

INTRO AL ANÁLISIS DE DATOS CON PANDAS





¡HOLA! YO SOY...

MARÍA CAMILA GUERRERO GIRALDO O
@MCGUEGI

ESTUDIO INGENIERÍA DE SISTEMAS EN
LA U.DISTRITAL .

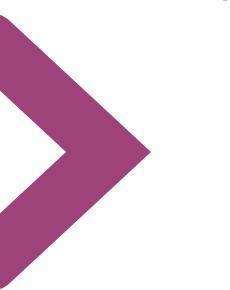
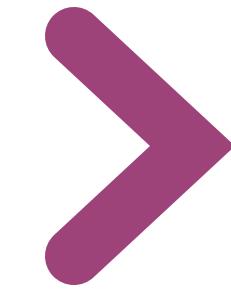
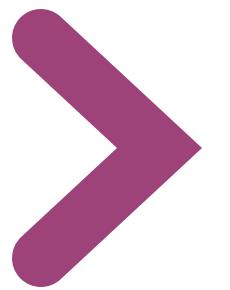
APRENDÍ A PROGRAMAR EN PYTHON
HACE 3 AÑOS Y UNO DE MIS HOBBIES
ES PRACTICAR LETTERING .





¿QUÉ ES EL ANÁLISIS
DE DATOS?

