Converting pandas dataframe to pyspark dataframe:

A screenshot of a computer program

Description automatically generated



Creating tables :

To create tables from dataframe we should use

Dataframe.createOrReplace TempView(“table”)

Then perform sql operations on it.

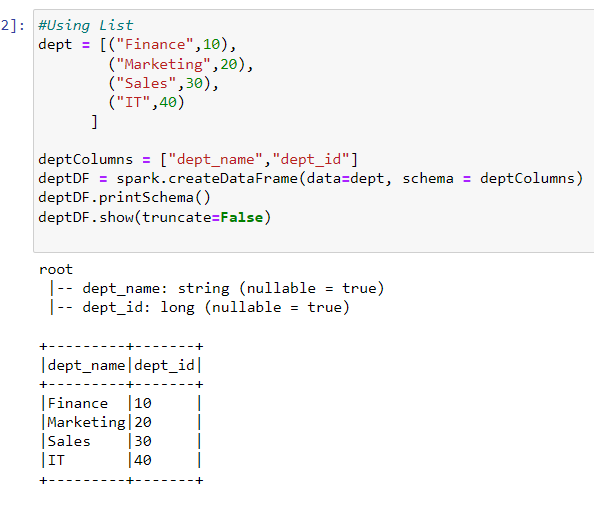
A screenshot of a computer

Description automatically generated

Aggregate:



Converting lists to dataframes and rdd:

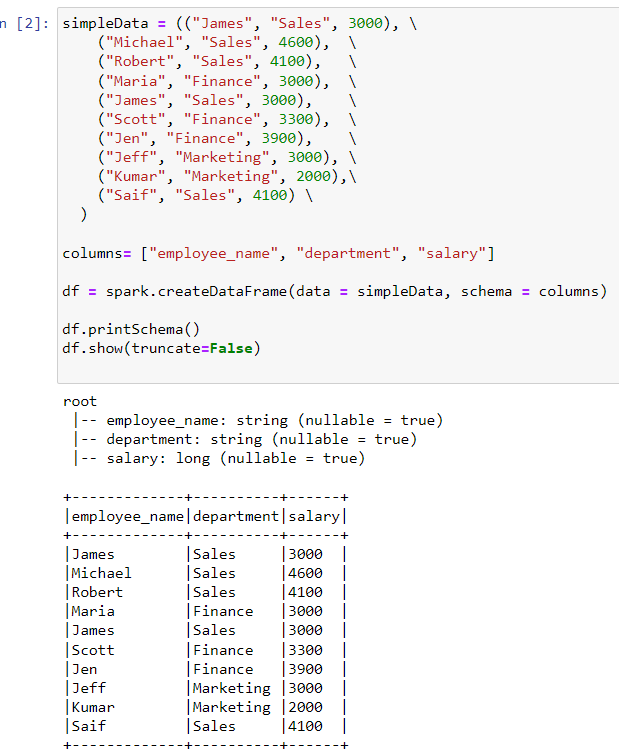


A screenshot of a computer program

Description automatically generated

Window functions:

Original Data:



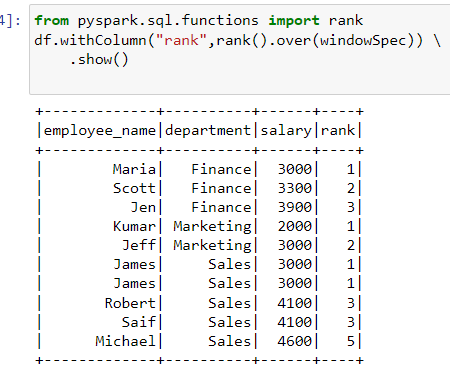
After performing using rownumber partitions ;

It arranges the data department wise orderby salary and gives rownumber to each department people.

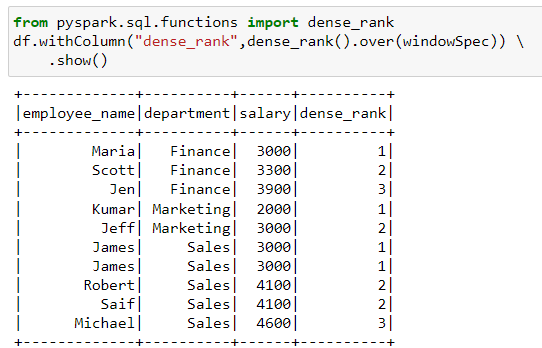
A screenshot of a computer code

Description automatically generated

Rank:



Denserank:



The **LAG()** window function facilitates access to previous rows based on the offset argument. It can be particularly useful when a comparison of a previous value is necessary without the use of a self join. There is a similarity to the LEAD() function with the difference being the accessible rows. LEAD() accesses subsequent rows while LAG() accesses previous rows

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer code

Description automatically generated

When otherwise:

