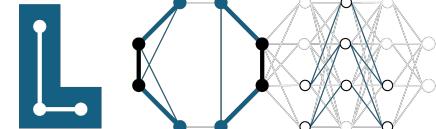


Ciencia y Tecnología

Secretaría de Ciencia, Humanidades,
Tecnología e Innovación



Laboratorio de Ciencia de Datos y Aprendizaje Automático

Curso IA desde Cero

Dr. Irvin Hussein López Nava
M.C. Joan M. Raygoza Romero

Departamento de Ciencias de la Computación
Centro de Investigación Científica y de Educación Superior de Ensenada
Edición 2025

CURSO VIRTUAL

IA desde Cero:
Un Curso Práctico

EDUCACIÓN CONTINUA | FÍSICA APLICADA | CICESE



Del 3 de Nov
al 1 de Dic

Dirigido al Público en general,
que desee aprender técnicas de IA.



Duración: 18 horas

Horario: Lunes y miércoles,
5:00 p.m. - 7:00 p.m.



Recursos

SCAN ME



Datos



DATA, DATA EVERY WHERE





Word cloud

<https://www.freewordcloudgenerator.com/generatewordcloud>

<https://www.wordclouds.com/>

<https://classic.wordclouds.com/>

¿Qué son los
datos?

Definiciones

ChatGPT

Son hechos, cifras u observaciones sin procesar que pueden ser recolectados, almacenados y analizados. Constituyen la base de la información y el conocimiento, a menudo en forma de números, texto, imágenes u otros formatos.

<https://chat.openai.com/>

Wikipedia

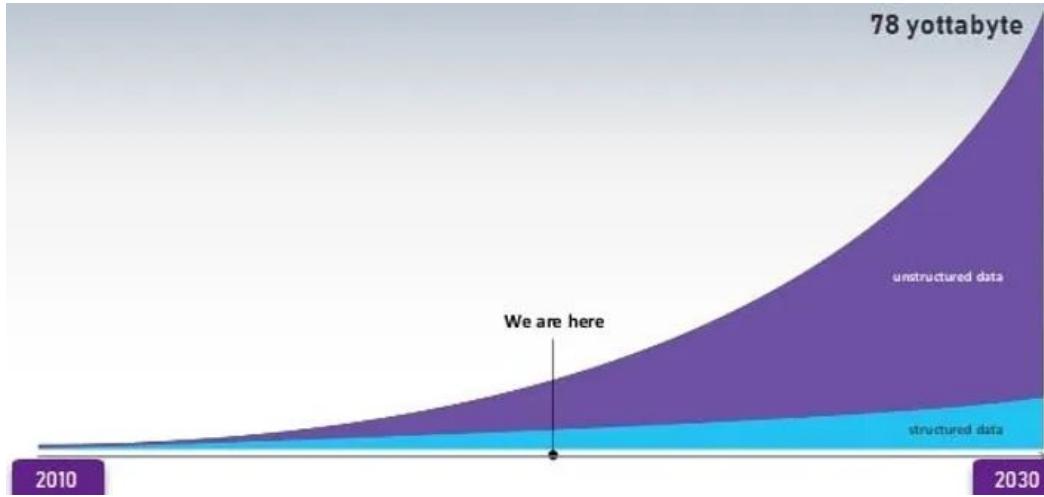
Son una representación simbólica (numérica, alfabética, algorítmica, espacial, etc.) de un atributo o variable cuantitativa o cualitativa. Los datos describen hechos empíricos, sucesos y entidades.

