SESSION 4: FOUNDATIONAL R PROGRAMMING-II

Assignment 1

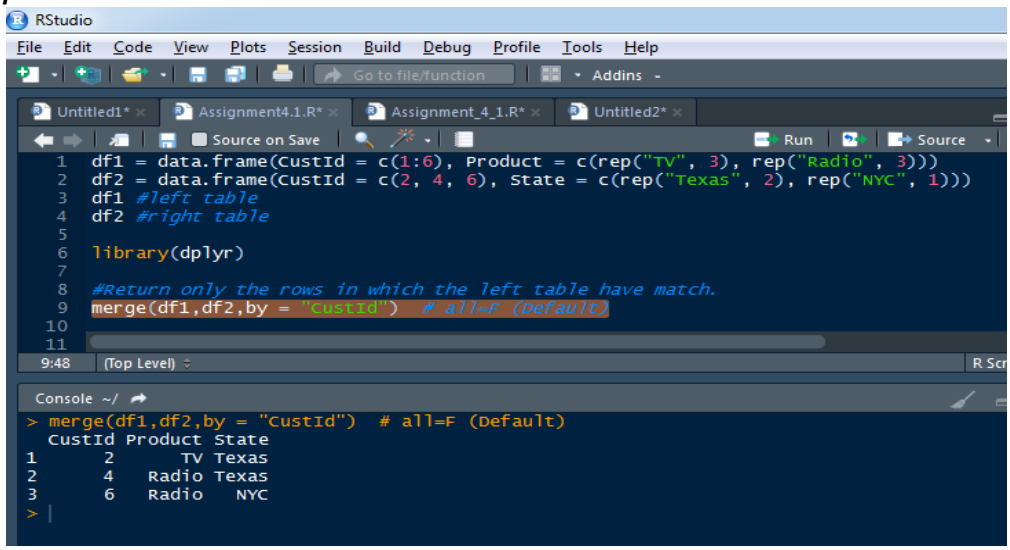
1. df1 = data.frame(CustId = c(1:6), Product = c(rep("TV", 3), rep("Radio", 3)))
   df2 = data.frame(CustId = c(2, 4, 6), State = c(rep("Texas", 2), rep("NYC", 1)))
   df1 #left table
   df2 #right table

   For the above given data frames and tables perform the following operations:
   • Return only the rows in which the left table have match.

Answer:

merge(df1,df2,by = "CustId") *#all=F (Default)*

Output:

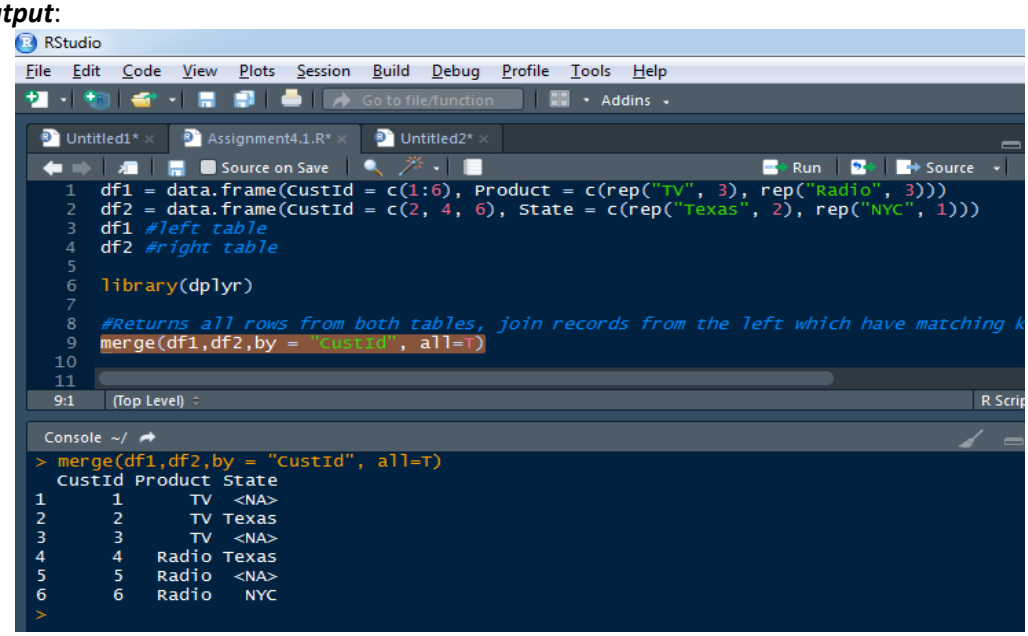


• Returns all rows from both tables, join records from the left which have matching keys in the right table.

Answer:

merge(df1,df2,by = "CustId", all=T)

Output:

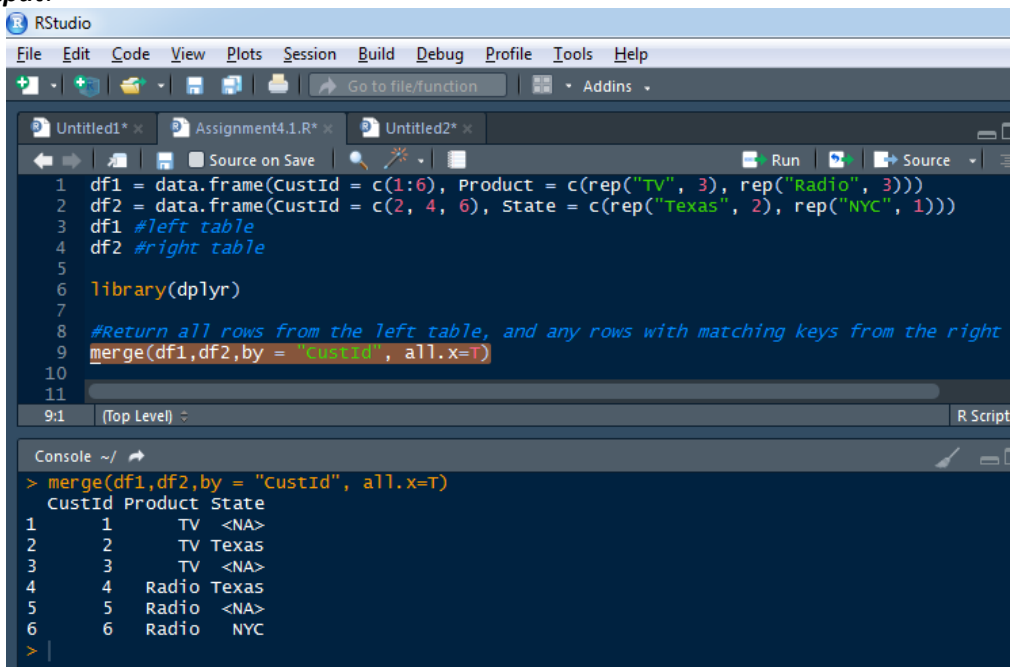


• Return all rows from the left table, and any rows with matching keys from the right table.

Answer:

merge(df1,df2,by = "CustId", all.x=T)

Output:

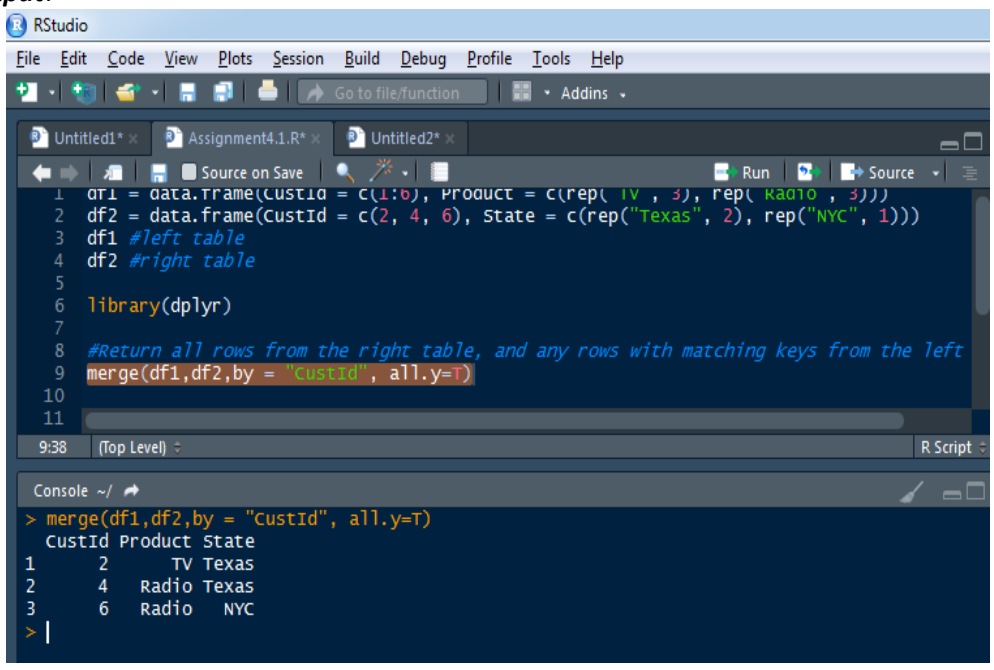


• Return all rows from the right table, and any rows with matching keys from the left table.

Answer:

merge(df1,df2,by = "CustId", all.y=T)

Output:

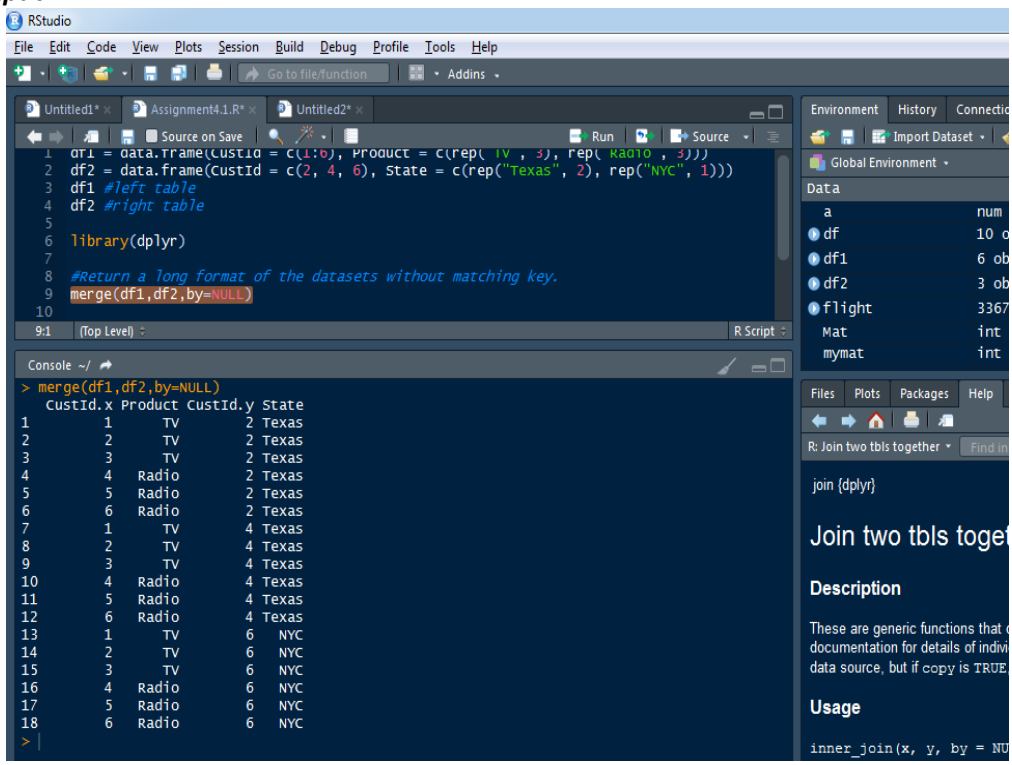


2. Perform the below operations on above given data frames and tables:
   - Return a long format of the datasets without matching key.

Answer:

merge(df1,df2, by=NULL)

Output:

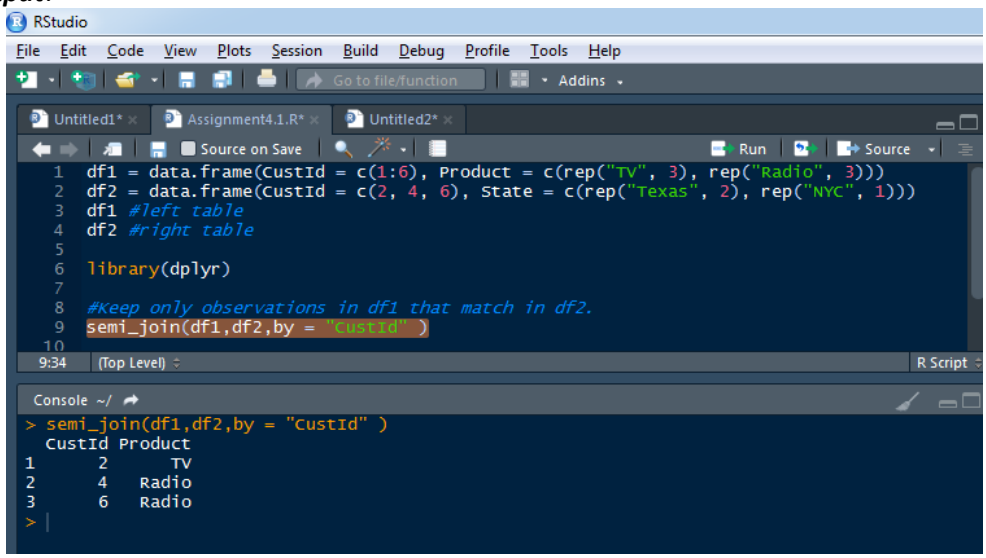


   - Keep only observations in df1 that match in df2.

Answer:

semi_join(df1,df2,by = "CustId")

Output:



- Drop all observations in df1 that match in df2.

Answer:

anti_join(df1,df2,by = "CustId")

Output: