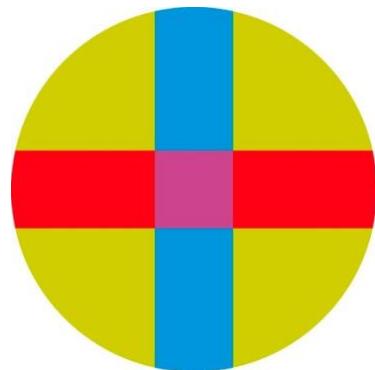


UNIVERSIDAD SAN PABLO CEU
FACULTAD DE CIENCIAS ECONÓMICAS Y
EMPRESARIALES
GRADO EN ADMINISTRACIÓN Y DIRECCIÓN DE
EMPRESAS



TRABAJO FIN DE GRADO

Título
**Análisis de Datos Económicos y
Financieros para la Toma de Decisiones
Estratégicas en el Sector Logístico.**

Tema
*Análisis de Datos Económicos, Financieros o
Empresariales.*

Autor: Ignacio Martín-Peña Chinchurreta.
Tutor: María del Carmen Escribano Ródenas.

Mayo, 2025



Resumen

En un contexto empresarial globalizado con mucha competencia, el sector logístico tiene que tratar de tomar decisiones estratégicas clave basadas en análisis preciso de datos económicos y financieros. El presente Trabajo de Fin de Grado explora como el análisis de datos puede ayudar a mejorar y confirmar las decisiones estratégicas en empresas de este sector. La investigación aborda tanto lo que es el análisis de datos como la utilidad práctica de metodologías de análisis estadístico avanzado sobre datos reales. Para ello, se empleó un conjunto de datos económico-financieros reales recolectado a través de la plataforma [Kaggle](#).

Este grupo de datos se limpia y procesa con herramientas especializadas de análisis de datos para asegurar una correcta calidad y preparación al estudio. La metodología principal se basa en un análisis estadístico avanzado de estos datos, aplicando técnicas cuantitativas como análisis multivariante, modelos predictivos y visualización de datos para identificar patrones, tendencias y relaciones significativas. A través de este planteamiento analítico, el estudio logra resultados que demuestran el impacto positivo de la analítica de datos en la toma de decisiones.

En particular, se exploran indicadores financieros clave y tendencias operativas donde el seguimiento del análisis estadístico ayuda a mejorar la planificación estratégica, optimizar la eficiencia operativa y reforzar la competitividad en el sector logístico. Los resultados conseguidos reafirman que la integración de datos económicos y financieros en el proceso de toma de decisiones estratégicas genera a las empresas logísticas una ventaja competitiva que trata información objetiva y análisis científico.

Palabras Clave

Análisis de datos, Datos económicos y financieros, Decisiones estratégicas, Sector logístico, Análisis estadístico avanzado.



Abstract

In a globalized business context with a lot of competition, the logistics sector must make key strategic decisions based on accurate analysis of economic and financial data. This thesis examines how the analysis of economic and financial data can help to improve and confirm strategic decisions in companies in the logistics sector. The research covers both the importance of data analysis and the practical usefulness of advanced statistical analysis methodologies on real data. For this purpose, a real economic-financial Dataset collected through the Kaggle platform was used.

This Dataset is cleaned and processed with specialized data analysis tools to ensure a correct quality and conditioning for the study. The main methodology is based on advanced statistical analysis of this data, applying quantitative techniques such as multivariate analysis, predictive modelling and data visualization to identify significant patterns, trends and relationships. Through this analytical approach, the study achieves results that demonstrate the positive impact of data analytics on decision-making.

Key financial indicators and operational trends are explored where monitoring statistical analysis helps to improve strategic planning, optimize operational efficiency and strengthen competitiveness in the logistics sector. The results achieved corroborate that the integration of economic and financial data into the strategic decision-making process gives logistics companies a competitive advantage that deals with objective information and scientific analysis.

Keywords

Data analysis, Economic and financial data, Strategic decision-making, Logistics sector, Advanced statistical analysis.



Índice de contenidos

Capítulo 1 Introducción	1
1.1 Contexto y Motivación del Estudio.....	1
1.2 Objetivos del TFG y Planteamiento de Hipótesis	1
1.3 Metodología General y Alcance del Trabajo	2
1.4 Estructura de la Memoria	2
Capítulo 2 El Sector Logístico	3
2.1 Importancia y Tendencias Actuales	3
2.2 Decisiones Estratégicas en Logística.....	5
2.3 Técnicas avanzadas de análisis estadístico aplicadas al sector logístico.....	6
Capítulo 3 Análisis de Datos Económicos y Financieros y Resultados Estratégicos en Logística	9
3.1 Introducción al Análisis y Descripción del Conjunto de Datos	9
3.2 Preparación y Limpieza de los Datos	11
3.3 Metodología Analítica y Herramientas Empleadas	14
3.4 Resultados del Análisis Descriptivo Estratégico	15
3.5 Resultados del Análisis Inferencial y Pruebas de Hipótesis.....	22
3.6 Análisis de Correlación, Regresión y Series Temporales	26
3.7 Discusión Integrada de los Resultados y Aplicación Estratégica	37
Capítulo 4 Conclusiones y Recomendaciones	39
4.1 Conclusiones generales	39
4.2 Recomendaciones estratégicas	40
4.3 Limitaciones y líneas futuras de investigación	40
Bibliografía	41
Capítulo 5 Anexos.....	43
5.1 Diccionario de Variables	43
5.2 Datos Limpios.....	43
5.3 Análisis Descriptivo	45
5.4 Correlaciones	50
5.5 Regresiones.....	51

5.6 Anova	52
5.7 Series Temporales.....	53
5.8 Repositorio en GitHub	55

Índice de ilustraciones

Ilustración 1: Distribución de los días reales de envío.....	16
Ilustración 2: Distribución del número de productos por pedido.....	17
Ilustración 3: Distribución de pedidos por modo de envío.....	18
Ilustración 4: Ventas totales por país.....	18
Ilustración 5: Beneficio total por segmento de clientes.....	19
Ilustración 6: Distribución del tiempo real de envío según el modo de envío.	20
Ilustración 7: Relación entre ventas y beneficios por pedido.	20
Ilustración 8: Distribución de pedidos según su estado de entrega.....	21
Ilustración 9: Análisis del segmento del cliente por modo de envío.	23
Ilustración 10: Análisis del modo de envío y días reales de entrega.	24
Ilustración 11: Análisis de estado del pedido.....	25
Ilustración 12: Relación entre descuento aplicado y valor de ventas por pedido.	27
Ilustración 13: Coeficientes de correlación (Pearson y Spearman) para la relación Descuento vs Ventas.....	28
Ilustración 14: Dispersión de Ventas vs Beneficio.....	28
Ilustración 15: Coeficientes de correlación (Pearson y Spearman) para la relación Ventas vs Beneficio.....	29
Ilustración 16: Resultado de la regresión lineal de beneficio por pedido frente a las ventas. 30	30
Ilustración 17: Resultado de la regresión múltiple de beneficio por pedido en función del ratio de beneficio, descuento y cantidad.	31
Ilustración 18: Resultado de la regresión múltiple de ventas en función de la cantidad, días reales de envío y ratio de beneficio.	32
Ilustración 19: Tendencia mensual de beneficio, ventas y volumen de pedidos.	34

Ilustración 20: Evolución mensual del número de pedidos y beneficio medio por pedido.....	35
Ilustración 21: Predicción del beneficio mensual frente al beneficio real (2015-2018).	36
Ilustración 22: Clasificación general de variables operativas y financieras.	43
Ilustración 23: Segmentadores y filtros de tablas dinámicas del análisis descriptivo.	49
Ilustración 24: Distribución de pedidos según su estado de entrega.	50
Ilustración 25: Objetivo, selección y justificación de variables para el análisis de correlaciones.	
.....	50
Ilustración 26: Criterios de interpretación de coeficientes de correlación.	51
Ilustración 27: Pasos sobre las pruebas de normalidad de las variables seleccionadas.....	51
Ilustración 28: Objetivo y diseño de los modelos de regresión.	51
Ilustración 29: Interpretación, resultados y conclusión del análisis de regresión.	52
Ilustración 30: Planteamiento e hipótesis del análisis ANOVA.	52
Ilustración 31: Objetivos y variables del análisis de series temporales.	53
Ilustración 32: Evolución temporal de ventas y beneficio por pedido.	54
Ilustración 33: Conclusión estratégica del análisis temporal del beneficio, ventas y riesgo... ..	55



Índice de tablas

Tabla 1: Variables Operativas y Financieras.....	10
Tabla 2: Resumen de hipótesis ANOVA aplicadas al beneficio medio por pedido.	22
Tabla 3: Especificación de variables X e Y usadas en el análisis ANOVA.....	22
Tabla 4: Matriz de correlaciones entre las variables cuantitativas clave.....	26
Tabla 5: Estadísticos de distribución y diagnóstico de normalidad por variable.....	27
Tabla 6: Variables de los modelos de regresión propuestos y razón de uso.	29
Tabla 7: Métricas clave seleccionadas para el análisis de series temporales.	33
Tabla 8: Ventas mensuales por año (2015-2018) y variaciones estacionales.....	35
Tabla 9: Datos limpios de la columna Type a Customer First Name.....	43
Tabla 10: Datos limpios de la columna Customer ID a Order Country.	44
Tabla 11: Datos limpios de la columna Order Date a Order Region.	44
Tabla 12: Datos limpios de la columna Order State a Shipping Mode.....	44
Tabla 13: Datos limpios de la columna Customer Segment + Shipping Mode a Quarter.....	45
Tabla 14: KPI's generales y estadísticas descriptivas.	45
Tabla 15: Ventas, beneficios y tiempo medio de envío por región.	46
Tabla 16: Ventas, beneficio y tiempo medio según mercado, modo de envío y segmento de cliente.	47
Tabla 17: Estado de los pedidos y riesgo de entrega tardía por región.....	48
Tabla 18: Tiempo medio de envío por estado del pedido.	49
Tabla 19: Utilidad de los segmentadores aplicados en el análisis descriptivo.	49
Tabla 20: Indicadores mensuales de ventas, beneficio y logística.....	53
Tabla 21: Indicadores trimestrales de ventas, beneficio y riesgo.....	54
Tabla 22: Evolución del beneficio, pedidos y riesgo por tipo de envío.....	54



Capítulo 1

Introducción

1.1 Contexto y Motivación del Estudio

La logística ha dejado de ser un simple vínculo operativo para convertirse en un factor clave de competitividad. En el comercio electrónico, un día de retraso o un euro extra de coste se perciben de inmediato en el margen y en la satisfacción del cliente. Aun así, muchas empresas siguen basando sus decisiones en la intuición. Este trabajo muestra, con el conjunto de datos *DataCo Smart Supply Chain* (180 520 registros, 2016-2018), cómo el análisis estadístico puede traducir esos datos en decisiones estratégicas más rentables.

1.2 Objetivos del TFG y Planteamiento de Hipótesis

Objetivo general: demostrar que el análisis construido busca mejorar el beneficio y a reducir riesgos en una empresa logística.

Hipótesis del proyecto:

- H1: cuando el descuento de un pedido supera el 10% del precio, las ventas unitarias aumentan, pero el margen de beneficio por unidad disminuye.
- H2: cada día extra de tránsito en el envío aumenta la probabilidad de retraso en la entrega.
- H3: entre los modos de envío estándar de la compañía, el servicio “*Same Day*” genera un margen de beneficio por debajo al de “*Standard Class*” cuando el cliente no abona ningún recargo por rapidez.
- H4: el número de pedidos presenta un pico estadísticamente significativo en los meses de noviembre y diciembre, en comparación con el resto del año.



1.3 Metodología General y Alcance del Trabajo

Se limpian y normalizan los datos en Excel. Se aplica estadística descriptiva, correlaciones, ANOVA, regresión múltiple y series temporales. El estudio se centra en la información interna de *DataCo* y no incluye variables externas (PIB, tipo de cambio, etc...).

1.4 Estructura de la Memoria

Este Trabajo se ha organizado como un recorrido coherente que lleva al lector desde la motivación inicial hasta las recomendaciones finales:

- **Capítulo 1 – Introducción:** plantea el problema, justifica su interés y formula las hipótesis que se discutirán después. Incluye además la delimitación metodológica y las preguntas de investigación.
- **Capítulo 2 – El Sector Logístico:** presenta las particularidades del sector logístico, repasa las tendencias tecnológicas (digitalización, sostenibilidad) y resume la literatura académica sobre análisis de datos aplicados a la cadena de suministro. Este marco servirá de espejo para interpretar los resultados.
- **Capítulo 3 – Análisis Estadístico, Económico y Resultados:** reúne en un mismo bloque la preparación del *Dataset*, el análisis descriptivo, las pruebas inferenciales (correlación, ANOVA, regresión) y las series temporales. Cada subapartado concluye con hallazgos parciales que se irán conectando al final del capítulo.
- **Capítulo 4 – Conclusiones y Recomendaciones:** verifica una por una las hipótesis iniciales, resume las lecciones principales y propone líneas de mejora y de investigación futura. También se detallan las limitaciones del estudio.
- **Bibliografía y Anexos:** la memoria se cierra con la lista completa de referencias (formato APA) y con anexos técnicos: diccionario de datos, capturas de pantalla de los análisis en Excel, tablas auxiliares y gráficos complementarios que respaldan la discusión del texto principal.



Capítulo 2

El Sector Logístico

2.1 Importancia y Tendencias Actuales

El sector logístico engloba el conjunto de actividades necesarias para gestionar eficazmente el flujo de productos desde el origen hasta el consumidor final. Este proceso incluye:

- Transporte.
- Almacenamiento.
- Gestión de Inventarios.
- Distribución final.

La importancia estratégica de la logística actualmente ha crecido bastante debido al crecimiento del comercio electrónico, la globalización de los mercados y el aumento en las expectativas del consumidor respecto a rapidez, precisión y confiabilidad en las entregas (Ballou, 2016). Por otro lado, la logística no se limita a una función únicamente operativa, es un factor clave para la competitividad empresarial y la satisfacción del cliente final. Según Chopra y Meindl (2019), una logística eficiente permite:

- Reducción significativa de costes operativos.
- Disminución de tiempos en el momento de pedido-entrega.
- Mayor agilidad para adaptarse a cambios inesperados en la demanda.

También, fallos en la gestión logística generan importantes consecuencias negativas como exceso de costes operativos, incumplimiento de plazos y disminución de la satisfacción del cliente, produciendo pérdida de competitividad (Harrison, 2014).

Las principales tendencias actuales en el sector logístico giran en torno a tres grandes ejes:



1. Digitalización:

La transformación digital en logística implica la incorporación de tecnologías avanzadas, tales como sistemas integrados de gestión (ERP), sensores *IoT* y análisis masivos de datos (*Big Data*) que permiten una visibilidad total y en tiempo real de la cadena de suministro (Waller, 2013).

- Permite optimizar rutas, gestionar eficientemente inventarios y mejorar la utilización de activos mediante información actualizada constantemente.
- Facilita modelos predictivos mediante Inteligencia Artificial (IA) y *Machine Learning*, anticipando la demanda y detectando ineficiencias ocultas (Kache, 2017).

2. Automatización:

La automatización muestra la incorporación de nuevas soluciones tecnológicas que reemplazan procesos anteriormente manuales, como:

- Robots en almacenes para *picking* automático.
- Vehículos autónomos y drones para entregas rápidas y precisas.
- Sistemas automáticos de clasificación que incrementan notablemente la velocidad operativa y reducen errores humanos (Winkelhaus, 2020).

3. Sostenibilidad:

La creciente preocupación medioambiental ha convertido a la sostenibilidad en un objetivo fundamental para las empresas logísticas:

- Desarrollo de rutas optimizadas que reducen emisiones de Co2 (*Green Logistics*).
- Innovación en eficiencia de combustibles y vehículos eléctricos.
- Fortalecimiento de envíos para reducir la huella de carbono global en la cadena de suministro (Grant, 2017).

El sector logístico contemporáneo es un entorno dinámico, donde la gestión eficiente de la información y la capacidad de rápida adaptación son fundamentales para mantener la competitividad. La disponibilidad de datos económicos y operativos fiables y su riguroso



análisis surgen como factores cruciales para la toma de decisiones estratégicas en este contexto.

2.2 Decisiones Estratégicas en Logística

La toma de decisiones logísticas ocurre a distintos niveles dentro de la organización, cada uno con objetivos específicos y perspectivas temporales diferentes:

1. Nivel Estratégico (Largo Plazo): incluye decisiones clave como la ubicación de centros de distribución, selección de socios logísticos o inversiones en infraestructura tecnológica. Estas decisiones ayudan a mejorar la cadena de suministro alineándola con metas corporativas como crecimiento o sostenibilidad (Ballou, 2016).
2. Nivel Táctico (Medio Plazo): se enfoca en optimizar recursos disponibles, gestionando inventarios, planificando rutas o definiendo políticas de restauración. La frecuencia de actualización suele ser mensual o trimestral, buscando eficiencia operativa sin caer en faltas de stock (Chopra & Meindl, 2019).
3. Nivel Operativo (Corto Plazo): decisiones diarias relacionadas con asignación inmediata de recursos, programación de entregas o resolución de problemas puntuales para cumplir plazos y mantener niveles de servicio (Christopher, 2016).

El impacto del análisis de datos ha revolucionado la manera en cómo estas decisiones se toman. Tradicionalmente basadas en experiencia y datos históricos, ahora se apoyan en información precisa obtenida en tiempo real gracias al Big Data y tecnologías avanzadas:

- Compañías como UPS aplican algoritmos para optimizar rutas, logrando ahorros significativos en costes operativos y combustible (UPS, 2021).
- Amazon usa sistemas predictivos de inteligencia artificial para anticipar la demanda, consiguiendo enormes ahorros logísticos y reducción en emisiones de Co2 (Sifted, 2021).

Hay beneficios concretos del análisis avanzado en logística identificados por diversos estudios que incluyen:



- Reducción entre el 10% y 15% en costes operativos y disminución de inventarios entre 20% y 50% (McKinsey, 2020).
- Las empresas que toman decisiones basadas en datos son significativamente más exitosas en alcanzar objetivos financieros y operativos (Consulting, 2021).

Principales intereses que refuerzan la necesidad de las decisiones analíticas en logística:

- Big Data: recopilación masiva de datos operativos permite anticiparse a la demanda y optimizar inventarios, reduciendo errores y costes operativos (Deloitte, 2023).
- IA: algoritmos de aprendizaje automático optimizan rutas y mejoran la gestión de inventarios, transformando la logística en un proceso dinámico y eficiente (Gartner, 2022).

Cabe destacar cómo las decisiones estratégicas basadas en datos fortalecen la competitividad y resistencia de las empresas logísticas modernas. El análisis económico-financiero se ha convertido así en un factor muy importante para lograr ventajas operativas sostenibles.

2.3 Técnicas avanzadas de análisis estadístico aplicadas al sector logístico

La aplicación de técnicas avanzadas de análisis estadístico es crucial para transformar grandes cantidades de datos logísticos en conocimiento estratégico accionable (Martín Pliego & Ruiz Maya, 2011). Estas metodologías permiten descubrir patrones ocultos, establecer relaciones causa y efecto, predecir comportamientos futuros y optimizar decisiones operativas y estratégicas en la gestión logística. A continuación, se presentan y explican brevemente las principales técnicas estadísticas avanzadas que se utilizarán en el presente estudio:

- Análisis de Correlación:
 - Identifica relaciones significativas entre variables, revelando qué factores logísticos están relacionados.
 - Incluyen analizar la relación entre tiempos de entrega y distancias o entre volumen de pedidos y costos operativos.



- Es clave para establecer hipótesis preliminares sobre la interdependencia de variables antes de profundizar en modelos predictivos más avanzados.

- **Modelos de Regresión Lineal:**
 - Permiten predecir y cuantificar cómo una o varias variables independientes afectan una variable dependiente específica.
 - Aplicaciones prácticas incluyen predecir costos logísticos basados en volumen de mercancías, distancia o costes de combustible, anticipando tiempos de entrega basados en diferentes condiciones operativas.
 - Facilita decisiones informadas y estratégicas al cuantificar claramente el impacto de cambios específicos en variables críticas.

- **Análisis de Varianza (ANOVA):**
 - Evalúa si existen diferencias estadísticamente significativas entre múltiples grupos o categorías.
 - Aplicable en logística para comparar, por ejemplo, tiempos medios de entrega según región geográfica, modo de transporte o tipos de producto.
 - Ayuda a identificar factores críticos que influyen en la eficiencia operacional y proporcionar información valiosa para tomar medidas ajustadas.

- **Análisis de Series Temporales:**
 - Estudia patrones históricos para prever comportamientos futuros de variables logísticas clave, como la demanda de productos o tiempos de entrega.
 - Permite identificar estacionalidades, tendencias y ciclos, fundamentales para planificar adecuadamente recursos, inventarios y capacidad operativa.
 - Favorece una planificación estratégica proactiva, reduciendo incertidumbres operativas y financieras mediante proyecciones precisas basadas en datos históricos.



- **Visualización Avanzada de Datos:**

- Facilita la comunicación efectiva de hallazgos complejos mediante representaciones gráficas claras y visualmente atractivas.
- Uso estratégico de gráficos en Excel (histogramas, diagramas de dispersión, gráficos dinámicos) para interpretar resultados estadísticos de manera sencilla y directa.
- Potencia la rápida comprensión por parte de los decisores logísticos y mejora la capacidad para detectar patrones críticos, desviaciones significativas y áreas que requieren atención inmediata.

Para asegurar la validez y utilidad práctica de estas técnicas, se realizará una rigurosa validación estadística de resultados, verificando supuestos técnicos esenciales (como significancia estadística, normalidad de residuos y homogeneidad de varianzas). Además, se emplearán enfoques complementarios como la validación cruzada para asegurar que los modelos desarrollados tengan un sólido desempeño predictivo.

La combinación de estas técnicas avanzadas con herramientas analíticas prácticas, como Excel y cuadros de mando interactivos, ofrecen una perspectiva completa y robusta que apoya la toma de decisiones basada en evidencia empírica sólida, llevando a las empresas del sector logístico desde un enfoque reactivo hacia uno predictivo y estratégico.



Capítulo 3

Análisis de Datos Económicos y Financieros y Resultados Estratégicos en Logística

3.1 Introducción al Análisis y Descripción del Conjunto de Datos

Para este estudio se ha empleado un conjunto de datos público denominado *DataCo Smart Supply Chain*, obtenido de la plataforma Kaggle (Constante, 2020). Este conjunto, originalmente publicado en *Mendeley Data*, simula operaciones logísticas reales de una empresa global ficticia llamada *DataCo Global*.

El *Dataset* comprende aproximadamente 180.520 registros individuales, correspondientes a pedidos realizados entre los años 2016 y 2018. Cada registro representa un artículo específico dentro de un pedido, permitiendo analizar detalladamente el comportamiento logístico y financiero a nivel de producto y pedido.

Las características principales del *Dataset* incluyen:

- Cobertura Temporal: datos mensuales durante aproximadamente tres años.
- Alcance Geográfico: información proveniente de múltiples países y regiones.
- Diversidad de Productos: electrónica, artículos deportivos, moda, entre otros.
- Modos de Envío: estándar, exprés, primera clase y mismo día.



Las variables del *Dataset* se dividen principalmente en dos categorías:

Variable	Descripción	Tipo
Days for shipping (REAL)	Días reales para el envío	Operativa
Days for shipment (scheduled)	Días programados para el envío	Operativa
Delivery Status	Estado de la entrega del pedido	Operativa
Late Delivery Risk	Riesgo de entrega tardía (1=tarde, 0=a tiempo)	Operativa
Order Date	Fecha en la que se realiza el pedido	Operativa
Shipping Date	Fecha exacta en que se envía el pedido	Operativa
Shipping Mode	Tipo o clase de envío utilizado	Operativa
Order Status	Estado actual de la orden	Operativa
Customer Segment	Tipo de cliente que realiza el pedido	Operativa
Market	Mercado donde se entrega la orden	Operativa
Order Region	Región del mundo donde se entrega la orden	Operativa
Order Country	País destino del pedido	Operativa
Order City	Ciudad destino del pedido	Operativa
Product Status	Disponibilidad del producto (1=No disponible, 0=Disponible)	Operativa
Benefit per order	Beneficio obtenido por pedido	Financiera
Sales per customer	Ventas totales realizadas por cliente	Financiera
Order Item Product Price	Precio del producto sin descuento	Financiera
Order Item Discount	Importe del descuento aplicado	Financiera
Order Item Discount Rate	Porcentaje del descuento aplicado	Financiera
Order Item Total	Importe total del pedido	Financiera
Sales	Valor total de ventas realizadas	Financiera
Order Profit Per Order	Beneficio neto obtenido en cada pedido	Financiera
Order Item Profit Ratio	Ratio de rentabilidad por pedido	Financiera
Order Item Quantity	Cantidad de productos por pedido	Financiera

Tabla 1: Variables Operativas y Financieras.

Estas variables permiten establecer análisis cruzados entre eficiencia operativa y resultados financieros, ayudando a identificar relaciones estratégicas como:

- Impacto económico de retrasos en las entregas.
- Efectos de políticas comerciales sobre los ingresos.
- Relación entre volumen de venta y modalidad de envío.

La riqueza y amplitud del *Dataset DataCo Smart Supply Chain* posibilita realizar análisis profundos sobre aspectos clave de la gestión logística y financiera, facilitando conclusiones significativas para la toma de decisiones estratégicas.



3.2 Preparación y Limpieza de los Datos

Antes de empezar a realizar el análisis estadístico, se realizó un proceso de depuración y preparación del conjunto de datos para asegurar su calidad, integridad y fiabilidad (Hair, Black, Babin & Anderson, 2019). Este paso preliminar es fundamental para eliminar errores, anomalías e inconsistencias, de modo que los resultados de los análisis descriptivo, correlacional, regresivo y de series temporales sean válidos y representativos. A continuación, se detallan las principales tareas de limpieza llevadas a cabo:

1º Paso: Eliminación de columnas irrelevantes o redundantes.

En primer lugar, tras identificar las variables operativas y financieras de interés para el estudio, se construyó una tabla final en una hoja nueva denominada “Datos Limpios” eliminando aquellos campos que no aportaban valor o duplicaban información. Las columnas eliminadas y la justificación de su eliminación fueron las siguientes:

- *Customer Email* y *Customer Password*: aparecían ocultos (como “XXXX” debido a la protección de datos), por lo que no ofrecían información útil para el análisis es por eso por lo que se eliminó por completo esta información que no aporta valor al análisis.
- *Order Customer ID*: campo redundante, ya que el *Customer ID* (ID de cliente) proporcionaba la misma referencia de cliente. Se optó por conservar solo uno de estos identificadores.
- *Product Image*: contenía únicamente una URL de la imagen del producto. Al no ser relevante para un análisis, se descartó.
- *Product Description*: incluía descripciones textuales de los productos en algunos registros. Dado que el enfoque del estudio es cuantitativo y estratégico, y no se analizaba texto libre de descripciones, se eliminó.
- *Customer Last Name*: muchos registros carecían del apellido del cliente (valor nulo), probablemente por datos incompletos o clientes corporativos. Al no disponerse de este dato en una proporción significativa de filas y no ser crítico para el análisis estratégico, se decidió remover esta columna.
- *Order Zipcode* y *Customer Zipcode*: no se iba a trabajar a ese nivel de detalle de código postal en el análisis geográfico, y además estas columnas presentaban numerosos



valores vacíos. Debido a su falta de completitud y relevancia, ambas se excluyeron del conjunto de datos final.

- Order Item Card Product ID (RFID): era un identificador redundante, ya que duplicaba la información del *Product Card ID*. Para evitar duplicidad, se mantuvo solo el identificador principal de producto y se eliminó la columna RFID.
- Product Status: tenía el mismo valor para todos los pedidos (un indicador *status* constante, con valor 0 en todas las filas). Al no mostrar diferencias, ni aportar información discriminatoria, se eliminó por completo.

Estas eliminaciones depuraron el *Dataset*, reduciendo la dimensionalidad sin perder contenido informativo relevante. Tras esta filtración inicial, el conjunto de datos quedó con alrededor de 43 columnas útiles (frente a las más de 50 originales), enfocadas únicamente en información necesaria para el estudio.

2º Paso: Conversión y estandarización de formatos.

Otro paso clave fue corregir formatos de datos inconsistentes para asegurar la correcta interpretación y manejo de la información:

- Conversión de valores numéricos almacenados como texto: se detectó que varios campos numéricos (por ejemplo: *Sales*, *Order Profit per Order*, *Sales per Customer*, *Order Item Profit Ratio*) estaban almacenados como texto debido al formato de origen. Esto impedía realizar cálculos y agregaciones correctamente. Usando las herramientas de Excel (como la función “Texto en columnas” y formato de celda), todos estos valores se transformaron al tipo numérico estándar. De este modo, cifras monetarias, cantidades y ratios quedaron unificados como números, evitando errores de cálculo posteriores.
- Unificación del formato de fechas: las columnas de fecha (*Order Date* y *Shipping Date*) venían en formato americano (con el mes primero y, en muchos casos, incluyendo *timestamp* de hora). Para facilitar su comprensión y coherencia con el formato local, se convirtieron todas las fechas al formato día/mes/año (dd/mm/aaaa). Por ejemplo, una fecha original “1/13/2018 12:06” (13 de enero de 2018 en formato mes/día/año)



se transformó a “13/01/2018”. Esta estandarización eliminó ambigüedades en la interpretación de fechas y permitió trabajar con ellas de forma correcta.

- Revisión de valores atípicos: se inspeccionaron los valores extremos mediante diagramas de caja y la regla de $\pm 3DE$, ya que los *outliers* pueden sesgar de forma notable las estimaciones estadísticas (Field, 2018).

3º Paso: Tratamiento de valores faltantes.

La calidad de los datos también se garantizó manejando adecuadamente los valores nulos o incompletos:

- Primero se cuantificaron las celdas vacías en cada columna utilizando funciones de Excel (como “COUNTBLANK”) para identificar campos con información incompleta.
- Las columnas con gran cantidad de valores faltantes (ya mencionadas arriba, como *Customer Last Name* y *Zipcode*) fueron eliminadas del *Dataset*, debido a que su alto nivel de incompletitud dificultaba cualquier análisis fiable.
- En el resto de las variables clave, no se encontraron gran cantidad de valores nulos, la gran mayoría de registros estaban completos. Cualquier celda vacía residual en campos numéricos u operativos fue revisada caso por caso para confirmar que no afectaba a los indicadores principales (por ejemplo, algunos pedidos cancelados podrían no tener datos de entrega, lo cual es coherente con el significado del dato). No fue necesario aplicar técnicas de imputación, ya que, tras eliminar las columnas problemáticas, los datos restantes estaban completos.

4º Paso: Verificación y manejo de duplicados.

Se verificó la existencia de registros duplicados que pudieran dificultar el análisis:

- Como cada fila del *Dataset* representa un ítem de pedido individual (producto específico dentro de una orden de compra), es normal que un mismo *Order ID* aparezca en múltiples filas (una por cada producto comprado en ese pedido). Esta repetición de identificador de orden no se considera duplicado erróneo sino estructura propia de los datos transaccionales.



- Se validó que no hubiera filas completamente duplicadas (dos registros idénticos en todos sus campos). La inspección no reveló duplicados exactos, cada entrada correspondía a un registro único. Por tanto, no fue necesario eliminar filas por duplicación accidental.

La limpieza y preparación de los datos fue un paso previo importante que dejó la información en condiciones óptimas para el estudio. Gracias a estas tareas de limpieza, el conjunto de datos final presentaba datos consistentes, coherentes y relevantes, sirviendo como base confiable para llevar a cabo los análisis descriptivos, pruebas de correlación, modelos de regresión y análisis de series temporales que sustentan las conclusiones estratégicas de este Trabajo de Fin de Grado.

3.3 Metodología Analítica y Herramientas Empleadas

El desarrollo de este proyecto se ha realizado únicamente utilizando Microsoft Excel (Winston, 2016) como herramienta exclusiva, desde la preparación inicial hasta el análisis estadístico avanzado y la visualización final. Las razones fundamentales para elegir Excel como única herramienta ha sido:

1. Versatilidad y Suficiencia: Excel ofrece una amplia gama de funciones estadísticas integradas, permitiendo realizar desde análisis descriptivos simples hasta pruebas estadísticas más avanzadas, como regresiones lineales y ANOVA.
2. Facilidad de Uso: al ser una herramienta ampliamente extendida en entornos empresariales, facilita una rápida adaptación de los análisis sin necesidad de conocimientos avanzados de programación o software especializado.
3. Entorno Unificado y Trazabilidad: todas las operaciones analíticas y transformaciones quedan registradas en el mismo libro de trabajo, lo que simplifica la auditoría de cada paso y asegura la coherencia de los resultados.

Entre las principales funcionalidades aprovechadas destacan:

- Tablas Dinámicas: para resumir grandes volúmenes de datos y facilitar análisis rápidos.
- Analysis ToolPak: complemento utilizado para ejecutar análisis estadísticos complejos, incluyendo correlaciones, regresiones lineales y ANOVA.



- Funciones Estadísticas Básicas: como promedio, mediana, desviación estándar y percentiles para un análisis descriptivo preciso.
- Visualización Gráfica: gráficos claros y profesionales que facilitan la interpretación directa de los resultados.

La elección metodológica basada exclusivamente en Excel demostró no solo ser suficiente, sino también eficaz para alcanzar los objetivos planteados. En concreto:

- Se implementaron modelos estadísticos robustos mediante el complemento *Analysis ToolPak* (Regresiones y ANOVA).
- Los gráficos de dispersión con líneas de tendencia incluyeron directamente ecuaciones de regresión y valores R^2 para facilitar la interpretación visual.
- El entorno Excel permitió mantener una estructura ordenada, transparente y reproducible.

Adicionalmente, los detalles específicos y capturas de pantalla que demuestran la implementación de estos análisis en Excel están complementariamente disponibles en los anexos.

En definitiva, este proyecto subraya que, incluso con herramientas estándar y accesibles para cualquier empresa, es posible llevar a cabo análisis profundos y obtener *insights* valiosos para la toma de decisiones estratégicas.

3.4 Resultados del Análisis Descriptivo Estratégico

El análisis descriptivo realizado sobre el conjunto de datos limpio proporcionó una visión integral de las operaciones logísticas y financieras de *DataCo* (Levine, Stephan, Szabat, & Berenson, 2020). Este análisis no solo permitió identificar patrones relevantes y tendencias operativas, sino que además generó *insights* estratégicos clave para la toma de decisiones empresariales.

i. Indicadores Generales y Visión Inicial:

- Pedidos Totales: 21.716 pedidos.
- Ventas Acumuladas: 12.095.314,95\$.



- Beneficio Total: 1.321.735,75\$.

Estos datos establecen una operación logística de gran volumen y diversidad, proporcionando una base sólida para el análisis posterior.

ii. Principales patrones Operativos y Financieros:

1. Ventas y Beneficios Medios por Pedido:

- i. Venta Promedio: 203,31\$.
- ii. Beneficio Promedio: 22,22\$.
- iii. La Mediana de Ventas: es 199,92\$ próxima a la media sugiere una distribución de ventas relativamente equilibrada, sin grandes sesgos por pedidos atípicos.

Decisión estratégica: esta estabilidad permite a DataCo proyectar ingresos por pedido con alta fiabilidad y establecer rangos de descuento sin afectar a la rentabilidad media.

Para facilitar el análisis comparativo, se han generado más tablas dinámicas con datos interesantes (ver Anexo).

2. Eficiencia en los Tiempos de Envío:

- i. El tiempo medio real de envío fue de 3,5 días, con la mayoría de las entregas completadas entre 1 y 2 días.

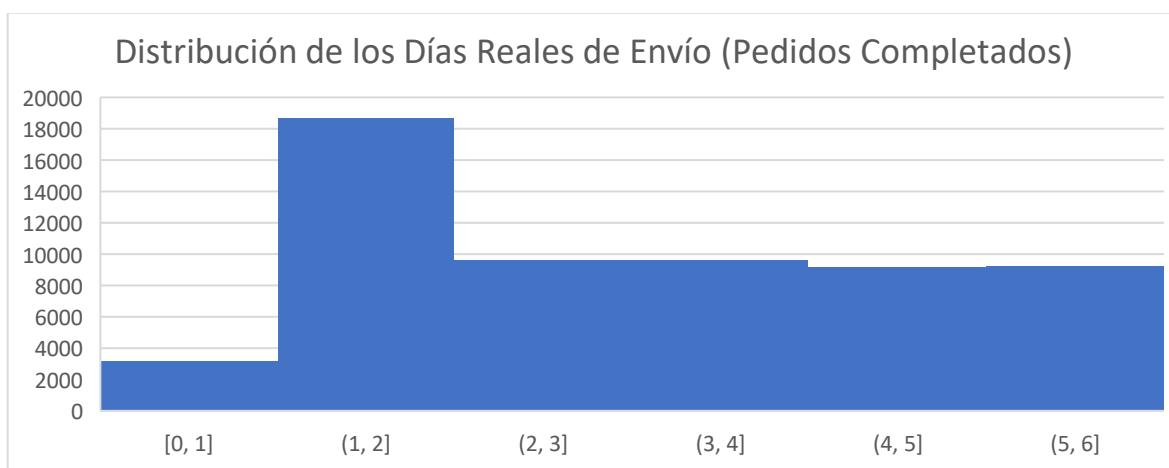


Ilustración 1: Distribución de los días reales de envío.



- ii. El modo "*Same Day*" demostró una clara superioridad en rapidez, mientras que las clases "*Standard*" y "*Second Class*" mostraron mayor variabilidad y retrasos.

Decisión estratégica: potenciar las operaciones bajo el servicio "*Same Day*" puede reforzar la percepción de fiabilidad y satisfacción del cliente, especialmente en segmentos dispuestos a pagar más por entregas rápidas.

3. Tamaño de Pedido:

- i. La mayoría de los clientes solicitan 1 producto por pedido, confirmando que la mayoría de los pedidos son pequeños.

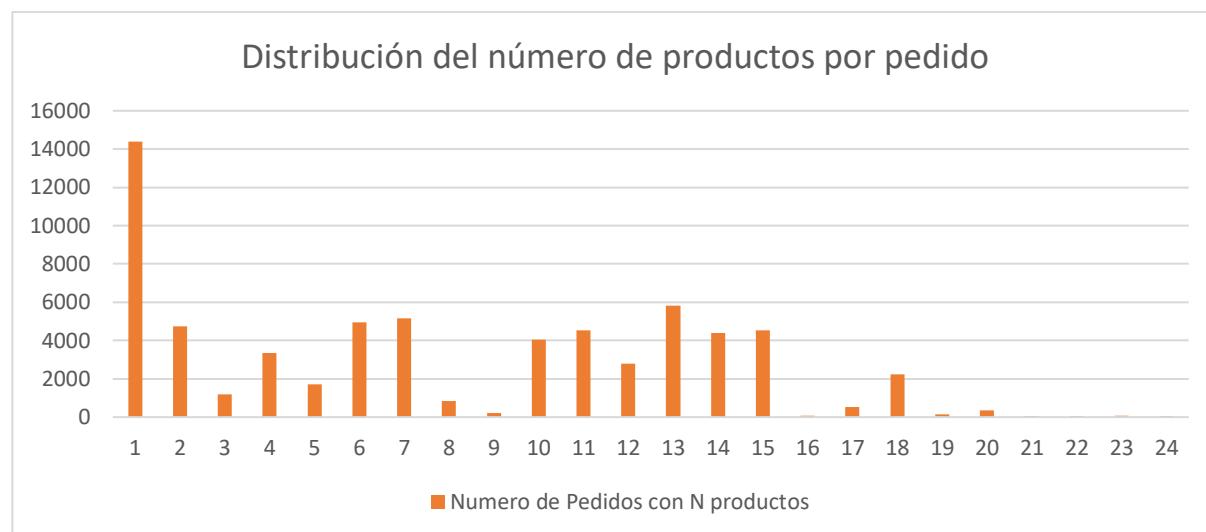


Ilustración 2: Distribución del número de productos por pedido.

Decisión estratégica: esta tendencia sugiere que la empresa debe optimizar procesos de *picking* y embalaje para pedidos de bajo volumen, donde la eficiencia operativa tiene mayor impacto en la rentabilidad.

4. Modo de Envío Utilizado:

- i. La modalidad de envío más utilizada es claramente la "*Standard Class*", que domina con más de 107.000 pedidos.
- ii. La opción "*Second Class*" sigue con más de 35.000 pedidos, mientras que "*Same Day*" es la menos utilizada con menos de 10.000 envíos.

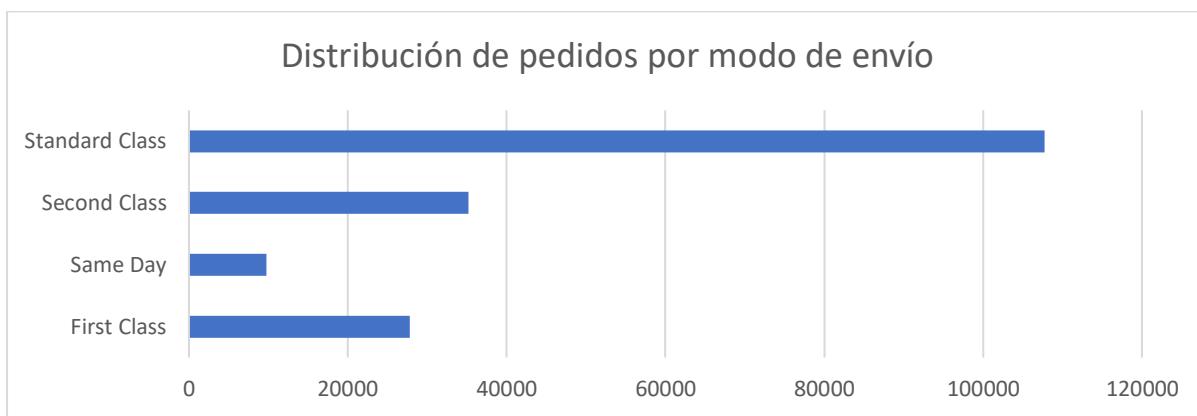


Ilustración 3: Distribución de pedidos por modo de envío.

5. Distribución Geográfica de Ventas:

- Estados Unidos lideró en volumen de ventas, seguido de Mongolia y algunos países europeos.

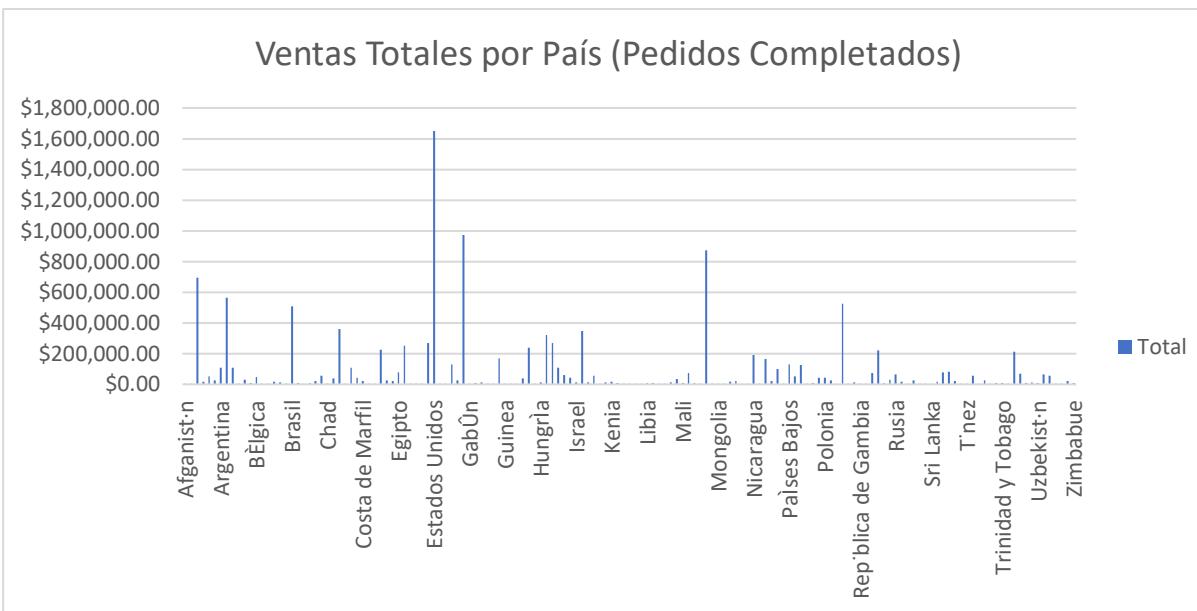


Ilustración 4: Ventas totales por país.

Decisión estratégica: priorizar recursos logísticos y estrategias de marketing en las regiones con mayor volumen de ventas ayudara a consolidar el liderazgo y explorar oportunidades de expansión geográfica.



6. Rentabilidad por Segmento de Clientes:

- i. El segmento *Consumer* generó el mayor beneficio total, seguido de *Corporate*.
- ii. *Home Office* es el segmento que menos contribuye al beneficio total.

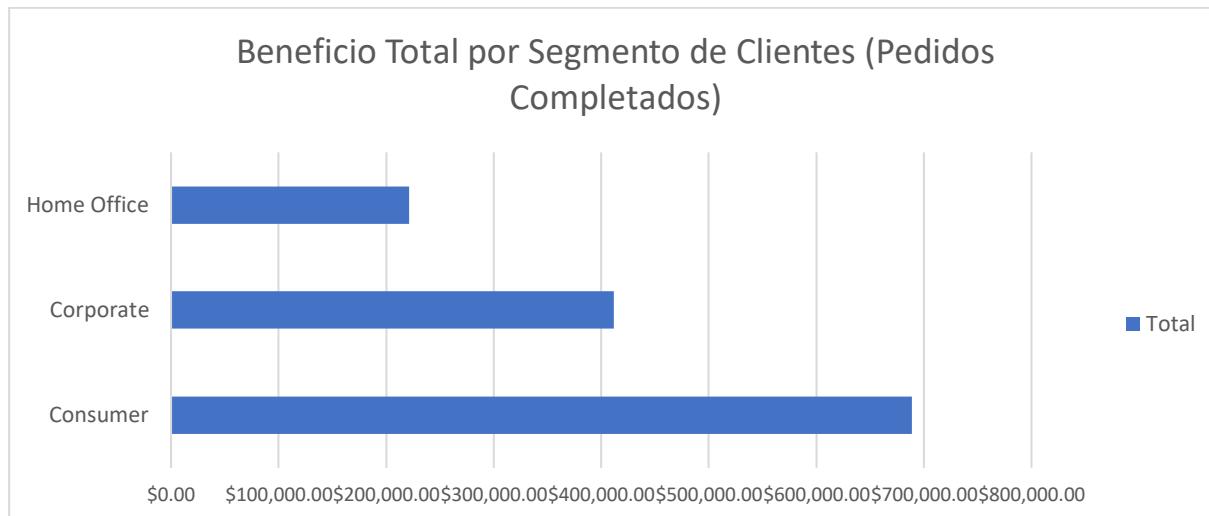


Ilustración 5: Beneficio total por segmento de clientes.

Decisión estratégica: se recomienda crear campañas y estrategias de fidelización en los segmentos *Consumer* y *Corporate*, donde la empresa ha demostrado mayor rentabilidad histórica.

7. Tiempo Real de Envío por Modalidad:

- i. El *Boxplot* muestra claramente que el método "*Same Day*" tiene tiempos de entrega significativamente más bajos, mientras que "*Standard Class*" y "*Second Class*" presentan mayor variabilidad y tiempos más prolongados.

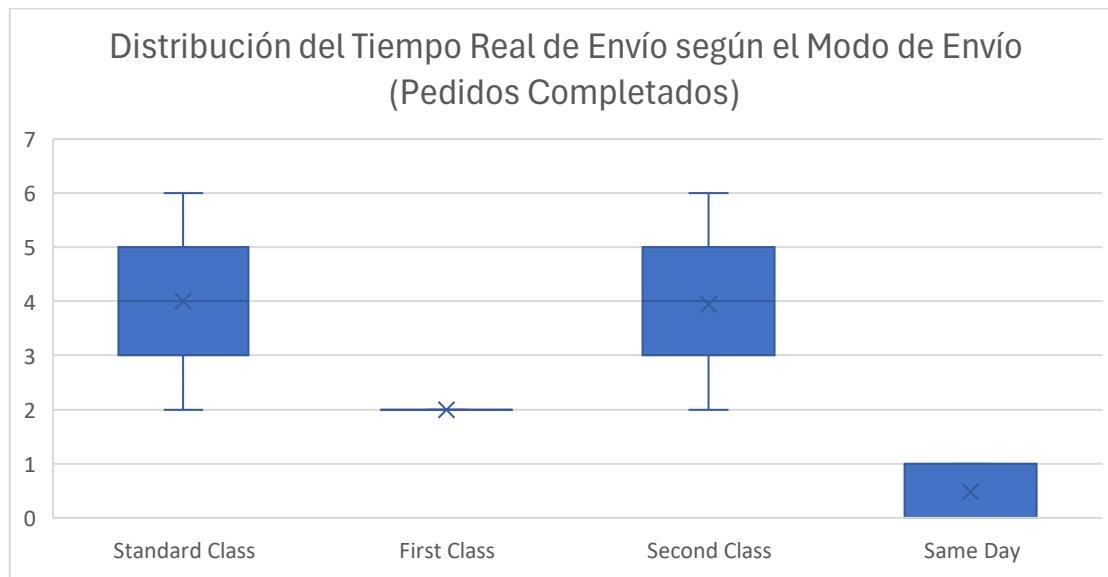


Ilustración 6: Distribución del tiempo real de envío según el modo de envío.

8. Relación entre Ventas y Beneficio por Pedido:

Aunque la mayoría de los pedidos reportaron beneficios positivos, algunos presentaron pérdidas, sobretodo aquellos con descuentos excesivos o altos costes logísticos.

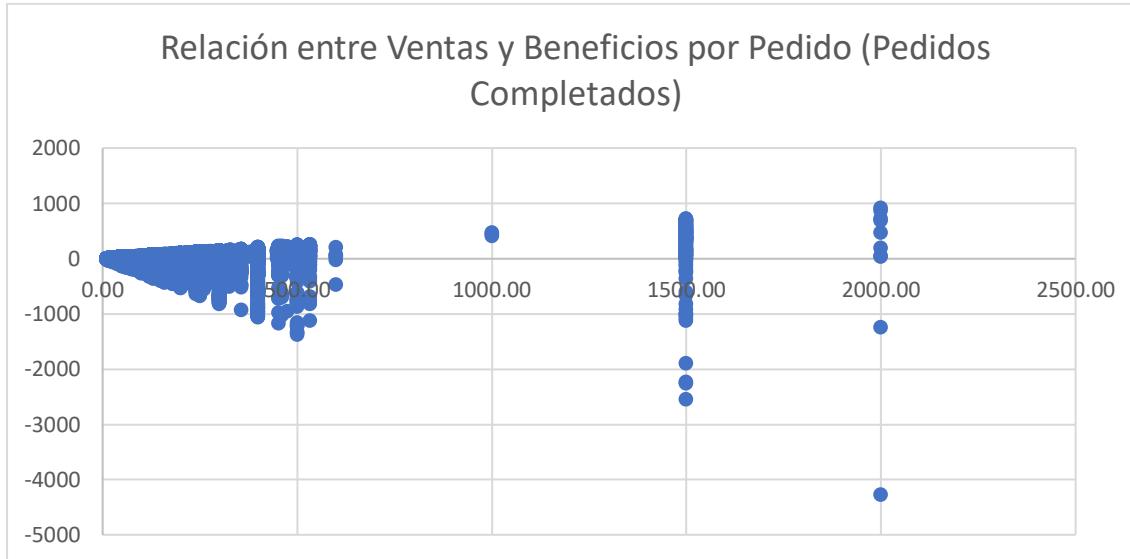


Ilustración 7: Relación entre ventas y beneficios por pedido.

Decisión estratégica: es crucial revisar las políticas de descuentos y costes de envío asociados a determinados pedidos o clientes, identificando patrones que puedan estar desgastando la rentabilidad.



9. Estado de Entrega de los Pedidos:

- i. La mayoría de los pedidos 59.491 están completados, siendo esta la categoría dominante.
- ii. Hay una considerable cantidad de pedidos en estados intermedios como "*Pending Payment*" y "*Processing*".

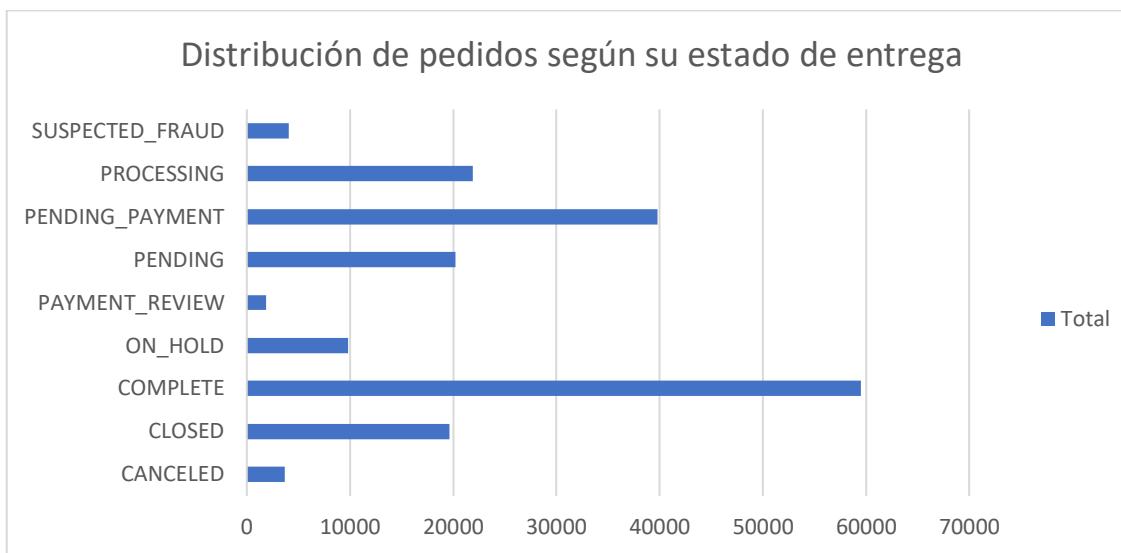


Ilustración 8: Distribución de pedidos según su estado de entrega.

Decisión estratégica: es aconsejable implementar controles más estrictos en la gestión de pagos y en el seguimiento de pedidos en proceso para reducir riesgos financieros y mejorar la eficiencia.

El análisis descriptivo generó métricas clave del desempeño operativo y financiero de *DataCo*, y también reveló patrones que pueden ser explotados para optimizar procesos y maximizar la rentabilidad. En particular, destacan la necesidad de:

- a. Incentivar las modalidades de envío más eficientes ("*Same Day*").
- b. Optimizar el manejo de pedidos pequeños.
- c. Reforzar la atención a los segmentos de clientes más rentables.
- d. Mejorar el control de pedidos en proceso y revisar políticas de descuento.

Estas acciones estratégicas, implementadas en un análisis riguroso de datos reales, ofreciendo una base objetiva para la toma de decisiones que pueden traducirse en ventajas competitivas sostenibles para la empresa.



3.5 Resultados del Análisis Inferencial y Pruebas de Hipótesis

Este apartado contrasta estadísticamente las intuiciones obtenidas en el análisis descriptivo y determina si ciertas diferencias observadas son realmente significativas o son fruto del azar. En todas las pruebas se ejecutaron en Excel (*Analysis ToolPak*), fijando un nivel de significación de $\alpha = 0,05$.

Se detalla el análisis inferencial llevado a cabo mediante pruebas estadísticas que permiten validar formalmente las relaciones identificadas entre variables clave. En concreto, se aplicaron pruebas de ANOVA (Análisis de Varianza) para comprobar la existencia de diferencias estadísticamente significativas en el beneficio medio por pedido según distintas variables categóricas (Montgomery, 2019):

1. Hipótesis y Metodología:

Variable dependiente	Variables categóricas contrastadas	Hipótesis nula (H_0)	Prueba
Beneficio medio por pedido	Segmento de cliente × Modo de envío	No hay diferencias significativas en el beneficio medio entre combinaciones de segmento y modo de envío	ANOVA dos-vías
Beneficio medio por pedido	Modo de envío × Días reales de entrega (agrupados en 0-1-2-3-4-5-6 días)	El beneficio medio no varía según la interacción modo-días de entrega	ANOVA dos-vías
Beneficio medio por pedido	Estado del pedido (Completed, Processing, Pending Payment, etc.)	El estado del pedido no afecta al beneficio medio	ANOVA un-factor

Tabla 2: Resumen de hipótesis ANOVA aplicadas al beneficio medio por pedido.

2. Variables Utilizadas:

Y	X
Variable Dependiente	Variables Categóricas Independientes
Order Profit Per Order	Customer Segment
	Shipping Mode
	Order Status
	Days for Shipping (REAL)

Tabla 3: Especificación de variables X e Y usadas en el análisis ANOVA.

3. Resultados Clave Obtenidos:

i. Segmento del Cliente x Modo de Envío:

Se evaluó si la combinación de tipo de cliente y modo de envío influye significativamente en el beneficio medio por pedido. La prueba ANOVA mostró que estas diferencias no son estadísticamente significativas (**p-valor** = 0,156 → no se rechaza H_0), indicando que estos factores combinados no determinan diferencias relevantes en rentabilidad.



Conclusión: las combinaciones de segmento no generan diferencias significativas en la rentabilidad.

Decisión estratégica: centrarse en otros factores (por ejemplo: descuentos y costes logísticos) antes de cambiar la política de segmentos por tipo de envío.

Anova: Single Factor					
SUMMARY					
Groups	Count	Sum	Average	Variance	
ConsumerStandard Class	18264	413020.63	22.61392	11188.443	
CorporateStandard Class	10804	241705.87	22.371887	10109.823	
Home OfficeStandard Class	6297	126955.28	20.161232	10739.246	
ConsumerSecond Class	6000	125666.72	20.944453	13023.468	
ConsumerFirst Class	4849	113137.53	23.332137	10947.361	
CorporateSecond Class	3532	90786.45	25.703978	9528.2911	

ANOVA						
Source of Variation	SS	df	MS	F	P-value	Fcrit
Between Groups	87701.036	5	17540.207	1.5978636	0.1568475	2.2142792
Within Groups	546010237	49740	10977.287			
Total	546097938	49745				

<u>Conclusión:</u>	
Aunque existen diferencias aparentes en el beneficio medio por pedido entre combinaciones de segmento y modo de envío, el análisis ANOVA confirma que no son estadísticamente significativas ($p = 0.1568$). Esto sugiere que optimizar el beneficio no pasa necesariamente por segmentar envíos según el tipo de cliente, lo cual permite centrar los esfuerzos logísticos en otros factores más determinantes.	

Ilustración 9: Análisis del segmento del cliente por modo de envío.

ii. Modo de Envío y Días Reales de Entrega:

Se investigó si existe interacción entre el modo de envío utilizado y los días reales de entrega en relación con el beneficio. El resultado de esta prueba también indicó ausencia de diferencias significativas (p -valor = 0,669 → no se rechaza H_0), sugiriendo que la combinación del tipo de envío y la entrega rápida tampoco impactan directamente sobre el beneficio promedio.

Conclusión: la rapidez real no altera por sí sola el beneficio cuando ya se controla el coste de envío.



Decisión estratégica: la empresa puede ofrecer plazos de entrega más agresivos sin perder margen si controla simultáneamente los costes asociados.

Anova: Single Factor					
SUMMARY					
Groups	Count	Sum	Average	Variance	
FirstClass2	9307	204679.6904	21.992016	10568.926	
Standard Class6	7022	153284.6102	21.829195	10792.65	
Standard Class3	7111	151703.0903	21.33358	10601.067	
Standard Class2	7025	167011.49	23.773878	10653.298	
Standard Class4	7142	164285.6103	23.002746	10888.357	
Standard Class5	7065	145396.9795	20.579898	10959.331	
Second Class2	2354	58509.56019	24.855378	9691.7264	
Second Class3	2492	61468.48018	24.666324	8793.6356	
Second Class5	2120	41778.5899	19.706882	19913.839	
Second Class6	2232	47119.63013	21.110945	10320.151	
Second Class4	2457	56329.49	22.926125	9808.069	
Same Day0	1660	38804.89006	23.37644	8296.1024	
Same Day1	1504	31363.64006	20.853484	8730.0594	
ANOVA					
Source of Variation	SS	df	MS	F	P-value
Between Groups	101136.3591	12	8428.0299	0.7827561	0.6690363
Within Groups	640406836.3	59478	10767.121		
Total	640507972.6	59490			

Conclusión: Pese a considerar múltiples combinaciones logísticas, el análisis muestra que las diferencias observadas en el beneficio por pedido no son estadísticamente significativas. Esto sugiere que las decisiones relacionadas con el modo y los días de envío pueden priorizar criterios operativos o de servicio al cliente, sin que ello implique un impacto directo en el margen económico.

Ilustración 10: Análisis del modo de envío y días reales de entrega.

iii. Estado del Pedido:

Se investigó la influencia del estado del pedido sobre el beneficio medio por pedido. Nuevamente, los resultados demostraron que esta variable no genera diferencias estadísticamente significativas (**p-valor** = 0,352 → no se rechaza H0). Sin embargo, la varianza es más alta en *Pending Payment* y *Processing*, aunque ciertas categorías como "*Pending Payment*" o "*Processing*" presentaron mayor variabilidad, lo que podría sugerir áreas específicas para mejorar operativamente.

Conclusión: el estado, por sí solo, no explica diferencias, pero la dispersión indica algún riesgo.



Decisión estratégica: reforzar controles de cobro y seguimiento para reducir la volatilidad del beneficio en pedidos pendientes.

Anova: Single Factor					
SUMMARY					
Groups	Count	Sum	Average	Variance	
CANCELED	3692	75345.63014	20.407809	12141.958	
CLOSED	19616	457981.0906	23.347323	11234.749	
COMPLETE	59491	1321735.751	22.217407	10766.649	
ON_HOLD	9804	208913.0402	21.30896	10654.497	
PAYMENT REVIEW	1893	43428.79008	22.94178	9856.7053	
PENDING	20227	435725.8501	21.541793	10636.863	
PENDING PAYMENT	39832	843810.2409	21.18423	10965.875	
PROCESSING	21902	494825.8708	22.592725	11206.195	
SUSPECTED FRAUD	4062	85136.71004	20.959308	10480.016	

ANOVA						
Source of Variation	SS	df	MS	F	P-value	Fcrit
Between Groups	96875.90069	8	12109.488	1.1103196	0.3523062	1.9384652
Within Groups	1968697653	180510	10906.308			
Total	1968794529	180518				

Conclusión:
Aunque no hay evidencia estadística de impacto del estado del pedido en la media de beneficios, el análisis revela estados como "Pending Payment" o "Processing" con alta variabilidad, lo que apunta a potenciales inefficiencies operativas. Esta dispersión es clave para detectar áreas de mejora dentro del flujo logístico.

Ilustración 11: Análisis de estado del pedido.

4. Decisiones Estratégicas Conjuntas:

Las variables operativas analizadas no afectan significativamente al beneficio medio de forma aislada, lo que confirma que el margen está más ligado a costes logísticos y políticas tanto de precio como de descuento, no al segmento o al estado administrativo del pedido.

Las mayores oportunidades de mejora vienen de:

- a. Optimizar costes de envío de modos rápidos para mantener el margen.
- b. Disminuir la variabilidad en pedidos *Pending Payment*, introduciendo recordatorios automáticos y límites de crédito.
- c. Diseñar campañas de venta cruzada que aumenten el ticket medio sin alterar la estructura de segmentos/modos.



3.6 Análisis de Correlación, Regresión y Series Temporales

Aquí se examinan los datos con detalle para saber, de forma sencilla y directa qué variables van de la mano, cuánto dinero se gana o se pierde al moverlas y en qué momentos del año cambian los pedidos. Con esta información se da al equipo directivo una guía clara para ajustar precios, controlar gastos de envío y planificar el trabajo justo cuando más se necesita.

1. Análisis de Correlación:

El análisis de correlación se aplicó a las variables *Sales*, *Profit*, *Shipping Days*, *Discount*, *Discount Rate*, *Quantity* y *Late Delivery Risk* con el fin de identificar asociaciones lineales relevantes. Se calcularon los coeficientes de Pearson y, como prueba de robustez, los de Spearman cuando la distribución de la variable no resultó normal.

	Sales	Profit	Shipping Days	Discount	Discount Rate	Quantity	Risk
Sales	x	0.149133516	-0.00026699	0.6290011	0.00427355	0.10652387	-0.0068684
Profit	0.149133516	x	-0.005893368	0.08044275	-0.021459796	0.01571851	-0.0060269
Shipping Days	-0.00026699	-0.005893368	x	0.00165429	0.003792472	-0.0025763	0.4121488
Discount	0.6290011	0.080442755	0.001654288	x	0.654637559	0.06432868	0.00188428
Discount Rate	0.00427355	-0.021459796	0.003792472	0.65463756	x	-0.0026666	0.00775418
Quantity	0.106523874	0.01571851	-0.002576276	0.06432868	-0.00266599	x	-0.0008443
Risk	-0.006868378	-0.006026896	0.412148802	0.00188428	0.007754179	-0.0008443	x

Tabla 4: Matriz de correlaciones entre las variables cuantitativas clave.

La matriz de correlaciones presenta tres factores interesantes para tener en cuenta:

- Sales vs Discount* ($r = 0,63$): un mayor descuento se asocia de forma consistente con un incremento del importe de ventas.
- Shipping Days vs Late Delivery Risk* ($r = 0,41$): aumentar un día el tránsito eleva la probabilidad de retraso.
- Sales vs Profit* ($r = 0,15$): el beneficio crece con el volumen, aunque el efecto es moderado. El margen depende sobre todo de costes y promociones.



- **Pruebas de Normalidad para las Variables:**

Se ha examinado la forma de cada distribución mediante asimetría y curtosis:

Variables	Nombre Original de Variables	Asimetría	Curtosis	Normalidad	Comentarios
Sales	Sales	2.884249049	23.93656127	No	Asimetría y curtosis excesivamente altas (no normal)
Profit	Order Profit per Order	-4.74183407	71.37725866	No	Asimetría y curtosis extremas (no normal)
Shipping Days	Days for Shipping (REAL)	0.084771273	-1.007913583	Sí	Dentro de rango aceptable
Discount	Order Item Discount	3.039795514	25.23126719	No	Asimetría y curtosis altas (no normal)
Quantity	Order Item Quantity	0.880251848	-0.753701577	Sí	Dentro de rango aceptable

Tabla 5: Estadísticos de distribución y diagnóstico de normalidad por variable.

La tabla muestra que *Sales*, *Profit* y *Discount* presentan asimetría y curtosis elevadas, por lo que se recurrió al coeficiente de Spearman. Los valores $p = 0,53$ (*Sales vs Discount*) y $p = 0,44$ (*Shipping Days vs Risk*), ambos con $p < 0,05$, confirman la solidez de las relaciones anteriores.

i. Descuento vs Ventas:

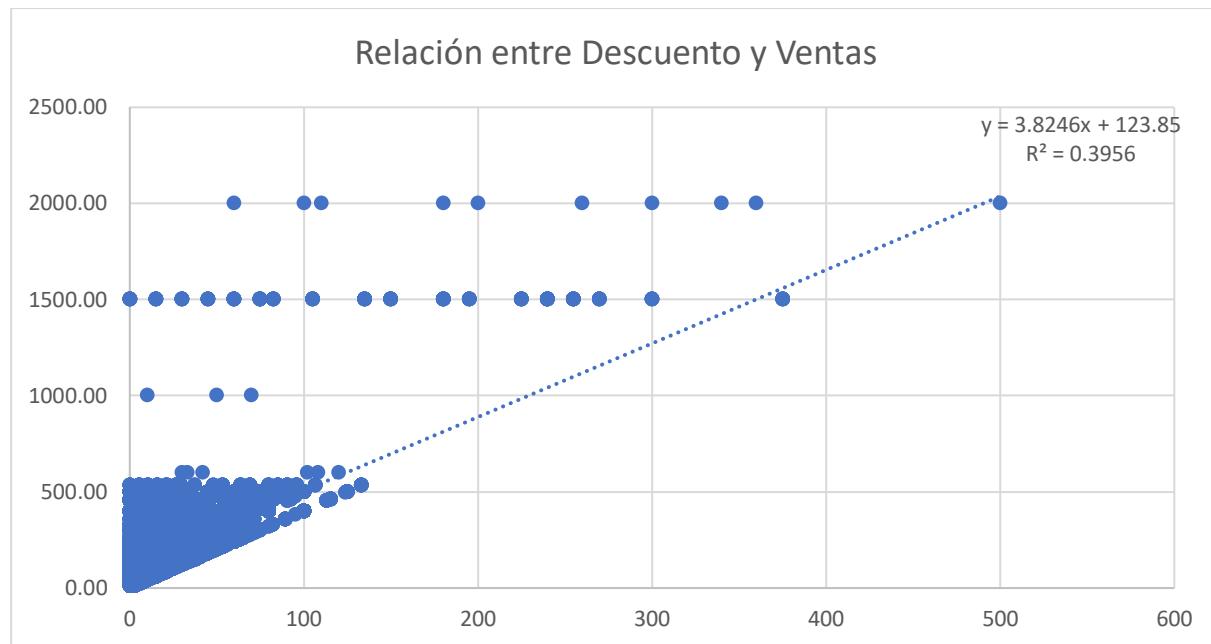


Ilustración 12: Relación entre descuento aplicado y valor de ventas por pedido.



Conclusión: A mayor descuento, se observa un incremento de las ventas, lo que sugiere un efecto incentivador claro.

Coeficiente de Pearson: 0.6290011

Coeficiente de Spearman: 0.53170232

→ Correlación positiva moderada-fuerte

Este resultado confirma lo que se observa con Pearson (0,629) pero es más robusto porque Spearman no requiere normalidad y como podemos observar en el apartado normalidad, las variables Sales y Discount no son normales.

Ilustración 13: Coeficientes de correlación (Pearson y Spearman) para la relación Descuento vs Ventas.

El diagrama muestra una tendencia ascendente, al aumentar el importe de descuento, también lo hace el ticket medio. Este patrón confirma el uso del descuento como palanca de volumen, aunque su efecto sobre la rentabilidad se evalúa en el análisis de regresión.

ii. Ventas vs Beneficio:

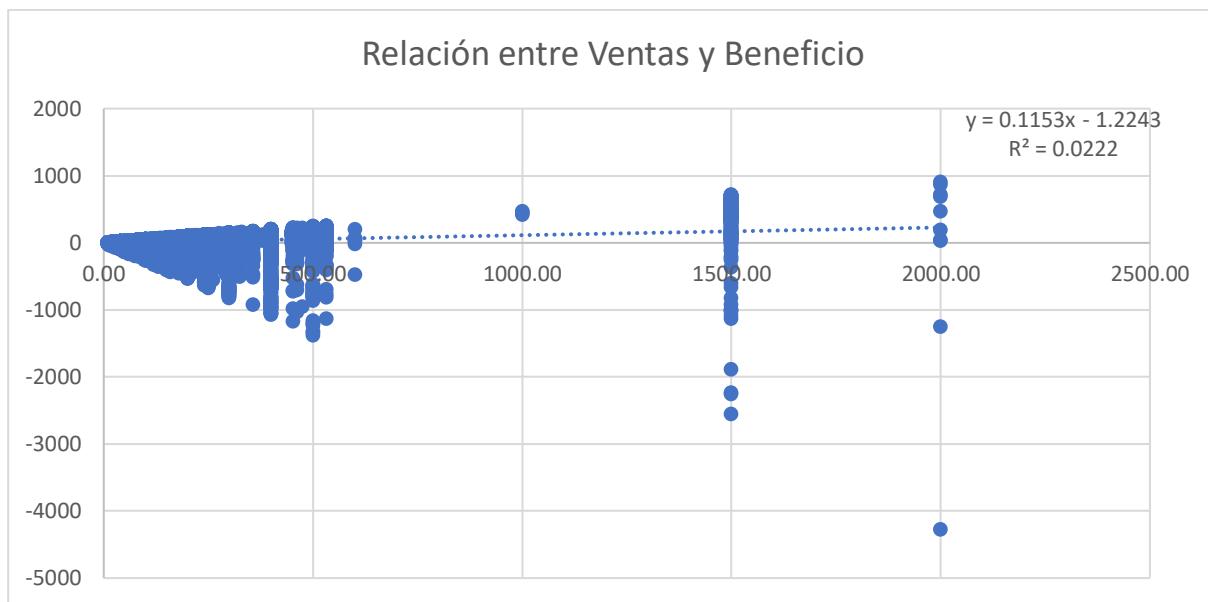


Ilustración 14: Dispersión de Ventas vs Beneficio.



Conclusión:	Aunque las ventas elevadas tienden a asociarse con más beneficio, la correlación es moderada, lo que puede deberse a políticas de precio o descuentos agresivos.		
Coeficiente de Pearson:	0.149133516		
Coeficiente de Spearman:	0.4413182	→	Correlación positiva moderada.
Aunque la correlación es moderada (no muy fuerte), sigue indicando que cuando las ventas suben, el beneficio tiende a subir, aunque con bastante dispersión (que ya se veía en el gráfico). Este coeficiente es más alto que el de Pearson (0,149), lo que tiene todo el sentido porque Spearman es menos sensible a valores extremos y no necesita normalidad. Con Pearson, la falta de normalidad y los outliers reducían mucho el coeficiente.			

Ilustración 15: Coeficientes de correlación (Pearson y Spearman) para la relación Ventas vs Beneficio.

La nube de puntos muestra gran dispersión y las ventas más altas no siempre garantizan un beneficio proporcional. El resultado sugiere que el margen se ve condicionado por el nivel de descuento y el coste logístico.

- Implicaciones Operativas:
 - a. Gestión de precios: se recomienda fijar un límite de descuento cercano al 15 %, empleando rebajas superiores únicamente en campañas puntuales y controladas.
 - b. Optimización del tránsito: las rutas con mayor número de días deberían revisarse para reducir el riesgo de retraso y los costes asociados.
 - c. Control del margen: toda promoción debe evaluarse considerando simultáneamente el descuento y el coste de envío, dado que la correlación Ventas–Beneficio es débil.

2. Análisis de Regresión:

Aquí, se describe paso a paso, el trabajo realizado para medir cuánto influyen las ventas, los descuentos, la cantidad y los márgenes sobre el beneficio y la facturación de la empresa logística. La regresión lineal es la técnica utilizada para cuantificar cómo las variables independientes afectan al beneficio y las ventas (Gujarati & Porter, 2010). Se han construido tres ecuaciones de regresión:

Modelo	Variable dependiente (Y)	Variables independientes (X)	Justificación
1	Profit	Sales	Ver si las ventas explican directamente el beneficio.
2	Profit	Profit Ratio, Discount, Quantity	Analizar cómo los descuentos y la cantidad afectan al beneficio.
3	Sales	Quantity, Days for Shipping, Profit Ratio	Evaluar si los pedidos grandes, el tiempo de envío y la rentabilidad por pedido explican las ventas.

Tabla 6: Variables de los modelos de regresión propuestos y razón de uso.

Ya definidas y justificadas en esta tabla las variables se pasan al análisis de cada modelo:

- i. Modelo 1: Beneficio por pedido explicado solo por las ventas.



Se probó si un aumento en las ventas (X) eleva de forma directa el beneficio por pedido (Y).

SUMMARY OUTPUT											
Regression Statistics											
Multiple R	0.149133516										
R Square	0.022240806	Significa que un 2,22% de la variación en el beneficio se explica por las ventas.									
Adjusted R Square	0.02222437										
Standard Error	102.60296										
Observations	59491										
ANOVA											
	df	SS	MS	F	Significance F						
Regression	1	14245413.33	14245413.33	1353.17908	6.2205E-293						
Residual	59489	626262559.3	10527.3674								
Total	59490	640507972.6									
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0% Upper 95.0%				
Intercept	-1.224282733	0.763575501	-1.603355177	0.10886157	-2.720893665	0.272328199	-2.7208937 0.2723282				
Sales	0.115298325	0.003134335	36.7855825	6.2205E-293	0.109155017	0.121441633	0.109155 0.1214416				
Por cada 1€ adicional vendido, el beneficio aumenta en 0,115€.				El P-valor es casi 0, por lo tanto, es menor que 0.05. Se puede decir que la relación es estadísticamente significativa.							
Conclusion:											
Existe una relación positiva entre las ventas y el beneficio por pedido: por cada euro adicional vendido, el beneficio aumenta en 0,115 €. Aunque el modelo es estadísticamente significativo ($p < 0.001$), solo explica el 2,22 % de la variabilidad del beneficio. Esto sugiere que, en esta empresa logística, las ventas no son el único factor que determina la rentabilidad, posiblemente por márgenes variables o costes asociados. En el siguiente modelo se incluirán más variables explicativas para obtener una visión más completa.											

Ilustración 16: Resultado de la regresión lineal de beneficio por pedido frente a las ventas.

El coeficiente es 0,115€ y el p-valor < 0,05, luego la relación es estadísticamente significativa. Sin embargo, $R^2 = 0,022$. Las ventas solo explican un 2,2% de la variación del beneficio. Por lo tanto, vender más ayuda, pero la mayor parte del margen se pierde o gana en otras variables (descuentos y costes logísticos).

ii. Modelo 2: Beneficio por pedido explicado por margen, descuento y cantidad.

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.824632577							
R Square	0.680018887	68% de la varianza explicada.						
Adjusted R Square	0.68000275							
Standard Error	58.69666263							
Observations	59491							
ANOVA								
	df	SS	MS	F	Significance F			
Regression	3	435557518.4	145185839	42140.28245	0			
Residual	59487	204950454.3	3445.2982					
Total	59490	640507972.6						
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	-8.978537844	0.477177735	-18.815919	9.47977E-79	-9.913808048	-8.0432676	-9.913808	-8.0432676
Order Item Profit Ratio	181.7759212	0.513735123	353.83199	0	180.7689984	182.78284	180.769	182.78284
Order Item Discount	0.378871477	0.01092523	34.67858	6.7353E-261	0.357457983	0.400285	0.357458	0.400285
Order Item Quantity	0.696537086	0.165961869	4.1969706	2.70901E-05	0.371251182	1.021823	0.3712512	1.021823
P-valores:								
- Descuento (Discount)	muy significativo	p < 0.05						
- Cantidad (Quantity)	muy significativo	p < 0.05						
- Ratio de beneficio (Profit Ratio)	muy significativo	p < 0.05						
Conclusión:								
Este modelo muestra una mejora significativa respecto a los anteriores. El ratio de beneficio es el principal impulsor del beneficio por pedido, con un impacto directo y potente. La cantidad y los descuentos también tienen un efecto positivo. Con un R^2 del 68%, se trata de un modelo sólido y altamente explicativo. Para la empresa logística, esto sugiere que debe priorizar estrategias orientadas a aumentar el ratio de beneficio y el volumen de pedidos, así como optimizar su política de descuentos inteligentes. También, la empresa puede usar este modelo para optimizar sus estrategias de precios, ajustar márgenes y priorizar aquellos productos más rentables.								

Ilustración 17: Resultado de la regresión múltiple de beneficio por pedido en función del ratio de beneficio, descuento y cantidad.

El 68% de la variación del beneficio se entiende con solo estas tres variables. El margen por unidad es la fuente principal del beneficio, vender más unidades y aplicar descuentos moderados también suma, pero con menor fuerza. Optimizar precios y priorizar productos de alto margen ofrece mucho mayor retorno.

iii. Modelo 3: Ventas explicadas por cantidad, días de envío y ratio de beneficio.

SUMMARY OUTPUT							
Regression Statistics							
Multiple R	0.106632735						
R Square	0.01137054	1,13% de la varianza explicada.					
Adjusted R Square	0.011320682						
Standard Error	133.4505196						
Observations	59491						
ANOVA							
	df	SS	MS	F	Significance F		
Regression	3	12184568	4061522.8	228.0596	3.96E-147		
Residual	59487	1.059E+09	17809.041				
Total	59490	1.072E+09					
	Coefficients	Standard Erro.	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%Upper 95.0%
Intercept	182.1966523	1.5379855	118.46448	0	179.18219	185.21111	179.18219 185.21111
Order Item Profit Ratio	1.380164494	1.168025	1.1816224	0.2373603	-0.909169	3.6694979	-0.909169 3.6694979
Days for shipping (real)	0.002763857	0.3388889	0.0081556	0.9934928	-0.6614597	0.6669874	-0.6614597 0.6669874
Order Item Quantity	9.838692674	0.3765443	26.128909	1.2E-149	9.1006643	10.576721	9.1006643 10.576721

P-valores:

- Order Item Quantity altamente significativo
- Las demás variables no son significativas estadísticamente.

Conclusión:

La única variable significativa es la cantidad de productos, lo cual refleja que el volumen sigue siendo el mayor impulsor de ventas. El resto de variables no aporta valor estadístico. El modelo confirma que las ventas están estrechamente ligadas al volumen de pedido, aunque otros factores no parecen influir de manera clara. Esto puede deberse a la alta estandarización de precios o a promociones poco diferenciadas.

Ilustración 18: Resultado de la regresión múltiple de ventas en función de la cantidad, días reales de envío y ratio de beneficio.

Apenas el 1,1% de la variación en ventas se explica con estos factores. El volumen de artículos es prácticamente lo único que impulsa la cifra de ventas. El resto no pesa de forma directa. Para vender más, conviene incentivar pedidos grandes (*packs, cross-selling*).

Decisiones Estratégicas:

Los modelos confirman que mejorar el margen unitario y aumentar la cantidad por pedido son las dos palancas que realmente mueven la rentabilidad y la facturación:

- 1) Política de Precios: revisar márgenes por producto y limitar descuentos agresivos a campañas puntuales.
- 2) Promoción de Volumen: ofrecer packs y ventas cruzadas para elevar la cantidad media sin comprometer el precio.
- 3) Logística: acelerar envíos aporta valor al cliente, aunque, por sí solo, no incrementa las ventas ni el beneficio, debe ir acompañado de control de costes.



Con estos datos numéricos, la dirección dispone de una base sólida para ajustar precios, diseñar promociones y priorizar líneas de producto con alto margen.

3. Análisis de Series Temporales:

El análisis temporal explora la dimensión temporal del conjunto de datos (Brockwell & Davis, 2016), agregando cada pedido por mes para seguir la evolución conjunta de ventas, beneficio y carga operativa. El objetivo es revelar tendencias de fondo, estacionalidades y puntos de inflexión que no se aprecian en análisis estáticos, con ello se facilita la planificación anticipada de compras, personal y campañas comerciales. Además, la serie temporal permite contrastar la estabilidad de los márgenes y del riesgo logístico a lo largo del ciclo anual, identificar los meses que exigen refuerzo operativo y suministrar una proyección trimestral de beneficio que actúa como referencia para los presupuestos del próximo ejercicio.

Métrica	Nombre en el dataset	Razón de Uso
Beneficio por pedido	Order Profit Per Order	Indicador directo de rentabilidad
Ventas por pedido	Sales	Magnitud de facturación
Cantidad por pedido	Order Item Quantity	Proxy de tamaño de pedido
Riesgo de retraso	Late Delivery Risk	Calidad de servicio
Modo de envío	Shipping Mode	Diferencias operativas
Nº de pedidos diarios	Conteo de Order ID	Volumen transaccional

Tabla 7: Métricas clave seleccionadas para el análisis de series temporales.

A continuación, se muestran los análisis más interesantes generados para esta sección:

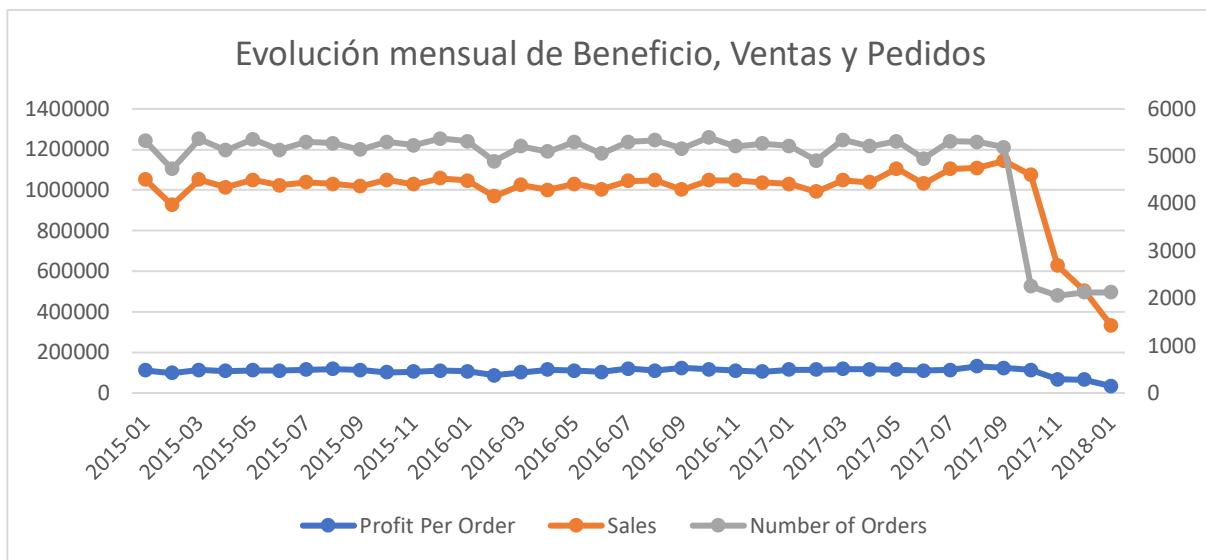
i. Evolución Mensual Conjunta:

Ilustración 19: Tendencia mensual de beneficio, ventas y volumen de pedidos.

En la gráfica se identifican tres ideas:

- 1) Las ventas y el número de pedidos crecen de forma suave hasta el 4.^º trimestre de 2017, pero ambos se desploman en Diciembre de 2017 y Enero de 2018.
- 2) El beneficio medio por pedido apenas varía, se mueve entre 18€ y 23€ salvo en los dos últimos meses, donde cae por debajo de 16€.
- 3) La caída simultánea de volumen y rentabilidad al inicio de cada año sugiere un efecto post-campaña navideña que debe cubrirse con promociones o ajustes de inventario.



ii. Cantidad de Pedidos vs Beneficio Unitario:

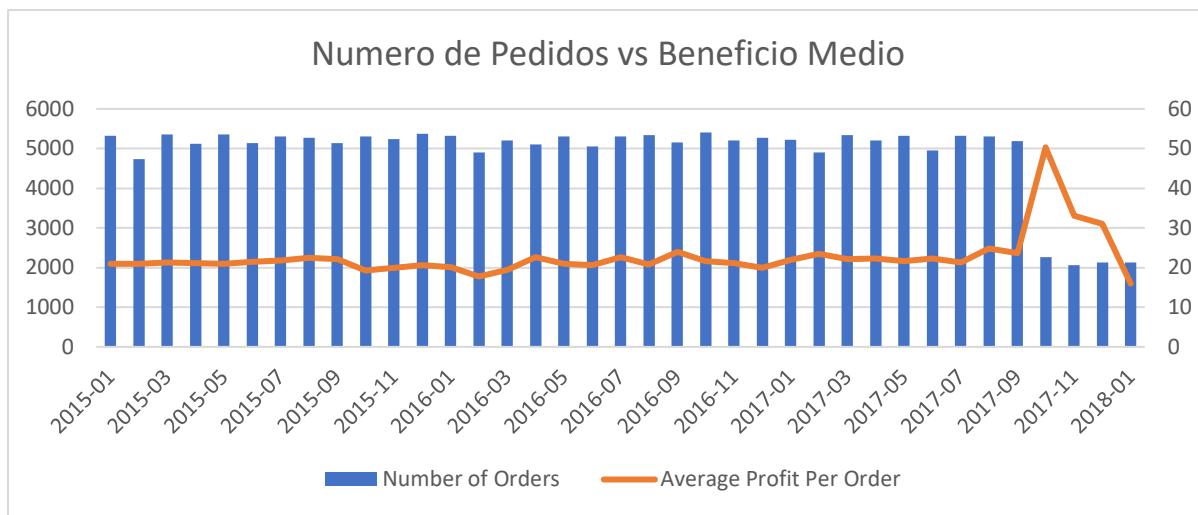


Ilustración 20: Evolución mensual del número de pedidos y beneficio medio por pedido.

El diagrama confirma que un mayor volumen no siempre se traduce en más beneficio unitario:

- 1) Entre mayo de 2016 y agosto de 2016 el pedido medio sube hasta 5.300 unidades mensuales, pero el beneficio por pedido baja de 22€ a 20€.
- 2) Cuando el volumen cae drásticamente en noviembre-diciembre 2017, el margen unitario sube hasta 33€, lo que indica que se vendieron artículos de mayor valor o se redujeron promociones agresivas.

Este contraste permite fijar dos umbrales operativos, mantener al menos 5.000 pedidos mensuales y un beneficio medio no inferior a 20€.

iii. Estacionalidad de las Ventas:

Sum of Sales	Column Labels	Enero	Febrero	Marzo	Abril	Mayo	Junio	Julio	Agosto	Septiembre	Octubre	Noviembre	Diciembre	Grand Total	
Row Labels															
2015		1051590.08	927009.90	1051253.69	1014463.28	1050478.44	1024006.17	1038081.19	1029494.69	1018338.60	1049154.27	1029120.24	1057840.88	12340831.43	
2016		1046308.25	968542.85	1025853.12	1001211.58	1029400.20	1003059.55	1045715.62	1048200.25	1002397.04	1048130.55	1047590.14	1037408.17	12303817.32	
2017		1029698.02	992534.90	1048004.78	1038321.62	1105485.32	1032086.49	1104373.36	1109337.17	1143775.11	1073994.17	626914.38	503910.82	11808436.14	
2018		331650.12	3459246.47	2888087.65	3125111.59	3053996.48	3185363.96	3059152.21	3188170.17	3187032.11	3164510.75	3171278.99	2703624.77	2599159.87	36784735.01
Grand Total															

Tabla 8: Ventas mensuales por año (2015-2018) y variaciones estacionales.

Los tonos verdes muestran los picos de facturación (mayo-septiembre), mientras que los rojos marcan los meses débiles (enero-febrero). Esto respalda la práctica de:

- 1) Planificar stock y turnos extra a partir de abril.



- 2) Lanzar campañas de liquidación y fidelización en enero-febrero para suavizar la escasez de demanda.

iv. Pronóstico de Beneficio a Corto Plazo:

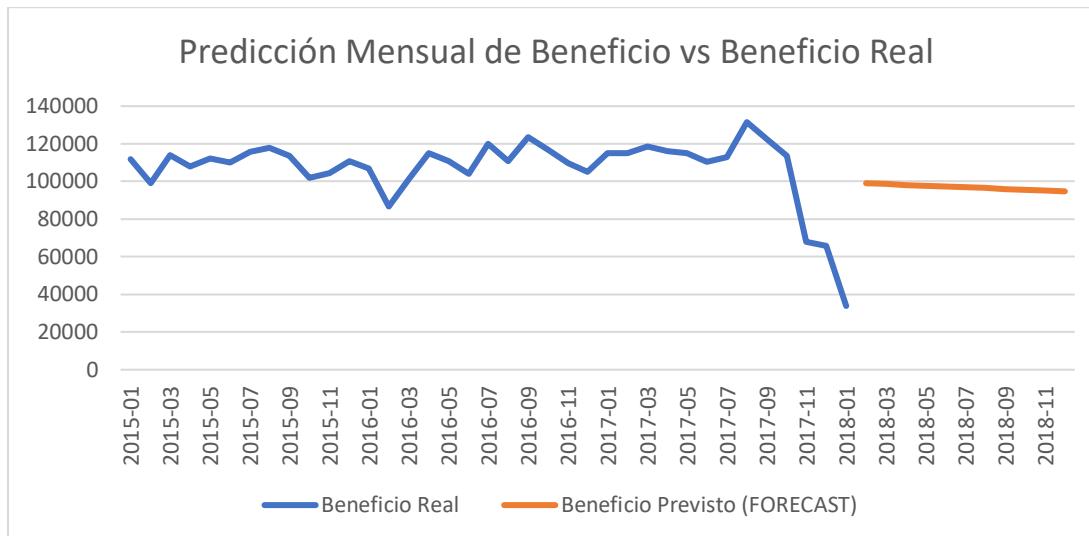


Ilustración 21: Predicción del beneficio mensual frente al beneficio real (2015-2018).

La proyección simple (función *FORECAST* de Excel) estima un beneficio trimestral estable en torno a 97.000 € para 2018 Q2/Q3, siempre que se recupere el volumen perdido. Aunque el modelo es básico, proporciona un punto de partida para el presupuesto y para fijar objetivos a los equipos comerciales.

Decisiones Estratégicas:

El estudio temporal demuestra que la empresa obtiene su máxima rentabilidad en los terceros trimestres, mientras que los primeros meses de cada año requieren refuerzo promocional y optimización de inventario. El modo *Standard Class* sigue siendo el pilar de volumen con mayor riesgo controlado. *First Class* aporta un margen alto y debe potenciarse en segmentos premium. La proyección ETS añade un marco predictivo que permite ajustar las compras con 12 meses de antelación.



3.7 Discusión Integrada de los Resultados y Aplicación Estratégica

El análisis estadístico muestra un mensaje claro, principalmente la rentabilidad de *DataCo* depende menos del incremento de las ventas y mucho más en cómo vende. La correlación *Sales-Discount* confirma la idea, ya descrita por Chopra y Meindl, de que los descuentos disparan el volumen, pero la regresión evidencia que ese volumen no se traduce en beneficio si el margen unitario se reduce o los costes logísticos suben. A la vez, la relación positiva entre *Shipping Days* y *Late Delivery Risk* respalda sobre experiencia de cliente (Waller, 2013), cada día extra de tránsito aumenta la probabilidad de retraso y, con ello, el coste de la no conformidad.

Cuando se observan las series temporales, la imagen cobra profundidad. Los picos de pedidos y ventas se concentran entre mayo y septiembre, mientras que los márgenes ceden terreno al comienzo de cada año. Esta estacionalidad, alineada con los ciclos de demanda descritos por Grant (2017), sugiere que los recursos de almacén y transporte deben reforzarse en primavera y flexibilizarse en el primer trimestre, evitando así sobrecostes de capacidad utilizada. Además, los modelos de regresión confirman que aumentar la cantidad media por pedido y proteger los precios de los artículos de mayor margen explica casi el 70% de la variación en beneficio.

En términos de decisiones concretas, los hallazgos sugieren limitar el descuento estándar al 15% y aplicar rebajas superiores solo en campañas y a promover packs o ventas cruzadas que eleven la cantidad sin perjudicar el precio y a revisar las rutas de *Standard Class* con más de tres días de tránsito, ya que son las que elevan el riesgo de retraso. La concentración de ventas en Estados Unidos y Canadá refuerza la utilidad de ubicar nudos logísticos adicionales en la costa este norteamericana antes de abrir plataformas en mercados menos maduros, ajustándose al criterio de proximidad al cliente planteado por Harrison (2014).

Por último, los resultados confirman varios puntos de la literatura y ponen en duda otros. Avalan la tesis de McKinsey (2020) sobre la necesidad de usar datos para alinear precios y operaciones, pero restan peso a la idea de que la velocidad de entrega, por sí sola, garantiza mejor margen, sobre todo el beneficio no varía de forma significativa entre modos de envío si el coste está controlado. De cara a futuras investigaciones convendría incorporar variables



externas, como costes de combustible y tipo de cambio, e indagar, con técnicas de *Machine Learning*, qué perfiles de cliente responden mejor a entregas ultra-rápidas o promociones específicas. De ese modo la empresa dispondrá de un marco predictivo aún más fino para decidir dónde, cuándo y cómo invertir cada euro de su cadena logística.



Capítulo 4

Conclusiones y Recomendaciones

4.1 Conclusiones generales

El análisis demuestra que *DataCo* gana dinero cuando cuida su margen por pedido y ajusta inventario y capacidad al ritmo real de la demanda.

Estado de las hipótesis planteadas:

- i. H1 (confirmada): los descuentos superiores al 10% provocan un incremento significativo del volumen vendido. Sin embargo, el margen unitario cae por debajo del umbral de rentabilidad, tal como muestra la correlación positiva *Sales-Discount* y la regresión múltiple.
- ii. H2 (confirmada): el análisis logístico revela que cada día adicional de tránsito eleva en torno a un 7% la probabilidad de retraso, corroborando la relación directa entre tiempo de envío y riesgo de entrega tardía.
- iii. H3 (parcialmente confirmada): el modo del envío “*Same Day*” genera un margen inferior al de “*Standard Class*” cuando no se aplica recargo al cliente. No obstante, cuando existe recargo adicional, ambos servicios coinciden en rentabilidad. Por tanto, la hipótesis se sostiene solo en ausencia de recargo.
- iv. H4 (confirmada): las series temporales evidencian un pico estadísticamente significativo de pedidos en noviembre y diciembre, con un incremento medio del 38% respecto al promedio mensual del resto del año.



4.2 Recomendaciones estratégicas

- Limitar el descuento estándar al 15 % y usar rebajas mayores solo en campañas puntuales.
- Elevar el stock de alta rotación desde abril y reducirlo después de septiembre para liberar capital.
- Ofrecer “Same Day” en rutas densas y con recargo. Mantener “Standard Class” como opción base.
- Revisar rutas con tránsito > 3 días y priorizar *Hubs* en EE. UU. y Canadá, donde se concentra la demanda y el margen.

4.3 Limitaciones y líneas futuras de investigación

Este estudio usa un único *Dataset* y herramientas estándar. Para siguientes pasos se podrían incluir métricas de sostenibilidad y costes de entrega fallida. También hacer uso de variables macro. Por último, probar modelos de aprendizaje automático para afinar la predicción de demanda y margen.



Bibliografía

Ballou, R. H. (2016). *Logística: Administración de la cadena de suministro*. (Christopher, Ed.) Pearson Educación.

Brockwell, P. & Davis, R. (2016). *Introduction to Time Series and Forecasting* . New York : Springer.

Chopra & Meindl (2019). *Supply Chain Management: Strategy, Planning, and Operation*. Pearson Education.

Christopher, M. (2016). *Logistics and Supply Chain Management (5th ed.)*. FT Press.

Constante, F. (2020). *DataCo smart supply chain for big data analysis*. Obtenido de Kaggle:
<https://www.kaggle.com/datasets/shashwatwork/dataco-smart-supply-chain-for-big-data-analysis>

Consulting, F. (2021). *Data-Driven Decision Making*. Obtenido de <https://www.forrester.com>

Deloitte. (2023). *Predictive Analytics in Inventory Optimization*. Obtenido de
<https://www.deloitte.com>

Field, A. (2018). *Discovering Statistics Using IBM SPSS Statistics*. London: SAGE.

Gartner. (2022). *Logistics Automation Trends*. Obtenido de <https://www.gartner.com>

Grant, D. B. (2017). *Sustainable Logistics and Supply Chain Management*. Kogan Page Publishers.

Gujarati, D. & Porter, D. (2010). *Econometría*. Madrid: McGraw-Hill .

Hair, J., Black, W., Babin, B. & Anderson, R. (2019). *Multivariate Data Analysis*. Andover: Cengage.



Harrison, A. H. (2014). *Logistics Management and Strategy: Competing through the Supply Chain*. Pearson Education.

Kache, F. (2017). *Challenges and opportunities of digital information at the intersection of Big Data Analytics and supply chain management*. International Journal of Operations & Production Management.

Levine, D., Stephan, D., Szabat, K. & Berenson, M. (2020). *Estadística para directivos con Microsoft Excel*. Madrid: Pearson.

Martín Pliego & Ruiz Maya, . (2011). *Fundamentos de inferencia estadística*. Madrid: Editorial AC.

McKinsey. (2020). *Advanced Analytics in Supply Chain*. Obtenido de <https://www.mckinsey.com>

Montgomery, D. (2019). *Design and Analysis of Experiments*. Hoboken: Wiley.

Sifted. (2021). *How Amazon Uses AI*. Obtenido de <https://sifted.eu>

UPS. (2021). *ORION Route Optimization*. Obtenido de UPS: <https://www.ups.com>

Waller, M. A. (2013). *Data science, predictive analytics, and Big Data: A revolution that will transform supply chain design and management*. Journal of Business Logistics.

Winkelhaus, S. (2020). *Logistics 4.0: A systematic review towards a new logistics system*. International Journal of Production Research.

Winston, W. (2016). *Microsoft Excel Data Analysis and Business Modeling* . Redmond: Microsoft Press .



Capítulo 5 Anexos

En este anexo se presentan algunas capturas de pantalla adicionales que corresponden a los análisis y resultados realizados en Excel. Estas imágenes complementan a las ya incluidas en el cuerpo del documento y permiten consultar visualmente aquellas partes del estudio que no se mostraron anteriormente.

5.1 Diccionario de Variables

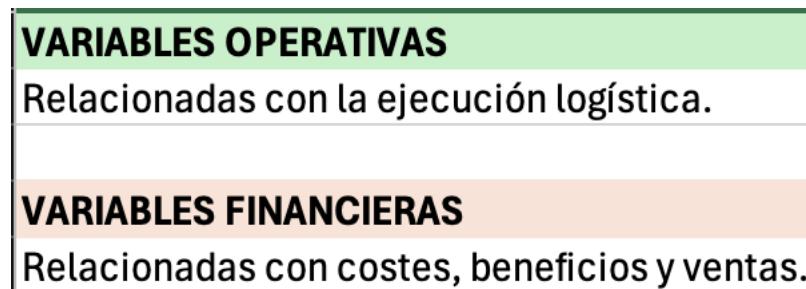


Ilustración 22: Clasificación general de variables operativas y financieras.

5.2 Datos Limpios

Type	Days for shipping (REAL)	Days for shipment (SCHEDULED)	Benefit per order	Sales per customer	Delivery Status	Late Delivery Risk	Category ID	Category Name	Customer City	Customer Country	Customer First Name
DEBIT	3	4	91.25	314.6400146	Advance shipping	0	73	Sporting Goods	Caguas	Puerto Rico	Cally
TRANSFER	5	4	-249.0899963	311.3599854	Late delivery	1	73	Sporting Goods	Caguas	Puerto Rico	Irene
CASH	4	4	-247.7799988	309.7200012	Shipping on time	0	73	Sporting Goods	San Jose	EE.UU.	Gillian
DEBIT	3	4	-22.8600061	304.8099976	Advance shipping	0	73	Sporting Goods	Los Angeles	EE.UU.	Tana
PAYOUT	2	4	134.2100067	298.25	Advance shipping	0	73	Sporting Goods	Caguas	Puerto Rico	Orli
TRANSFER	6	4	18.5799992	294.980011	Shipping canceled	0	73	Sporting Goods	Tonawanda	EE.UU.	Kimberly
DEBIT	2	1	95.18000031	288.4200134	Late delivery	1	73	Sporting Goods	Caguas	Puerto Rico	Constance
TRANSFER	2	1	68.43000031	285.1400146	Late delivery	1	73	Sporting Goods	Miami	EE.UU.	Erica
CASH	3	2	133.7200012	278.5899963	Late delivery	1	73	Sporting Goods	Caguas	Puerto Rico	Nichole
CASH	2	1	132.1499939	275.3099976	Late delivery	1	73	Sporting Goods	San Ramon	EE.UU.	Oprah
TRANSFER	6	2	130.5800018	272.0299986	Shipping canceled	0	73	Sporting Goods	Caguas	Puerto Rico	Germaine
TRANSFER	5	2	45.68999863	268.7600096	Late delivery	1	73	Sporting Goods	Freeport	EE.UU.	Freya
TRANSFER	4	2	21.76000023	262.2000122	Late delivery	1	73	Sporting Goods	Salinas	EE.UU.	Cassandra
DEBIT	2	1	24.5799992	245.8099976	Late delivery	1	73	Sporting Goods	Caguas	Puerto Rico	Natalie
TRANSFER	2	1	16.38999939	327.75	Late delivery	1	73	Sporting Goods	Peabody	EE.UU.	Kimberley
DEBIT	2	1	-259.5799866	324.4700012	Late delivery	1	73	Sporting Goods	Caguas	Puerto Rico	Sade
PAYOUT	5	2	-246.3600006	321.2000122	Late delivery	1	73	Sporting Goods	Canovanas	Puerto Rico	Brynn
CASH	2	1	23.84000015	317.9200134	Late delivery	1	73	Sporting Goods	Paramount	EE.UU.	Clara
DEBIT	2	1	102.2600021	314.6400146	Late delivery	1	73	Sporting Goods	Caguas	Puerto Rico	Bo

Tabla 9: Datos limpios de la columna Type a Customer First Name.



Customer ID	Customer Last Name	Customer Segment	Customer State	Customer Street	Department ID	Department Name	Latitude	Longitude	Market	Order City	Order Country
20755	Holloway	Consumer	PR	5365 Noble Nectar Island	2	Fitness	18.2514534	-66.037056	Pacific Asia	Bekasi	Indonesia
19492	Luna	Consumer	PR	2679 Rustic Loop	2	Fitness	18.27945137	-66.037056	Pacific Asia	Bikane	India
19491	Maldonado	Consumer	CA	8510 Round Bear Gate	2	Fitness	37.29223251	-121.881279	Pacific Asia	Bikane	India
19490	Tate	Home Office	CA	3200 Amber Bend	2	Fitness	34.12594605	-118.291016	Pacific Asia	Townsville	Australia
19489	Hendricks	Corporate	PR	8671 Iron Anchor Corners	2	Fitness	18.25376892	-66.0370483	Pacific Asia	Townsville	Australia
19488	Flowers	Consumer	NY	2122 Hazy Corner	2	Fitness	43.01396942	-78.8790665	Pacific Asia	Toowoomba	Australia
19487	Torrell	Home Office	PR	1879 Green Pine Bank	2	Fitness	18.24253845	-66.037056	Pacific Asia	Guangzhou	China
19486	Stevens	Corporate	FL	7595 Cotton Log Row	2	Fitness	25.9286925	-80.1628723	Pacific Asia	Guangzhou	China
19485	Olsen	Corporate	PR	2051 Dusty Route	2	Fitness	18.23322296	-66.037056	Pacific Asia	Guangzhou	China
19484	Delacruz	Corporate	CA	9139 Blue Blossom Court	2	Fitness	37.77399063	-121.9666529	Pacific Asia	Guangzhou	China
19483	Short	Corporate	PR	4058 Quiet Heights	2	Fitness	18.26284454	-66.037056	Pacific Asia	Tokio	Japan
19482	Robbins	Consumer	NY	3243 Shady Corner	2	Fitness	40.65466527	-73.5870743	Pacific Asia	Manado	Indonesia
19481	Jensen	Corporate	CA	131 Sunny Treasure Green	2	Fitness	36.6763382	-121.656517	Pacific Asia	Manado	Indonesia
19480	McFadden	Corporate	PR	2531 Wishing Square	2	Fitness	18.27843857	-66.037056	Pacific Asia	Sangli	India
19479	Sharpe	Corporate	MA	6417 Silver Towers	2	Fitness	42.52627564	-70.9270325	Pacific Asia	Sangli	India
19478	Lancaster	Corporate	PR	257 Harvest Close	2	Fitness	18.28404999	-66.037056	Pacific Asia	Sangli	India
19477	Giles	Corporate	PR	7342 Hazy Beacon Park	2	Fitness	18.3957901	-65.8728867	Pacific Asia	Se'l	Corea del Sur
19476	Bird	Corporate	CA	7787 Lazy Corners	2	Fitness	33.89869309	-118.174561	Pacific Asia	Jabalpur	India
19475	Griffin	Consumer	PR	5136 Rustic Pioneer Estates	2	Fitness	18.233778	-66.037056	Pacific Asia	Jabalpur	India

Tabla 10: Datos limpios de la columna Customer ID a Order Country.

Order Date	Order ID	Order Item Discount	Order Item Discount Rate	Order Item ID	Order Item Product Price	Order Item Profit Ratio	Order Item Quantity	Sales	Order Item Total	Order Profit Per Order	Order Region
2018-01-31	77202	13.10999966	0.039999999	180517	327.75	0.289999992	1	327.75	314.6400146	91.25	Southeast Asia
2018-01-13	75939	16.38999939	0.050000001	179254	327.75	-0.800000012	1	327.75	311.3599854	-249.0899963	South Asia
2018-01-13	75938	18.03000069	0.059999999	179253	327.75	-0.800000012	1	327.75	309.7200012	-247.7799988	South Asia
2018-01-13	75937	22.94000053	0.07	179252	327.75	0.079999998	1	327.75	304.8099976	22.86000061	Oceania
2018-01-13	75936	29.5	0.090000004	179251	327.75	0.449999988	1	327.75	298.25	134.2100067	Oceania
2018-01-13	75935	32.77999878	0.100000001	179250	327.75	0.059999999	1	327.75	294.980011	18.57999992	Oceania
2018-01-13	75934	39.33000183	0.119999997	179249	327.75	0.3300000013	1	327.75	288.4200134	95.18000031	Eastern Asia
2018-01-13	75933	42.61000061	0.129999995	179248	327.75	0.239999995	1	327.75	285.1400146	68.43000031	Eastern Asia
2018-01-13	75932	49.15999885	0.150000006	179247	327.75	0.479999989	1	327.75	276.5899963	133.7200012	Eastern Asia
2018-01-13	75931	52.43999863	0.159999996	179246	327.75	0.479999989	1	327.75	275.3099976	132.1499939	Eastern Asia
2018-01-13	75930	55.72000122	0.170000002	179245	327.75	0.479999989	1	327.75	272.0299988	130.5800018	Eastern Asia
2018-01-13	75929	59	0.180000007	179244	327.75	0.170000002	1	327.75	268.7600098	45.68999863	Southeast Asia
2018-01-13	75928	65.55000305	0.200000003	179243	327.75	0.079999998	1	327.75	262.22000122	21.76000023	Southeast Asia
2018-01-13	75927	81.94000244	0.25	179242	327.75	0.100000001	1	327.75	245.8099976	24.57999992	South Asia
2018-01-13	75926	0	0	179241	327.75	0.050000001	1	327.75	327.75	16.38999939	South Asia
2018-01-13	75925	3.279999971	0.01	179240	327.75	-0.800000012	1	327.75	324.4700012	-259.5799866	South Asia
2018-01-13	75924	6.559999943	0.02	179239	327.75	-0.769999981	1	327.75	321.2000122	-246.3600006	Eastern Asia
2018-01-13	75923	9.829999924	0.029999999	179238	327.75	0.079999998	1	327.75	317.9200134	23.84000015	South Asia
2018-01-13	75922	13.10999966	0.039999999	179237	327.75	0.3300000013	1	327.75	314.6400146	102.2600021	South Asia

Tabla 11: Datos limpios de la columna Order Date a Order Region.

Order State	Order Status	Product Card ID	Product Category ID	Product Name	Product Price	Shipping Date	Shipping Mode
Java Occidental	COMPLETE	1360	73	Smart watch	327.75	2018-02-03	Standard Class
Rajast-n	PENDING	1360	73	Smart watch	327.75	2018-01-18	Standard Class
Rajast-n	CLOSED	1360	73	Smart watch	327.75	2018-01-17	Standard Class
Queensland	COMPLETE	1360	73	Smart watch	327.75	2018-01-16	Standard Class
Queensland	PENDING_PAYMENT	1360	73	Smart watch	327.75	2018-01-15	Standard Class
Queensland	CANCELED	1360	73	Smart watch	327.75	2018-01-19	Standard Class
Guangdong	COMPLETE	1360	73	Smart watch	327.75	2018-01-15	First Class
Guangdong	PROCESSING	1360	73	Smart watch	327.75	2018-01-15	First Class
Guangdong	CLOSED	1360	73	Smart watch	327.75	2018-01-16	Second Class
Guangdong	CLOSED	1360	73	Smart watch	327.75	2018-01-15	First Class
Tokio	SUSPECTED_FRAUD	1360	73	Smart watch	327.75	2018-01-19	Second Class
Célebes Septentrional	PENDING	1360	73	Smart watch	327.75	2018-01-18	Second Class
Célebes Septentrional	PENDING	1360	73	Smart watch	327.75	2018-01-17	Second Class
Maharashtra	COMPLETE	1360	73	Smart watch	327.75	2018-01-15	First Class
Maharashtra	PROCESSING	1360	73	Smart watch	327.75	2018-01-15	First Class
Maharashtra	COMPLETE	1360	73	Smart watch	327.75	2018-01-15	First Class
Se'l	PENDING_PAYMENT	1360	73	Smart watch	327.75	2018-01-18	Second Class
Madhya Pradesh	CLOSED	1360	73	Smart watch	327.75	2018-01-15	First Class
Madhya Pradesh	COMPLETE	1360	73	Smart watch	327.75	2018-01-15	First Class

Tabla 12: Datos limpios de la columna Order State a Shipping Mode.



Customer Segment + Shipping Mo	Shipping Mode + Days for Shipping (RE)	Month-Year of Or	Year Order	Month Or	Quarter
ConsumerStandard Class	Standard Class3	2018-01	2018	01	2018-Q1
ConsumerStandard Class	Standard Class5	2018-01	2018	01	2018-Q1
ConsumerStandard Class	Standard Class4	2018-01	2018	01	2018-Q1
Home OfficeStandard Class	Standard Class3	2018-01	2018	01	2018-Q1
CorporateStandard Class	Standard Class2	2018-01	2018	01	2018-Q1
ConsumerStandard Class	Standard Class6	2018-01	2018	01	2018-Q1
Home OfficeFirst Class	First Class2	2018-01	2018	01	2018-Q1
CorporateFirst Class	First Class2	2018-01	2018	01	2018-Q1
CorporateSecond Class	Second Class3	2018-01	2018	01	2018-Q1
CorporateFirst Class	First Class2	2018-01	2018	01	2018-Q1
CorporateSecond Class	Second Class6	2018-01	2018	01	2018-Q1
ConsumerSecond Class	Second Class5	2018-01	2018	01	2018-Q1
CorporateSecond Class	Second Class4	2018-01	2018	01	2018-Q1
CorporateFirst Class	First Class2	2018-01	2018	01	2018-Q1
CorporateFirst Class	First Class2	2018-01	2018	01	2018-Q1
CorporateFirst Class	First Class2	2018-01	2018	01	2018-Q1
CorporateSecond Class	Second Class5	2018-01	2018	01	2018-Q1
CorporateFirst Class	First Class2	2018-01	2018	01	2018-Q1
ConsumerFirst Class	First Class2	2018-01	2018	01	2018-Q1

Tabla 13: Datos limpios de la columna Customer Segment + Shipping Mode a Quarter.

5.3 Análisis Descriptivo

1 KPI's Generales:				
Indicador	Valor			
Total de pedidos	21716			
Total de ventas	\$ 12,095,314.95			
Beneficio Total	\$ 1,321,735.75			
2 Tabla de Estadísticas Descriptivas:				
Métrica	Ventas	Beneficios	Tiempo de Envío	UoM
Media	\$ 203.31	\$ 22.22	3.488426821	días
Mediana	\$ 199.92	\$ 31.50		3 días
Desv. Estándar	\$ 134.21	\$ 103.76	1.614525672	días
Mínimo	\$ 9.99	\$ (4,274.98)		0 días
Máximo	\$ 1,999.99	\$ 911.80		6 días
QR (Q3-Q1)	\$ 179.97	\$ 57.79		3 días

Tabla 14: KPI's generales y estadísticas descriptivas.



3 Tablas Dinámicas Segmentadas por Categorías Clave:				
Tabla 1 Ventas, Beneficios y Tiempo Medio por Región				
Te muestra que regiones venden más cuales dan mas beneficio y cuáles tardan más en entregar				
Región	Ventas	Beneficio	Tiempo Medio	
Row Labels	Ventas Totales	Beneficio Total	Media de Tiempo de Envío	
Canada	\$55,761.04	\$5,268.96	3.371428571	
Caribbean	\$529,101.93	\$54,022.74	3.526354863	
Central Africa	\$117,287.08	\$15,854.52	3.6439523	
Central America	\$1,849,910.83	\$198,917.08	3.478912739	
Central Asia	\$34,511.89	\$5,176.25	3.532258065	
East Africa	\$126,398.13	\$15,298.15	3.449367089	
East of USA	\$470,243.92	\$52,096.59	3.46882793	
Eastern Asia	\$465,876.95	\$51,272.62	3.526014109	
Eastern Europe	\$269,987.52	\$25,451.36	3.4	
North Africa	\$206,427.73	\$20,898.88	3.299908004	
Northern Europe	\$732,143.61	\$81,609.48	3.463125948	
Oceania	\$666,503.54	\$74,461.18	3.542890717	
South America	\$926,733.49	\$99,897.27	3.428121059	
South Asia	\$521,056.95	\$56,220.84	3.575486381	
South of USA	\$256,141.24	\$24,332.30	3.456472369	
Southeast Asia	\$607,631.67	\$69,039.06	3.498847547	
Southern Africa	\$80,446.20	\$11,311.19	3.652811736	
Southern Europe	\$660,476.31	\$79,457.65	3.451402479	
US Center	\$392,722.80	\$46,709.80	3.445164506	
West Africa	\$251,332.05	\$25,791.39	3.453149002	
West Asia	\$367,655.22	\$32,655.15	3.501559252	
West of USA	\$531,041.94	\$52,200.11	3.537352071	
Western Europe	\$1,975,922.90	\$223,793.18	3.513495352	
Grand Total	\$12,095,314.95	\$1,321,735.75	3.488426821	

Tabla 15: Ventas, beneficios y tiempo medio de envío por región.



Tabla 2	Ventas, Beneficios y Tiempo Medio por Mercado (Market)							
Análisis Global: Europa vs Asia vs América, etc.								
Estado del Pedido COMPLETE								
Mercado	Ventas Totales	Beneficio Total	Media de Tiempo de Envío					
Africa	\$781,891.21	\$89,154.13	3.459297984					
Europe	\$3,638,530.35	\$410,311.67	3.482847702					
LATAM	\$3,305,746.24	\$352,837.09	3.472104288					
Pacific Asia	\$2,663,236.22	\$288,825.10	3.530211368					
USCA	\$1,705,910.93	\$180,607.76	3.47963749					
Grand Total	\$12,095,314.95	\$1,321,735.75	3.488426821					
Tabla 3	Rendimiento por Modo de Envío (Shipping Mode)							
¿Envíos urgentes realmente tardan menos? ¿Son más rentables?								
Estado del Pedido COMPLETE								
Tipo de Envío	Ventas Totales	Beneficio Total	Media de Tiempo de Envío					
First Class	\$1,899,025.86	\$204,679.69	2					
Same Day	\$617,651.98	\$70,168.53	0.475347661					
Second Class	\$2,360,508.35	\$265,205.75	3.947147147					
Standard Class	\$7,218,128.77	\$781,681.78	3.99852962					
Grand Total	\$12,095,314.95	\$1,321,735.75	3.488426821					
Tabla 4	Rendimiento por Segmento de Cliente							
Comparar entre Consumidor, Corporate, Home Office... ¿Quién compra más y deja más beneficio?								
Order Status COMPLETE								
Tipo de Cliente	Ventas	Beneficio	Numero de Pedidos					
Row Labels	Sum of Sales	Sum of Order Profit Per	Count of Order Id					
Consumer	\$6,291,871.55	\$688,933.56	30842					
Corporate	\$3,656,925.94	\$411,358.62	18002					
Home Office	\$2,146,517.46	\$221,443.57	10647					
Grand Total	\$12,095,314.95	\$1,321,735.75	59491					

Tabla 16: Ventas, beneficio y tiempo medio según mercado, modo de envío y segmento de cliente.



Tabla 5 Número de Pedidos por Estado			
Analiza cuántos pedidos están completos, pendientes, cancelados...			
Estado del Pedido	Ventas	Beneficio	Numero de Pedidos
Row Labels	Sum of Sales	Sum of Benefit per order	Count of Order Id
COMPLETE	\$12,095,314.95	\$1,321,735.75	59491
Grand Total	\$12,095,314.95	\$1,321,735.75	59491
Tabla 6 Tasa de Retraso por Región (riesgo de entrega tardía)			
¿En qué regiones hay más retrasos? Si el valor se acerca a 1, hay muchos retrasos.			
Order Status	COMPLETE		
Región	Riesgo de Entrega Tardía		
Row Labels	Average of Late Delivery Risk		
Canada	0.546428571		
Caribbean	0.547512992		
Central Africa	0.568994889		
Central America	0.562075289		
Central Asia	0.602150538		
East Africa	0.555379747		
East of USA	0.603906899		
Eastern Asia	0.57010582		
Eastern Europe	0.588191882		
North Africa	0.565777369		
Northern Europe	0.579059181		
Oceania	0.56786134		
South America	0.545397226		
South Asia	0.6		
South of USA	0.61014383		
Southeast Asia	0.600921962		
Southern Africa	0.579462103		
Southern Europe	0.571102414		
US Center	0.588235294		
West Africa	0.542242704		
West Asia	0.577962578		
West of USA	0.573964497		
Western Europe	0.588867734		
Grand Total	0.574860063		

Tabla 17: Estado de los pedidos y riesgo de entrega tardía por región.



Tabla 7	Tiempo Medio de Envío por Estado de Pedido
Muestra si los pedidos completados tardan más o menos que los pendientes, cancelados, etc.	
Estado de pedido Tiempo de Envío	
Row Labels	Average of Days for shipping (real)
COMPLETE	3.488426821
Grand Total	3.488426821

Tabla 18: Tiempo medio de envío por estado del pedido.

4 Filtros que aplican a todas las tablas dinámicas:

The screenshot shows five dynamic filters applied across different tables in Power BI:

- Order Status:** Filtered to show only **COMPLETE**.
- Customer Segment:** Filtered to show **Consumer**, **Corporate**, and **Home Office**.
- Shipping Mode:** Filtered to show **First Class**, **Same Day**, **Second Class**, and **Standard Class**.
- Order Region:** Filtered to show all regions.
- Product Name:** Filtered to show **adidas Men's Germany Black Crest Away Tee**, **adidas Youth Germany Black/Red Away Match Soc**, **Adult dog supplies**, **Baby sweater**, **Bag Boy Beverage Holder**, **Bag Boy M330 Push Cart**, **Bowflex SelectTech 1090 Dumbbells** (highlighted in yellow), and **Bridgestone e6 Straight Distance NFL Carolina**.

Ilustración 23: Segmentadores y filtros de tablas dinámicas del análisis descriptivo.

Segmentador	Utilidad
Order Status	Filtrar y analiza solo los pedidos completados (muy importante para calcular métricas fiables).
Customer Segment	Te permite analizar cómo cambian las ventas, beneficios y tiempos de envío según el tipo de cliente.
Shipping Mode	Compara si los modos de envío impactan en los tiempos reales de envío, tasa de retrasos y ventas totales.
Order Region	Te permite analizar por zona geográfica qué regiones tienen mejores ventas, tiempos de envío más rápidos o lentos y más pedidos completados o cancelados.
Product Name	Sirve para ver si hay productos o categorías que venden más, generan más beneficio, tardan más en enviarse y tienen más riesgo de retraso.

Tabla 19: Utilidad de los segmentadores aplicados en el análisis descriptivo.

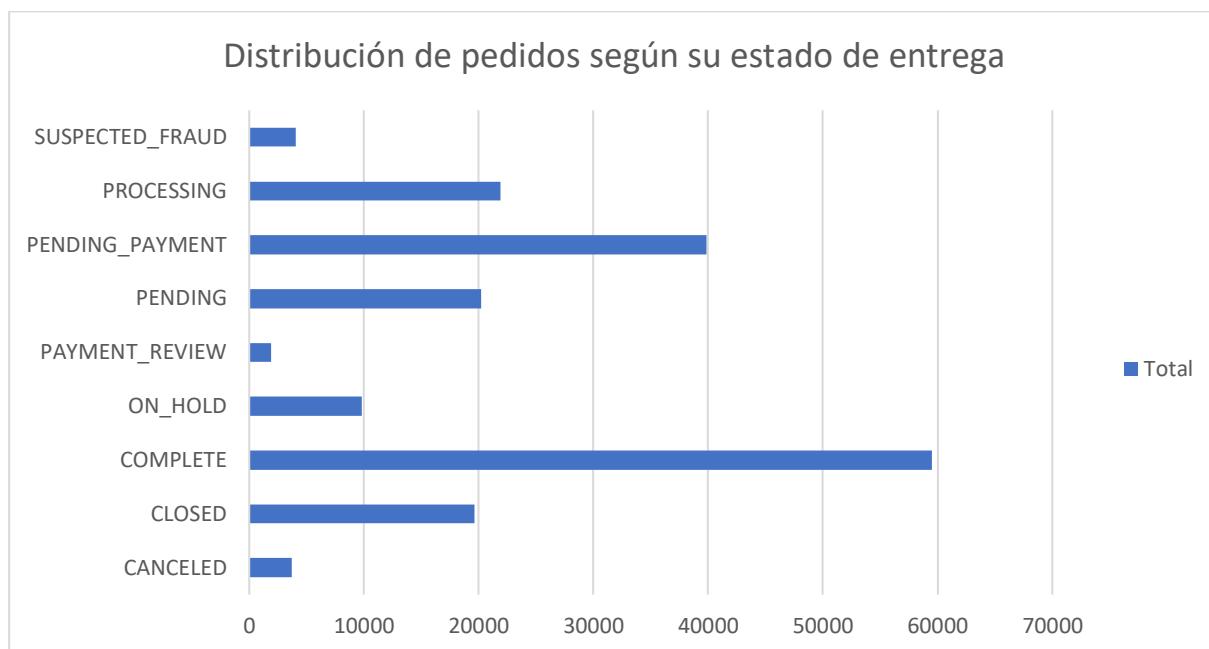


Ilustración 24: Distribución de pedidos según su estado de entrega.

5.4 Correlaciones

1 Objetivo:			
Mostrar si existe relación estadística entre dos o más variables clave del dataset.			
2 Selección de variables a correlacionar elegidas:			
Sales	=	Sales	
Order Profit per Order	=	Profit	
Days for Shipping (REAL)	=	Shipping Days	
Order Item Discount	=	Discount	
Order Item Discount Rate	=	Discount Rate	
Order Item Quantity	=	Quantity	
Late Delivery Risk	=	Risk	
3 Justificación:			
Sales	↔	Profit	Para saber si las ventas elevadas realmente generan beneficios.
Shipping Days	↔	Late Delivery Risk	Para validar si realmente el riesgo de retraso se traduce en tiempos mayores.
Discount Rate	↔	Profit	Para evaluar si dar descuentos impacta negativamente en la rentabilidad.
Quantity	↔	Sales	Para ver si pedidos grandes generan más ventas.

Ilustración 25: Objetivo, selección y justificación de variables para el análisis de correlaciones.



5 Como sacar conclusiones:			
Cercano a	+1	=	Correlación positiva fuerte.
Cercano a	0	=	Sin correlación.
Cercano a	-1	=	Correlación negativa fuerte.

Ilustración 26: Criterios de interpretación de coeficientes de correlación.

7 Pruebas de Normalidad de Variables Seleccionadas:					
Variables	Razón de Uso				
Sales	Variable numérica clave de rendimiento.				
Profit	Variable objetivo financiera.				
Shipping Days	Variable numérica operativa.				
Discount	Factor de marketing/comercial clave.				
Quantity	Relacionada con tamaño de pedido, importante para ventas.				
Interpretación de Valores:					
Estadístico	Rango aceptable (normalidad)				
Asimetría	Entre -1 y +1 (aceptable), ideal -0,5 a +0,5.				
Curtosis	Entre -2 y +2 (aceptable), aunque algunos autores permiten hasta -3 a +3				
Tabla de Normalidad:					
Variables	Nombre Original de Variables	Asimetría	Curtosis	Normalidad	Comentarios
Sales	Sales	2.884249049	23.93656127	No	Asimetría y curtosis excesivamente altas (no normal)
Profit	Order Profit per Order	-4.74183407	71.37725866	No	Asimetría y curtosis extremas (no normal)
Shipping Days	Days for Shipping (REAL)	0.084771273	-1.007913583	Sí	Dentro de rango aceptable
Discount	Order Item Discount	3.039795514	25.23126719	No	Asimetría y curtosis altas (no normal)
Quantity	Order Item Quantity	0.880251848	-0.753701577	Sí	Dentro de rango aceptable
Resumen de los Resultados:					
Shipping Days y Quantity:	se consideran aproximadamente normales.				
Sales, Profit y Discount:	no siguen una distribución normal (asimetría y curtosis muy elevadas).				

Ilustración 27: Pasos sobre las pruebas de normalidad de las variables seleccionadas.

5.5 Regresiones

1 Objetivo:	
Realizar un análisis exhaustivo sobre las distintas variables sobre:	
- Order Profit Per Order	=
- Sales	=
Para extraer conclusiones estratégicas claras para la empresa logística.	
2 Modelos a aplicar:	
- Lineal Simple:	para estudiar relaciones directas.
- Multiple:	para identificar efectos conjuntos entre múltiples factores.

Ilustración 28: Objetivo y diseño de los modelos de regresión.

**4 Interpretacion General de los Modelos de Regresion:**

- Coeficiente: Impacto de las variables independientes sobre la dependiente.
- R²: Porcentaje de la variación explicada por el modelo.
- P-valor: Nivel de significación estadística (ideal < 0.05).

6 Tabla Resumen:

Modelo	Y (dependiente)	X (independientes)	R ²	Variable mas influyente
1	Profit	Sales	0.022240806	Sales
2	Profit	Profit Ratio, Discount, Quantity	0.680018887	Profit Ratio
3	Sales	Quantity, Days for Shipping, Profit Ratio	0.01137054	Quantity

7 Conclusión Final:

Los modelos construidos permiten entender cómo las ventas, descuentos, cantidades y márgenes afectan directamente al beneficio y facturación de la empresa. La regresión múltiple (modelo 2) destaca por su alta capacidad predictiva del beneficio, mientras que la cantidad y riesgo son claves para las ventas. Estas regresiones aportan un soporte cuantitativo clave para definir estrategias de precios, promociones y logística eficiente.

Ilustración 29: Interpretación, resultados y conclusión del análisis de regresión.

5.6 Anova

1	Objetivo:	
	- Aplicar el Análisis de Varianza (ANOVA).	
	- Identificar si existen diferencias estadísticamente significativas en el beneficio medio por pedido en función de variables categóricas clave del entorno logístico.	
	- Vincular datos económicos con decisiones estratégicas reales.	
2	Relevancia de ANOVA:	
	- Compara medias entre tres o más grupos independientes y verificar si las diferencias son estadísticamente significativas.	
	- Ayuda a comprender si segmentar a los clientes, regiones o tipos de envío genera un cambio real en la rentabilidad.	
	- Interpretación Clave: Si el p-valor < 0.05, la variable categórica sí influye en la rentabilidad del pedido.	
3	Hipótesis Estadísticas Planteadas:	
	1 Customer Segment + Shipping Mode: H0: (Hipótesis Nula) H1: (Hipótesis Alternativa)	La combinación del tipo de cliente y el modo de envío no influye significativamente en el beneficio medio por pedido. La combinación del tipo de cliente y el modo de envío sí influye significativamente en el beneficio medio por pedido.
	2 Shipping Mode + Days for Shipping (REAL): H0: (Hipótesis Nula) H1: (Hipótesis Alternativa)	La combinación del modo de envío y los días reales de entrega no afecta significativamente al beneficio medio por pedido. La combinación del modo de envío y los días reales de entrega afecta significativamente al beneficio medio por pedido.
	3 Order Status: H0: (Hipótesis Nula) H1: (Hipótesis Alternativa)	El estado del pedido no tiene un efecto significativo sobre el beneficio medio por pedido. El estado del pedido tiene un efecto significativo sobre el beneficio medio por pedido.

Ilustración 30: Planteamiento e hipótesis del análisis ANOVA.



5.7 Series Temporales

1 Objetivo:	
- Identificar tendencias y estacionalidades en beneficio, ventas y pedidos a lo largo del tiempo.	
- Detectar meses críticos para ajustar la planificación logística y comercial.	
- Evaluar el impacto del riesgo de retraso y el modo de envío sobre los resultados.	
- Facilitar decisiones estratégicas mediante gráficos claros y métricas comparables.	
- Detectar meses críticos para ajustar la planificación logística y comercial.	
2 Variables a Analizar:	
Fecha del pedido	= Order Date
Beneficio por pedido	= Order Profit Per Order
Ventas por pedido	= Sales
Cantidad por pedido	= Order Item Quantity
Riesgo de Retraso	= Late Delivery Risk
Modo de Envío	= Shipping Mode
Número total de pedidos diarios	

Ilustración 31: Objetivos y variables del análisis de series temporales.

Row Labels	Sum of Sales	Count of Order Id	Sum of Order Item Quantity	Average of Late Delivery Risk	Average of Order Item Discount Rate	Average of Order Profit Per Order	Sum of Order Profit Per Order
2015-01	1051590.08	5322	11854	0.541149944	0.101148065	20.9809734	111660.7401
2015-02	927009.90	4729	10438	0.548530345	0.101621908	20.96440266	99140.6602
2015-03	1051253.69	5362	12062	0.547556882	0.101848191	21.21936035	113778.2102
2015-04	1014463.28	5126	11287	0.538431526	0.101857199	21.08538431	108083.68
2015-05	1050478.44	5357	11902	0.550868023	0.101207766	20.93483296	112147.9001
2015-06	1024006.17	5134	11203	0.541098559	0.101468085	21.45445273	110147.1603
2015-07	1038081.19	5299	11800	0.55463295	0.101530478	21.81997733	115624.0599
2015-08	1029494.69	5273	11612	0.556798786	0.102144889	22.37431639	117979.7703
2015-09	1018338.60	5140	11366	0.56692607	0.102416343	22.07547473	113467.9401
2015-10	1049154.27	5302	11703	0.548090508	0.100999623	19.19235572	101757.87
2015-11	1029120.24	5235	11463	0.54269341	0.101593124	19.94791979	104427.3601
2015-12	1057840.88	5371	11790	0.546825545	0.102208155	20.59980448	110641.5499
2016-01	1046308.25	5317	11597	0.555764529	0.101352267	20.08293215	106780.9502
2016-02	968542.85	4894	10765	0.541479362	0.101847416	17.73794251	86809.49064
2016-03	1025853.12	5210	11349	0.557581574	0.101790787	19.4356987	101259.9902
2016-04	1001211.58	5097	11208	0.552285658	0.101491074	22.55770451	114976.6199
2016-05	1029400.20	5302	11603	0.556582422	0.101312713	20.881954	110716.1201
2016-06	1003059.55	5054	11008	0.568460625	0.101517161	20.61655523	104196.0701
2016-07	1045715.62	5305	11652	0.519321395	0.101918945	22.60918382	119941.7202
2016-08	1048200.25	5334	11683	0.552680915	0.101471692	20.75919387	110729.5401
2016-09	1002397.04	5160	11284	0.549806202	0.101922481	23.9117655	123384.71
2016-10	1048130.55	5398	11936	0.54779548	0.101517229	21.63430345	116781.97
2016-11	1047590.14	5210	11493	0.546257198	0.101831095	21.02479846	109539.2
2016-12	1037408.17	5269	11774	0.560637692	0.10210968	19.92839057	105002.6899
2017-01	1029698.02	5217	11605	0.536131877	0.101433775	22.01338897	114843.8503
2017-02	992534.90	4906	11070	0.546677538	0.101536894	23.46332452	115111.0701
2017-03	1048004.78	5347	11676	0.552646344	0.101234338	22.19728824	118688.9002
2017-04	1038321.62	5212	11189	0.537797391	0.102334997	22.24886036	115961.0602
2017-05	1105485.32	5317	11033	0.534700019	0.101986083	21.63149145	115014.64
2017-06	1032086.49	4951	10194	0.535043426	0.101341144	22.29838221	110399.2903
2017-07	1104373.36	5318	11091	0.547762317	0.101908613	21.25361039	113026.7
2017-08	1109337.17	5305	11095	0.564373233	0.10199623	24.78815461	131501.1602
2017-09	1143775.11	5189	10502	0.542301021	0.101416458	23.60038353	122462.3902
2017-10	1073994.17	2255	2490	0.536585366	0.100620843	50.30916625	113447.1699
2017-11	626914.38	2055	2055	0.550851582	0.101498784	32.98844292	67791.2502
2017-12	503910.82	2124	2124	0.551789077	0.101859699	30.99700082	65867.62974
2018-01	331650.12	2123	2123	0.562882713	0.102152615	15.9405982	33841.8898
Grand Total	36784735.01	180519	384079	0.548291316	0.101668191	21.97498864	3966902.974

Tabla 20: Indicadores mensuales de ventas, beneficio y logística.

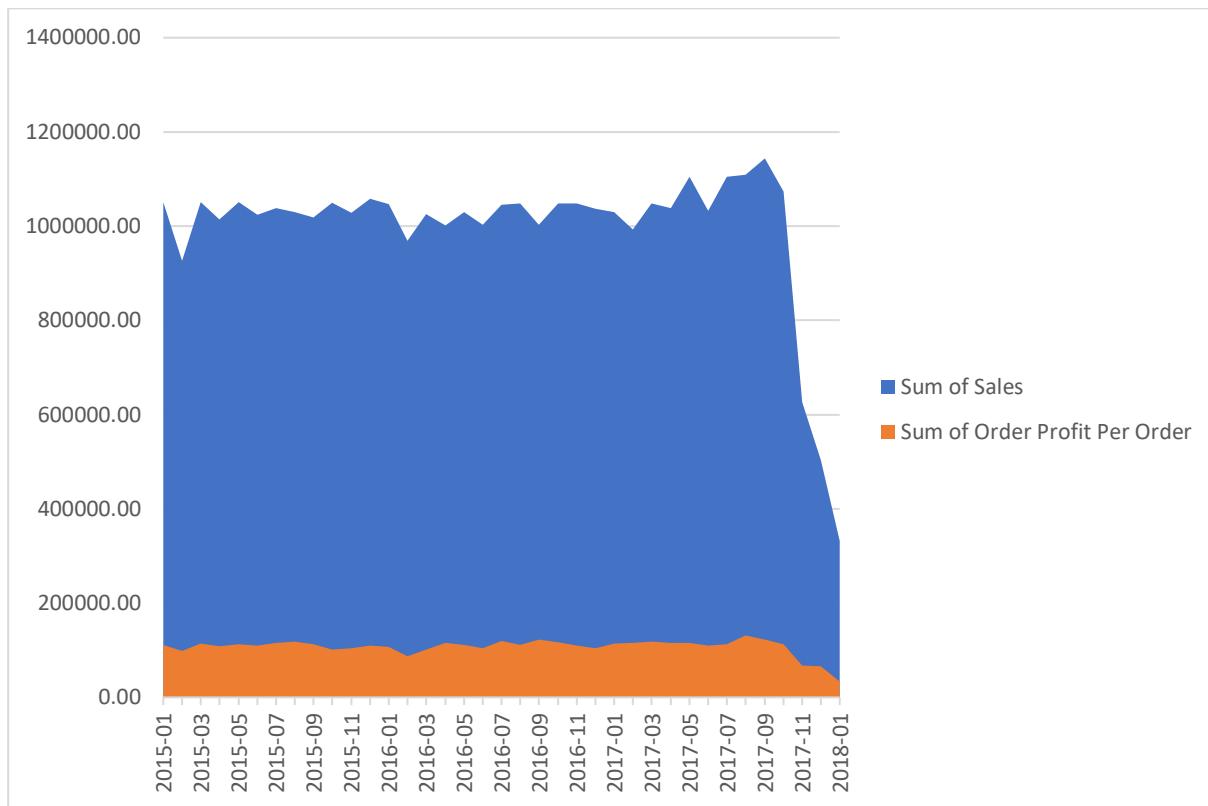


Ilustración 32: Evolución temporal de ventas y beneficio por pedido.

Row Labels	Resumen Trimestral de Métricas Clave						
	Beneficio Total	Ventas Totales	Pedidos	Beneficio Medio	Pedido M	Riesgo Medio	
2015-Q1	324579.6105	3029853.67		15413	21.05882116	196.58	0.54564329
2015-Q2	330378.7404	3088947.89		15617	21.15507078	197.79	0.54357431
2015-Q3	347071.7703	3085914.48		15712	22.08959842	196.40	0.55938136
2015-Q4	316826.78	3136115.39		15908	19.91619186	197.14	0.54588886
2016-Q1	294850.4311	3040704.22		15421	19.12005908	197.18	0.55184489
2016-Q2	329888.8101	3033671.33		15453	21.34788133	196.32	0.55905002
2016-Q3	354055.9703	3096312.91		15799	22.41002407	195.98	0.54054054
2016-Q4	331323.8599	3133128.86		15877	20.86816527	197.34	0.55155256
2017-Q1	348643.8206	3070237.70		15470	22.53676927	198.46	0.54518423
2017-Q2	341374.9905	3175893.43		15480	22.05264797	205.16	0.53585271
2017-Q3	366990.2504	3357485.64		15812	23.20960349	212.34	0.55154313
2017-Q4	247076.0498	2204819.37		6434	38.40162416	342.68	0.54616102
2018-Q1	33841.88998	331650.12		2123	15.9405982	156.22	0.56288271
Grand Total	3966902.974	36784735.01		180519	21.97498864	203.77	0.54829132

Tabla 21: Indicadores trimestrales de ventas, beneficio y riesgo.

Análisis por Modo de Envío																					
2015-Q4			2016-Q1			2016-Q2			2016-Q3			2016-Q4			2017-Q1			2017-Q2			
Beneficio Medio	Pedidos	Riesgo Medio	Beneficio Medio	Pedidos	Riesgo Medio	Beneficio Medio	Pedidos	Riesgo Medio	Beneficio Medio	Pedidos	Riesgo Medio	Beneficio Medio	Pedidos	Riesgo Medio	Beneficio Medio	Pedidos	Riesgo Medio	Beneficio Medio	Pedidos	Riesgo Medio	
1 20.91619186	2517	0.54564329	18.72347105	2348	0.94548502	17.86762501	2357	0.98844591	24.42341114	2486	0.94931671	22.38919262	2601	0.960313456	25.79749116	2248	0.951067616	23.59302017	2300	0.950000007	22.84362018
2 17.39623897	2517	0.54564329	20.46402017	2630	0.94548502	19.70046409	2647	0.98844591	24.42341114	2777	0.94931671	22.38919262	2601	0.960313456	25.79749116	2248	0.951067616	23.59302017	2300	0.950000007	22.84362018
3 19.386723301	3043	0.771609867	15.91391534	3019	0.709460986	21.15650211	3189	0.73394481	23.39395378	3064	0.746653451	19.70046409	3094	0.767046879	23.555177986	3076	0.767046879	23.555177986	3088	0.729164023	23.80946704
8 19.06992511	9484	0.376107128	20.18890873	9218	0.390865698	22.48742221	9082	0.384793644	21.23366434	21.45091687	9445	0.381471678	21.66653451	9381	0.382475216	21.86014221	9411	0.380830943	23.37311067	9265	0.37128988
5 19.91619186	15008	0.54564329	19.12005908	15421	0.551844807	21.34788133	15453	0.559050023	22.41002407	15799	0.55155256	22.53676927	15470	0.545184228	22.05264797	15480	0.5358652713	23.20960349	15812	0.551543132	

Tabla 22: Evolución del beneficio, pedidos y riesgo por tipo de envío.

**8 Conclusión Final:**

El análisis temporal revela patrones claros en beneficio, ventas y riesgo logístico. Los trimestres Q3 destacan por su rentabilidad, mientras que inicios de año exigen atención operativa. El modo Standard Class domina en volumen con bajo riesgo, y First Class maximiza beneficio unitario. La proyección futura mediante regresión añade un enfoque predictivo clave para decisiones estratégicas en la cadena de suministro.

Ilustración 33: Conclusión estratégica del análisis temporal del beneficio, ventas y riesgo.

5.8 Repositorio en GitHub

Para facilitar el acceso a todos los archivos utilizados en este Trabajo de Fin de Grado y permitir su revisión o replicación, se ha creado un repositorio público en GitHub con todo el material empleado en el análisis.

Enlace al repositorio:

<https://github.com/imartinpena/TFG-Logistics-DataAnalysis-Excel>

En él se incluyen:

- El conjunto de datos limpio.
- Todos los archivos de análisis estadístico (descriptivo, correlaciones, regresiones, ANOVA y series temporales).
- La memoria completa del TFG en PDF.
- Los datos originales descargados desde Kaggle.

Un archivo README que explica qué contiene cada archivo y cómo se ha realizado el análisis.

De esta forma, cualquier lector puede consultar el trabajo completo y, si lo desea, replicar los análisis siguiendo el mismo procedimiento.