

34

LA REGRESIÓN LINEAL

**EN EL PROCESO DE VENTAS Y SU INFLUENCIA EN LA
COMPETITIVIDAD ORGANIZACIONAL EN ENTORNOS BIG
DATA**



© 2025; Los autores. Este es un artículo en acceso abierto, distribuido bajo los términos de una licencia Creative Commons que permite el uso, distribución y reproducción en cualquier medio siempre que la obra original sea correctamente citada.

LA REGRESIÓN LINEAL

EN EL PROCESO DE VENTAS Y SU INFLUENCIA EN LA COMPETITIVIDAD ORGANIZACIONAL EN ENTORNOS BIG DATA

LINEAR REGRESSION IN THE SALES PROCESS AND ITS INFLUENCE ON ORGANIZATIONAL COMPETITIVENESS IN BIG DATA ENVIRONMENTS

Bolívar Solano-Morales¹

E-mail: bolivarenrique@hotmail.com

ORCID: <https://orcid.org/0009-0007-9641-3561>

¹ Instituto Superior Tecnológico Bolivariano de Tecnología. Ecuador.

Cita sugerida (APA, séptima edición)

Solano-Morales, B. (2025). La regresión lineal en el proceso de ventas y su influencia en la competitividad organizacional en entornos Big Data. *Revista Metropolitana de Ciencias Aplicadas*, 8(4), 324-338.

Fecha de presentación: 09/07/2025

Fecha de aceptación: 14/08/2025

Fecha de publicación: 01/10/25

RESUMEN

En el competitivo sector del retail de moda, la optimización de las ventas mediante el uso de Big Data se ha convertido en un elemento esencial para sostener la competitividad. En este estudio se aplicaron modelos de regresión lineal con el propósito de analizar la influencia de variables clave en el rendimiento de las tiendas de ropa, empleando datos simulados procesados en Apache Spark. A partir de un proceso de eliminación hacia atrás, el modelo inicial fue depurado y se identificó que la tasa de conversión y la satisfacción con el precio guardan una correlación positiva con la competitividad, mientras que el precio competitivo muestra una relación negativa. Estos hallazgos revelan que, más allá de la reducción de precios, los factores vinculados a la conversión del interés en compra y la percepción de satisfacción resultan determinantes en la propuesta de valor de las organizaciones. En consecuencia, la regresión lineal se presenta como una herramienta analítica valiosa para orientar decisiones estratégicas que fortalezcan la competitividad en el retail de moda en un entorno dominado por el Big Data.

Palabras clave:

Regresión lineal, competitividad organizacional, comercio minorista de moda, análisis de Big Data, proceso de ventas.

ABSTRACT

In the competitive fashion retail sector, sales optimization through the use of Big Data has become an essential element for sustaining competitiveness. This study applied linear regression models to analyze the influence of key variables on the performance of clothing stores, using simulated data processed in Apache Spark. Through a backward elimination process, the initial model was refined, and it was identified that conversion rate and price satisfaction are positively correlated with competitiveness, while competitive pricing shows a negative relationship. These findings reveal that beyond price reduction, factors related to the conversion of interest into purchase and the perception of satisfaction are decisive in shaping the value proposition of organizations. Consequently, linear regression emerges as a valuable analytical tool to guide strategic decisions that strengthen competitiveness in fashion retail within a Big Data-driven environment.

Keywords:

Linear regression, organizational competitiveness, fashion retail, Big Data analysis, sales process.

INTRODUCCIÓN

En el competitivo sector del retail de moda, la optimización de los procesos de venta mediante el análisis de grandes volúmenes de datos (Big Data) se ha vuelto crucial para la competitividad organizacional (Shankar, 2019), utilizando datasets extensos como el simulado de 10,000 registros de tiendas. La capacidad de extraer conocimiento valioso de diversas fuentes de datos presenta una oportunidad para mejorar la toma de decisiones estratégicas, aunque el desafío radica en analizar esta información de manera efectiva (Verhagen et al., 2020). Esta investigación se centra en la regresión lineal como herramienta analítica para cuantificar la influencia de variables operativas clave –efectividad comunicacional (Canales_Promedio, Tasa_Conversion, Tiempo_Seguimiento), estrategia de precios (Precio_Competitivo, Satisfaccion_Precio, Tasa_Recompra) y eficiencia operativa (Rotacion_Inventario, Tiempo_Reposicion y Merma)– sobre la competitividad de las tiendas de ropa en un entorno Big Data, abordando así el problema de conocimiento sobre cómo maximizar el potencial de estos datos para obtener una ventaja competitiva.

