**Q. Execute Manipulating, Droping, Sorting, Aggregations, Joining, GroupeBy  DataFrames**

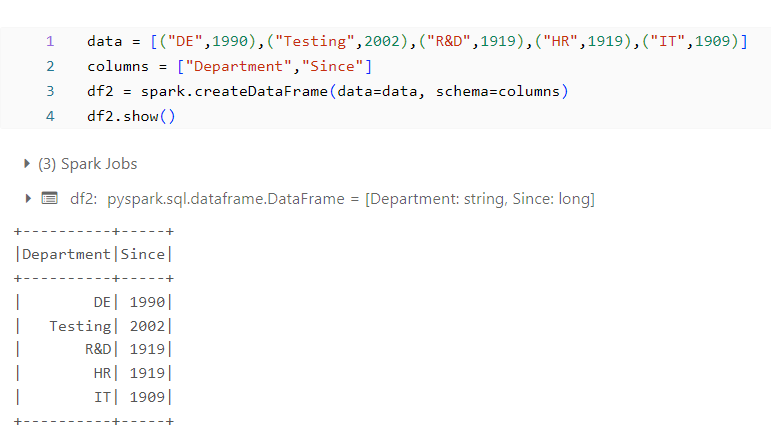
A. Initially we imported the Spark Session



Then we created a sample dataframe ‘df1’ with some sample data

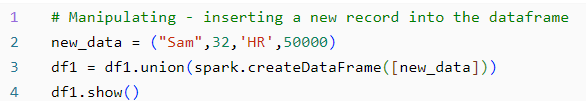


We also created a second dataframe ‘df2’ with sample data

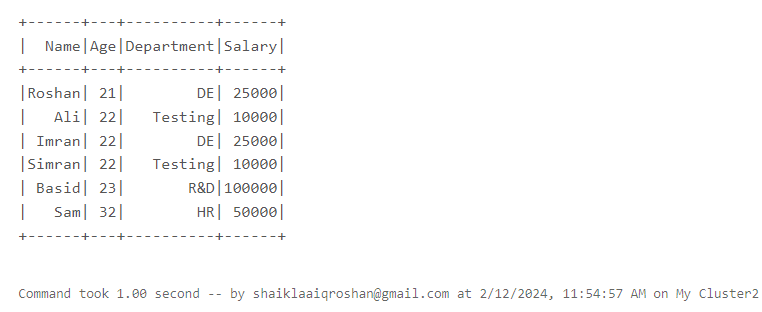


**Manipulating:**

We did manipulating operation by inserting a new record into the dataframe ’df1’

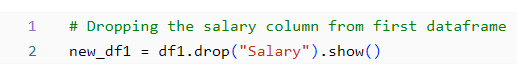


Output:

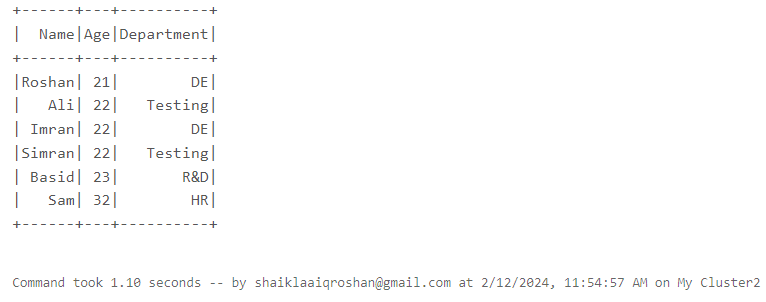


**Dropping:**

We performed the dropping operation by creating a new dataframe in which we dropped the Salary column from df1.

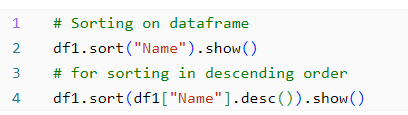


Output:

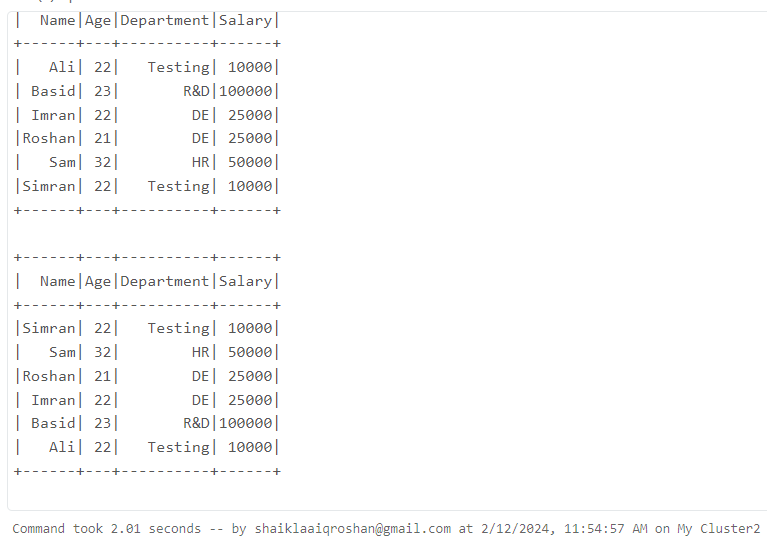


**Sorting():**

We sorted the data based on the name using the sort function. We sorted the data based on name in both ascending and descending order.



Output:



**Group by aggregations :**

We performed the group by aggregations on the data,

**Sum()** – used to calculate the total of data

**Avg()** – used to calculate the average of data

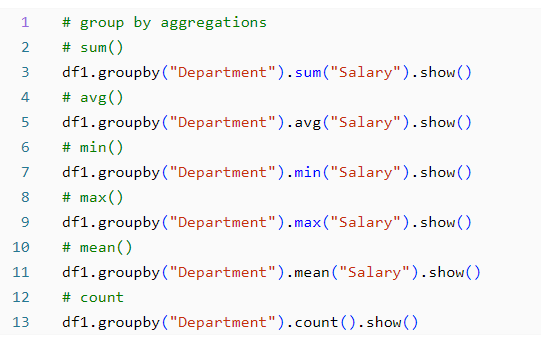
**Mean()** – used to calculate the mean of data

**Min()** – used to calculate the minimum of data

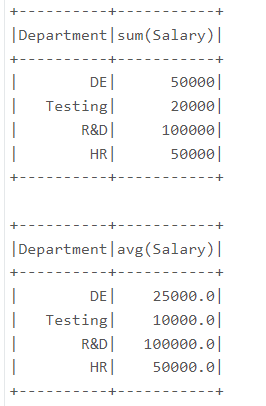
**Max()** – used to calculate the maximum of data

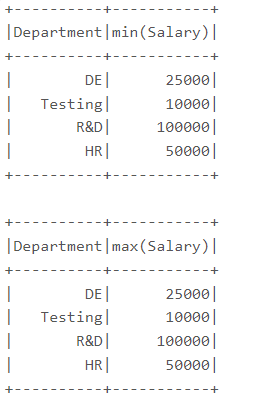
**Count()** – returns the number of records

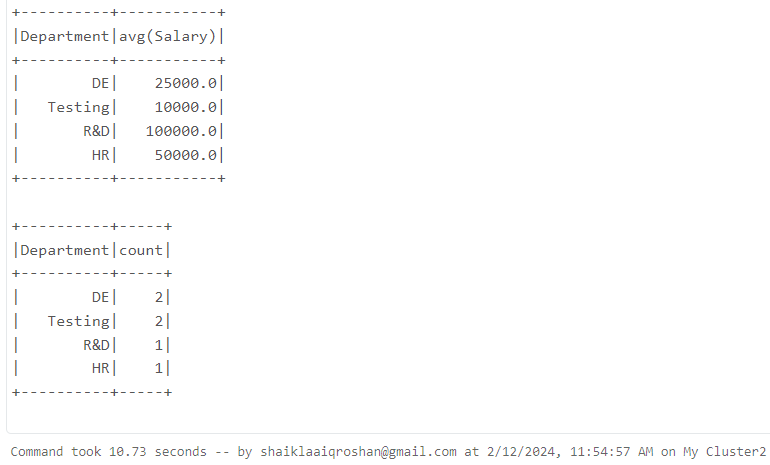
Here we grouped according to the department and salary column is used in aggregation functions.



Output:

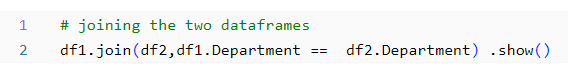






**Join:**

We joined the two dataframes using the join operation by taking the foreign key as ‘Department’. The common values between the dataframes will be printed.



Output:

