

ProyectoTD2024

Tratamiento de Datos. Grado en Ciencia de Datos- UV

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

30/03/2024

Índice

1	ABSTRACT	1
2	Introducción	2
2.1	Instalación automática de paquetes	2
2.2	Importación Datos	2
3	Preguntas	3
3.1	¿cuáles son los productos más comprados?	3
3.2	¿A qué hora se realizan más compras?	4
3.3	¿Cual es el valor medio de ingresos por cada tipo de porcentaje del Impuesto al Valor Añadido (IVA)?	7
3.4	¿Relación productos más y menos vendidos y su precio?	7

1 ABSTRACT

Los patrones de compra en la cadena de supermercados Mercadona en España se investigan en este proyecto a través de análisis de tickets de compra realizados entre 2023 y 2024. Se automatizó la preparación de datos y se utilizaron librerías especializadas para la manipulación y visualización de datos, como dplyr, lubridate, tidyverse y ggplot2, utilizando el entorno de programación R.

El objetivo del estudio es determinar las tendencias de consumo, los productos preferidos y las fechas de compra, así como evaluar los métodos de pago.

El código R inicial establece un flujo de trabajo limpio y centrado en la visualización eliminando mensajes innecesarios y automatizando el redondeo de resultados numéricos. Los datos se extraen de archivos PDF convertidos a texto, que están estructurados en un dataframe con variables detalladas como la ubicación del supermercado, la fecha y hora de compra y los detalles fiscales. Después de eso, se realiza un análisis exploratorio para identificar los productos más comprados y examinar otras métricas relevantes. Las imágenes ayudan a comprender los patrones de consumo y mejoran el análisis de los datos. Este método no solo ofrece una comprensión detallada de las preferencias de los clientes de Mercadona, sino que también sugiere métodos para optimizar la publicidad y la atención al cliente.

2 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.

2.1 Instalación automática de paquetes

Para la preparación de nuestro entorno de trabajo y la manipulación de los datos, hemos instalado y cargado varias librerías relevantes.

dplyr: Esta librería es fundamental para el procesamiento y manipulación de datos. Nos proporciona una amplia gama de funciones para realizar operaciones de filtrado, selección, agrupación y resumen de datos de manera eficiente.

lubridate: Utilizamos esta librería para trabajar con fechas y horas de forma más sencilla. Nos permite realizar operaciones como extracción de partes de fecha, cálculos de diferencia entre fechas, etc.

tidyverse: Tidyverse es un conjunto de paquetes de R diseñados para la ciencia de datos. Incluye varias librerías útiles, como **ggplot2** para visualización de datos, **tidyr** para manipulación de datos en formato “tidy” y **purrr** para programación funcional.

ggplot2: Es una librería ampliamente utilizada para la visualización de datos en R. Proporciona una sintaxis intuitiva y poderosa para crear una variedad de gráficos estadísticos.

Estas librerías nos brindan las herramientas necesarias para llevar a cabo el análisis exploratorio de los datos de los tickets de supermercado, desde la carga y limpieza de los datos hasta la visualización y extracción de patrones significativos.

2.2 Importación Datos

Para este análisis hemos recibido los datos en forma de ticket con formato pdf que hemos tenido que pasar a txt.

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.

Una vez que hemos importado los datos a un dataframe, estamos listos para realizar análisis exploratorio de datos, visualizaciones y responder preguntas específicas sobre los patrones y tendencias presentes en los datos de los tickets de supermercado. Esta etapa generalmente implica la aplicación de diversas técnicas de análisis de datos, como agregación, filtrado, agrupación, y visualización de datos.

3 Preguntas

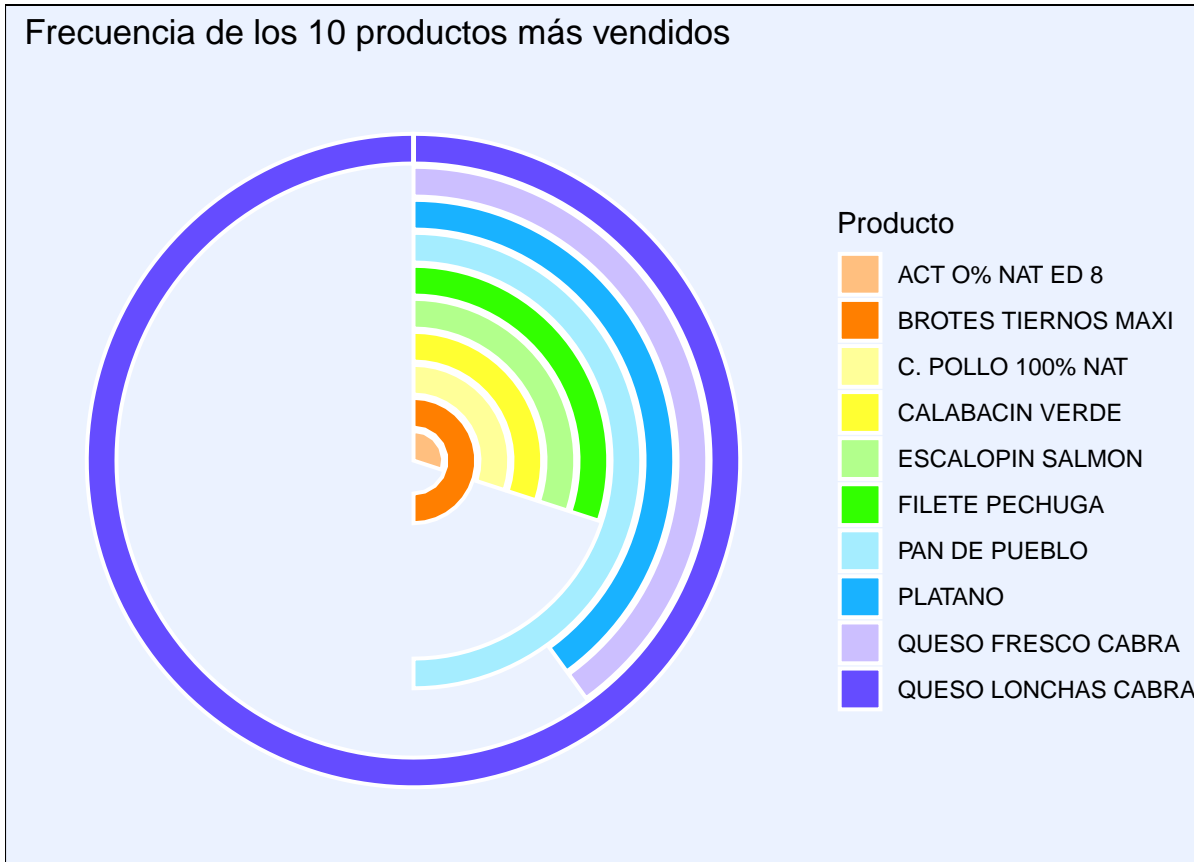
Una vez establecemos nuestros datos en forma de data frame, con etiquetas precisas y valores correctamente alineados, nos sumergimos en la exploración de posibles patrones dentro de las instancias y entre las características. En este punto, nos vemos inmersos en un proceso creativo, donde surgen preguntas innovadoras mientras exploramos qué aspectos de los datos capturan nuestra atención con mayor intensidad. Esto nos lleva a plantearnos las siguientes preguntas.

3.1 ¿cuáles son los productos más comprados?

La pregunta sobre los productos más comprados en un supermercado es fundamental por varias razones. En primer lugar, proporciona información valiosa sobre las preferencias y hábitos de compra de los clientes. Conocer qué productos son los más populares permite a los supermercados adaptar su inventario y estrategias de marketing para satisfacer mejor las necesidades de sus clientes, lo que puede conducir a un aumento de las ventas y la fidelidad del cliente.

Además, el análisis de los productos más comprados puede ayudar a identificar tendencias de consumo y patrones de compra a lo largo del tiempo. Esto puede ser especialmente útil para prever la demanda de ciertos productos, planificar promociones o descuentos, y optimizar la gestión de inventario.

Otra razón importante para estudiar los productos más comprados es que puede proporcionar información útil para la toma de decisiones estratégicas en cuanto a la colocación de productos en las estanterías y la disposición del espacio en la tienda. Saber qué productos tienen una alta demanda puede influir en la organización del layout del supermercado para mejorar la experiencia de compra del cliente y aumentar las ventas.



Observamos que los tres productos más comprados son el Queso en lonchas de cabra, los brotes tiernos, y el pan de pueblo. Este hallazgo revela tendencias interesantes en cuanto a las preferencias de los clientes. Por ejemplo, el hecho de que el queso en lonchas de cabra sea uno de los productos más comprados sugiere que existe una demanda significativa de productos lácteos específicos, posiblemente debido a una creciente preocupación por la calidad y la procedencia de los alimentos.

Por otro lado, la popularidad de los brotes tiernos podría reflejar una mayor conciencia sobre la alimentación saludable y la inclusión de ingredientes frescos y nutritivos en la dieta. Este hallazgo podría ser útil para los supermercados en términos de destacar y promocionar productos frescos y saludables en sus estanterías.

El pan de pueblo, como uno de los productos más comprados, sugiere una preferencia por productos de panadería tradicionales y de calidad. Esto podría indicar una demanda de productos horneados frescos y artesanales, lo que podría influir en las decisiones de los supermercados sobre su oferta de pan y productos de panadería.

En conjunto, estos hallazgos subrayan la importancia de entender las preferencias y tendencias de compra de los clientes para adaptar la oferta de productos y mejorar la experiencia de compra en el supermercado. Además, proporcionan insights valiosos para estrategias de marketing, promoción y gestión de inventario que podrían ayudar a aumentar las ventas y la satisfacción del cliente.

3.2 ¿A qué hora se realizan más compras?

Esta pregunta es fundamental en el análisis de datos de un supermercado por varias razones. En primer lugar, entender los patrones de compra horaria proporciona información valiosa sobre el comportamiento del consumidor. Conocer las horas pico de actividad en el supermercado permite a los gerentes y propietarios optimizar la distribución de personal y recursos durante esos períodos, asegurando así una experiencia de compra más fluida y satisfactoria para los clientes.

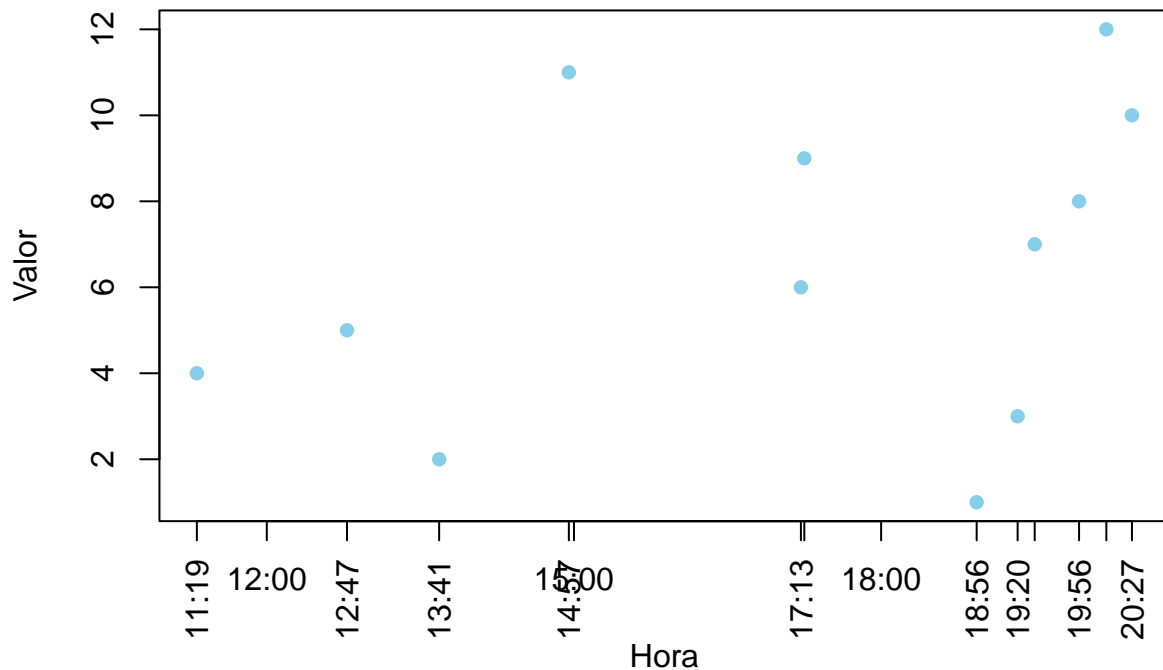
Además, comprender los horarios de mayor afluencia también es crucial para la planificación del inventario y la gestión de la cadena de suministro. Saber cuándo se esperan más compras ayuda a los minoristas a ajustar sus niveles de stock para satisfacer la demanda esperada y evitar escaseces o excedentes que puedan afectar negativamente a las ventas y la rentabilidad.

Aparte, también nos puede servir para ajustar el número de personal que necesitamos en un determinado momento.

```
[1] "14:22" "16:48" "10:20" "08:25" "18:56" "13:41"
```



Gráfico de Dispersión con Horas

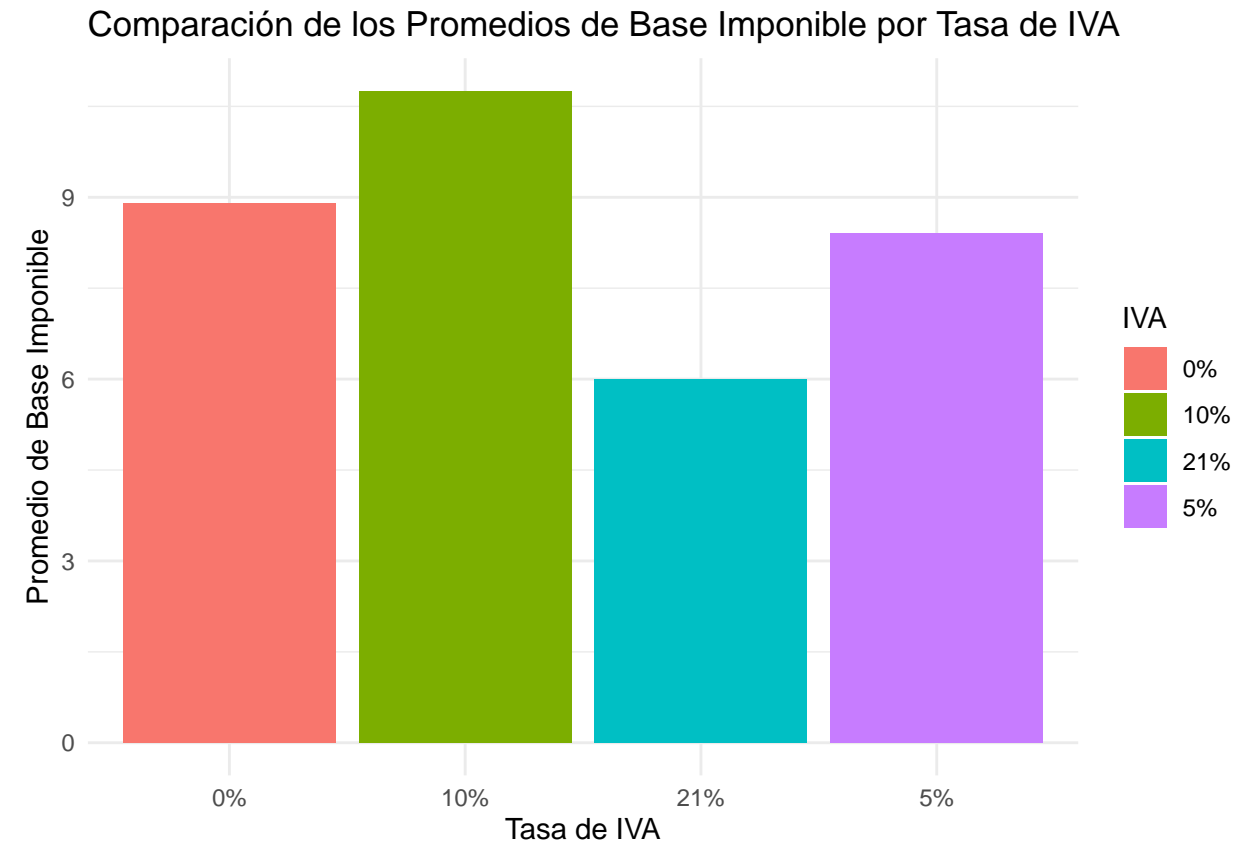


Estas observaciones nos brindan una visión clara de los momentos críticos del día en los que se registra un mayor flujo de clientes en el supermercado. Estas horas crítica son entre las 10:00 y las 12:00 , a las 17:00 y sobretodo a las 14:00 del medio día, que es el momento en el que se observa más personal. Aprovechar esta información estratégicamente puede tener un impacto significativo en la experiencia general del cliente y en la eficiencia operativa del establecimiento.

Al aumentar el personal durante las horas pico, podemos garantizar un servicio más ágil y personalizado, lo que contribuye a reducir los tiempos de espera en las cajas y mejora la atención al cliente. Esto no solo incrementa la satisfacción del cliente, sino que también puede influir en la fidelización y recompra.

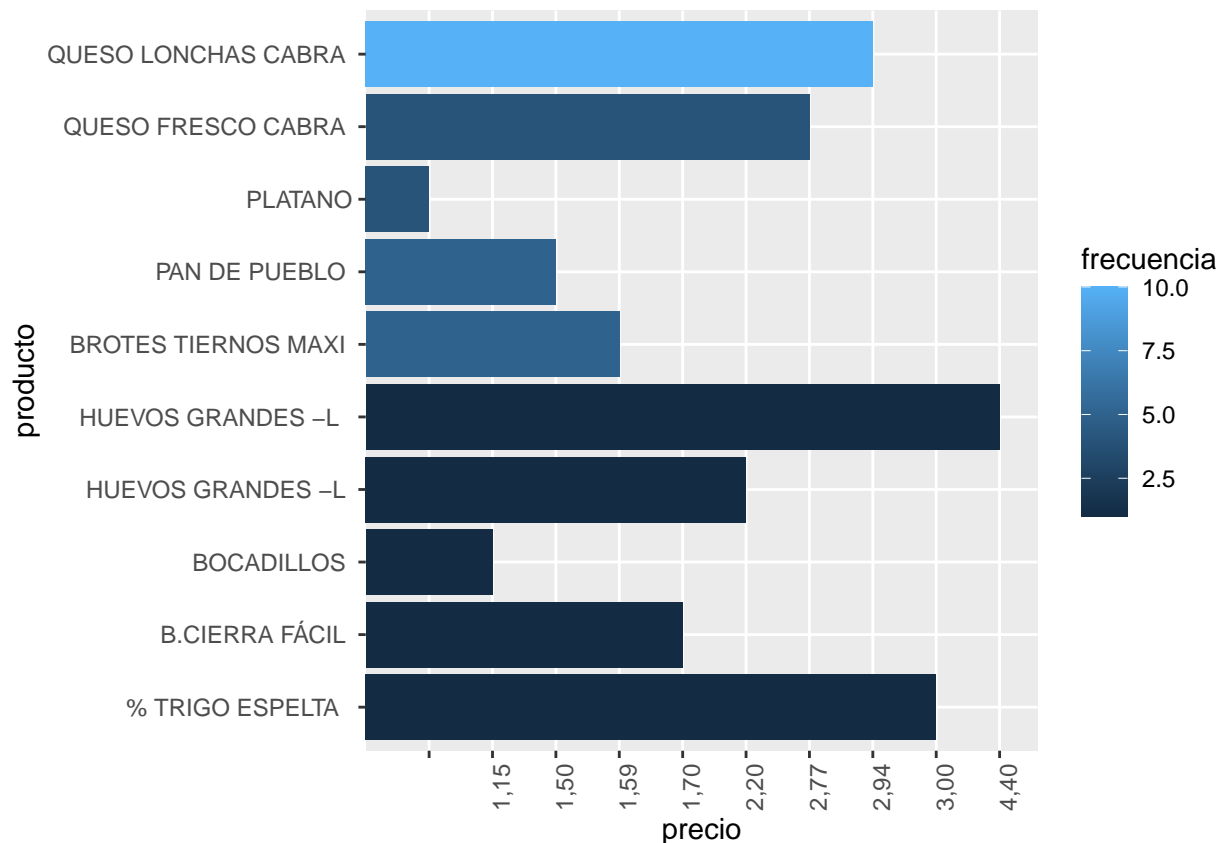
Por otro lado, esta información también puede ser útil para optimizar las estrategias de marketing y promoción. Al dirigir campañas publicitarias o promociones especiales específicamente durante estas horas de alto tráfico, podemos maximizar su impacto y efectividad, ya que llegarán a un mayor número de clientes potenciales.

3.3 ¿Cual es el valor medio de ingresos por cada tipo de porcentaje del Impuesto al Valor Añadido (IVA)?



3.4 ¿Relación productos más y menos vendidos y su precio?

La pregunta sobre la relación entre los productos más y menos vendidos y su precio es fundamental por varias razones. En primer lugar, nos permite evaluar la percepción de valor de los clientes en relación con los productos ofrecidos. Al comparar los productos más populares con los menos vendidos y sus respectivos precios, podemos identificar si existe una correlación entre la demanda y el precio. Esta información nos ayuda a determinar si los precios actuales están alineados con las expectativas del mercado y si existen oportunidades para ajustar estratégicamente los precios para maximizar las ventas y la rentabilidad.



Basándonos en los resultados obtenidos, observamos una clara diferencia en la demanda de productos dentro de nuestro supermercado. El queso lonchas de cabra destaca como el producto más vendido, con un precio de 2,94 euros. Este hallazgo sugiere que este producto es altamente popular entre nuestros clientes y que su precio actual puede estar alineado con su valor percibido.

Por otro lado, identificamos varios productos que tienen una cantidad media de ventas, como el queso fresco de cabra, el plátano, el pan de pueblo y los brotes tiernos maxi. Estos productos, aunque no alcanzan el nivel de ventas del queso lonchas de cabra, mantienen una demanda constante, lo que sugiere que tienen un atractivo para la mayoría de clientes. Esto es coherente con el tipo de productos puesto que se tratan de productos de alimentación muy importantes.

Sin embargo, resulta preocupante observar que algunos productos, como los huevos grandes, los bocadillos, el cierra fácil y el trigo espelta, son los menos vendidos. Estos hallazgos pueden indicar que estos productos no están cumpliendo con las expectativas de los clientes en términos de precio o calidad, lo que podría estar afectando su demanda.

Observamos dos cosas diferente, por una parte parece que los huevos grandes y el trigo de espelta podrían estar sufriendo de una baja frecuencia de compras debido a un precio excesivamente alto. Sin embargo po otro lado, con los bocadillos y el cierre facil el problema no parece tanto de precio sino de demanda e interés de los consumidores.

Con esta información diseñar estrategias específicas para abordar las distintas situaciones identificadas. En el caso de los huevos grandes y el trigo de espelta, donde parece que el precio puede estar afectando negativamente su demanda, podríamos considerar ajustes en los precios para alinearlos mejor con las expectativas de los clientes y mejorar su competitividad en el mercado.

Por otro lado, para los productos como los bocadillos y el cierre fácil, donde el problema parece ser más relacionado con la falta de interés o demanda por parte de los consumidores, podríamos explorar otras estrategias. Esto podría incluir realizar encuestas o estudios de mercado para comprender mejor las preferencias

y necesidades de los clientes en relación con estos productos. Con esta información, podríamos ajustar el surtido de productos, mejorar su calidad, desarrollar campañas de marketing específicas o incluso explorar la posibilidad de reemplazar estos productos por otros con mayor demanda y potencial de ventas.