" If you torture the data long enough, it will confess."
- Ronald Coase



FASES

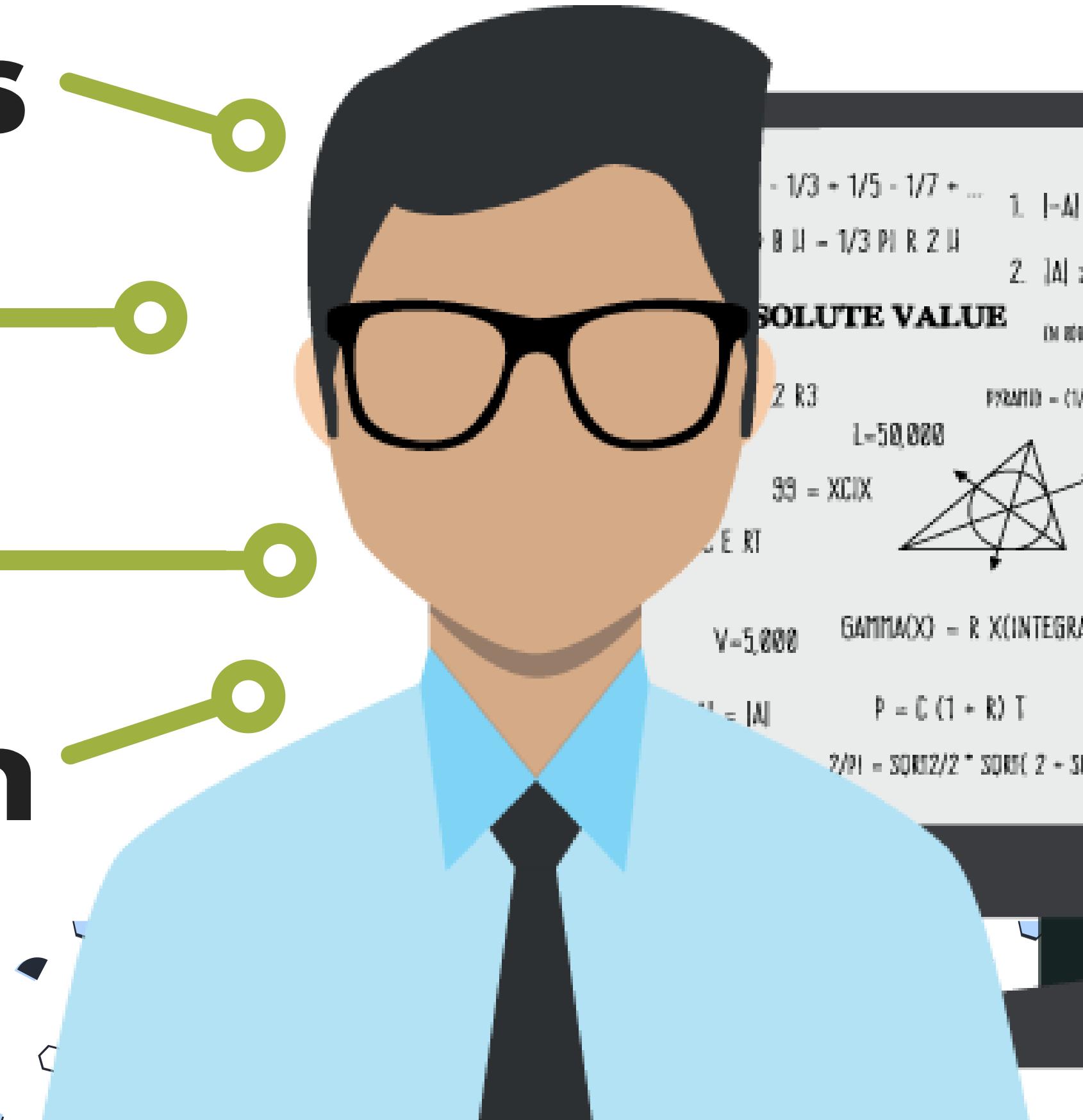


Matemáticas

Estadística

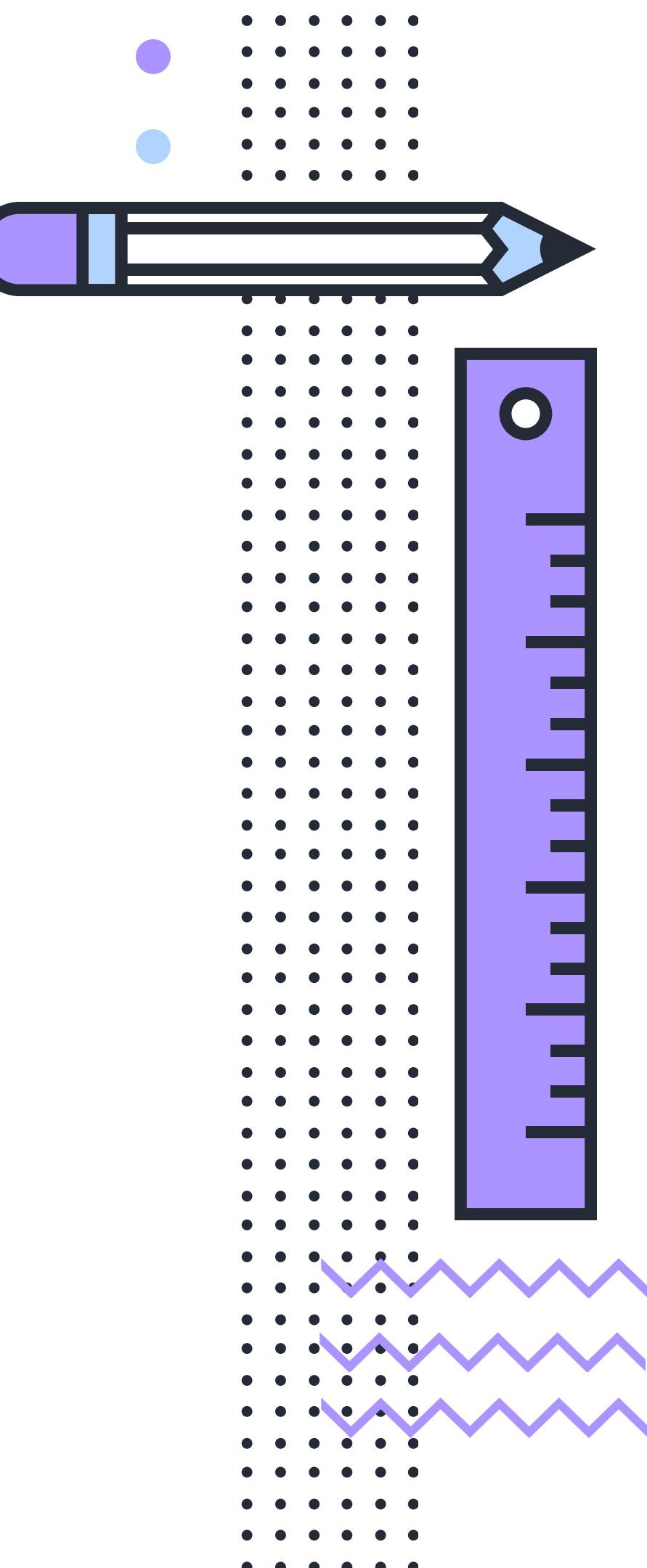
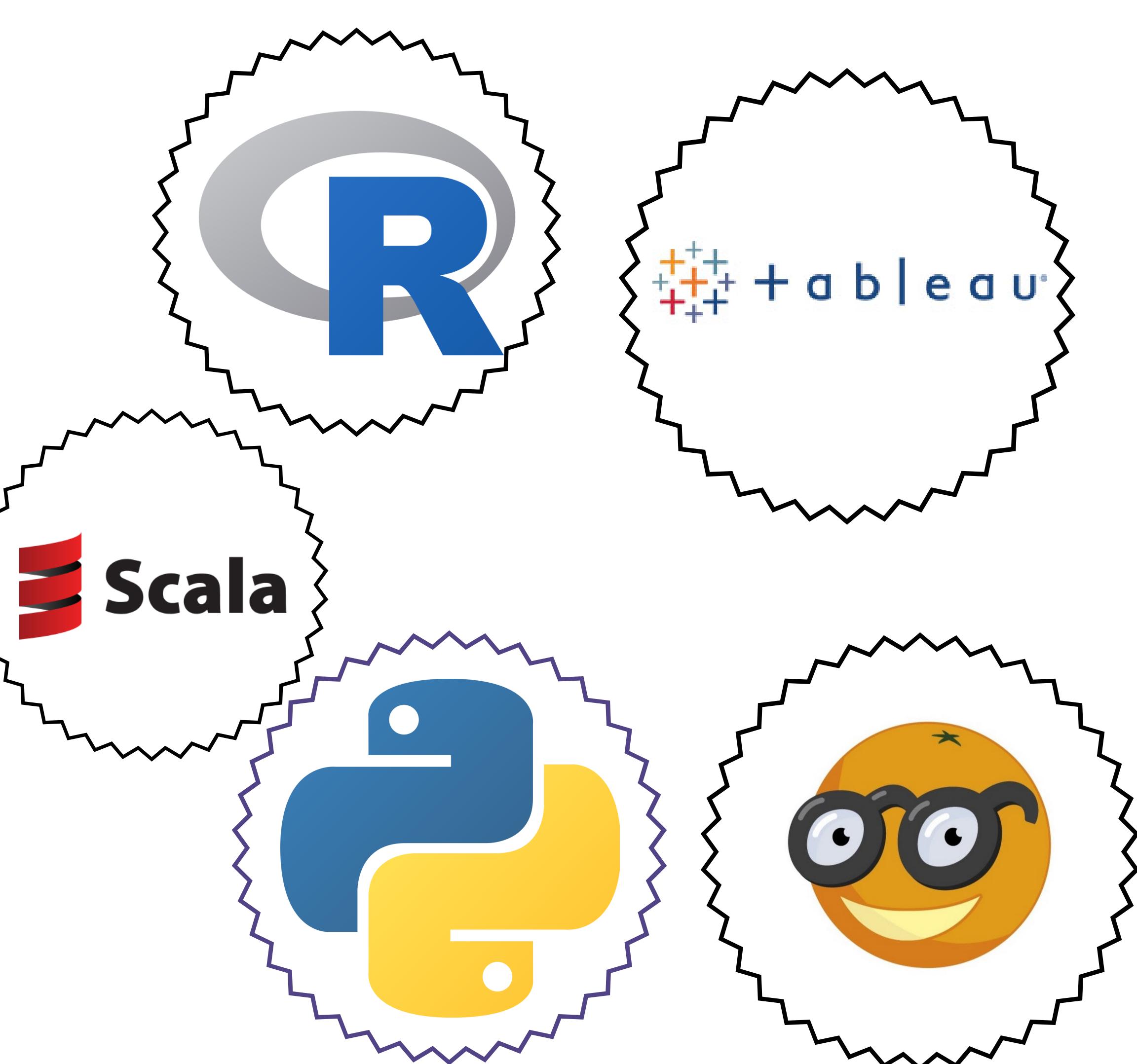
Análisis

Programación





ALGUNAS HERRAMIENTAS

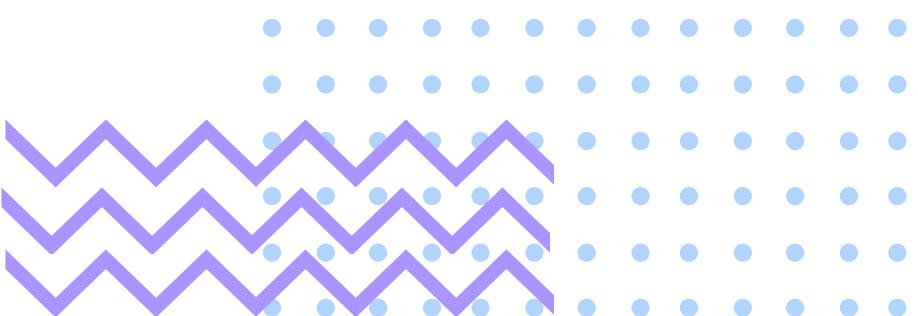




APRENDEMOS SOBRE
PANDAS

DATOS CURIOSOS LIBRERIA PANDAS

- 01 **PANEL DATA** : ES EL TÉRMINO QUE ORIGINA EL NOMBRE DE LA LIBRERIA
- 02 ES UNA BIBLIOTECA DE SOFTWARE DE PYTHON .
ES UNA EXTENSIÓN DE NUMPY PARA MANIPULACIÓN Y ANÁLISIS DE DATOS.
- 03 DESARROLLADA POR WES MCKINNEY
- 04 TRABAJA JUNTO A OTRAS LIBRERIAS . EJEMPLO :
MATPLOTLIB Y SEABORN

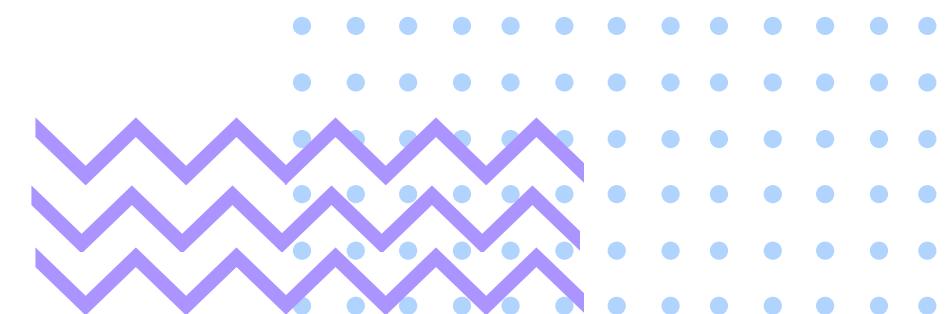


Pandas



CARACTERÍSTICAS PRINCIPALES

- SOPORTE PARA CSV, EXCEL, JSON, SQL, SAS, CLIPBOARD, HDF5,...
- LIMPIEZA DE DATOS
- JOINS, MERGE Y PIOTES
- VISUALIZACIÓN DE DATOS
- OPERACIONES SIMILARES A LAS DE BD





RECOLECCIÓN DE DATOS Y ESTRUCTURAS

Serie

index	values
A	5
B	6
C	12
D	-5
E	6.7

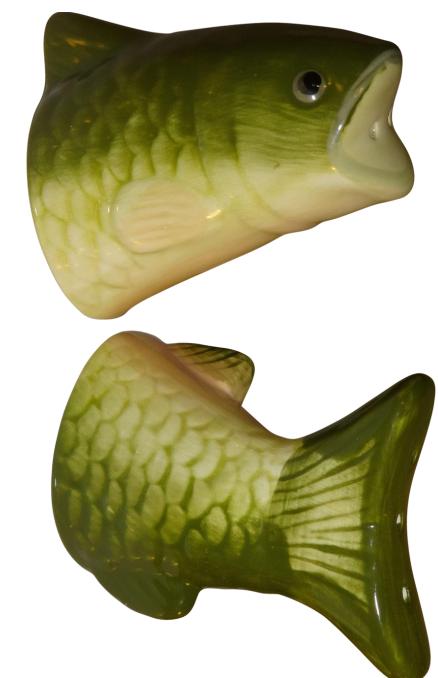
Dataframe

Índice	Columnas		
	A	B	C
0	'Hello'	'Column B'	NaN
1	'NO INFO'	'NO INFO'	'NO INFO'
2	'A'	'Column B'	NaN
3	'A'	'Column B'	NaN
4	'A'	'Column B'	NaN

Datos



PROCESAMIENTO Y LIMPIEZA



head()

tail()

**Operaciones
de emparejamiento**

add()

sub()

mul()

div()

+ -
x /

**Comparaciones
flexibles**

eq() ne() lt() gt() le()

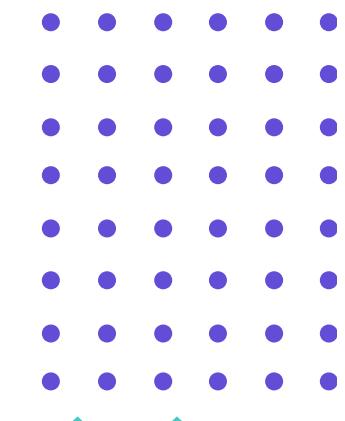
Reducciones booleanas

empty()

any()

all()

bool()



ANÁLISIS EXPLORATORIO



Function	Description
count	Number of non-NA observations
sum	Sum of values
mean	Mean of values
mad	Mean absolute deviation
median	Arithmetic median of values
min	Minimum
max	Maximum
mode	Mode
abs	Absolute Value
prod	Product of values
std	Bessel-corrected sample standard deviation
var	Unbiased variance
sem	Standard error of the mean
skew	Sample skewness (3rd moment)
kurt	Sample kurtosis (4th moment)
quantile	Sample quantile (value at %)
cumsum	Cumulative sum
cumprod	Cumulative product
cummax	Cumulative maximum
cummin	Cumulative minimum

