

1^a
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

El Concepto de DataFrame y su API

Manipulación y visualización de datos con Python

Ricardo Alanís

Presentación

En este tema seguiremos adentrándonos en las características de la estructura de datos más común en Pandas: El DataFrame. Aprenderemos a utilizarlo para explorar y manipular la información, conoceremos sus limitaciones para luego practicar con él.

Objetivo

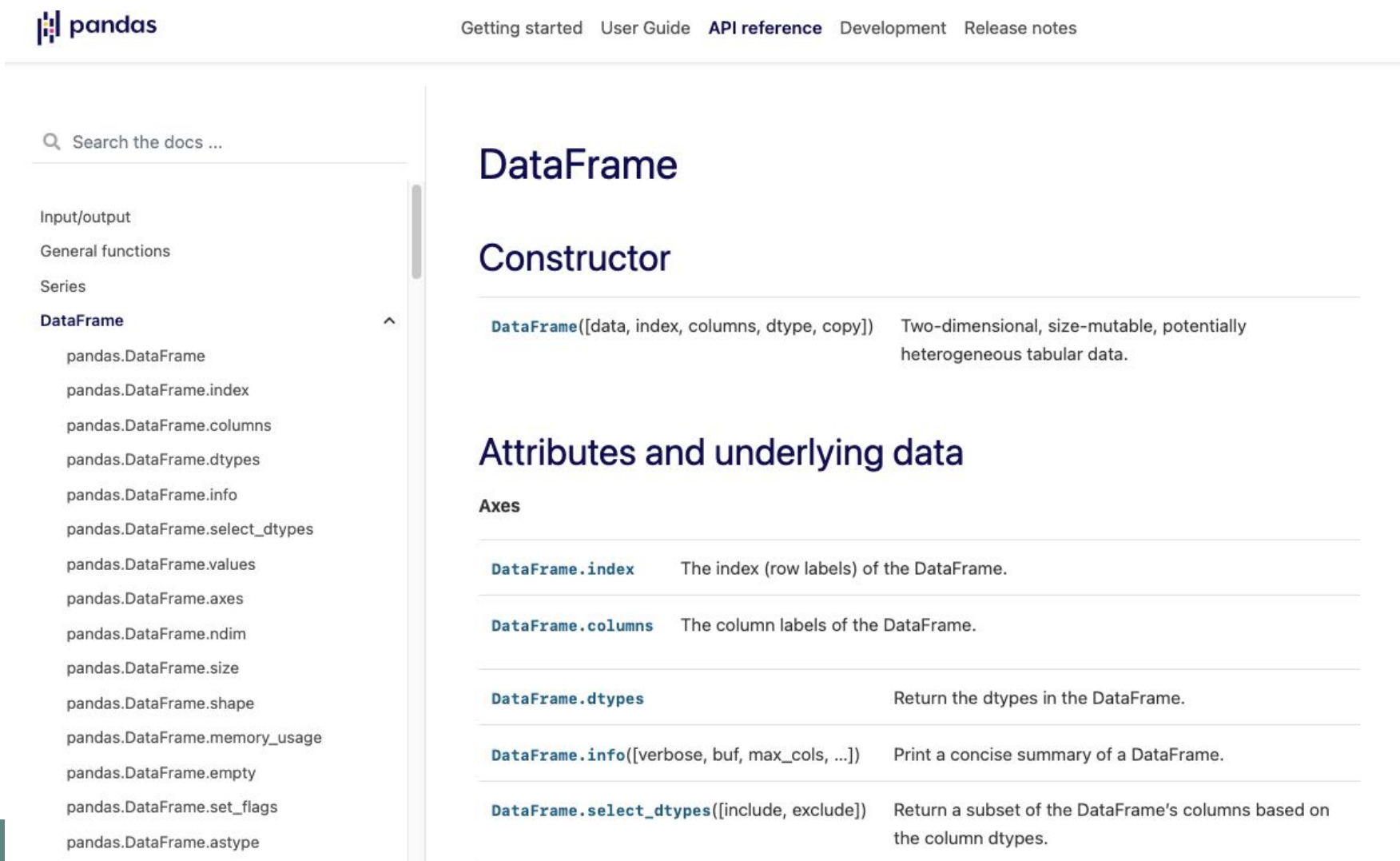
El participante identificará el valor de la abstracción del Data Frame como componente importante Pandas.

