# INDEXACIÓN Y SLICING EN PANDAS

**Selección de datos de Pandas**

Hay varias formas de seleccionar e indexar filas y columnas en Pandas :

# •Seleccionar datos por posición (.iloc)

# •Seleccionar datos por etiqueta o por una declaración condicional (.loc)

Para verificar la estructura que devuelve la selección (Series o Dataframe) asignar a una variable la selección y posteriormente aplicar type().

# SLICING EN PANDAS CON ILOC

# Indexación en Pandas

import pandas as pd

datos = pd.read_csv('Salaries.csv')

print(datos)

```
>>> import pandas as pd
>>> datos = pd.read_csv('Salaries.csv')
>>> print(datos)
     order       rank discipline  yrs.since.phd  yrs.service   sex  salary
0        1       Prof          B             19           18  Male  139750
1        2       Prof          B             20           16  Male  173200
2        3   AsstProf          B              4            3  Male   79750
3        4       Prof          B             45           39  Male  115000
4        5       Prof          B             40           41  Male  141500
..     ...        ...        ...            ...          ...   ...     ...
392    393       Prof          A             33           30  Male  103106
393    394       Prof          A             31           19  Male  150564
394    395       Prof          A             42           25  Male  101738
395    396       Prof          A             25           15  Male   95329
396    397   AsstProf          A              8            4  Male   81035

[397 rows x 7 columns]
```

| | Índice de columnas | | | | | |
|---|---|---|---|---|---|---|
| 0 | 1 | 2 | 3 | 4 | 5 | 6 |

| | order | rank | discipline | yrs.since.phd | yrs.service | sex | salary |
|---|---|---|---|---|---|---|---|
| 0 | 1 | Prof | B | 19 | 18 | Male | 139750 |
| 1 | 2 | Prof | B | 20 | 16 | Male | 173200 |
| 2 | 3 | AsstProf | B | 4 | 3 | Male | 79750 |
| 3 | 4 | Prof | B | 45 | 39 | Male | 115000 |
| 4 | 5 | Prof | B | 40 | 41 | Male | 141500 |
| 5 | 6 | AssocProf | B | 6 | 6 | Male | 97000 |
| 6 | 7 | Prof | B | 30 | 23 | Male | 175000 |
| 7 | 8 | Prof | B | 45 | 45 | Male | 147765 |
| 8 | 9 | Prof | B | 21 | 20 | Male | 119250 |
| 9 | 10 | Prof | B | 18 | 18 | Female | 129000 |
| 10 | 11 | AssocProf | B | 12 | 8 | Male | 119800 |
| 11 | 12 | AsstProf | B | 7 | 2 | Male | 79800 |
| 12 | 13 | AsstProf | B | 1 | 1 | Male | 77700 |
| 13 | 14 | AsstProf | B | 2 | 0 | Male | 78000 |
| 14 | 15 | Prof | B | 20 | 18 | Male | 104800 |

Índice de filas

El dataset que se utuliza en los ejemplos tiene 397 filas

# Seleccionar una fila

```
import pandas as pd

datos = pd.read_csv('Salaries.csv')
print(datos)

datos.iloc[4]
```

En esta selección pandas
devuelve una Serie

| | Índice de columnas | | | | | | |
|---|---|---|---|---|---|---|---|
| | 0 | 1 | 2 | 3 | 4 | 5 | 6 |

| Índice de filas | | order | rank | discipline | yrs.since.phd | yrs.service | sex | salary |
|---|---|---|---|---|---|---|---|---|
| | 0 | 1 | Prof | B | 19 | 18 | Male | 139750 |
| | 1 | 2 | Prof | B | 20 | 16 | Male | 173200 |
| | 2 | 3 | AsstProf | B | 4 | 3 | Male | 79750 |
| | 3 | 4 | Prof | B | 45 | 39 | Male | 115000 |
| | 4 | 5 | Prof | B | 40 | 41 | Male | 141500 |
| | 5 | 6 | AssocProf | B | 6 | 6 | Male | 97000 |
| | 6 | 7 | Prof | B | 30 | 23 | Male | 175000 |
| | 7 | 8 | Prof | B | 45 | 45 | Male | 147765 |
| | 8 | 9 | Prof | B | 21 | 20 | Male | 119250 |
| | 9 | 10 | Prof | B | 18 | 18 | Female | 129000 |
| | 10 | 11 | AssocProf | B | 12 | 8 | Male | 119800 |
| | 11 | 12 | AsstProf | B | 7 | 2 | Male | 79800 |
| | 12 | 13 | AsstProf | B | 1 | 1 | Male | 77700 |
| | 13 | 14 | AsstProf | B | 2 | 0 | Male | 78000 |
| | 14 | 15 | Prof | B | 20 | 18 | Male | 104800 |

# Seleccionar última fila

```python
import pandas as pd

datos = pd.read_csv('Salaries.csv')
print(datos)

datos.iloc[-1]

# podemos ingresar una lista, con un solo índice entero,
# cuando usamos iloc. Esto indexará una fila, pero la salida
# será diferente en comparación con el ejemplo anterior:
datos.iloc[[-1]]
```

| | Índice de columnas | | | | | | |
|---|---|---|---|---|---|---|---|
| | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
| | order | rank | discipline | yrs.since.phd | yrs.service | sex | salary |
| 0 | 383 | AssocProf | A | 8 | 5 | Male | 86895 |
| 1 | 384 | Prof | A | 44 | 44 | Male | 105000 |
| 2 | 385 | Prof | A | 27 | 21 | Male | 125192 |
| 3 | 386 | Prof | A | 15 | 9 | Male | 114330 |
| 4 | 387 | Prof | A | 29 | 27 | Male | 139219 |
| 5 | 388 | Prof | A | 29 | 15 | Male | 109305 |
| 6 | 389 | Prof | A | 38 | 36 | Male | 119450 |
| 7 | 390 | Prof | A | 33 | 18 | Male | 186023 |
| 8 | 391 | Prof | A | 40 | 19 | Male | 166605 |
| 9 | 392 | Prof | A | 30 | 19 | Male | 151292 |
| 10 | 393 | Prof | A | 33 | 30 | Male | 103106 |
| 11 | 394 | Prof | A | 31 | 19 | Male | 150564 |
| 12 | 395 | Prof | A | 42 | 25 | Male | 101738 |
| 13 | 396 | Prof | A | 25 | 15 | Male | 95329 |
| 14 | 397 | AsstProf | A | 8 | 4 | Male | 81035 |

