

# ProyectoTD2024

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

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 P. Unit Importe", linea)
  indicefinal <- grep("SE ADMITEN DEVOLUCIONES CON TICKET", linea)
  productos <- linea[(indiceinicio + 2):(indicefinal - 14)]
  iVA_10_idx <- grep("10%", linea)
  iVA_21_idx <- grep("21%", linea)
  iVA_5_idx <- grep("\\b5%", linea)
  candidatos_sinIVA_idx <- grep("\\b0,00\\b", linea)
  sinIVA_idx <- candidatos_sinIVA_idx[!grepl("PARKING", linea[candidatos_sinIVA_idx])]
  iVA_10_value <- ifelse(length(iVA_10_idx) > 0, linea[iVA_10_idx[1]], NA)
  iVA_21_value <- ifelse(length(iVA_21_idx) > 0, linea[iVA_21_idx[1]], NA)
  iVA_5_value <- ifelse(length(iVA_5_idx) > 0, linea[iVA_5_idx[1]], NA)
  sinIVA_value <- ifelse(length(sinIVA_idx) > 0, linea[sinIVA_idx[1]], NA)
#Creamos un data frame con los datos asignados
  datos <- data.frame(
    Supermercado = supermercado,
    Dirección = direccion,
    Ciudad = ciudad,
    Telefono = telefono,
    Fecha = fecha,
    Productos = productos,
    Factura = factura,
    Iva_10 = iVA_10_value,
    Iva_21 = iVA_21_value,
    Iva_5 = iVA_5_value,
    SinIVA = sinIVA_value,
    stringsAsFactors = FALSE
  )
}
```

```

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

#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)
}

```

## ##INTRODUCCIÓN

Este proyecto aspira a realizar un análisis exhaustivo de los patrones de compra de los clientes de Mercadona, una destacada cadena de supermercados en España, basándose en los tickets de compra recogidos durante el período 2023-2024. Con el objetivo de identificar tendencias de consumo, determinar los productos más y menos comprados, examinar las preferencias horarias de compra, y evaluar los métodos de pago preferidos por los consumidores, este estudio profundo busca no solo comprender los hábitos de compra en Mercadona, sino también ofrecer recomendaciones estratégicas para optimizar tanto las estrategias de marketing como la gestión de inventario de la cadena. Este enfoque analítico promete revelar insights valiosos para mejorar la experiencia de compra de los clientes y fortalecer la posición de Mercadona en el mercado.

### #1.1 Material y Métodos

Detallaremos los materiales y métodos empleados en el análisis, abordando los desafíos encontrados, como la necesidad de convertir los datos de su forma original (“Raw data”) a un formato técnicamente adecuado que permita un análisis eficiente. Este proceso incluirá la codificación manual de ciertas variables y la normalización de formatos, especialmente en lo que respecta a las fechas y métodos de pago, así como el desglose de los porcentajes de IVA para asegurar coherencia a lo largo de todo el análisis.

**#1.1.1 Carga de librerías y datos:** Primero hemos de descargar las librerías que sean necesarias para la carga y análisis de los datos. En este caso hemos instalado tres: ‘readr’, ‘dplyr’ y ‘rlist’. La librería readr sirve para leer datos de archivos de manera eficiente. La librería dplyr nos servirá para el procesamiento y manipulación de los datos cargados. Y por último, rlist manipula las listas en R. Los datos utilizados en este análisis provienen exclusivamente de los archivos de texto que contienen los registros detallados de cada compra (tickets), almacenados en la carpeta data de nuestro proyecto. Este conjunto de datos incluye una amplia gama de información, como el detalle de los productos comprados, la fecha y hora de la compra, el método de pago, datos de la sucursal donde se está efectuando la compra, desglose del ‘IVA’. A través de la exploración de 12 tickets, se ha conformado un dataframe que nos permite realizar un análisis significativo sobre las tendencias de consumo. Para la lectura de datos hemos hecho uso de la primera librería mencionada (readr), iterando con un bucle sobre files (que son los archivos de texto) y creando un dataframe por cada ticket.

**#1.2. Características generales de los datos Variables:** En nuestro dataframe creado, llamado datos, se le asigna un total de 9 variables: supermercado, direccion, ciudad, teléfono, fecha, factura, IVA 10%, IVA 21% e IVA 5%, Sin IVA e IVA. Descripción detallada de las Variables: Supermercado: Nombre del supermercado donde se realizó la compra. Este dato es importante para identificar el punto de venta específico dentro de la cadena de supermercados. Dirección: Dirección física del supermercado. Proporciona un nivel adicional de detalle geográfico que podría ser útil para análisis relacionados con la ubicación y así saber que afluencia de personas tiene cada supermercado y de esta manera considerar el incremento de infraestructura en esa zona, mayor manejo de productos, entre otros. Ciudad: La ciudad en la que se encuentra el supermercado. Esta variable permite realizar análisis por ciudad, identificando patrones de consumo provincial y regional. Teléfono: Número de teléfono del supermercado. Aunque esta información puede no ser directamente relevante para el análisis de patrones de compra, sirve como dato de contacto. Fecha: Fecha y hora en que se realizó la compra. Es fundamental para análisis temporales, como identificar picos de compra durante el

día, la semana o tendencias dependiendo de la temporada del año/mes. Factura: Número o código de la factura asociada a cada compra. Un identificador único para cada transacción que permite el seguimiento de las compras individuales. IVA (10%, 21%, 5%): Porcentajes de IVA aplicados a los productos adquiridos. Estos datos son cruciales para análisis financieros y para entender cómo los diferentes tipos de productos contribuyen a las ventas totales. Sin IVA: Monto total de los productos exentos de IVA. Ofrece una visión de la proporción de productos no gravados en las compras. IVA: Monto total del IVA cobrado en la compra. Un indicador importante de la carga fiscal de las transacciones. Para finalizar, se crea un dataframe vacío para almacenar la información de todos los tickets, es decir, un dataframe que almacene todos los anteriores dataframes, llamado datoscompletos.

#1.3 Resultados Presentaremos los hallazgos y logros clave derivados del análisis, proporcionando una base sólida para decisiones estratégicas informadas y mejoras operativas.

## 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?