

Informe del Proyecto: Análisis exploratorio de Datos

Tratamiento de los Datos - Curso 2023/2024

Grupo D

2024-02-09

Índice

1. Introducción	2
2. Materiales	2
2.1. Métodos de importación	2
2.2. Método ejemplo	10
3. Análisis	10
4. Resultados	11
5. Conclusiones	11
6. Ejemplos provisionales de figuras con pie de página	11

1. Introducción

Este proyecto tiene por finalidad realizar un análisis exploratorio de los datos que se han recopilado de un supermercado, en concreto, Mercadona.

A partir de estos datos nos plantearemos unas cuestiones a resolver y realizaremos un estudio a partir de ellas, tratando de extraer conclusiones.

2. Materiales

El material utilizado son diferentes tickets de distintos supermercados de una misma cadena.

Una parte de los datos con los que hemos trabajado han sido proporcionados por el profesorado y otra recopilados por los autores del proyecto.

2.1. Métodos de importación

Realizamos la importación de los datos, para ello crearemos una lista con las cabeceras, otra con los productos (con sus cantidades y precios) y una última con las últimas líneas.

```
# Cargar el paquete readr
```

```
library(readr)
```

```
## Warning: package 'readr' was built under R version 4.2.3
```

```
# Ruta de la carpeta que contiene los archivos de tickets en formato texto
```

```
ruta <- "data"
```

```
# Obtener la lista de archivos en la carpeta
```

```
archivos <- list.files(path = ruta, full.names = TRUE)
```

```
# Crear listas para almacenar las cabeceras, productos y últimas líneas de cada archivo
```

```
cabeceras <- list()
```

```
productos <- list()
```

```
ultimas_lineas <- list()
```

```
# Recorrer cada archivo y procesar su contenido
```

```
for (archivo in archivos) {
```

```
  # Leer el contenido del archivo
```

```
contenido <- read_lines(archivo)

# Encontrar el índice de la palabra "TOTAL"
indice_total <- grep("TOTAL", contenido)

# Extraer las cabeceras (primeras 7 líneas)
cabecera <- contenido[1:min(7, length(contenido))]

# Extraer los productos
productos_temp <- contenido[8:(indice_total - 1)]

# Extraer las últimas líneas (después de "TOTAL")
ultimas_lineas_temp <- contenido[(indice_total + 1):length(contenido)]

# Almacenar las líneas en las listas correspondientes
cabeceras[[archivo]] <- cabecera
productos[[archivo]] <- productos_temp
ultimas_lineas[[archivo]] <- ultimas_lineas_temp

}

## Warning in 8:(indice_total - 1): numerical expression has 2 elements: only the
## first used

## Warning in (indice_total + 1):length(contenido): numerical expression has 2
## elements: only the first used

## Warning in 8:(indice_total - 1): numerical expression has 2 elements: only the
## first used

## Warning in (indice_total + 1):length(contenido): numerical expression has 2
## elements: only the first used

## Warning in 8:(indice_total - 1): numerical expression has 2 elements: only the
## first used

## Warning in (indice_total + 1):length(contenido): numerical expression has 2
## elements: only the first used
```

```
## Warning in 8:(indice_total - 1): numerical expression has 2 elements: only the  
## first used
```

```
## Warning in (indice_total + 1):length(contenido): numerical expression has 2  
## elements: only the first used
```

```
## Warning in 8:(indice_total - 1): numerical expression has 2 elements: only the  
## first used
```

```
## Warning in (indice_total + 1):length(contenido): numerical expression has 2  
## elements: only the first used
```

```
## Warning in 8:(indice_total - 1): numerical expression has 2 elements: only the  
## first used
```

```
## Warning in (indice_total + 1):length(contenido): numerical expression has 2  
## elements: only the first used
```

```
## Warning in 8:(indice_total - 1): numerical expression has 2 elements: only the  
## first used
```

```
## Warning in (indice_total + 1):length(contenido): numerical expression has 2  
## elements: only the first used
```

```
## Warning in 8:(indice_total - 1): numerical expression has 2 elements: only the  
## first used
```

```
## Warning in (indice_total + 1):length(contenido): numerical expression has 2  
## elements: only the first used
```

```
## Warning in 8:(indice_total - 1): numerical expression has 2 elements: only the  
## first used
```

```
## Warning in (indice_total + 1):length(contenido): numerical expression has 2  
## elements: only the first used
```

```
## Warning in 8:(indice_total - 1): numerical expression has 2 elements: only the  
## first used
```

```
## Warning in (indice_total + 1):length(contenido): numerical expression has 2
## elements: only the first used
```

```
## Warning in 8:(indice_total - 1): numerical expression has 2 elements: only the
## first used
```

```
## Warning in (indice_total + 1):length(contenido): numerical expression has 2
## elements: only the first used
```

```
## Warning in 8:(indice_total - 1): numerical expression has 2 elements: only the
## first used
```

```
## Warning in (indice_total + 1):length(contenido): numerical expression has 2
## elements: only the first used
```

```
## Warning in 8:(indice_total - 1): numerical expression has 2 elements: only the
## first used
```

```
## Warning in (indice_total + 1):length(contenido): numerical expression has 2
## elements: only the first used
```

```
## Warning in 8:(indice_total - 1): numerical expression has 2 elements: only the
## first used
```

```
## Warning in (indice_total + 1):length(contenido): numerical expression has 2
## elements: only the first used
```

```
## Warning in 8:(indice_total - 1): numerical expression has 2 elements: only the
## first used
```

```
## Warning in (indice_total + 1):length(contenido): numerical expression has 2
## elements: only the first used
```

```
## Warning in 8:(indice_total - 1): numerical expression has 2 elements: only the
## first used
```

```
## Warning in (indice_total + 1):length(contenido): numerical expression has 2
## elements: only the first used
```

```
## Warning in 8:(indice_total - 1): numerical expression has 2 elements: only the  
## first used
```

```
## Warning in (indice_total + 1):length(contenido): numerical expression has 2  
## elements: only the first used
```

```
## Warning in 8:(indice_total - 1): numerical expression has 2 elements: only the  
## first used
```

```
## Warning in (indice_total + 1):length(contenido): numerical expression has 2  
## elements: only the first used
```

```
## Warning in 8:(indice_total - 1): numerical expression has 2 elements: only the  
## first used
```

```
## Warning in (indice_total + 1):length(contenido): numerical expression has 2  
## elements: only the first used
```

```
## Warning in 8:(indice_total - 1): numerical expression has 2 elements: only the  
## first used
```

```
## Warning in (indice_total + 1):length(contenido): numerical expression has 2  
## elements: only the first used
```

```
## Warning in 8:(indice_total - 1): numerical expression has 2 elements: only the  
## first used
```

```
## Warning in (indice_total + 1):length(contenido): numerical expression has 2  
## elements: only the first used
```

```
## Warning in 8:(indice_total - 1): numerical expression has 2 elements: only the  
## first used
```

```
## Warning in (indice_total + 1):length(contenido): numerical expression has 2  
## elements: only the first used
```

```
## Warning in 8:(indice_total - 1): numerical expression has 2 elements: only the  
## first used
```

```
## Warning in (indice_total + 1):length(contenido): numerical expression has 2
## elements: only the first used
```

```
## Warning in 8:(indice_total - 1): numerical expression has 2 elements: only the
## first used
```

```
## Warning in (indice_total + 1):length(contenido): numerical expression has 2
## elements: only the first used
```

```
## Warning in 8:(indice_total - 1): numerical expression has 2 elements: only the
## first used
```

```
## Warning in (indice_total + 1):length(contenido): numerical expression has 2
## elements: only the first used
```

```
## Warning in 8:(indice_total - 1): numerical expression has 2 elements: only the
## first used
```

```
## Warning in (indice_total + 1):length(contenido): numerical expression has 2
## elements: only the first used
```

```
## Warning in 8:(indice_total - 1): numerical expression has 2 elements: only the
## first used
```

```
## Warning in (indice_total + 1):length(contenido): numerical expression has 2
## elements: only the first used
```

```
## Warning in 8:(indice_total - 1): numerical expression has 2 elements: only the
## first used
```

```
## Warning in (indice_total + 1):length(contenido): numerical expression has 2
## elements: only the first used
```

```
## Warning in 8:(indice_total - 1): numerical expression has 2 elements: only the
## first used
```

```
## Warning in (indice_total + 1):length(contenido): numerical expression has 2
## elements: only the first used
```

```
## Warning in 8:(indice_total - 1): numerical expression has 2 elements: only the  
## first used
```

```
## Warning in (indice_total + 1):length(contenido): numerical expression has 2  
## elements: only the first used
```

```
## Warning in 8:(indice_total - 1): numerical expression has 2 elements: only the  
## first used
```

```
## Warning in (indice_total + 1):length(contenido): numerical expression has 2  
## elements: only the first used
```

```
## Warning in 8:(indice_total - 1): numerical expression has 2 elements: only the  
## first used
```

```
## Warning in (indice_total + 1):length(contenido): numerical expression has 2  
## elements: only the first used
```

```
## Warning in 8:(indice_total - 1): numerical expression has 2 elements: only the  
## first used
```

```
## Warning in (indice_total + 1):length(contenido): numerical expression has 2  
## elements: only the first used
```

```
## Warning in 8:(indice_total - 1): numerical expression has 2 elements: only the  
## first used
```

```
## Warning in (indice_total + 1):length(contenido): numerical expression has 2  
## elements: only the first used
```

```
## Warning in 8:(indice_total - 1): numerical expression has 2 elements: only the  
## first used
```

```
## Warning in (indice_total + 1):length(contenido): numerical expression has 2  
## elements: only the first used
```

```
## Warning in 8:(indice_total - 1): numerical expression has 2 elements: only the  
## first used
```



```
## Warning in (indice_total + 1):length(contenido): numerical expression has 2
## elements: only the first used
```

```
## Warning in 8:(indice_total - 1): numerical expression has 2 elements: only the
## first used
```

```
## Warning in (indice_total + 1):length(contenido): numerical expression has 2
## elements: only the first used
```

```
## Warning in 8:(indice_total - 1): numerical expression has 2 elements: only the
## first used
```

```
## Warning in (indice_total + 1):length(contenido): numerical expression has 2
## elements: only the first used
```

```
## Warning in 8:(indice_total - 1): numerical expression has 2 elements: only the
## first used
```

```
## Warning in (indice_total + 1):length(contenido): numerical expression has 2
## elements: only the first used
```

```
## Warning in 8:(indice_total - 1): numerical expression has 2 elements: only the
## first used
```

```
## Warning in (indice_total + 1):length(contenido): numerical expression has 2
## elements: only the first used
```

```
## Warning in 8:(indice_total - 1): numerical expression has 2 elements: only the
## first used
```

```
## Warning in (indice_total + 1):length(contenido): numerical expression has 2
## elements: only the first used
```

```
## Warning in 8:(indice_total - 1): numerical expression has 2 elements: only the
## first used
```

```
## Warning in (indice_total + 1):length(contenido): numerical expression has 2
## elements: only the first used
```

```
#Mostramos las primeras líneas para ver el formato de cada lista
```

```
head(cabeceras[1])
```

```
## $`data/20231213 Mercadona 4,85 .txt`  
## [1] "MERCADONA, S.A. A-46103834"  
## [2] "C/ PROFESOR BELTR\x1c1N B\x1c1GUENA S/N"  
## [3] "46009 VALENCIA"  
## [4] "TEL\x9FONO: 963470267"  
## [5] "13/12/2023 18:32 OP: 348873"  
## [6] "FACTURA SIMPLIFICADA: 2427-012-371161"  
## [7] "Descripci\xf3n P. Unit Importe"
```

```
head(productos[1])
```

```
## $`data/20231213 Mercadona 4,85 .txt`  
## [1] "1TORTITAS ARROZ YOGUR 1,60" "1CAMELOS VITA-C 2,35"  
## [3] "1CAFE LIGHT S/LAC 0,90"
```

```
head(ultimas_lineas[1])
```

```
## $`data/20231213 Mercadona 4,85 .txt`  
## [1] "TARJETA BANCARIA 4,85"  
## [2] "IVA BASE IMPONIBLE (\x80) CUOTA (\x80)"  
## [3] "10% 4,41 0,44"  
## [4] "TOTAL 4,41 0,44"  
## [5] "TARJ. BANCARIA: **** * 8438"  
## [6] "N.C: 098100860 AUT: 154052"  
## [7] "AID: A0000000041010 ARC: 3030"  
## [8] "MASTERCARD"  
## [9] "Importe: 4,85 \x80 Mastercard"  
## [10] "SE ADMITEN DEVOLUCIONES CON TICKET"
```

