

NOMBRE COMPLETO: Ing. Hilaría Adima Vásquez Durán

NAO ID: 3033

FECHA: 2 de agosto de 2024

NOMBRE DE LA TRAYECTORIA

EN LA QUE ESTÁS ENROLADO: DATA ANALYST CORE

Título del Reto:

Programación R al elaborar reportes estadísticos

SPRINT 1 DESARROLLO R1

Crea un conjunto de programas en R para explorar los datos de Rossmann Pharma a través de tablas con resúmenes estadísticos y visualizaciones que cumplan con las siguientes directrices:

EJERCICIO 2

Crea un programa("prop_changes_sales_by_country_by_year_month.R") que construya una tabla con el cambio porcentual del crecimiento de ventas de cada país a nivel mensual (con respecto al mes anterior). El resultado de este script deberá ser una tabla denominada

"prop_changes_sales_by_country_by_year_month.csv".

Desarrollo del ejercicio

Carga las bibliotecas necesarias para la manipulación de datos (tidyverse), la lectura de archivos Excel (readxl) y la manipulación de fechas (lubridate).

```
```{r}
library(tidyverse)
library(readxl)
library(lubridate)
```
```

```

> library(tidyverse)
— Attaching core tidyverse packages — tidyverse 2.0.0 —
✓ dplyr 1.1.4 ✓ readr 2.1.5
✓ forcats 1.0.0 ✓ stringr 1.5.1
✓ ggplot2 3.5.1 ✓ tibble 3.2.1
✓ lubridate 1.9.3 ✓ tidyr 1.3.1
✓ purrr 1.0.2
— Conflicts — tidyverse_conflicts() —
✖ dplyr::filter() masks stats::filter()
✖ dplyr::lag() masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
> library(readxl)
> library(lubridate)

```

#Establecer el directorio de trabajo a la ruta especificada en

DATA_PATH

```

DATA_PATH <- "/Users/Usuario/Desktop/Help ME"
setwd(DATA_PATH)

```

Cargar los datos correspondientes de para el desarrollo

```

SALES_DATA <- 'sales.csv'
sales <- read.csv2(SALES_DATA, sep=";", header=T)

STORES_INFO_DATA <- 'stores_info.xlsx'
stores <- read_excel(STORES_INFO_DATA, sheet = "data")

```

```

>
> DATA_PATH <- "/Users/Usuario/Desktop/Help ME"
> setwd(DATA_PATH)
> SALES_DATA <- 'sales.csv'
> sales <- read.csv2(SALES_DATA, sep=";", header=T)
> STORES_INFO_DATA <- 'stores_info.xlsx'
> stores <- read_excel(STORES_INFO_DATA, sheet = "data")
>

```

Unión de los datos de ventas (**sales**) con la información de las tiendas (**stores**) usando la columna **store** como clave.

```
sales_all <- sales %>% left_join(stores, by=c("store"))
```

#Convertir la columna

```
sales_all$date <- ymd(sales_all$date)
```

**date a un objeto de fecha**

#Aquí Se Agregan Columnas Para El Año Y El Mes, Y Se Filtran Los Datos Para Los Años 2013 A 2015.

```
sales_all <- sales_all %>%  
  mutate(year = year(date), month = month(date)) %>%  
  filter(year >= 2013 & year <= 2015)
```

#Agrupar Los Datos Por País, Año Y Mes, Y Calcula La Suma Total De Ventas Para Cada Grupo.

```
monthly_sales <- sales_all %>%  
  group_by(country, year, month) %>%  
  summarise(total_sales = sum(sales), .groups = 'drop')
```

```
>  
> sales_all$date <- ymd(sales_all$date)  
> sales_all <- sales_all %>%  
+   mutate(year = year(date), month = month(date)) %>%  
+   filter(year >= 2013 & year <= 2015)  
> monthly_sales <- sales_all %>%  
+   group_by(country, year, month) %>%  
+   summarise(total_sales = sum(sales), .groups = 'drop')  
> 
```

#Ordenan los datos por país, año y mes, se agrupan por país y se calcula la venta del mes anterior (

`prev_month_sales`) usando la función `lag()` . Luego, se calcula el cambio porcentual de las ventas (`pct_change`) y se seleccionan las columnas `country` , `year` , `month` y `pct_change` .

```
monthly_sales <- monthly_sales %>%   arrange(country,  
year, month) %>%   group_by(country) %>%  
mutate(prev_month_sales = lag(total_sales),  
pct_change = (total_sales - prev_month_sales) / prev_  
ungroup() %>%  
  select(country, year, month, pct_change)
```

```
> monthly_sales <- monthly_sales %>%  
+   arrange(country, year, month) %>%  
+   group_by(country) %>%  
+   mutate(prev_month_sales = lag(total_sales),  
+         pct_change = (total_sales - prev_month_sales) / prev_month_sales * 100)  
+   ungroup() %>%  
+   select(country, year, month, pct_change)  
> 
```

#Imprimir En Consola

```
print(monthly_sales)
```

```
# A tibble: 93 x 4
  country year month pct_change
  <chr>   <dbl> <dbl>   <dbl>
1 france  2013     1      NA
2 france  2013     2    -5.80
3 france  2013     3    15.9
4 france  2013     4    -7.68
5 france  2013     5   -0.0801
6 france  2013     6    -2.99
7 france  2013     7    15.4
8 france  2013     8    -6.86
9 france  2013     9    -8.82
10 france 2013    10     6.00
# i 83 more rows
# i Use `print(n = ...)` to see more rows
> 
```

#Crear el archivo CSV llamado

prop_changes_sales_by_country_by_year_month.csv

```
write.csv(monthly_sales,
"prop_changes_sales_by_country_by_year_month.csv")
```

```
> write.csv(monthly_sales, "prop_changes_sales_by_country_by_year_month.csv", $
> cat("La tabla 'prop_changes_sales_by_country_by_year_month.csv' ha sido crea$
La tabla 'prop_changes_sales_by_country_by_year_month.csv' ha sido creada exitosamente.
> 
```

prop_changes_sales_by_country_by_year_month.csv