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

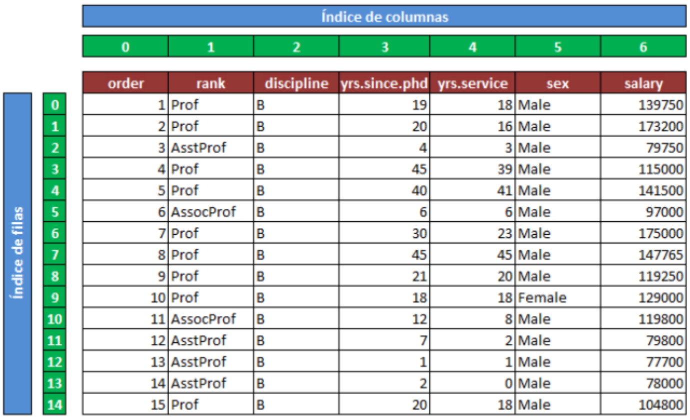
Indexación en Pandas

In [1]:

```
import pandas as pd
datos = pd.read_csv('archs/Salaries.csv')
datos.head()
```

Out[1]:

	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



Índice de columnas

Índice de filas

	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

Seleccionar una fila

In [2]:

```
datos.head(10)
```

Out[2]:

	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

In [3]:

```
datos.iloc[4]
```

Out[3]:

```
order                 5
rank               Prof
discipline            B
yrs.since.phd        40
yrs.service          41
sex                Male
salary           141500
Name: 4, dtype: object
```

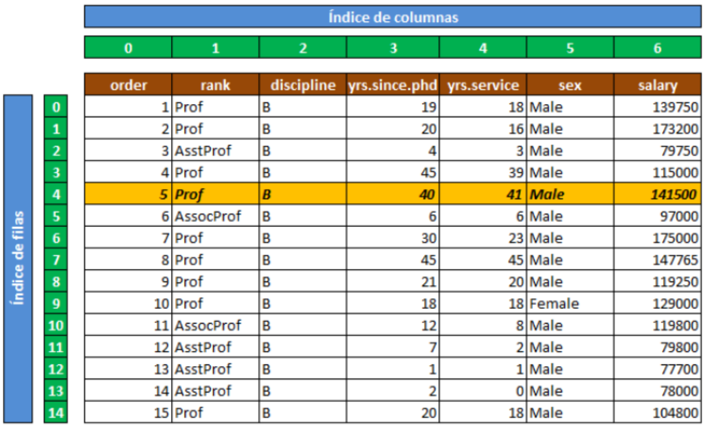
In [4]:

```
type(datos.iloc[4])
```

Out[4]:

```
pandas.core.series.Series
```

- En esta selección pandas devuelve una Serie



Índice de columnas

Índice de filas

	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

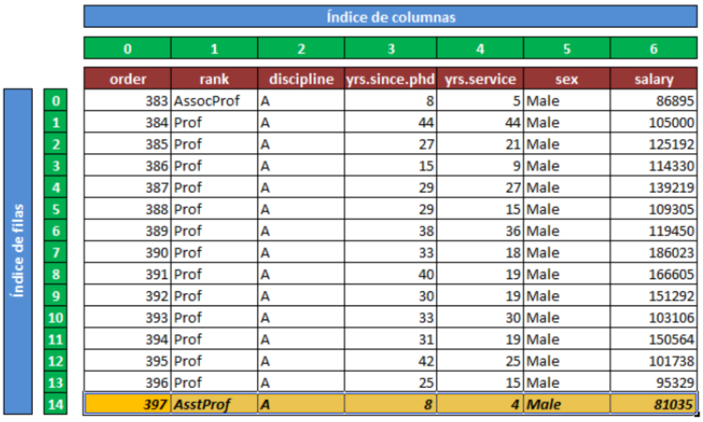
In [5]:
```
datos.iloc[-1]
```

Out[5]:
```
order                 397
rank             AsstProf
discipline              A
yrs.since.phd           8
yrs.service             4
sex                  Male
salary              81035
Name: 396, dtype: object
```

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:

In [6]:
```
datos.iloc[[-1]]
```

Out[6]:
```
     order     rank  discipline  yrs.since.phd  yrs.service   sex  salary
396    397  AsstProf          A              8            4  Male   81035
```



