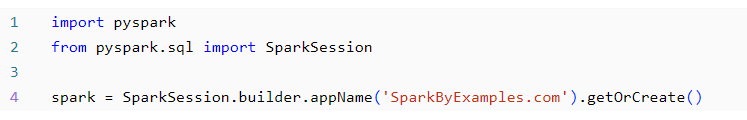
**Unions;**

PySpark union() and unionAll() transformations are used to merge two or more DataFrame’s of the same schema or structure.

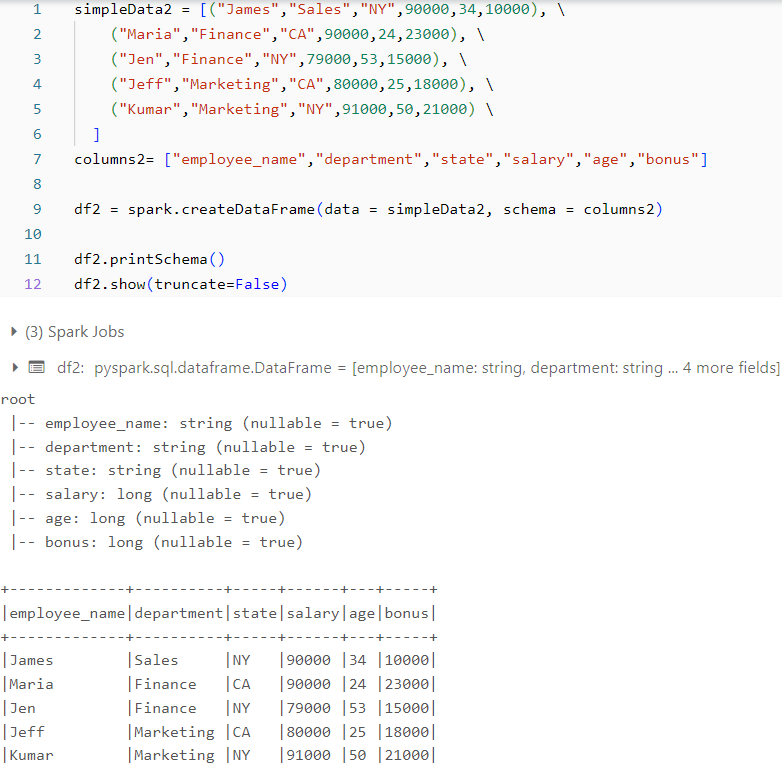
Dataframe union() – union() method of the DataFrame is used to merge two DataFrame’s of the same structure/schema.

****

**First data frame:**

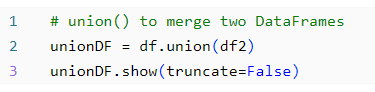
****

**Second data frame:**

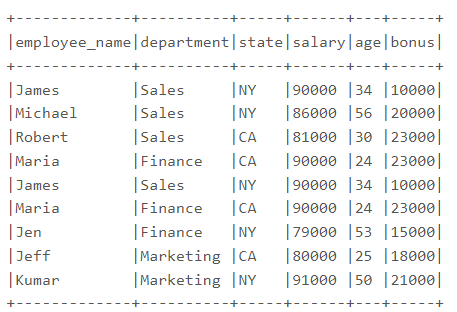
****

**Union():**

DataFrame union() method merges two DataFrames and returns the new DataFrame with all rows from two Dataframes regardless of duplicate data.

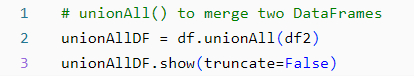
****

**Output:**

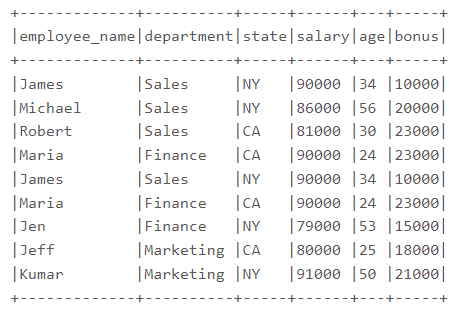
****

**Union all():**

DataFrame unionAll() method is deprecated since PySpark “2.0.0” version and recommends using the union() method.

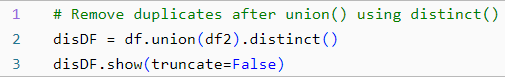
****

**Output:**

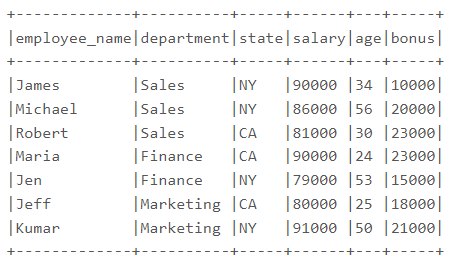
****

**Removing duplicates by using distinct() along with union():**

Since the union() method returns all rows without distinct records, we will use the distinct() function to return just one record when a duplicate exists.

****

**Output:**

****