

Introducción a datos abiertos

Clase 03

Tabla de contenidos

¿Qué es un lenguaje de programación?

Lenguaje de programación Python

Entorno de desarrollo y Google Colaboratory

Descargar y abrir datos con Google Colaboratory



¿Qué es un lenguaje de programación?

Una herramienta que utilizamos para escribir tareas: secuencias y comandos que queremos que se ejecuten relacionadas a nuestros datos. Es un medio para escribir algo que queremos que suceda.



Lenguaje de programación Python

- Lenguaje de programación interactivo, interpretado y orientado a objetos
- Sintaxis amigable, fácil de interpretar
- Distintos tipos de datos (numeros, texto, booleanos, fecha y hora)



¿Por qué hablamos de sintaxis?

Al escribir código hay que respetar la estructura sintáctica en la cual nosotros escribimos las tareas, para que éstas se ejecuten. Por eso es importante aprender bien las reglas que tiene el lenguaje, para que a la hora de ejecutarlo lo hagamos exitosamente.



Estructura de los datos

Atributos 

FECHA	DESDE	HASTA	LINEA	MOLINETE	ESTACION	pax_pagos	pases_pa	pax_franq	pax_TOTAL
1/8/2021	08:00:00	08:15:00	LineaA	LineaA_Ca	Carabobo	0	0	1	1
1/8/2021	08:00:00	08:15:00	LineaA	LineaA_Flc	Flores	1	0	0	1
1/8/2021	08:00:00	08:15:00	LineaA	LineaA_Lo	Loria	1	0	0	1
1/8/2021	08:00:00	08:15:00	LineaA	LineaA_Pe	Peru	0	0	1	1
1/8/2021	08:00:00	08:15:00	LineaA	LineaA_PJ	Primera Ju	3	0	0	3
1/8/2021	08:00:00	08:15:00	LineaA	LineaA_PN	Plaza de M	1	0	0	1
1/8/2021	08:00:00	08:15:00	LineaA	LineaA_Sa	San Pedrit	1	0	1	2
1/8/2021	08:00:00	08:15:00	LineaB	LineaB_Ak	Leandro N	2	0	0	2
1/8/2021	08:00:00	08:15:00	LineaB	LineaB_Dc	Dorrego	1	0	0	1
1/8/2021	08:00:00	08:15:00	LineaB	LineaB_Dc	Dorrego	2	0	0	2
1/8/2021	08:00:00	08:15:00	LineaB	LineaB_Mi	Medrano	3	0	0	3
1/8/2021	08:00:00	08:15:00	LineaB	LineaB_Pe	Carlos Pell	5	0	0	5
1/8/2021	08:00:00	08:15:00	LineaB	LineaB_Pu	Pueyrredo	2	0	0	2
1/8/2021	08:00:00	08:15:00	LineaC	LineaC_Co	Constituci	24	0	0	24
1/8/2021	08:00:00	08:15:00	LineaC	LineaC_La	Lavalle	1	0	0	1
1/8/2021	08:00:00	08:15:00	LineaC	LineaC_Sa	General Sa	1	0	0	1
1/8/2021	08:00:00	08:15:00	LineaA	LineaA_Flc	Flores	0	0	1	1
1/8/2021	08:00:00	08:15:00	LineaA	LineaA_Mi	Plaza Mise	1	0	0	1
1/8/2021	08:00:00	08:15:00	LineaA	LineaA_Pe	Peru	0	0	1	1
1/8/2021	08:00:00	08:15:00	LineaA	LineaA_PN	Plaza de M	2	0	0	2
1/8/2021	08:00:00	08:15:00	LineaA	LineaA_PN	Plaza de M	0	0	1	1
1/8/2021	08:00:00	08:15:00	LineaB	LineaB_Ak	Leandro N	2	0	0	2
1/8/2021	08:00:00	08:15:00	LineaB	LineaB_Ga	Angel Gall	1	0	0	1

Registros



Vamos a usar Python para analizar los conjuntos de datos: las tablas. El programa nos va a permitir escribir las secuencias, los comandos, las tareas que queremos que se ejecuten en ese conjunto de datos.

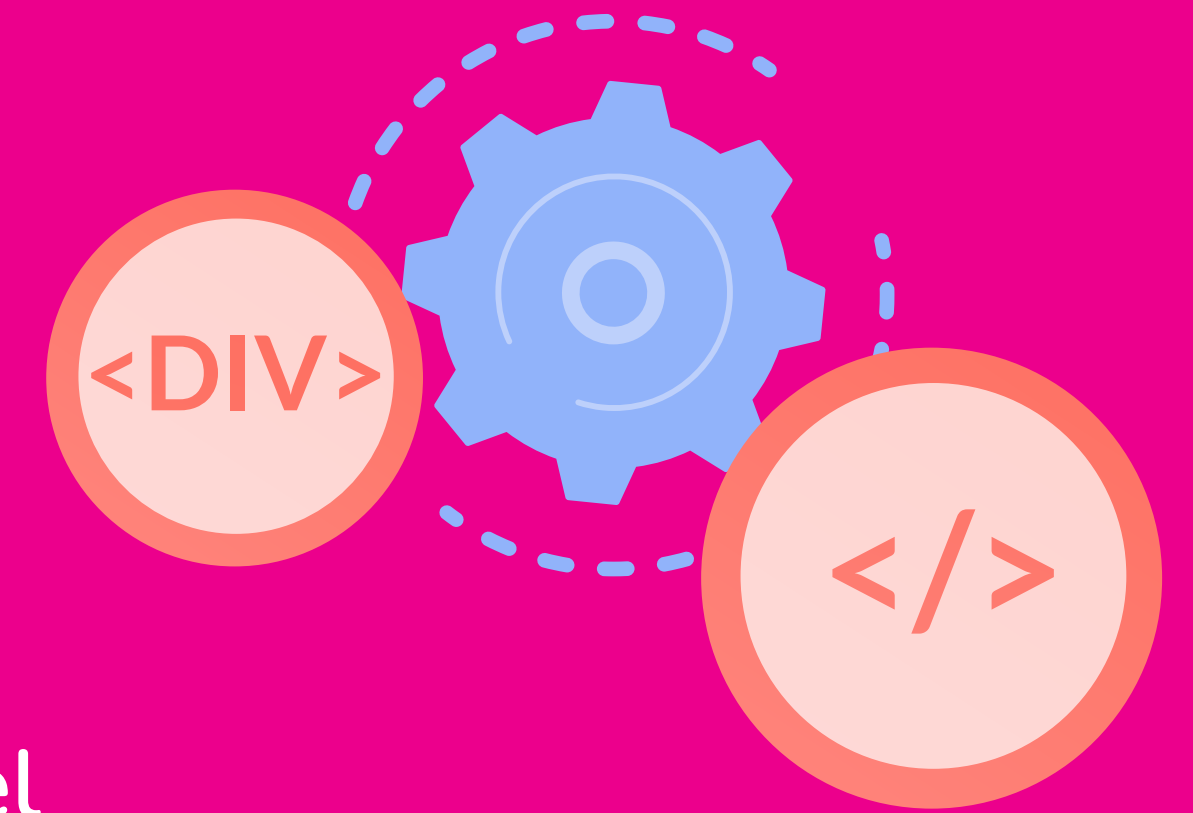
Por ejemplo: sacar un promedio de una columna.

Metadatos

Nos dan información sobre las tablas y los conjuntos de datos sobre los que estamos trabajando.

Información adicional	
Temas	
Etiquetas	cambio climático subte transito transporte urbanismo
Licencia	CC-BY-2.5-AR
Frecuencia de actualización	Mensualmente
Responsable	Jefatura de Gabinete de Ministros. Secretaria de Transporte y Obras Públicas. Subterráneos Buenos Aires (SBASE) 
Mantenedor	Secretaría General y Relaciones Internacionales. Subsecretaría Gestión Estratégica y Calidad Institucional. Dirección General Calidad Institucional y Gobierno Abierto 
Fecha de publicación	10 de Mayo de 2021
Fecha de actualización	5 de Octubre de 2021

