

25 febrero del 2026

Geiler Chable

IINFORME DE LAS VENTAS DE LA BASE DE DATOS:

E-Commerce Business Analysis (2022–2024)

Descripción de la Base de Datos

Este conjunto de datos presenta un análisis exhaustivo de la rentabilidad de un negocio de comercio electrónico utilizando datos transaccionales que abarcan desde 2022 hasta 2024. El conjunto de datos contiene más de 3500 pedidos en categorías de productos clave, como electrónica, accesorios y material de oficina, y cuatro regiones geográficas principales: Norte, Este, Sur y Oeste.

Informe

Se trabajó con la base de datos utilizando power bi para la visualización de los datos y sus relaciones entre columnas de la base de datos y con el lenguaje de programación Python para obtener información descriptiva e inferencial de los datos. Se pudieron abstraer las siguientes conclusiones.

Gasto Promedio de Cliente por Orden

Se analizó el gasto promedio por todas las regiones de manera global e individual, los resultados que se obtuvieron fueron los siguientes:

Promedio de Ventas Totales: 3047.966

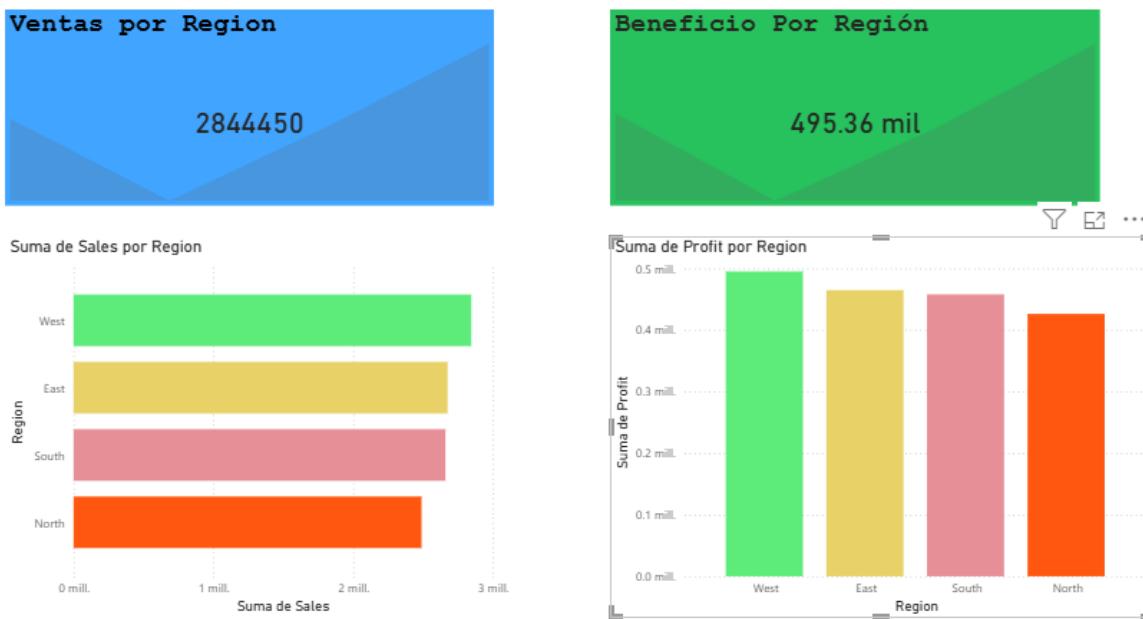
Region	Promedio de Ventas
0 East	3106.980256
1 North	2900.667832
2 South	3011.945640
3 West	3167.538976

En general, se observa que el promedio de las ventas globales está muy cercano a la venta promedio por región. Por lo que, es conveniente analizar si hay diferencias entre los beneficios netos que hay por región. En la siguiente salida de código a través de Python se visualiza dicho beneficio.

...	Region	Sales	Profit	Margin %
0	East	2675110	464888.46	17.507327
1	North	2488773	426314.75	17.259623
2	South	2659548	458103.27	17.228104
3	West	2844450	495358.73	17.483003

De igual manera, visualizamos que los beneficios y el margen de beneficio neto en las 4 regiones son similares.

En general, no vemos discrepancias entre las Regiones, esto lo podemos deducir en el análisis descriptivo en relación a las regiones, sus promedios y márgenes de beneficios neto y por medio de las visualizaciones de Power BI que se ilustran en la siguiente imagen:



Sin embargo, para asegurarnos de un comportamiento similar, se realiza una prueba de hipótesis de homogeneidad, las cuales nos dan los siguientes resultados obtenidos con el lenguaje de programación Python:

La tabla de contingencia para Región y Categoría es:

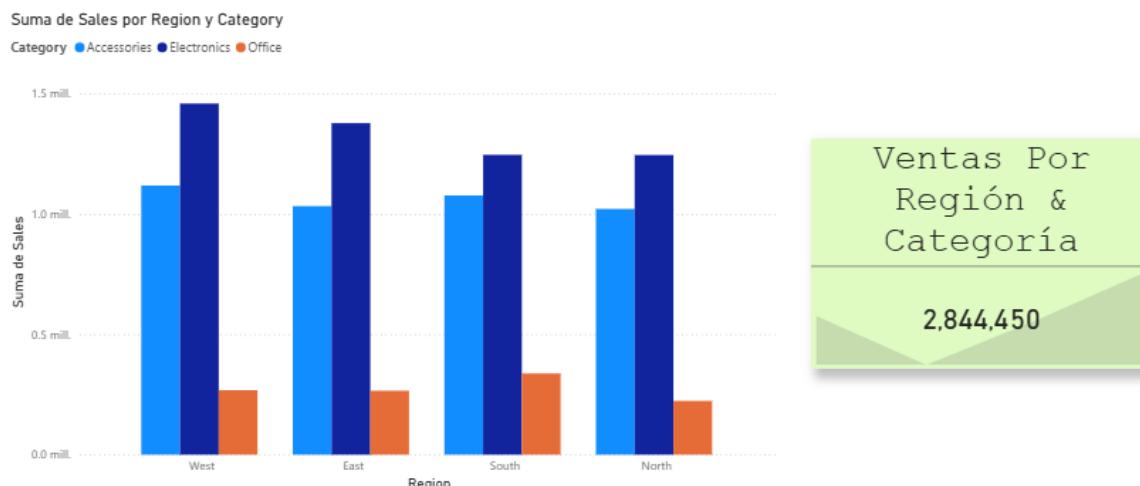
Category	Accessories	Electronics	Office
Region			
East	338	442	81
North	348	429	81
South	370	416	97
West	345	455	98

Estadístico Ji-cuadrada: 5.373357148469919

p-valor: 0.496893377129886

Resultado: El consumo es homogéneo; las regiones se comportan igual.

Por otra parte, se toma importancia analizar las categorías de los artículos que se vendieron y se observó que, en las 4 regiones, las ventas por categoría de artículos, son similares.



Se realizó una prueba ji cuadrada de independencia para verificar el tipo de relación que hay entre las regiones y las categorías de venta. Lo que se obtuvo la siguiente salida con Python.

Estadístico Chi-cuadrado: 5.3734

P-valor: 4.9689e-01

Grados de libertad: 6

Dado a que nuestro p valor es aproximadamente 0.496 y dicho valor es mayor a un nivel de significancia alpha de 0.05, hay evidencia estadística suficiente para no rechazar la hipótesis nula, es decir, las regiones son independientes de las categorías de los productos e-commerce analizados. O en otras palabras:

No podemos afirmar que en el Norte prefieren más la electrónica que en el Sur basándote solo en estos datos, ya que la distribución es estadísticamente similar en todas partes.

Conclusiones

La empresa e-commerce que realizó ventas durante el 2022 al 2024 en las 4 regiones, tuvo un comportamiento similar en cada una de éstas, donde no se obtuvo información relevante para excluir alguna región para futuras ventas, sin embargo, se considera focalizar como área de oportunidad la región Norte, ya que es la zona donde menos ventas y margen de beneficio hubo.