

MVLU COLLEGE

PRACTICAL NO. 8

AIM: Applying basic data cleaning functions: handling missing values using na.omit()/replace_na() in R. import dataset.

The screenshot shows the RStudio interface with the following details:

- File Menu:** File, Edit, Code, View, Plot, Session, Build, Debug, Profile, Tools, Help.
- Source Tab:** Displays R code for data cleaning and analysis. The code reads a CSV file, installs tidyverse packages, and performs various data manipulation steps like filtering, replacing missing values, and calculating averages.
- Global Environment Tab:** Shows the current state of the R environment with 298 objects. Key objects include:
 - retail_df: 298 obs. of 17 variables
 - clean_omit: 540 obs. of 5 variables
 - clean_replace: 4362 obs. of 5 variables
 - Cleaned_BMW_Sales: 99 obs. of 11 variables
 - Cleaned_Car_Price: 199 obs. of 10 variables
 - df1: 99 obs. of 11 variables
 - df2: 199 obs. of 10 variables
 - dropped_multiple: 4573 obs. of 13 variables
 - dropped_one: 4573 obs. of 14 variables
 - dropped_range: 4573 obs. of 11 variables
 - housing: 4573 obs. of 15 variables
 - merged_full: 294 obs. of 17 variables
 - merged_inner: 4 obs. of 17 variables
 - merged_left: 99 obs. of 17 variables
 - range_cols: 4573 obs. of 6 variables
 - retail_df: 4362 obs. of 5 variables
 - Retail_product_re: 4362 obs. of 5 variables
 - selected_cols: 4573 obs. of 3 variables
 - spotify: 4573 obs. of 15 variables
 - spotify_data_clean: 8573 obs. of 15 variables
 - starts_with_track: 4573 obs. of 5 variables
- Console Tab:** Shows the command history and output.
- Bottom Bar:** Includes tabs for Files, Plots, Packages, Help, Viewer, and Presentation, along with system status icons.

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help

Source

```
R> library(readr)
> Retail_Product_Retail_Product <- read_csv("S105/Retail Product - Retail Product.csv")
Rows: 4362 Columns: 5
  _____
  - Column specification -
  _____
  delimeter: ","
chr (2): Category, stock
dbl (3): Price, Rating, Discount

i use 'spec()' to retrieve the full column specification for this data.
i specify the column types or set 'show_col_types = FALSE' to quiet this message.
> view(Retail_Product_Retail_Product)
> install.packages("tidyverse")
WARNING: Rtools is required to build R packages but is not currently installed. Please download and install the appropriate version of Rtools before proceeding.

https://cran.rstudio.com/bin/windows/Rtools/
Installing package into 'C:/users/itlab/AppData/Local/R/win-library/4.5'
(as 'lib' is unspecified)

also installing the dependencies 'stringi', 'purrr', 'stringr'

trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.5/stringi_1.8.7.zip'
trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.5/purrr_1.0.0.zip'
trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.5/stringr_1.6.0.zip'
trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.5/tidyr_1.3.1.zip'
package 'stringi' successfully unpacked and MD5 sums checked
package 'purrr' successfully unpacked and MD5 sums checked
package 'stringr' successfully unpacked and MD5 sums checked
package 'tidy' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
  C:/users/itlab/AppData/Local/Temp/RtmpQVEM1/downloaded_packages
> library(dplyr)
> library(tidy)
> retail_df <- read.csv("Retail Product.csv", na.strings = c("", "NA"))

Error in file(file, "rt") : cannot open the connection

In addition: warning message:
In file(file, "rt") :
  No such file or directory
```

Environment History Connections Tutorial Project: (None)

Console Background Jobs

Source

Files Plots Packages Help Viewer Presentation

ENG IN 01-12-2024

MVLU COLLEGE

PRACTICAL NO. 8

The screenshot shows the RStudio interface with the following details:

- Console:** Displays R code and its output. The user is trying to read a CSV file named "Retail Product.csv" but receives an error message: "Error in file(file, "rt") : cannot open the connection".
- Data View:** Shows a list of objects in the global environment, including various data frames like "retail_df", "cleaned_BMW_Sales", and "df1", along with their dimensions and variable counts.
- Environment View:** Shows the current project structure, which is currently empty ("Project: (None)").
- File Bar:** Includes options for Files, Plots, Packages, Help, Viewer, and Presentation.
- System Tray:** Shows system status icons and the date/time: 01-12-2025 at 12:35.