2.2. Método ejemplo

(Aquí el material que hemos obtenido y que métodos vamos a usar para tratarlos)

3. Análisis

(Aquí la realización del estudio y las preguntas a plantear)

4. Resultados

(Aquí los resultados que hemos obtenido)

5. Conclusiones

(Aquí las conclusiones que hemos sacado sobre el proyecto)

6. Ejemplos provisionales de figuras con pie de página

Ejemplo 1: Vamos a observar un gráfico hecho con código, la [Figura 1](#) (si se le da lleva a la figura, es un enlace).

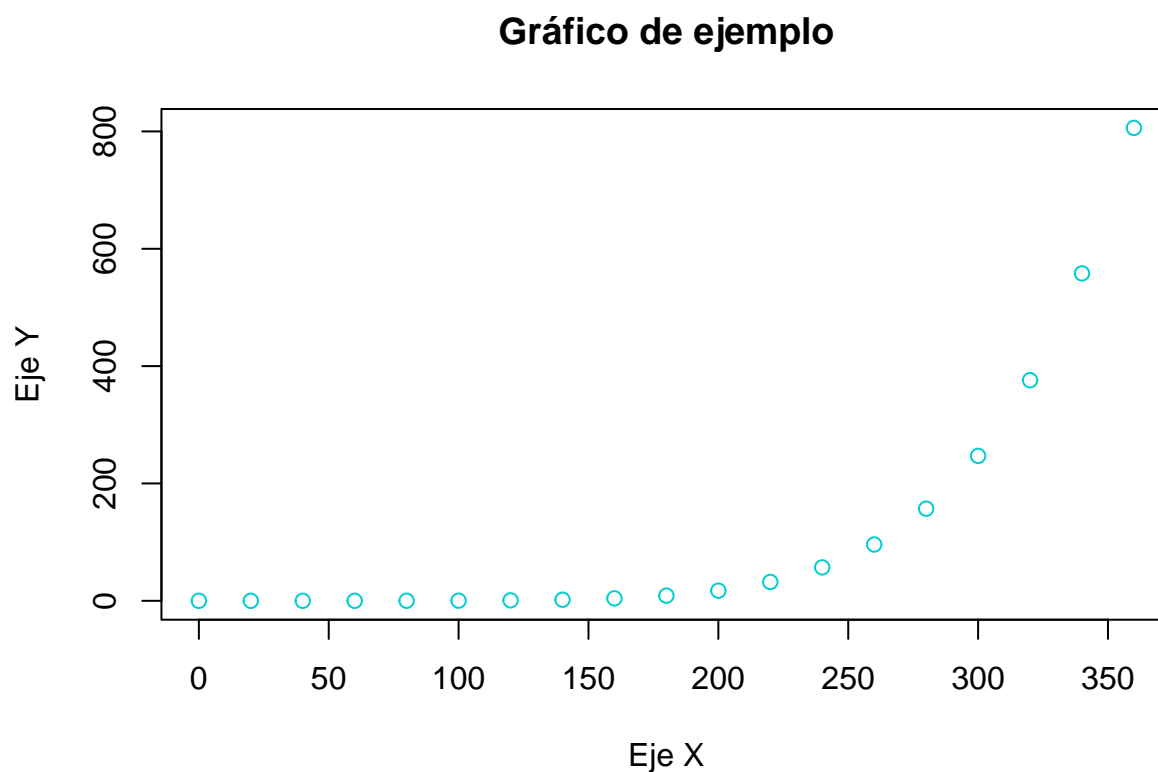


Figura 1: Gráfico de ejemplo para mostrar un pie de imagen

Ejemplo 2: Vamos a observar una imagen, la [Figura 2](#) (si se le da lleva a la figura, es un enlace).

Ejemplo 3: Vamos a observar una figura hecha con *l^atex*, por ejemplo, veremos mejor una ecuación, la [Ecuacion 1](#) (si se le da lleva a la figura, es un enlace).



Figura 2: Pie de página de la imagen de ejemplo

$$e^{i\pi} + 1 = 0$$

(1)