Índice de filas

# Seleccionar una celda específica

import pandas as pd

datos = pd.read_csv('Salaries.csv')

print(datos)

datos.iloc[9,5]

| | Índice de columnas | | | | | |
|---|---|---|---|---|---|---|
| 0 | 1 | 2 | 3 | 4 | 5 | 6 |
| order | rank | discipline | yrs.since.phd | yrs.service | sex | salary |
| 1 | Prof | B | 19 | 18 | Male | 139750 |
| 2 | Prof | B | 20 | 16 | Male | 173200 |
| 3 | AsstProf | B | 4 | 3 | Male | 79750 |
| 4 | Prof | B | 45 | 39 | Male | 115000 |
| 5 | Prof | B | 40 | 41 | Male | 141500 |
| 6 | AssocProf | B | 6 | 6 | Male | 97000 |
| 7 | Prof | B | 30 | 23 | Male | 175000 |
| 8 | Prof | B | 45 | 45 | Male | 147765 |
| 9 | Prof | B | 21 | 20 | Male | 119250 |
| 10 | Prof | B | 18 | 18 | Female | 129000 |
| 11 | AssocProf | B | 12 | 8 | Male | 119800 |
| 12 | AsstProf | B | 7 | 2 | Male | 79800 |
| 13 | AsstProf | B | 1 | 1 | Male | 77700 |
| 14 | AsstProf | B | 2 | 0 | Male | 78000 |
| 15 | Prof | B | 20 | 18 | Male | 104800 |

Índice de filas: 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14

# Seleccionar múltiples filas

import pandas as pd

datos = pd.read_csv('Salaries.csv')

print(datos)

datos.iloc[[7, 2, 0]]

En esta selección pandas devuelve un Dataframe

| | Índice de columnas | | | | | | |
|---|---|---|---|---|---|---|---|
| | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
| | order | rank | discipline | yrs.since.phd | yrs.service | sex | salary |
| 0 | 1 | Prof | B | 19 | 18 | Male | 139750 |
| 1 | 2 | Prof | B | 20 | 16 | Male | 173200 |
| 2 | 3 | AsstProf | B | 4 | 3 | Male | 79750 |
| 3 | 4 | Prof | B | 45 | 39 | Male | 115000 |
| 4 | 5 | Prof | B | 40 | 41 | Male | 141500 |
| 5 | 6 | AssocProf | B | 6 | 6 | Male | 97000 |
| 6 | 7 | Prof | B | 30 | 23 | Male | 175000 |
| 7 | 8 | Prof | B | 45 | 45 | Male | 147765 |
| 8 | 9 | Prof | B | 21 | 20 | Male | 119250 |
| 9 | 10 | Prof | B | 18 | 18 | Female | 129000 |
| 10 | 11 | AssocProf | B | 12 | 8 | Male | 119800 |
| 11 | 12 | AsstProf | B | 7 | 2 | Male | 79800 |
| 12 | 13 | AsstProf | B | 1 | 1 | Male | 77700 |
| 13 | 14 | AsstProf | B | 2 | 0 | Male | 78000 |
| 14 | 15 | Prof | B | 20 | 18 | Male | 104800 |

Índice de filas

# Seleccionar parte de los datos de una fila

Import pandas as pd

datos = pd.read_csv('Salaries.csv')

print(datos)

datos.iloc[3, [1, 2, 3]]

En esta selección pandas devuelve una Serie

| | | Índice de columnas | | | | | |
|---|---|---|---|---|---|---|---|
| | 0 | 1 | 2 | 3 | 4 | 5 | 6 |

| | | order | rank | discipline | yrs.since.phd | yrs.service | sex | salary |
|---|---|---|---|---|---|---|---|---|
| Índice de filas | 0 | 1 | Prof | B | 19 | 18 | Male | 139750 |
| | 1 | 2 | Prof | B | 20 | 16 | Male | 173200 |
| | 2 | 3 | AsstProf | B | 4 | 3 | Male | 79750 |
| | 3 | 4 | *Prof* | *B* | *45* | 39 | Male | 115000 |
| | 4 | 5 | Prof | B | 40 | 41 | Male | 141500 |
| | 5 | 6 | AssocProf | B | 6 | 6 | Male | 97000 |
| | 6 | 7 | Prof | B | 30 | 23 | Male | 175000 |
| | 7 | 8 | Prof | B | 45 | 45 | Male | 147765 |
| | 8 | 9 | Prof | B | 21 | 20 | Male | 119250 |
| | 9 | 10 | Prof | B | 18 | 18 | Female | 129000 |
| | 10 | 11 | AssocProf | B | 12 | 8 | Male | 119800 |
| | 11 | 12 | AsstProf | B | 7 | 2 | Male | 79800 |
| | 12 | 13 | AsstProf | B | 1 | 1 | Male | 77700 |
| | 13 | 14 | AsstProf | B | 2 | 0 | Male | 78000 |
| | 14 | 15 | Prof | B | 20 | 18 | Male | 104800 |

# Seleccionar rango de filas y todas las columnas

import pandas as pd

datos = pd.read_csv('Salaries.csv')

print(datos)

datos.iloc[8:13]

