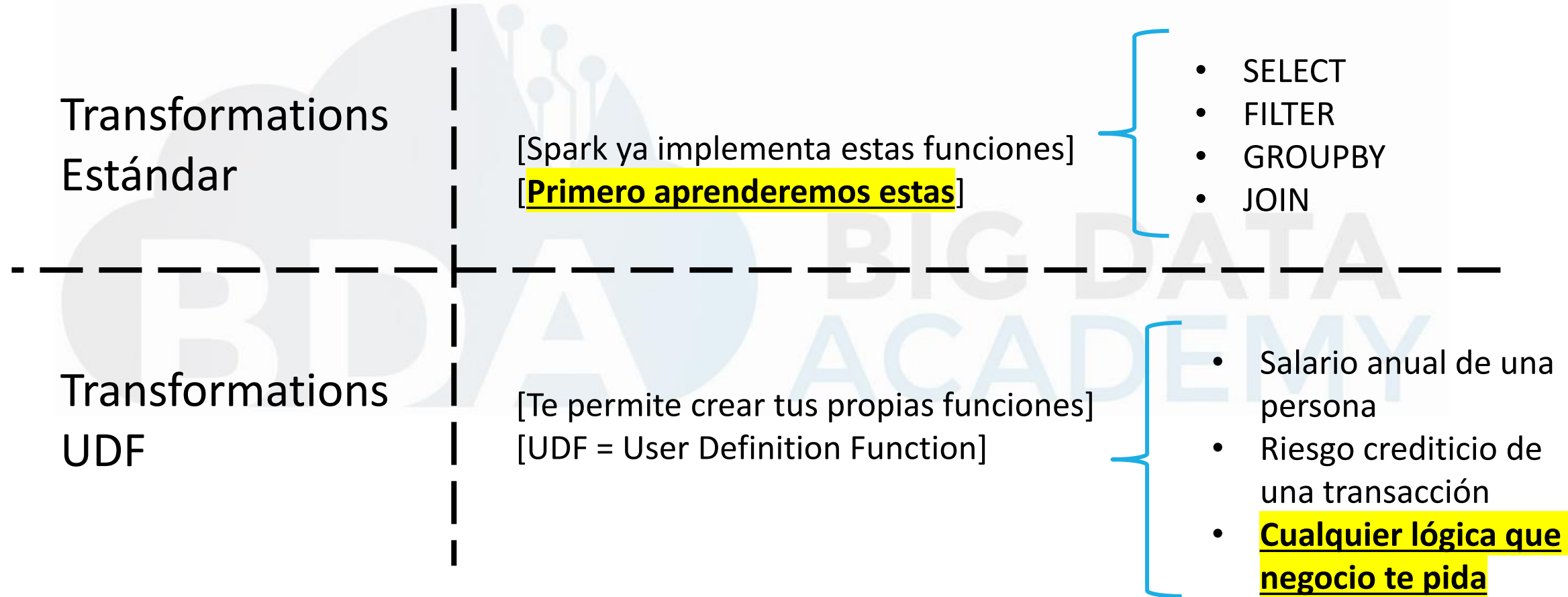


Laboratorio 11

PROCESAMIENTO ESTRUCTURADO

Transformations

Un transformation es una función que permite ejecutar cierta lógica sobre otro dataframe, por ejemplo un filtrado según cierta condición



Transformations SELECT y FILTER

SELECT [ID, EDAD, SALARIO]



Crea un dataframe seleccionando ciertos campos de otro dataframe

ID	NOMBRE	EDAD
1	Carl	32
2	Priscilla	34
3	Jocelyn	27
4	Aidan	29
5	Leandra	41
6	Bert	70
7	Mark	52
8	Jonah	23
9	Hanae	69
10	Cadman	19
11	Melyssa	48
12	Tanner	24
13	Trevor	34

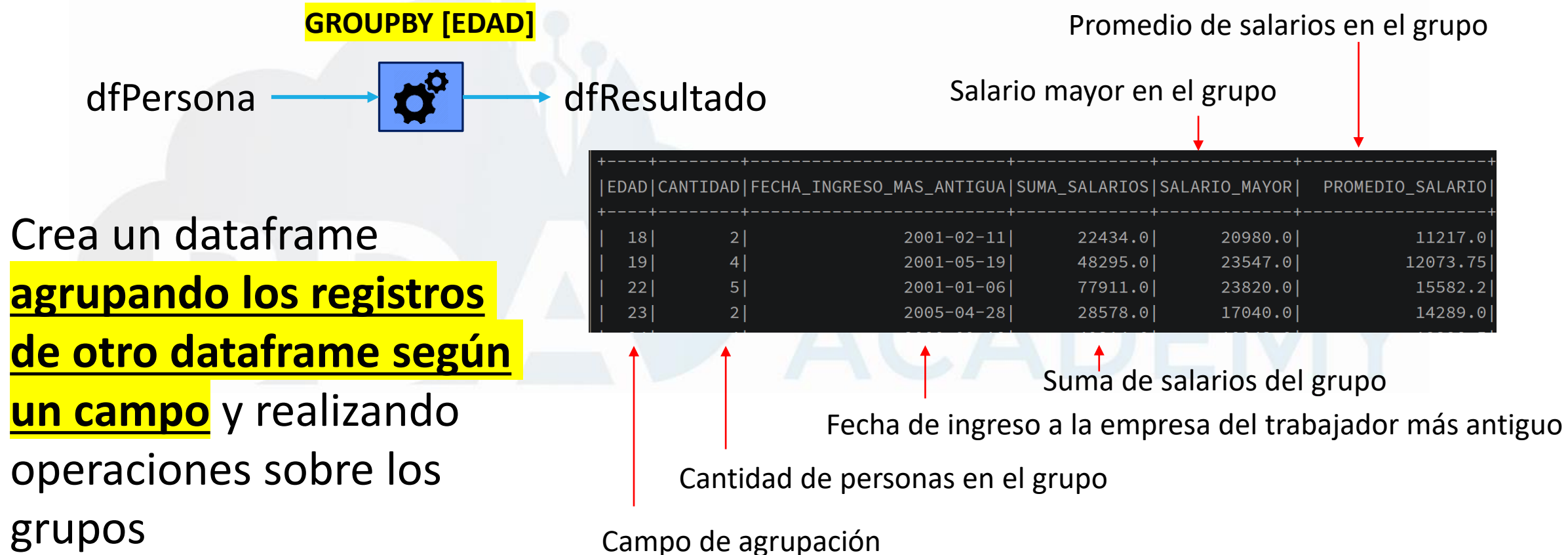
FILTER [EDAD > 60]



Crea un dataframe filtrando los registros de otro dataframe según cierta condición

CHA_INGRESO	EDAD
2017-04-25	70
2003-05-25	69
2017-03-09	70
2002-08-21	67
2013-05-29	64
2009-07-27	67
2008-03-24	61
2008-03-07	63
2011-10-19	64
2012-04-05	70
2008-12-13	67

Transformation GROUPBY



Transformation JOIN

dfTransaccion

ID_PERSONA	ID_EMPRESA	MONTO	FECHA
18	3	1383.0	2018-01-21
30	6	2331.0	2018-01-21
47	2	2280.0	2018-01-21
28	1	730.0	2018-01-21
91	4	3081.0	2018-01-21
74	8	2409.0	2018-01-21

dfPersona

ID	NOMBRE	TELEFONO	CORREO	FECHA_INGRESO	EDAD	SALARIO	ID_EMPRESA
15	Wanda	359-6973	Nam.nulla.magna@I...	2005-08-21	27	1539.0	5
16	Alden	341-8522	odio@morbitristiq...	2006-12-05	26	3377.0	2
17	Omar	720-1543	Phasellus.vitae.m...	2014-06-24	60	6851.0	6
18	Owen	1-167-335-7541	sociis@erat.com	2002-04-09	34	4759.0	7

JOIN [ID_PERSONA = ID]

Campos de dfPersona
[Datos de la persona que
hizo la transacción]

Campos de dfTransaccion

ID_PERSONA	NOMBRE	EDAD	SALARIO	ID_EMPRESA_TRABAJO	ID_EMPRESA_TRANSACCION	MONTO	FECHA
18	Owen	34	4759.0	7	3	1383.0	2018-01-21
30	Clayton	52	9505.0	8	6	2331.0	2018-01-21
47	Vernon	35	7109.0	10	2	2280.0	2018-01-21
28	Stephen	53	9469.0	8	1	730.0	2018-01-21
91	Erica	32	8934.0	6	4	3081.0	2018-01-21
74	Kaitlin	56	6515.0	10	8	2409.0	2018-01-21
41	Wynne	31	19522.0	4	2	3754.0	2018-01-22
42	Wanda	42	5419.0	9	9	4079.0	2018-01-22
24	Amaya	24	1801.0	8	6	4475.0	2018-01-22

Cruza dos dataframes en función de un campo de coincidencia, **si la coincidencia se cumple, se extraen en un nuevo dataframe ciertos campos de ambos dataframe**