Índice de columnas

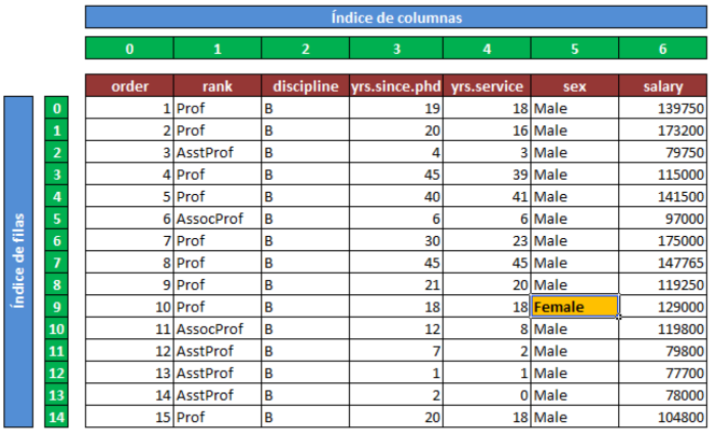
Índice de filas

	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

Seleccionar una celda específica

In [7]:
```
datos.iloc[9,5]
```

Out[7]:
```
'Female'
```



Índice de columnas

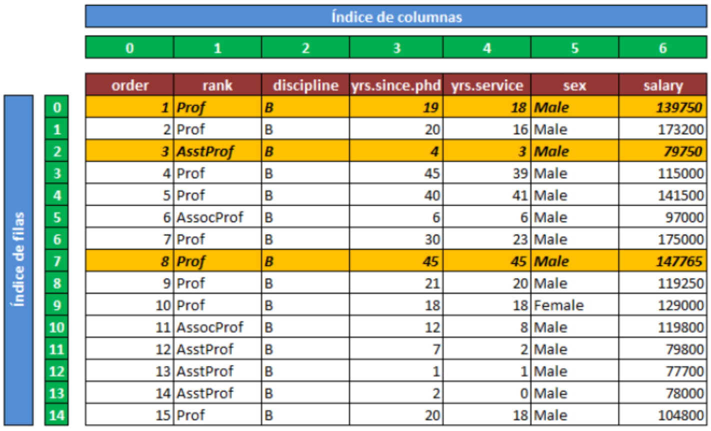
Índice de filas

	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

Seleccionar múltiples filas

In [8]:
```
datos.iloc[[7, 2, 0]]
```

Out[8]:
```
   order      rank  discipline  yrs.since.phd  yrs.service   sex  salary
7      8      Prof           B             45           45  Male  147765
2      3  AsstProf           B              4            3  Male   79750
0      1      Prof           B             19           18  Male  139750
```



Índice de columnas

Índice de filas

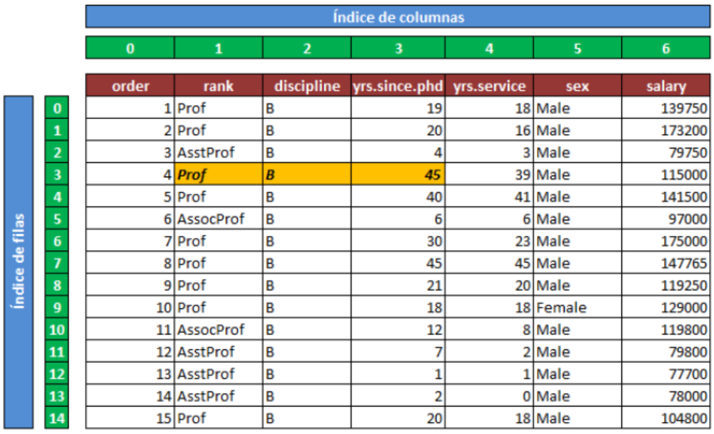
	0	1	2	3	4	5	6
	order	rank	discipline	yrs.since.phd	yrs.service	sex	salary
0	***1***	***Prof***	***B***	***19***	***18***	***Male***	***139750***
1	2	Prof	B	20	16	Male	173200
2	***3***	***AsstProf***	***B***	***4***	***3***	***Male***	***79750***
3	4	Prof	B	45	39	Male	115000
4	5	Prof	B	40	41	Male	141500
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

- En esta selección pandas devuelve un Dataframe

Seleccionar parte de los datos de una fila

In [9]:
```
datos.iloc[3, [1, 2, 3]]
```

Out[9]:
```
rank             Prof
discipline          B
yrs.since.phd      45
Name: 3, dtype: object
```



Índice de columnas

Índice de filas

	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

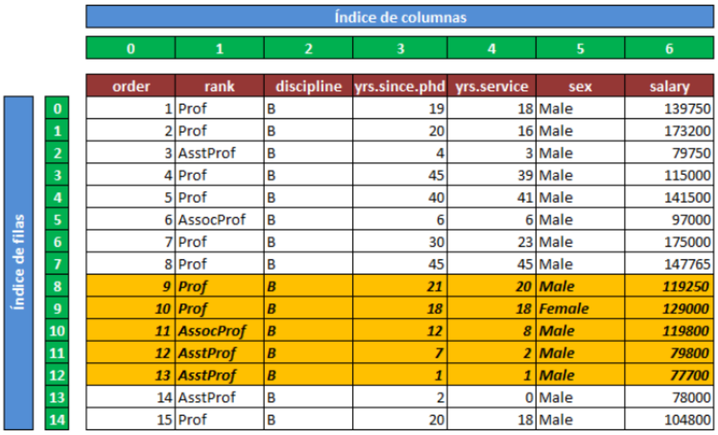
- En esta selección pandas devuelve una Serie

