**Assignment 14**

**Name:** Jeevan Sai Badana

**Mail:** [jeevansai100@gmail.com](mailto:jeevansai100@gmail.com)

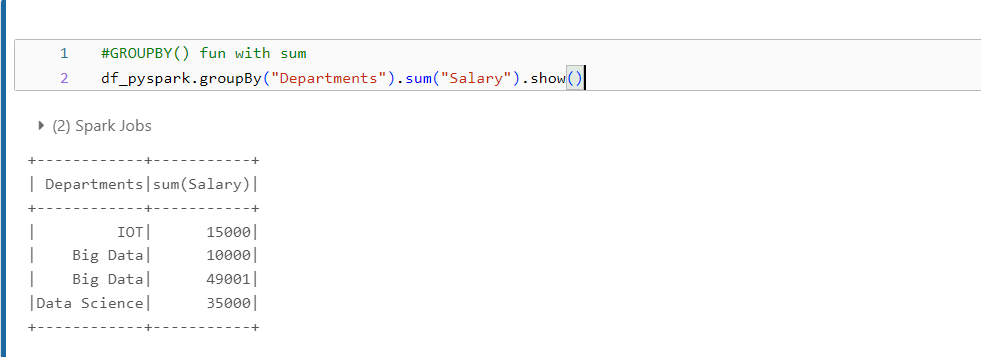
**Topic:** Aggregations and Joins in spark

Date: 09-02-2024

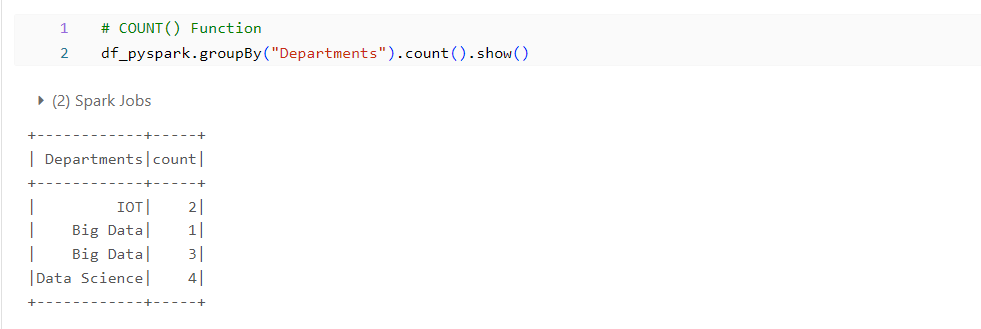
First we have to create one csv file for this and we read data in that using read.csv as bellow



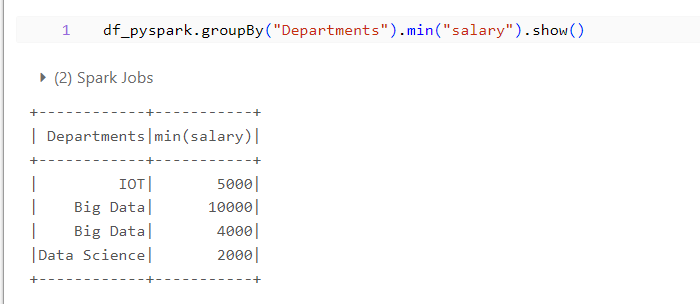
GROUPBY() fun with sum:

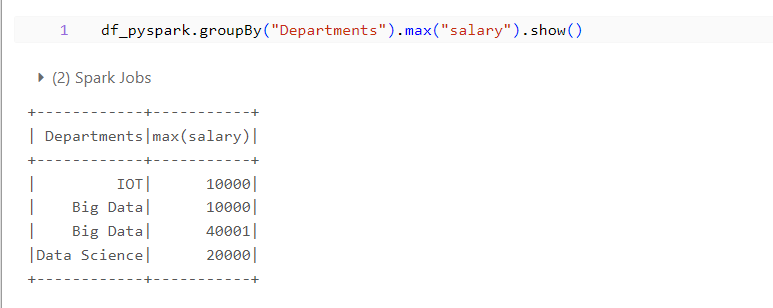


COUNT() Function:

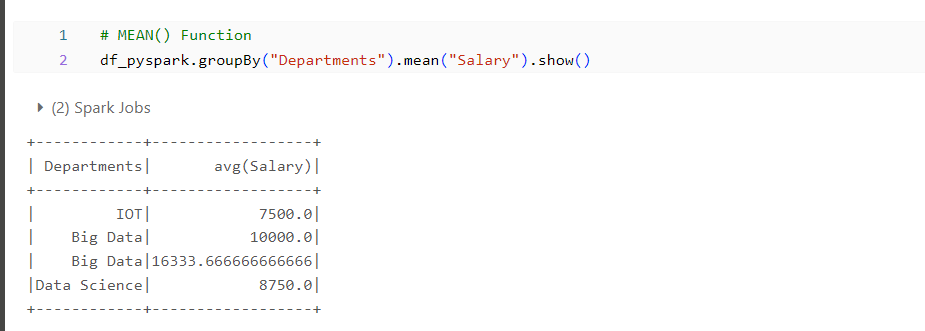


Group by with min and max:

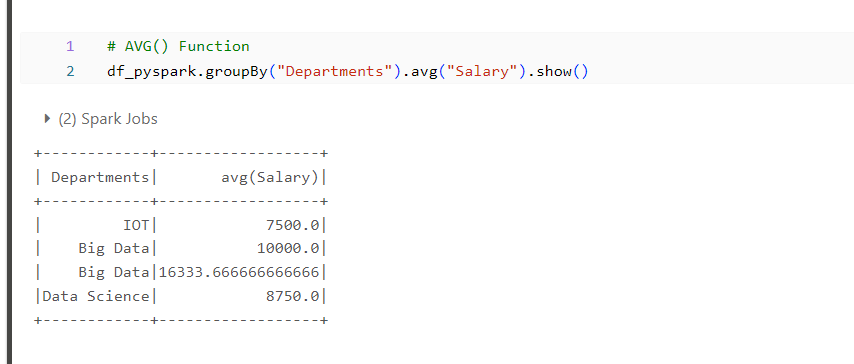




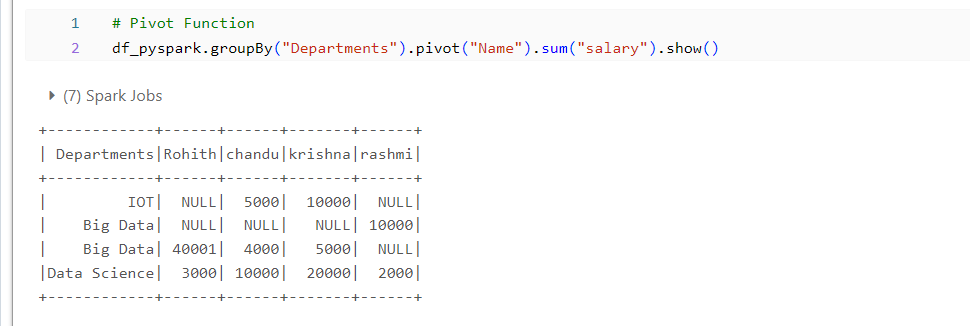
MEAN() Function:



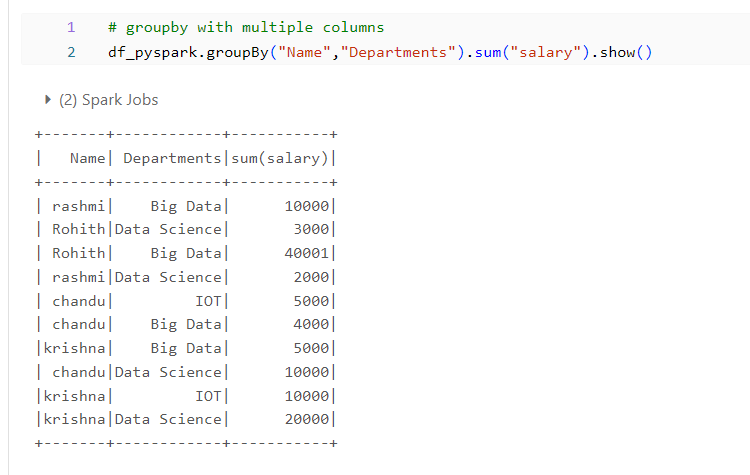
AVG() Function:



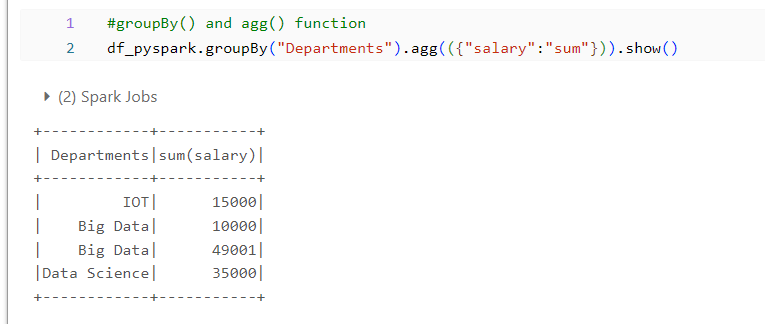
Pivot Function:

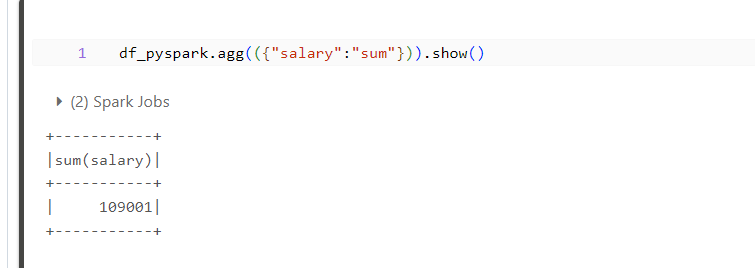


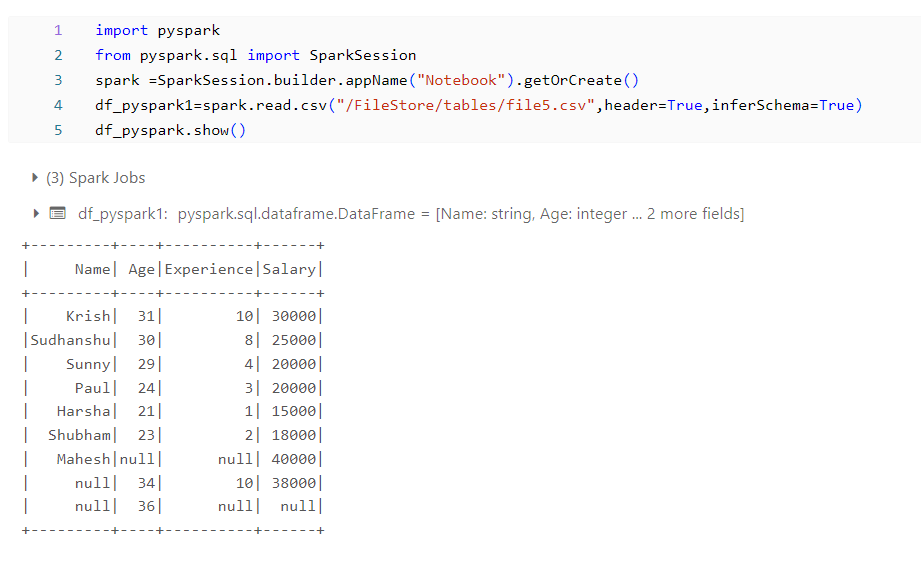
groupby with multiple columns:



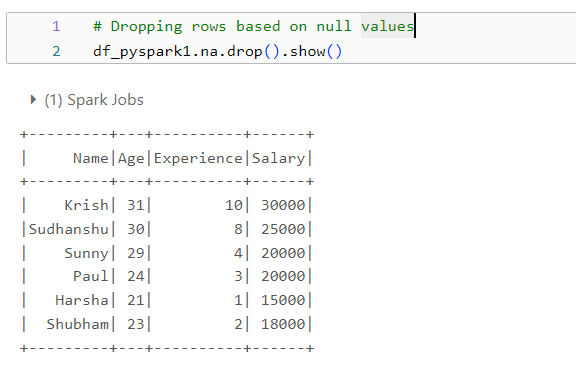
groupBy() and agg() function:



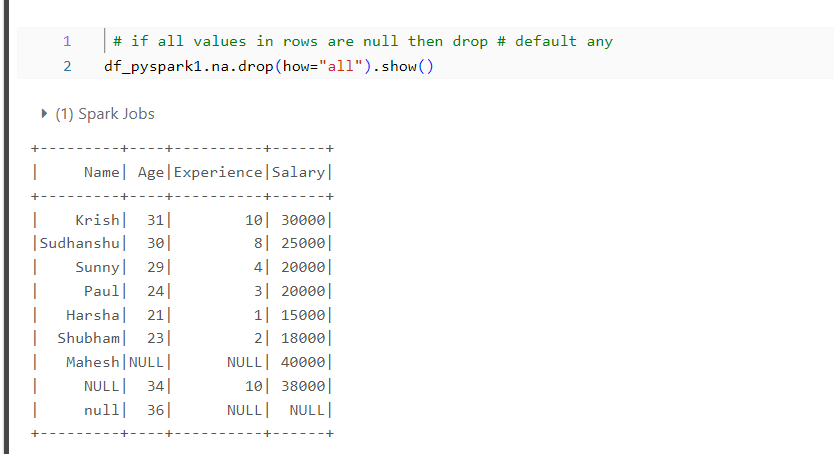


Now we create another csv file with null values  


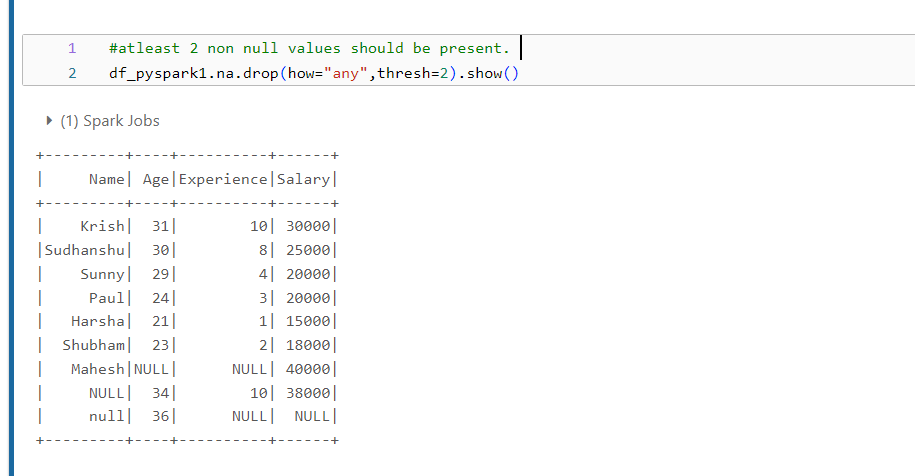
Dropping rows based on null values:



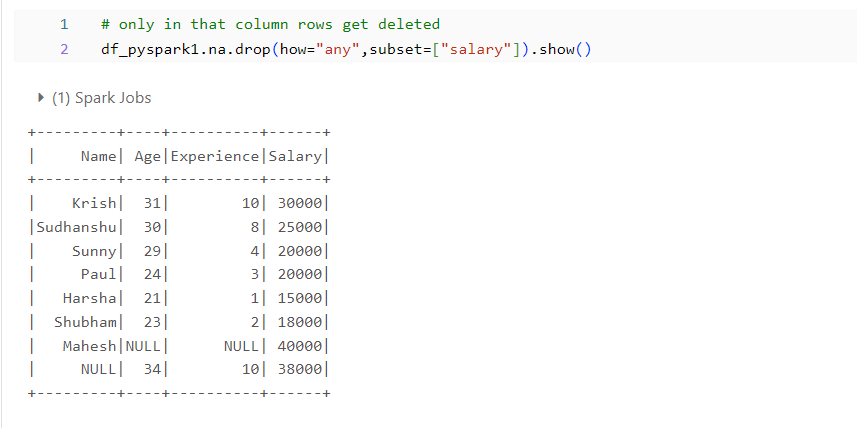
if all values in rows are null then drop:



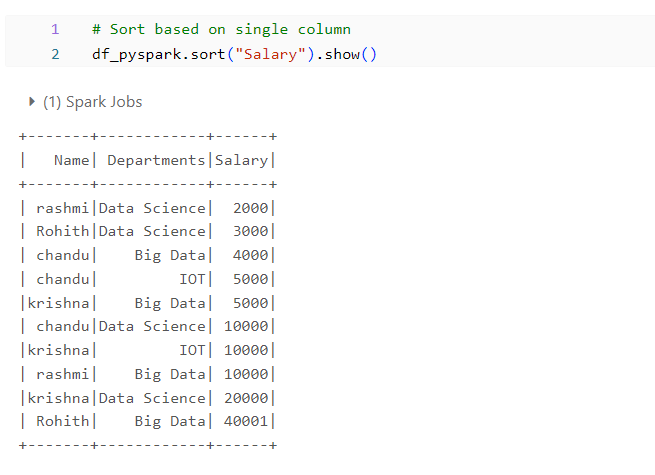
atleast 2 non null values should be present.



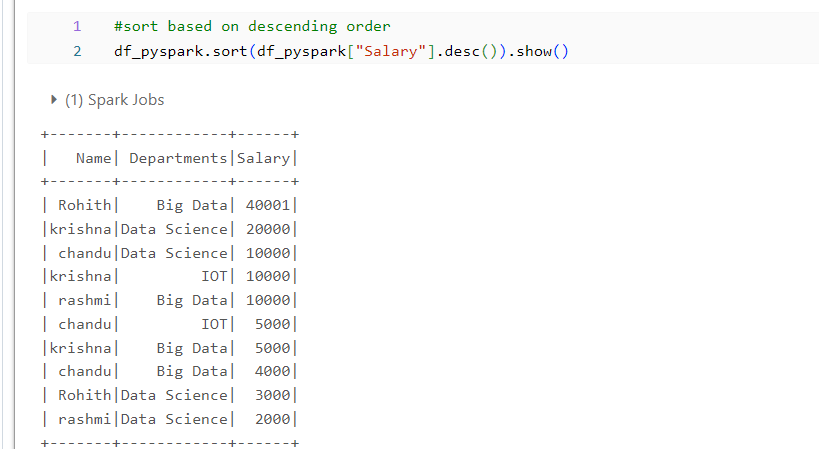
only in that column rows get deleted:



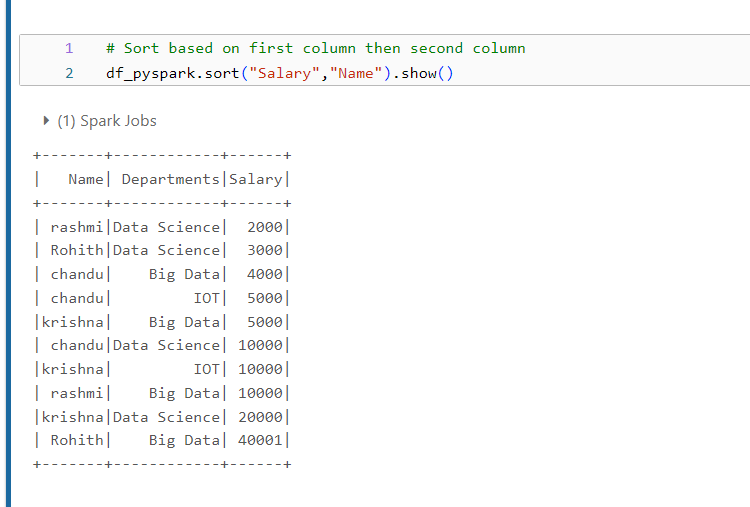
Sort based on single column:



sort based on descending order:

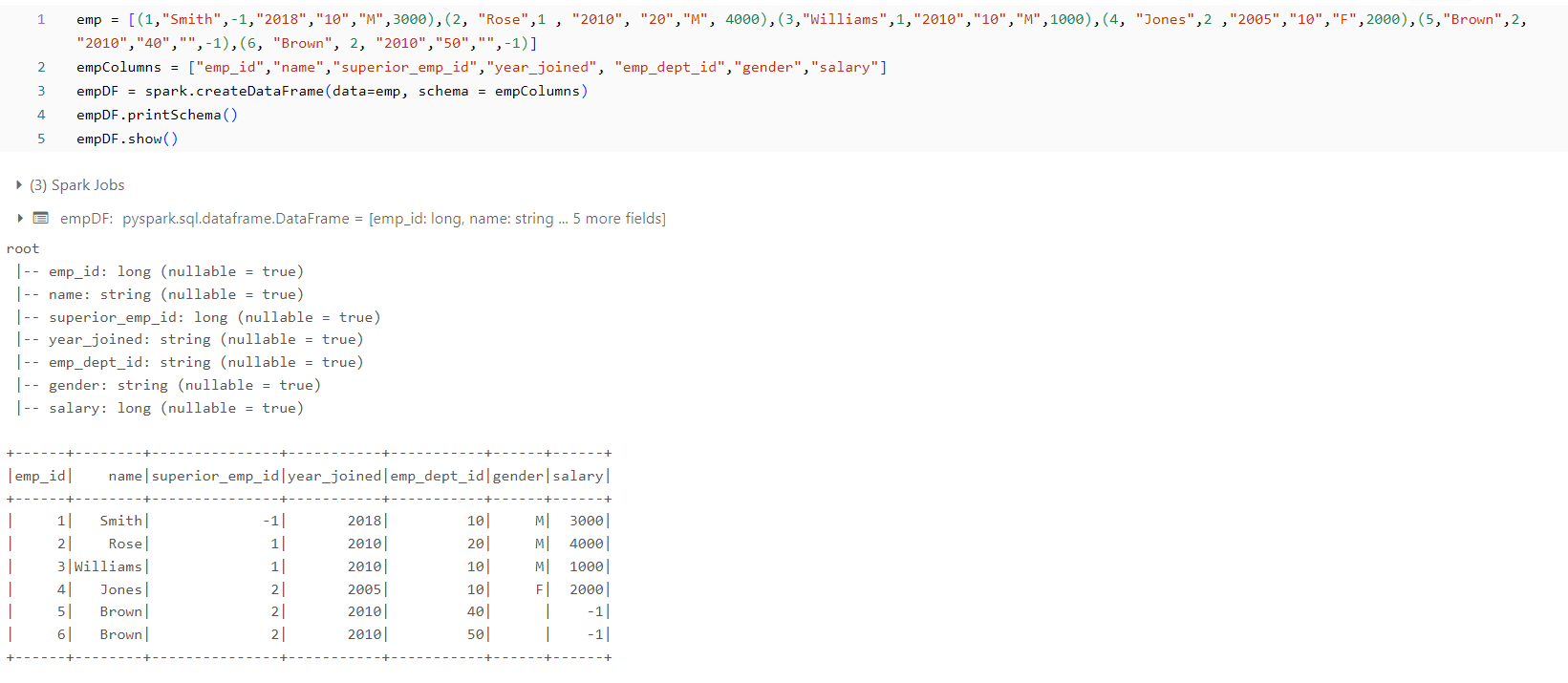


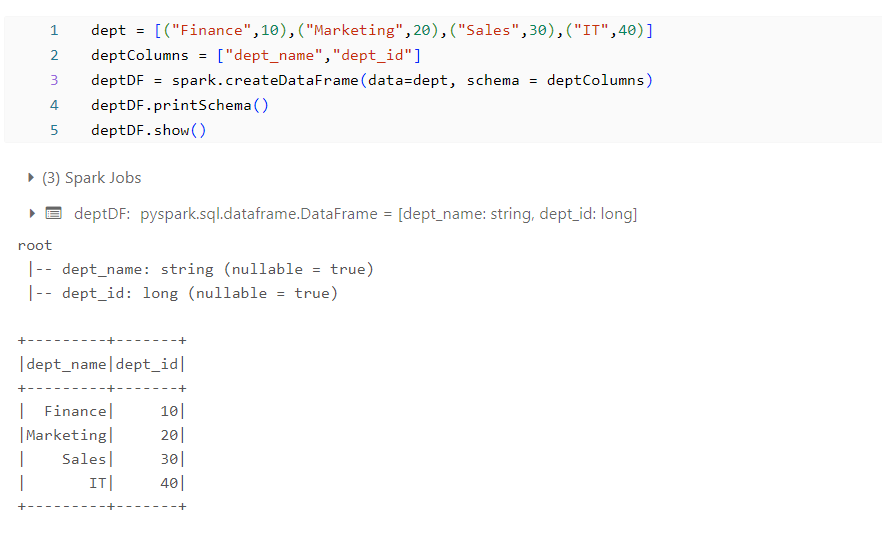
Sort based on first column then second column:



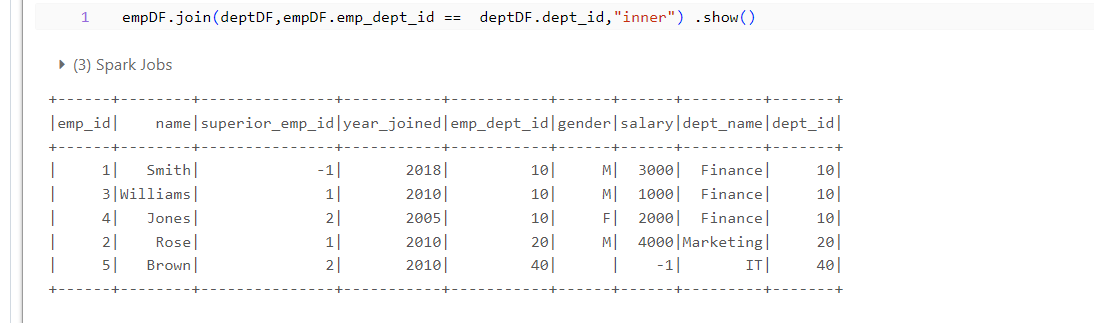
JOINS

We are create two tables emp and dept first for this purpose

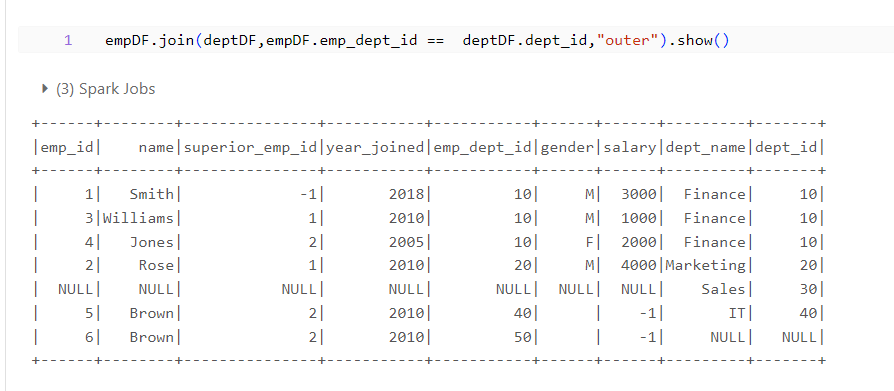




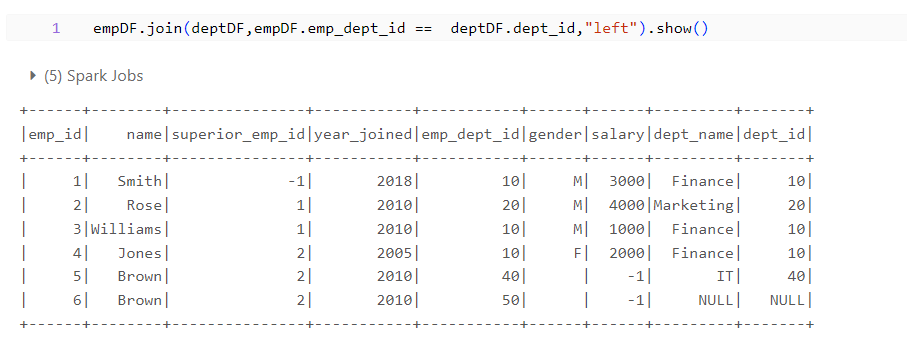
Inner join:



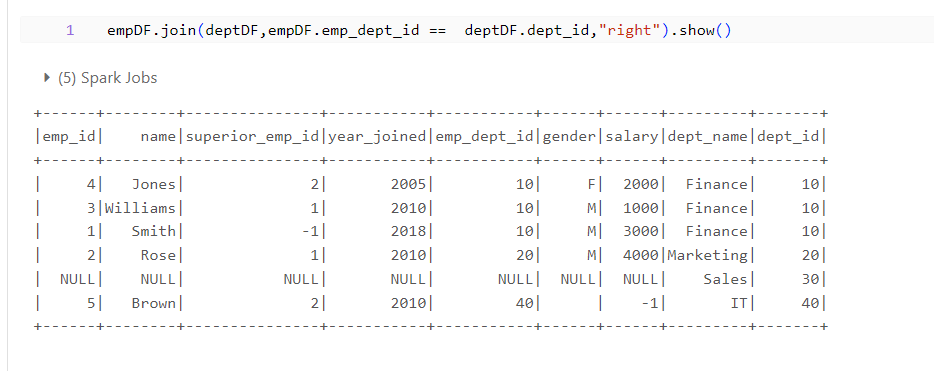
Outer join:



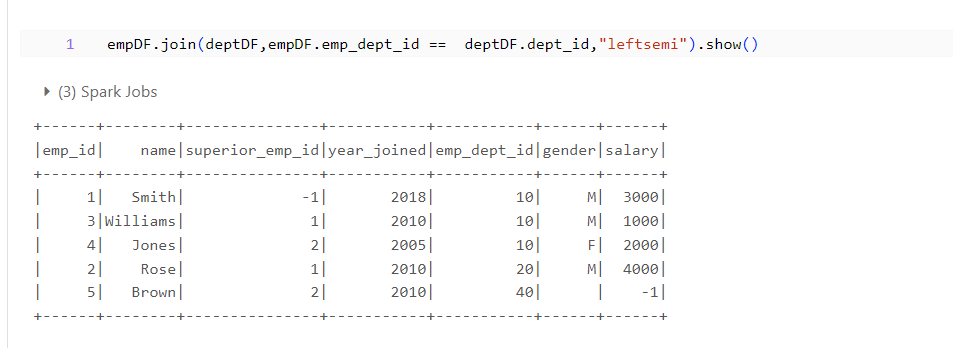
Left join:



Right join:



Leftsemi join:



Leftanti join:

