**Name: Kanak Agrawal**

**Email:** [**Kanakagrawal2001@gmail.com**](mailto:Kanakagrawal2001@gmail.com)

**Date: 09-02-2024**

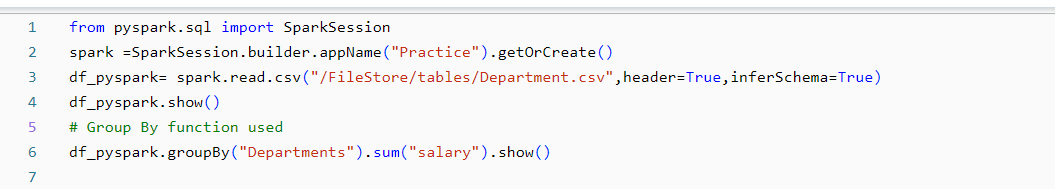
**Day-14 Assignment (Pyspark)**

**Question: Manipulating, Dropping, Sorting, Aggregations, Joins**

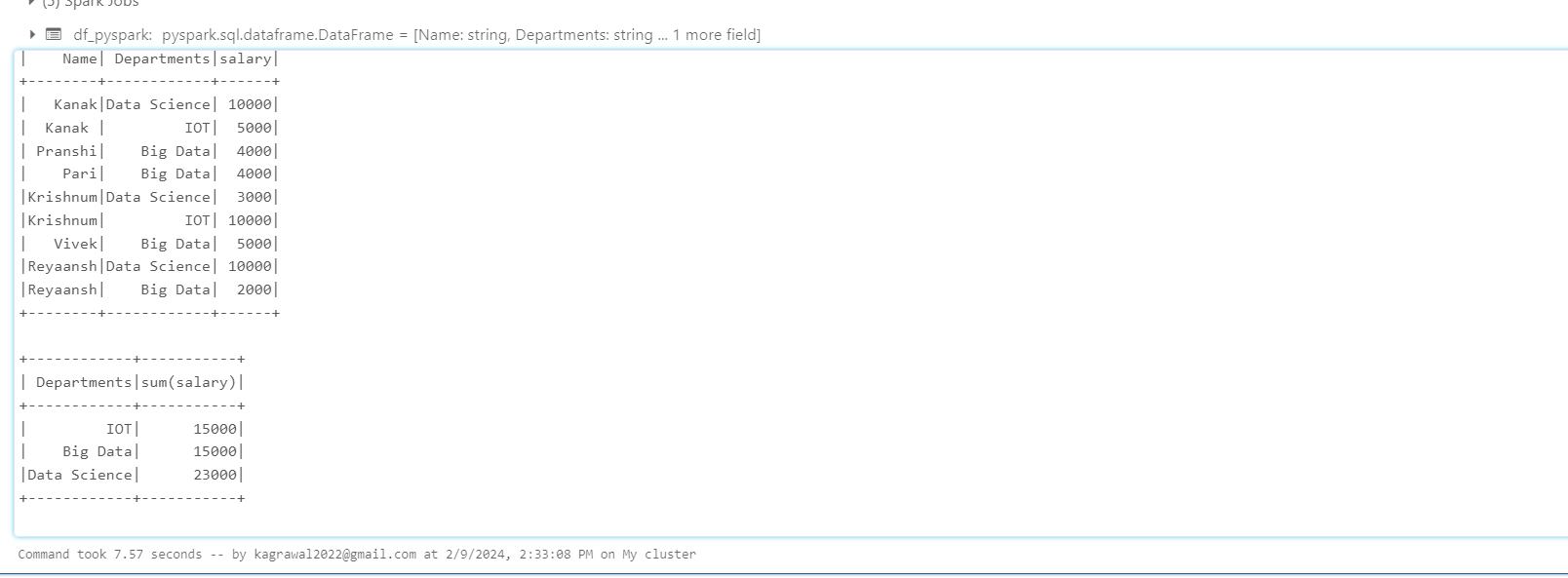
1. **Manipulating**

* **Group By and Aggregate Functions**

**Code:**

****

**Ouput:**

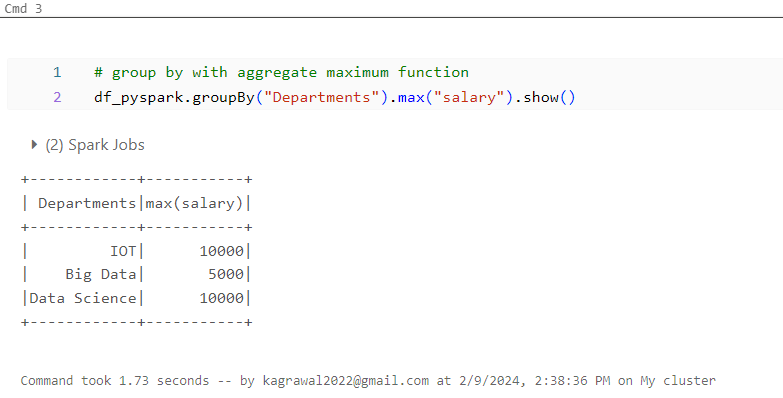
****

* **Aggregate Function**

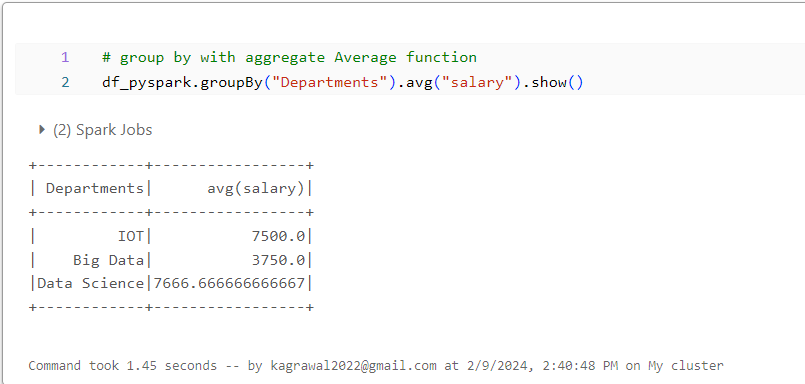
1. **Minimum**

****

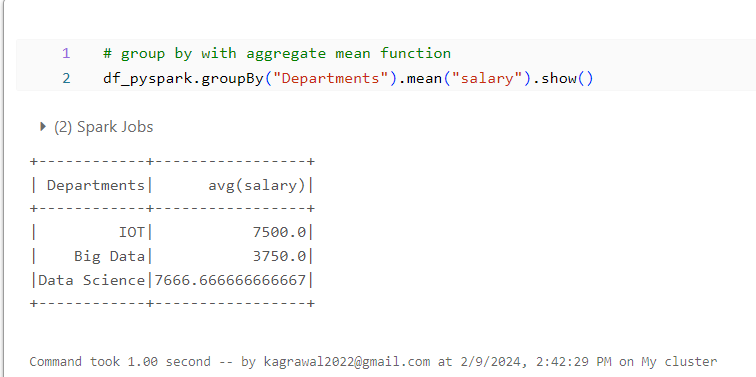
1. **Maximum**

****

1. **Average**

****

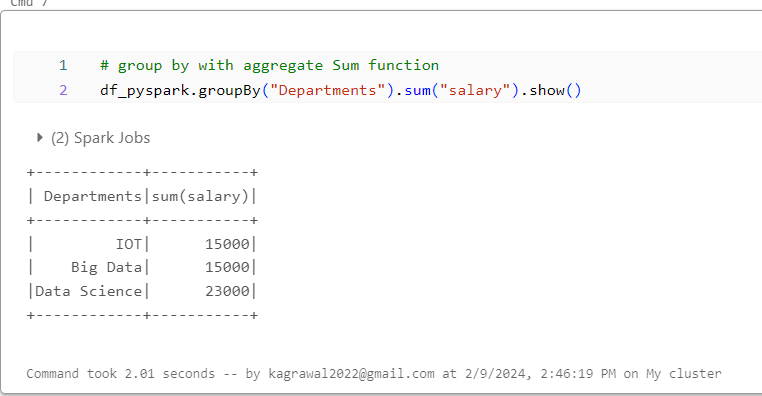
1. **Mean**

****

1. **Count**

****

1. **Sum**

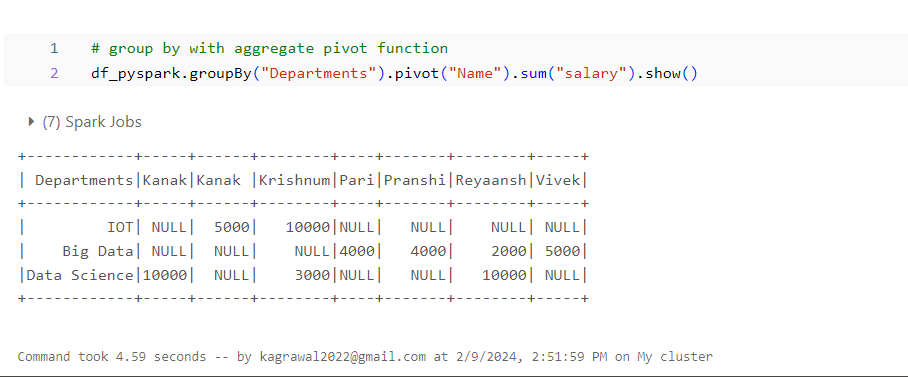
****

1. **Agg**

****

1. **Pivot**

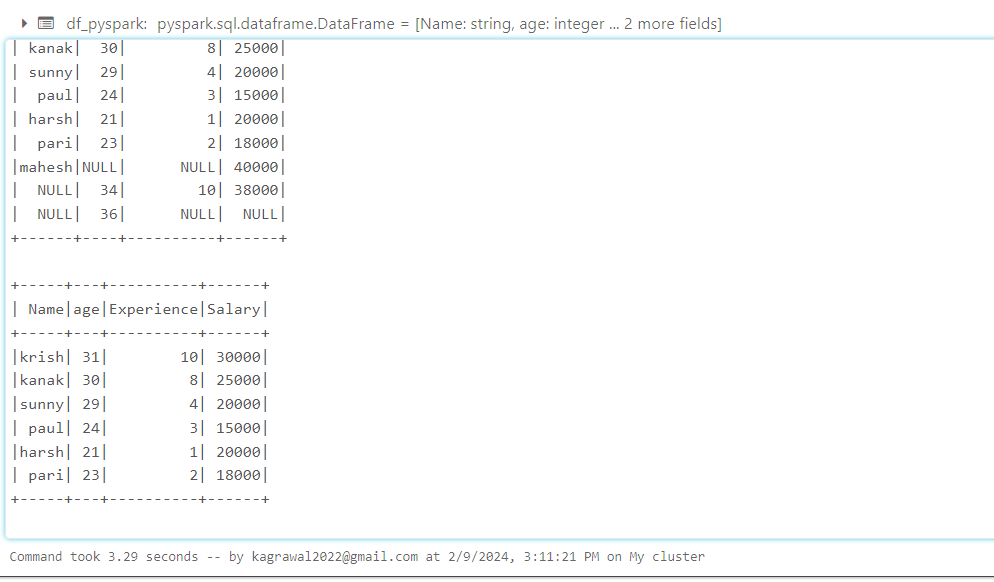
This function is used to convert one column into multiple column(transpose of row to column).



1. **Dropping**

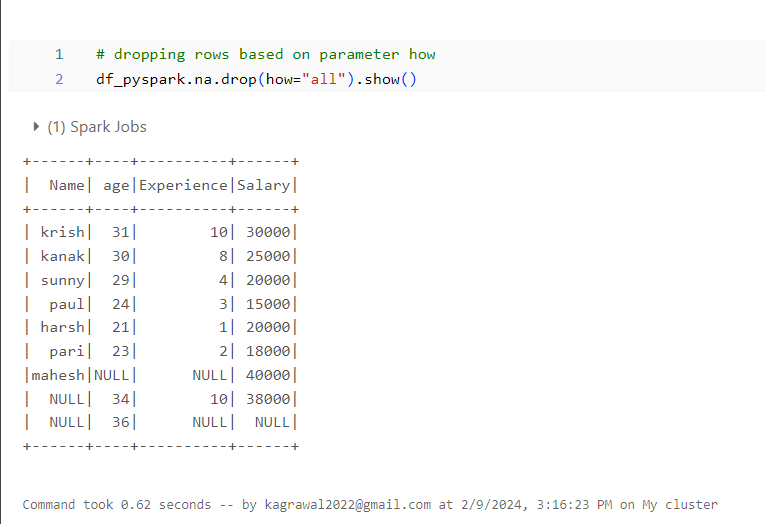
* **Dropping rows based on null values**

****

****

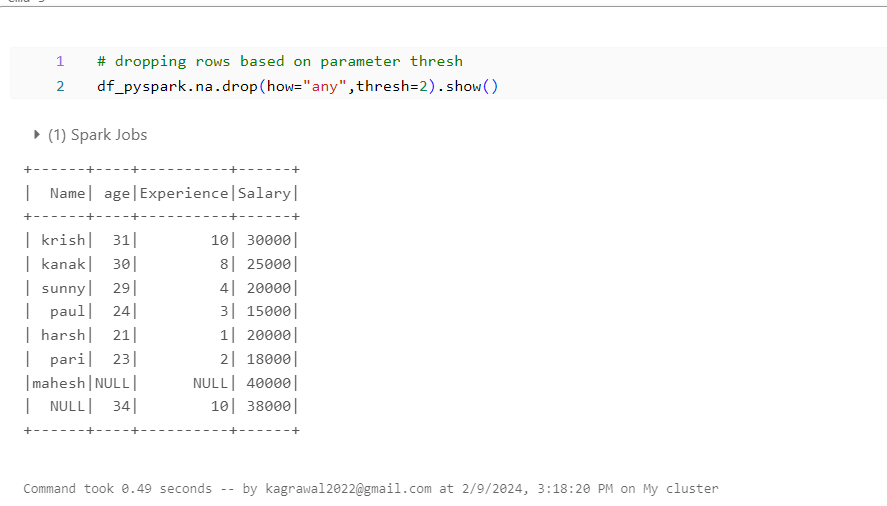
* **Dropping rows based on “how”**

If all values in a row are null then



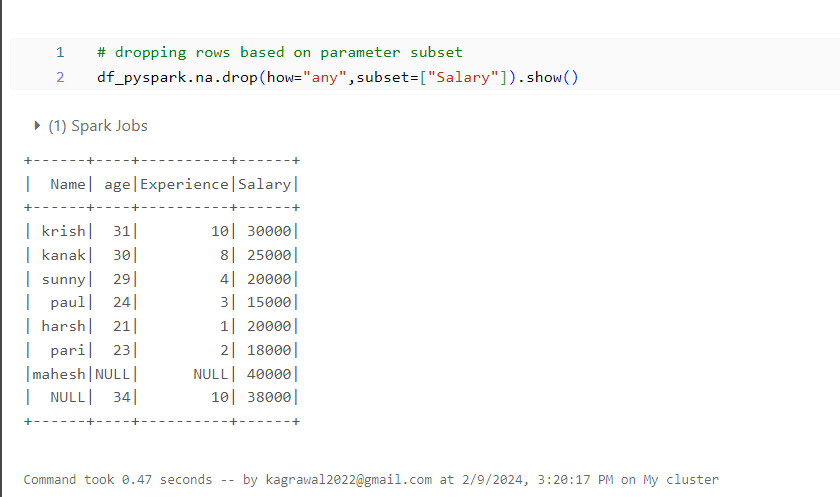
* **Dropping rows based on “thresh”**

Atleast 2 non null values should be present



* **Dropping rows based on “subset”**

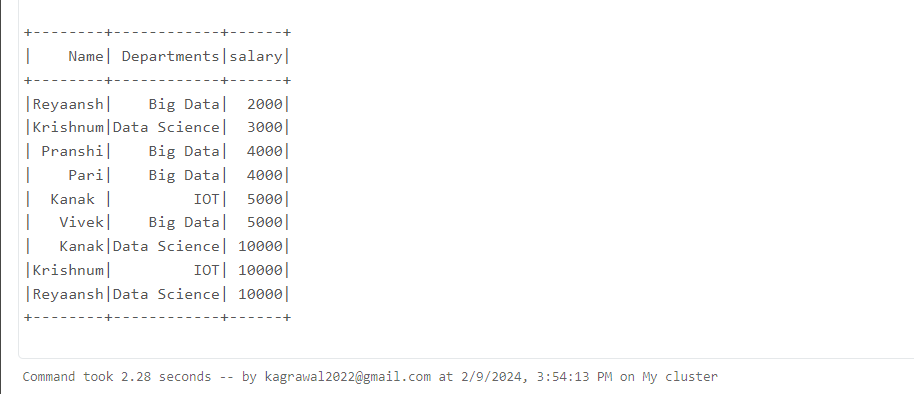
Only in that column rows get deleted



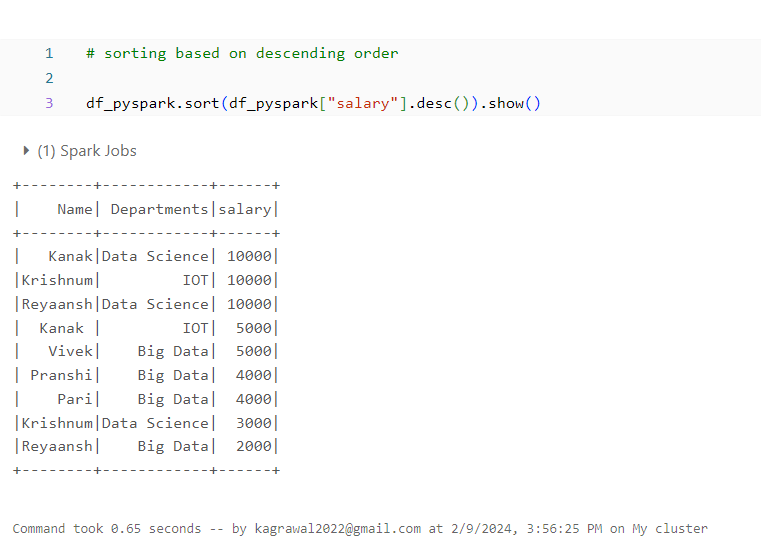
1. **Sorting and order by**

* **Sorting based on single column**

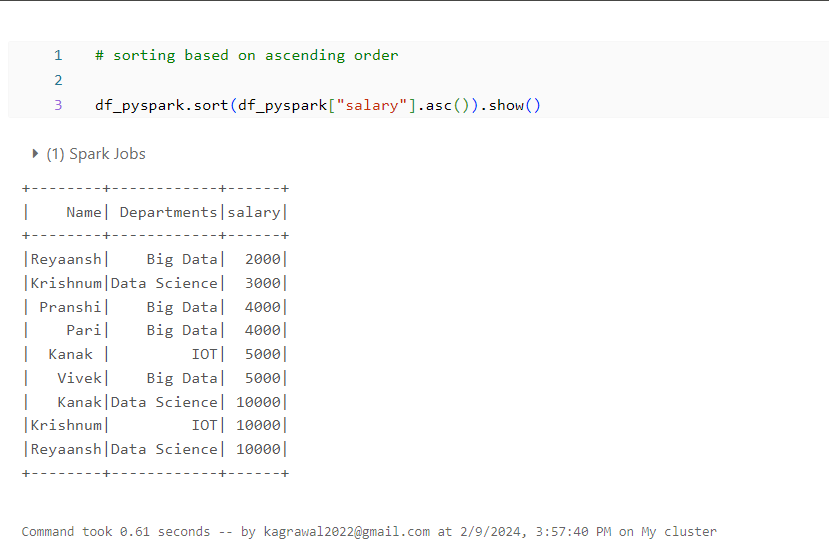
****

****

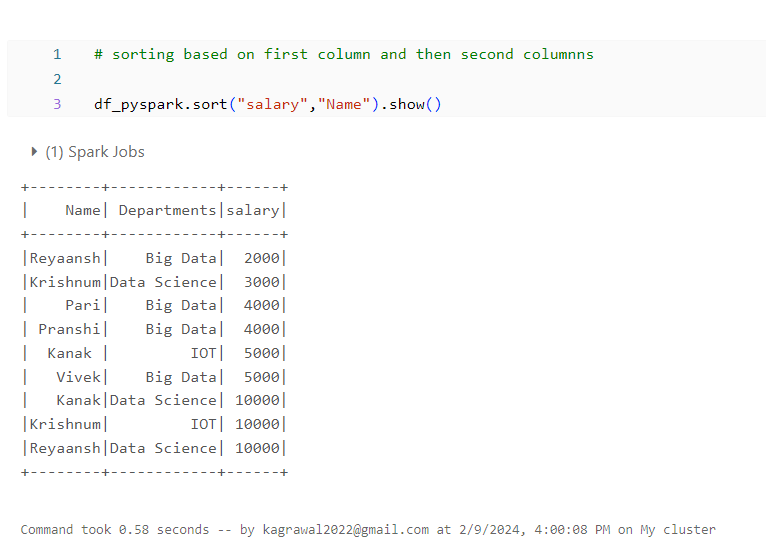
* **Sorting based on descending order**

****

* **Sorting based on Ascending order**

****

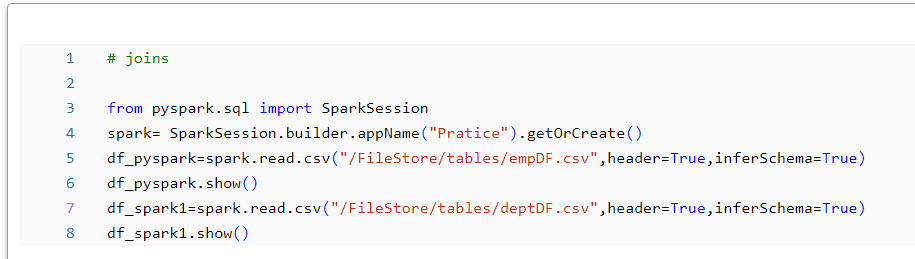
* **Sorting based on first column and then second column**

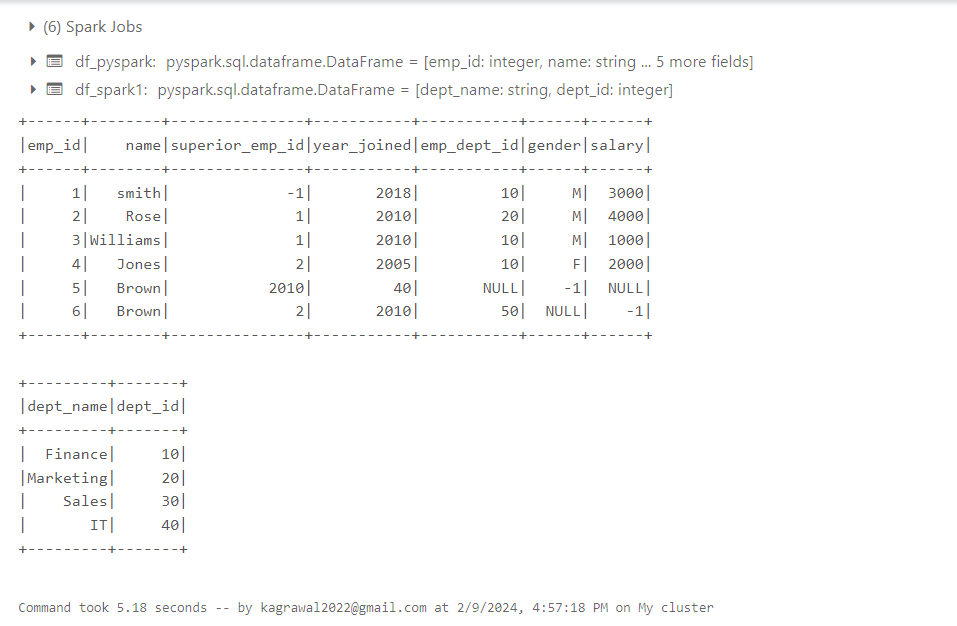
****

* **Order By**

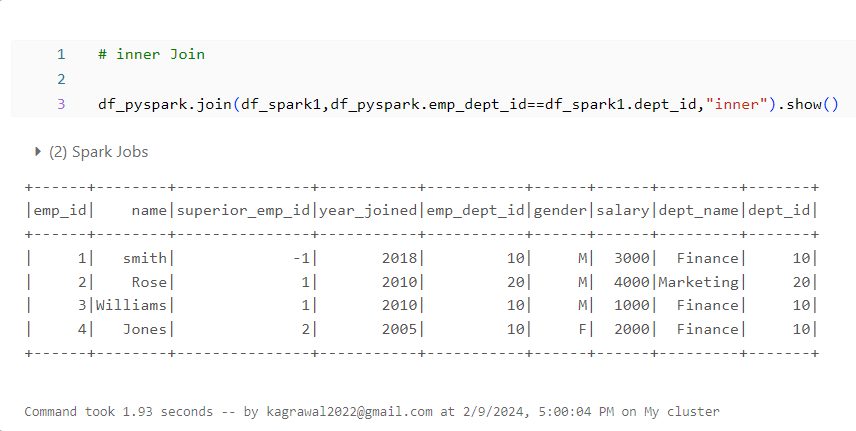
****

1. **Joins**

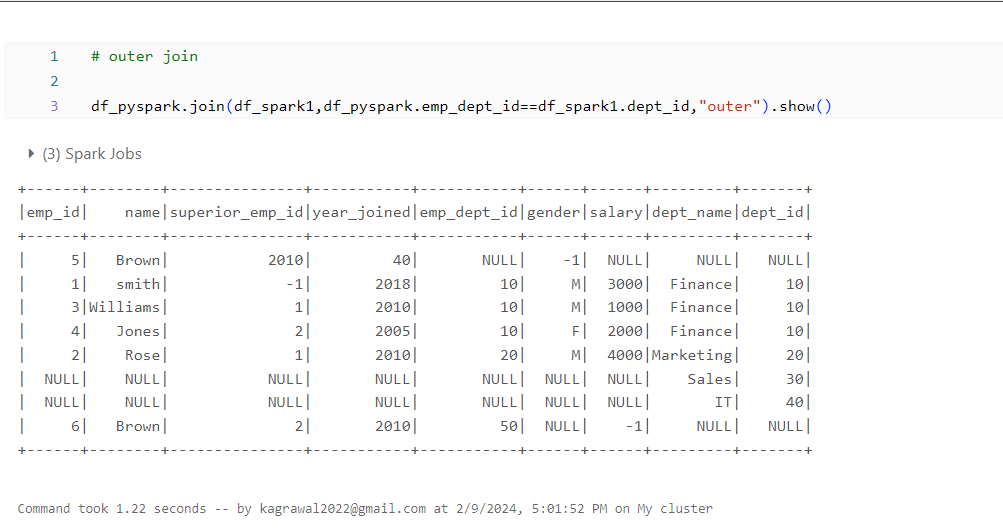