Seleccionar rango de filas y todas las columnas

In [10]:
```
datos.iloc[8:13]
```

Out[10]:

	order	rank	discipline	yrs.since.phd	yrs.service	sex	salary
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



Índice de columnas

Índice de filas

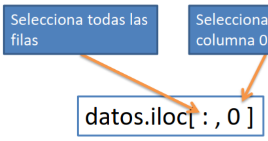
	0	1	2	3	4	5	6
	order	rank	discipline	yrs.since.phd	yrs.service	sex	salary
0	1	Prof	B	19	18	Male	139750
1	2	Prof	B	20	16	Male	173200
2	3	AsstProf	B	4	3	Male	79750
3	4	Prof	B	45	39	Male	115000
4	5	Prof	B	40	41	Male	141500
5	6	AssocProf	B	6	6	Male	97000
6	7	Prof	B	30	23	Male	175000
7	8	Prof	B	45	45	Male	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

- 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.

Seleccionar columnas

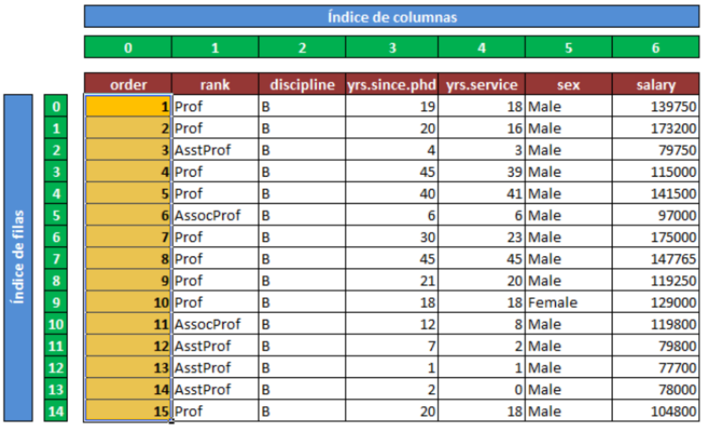
```
In [11]:  datos.iloc[:, 0]
```

```
Out[11]: 0        1
         1        2
         2        3
         3        4
         4        5
                 ...
         392    393
         393    394
         394    395
         395    396
         396    397
         Name: order, Length: 397, dtype: int64
```


Selecciona todas las filas

Selecciona la columna 0

datos.iloc[: , 0]

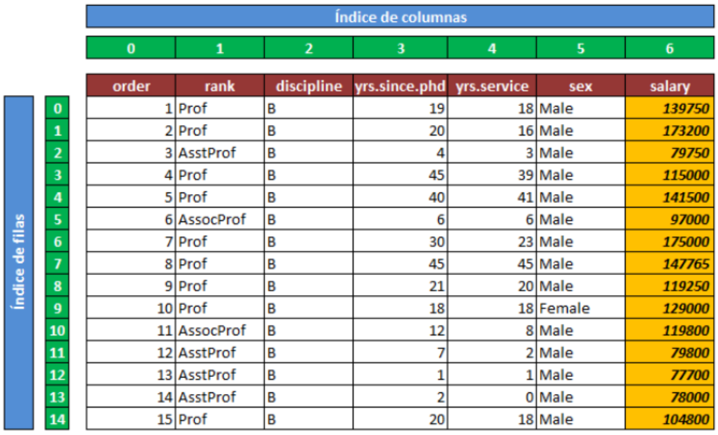

Índice de columnas

Índice de filas

	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

```
In [12]:  datos.iloc[:, -1]
```

```
Out[12]: 0      139750
         1      173200
         2       79750
         3      115000
         4      141500
                 ...
         392    103106
         393    150564
         394    101738
         395     95329
         396     81035
         Name: salary, Length: 397, dtype: int64
```


Índice de columnas

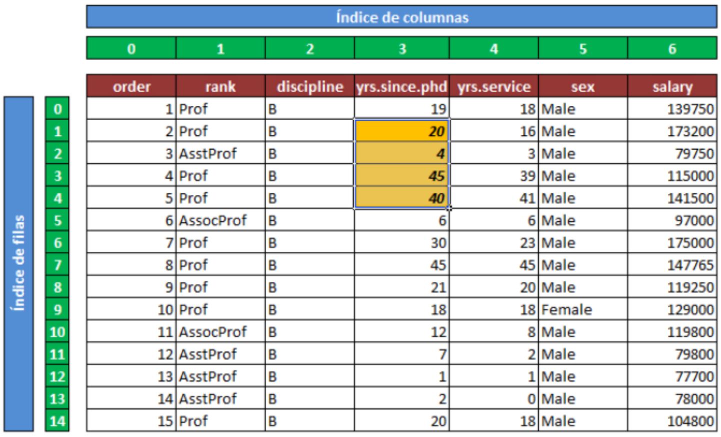
Índice de filas

	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

Seleccionar parte de filas y una columna

```
In [13]:  datos.iloc[1:5, 3]
```

```
Out[13]: 1    20
         2     4
         3    45
         4    40
         Name: yrs.since.phd, dtype: int64
```



