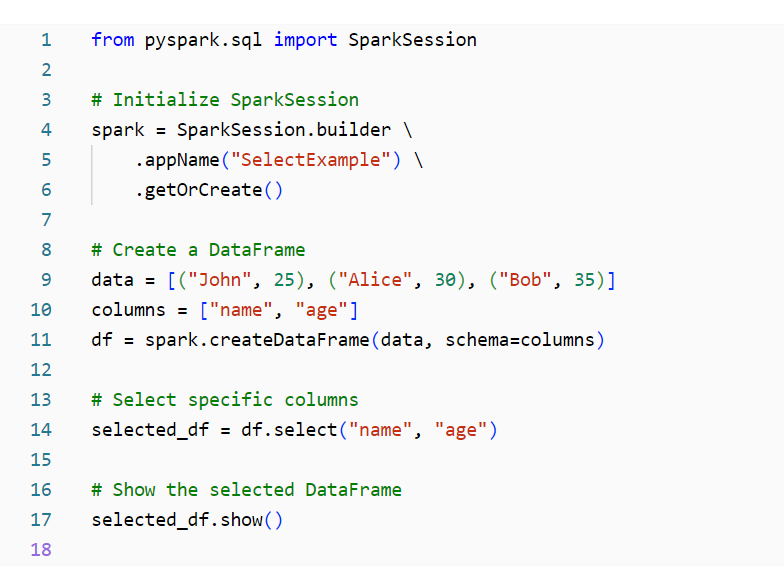
Name : Vishal Jha

Batch : Data Engineering

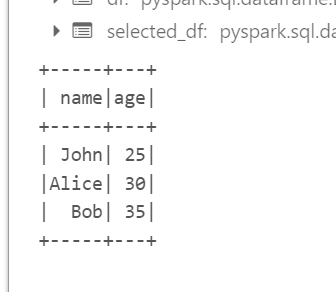
Date : 09-02-2024

Topic : PySpark – 4

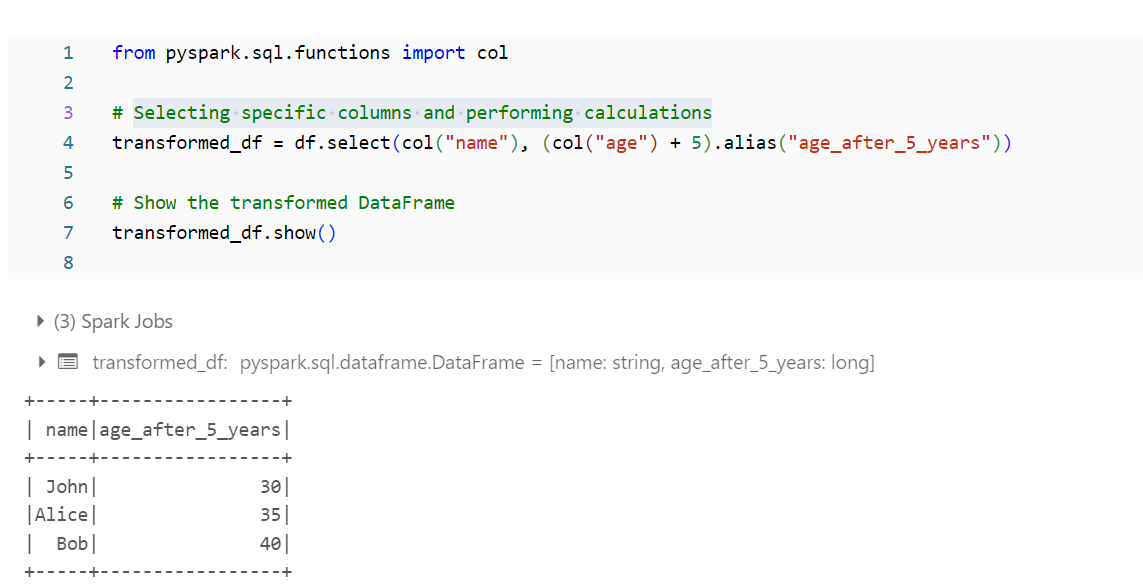
1. Select statement –



Output :



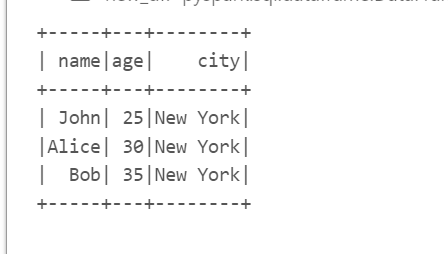
1. Selecting specific columns and performing calculations:



1. Add new column with constant value :



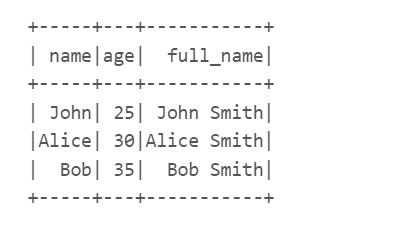
Output:



1. Concat-



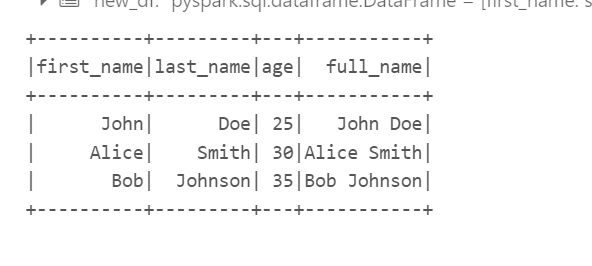
Output :



1. Concat-ws:



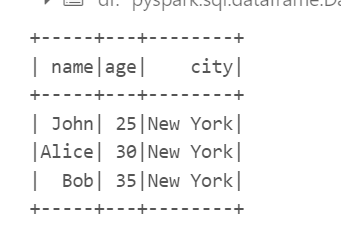
Output:



1. Add column if not exist in dataframe:



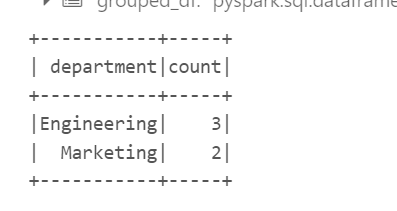
Output:



1. Group by –



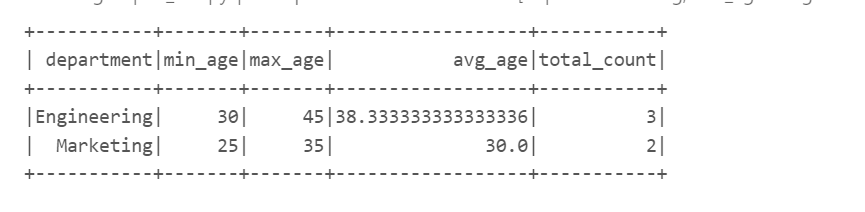
Output :



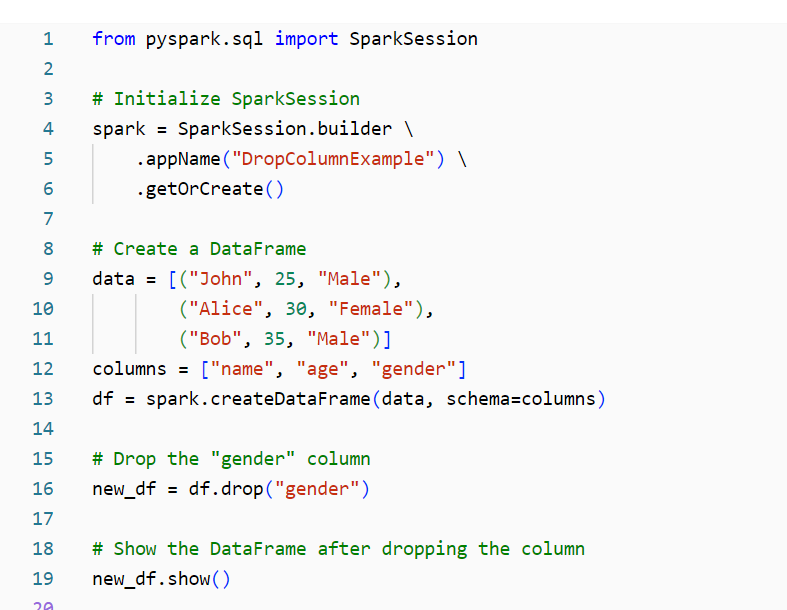
1. Aggregation functions- Min,Max,Avg,Sum,Count:



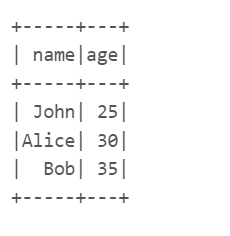
Output :



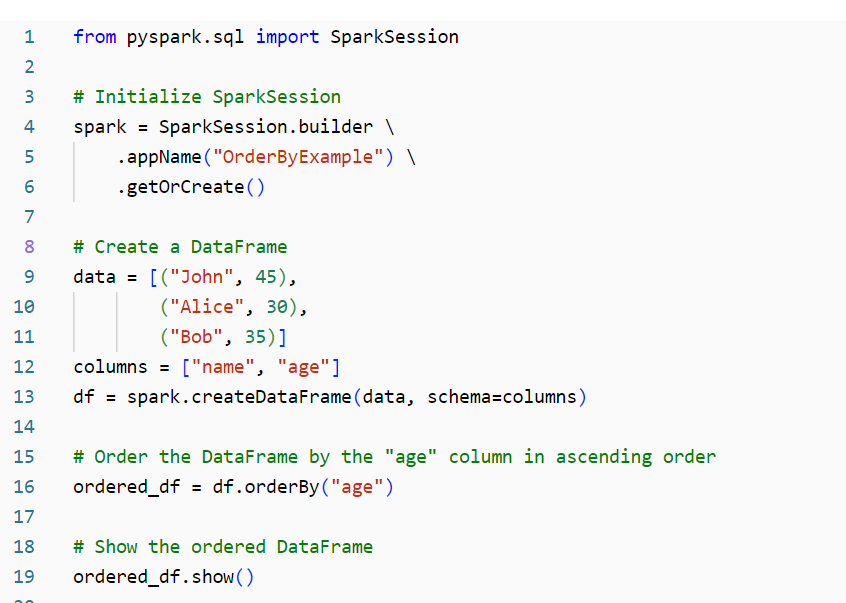
1. Drop :