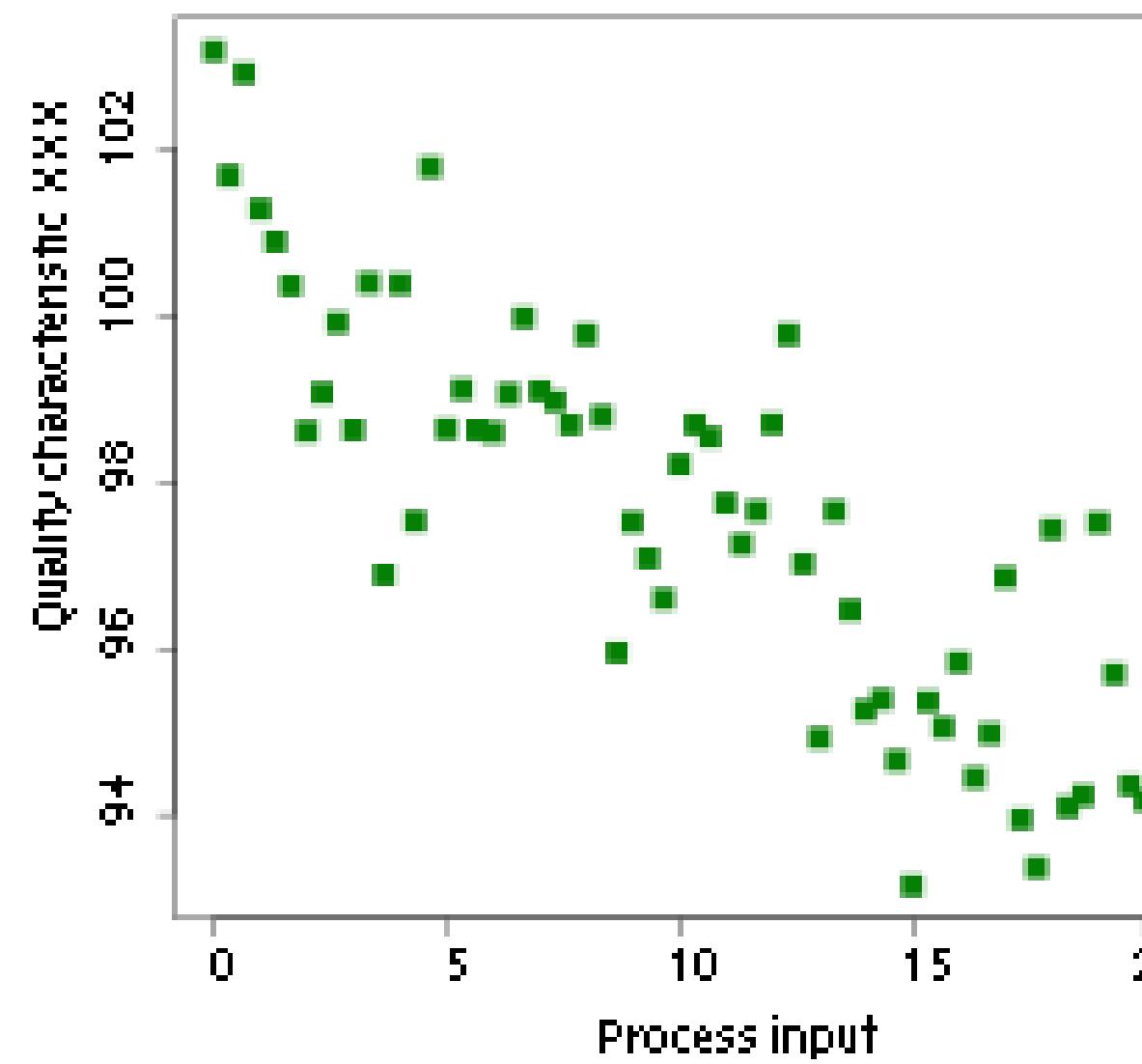
describe()
value_counts()