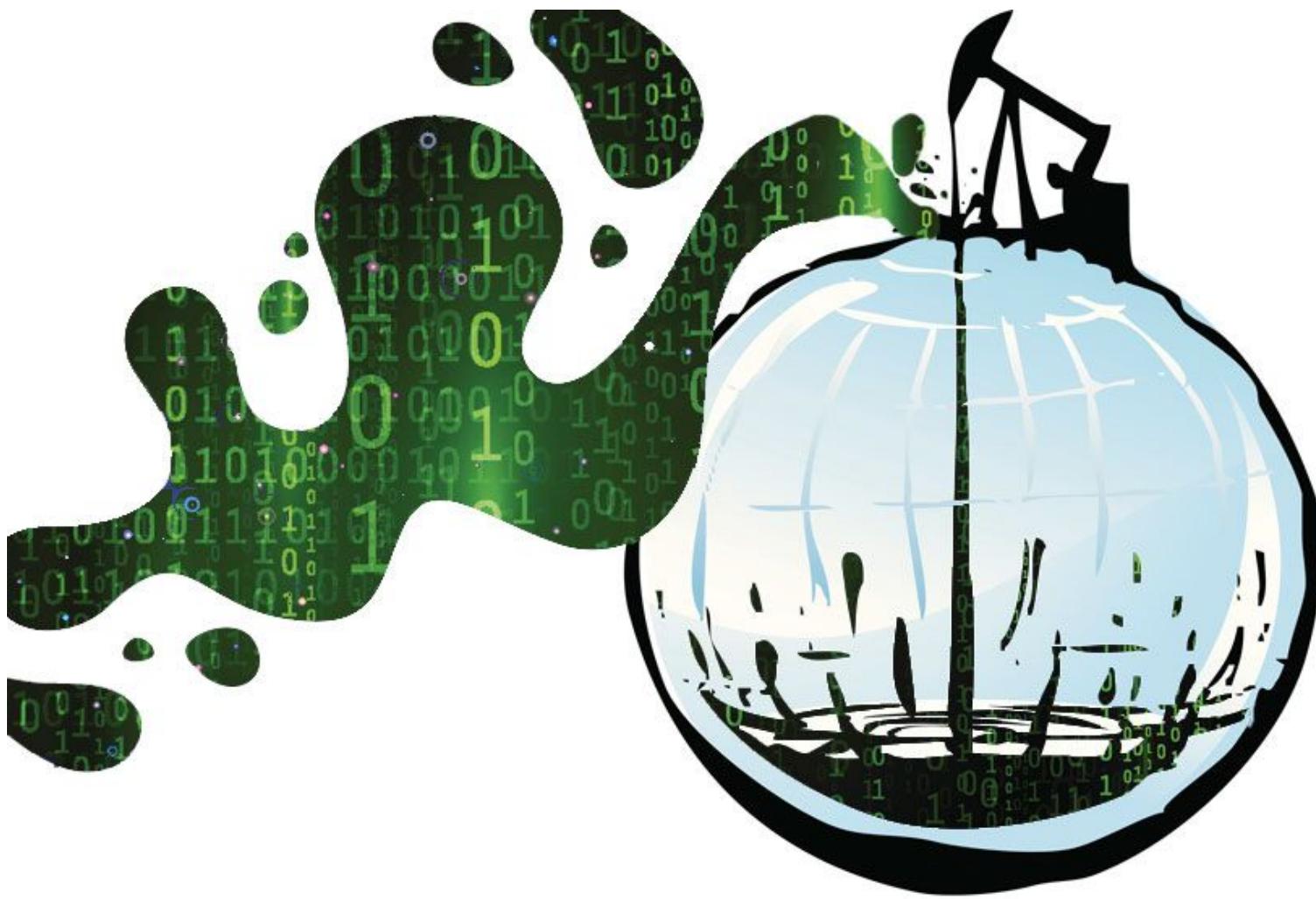
[https://es.wikipedia.org/wiki/
Dato](https://es.wikipedia.org/wiki/Dato)

Cambridge dic.

Información, especialmente hechos o cifras, recopilada para ser examinada, considerada y utilizada como ayuda para la toma de decisiones, o información en formato electrónico que puede ser almacenada y utilizada por un ordenador.

[https://dictionary.cambridge.org/
dictionary/english/data](https://dictionary.cambridge.org/dictionary/english/data)





Datos

VS

Petróleo

Internet de las cosas (IoT)

Red

Almacenamiento en la nube

Inteligencia Artificial

Teléfono Inteligente

Bomba de extracción

Tubería

Tanque

Refinería

Surtidor de Gasolina



MÉTODO CUANTITATIVO



MÉTODO CUALITATIVO



Tipos de datos

Cuantitativos

Se trata de números, información referente de una determinada “**cantidad**”.

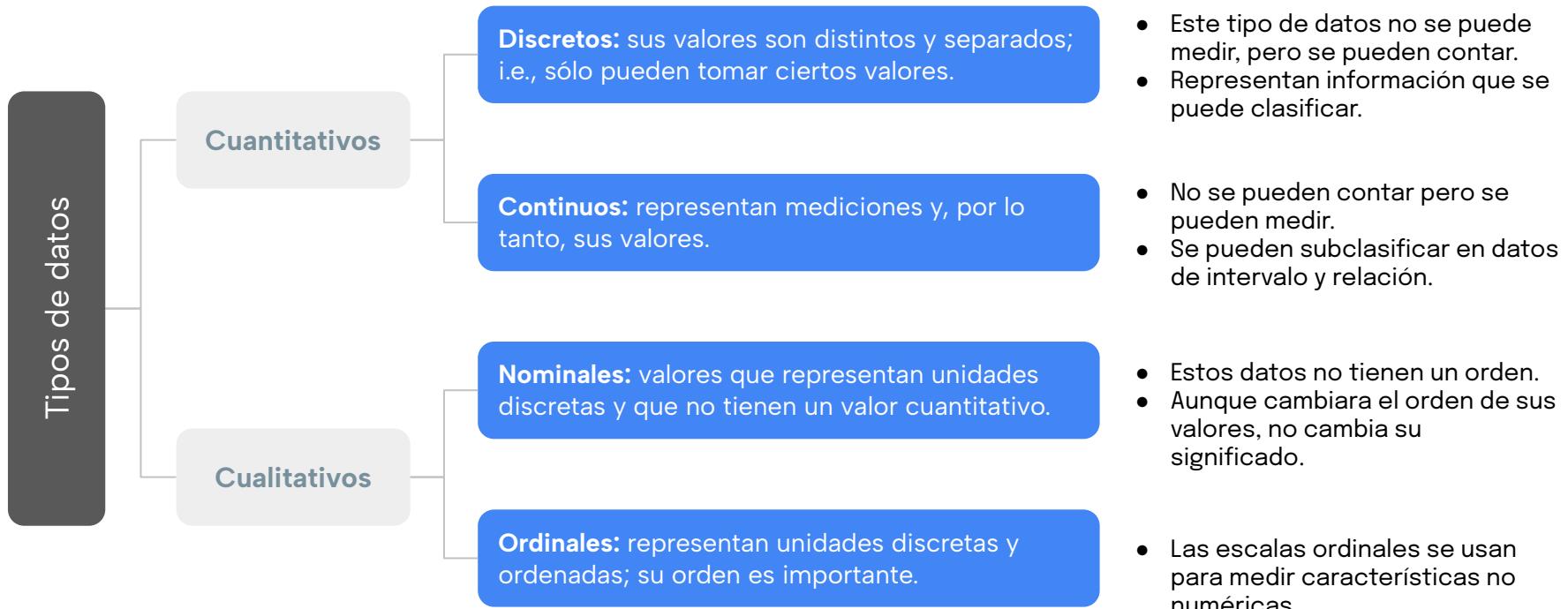
Por ejemplo, al medir la altura o el peso de una persona, o el tiempo que lleva ir de un lugar a otro.

Cualitativos

Relativos a las “**cualidades**”, este tipo de información está relacionada con los adjetivos.

Por ejemplo, al describir experiencias de compra es posible utilizar los términos: rápido, práctico, agradable, caro o inútil.

Escalas de medición

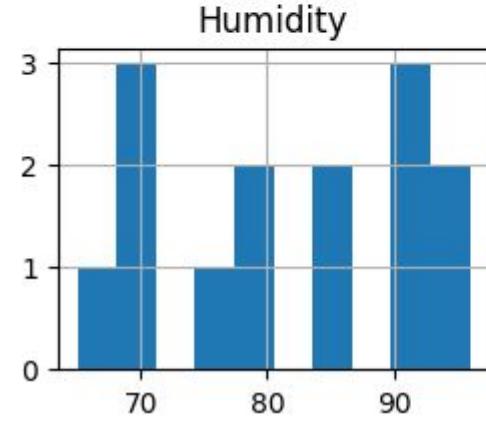
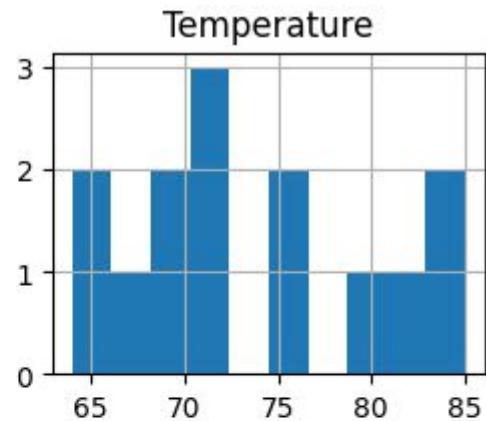
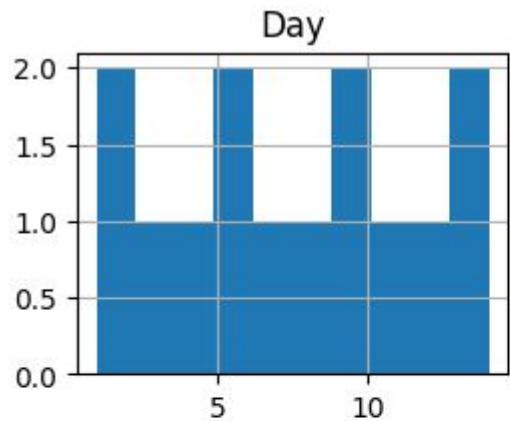


