

ANALISIS EXPLORATORIO DE DATOS

LEONARDO JAVIER TORRES VELILLA

UNIVERSIDAD PILOTO DE COLOMBIA

**ASIGNATURA METODOLOGIAS Y HERRAMIENTAS PARA EL
TRATAMIENTO Y LA ANALITICA DE DATOS**

BOGOTA D.C

TABLA DE CONTENIDO

1. Introducción	3
2. Procesamiento y Limpieza de Datos.....	3
3. Análisis Estadístico	4
4. Análisis de Desempeño	5
4.1 Promedio de la tabla datos_combinados.xlsx para las columnas presentadas	6
4.2 Conteo de duplicados, por mayor frecuencia	6
4.3 Índice de correlación de las variables.....	6
4.4 Indicadores Clave Calculados	7
4.5 Clasificación por zonas	7
5. Patrones y Tendencias Identificados	8
6. Interpretación de gráficos	8
7. Conclusiones	24
7.1 Conclusiones Detalladas.....	24
8. Recomendaciones y Análisis Prospectivo.....	25
9. Tabla de gráficos e ilustraciones.....	27

Informe Gerencial: Análisis de Ejercicio unidad 2

1. Introducción

Este informe busca analizar el comportamiento de ventas, ingresos, gastos operativos e inventarios de una organización, agrupando la información por zona y vendedor. El análisis se realizó con base en un archivo de datos transaccionales depurado y consolidado, empleando herramientas estadísticas y visuales.

2. Procesamiento y Limpieza de Datos

Se importaron los datos desde un archivo Excel llamado 20250528_BD_ejercicio_Unidad2. Para este proceso usamos el entorno de Visual Studio Code, para modelar la base, analizar, depurar y dejar lo más limpio posible para un efectivo análisis.

Se ajustaron los nombres de las columnas para facilitar su manejo (UNIDADES VENDIDAS → UNIDADES_VENDIDAS, GASTOS DE OPERACIÓN → GASTOS_OP).

Se aseguraron los tipos de datos numéricos, transformando los valores erróneos o faltantes.

Se agruparon los registros por **ZONA** y **VENDEDOR**, sumando los registros duplicados obteniendo:

- **Suma total** de unidades vendidas, ingresos, gastos operativos e inventarios.
- **Promedios redondeados** de estas variables por agrupación.
- **Frecuencia de registros duplicados** se contaron los registros agrupados por zona y vendedor para identificar cuantas veces se repetían estas variables.
- **Índice de Correlación** se halló el índice de correlación cuantificar la intensidad y dirección de la relación de las variables contenidas en la tabla.

Se creó una nueva base llamada datos_combinados.xlsx de la cual importamos a Python para realizar los respectivos análisis estadísticos y medidas de resumen. La base quedó distribuida de la siguiente forma:

df_Medidas						
	ZONA	VENDEDOR	UNIDADES_VENDIDAS	INGRESOS	GASTOS_OP	INVENTARIOS
0	Zona 1	Vendedor 1	48913	7092385000	244565000	11113
1	Zona 1	Vendedor 2	19038	2760510000	95190000	19473
2	Zona 1	Vendedor 3	24370	3533650000	121850000	14697
3	Zona 1	Vendedor 4	27263	3953135000	136315000	17305
4	Zona 2	Vendedor 1	29496	4276920000	147480000	11910
5	Zona 2	Vendedor 2	17587	2550115000	87935000	11710
6	Zona 2	Vendedor 3	32611	4728595000	163055000	12462
7	Zona 2	Vendedor 4	46690	6770050000	233450000	17723
8	Zona 3	Vendedor 1	25821	3744045000	129105000	18267
9	Zona 3	Vendedor 2	44258	6417410000	221290000	12348
10	Zona 3	Vendedor 3	22265	3228425000	111325000	11135
11	Zona 3	Vendedor 4	27575	3998375000	137875000	19968
12	Zona 4	Vendedor 1	34364	4982780000	171820000	18275
13	Zona 4	Vendedor 2	25066	3634570000	125330000	19687
14	Zona 4	Vendedor 3	23621	3425045000	118105000	10166
15	Zona 4	Vendedor 4	29767	4316215000	148835000	9393

Tabla 1. Base de datos. Fuente propia

3. Análisis Estadístico

3.1 Medidas de Resumen

- **Media, mediana y cuartiles** se calcularon para las columnas:

Métrica	Media	Mediana	Q1	Q2	Q3
Unidades Vendidas	29919	27419	24182	27419	33049
Ingresos	\$ 4.338.264.062	\$ 3.975.755.000	\$ 3.506.498.750	\$ 3.975.755.000	\$ 4.792.141.250
Gastos OP	\$ 149.595.312	\$ 137.095.000	\$ 120.913.750	\$ 137.095.000	\$ 165.246.250
Inventarios	14727	13579	11566	13579	18269

Tabla 2. Medidas de resumen base de datos. Fuente propia

Las cifras son representativas del comportamiento agregado por vendedor y zona.

- **Desviación estándar.**

UNIDADES_VENDIDAS	9406
INGRESOS	\$ 1.363.956.310
GASTOS_OP	\$ 47.032.976
INVENTARIOS	3815

Tabla 3. Desviación estándar medidas base de datos. Fuente propia

4. Análisis de Desempeño

Se identificaron los **vendedores con mayores ingresos por zona**, lo cual sirve como insumo clave para la evaluación de desempeño y políticas de incentivos.

ZONA	VENDEDOR	INGRESOS
ZONA 1	Vendedor 1	\$ 7.092.385.000
ZONA 2	Vendedor 4	\$ 6.770.050.000
ZONA 3	Vendedor 2	\$ 6.417.410.000
ZONA 4	Vendedor 1	\$ 4.982.780.000

Igualmente se identificó a través del promedio las unidades vendidas por zona.

Promedio de Unidades Vendidas por zona	
ZONA 1	29896
ZONA 2	31596
ZONA 3	29979
ZONA 4	28204

4.1 Promedio de la tabla datos_combinados.xlsx para las columnas presentadas

	ZONA	VENDEDOR	UNIDADES_VENDIDAS	INGRESOS	GASTOS_OP	INVENTARIOS
0	Zona 1	Vendedor 1	48913	7092385000	244565000	11113
1	Zona 1	Vendedor 2	19038	2760510000	95190000	19473
2	Zona 1	Vendedor 3	24370	3533650000	121850000	14697
3	Zona 1	Vendedor 4	27263	3953135000	136315000	17305
4	Zona 2	Vendedor 1	29496	4276920000	147480000	11910
5	Zona 2	Vendedor 2	17587	2550115000	87935000	11710
6	Zona 2	Vendedor 3	32611	4728595000	163055000	12462
7	Zona 2	Vendedor 4	46690	6770050000	233450000	17723
8	Zona 3	Vendedor 1	25821	3744045000	129105000	18267
9	Zona 3	Vendedor 2	44258	6417410000	221290000	12348
10	Zona 3	Vendedor 3	22265	3228425000	111325000	11135
11	Zona 3	Vendedor 4	27575	3998375000	137875000	19968
12	Zona 4	Vendedor 1	34364	4982780000	171820000	18275
13	Zona 4	Vendedor 2	25066	3634570000	125330000	19687
14	Zona 4	Vendedor 3	23621	3425045000	118105000	10166
15	Zona 4	Vendedor 4	29767	4316215000	148835000	9393

Tabla 4. Promedio variables base de datos. Fuente propia

4.2 Conteo de duplicados, por mayor frecuencia

	ZONA	VENDEDOR	FRECUENCIA
0	Zona 1	Vendedor 1	7
1	Zona 1	Vendedor 2	6
2	Zona 1	Vendedor 3	6
3	Zona 1	Vendedor 4	6
4	Zona 2	Vendedor 1	6
5	Zona 2	Vendedor 2	6
6	Zona 2	Vendedor 3	6
7	Zona 2	Vendedor 4	6
8	Zona 3	Vendedor 1	5
9	Zona 3	Vendedor 2	6
10	Zona 3	Vendedor 3	6
11	Zona 3	Vendedor 4	6
12	Zona 4	Vendedor 1	6
13	Zona 4	Vendedor 2	6
14	Zona 4	Vendedor 3	5
15	Zona 4	Vendedor 4	5

Tabla 5. Duplicados base de datos. Fuente propia

4.3 Índice de correlación de las variables.

=== Matriz de correlación para las columnas numéricas ===

	UNIDADES_VENDIDAS	INGRESOS	GASTOS_OP	INVENTARIOS
UNIDADES_VENDIDAS	1.000000	1.000000	1.000000	-0.105123
INGRESOS	1.000000	1.000000	1.000000	-0.105123
GASTOS_OP	1.000000	1.000000	1.000000	-0.105123
INVENTARIOS	-0.105123	-0.105123	-0.105123	1.000000

Tabla 6. Índice correlación de las variables de la base de datos. Fuente propia

4.4 Indicadores Clave Calculados

Se agruparon los datos por cada zona, se sumaron los ingresos, gastos operativos y unidades vendidas por zona.

Agregamos dos nuevas columnas nuevas llamadas **INGRESOS_POR_GASTO** la cual calcula cuántos ingresos se generan por cada peso invertido en operación; y **INGRESOS_POR_UNIDAD**, que son los ingresos generados por cada unidad vendida.

ZONA	UNIDADES_VENDIDAS	INGRESOS	GASTOS_OP	INGRESOS_POR_GASTO	INGRESOS_POR_UNIDAD
Zona 1	119584	\$ 17.339.680.000	\$ 597.920.00	29.0	145000
Zona 2	126384	\$ 18.325.680.000	\$ 631.920.00	29.0	145000
Zona 3	119919	\$ 17.388.255.000	\$ 599.595.00	29.0	145000
Zona 4	112818	\$ 16.358.610.000	\$ 564.090.00	29.0	145000

Tabla 7. Indicadores clave de las variables de la base de datos. Fuente propia

4.5 Clasificación por zonas

Se agruparon las zonas, calculamos la eficiencia operativa y por unidad para establecer zonas como: líderes, intermedia, crítica.

ZONA	UNIDADES_VENDIDAS	INGRESOS	GASTOS_OP	INGRESOS_POR_GASTO	INGRESOS_POR_UNIDAD	CLASIFICACION
Zona 2	126384	\$ 18.325.680.000	\$ 631.920.00	29.0	145000	Líder
Zona 3	119919	\$ 17.388.255.000	\$ 599.595.00	29.0	145000	Intermedia
Zona 1	119584	\$ 17.339.680.000	\$ 597.920.00	29.0	145000	Intermedia
Zona 4	112818	\$ 16.358.610.000	\$ 564.090.00	29.0	145000	Crítica

Tabla 8. Clasificación por zonas de las variables de la base de datos. Fuente propia

5. Patrones y Tendencias Identificados

- **Relación directa** entre ingresos y unidades vendidas (lógica comercial esperada).
- **Zonas con mayor rotación de inventario** muestran también mayores ingresos.
- **Gastos operativos altos** en algunos casos no se traducen en mayores ventas, lo cual sugiere oportunidad de eficiencia.
- Existe **variabilidad significativa** entre zonas y vendedores, especialmente en ingresos e inventarios.

6. Interpretación de gráficos

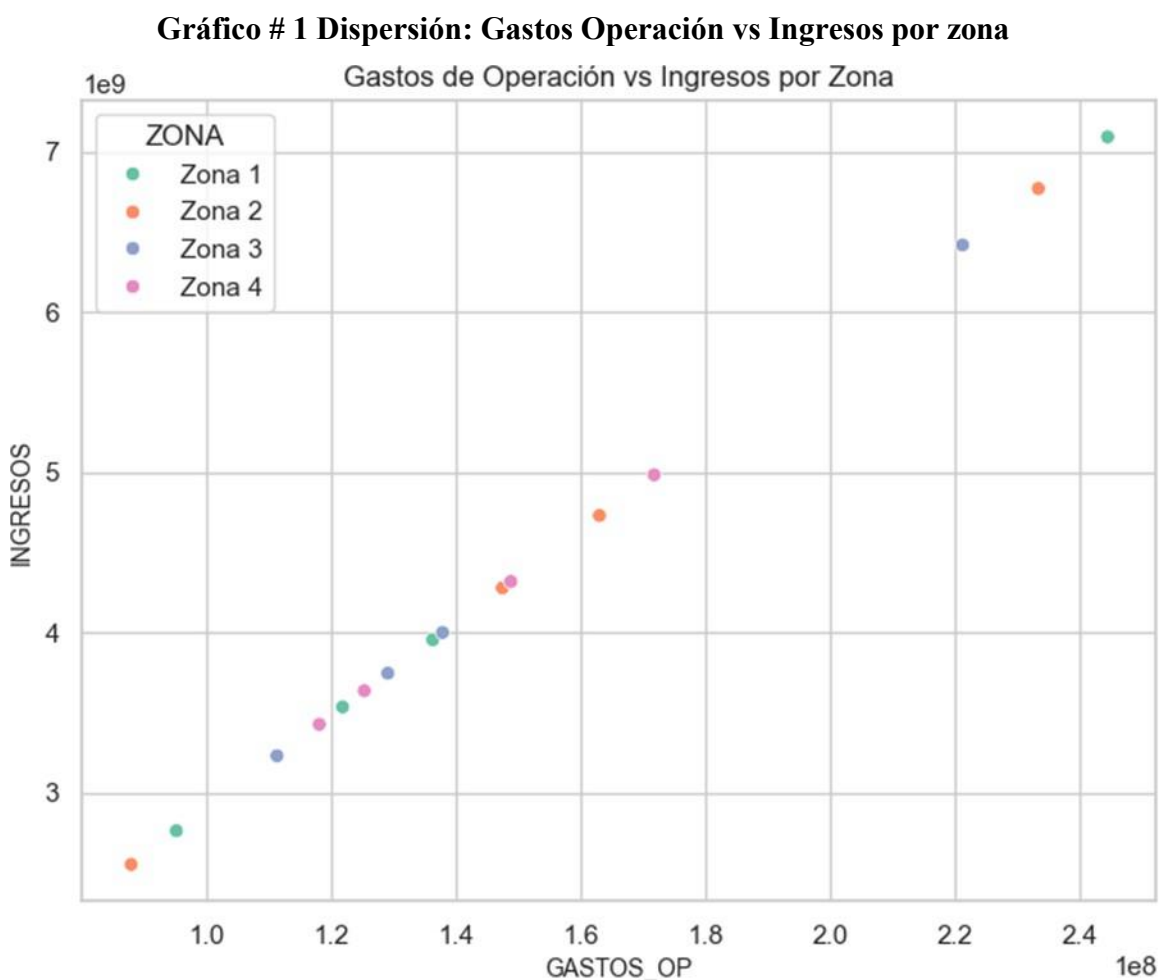


Ilustración 1. Gráfico # 1. (Elaboración propia)

Tipo: Gráfico de Dispersión.

Variables: GASTOS_OP vs INGRESOS por zona(hue).

Implicación:

Este grafico es útil para evaluar el retorno de la inversión operativa según zonas y detectar zonas que requieren intervención o replicar estrategias exitosas.

Observación:

Existe una fuerte correlación positiva a medida que los gastos operativos aumentan, también lo hacen los ingresos. Se puede observar que los puntos se agrupan linealmente, lo que sugiere que la inversión en operaciones genera retorno.

Diferencias por zona: algunas zonas (como Zona 1), tienen una **alta densidad de puntos**, lo que podría indicar una mayor concentración de vendedores o transacciones. De igual manera en otras zonas, aunque con menos puntos, se presentan niveles más altos de ingresos con gastos relativamente similares, lo cual es señal de una mayor **eficiencia operativa**.

Análisis:

En el gráfico podemos observar la relación entre gastos operativos e ingresos en distintas zonas, al usar color por zona ('ZONA=hue') permite identificar la existencia de patrones distintos. Se puede evaluar eficiencia de más ingresos con menos gasto.

Interpretación:

1. Esta relación indica que puede llegar a ser rentable aumentar el gasto, siempre y cuando se administre correctamente (más gasto = más ingreso).
2. Las zonas que logran altos ingresos con menores gastos deben ser estudiadas para replicar su estrategia en otras áreas.

