



# Argentina Programa

Introducción Pandas



# Que es un Data Set?

Un data set, también conocido como conjunto de datos, es una colección de datos organizados y estructurados que se utilizan en el análisis, la investigación o la creación de modelos y algoritmos en diversas disciplinas, como la ciencia de datos, la inteligencia artificial, el aprendizaje automático y la estadística.

Un data set puede contener diferentes tipos de información, como números, texto, imágenes, audio u otros formatos de datos. Estos datos se recopilan y almacenan en un formato coherente y estructurado para que puedan ser procesados y analizados de manera eficiente.

# Pandas

Pandas es una biblioteca de Python que proporciona estructuras de datos y herramientas de análisis de datos de alto rendimiento. Es ampliamente utilizada en el campo de la ciencia de datos y análisis para manipular y analizar datos de manera eficiente.

Pandas



# Tipos de Datos De Pandas

- Series: Es una estructura de datos unidimensional similar a un array o una lista con etiquetas de datos, llamadas índices. Puedes pensar en ella como una columna en una hoja de cálculo.
- DataFrame: Es una estructura de datos bidimensional que se asemeja a una tabla de bases de datos o una hoja de cálculo con filas y columnas. Un DataFrame puede contener múltiples series y cada una de ellas puede tener un tipo de dato diferente.

The diagram illustrates a DataFrame structure. It features a table with 7 rows and 5 columns. The columns are labeled 'Name', 'Team', 'Number', 'Position', and 'Age'. The rows are indexed from 0 to 6. Annotations include: 'Columns' with arrows pointing to the column headers; 'Rows' with arrows pointing to the row indices; and 'Data' with a box highlighting a specific cell (Jonas Jerebko, 8.0) and its corresponding row and column headers.

	Name	Team	Number	Position	Age
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



# Funciones para Exploración de Datos

`Head()` y `tail()`: Muestra las primeras (por defecto 5) o las últimas filas del DataFrame

`shape`: Devuelve una tupla con el número de filas y columnas del DataFrame.

`info()`: Proporciona información sobre el DataFrame, incluyendo el tipo de datos de cada columna y si hay valores nulos.

`describe()`: Proporciona estadísticas resumidas de las columnas numéricas, como la media, la mediana, el mínimo, el máximo y los percentiles.

`value_counts()`: Cuenta las ocurrencias únicas en una columna.



# Funciones Para Exploración de Datos

`isnull()` y `sum()`: Verifica si hay valores nulos y cuenta cuántos hay por columna.

`corr()`: Calcula la matriz de correlación entre las columnas numéricas del DataFrame.