Al seleccionar varias columnas o varias filas , las filas / columnas seleccionadas se ejecutarán desde el primer número hasta *uno menos* del segundo valor, por ejemplo, [1: 5] será 1, 2, 3, 4.

| | Índice de columnas | | | | | | |
|---|---|---|---|---|---|---|---|
| | 0 | 1 | 2 | 3 | 4 | 5 | 6 |

| Índice de filas | | order | rank | discipline | yrs.since.phd | yrs.service | sex | salary |
|---|---|---|---|---|---|---|---|---|
| | 0 | 1 | Prof | B | 19 | 18 | Male | 139750 |
| | 1 | 2 | Prof | B | 20 | 16 | Male | 173200 |
| | 2 | 3 | AsstProf | B | 4 | 3 | Male | 79750 |
| | 3 | 4 | Prof | B | 45 | 39 | Male | 115000 |
| | 4 | 5 | Prof | B | 40 | 41 | Male | 141500 |
| | 5 | 6 | AssocProf | B | 6 | 6 | Male | 97000 |
| | 6 | 7 | Prof | B | 30 | 23 | Male | 175000 |
| | 7 | 8 | Prof | B | 45 | 45 | Male | 147765 |
| | 8 | 9 | Prof | B | 21 | 20 | Male | 119250 |
| | 9 | 10 | Prof | B | 18 | 18 | Female | 129000 |
| | 10 | 11 | AssocProf | B | 12 | 8 | Male | 119800 |
| | 11 | 12 | AsstProf | B | 7 | 2 | Male | 79800 |
| | 12 | 13 | AsstProf | B | 1 | 1 | Male | 77700 |
| | 13 | 14 | AsstProf | B | 2 | 0 | Male | 78000 |
| | 14 | 15 | Prof | B | 20 | 18 | Male | 104800 |

**Seleccionar columnas**

```
import pandas as pd

datos = pd.read_csv('Salaries.csv')

print(datos)

datos.iloc[:, 0]
```

Selecciona todas las filas

Selecciona la columna 0

datos.iloc[ : , 0 ]

| | | Índice de columnas | | | | | |
|---|---|---|---|---|---|---|---|
| | | 0 | 1 | 2 | 3 | 4 | 5 | 6 |

| Índice de filas | | order | rank | discipline | yrs.since.phd | yrs.service | sex | salary |
|---|---|---|---|---|---|---|---|---|
| | 0 | 1 | Prof | B | 19 | 18 | Male | 139750 |
| | 1 | 2 | Prof | B | 20 | 16 | Male | 173200 |
| | 2 | 3 | AsstProf | B | 4 | 3 | Male | 79750 |
| | 3 | 4 | Prof | B | 45 | 39 | Male | 115000 |
| | 4 | 5 | Prof | B | 40 | 41 | Male | 141500 |
| | 5 | 6 | AssocProf | B | 6 | 6 | Male | 97000 |
| | 6 | 7 | Prof | B | 30 | 23 | Male | 175000 |
| | 7 | 8 | Prof | B | 45 | 45 | Male | 147765 |
| | 8 | 9 | Prof | B | 21 | 20 | Male | 119250 |
| | 9 | 10 | Prof | B | 18 | 18 | Female | 129000 |
| | 10 | 11 | AssocProf | B | 12 | 8 | Male | 119800 |
| | 11 | 12 | AsstProf | B | 7 | 2 | Male | 79800 |
| | 12 | 13 | AsstProf | B | 1 | 1 | Male | 77700 |
| | 13 | 14 | AsstProf | B | 2 | 0 | Male | 78000 |
| | 14 | 15 | Prof | B | 20 | 18 | Male | 104800 |

# Seleccionar columnas

```python
import pandas as pd

datos = pd.read_csv('Salaries.csv')

print(datos)

datos.iloc[:, -1]
```

| | Índice de columnas | | | | | |
|---|---|---|---|---|---|---|
| 0 | 1 | 2 | 3 | 4 | 5 | 6 |

| | order | rank | discipline | yrs.since.phd | yrs.service | sex | salary |
|---|---|---|---|---|---|---|---|
| 0 | 1 | Prof | B | 19 | 18 | Male | 139750 |
| 1 | 2 | Prof | B | 20 | 16 | Male | 173200 |
| 2 | 3 | AsstProf | B | 4 | 3 | Male | 79750 |
| 3 | 4 | Prof | B | 45 | 39 | Male | 115000 |
| 4 | 5 | Prof | B | 40 | 41 | Male | 141500 |
| 5 | 6 | AssocProf | B | 6 | 6 | Male | 97000 |
| 6 | 7 | Prof | B | 30 | 23 | Male | 175000 |
| 7 | 8 | Prof | B | 45 | 45 | Male | 147765 |
| 8 | 9 | Prof | B | 21 | 20 | Male | 119250 |
| 9 | 10 | Prof | B | 18 | 18 | Female | 129000 |
| 10 | 11 | AssocProf | B | 12 | 8 | Male | 119800 |
| 11 | 12 | AsstProf | B | 7 | 2 | Male | 79800 |
| 12 | 13 | AsstProf | B | 1 | 1 | Male | 77700 |
| 13 | 14 | AsstProf | B | 2 | 0 | Male | 78000 |
| 14 | 15 | Prof | B | 20 | 18 | Male | 104800 |

Índice de filas

# Seleccionar parte de filas y una columna

import pandas as pd

datos = pd.read_csv('Salaries.csv')

print(datos)

datos.iloc[1:5, 3]

En esta selección pandas devuelve una Serie

| Índice de columnas | | | | | | |
|---|---|---|---|---|---|---|
| 0 | 1 | 2 | 3 | 4 | 5 | 6 |
| order | rank | discipline | yrs.since.phd | yrs.service | sex | salary |
| 1 | Prof | B | 19 | 18 | Male | 139750 |
| 2 | Prof | B | 20 | 16 | Male | 173200 |
| 3 | AsstProf | B | 4 | 3 | Male | 79750 |
| 4 | Prof | B | 45 | 39 | Male | 115000 |
| 5 | Prof | B | 40 | 41 | Male | 141500 |
| 6 | AssocProf | B | 6 | 6 | Male | 97000 |
| 7 | Prof | B | 30 | 23 | Male | 175000 |
| 8 | Prof | B | 45 | 45 | Male | 147765 |
| 9 | Prof | B | 21 | 20 | Male | 119250 |
| 10 | Prof | B | 18 | 18 | Female | 129000 |
| 11 | AssocProf | B | 12 | 8 | Male | 119800 |
| 12 | AsstProf | B | 7 | 2 | Male | 79800 |
| 13 | AsstProf | B | 1 | 1 | Male | 77700 |
| 14 | AsstProf | B | 2 | 0 | Male | 78000 |
| 15 | Prof | B | 20 | 18 | Male | 104800 |

Índice de filas: 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14

# Seleccionar todas las filas y un rango de columnas

import pandas as pd

datos = pd.read_csv('Salaries.csv')

print(datos)

datos.iloc[ : , 1: 6]

En esta selección pandas
devuelve un Dataframe

| | Índice de columnas | | | | | | |
|---|---|---|---|---|---|---|---|
| | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
| | order | rank | discipline | yrs.since.phd | yrs.service | sex | salary |
| 0 | 1 | Prof | B | 19 | 18 | Male | 139750 |
| 1 | 2 | Prof | B | 20 | 16 | Male | 173200 |
| 2 | 3 | AsstProf | B | 4 | 3 | Male | 79750 |
| 3 | 4 | Prof | B | 45 | 39 | Male | 115000 |
| 4 | 5 | Prof | B | 40 | 41 | Male | 141500 |
| 5 | 6 | AssocProf | B | 6 | 6 | Male | 97000 |
| 6 | 7 | Prof | B | 30 | 23 | Male | 175000 |
| 7 | 8 | Prof | B | 45 | 45 | Male | 147765 |
| 8 | 9 | Prof | B | 21 | 20 | Male | 119250 |
| 9 | 10 | Prof | B | 18 | 18 | Female | 129000 |
| 10 | 11 | AssocProf | B | 12 | 8 | Male | 119800 |
| 11 | 12 | AsstProf | B | 7 | 2 | Male | 79800 |
| 12 | 13 | AsstProf | B | 1 | 1 | Male | 77700 |
| 13 | 14 | AsstProf | B | 2 | 0 | Male | 78000 |
| 14 | 15 | Prof | B | 20 | 18 | Male | 104800 |

Índice de filas

**Seleccionar subconjuntos de celdas**

```python
import pandas as pd

datos = pd.read_csv('Salaries.csv')

print(datos)

datos.iloc[2:5, 3:6]
```

Indexación de filas

Indexación de columnas

datos.iloc[ 2:5 , 3:6 ]

| | | Índice de columnas | | | | | |
|---|---|---|---|---|---|---|---|
| | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
| | order | rank | discipline | yrs.since.phd | yrs.service | sex | salary |
| 0 | 1 | Prof | B | 19 | 18 | Male | 139750 |
| 1 | 2 | Prof | B | 20 | 16 | Male | 173200 |
| 2 | 3 | AsstProf | B | **4** | **3** | **Male** | 79750 |
| 3 | 4 | Prof | B | **45** | **39** | **Male** | 115000 |
| 4 | 5 | Prof | B | **40** | **41** | **Male** | 141500 |
| 5 | 6 | AssocProf | B | 6 | 6 | Male | 97000 |
| 6 | 7 | Prof | B | 30 | 23 | Male | 175000 |
| 7 | 8 | Prof | B | 45 | 45 | Male | 147765 |
| 8 | 9 | Prof | B | 21 | 20 | Male | 119250 |
| 9 | 10 | Prof | B | 18 | 18 | Female | 129000 |
| 10 | 11 | AssocProf | B | 12 | 8 | Male | 119800 |
| 11 | 12 | AsstProf | B | 7 | 2 | Male | 79800 |
| 12 | 13 | AsstProf | B | 1 | 1 | Male | 77700 |
| 13 | 14 | AsstProf | B | 2 | 0 | Male | 78000 |
| 14 | 15 | Prof | B | 20 | 18 | Male | 104800 |

Índice de filas

# SLICING EN PANDAS CON LOC

# Seleccionar fila

```
import pandas as pd

datos = pd.read_csv('Salaries.csv')

print(datos)

datos.loc[3]
datos.loc[[3]]
```

| | Índice de columnas | | | | | | |
|---|---|---|---|---|---|---|---|
| | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
| | order | rank | discipline | yrs.since.phd | yrs.service | sex | salary |
| 0 | 1 | Prof | B | 19 | 18 | Male | 139750 |
| 1 | 2 | Prof | B | 20 | 16 | Male | 173200 |
| 2 | 3 | AsstProf | B | 4 | 3 | Male | 79750 |
| 3 | 4 | Prof | B | 45 | 39 | Male | 115000 |
| 4 | 5 | Prof | B | 40 | 41 | Male | 141500 |
| 5 | 6 | AssocProf | B | 6 | 6 | Male | 97000 |
| 6 | 7 | Prof | B | 30 | 23 | Male | 175000 |
| 7 | 8 | Prof | B | 45 | 45 | Male | 147765 |
| 8 | 9 | Prof | B | 21 | 20 | Male | 119250 |
| 9 | 10 | Prof | B | 18 | 18 | Female | 129000 |
| 10 | 11 | AssocProf | B | 12 | 8 | Male | 119800 |
| 11 | 12 | AsstProf | B | 7 | 2 | Male | 79800 |
| 12 | 13 | AsstProf | B | 1 | 1 | Male | 77700 |
| 13 | 14 | AsstProf | B | 2 | 0 | Male | 78000 |
| 14 | 15 | Prof | B | 20 | 18 | Male | 104800 |

