



Reporte de Análisis de Ventas

Jesús David Silva Rangel.
Científico de datos.



Descripción del Proyecto

Presentación del proyecto de análisis de datos de ventas, centrado en la exploración, comprensión y hallazgos principales del comportamiento comercial.

Este informe examina el comportamiento de las ventas, ganancias y descuentos en una tienda global. Se busca identificar patrones, correlaciones y oportunidades de mejora a partir de datos históricos de más de 51.000 transacciones.



Objetivo del Proyecto

El propósito de este análisis es evaluar el rendimiento comercial de tres grandes categorías –*Technology*, *Furniture* y *Office Supplies*– mediante métricas clave de ventas, cantidad, ganancia y descuentos.

El estudio se centra en responder:

- ¿Qué categorías generan más ingresos y rentabilidad?
- ¿Qué meses concentran los mayores volúmenes de venta?
- ¿Qué relación existe entre descuentos, costos de envío y ganancias?



Librerías y Configuración

Se emplearon librerías estándar de análisis de datos:

- **NumPy y Pandas** para manipulación de datos.
- **Matplotlib y Seaborn** para visualizaciones.
- **SciPy y StatsModels** para estadística inferencial.

El entorno fue desarrollado en **Jupyter Notebook**, permitiendo reproducibilidad y trazabilidad de los resultados.



Estructura del Conjunto de Datos

- **Filas:** 51.290 registros
- **Columnas:** 17 variables
- **Principales variables:**
- `order_id`, `order_date`, `ship_mode`, `category`, `sales`, `quantity`, `discount`,
`profit`, `shipping_cost`
- **Fuente:** dataset comercial en Kaggle.

Descripción del Conjunto de Datos

Exploración inicial

```
In [2]: # cargar el dataset
df = pd.read_csv('/kaggle/input/sales-data/sales_data.csv')

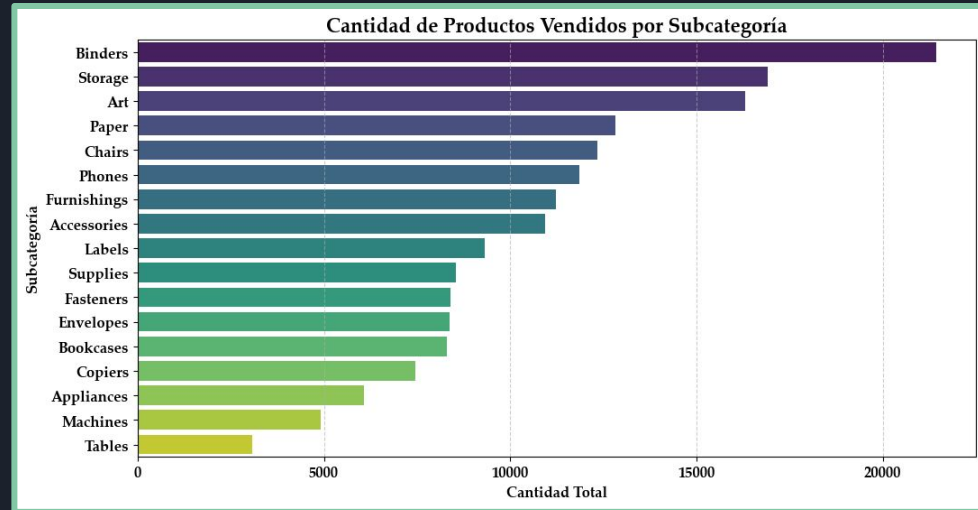
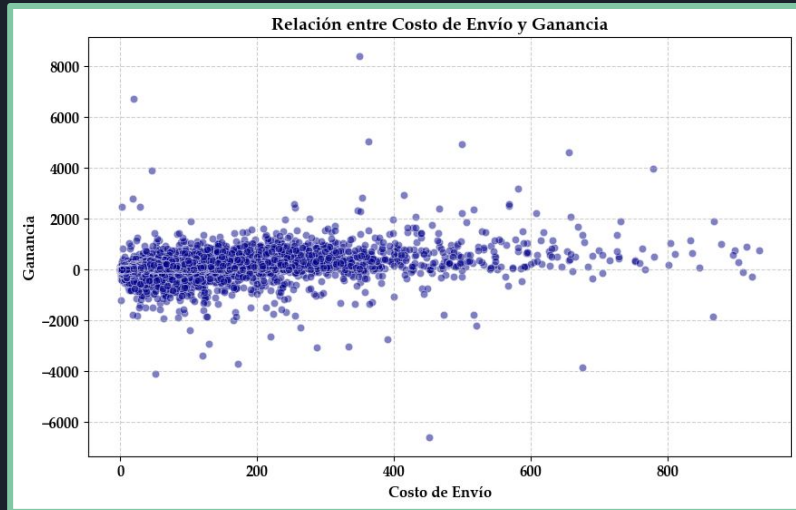
# imprime las primeras 8 filas
df.head(8)
```

```
Out[2]:
```

	row_id	order_id	order_date	ship_date	ship_mode	customer_id	product_id	category	sub_category	product_name	sales	q
0	1	MX-2014-143658	02/10/2014	06/10/2014	Standard Class	SC-20575	OFF-LA-10002782	Office Supplies	Labels	Hon File Folder Labels, Adjustable	13.08	
1	2	MX-2012-155047	15/10/2012	20/10/2012	Standard Class	KW-16570	FUR-FU-10004015	Furniture	Furnishings	Tenex Clock, Durable	252.16	
2	3	MX-2012-155047	15/10/2012	20/10/2012	Standard Class	KW-16570	FUR-BO-10002352	Furniture	Bookcases	Ikea 3-Shelf Cabinet, Mobile	193.28	
3	4	MX-2012-155047	15/10/2012	20/10/2012	Standard Class	KW-16570	OFF-BI-10004428	Office Supplies	Binders	Cardinal Binder, Clear	35.44	
4	5	MX-2012-155047	15/10/2012	20/10/2012	Standard Class	KW-16570	OFF-AR-10004594	Office Supplies	Art	Sanford Canvas, Water Color	71.60	
5	6	MX-2012-155047	15/10/2012	20/10/2012	Standard Class	KW-16570	OFF-EN-10001375	Office Supplies	Envelopes	GlobeWeis Mailers, with clear poly window	56.12	
		MX-					OFF-EN-	OFF-		GlobeWeis		

Análisis Exploratorio de Datos (EDA)

El análisis exploratorio permite conocer la distribución de los datos, detectar valores atípicos y obtener una visión general del comportamiento de las ventas y las cantidades vendidas.





Ventas por Categoría

Se agruparon las ventas por categoría de producto para identificar los grupos que generan mayores ingresos.

El resultado permite conocer cuáles son los más rentables.

Se basa en el cálculo:

```
df.groupby('category')['sales'].sum().round(2).sort_values(ascending=False)
```

```
In [7]: # agrupa por categoría y suma el total de ventas
        print(df.groupby('category')['sales'].sum().round(2).sort_values(ascending = False))
```

category	
Technology	4744557.50
Furniture	4110874.19
Office Supplies	3787070.23

Name: sales, dtype: float64



Panorama General de Ventas

Resumen de ventas por categoría

Categoría	Ventas Totales (USD)	Ganancia Total (USD)	Unidades Vendidas
Tecnología	4.744.557,50	663.778,73	35.176
Muebles	4.110.874,19	285.204,72	34.954
Oficina	3.787.070,23	518.473,83	108.182

Tecnología lidera en ingresos y rentabilidad, mientras que *Oficina* concentra el mayor número de unidades vendidas (61% de los pedidos). El margen más alto corresponde a Technology, con una ganancia promedio por producto de \$65,45.



Panorama General de Ventas

Promedio de ganancias por categoría.

Categoría	Promedio de Ganancia
Tecnología	65.45
Muebles	28.88
Oficina	16.58

La categoría de Tecnología representa una proporción del 59.01% del promedio de ganancias totales.



Cantidad Vendida por Categoría

Este análisis permite identificar los productos con mayor rotación, independientemente del valor monetario de las ventas.

```
In [8]: # agrupa por categoría y suma el total de la cantidad
        print(df.groupby('category')['quantity'].sum().round(2).sort_values(ascending = False))
```

category	
Office Supplies	108182
Technology	35176
Furniture	34954

Name: quantity, dtype: int64



Indicadores Generales

El dataset contiene 51.290 pedidos.

Venta promedio por pedido	Ganancia total	Ganancia promedio por pedido	Órdenes con descuento	Correlación ventas-ganancias
\$ 246.49	\$1,467,457.29	\$ 28.61	43.44%	0.48 correlación positiva moderada

A mayor venta, suele haber mayor ganancia, aunque no siempre proporcional. No se registran envíos gratuitos. El 43,44 % de los pedidos incluyeron algún descuento.



Análisis Temporal

Top 3 meses con más ventas

Marzo	Noviembre	Agosto
\$ 443,014.04	\$ 439,218.48	\$ 435,505.90

Meses con menos ventas

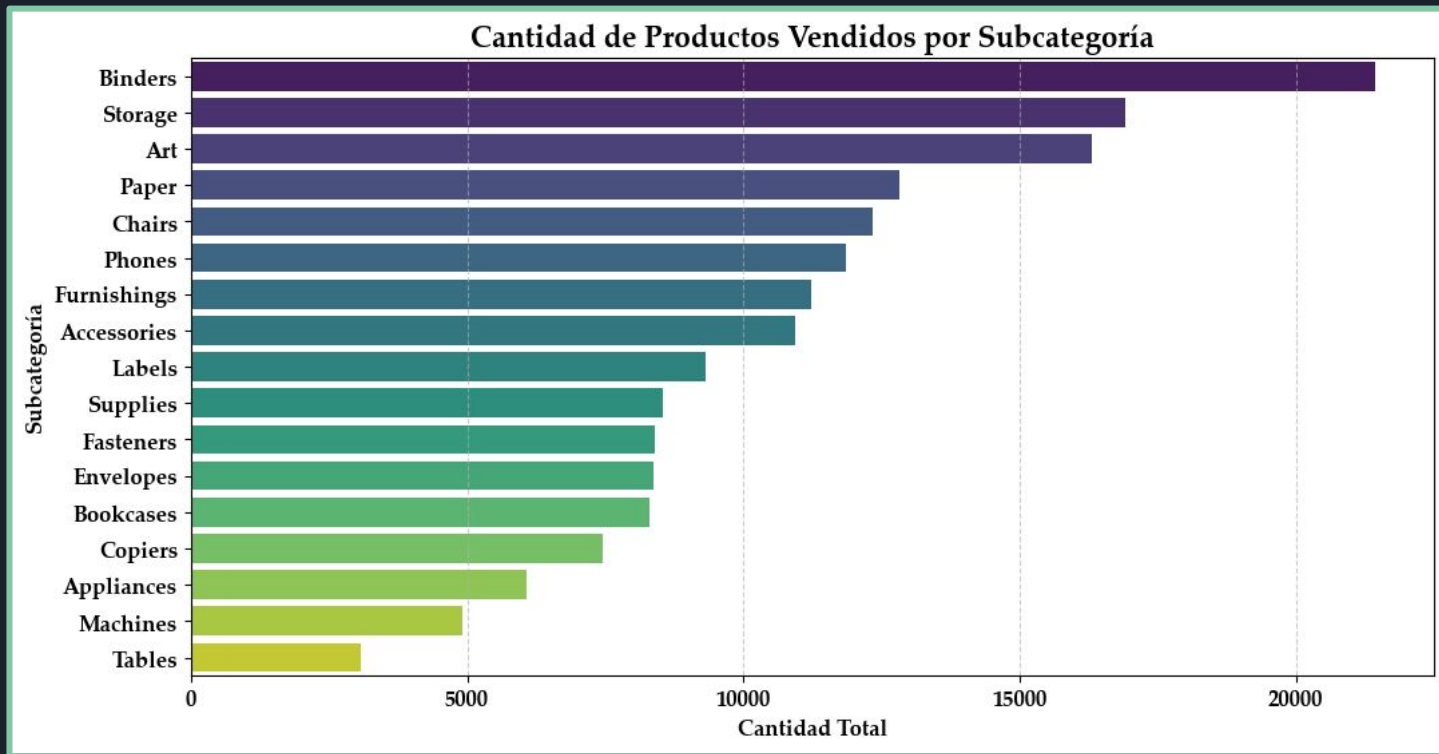
Enero	Septiembre	Junio
\$ 364,011.36	\$ 373,335.36	\$ 382,443.27



Análisis Temporal

El comportamiento mensual muestra una tendencia oscilante, con picos en el primer y último trimestre, lo que sugiere estacionalidad en campañas promocionales.

Subcategorías Destacadas





Subcategorías Destacadas

Las subcategorías con mayor cantidad vendida fueron:

1. **Binders** (más de 22.000 unidades)
2. **Storage**
3. **Art**
4. **Paper**
5. **Chairs**

Estas líneas reflejan alta rotación, aunque no necesariamente los mayores márgenes de ganancia.



Relaciones y Correlaciones

La correlación entre ventas y ganancias fue de 0,48, moderadamente positiva.

La correlación entre descuento y ganancia fue de $-0,316$, lo que indica que los descuentos tienden a reducir la rentabilidad.

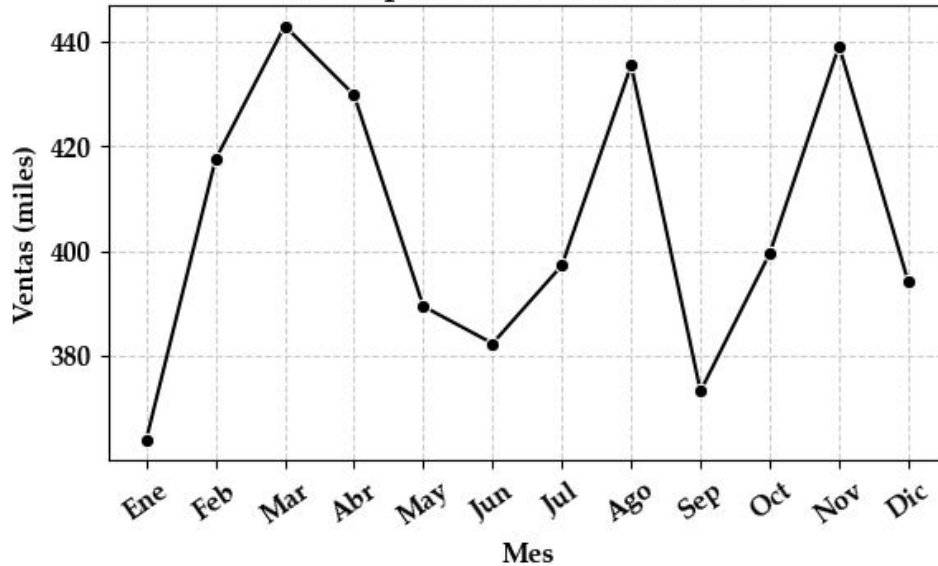
No existe relación significativa entre la categoría de producto y el modo de envío ($\text{Chi}^2 = 2,06$, $p = 0,9143$).

La regresión lineal simple confirma este efecto: cada aumento unitario en el descuento reduce la ganancia en \$259,93 ($p < 0,001$, $R^2 = 0,10$).

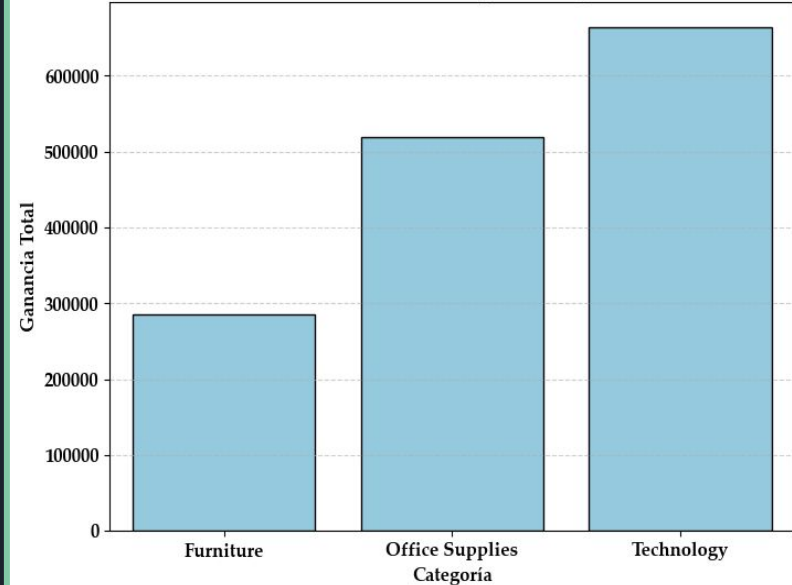
Tendencias de Ventas y Ganancias

Marzo es el mes con mayor proyección de ventas.
Tecnología lidera las ganancias totales por categoría.

Ventas por Mes (total acumulado)



Ganancias Totales por Categoría





Estadísticas Descriptivas

Las ventas presentan alta variabilidad: la mayoría de los pedidos son pequeños (mediana \$85,05), pero unas pocas operaciones elevan el promedio. Solo una de cada cuatro ventas supera la media, lo que refleja una dependencia parcial de pedidos grandes.

El intervalo de confianza (\$242–\$251) confirma la estabilidad del promedio pese a la dispersión.



Pruebas Estadísticas

El negocio mantiene ganancias positivas y significativas, pero los descuentos reducen directamente la rentabilidad (explican un 10 % de su variación).

No hay relación entre categoría y modo de envío, lo que sugiere logística uniforme.

Se recomienda controlar descuentos, diversificar ventas menores y optimizar envíos por tipo de producto.



Conclusión

El negocio es rentable, aunque con márgenes ajustados. *Technology* aporta el mayor valor económico, *Office Supplies* volumen de ventas, y los descuentos excesivos erosionan la ganancia. Se recomienda:

1. Controlar descuentos en categorías de alto valor.
2. Promover ventas cruzadas entre tecnología y suministros.
3. Optimizar costos logísticos para mejorar el margen operativo.



¡Gracias por tu Atención!

Jesús David Silva Rangel.
Científico de datos.