****

****

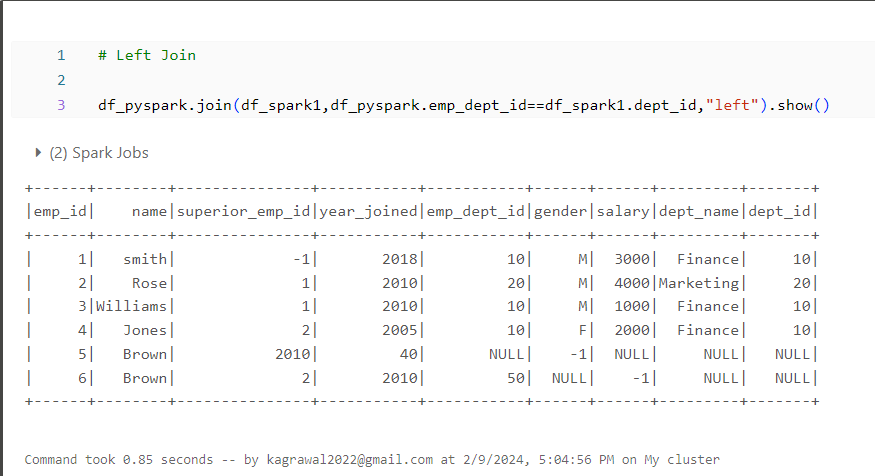
1. **Inner Join**

****

1. **Outer join**

****

1. **Left Join**

****

1. **Right Join**

****