Join in R: How to join (merge) data frames (inner, outer, left, right) in R

We can merge two data frames in R by using the **merge()** function. The data frames must have same column names on which the merging happens. Merge() Function in R is similar to database join operation in SQL. The different arguments to merge() allow you to perform natural joins, as well as left, right, and full outer joins. We can perform Join in R using R **merge()** Function

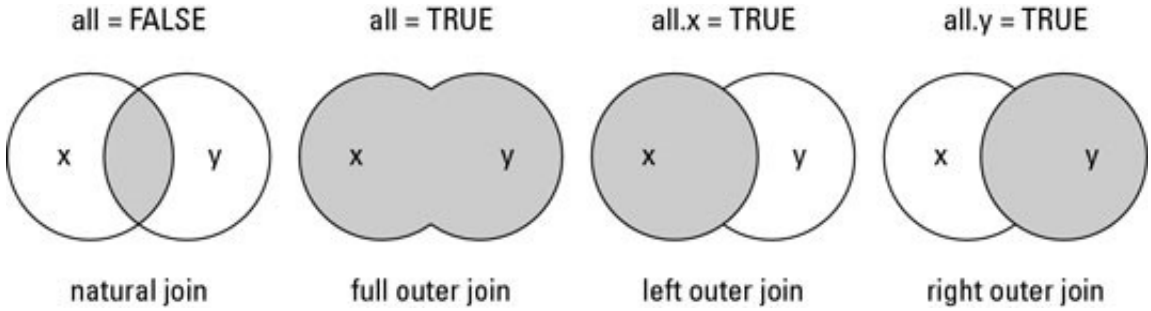
Arguments of merge() function in R are

- **x:**data frame1.
- **y:**data frame2.
- **by,x, by.y:** The names of the columns that are common to both x and y. The default is to use the columns with common names between the two data frames.
- **all, all.x, all.y:**Logical values that specify the type of merge. The default value is all=FALSE (meaning that only the matching rows are returned).

UNDERSTANDING THE DIFFERENT TYPES OF MERGE IN R:

- **Natural join**: To keep only rows that match from the data frames, specify the argument all=FALSE.
- **Full outer join:**To keep all rows from both data frames, specify all=TRUE.
- **Left outer join:**To include all the rows of your data frame x and only those from y that match, specify x=TRUE.

- **Right outer join:**To include all the rows of your data frame y and only those from x that match, specify y=TRUE.



Lets look at with some examples

```
1
2  df1 = data.frame(CustomerId = c(1:6), Product = c(rep("Oven", 3),
   rep("Television", 3)))
3  df1
4  df2 = data.frame(CustomerId = c(2, 4, 6), State = c(rep("California",
   2), rep("Texas", 1)))
5  df2
6
```

so we will get following two data frames

df1:

CustomerId Product
1 1 Oven
2 2 Oven
3 3 Oven
4 4 Television
5 5 Television
6 6 Television

df2:

CustomerId	State
2	California
4	California
6	Texas

Inner join in R: Return only the rows in which the left table have matching keys in the right table.

```
df<-merge(x=df1,y=df2,by="CustomerId")
```

the resultant data frame df will be

CustomerId	Product	State
2	Oven	California
4	Television	California
6	Television	Texas

Outer join in R: Returns all rows from both tables, join records from the left which have matching keys in the right table.

```
df<-merge(x=df1,y=df2,by="CustomerId",all=TRUE)
```

the resultant data frame df will be

	CustomerId	Product	State
1	1	Oven	<NA>
2	2	Oven	California
3	3	Oven	<NA>
4	4	Television	California
5	5	Television	<NA>

6 6 Television Texas

Left outer join in R: Return all rows from the left table, and any rows with matching keys from the right table.

```
1 df<-merge(x=df1,y=df2,by="CustomerId",all.x=TRUE)
```

the resultant data frame df will be

CustomerId Product State
1 1 Oven <NA>
2 2 Oven California
3 3 Oven <NA>
4 4 Television California
5 5 Television <NA>
6 6 Television Texas

Right outer join in R: Return all rows from the right table, and any rows with matching keys from the left table.

```
1 df<-merge(x=df1,y=df2,by="CustomerId",all.y=TRUE)
```

the resultant data frame df will be

CustomerId Product State
1 2 Oven California
2 4 Television California
3 6 Television Texas

Cross join in R: A Cross Join (also sometimes known as a Cartesian Join) results in every row of one table being joined to every row of another table

```
df<-merge(x = df1, y = df2, by = NULL)
```

the resultant data frame df will be

CustomerId.x	Product	CustomerId.y	State
1 1	Oven	2	California
2 2	Oven	2	California
3 3	Oven	2	California
4 4	Television	2	California
5 5	Television	2	California
6 6	Television	2	California
7 1	Oven	4	California
8 2	Oven	4	California
9 3	Oven	4	California
10 4	Television	4	California
11 5	Television	4	California
12 6	Television	4	California
13 1	Oven	6	Texas
14 2	Oven	6	Texas
15 3	Oven	6	Texas
16 4	Television	6	Texas
17 5	Television	6	Texas
18 6	Television	6	Texas