

# ¿Preguntas?

**1<sup>a</sup>**  
**Emisión**

# DATA SCIENCE

## **Módulo 05** Manipulación y visualización de datos con Python

*Mtro. Ricardo Daniel Alanis Tamez*



**UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO**  
Dirección General de Cómputo y de Tecnologías de información y Comunicación  
Dirección de Docencia en TIC



Educación  
Continua  
1971 - 2021

# Filtrado y selección de datos

## **Manipulación y visualización de datos con Python**

Ricardo Alanís

# Presentación

En este módulo tendremos la oportunidad de ver las diferentes maneras de acceder y filtrar la información de un dataframe

# Objetivo

El participante identificará los diferentes métodos para filtrar y seleccionar datos

# Agenda de Hoy

- 1. Paseo por los filtros posibles disponibles**
- 2. Ejemplos de filtrados**
- 3. Exposición de la selección de datos**
- 4. Ejemplos de selección de datos y temas para estar alertas**

### By Position

```
>>> df.iloc([0],[0])  
'Belgium'  
>>> df.iat([0],[0])  
'Belgium'
```

### By Label

```
>>> df.loc([0], ['Country'])  
'Belgium'  
>>> df.at([0], ['Country'])  
'Belgium'
```

### By Label/Position

```
>>> df.ix[2]  
Country      Brazil  
Capital      Brasília  
Population    207847528  
  
>>> df.ix[:, 'Capital']  
0      Brussels  
1    New Delhi  
2    Brasília
```

Select single value by row & column

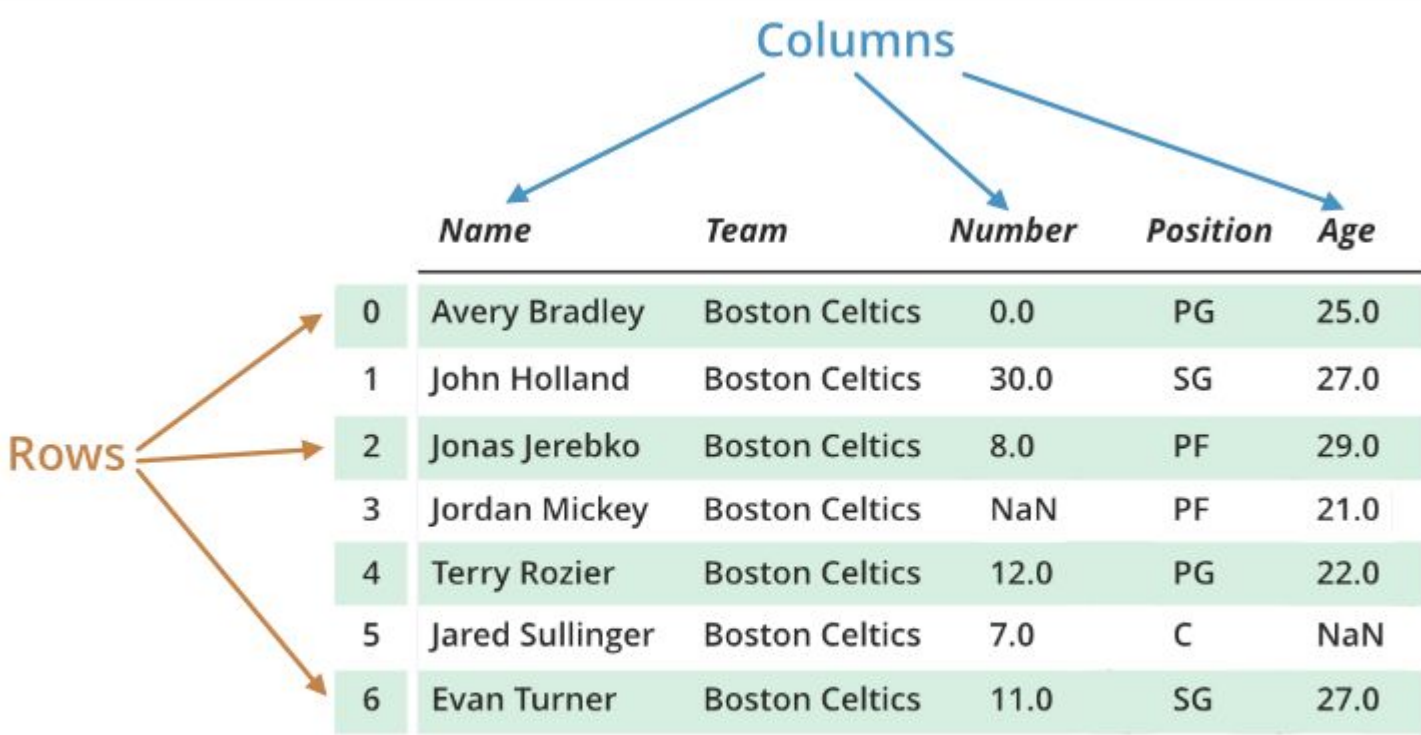
Select single value by row & column labels