Índice de columnas
Índice de filas

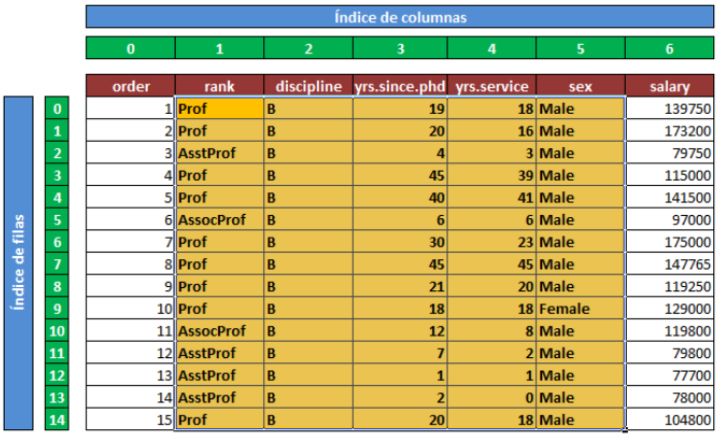
	0	1	2	3	4	5	6
	order	rank	discipline	yrs.since.phd	yrs.service	sex	salary
0	1	Prof	B	19	18	Male	139750
1	2	Prof	B	20	16	Male	173200
2	3	AsstProf	B	4	3	Male	79750
3	4	Prof	B	45	39	Male	115000
4	5	Prof	B	40	41	Male	141500
5	6	AssocProf	B	6	6	Male	97000
6	7	Prof	B	30	23	Male	175000
7	8	Prof	B	45	45	Male	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

- En esta selección pandas devuelve una Serie

Seleccionar todas las filas y un rango de columnas

In [14]:
```
datos.iloc[ : , 1: 6]
```

Out[14]:
```
       rank  discipline  yrs.since.phd  yrs.service   sex
  0    Prof           B             19           18  Male
  1    Prof           B             20           16  Male
  2 AsstProf          B              4            3  Male
  3    Prof           B             45           39  Male
  4    Prof           B             40           41  Male
...     ...         ...            ...          ...   ...
392    Prof           A             33           30  Male
393    Prof           A             31           19  Male
394    Prof           A             42           25  Male
395    Prof           A             25           15  Male
396 AsstProf          A              8            4  Male

397 rows × 5 columns
```



Índice de columnas
Índice de filas

	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

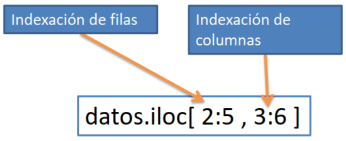
La tabla sigue...

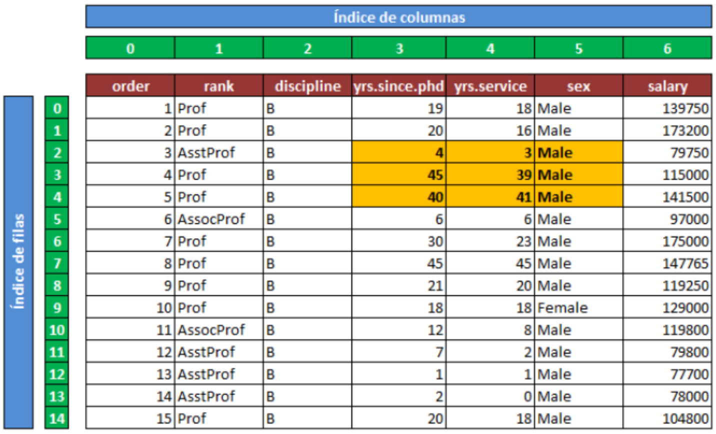
- En esta selección pandas devuelve un Dataframe

Seleccionar subconjuntos de celdas

In [15]:
```
datos.iloc[2:5, 3:6]
```

Out[15]:
```
   yrs.since.phd  yrs.service   sex
2              4            3  Male
3             45           39  Male
4             40           41  Male
```





Índice de columnas

Índice de filas	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

Slicing en Pandas con loc

In [16]:
```
import pandas as pd
datos = pd.read_csv('archs/Salaries.csv')
datos.loc[3]
```

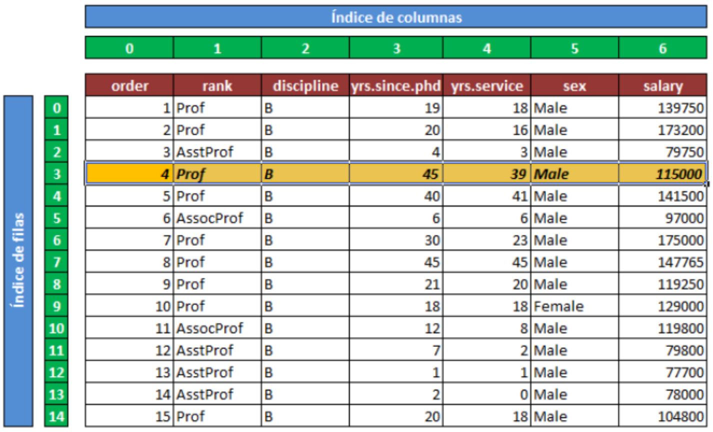
Out[16]:
```
order                 4
rank               Prof
discipline            B
yrs.since.phd        45
yrs.service          39
sex                Male
salary           115000
Name: 3, dtype: object
```

In [18]:
```
datos.loc[[3]]
```

Out[18]:
```
   order  rank  discipline  yrs.since.phd  yrs.service   sex  salary
3      4  Prof           B             45           39  Male  115000
```

In [19]:
```
type(datos.loc[[3]])
```

Out[19]:
```
pandas.core.frame.DataFrame
```



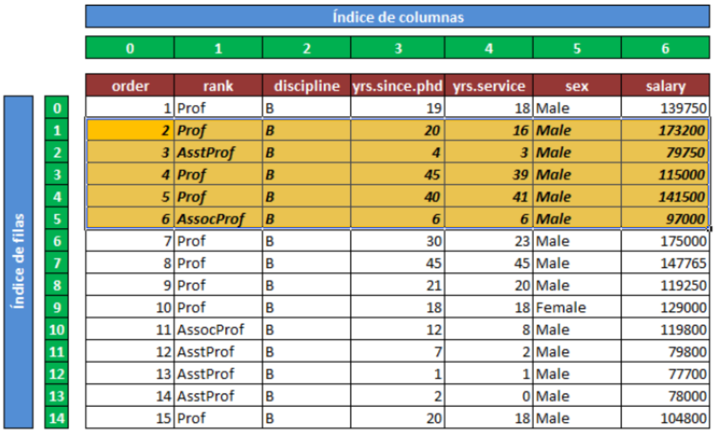
Índice de columnas

Índice de filas	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

Seleccionar subconjunto

In [20]:
```
datos.loc[1:5]
```

Out[20]:
```
   order       rank  discipline  yrs.since.phd  yrs.service   sex  salary
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
```



Índice de columnas

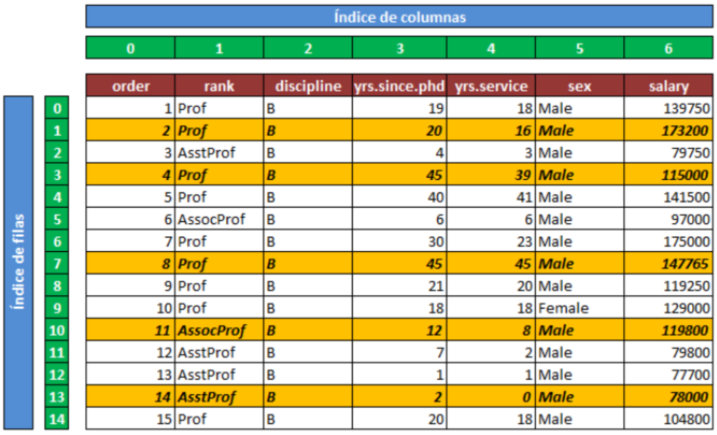
Índice de filas	0	1	2	3	4	5	6
	order	rank	discipline	yrs.since.phd	yrs.service	sex	salary
0	1	Prof	B	19	18	Male	139750
1	***2***	***Prof***	***B***	***20***	***16***	***Male***	***173200***
2	***3***	***AsstProf***	***B***	***4***	***3***	***Male***	***79750***
3	***4***	***Prof***	***B***	***45***	***39***	***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

Seleccionar filas alternadas

```
In [21]:  datos.loc[[1, 3, 7, 10, 13]]
```

Out[21]:

	order	rank	discipline	yrs.since.phd	yrs.service	sex	salary
1	2	Prof	B	20	16	Male	173200
3	4	Prof	B	45	39	Male	115000
7	8	Prof	B	45	45	Male	147765
10	11	AssocProf	B	12	8	Male	119800
13	14	AsstProf	B	2	0	Male	78000



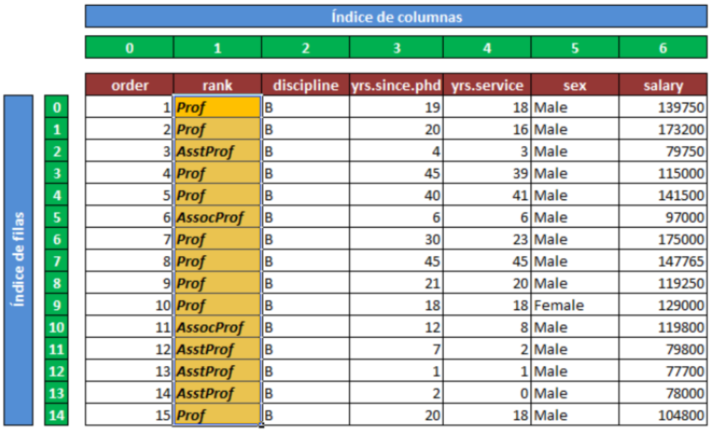
Índice de columnas

Índice de filas	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

Seleccionar con nombres de columnas

```
In [22]:  datos.loc[:, 'rank']
```

```
Out[22]: 0           Prof
         1           Prof
         2       AsstProf
         3           Prof
         4           Prof
                   ...
         392         Prof
         393         Prof
         394         Prof
         395         Prof
         396     AsstProf
         Name: rank, Length: 397, dtype: object
```



Índice de columnas

Índice de filas	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

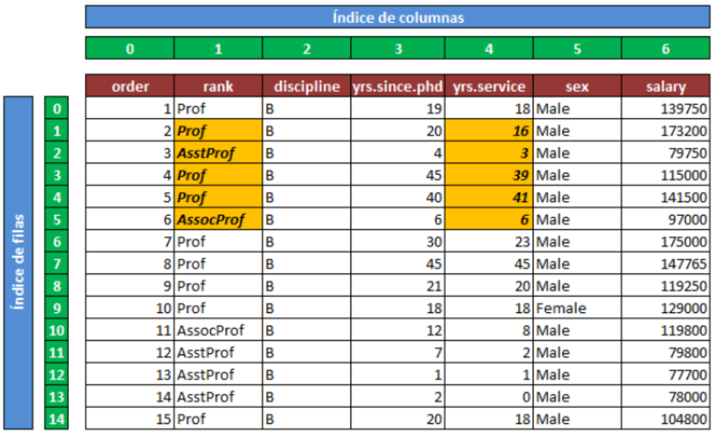
La tabla sigue...

In [23]:

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

Out[23]:

	rank	yrs.service
1	Prof	16
2	AsstProf	3
3	Prof	39
4	Prof	41
5	AssocProf	6



Índice de columnas

Índice de filas

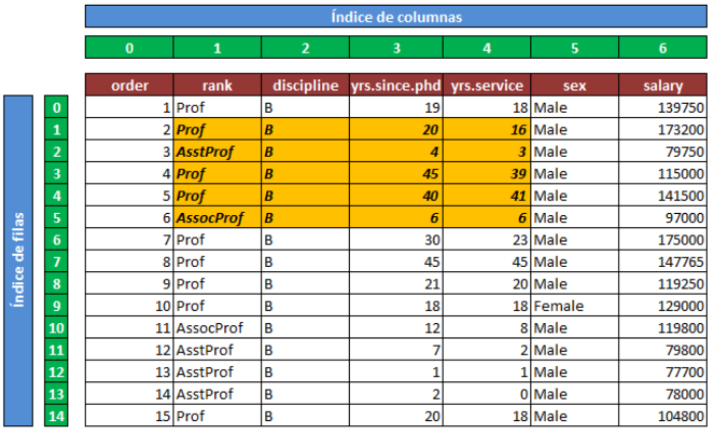
	0	1	2	3	4	5	6
	order	rank	discipline	yrs.since.phd	yrs.service	sex	salary
0	1	Prof	B	19	18	Male	139750
1	2	***Prof***	B	20	***16***	Male	173200
2	3	***AsstProf***	B	4	***3***	Male	79750
3	4	***Prof***	B	45	***39***	Male	115000
4	5	***Prof***	B	40	***41***	Male	141500
5	6	***AssocProf***	B	6	***6***	Male	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

In [24]:

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

Out[24]:

	rank	discipline	yrs.since.phd	yrs.service
1	Prof	B	20	16
2	AsstProf	B	4	3
3	Prof	B	45	39
4	Prof	B	40	41
5	AssocProf	B	6	6



