## **SESSION 4 - ASSIGNMENT 4.1**

Date: 7th Jan 2019

```
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.

```
#use of merge function for inner join
dfa<-merge(x=df1,y=df2,by="CustId")
dfa</pre>
```

```
Console Terminal ×
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
   CustId Product
                    TV
               Radio
               Radio
               Radio
> df2 #right table
  CustId State
2 Texas
          4 Texas
              NYC
> dfa<-merge(x=df1,y=df2,by="CustId")
  CustId Product State

2 TV Texas
                Radio Texas
3
          6 Radio NYC
```

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

#use of merge function for outer join

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

df

```
> #use of merge function
> df<-merge(x=df1,y=df2,by="CustId",all=TRUE)</pre>
> df
 CustId Product State
      1
            TV <NA>
             TV Texas
2
      2
3
       3
             TV <NA>
      3 TV <NA>
4 Radio Texas
5 Radio <NA>
4
5
     6 Radio NYC
6
> |
```

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

```
#left join
```

```
library(dplyr)
dfb<-left_join(df1,df2)
dfb
```

```
Console Terminal ×
> 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
   CustId Product
                       TV
           4 Radio
                 Radio
                 Radio
 > df2 #right table
   CustId State
         2 Texas
4 Texas
2 6 NYC
> #left outer join
> library(dplyr)
> dfb<-left_join(df1,df2)
Joining, by = "CustId"
> dfb
> dfb
   CustId Product State
               TV <NA>
TV Texas
                       TV <NA>
           4 Radio Texas
           5 Radio <NA>
6 Radio NYC
6
```

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

## #left join

```
library(dplyr)
dfb<-right_join(df1,df2)
```

dfb

```
Console Terminal ×
 ~/ @
> 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
  CustId Product
          1
                     TV
           2
                     TV
3
           3
                     TV
           4 Radio
4
5
          5 Radio
6 Radio
6
> df2 #right table
  CustId State
2 Texas
           4 Texas
2 4 Texas

3 6 NYC

> #right join

> dfc<-right_join(df1,df2)

Joining, by = "CustId"
> dfc
  CustId Product State
   2 TV Texas
4 Radio Texas
6 Radio NYC
3
```

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

```
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
dfg<-merge(x=df1,y=df2,all=TRUE)
dfg</pre>
```

```
Terminal ×
Console
~/ @
> dfg<-merge(x=df1,y=df2,all=TRUE)</pre>
> dfg
  CustId Product State
1
       1
               TV <NA>
2
       2
               TV Texas
3
       3
               TV <NA>
4
       4 Radio Texas
5
       5
           Radio <NA>
6
            Radio
       6
                    NYC
> |
```

• Keep only observations in df1 that match in df2.

```
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
semi_join(df1, df2)
```

```
Console Terminal ×
> 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
   CustId Product
1
                   T۷
                   T۷
3
                    TV
               Radio
               Radio
               Radio
> df2 #right table
   CustId State
          2 Texas
          4 Texas
          6 NYC
> semi_join(df1, df2)
Joining, by = "CustId"
   CustId Product
               Radio
3
               Radio
```

Drop all observations in df1 that match in df2.

```
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
anti_join(df1, df2)
```

```
Console
        Terminal ×
~/@
> 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
  CustId Product
1
       1
              TV
              ΤV
2
3
              TV
       3
4
       4
           Radio
5
           Radio
6
       6
           Radio
> df2 #right table
  CustId State
1
       2 Texas
       4 Texas
3
          NYC
> anti_join(df1,df2)
Joining, by = "CustId"
  CustId Product
              TV
       1
2
       3
              TV
3
          Radio
> |
```