Índice de filas

# Seleccionar subconjunto

import pandas as pd

datos = pd.read_csv('Salaries.csv')

print(datos)

datos.loc[1:5]

| | | Índice de columnas | | | | | |
|---|---|---|---|---|---|---|---|
| | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
| | order | rank | discipline | yrs.since.phd | yrs.service | sex | salary |
| 0 | 1 | Prof | B | 19 | 18 | Male | 139750 |
| 1 | 2 | Prof | B | 20 | 16 | Male | 173200 |
| 2 | 3 | AsstProf | B | 4 | 3 | Male | 79750 |
| 3 | 4 | Prof | B | 45 | 39 | Male | 115000 |
| 4 | 5 | Prof | B | 40 | 41 | Male | 141500 |
| 5 | 6 | AssocProf | B | 6 | 6 | Male | 97000 |
| 6 | 7 | Prof | B | 30 | 23 | Male | 175000 |
| 7 | 8 | Prof | B | 45 | 45 | Male | 147765 |
| 8 | 9 | Prof | B | 21 | 20 | Male | 119250 |
| 9 | 10 | Prof | B | 18 | 18 | Female | 129000 |
| 10 | 11 | AssocProf | B | 12 | 8 | Male | 119800 |
| 11 | 12 | AsstProf | B | 7 | 2 | Male | 79800 |
| 12 | 13 | AsstProf | B | 1 | 1 | Male | 77700 |
| 13 | 14 | AsstProf | B | 2 | 0 | Male | 78000 |
| 14 | 15 | Prof | B | 20 | 18 | Male | 104800 |

Índice de filas

# Seleccionar filas alternadas

import pandas as pd

datos = pd.read_csv('Salaries.csv')

print(datos)

datos.loc[[1, 3, 7, 10, 13]]

| | Índice de columnas | | | | | | |
|---|---|---|---|---|---|---|---|
| | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
| | order | rank | discipline | yrs.since.phd | yrs.service | sex | salary |
| 0 | 1 | Prof | B | 19 | 18 | Male | 139750 |
| 1 | 2 | Prof | B | 20 | 16 | Male | 173200 |
| 2 | 3 | AsstProf | B | 4 | 3 | Male | 79750 |
| 3 | 4 | Prof | B | 45 | 39 | Male | 115000 |
| 4 | 5 | Prof | B | 40 | 41 | Male | 141500 |
| 5 | 6 | AssocProf | B | 6 | 6 | Male | 97000 |
| 6 | 7 | Prof | B | 30 | 23 | Male | 175000 |
| 7 | 8 | Prof | B | 45 | 45 | Male | 147765 |
| 8 | 9 | Prof | B | 21 | 20 | Male | 119250 |
| 9 | 10 | Prof | B | 18 | 18 | Female | 129000 |
| 10 | 11 | AssocProf | B | 12 | 8 | Male | 119800 |
| 11 | 12 | AsstProf | B | 7 | 2 | Male | 79800 |
| 12 | 13 | AsstProf | B | 1 | 1 | Male | 77700 |
| 13 | 14 | AsstProf | B | 2 | 0 | Male | 78000 |
| 14 | 15 | Prof | B | 20 | 18 | Male | 104800 |

Índice de filas

# Seleccionar con nombres de columnas

import pandas as pd

datos = pd.read_csv('Salaries.csv')

datos.loc[:, 'rank']

| | Índice de columnas | | | | | | |
|---|---|---|---|---|---|---|---|
| | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
| | order | rank | discipline | yrs.since.phd | yrs.service | sex | salary |
| 0 | 1 | Prof | B | 19 | 18 | Male | 139750 |
| 1 | 2 | Prof | B | 20 | 16 | Male | 173200 |
| 2 | 3 | AsstProf | B | 4 | 3 | Male | 79750 |
| 3 | 4 | Prof | B | 45 | 39 | Male | 115000 |
| 4 | 5 | Prof | B | 40 | 41 | Male | 141500 |
| 5 | 6 | AssocProf | B | 6 | 6 | Male | 97000 |
| 6 | 7 | Prof | B | 30 | 23 | Male | 175000 |
| 7 | 8 | Prof | B | 45 | 45 | Male | 147765 |
| 8 | 9 | Prof | B | 21 | 20 | Male | 119250 |
| 9 | 10 | Prof | B | 18 | 18 | Female | 129000 |
| 10 | 11 | AssocProf | B | 12 | 8 | Male | 119800 |
| 11 | 12 | AsstProf | B | 7 | 2 | Male | 79800 |
| 12 | 13 | AsstProf | B | 1 | 1 | Male | 77700 |
| 13 | 14 | AsstProf | B | 2 | 0 | Male | 78000 |
| 14 | 15 | Prof | B | 20 | 18 | Male | 104800 |

Índice de filas

# Seleccionar con nombres de columnas

```python
import pandas as pd

datos = pd.read_csv('Salaries.csv')

datos.loc[1:5, ['rank', 'yrs.service']]
```

| | Índice de columnas | | | | | | |
|---|---|---|---|---|---|---|---|
| | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
| | order | rank | discipline | yrs.since.phd | yrs.service | sex | salary |
| 0 | 1 | Prof | B | 19 | 18 | Male | 139750 |
| 1 | 2 | *Prof* | B | 20 | *16* | Male | 173200 |
| 2 | 3 | *AsstProf* | B | 4 | *3* | Male | 79750 |
| 3 | 4 | *Prof* | B | 45 | *39* | Male | 115000 |
| 4 | 5 | *Prof* | B | 40 | *41* | Male | 141500 |
| 5 | 6 | *AssocProf* | B | 6 | *6* | Male | 97000 |
| 6 | 7 | Prof | B | 30 | 23 | Male | 175000 |
| 7 | 8 | Prof | B | 45 | 45 | Male | 147765 |
| 8 | 9 | Prof | B | 21 | 20 | Male | 119250 |
| 9 | 10 | Prof | B | 18 | 18 | Female | 129000 |
| 10 | 11 | AssocProf | B | 12 | 8 | Male | 119800 |
| 11 | 12 | AsstProf | B | 7 | 2 | Male | 79800 |
| 12 | 13 | AsstProf | B | 1 | 1 | Male | 77700 |
| 13 | 14 | AsstProf | B | 2 | 0 | Male | 78000 |
| 14 | 15 | Prof | B | 20 | 18 | Male | 104800 |

Índice de filas

# Seleccionar con nombres de columnas

```
import pandas as pd

datos = pd.read_csv('Salaries.csv')

datos.loc[1:5, 'rank':'yrs.service']
```

| | Índice de columnas | | | | | | |
|---|---|---|---|---|---|---|---|
| | **0** | **1** | **2** | **3** | **4** | **5** | **6** |
| | order | rank | discipline | yrs.since.phd | yrs.service | sex | salary |
| **0** | 1 | Prof | B | 19 | 18 | Male | 139750 |
| **1** | 2 | *Prof* | *B* | *20* | *16* | Male | 173200 |
| **2** | 3 | *AsstProf* | *B* | *4* | *3* | Male | 79750 |
| **3** | 4 | *Prof* | *B* | *45* | *39* | Male | 115000 |
| **4** | 5 | *Prof* | *B* | *40* | *41* | Male | 141500 |
| **5** | 6 | *AssocProf* | *B* | *6* | *6* | Male | 97000 |
| **6** | 7 | Prof | B | 30 | 23 | Male | 175000 |
| **7** | 8 | Prof | B | 45 | 45 | Male | 147765 |
| **8** | 9 | Prof | B | 21 | 20 | Male | 119250 |
| **9** | 10 | Prof | B | 18 | 18 | Female | 129000 |
| **10** | 11 | AssocProf | B | 12 | 8 | Male | 119800 |
| **11** | 12 | AsstProf | B | 7 | 2 | Male | 79800 |
| **12** | 13 | AsstProf | B | 1 | 1 | Male | 77700 |
| **13** | 14 | AsstProf | B | 2 | 0 | Male | 78000 |
| **14** | 15 | Prof | B | 20 | 18 | Male | 104800 |

Índice de filas

# Seleccionar con el nombre de una columna un dato determinado

import pandas as pd

datos = pd.read_csv('Salaries.csv')

datos.loc[datos['rank'] == 'Prof']

| | Índice de columnas | | | | | | |
|---|---|---|---|---|---|---|---|
| | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
| | order | rank | discipline | yrs.since.phd | yrs.service | sex | salary |
| 0 | 1 | Prof | B | 19 | 18 | Male | 139750 |
| 1 | 2 | Prof | B | 20 | 16 | Male | 173200 |
| 2 | 3 | AsstProf | B | 4 | 3 | Male | 79750 |
| 3 | 4 | Prof | B | 45 | 39 | Male | 115000 |
| 4 | 5 | Prof | B | 40 | 41 | Male | 141500 |
| 5 | 6 | AssocProf | B | 6 | 6 | Male | 97000 |
| 6 | 7 | Prof | B | 30 | 23 | Male | 175000 |
| 7 | 8 | Prof | B | 45 | 45 | Male | 147765 |
| 8 | 9 | Prof | B | 21 | 20 | Male | 119250 |
| 9 | 10 | Prof | B | 18 | 18 | Female | 129000 |
| 10 | 11 | AssocProf | B | 12 | 8 | Male | 119800 |
| 11 | 12 | AsstProf | B | 7 | 2 | Male | 79800 |
| 12 | 13 | AsstProf | B | 1 | 1 | Male | 77700 |
| 13 | 14 | AsstProf | B | 2 | 0 | Male | 78000 |
| 14 | 15 | Prof | B | 20 | 18 | Male | 104800 |

Índice de filas

# Seleccionar filas usando múltiples condiciones

import pandas as pd

datos = pd.read_csv('Salaries.csv')

datos.loc[(datos['yrs.service'] > 25) & (datos['rank'] == 'AssocProf')]



| | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| | order | rank | discipline | yrs.since.phd | yrs.service | sex | salary |
| | 187 | AssocProf | B | 13 | 10 | Female | 103750 |
| | 188 | Prof | B | 18 | 10 | Male | 107500 |
| 188 | 189 | AssocProf | B | 28 | 28 | Male | 106300 |
| | 190 | Prof | B | 25 | 19 | Male | 153750 |
| | 193 | Prof | B | 19 | 18 | Male | 122100 |
| | 194 | AssocProf | B | 19 | 19 | Male | 86250 |
| 194 | 195 | AssocProf | B | 48 | 53 | Male | 90000 |
| | 196 | AssocProf | B | 9 | 7 | Male | 113600 |
| | 197 | AsstProf | B | 4 | 4 | Male | 92700 |
| | 259 | AsstProf | A | 9 | 3 | Male | 73800 |
| | 260 | Prof | A | 32 | 30 | Male | 92550 |
| 260 | 261 | AssocProf | A | 41 | 33 | Male | 88600 |
| | 262 | Prof | A | 45 | 45 | Male | 107550 |
| | 263 | Prof | A | 31 | 26 | Male | 121200 |
| | 284 | Prof | A | 45 | 43 | Male | 155865 |
| | 285 | AssocProf | A | 8 | 6 | Male | 88650 |
| 285 | 286 | AssocProf | A | 49 | 49 | Male | 81800 |
| | 287 | Prof | A | 28 | 27 | Male | 115800 |
| | 288 | AsstProf | A | 2 | 0 | Male | 85000 |
| | 298 | Prof | A | 17 | 11 | Male | 148800 |
| | 299 | Prof | A | 49 | 43 | Male | 72300 |
| 299 | 300 | AssocProf | A | 45 | 39 | Male | 70700 |
| | 301 | Prof | A | 39 | 36 | Male | 88600 |
| | 302 | Prof | A | 27 | 16 | Male | 127100 |

Índice de columnas

Índice de filas

# Seleccionar filas usando múltiples condiciones

import pandas as pd

datos = pd.read_csv('Salaries.csv')

datos.loc[(datos['yrs.service'] > 25) & (datos['rank'] == 'AssocProf'), 'yrs.since.phd' : 'sex']

**Índice de columnas**

| | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| | order | rank | discipline | yrs.since.phd | yrs.service | sex | salary |
| | 187 | AssocProf | B | 13 | 10 | Female | 103750 |
| | 188 | Prof | B | 18 | 10 | Male | 107500 |
| 188 | 189 | AssocProf | B | 28 | 28 | Male | 106300 |
| | 190 | Prof | B | 25 | 19 | Male | 153750 |
| | 193 | Prof | B | 19 | 18 | Male | 122100 |
| | 194 | AssocProf | B | 19 | 19 | Male | 86250 |
| 194 | 195 | AssocProf | B | 48 | 53 | Male | 90000 |
| | 196 | AssocProf | B | 9 | 7 | Male | 113600 |
| | 197 | AsstProf | B | 4 | 4 | Male | 92700 |
| | 259 | AsstProf | A | 9 | 3 | Male | 73800 |
| | 260 | Prof | A | 32 | 30 | Male | 92550 |
| 260 | 261 | AssocProf | A | 41 | 33 | Male | 88600 |
| | 262 | Prof | A | 45 | 45 | Male | 107550 |
| | 263 | Prof | A | 31 | 26 | Male | 121200 |
| | 284 | Prof | A | 45 | 43 | Male | 155865 |
| | 285 | AssocProf | A | 8 | 6 | Male | 88650 |
| 285 | 286 | AssocProf | A | 49 | 49 | Male | 81800 |
| | 287 | Prof | A | 28 | 27 | Male | 115800 |
| | 288 | AsstProf | A | 2 | 0 | Male | 85000 |
| | 298 | Prof | A | 17 | 11 | Male | 148800 |
| | 299 | Prof | A | 49 | 43 | Male | 72300 |
| 299 | 300 | AssocProf | A | 45 | 39 | Male | 70700 |
| | 301 | Prof | A | 39 | 36 | Male | 88600 |
| | 302 | Prof | A | 27 | 16 | Male | 127100 |

Índice de filas

# Agregando columna y dato según una condición

```python
import pandas as pd

datos = pd.read_csv('Salaries.csv')

datos.loc[datos['yrs.service'] > 25, 'Antigüedad'] = 'Jubilable'
datos.loc[datos['yrs.service'] > 25, 'yrs.since.phd':'Antigüedad']
```

| | Índice de columnas | | | | | | |
|---|---|---|---|---|---|---|---|
| | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
| | order | rank | discipline | yrs.since.phd | yrs.service | sex | salary | Antigüedad |
| 0 | 1 | Prof | B | 19 | 18 | Male | 139750 | NaN |
| 1 | 2 | Prof | B | 20 | 16 | Male | 173200 | NaN |
| 2 | 3 | AsstProf | B | 4 | 3 | Male | 79750 | NaN |
| 3 | 4 | Prof | B | 45 | 39 | Male | 115000 | Jubilable |
| 4 | 5 | Prof | B | 40 | 41 | Male | 141500 | Jubilable |
| 5 | 6 | AssocProf | B | 6 | 6 | Male | 97000 | NaN |
| 6 | 7 | Prof | B | 30 | 23 | Male | 175000 | NaN |
| 7 | 8 | Prof | B | 45 | 45 | Male | 147765 | Jubilable |
| 8 | 9 | Prof | B | 21 | 20 | Male | 119250 | NaN |
| 9 | 10 | Prof | B | 18 | 18 | Female | 129000 | NaN |
| 10 | 11 | AssocProf | B | 12 | 8 | Male | 119800 | NaN |
| 11 | 12 | AsstProf | B | 7 | 2 | Male | 79800 | NaN |
| 12 | 13 | AsstProf | B | 1 | 1 | Male | 77700 | NaN |
| 13 | 14 | AsstProf | B | 2 | 0 | Male | 78000 | NaN |
| 14 | 15 | Prof | B | 20 | 18 | Male | 104800 | NaN |

