

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

Lectura de Datos

Manipulación y visualización de datos con Python

Ricardo Alanís

Presentación

En este tema tendremos la oportunidad de aprender la interacción de Python con distintas fuentes de datos; y entraremos a detalles con algunas ventajas y desventajas de distintas estructuras de información disponibles.

Objetivo

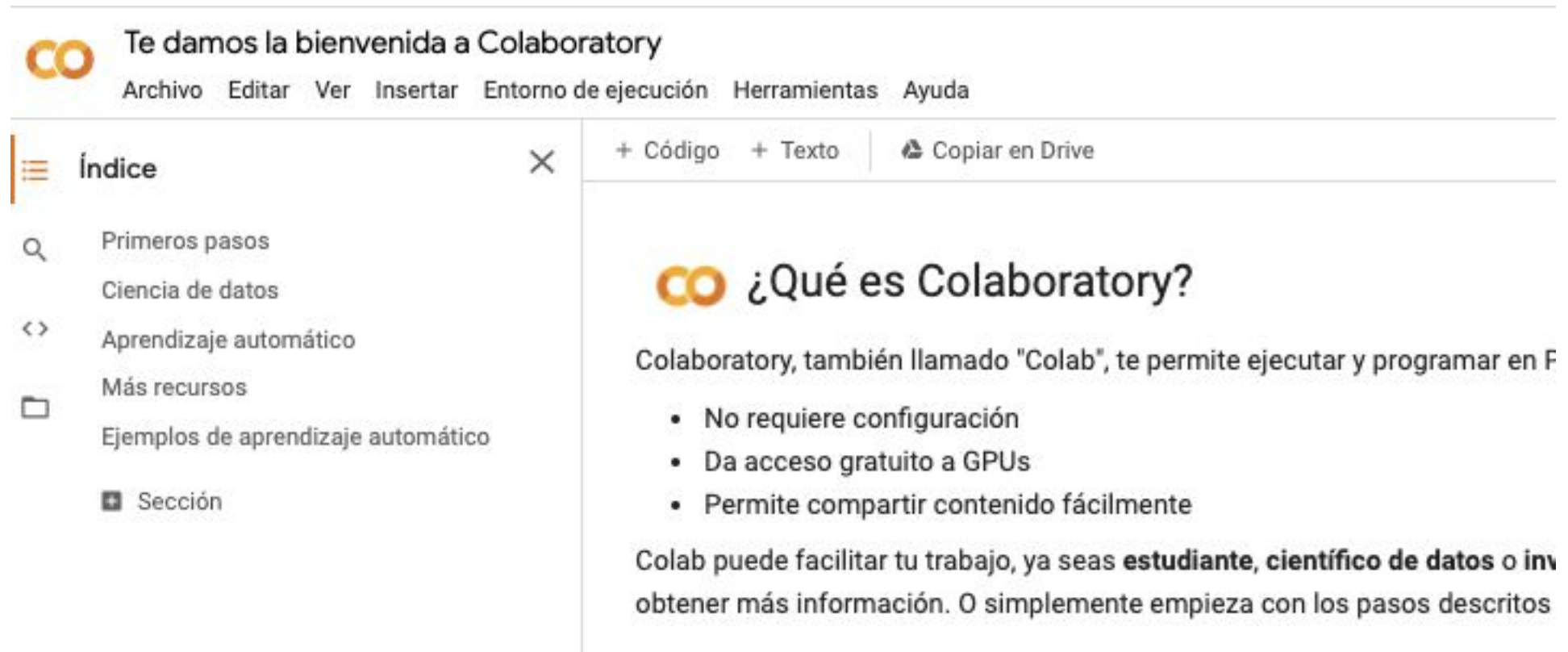
El participante identificará el rol de Pandas como un lector universal de datos.

Además, tendrá exposición a los distintos tipos de datos disponibles y su utilidad.

Agenda de hoy

- 1. Cómo leer un conjunto de datos usando la API de pandas**
- 2. Reto de lectura de datos - fuentes de datos disponibles y su relativa utilidad**
- 3. Exposición de refuerzo de los distintos tipos de datos**

Usemos el material de la clase



Te damos la bienvenida a Colaboratory

Archivo Editar Ver Insertar Entorno de ejecución Herramientas Ayuda

Índice

- Primeros pasos
- Ciencia de datos
- Aprendizaje automático
- Más recursos
- Ejemplos de aprendizaje automático
- Sección

+ Código + Texto Copiar en Drive

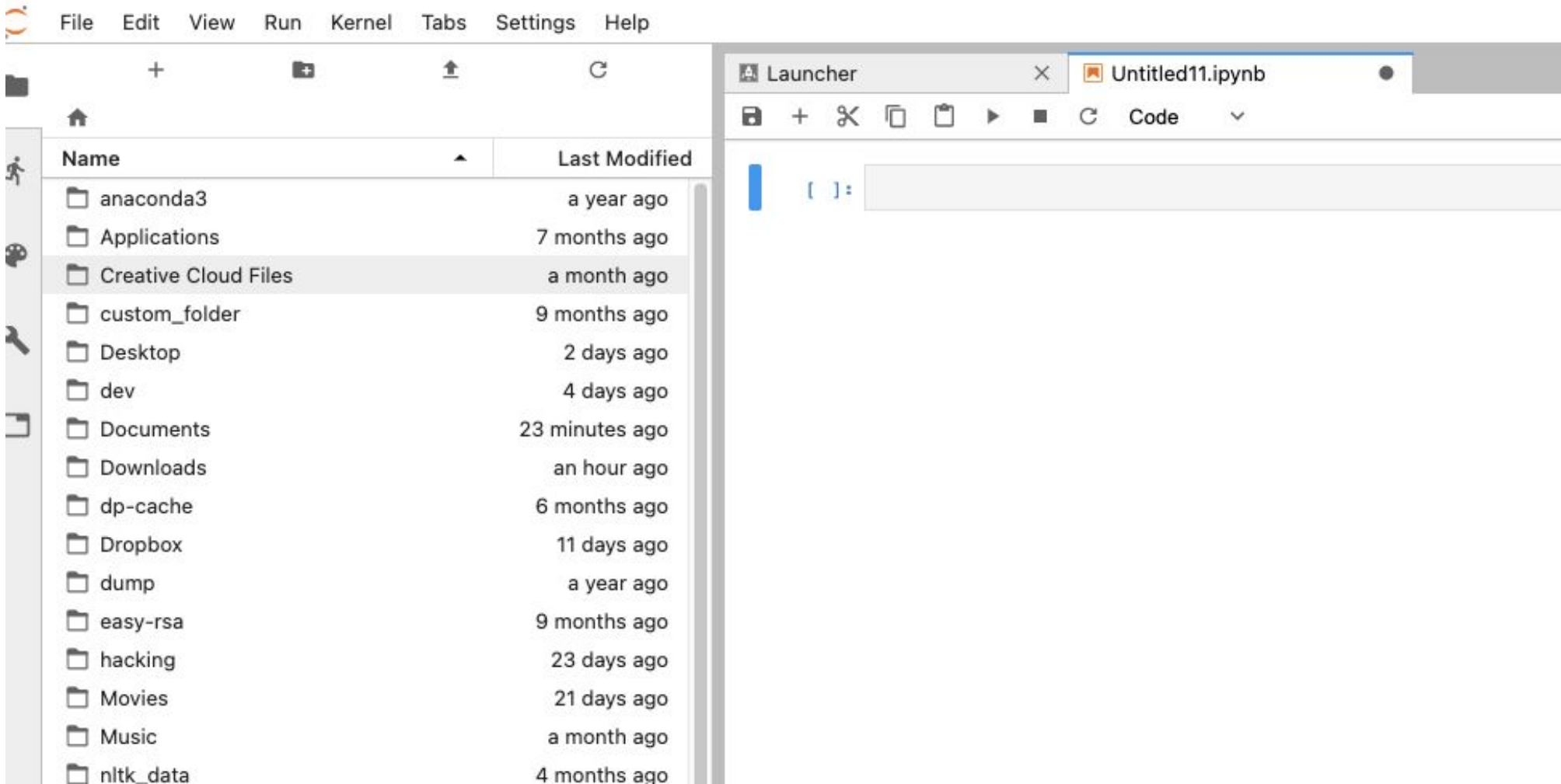
¿Qué es Colaboratory?

Colaboratory, también llamado "Colab", te permite ejecutar y programar en F

- No requiere configuración
- Da acceso gratuito a GPUs
- Permite compartir contenido fácilmente

Colab puede facilitar tu trabajo, ya seas **estudiante, científico de datos o inv** obtener más información. O simplemente empieza con los pasos descritos

Definiendo tu espacio de trabajo virtual



Demo: Cargando tu primer dataset

```
[1] import pandas as pd
```

Double-click (or enter) to edit

```
[12] # Cargando datos desde una estructura de datos de python
pd.DataFrame({'a':[1,2,3], 'b':[1,2,3]})
```

	a	b
0	1	1
1	2	2
2	3	3

```
# Cargando datos desde otra estructura de datos de python
pd.DataFrame({'a':1, 'b':2})
```

#Porqué crees que esto es importante o útil



	a	b
0	1	2

```
[13] # Cargando datos desde un csv de una liga de internet
df = pd.read_csv('https://raw.githubusercontent.com/ricalanis/cerouno-final/master/data.csv')
```

/usr/local/lib/python3.7/dist-packages/IPython/core/interactiveshell.py:2718: DtypeWarning: Columns (9,12,13,16,21,28,32,33,34,35,36,37,40,41,42) have mixed types. Specify dtype option on import or setting with pd.read_csv(..., dtype={...})

```
[14] df.head(5)
```

	Column	Calle	Causa	Ciudad	Cruce	Fecha	Hora	ID	Latitud	Lesionados	Longitud	Mes	Muertos	No. De Infracción
0	0	CAMINO REAL	NO RESPETAR SEÑALAMIENTOS	Monterrey	SANTOS PALOMO	2016-05-01 00:00:00	1899-12-30 00:00:00	MTY1	25.77443	No	-100.371506	Mayo	No	NaN
1	1	BERNARDO REYES	Otros	Monterrey	FIDEL VELAZQUEZ	2016-05-01 00:00:00	1899-12-30 00:00:00	MTY10	25.724161	No	-100.327445	Mayo	No	NaN
2	2	BERNARDO REYES	Otros	Monterrey	REFORMA	2016-05-01 00:00:00	1899-12-30 00:00:00	MTY11	25.685874	No	-100.322931	Mayo	No	NaN
					CUMBRES	2016-	1899-							

Cargando distintas fuentes de datos

Podemos cargar datos de distintas fuentes:

- CSV
- JSON
- Pickle
- Desde una base de datos usando SQL
- Entre muchas otras fuentes de datos

Ejercicio 2.1 - Tipos de datos

- Elige una de las fuentes de datos y encuentra donde puedes encontrar on the wild información disponible para consultarlas
 - CSV - Open Data
 - JSON - API
 - Pickle - Crea el archivo tu y vuélvelo a cargar en Python
 - SQL - Un tip: portal.mammals
- Presentalas a tus compañeros
- Discutamos de los distintos tipos de datos

Investigar tipo de dato, cargarlo, ver usos del tipo de dato, y utilidad/retos al utilizar pandas para cargar ese tipo de dato.

Importancia de las fuentes

- CSV - datos planos sencillos
- JSON - Conexión con internet
- Pickle - datos guardados de proceso
- SQL - Datos desde una base de datos

Discutamos el valor de cada uno de ellos, limitaciones y ventajas

Índices, Accesos y columnas

Considera que puedes acceder a columnas de datos, pero también a filas.

`df['col']` se trae la columna del dataframe

`df.loc[1]` se trae la fila de índice 1

`df.iloc[1]` se trae la primera fila del DataFrame (comúnmente una operación lenta - relativamente hablando)

REGRESAMOS 1:15

Referencias

- Carmona, Patricia., “Data Analysis con Pandas. Primeros pasos.” Ironhack. 10 Feb. 2020.
medium.com/ironhack/data-analysis-con-pandas-primeros-pasos-6234b1234b9c.
- “pandas.Index.” pandas.
pandas.pydata.org/docs/reference/api/pandas.Index.html

Contacto

Mtro. Ricardo Daniel Alanis Tamez

ricardo@codeandomexico.org

LinkedIn: Ricardo Alanís