GRAFICO # 2 Tendencia de Ingresos a lo Largo de las Zonas

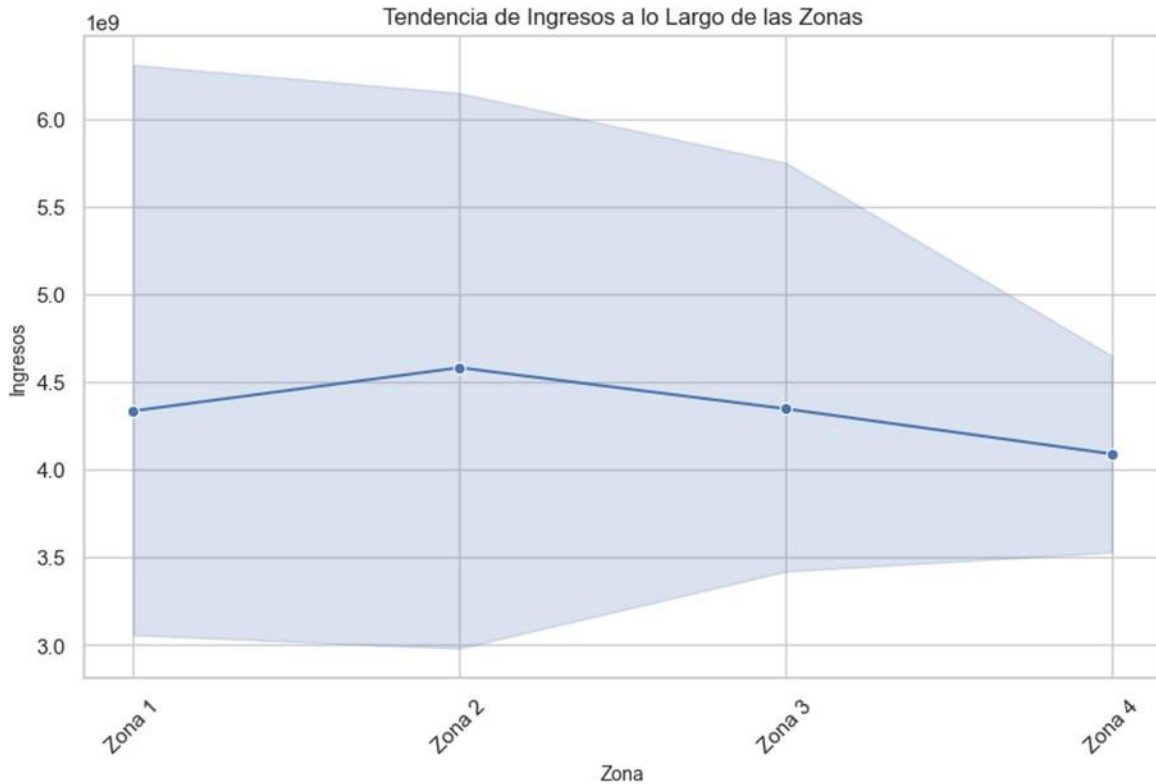


Ilustración 2. Gráfico # 2 (Elaboración propia).

Tipo: Gráfico Lineal

Variables: ZONA vs INGRESOS

Implicación:

Este gráfico sirve para identificar zonas de alto rendimiento o aquellas que necesitan mejoras. De la misma manera puede orientar políticas de redistribución o estrategias de desarrollo por región o zona.

Análisis:

Este gráfico muestra cómo varían los ingresos promedio en cada zona. Cada punto representa una zona, y su altura indica el nivel de ingresos correspondiente. La línea conecta estos puntos para visualizar la tendencia general.

Interpretación:

1. Importancia del orden:
 - ✓ Si las zonas están ordenadas alfabéticamente, el gráfico puede parecer irregular

- ✓ Para una mejor interpretación conviene ordenar las zonas de mayor a menor ingreso promedio, así se observa cuáles son las más rentables.
2. Variación entre zonas:
- ✓ Las zonas no tienen ingresos similares entre sí.
 - ✓ Hay algunas con ingresos significativamente más altos que otras.
3. Análisis de patrones:
- ✓ Al observarse una caída progresiva, puede sugerir una desigualdad estructural en el desempeño por zona.
 - ✓ Si hay subidas y bajadas abruptas, puede indicar la irregularidad, lo que justifica revisar factores como liderazgo, estrategias locales o condiciones del mercado.

GRAFICO # 3 Desviación Estándar de Variables Numéricas

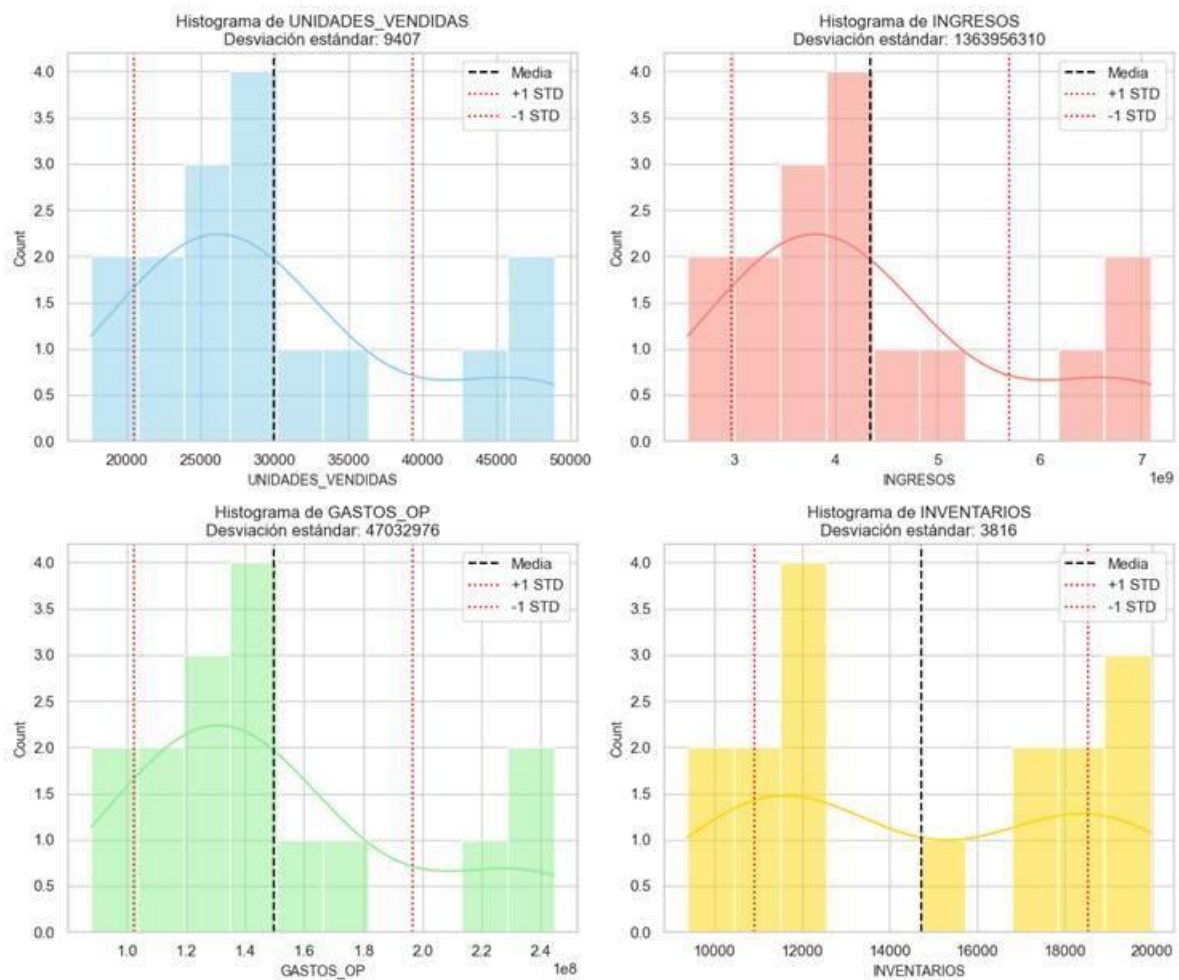


Ilustración 3. Gráfico # 3 (Elaboración propia).

Tipo: Histograma con KDE y línea de desviación estándar

Variables: Desviación estándar de UNIDADES VENDIDAS, INGRESOS, GASTOS_OP, INVENTARIOS con línea de media y su desviación estándar.

Implicación:

Este gráfico ayuda a entender la estabilidad de diferentes indicadores, por ejemplo: ingresos, costos, utilidades. Es especialmente útil para evaluar el riesgo, variables con alta desviación podrían representar áreas donde se requiere más control.

Análisis:

1. UNIDADES_VENDIDAS:
 - Se observa una ligera simetría.
 - Desviación estándar moderada: hay una variedad razonable entre vendedores o zonas, pero sin extremos.
2. INGRESOS:
 - Distribución sesgada a la derecha con algunos casos muy altos.
 - Desviación estándar alta: variabilidad importante entre ingresos, posiblemente por diferencias entre zonas o vendedores.
3. GASTOS_OP:
 - Distribución sesgada: algunos puntos con gastos muy altos.
 - Alta desviación estándar, aunque con un núcleo concentrado alrededor de la media.
4. INVENTARIOS:
 - Distribución más simétrica y controlada.
 - Baja desviación estándar, lo que sugiere que los niveles de inventario son más estandarizados o regulados.

Interpretación:

- Las variables de ingresos y gastos muestran alta variabilidad, lo que indica oportunidades para optimizar.
- Las unidades vendidas son algo más homogéneas, y los inventarios parecen estar bien controlados.
- Las áreas con más dispersión deben ser foco de análisis para reducir variabilidad no deseada o capitalizar eficiencias.

GRAFICO # 4 INGRESOS_POR_GASTO e INGRESOS_POR_UNIDAD en zonas

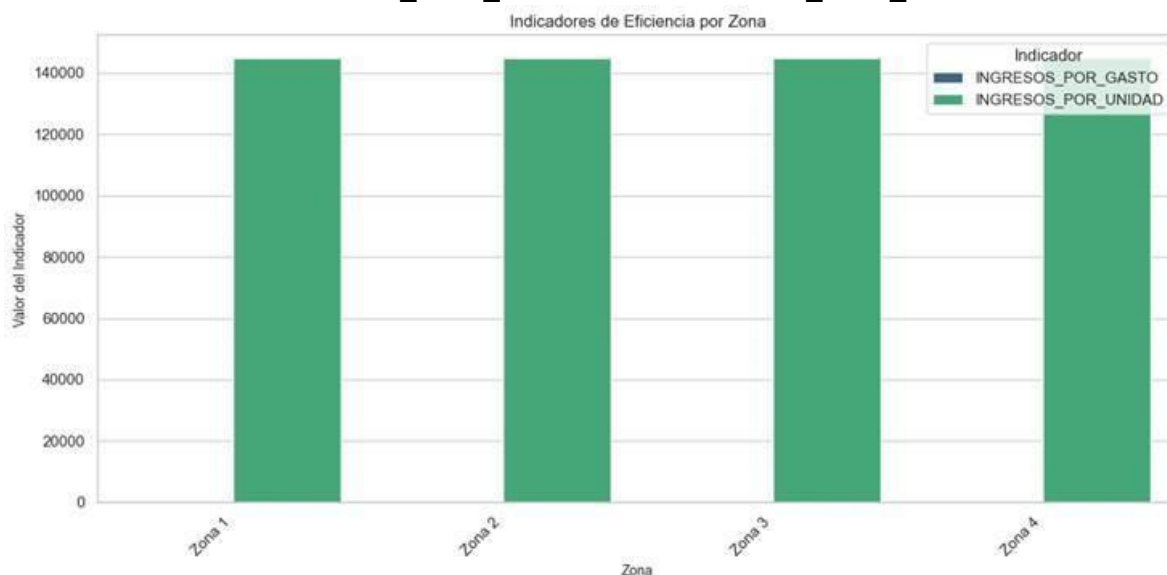


Ilustración 4. Gráfico # 4 (Elaboración propia).

Tipo: Grafico de barras

Variables: Ingresos por gasto, ingresos por unidad, zonas

Observación:

1. El gráfico muestra que todos los valores de INGRESOS_POR_GASTO son iguales (29.0) para las cuatro zonas.
2. De igual manera, los valores de INGRESOS_POR_UNIDAD son constantes en 145.000 en todas las zonas.
3. No existe variabilidad entre zonas para ninguno de los dos indicadores presentados.

Implicación:

La homogeneidad absoluta de estos indicadores implica que:

- Los procesos operativos están estandarizados, lo que puede ser positivo desde una perspectiva de control de calidad y eficiencia.
- Las zonas no tienen autonomía en precios ni en estrategias de optimización de gastos, o si la tienen, no la están aprovechando para mejorar sus resultados. También indica que las diferencias de desempeño entre zonas no se deben a eficiencia ni precio, sino exclusivamente al volumen de ventas.

Interpretación:

- **INGRESOS_POR_GASTO = 29.0** significa que, por cada peso gastado en operación, cada zona genera 29 pesos en ingresos. Esto indica una eficiencia operativa muy alta y homogénea en toda la organización.

- **INGRESOS_POR_UNIDAD = 145,000** sugiere que el ingreso generado por cada unidad vendida es exactamente el mismo en todas las zonas. Esto puede reflejar una política de precios uniforme a nivel nacional o por segmento.

Análisis:

Si el objetivo es mejorar la rentabilidad total, no hay margen para mejorar eficiencia operativa o precio por unidad dado a que ya están maximizados o fijados. Por lo tanto:

- ✓ La empresa debería enfocarse en aumentar las unidades vendidas, especialmente en zonas con menor volumen, por ejemplo, la Zona 4.
- ✓ Sería útil investigar barreras de mercado específicas en cada zona (competencia, distribución, demanda, etc.) ya que el problema no es interno (costos/precios) sino externo (ventas).

GRAFICO # 5 Gráfico de barras para mostrar la clasificación de zonas

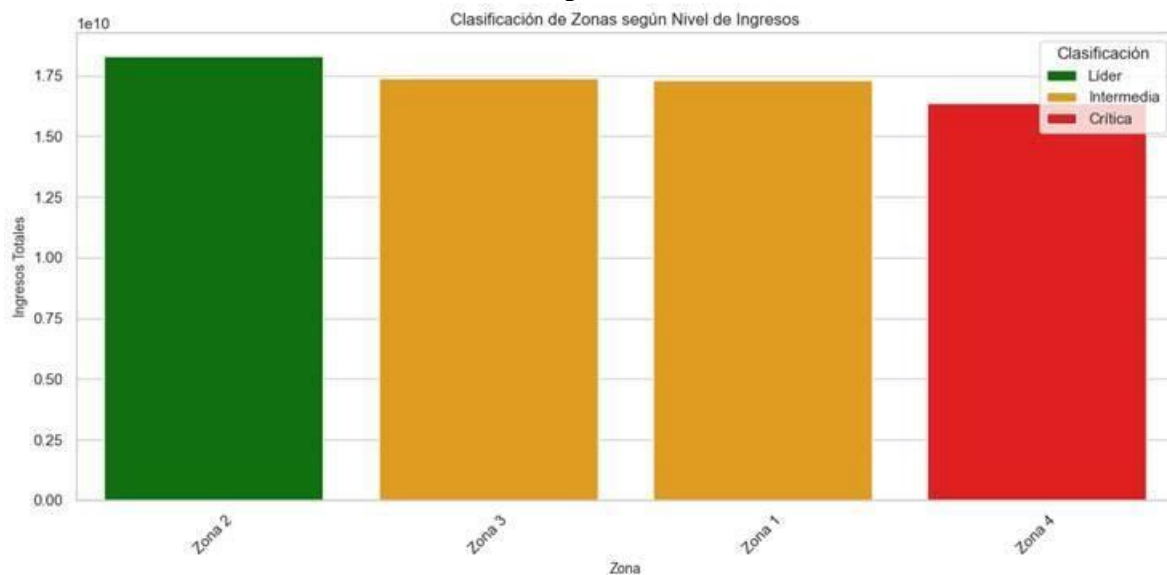


Ilustración 5. Gráfico # 5 (Elaboración propia).

Tipo: Grafico de barras.

Variables: Ingresos Totales, clasificación, zonas.

Observación:

Las zonas están ordenadas de mayor a menor ingreso total. Se destacan claramente tres grupos de zonas:

- **Zonas líderes** (en verde): con ingresos significativamente más altos.
- **Zonas intermedias** (naranja): con desempeño aceptable, pero sin destacar.
- **Zonas críticas** (rojo): con los ingresos más bajos del conjunto.

Análisis:

1. Concentración de desempeño:
 - ✓ Un pequeño grupo de zonas genera una porción desproporcionada de los ingresos.
 - ✓ Esto implica dependencia: si una zona líder falla, puede afectar significativamente el total.
2. Homogeneidad en eficiencia, heterogeneidad en volumen:
 - ✓ Como vimos antes, el ingreso por unidad y por gasto es similar en todas las zonas.
 - ✓ Esto indica que las diferencias no se deben a ineficiencia operativa, sino a cantidad de ventas o tamaño de mercado.
3. Clasificación útil para estrategia:
 - ✓ Puedes usar esta segmentación para aplicar políticas comerciales diferenciadas.
 - ✓ Por ejemplo, apoyar zonas críticas con campañas, capacitación o revisión de oferta.

Interpretación:

Las zonas líderes concentran gran parte de los ingresos, lo que sugiere un mayor volumen de ventas. Posiblemente mayor demanda, mejor logística o ejecución comercial más eficiente.

Las zonas críticas están considerablemente por debajo del resto. Esto puede deberse a factores internos (bajo desempeño de equipo) o externos (mercado pequeño, problemas de acceso, etc.).

La diferencia entre zonas líderes y críticas es marcada y estructural, no aleatoria.

GRAFICO # 6 Promedio De Unidades Vendidas Por Zona.

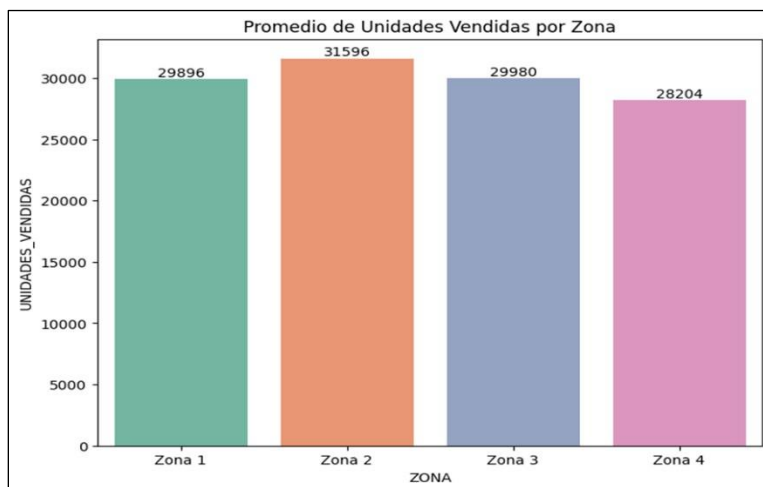


Ilustración 6. Gráfico# 6 (Elaboración propia).

Tipo: Barplot (Grafica de barras).

Variables: ZONA vs UNIDADES VENDIDAS (promedio).

Interpretación:

- ✓ Permite comparar el desempeño comercial entre zonas.
- ✓ Zonas con mayor promedio indican mayor demanda o mejor gestión de ventas.
- ✓ Útil para reasignación de recursos o metas.

En este grafico podemos evidenciar el promedio de la cantidad de unidades vendidas (Eje Y) por cada Zona (Eje X). Donde podemos concluir que la zona que más tuvo ventas fue la Zona 2, con un promedio de Unidades Vendidas de 31.596. Como producto de la eficiente labor que están haciendo sobre los clientes para incrementar las ventas. Se puede observar que el promedio de unidades vendidas por zona es el siguiente, Zona 1 = 29.896, Zona 3 = 29.979 y Zona 4 = 28.204. En la Zona 4 se evidencia que hay un bajo promedio de unidades vendidas, lo que genera un desafío y por consiguiente se debe realizar un análisis exhaustivo para encontrar las diferentes razones y proponer indicadores que mejoren el promedio.

GRAFICO # 7 Vendedor Con Más Ingresos Por Zona.

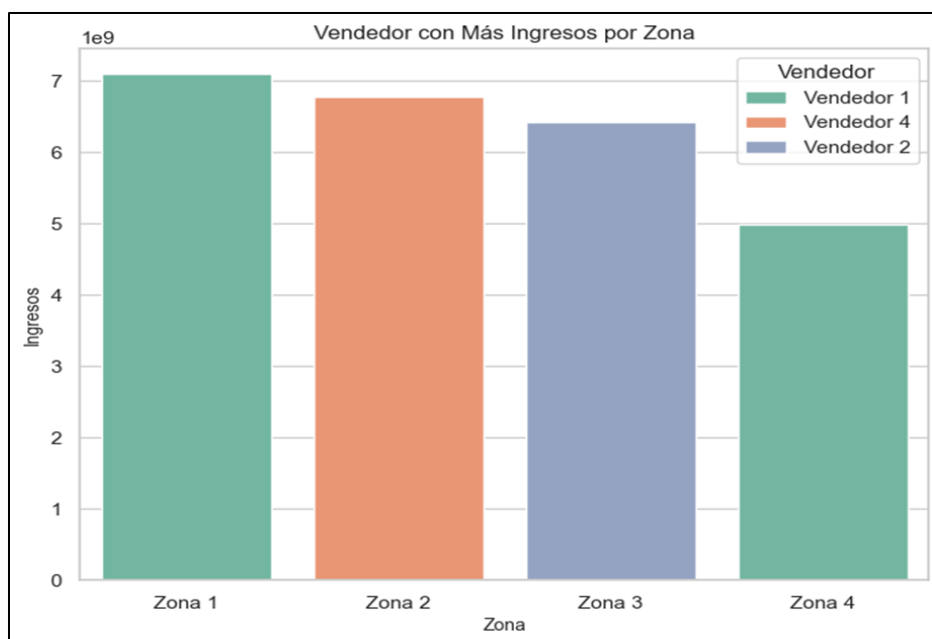


Ilustración 7. Gráfico # 7 (Elaboración propia).

Tipo: Barplot (Grafico de Barras).

Variables: ZONA, INGRESOS y VENDEDOR.

Interpretación:

- ✓ Destaca al mejor vendedor en cada zona.
- ✓ Puede ayudar a identificar buenas prácticas comerciales.

Se evidencia que el Vendedor 1 fue quien más tuvo ingresos en la Zona 1, con un total de \$ 7.092.385.000. Se puede inferir que el Vendedor 1 es el más representativo para la empresa, ya que también ocupa el primer lugar en la Zona 4 con ingresos igual a \$ 4.982.780.000. En la zona 2 el mejor fue el Vendedor 4 con un total de ingresos igual a \$ 6.770.050.000, y en la Zona 3 el mejor fue el Vendedor 2 con un total de ingresos igual a \$ 6.417.410.000. Se observa que hay inconvenientes para las ventas en la Zona 4 con una diferencia amplia en la cantidad de ingresos, comparado con las otras zonas.

GRAFICO # 8 Inventario de Cada Vendedor por Zona.

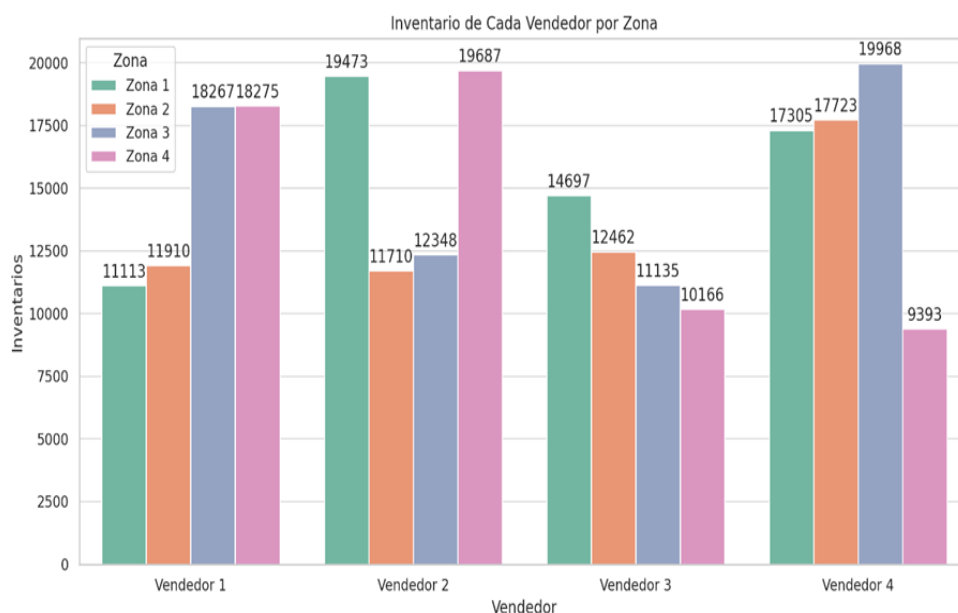


Ilustración 8. Gráfico # 8 (Elaboración propia).

Tipo: Barplot (Grafico de Barras).

Variables: VENDEDOR, INVENTARIOS y ZONA.

Interpretación:

- ✓ Se aprecian diferencias claras en el volumen de inventario manejado, tanto entre vendedores como entre zonas.
- ✓ Permite identificar los vendedores con inventarios excesivos o insuficientes. Y la Diferencia de stock entre zonas.
- ✓ En las diferentes zonas se puede estar aplicando estrategias logísticas distintas.
- ✓ Es fundamental alinear los inventarios con las proyecciones reales de ventas.
- ✓ Se pueden establecer **indicadores de rotación de inventario y políticas de reposición más precisas** por vendedor y zona.

Al observar el grafico de barras se puede deducir que el vendedor 3 es quien tiene la cantidad total más baja de inventarios con una sumatoria total de 48.460 unidades, dividido de la siguiente forma, en la zona 1 = 14.697, zona 2 = 12.462, zona 3 = 11.135 y zona 4 = 10.166. Respecto a los otros vendedores, es quien tiene menos cantidad de inventarios.

Otro factor relevante es que el menor número de unidades de inventario la tiene el vendedor 4 en su zona 4, con un total de 9393 unidades.

GRAFICO # 9 Histogramas de Variables Numéricas

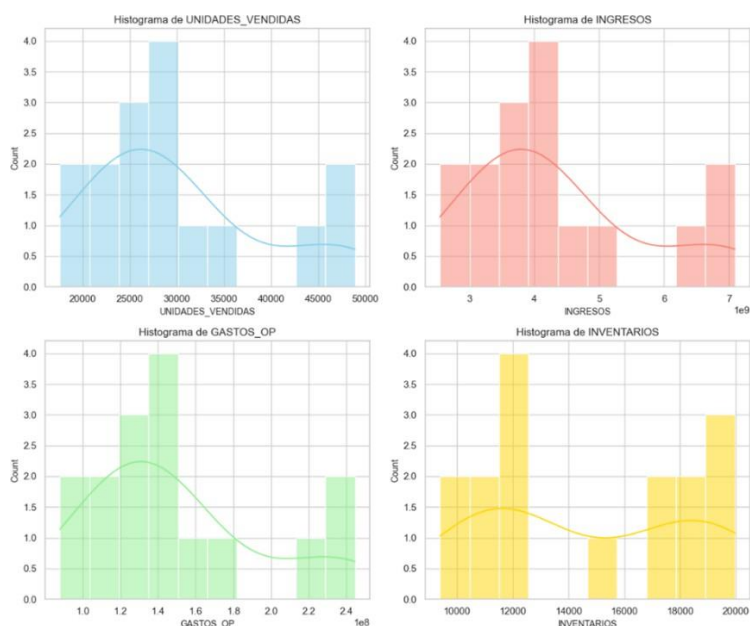


Ilustración 9. Gráfico # 9 (Elaboración propia).

Tipo: Histogramas con KDE (curva de densidad)

Variables: Cada una de las 4 variables cuantitativas.

Análisis:

- Muestran la distribución de datos (asimetría, curtosis).
- Ayudan a detectar valores atípicos o sesgos.

Interpretación:

- Algunas variables presentan distribución sesgada a la derecha (por ejemplo, ingresos).
- Posible existencia de valores extremos (outliers) en inventarios.

UNIDADES VENDIDAS:

- Distribución: Levemente sesgada a la derecha (asimetría positiva).
- Media: Alrededor de 31,000 (línea negra punteada).
- Desviación estándar: 9,407 (indicada en la parte superior).
- La mayoría de los datos se agrupan entre 20,000 y 35,000.
- Algunos valores altos (cerca de 50,000) podrían considerarse atípicos.
- Aunque el histograma no es perfectamente normal, sigue una forma razonablemente simétrica con cierta dispersión.

INGRESOS

- Distribución: Claramente sesgada a la derecha (asimetría positiva).
- Media: Aproximadamente 4.1×10^9 .
- Desviación estándar: Muy alta, 1.36×10^9 .
- Hay varios valores altos extremos (hasta 7×10^9), lo que sugiere la presencia de outliers.
- La mayoría de los datos está concentrada en el rango bajo (entre 3 y 5×10^9), con algunos ingresos atípicos que elevan la media.
- Hay pocas unidades de negocio o periodos excepcionales con ingresos muy altos.
- Estos valores extremos aumentan artificialmente la media, reduciendo su representatividad como indicador central.
- La gran dispersión (STD alta) limita la capacidad de predicción si se asume una distribución normal.

GASTOS_OP (Gastos Operativos)

- Distribución: Leve sesgo a la derecha.
- Media: Cerca de 1.55×10^8 .
- Desviación estándar: 4.7×10^7 .
- La mayoría de los valores están en el rango de 1.2×10^8 a 1.8×10^8 .
- Hay algunos valores extremos por encima de 2×10^8 , que podrían influir en la media.

- La dispersión es considerable, pero sin ser tan extrema como en ingresos.

INVENTARIOS

- Distribución: Bimodal o uniforme, no se ajusta a una curva normal típica.
- Media: Alrededor de 14,000.
- Desviación estándar: 3,816.
- Distribución más plana y dispersa.
- Hay agrupamientos claros alrededor de los 12,000 y 18,000, lo que podría indicar estacionalidad o diferentes políticas de inventario entre unidades.

GRAFICO # 10 Distribución de Inventarios por Zona (Boxplot)

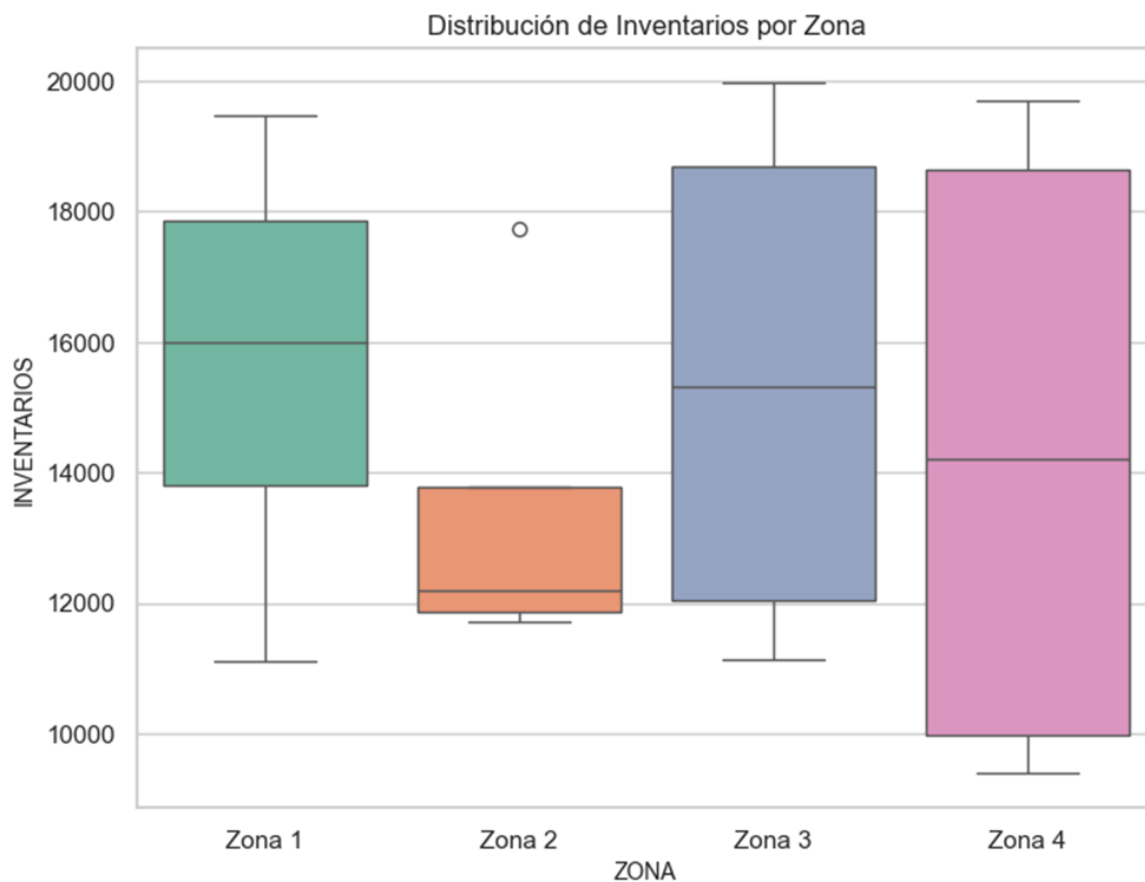


Ilustración 10. Gráfico # 10 (Elaboración propia).

Tipo: Boxplot

Variables: INVENTARIOS por ZONA.

Análisis:

- Visibiliza la mediana, cuartiles y valores extremos.
- Eje Y: Cantidad de inventarios.
- Eje X: Zonas (1, 2, 3 y 4).
- Cada caja representa la distribución del inventario en cada zona:
- Línea horizontal dentro de la caja: Mediana (valor central).
- Límites de la caja: Primer y tercer cuartil (Q1 y Q3).
- Bigotes, permiten visualizar el rango mínimo y máximo dentro de 1.5 veces.
- Punto aislado (Zona 2): Valor atípico (outlier).

Interpretación:

- Algunas zonas tienen inventarios con alta dispersión o valores atípicos.
- Ideal para ajustes logísticos y de almacenamiento.

Zona 1

- Rango intercuartílico amplio, indicando alta variabilidad interna.
- Mediana: Aproximadamente 16,000.
- Distribución equilibrada, con pocos valores extremos.
- Es una zona con inventarios altos y cierta estabilidad, aunque con dispersión moderada.

Zona 2

- La más baja en inventarios, con mediana cerca de 12,000.
- Caja estrecha: Poca variabilidad (inventarios más consistentes).
- Inventario inusualmente alto (17,500),
- Es una zona más controlada en inventarios, pero con una excepción notable que debe investigarse.

Zona 3

- Mediana: Aproximadamente 15,000.
- Rango muy amplio: Variabilidad alta (desde 11,500 hasta 20,000).
- Caja amplia y bigotes extendidos: Indica comportamientos heterogéneos.
- Es una zona con comportamientos mixtos; algunas regiones podrían estar acumulando exceso de inventario o enfrentando desabasto.

Zona 4

- Distribución más dispersa de todas.
- Mediana: Alrededor de 14,000.
- Valores van desde casi 9,000 hasta 20,000.
- Es una zona con alto grado de inestabilidad en inventarios. Podría haber problemas de planificación o diferencias logísticas significativas entre subregiones.

GRAFICO # 11 Matriz de Correlación entre Variables Numéricas

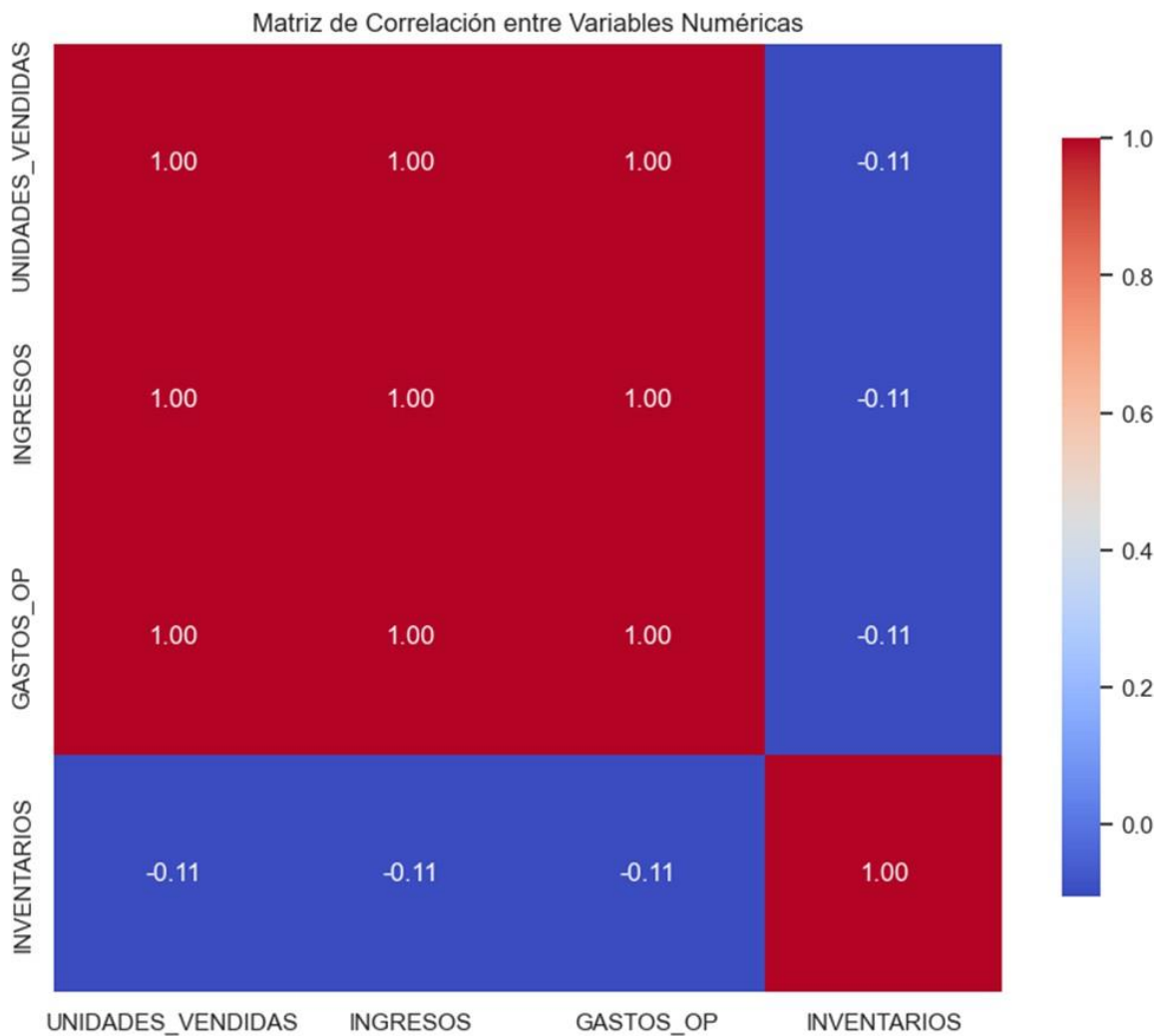


Ilustración 11. Gráfico # 11 (Elaboración propia).

Tipo: Heatmap

Variables: Correlación entre UNIDADES_VENDIDAS, INGRESOS, GASTOS_OP, INVENTARIOS.

Análisis:

- Identifica relaciones lineales entre variables.
- Valores cercanos a 1 o -1 indican alta correlación.