El estudio tiene como objetivo general investigar cómo la aplicación de la regresión lineal, en el contexto del análisis de Big Data, influye en la competitividad organizacional de las tiendas de ropa. Para lograrlo, se analizará la relación lineal entre las variables mencionadas, se implementará una selección de variables backward con Apache Spark, y se evaluarán los supuestos del modelo resultante mediante pruebas estadísticas y Figuras de residuos. Además, se calcularán coeficientes estandarizados para identificar las variables con mayor impacto en la competitividad, lo que se relaciona con el principio de Pareto o regla 80/20, generando insights para la toma de decisiones estratégicas. Metodológicamente, se adoptará un enfoque cuantitativo y aplicado, utilizando datos simulados procesados con Apache Spark (Varma, 2022). La regresión lineal se fundamenta en la teoría de la ventaja competitiva (Kumar et al., 2011), permitiendo modelar las relaciones entre variables y proporcionar una base empírica para la estrategia empresarial.

Para optimizar el proceso de ventas en el competitivo retail de moda mediante el análisis de Big Data y la regresión lineal, es fundamental comprender inicialmente los pilares de un proceso de venta efectivo: comunicación, precios e inventario (Firework Staff, 2024). Este artículo explorará cómo la aplicación de la regresión lineal en entornos de Big Data permite analizar estos conceptos para fortalecer la competitividad organizacional. A *continuación*, se presenta una revisión concisa de los conceptos clave que sustentan esta investigación, abarcando tanto el proceso de ventas, los fundamentos de la regresión lineal como las herramientas de procesamiento de Big Data.

Para un proceso de venta competitivo en tiendas de ropa, es crucial enfocarse en una comunicación efectiva que

conecte con el cliente y genere lealtad a través de interacciones personalizadas y un seguimiento post-venta atento (Firework Staff, 2024); en una estrategia de precios justos que los clientes perciban como valiosos en relación con la calidad y las alternativas (Serrano et al., 2025); fomentando así la satisfacción y la recompra; y en una gestión de inventario ágil que asegure la disponibilidad de los productos deseados en el momento oportuno, optimizando la experiencia de compra y evitando la pérdida de ventas. La integración de estos tres pilares en cada interacción de venta no solo facilita la transacción inmediata, sino que también construye relaciones duraderas y fortalece la posición de la tienda en el mercado actual.

En el análisis estadístico, la regresión lineal es un método para modelar la relación lineal entre una variable dependiente y una única variable independiente. El objetivo es encontrar la línea recta que mejor se ajuste a los datos observados, permitiendo predecir el valor de la variable dependiente basándose en el valor de la variable independiente (Anderson et al., 2024). Esta línea se define por un intercepto, que es el valor esperado de la variable dependiente cuando la variable independiente es cero, y un coeficiente, que indica el cambio en la variable dependiente por cada unidad de cambio en la variable independiente.

Cuando se utilizan dos o más variables independientes para predecir o explicar la variable dependiente, el método se conoce como regresión lineal múltiple (Wooldridge, 2020). En este caso, el modelo incluye un intercepto y un coeficiente para cada variable independiente, representando el cambio en la variable dependiente asociado con un cambio unitario en esa variable independiente, manteniendo constantes las demás. La significancia estadística de estos coeficientes se evalúa mediante el p-valor, donde un valor bajo (generalmente < 0.05) sugiere que la variable independiente tiene un efecto significativo en la variable dependiente. La proporción de la varianza en la variable dependiente explicada por el modelo se mide con el R-cuadrado, mientras que el Error Cuadrático Medio (RMSE) cuantifica la magnitud de los errores de predicción (Gujarati, 2023). La técnica de Backward Elimination puede utilizarse para seleccionar las variables más relevantes para el modelo basándose en su significancia estadística.

Para asegurar la validez del modelo lineal (ya sea simple o múltiple), se deben cumplir varios supuestos del modelo lineal, incluyendo la linealidad de la relación entre las variables, la independencia de los errores, la homocedasticidad (varianza constante de los errores) y la normalidad de los errores. Los residuos, que son las diferencias entre los valores observados y predichos, son fundamentales para verificar estos supuestos a través de Figuras de diagnóstico como Residuals vs Fitted, Q-Q Normal, Scale-Location y Residuals vs Leverage, así como pruebas estadísticas como la Prueba de Shapiro-Wilk para

evaluar la normalidad de los errores. La presencia de multicolinealidad, una alta correlación entre variables independientes en la regresión múltiple, puede afectar la interpretación de los coeficientes. Un modelo que se ajusta demasiado a los datos de entrenamiento puede sufrir de sobreajuste, generalizando pobremente a nuevos datos. La significancia global del modelo se evalúa con el Estadístico F. La Correlación de Pearson mide la relación lineal entre pares de variables, presentándose en una Matriz de Correlación, y es crucial para detectar la multicolinealidad (Anderson et al., 2024). La violación de la homocedasticidad lleva a la heterocedasticidad, y la evaluación de todos estos aspectos es vital para la construcción de un modelo de regresión robusto y confiable. La varianza es una medida general de la dispersión de los datos, útil para entender la variabilidad de la variable dependiente y los residuos.

En el entorno del Big Data Processing, herramientas como Apache Spark, un framework de computación distribuida de código abierto, son fundamentales para el procesamiento eficiente de datos a gran escala en clústeres de Apache Spark. Para usuarios de R, el paquete sparklyr proporciona una interfaz intuitiva para interactuar con Spark, permitiendo aprovechar sus capacidades de procesamiento distribuido. Una Conexión Local a Apache Spark facilita la simulación de un entorno distribuido para pruebas o análisis de conjuntos de datos no masivos. El Procesamiento Distribuido de Datos implica dividir grandes conjuntos de datos y tareas en múltiples nodos para acelerar el análisis. Los datos en Spark se organizan en Figuras en Spark, estructuras similares a los Data Frame de R. La Ingesta de Datos es el proceso de cargar información desde diversas fuentes a Spark, y la función `copy_to` de sparklyr permite transferir un data frame de R a Spark, creando una Figura Spark equivalente. Para el modelado, las variables independientes se transforman en un Vector de Características, que sirve como entrada a algoritmos de machine learning en Spark, y la función `ml_predict` de sparklyr se utiliza para obtener predicciones de modelos entrenados (Potla, 2022). Una Sesión de Spark es la instancia de la aplicación que permite la interacción con los datos, y los Logs de Spark registran las actividades para monitoreo y depuración.

Modelos Analíticos para la Competitividad en la Industria de la Moda

En el dinámico y altamente competitivo sector de la moda, las empresas buscan constantemente estrategias innovadoras para diferenciarse y prosperar. La adopción de modelos analíticos avanzados emerge como un factor crucial para obtener una ventaja sostenible, permitiendo una comprensión profunda del cliente, la optimización de las operaciones y la anticipación de las tendencias del mercado (Gong, 2024).

Personalización de la Experiencia del Cliente: Este modelo, fundamentado en el análisis exhaustivo del comportamiento de compra, preferencias individuales y datos demográficos, permite a las empresas de moda ofrecer interacciones y ofertas altamente relevantes. A través de sistemas de recomendación personalizados, correos electrónicos dirigidos y experiencias de compra adaptadas, se fomenta la lealtad del cliente y se incrementa el valor del ciclo de vida del cliente (Liu & Ding, 2022). La implementación de algoritmos de segmentación y aprendizaje automático permite identificar patrones ocultos en los datos del cliente, lo que resulta en estrategias de marketing más efectivas y una mayor satisfacción del consumidor.

Optimización de Precios Dinámicos: La implementación de estrategias de precios dinámicos posibilita el ajuste en tiempo real de los precios de las prendas y accesorios en función de variables como la demanda actual, los precios de la competencia, la estacionalidad y el inventario disponible (Polacek et al., 2024). Mediante el uso de algoritmos de optimización y modelos econométricos, las empresas pueden maximizar sus ingresos y rentabilidad, respondiendo ágilmente a las fluctuaciones del mercado y liquidando el inventario de manera eficiente. Esta estrategia requiere una monitorización constante de los datos del mercado y una capacidad de respuesta rápida para ajustar los precios de manera efectiva.

Optimización de la Cadena de Suministro: La eficiencia en la cadena de suministro es un pilar fundamental para la competitividad en la industria de la moda. La predicción precisa de la demanda, basada en el análisis de datos históricos de ventas, tendencias de la moda y factores externos, permite una gestión del inventario más eficiente, reduciendo costos de almacenamiento y minimizando el riesgo de obsolescencia (Dolgui et al., 2018). La implementación de modelos de optimización y sistemas de planificación de recursos empresariales (ERP) avanzados facilita la coordinación entre proveedores, fabricantes y distribuidores, mejorando la capacidad de respuesta ante las demandas del mercado y reduciendo los plazos de entrega (Akand, 2024).

Análisis de Tendencias de Moda y Predicción de la Demanda: Anticipar los gustos futuros de los consumidores es crucial en la industria de la moda, caracterizada por ciclos de vida de los productos cortos y cambios rápidos en las preferencias. El análisis de tendencias emergentes en redes sociales, blogs de moda, datos de búsqueda en línea y pasarelas, combinado con modelos de predicción de la demanda basados en aprendizaje automático y series temporales, permite a las empresas adaptar sus colecciones de manera proactiva (Shin et al., 2022). Esto minimiza el riesgo de inventario no vendido, fortalece la imagen de la marca como referente de moda y optimiza la planificación de la producción.

La Figura 1 muestra cómo el análisis de datos, un componente clave del Big Data, permite la selección de modelos para optimizar el proceso de ventas. La personalización de la experiencia del cliente, la optimización de precios dinámicos y el análisis de tendencias de moda y predicción de la demanda son aplicaciones directas del análisis de datos que impactan las variables de tu estudio (comunicación, precio, gestión de inventario) y, por ende, la competitividad organizacional. La regresión lineal sería uno de los modelos seleccionados para cuantificar estas relaciones.



Figura 1. Justificación de Modelos Analíticos.

Fuente: Elaboración propia con base en Liu & Ding, 2022; Shin (2022); y Ye (2024).

En la Figura 2 conceptualiza la Competitividad como resultado de la Comunicación, el Precio y el Inventario. Cada variable independiente se desglosa en dimensiones con indicadores medibles. Para tu tema, sugiere variables clave del proceso de ventas (como la tasa de conversión y la percepción del precio) que podrían influir linealmente en la competitividad. Los indicadores se convierten en variables para tu análisis de regresión lineal. En un entorno Big Data, estos indicadores pueden enriquecerse con datos masivos para comprender mejor su impacto en la competitividad organizacional.

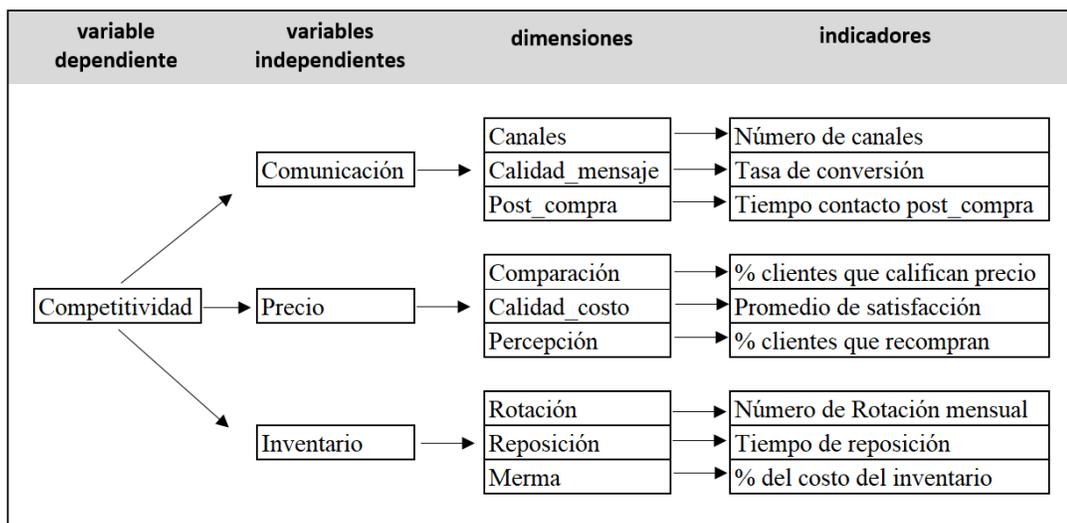


Figura 2. Resumen de las variables con sus dimensiones e indicadores.

Fuente: Elaboración propia con base en Liu & Ding, 2022; Shin (2022); y Ye (2024).

METODOLOGÍA

La presente investigación, de enfoque cuantitativo y alcance explicativo, busca elucidar la influencia de variables operativas clave en la competitividad organizacional de una tienda de ropa en un contexto de análisis de datos a gran escala (Big Data). Para ello, se cuantificará el impacto de la comunicación con el cliente (medida a través de Canales_Promedio, Tasa_Conversion y Tiempo_Seguimiento), la percepción del precio (medida a través de Precio_Competitivo, Satisfaccion_Precio y Tasa_Recompra) y la gestión de inventario (medida a través de Rotacion_Inventario, Tiempo_Reposicion y Merma) sobre la competitividad, evaluando su contribución a la toma de decisiones y eficiencia operativa.

Metodológicamente, se integrará un análisis cuantitativo robusto, complementado por una interpretación cualitativa de patrones y tendencias para una comprensión exhaustiva. El análisis cuantitativo construirá y validará modelos de regresión lineal con datos históricos para identificar y cuantificar la relación entre las variables independientes detalladas anteriormente y la competitividad, evaluando su ajuste y significancia estadística (Hair et al., 2022). Este análisis se llevará a cabo principalmente utilizando el Lenguaje de Programación R (Wickham et al., 2019) y diversas Librerías de R: `sparklyr` para la conexión y manipulación de datos en un entorno Big Data simulado con Apache Spark, `dplyr` para la transformación eficiente de datos, `ggplot2` para la visualización de resultados y la evaluación de supuestos, `lmtest` y `car` para la realización de pruebas de diagnóstico del modelo (homocedasticidad, multicolinealidad, autocorrelación), y `scales` para el cálculo de coeficientes estandarizados. Se priorizarán modelos parsimoniosos que expliquen la mayor varianza en la competitividad con el menor número de predictores significativos.

Para la construcción del modelo final, se empleará una estrategia de eliminación hacia atrás (*backward elimination*) implementada en R. Inicialmente, se ajustará un modelo completo incluyendo todas las variables independientes definidas. Posteriormente, se evaluará la significancia estadística de cada predictor. Las variables con los valores *p* más altos serán eliminadas iterativamente hasta que todas las variables restantes sean estadísticamente significativas (con un nivel alfa predefinido). Se calcularán los coeficientes estandarizados para evaluar la magnitud relativa del impacto de cada variable predictora significativa en la competitividad, facilitando la interpretación de la regla 80/20. Adicionalmente, se realizará una evaluación exhaustiva de los supuestos del modelo de regresión lineal utilizando pruebas estadísticas y visualizaciones en R.

Adicionalmente, para la fase de prueba y la presentación de resultados preliminares, se utilizará la herramienta Microsoft Excel (Alexander et al., 2018). Excel permitirá la realización de análisis de regresión lineal básicos, la creación de visualizaciones iniciales y la comprobación de la consistencia de los resultados obtenidos con R en una muestra más pequeña de los datos. De forma complementaria, se realizará un análisis cualitativo para interpretar patrones y tendencias en los datos históricos, buscando conexiones temporales, comparando segmentos de clientes y analizando el impacto de eventos significativos, triangulando indicadores internos para fortalecer la comprensión de la competitividad (Saldana, 2021). La documentación y presentación de los hallazgos se realizarán utilizando R Markdown, permitiendo la creación de informes dinámicos y reproducibles.

Respecto a la población y muestra, la población de este estudio hipotético está constituida por la totalidad de los datos generados por la tienda de ropa ubicada en un

centro comercial durante un periodo de un año. Estos datos hipotéticos incluyen transacciones de venta, registros de inventario, interacciones con clientes (si se simulan), y datos relacionados con actividades de marketing y comunicación implementadas durante ese año. Para la muestra cuantitativa, dada la naturaleza hipotética y el periodo de un año, se considerará la totalidad de los datos generados como la muestra para el análisis de regresión lineal, asumiendo una cobertura completa de las operaciones simuladas. El Lenguaje de Programación R permitirá gestionar eficientemente este volumen de datos para la construcción y evaluación de los modelos, potencialmente utilizando Librerías de R como `dplyr` para la manipulación previa de los datos. Microsoft Excel se utilizará en una submuestra para la verificación. En un escenario de Big Data, si el volumen de datos fuera extremadamente grande y requiriera procesamiento distribuido, se considerará la integración con el Programa de Procesamiento de Datos: Apache Spark a través de la Librería de R: `sparklyr`, facilitando la aplicación de modelos de regresión lineal (`ml_linear_regression`) en un entorno de procesamiento paralelo. Para la muestra cualitativa, se seleccionarán intencionalmente elementos informativos hipotéticos representativos del periodo de un año. Esto podría incluir comentarios de clientes simulados, informes de gestión internos hipotéticos relevantes para diversas etapas del año, y datos relacionados con eventos o promociones específicas simuladas en diferentes momentos del año.

En cuanto a la recolección de información, se adoptará una metodología mixta, aprovechando la base de datos existente de la tienda (Herramienta de Almacenamiento y Organización de Datos: Microsoft Excel, aunque en un contexto real podría ser una base de datos más robusta). Cuantitativamente, la fuente principal será la base de datos (ventas, precios, inventario, interacciones), aplicando técnicas de extracción, depuración, análisis descriptivo y modelado de regresión lineal para cuantificar las relaciones, validando los modelos con métricas de ajuste y pruebas de significancia (Hair et al., 2022). El uso del Lenguaje de Programación R optimizará este proceso, permitiendo la automatización de tareas, la aplicación de algoritmos de regresión avanzados y la generación de Figuras para la interpretación de los resultados, posiblemente utilizando Librerías de R como `ggplot2` para la visualización. Microsoft Excel complementará en la fase de prueba con análisis básicos y visualizaciones iniciales. Cualitativamente, las fuentes primarias incluirán registros de comentarios, informes de gestión y observación no participante, mientras que las secundarias serán estudios de mercado y publicaciones académicas; las técnicas se centrarán en el análisis de contenido, patrones temporales, comparación de segmentos, impacto de eventos y triangulación para una comprensión profunda de la competitividad (Saldana, 2021). La creación de informes detallados y reproducibles se realizará utilizando R Markdown.

A través de la lente de la regresión lineal, una poderosa herramienta estadística implementada eficientemente en el Lenguaje de Programación R (utilizando paquetes base y librerías como lm para el modelado tradicional y sparklyr para la implementación en un entorno Big Data simulado con Apache Spark), se busca cuantificar cómo la comunicación con el cliente, las estrategias de precios y la eficiencia en la gestión del inventario influyen directamente en la competitividad organizacional. El análisis se fundamenta en la visualización inicial de las relaciones individuales entre estas variables y la competitividad, posiblemente utilizando las capacidades gráficas del Lenguaje de Programación R (con Librerías de R como ggplot2), seguido por el ajuste y diagnóstico de un modelo de regresión lineal (Rawling et al., 1998). El Lenguaje de Programación R proporcionará las herramientas necesarias para realizar estos análisis, incluyendo la evaluación de supuestos del modelo (utilizando librerías como lmtest para pruebas de homocedasticidad y autocorrelación, y visualizaciones con ggplot2), la detección de multicolinealidad (con la librería car) y la identificación de valores atípicos. La aplicación de la eliminación hacia atrás en R permitirá refinar el modelo, asegurando que solo las variables predictoras más relevantes y significativas se incluyan en la explicación de la competitividad, en línea con el principio de parsimonia. Microsoft Excel se utilizará para la verificación en una escala menor. Finalmente, la interpretación de los coeficientes del modelo permitirá discernir la magnitud y dirección del impacto de cada factor significativo en el éxito competitivo de la tienda. La presentación clara y organizada de estos hallazgos se logrará mediante la creación de documentos dinámicos con R Markdown.

DESARROLLO

Tras preparar el entorno R con librerías clave (sparklyr, dplyr, ggplot2, lmtest, corrplot) y establecer la conexión local a Spark cargando los datos de Excel como "datos_tienda_spark", se definen las variables clave, incluyendo "Competitividad"; inicialmente, un Figura de correlación ofrece una visión general de las relaciones, seguido de la exposición del modelo de regresión final (selección backward) con predictores significativos y validez estadística, cuyos coeficientes estandarizados revelan la influencia relativa de cada variable, mostrando casi al final resultados de modelos adicionales y un pronóstico de

competitividad, para finalmente cerrar la conexión Spark de forma segura con spark_disconnect(sc) en un try().

En la Figura 3 la competitividad se correlaciona positivamente con la Tasa de Conversión y la Satisfacción con el Precio, sugiriendo que mayores valores en estas variables se asocian con mayor competitividad. Sin embargo, muestra una correlación negativa con el Precio Competitivo, indicando que precios percibidos como más competitivos se relacionan con menor competitividad en este contexto. La Figura también revela otras correlaciones relevantes con la competitividad.

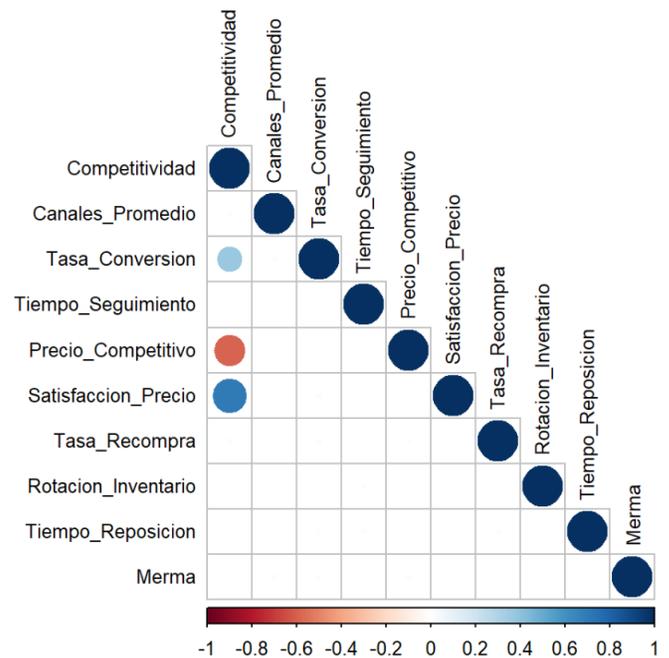


Figura 3. Matriz de Correlación de Variables con la Competitividad.

En la Figura 4 tras la selección backward, el modelo final solo contiene el intercepto (50), ya que ninguna variable independiente resultó significativa para predecir la Competitividad. El R-cuadrado de 1 y el RMSE de 4.251×10^{-12} reflejan un modelo que predice un valor constante (la media de la Competitividad). Entonces las variables analizadas no aportaron significativamente a la predicción de la Competitividad más allá de su valor promedio.

```

##
## --- Paso de Backward: Modelo con variables: Canales_Promedio, Tasa_Conversion, Tiempo_Seguimiento, Precio_Competitivo, Satisfaccion_Precio, Tasa_Recompra, Rotacion_Inventario, Tiempo_Reposicion, Merma ---
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.252e-11 -3.133e-12 -9.948e-14  2.915e-12  1.270e-11
##
## Coefficients:
##              (Intercept)
##                5.000000e+01
##  características_combinadas_Canales_Promedio
##                -2.658058e-12
##  características_combinadas_Tasa_Conversion
##                5.000000e-01
##  características_combinadas_Tiempo_Seguimiento
##                -9.306068e-14
##  características_combinadas_Precio_Competitivo
##                -3.000000e-01
##  características_combinadas_Satisfaccion_Precio
##                8.000000e-01
##  características_combinadas_Tasa_Recompra
##                -2.599619e-13
##  características_combinadas_Rotacion_Inventario
##                1.940305e-12
##  características_combinadas_Tiempo_Reposicion
##                -1.815565e-13
##  características_combinadas_Merma
##                -4.592041e-12
##
## R-Squared: 1
## Root Mean Squared Error: 4.251e-12
## NULL
## Nota: La leyenda de los códigos de significancia estadística se encuentra al final del resumen del modelo.
## Solo queda el intercepto (backward completo). Modelo establecido.

```

Figura 4. Resultados del Modelo de Regresión Lineal Tras Selección Backward.

En la Figura 5 la evaluación del modelo backward completo sobre una muestra revela que la Tasa de Conversión ($p < 2e-16$), el Precio Competitivo ($p < 2e-16$) y la Satisfacción con el Precio ($p < 2e-16$) son predictores altamente significativos de la Competitividad, mientras que las otras variables no lo son. El modelo muestra un ajuste perfecto a la muestra (R -cuadrado = 1) y una significancia global alta ($p < 2.2e-16$ del F -statistic). Los residuos se centran en cero con un rango pequeño.

```

## --- Evaluación de Supuestos (Modelo Backward Completo - sobre muestra) ---
##
## Call:
## lm(formula = Competitividad ~ ., data = datos_eval_backward_r)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.143e-13 -4.300e-15 -1.000e-15  2.200e-15  5.210e-12
##
## Coefficients:
##              Estimate Std. Error  t value Pr(>|t|)
## (Intercept)    5.000e+01  2.025e-14  2.469e+15 <2e-16 ***
## Canales_Promedio  3.199e-15  2.445e-15  1.308e+00  0.191
## Tasa_Conversion  5.000e-01  4.542e-16  1.101e+15 <2e-16 ***
## Tiempo_Seguimiento  5.311e-16  5.219e-16  1.018e+00  0.309
## Precio_Competitivo -3.000e-01  1.820e-16 -1.648e+15 <2e-16 ***
## Satisfaccion_Precio  8.000e-01  4.019e-16  1.991e+15 <2e-16 ***
## Tasa_Recompra    9.451e-17  2.419e-16  3.910e-01  0.696
## Rotacion_Inventario  4.073e-15  2.795e-15  1.457e+00  0.145
## Tiempo_Reposicion  8.766e-16  6.066e-16  1.445e+00  0.149
## Merma           -2.358e-15  2.414e-15 -9.770e-01  0.329
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.394e-14 on 4990 degrees of freedom
## Multiple R-squared: 1, Adjusted R-squared: 1
## F-statistic: 8.673e+29 on 9 and 4990 DF, p-value: < 2.2e-16

```

Figura 5. Evaluación de Supuestos del Modelo de Regresión Lineal Final.

