

# ProyectoTD2024

Eric Villaescusa, Isaac Pazmiño, Alexandra Estela, Oscar Ibañez, Adrià Reyes

2024-03-30

```
#Cargamos los nombres de cada tiket (guardados en la carpeta data)
files <- list.files(path = "./data", pattern = ".txt", full.names = TRUE)

#Creamos una lista que contendrá cada ticket.
tickets <- list()

#Asignamos las líneas de los archivos .txt
for (file in files) {
  linea <- read_lines(file, locale = locale(encoding = 'latin1'))
  supermercado <- linea[1]
  direccion <- linea[2]
  ciudad <- linea[3]
  telefono <- linea[4]
  fecha <- linea[5]
  factura <- linea[6]
  indiceinicio <- grep("DESCRIPCIÓN", linea)
  indicefinal <- grep("TOTAL", linea)
  productos <- linea[indiceinicio + 1:indicefinal - 1]
  #Falta IVA
  tarjetabancaria <- linea[indicefinal + 7]

#Creamos un data frame con los datos asignados
  datos <- data.frame(
    Supermercado <- supermercado, #Falta distinguir entre los distintos productos porque tienen distintos
    Dirección <- direccion,
    Ciudad <- ciudad,
    Telefono <- telefono,
    Fecha <- fecha,
    Factura <- factura,
    #Falta añadir productos, error con el número de columnas (ARREGLAR)
    stringsAsFactors = FALSE
  )
  #Falta asignar el IVA
}

#Le asigno cada ticket a cada elemento de la lista
tickets <- list.append(tickets, datos)
}
```

```
## Warning in 1:indicefinal: numerical expression has 2 elements: only the first
## used
```

```
## Warning in 1:indicefinal: numerical expression has 2 elements: only the first
## used
```

```
## Warning in 1:indicefinal: numerical expression has 2 elements: only the first
## used

## Warning in 1:indicefinal: numerical expression has 2 elements: only the first
## used

## Warning in 1:indicefinal: numerical expression has 2 elements: only the first
## used

## Warning in 1:indicefinal: numerical expression has 2 elements: only the first
## used

## Warning in 1:indicefinal: numerical expression has 2 elements: only the first
## used

## Warning in 1:indicefinal: numerical expression has 2 elements: only the first
## used

## Warning in 1:indicefinal: numerical expression has 2 elements: only the first
## used

## Warning in 1:indicefinal: numerical expression has 2 elements: only the first
## used
```

```
#Creamos un dataframe vacio donde añadiremos la información contenida en cada ticket.
datoscompletos <- data.frame()

#Le añadimos a cada columna del data frame un ticket.
for (ticket in tickets) {
  datoscompletos <- rbind(datoscompletos,ticket)
}
```

## Preguntas

Ahora que tenemos los tickets cargados correctamente en un data frame, nos planteamos las preguntas que queremos responder mediante el análisis exploratorio.

1. ¿cuáles son los productos más comprados?
2. ¿A qué hora se realizan más compras?
3. ¿Qué método de pago es el más popular?
4. ¿Cuáles son los productos menos comprados?