MVLU COLLEGE

PRACTICAL NO. 8



RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Source

```
R > 8.452 - ~ /
```

```
[1] "Rows remaining: 540"
> Print(head(clean omit))
Category Price Rating Stock Discount
7 C 667 3.668341 In Stock 41
8 A 7125 4.983998 Out of Stock 7
9 A 2777 2.678384 In Stock 6
12 A 3772 4.890750 In Stock 45
16 A 7932 3.632832 In Stock 44
17 B 9319 3.479064 In Stock 28
> avg_price <- mean(retail_df$Price, na.rm = TRUE)
>
> clean_replace <- retail_df %>%
+   replace_na(list(
+     Category = "Unknown",
+     Discount = 0,
+     Stock = "Check warehouse",
+     Price = avg_price
+   ))
Error in `vec_assign()`:
! Can't convert from `replace$Price` <double> to `data$Price` <integer> due to loss of precision.
• Locations: 1
Run rlang::last_trace() to see where the error occurred.

> retail_df$Price <- as.numeric(retail_df$Price)
> avg_price <- mean(retail_df$Price, na.rm = TRUE)
> Clean_replace <- retail_df %>%
+   replace_na(list(
+     Category = "Unknown",
+     Discount = 0,
+     Stock = "Check warehouse",
+     Price = avg_price
+   ))
>
> cat("\n--- 3. Data after replace_na() ---\n")
--- 3. Data after replace_na() ---
>
> print(clean_replace[3, ])
Category Price Rating Stock Discount
3 Unknown 4004 NA In Stock 0
> print(head(clean_replace))
```

Air: Moderate Now

Search DELL R 12:35 01-12-2025

Files Plots Packages Help Viewer Presentation



RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Source

```
R > 8.452 - ~ /
```

```
> retail_df$Price <- as.numeric(retail_df$Price)
> avg_price <- mean(retail_df$Price, na.rm = TRUE)
> Clean_replace <- retail_df %>%
+   replace_na(list(
+     Category = "Unknown",
+     Discount = 0,
+     Stock = "Check warehouse",
+     Price = avg_price
+   ))
>
> cat("\n--- 3. Data after replace_na() ---\n")
--- 3. Data after replace_na() ---
>
> print(clean_replace[3, ])
Category Price Rating Stock Discount
3 Unknown 4004 NA In Stock 0
> print(head(clean_replace))
Category Price Rating Stock Discount
1 Unknown 5548 1.870322 Check warehouse 0
2 Unknown 3045 4.757798 Check warehouse 38
3 Unknown 4004 NA In Stock 0
4 Unknown 1.492085 Check warehouse 33
5 Unknown 1817 NA Out of Stock 232
6 Unknown 3522 NA Check warehouse 0
> print(" --- Remaining NAs after replacement --- ")
[1] " --- Remaining NAs after replacement --- "
> print(cols(is.na(clean_replace)))
> print(cols(is.na(clean_replace)))
Error: unexpected symbol in:
"print(cols(is.na(clean_replace)))"
print"
```

Show Traceback Run with Debug

```
> cat("\n--- Remaining NAs after Replacement ---\n")
--- Remaining NAs after Replacement ---
> print(cols(is.na(clean_replace)))
Category Price Rating Stock Discount
0 0 2050 0 0
```

Trending videos Stranger Things... Search DELL R 12:35 01-12-2025

Files Plots Packages Help Viewer Presentation