Índice de filas

# Gráfico agrupando datos

```python
import pandas as pd
import matplotlib.pyplot as plt

datos = pd.read_csv('Salaries.csv')

datos.groupby('rank')['salary'].mean().plot(kind='barh',
                    color='salmon',
                    label='Promedios por cargo')

plt.xlabel('Promedio de Salarios', weight='bold')
plt.ylabel('Cargos Docentes')
plt.title('Salarios',
      weight='bold',
      size=10)
plt.legend(title='Sueldos')
plt.plot(data=None)
plt.show()
```
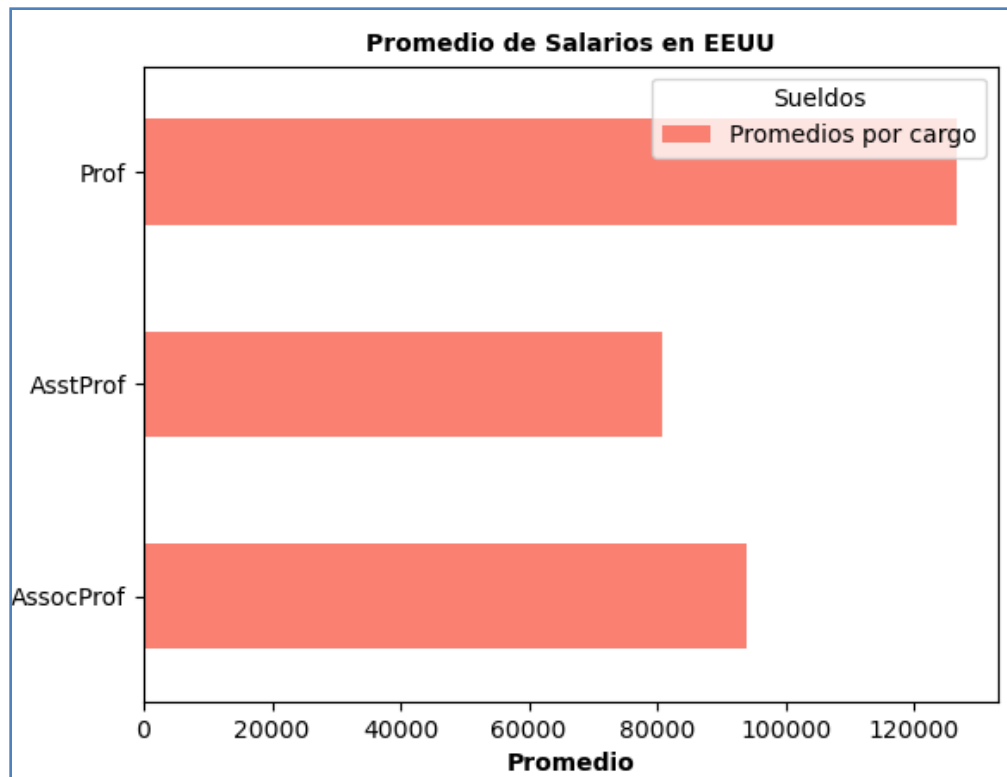
# Graficando con valores de columnas aplicando funciones

```python
import pandas as pd
import matplotlib
import matplotlib.pyplot as plt
import numpy as np

datos = pd.read_csv('Salaries.csv')
cargos = ['AssocProf', 'AsstProf','Prof']
phd_means = pd.Series(datos.groupby('rank')['yrs.since.phd'].mean())
serv_means = pd.Series(datos.groupby('rank')['yrs.service'].mean())

#Obtenemos la posicion de cada etiqueta en el eje de X
x = np.arange(len(cargos))
fig, ax = plt.subplots()
width=0.25
```
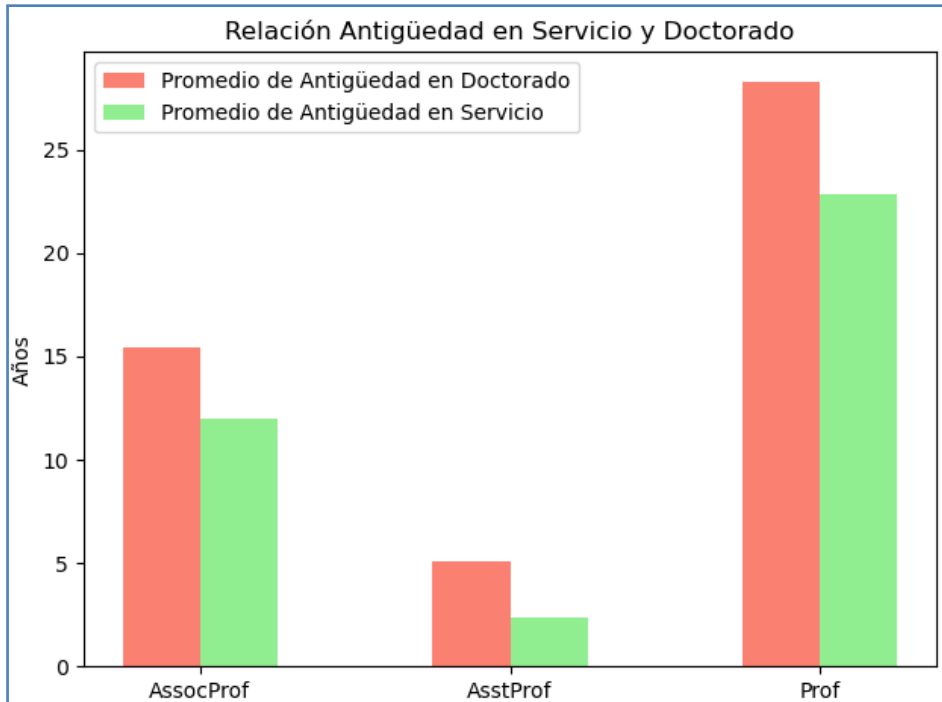
```python
#Generamos las barras para el conjunto de promedios de salarios
ax.bar(x -
 width/2, phd_means, width, label='Promedio de Antigüedad en Doctorado',color='salmon')
#Generamos las barras para el conjunto de promedios de antigüedad
ax.bar(x + width/2, serv_means, width, label='Promedio de Antigüedad en Servicio',color='lightgreen')

#Agregamos las etiquetas de identificación de valores en el grafico
ax.set_ylabel('Años')
ax.set_title('Relación Antigüedad en Servicio y Doctorado')
ax.set_xticks(x)
ax.set_xticklabels(cargos)

#Agregamos legen() para mostrar con colores a que pertence cada valor.
ax.legend()
fig.tight_layout()
plt.show()
```

https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.DataFrame.html

https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html

https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.DataFrame.plot.barh.html