Índice de columnas

Índice de filas

	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

Seleccionar con el nombre de una columna un dato determinado

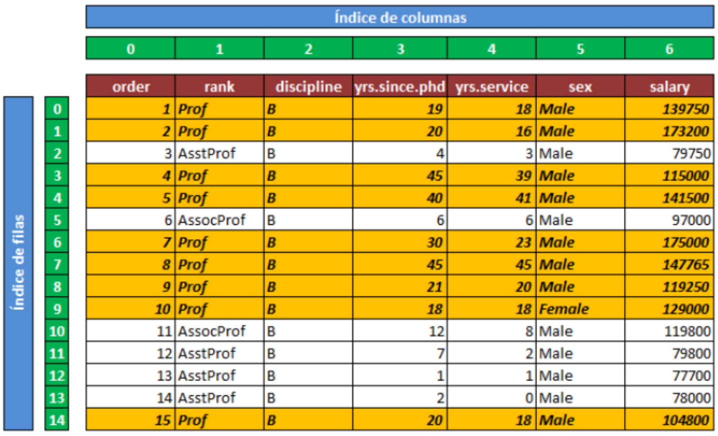
In [25]:

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

Out[25]:

	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
3	4	Prof	B	45	39	Male	115000
4	5	Prof	B	40	41	Male	141500
6	7	Prof	B	30	23	Male	175000
...	...	...	...	...	...	...	...
391	392	Prof	A	30	19	Male	151292
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

266 rows × 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

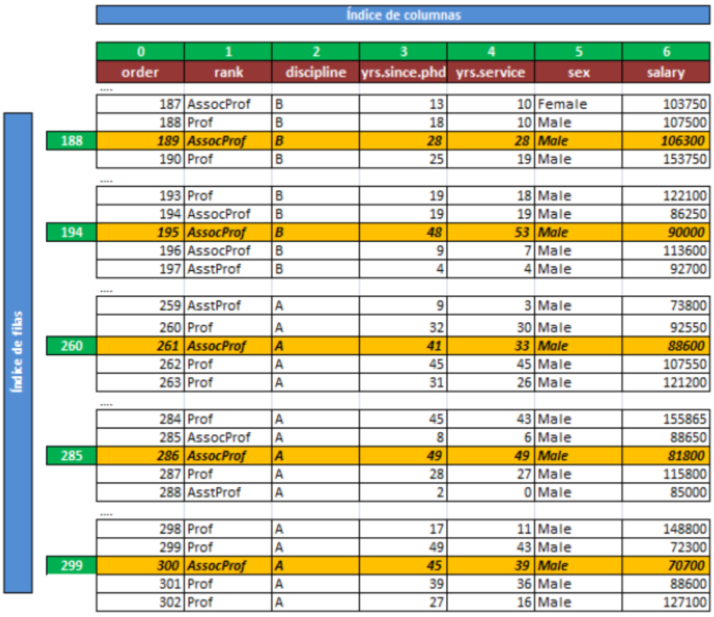
La tabla sigue...

Seleccionar filas usando múltiples condiciones

```
In [26]:  datos.loc[(datos['yrs.service'] > 25) & (datos['rank'] == 'AssocProf')]
```

Out[26]:

	order	rank	discipline	yrs.since.phd	yrs.service	sex	salary
188	189	AssocProf	B	28	28	Male	106300
194	195	AssocProf	B	48	53	Male	90000
260	261	AssocProf	A	41	33	Male	88600
285	286	AssocProf	A	49	49	Male	81800
299	300	AssocProf	A	45	39	Male	70700



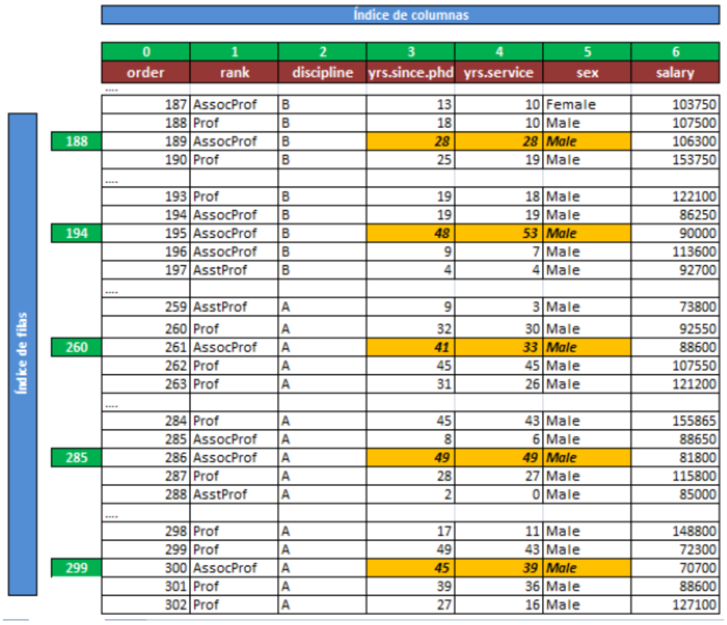
	Índice de columnas						
	0	1	2	3	4	5	6
	order	rank	discipline	yrs.since.phd	yrs.service	sex	salary
	....						
	187	AssocProf	B	13	10	Female	103750
	188	Prof	B	18	10	Male	107500
188	189	AssocProf	B	28	28	Male	106300
	190	Prof	B	25	19	Male	153750
	....						
	193	Prof	B	19	18	Male	122100
	194	AssocProf	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

```
In [27]:  datos.loc[(datos['yrs.service'] > 25) & (datos['rank'] == 'AssocProf'), 'yrs.since.phd' : 'sex']
```

