` дозволяють отримувати доступ до елементів `DataFrame` та `Series` за їх мітками та позиціями відповідно. Метод `.loc` підтримує вибір даних за чітко заданими іменами рядків і стовпців, що робить його зручним для роботи з семантично визначеними індексами. Водночас `.iloc` забезпечує доступ до елементів за їх числовими індексами, подібно до класичного масиву, що особливо корисно при роботі з даними, де імена стовпців чи рядків не мають особливого значення.

Крім того, `pandas` пропонує функцію `.at` і `.iat` для швидкого довільного доступу до окремих елементів. Ці функції є оптимізованими версіями `.loc` і `.iloc` і забезпечують найбільш швидкий і ефективний доступ до конкретних даних у випадках, коли необхідно працювати з окремими значеннями, а не цілими рядками чи стовпцями.

Пакет також підтримує умовний доступ до даних за допомогою логічної індексації, що дозволяє швидко вибирати дані, які відповідають певним критеріям чи умовам. Це надає додаткову гнучкість при роботі зі складними запитам та фільтрами.

Таким чином, пакет `pandas` забезпечує потужні та гнучкі механізми для довільного доступу до даних, що є незамінним у науковому та прикладному аналізі. Ці можливості роблять `pandas` ефективним інструментом для швидкої та зручної роботи з великими обсягами структурованих даних.

Доступ до стовпців

```
print(type(df['Name'])) # pandas.Series, Один стовець
```

```
print(type(df[['Name', 'Salary']])) # pandas.DataFrame, Декілька стовпців
```

✅ `df['Name']` повертає **Series**, а `df[['Name']]` – **DataFrame**.

Доступ до рядків

```
print(df.iloc[0]) # Перший рядок
print(df.iloc[:2]) # Перші два рядки
print(df.loc[0, 'Salary']) # Конкретне значення
```

`df.loc` – це метод `pandas`, який використовується для доступу до рядків та стовпців у `DataFrame` за мітками (label-based indexing).

Синтаксис:

```
df.loc[row_label, column_label]
```

- `row_label` – назва або індекс рядка.
- `column_label` – назва або список назв стовпців.

Приклад:

```
print(df.loc[0, 'Salary']) # Отримати зарплату першого співробітника

print(df.loc[0, ['Name', 'Salary']]) # Отримати ім'я та зарплату першого співробітника

print(df.loc[:2, 'Name']) # Імена перших трьох співробітників
```

Зверніть увагу:

`loc[:2, 'Name']` включає рядок 2, а `iloc[:2]` – ні

Метод	Використання	Приклад
<code>iloc[]</code>	Доступ за номером рядка/стовпця	<code>df.iloc[0]</code> – перший рядок
<code>loc[]</code>	Доступ за назвою індексу та/або назвою стовпця	<code>df.loc[0, "Salary"]</code> – значення "Salary" у першому рядку

Приклад роботи з рядковими індексами:

```
df.set_index('Name', inplace=True) # Встановити 'Name' як індекс

print(df.loc['Alice']) # Отримати всі дані про Alice
```

```
print(df.iloc[0]) # Отримати перший рядок (якщо відлік від 0)
```

✓ `df.at[]` і `df.iat[]` працюють швидше за `loc[]` і `iloc[]`, якщо потрібно отримати лише одне значення.

```
print(df.at[0, 'Salary']) # Швидший доступ до значення у першому  
рядку, стовпець 'Salary'
```

```
print(df.iat[0, 2]) # Швидший доступ за числовим індексом (перший  
рядок, третій стовпець)
```

Література для поглибленого вивчення доступу до даних (індексації) у Pandas

1. Pandas documentation: Data structures. (б.д.). Отримано 4 липня 2025 з https://pandas.pydata.org/docs/user_guide/dsintro.html
2. MCKINNEY, W. (2017). *Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython* (2nd ed.). O'Reilly Media. URL: <https://www.oreilly.com/library/view/python-for-data/9781491957653/>
3. VANDERPLAS, J. (2016). *Python Data Science Handbook: Essential Tools for Working with Data*. O'Reilly Media. URL: <https://jakevdp.github.io/PythonDataScienceHandbook/>