

SHETH L.U.J AND SIR M.V. COLLEGE

SUBJECT :- DATA ANALYSIS WITH SAS/SPSS/R

PRACTICAL – 7

AIM:- Selecting and dropping variables using select() in R. import dataset.

OUTPUT:-

```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins
Source
Console Terminal Background Jobs
R - R452 - ~/
Run "clang::last_trace()" to see where the error occurred.
> library(dplyr)
> #1. Import the csv file
> housing <- read.csv("C:\\Users\\info\\desktop\\shopping_behavior_updated.csv")
> print("--- Original Dataset (first 3 rows) ---")
[1] "--- Original Dataset (first 3 rows) ---"
> print(head(shopping, 3))
  Customer.ID Age Gender Item.Purchased Category Purchase.Amount .USD Location Size Color Season Review.Rating
1      1  55   Male   blouse Clothing           53      Kentucky L Gray winter      3.1
2      2  19   Male   Sweater Clothing           64      Maine L Maroon winter      3.1
3      3  50   Male   Jeans Clothing           73 Massachusetts S Maroon Spring      3.1
Subscription.Status Shipping.Type Discount.Applied Promo.Code.Used Previous.Purchases Payment.Method Frequency.of.Purchases
1      Yes Express Yes Yes 14 Verno Fortnightly Cash
2      Yes Express Yes Yes 2 Cash
3      Yes Free Shipping Yes Yes 23 Credit Card weekly
> #2. SELECTING VARIABLES
> # Method A: Select specific columns by name
> selected_cols <- shopping %>%
+   select(Age, Gender, category)
>
> print("--- Selected specific columns ---")
[1] "--- Selected specific columns ---"
> print(head(selected_cols, 3))
  Age Gender category
1  55   Male Clothing
2  19   Male Clothing
3  50   Male Clothing
> # Method B: Select a range of adjacent columns
> range_cols <- housing %>%
+   select(shipping.Type:Size)
>
> print("--- Selected Range of Columns ---")
[1] "--- Selected Range of columns ---"
> print(head(range_cols, 3))
  shipping.Type Subscription.Status Review.Rating Season Color Size
1      Express Yes 3.1 winter Gray L
2      Express Yes 3.1 winter Maroon L
3 Free shipping Yes 3.1 Spring Maroon S
> # Method C: Select using helper functions
> starts_with_r <- shopping %>%
+   select(starts_with("r"))
>
> print("--- Selected columns starting with 'r' ---")
[1] "--- Selected columns starting with 'r' ---"
> print(head(starts_with_r, 3))
  Review.Rating
1      3.1
2      3.1
3      3.1
> #3. DROPPING VARIABLES
> # Method A: Drop a single specific column
> dropped_one <- shopping %>%
+   select(-Location)
>
> print("--- Dataset with 'Location' dropped ---")
[1] "--- Dataset with 'Location' dropped ---"
> print(names(dropped_one))
[1] "Customer.ID" "Age" "Gender" "Item.Purchased" "Category"
[6] "Purchase.Amount..USD." "Size" "Color" "Season" "Review.Rating"
[11] "Subscription.Status" "Shipping.Type" "Discount.Applied" "Promo.Code.Used" "Previous.Purchases"
[16] "Payment.Method" "Frequency.of.Purchases"
> # Method B: Drop multiple columns
> dropped_multiple <- shopping %>%
+   select(-color, -season)
>
> print("--- Dataset with 'color' and 'season' dropped ---")
[1] "--- Dataset with 'color' and 'season' dropped ---"
> print(names(dropped_multiple))
[1] "Customer.ID" "Age" "Gender" "Item.Purchased" "Category"
[6] "Purchase.Amount..USD." "Size" "Location" "Review.Rating" "Subscription.Status"
[11] "Shipping.Type" "Discount.Applied" "Promo.Code.Used" "Previous.Purchases" "Payment.Method"
[16] "Frequency.of.Purchases"
```

```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins
Source
Console Terminal Background Jobs
R - R452 - ~/
[1] --- Selected Range or columns ---
> print(head(range_cols, 3))
  shipping.Type Subscription.Status Review.Rating Season Color Size
1      Express Yes 3.1 winter Gray L
2      Express Yes 3.1 winter Maroon L
3 Free shipping Yes 3.1 Spring Maroon S
> # Method C: Select using helper functions
> starts_with_r <- shopping %>%
+   select(starts_with("r"))
>
> print("--- Selected columns starting with 'r' ---")
[1] --- Selected columns starting with 'r' ---
> print(head(starts_with_r, 3))
  Review.Rating
1      3.1
2      3.1
3      3.1
> #3. DROPPING VARIABLES
> # Method A: Drop a single specific column
> dropped_one <- shopping %>%
+   select(-Location)
>
> print("--- Dataset with 'Location' dropped ---")
[1] --- Dataset with 'Location' dropped ---
> print(names(dropped_one))
[1] "Customer.ID" "Age" "Gender" "Item.Purchased" "Category"
[6] "Purchase.Amount..USD." "Size" "Color" "Season" "Review.Rating"
[11] "Subscription.Status" "Shipping.Type" "Discount.Applied" "Promo.Code.Used" "Previous.Purchases"
[16] "Payment.Method" "Frequency.of.Purchases"
> # Method B: Drop multiple columns
> dropped_multiple <- shopping %>%
+   select(-color, -season)
>
> print("--- Dataset with 'color' and 'season' dropped ---")
[1] --- Dataset with 'color' and 'season' dropped ---
> print(names(dropped_multiple))
[1] "Customer.ID" "Age" "Gender" "Item.Purchased" "Category"
[6] "Purchase.Amount..USD." "Size" "Location" "Review.Rating" "Subscription.Status"
[11] "Shipping.Type" "Discount.Applied" "Promo.Code.Used" "Previous.Purchases" "Payment.Method"
[16] "Frequency.of.Purchases"
```

SUBJECT :- DATA ANALYSIS WITH SAS/SPSS/R

