ACADGILD

Session 3: Foundational R Programming-II

Assignment 4.1

1. Problem Statement

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.
- Returns all rows from both tables, join records from the left which have matching keys in the right table.
- Return all rows from the left table, and any rows with matching keys from the right table.
- Return all rows from the right table, and any rows with matching keys from the left table.

2. Solution

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

```
> df1 #left table
  CustId Product
1      1      TV
2      2      TV
3      3      TV
4      4   Radio
5      5   Radio
6      6   Radio
```

df2 #right table

```
> df2 #right table
  CustId State
1      2 Texas
2      4 Texas
3      6   NYC
```

Part 1: For the above given data frames and tables perform the following operations:

I - Return only the rows in which the left table have match.

```
> merge(df1,df2, by="CustId")

 CustId Product State
1      2      TV Texas
2      4   Radio Texas
3      6   Radio   NYC
```

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

```
> merge(df1,df2,by="CustId", all=TRUE)

  CustId Product State
1      1      TV  <NA>
2      2      TV Texas
3      3      TV  <NA>
4      4   Radio Texas
5      5   Radio  <NA>
6      6   Radio   NYC
```

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

```
> merge(df1,df2,by="CustId", all.x = TRUE)
```

```
  CustId Product State
1      1      TV  <NA>
2      2      TV Texas
3      3      TV  <NA>
4      4   Radio Texas
5      5   Radio  <NA>
6      6   Radio   NYC
```

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

```
> df3<- merge(df1,df2,by="CustId", all.y = TRUE)
> df3
  CustId Product State
1      2      TV Texas
2      4   Radio Texas
3      6   Radio   NYC
```

Part 2: Perform the below operations on above given data frames and tables:

I - Return a long format of the datasets without matching key.

```
> df1$CustId%in%df2$CustId
[1] FALSE  TRUE FALSE  TRUE FALSE  TRUE
> select_index=which(df1$CustId%in%df2$CustId)
> df1[-select_index]
  CustId
1      1
2      2
3      3
4      4
5      5
6      6
> df1[-select_index,]
  CustId Product
1      1      TV
3      3      TV
5      5   Radio
```