

Curso Pandas 1.x en español

Kevin Farinango (Asuskf)

2020-08-26

Contents

1	Introducción	5
1.1	Para quién es esta guía	5
1.2	Qué cubre esta guía	5
2	Fundamentos de Pandas	7
3	Literature	9
4	Creación y persistencia de un Dataframe	11
5	Empezando con la ciencia de datos	13
5.1	Python	13
6	Explorando el Dataframe	15

Chapter 1

Introducción

Pandas es una librería para creación y manipulación de datos con Python. Está librería manipula datos tabulares en filas y columnas cómo encontrarías en una base de datos o en una hoja de cálculo. Pandas es aprovechado por: **Científicos de datos, analistas, programadores, ingenieros que lo usan para moldear sus datos.**

Pandas se limita a “**datos pequeños**” (datos que pueden caber en la memoria de una sola máquina). Sin embargo, la sintaxis y las operaciones se han adoptado o inspirado a otros proyectos: PySpark, Dask, Modin, cuDF, Baloo, Dexplo, Tabel, StaticFrame, entre otros.

1.1 Para quién es esta guía

Esta guía fue pensada en principiantes de habla hispana que estén empezando en el mundo de la ciencia de datos. Esta guía tiene la intención de dotar al principiante de los conocimientos necesarios para manipular y analizar sus datos con una herramienta potente y de fácil uso como es Pandas.

La presente guía tiene más de 50 ejercicios que van de básico a intermedio con los cuales el lector podrá poner en práctica e interiorizar lo aprendido. Para mejor comprensión cada ejercicio se detalla pasos a paso.

1.2 Qué cubre esta guía

- **Capítulo 1, Introducción:** Descripción de lo que tiene el libro y para quien va dirigido.
- **Capítulo 2, Fundamentos de Pandas:** Presenta la anatomía y el vocabulario utilizado para identificar las dos estructuras de datos principales de pandas, la Serie y el DataFrame con sus respectivos métodos.

- **Capítulo 3, Operaciones esenciales en un Dataframe:** Se centra en las operaciones comunes que se realizan durante el análisis de datos.
- **Capítulo 4, Importación y creación de Dataframes:** Analiza las diversas formas de importar datos y crea DataFrames.
- **Capítulo 5, Empezando con la ciencia de datos:** Presenta técnicas de análisis básicas que sirven para comparar números y datos categóricos. Este capítulo también demostrará técnicas de visualización comunes.
- **Capítulo 6, Trabajando con subconjuntos:** Cubre las muchas formas variadas y potencialmente confusas de seleccionar diferentes subconjuntos de datos.
- **Capítulo 7, Filtrando los subconjuntos:** Explica varias formas de cómo seleccionar diferentes subconjuntos de datos.
- **Capítulo 8, Consultas booleanas:** Consultas basadas en condiciones booleanas
- **Capítulo 9, el uso de index en Pandas:** El uso correcto de index para evitar resultados erróneos.
- **Capítulo 10, Agrupación:** Presenta métodos de agrupación y la construcción de funciones personalizadas para aplicar a diferentes grupos
- **Capítulo 11, Ordenando el dataframe:** Explica qué son los datos ordenados, por qué son tan importantes y que métodos pueden ser usados para transformar conjuntos de datos desordenados en ordenados.
- **Capítulo 12, Combinando objetos Pandas:** Presenta varios métodos para combinar DataFrames y Series vertical u horizontalmente.
- **Capítulo 13, Trabajando con las series de tiempo:** Presenta distintos métodos para trabajar con las series de tiempo en cualquier dimensión de tiempo posible
- **Capítulo 14, Visualización de datos:** Tips para crear visualizaciones estéticamente agradables usando: Pandas, Seaborn y Matplotlib
- **Capítulo 15, Test del dataframe:** Explora los mecanismos de test de los DataFrames y los resultados que presenta Pandas.

Chapter 2

Fundamentos de Pandas

You can label chapter and section titles using `{#label}` after them, e.g., we can reference Chapter 2. If you do not manually label them, there will be automatic labels anyway, e.g., Chapter ??.

Figures and tables with captions will be placed in `figure` and `table` environments, respectively.

```
par(mar = c(4, 4, .1, .1))  
plot(pressure, type = 'b', pch = 19)
```

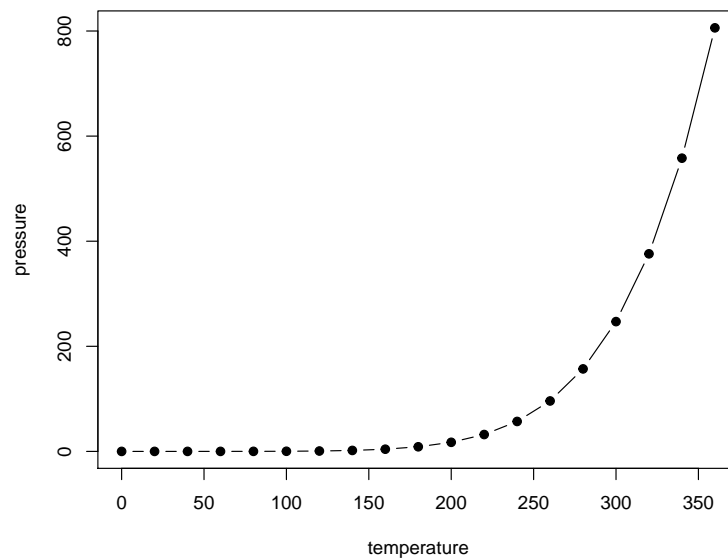


Figure 2.1: Here is a nice figure!

Table 2.1: Here is a nice table!

Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
5.1	3.5	1.4	0.2	setosa
4.9	3.0	1.4	0.2	setosa
4.7	3.2	1.3	0.2	setosa
4.6	3.1	1.5	0.2	setosa
5.0	3.6	1.4	0.2	setosa
5.4	3.9	1.7	0.4	setosa
4.6	3.4	1.4	0.3	setosa
5.0	3.4	1.5	0.2	setosa
4.4	2.9	1.4	0.2	setosa
4.9	3.1	1.5	0.1	setosa
5.4	3.7	1.5	0.2	setosa
4.8	3.4	1.6	0.2	setosa
4.8	3.0	1.4	0.1	setosa
4.3	3.0	1.1	0.1	setosa
5.8	4.0	1.2	0.2	setosa
5.7	4.4	1.5	0.4	setosa
5.4	3.9	1.3	0.4	setosa
5.1	3.5	1.4	0.3	setosa
5.7	3.8	1.7	0.3	setosa
5.1	3.8	1.5	0.3	setosa

Reference a figure by its code chunk label with the `fig:` prefix, e.g., see Figure 2.1. Similarly, you can reference tables generated from `knitr::kable()`, e.g., see Table 2.1.

```
knitr::kable(
  head(iris, 20), caption = 'Here is a nice table!',
  booktabs = TRUE
)
```

You can write citations, too. For example, we are using the **bookdown** package (Xie, 2020) in this sample book, which was built on top of R Markdown and **knitr** (Xie, 2015).

Chapter 3

Literature

Here is a review of existing methods.

Chapter 4

Creación y persistencia de un Dataframe

We describe our methods in this chapter.

Chapter 5

Empezando con la ciencia de datos

Some *significant* applications are demonstrated in this chapter.

5.1 Python

Chapter 6

Explorando el Dataframe

We have finished a nice book.

Bibliography

Xie, Y. (2015). *Dynamic Documents with R and knitr*. Chapman and Hall/CRC, Boca Raton, Florida, 2nd edition. ISBN 978-1498716963.

Xie, Y. (2020). *bookdown: Authoring Books and Technical Documents with R Markdown*. R package version 0.20.