

Sheth L.U.J & Sir M.V College Of Science
Subject :- Data Analysis with SAS / SPSS / R
Practical No 6

Aim :- Combining and appending datasets using merge() or bind_rows() in R.

```
Console Terminal Background Jobs
R - R4.5.2 - ~/R/
> print("--- Fraud Dataset ---")
[1] "--- Fraud Dataset ---"
> print(head(fraud))
# A tibble: 6 x 10
  transaction_id user_id amount transaction_type merchant_category country hour
  <dbl> <dbl> <dbl> <chr> <chr> <chr> <dbl>
1 2608 363 4923. ATM Travel TR 12
2 456 692 48.0 QR Food US 21
3 4747 587 137. Online Travel TR 14
4 6934 445 80.5 POS Clothing TR 23
5 1646 729 120. Online Grocery FR 16
6 2183 944 97.1 POS Clothing DE 17
# i 3 more variables: device_risk_score <dbl>, ip_risk_score <dbl>,
# is_fraud <dbl>
> fraud_A <- fraud[1:5000, ]
> fraud_B <- fraud[5001:10000, ]
> print("--- Fraud_A (First Part) ---")
[1] "--- Fraud_A (First Part) ---"
> print(head(fraud_A))
# A tibble: 6 x 10
  transaction_id user_id amount transaction_type merchant_category country hour
  <dbl> <dbl> <dbl> <chr> <chr> <chr> <dbl>
1 2608 363 4923. ATM Travel TR 12
2 456 692 48.0 QR Food US 21
3 4747 587 137. Online Travel TR 14
4 6934 445 80.5 POS Clothing TR 23
5 1646 729 120. Online Grocery FR 16
6 2183 944 97.1 POS Clothing DE 17
# i 3 more variables: device_risk_score <dbl>, ip_risk_score <dbl>,
# is_fraud <dbl>
> print("--- Fraud_B (Second Part) ---")
[1] "--- Fraud_B (Second Part) ---"
> print(head(fraud_B))
# A tibble: 6 x 10
  transaction_id user_id amount transaction_type merchant_category country hour
  <dbl> <dbl> <dbl> <chr> <chr> <chr> <dbl>
1 8092 593 79.4 POS Electronics US 15
2 8151 496 89.8 Online Food DE 23
3 474 158 62.3 POS Grocery DE 16
4 8876 616 40.4 ATM Electronics FR 13
5 2870 252 111. Online Food US 18
6 5591 209 133. QR Grocery UK 12
# i 3 more variables: device_risk_score <dbl>, ip_risk_score <dbl>,
# is_fraud <dbl>
> merged_data <- merge(
+   fraud_A,
+   fraud_B,
+   by = "transaction_id",
+   suffixes = c("_A", "_B")
+ )
> print("--- Merged Data (Columns Added) ---")
[1] "--- Merged Data (Columns Added) ---"
> print(head(merged_data))
[1] transaction_id user_id_A amount_A
[4] transaction_type_A merchant_category_A country_A
[7] hour_A device_risk_score_A ip_risk_score_A
[10] is_fraud_A user_id_B amount_B
[13] transaction_type_B merchant_category_B country_B
[16] hour_B device_risk_score_B ip_risk_score_B
[19] is_fraud_B
<0 rows> (or 0-length row.names)
> fraud_A$Group <- "A"
> fraud_B$Group <- "B"
> combined_data <- bind_rows(fraud_A, fraud_B)

Error in bind_rows(fraud_A, fraud_B) :
  could not find function "bind_rows"

> print("--- Combined Data (Rows Added from A & B) ---")
[1] "--- Combined Data (Rows Added from A & B) ---"
> print(head(combined_data))

Error: object 'combined_data' not found

> View(fraud_A)
> View(fraud_B)
> |
```

Name :- Vinod Mali
Roll No :- S092