ACADGILD

SESSION 4: FOUNDATIONAL R PROGRAMMING-II

Assignment 1

Table of Contents

1. Introduction 3
2. Objective 3
3. Prerequisites 3
4. Associated Data Files 3
5. Problem Statement 3
6. Expected Output 3
7. Approximate Time to Complete Task 3

1. Introduction

This assignment will help you understand the concepts learnt in the session.

2. Objective

This assignment will test your skills on foundational R Programming- writing functions.

3. Prerequisites

Not applicable.

4. Associated Data Files

Not applicable.

5. Problem Statement

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.
- 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.

ANS: 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

```
> 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
  CustId State
1      2 Texas
2      4 Texas
3      6   NYC
```

#Return only the rows in which the left table have match.
df=merge(df1, df2, by = "CustId")
df

```
> df
  CustId Product State
1      2      TV Texas
2      4   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.
df0=merge(df1, df2, by = "CustId", all = T)
df0

```
> df0
  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
>
```

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

```
df3=merge(df1, df2, all.x = T)
df3
```

```
> df3
  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
>
```

#Return all rows from the right table, and any rows with matching keys from the lefttable.

```
df4=merge(df1, df2, all.y  = T)
df4
```

```
> df4
  CustId Product State
1      2      TV Texas
2      4   Radio Texas
3      6   Radio   NYC
>
```

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

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

```
Df=merge(df1,df2, all = T)
Df
```

```
> Df
  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
>
```

- **Keep only observations in df1 that match in df2.**

```
library(dplyr)
Df1=semi_join(df1,df2)
Df1
```

```
> Df1=semi_join(df1,df2)
Joining, by = "CustId"
> Df1
  CustId Product
1      2      TV
2      4   Radio
3      6   Radio
>
```

- **Drop all observations in df1 that match in df2.**

```
Df2=anti_join(df1,df2)
Df2
```

```
> Df2=anti_join(df1,df2)
Joining, by = "CustId"
> Df2
  CustId Product
1      1      TV
2      3      TV
3      5   Radio
>
```