

# Mini Proyecto de Análisis Exploratorio de Datos

Tratamiento de Datos. Grado en Ciencia de Datos- UV

Laura Beltran, Victor Cardama, Mara Ferrús, Jorge García, Sandra López

2024-05-14

## Índice

<b>1 INTRODUCCIÓN PROYECTO.</b>	<b>2</b>
<b>2 CARGA DE LIBRERÍAS NECESARIAS PARA EL ANÁLISIS.</b>	<b>2</b>
<b>3 PROCESAMIENTO DE DATOS.</b>	<b>3</b>
<b>4 DESCRIPCIÓN DE DATOS USADOS Y CARACTERÍSTICAS.</b>	<b>3</b>
<b>5 DATOS FALTANTES EN LOS TICKETS DE MERCADONA.</b>	<b>4</b>
<b>6 PREGUNTAS PLANTEADAS CARA AL ANÁLISIS.</b>	<b>5</b>
<b>7 EXPLORACIÓN Y VISUALIZACIÓN.</b>	<b>6</b>
<b>8 VISUALIZACIÓN OUTLIERS</b>	<b>7</b>
8.1 Análisis univariante. . . . .	8
8.2 Análisis multivariante. . . . .	9
<b>9 RESULTADOS.</b>	<b>9</b>
<b>10 CONCLUSIONES.</b>	<b>9</b>

# 1 INTRODUCCIÓN PROYECTO.

El comercio minorista ha experimentado una transformación significativa en las últimas décadas, con la digitalización y la adopción de nuevas tecnologías que han impactado en la forma en que los consumidores realizan sus compras. En este contexto, la disponibilidad de datos detallados sobre las transacciones de compra se ha convertido en un activo invaluable para las empresas minoristas, permitiéndoles comprender mejor el comportamiento del cliente, optimizar la gestión de inventario y desarrollar estrategias de marketing más efectivas.

En este sentido, Mercadona, una de las cadenas de supermercados más grandes y reconocidas en España, se encuentra en una posición privilegiada para aprovechar los datos generados por las transacciones de sus clientes. Con una amplia red de tiendas y una amplia variedad de productos, Mercadona está constantemente buscando formas de mejorar su oferta y satisfacer las necesidades cambiantes de sus clientes.

En línea con esta visión, el presente estudio se centra en el análisis de los tickets de compra obtenidos en Mercadona, con el objetivo de extraer información valiosa sobre los patrones de compra, preferencias de productos, tendencias de gasto y otros aspectos relevantes relacionados con la experiencia del cliente. Al analizar estos datos de manera sistemática, podemos obtener una visión más clara y detallada de los hábitos de compra de los clientes, lo que a su vez puede informar decisiones estratégicas importantes para Mercadona.

Este informe se estructura en torno a una serie de preguntas de análisis específicas, que abarcan desde la relación entre las compras individuales hasta el análisis de tendencias a nivel agregado. A través de un enfoque riguroso de procesamiento de datos y análisis estadístico, buscamos proporcionar información relevante y perspicaz que pueda ser utilizada por Mercadona para mejorar la eficiencia operativa, optimizar la gestión de inventario y mejorar la experiencia general del cliente.

## 2 CARGA DE LIBRERÍAS NECESARIAS PARA EL ANÁLISIS.

En nuestro proyecto de análisis de tickets de Mercadona, hemos hecho uso de varias librerías en el lenguaje de programación *R* para llevar a cabo diferentes tareas de procesamiento, análisis y visualización de datos.

Vamos a explicarla una por una:

1. *dplyr*: Esta es una de las librerías más populares para el manejo de datos en *R*. Proporciona una gramática consistente y fácil de usar para manipular conjuntos de datos, permitiendo realizar operaciones como filtrado, selección, agrupación y resumen de datos de forma eficiente.
2. *ggplot2*: Se trata de una librería ampliamente utilizada para la visualización de datos en *R*. Permite crear gráficos de alta calidad y altamente personalizables mediante una sintaxis declarativa. Con esta librería, podemos crear una amplia variedad de gráficos, incluyendo histogramas, gráficos de dispersión, diagramas de caja y diagramas de barras, entre otros.
3. *tidyr*: Esta librería se utiliza para realizar operaciones de limpieza y transformación de datos en formato "*tidy*" u "*ordenado*". Proporciona funciones para pivotar datos, separar y unir columnas, y reorganizar conjuntos de datos de manera que sea más fácil trabajar con ellos en el análisis posterior.
4. *lubridate*: Cuando trabajamos con fechas y horas en *R*, la librería *lubridate* resulta muy útil. Esta librería simplifica el manejo de fechas y horas, permitiendo realizar operaciones como extracción de componentes de fechas, cálculo de diferencias entre fechas y manipulación de formatos de fecha y hora de manera sencilla.
5. *stringr*: Para el manejo de cadenas de texto en *R*, la librería *stringr* es una opción excelente. Proporciona funciones para manipular y analizar cadenas de texto, incluyendo operaciones de búsqueda, reemplazo, división y extracción de patrones.

6. *knitr*: Esta librería es fundamental para la generación de documentos dinámicos en R, permitiendo combinar código R con texto explicativo y resultados de análisis en un único documento. Con *knitr*, podemos crear informes, presentaciones o documentos en formato *HTML*, *PDF* o *Word* que incluyan tanto el código utilizado en el análisis como los resultados obtenidos, lo que facilita la reproducibilidad y la comunicación de los resultados del análisis de datos.

### 3 PROCESAMIENTO DE DATOS.

El “Procesamiento de Datos” es una etapa fundamental en cualquier proyecto analítico, y en el contexto de nuestro proyecto de tickets de Mercadona, no es diferente.

1. Importación de Datos: En esta fase, los datos de los tickets de Mercadona se cargan desde su fuente original, que podría ser una base de datos, un archivo CSV u otro formato de archivo. Es crucial asegurarse de que la importación se realice correctamente para garantizar la integridad de los datos.
2. Limpieza de Datos: Una vez que los datos se cargan, es común encontrarse con problemas como valores faltantes, formatos incorrectos o datos erróneos. Durante la limpieza de datos, se aplican diversas técnicas para abordar estos problemas.
  - Extracción de Cantidades y Precios: Los productos en los tickets pueden tener información asociada, como la cantidad y el precio. Se necesitará extraer esta información para análisis posteriores.
  - Separación de Fecha y Hora: Si la fecha y la hora están combinadas en una sola columna, es necesario separarlas para facilitar su análisis temporal.
  - Corrección de Formatos: Los datos pueden estar en formatos incorrectos o inconsistentes. Es esencial estandarizar estos formatos para evitar confusiones.
3. Estructuración de Datos: Una vez que los datos están limpios, se procede a estructurarlos de manera que sean adecuados para el análisis. Esto puede implicar la creación de nuevas variables, la combinación de datos de diferentes fuentes o la agregación de datos a diferentes niveles de granularidad.

### 4 DESCRIPCIÓN DE DATOS USADOS Y CARACTERÍSTICAS.

La descripción de los datos establece la información recopilada de los tickets de compra de Mercadona. Estos tickets, que se corresponden con el material de nuestro análisis, representan las transacciones comerciales realizadas por los clientes en las tiendas de Mercadona. Se encuentran almacenados en archivos de texto (*.txt*) dentro de la carpeta designada como *'/data'*, cada ticket muestra información detallada de la compra de cada cliente.

La información de cada ticket radica en tanto detalles básicos como la ubicación del supermercado y la fecha de la transacción, como detalles específicos: productos adquiridos y el método de pago utilizado. La inclusión de estos detalles es fundamental para comprender los hábitos de compra de los clientes y las tendencias de consumo en Mercadona.

Entre los datos clave incluidos en cada ticket se encuentran:

- Información del Supermercado: Esto incluye el nombre del supermercado, su identificación fiscal y su ubicación específica, lo que proporciona contexto sobre la tienda en la que se realizó la compra.
- Dirección y Datos de Contacto: Los tickets también suelen incluir la dirección completa del supermercado, lo que permite la localización precisa de la tienda. Además, es común encontrar información de contacto, como números de teléfono, que pueden ser útiles.

- **Detalles de la Transacción:** Cada ticket registra la fecha y la hora exactas en que se realizó la compra, lo que permite un análisis detallado de los patrones de compra a lo largo del tiempo. Además, se proporciona un número de factura único para cada transacción, lo que facilita la identificación y el seguimiento de transacciones individuales.
- **Productos Adquiridos:** Esta sección del ticket enumera todos los productos comprados durante la transacción, junto con detalles como la cantidad adquirida y el precio. Esta información es fundamental para comprender las preferencias de los clientes y las tendencias de compra de productos específicos.
- **Método de Pago:** Finalmente, cada ticket indica el método de pago utilizado para la transacción, lo que puede incluir efectivo, tarjeta de crédito, débito u otros métodos de pago electrónico. Este detalle es importante para comprender las preferencias de pago de los clientes y la evolución de los métodos de pago en el tiempo.

*Tabla 1: VARIABLES DE INTERÉS PARA NUESTRO ESTUDIO.*

VARIABLE	DESCRIPCIÓN
list_tickets	Lista de tickets guardados en la carpeta ‘data’.
Supermercado	El supermercado donde se ha hecho la compra.
Dirección	La dirección donde se encuentra el supermercado.
codigo_postal_ciudad	El código postal donde se encuentra la ciudad del supermercado.
Teléfono	El teléfono del supermercado.
Fecha_hora_operador	La fecha y la hora en que se produce la compra.
Factura	La factura donde se registran todas las características de la compra.
Productos	Todos los productos que se han comprado.
Total	El precio total que hay que pagar por los productos comprados.
Metodo_pago	El diferente método de pago que se usa, es decir, si es en tarjeta o en efectivo.
Importe	Precio total a pagar.

Debemos asegurarnos de que cada una de estas variables esté correctamente establecida en nuestros datos, que no hayan datos faltantes, que tengan valores correctos... Para ello efectuamos un análisis exploratorio de estos datos, este se utiliza cuando no tenemos información de los datos como para plantear hipótesis. El objetivo es buscar estructuras dentro de los datos que puedan resultar de interés.

## 5 DATOS FALTANTES EN LOS TICKETS DE MERCADONA.

Las etapas en un problema de análisis de datos radican en una importación de los datos a analizar, seguida de una preparación de dichos datos; tras esto aplicaríamos las transformaciones necesarias para tratar correctamente nuestros datos, cuando ya los tenemos en la forma adecuada, visualizamos para percibir resultados de una manera más precisa, por último, incidimos en un modelado para la posterior comunicación de los resultados.

Para realizar un análisis de los datos faltantes en los tickets de Mercadona, primero necesitaríamos conocer su estructura. Procedemos a la explicación de como abordar este análisis.

1. **Exploración de los datos:** Lo primero sería cargar los datos de los tickets de Mercadona y explorar su estructura. Esto nos permitiría identificar las variables presentes en los tickets y entender cómo se almacenan los datos.
2. **Identificación de datos faltantes:** Una vez que tengamos los datos cargados, podríamos identificar las variables que contienen valores faltantes. Esto se puede hacer calculando el porcentaje de valores faltantes para cada variable y visualizando la distribución de los valores faltantes en un gráfico.

3. Análisis de patrones de datos faltantes: Después de identificar las variables con datos faltantes, analizamos los posibles patrones en los datos faltantes. Por ejemplo, podríamos investigar si hay algún patrón temporal en los datos faltantes o si ciertas variables tienden a tener más datos faltantes que otras.
4. Imputación de datos faltantes: Dependiendo del contexto y la naturaleza de los datos, podríamos optar por imputar los valores faltantes utilizando técnicas como la imputación media, la imputación de vecinos más cercanos o modelos de imputación más avanzados. Sin embargo, es importante tener en cuenta que la imputación de datos faltantes debe realizarse con cuidado y teniendo en cuenta el impacto en los resultados del análisis.
5. Evaluación del impacto de los datos faltantes: Finalmente, podríamos evaluar el impacto de los datos faltantes en nuestros análisis. Esto podría implicar comparar los resultados de los análisis con y sin imputación de datos faltantes, o realizar análisis de sensibilidad para evaluar cómo varían los resultados bajo diferentes escenarios de imputación.

Representamos la gráfica en R:

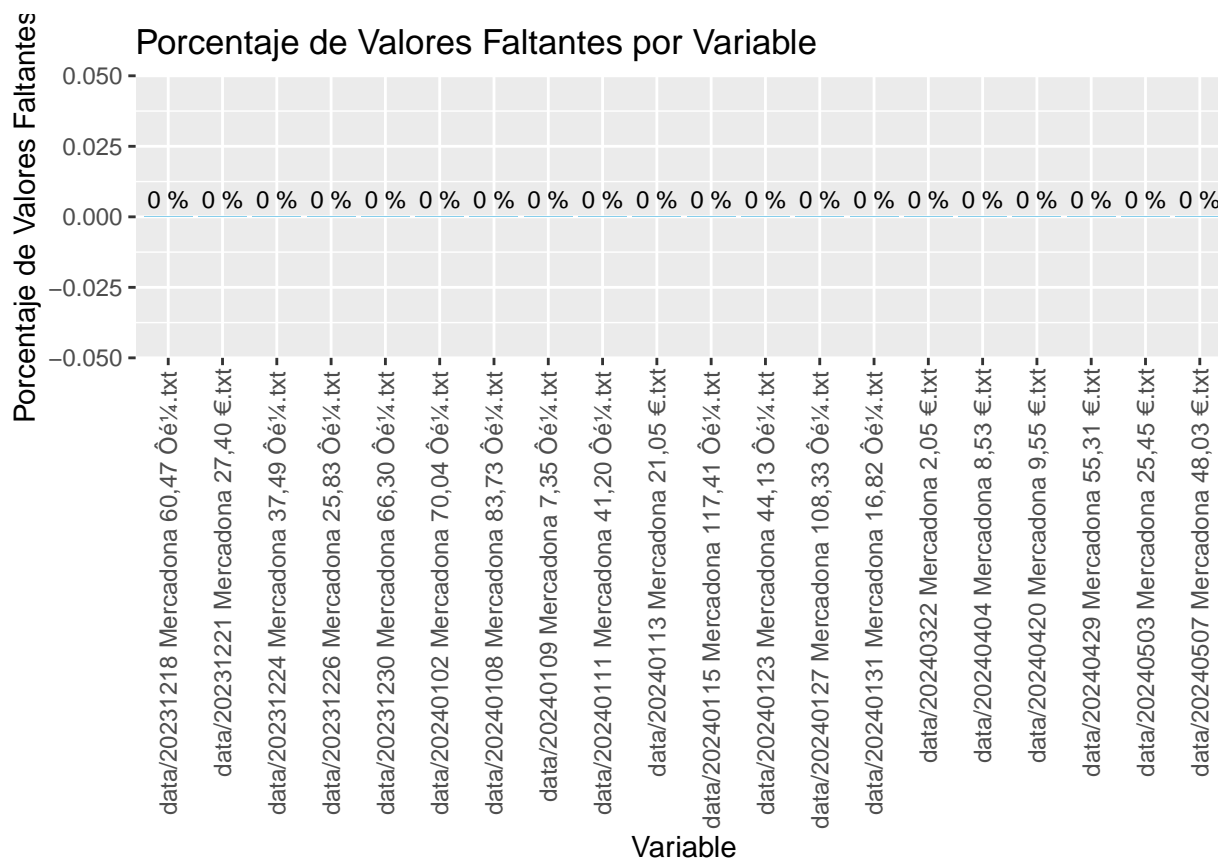


Figura 1: Figura 1: Porcentaje valores faltantes

Observamos que en ningún ticket se encuentra algún “dato faltante”. Esto es debido a que no todos los tickets presentan los mismos productos, pero en sí sí que siguen una estructura en concreto la cual todos los tickets deben seguir para un procesamiento correcto.

## 6 PREGUNTAS PLANTEADAS CARA AL ANÁLISIS.

1. ¿Cuál es el promedio de gasto de los tickets?

2. ¿Categoría y producto más comprados en general?
3. ¿Cuál es el día de la semana más habitual en que se han producido los tickets registrados?
4. ¿Cual ha sido el Mercadona con más productos comprados? ¿Y en el que más compras se han hecho?

## 7 EXPLORACIÓN Y VISUALIZACIÓN.

Una vez tengamos nuestros datos en la estructura de `data.frame` y estén correctamente almacenados, podemos empezar a buscar posibles patrones en las instancias; esto lo podemos hacer en base a ciertas preguntas planteadas por nosotros.

1. **Análisis de Missing Data:** Cuando nos adentramos en el análisis de los datos de tickets, nos enfrentamos a la tarea de comprender la presencia y distribución de valores faltantes. Por otro lado, los problemas de integración de datos pueden surgir cuando se combinan diferentes fuentes de información. Las discrepancias en la estructura o el formato de los datos pueden dificultar la integración completa, lo que a su vez puede generar valores faltantes en ciertas áreas de nuestros registros de tickets.
2. **Estrategia de manejo de valores faltantes:** La estrategia de manejo de valores faltantes que implementamos se diseñó para abordar la complejidad de los datos incompletos de manera eficiente y precisa. Comenzamos este proceso calculando el porcentaje de valores faltantes para cada ticket individualmente, en lugar de simplemente observar la completitud de las columnas. Este enfoque nos da una visión más detallada de la integridad de nuestros datos, ya que nos permite evaluar la cantidad de información ausente en cada entrada de la base de datos de tickets. Al calcular el porcentaje de valores faltantes para cada ticket, podemos identificar patrones específicos de ausencia de datos, como aquellos tickets que carecen de información en múltiples campos o aquellos que están casi completos pero tienen uno o dos campos importantes faltantes. Esta información detallada nos ayuda a comprender mejor la naturaleza y la extensión de los datos faltantes en nuestro conjunto de datos, lo que a su vez nos permite tomar decisiones más informadas sobre cómo abordar estos vacíos. Además, al centrarnos en los tickets individualmente, podemos adaptar nuestras estrategias de manejo de valores faltantes para abordar las necesidades específicas de cada entrada de la base de datos.
3. **Visualización de Valores Faltantes:** Es una parte fundamental del proceso de análisis de datos, ya que nos ayuda a comprender la distribución y la magnitud de los vacíos en nuestro conjunto de datos. Después de calcular los porcentajes de valores faltantes para cada ticket, buscamos representar esta información de manera clara y efectiva para facilitar su interpretación. Optamos por utilizar un gráfico de barras para visualizar los porcentajes de valores faltantes en cada ticket. Este tipo de gráfico nos permite comparar fácilmente la cantidad de valores faltantes entre diferentes tickets y proporciona una representación visual intuitiva de la distribución de los vacíos en nuestros datos. Cada barra en el gráfico representa un ticket específico, y la altura de la barra indica el porcentaje de valores faltantes para ese ticket en particular. Las barras más altas indican un mayor porcentaje de valores faltantes en un ticket, mientras que las barras más cortas representan tickets con menos valores faltantes. Al observar este gráfico, podemos identificar fácilmente aquellos tickets que presentan una alta proporción de valores faltantes. Estos tickets pueden señalar áreas problemáticas en nuestros datos, como transacciones incompletas o campos que no se completaron correctamente durante la recopilación de datos.
4. **Interpretación Detallada:** Al profundizar en la interpretación de la visualización de valores faltantes, nos adentramos en un análisis detallado para comprender las implicaciones de estos vacíos en nuestros datos. Al identificar los tickets con la mayor proporción de valores faltantes, dirigimos nuestra atención hacia estos casos específicos para entender mejor las posibles razones detrás de la falta de datos. Podríamos encontrarnos con varias situaciones al examinar estos tickets con una alta proporción de valores faltantes. Esto podría indicar problemas en los sistemas de registro de ventas o puntos de venta, lo que requeriría una investigación adicional para comprender y corregir estos errores. Además, podríamos encontrarnos con tickets que tienen valores faltantes debido a errores en la entrada de datos.

Cuadro 2: (#tab:tab:tabla)Tabla 2: Ejemplo de Muestra

variables	descripcion	elemento
Supermercado	Id del supermercado	MERCADONA, S.A A-46103834
Direccion	Dirección	C/ MENORCA (C.C. AQUA) 19
Codigo Postal	Código Postal del supermercado	46023
Ciudad	Ciudad	Valencia
Telefono	Telefono del supermercado	963319378
Factura	Número de factura	3075010680549
Productos	Productos adquiridos	PATATA
Metodo de Pago	Método de Pago	TARJETA BANCARIA
Número de Cliente	Id del cliente	077763746
Cantidad	Cantidad de productos	1
Precio	Precio del producto	6,49
Importe	Importe final	60,47
Fecha	Fecha de la compra	2023/12/18
Hora	Hora de la compra	18H 56M 0S

o inconsistencias en la forma en que se registraron ciertas transacciones. En tales casos, sería necesario revisar los procedimientos de entrada de datos y garantizar una mayor precisión y consistencia en la recopilación de información para evitar futuros problemas de integridad de datos. Al realizar esta interpretación detallada, no solo identificamos áreas problemáticas en nuestros datos, sino que también obtenemos información valiosa sobre posibles mejoras en los procesos de recopilación, almacenamiento y gestión de datos. Esto nos permite tomar medidas de corrección y prevención para garantizar la calidad y la integridad de nuestros datos a largo plazo.

5. Acciones Sigüientes: Basándonos en nuestra comprensión de la distribución de valores faltantes, pudimos tomar decisiones sobre las acciones sigüientes. Estas acciones podrían incluir la imputación de datos para ciertos tickets, la eliminación de transacciones problemáticas o la recopilación adicional de información para completar los registros. Además, esta exploración nos permitió establecer prioridades y desarrollar estrategias efectivas para garantizar la integridad y calidad de nuestros datos antes de continuar con análisis más avanzados o la implementación de modelos predictivos.

## 8 VISUALIZACIÓN OUTLIERS

La detección de anomalías y outliers en los tickets de Mercadona implica identificar transacciones o patrones de compra que se desvían significativamente de lo que sería considerado normal o típico.

1. Preprocesamiento de datos: Antes de detectar anomalías y outliers, es fundamental realizar un preprocesamiento de los datos de los tickets de Mercadona. Esto implica limpiar los datos, estructurarlos adecuadamente y realizar transformaciones necesarias para prepararlos para el análisis. Algunas técnicas comunes incluyen la extracción de información relevante de los tickets, como productos comprados, precios, fechas y horas de compra.
2. Identificación de características relevantes: Para detectar anomalías y outliers, es importante identificar las características relevantes de los tickets que se utilizarán en el análisis. Estas características pueden incluir el total gastado en una transacción, la cantidad de productos comprados, la frecuencia de compra de ciertos productos o cualquier otra información relevante que pueda indicar comportamientos inusuales.