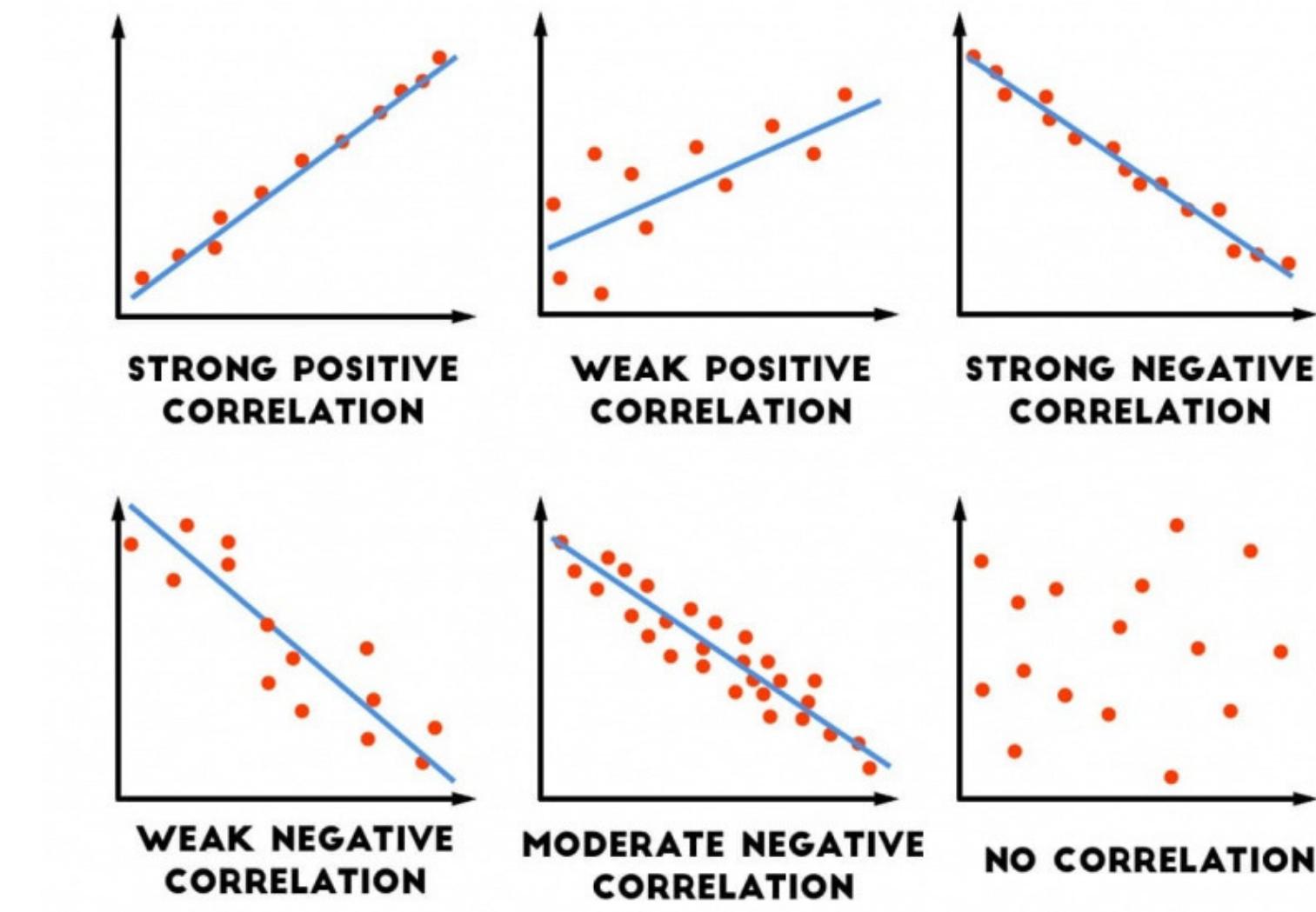
MODELADO

Scatterplot for quality characteristic XXX

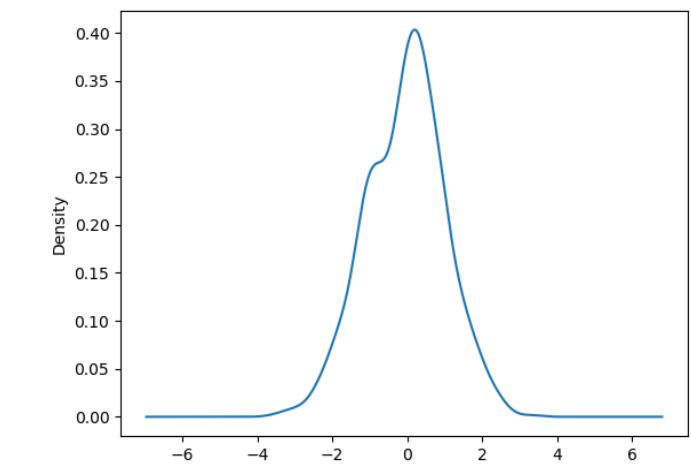
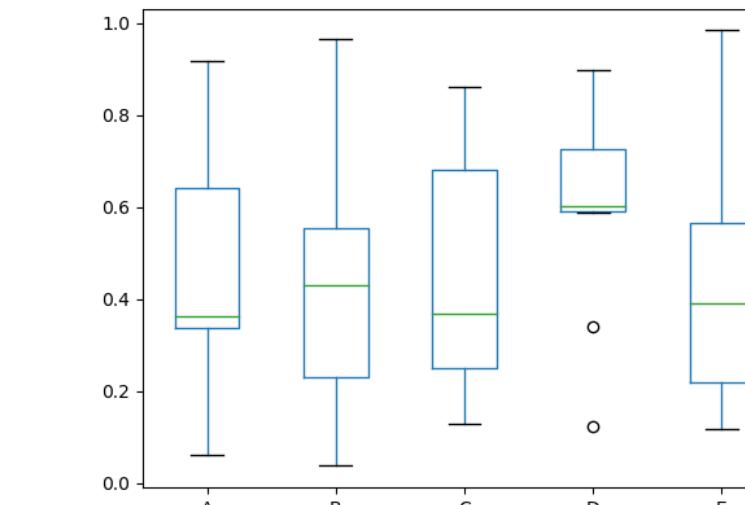
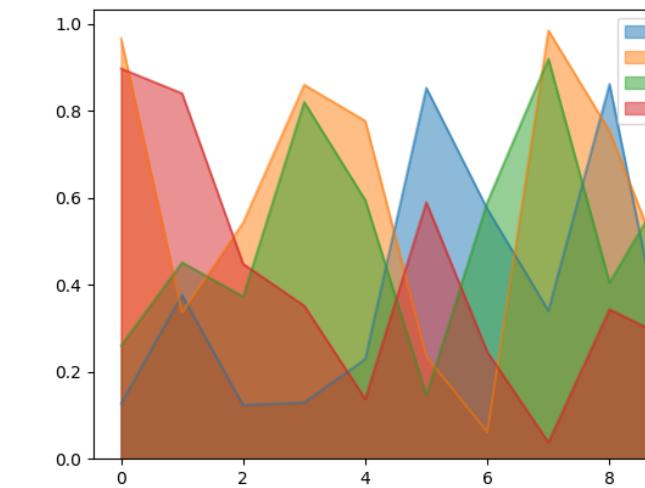
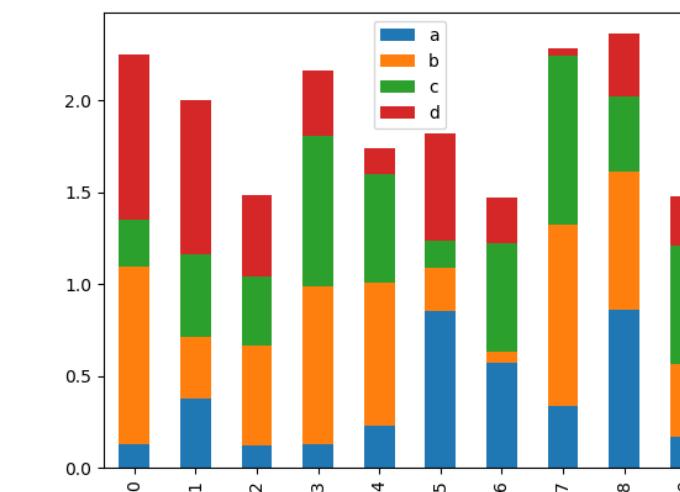
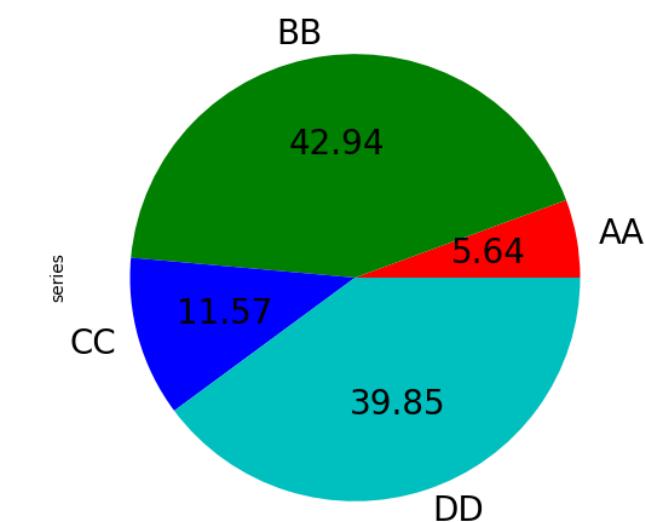
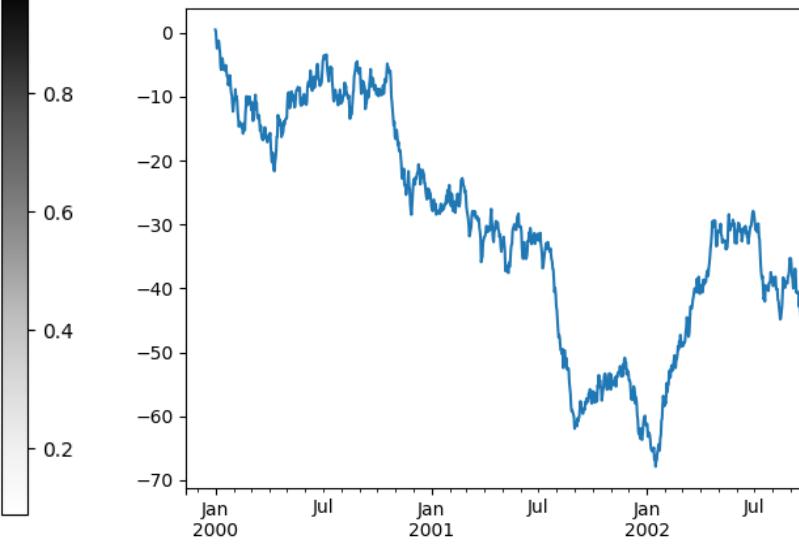
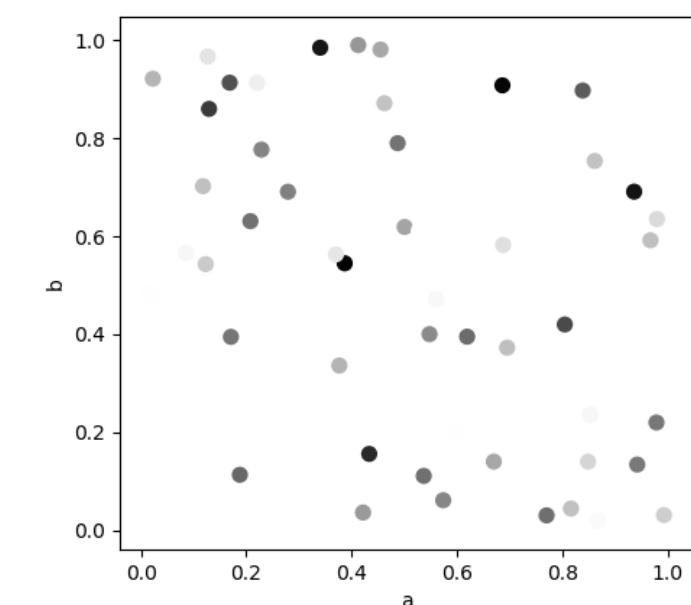
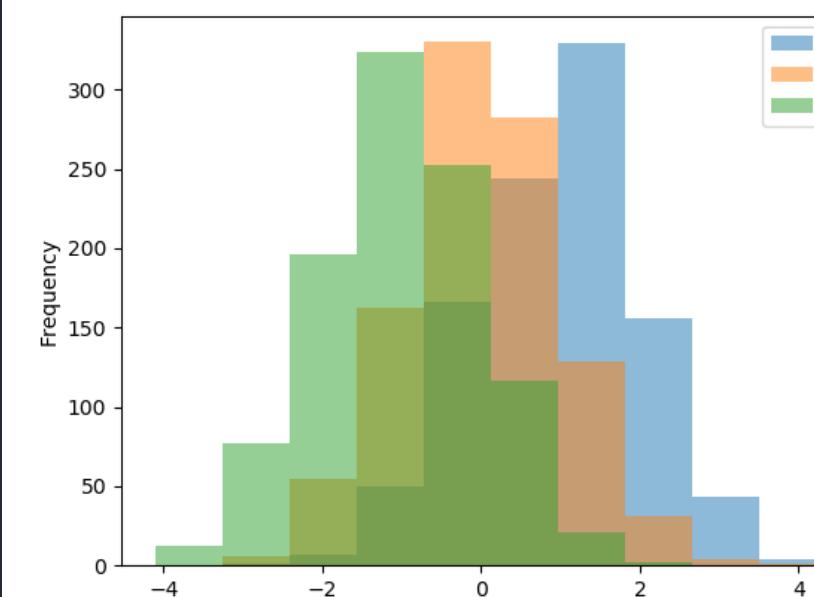


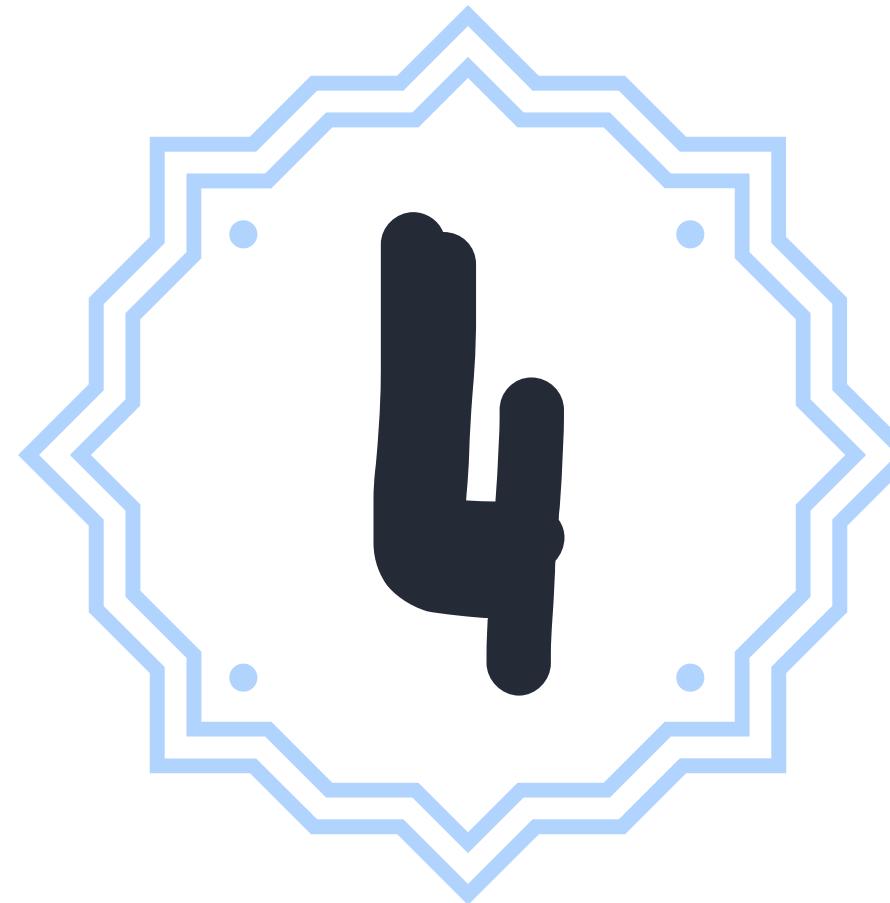
CORRELATION

(INDICATES THE RELATIONSHIP BETWEEN TWO SETS OF DATA)



PRESENTACIÓN Y VISUALIZACIÓN





iMANOS A LA
OBRA!



2 MANERAS DIFERENTES DE ACCEDER AL MATERIAL

1) Descargar los archivos desde Google Drive y correrlo en Jupyter Notebook o en Google Collab :

[https://drive.google.com/drive/folders/1dpae_kU08ITSBpBK9GYxaMqlWKryj63O?
usp=sharing](https://drive.google.com/drive/folders/1dpae_kU08ITSBpBK9GYxaMqlWKryj63O?usp=sharing)

2) Clonar el repositorio con el link : <https://github.com/macaguegi/pyladies-intro-Pandas.git>



FEEDBACK Y
CIERRE

¡Muchas Gracias!

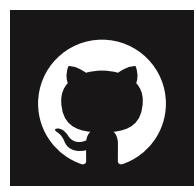
¿TIENES PREGUNTAS? CONTÁCTAME POR :



macaguegi@yahoo.es



/in/guerrerogiraldo



<https://github.com/macaguegi>