

PRACTICAL NO 11

AIM Reshaping data using pivot_longer()/pivot_wider() (R).

```

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins Project: (None)
Source
Console Terminal Background Jobs
R > library(dplyr)
> library(tidyverse)
> # use read_csv (explicit) and correct 'na' argument
> df <- read_csv("R/flipkart_com-e-commerce_sample.csv", na = c("", "NA")) %>%
+   mutate(invoiceindex = row_number())
Rows: 20000 Columns: 15
--- Column specification ---
Delimiter: ","
chr (12): uniq_id, crawl_timestamp, product_url, product_name, product_category_tree, pid, image, description, product_rating, overall_rating, brand, produ...
dbl (2): retail_price, discounted_price
lgl (1): is_FK_Advantage_product
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.
> # Print column names so you can confirm what's actually in the file
> cat("--- Columns in the imported CSV ---\n")
--- Columns in the imported CSV ---
> print(colaNames(df))
[1] "uniq_id"           "crawl_timestamp"    "product_url"        "product_name"       "product_category_tree"
[6] "pid"               "retail_price"       "discounted_price"   "image"             "brand"
[11] "description"       "product_rating"     "overall_rating"     "is_FK_Advantage_product"
[16] "InvoiceIndex"
> # Try to keep only the columns of interest if they exist (avoids select() errors)
> df <- df %>% select(invoiceindex, any_of(c("Product.line", "Unit.price", "Tax.5.")))
> cat("--- 1. original wide Data ---\n")
--- 1. Original wide Data ---
> print(head(df))
# A tibble: 6 x 1
  InvoiceIndex
  <int>
1 1
2 2
3 3
4 4
5 5
6 6

```

```

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins Project: (None)
Source
Console Terminal Background Jobs
R > pivot_cols <- intersect(names(df), c("Unit.price", "Tax.5."))
> if (length(pivot_cols) == 0) {
+   cat("Warning: No pivot columns (Unit.price or Tax.5.) found - skipping pivot_longer\n")
+   long_df <- df  # fallback: keep original
+ } else {
+   long_df <- df %>%
+     pivot_longer(
+       cols = any_of(pivot_cols),
+       names_to = "Metric",
+       values_to = "Value"
+     )
+   cat("--- 2. Long Format (pivot_longer) ---\n")
+   print(head(long_df, 6))
+ }
Warning: No pivot columns (Unit.price or Tax.5.) found - skipping pivot_longer

> if ("Metric" %in% names(long_df)) {
+   wide_df <- long_df %>%
+     pivot_wider(
+       names_from = Metric,
+       values_from = Value
+     )
+   cat("--- 3. wide Format (Back to original) ---\n")
+   print(head(wide_df))
+ } else {
+   cat("Skipping pivot_wider because long format was not created.\n")
+ }
Skipping pivot_wider because long format was not created.
> if (!("Product.line" %in% names(df))) f

```

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

The screenshot shows the RStudio interface with the following details:

- File Menu:** File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, Help.
- Project:** Project: (None)
- Source Tab:** Contains R code for data cleaning and pivoting. The code includes handling of 'Product.line' and 'unit.price' columns, creating placeholder values for NA, and performing a category pivot.
- Console Tab:** Displays the output of the R code, showing the resulting tibble 'category_pivot' with 6 rows and 2 columns: 'InvoiceIndex' and 'Unknown'. All values in the 'Unknown' column are 'NA'.
- Terminal Tab:** Not visible in the screenshot.
- Background Jobs Tab:** Not visible in the screenshot.
- Taskbar:** Shows the Windows taskbar with various pinned icons (Calculator, File Explorer, Edge, etc.) and system status (32°C, Sunny, ENG, 08-12-2025, 10:38).

```
R - R4.12 : ~/...  
+ }  
Skipping pivot_wider because long format was not created.\n}  
+ }  
Skipping pivot_wider because long format was not created.  
> if (!("Product.line" %in% names(df))) {  
+ cat("Note: 'Product.line' not present - creating placeholder 'Product.line' = 'Unknown'\n")  
+ df <- df %>% mutate(Product.line = "Unknown")  
+ }  
Note: 'Product.line' not present - creating placeholder 'Product.line' = 'unknown'  
> if (!("unit.price" %in% names(df))) {  
+ cat("Note: 'unit.price' not present - creating placeholder 'unit.price' = NA_real_`\n")  
+ df <- df %>% mutate(unit.price = NA_real_)  
+ }  
Note: 'unit.price' not present - creating placeholder 'unit.price' = NA_real_  
> df_clean <- df %>%  
+ mutate(Product.line = ifelse(is.na(Product.line) | Product.line == "", "Unknown", Product.line))  
> category_pivot <- df_clean %>%  
+ select(InvoiceIndex, Product.line, unit.price) %>%  
+ pivot_wider(  
+   names_from = Product.line,  
+   values_from = unit.price  
+ )  
> cat("--- 4. Category Pivot (Spreading Categories) ---\n")  
-- 4. Category Pivot (Spreading Categories) -->  
> print(head(category_pivot))  
# A tibble: 6 x 2  
  InvoiceIndex Unknown  
        <int>    <dbl>  
1             1     NA  
2             2     NA  
3             3     NA  
4             4     NA  
5             5     NA  
6             6     NA  
>  
>
```