Además, explorará de manera Individual al Data Frame usando un ejercicio dirigido.

Agenda de hoy

- 1. Lectura en equipo de documentación de DataFrame**
- 2. Exposición del concepto de Dataframe**
- 3. Ventajas y desventajas del dataframe**
- 4. Ejercicio individual de exploración de un dataframe**
- 5. Compartir resultados y cerrar clase.**

Entendamos qué es el DataFrame



The screenshot shows the pandas documentation website. The top navigation bar includes links for 'Getting started', 'User Guide', 'API reference', 'Development', and 'Release notes'. The left sidebar contains a search bar and a list of categories: 'Input/output', 'General functions', 'Series', and 'DataFrame'. The 'DataFrame' category is selected, showing a list of attributes and methods. The main content area is titled 'DataFrame' and 'Constructor', followed by a description of the `DataFrame` constructor. Below this, the 'Attributes and underlying data' section lists several attributes and methods with their descriptions.

pandas Getting started User Guide **API reference** Development Release notes

Search the docs ...

Input/output
General functions
Series
DataFrame

- `pandas.DataFrame`
- `pandas.DataFrame.index`
- `pandas.DataFrame.columns`
- `pandas.DataFrame.dtypes`
- `pandas.DataFrame.info`
- `pandas.DataFrame.select_dtypes`
- `pandas.DataFrame.values`
- `pandas.DataFrame.axes`
- `pandas.DataFrame.ndim`
- `pandas.DataFrame.size`
- `pandas.DataFrame.shape`
- `pandas.DataFrame.memory_usage`
- `pandas.DataFrame.empty`
- `pandas.DataFrame.set_flags`
- `pandas.DataFrame.astype`

DataFrame

Constructor

`DataFrame([data, index, columns, dtype, copy])` Two-dimensional, size-mutable, potentially heterogeneous tabular data.

Attributes and underlying data

Axes

<code>DataFrame.index</code>	The index (row labels) of the DataFrame.
<code>DataFrame.columns</code>	The column labels of the DataFrame.
<code>DataFrame.dtypes</code>	Return the dtypes in the DataFrame.
<code>DataFrame.info([verbose, buf, max_cols, ...])</code>	Print a concise summary of a DataFrame.
<code>DataFrame.select_dtypes([include, exclude])</code>	Return a subset of the DataFrame's columns based on the column dtypes.

Ejercicio 3.1: Entendamos qué es el DataFrame

- Vamos a separarnos en grupos
- Cada quien va a leer la documentación (10 minutos), encuentra:
 - Propiedades más importantes
 - Funciones más importantes
 - Agrupaciones que puedes detectar
 - Qué cambiarías de la estructura de la implementación del DF
- Cada quien expone las partes más importantes
- Tomen notas para compartir con los demás miembros y entregar al instructor
- Discutamos todos en el grupo al terminar

La potente abstracción del DataFrame

- Discutamos:
 - Por qué hacer estas operaciones nos habla de la potencia de Pandas
 - Tema de programación: La abstracción correcta

Demos un paseo por las funciones disponibles para entender su poder.

Ventajas y Desventajas del DataFrame

Ventajas:

- Forma práctica de trabajar la data
- Abstracción con muchas facilidades
- Carga de datos (no lazy, en pandas)
- Manipulación posible por fila y columna

Desventajas:

- Data Rectangular
- Abstracción con funciones a distintas utilidades
- Cuando los datos son más que la memoria, limitaciones
- Overhead en uso de memoria

Práctica 3.1: Manos a la obra

Completa los siguientes retos:

- Carga el conjunto de datos de mortalidad materna datos.gob.mx
- Traete la serie de una columna de datos
- Traete la primera fila de observaciones
- Filtra las observaciones con respecto al valor de una columna
- Aplica una función a una columna
- Concatenar dos columnas
- Nivel Experto: Limpia el conjunto de datos

Deadline 2:30 PM

Comparte con el resto
del equipo

¿Preguntas?

Referencias

- “DataFrame.” Pandas. pandas.pydata.org/docs/reference/frame.html.

Contacto

Mtro. Ricardo Daniel Alanis Tamez

ricardo@codeandomexico.org

LinkedIn: Ricardo Alanís