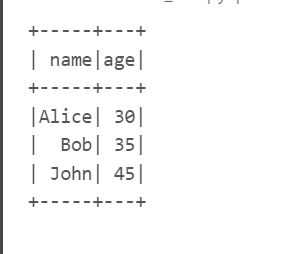
Output:



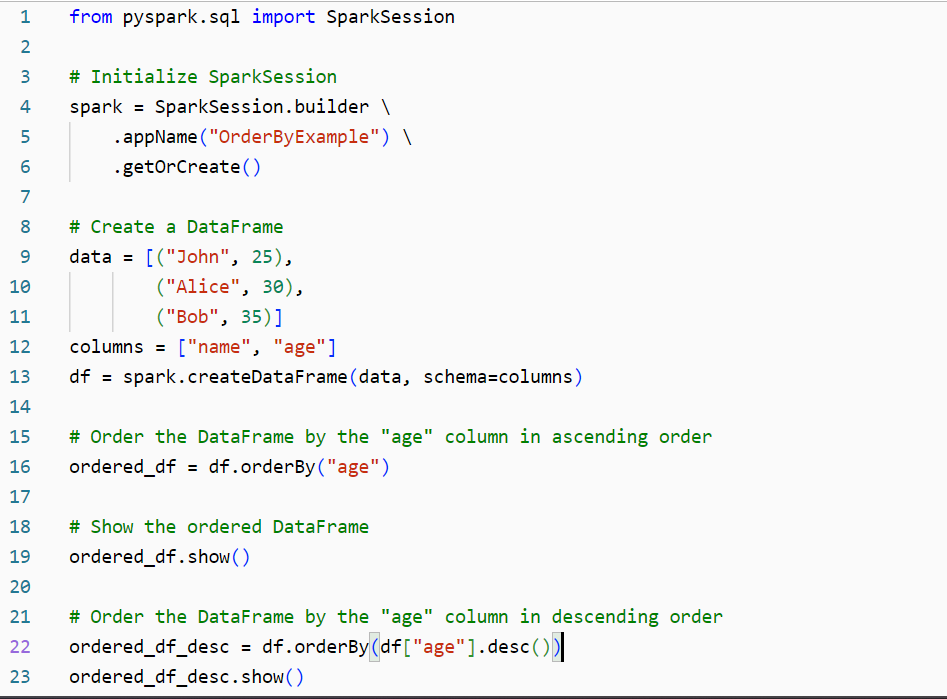
1. Order BY and Sort:



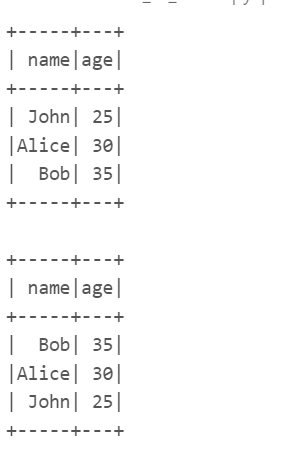
Output:



1. Sort in Descending order:



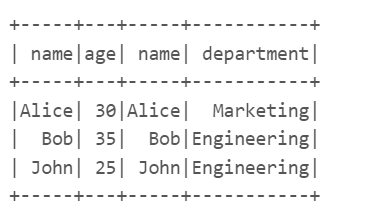
Output:



1. Inner Join:



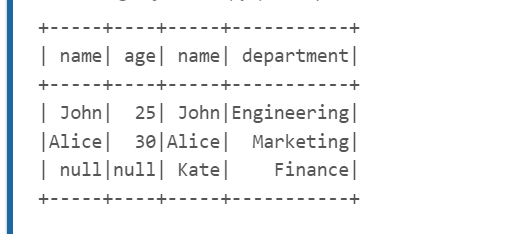
Output:



1. Right Join:



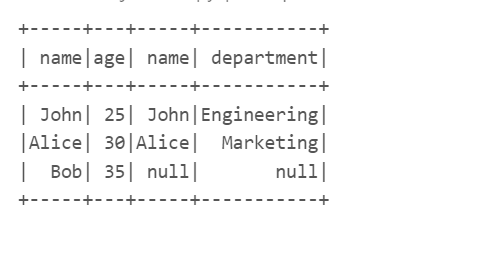
Output:



1. Left Outer Join:



Output:



1. Union:



Output:

