

Análisis exploratorio y descriptivo de las ventas de Amazon + Dashboard (2022–2023)

1. INTRODUCCIÓN	1
2. DESCRIPCIÓN DEL CONJUNTO DE DATOS	2
3. TRANSFORMACIÓN Y LIMPIEZA DE LOS DATOS	2
4. ANÁLISIS DESCRIPTIVO Y EXPLORATORIO DE LOS DATOS	3
5. DASHBOARD	6
6. CONCLUSIONES DEL PROYECTO	7

1. INTRODUCCIÓN

El presente proyecto tiene como objetivo realizar un análisis exploratorio y descriptivo de un conjunto de datos de ventas procedente de la plataforma de comercio electrónico Amazon. A través del análisis de estos datos, se pretende identificar patrones relevantes en el comportamiento de las ventas, así como extraer conclusiones que puedan resultar útiles para la toma de decisiones empresariales.

El proyecto se centra en el estudio de las ventas correspondientes a los años 2022 y 2023, abordando distintas dimensiones del negocio como la evolución temporal de las ventas, la distribución por categorías de producto, regiones, métodos de pago y el impacto de los descuentos. Para ello, se ha llevado a cabo un proceso completo de transformación y limpieza de los datos, seguido de un análisis descriptivo mediante tablas dinámicas y la elaboración de un dashboard interactivo en Excel.

La herramienta principal utilizada para el desarrollo del proyecto ha sido Microsoft Excel, apoyándose en Power Query para la preparación de los datos y en tablas dinámicas para el análisis exploratorio y la visualización de la información.

2. DESCRIPCIÓN DEL CONJUNTO DE DATOS

El conjunto de datos utilizado en este proyecto ha sido obtenido a través de la plataforma Kaggle y corresponde a un registro de ventas de Amazon. Dicho dataset recoge información detallada sobre los pedidos realizados por los clientes, incluyendo variables relacionadas con la fecha del pedido, el producto, la categoría, la región, el método de pago, el descuento aplicado, la cantidad vendida y los valores económicos asociados a cada transacción.

El dataset está compuesto por más de 50.000 registros y un total de 13 variables, lo que proporciona un volumen de información suficiente para llevar a cabo un análisis exploratorio completo. La estructura del conjunto de datos es clara y presenta variables adecuadas para el estudio del comportamiento de las ventas desde distintas perspectivas, tanto temporales como categóricas.

No obstante, antes de proceder al análisis, fue necesario realizar un proceso de transformación y limpieza de los datos con el fin de garantizar su calidad, coherencia y correcta interpretación.

3. TRANSFORMACIÓN Y LIMPIEZA DE LOS DATOS

Una vez seleccionado el conjunto de datos, se procedió a llevar a cabo un proceso de transformación y limpieza con el objetivo de garantizar la calidad, coherencia y fiabilidad de la información antes de realizar el análisis descriptivo. Esta fase resulta fundamental en cualquier proyecto de análisis de datos, ya que permite evitar errores de interpretación y asegurar que los resultados obtenidos sean consistentes.

En primer lugar, el archivo original en formato CSV fue importado a Microsoft Excel utilizando la herramienta Power Query, lo que permitió gestionar de forma controlada los distintos procesos de limpieza y transformación. Durante esta etapa inicial, se detectaron discrepancias en el formato numérico de algunas variables cuantitativas, debido a que el archivo original empleaba el punto como separador decimal, mientras que la configuración regional de Excel utiliza la coma decimal. Para corregir este problema, se realizó la conversión del tipo de dato de las variables numéricas mediante la opción de “Tipo de datos con configuración regional”, seleccionando el formato correspondiente a Inglés (Estados Unidos). De este modo, se aseguró una correcta interpretación de los valores decimales en variables económicas como el precio, el precio con descuento y los valores monetarios asociados a cada pedido.

Posteriormente, se revisó el significado de las variables económicas incluidas en el conjunto de datos. Durante este proceso, se identificó que la variable denominada originalmente como “beneficio” no representaba el beneficio neto de las ventas, sino los ingresos totales generados por cada pedido, calculados a partir del precio con descuento multiplicado por la cantidad vendida, sin considerar los costes asociados. Con el fin de evitar interpretaciones

erróneas, esta variable fue tratada a lo largo del proyecto como ingresos totales, ajustando su interpretación en todos los análisis posteriores.

A continuación, se procedió a la identificación y eliminación de posibles registros duplicados. Para ello, se utilizó la variable correspondiente al identificador único de cada pedido (Order ID), dado que cada pedido debe aparecer una única vez en el conjunto de datos. Se eliminaron los registros repetidos, conservando una sola observación por pedido. Este paso contribuyó a mejorar la calidad del conjunto de datos y a garantizar la fiabilidad de los resultados obtenidos.

Seguidamente, se revisó el formato de la variable correspondiente a la fecha del pedido. Se comprobó que dicha variable fuese reconocida correctamente como un valor de tipo fecha por Excel y, en los casos necesarios, se realizó su conversión utilizando Power Query y la configuración regional adecuada. Una vez validado el formato, los registros fueron ordenados cronológicamente según la fecha del pedido, facilitando así el análisis temporal de las ventas durante los años 2022 y 2023.

Finalmente, como parte del proceso de preparación de los datos, se procedió a la creación de nuevas variables derivadas con el objetivo de enriquecer el análisis. A partir de la fecha del pedido, se generaron las variables año, trimestre y mes, lo que permitió realizar análisis temporales a distintos niveles de agregación. Asimismo, se creó una variable categórica que distingue entre pedidos con descuento y pedidos sin descuento, en función de si el descuento aplicado era superior a cero. La incorporación de estas variables facilitó la segmentación de la información y aportó mayor profundidad al análisis descriptivo y a la posterior elaboración del dashboard.

4. ANÁLISIS DESCRIPTIVO Y EXPLORATORIO DE LOS DATOS

Una vez finalizado el proceso de transformación y limpieza de los datos, se procedió a realizar un análisis descriptivo y exploratorio con el objetivo de comprender el comportamiento de las ventas de Amazon durante los años 2022 y 2023. Para ello, se emplearon tablas dinámicas y segmentadores en Microsoft Excel, lo que permitió analizar la información de forma flexible y dinámica según distintos criterios seleccionados por el usuario.

El análisis se centró en el estudio de los ingresos totales, la cantidad de productos vendidos, el número de pedidos y el ticket medio, así como en su distribución temporal, por categoría de producto, región, método de pago y según la aplicación de descuentos.

4.1 Indicadores generales de ventas

En primer lugar, se calcularon una serie de indicadores generales con el fin de obtener una visión global del comportamiento del negocio durante el periodo analizado. El volumen total de ingresos asciende a 32.866.574 €, correspondiente a un total de 50.000 pedidos y 149.970 productos vendidos. A partir de estos valores, se obtiene un ticket medio de 657 € por pedido, lo que refleja el importe medio gastado por los clientes en cada transacción.

Estos indicadores permiten contextualizar el resto del análisis y sirven como referencia para interpretar los resultados obtenidos en los distintos apartados posteriores.

4.2 Análisis temporal de las ventas

A continuación, se analizó la evolución temporal de los ingresos a lo largo del año, utilizando la variable mes del pedido. Los resultados muestran que las ventas se distribuyen de manera relativamente homogénea a lo largo de los distintos meses, aunque se observan ligeras variaciones entre ellos. Meses como enero, marzo, mayo y agosto presentan niveles de ingresos algo superiores, mientras que otros como febrero o noviembre registran valores ligeramente inferiores.

Este análisis permite identificar posibles patrones estacionales en el comportamiento de las ventas, aunque no se aprecia una concentración excesiva de los ingresos en un periodo concreto del año, lo que sugiere una demanda relativamente estable a lo largo del tiempo. El uso de segmentadores permite, además, analizar esta evolución de forma independiente para cada año o trimestre.

4.3 Ventas por categoría de producto

El análisis por categoría de producto se realizó a partir de la cantidad total de unidades vendidas. Los resultados indican que las distintas categorías presentan volúmenes de ventas similares, sin grandes diferencias entre ellas. Categorías como Books, Beauty, Electronics, Fashion, Home & Kitchen y Sports muestran valores de unidades vendidas muy próximos, lo que sugiere una demanda equilibrada entre los distintos tipos de productos ofrecidos.

Este comportamiento indica que no existe una única categoría dominante en términos de volumen de ventas, sino que el negocio presenta una diversificación adecuada de sus productos, reduciendo la dependencia de una sola línea de producto.

4.4 Impacto del descuento en las ventas

Uno de los aspectos analizados fue el impacto de la aplicación de descuentos sobre los ingresos obtenidos. Los resultados muestran que una parte muy significativa de los ingresos procede de pedidos con descuento, que alcanzan aproximadamente 26,7 millones de euros, frente a 6,15 millones de euros generados por pedidos sin descuento.

Este resultado pone de manifiesto la importancia de las estrategias de descuento en el volumen total de ingresos, sugiriendo que los descuentos actúan como un factor clave para

incentivar la compra. No obstante, este análisis se limita a los ingresos totales, ya que no se dispone de información sobre los costes asociados, por lo que no es posible evaluar el impacto real sobre el beneficio neto.

4.5 Ventas por región

El análisis geográfico de las ventas se realizó agrupando los ingresos por región. Los resultados muestran una distribución bastante equilibrada entre las distintas regiones analizadas, con valores de ingresos similares en Asia, Europa, Oriente Medio y Norteamérica. Ninguna región destaca de forma significativa sobre las demás, lo que indica una presencia internacional bien distribuida y un comportamiento homogéneo del mercado en las distintas zonas geográficas.

Este resultado sugiere que la empresa no depende en exceso de una única región para generar ingresos, lo que puede considerarse un aspecto positivo desde el punto de vista de la diversificación del riesgo.

4.6 Ingresos por método de pago

Finalmente, se analizaron los ingresos en función del método de pago utilizado por los clientes. Los resultados muestran que los distintos métodos de pago generan volúmenes de ingresos similares, destacando ligeramente opciones como Cash on Delivery, Credit Card y Debit Card. Métodos digitales como UPI y Wallet también presentan un peso relevante en el total de ingresos. Esta distribución indica que los clientes utilizan una amplia variedad de métodos de pago, lo que refuerza la importancia de ofrecer múltiples opciones para facilitar el proceso de compra y mejorar la experiencia del usuario.

4.7 Uso de segmentadores y análisis interactivo

Todas las tablas y métricas analizadas se encuentran conectadas a segmentadores de año, trimestre y mes, lo que permite al usuario filtrar la información de forma dinámica y analizar los resultados desde distintas perspectivas temporales. Esta funcionalidad facilita la exploración de los datos y refuerza el carácter interactivo del análisis, permitiendo adaptar la visualización de los resultados a las necesidades del usuario final.

5. DASHBOARD

Con el objetivo de sintetizar y visualizar de forma clara los principales resultados obtenidos en el análisis descriptivo, se ha diseñado un dashboard interactivo en Microsoft Excel. Este panel permite explorar los datos de manera dinámica y facilita la interpretación de los patrones de comportamiento de las ventas.

El dashboard se ha construido a partir de tablas dinámicas, garantizando que todas las visualizaciones y métricas se actualicen automáticamente al aplicar filtros mediante segmentadores.

5.1 Estructura general del dashboard

El dashboard se organiza en tres niveles principales: indicadores clave en la parte superior, análisis temporal de los ingresos en la zona central y análisis por dimensiones de negocio en la parte inferior. Esta estructura permite una lectura jerárquica, comenzando por una visión global y descendiendo hacia análisis más específicos.

5.2 Indicadores clave

En la parte superior del dashboard se presentan los principales indicadores agregados del conjunto de datos: ingresos totales, productos vendidos, número de pedidos y ticket medio. Todos los indicadores están vinculados a tablas dinámicas y responden a los filtros aplicados, lo que permite analizar su evolución por año, trimestre o mes.

5.3 Evolución temporal de los ingresos

El núcleo del dashboard lo constituye un gráfico de líneas que muestra la evolución mensual de los ingresos. Esta visualización permite identificar tendencias generales a lo largo del tiempo, posibles patrones de estacionalidad y comparaciones entre distintos periodos temporales. El gráfico está conectado a los segmentadores de año, trimestre y mes, facilitando el análisis comparativo entre ejercicios.

5.4 Análisis por categoría de producto

Para analizar la distribución del volumen de ventas, se incluye un gráfico de barras que muestra las unidades vendidas por categoría de producto. Este gráfico permite evaluar el peso relativo de cada categoría dentro del total de ventas y detectar posibles diferencias en la demanda entre tipos de productos.

5.5 Análisis geográfico

El dashboard incorpora un gráfico de columnas que representa los ingresos por región, lo que permite comparar el rendimiento comercial entre distintas áreas geográficas y obtener una visión territorial del negocio.

5.6 Impacto de la aplicación de descuentos

Con el fin de analizar el uso de estrategias promocionales, se incluye un gráfico circular que muestra el porcentaje de ventas realizadas con y sin descuento. Esta visualización permite observar el peso de los descuentos en el volumen total de ventas.

5.7 Ingresos según método de pago

Finalmente, se presenta un gráfico de barras horizontales que recoge los ingresos según el método de pago utilizado. Este análisis proporciona información sobre las preferencias de los clientes y el peso relativo de cada método dentro del total de ingresos.

5.8 Interactividad del dashboard

El dashboard incorpora segmentadores que permiten filtrar la información por año, trimestre y mes. Estos segmentadores afectan de forma simultánea a todos los indicadores y gráficos, garantizando coherencia en el análisis y permitiendo explorar los datos desde diferentes perspectivas temporales.

6. CONCLUSIONES DEL PROYECTO

Conclusiones generales (visión global)

El análisis agregado del periodo 2022–2023 muestra un volumen de negocio elevado y estable, con unos ingresos totales de 32,87 millones de euros, derivados de cerca de 50.000 pedidos y casi 150.000 productos vendidos. El ticket medio se sitúa en torno a los 657 €, lo que indica una estrategia comercial orientada a pedidos de valor medio-alto, más que a un gran volumen de transacciones de bajo importe.

La evolución temporal de las ventas evidencia un comportamiento estacional moderado, con picos recurrentes en los meses centrales del año, especialmente en verano, y descensos puntuales a comienzos de año y en noviembre, lo que sugiere una demanda relativamente predecible y bien distribuida a lo largo del ejercicio.

En cuanto a la distribución por categorías, no se observan desequilibrios extremos, ya que todas las categorías presentan volúmenes de unidades vendidos similares. Este resultado refleja una cartera de productos diversificada y una dependencia limitada de una única línea de negocio.

Desde el punto de vista geográfico, Middle East y North America se consolidan como las regiones con mayor aportación a las ventas, mientras que Europa presenta un peso relativo menor, lo que podría abrir oportunidades de crecimiento futuro en este mercado.

Por último, el análisis del impacto de los descuentos muestra que aproximadamente el 81% de las ventas se realizan con descuento, lo que indica una clara dependencia de la política promocional para estimular la demanda. Este aspecto es especialmente relevante desde el punto de vista estratégico, ya que puede tener implicaciones directas sobre los márgenes.

Proyecto formativo de análisis de datos

Excel & Power Query — The Power MBA

Análisis y conclusiones del año 2022

Durante 2022, la empresa generó 16,39 millones de euros en ingresos, con 24.926 pedidos y 74.738 productos vendidos, manteniendo un ticket medio de 658 €, ligeramente superior al promedio global.

La evolución mensual refleja una mayor volatilidad, con un inicio de año fuerte, una caída significativa en febrero y una recuperación progresiva hasta alcanzar el máximo anual en los meses de verano. Este patrón apunta a una sensibilidad estacional más acusada durante este ejercicio.

A nivel regional, destacan North America y Middle East como principales motores del crecimiento, mientras que Europa se mantiene como la región con menor contribución. En cuanto a los métodos de pago, los ingresos están muy equilibrados entre las distintas alternativas, lo que reduce el riesgo operativo asociado a la dependencia de un único medio de pago.

El uso de descuentos ya es elevado en 2022, lo que sugiere que la estrategia promocional no es coyuntural, sino estructural dentro del modelo de negocio.

Análisis y conclusiones del año 2023

En 2023 se observa una ligera mejora respecto a 2022, alcanzando 16,48 millones de euros en ingresos, con 25.074 pedidos y 75.232 productos vendidos, mientras que el ticket medio se mantiene estable en 657 €. Esto indica que el crecimiento proviene principalmente del aumento del volumen de pedidos y no de un incremento del valor medio de cada compra.

La evolución de las ventas es más homogénea y estable que en 2022, reduciéndose la volatilidad mensual. Este comportamiento sugiere una mayor madurez operativa y comercial, así como una posible mejora en la planificación de campañas y en la gestión de la demanda.

Por regiones, Middle East refuerza su posición como principal mercado, mientras que North America mantiene un nivel elevado de ventas. Europa experimenta una leve mejora, aunque continúa siendo la región con menor peso relativo.

El patrón de ventas con descuento se mantiene prácticamente inalterado respecto al año anterior, confirmando que la empresa sostiene su volumen de negocio en una estrategia intensiva de promociones. Este resultado refuerza la necesidad de analizar su impacto sobre la rentabilidad a largo plazo.

Conclusión final

En conjunto, el proyecto pone de manifiesto que la empresa presenta un modelo de negocio sólido, estable y diversificado, con un crecimiento moderado y controlado entre 2022 y 2023. No obstante, se identifican dos elementos clave para la toma de decisiones futuras: la alta dependencia de los descuentos y el potencial de crecimiento en regiones menos explotadas, especialmente Europa.

Estas conclusiones ofrecen una base cuantitativa clara para el diseño de estrategias orientadas a mejorar los márgenes, optimizar las políticas promocionales y equilibrar el crecimiento geográfico.