

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:-

The screenshot shows an RStudio interface with the following details:

- Console:** Displays an R session with code and data frames. The code includes importing a CSV file, creating a shopping behavior dataset, and printing specific columns. It also demonstrates selecting variables using methods A and B.
- Environment:** Shows the global environment with various objects like `final_list`, `housing`, `mental_health_`, etc.
- Files:** Shows a file browser with several files and folders, including `STUDENTS DETAILS.accdb`, `.RData`, `Rhistory`, `ADPS PRAC3.pdf`, `BOOK STORE.accdb`, `cleaned_house_prices.csv`, `Custom Office Templates`, `Database1.accdb`, `Database2.accdb`, `desktop.ini`, `excel data`, `GIS DataBase`, `house_prices_Texas_countries`, `iSExpress`, `My Music`, `My Pictures`, and `My Videos`.

The screenshot shows an RStudio interface with a large R script in the Source pane. The script performs several operations on a dataset, including selecting columns, dropping variables, and dropping multiple columns. It also prints the names of the datasets and variables at various points. The Environment pane on the right lists numerous global variables and their sizes. The bottom status bar shows the current session details.

```
R - R4.5.2 - ~/r
[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"
[5] "Purchase.Amount..USD." "Size"         "Color"         "Season"        "Review.Rating"
[9] "Subscription.Status" "Shipping.Type" "Discount.Applied" "Promo.Code.Used" "Previous.Purchases"
[13] "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"
[5] "Purchase.Amount..USD." "Location"    "Size"          "Review.Rating"  "Subscription.Status"
[9] "Shipping.Type"    "Discount.Applied" "Promo.Code.Used" "Previous.Purchases" "Payment.Method"
[13] "Frequency.of.Purchases"
```

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

SUBJECT :- DATA ANALYSIS WITH SAS/SPSS/R



RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Source Environment History Connections Tutorial
Project: (None)
Console Terminal Background Jobs
R 4.5.2 - ~/
> 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"
[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" "Location" "Size" "Review.Rating" "Subscription.Status"
[11] "Shipping.Type" "Discount.Applied" "Promo.Code.Used" "Previous.Purchases"
[16] "Frequency.of.Purchases"
>
> # Method C: Drop a range of columns
> dropped_range <- shopping %>%
> select(-(Payment.Method:Discount.Applied))
>
> print("... dataset with range 'Payment' to 'Discount' dropped ...")
[1] "... dataset with range 'Payment' to 'Discount' dropped ..."
> print(names(dropped_range))
[1] "Customer.ID" "Age" "Gender" "Item.Purchased" "Category"
[6] "Purchase.Amount.USD" "Location" "Size" "Color" "Season"
[11] "Review.Rating" "Subscription.Status" "Shipping.Type" "Frequency.of.Purchases"
[16] "Payment.Method" "Discount.Applied"
>
|