Out[27]:

	yrs.since.phd	yrs.service	sex
188	28	28	Male
194	48	53	Male
260	41	33	Male
285	49	49	Male
299	45	39	Male



Índice de columnas

Índice de filas

	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

Agregando columna y dato según una condición

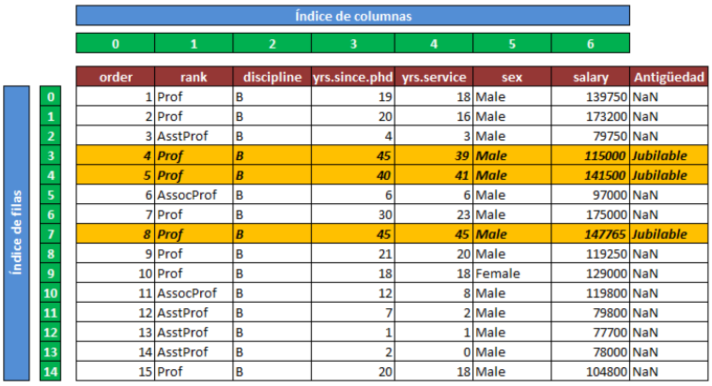
```
In [28]:  datos.loc[datos['yrs.service'] > 25, 'Antigüedad'] = 'Jubilable'
```

```
In [29]:  datos.loc[datos['yrs.service'] > 25, 'yrs.since.phd':'Antigüedad']
          datos
```

Out[29]:

	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
...	...	...	...	...	...	...	...	...
392	393	Prof	A	33	30	Male	103106	Jubilable
393	394	Prof	A	31	19	Male	150564	NaN
394	395	Prof	A	42	25	Male	101738	NaN
395	396	Prof	A	25	15	Male	95329	NaN
396	397	AsstProf	A	8	4	Male	81035	NaN

397 rows × 8 columns



Índice de columnas

Índice de filas

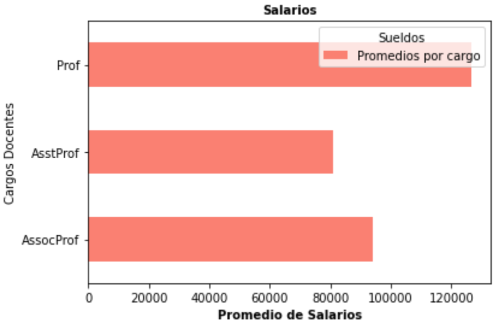
	0	1	2	3	4	5	6	
	order	rank	discipline	yrs.since.phd	yrs.service	sex	salary	Antigüedad
0	1	Prof	B	19	18	Male	139750	NaN
1	2	Prof	B	20	16	Male	173200	NaN
2	3	AsstProf	B	4	3	Male	79750	NaN
3	*4*	*Prof*	*B*	*45*	*39*	*Male*	*115000*	*Jubilable*
4	*5*	*Prof*	*B*	*40*	*41*	*Male*	*141500*	*Jubilable*
5	6	AssocProf	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

La tabla sigue...

Gráfico agrupando datos

In [30]:

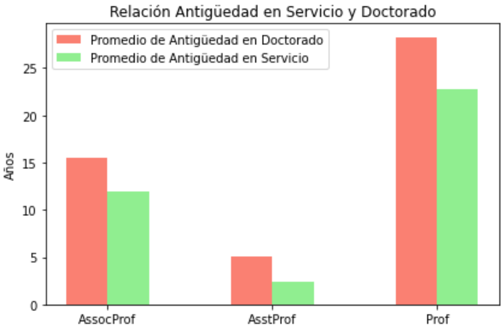
```python
import pandas as pd
import matplotlib.pyplot as plt

datos = pd.read_csv('archs/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)
```



Graficando con valores de columnas aplicando funciones

In [31]:

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

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

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
cargos = ['AssocProf', 'AsstProf','Prof']
x = np.arange(len(cargos))
fig, ax = plt.subplots()
width=0.25

#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 gráfico
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 pertenece cada valor.
ax.legend()
fig.tight_layout()
```



In [ ]:

```python

```