Dataframes métodos usuales

agg Devuelve un data frame con una única fila con varias columnas con valores que provienen de funciones de agregación tales como: avg, count, first, last, max, mean, min, sum, sumDistinct

columns Atributo que devuelve una lista con los nombres de las columnas

count Devuelve el numero de filas del DataFrame

distinct Retorna un nuevo DataFrame que contiene las filas que son distintas

dtypes un atributo que devuelve el nombre y tipo de dato de cada columna

drop (col) Devuelve un nuevo del que se ha eliminado la columna col

dropDuplicates Retorna un nuevo DataFrame del que se han eliminado filas duplicadas A diferencia distinct se puede especificar las columnas que se usarán para comprobar la duplicidad. `df.dropDuplicates(Array("a", "b"))`

na Para el manejo de valores nulos. Se puede eliminar, llenar o reemplazar valores nulos
drop elimina

fill	Llena los nulos con un valor especifico
replace	Reemplaza nulos con un valor de una columna especifica

filter o where	Filtra las filas que cumplen una condicion dada
groupBy	Agrupa las filas del DataFrame por una o más columnas Sobre el dataframe resultante se puede realizar funcione de agregación (avg, min, max, sum, ...)
intersect union union All union By Name	Devuelven un dataframe con la intersección y unión, es decir intersect devuelve un nuevo dataframe con las filas comunes entre dos dataframes union devuelve todas las filas de ambos dataframes, sin eliminar duplicados union All igual a union unionByName une 2 dataframes por el nombre en lugar de la posición
limit	Limita el resultado a n filas
orderBy	ordena el dataframe según una o más columnas
select	Selecciona columnas especificas del dataframe

selectExpr — Útil para realizar selecciones complejas y transformaciones en un dataframe usando sintaxis familiar de SQL

```
df.selectExpr("a*2", "b<18 as M")
```

withColumn — Se usa para crear una nueva columna o reemplazar una existente. Se puede usar para aplicar transformaciones o agregar columnas derivadas de otras

withColumnRenamed — Se usa para cambiar el nombre de una columna existente

join — Permite unir filas de dos dataframes basándose en una condición específica

Inner Join	filas que coinciden en ambos dataframes
Left outer Join	todas las filas del df izq. y las coincidencias del de dere.
Rigth Outer Join	todas las filas del df dere. y las coincidencias del de la izq.
Full outer Join	todas las filas cuando hay una coincidencia en uno de los df
Cross Join	Devuelve el producto cartesiano de los 2 df.

by: /jorgaf
@jorgaf