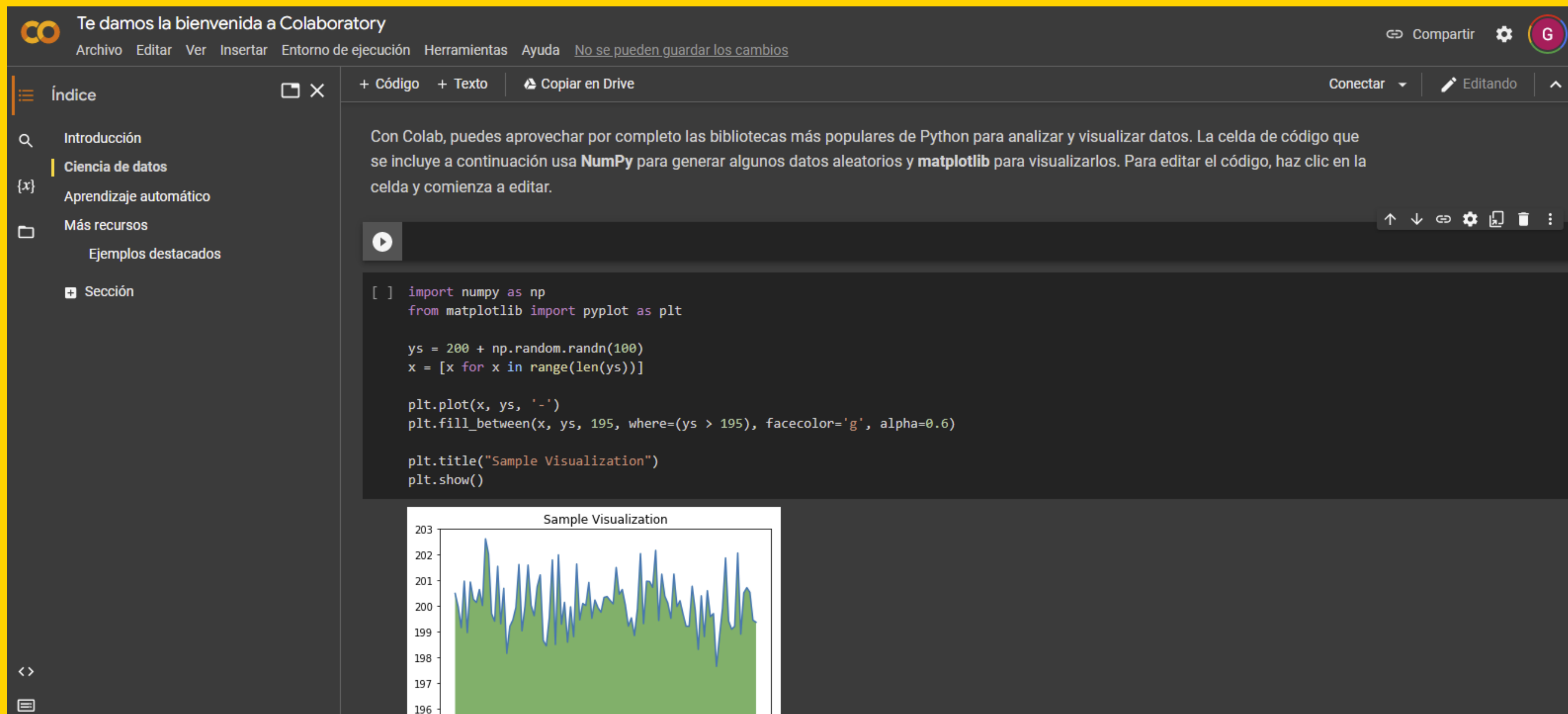
Introducción a Python



- Librerías (library): caja de herramientas que tenemos por fuera para el problema particular que tenemos.
- Tipos de datos
- Objetos de Python
- Librería Pandas
- Manipulación de un Data Frame: conjunto de datos tabulado. Tiene un formato particular de almacenado cuando nosotros lo queremos exportar (csv, excel)
- Consultas al Data Frame
- Librería Numpy
- Comando Groupby

Google Colaboratory

Es un entorno de desarrollo dentro del Google Workplace que permite escribir código en Python en una notebook: un documento en el cual vamos a poder escribir objetos tipo código y objetos tipo texto.



The screenshot displays the Google Colaboratory web interface. At the top, there is a navigation bar with the Google Colaboratory logo, a welcome message "Te damos la bienvenida a Colaboratory", and a menu with options like "Archivo", "Editar", "Ver", "Insertar", "Entorno de ejecución", "Herramientas", and "Ayuda". A warning message "No se pueden guardar los cambios" is visible. On the right side of the top bar, there are icons for "Compartir", "Configuración", and a user profile icon.

Below the navigation bar, there is a sidebar on the left with a search icon and a list of sections: "Índice", "Introducción", "Ciencia de datos", "Aprendizaje automático", "Más recursos", "Ejemplos destacados", and "Sección".

The main workspace contains a code cell with the following Python code:

```
[ ] import numpy as np
from matplotlib import pyplot as plt

ys = 200 + np.random.randn(100)
x = [x for x in range(len(ys))]

plt.plot(x, ys, '-')
plt.fill_between(x, ys, 195, where=(ys > 195), facecolor='g', alpha=0.6)

plt.title("Sample Visualization")
plt.show()
```

Below the code cell, there is a plot titled "Sample Visualization". The plot shows a line graph with a blue line representing the data points and a green shaded area representing the region where the data points are greater than 195. The x-axis ranges from 0 to 100, and the y-axis ranges from 196 to 203.

Ejercicio de clase:

- Cargar datos a colab
- Uso de pandas en colab



Ejercicio práctico:

1- Descargar el recurso Molinetes - Agosto 2021 (XLSX) del dataset Subte: Viajes Molinetes y abrir en colab. (Cuidado con el encoding del archivo y el separador)

2- Identificar e interpretar metadatos y responder: ¿qué preguntas o problemas podrían ser resueltas con este recurso?

