

INFORME ANALISIS EXPLORATORIO DE DATOS

- **Contexto:** Análisis del Integral de Ventas de Supermercado
- **Fecha de Elaboración:** 31/03/2024
- **Elaborado por:** Víctor Manuel Rocha Ramos

RESUMEN EJECUTIVO

El presente informe proporciona un análisis exhaustivo de los datos de ventas de un supermercado durante un período de tres meses, con el objetivo de comprender mejor el desempeño comercial y las tendencias del mercado. A través de una serie de preguntas objetivo, se exploraron diversas dimensiones de las ventas, incluyendo la distribución de ingresos por sucursal y ciudad, las preferencias de producto, el comportamiento de los clientes, la dinámica de precios y cantidad, los métodos de pago y la satisfacción del cliente.

METODOLOGIA

1. **Extracción de la Base de Datos:** Se inició con la identificación y extracción de datos de ventas de supermercado a partir de una fuente de datos abierta y confiable. Esta etapa fue crucial para asegurar que el análisis se basara en datos precisos y representativos de las operaciones reales del supermercado, incluyendo transacciones de ventas, detalles de productos, información de clientes, y performance de las diferentes sucursales.
2. **Entendimiento del Negocio y Formulación de Preguntas Objetivo:** Se llevó a cabo una inmersión profunda en el modelo de negocio del supermercado, analizando el entorno de mercado, la estructura operacional, y las estrategias de ventas y marketing.
3. **Importación de la Base de Datos a Herramientas de Análisis:** Utilizando Jupyter Notebook, una plataforma interactiva de ciencia de datos, se importaron los datos para su manipulación y análisis.
4. **Análisis Exploratorio de Datos (AED):** En esta fase crítica, se examinaron las distribuciones de las variables, se identificaron valores atípicos, y se exploraron las relaciones entre diferentes variables del conjunto de datos. El AED ayudó a construir una comprensión sólida de los datos a disposición, destacando las características principales, las anomalías, y las posibles correlaciones entre variables que podrían influir en las decisiones de negocio.
5. **Limpieza de la Base de Datos:** La calidad de los datos se aseguró mediante la limpieza, que incluyó la corrección de errores de entrada, el tratamiento de valores faltantes a través de técnicas como la imputación o la eliminación, y la estandarización de formatos para garantizar consistencia a través de la base de datos. Este paso fue esencial para preparar los datos para análisis más avanzados y precisos.
6. **Tratamiento de Valores Extremos (Outliers):** Con el uso de la librería Matplotlib se emplearon métodos estadísticos y visualizaciones para identificar y manejar los valores extremos que podrían distorsionar los análisis y conclusiones. Dependiendo

del contexto y la importancia de estos valores para el análisis, se tomaron decisiones informadas sobre si ajustarlos, eliminarlos o mantenerlos, basándose en su impacto potencial en los resultados del estudio.

7. **Análisis Estadístico:** Esta etapa abarcó el análisis cuantitativo de los datos, utilizando estadísticas descriptivas y inferenciales para comprender las tendencias centrales, la variabilidad, y las relaciones entre variables. Este análisis proporcionó insights valiosos sobre las dinámicas de ventas, la efectividad de las estrategias de precios, y las preferencias de los clientes.
8. **Extracción de Información:** Con los datos ya limpios y una comprensión clara de las necesidades de análisis, se procedió a extraer información clave para responder a las preguntas objetivas formuladas inicialmente. Esta extracción se llevó a cabo mediante la ejecución de scripts de análisis en Python y consultas SQL, aprovechando la capacidad de estas herramientas para manejar grandes volúmenes de datos y realizar cálculos complejos.
9. **Exportación de la Base de Datos y Visualización de Datos:** Los datos analizados y las conclusiones clave se prepararon para la visualización, exportándolos a formatos compatibles con herramientas de visualización como Excel y Power BI. Esto permitió la creación de representaciones visuales comprensibles de los datos, facilitando la interpretación de los resultados y la comunicación de los insights obtenidos.
10. **Creación de Dashboards y Visualizaciones Avanzadas:** Se desarrollaron dashboards en Excel y Power BI para presentar los resultados del análisis de manera interactiva y comprensible, permitiendo explorar las tendencias de ventas, preferencias de clientes, y la eficacia de diferentes productos y estrategias de marketing.

Este proceso metodológico proporcionó una comprensión detallada y multifacética del funcionamiento interno y las estrategias comerciales del supermercado, apoyando la toma de decisiones informadas y la implementación de mejoras basadas en datos.

INFORMACION GENERAL

DESCRIPCION DE LAS COLUMNAS DE LA BASE DE DATOS:

1. **Branch:** Identifica la sucursal donde se realizó la venta.
2. **City:** Ciudad de la sucursal.
3. **Customer_type:** Diferencia entre miembros y clientes regulares.
4. **Gender:** Género del cliente.
5. **Product_line:** Categoría del producto vendido.
6. **Unit_price:** Precio por unidad del producto.
7. **Quantity:** Número de unidades vendidas.
8. **Total:** Monto total de la venta.
9. **Date:** Fecha de la transacción.
10. **Month:** Mes en que se realizó la venta.
11. **Payment:** Método de pago utilizado.
12. **Rating:** Calificación de satisfacción del cliente.

Cada columna proporciona una pieza clave para el entendimiento completo de las operaciones de venta del supermercado, desde la popularidad de productos hasta las preferencias de los consumidores y la eficacia de las estrategias de negocio.

ANALISIS ESTADISTICO

VARIABLES CUANTITATIVAS:

1. **Unit_price (Precio Unitario):**

- La media del precio unitario es de aproximadamente \$55.67, lo que indica el precio promedio por unidad de producto.
- El rango de precios va desde \$10.08 hasta \$99.96.
- El 25% de los precios unitarios están por debajo de \$32.88, mientras que el 75% están por debajo de \$77.93.

2. **Quantity (Cantidad):**

- La media de la cantidad de productos comprados por transacción es de alrededor de 5.51 unidades.
- El rango de cantidad comprada va desde 1 hasta 10 unidades.
- El 25% de las transacciones involucran la compra de hasta 3 unidades, mientras que el 75% implican la compra de hasta 8 unidades.

3. **Total (Total de la Venta):**

- La media del monto total de la venta por transacción es de aproximadamente \$307.59.
- El rango del monto total de la venta va desde \$10.17 hasta \$993.00.
- El 25% de las transacciones tienen un monto total de venta inferior a \$118.50, mientras que el 75% tienen un monto total inferior a \$448.91.

4. **Rating (Calificación):**

- La media de la calificación de satisfacción del cliente es de aproximadamente 6.97 sobre 10.
- Las calificaciones van desde 4 hasta 10.
- El 25% de las calificaciones están por debajo de 5.50, mientras que el 75% están por debajo de 8.50.

Este análisis proporciona una comprensión detallada de la distribución y las tendencias de las variables cuantitativas en el conjunto de datos de ventas del supermercado.

INTRODUCCION

Este informe presenta un análisis conciso del conjunto de datos de ventas de un supermercado, con el propósito de extraer información relevante para la toma de decisiones estratégicas. Se explorarán diversos aspectos del negocio, incluyendo la distribución geográfica de las ventas, las preferencias de los clientes y la satisfacción del cliente.

El análisis se realizará utilizando técnicas estadísticas y herramientas de visualización de datos para identificar patrones y tendencias significativas en el comportamiento de compra. El objetivo es proporcionar insights prácticos que puedan utilizarse para mejorar la eficiencia operativa y la experiencia del cliente, impulsando así el crecimiento y la rentabilidad del supermercado.

ANALISIS DE DATOS Y RESULTADOS

PREGUNTAS OBJETIVO:

- Análisis General de Ventas

- 1. Pregunta: ¿Cuál es el ingreso total generado por las ventas para cada sucursal?**

Respuesta:

- Sucursal A: \$101,143.21
- Sucursal B: \$101,140.64
- Sucursal C: \$105,303.53

- Comportamiento de Compra por Ciudad y Sucursal

- 2. Pregunta: ¿Cómo varían las ventas (en términos de ingresos totales) entre las diferentes ciudades donde se encuentran las sucursales?**

Respuesta:

- Naypyitaw: \$105,303.53
- Yangon: \$101,143.21
- Mandalay: \$101,140.64

- Preferencias de Producto

- 3. Pregunta: ¿Cuáles son las líneas de productos más vendidas?**

Respuesta:

- Accesorios electrónicos: 971 unidades
- Alimentos y bebidas: 952 unidades
- Deportes y viajes: 920 unidades
- Hogar y estilo de vida: 911 unidades
- Accesorios de moda: 902 unidades
- Salud y belleza: 854 unidades

- Preferencias de Producto

- 4. Pregunta: ¿Cuáles son las líneas de productos que generan más ingresos?**

Respuesta:

- Alimentos y bebidas: \$53,471.28
- Deportes y viajes: \$52,497.93
- Accesorios electrónicos: \$51,750.03
- Accesorios de moda: \$51,719.90

- Hogar y estilo de vida: \$51,297.06
- Salud y belleza: \$46,851.18
- **Dinámica de Precios y Cantidad**
 5. **Pregunta: ¿Existe alguna correlación entre el precio unitario de los productos y la cantidad vendida?**

Respuesta: Hay una correlación positiva muy débil de 0.010777564342497267 entre el precio unitario de los productos y la cantidad vendida.

- **Comportamiento de los Clientes**
 6. **Pregunta: ¿Difieren los patrones de compra entre miembros y no miembros?**

Respuesta:

- Miembros: \$156,403.28
- No miembros: \$151,184.10

- **Comportamiento de los Clientes**
 7. **Pregunta: ¿Difieren los patrones de compra entre géneros?**

Respuesta:

- Femenino: \$159,888.50
- Masculino: \$147,698.88

- **Análisis Temporal de las Ventas**
 8. **Pregunta: ¿Cómo varían las ventas a lo largo de los tres meses? ¿Hay algún patrón o tendencia visible?**

Respuesta:

- Enero: \$110,754.16
- Febrero: \$92,589.88
- Marzo: \$104,243.34

2. Métodos de Pago

9. **Pregunta: ¿Qué método de pago es el más utilizado por los clientes en cada sucursal?**

Respuesta:

- Ewallet: 345 transacciones
- Efectivo: 344 transacciones
- Tarjeta de crédito: 311 transacciones

3. Satisfacción del Cliente

10. Pregunta: ¿Cómo se relaciona la calificación de satisfacción del cliente con el volumen de ventas?

Respuesta: Existe una correlación negativa muy débil de -0.036441704997018334 entre la calificación de satisfacción del cliente y el volumen de ventas.

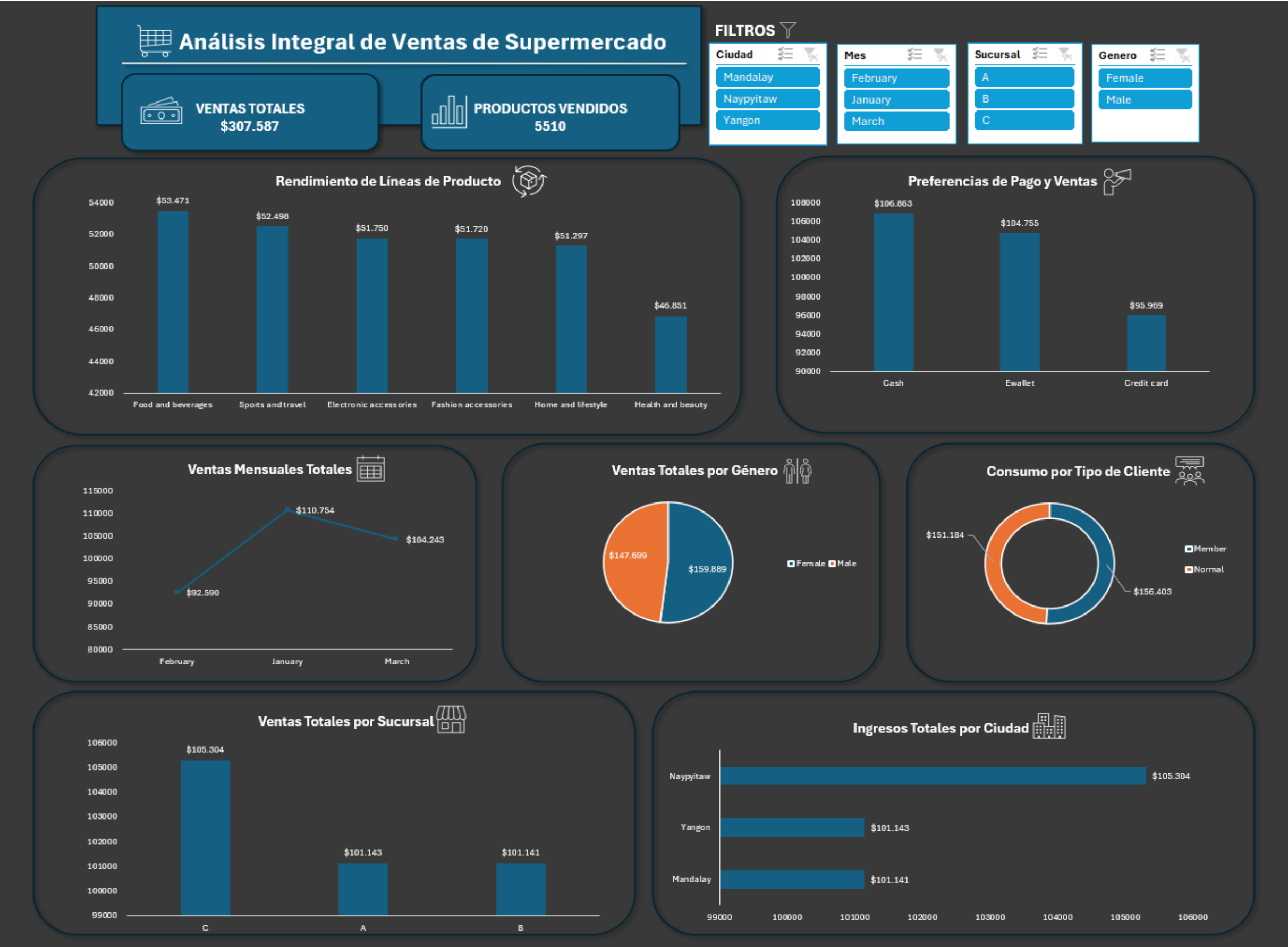
4. Satisfacción del Cliente

11. Pregunta: ¿Los productos con mayores ventas tienen mejores calificaciones?

Respuesta:

- Alimentos y bebidas: Promedio de calificación: 7.113218
- Accesorios de moda: Promedio de calificación: 7.029213
- Salud y belleza: Promedio de calificación: 7.003289
- Accesorios electrónicos: Promedio de calificación: 6.924706
- Deportes y viajes: Promedio de calificación: 6.916265
- Hogar y estilo de vida: Promedio de calificación: 6.837500

DASHBOARD EXCEL



DASHBOARD POWER BI



Análisis Integral de Ventas de Supermercado

TOTAL DE VENTAS



\$307,59 mil

PRODUCTOS VENDIDOS



5510

FILTROS



Ciudad

- ☐ Mandalay
- ☐ Naypyitaw
- ☐ Yangon

Mes

- ☐ February
- ☐ January
- ☐ March

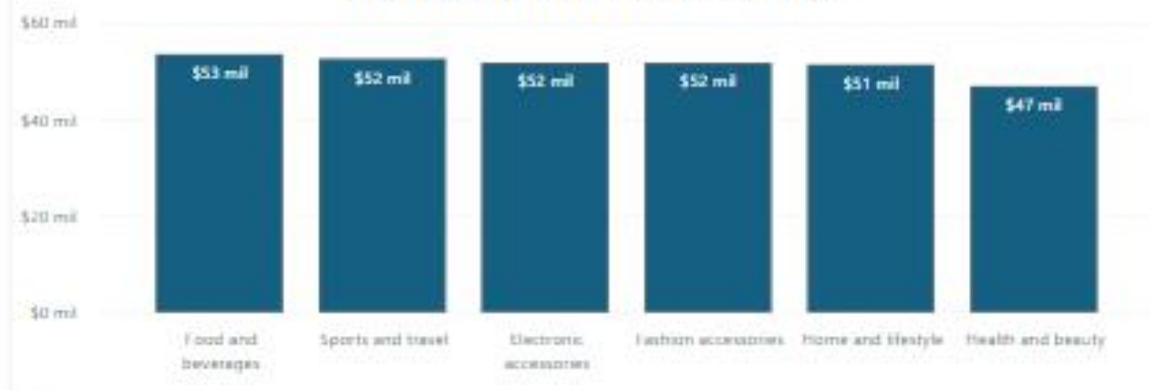
Sucursal

- ☐ A
- ☐ B
- ☐ C

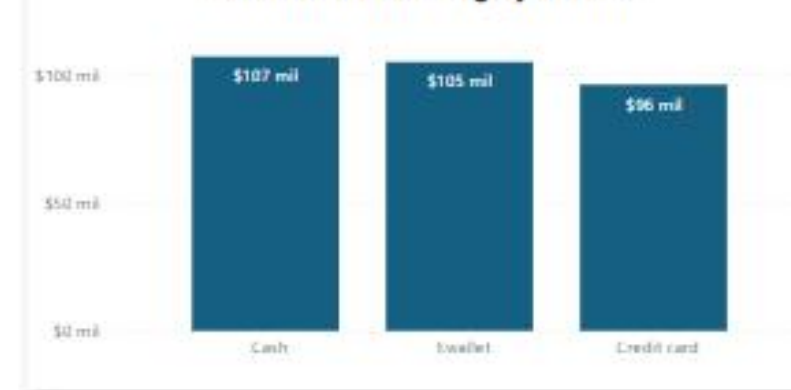
Genero

- ☐ Female
- ☐ Male

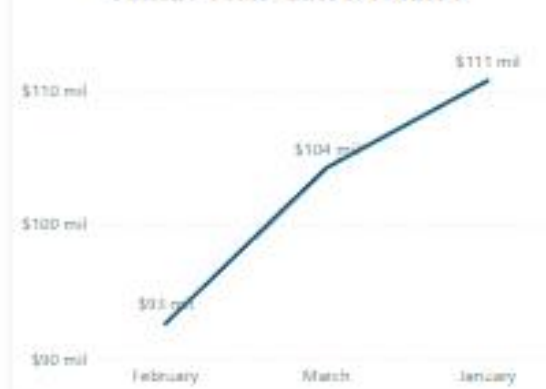
Rendimiento de Líneas de Producto



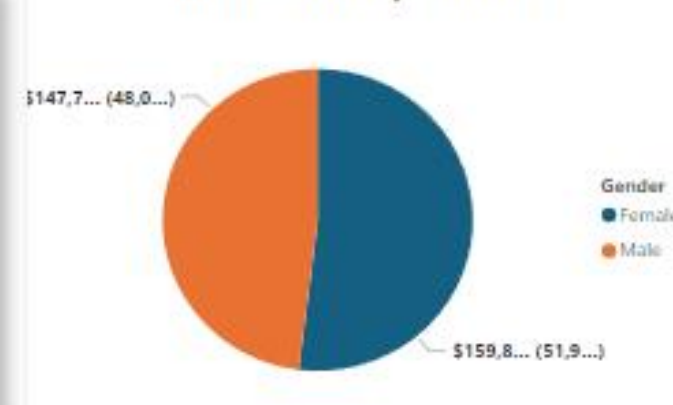
Preferencias de Pago y Ventas



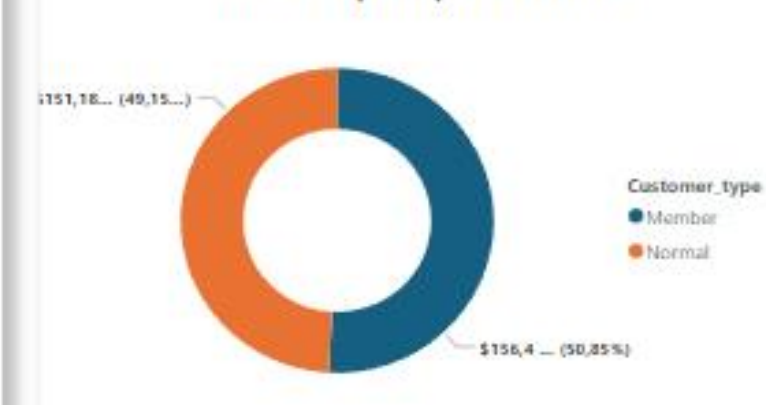
Ventas Mensuales Totales



Ventas Totales por Género



Consumo por Tipo de Cliente



Ventas Totales por Sucursal



Ventas Totales por Ciudad



Ventas Totales por Ciudad (Mapa)



CONCLUSIONES

Las conclusiones extraídas del análisis del conjunto de datos del supermercado proporcionan una visión detallada del rendimiento de este, destacando las siguientes observaciones clave:

1. Desempeño de las Sucursales:

- Las ventas totales muestran una distribución relativamente equitativa entre las sucursales, con ingresos generados cercanos entre sí. Sin embargo, la Sucursal C destaca ligeramente con el ingreso total más alto.

2. Varianza en Ventas por Ciudad:

- La ciudad de Naypyitaw lidera en términos de ingresos totales generados por ventas, seguida de Yangon y Mandalay.

3. Preferencias de Producto:

- Las líneas de productos más vendidas incluyen accesorios electrónicos, alimentos y bebidas, y artículos deportivos y de viaje.
- En términos de ingresos, los productos alimenticios y las categorías relacionadas con el estilo de vida generan los mayores ingresos.

4. Correlación Precio-Cantidad:

- Existe una correlación muy débil entre el precio unitario de los productos y la cantidad vendida, lo que sugiere que el precio no es un factor determinante en la cantidad de productos vendidos.

5. Patrones de Compra por Segmento de Cliente:

- Los miembros del supermercado parecen gastar ligeramente más que los no miembros, lo que indica una posible preferencia por servicios adicionales o promociones exclusivas.
- No se observan diferencias significativas en los patrones de compra entre géneros.

6. Tendencias Temporales:

- Las ventas muestran una fluctuación mensual, con un pico en enero seguido de una disminución en febrero y un ligero aumento en marzo. Se necesita un análisis adicional para identificar las razones detrás de estas variaciones.

7. Preferencias de Pago:

- El método de pago más utilizado varía entre Ewallet y efectivo, siendo estos los métodos de elección en las transacciones del supermercado.

8. Satisfacción del Cliente:

- Existe una correlación negativa muy débil entre la calificación de satisfacción del cliente y el volumen de ventas, lo que sugiere que los clientes no siempre califican la satisfacción en función del volumen de sus compras.
- No hay una relación clara entre las calificaciones de los productos y sus ventas, lo que indica que la popularidad de un producto no siempre coincide con su calificación.

Estas conclusiones proporcionan una visión integral del rendimiento del supermercado y ofrecen puntos de partida para futuros análisis y estrategias comerciales.

REFERENCIAS:

Fuente de datos extraída de: <https://www.kaggle.com/datasets/aungpyaeap/supermarket-sales>