



ESCOLA TÈCNICA SUPERIOR
D'ENGINYERIA
Universitat Rovira i Virgili



Práctica 1

Alumnos: Alfonso Sanchez, Antoni Llop
Profesor: Edgar Batista
Curso: 2025-2026

Introducción.....	4
Herramientas y metodología.....	4
Fase 1.....	5
 Gráfica 1: Análisis: Días reales de envío según estado de entrega.....	5
Pregunta planteada.....	5
Procedimiento.....	5
Resultados (gráfico).....	6
Interpretación.....	6
Análisis estadístico.....	6
 Gráfica 2: Análisis: Relación entre precio del producto y total del pedido.....	7
Pregunta planteada.....	7
Procedimiento.....	7
Resultados (gráfico).....	8
Interpretación.....	8
Análisis estadístico.....	9
 Gráfica 3: Análisis: Tiempo de envío por modo de transporte.....	9
Pregunta planteada.....	9
Procedimiento.....	9
Resultados (gráfico).....	10
Interpretación.....	10
Análisis estadístico.....	11
 Gráfica 4: Número de Clientes por Segmento y País.....	11
Pregunta planteada.....	11
Procedimiento.....	11
Resultados (gráfico).....	12
Interpretación.....	12
Análisis estadístico.....	13
 Gráfica 5: Clientes con más compras realizadas.....	14
Pregunta planteada.....	14
Procedimiento.....	14
Resultados (gráfico).....	14
Interpretación.....	15
 Gráfica 6: Evolución de ventas a lo largo del tiempo.....	15
Pregunta planteada.....	15
Procedimiento.....	15
Resultados (gráfico).....	15
Interpretación.....	16
Análisis estadístico.....	16
 Gráfica 7 - Categoría más vendida.....	17
Pregunta planteada.....	17
Procedimiento.....	17

Resultados (gráfico).....	17
Interpretación.....	17
Análisis Estadístico.....	18
Gráfica 8 - Beneficio para cada tipo de cliente.....	19
Pregunta planteada.....	19
Procedimiento.....	19
Resultados (gráfico).....	20
Interpretación.....	20
Análisis Estadístico.....	21
Gráfica 9 - Ratio de beneficio.....	21
Pregunta planteada.....	21
Procedimiento.....	21
Resultados (gráfico).....	22
Interpretación.....	22
Análisis Estadístico.....	22
Gráfica 10 - Mayor porcentaje de retraso.....	23
Pregunta planteada.....	23
Procedimiento.....	23
Resultados (gráfico).....	24
Interpretación.....	24
Análisis Estadístico.....	24
Estudio - Bajada de ventas en 2018.....	25
Motivación.....	25
Enfoque.....	25
Datos encontrados.....	26
Conclusión del estudio.....	28
Estudio - Relación entre descuentos, beneficios y márgenes en las ventas.....	28
Motivación.....	28
Enfoque.....	28
Gráficas realizadas.....	29
Gráfica 1.....	29
Gráfica 2.....	29
Gráfica 3.....	30
Gráfica 4.....	31
Conclusión del estudio.....	32
Estudio - Relación entre modo de envío, tiempos de entrega y rentabilidad.....	32
Motivación.....	32
Enfoque.....	33
Gráficas realizadas.....	33
Gráfica 1.....	33
Gráfica 2.....	34
Gráfica 3.....	34
Gráfica 4.....	35
Conclusión del estudio.....	36

Fase 2.....	37
Introducción.....	37
Gráfica 1 - Beneficio por país.....	37
Gráfica 2 - Retrasos.....	38
Gráfica 3 - Ciudades con vendedores destacables.....	39
Gráfica 4 - Mapa de clústeres Categórico métodos de envío por destino.....	40
Gráfica 5 - Mapa de flujo de rutas comerciales principales.....	42

Introducción

En esta práctica trabajamos con datos reales de una cadena de suministro para aprender a analizarlos y sacar conclusiones útiles. La idea es entender cómo usar las herramientas de análisis exploratorio de datos para descubrir patrones, relaciones y posibles problemas dentro del conjunto de datos.

El archivo que usamos contiene información de pedidos realizados entre 2015 y 2018 por empresas de todo el mundo.

Herramientas y metodología

Para realizar todos los análisis de la práctica se ha utilizado el lenguaje de programación Python, por ser una herramienta muy completa, flexible y ampliamente utilizada en el ámbito del análisis de datos. Python permite combinar fácilmente el tratamiento estadístico con la visualización, lo que facilita interpretar los resultados de forma clara y visual.

Las principales librerías empleadas han sido:

- **pandas**, para la carga, limpieza y manipulación de datos, ya que nos permite acceder al csv de nuestra práctica de manera bastante sencilla sin tener que parsearlo.
- **matplotlib**, utilizada para la representación gráfica de los resultados mediante diferentes tipos de gráficos (barras, dispersión, boxplot, líneas, etc.), escogiendo en cada caso la opción más adecuada según el tipo de variables analizadas.
- **scipy.stats**, empleada para realizar los análisis estadísticos necesarios, como t-student, ANOVA, chi-cuadrado etc... según la naturaleza de los datos.

La metodología seguida en todos los estudios fue similar:

1. **Selección de variables relevantes** según el objetivo del análisis.
2. **Limpieza de los datos**, eliminando valores nulos o erróneos para garantizar la fiabilidad de los resultados.
3. **Aplicación del método estadístico o gráfico** más adecuado a cada tipo de variables.
4. **Interpretación de los resultados**, tanto de los valores numéricos como de las visualizaciones, para extraer conclusiones claras y fundamentadas.

Fase 1

Gráfica 1: Análisis: Días reales de envío según estado de entrega

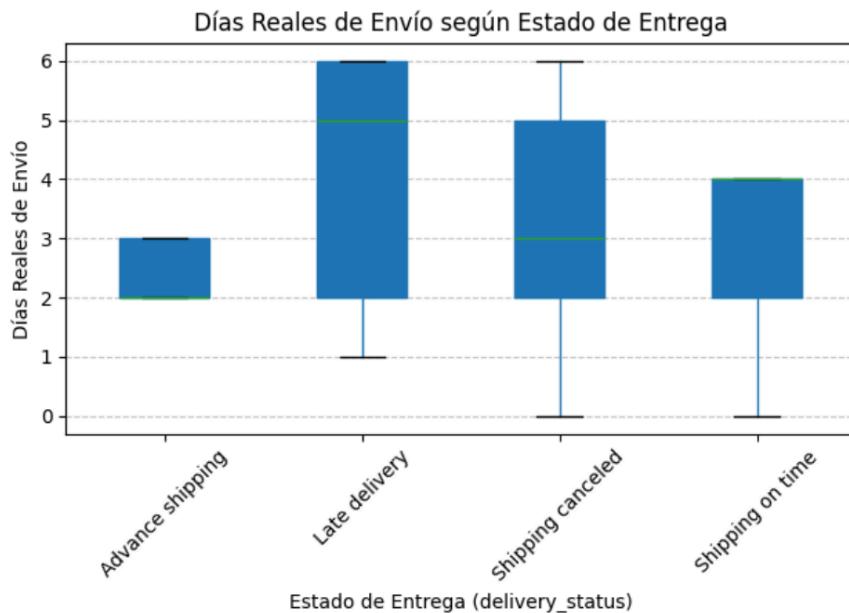
Pregunta planteada

¿Existen diferencias en los días reales de envío dependiendo del estado de entrega de la orden?

Procedimiento

1. Se seleccionaron las variables `days_for_shipping_real` (días reales de envío) y `delivery_status` (estado de la entrega).
2. Para representar los resultados se utilizó un boxplot. Lo hemos usado porque:
 - Permite comparar la distribución de una variable numérica (días reales de envío) entre diferentes categorías (estados de entrega).
 - Muestra de manera clara la mediana, la dispersión (rango intercuartílico) y la presencia de valores extremos (outliers).
 - Es más informativo que una media o un histograma, ya que destaca las diferencias en la variabilidad y no solo en los promedios.

Resultados (gráfico)



Advance shipping: valores concentrados entre 2 y 3 días, con poca dispersión.

Late delivery: distribución más amplia (1–6 días), con una mediana elevada (≈ 5 días).

Shipping canceled: gran variabilidad (0–6 días), mediana en torno a 3.

Shipping on time: valores entre 0 y 4 días, mediana cercana a 4.

Interpretación

1. **Entregas adelantadas (advance shipping):** los pedidos se envían en muy pocos días y con poca variación. Esto refleja eficiencia y rapidez en la logística.
2. **Entregas a tiempo (shipping on time):** suelen tardar entre 3 y 4 días, con una dispersión moderada. Indican un proceso estable y controlado.
3. **Entregas tardías (late delivery):** concentran los mayores tiempos de envío y la mayor variabilidad. Son el grupo más problemático porque representan retrasos claros en la cadena.
4. **Envíos cancelados (shipping canceled):** muestran un comportamiento irregular, con casos rápidos y otros muy lentos. Esto sugiere que las cancelaciones pueden darse en diferentes fases del proceso.

Este análisis nos indica que el estado de entrega influye directamente en los días reales de envío, lo que lo convierte en un indicador clave para identificar ineficiencias y proponer mejoras logísticas.

Análisis estadístico

Para hacer un análisis estadístico de nuestra boxplot lo que hemos hecho es hacer una ANOVA, ya que permite comparar la igualdad de medias entre más de dos grupos. Para hacerlo lo hemos hecho con la librería de python scipy concretamente importando stats. Hemos agrupado todos los grupos y hemos realizado el estudio con la función f_oneway.

El ANOVA nos devuelve dos valores: la estadística F, que mide la variación entre grupos frente a la variación dentro de cada grupo, y el p-valor, que indica la probabilidad de observar esas diferencias si en realidad las medias fueran iguales.

En nuestro caso, obtuvimos $F = 13079.9$ y $p = 0.0$, lo cual muestra que existen diferencias muy significativas entre los estados de entrega. Esto significa que al menos dos estados difieren claramente en el número de días de envío.

Resultados del ANOVA:
La F es el valor estadístico de la prueba y la p es el valor
Si la F es alta y la p es menor que 0,05, significa que hay diferencias significativas entre los grupos.
 $F = 13079.903$, $p = 0.0000$
Hay diferencias significativas entre al menos dos estados.

Significa que hay diferencias significativas entre al menos dos estados, pero para saber entre qué estados hay tanta diferencia tendríamos que hacer una T-student entre todos los grupos. Por ejemplo si son 4 grupos, el primero tendría que hacerlo con el segundo,tercero,cuarto, luego el segundo con el primero, tercero y cuarto y así sucesivamente...

En este caso hemos decidido hacer un estudio más simple sólo con un ANOVA sin entrar tanto en el trapo.

Gráfica 2: Análisis: Relación entre precio del producto y total del pedido

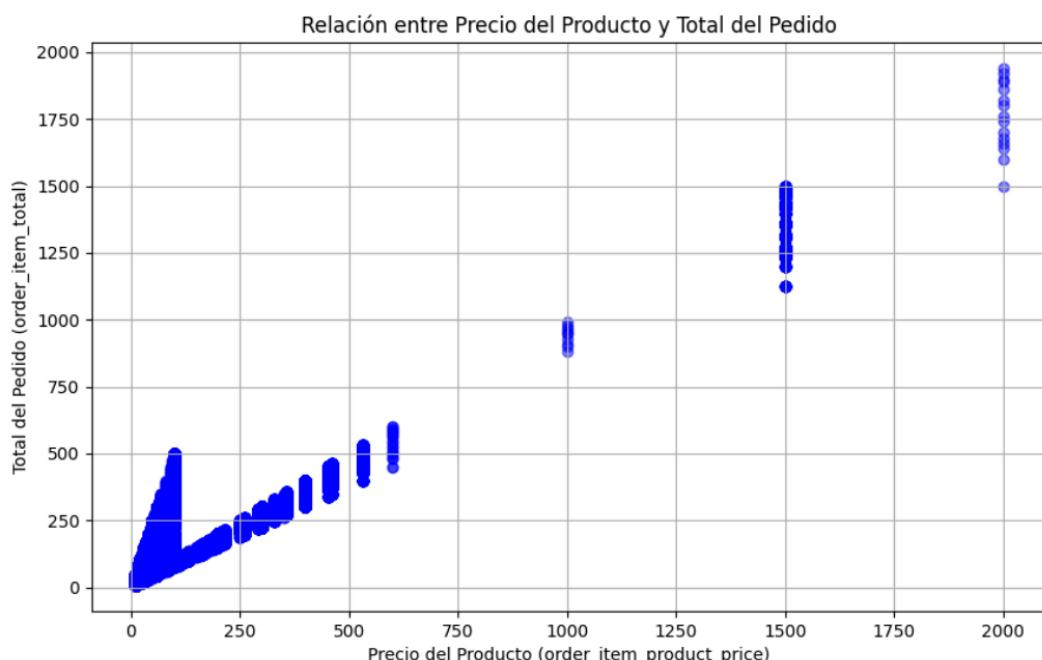
Pregunta planteada

¿Existe relación entre el precio unitario de los productos y el valor total del pedido?

Procedimiento

1. Se seleccionaron las variables `order_item_product_price` (precio unitario del producto) y `order_item_total` (importe total del pedido).
2. Se representaron los datos mediante un scatter plot, porque es mejor para:
 - Analizar la relación entre dos variables numéricas.
 - Detectar patrones de correlación positiva, negativa o inexistente.
 - Identificar posibles grupos de valores atípicos o concentraciones de puntos.

Resultados (gráfico)



A medida que aumenta el precio unitario del producto, el total del pedido también tiende a crecer.

Se observan varios bloques verticales de puntos, lo que indica que ciertos precios unitarios se repiten con frecuencia (ej.: productos de 1000, 1500 o 2000).

En los precios bajos (menores a 200), los totales de pedido varían mucho, pero rara vez superan los 500.

En los precios más altos (≥ 1000), los totales de pedido se concentran en valores elevados, normalmente por encima de 1000.

Interpretación

Existe una correlación positiva clara: los productos más caros generan pedidos de mayor valor total.

La presencia de franjas verticales refleja que algunos productos estándar se venden con mucha frecuencia a precios fijos.

En productos baratos, el valor total depende principalmente de la cantidad comprada, mientras que en productos caros el precio unitario domina el valor del pedido.

Este análisis nos ayuda a entender cómo diferentes segmentos de precios impactan en la facturación total.

Análisis estadístico

Para hacer un análisis estadístico de nuestro scatter plot lo que hemos hecho es calcular la **correlación de Pearson**, ya que es la medida que permite cuantificar la fuerza y dirección de la relación lineal entre dos variables numéricas.

El coeficiente de Pearson nos devuelve dos valores: el coeficiente R, que mide la intensidad de la relación lineal (cercano a +1 indica una relación positiva fuerte y cercano a -1 una relación negativa fuerte)

En nuestro caso:

Si la correlación es próxima a 1 o -1 indica una relación lineal fuerte entre las variables.
Correlación de Pearson (R) = 0,782
Hay una relación lineal fuerte entre el precio del producto y el total del pedido.

En nuestro caso, obtuvimos R = 0,782, lo cual muestra una correlación positiva fuerte entre el precio unitario y el total del pedido. Esto significa que, en general, cuando aumenta el precio de un producto, también tiende a aumentar el importe total del pedido.

Gráfica 3: Análisis: Tiempo de envío por modo de transporte

Pregunta planteada

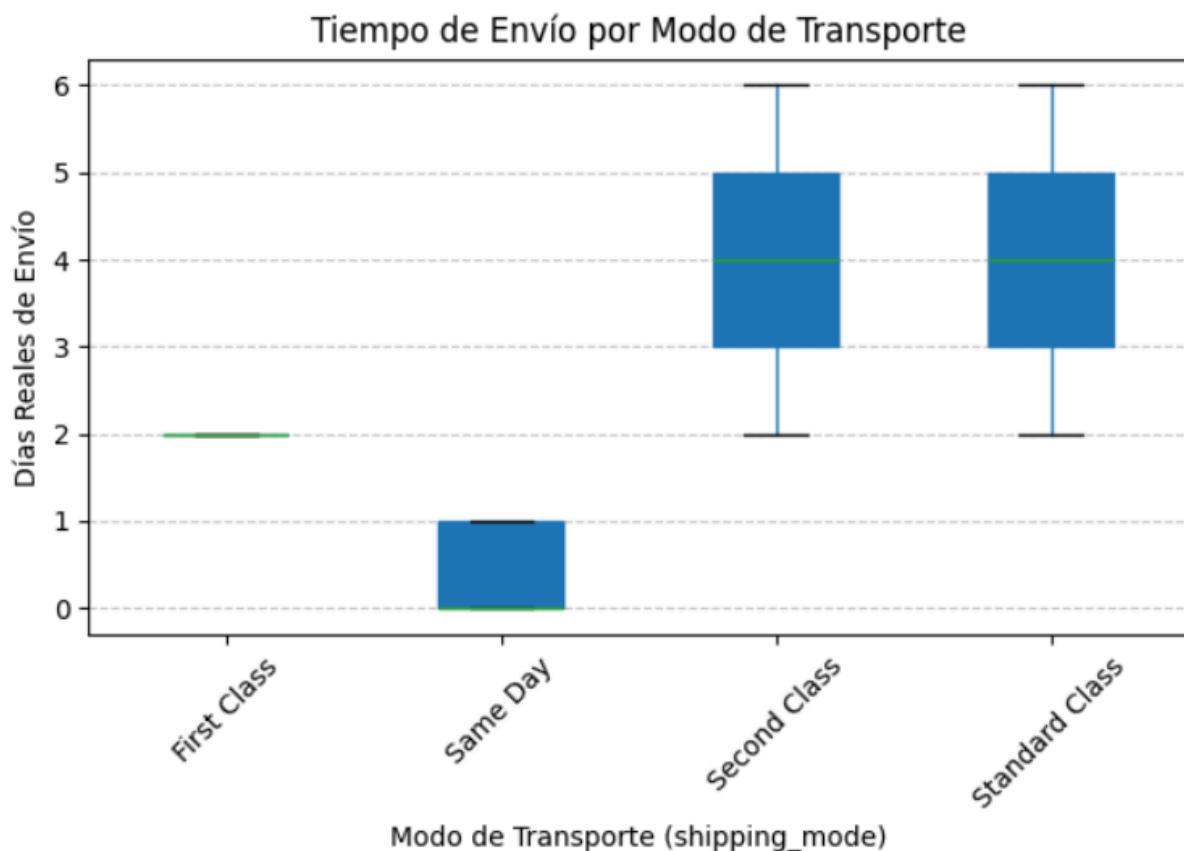
¿Los envíos en First Class son realmente más rápidos que los realizados en Standard Class?

Procedimiento

1. Se seleccionaron las variables `days_for_shipping_real` (días reales de envío) y `shipping_mode` (modo de transporte).
2. Se representó la distribución de los días de envío mediante un boxplot, ya que este tipo de gráfico permite:

- Comparar de forma clara la distribución de tiempos entre diferentes categorías de transporte.
- Visualizar la mediana, la dispersión y la posible existencia de outliers.
- Analizar si los modos de transporte cumplen con lo esperado en términos de rapidez.

Resultados (gráfico)



Same Day: muestra los valores más bajos (0–1 día), confirmando que es el método más rápido.

First Class: tiempos reducidos y estables, con una mediana de 2 días.

Second Class: mayor variabilidad, con una mediana de 4 días.

Standard Class: tiempos similares a Second Class, con una mediana también en torno a 4 días y dispersión hasta 6 días.

Interpretación

Los resultados confirman que First Class es más rápido que Standard Class, con menos días de envío y menor variabilidad.

Same Day es claramente el más rápido, aunque se trata de un servicio específico y menos frecuente.

Second Class y Standard Class presentan tiempos de entrega similares, ambos más largos y menos eficientes que First Class.

Esto demuestra que el modo de transporte es un factor clave en la duración de los envíos, y válida la diferenciación entre los diferentes servicios.

Análisis estadístico

Para hacer un análisis estadístico de nuestro boxplot hemos aplicado el **test t de Student**, ya que permite comparar si existen diferencias significativas entre las medias de varios grupos independientes. En este caso, los grupos corresponden a los distintos modos de transporte utilizados en los envíos.

Como ya habíamos realizado un ANOVA (en otro caso) que nos decía que había diferencias significativas entre los diferentes estados, hemos decidido hacer un t-student entre todos los posibles estados para así saber entre qué estados había una diferencia significativa.

El test t de Student devuelve dos valores principales: la estadística t, que mide la magnitud y dirección de la diferencia entre medias, y el p-valor, que nos indica si la diferencia es estadísticamente significativa ($p < 0.05$).

Resultados:							
	Grupo A	Grupo B	n_A	n_B	t	p_valor	significativo_0.05
0	Standard Class	First Class	107752	27814	462.276428	0.00000	True
1	Standard Class	Same Day	107752	9737	528.678185	0.00000	True
2	First Class	Second Class	27814	35216	-263.858623	0.00000	True
3	First Class	Same Day	27814	9737	300.583812	0.00000	True
4	Second Class	Same Day	35216	9737	386.584717	0.00000	True
5	Standard Class	Second Class	107752	35216	0.584287	0.55903	False

Interpretación rápida

Hay 5 pares con $p < 0.05$. Ejemplos:

	Grupo A	Grupo B
0	Standard Class	First Class
1	Standard Class	Same Day
2	First Class	Second Class
3	First Class	Same Day
4	Second Class	Same Day

En nuestro caso, los resultados mostraron diferencias significativas en casi todos los pares de grupos, excepto entre Standard Class y Second Class, cuyos tiempos de envío son similares ($p = 0.56$). Esto confirma que el modo de transporte influye directamente en el tiempo real de envío, siendo algunos métodos (como Same Day o First Class) considerablemente más rápidos que otros.

Gráfica 4: Número de Clientes por Segmento y País

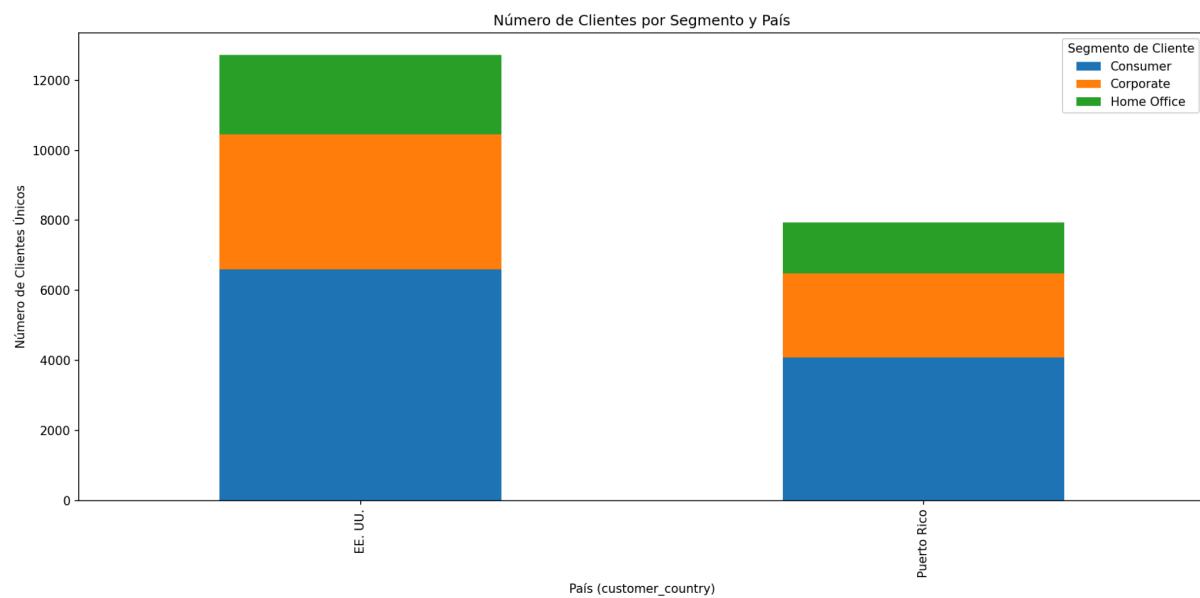
Pregunta planteada

¿Cómo se distribuyen los clientes por segmento en cada país, y en qué países se concentra cada tipo de cliente?

Procedimiento

1. Se seleccionaron las variables, `customer_country`: país del cliente y `customer_segment`: tipo de cliente (Consumer, Corporate, Home Office).
2. Se realizó un conteo de clientes únicos por país y segmento utilizando `customer_id`.
3. Para representar los resultados, se utilizó un gráfico de barras apiladas, porque permite:
 - Comparar el número total de clientes por país (altura de la barra).
 - Visualizar la composición interna por segmento (partes apiladas con distintos colores).
 - Detectar qué segmentos son predominantes en cada país.

Resultados (gráfico)



En **EE. UU.** hay el mayor número de clientes, con predominio del tipo Consumer, seguido por Corporate y Home Office.

En **Puerto Rico**, aunque hay menos clientes en total, la distribución relativa por segmentos es parecida: Consumer es el más numeroso, pero con mayor peso proporcional de Corporate respecto a EE. UU.

No aparecen otros países en la variable `customer_country`, lo que indica que la base de clientes únicos está principalmente localizada en estas dos regiones.

Interpretación

La mayoría de los clientes pertenecen al tipo Consumer, tanto en EE. UU. como en Puerto Rico, lo que sugiere que gran parte del negocio se centra en ventas al consumidor final.

El tipo Corporate tiene un peso considerable, especialmente en Puerto Rico, donde representa una parte más equilibrada frente a Consumer.

El tipo Home Office es el menos representado, aunque también tiene presencia estable en ambos países.

Análisis estadístico

Para hacer un análisis estadístico de la distribución de clientes hemos utilizado la **prueba Chi-cuadrado de independencia**, ya que permite comprobar si dos variables categóricas (en este caso, país del cliente y segmento de cliente) están asociadas entre sí.

La prueba nos devuelve tres valores principales: el estadístico Chi², que mide la discrepancia entre frecuencias observadas y esperadas; los grados de libertad, que dependen del número de categorías consideradas; y el p-valor, que indica si la relación observada es estadísticamente significativa.

En nuestro caso:

```
Chi-square test
Chi2 = 1.073
Grados de libertad = 2
p-valor = 0.5847
No hay evidencia suficiente para afirmar que la distribución de segmentos dependa del país.
```

En nuestro caso, obtuvimos $\text{Chi}^2 = 1.073$, $gI = 2$ y $p = 0.585$. Dado que el p-valor es mucho mayor que 0.05, no hay evidencia suficiente para rechazar la hipótesis nula de independencia. Esto significa que, según los datos disponibles, la distribución de segmentos de clientes no varía de forma significativa entre países.

Gráfica 5: Clientes con más compras realizadas

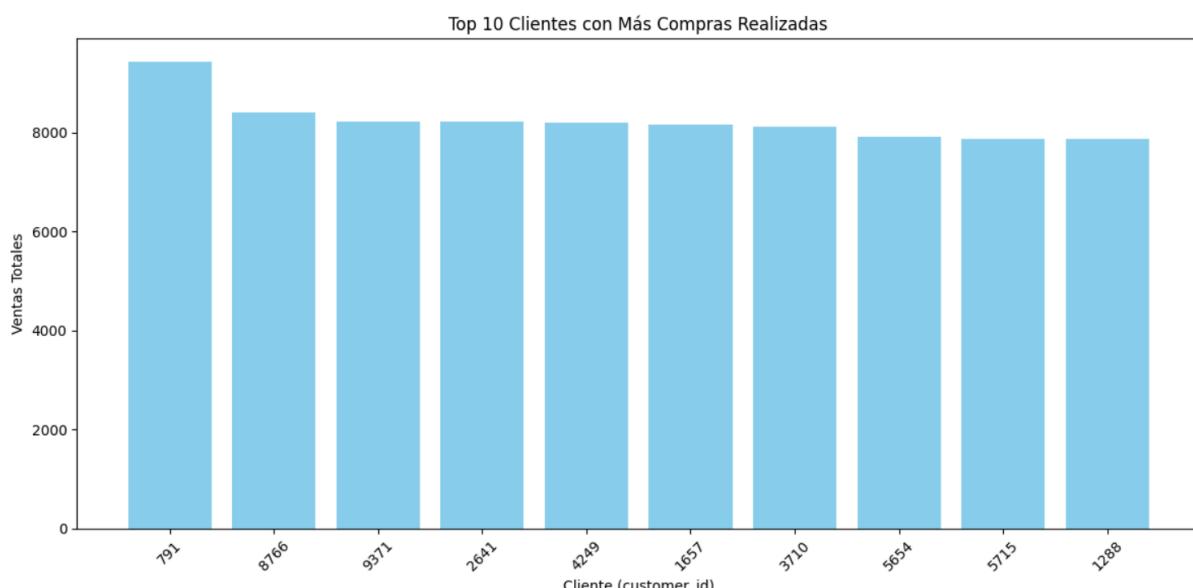
Pregunta planteada

¿Quienes son los clientes con más compras realizadas que se podrían llegar a considerar clientes VIP?

Procedimiento

1. Se seleccionaron las variables, `customer_id`: identificador único del cliente y `order_item_total`: importe total de cada pedido.
2. Se agruparon las compras por cliente y se calculó el total acumulado de ventas.
3. Se ordenaron los resultados de mayor a menor y se seleccionaron los 10 clientes con más compras realizadas.
4. Se utilizó un gráfico de barras, ya que este tipo de visualización permite:
 - Comparar rápidamente los valores entre un número reducido de categorías (clientes).
 - Identificar cuáles destacan sobre el resto en volumen de compras.

Resultados (gráfico)



El cliente con **ID 791** es el que más compras acumula, superando los 9.000 en ventas totales.

El resto de clientes en el top 10 tienen ventas totales que oscilan entre los 7.800 y 8.600, con diferencias menos marcadas entre ellos.

Todos los clientes del top se mantienen en un nivel muy superior al resto de clientes de la base de datos.

Interpretación

Este grupo de clientes representa a los principales generadores de ingresos.

El cliente 791 es el que más destaca como un VIP, ya que concentra más compras que los demás.

Conocer cuales son los clientes con más compras realizadas es una buena idea porque así puedes seguirlos e intentar retenerlos, esto a veces genera más ingresos que buscar a nuevos clientes, por eso creemos que es importante mantener un seguimiento de estos clientes VIP.

Gráfica 6: Evolución de ventas a lo largo del tiempo

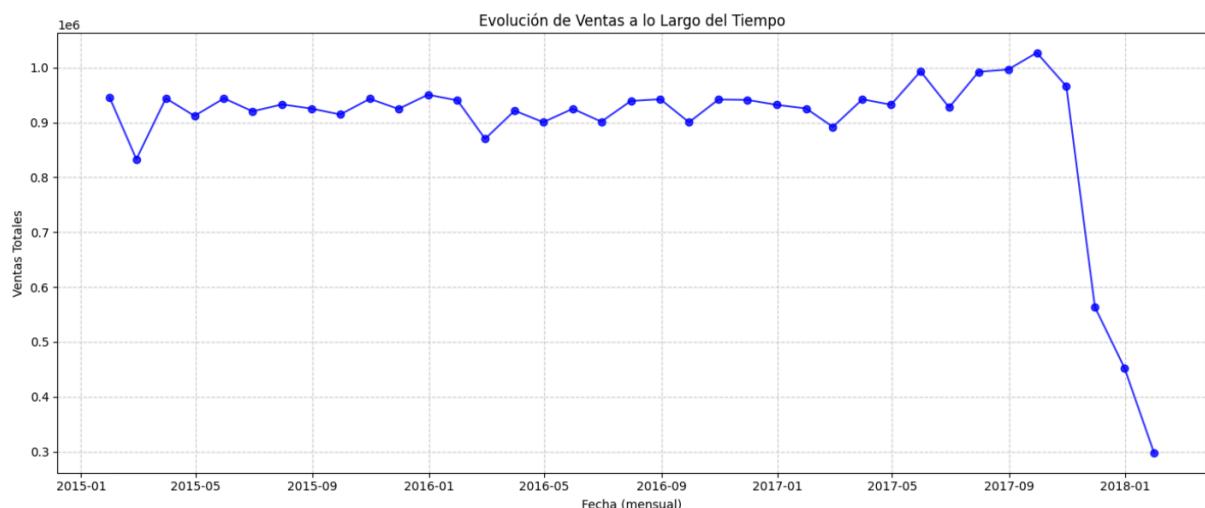
Pregunta planteada

¿Cómo evolucionan las ventas a lo largo del periodo 2015–2018?

Procedimiento

1. Se seleccionaron las variables , order_date_dateorders: fecha del pedido (convertida a tipo fecha) y order_item_total: importe total del pedido.
2. Se agruparon las ventas por mes, para reducir la variabilidad diaria y poder observar tendencias claras en el tiempo.
3. Se utilizó un gráfico de línea, ya que:
 - Permite representar datos temporales en orden cronológico.
 - Facilita la detección de tendencias generales y variaciones estacionales.
 - Muestra con claridad los picos y descensos de ventas en el tiempo.

Resultados (gráfico)



Entre 2015 y mediados de 2017, las ventas se mantuvieron relativamente estables, con pequeñas oscilaciones.

A partir del final de 2017, se observa una caída pronunciada en las ventas, que continúan descendiendo de forma acelerada hasta principios de 2018.

Interpretación

En los primeros años se muestra estabilidad en la evolución de las ventas, sin grandes altibajos.

Al final de la gráfica podemos observar una caída drástica en las ventas, lo que nos permite ver un punto crítico a finales de 2017.

Esta gráfica resulta muy útil para visualizar cómo ha evolucionado el mercado a lo largo del tiempo y permite identificar momentos clave, como esta caída final, que podrían analizarse con más detalle para conocer los motivos de esta caída.

Análisis estadístico

Para analizar la evolución de las ventas a lo largo del tiempo, se realizó un **análisis estadístico descriptivo** de la variable Ventas Totales. Este estudio permite resumir y entender el comportamiento general de una variable cuantitativa continua.

Se calcularon las medidas de tendencia central y dispersión (**media, mediana y desviación estándar**), así como los indicadores de forma asimetría (**skewness**) y kurtosis. Además, se estimó la **correlación de Pearson** entre el tiempo (en meses) y las ventas totales para identificar posibles tendencias.

```
Análisis estadístico de la variable 'Ventas Totales'
```

```
Estadísticos descriptivos:
```

```
count    3.700000e+01
mean     8.933622e+05
std      1.448069e+05
min      2.979521e+05
25%     9.120044e+05
50%     9.321129e+05
75%     9.431311e+05
max      1.027112e+06
Name: order_item_total, dtype: float64
```

```
Asimetría (skewness): -3.016
```

```
Curtosis (kurtosis): 8.426
```

```
Correlación de Pearson tiempo-ventas: R = -0.349, p = 0.0345
```

Interpretación:

Existe una tendencia decreciente significativa en las ventas a lo largo del tiempo.

La distribución presenta asimetría negativa: algunos meses con ventas muy bajas.

En nuestro caso, se obtuvo un coeficiente de correlación $R = -0.349$ ($p = 0.0345$), lo que indica una tendencia decreciente significativa en las ventas a lo largo del tiempo. La distribución presenta una asimetría negativa pronunciada ($skewness = -3.016$), reflejando la existencia de algunos meses con ventas especialmente bajas.

Esto confirma una caída clara en las ventas hacia el final del periodo analizado, tal como se observa en la gráfica de evolución temporal

Gráfica 7 - Categoría más vendida.

Pregunta planteada

¿Cuál ha sido la categoría que ha generado más ventas?

Procedimiento

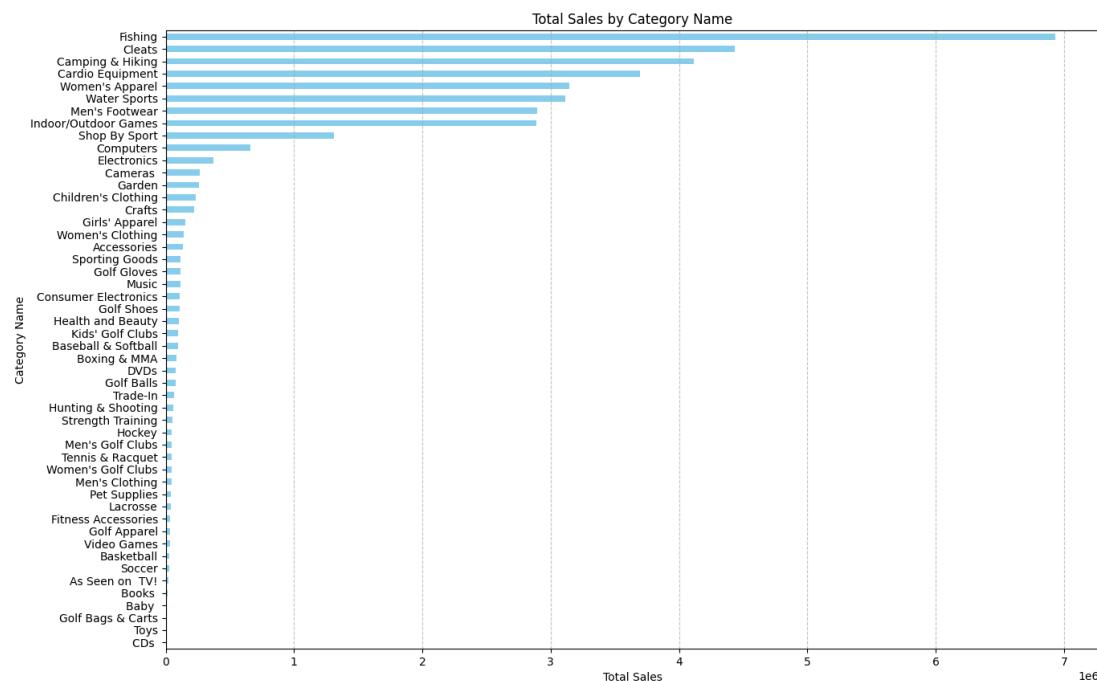
Variables escogidas:

- Número de ventas totales.
- Categoría.

Tipo de gráfico:

Barras horizontales de cada categoría ordenadas por el número de ventas de mayor a menor. Permite ver cuál ha sido la categoría más vendida.

Resultados (gráfico)



Interpretación

Se observa que la pesca es la categoría más vendida de la cadena. Esta gráfica es útil para poder analizar qué categorías son más vendidas o tomar decisiones para motivar las ventas en otras categorías de la cadena.

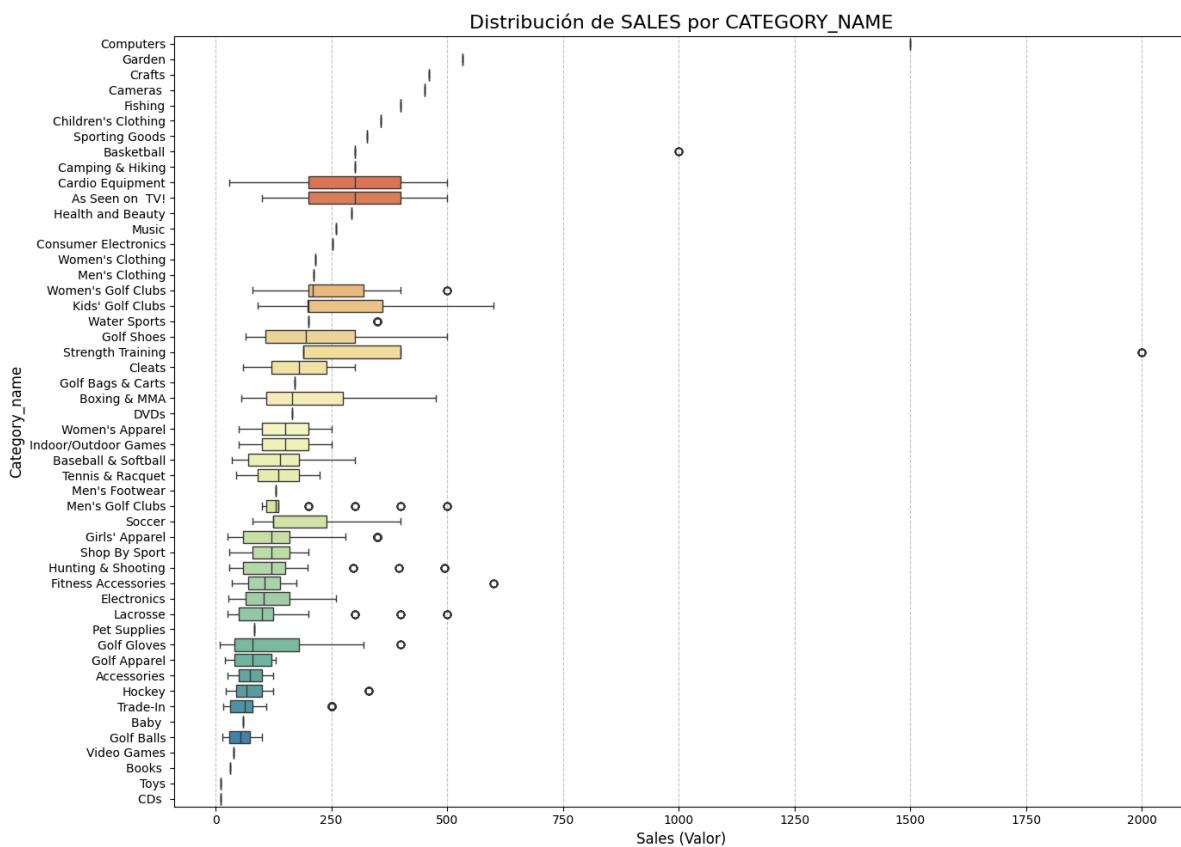
Análisis Estadístico

```
(P1-SIO) ubuntu@Alfonso:/mnt/c/Users/sanch/Documents/GitHub/P1-SIO$ python PART1/G7/statistic.py
=====
ESTADÍSTICAS DESCRIPTIVAS UNIVARIADAS para 'sales'
=====
           sales
count    180519.000000
mean      203.772096
std       132.273077
min       9.990000
25%     119.980003
50%     199.919998
75%     299.950012
max     1999.989990
skewness   2.884249
kurtosis   23.936561
IQR      179.970009
```

Las ventas muestran una dispersión considerable, con una desviación estándar de 132, lo que indica variaciones significativas entre los valores. El promedio de ventas fue de 203, mientras que el 75% de las compras tuvieron un valor de 119.

La distribución es asimétrica positiva (skewness), sugiriendo que la mayoría de las ventas son pequeñas, pero las grandes ventas elevan la media sustancialmente. La curtosis también indica que, si bien la mayoría de las ventas son pequeñas, existen "colas pesadas" o ventas grandes más frecuentes de lo esperado, afectando el promedio.

El rango intercuartílico, que abarca el 50% central de las transacciones, se sitúa en 179. Se espera que, a excepción del skewness y la curtosis, los demás valores se expresen en la moneda de las transacciones.



Este gráfico presenta diagramas de caja (boxplots) que ilustran las desviaciones en las ventas de cada producto. Se puede observar que algunos productos no muestran desviación, mientras que otros exhiben una mayor variabilidad en sus ventas, lo que resulta útil para analizar el comportamiento individual de cada producto.

Gráfica 8 - Beneficio para cada tipo de cliente.

Pregunta planteada

¿Qué tipo de clientes están generando más beneficios en la cadena?

Procedimiento

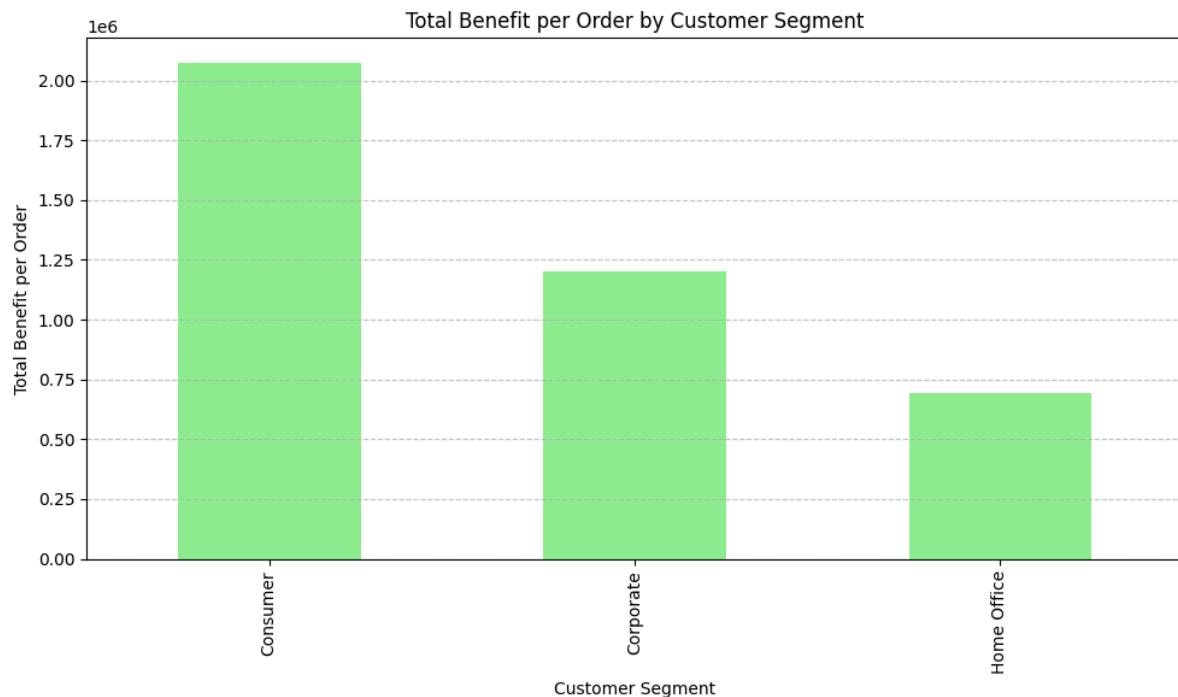
Variables seleccionadas:

- Beneficio total.
- Segmento de cliente.

Tipo de gráfico:

Se ha escogido barra vertical por tipo de cliente mostrando en la barra el beneficio que ha ofrecido cada segmento de cliente

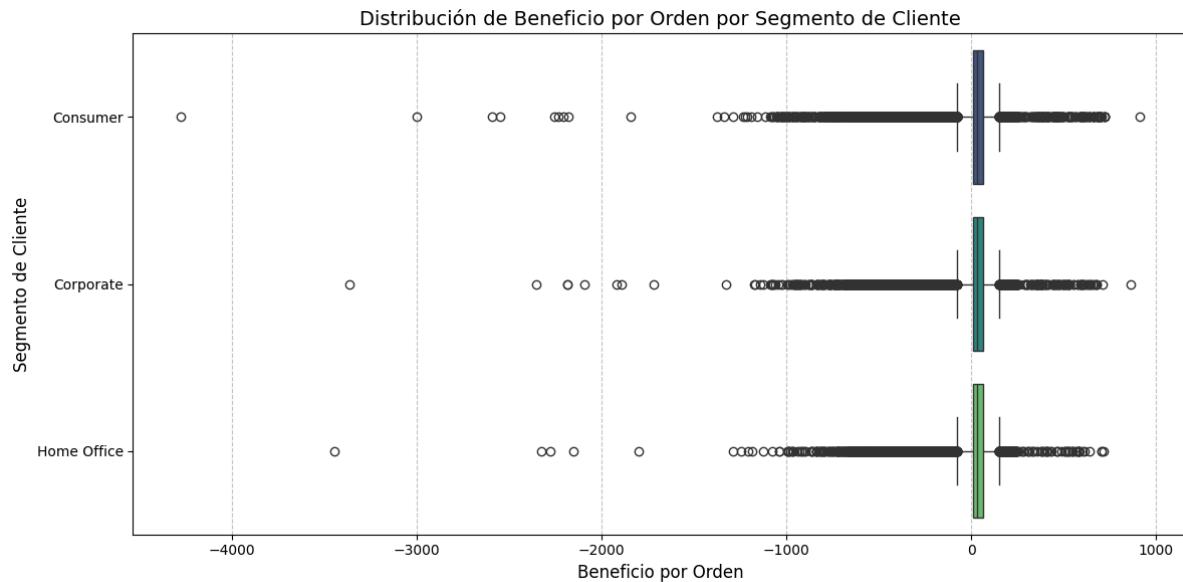
Resultados (gráfico)



Interpretación

Se observa que los consumidores de la cadena son los que mayores beneficios generan mientras que las empresas y las 'home office' generan un menor número de beneficios. Esta gráfica permite tomar decisiones para motivar las ventas en aquellos clientes que mayores beneficios dan, sin embargo antes de ello se tendría que revisar los datos con otras gráficas para poder concluir el motivo de este beneficio, es probable que la mayor frecuencia de cliente que se tenga en la cadena está motivando este resultado.

Análisis Estadístico



Al analizar los diagramas de caja, se observa una variabilidad significativa en los datos de beneficio entre los tres tipos de clientes.

ESTADÍSTICAS CONDICIONALES: 'benefit_per_order' por 'customer_segment'						
	Total_Beneficio	Ordenes_Contadas	Media_Beneficio	Mediana_Beneficio	Desviacion_Estandar	IQR
customer_segment						
Consumer	2.073488e+06	93504	22.175390	31.850000	104.537546	57.690000
Corporate	1.202575e+06	54789	21.949204	31.320000	103.874259	58.250000
Home Office	6.908403e+05	32226	21.437359	31.110001	105.079155	57.420001

Obteniendo la mediana, media, la desviación estandar y el IQR nos permite observar que estamos ante una serie de datos con mucha variabilidad. Los clientes compran en diferentes cantidades y precios. El análisis de los datos de compra (mediana, media, desviación estándar, IQR) revela una gran heterogeneidad en el comportamiento del cliente, con amplias variaciones en cantidades y precios. La diferencia entre mediana y media sugiere asimetría, y la alta desviación estandar confirma la dispersión de las compras. Un IQR amplio indica variabilidad incluso en compras "típicas".

Gráfica 9 - Ratio de beneficio

Pregunta planteada

¿Cuánto ratio de beneficio tiene la cadena? ¿Existe algún ratio de beneficio negativo?

Procedimiento

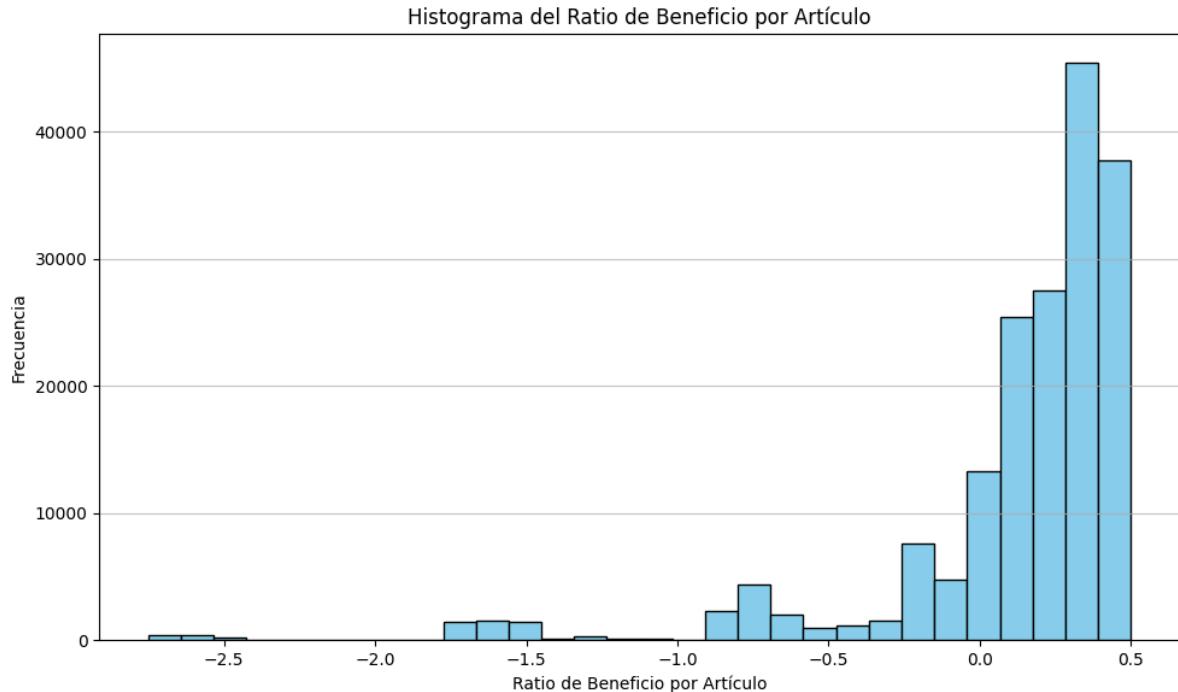
Variables:

- Order item profit ratio

Tipo de gráfico:

Se ha realizado un histograma de esta variable para observar donde se encuentra la mayoría del profit de la cadena.

Resultados (gráfico)



Interpretación

Observamos que la mayoría de órdenes han generado un margen de beneficio entre el 0% y el 50% mientras que otras órdenes han generado pérdidas. Se deberá analizar cuáles órdenes han generado pérdidas, que productos se han vendido y si existe alguna oferta que genere pérdidas en la cadena o un error de precio. Esta gráfica permite enfocar la vista en el beneficio de cada venta, se debe lograr que todas las ventas se encuentren al lado positivo es decir que den beneficio.

Análisis Estadístico

```
(P1-SIO) ubuntu@Alfonso:/mnt/c/Users/sanch/Documents/GitHub/P1-SIO$ python PART1/G9/analisis.py
=====
ANÁLISIS ESTADÍSTICO para 'order_item_profit_ratio'
=====

Estadísticas Descriptivas Completas:
|           | count   | mean    | std     | min    | 25%   | 50%   | 75%   | max    | Skewness | Kurtosis |
|:-----|:-----|:-----|:-----|:-----|:-----|:-----|:-----|:-----|:-----|:-----|
| order_item_profit_ratio | 180519 | 0.120647 | 0.466796 | -2.75 | 0.08 | 0.27 | 0.36 | 0.5 | -2.89353 | 10.1572 |
=====

ANÁLISIS CONDICIONAL: Ratio de Beneficio vs Umbral de 0.0
=====
Total de Órdenes Analizadas: 180519
Porcentaje de Órdenes Rentables (> 0): 80.63%
Porcentaje de Órdenes No Rentables (<= 0): 19.37%
=====

🟡 Tres cuartas partes de los artículos son rentables. Revisar las categorías o artículos en el 25% inferior.
```

Primero, la media es superior a la mediana, lo que indica un sesgo negativo, esto sugiere que la mayoría de los ratios son altos, pero existen pérdidas significativas que reducen el promedio. La curtosis muestra que la distribución está agrupada y presenta colas pesadas, lo que implica una mayor probabilidad de valores extremos. La elevada desviación estandar también apunta a una gran variabilidad, causada por estos valores extremos.

Se observa que el 19.37% de los artículos generan pérdidas. Además, estas pérdidas son tan considerables que un pequeño porcentaje de transacciones no rentables tiene un impacto masivo en el rendimiento promedio.

En resumen, aunque muchos artículos generan beneficios, algunas de estas transacciones están provocando grandes pérdidas.

Gráfica 10 - Mayor porcentaje de retraso

Pregunta planteada

¿Cuáles son los productos que tienen mayor porcentaje de retraso?

Procedimiento

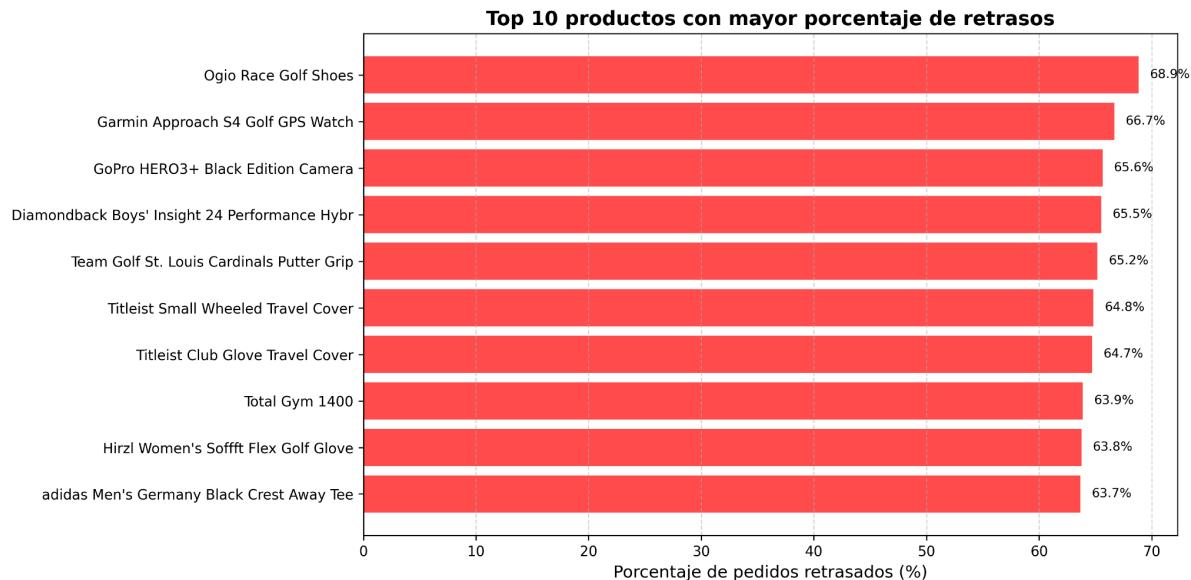
Variables:

- Nombre de producto.
- Días de entrega real.
- Días de entrega programada.
- Porcentaje de retraso por producto (calculado).

Tipo de gráfico:

Se ha decidido mostrar barras horizontales que muestran el % de retraso de cada producto clasificado de mayor a menor y mostrando solo los 10 primeros productos.

Resultados (gráfico)



Interpretación

Esta gráfica nos indica qué productos están teniendo más retrasos en las entregas. Esto permite a la cadena enfocarse en mejorar la entrega de los productos que están en este gráfico. A partir de estos datos se puede realizar un estudio de por qué estos productos están generando retrasos en las entregas.

Análisis Estadístico

ANÁLISIS ESTADÍSTICO DE LA SEVERIDAD DEL RETRASO ('days_shipping_diff')									
Estadísticas (solo para pedidos RETRASADOS - días > 0):									
count mean std min 25% 50% 75% max Skewness Kurtosis									
:-----: :-----: :-----: :-----: :-----: :-----: :-----: :-----: :-----: :-----:									
days_shipping_diff 103400 1.61674 0.882034 1 1 1 2 4 1.4176 1.16252									
Interpretación Clave:									
-> Media del Retraso: 1.62 días.									
-> Desv. Estándar: 0.88. Indica la consistencia del retraso.									
-> Máximo Retraso: 4 días (Outlier de servicio).									

Se han identificado 103 400 pedidos retrasados, lo que requiere atención. El retraso habitual es de 1 día (mediana), aunque el promedio se sitúa en 1.62 días.

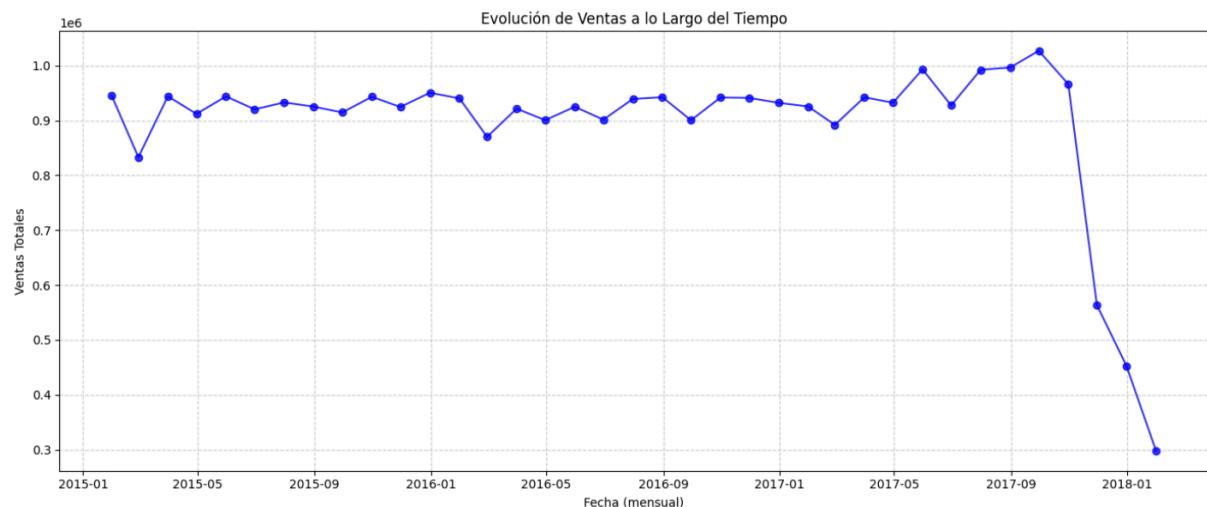
La asimetría positiva indica que la mayoría de los retrasos son de un día. Sin embargo, una cola de distribución con retrasos más largos (2-4 días) eleva el promedio a 1.62 días. La curtosis sugiere una alta concentración de datos alrededor de la mediana (1 día), pero con una probabilidad mayor de lo esperado de retrasos prolongados en comparación con una distribución normal.

La baja desviación estándar confirma que los retrasos son bastante consistentes y rara vez superan los 2 días.

Estudio - Bajada de ventas en 2018.

Motivación

Hemos observado que las ventas/beneficios han descendido bruscamente en 2018. ¿Por qué ha ocurrido esto?



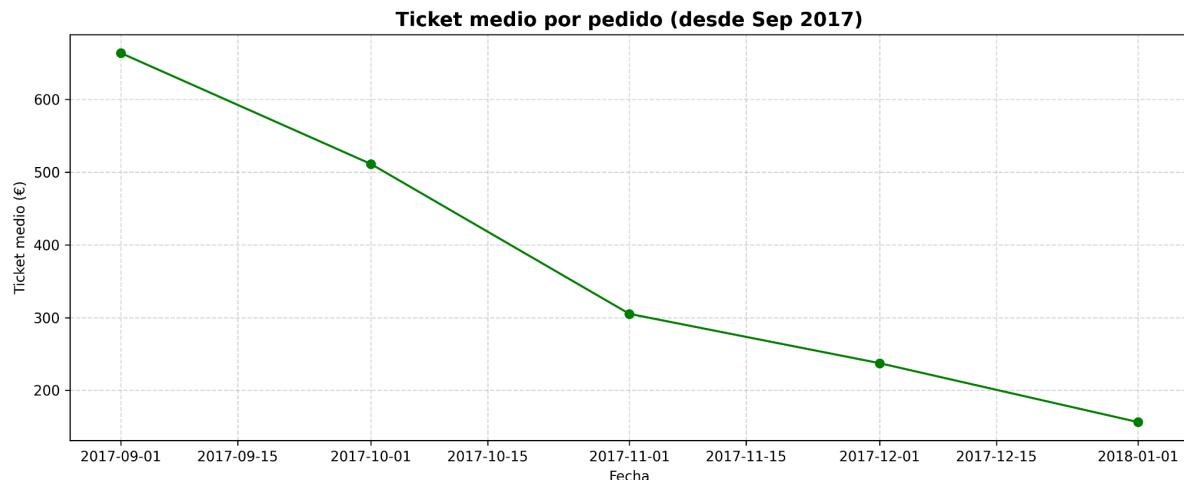
Enfoque

Se decide investigar los siguientes datos especialmente entre el año 2017 y los datos a partir de septiembre de 2017 cuando inició la caída aproximadamente.

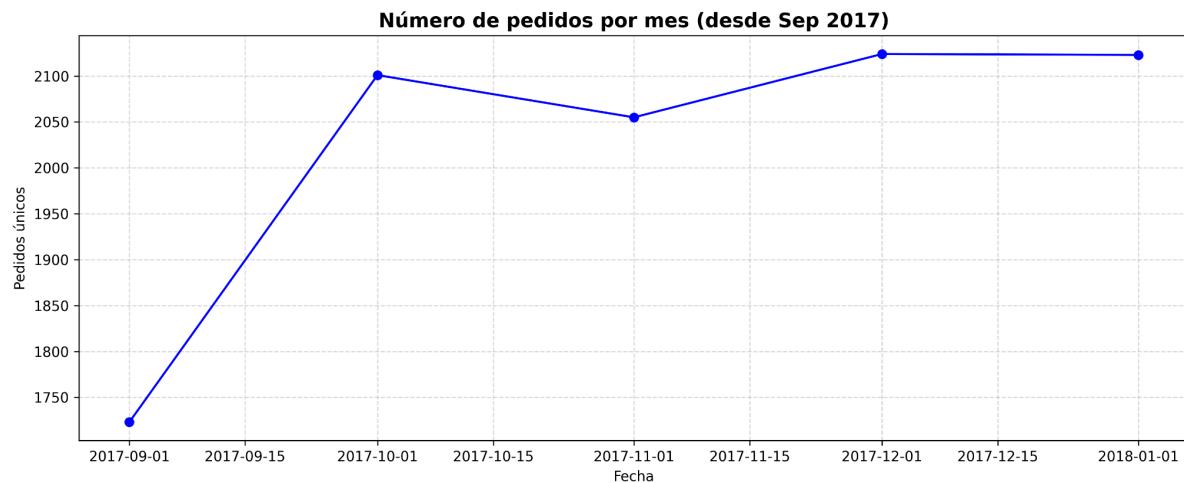
Datos que se investigan:

- Ticket medio por pedido.
- Pedidos por mes.
- Top 10 categorías vendidas entre un periodo y otro.
- Distribución del estado de los pedidos.

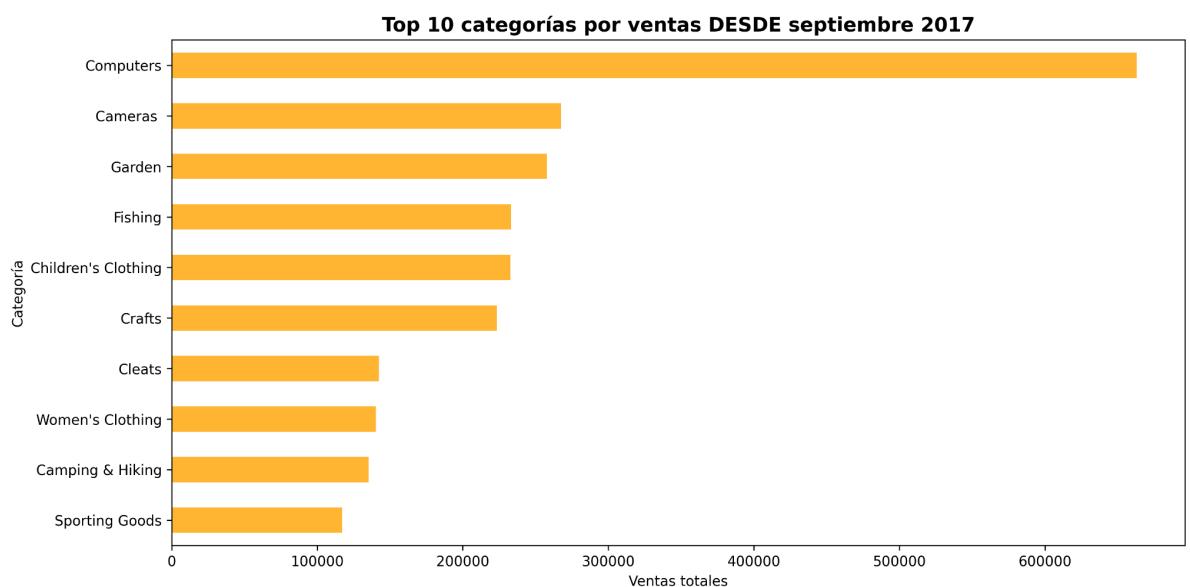
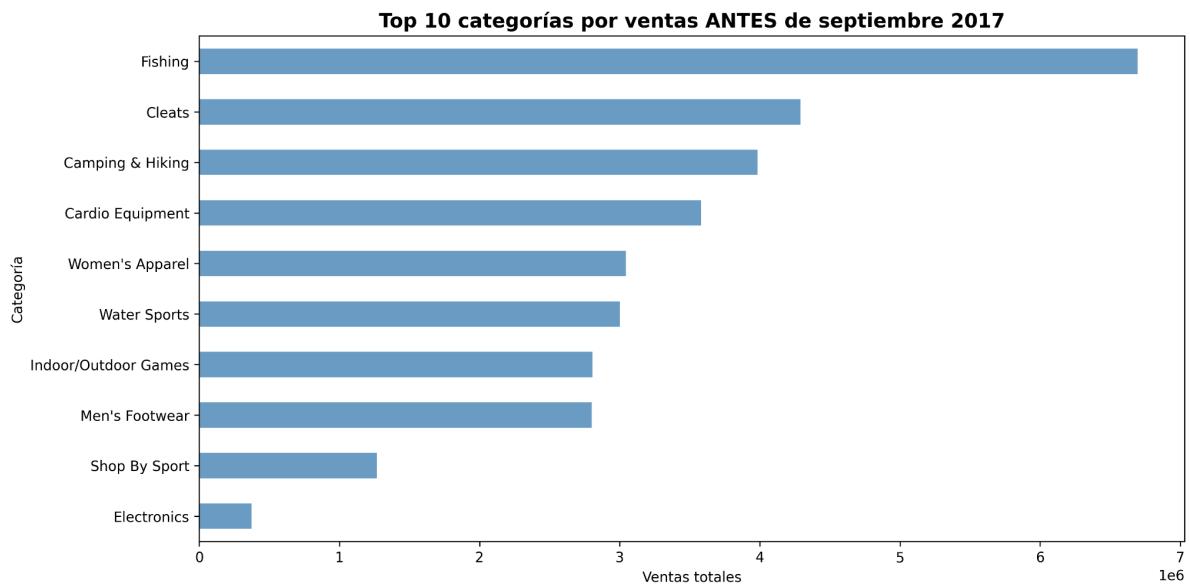
Datos encontrados



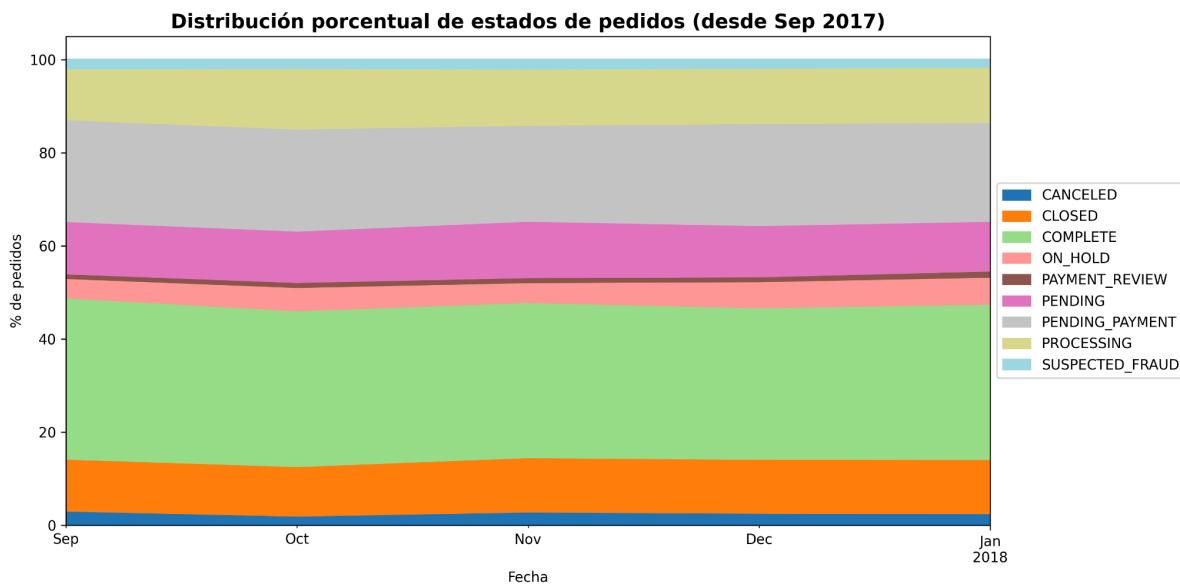
Se observa que el ticket medio por pedido ha descendido indicando que los clientes realizan compras más pequeñas generando menores ventas y por tanto menores beneficios.



Sin embargo, se ha encontrado que el número de pedidos por mes ha ascendido aunque no ha sido suficiente para recompensar la pérdida de ticket medio visto anteriormente.



Las categorías más vendidas entre cada periodo ha variado. Se puede asociar a que la caída de ventas ha sido provocada por un cambio de categoría siendo la pesca más cara que los ordenadores. Posiblemente se ha perdido atracción con los productos de pesca que generaban mayor cantidad de beneficio y tickets más voluminosos.



Los pedidos se han comportado prácticamente igual en el periodo de caída, por lo que no se puede asociar a problemas de envío.

Conclusión del estudio

Se concluye que el cambio de atracción de los clientes sustituyendo la categoría de pesca por la de ordenadores ha motivado esta caída de beneficio. La pesca asocia productos más caros que los ordenadores y a pesar de que se realizan más ventas no son suficientes para recompensar lo obtenido con la pesca. Es probable que el no tener ofertas de pesca o dejar de motivar a los clientes de la cadena a comprar productos de pesca haya motivado este cambio.

Estudio - Relación entre descuentos, beneficios y márgenes en las ventas

Motivación

Se quiere analizar cómo los descuentos influyen en el margen de beneficio y el rendimiento general de la cadena, así como identificar qué categorías o modos de envío son más sensibles a este efecto.

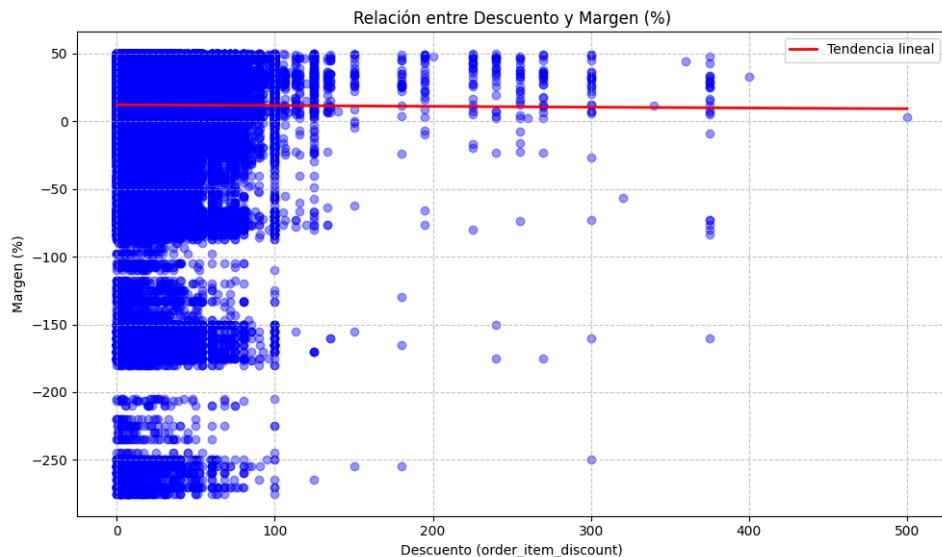
Enfoque

Se relacionan variables comerciales y logísticas como descuentos, márgenes, categorías de producto, modos de envío y beneficio temporal. El objetivo es comprender cómo las políticas de precios y las decisiones logísticas influyen en la rentabilidad y eficiencia global de la cadena.

Gráficas realizadas

Gráfica 1

En primer lugar, se analizó la relación entre el descuento aplicado y el margen de beneficio . Para ello, se representaron ambas variables mediante un gráfico de dispersión con una línea de tendencia lineal. El objetivo era comprobar si los descuentos influyen directamente en la rentabilidad.

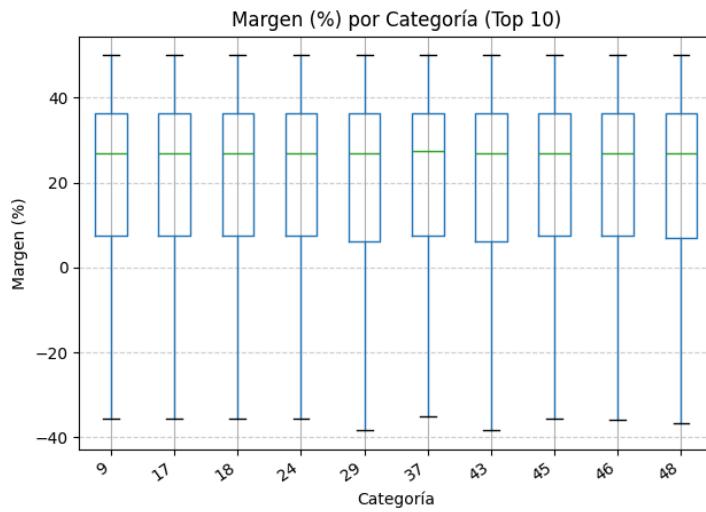


Sin embargo, los resultados obtenidos ($R = -0.003$, $p = 0.2346$) muestran que no existe una relación lineal significativa entre el descuento y el margen. Esto indica que los descuentos aplicados no afectan de forma clara o consistente a los márgenes, por lo que otros factores pueden tener mayor peso en la rentabilidad del pedido.

```
PS C:\Users\Antoni\Desktop\SI0 P1> & C:/Users/Antoni/AppData/Local/Programs/Python/Python313/python.exe "c:/Users/Antoni/Desktop/SI0 P1/P1-SI0/PART1/Estudio 2/descuento_vs_margen.py"
Correlación de Pearson: R = -0.003, p = 0.2346
No se observa una relación lineal significativa entre descuento y margen.
```

Gráfica 2

A continuación, se realizó un segundo análisis centrado en el margen (%) por categoría de producto. Con esta comparación se buscaba observar si existen diferencias relevantes en la rentabilidad según el tipo de producto. En el gráfico se aprecia que las distribuciones de margen son bastante similares entre categorías, sin destacar ninguna de manera clara.



Para confirmar esta observación se aplicó un ANOVA, que dio como resultado $F = 1.05$ y $p = 0.394$, indicando que no existen diferencias estadísticamente significativas entre los márgenes medios de las distintas categorías.

```
PS C:\Users\Antoni\Desktop\SIO P1> & C:/Users/Antoni/AppData/Local/Programs/Python/Python313/python.exe "c:/Users/Antoni/Desktop/SIO P1/P1-SIO/PART1/Estudio 2/margen_por_categoria.py"
ANOVA: F=1.05, p=3.9368e-01
No hay diferencias significativas.
```

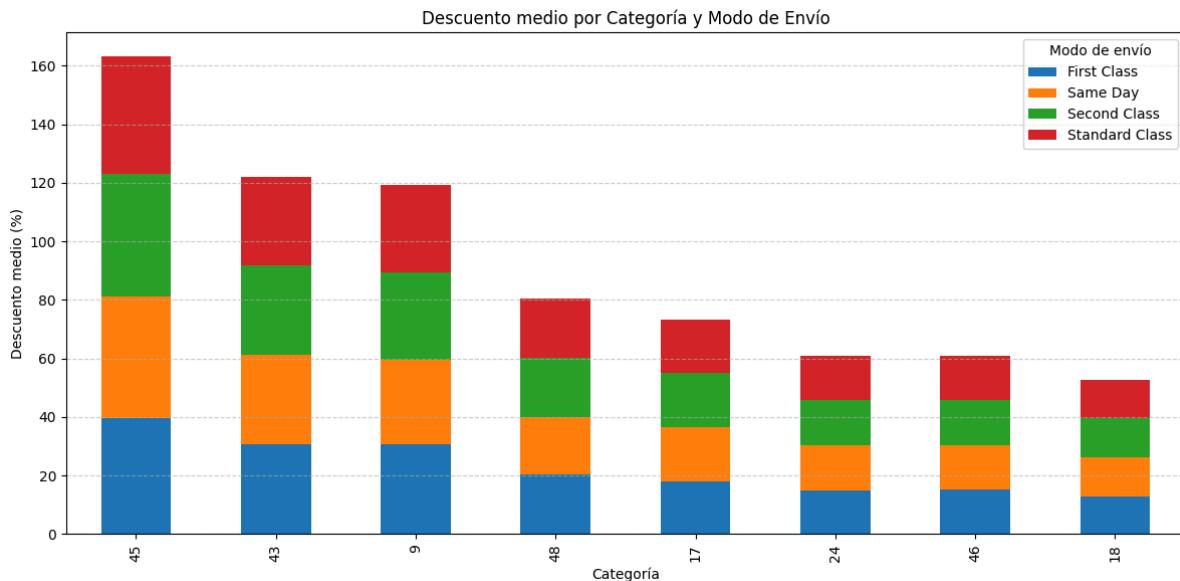
Gráfica 3

Después de comprobar que ni el descuento ni la categoría afectan directamente al margen, se decidió analizar cómo se distribuyen los descuentos medios según la categoría de producto y el modo de envío

Este paso permite observar si las estrategias de descuento están ligadas a la logística o al tipo de producto.

En el gráfico de barras apiladas se puede ver claramente que algunas categorías reciben descuentos medios mucho más altos que otras, destacando la categoría 45 con un descuento promedio del 40.56%, seguida de las categorías 43 y 9, con valores cercanos al 30%.

Además, los distintos colores dentro de cada barra muestra el reparto de descuentos por modo de transporte. En general, los modos de envío más rápidos, como Same Day o First Class, concentran una parte importante de los descuentos, mientras que los envíos estándar presentan valores más moderados.



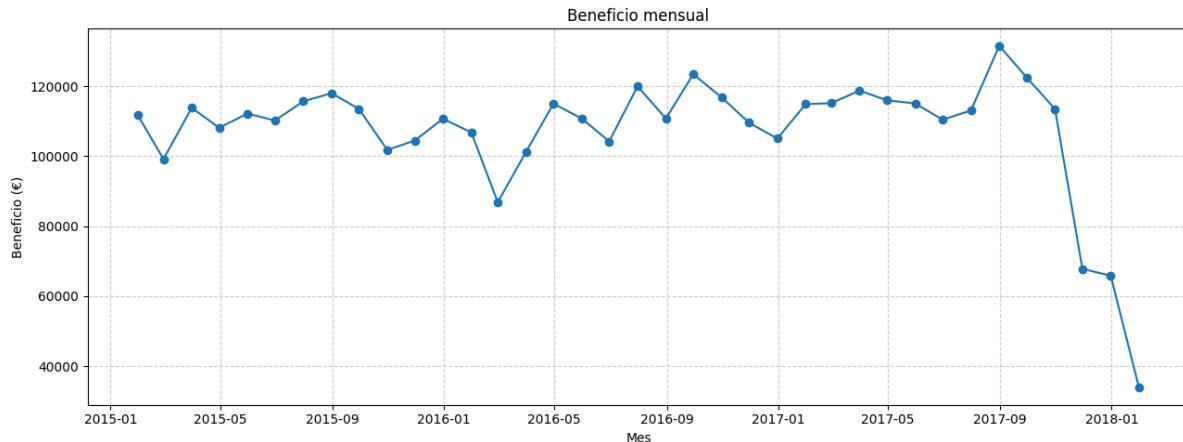
```
PS C:\Users\Antoni\Desktop\SIO P1 & C:/Users/Antoni/AppData/Local/Programs/Python/Python313/python.exe "c:/Users/Antoni/Desktop/SIO P1/P1-SIO/PART1/Estudio 2/descuento_envio.py"
.
.
.
Top categorías por descuento medio (promedio sobre modos):
category_id
45    40.561501
43    30.420970
9     29.998814
48    20.321782
17    18.292195
Name: order_item_discount, dtype: float64
PS C:\Users\Antoni\Desktop\SIO P1> [
```

Estos resultados indican que las políticas de descuento varían tanto por tipo de producto como por modalidad de envío, posiblemente como estrategia comercial para incentivar la compra de determinadas categorías o compensar los mayores costes logísticos de los transportes urgentes.

Gráfica 4

Tras analizar los descuentos y márgenes en función de las categorías y modos de envío, el siguiente paso fue observar cómo evoluciona el beneficio total a lo largo del tiempo.

En la gráfica se aprecia que, durante la mayor parte del tiempo, el beneficio se mantiene relativamente estable, con pequeñas oscilaciones normales entre meses. Sin embargo, a partir de finales de 2017 se observa una caída pronunciada en los beneficios.



```
PS C:\Users\Antoni\Desktop\SIO P1> & C:/Users/Antoni/AppData/Local/Programs/Python/Python313/python.exe "c:/Users/Antoni/Desktop/SIO P1/P1-SIO/PART1/Estudio 2/beneficio_mensual.py"
Correlación tiempo-beneficio: R=-0.261, p=1.1820e-01
Sin tendencia lineal significativa.
PS C:\Users\Antoni\Desktop\SIO P1> 
```

Para complementar la observación visual, se calculó la correlación de Pearson entre el tiempo y el beneficio mensual, obteniéndose un valor $R = -0.261$ y $p = 0.118$, lo que indica que no existe una tendencia lineal significativa en el conjunto total del periodo.

El análisis de esta caída lo hemos realizado anteriormente.

Conclusión del estudio

Los resultados muestran que ni los descuentos ni el tipo de producto influyen directamente en el margen, pero sí se observan diferencias en cómo se aplican los descuentos según la categoría y el modo de transporte.

El beneficio mensual se mantiene estable durante la mayor parte del periodo, aunque presenta una fuerte caída al final de 2017.

En general, se concluye que la rentabilidad de la empresa es homogénea entre productos, y que las políticas de descuento y la gestión logística son los factores que más afectan al rendimiento global.

Estudio - Relación entre modo de envío, tiempos de entrega y rentabilidad

Motivación

Queremos analizar si el modo de transporte y la categoría del producto influyen conjuntamente en la eficiencia de envío (tiempo real) y en la rentabilidad (beneficio). En otras palabras, ¿ciertos tipos de producto se envían más rápido o generan más beneficio dependiendo del modo de transporte utilizado?

Enfoque

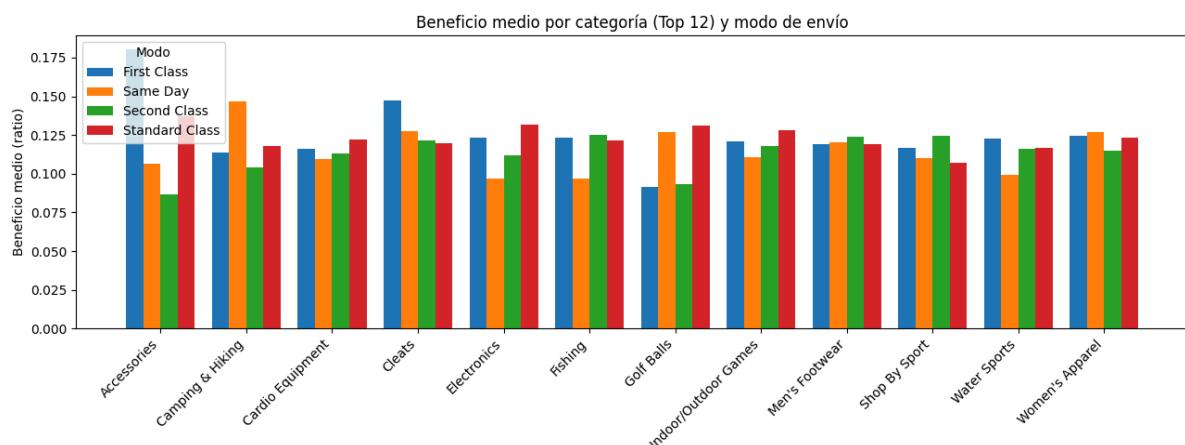
En este estudio comparamos variables como el beneficio medio, los tiempos y la puntualidad de las entregas, así como la relación entre los días reales de envío y el margen obtenido. Con este análisis se busca entender si los distintos tipos de transporte afectan al rendimiento económico y operativo de la cadena de suministro.

Gráficas realizadas

Gráfica 1

En primer lugar, se analizó el beneficio medio (ratio) obtenido en función de la categoría del producto y del modo de envío.

El objetivo era determinar si ciertos tipos de productos o métodos logísticos generan mayores márgenes de rentabilidad que otros.



En el gráfico se aprecia que las diferencias entre modos y categorías son reducidas. Algunas categorías, como *Cleats* o *Accessories*, presentan ratios de beneficio ligeramente superiores, especialmente en envíos Same Day y First Class, pero en general los valores se mantienen en un rango muy similar entre 0.09 y 0.17. Esto sugiere que los márgenes de beneficio se distribuyen de forma bastante uniforme entre los diferentes productos y servicios logísticos.

```
PS C:\Users\Antoni\Desktop\SIO P1> & c:/Users/Antoni/AppData/Local/Programs/Python/Python313/python.exe "c:/Users/Antoni/Desktop/SIO P1/P1-SIO/PART1/Estudio 3/benef_cat_modo.py"
ANOVA categoría: F=0.983, p=4.593e-01
ANOVA modo: F=1.243, p=2.922e-01
Interpretación rápida: p<0.05 => diferencias significativas en el beneficio medio.
PS C:\Users\Antoni\Desktop\SIO P1>
```

Para verificar si las diferencias observadas son estadísticamente significativas, se aplicó un **ANOVA** de un factor tanto por categoría como por modo de envío.

Los resultados obtenidos fueron $F = 0.983$, $p = 0.459$ para las categorías y $F = 1.243$, $p = 0.292$ para los modos, lo que indica que no existen diferencias significativas ($p > 0.05$) entre los grupos analizados.

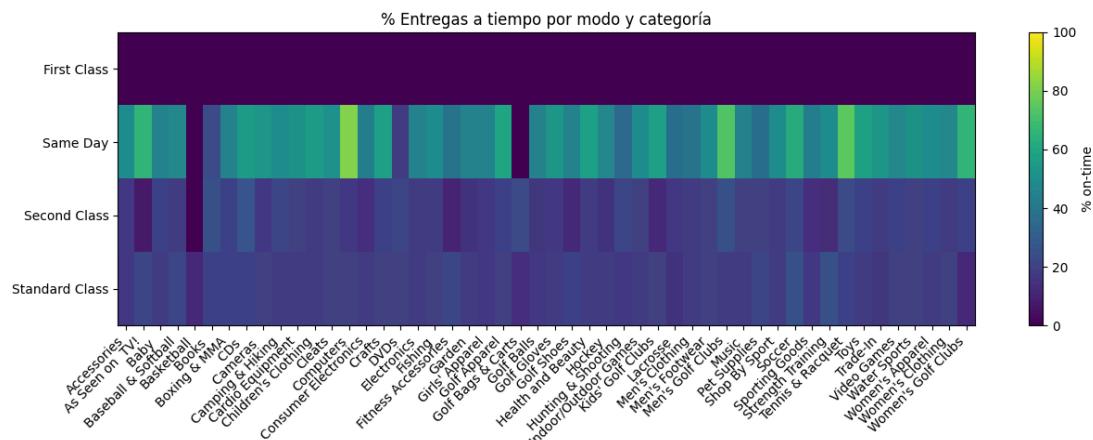
En conclusión, el estudio muestra que la rentabilidad media no depende de manera significativa ni del tipo de producto ni del modo de envío.

Gráfica 2

En esta segunda parte, se analizó el porcentaje de entregas realizadas a tiempo en función del modo de envío y la categoría del producto.

El objetivo era comprobar si existen diferencias en la puntualidad de las entregas según el tipo de transporte utilizado y si ciertas categorías de productos presentan más retrasos que otras.

Se generó una nueva variable binaria llamada `on_time` (1 = entregado a tiempo, 0 = no entregado a tiempo) a partir de la información de `delivery_status`. Se calculó el porcentaje medio de entregas a tiempo agrupando los datos por modo y categoría, y se representó mediante un heatmap



En el gráfico se observa que los envíos Same Day son, en general, los que presentan los mayores porcentajes de entregas puntuales, con valores muy cercanos al 100 %.

Los envíos First Class también mantienen una buena puntualidad, mientras que los modos Second Class y especialmente Standard Class concentran los porcentajes más bajos, evidenciando tiempos de entrega más largos y mayor variabilidad entre categorías.

No se aprecian diferencias claras entre categorías de producto: todas muestran una tendencia similar en cada modo de transporte.

Se aplicó una prueba **Chi-cuadrado** de independencia, comparando la variable `on_time` frente a los factores `shipping_mode` y `category_name`.

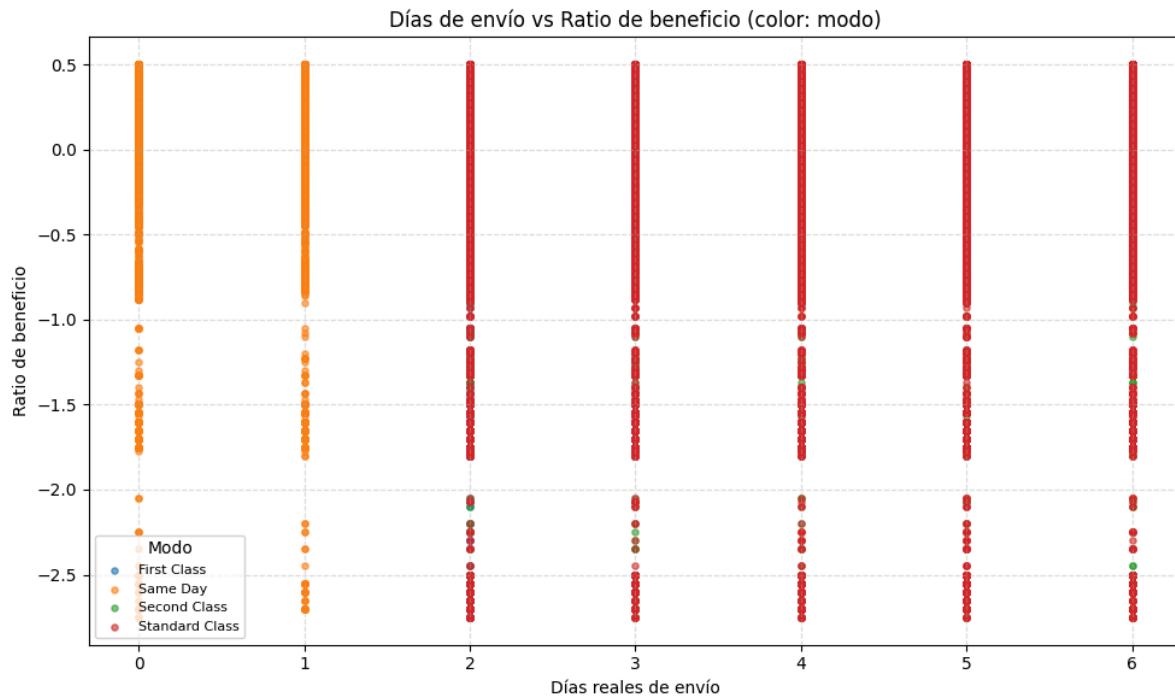
```
PS C:\Users\Antoni\Desktop\SIO P1> & C:/Users/Antoni/AppData/Local/Programs/Python/Python313/python.exe "c:/Users/Antoni/Desktop/SIO P1/P1-SIO/PART1/Estud... 3/entregas_tiempo_modo.py"
Chi^2 on_time ~ modo: p=0.000e+00
Chi^2 on_time ~ categoría: p=2.140e-01
PS C:\Users\Antoni\Desktop\SIO P1>
```

Los resultados obtenidos fueron $p = 0.000$ para el modo de envío y $p = 0.214$ para la categoría, lo que indica que existe una relación significativa entre la puntualidad y el modo de envío, pero no con la categoría del producto.

Gráfica 3

A continuación, se analizó la relación entre los días reales de envío y el ratio de beneficio obtenido en cada pedido.

El objetivo era determinar si los pedidos con tiempos de envío más largos presentan márgenes de rentabilidad diferentes respecto a los envíos más rápidos, y comprobar si la logística influye de forma directa en la rentabilidad de la cadena.



En el gráfico se observa que los pedidos con entregas más rápidas (Same Day) concentran la mayor densidad de puntos en torno a ratios de beneficio cercanos a cero, con poca dispersión.

Los envíos Standard Class y Second Class, que implican más días de transporte, muestran una distribución más amplia, incluyendo valores negativos de beneficio.

Esto sugiere que los envíos más lentos podrían estar asociados a mayores costes logísticos o a descuentos aplicados para compensar la demora.

```
PS C:\Users\Antoni\Desktop\SIO P1> & c:/Users/Antoni/AppData/Local/Programs/Python/Python313/python.exe "c:/Users/Antoni/Desktop/SIO P1/P1-SIO/PART1/Estudio 3/envio_vs_beneficio.py"
c:/Users/Antoni/Desktop/SIO P1/P1-SIO/PART1/Estudio 3/envio_vs_beneficio.py:18: UserWarning: Creating legend with loc="best" can be slow with large amounts of data.
plt.tight_layout(); plt.savefig("g3_scatter_dias_vs_ratio.png"); plt.show()
C:\Users\Antoni\AppData\Local\Programs\Python\Python313\lib\tkinter\_init__.py:862: UserWarning: Creating legend with loc="best" can be slow with large amounts of data.
func(*args)
Correlación Pearson (global): R=-0.005, p=4.877e-02
PS C:\Users\Antoni\Desktop\SIO P1> 
```

Para cuantificar la relación entre ambas variables, se aplicó un **análisis de correlación de Pearson**, obteniendo un coeficiente $R = -0.005$ con un p -valor $= 0.0487$.

Aunque el valor de R es muy cercano a cero, el p -valor indica una relación negativa débil pero significativa entre el tiempo de envío y el ratio de beneficio.

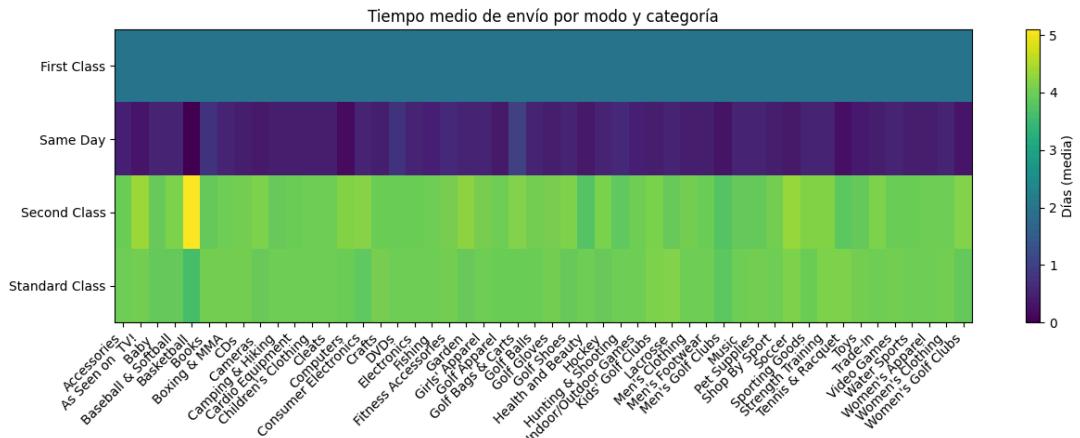
Esto significa que, a medida que aumentan los días de envío, tiende a disminuir ligeramente el margen de rentabilidad.

En conclusión, el estudio evidencia que los pedidos con mayor tiempo de envío presentan una tendencia leve hacia menores beneficios.

Gráfica 4

En esta última gráfica se estudió el tiempo medio de envío según el modo de transporte y la categoría del producto.

El propósito era comprobar si determinados tipos de productos tardan más en enviarse y, sobre todo, confirmar si los distintos modos de envío cumplen con los tiempos esperados.



El gráfico deja muy clara la diferencia entre los modos de envío.

Los pedidos enviados en Same Day son, como era de esperar, los más rápidos, con tiempos de entrega cercanos a cero días.

Le sigue First Class, con tiempos medios muy estables en torno a los 2 días.

En cambio, los envíos Second Class y Standard Class son considerablemente más lentos, con tiempos que rondan los 4 a 5 días, y una ligera variabilidad según la categoría.

No se aprecian diferencias destacables entre productos: todas las categorías presentan patrones muy similares dentro de cada modo.

```
PS C:\Users\Antoni\Desktop\S10_P1 & C:/Users/Antoni/AppData/Local/Programs/Python/Python313/python.exe "C:/Users/Antoni/Desktop/S10_P1-S10/PART1/Estudio_3/tenvio_modo_categoria.py"
...
ANOVA días ~ modo: F=38762.414, p=0.000e+00
ANOVA días ~ categoría: F=1.152, p=2.161e-01
```

Para comprobar si estas diferencias son estadísticamente significativas, se realizó un

ANOVA tanto por modo de envío como por categoría de producto.

Los resultados mostraron un valor $F = 38762.414$, $p = 0.000$ para el modo y $F = 1.152$, $p = 0.216$ para la categoría.

Esto confirma que el modo de envío influye de forma significativa en el tiempo de entrega, mientras que la categoría del producto no tiene un impacto relevante.

Conclusión del estudio

El análisis muestra que la cadena de suministro funciona de forma estable y eficiente.

Los beneficios son parecidos entre las distintas categorías y modos de envío, lo que indica una buena gestión de precios y costes.

El modo de transporte es el factor que más influye en los tiempos de entrega: los envíos Same Day y First Class llegan mucho más rápido y con más puntualidad, mientras que los modos Standard y Second Class tardan más.

También se observa que los pedidos con más días de envío tienden a generar un poco menos de beneficio, aunque la diferencia es pequeña.

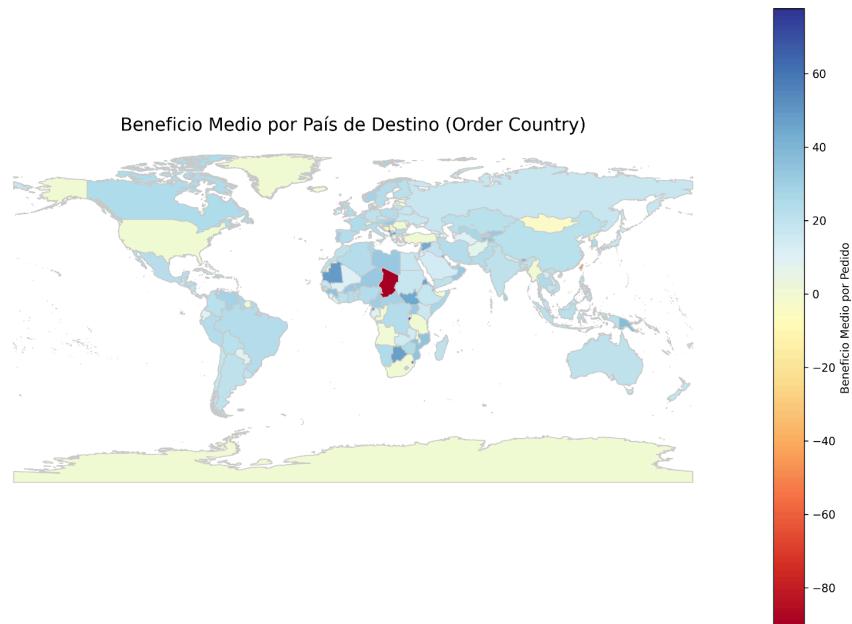
En general, se mantiene una logística equilibrada, con tiempos de entrega y beneficios bastante uniformes en todos sus servicios.

Fase 2

Introducción

En esta etapa, llevaremos a cabo análisis geográficos utilizando diversos tipos de gráficos. Esto nos permitirá extraer información geográfica relevante de los datos, posibilitando la conexión de dichos datos con características geográficas específicas, tales como la economía o la cultura de un país determinado.

Gráfica 1 - Beneficio por país



Al analizar este gráfico, observamos que la mayoría de los países generan beneficios. Sin embargo, Chad muestra pérdidas, mientras que países como Estados Unidos se mantienen aproximadamente neutrales.

Esta visualización es crucial para identificar desequilibrios operativos o estratégicos. La empresa debería investigar varias posibles causas, como:

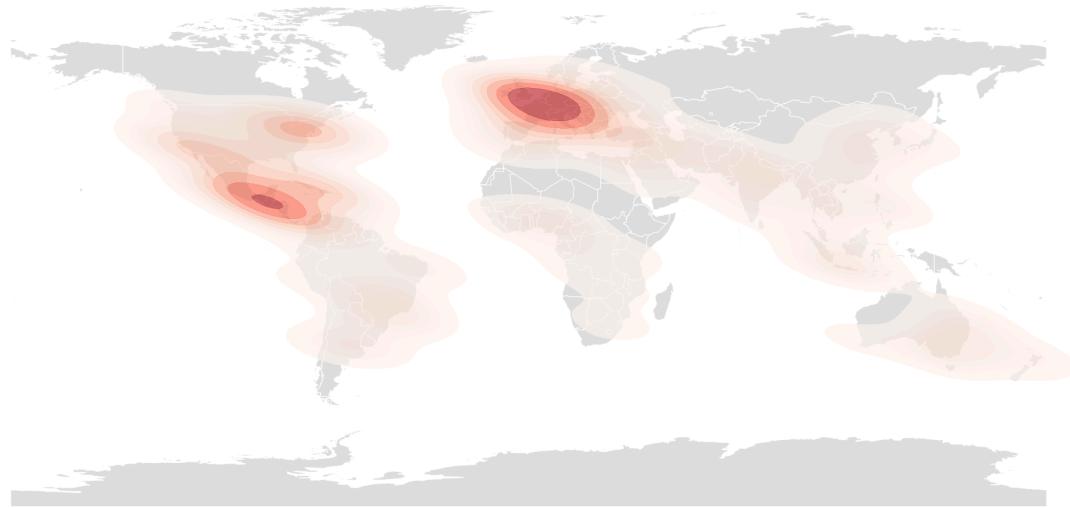
- ¿Son las políticas de descuento demasiado agresivas en los países con menor beneficio?
- ¿Los costos de beneficio en Estados Unidos son más altos, lo que reduce el margen por pedido?
- ¿Podría la competencia tener precios más elevados en los mercados de países con

beneficio nulo o negativo?

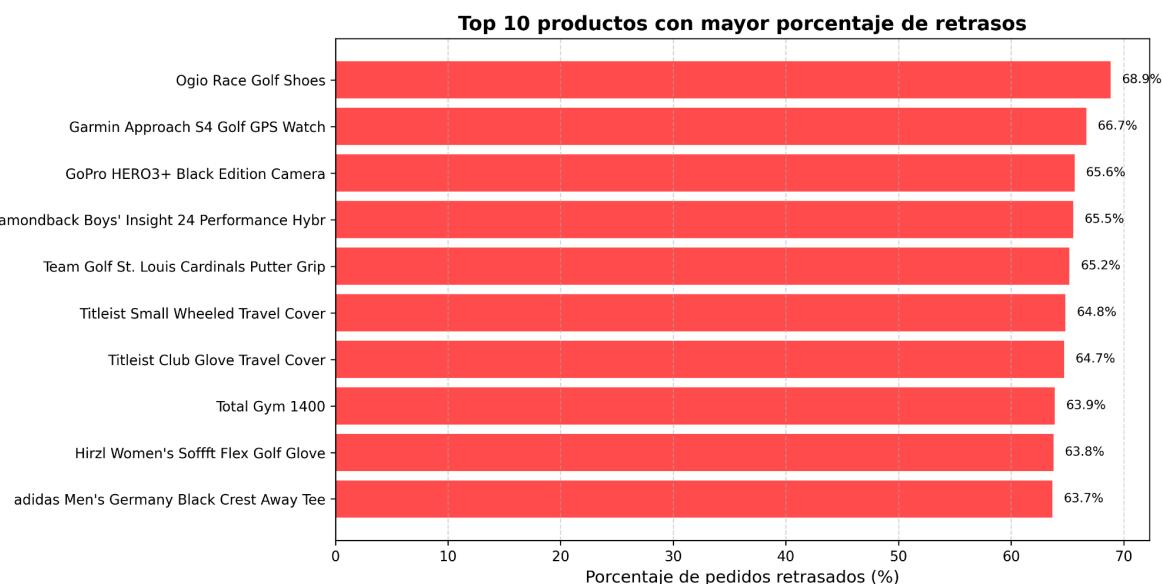
Gráfica 2 - Retrasos

Con esta gráfica, buscamos identificar las áreas geográficas con mayor número de retrasos.

Mapa de Calor: Concentración de Pedidos con Retraso



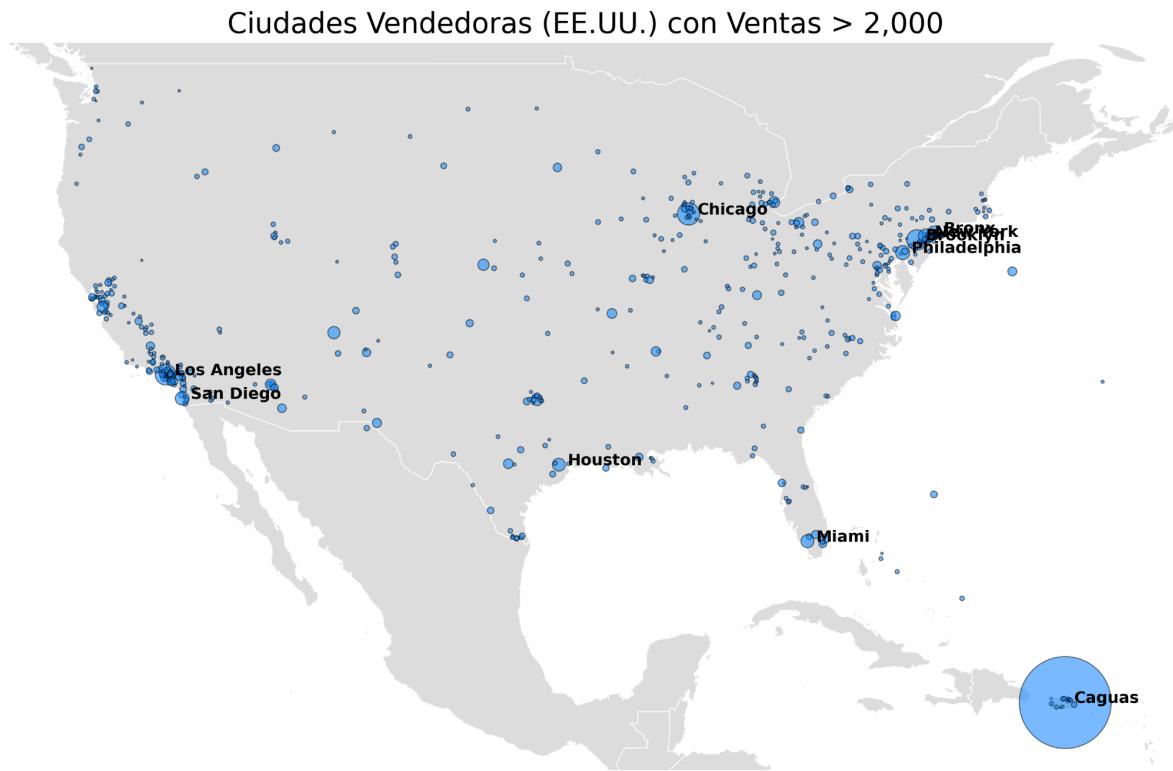
Hemos observado que los retrasos se concentran principalmente en México, Estados Unidos y Europa Central. Esta información es crucial para la empresa, ya que es fundamental investigar las causas de estos retrasos en dichas regiones.



Es probable que el porcentaje de retrasos esté influenciado por el tipo de compra y el país del usuario. Esto podría establecer una relación entre el precio del producto, el retraso y el tipo de país. Por ejemplo, si la mayoría de los retrasos ocurren en productos de mayor precio y que se adquieren más en países desarrollados, ese país podría ser identificado con problemas de envío. Asimismo, esta situación podría vincularse con las empresas de transporte/logística de los países afectados.

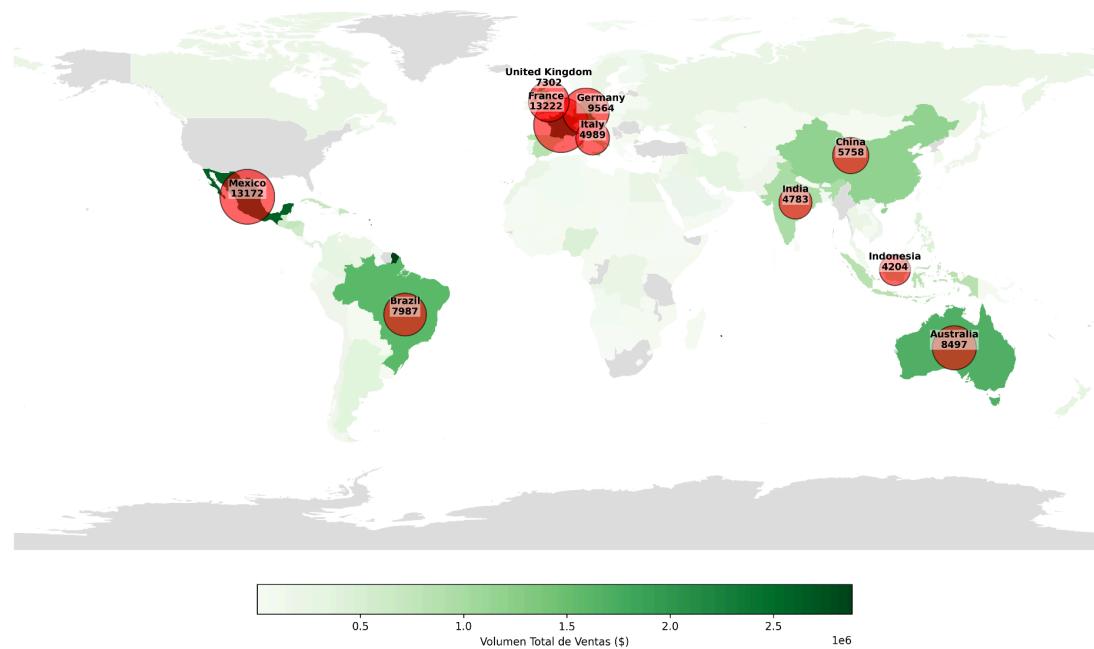
Gráfica 3 - Ciudades con vendedores destacables

Este análisis destaca las ciudades con mayores ventas, observando que los datos se limitan a Estados Unidos, con Caguas como ciudad principal. Esta información es crucial para identificar la ubicación de nuestros vendedores clave, lo que a su vez nos permite comprender mejor los productos que ofrecen a los clientes y la procedencia de los materiales que comercializan.



Si analizamos luego las ventas por país destacando los 10 países que más venden tenemos:

Ventas Totales (Color) y Top 10 Países por Nº de Pedidos (Burbujas)



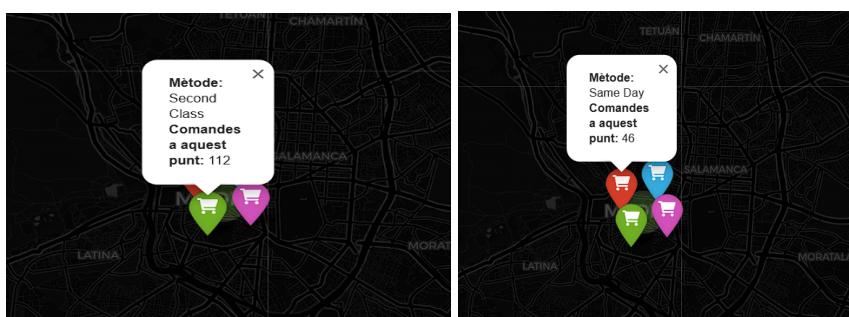
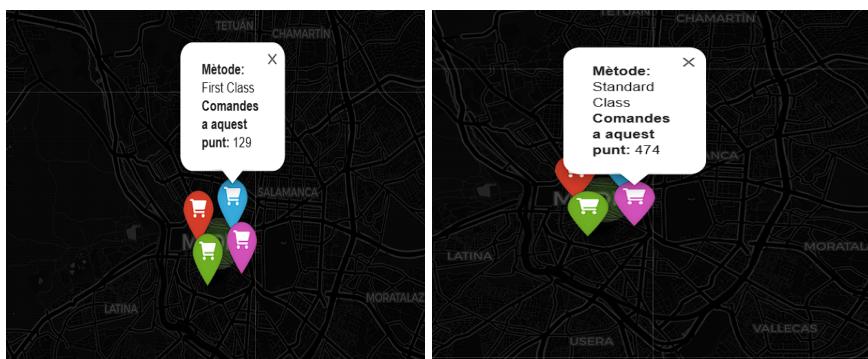
Se observa que la mayoría de los pedidos proceden de Francia, México y Alemania, entre otros, mientras que el resto de los países presenta un número de pedidos bajo o nulo. Esto permite identificar a los principales clientes (compradores) de la empresa dentro de estos datos. Sin embargo, la suma de los datos de Europa Central indica que son los clientes principales en este conjunto de datos.

Gráfica 4 - Mapa de clústeres Categórico métodos de envío por destino

¿Existen patrones geográficos en la distribución de los modos de envío (Same Day, First Class, Second Class y Standard Class)?



- A escala mundial se observa una alta concentración de pedidos en Europa y Norteamérica.
- Las burbujas naranjas/amarillas más grandes indican volúmenes de pedidos altos, coincidiendo con centros logísticos principales.
- En regiones como África o Asia oriental, la densidad es menor, mostrando menor actividad comercial o presencia logística.



Al acercarse a España (en este caso Madrid), se distinguen claramente los cuatro tipos de envío representados con íconos de carrito de colores:

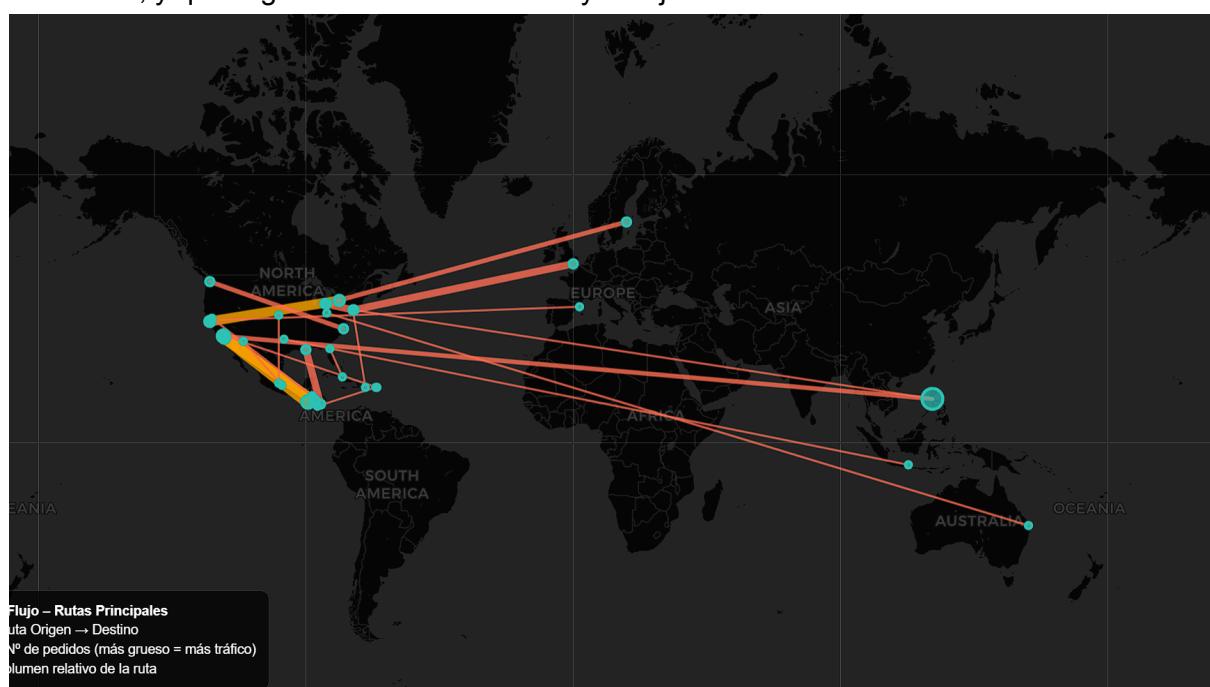
- **Same Day (rojo)** → volúmenes pequeños, pero de entrega inmediata.
- **First Class (azul)** → frecuentes, con buena cobertura urbana.
- **Second Class (verde)** → nivel medio, usado en pedidos interregionales.
- **Standard Class (morado)** → el más común en áreas periféricas y pedidos internacionales.

En conclusión, el mapa de Clústeres Categórico permite visualizar de forma geográfica lo que los análisis estadísticos ya apuntaban:

- El modo de transporte influye directamente en la eficiencia y tiempo de envío.
- Los modos rápidos se concentran en regiones desarrolladas y con alta conectividad logística.
- Los modos estándar o de segunda clase tienen mayor alcance global, pero tiempos más largos y menor margen de beneficio.
- En conjunto, el análisis reafirma la relación entre la infraestructura geográfica y la eficiencia logística, evidenciando una red global bien estructurada, aunque con margen de mejora en regiones menos conectadas.

Gráfica 5 - Mapa de flujo de rutas comerciales principales

¿Cuáles son las rutas de envío (origen → destino) más frecuentes dentro de la cadena de suministro, y qué regiones concentran el mayor flujo de comercio?

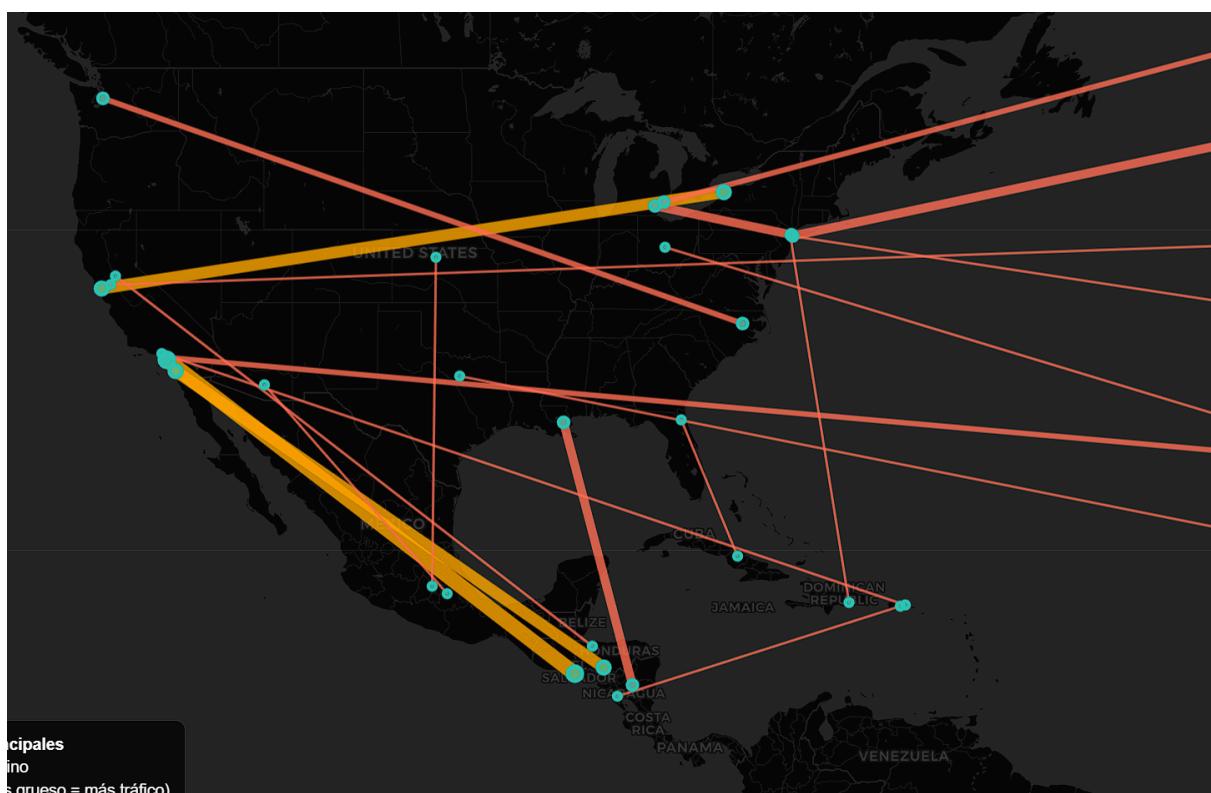


Las rutas más destacadas se concentran entre Norteamérica y Europa, con especial intensidad en los Estados Unidos y México, que aparecen como centros neurálgicos del comercio.

Varias líneas de color naranja y amarillo más gruesas indican un alto volumen de tráfico comercial entre:

- Estados Unidos ↔ México
- Estados Unidos (Long Beach sobretodo) ↔ El Salvador
- Estados Unidos (Long Beach sobretodo) ↔ Honduras
- Estados Unidos ↔ países de Europa Occidental (Reino Unido, Barcelona, Estocolmo)

También se observan conexiones relevantes hacia Asia y Oceanía, principalmente con Australia y el Sudeste Asiático, aunque con menor densidad de pedidos. A nivel regional, dentro de Norteamérica se distinguen rutas internas intensas, que conectan los principales estados de EE. UU. (California, Texas, Illinois, Nueva York) con destinos en México y Centroamérica.



En conclusión, este mapa permite ver claramente cuáles son las rutas comerciales más importantes dentro de la cadena de suministro global.

La mayoría de las rutas empiezan desde Estados Unidos, lo que demuestra que este país es el punto principal de salida de los envíos hacia otras partes del mundo.

Esto refleja su importancia en el comercio internacional y cómo concentra gran parte del movimiento de productos.

En general, el mapa muestra que Estados Unidos es el eje principal del flujo comercial, y que a partir de él se extienden las rutas que mantienen activa la red global de envíos.