3. Selección de métodos de detección: Existen varios métodos y técnicas para detectar anomalías y outliers en los datos de los tickets de Mercadona. Algunos de los métodos comunes incluyen:
  - Estadísticas descriptivas: Calculando medidas estadísticas como la media, la mediana, la desviación estándar y la varianza para identificar valores extremadamente altos o bajos.
  - Gráficos de caja (boxplots): Visualizando la distribución de las características relevantes mediante gráficos de caja para identificar valores atípicos.
  - Métodos basados en umbrales: Estableciendo umbrales o límites específicos para las características y identificando aquellos puntos que caen fuera de estos límites como anomalías.
  - Algoritmos de aprendizaje automático: Utilizando algoritmos de detección de anomalías supervisados o no supervisados, como Isolation Forest, Local Outlier Factor (LOF) o One-Class SVM, para identificar patrones inusuales en los datos.
4. Evaluación y validación de resultados: Una vez aplicados los métodos de detección de anomalías y outliers, es importante evaluar y validar los resultados obtenidos. Esto puede implicar la comparación de las anomalías detectadas con información adicional o la retroalimentación de expertos en el dominio para determinar la validez de las detecciones.
5. Interpretación de resultados: Finalmente, se interpreta y se analiza la naturaleza de las anomalías y outliers detectados en los tickets de Mercadona. Esto puede implicar identificar posibles causas detrás de estos comportamientos inusuales, como errores de registro, fraudes, promociones especiales o cambios en el comportamiento del cliente.

Primero, necesitamos realizar un análisis para detectar los outliers y las anomalías. Una forma muy común de identificarlos es calculando estadísticas descriptivas como el rango intercuartílico (IQR) y luego definir un criterio para identificar valores que están por encima o por debajo de ciertos umbrales basados en el IQR.

Podemos afirmar que no existe ningún outlier en los precios totales, sin embargo si hacemos el estudio producto a producto podemos observar varios outliers muy por encima de los percentiles.

## 8.1 Análisis univariante.

Diferenciaremos el análisis entre las variables de tipo numérico y las de tipo categórico presentes, ya que ciertos estadísticos descriptivos carecen de significado en las pertenecientes al último tipo.

Para variables de tipo numérico (cuantitativas): Establecemos estadísticos como el máximo, mínimo, media, mediana, quantiles, sesgo, curtosis tipo distribución. Por otra parte, dichos estadísticos se pueden representar en gráficos como: Boxplots, histogramas, q-q plot.

Para variables de tipo categórico (cualitativas): Establecemos estadísticos como tablas de frecuencias, porcentajes. . . Para representar estos estadísticos, lo hacemos en gráficos como sectores, barras, etc.

-*Variables de tipo numérico.* Hemos desarrollado una función para calcular estadísticas descriptivas de las variables numéricas clave. Estas estadísticas se presentan en la Tabla que haremos a continuación, que incluye la media, la mediana, la desviación estándar, la asimetría y la curtosis para cada variable numérica relevante en nuestro conjunto de datos:

Tabla 3: ESTADÍSTICOS DE CARACTERÍSTICAS DE LOS TICKETS DE MERCADONA.

VARIABLE	MEDIA	MEDIANA	DESVEST	ASIMETRÍA	CURTOSIS
Total productos	23.56	24	5.76	-0.03	-0.12
Total de clientes	4.78	5	1.92	0.67	1.31



VARIABLE	MEDIA	MEDIANA	DESVEST	ASIMETRÍA	CURTOSIS
Total de compras	6.91	7	2.10	-0.42	-0.19
Total de importe	62.05	61	15.24	0.21	-0.49

Hemos podido comprobar que la media de gasto de los tickets es de 62.05 euros.

*-Variables tipo categórico.*

En 3 vemos como el producto más comprado, con diferencia, es el “queso lonchas cabra”, seguido por “brotes tiernos maxi” y “queso fresco cabra” con la mitad de apariciones. Previamente se ha filtrado los productos para únicamente mostrar los productos que han sido comprados más de dos veces.

Como podemos ver en la tabla de todos los productos, el producto más comprado es queso lonchas cabra. Las categorías más compradas son frutas y verduras, pescado, hogar y limpieza, bebidas, drogueria y charcutería.

Los días que más se ha comprado son tanto miércoles como sábado, con un 23.5% de los tickets para cada día. Esto puede deberse a que los miércoles, como se trata de mitad de semana, los clientes van a comprar más frecuentemente para “llegar” a final de semana, y los sábados es el día de la semana no laborable con muchos supermercados abiertos, por tanto la gente puede ir a comprar con más frecuencia.

En esta matriz de gráficos tenemos, en la parte izquierda un gráfico donde se muestra el número de compras que se han hecho en cada mercadona. Podemos observar claramente como el del C.C AQUA es en el que más compras se hacen. Teniendo esto en cuenta, la empresa debería tomar medidas como reforzar el personal o agrandar los pedidos. Por lo contrario vemos como tanto el supermercado de la Calle Periodista Azzati, como el de la Calle Valencia son en los que menos se compra.

Por otra parte, en el gráfico de la derecha observamos el número de productos comprados en cada supermercado. Observamos como el del C.C AQUA es en el que más productos se compran lo que reafirma que se deberían mejorar los servicios en dicho establecimiento.

## 8.2 Análisis multivariante.

Ahora vamos a incidir en un análisis de las medias de cada variable establecida en la tabla:

En @ref(fig: Figura5) vemos como el Q-Q plot establece que la distribución de la media de los importes de compra en Mercadona está generalmente cerca de una distribución normal, pero podría haber algunas desviaciones en ciertos rangos de valores. Dado que la línea diagonal es una referencia de una distribución normal estándar, si los puntos caen cerca de esta diagonal se ajustan bien a una distribución normal. Aquí, parece que la mayoría de los puntos están cerca de la línea, lo que indica que la distribución de los importes de compra en Mercadona se aproxima a una distribución normal.