Interpretación:

- Posible alta correlación entre ingresos y unidades vendidas.

- Débil correlación entre inventarios y gastos en algunas zonas.
- UNIDADES_VENDIDAS, INGRESOS y GASTOS_OP tienen una correlación perfecta positiva (1.00) entre sí. Esto indica que cuando aumentan las unidades vendidas, también aumentan los ingresos y los gastos operativos en la misma proporción.
- INVENTARIOS tiene una correlación negativa débil con todas las demás variables (-0.11), lo que indicaría o podría optimizarse es que los niveles de inventario no están directamente relacionados con ventas, ingresos o gastos operativos.
- Los inventarios no se ajustan de forma dinámica a la demanda o que hay ineficiencias en la gestión de stock.

GRAFICO # 12 Boxplot Comparativo de Variables Numéricas

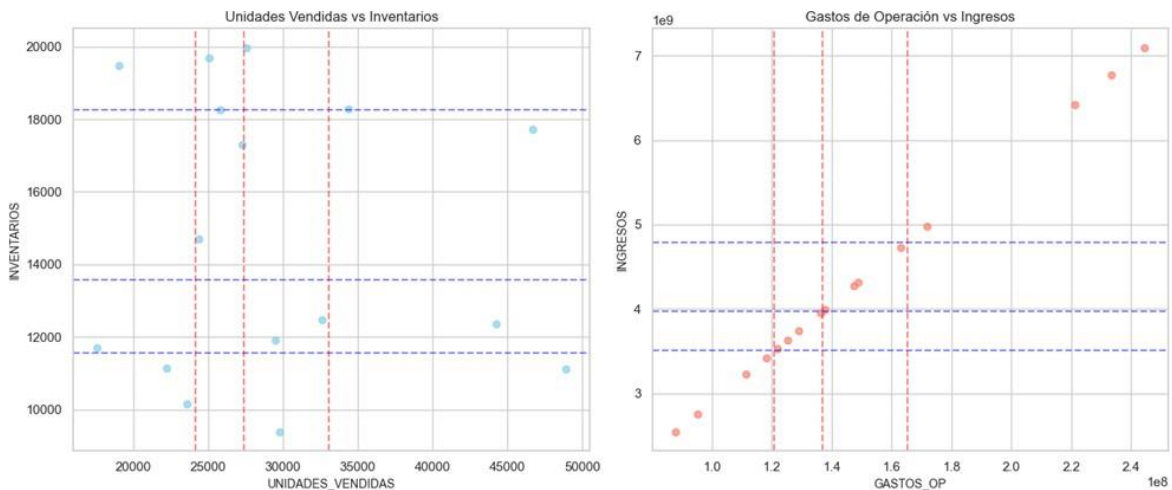


Ilustración 12. Gráfico # 12 (Elaboración propia).

Tipo: Boxplot múltiple

Variables: Todas las variables numéricas.

Análisis:

- Muestra la dispersión general de cada métrica.
- Comparación directa del rango y la centralidad.

Interpretación:

- Útil para entender cuál variable tiene más inestabilidad o dispersión en general.
- El grafico de la izquierda (unidades vendidas vs inventarios) no presentan una tendencia clara lineal. Muchos de los puntos están fuera del rango típico, lo que

indica que no existe una relacion entre la oferta del inventario con la demanda de las ventas.

- El grafico de la derecha (ingresos vs. Gastos op) existe una clara relación positiva: a mayor gasto operativo, mayores ingresos. También, la mayoría de los puntos siguen una tendencia ascendente

7. Conclusiones

- El desempeño comercial no es homogéneo entre zonas ni entre vendedores.
- Algunos vendedores destacan significativamente en ingresos, lo que sugiere buenas prácticas que podrían ser replicadas.
- Se identifican desequilibrios en inventarios que deben evaluarse en función de la demanda real.
- Existen oportunidades de optimización en gastos operativos sin afectar ingresos.
- La mayoría de los ingresos y ventas se concentran en ciertos vendedores o zonas, lo cual sugiere desigualdad en el rendimiento.
- Se evidencia una correlación esperada entre gasto e ingreso, pero podría optimizarse para lograr mayor eficiencia.
- Inventarios no muestran una correlación clara con los ingresos, lo que podría indicar exceso o escasez no alineada con la demanda.

7.1 Conclusiones Detalladas

- **Desigualdad Comercial por Zona y Vendedor**
Existe una concentración significativa del ingreso en ciertos vendedores y zonas (por ejemplo, Zona 2 y Vendedor 1 en varias zonas). La dependencia de pocos actores representa un riesgo estratégico si alguno falla.
- **Inventario con Variabilidad No Justificada por Ingresos**
Algunas zonas y vendedores presentan altos inventarios sin un reflejo proporcional en ventas. Esto podría implicar exceso de stock, falta de rotación, o problemas logísticos.
- **Gasto Operativo e Ingreso Positivamente Correlacionados**

El análisis de dispersión muestra que un aumento en gasto tiende a aumentar ingresos, pero no en todos los casos.
Hay oportunidad de ajustar inversión operativa en zonas donde el retorno es bajo.

- Las variables INGRESOS y GASTOS_OP presentan valores extremos que podrían estar afectando las medidas de tendencia central.
- UNIDADES_VENDIDAS es la variable con una distribución más cercana a la normal.
- INVENTARIOS muestra una posible bimodalidad que sugiere la existencia de subgrupos o una distribución no homogénea en el inventario.
- Las desviaciones estándar permiten visualizar la dispersión y la posible presencia de valores atípicos.

8. Recomendaciones y Análisis Prospectivo

1. Se debe documentar los procesos de venta de los vendedores Top para realizar un modelo de **buenas prácticas** que sirva como formación y mentoría.
2. **Optimizar niveles de inventario** en zonas con exceso mediante modelos de rotación ajustados a la demanda.
3. **En Las zonas críticas como la 4** deben recibir atención con: Campañas de marketing local, incentivos de venta adicionales y reestructuración de rutas o equipos comerciales.
4. Implementar un modelo predictivo de inventarios basado en: **ventas históricas por vendedor/zona y proyección de demanda estacional.**
5. Dado que el gasto e ingreso están correlacionados, auditar zonas con **gastos elevados y bajo retorno** puede revelar: Ineficiencias logísticas y gastos innecesarios.
6. Implementación de **Dashboard** automatizados con KPIs por zona y vendedor, incluyendo unidades vendidas, ingresos e inventarios, entre otros.

7. Es recomendable proyectar escenarios futuros aplicando modelos predictivos (ej. regresión, series temporales) sobre ventas e inventarios.
8. Monitorear los valores por encima de 2.2×10^8 para verificar si hay **ineficiencias o gastos excepcionales**.
9. Hacer un **análisis de variabilidad entre unidades de negocio o departamentos** para ver si esta dispersión es estructural o corregible.
10. Evaluar el comportamiento del gasto relativo a ingresos (margen operativo), para detectar **posibles pérdidas o desviaciones del margen esperado**.

9. Tabla de gráficos e ilustraciones

Tabla 1. Base de datos. Fuente propia	4
Tabla 2. Medidas de resumen base de datos. Fuente propia	4
Tabla 3. Desviación estándar medidas base de datos. Fuente propia.....	5
Tabla 4. Promedio variables base de datos. Fuente propia	6
Tabla 5. Duplicados base de datos. Fuente propia.....	6
Tabla 6. Índice correlación de las variables de la base de datos. Fuente propia	6
Tabla 7. Indicadores clave de las variables de la base de datos. Fuente propia	7
Tabla 8. Clasificación por zonas de las variables de la base de datos. Fuente propia.....	7
Ilustración 1. Gráfico # 1. (Elaboración propia).....	8
Ilustración 2. Gráfico # 2 (Elaboración propia)	10
Ilustración 3. Gráfico # 3 (Elaboración propia)	11
Ilustración 4. Gráfico # 4 (Elaboración propia)	13
Ilustración 5. Gráfico # 5 (Elaboración propia)	14
Ilustración 6. Gráfico# 6 (Elaboración propia).....	15
Ilustración 7. Gráfico # 7 (Elaboración propia)	16
Ilustración 8. Gráfico # 8 (Elaboración propia)	17