

**2° Actividad de evaluación**

**Análisis y Visualización de datos**

Materia Aprendizaje Automatico I

Prof. Javier Di Salvo

1°cuatrimestre 2024

Contenido

[Introducción 3](#_Toc151041796)

[Desarrollo 4](#_Toc151041797)

[Análisis de los datos 4](#_Toc151041798)

[Estructura del Conjunto de Datos: 4](#_Toc151041799)

[Integridad de los datos y Confiabilidad de los datos. 4](#_Toc151041800)

[Métricas para medir el desempeño 5](#_Toc151041801)

[Ventas Totales por Región 6](#_Toc151041802)

[Promedio de Descuento por Región 7](#_Toc151041803)

[Efectividad de Descuentos en Ventas por Región 7](#_Toc151041804)

[Dashboard 9](#_Toc151041805)

[Conclusión 10](#_Toc151041806)

[Bibliografía 11](#_Toc151041807)

# Introducción

Este informe se basa en un extenso conjunto de datos recopilados con el objetivo de arrojar luz sobre las transacciones de los clientes y, en particular, analizar el impacto de las estrategias de descuento en las ventas.

El conjunto de datos que estamos explorando tiene una visión detallada de cada transacción de ventas y envíos de una empresa. Esta riqueza de información ofrece la oportunidad de explorar patrones de compra, evaluar el rendimiento de los productos y, en última instancia, mejorar la toma de decisiones empresariales.

Cada fila representa una transacción única, incluyendo detalles como el tipo de envío, la fecha del pedido, la fecha de envío, la categoría del producto, la cantidad, el descuento aplicado, las ganancias y más.

En la primera sección, se aborda la integridad y confiabilidad de los datos, destacando la consistencia estructural, la ausencia de duplicados y de datos faltantes.

Posteriormente, se propone una métrica de desempeño centrada en evaluar el impacto de las estrategias de descuento, facilitando a los departamentos de ventas y análisis de datos una herramienta efectiva para medir la eficacia de las campañas promocionales.

# Desarrollo

## Análisis de los datos

Este conjunto de datos, extraído de [Kaggle](https://www.kaggle.com/datasets/vivek468/superstore-dataset-final), brinda una visión completa de las transacciones realizadas por los clientes en un supermercado. Con un total de 21 variables y 9993 observaciones, esta información se presenta de manera estructurada para facilitar un análisis detallado.

## Estructura del Conjunto de Datos:

La tabla central de nuestro análisis detalla cada transacción, incluyendo información esencial como la ID de la orden, fechas relevantes, modo de envío, ID del cliente, nombre del cliente, segmento, país, ciudad, entre otros campos significativos. La amplitud de las variables proporciona una visión completa de las interacciones de compra.

Descripción de las Columnas Principales:

**Row ID**: Identificador único de la fila.

**Order ID**: Identificador único de la orden de compra.

**Order Date**: Fecha de realización de la orden de compra. Entre los años 2013 a 2017

**Ship Date**: Fecha de envío de la orden.

**Ship Mode**: Modo de envío de la orden (por ejemplo, Second Class, Standard Class, etc.).

**Customer ID**: Identificación única del cliente que realizó la compra.

**Customer** Name: Nombre del cliente.

**Segment**: Segmento al que pertenece el cliente (por ejemplo, Consumer, Corporate, etc.).

**Country**: País donde se realizó la compra.

**City:** Ciudad donde se realizó la compra.

... (otros campos): Columnas adicionales contienen información como código postal, región, ID de producto, categoría de producto, subcategoría de producto, nombre del producto, ventas, cantidad, descuento y beneficio.

Significado de los Datos:

Cada fila representa una transacción de compra específica, y las columnas ofrecen detalles clave relacionados con la transacción, clientes y productos. Estos datos son fundamentales para analizar patrones de compra, rendimiento de productos y la rentabilidad general de las transacciones.

Este conjunto de datos será sometido a un análisis exhaustivo para comprender mejor los patrones de compra, y se explorará la posibilidad de construir modelos predictivos para mejorar la toma de decisiones empresariales.

## Integridad de los datos y Confiabilidad de los datos.

**Estructura Consistencia:** La tabla exhibe una consistencia notable en su estructura, donde cada columna presenta datos claros y bien definidos. Todas las entradas esenciales, como "Order ID," "Order Date," "Ship Date," "Customer ID," "Sales," "Quantity," "Discount," y "Profit," muestran coherencia en sus formatos y no contienen valores nulos. Esta uniformidad facilita la interpretación y el análisis preciso de los datos.

**Duplicados:** Cabe destacar la ausencia de datos duplicados en las columnas clave, como "Row ID" y "Order ID." La inexistencia de registros idénticos refuerza la integridad de la tabla, asegurando que cada entrada sea única y evitando posibles inconsistencias en los análisis derivados de estos datos. Esta combinación de consistencia estructural y la ausencia de duplicados contribuye a la confiabilidad y solidez de la tabla para su uso en diversas aplicaciones analíticas.

**Datos faltantes**

Como podemos observar no datos faltantes

RangeIndex: 9994 entries, 0 to 9993

Data columns (total 21 columns):

# Column Non-Null Count Dtype

--- ------ -------------- -----

0 Row ID 9994 non-null int64

1 Order ID 9994 non-null object

2 Order Date 9994 non-null datetime64[ns]

3 Ship Date 9994 non-null datetime64[ns]

4 Ship Mode 9994 non-null object

5 Customer ID 9994 non-null object

6 Customer Name 9994 non-null object

7 Segment 9994 non-null object

8 Country 9994 non-null object

9 City 9994 non-null object

10 State 9994 non-null object

11 Postal Code 9994 non-null int64

12 Region 9994 non-null object

13 Product ID 9994 non-null object

14 Category 9994 non-null object

15 Sub-Category 9994 non-null object

16 Product Name 9994 non-null object

17 Sales 9994 non-null float64

18 Quantity 9994 non-null int64

19 Discount 9994 non-null float64

20 Profit 9994 non-null float64

dtypes: datetime64[ns](2), float64(3), int64(3), object(13)

**Validez:** No hay valores extremadamente atípicos ni datos que claramente no tengan sentido en el contexto.

**Actualidad:** La tabla contiene una columna de fecha que nos permite evaluar la actualidad de los datos.

Notebook del análisis https://github.com/Marian2057/Istea-Modelos\_Analiticos

## Métricas para medir el desempeño

**Tipo:** Métrica de Desempeño

**Propósito:** Evaluar el impacto de las estrategias de descuentos y su impacto en las ventas y ganancias.

**Objetivo:** Medir la proporción de ventas con descuentos superiores al promedio.

**Meta:** Proveer a los departamentos de ventas información del impacto de sus campañas de descuentos

**Unidad de Medida:** Puntos porcentuales (%)

**Frecuencia:** Reporte mensual.

**Fuente de Datos:** Base de datos de transacciones de ventas con información detallada sobre descuentos.

**Responsables:** Departamento de Ventas y Analistas de Datos.

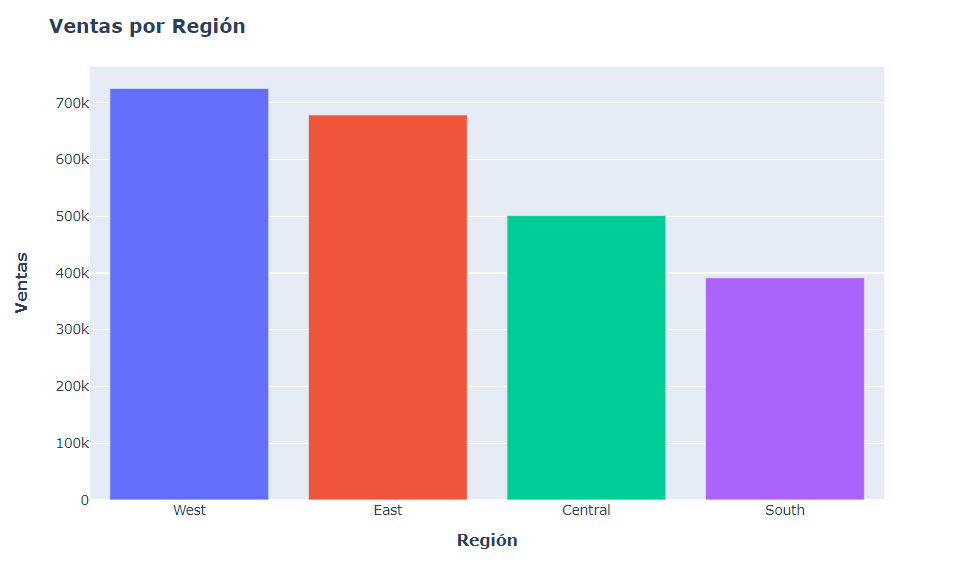
### Ventas Totales por Región

Descripción: *suma total de las ventas en una región específica.*

Fórmula DAX: **Ventas Totales por Región = SUM('Tabla'[Sales])**

****

|  |  |  |
| --- | --- | --- |
|  | **Región** | **Ventas Totales por Región** |
| 0 | Central | 501.239,89 |
| 1 | East | 678.781,24 |
| 2 | South | 391.721,91 |
| 3 | West | 725.457,82 |



El análisis revela que la región West lidera en ventas totales, superando significativamente a las demás regiones, mientras que la región South muestra el nivel más bajo de ventas, indicando disparidades regionales en el rendimiento de las ventas.

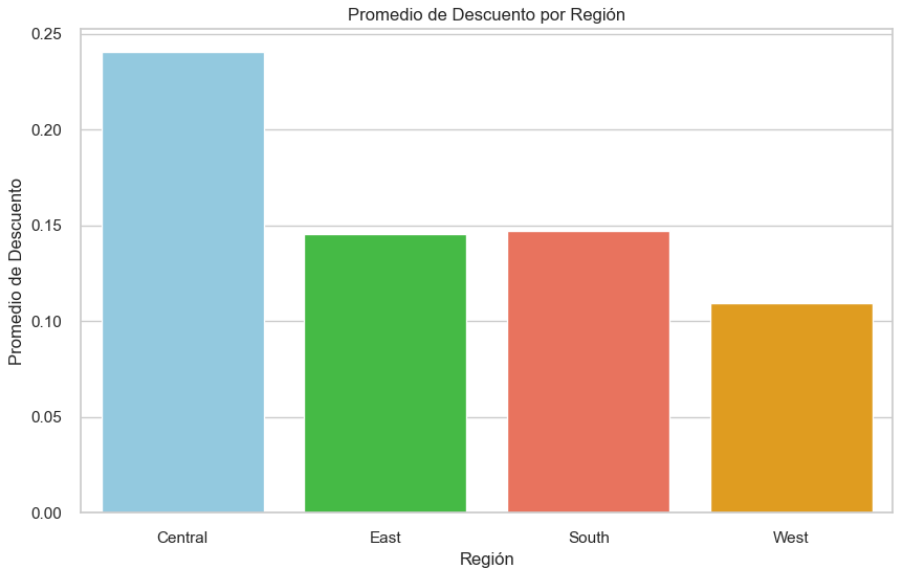
### Promedio de Descuento por Región

Descripción: *descuento promedio aplicado a las ventas en una región.*

Fórmula DAX: **Descuento Promedio por Región = AVERAGE('Tabla'[Discount])**



|  |  |  |
| --- | --- | --- |
|  | **Región** | **Descuento** |
| 0 | Central | 0.24 |
| 1 | East | 0.15 |
| 2 | South | 0.15 |
| 3 | West | 0.11 |



El gráfico muestra que, en promedio, las regiones Central y East aplican descuentos más altos en comparación con las regiones South y West, sugiriendo variaciones en las estrategias de precios entre las diferentes áreas geográficas.

### Efectividad de Descuentos en Ventas por Región

Descripción: *proporción de las ventas realizadas con un descuento superior al promedio en una región.*

Fórmula DAX para Descuento Promedio por Región: **Descuento Promedio por Región = CALCULATE(AVERAGE('Tabla'[Discount]), ALLEXCEPT('Tabla', 'Tabla'[Region]))**

Fórmula DAX para Ventas con Descuento Superior al Promedio: Ventas con Descuento Superior al Promedio = **CALCULATE(SUM('Tabla'[Sales]), FILTER('Tabla', 'Tabla'[Discount] > [Descuento Promedio por Región]))**

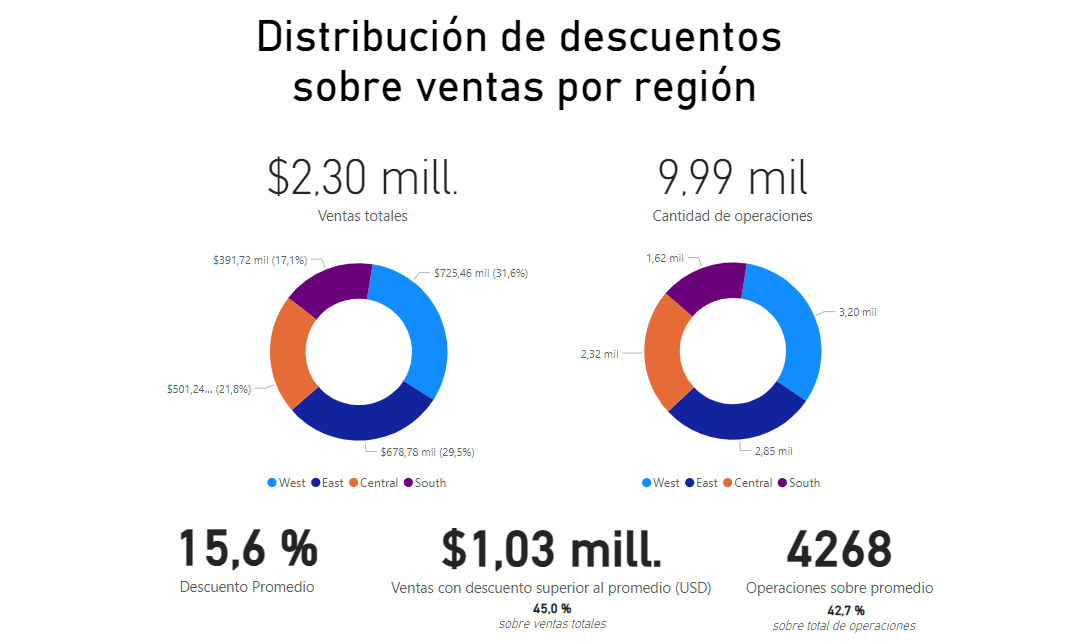
Fórmula DAX para Efectividad de Descuentos en Ventas: Efectividad de Descuentos en Ventas = **DIVIDE([Ventas con Descuento Superior al Promedio], [Ventas Totales por Región])**

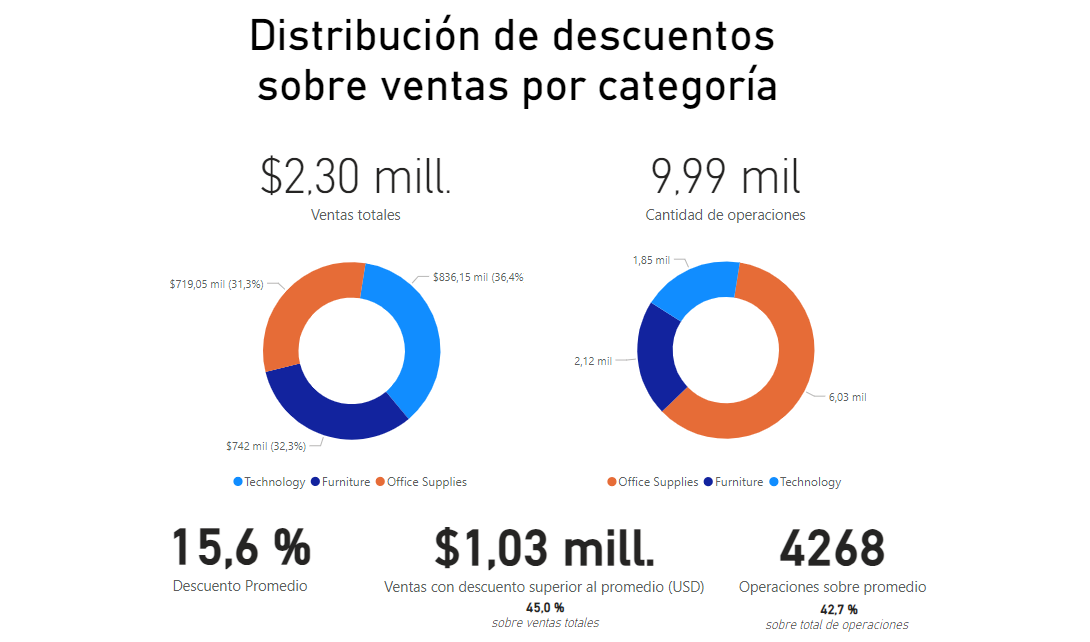


|  |  |  |
| --- | --- | --- |
|  | **Región** | **Ventas** |
| 0 | Central | 0.26 |
| 1 | East | 0.43 |
| 2 | South | 0.45 |
| 3 | West | 0.60 |

El análisis destaca que la región West exhibe la mayor efectividad en descuentos en ventas, alcanzando un notable 60%, mientras que las regiones South y East también muestran una sólida eficacia, con 45% y 43%, respectivamente

### Dashboard





# Conclusión

**Patrones de Compra por Región:** Se observa una clara variación en las ventas totales y promedio de descuentos entre las regiones. La región West lidera en ventas, mientras que las regiones Central y East aplican descuentos más altos en promedio. Este patrón puede influir en las estrategias de precios y marketing adaptadas a cada área geográfica.

**Estrategias Efectivas de Descuentos:** La región West destaca por su alta efectividad en descuentos, alcanzando el 60% de las ventas con descuento superior al promedio. Esto sugiere que las estrategias de descuentos implementadas en esta región son particularmente exitosas. Las regiones South y East también muestran una buena eficacia en este aspecto.

**Desafíos en la Región Central:** Aunque la región Central tiene un alto promedio de descuentos, su efectividad en ventas con descuentos superiores al promedio es menor en comparación con otras regiones. Esto podría indicar la necesidad de revisar y ajustar las estrategias de descuentos para mejorar el rendimiento de las ventas en esta región.

**Enfoque en la Región West:** Dada la alta efectividad en descuentos y las sólidas ventas totales, la región West podría ser un área de enfoque estratégico. Explorar y comprender las prácticas comerciales y preferencias del cliente en esta región puede proporcionar información valiosa para optimizar estrategias a nivel nacional.

**Oportunidades para Personalización:** La variabilidad en los patrones de compra entre regiones sugiere que personalizar las estrategias de marketing y precios según la ubicación geográfica puede ser beneficioso. Adaptar ofertas y descuentos a las preferencias y comportamientos de compra específicos de cada región puede impulsar la rentabilidad y la satisfacción del cliente.

# Bibliografía

dataset https://www.kaggle.com/datasets/vivek468/superstore-dataset-final