Select single row of subset of rows

Select a single column of subset of columns



# Estructura de un dataframe



The diagram illustrates the structure of a dataframe. It features a table with 7 rows and 6 columns. The columns are labeled *Name*, *Team*, *Number*, *Position*, and *Age*. The rows are indexed from 0 to 6. Blue arrows point from the word "Columns" to the column headers. Orange arrows point from the word "Rows" to the row indices. A green logo is located in the bottom right corner of the diagram area.

	<i>Name</i>	<i>Team</i>	<i>Number</i>	<i>Position</i>	<i>Age</i>
0	Avery Bradley	Boston Celtics	0.0	PG	25.0
1	John Holland	Boston Celtics	30.0	SG	27.0
2	Jonas Jerebko	Boston Celtics	8.0	PF	29.0
3	Jordan Mickey	Boston Celtics	NaN	PF	21.0
4	Terry Rozier	Boston Celtics	12.0	PG	22.0
5	Jared Sullinger	Boston Celtics	7.0	C	NaN
6	Evan Turner	Boston Celtics	11.0	SG	27.0



# Obtener datos

Obtener un elemento **df['col']**

Obtener un subconjunto de un dataframe **df[1:]**

# Seleccionando Datos

Por posición **`df.iloc([0],[0])`** y **`df.iat([0],[0])`**

Por etiqueta **`df.loc([0], ['Country'])`** y **`df.at([0], ['Country'])`**

Indexación Booleana **`s[-(s>1)]`**, **`s[(s<1) | (s>2)]`**, **`df[df['Population']>2]`**

Asignando datos **`s['a'] = 6`**

# Manipulación de Datos, alertas

Cuando asignas datos a una columna, se aplica a todas las observaciones.

No soy muy fanático del reemplazo de datos “inplace” en un lugar, ya que puede llevar a confusiones - ya que los datos se están filtrando para modificarlos

Mi preferencia está en la aplicación de funciones sobre columnas (o en el peor de los casos filas) con condiciones que permitan generar un nuevo conjunto de datos

Es más costoso en memoria, pero es más certero para la generación de datos y su orden, pero esa es una opinión.

# Referencias

- Moreno, Amanda Iglesias. “Data Filtering in Pandas.” Towards data science. 16 Sep 2020. [towardsdatascience.com/filtering-data-frames-in-pandas-b570b1f834b9](https://towardsdatascience.com/filtering-data-frames-in-pandas-b570b1f834b9)
- “pandas.DataFrame.filter.” pandas. [pandas.pydata.org/docs/reference/api/pandas.DataFrame.filter.html](https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.filter.html)

# Contacto

*Mtro. Ricardo Daniel Alanis Tamez*

[ricardo@codeandomexico.org](mailto:ricardo@codeandomexico.org)

*LinkedIn: Ricardo Alanís*