



DAJ 10

2024

Proyecto de Evaluación:

"Análisis de datos para optimización de ventas en e-commerce"

Presentado por:

Barrientos, Sandra k00002692

Herrera, Elizabeth k00002712

Martel, Rodrigo k00002721

Rivera, Edwin k00002715

ÍNDICE

| | |
|---|-----------|
| Resumen ejecutivo | 1 |
| Introducción | 2 |
| Objetivos del proyecto | 2 |
| Alcance del proyecto | 2 |
| Metodología | 3 |
| Diseño de la base de datos | 3 |
| Modelo Snowflake | 3 |
| Utilidad de este modelo en el proceso | 3 |
| Lucidchart | 3 |
| PostgreSQL como DBMS | 4 |
| Puntos considerados para trabajar con PostgreSQL | 4 |
| Recopilación, preparación y manipulación de datos | 5 |
| Importación de datos desde archivos CSV | 5 |
| Validación de datos | 5 |
| Extracción y manipulación de datos | 5 |
| Análisis exploratorio de datos | 6 |
| Jupyter Notebook | 6 |
| PowerBI | 6 |
| Visual Studio | 7 |
| Librerías como herramientas utilizadas | 7 |
| Resultados principales | 8 |
| KPIs clave | 8 |
| Tendencias de ventas | 8 |
| Análisis de clientes | 8 |
| Análisis de productos | 8 |
| Conclusiones y recomendaciones | 8 |
| Anexos | 10 |

Resumen ejecutivo

Hemos realizado un análisis detallado de las ventas de una empresa de comercio electrónico para ayudarles a mejorar su rendimiento comercial. Como equipo de analistas de datos, nuestro objetivo era descubrir información útil que pueda ayudar a la empresa a tomar decisiones inteligentes.

Utilizamos herramientas tecnológicas para:

- Organizar los datos de manera eficiente
- Examinar patrones de ventas
- Identificar oportunidades de crecimiento

Principales hallazgos:

- Algunos productos se venden más en ciertas regiones
- Las ventas aumentan en períodos específicos del año
- Algunos clientes compran con más frecuencia que otros

Nuestras recomendaciones buscan:

- Optimizar el inventario de productos
- Ajustar los precios de manera estratégica
- Crear programas para fidelizar clientes
- Diseñar campañas de marketing más efectivas

El objetivo final es ayudar a la empresa a incrementar sus ventas y mejorar su desempeño comercial de manera sostenible.

Introducción

Objetivos del proyecto

El objetivo principal de este proyecto es analizar los datos de ventas de un cliente especializado en comercio electrónico para optimizar sus estrategias de ventas. Utilizando diversas técnicas de análisis de datos, buscamos identificar patrones y tendencias clave que permitan proponer recomendaciones basadas en datos para mejorar las ventas del cliente.

Además, el proyecto tiene como objetivo aplicar los conocimientos adquiridos durante el bootcamp en un entorno real, abordando todas las fases del análisis de datos, desde el diseño de la base de datos hasta la creación de un dashboard interactivo.

Alcance del proyecto

El alcance del proyecto abarca las siguientes fases:

1. **Diseño de base de datos:** creación de un diagrama entidad-relación y su implementación en un DBMS.
2. **Extracción y manipulación de datos:** importación de datos desde archivos CSV y ejecución de consultas SQL para la extracción de información relevante.
3. **Análisis exploratorio de datos:** realización de análisis estadístico descriptivo, identificación de patrones y tendencias y generación de visualizaciones.
4. **Creación de dashboard:** desarrollo de un dashboard interactivo en PowerBI para presentar los hallazgos y las métricas clave.
5. **Reporte y presentación:** preparación de un informe escrito detallado y una presentación clara y visual de los hallazgos y recomendaciones.
6. **Consideraciones adicionales:** aplicación de buenas prácticas de codificación, optimización del código y consultas SQL y provisión de código comentado.

Metodología

Diseño de la base de datos

Modelo Snowflake

Se seleccionó el modelo Snowflake para trabajar con la base de datos en estudio. Este es un modelo variante del modelo estrella, que permite ser herramienta visual para diseñar, entender y optimizar las bases de datos.

En el modelo Snowflake las tablas de dimensiones (como venta y producto en este caso) se dividen en subtablas adicionales para normalizarlas. La estructura resultante tiene un aspecto de “copo de nieve” debido a las múltiples ramas que surgen las tablas normalizadas.

Utilidad de este modelo en el proceso

- Representar relaciones complejas: muestra las conexiones entre las tablas de hechos y las múltiples capas de tablas de dimensión normalizadas.
- Optimización del almacenamiento: ayuda a visualizar cómo la normalización reduce la redundancia al descomponer las dimensiones.
- Facilitar la consulta de datos: permite identificar cómo se deben unir las tablas para consultas analíticas en el almacén de datos.
- Diseño eficiente del esquema: ayuda a identificar áreas donde el modelo puede ser optimizado o simplificado sin perder integridad de los datos.
- Mejor comprensión y documentación: sirve como documentación visual para desarrolladores, analistas y administradores de bases de datos.

Lucidchart

Es una herramienta en línea ampliamente versátil y utilizada para crear diagramas, flujos de trabajo y representaciones visuales de procesos y sistemas.

Se seleccionó esta herramienta por varias razones:

- ❖ Interfaz intuitiva y fácil de usar: facilita la creación de diagramas, es amigable y accesible para usuarios tanto principiantes como avanzados.

- ❖ Variedad de plantillas y formas: ofrece una amplia biblioteca de plantillas prediseñadas y elementos personalizables para diferentes tipos de diagramas, que se necesitan para modelar bases de datos.
- ❖ Colaboración en tiempo real: permite que varios usuarios trabajen en el mismo proyecto de forma simultánea, logrando aportaciones de los integrantes del equipo.
- ❖ Integración con otras herramientas: es posible compartir el proyecto mediante otras plataformas y aplicaciones que son utilizadas en las empresas. En este caso, se utilizaron las aplicaciones colaborativas de Google (Drive) y Office 365.
- ❖ Accesibilidad multiplataforma: al ser una herramienta basada en la nube se puede acceder al proyecto desde cualquier dispositivo con conexión a internet.
- ❖ Exportación y compatibilidad: permite exportar diagramas en múltiples formatos, como PDF, PNG, SVG y otros. Esto facilita la integración de los diagramas en presentaciones, informes y documentación técnica.

PostgreSQL como DBMS

PostgreSQL es un Sistema de Gestión de Bases de Datos relacional (DBMS, por sus siglas en inglés) de código abierto, reconocido por su potencia, flexibilidad y características avanzadas. Adicionalmente, se seleccionó por ser un DBMS con el que se trabajó a lo largo del bootcamp.

Puntos considerados para trabajar con PostgreSQL

- Es completamente gratuito y no requiere licencias costosas, lo que lo hace ideal para proyectos de cualquier escala.
- Cumple con el estándar SQL y admite características avanzadas, como subconsultas, funciones de ventana, joins complejos, etc.
- Maneja desde pequeñas bases de datos, como en este caso, hasta grandes volúmenes de datos de manera eficiente, permitiendo el fácil manejo si la información en la base de datos crece.
- Es compatible con arquitecturas de nube (como AWS, Google Cloud, Azure) y entornos locales.
- Es altamente extensible, lo que permite personalizarlo y adaptarlo a necesidades específicas, como es el caso de este proyecto.

- Permite crear funciones personalizadas en lenguajes como Python, Java, C, o PL/SQL.
- Ofrece seguridad robusta.

Desarrollo de Software (Visual Studio)

Esta herramienta nos ayudó a escribir código de manera más eficiente, como un asistente que nos sugiere la mejor forma de resolver problemas.

Recopilación, preparación y manipulación de datos

Importación de datos desde archivos CSV

La importación de los datos se realizó desde los archivos CSV provistos por el cliente, utilizando consultas SQL específicas para cargar la información en las tablas temporales creadas en PostgreSQL.

Un ejemplo de las consultas utilizadas es la siguiente:

```
-- importar el archivo clientes.csv
COPY tmp.tmp_cliente(cliente_id, nombre_cliente, email, telefono, direccion)
FROM 'C:/clientes.csv' DELIMITER ',' CSV HEADER;
```

En esta consulta, se está utilizando el comando `COPY` para importar los datos desde el archivo `clientes.csv` ubicado en la ruta especificada. El delimitador de los campos es una coma (,) y se incluye la opción `CSV HEADER` para indicar que la primera fila del archivo contiene los nombres de las columnas.

Validación de datos

Como parte del proceso de importación, se realizaron diversas validaciones para asegurar la integridad y la calidad de los datos. Algunas de estas validaciones incluyeron:

- Verificación de la existencia de valores nulos en campos obligatorios
- Confirmación de que los valores numéricos estén dentro de rangos esperados
- Revisión de la consistencia y unicidad de los identificadores clave

Extracción y manipulación de datos

Una vez importados y validados los datos, se procedió a la ejecución de consultas SQL para extraer y manipular la información relevante. Algunas de las operaciones realizadas incluyeron:

- Filtrado de datos: selección de registros que cumplieran con ciertos criterios específicos.
- Agregación de datos: cálculo de métricas como sumas, promedios y conteos.
- Uniones de tablas: combinación de datos provenientes de diferentes tablas para obtener una visión integrada.

Estos pasos aseguraron que los datos estuvieran listos para el análisis exploratorio y la generación de visualizaciones en la siguiente fase del proyecto.

Análisis exploratorio de datos

Jupyter Notebook

Es una herramienta interactiva basada en la web que permite combinar código, texto, visualizaciones y resultados en un mismo documento. Para los integrantes del proyecto es de fácil acceso e ideal para experimentación, análisis y comunicación de datos.

Otras consideraciones para usar Jupyter Notebook:

- Interactividad: ejecuta y muestra resultados inmediatamente.
- Documentación integrada: incluye texto enriquecido (markdown) junto con el código.
- Visualizaciones: fácil integración con gráficos y bibliotecas como Matplotlib y Seaborn.
- Uso compartido: exporta notebooks como HTML o PDF para compartir ideas.
- Versatilidad: compatible con Python y otros lenguajes mediante kernels.

PowerBI

Es una herramienta de inteligencia de negocios desarrollada por Microsoft que permite analizar datos y crear informes interactivos y visualizaciones. Está diseñada para ayudar a empresas y usuarios a tomar decisiones informadas basadas en datos, es una solución poderosa, flexible y accesible para transformar datos en decisiones estratégicas.

Consideraciones para utilizar PowerBI:

- Amplia biblioteca de gráficos y visualizaciones interactivas.
- Personalización sencilla para adaptar informes a tus necesidades.
- Compatible con cientos de fuentes como Excel, SQL Server, SharePoint, Google Analytics, APIs, etc.
- Uso de DAX (Data Analysis eXpressions) para cálculos personalizados.
- Soporte para modelado avanzado de datos.
- Informes accesibles en la nube, escritorio y dispositivos móviles.
- Publicación y uso compartido de informes con equipos a través del servicio PowerBI o Microsoft Teams.
- Planes gratuitos y opciones de pago asequibles según las necesidades de los usuarios.

Visual Studio

Visual Studio es un entorno de desarrollo integrado (IDE, por sus siglas en inglés) de Microsoft, diseñado para crear, depurar y desplegar aplicaciones. Es ideal para desarrollar proyectos de software en varios lenguajes como C#, C++, Python, JavaScript, entre otros.

Puntos adicionales por los cuales se utilizó Visual Studio:

- Soporte para múltiples lenguajes y plataformas.
- Herramientas avanzadas para desarrollo y depuración.
- Integración con servicios en la nube y control de versiones.
- Amplia comunidad y recursos de aprendizaje.

Librerías como herramientas utilizadas

Para el proyecto, fue fundamental utilizar librerías para aprovechar al máximo las capacidades del lenguaje y mejorar la eficiencia del desarrollo.

Los puntos que ayudan al trabajar con librerías son:

- ✓ Productividad
- ✓ Ahorro de tiempo
- ✓ Código reutilizable
- ✓ Calidad y confiabilidad

- ✓ Facilidad de uso

Librerías estándar: incluidas en el lenguaje de programación (por ejemplo, math o datetime en Python, os).

Librerías externas: Pandas, Numpy o Matplotlib en Python, dotenv.

Librerías para áreas concretas como:

- Análisis de datos: Pandas
- Visualización: Matplotlib, Seaborn
- Bases de datos: SQLAlchemy

Resultados principales

KPIs clave

- Ventas totales: las ventas totales representan la suma de todas las ventas realizadas durante el período disponible y analizado. Este KPI proporciona una visión general del volumen de ventas y es fundamental para evaluar el rendimiento general.
- Número de transacciones: refleja la cantidad total de ventas realizadas, lo que ayuda a entender la actividad comercial y la frecuencia de las compras.

Tendencias de ventas

- Se analizaron las ventas a lo largo del tiempo, segmentadas por trimestre, utilizando la tabla calendario creada. Esto permitió identificar patrones estacionales y tendencias a lo largo del año.
- Las ventas segmentadas por región mostraron que la región Este lidera en términos de volumen de ventas.
- Las ventas segmentadas por categoría mostraron que Hogar lidera en términos de volumen de ventas.

Análisis de clientes

- Clientes con mayor número de compras: se identificaron los 8 clientes que más compras realizaron, tanto en términos de cantidad como de valor, proporcionando una visión de los clientes más frecuentes y valiosos en términos de ingresos, respectivamente.

- Clientes con menor número de compras: se identificaron también los clientes con menor actividad para identificar posibles oportunidades de mejora.

Análisis de productos

- Productos más vendidos: se identificaron los productos más vendidos por categoría y región, identificando los productos más populares entre los clientes.
- Productos menos vendidos: se analizaron los productos con menor rotación para ajustar estrategias de inventario.

Conclusiones y recomendaciones

El análisis permitió identificar patrones clave en las ventas y el comportamiento de los clientes. Las ventas totales proporcionan una visión clara del desempeño general, mientras que el análisis geográfico y de productos ofrece información detallada para la toma de decisiones.

Entre las recomendaciones que podemos ofrecer están:

Optimización de inventario

Basado en el análisis de las ventas por producto y región, se recomienda ajustar el inventario para asegurar la disponibilidad de productos de alta demanda y reducir el exceso de stock en productos con baja rotación.

Estrategias regionales

Las regiones con menores ventas pueden beneficiarse de campañas de marketing específicas.

Fidelización de clientes

Se deben implementar programas de fidelización para los clientes más frecuentes y estrategias para reactivar a los clientes con menos compras.

Tendencias de ventas

Las ventas mostraron un patrón estacional con picos en la segunda mitad del año. Esto sugiere que las estrategias de marketing y promoción deben ajustarse para aprovechar estos picos de demanda.

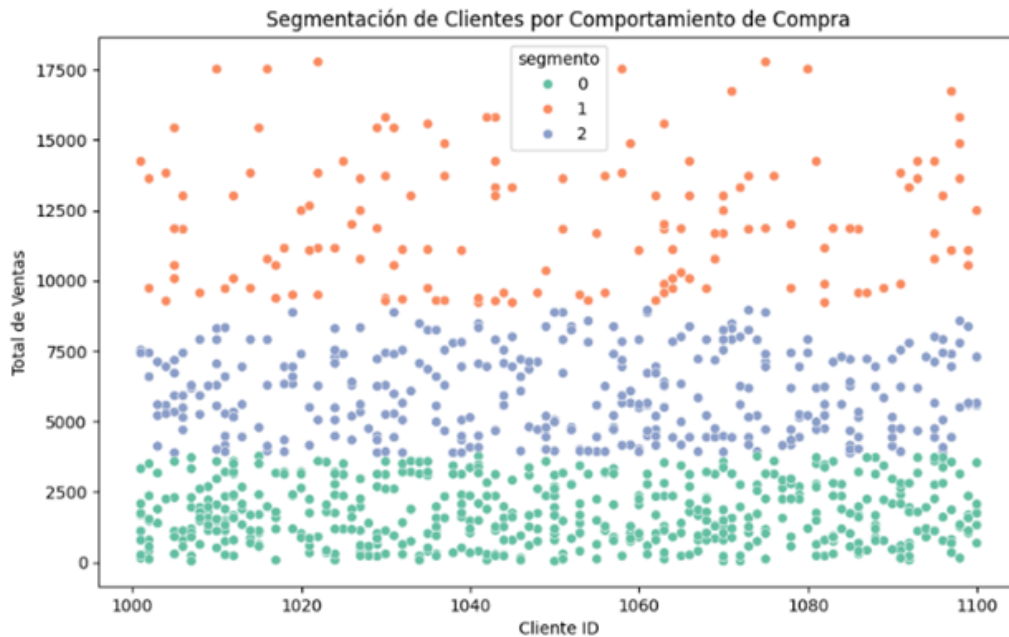
Distribución geográfica

La región Este lidera en términos de volumen de ventas, seguida por la región Norte. Esto indica una fuerte presencia comercial en esa región, lo que puede ser aprovechado para expandir operaciones o mejorar la distribución.

Anexos

" Análisis de Ventas y Segmentación de Clientes"

1. SEGMENTACIÓN DE CLIENTES POR COMPORTAMIENTO DE COMPRA



El análisis de segmentación divide a los clientes en tres grupos según su comportamiento de compra. A continuación, se detallan los hallazgos clave:

- **Cluster 0: Volumen Bajo, Alta Densidad**
 - Volumen de compras: Inferior al promedio.
 - Densidad: Mayor número de clientes concentrados.
 - Insight: Estos clientes realizan compras pequeñas y esporádicas, pero su volumen numérico puede representar un importante potencial de crecimiento si se los fideliza.
- **Cluster 1: Ticket Alto, Frecuencia Baja**
 - Valor transaccional: Superior al promedio.
 - Frecuencia: Compras poco frecuentes pero con un alto ticket promedio.
 - Insight: Este grupo es clave para el margen de ganancia debido al alto gasto por transacción. Sin embargo, la baja frecuencia sugiere que no son clientes regulares.
- **Cluster 2: Comportamiento Intermedio**
 - Frecuencia y ticket promedio: Moderados.
 - Potencial de conversión: Clientes con un comportamiento equilibrado que pueden evolucionar hacia más compras recurrentes o un ticket promedio más alto.
 - Insight: Este segmento es versátil y ofrece oportunidades para escalar su comportamiento hacia clusters de mayor rentabilidad o fidelización.

Recomendaciones

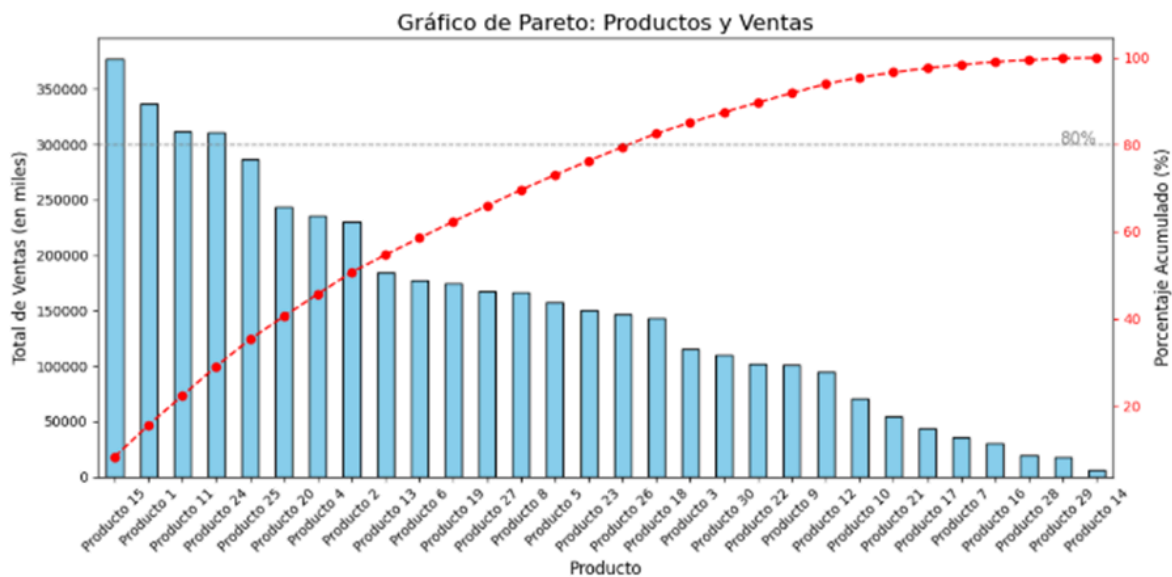
2. Cluster 0: Estrategias de Engagement

- a. **Promociones personalizadas:** Ofrecer descuentos o programas de lealtad para incentivar el volumen de compras.
 - b. **Campañas educativas:** Mostrar el valor del producto/servicio mediante contenido atractivo que despierte interés.
 - c. **Remarketing:** Implementar anuncios que refuercen la conexión emocional y las promociones, enfocándose en sus necesidades específicas.
- 3. **Cluster 1: Optimización del Alto Ticket**
 - a. **Experiencias premium:** Ofrecer beneficios exclusivos como membresías, envíos gratuitos o productos personalizados para fomentar lealtad.
 - b. **Upselling y cross-selling:** Presentar complementos o productos relacionados que maximicen el valor de cada transacción.
 - c. **Programas de retención:** Crear estrategias para aumentar la frecuencia de compra, como suscripciones o descuentos en compras repetitivas.
- 4. **Cluster 2: Conversión a Clientes Frecuentes**
 - a. **Incentivos recurrentes:** Ofrecer bonos por volumen o recompensas por alcanzar cierta frecuencia.
 - b. **Optimización del ticket promedio:** Utilizar técnicas de empaquetamiento o promociones para aumentar el gasto en cada compra.
 - c. **Segmentación dinámica:** Monitorear los comportamientos y ajustar estrategias según su evolución hacia otros clusters.

Factores a Considerar

- 5. **Datos y Métricas**
 - a. Monitorear constantemente la evolución de cada cluster mediante KPIs como frecuencia de compra, ticket promedio y valor del cliente a largo plazo.
 - b. Revisar si las estrategias implementadas generan migración de clientes entre clusters.
- 6. **Personalización**
 - a. La efectividad de las campañas dependerá de su capacidad para adaptarse a las necesidades específicas de cada segmento.
 - b. Evitar generalizar estrategias que no respondan a los patrones de compra detectados.
- 7. **Canales y Recursos**
 - a. Priorizar los canales digitales como email marketing, redes sociales y anuncios personalizados para llegar a los clusters más receptivos.
 - b. Asignar recursos equilibrados a cada cluster según su potencial de retorno.
- 8. **Factores Externos**
 - a. Considerar estacionalidades, cambios económicos y tendencias de mercado que puedan influir en el comportamiento de los clientes.
 - b. Realizar análisis competitivos para entender cómo otros negocios abordan segmentos similares.

2. ANÁLISIS DE PRODUCTOS - GRÁFICO DE PARETO



Distribución de Ventas por Producto:

El gráfico de Pareto muestra que una pequeña cantidad de productos generan la mayor parte de las ventas. Estos productos clave, que corresponden aproximadamente al 20% del total, contribuyen significativamente al 80% de las ventas. Los productos más destacados son los siguientes:

1. **Producto 15:** Más de 350,000 en ventas.
2. **Producto 1:** Más de 300,000 en ventas.
3. **Producto 11:** Más de 300,000 en ventas.
4. **Producto 24:** Más de 300,000 en ventas.
5. **Producto 25:** Más de 200,000 en ventas.
6. **Producto 20:** Más de 200,000 en ventas.
7. **Producto 4:** Más de 200,000 en ventas.
8. **Producto 2:** Más de 150,000 en ventas.
9. **Producto 13:** Más de 150,000 en ventas.

Análisis de la Curva de Pareto

La curva de Pareto sigue el principio 80/20, indicando que el 20% de los productos representan el 80% de las ventas totales. La curva muestra un aumento rápido en las ventas de los primeros productos, seguido por una disminución significativa, lo que señala que la mayoría de los ingresos provienen de unos pocos productos.

Recomendaciones

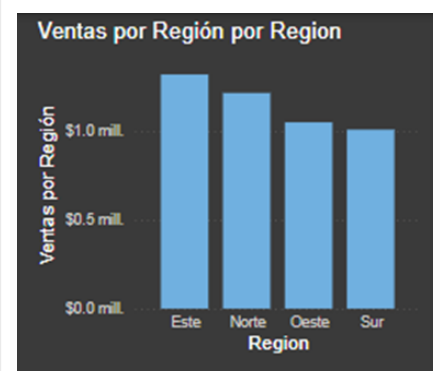
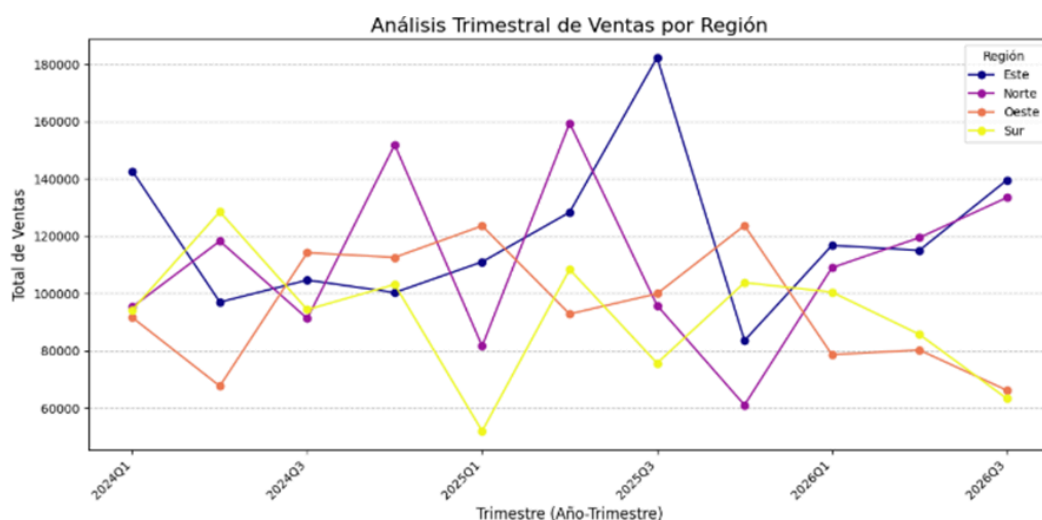
1. **Enfocar Estrategias en Productos de Alto Rendimiento:**

- Incrementar esfuerzos de marketing y ventas para los productos más vendidos. Diseñar campañas promocionales y anuncios centrados en estos productos.
 - Asegurar que estos productos estén siempre en stock para evitar pérdidas de ventas.
2. **Optimización de Inventarios y Costos:**
- Minimizar el inventario de productos de bajo rendimiento y considerar estrategias como el dropshipping para reducir costos operativos.
 - Negociar con proveedores para obtener mejores precios en los productos más vendidos.
3. **Mejora de la Experiencia del Cliente:**
- Mejorar la descripción, las imágenes y las reseñas de los productos más vendidos para aumentar la conversión.
 - Implementar programas de fidelización, ofreciendo descuentos y recompensas para clientes recurrentes.
4. **Estrategias de Cross-Selling:**
- Utilizar productos de alto rendimiento para promover productos de menor rendimiento mediante estrategias de cross-selling, lo que puede ayudar a aumentar las ventas de estos productos menos populares.

Factores a Considerar

5. **Análisis de Mercado:**
- Monitorear continuamente las tendencias del mercado para ajustar la oferta de productos según las demandas cambiantes de los consumidores.
6. **Feedback del Cliente:**
- Recopilar y analizar el feedback de los clientes para identificar áreas de mejora en productos de bajo rendimiento.
7. **Estrategias Competitivas:**
- Evaluar el desempeño de los productos de la competencia para asegurarse de que la empresa se mantenga competitiva.
8. **Innovación de Productos:**
- Considerar la innovación en los productos de alto rendimiento y en los de bajo rendimiento para mantener el interés del mercado.

3. COMPORTAMIENTO DE VENTAS POR REGIÓN



Análisis

1. Regiones en Crecimiento: Este y Norte

Tendencias: Estas regiones presentan un incremento constante en las ventas, lo que puede ser atribuido a factores como:

- Incremento market share y/o poder adquisitivo.
- Campañas de marketing y estrategias de distribución efectivas.

Insight: Estas áreas tienen un gran potencial para consolidarse como pilares estratégicos de crecimiento empresarial.

2. Regiones en Descenso: Sur y Oeste

Tendencias: Ventas decrecientes podrían ser indicativas de:

- Cambios en las necesidades del mercado.
- Competencia más fuerte o ineficiencias en la operación logística y publicitaria.

Insight: Necesitan una intervención estratégica inmediata para revertir la caída y recuperar el rendimiento.

3. Comparación General de Ventas

Región Este: Lidera en ingresos con \$1.3 millones en ventas, lo que la posiciona como la más fuerte y eficiente.

Región Sur: Con \$1.01 millones, muestra un desempeño débil en comparación. Este comportamiento podría reflejar limitaciones estructurales o desalineación con la demanda local.

Recomendaciones

1. Regiones en Crecimiento: Este y Norte

Refuerzo de Estrategias Exitosas:

- Incrementar inversiones en marketing digital y tradicional para mantener la fidelidad y atraer nuevos clientes.
- Explorar oportunidades de expansión, como abrir nuevos puntos de venta o fortalecer el comercio electrónico.

Replicación de Éxitos:

- Analizar detalladamente las estrategias que han impulsado el crecimiento (promociones, productos más vendidos, o canales efectivos) y probar su aplicación en otras regiones.

2. Regiones en Descenso: Sur y Oeste

Diagnóstico Localizado:

- Realizar estudios de mercado y encuestas para entender las causas del bajo rendimiento (precios, productos, o experiencia del cliente).
- Identificar a los competidores principales en estas regiones y analizar sus propuestas de valor.

Medidas Correctivas:

- Lanzar campañas promocionales dirigidas, como descuentos específicos o eventos locales.
- Ajustar la oferta de productos para alinearse mejor con las preferencias locales, como empaques más pequeños, productos económicos, o nuevos lanzamientos.

Mejora del Servicio y la Experiencia del Cliente:

- Capacitar al personal local para mejorar la atención al cliente.
- Fortalecer las plataformas digitales y el servicio post-venta.

Factores a Considerar

Dinámicas Demográficas

Evaluar la densidad poblacional, niveles de ingreso y patrones de consumo en cada región para personalizar estrategias.

Infraestructura y Logística

Analizar la efectividad de la cadena de suministro en las regiones de bajo rendimiento para detectar posibles ineficiencias.

Competencia y Mercado

Estudiar cómo los competidores locales están ganando cuota de mercado en las regiones en descenso.

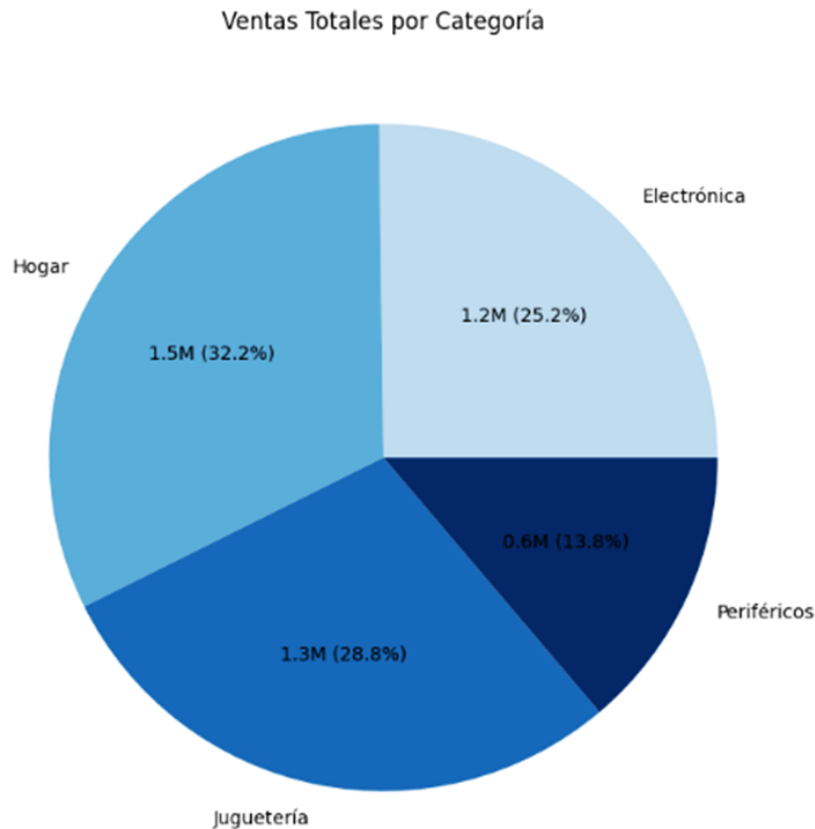
Identificar tendencias generales en las preferencias del consumidor.

Medir el impacto de las campañas actuales en cada región, tanto en términos de alcance como de conversión.

Presupuesto y Priorización

Dirigir mayores recursos a las regiones con más potencial (Este y Norte), pero sin descuidar las intervenciones estratégicas necesarias en las regiones en declive (Sur y Oeste).

4. COMPORTAMIENTO DE VENTAS POR CATEGORÍA



1. Categoría Hogar

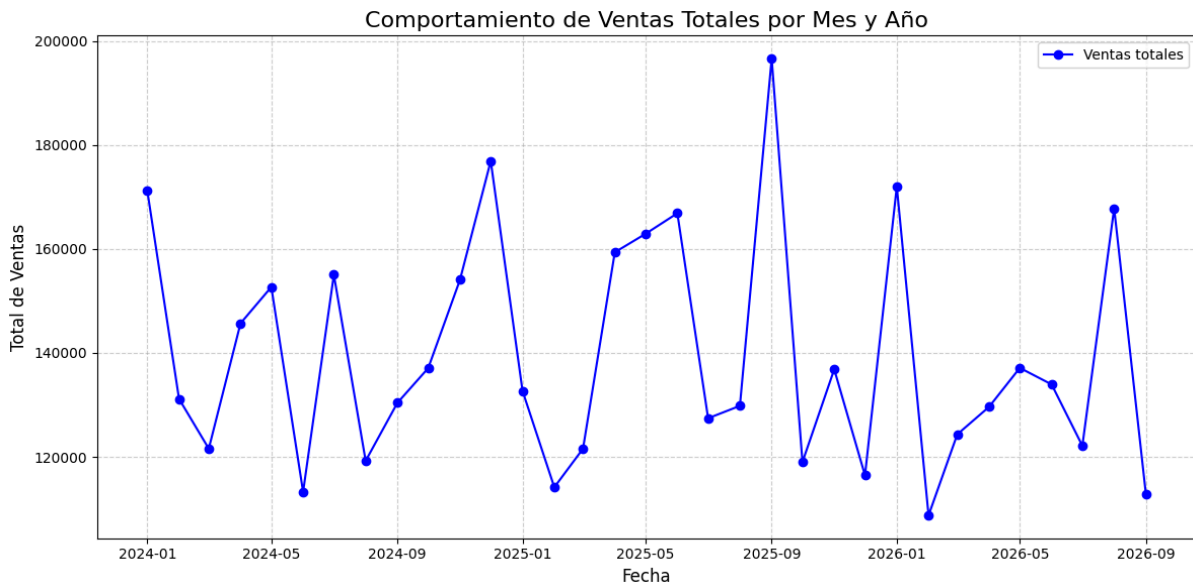
- **Análisis:** Representando un **32.2%** del total de ventas (1.5 millones), esta categoría lidera el mercado. La alta demanda puede estar impulsada por su relevancia diaria y la percepción de utilidad y valor entre los consumidores.
- **Recomendaciones:**
 - Expandir la oferta con productos complementarios y relacionados para capitalizar el interés existente.
 - Diseñar campañas de marketing basadas en el uso diario y la calidad de vida, resaltando los beneficios prácticos de los productos.
- **Factores a Considerar:** Analizar tendencias específicas dentro de esta categoría, como la sostenibilidad o el diseño de los productos, para desarrollar productos que refuercen la fidelidad del consumidor.

2. Categoría Periféricos

- **Análisis:** Con **13.8%** de las ventas totales (0.6 millones), esta categoría tiene el menor rendimiento, probablemente debido a una percepción de baja necesidad, alta competencia o promoción insuficiente.
- **Recomendaciones:**
 - Investigar las necesidades específicas del mercado, identificando posibles barreras de compra.

- Introducir promociones como combos o descuentos estratégicos para aumentar el interés en la categoría.
 - Fortalecer la presencia en plataformas de comercio electrónico mediante mejores descripciones, reseñas de usuarios y campañas orientadas a destacar beneficios clave.
 - **Factores a Considerar:** Evaluar la experiencia del cliente (CX) en esta categoría para detectar posibles fricciones, como complejidad en la elección o falta de confianza en el producto.
3. **Categoría Juguetería**
- **Análisis:** Contribuyendo con **28.8%** de las ventas totales (1.3 millones), esta categoría es fuerte, probablemente impulsada por factores estacionales (vacaciones, eventos) y productos innovadores.
 - **Recomendaciones:**
 - Desarrollar estrategias basadas en la estacionalidad, como lanzamientos de productos en temporadas clave.
 - Diversificar la oferta con juguetes educativos o de tendencia que atraigan a un público más amplio (niños y adultos).
 - **Factores a considerar:** Analizar patrones de compra estacional y monitorear competidores para mantener una ventaja competitiva en el diseño y promoción de productos.
4. **Categoría Electrónica**
- **Análisis:** Representa un **25.2%** del total (1.2 millones), con alta demanda debido a la constante innovación tecnológica y la necesidad de actualización de dispositivos.
 - **Recomendaciones:**
 - Promover lanzamientos de productos con características innovadoras y opciones de financiamiento para hacerlos más accesibles.
 - Implementar campañas de marketing que destaquen tendencias tecnológicas y aplicaciones prácticas.
 - **Factores a considerar:** Asegurarse de mantener disponibilidad y precios competitivos, especialmente en productos más demandados como smartphones y dispositivos inteligentes.

5. CORRELACIÓN DE PERIODOS



1. Patrones Generales:

- Las ventas parecen seguir un comportamiento **altamente variable**, con periodos cíclicos de incremento y disminución.
- Los picos más destacados se observan cerca de **2025-09** y **2026-01**, mientras que los valles más profundos ocurren luego del **2026-01**.

2. Factores Posibles:

- Estacionalidad:** Los picos en meses específicos podrían coincidir con eventos importantes como promociones de fin de año, días festivos o lanzamientos de productos.
- Valles:** Podrían relacionarse con periodos de baja demanda estacional, ajustes en estrategias de mercado o saturación de productos.

3. Relación con el Análisis Anterior:

- Categorías dominantes:** La categoría de Hogar, al ser la más fuerte, puede haber contribuido significativamente a los meses de ventas altas. Campañas estratégicas o necesidades específicas podrían haber impulsado estos picos.
- Electrónica:** Dado que esta categoría sigue tendencias tecnológicas, los incrementos podrían estar alineados con lanzamientos de dispositivos populares.
- Juguetería:** Es probable que contribuya a los aumentos estacionales, como en el cuarto trimestre de 2025, coincidiendo con festividades.

4. Oportunidades de Mejora:

- Promociones:** Identificar los meses con menor rendimiento y diseñar campañas para estimular las ventas (por ejemplo, descuentos en los meses de 2026-05).
- Análisis Predictivo:** Usar este historial de datos para prever periodos de alta o baja demanda y ajustar la estrategia de inventarios, marketing y recursos.

Factores a considerar

5. Estacionalidad:

- Aprovechar los meses de mayor venta con campañas intensivas que refuercen la lealtad del cliente, promociones cruzadas y eventos destacados.
- Planificar estrategias específicas para los meses de menor rendimiento, como introducción de nuevos productos o incentivos para clientes recurrentes.

6. Segmentación de Datos:

- Analizar las categorías de producto por separado para comprender cómo contribuyen al comportamiento mensual general.
- Evaluar factores externos que pudieran influir en la variación (competencia, economía, promociones).

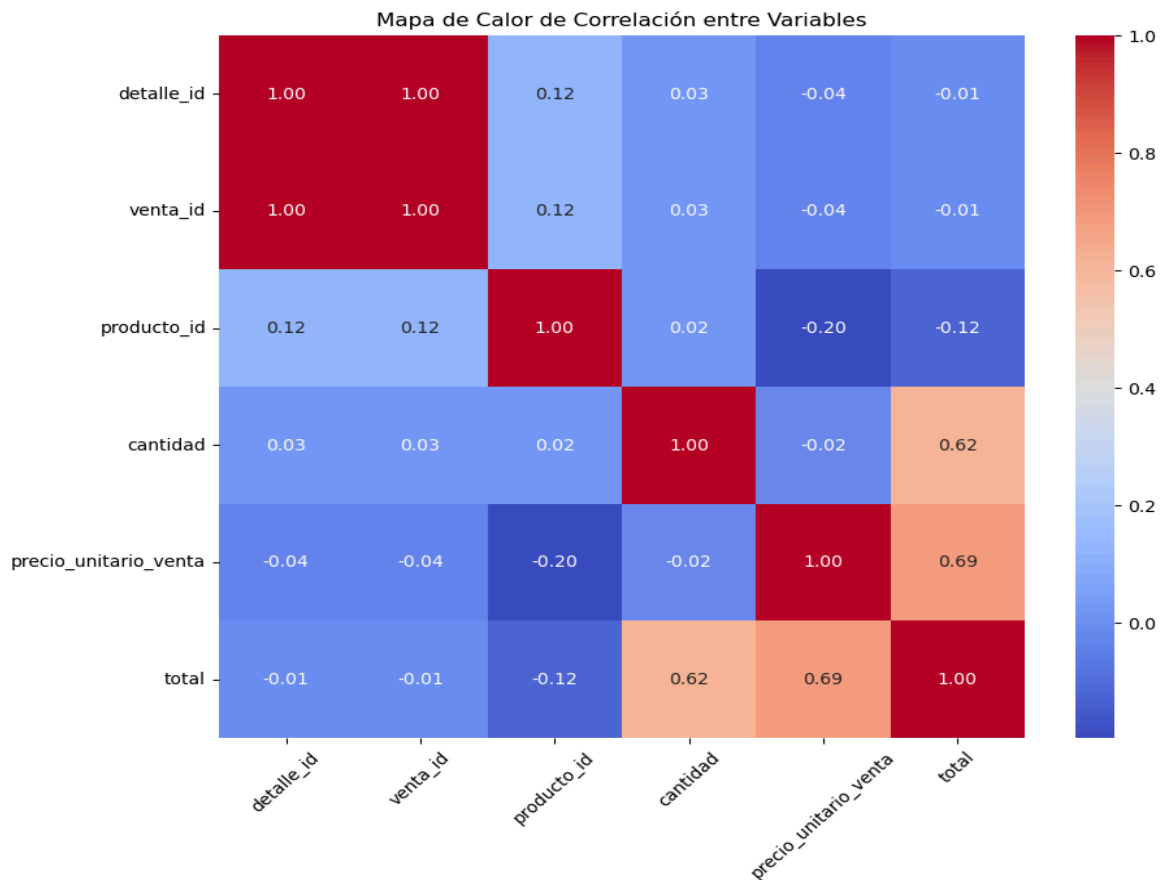
7. Optimización de Inventario:

- Garantizar una disponibilidad adecuada de los productos más demandados en meses clave para evitar rupturas de stock.
- Reducir costos operativos en meses de baja demanda ajustando niveles de inventario y logística.

8. Marketing Inteligente:

- Implementar campañas dirigidas basadas en la segmentación de clientes para maximizar las ventas en meses clave.
- Utilizar canales digitales para captar audiencias más amplias durante los periodos de disminución en ventas.

6. MAPA DE CALOR Y CORRELACIÓN DE VARIABLES



El gráfico presentado es un mapa de calor de correlación que muestra la relación entre diferentes variables en una base de datos. El valor de correlación está representado en una escala de colores y numérica (de -1 a 1), con las siguientes implicaciones:

1. Rangos de Correlación:
 - 1: Correlación positiva perfecta; las variables están directamente relacionadas.
 - 0: Sin correlación; no hay relación lineal entre las variables.
 - -1: Correlación negativa perfecta; las variables están inversamente relacionadas.
2. Colores:
 - Rojo oscuro: Correlaciones altas (positivas).
 - Azul oscuro: Correlaciones bajas (cercanas a 0 o negativas).

Analysis

Relación entre cantidad y total:

- **Correlación positiva (0.62):** Cuando se vende más cantidad de productos, el total de ventas aumenta. Esto es lógico porque más unidades generan más ingresos.

Relación entre precio unitario y total:

- **Correlación positiva (0.69):** A mayores precios, el total de ventas también aumenta. Esto muestra que los precios son un factor clave para los ingresos.

Producto y otras variables:

- La relación entre el tipo de producto (`producto_id`) y otras variables como `cantidad` o `total` es débil o negativa. Esto significa que no todos los productos tienen un impacto directo en las ventas totales.

Cantidad y precio unitario:

- Relación casi inexistente (-0.02): Esto podría significar que vender más no siempre depende del precio. Las promociones o estrategias comerciales pueden influir.

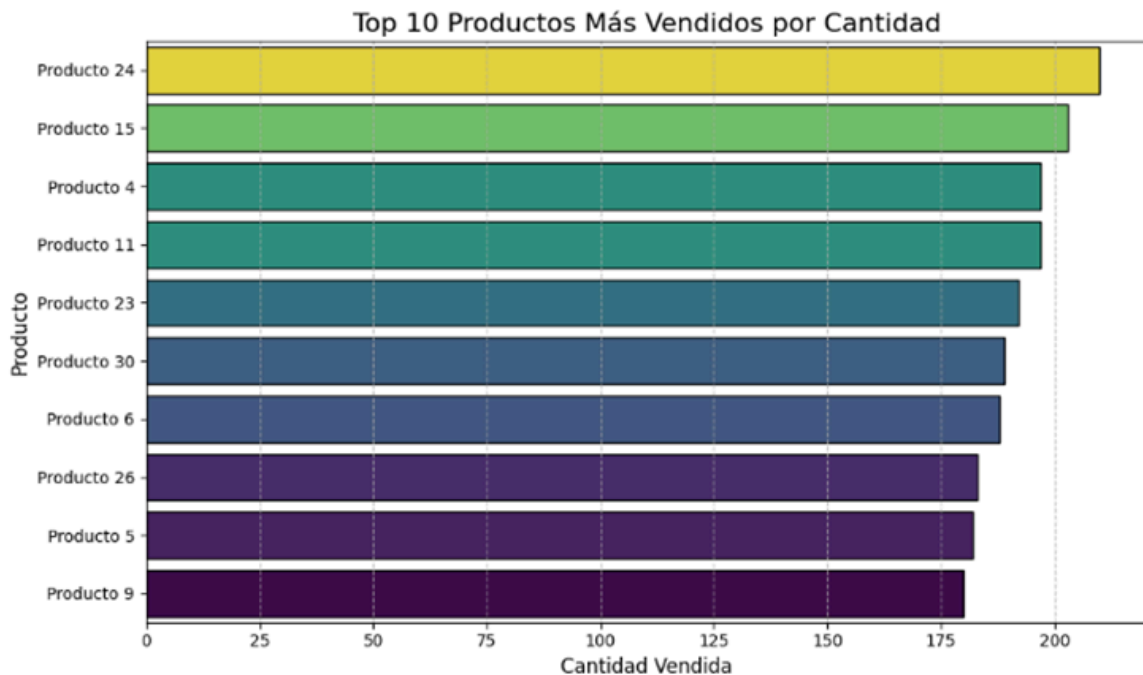
Recomendaciones

1. **Aumentar ventas estratégicamente:**
 - Focalizar las promociones en los productos que más contribuyen al total de ventas.
 - Analizar si bajar el precio puede aumentar la cantidad vendida en productos clave.
2. **Revisar la estrategia de precios:**
 - Dado que los precios tienen un impacto directo en las ventas totales, se pueden explorar descuentos, ofertas, o precios dinámicos para maximizar los ingresos.
3. **Entender el impacto de cada producto:**
 - Realizar un análisis más detallado para identificar cuáles productos tienen mayor impacto en las ventas totales y cuáles tienen menor relevancia.
4. **Monitorear los datos:**
 - Revisar si los datos de identificación (`detalle_id` y `venta_id`) son correctos y no duplicados, ya que tienen una relación perfecta (1.00) que podría ser técnica más que funcional.

Factores a Considerar

- **Clientes y mercado:** Asegurar que las estrategias de precios y promociones sean atractivas para el público objetivo.
- **Temporada o eventos:** Las variaciones en ventas pueden depender de factores externos como temporadas altas o promociones específicas.

6. TOP DE PRODUCTOS MÁS VENDIDOS



Analisis

1. **Producto más vendido:** El **Producto 24** lidera las ventas con respecto a los demás. Esto sugiere una preferencia por este producto o una estrategia comercial efectiva detrás de él (como promociones, calidad o demanda del mercado).
2. **Distribución uniforme:** A partir del segundo lugar, los productos muestran cantidades vendidas más cercanas entre ellos, lo que indica un comportamiento equilibrado en la oferta o demanda para estos productos.
3. **Productos destacados:** Además del Producto 24, el **Producto 15 y el Producto 4 y 11** son los siguientes más populares, lo que puede sugerir oportunidades de optimización para reforzar sus ventas.
4. **Relación con precios:**
 - Según el gráfico de correlación previo, los productos con mayor cantidad tienden a influir directamente en el "Total de Ventas". El Producto 24, al liderar en cantidad, probablemente contribuye significativamente a los ingresos generales.

Recomendaciones:

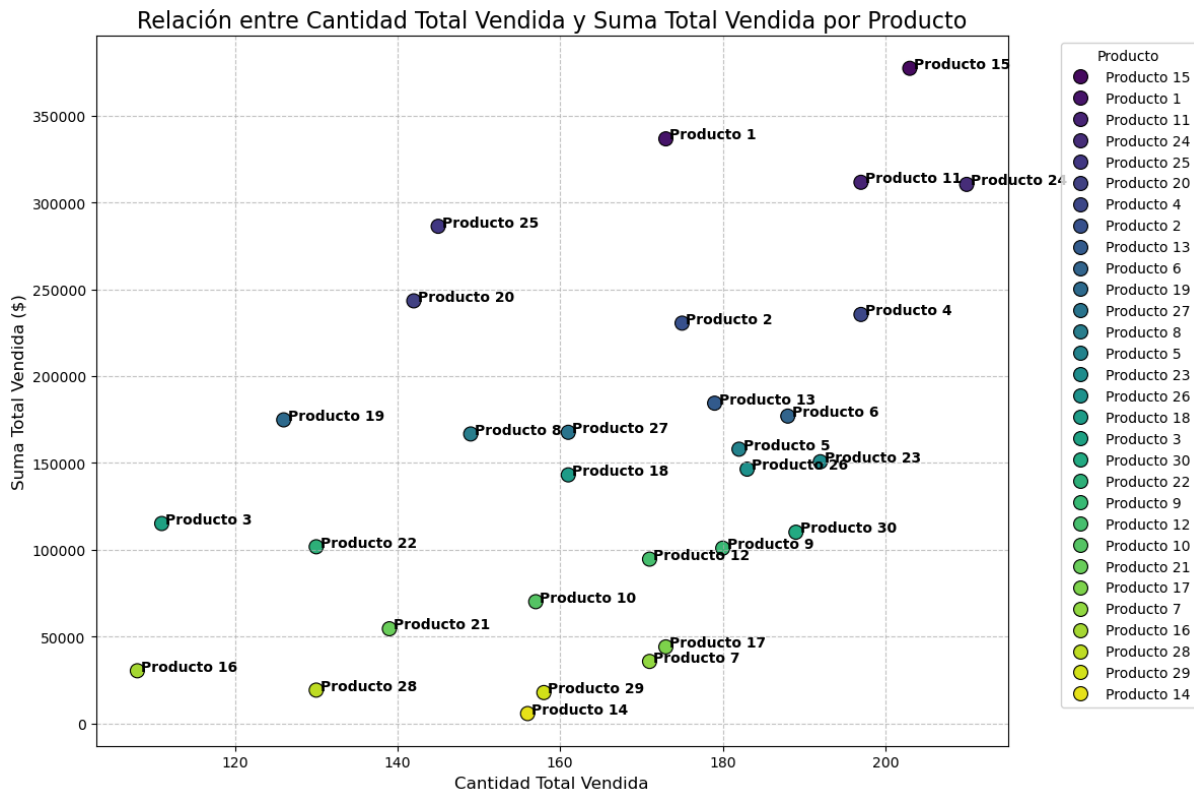
1. **Estrategias específicas para el Producto 24:**
 - Mantener su disponibilidad y analizar la razón detrás de su demanda (calidad, promociones, precios accesibles, etc.).
 - Diseñar campañas que refuercen la fidelización hacia este producto, evitando una posible saturación del mercado.
2. **Reforzar ventas de productos del Top 3:**

- Impulsar los productos 15 y 4 con estrategias similares al Producto 24 (ofertas, promociones cruzadas o cambios en la presentación).
- 3. **Optimizar productos menos destacados del Top 10:**
 - Revisar los **Productos 9 y 5**, los cuales ocupan las últimas posiciones, para identificar posibles barreras (precio, distribución o comunicación de valor).
- 4. **Relación con Precios y Total de Ventas:**
 - Dado que el "precio_unitario_venta" y la "cantidad" tienen una correlación positiva con el "Total", productos como el 24 deberían mantenerse con precios competitivos y estrategias que aumenten la rotación de inventario.

Factores a considerar:

- **Saturación del mercado:** Analizar si la alta rotación del Producto 24 afecta negativamente las ventas de otros productos similares.
- **Margen de ganancias:** Evaluar si los productos más vendidos generan un margen de beneficio suficiente o si las promociones están reduciendo excesivamente las utilidades.
- **Tendencias de consumo:** Realizar encuestas o análisis de datos históricos para comprender patrones de consumo que justifiquen estos resultados.

7. Gráfica Adicional de Correlación de ventas y Cantidad



Query Usada para la extraccion, se actualizo en el notebook de analisis_exploratorio

```
# Crear el gráfico de dispersión para visualizar la correlación entre
cantidad total vendida y suma total vendida
plt.figure(figsize=(12, 8))
scatter_plot = sns.scatterplot(x='cantidad_total_vendida',
y='suma_total_vendida', data=productos_vendidos, hue='nombre_producto',
palette='viridis', edgecolor='black', s=100)

# Añadir etiquetas para cada punto
for i in range(productos_vendidos.shape[0]):
    scatter_plot.text(productos_vendidos.cantidad_total_vendida.iloc[i]
+ 0.5,
                    productos_vendidos.suma_total_vendida.iloc[i] +
0.5,
                    productos_vendidos.nombre_producto.iloc[i],
                    horizontalalignment='left',
                    size='medium',
                    color='black',
                    weight='semibold')

plt.title('Relación entre Cantidad Total Vendida y Suma Total Vendida
por Producto', fontsize=16)
```

```
plt.xlabel('Cantidad Total Vendida', fontsize=12)
plt.ylabel('Suma Total Vendida ($)', fontsize=12)
plt.xticks(fontsize=10)
plt.yticks(fontsize=10)
plt.grid(True, linestyle='--', alpha=0.7)
plt.legend(title='Producto', bbox_to_anchor=(1.05, 1), loc='upper
left')
plt.tight_layout()
plt.show()
```

Análisis

1. Segmentación de Productos según Desempeño:

- **Productos de alto valor y volumen:**
 - Productos como el **Producto 15** y el **Producto 1** destacan tanto en cantidad total vendida como en la suma total vendida. Estos son los líderes del mercado, probablemente productos estrella en la matriz BCG.
- **Productos con alta cantidad vendida pero menores ingresos:**
 - Algunos productos (e.g., Producto 30 y Producto 9) tienen una cantidad vendida relativamente alta pero generan menos ingresos totales. Esto podría indicar que tienen un precio unitario bajo.
- **Productos con bajo volumen y valor:**
 - Los productos como el **Producto 16** y el **Producto 28** se ubican en el cuadrante inferior izquierdo, lo que sugiere baja rotación e ingresos marginales. Podrían ser candidatos para revisión estratégica.

2. Relación Precio-Cantidad:

- Existe una relación positiva entre cantidad total vendida y suma total vendida, pero no completamente lineal. Esto indica que los precios y volúmenes varían ampliamente entre los productos.
- Productos como el **Producto 15** tienen un precio unitario alto, ya que, a pesar de no ser los más vendidos en cantidad, generan ingresos significativos.

3. Concentración de Ventas:

- Un pequeño grupo de productos genera la mayor parte de los ingresos. Esto sugiere una estrategia de **pareto** (80-20), donde un 20% de los productos probablemente contribuyen al 80% de los ingresos.

Recomendaciones

1. Estrategias para Productos Estrella:

- Mantener altos niveles de inventario para los productos líderes (e.g., Producto 15, Producto 1).

- Realizar campañas que refuercen la fidelización y promuevan la rotación continua de estos productos.
- 2. **Productos con Alto Volumen pero Baja Contribución al Ingreso:**
 - Revisar los márgenes de ganancia de productos como el **Producto 30** y **Producto 9**.
 - Evaluar si es posible aumentar su precio sin afectar la demanda significativamente, considerando su elasticidad.
- 3. **Productos de Baja Rotación e Ingreso:**
 - Analizar barreras para productos como el **Producto 16** y el **Producto 28** (e.g., falta de promoción, baja demanda, competencia directa).
 - Considerar estrategias de liquidación o reposicionamiento para estos productos.
- 4. **Optimización de Precios y Campañas:**
 - Implementar análisis de precios para garantizar un balance entre competitividad y rentabilidad.
 - Utilizar promociones cruzadas para productos complementarios, especialmente aquellos que no se venden tanto.
- 5. **Logística y Disponibilidad:**
 - Asegurar la disponibilidad de los productos más vendidos en todos los canales, optimizando la cadena de suministro para reducir costos y tiempos de entrega.
 - Utilizar herramientas de pronóstico para anticipar picos de demanda en los productos estrella.
- 6. **Toma de Decisiones Basada en Datos:**
 - Se recomienda profundizar más y realizar un análisis de elasticidad de precios para determinar qué productos tienen margen para ajustes.
 - Investigar patrones de consumo a través de encuestas o analítica de datos para ajustar la oferta.