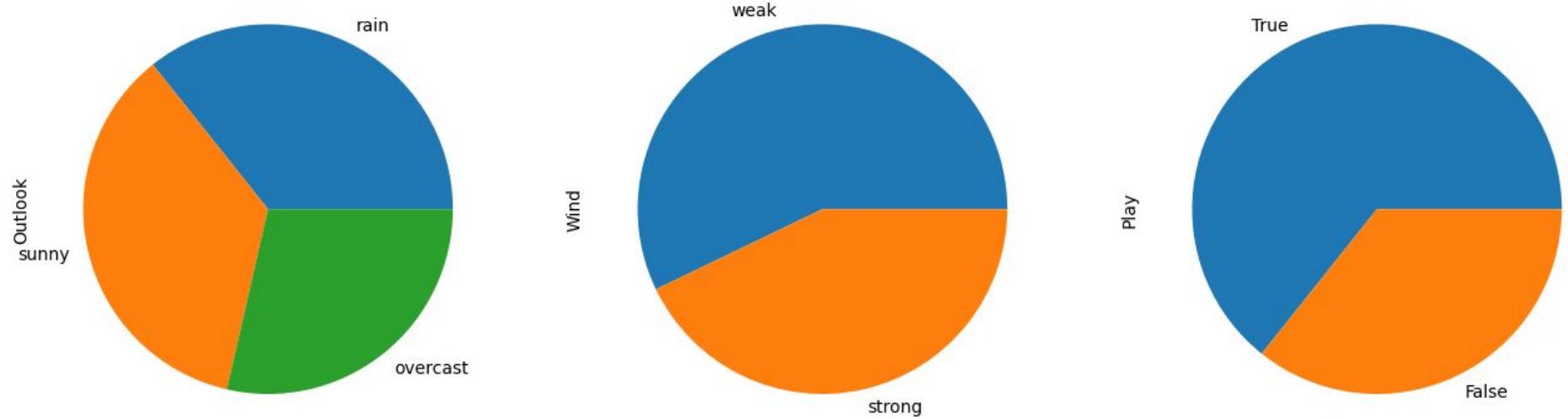
Dataset: weather.numeric

Day	Outlook	Temperature	Humidity	Wind	Play
1	sunny	85	85	weak	FALSE
2	sunny	80	90	strong	FALSE
3	overcast	83	86	weak	TRUE
4	rain	70	96	weak	TRUE
5	rain	68	80	weak	TRUE
6	rain	65	70	strong	FALSE
7	overcast	64	65	strong	TRUE
8	sunny	72	95	weak	FALSE
9	sunny	69	70	weak	TRUE
10	rain	75	80	weak	TRUE
11	sunny	75	70	strong	TRUE
12	overcast	72	90	strong	TRUE
13	overcast	81	75	weak	TRUE
14	rain	71	91	strong	FALSE

Relaciones

Day	Outlook	Temperature	Humidity	Wind	Play
1	sunny	85	85	weak	FALSE
2	sunny	80	90	strong	FALSE
3	overcast	83	86	weak	TRUE
4	rain	70	96	weak	TRUE
5	rain	68	80	weak	TRUE
6	rain	65	70	strong	FALSE
7	overcast	64	65	strong	TRUE
8	sunny	72	95	weak	FALSE
9	sunny	69	70	weak	TRUE
10	rain	75	80	weak	TRUE
11	sunny	75	70	strong	TRUE
12	overcast	72	90	strong	TRUE
13	overcast	81	75	weak	TRUE
14	rain	71	91	strong	FALSE





Tipos de datos (CS)

- En computación, un **tipo de datos** es una colección o agrupación de **valores de datos**, normalmente especificados por un conjunto de valores posibles, un conjunto de **operaciones** permitidas sobre estos valores, y/o una representación de estos valores como tipos de máquina.
- La mayoría de los lenguajes de programación admiten tipos de datos básicos de números enteros (de distintos tamaños), números de coma flotante (que se aproximan a los números reales), caracteres y booleanos.

Tipos de datos (CS)

- Diferentes lenguajes pueden utilizar diferentes tipos de datos o tipos similares con diferente semántica.
 - En Python, *int* representa un entero de precisión arbitraria que tiene las operaciones numéricas tradicionales como suma, resta y multiplicación.
 - En Java, el tipo *int* representa el conjunto de enteros de 32 bits cuyo valor oscila entre -2.147.483.648 y 2.147.483.647, con operaciones aritméticas que se desbordan.

Tipos de datos de máquina

- Todos los datos de las computadoras basados en la electrónica digital se representan como **bits** (0 y 1) en el nivel más bajo. La unidad de datos direccionable más pequeña suele ser un grupo de bits llamado **byte** (8 bits). La unidad procesada por instrucciones de código máquina se denomina **palabra** (a partir de 2011: 32 o 64 bits).
- Se utilizan principalmente en programación de sistemas o en lenguajes de programación de bajo nivel (ensamblador).

```
10100001 10111100 10010011 00000100  
00001000 00000011 00000101 11000000  
10010011 00000100 00001000 10100011  
11000000 10010100 00000100 00001000
```

It means: $z = x + y;$

Tipo booleano

- El tipo booleano representa los valores **verdadero** y **falso**. Aunque sólo son posibles dos valores, a menudo se representan como una palabra y no como un solo bit.
- Muchos lenguajes de programación no tienen un tipo booleano explícito, sino que utilizan un tipo entero e interpretan 0 como **FALSE** y otros valores como **TRUE**.

True	False
<input checked="" type="checkbox"/>	<input type="checkbox"/>

Tipos numéricos

- Todos los lenguajes de programación proporcionan **uno o más tipos de datos enteros**.
- Pueden proporcionar una pequeña cantidad de subtipos predefinidos restringidos a ciertos rangos; o permitir a los usuarios definir libremente subrangos.
- Los tipos de datos de **punto flotante** representan ciertos valores fraccionarios (números racionales, matemáticamente).

```
byte b = 1;
short s = 200;
int i = 3000;
long l = 10000;
float f = 10.0f;
double d = 10.1d;
```

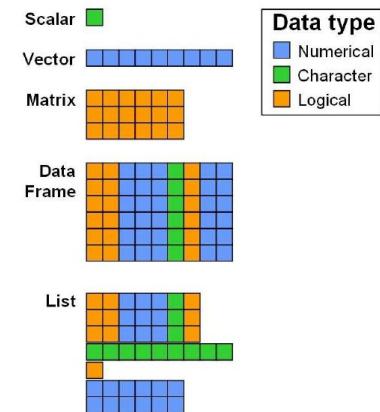
Tipos de texto y cadenas

- Los **caracteres** pueden ser una letra de algún alfabeto, un dígito, un espacio en blanco, un signo de puntuación, etc. Se extraen de un conjunto predefinido como ASCII.
- Las **cadenas** son una secuencia de caracteres que se utilizan para almacenar palabras o texto sin formato.
- Los tipos de caracteres y cadenas pueden tener diferentes subtipos según su codificación.
 - El ASCII original de 7 bits era limitado y fue reemplazado por conjuntos de 8, 16 y 32 bits, para codificar alfabetos no latinos (como el hebreo y el chino) y otros símbolos.

b ₇	b ₆	b ₅	b ₄	b ₃	b ₂	b ₁	Columna	0_0	0_0_1	0_1_0	0_1_1	1_0_0	1_0_1	1_1_0	1_1_1
0	0	0	0	0	0	0	NUL	DLE	SP	0	@	P	`	p	7
0	0	0	1	1	1	1	SOH	DC1	!	!	A	Q	a	q	
0	0	1	0	2	2	2	STX	DC2	"	2	B	R	b	r	
0	0	1	1	3	3	3	ETX	DC3	#	3	C	S	c	s	
0	1	0	0	4	4	4	EOT	DC4	\$	4	D	T	d	t	
0	1	0	1	5	5	5	ENQ	NAK	%	5	E	U	e	u	
0	1	1	0	6	6	6	ACK	SYN	8	6	F	V	f	v	
0	1	1	1	7	7	7	BEL	ETB	'	7	G	W	g	w	
1	0	0	0	8	BS	CAN	(8	H	X	h	x			
1	0	0	1	9	HT	EM)	9	I	Y	i	y			
1	0	1	0	10	LF	SUB	*	:	J	Z	j	z			
1	0	1	1	11	VT	ESC	+	:	K	[k	{			
1	1	0	0	12	FF	FS	,	<	L	\	l	-			
1	1	0	1	13	CR	GS	-	=	M]	m	}			
1	1	1	0	14	SO	RS	.	>	N	^	n	~			
1	1	1	1	15	SI	US	/	?	O	—	o	DEL			

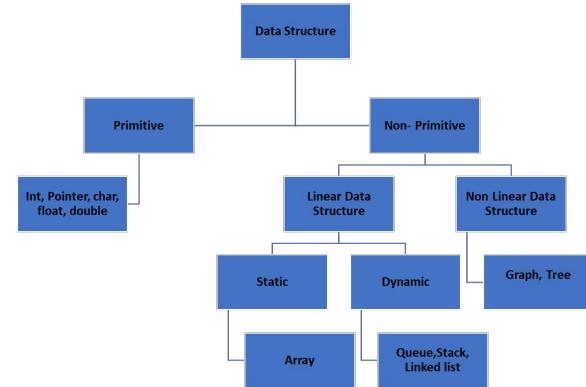
Estructuras de datos

- Son útiles para almacenar y recuperar conjuntos de datos.
- Una **matriz** (vector, lista, secuencia) almacena una cantidad de elementos y proporciona acceso a sus elementos individuales.
- Normalmente, se requiere que los elementos de una matriz sean del mismo tipo.
- Pueden ser de longitud fija o dinámica.
- Los índices de una matriz pueden ser números enteros.



Estructuras de datos

- Un **objeto** contiene campos de datos (variables), y también ofrece una cantidad de subrutinas para acceder a ellos o modificarlos, llamadas métodos.
- Las **listas enlazadas** (simples o dobles), se pueden utilizar para implementar estructuras especializadas, como: pilas, colas, árboles y grafos.



```

0 2 15 0 0 11 10 0 0 0 0 9 9 9 0 0 0
0 0 0 4 60 157 236 255 255 177 95 61 32 0 0 29
0 10 16 119 238 255 244 245 243 250 249 255 222 103 10 0
0 14 170 255 255 244 254 255 253 245 255 249 253 251 124 1
2 98 255 228 255 261 254 211 141 116 122 215 251 238 255 49
13 217 243 255 155 33 226 62 2 0 10 13 232 255 36
16 229 252 254 49 12 0 0 7 7 0 70 237 252 235 62
6 141 245 255 212 25 11 9 3 0 115 236 243 255 137 0
0 87 252 250 248 215 60 0 1 121 252 255 248 144 6 0
0 13 113 255 255 245 255 182 181 248 252 242 208 36 0 19
1 0 5 117 251 255 241 255 247 255 241 162 17 0 7 0
0 0 0 4 58 251 255 246 254 253 255 120 11 0 1 0
0 0 4 97 255 255 255 248 252 255 244 255 182 10 0 4
0 22 206 252 246 251 241 100 24 113 255 245 255 194 9 0
0 111 255 242 255 158 20 0 0 6 39 255 232 230 56 0
0 218 251 250 137 7 11 0 0 0 2 62 255 250 125 3
0 173 255 255 101 9 20 0 13 3 13 182 251 245 61 0
0 107 251 241 255 230 98 55 19 118 217 248 253 255 52 4
0 18 146 250 255 247 255 255 255 249 255 240 255 129 0 5
0 0 23 113 215 255 250 248 255 255 248 248 118 14 12 0
0 0 6 1 0 52 153 233 255 252 147 37 0 0 4 1
0 0 5 5 0 0 0 0 0 14 1 0 6 6 0 0 0

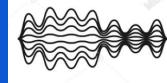
```