## 9 RESULTADOS.

Los resultados obtenidos de este análisis ofrecen una visión detallada de los comportamientos de compra en Mercadona. Observamos patrones claros en cuanto a los hábitos de gasto, con ciertos productos destacándose como favoritos entre los clientes. Además, identificamos preferencias en cuanto al horario de compra, con picos de actividad en momentos específicos del día. Asimismo, el análisis revela las ubicaciones más frecuentadas por los clientes, lo que proporciona información útil sobre la distribución geográfica de la clientela de Mercadona.

## 10 CONCLUSIONES.

Basándonos en los resultados obtenidos, podemos extraer varias conclusiones significativas. En primer lugar, Mercadona puede aprovechar estos insights para optimizar su estrategia de productos, asegurando que los

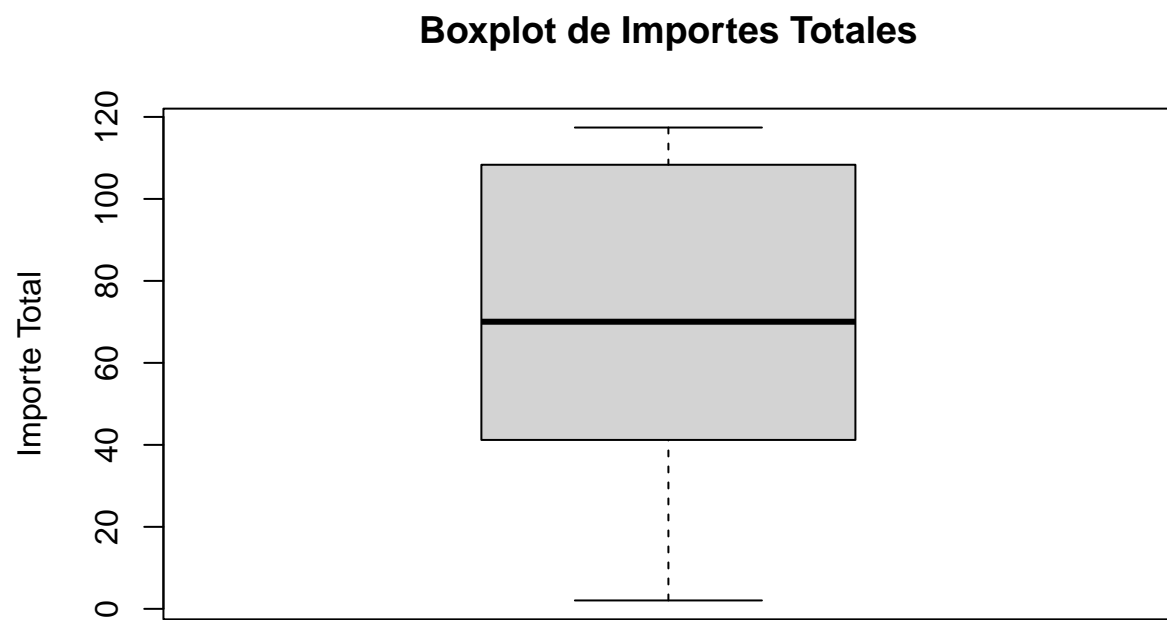


Figura 2: Figura2 : Outliers

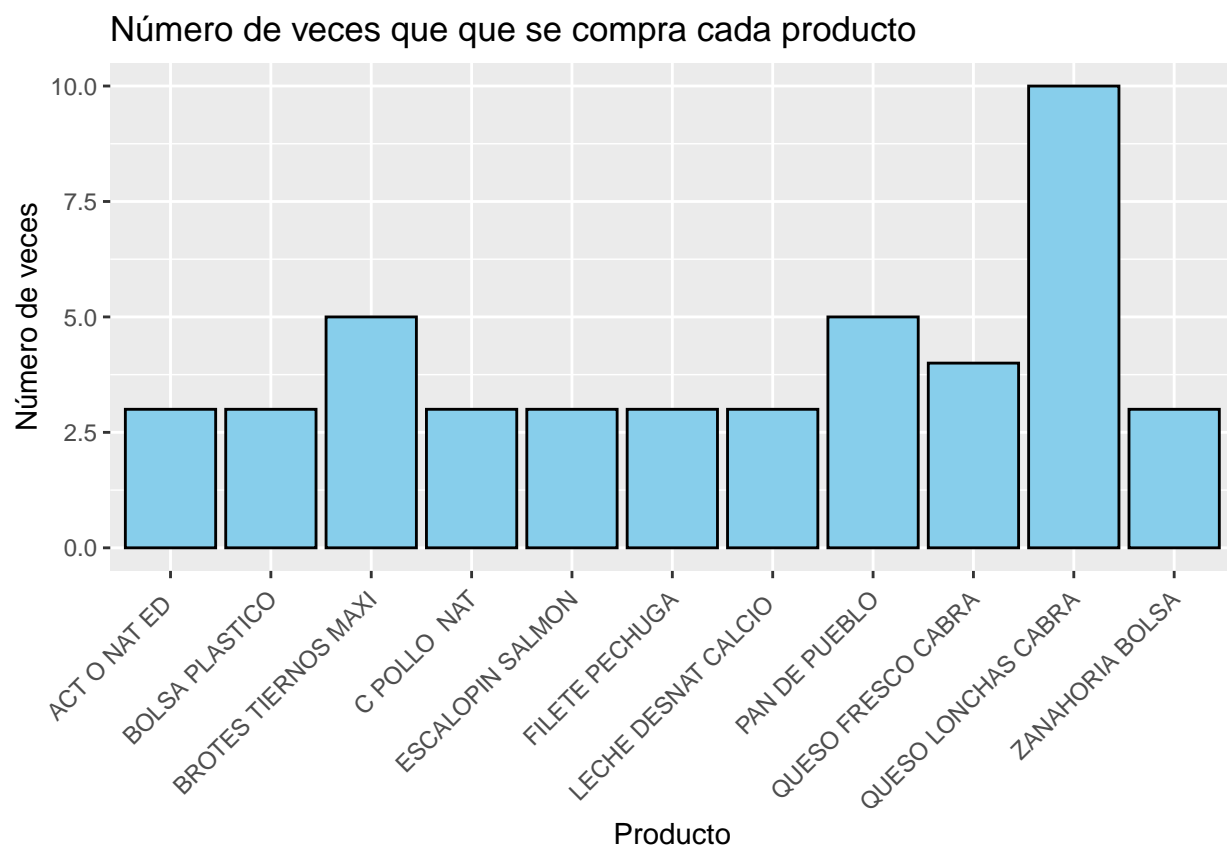


Figura 3: Frecuencia por producto

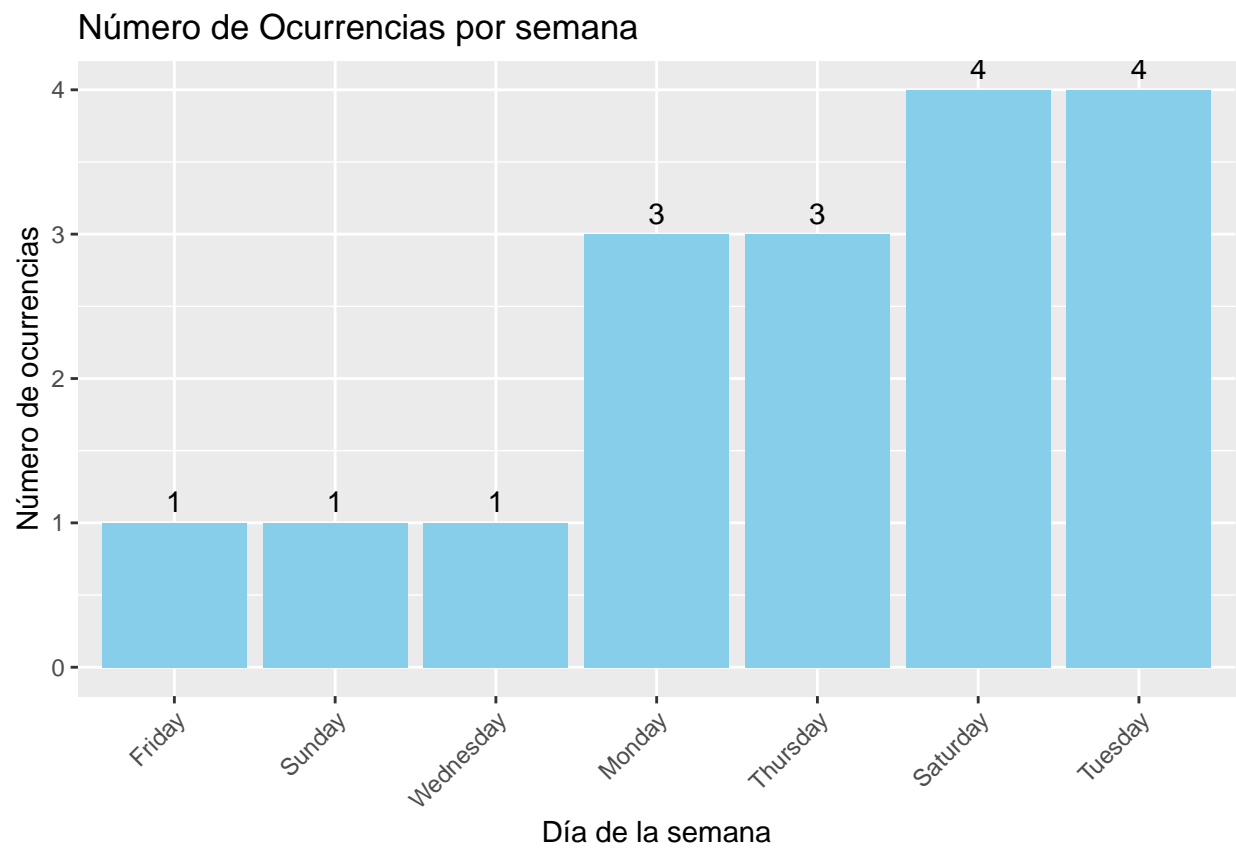


Figura 4: Figura 3: Ocurrencias por semana

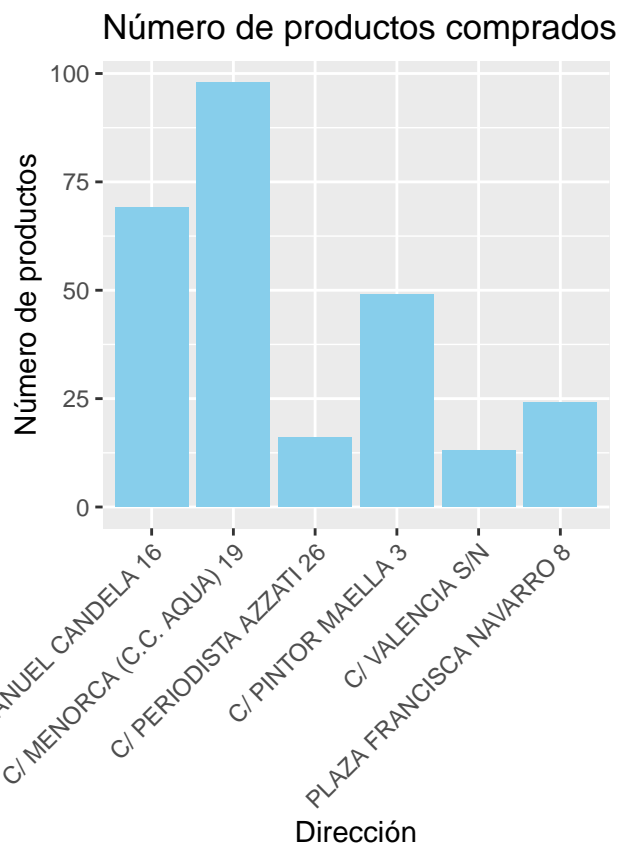
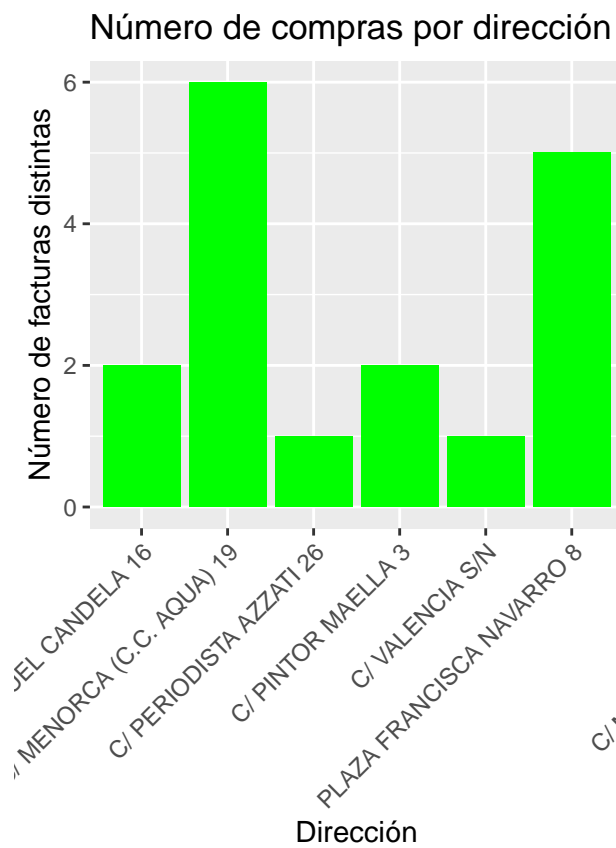


Figura 5: Figura 4: N° Compras

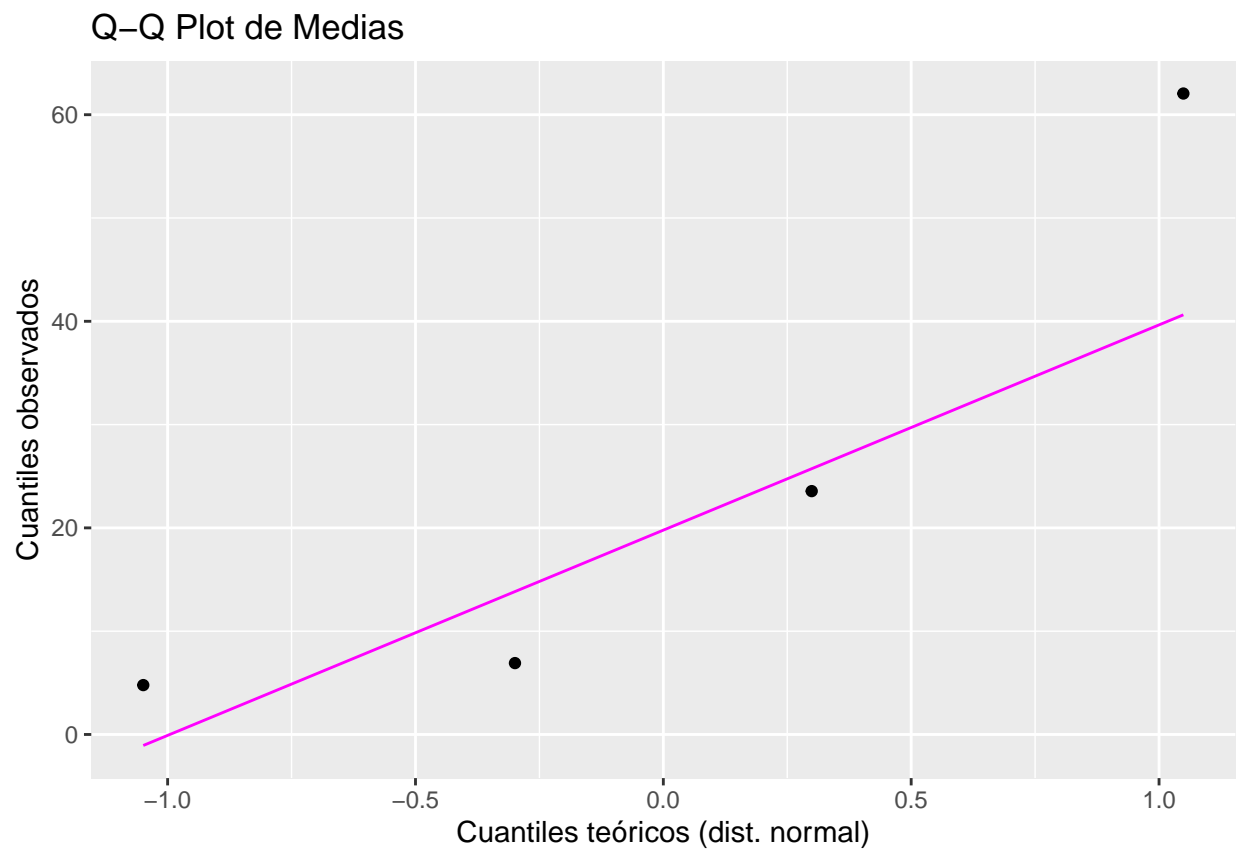


Figura 6: Figura 5: Medias

productos más populares estén siempre disponibles y promocionando aquellos que podrían tener un menor rendimiento. Además, el conocimiento de los horarios de mayor actividad puede ayudar a la gestión de inventario y la asignación de personal en las tiendas. Por último, el análisis de las ubicaciones más concurridas puede respaldar decisiones relacionadas con la expansión de tiendas o la reubicación de las existentes para maximizar su alcance y conveniencia para los clientes.