Imágenes



Audio



Video



Señales



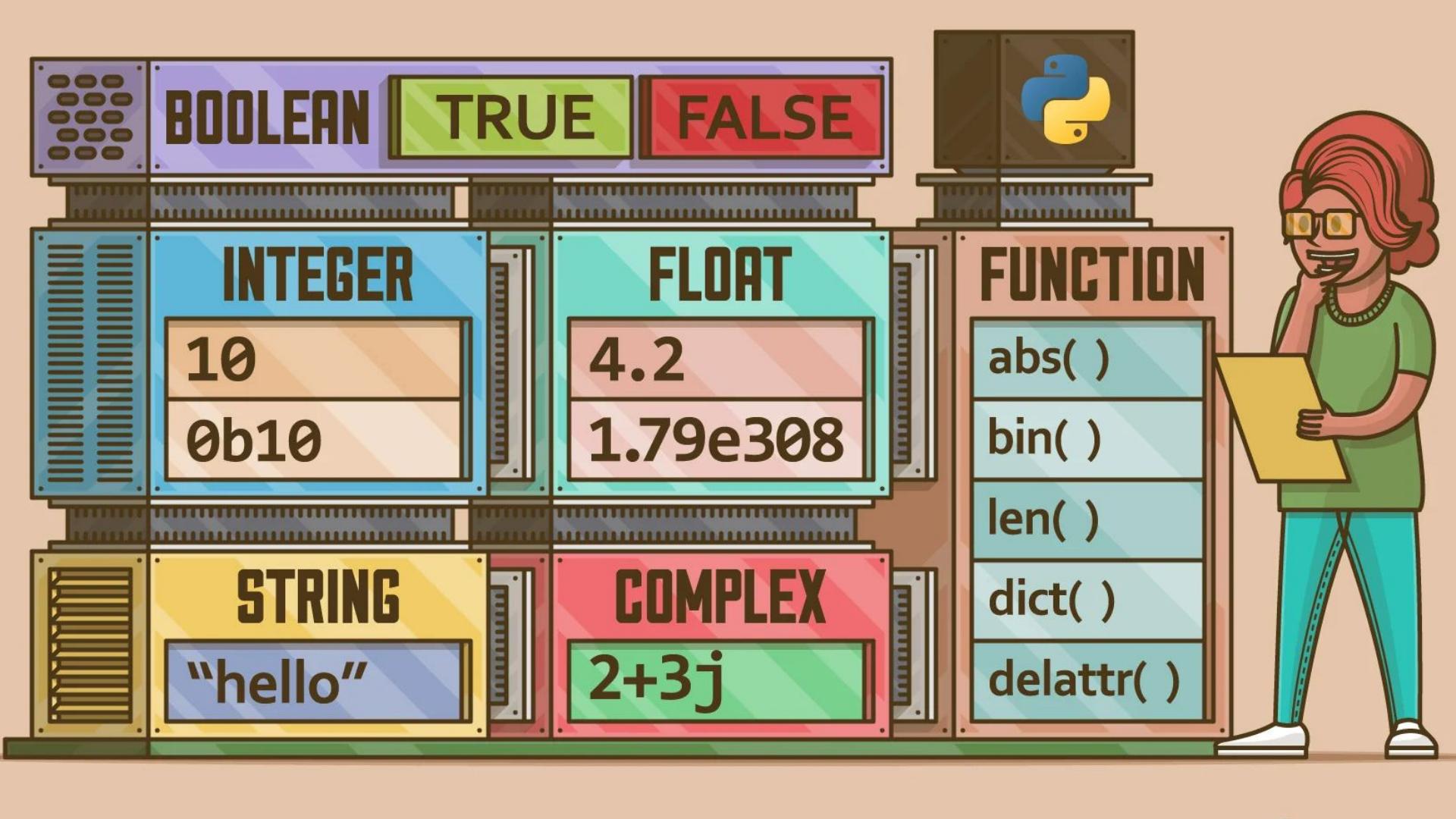
Texto

Lorem ipsum dolor sit
 dolor amet, consectetur
 nec adipiscing elit, sed
 do eiusmod tempor.

it	6
I	5
the	4
to	3
and	3
seen	2
yet	1
would	1
whimsical	1
times	1
sweet	1
satirical	1
adventure	1
genre	1
fairy	1
humor	1
have	1
great	1
...	...



Datos en Python



BOOLEAN

TRUE

FALSE

INTEGER

10

0b10

FLOAT

4.2

1.79e308

FUNCTION

abs()

bin()

len()

dict()

delattr()

STRING

"hello"

COMPLEX

2+3j

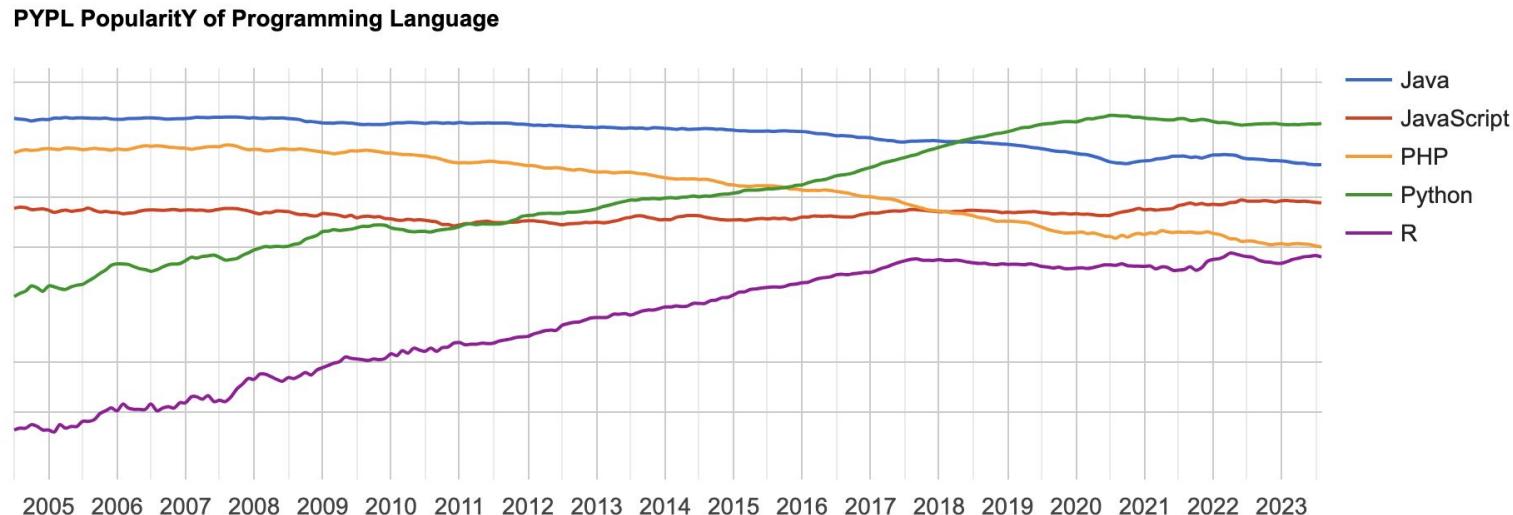
Python es el **lenguaje** de la Ciencia de Datos

- Especialmente cierto si las tareas de ciencia de datos implican mucho **procesamiento** de datos y/o aprendizaje automático.
- Menos cierto si las tareas son más "puramente **estadísticas**" (entonces R es más estándar)

El crecimiento y la popularidad de Python coinciden con el auge del aprendizaje automático.

Python 3 es uno de los lenguajes de programación de más rápido crecimiento. Es un lenguaje más fácil y eficiente que Python 2, o incluso que C#, R o Java. Python 3 es un lenguaje más limpio con una sintaxis más sencilla.

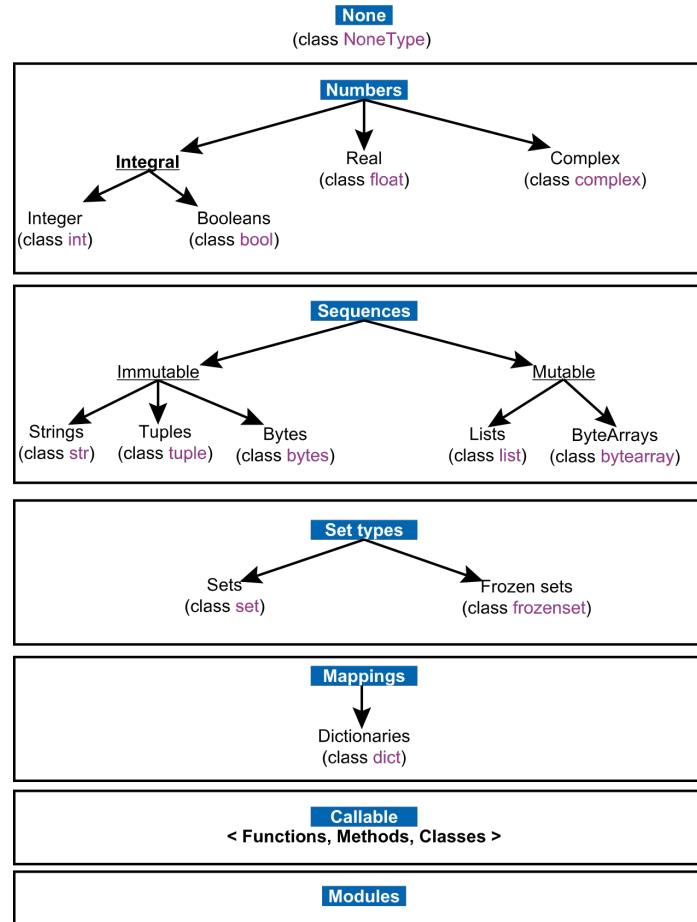
Python growth



Python growth

Rank	Change	Language	Share	1-year trend
1		Python	28.04 %	+0.3 %
2		Java	15.78 %	-1.3 %
3		JavaScript	9.27 %	-0.2 %
4		C#	6.77 %	-0.2 %
5		C/C++	6.59 %	+0.4 %
6		PHP	5.01 %	-0.4 %
7		R	4.35 %	+0.0 %
8		TypeScript	3.09 %	+0.3 %
9	↑↑	Swift	2.54 %	+0.5 %
10		Objective-C	2.15 %	+0.1 %

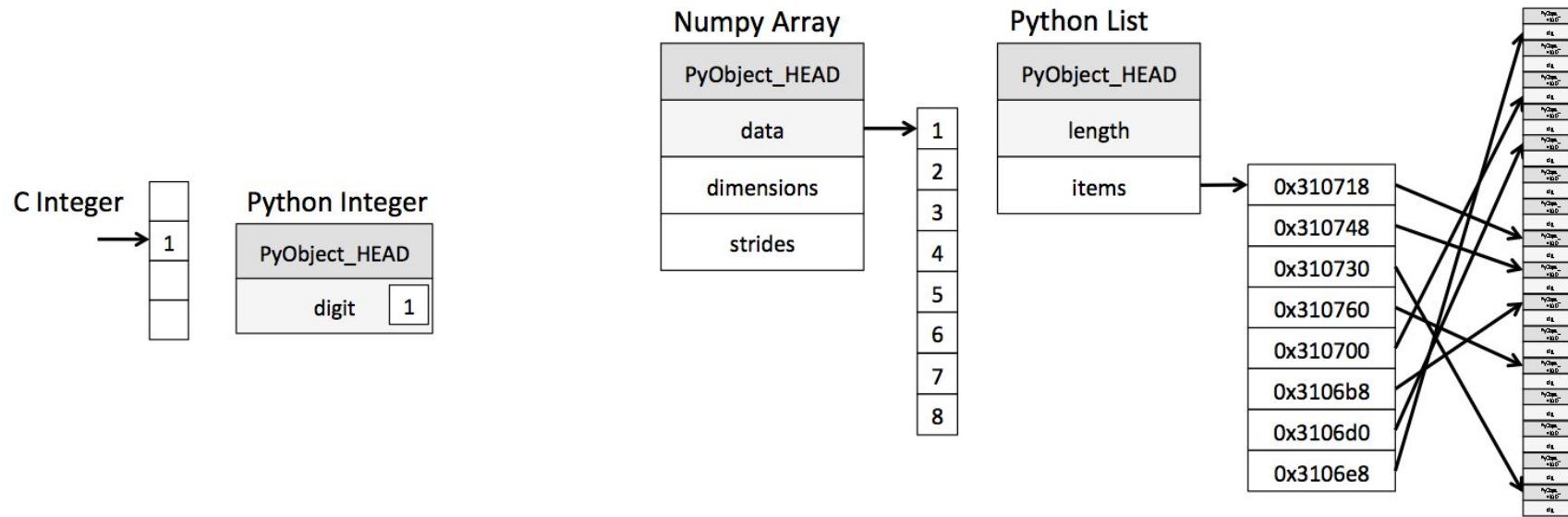
Python 3
The standard type hierarchy



Tipos de datos de Python

Name	Type	Description
Integers	int	Whole numbers, such as: 3 300 200
Floating point	float	Numbers with a decimal point: 2.3 4.6 100.0
Strings	str	Ordered sequence of characters: "hello" 'Sammy' "2000" "楽しい"
Lists	list	Ordered sequence of objects: [10,"hello",200.3]
Dictionaries	dict	Unordered Key:Value pairs: {"mykey": "value", "name": "Frankie"}
Tuples	tup	Ordered immutable sequence of objects: (10,"hello",200.3)
Sets	set	Unordered collection of unique objects: {"a","b"}
Booleans	bool	Logical value indicating True or False

Tipos de datos de Python



¿Preguntas?

hussein@cicese.mx