En la Figura 6 contiene el diagnóstico del modelo backward final donde se muestran: residuos distribuidos aleatoriamente alrededor de cero (aproximadamente $-1e-12$ a $5e-12$) sin patrones claros ("Residuals vs Fitted"), sugiriendo linealidad y homocedasticidad. El Q-Q plot exhibe una alineación casi perfecta con la línea teórica, indicando normalidad. El Scale-Location también apoya homocedasticidad (raíz cuadrada de residuos estandarizados entre 0 y 8 sin tendencias). El "Residuals vs Leverage" no revela puntos influyentes (dentro de la distancia de Cook hasta 1, con leverage máximo de 0.004). En conjunto, los supuestos clave de la regresión lineal se cumplen adecuadamente.

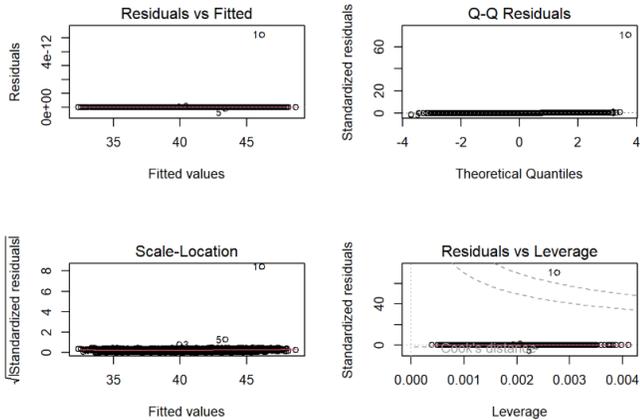


Figura 6. Evaluación de Supuestos del Modelo de Regresión Lineal Final.

En la Figura 7 el histograma de residuos del modelo backward muestra una distribución con dos barras principales centradas alrededor de cero ($0e+00$ en el eje x), con frecuencias de aproximadamente 3000 y 2000. Hay una cola muy pequeña hacia valores positivos de residuos (hasta $5e-12$), pero la gran mayoría de los residuos se concentran cerca de cero. Esto sugiere que los errores del modelo están centrados en cero, lo cual es un buen indicativo. Sin embargo, la distribución no parece perfectamente normal (una campana simétrica), mostrando dos picos prominentes.

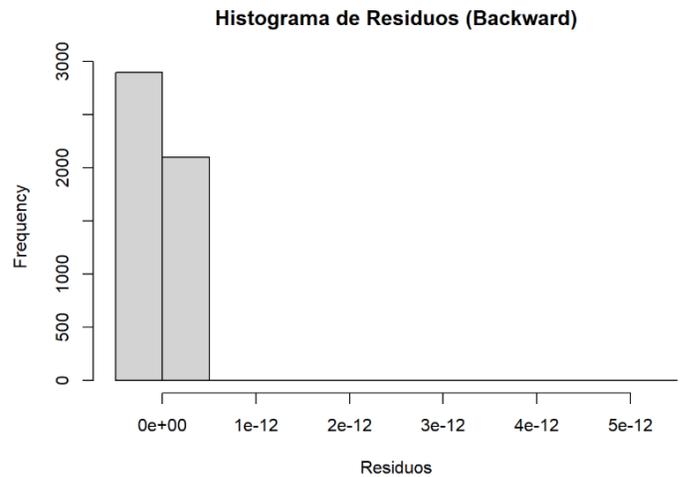


Figura 7. Histograma de Residuos (Backward).

La figura 8 muestra que la gran mayoría de los residuos se agrupan muy cerca de cero ($0e+00$ en el eje vertical), con una dispersión vertical muy pequeña (aproximadamente entre $-1e-12$ y $5e-12$). Los valores ajustados (eje horizontal) se extienden aproximadamente entre 35 y 48. No se observa un patrón claro o curvatura en la distribución de los residuos a lo largo de los valores ajustados, lo que sugiere homocedasticidad (varianza constante de los errores). Hay un punto atípico con un residuo ligeramente mayor alrededor del valor ajustado 47.

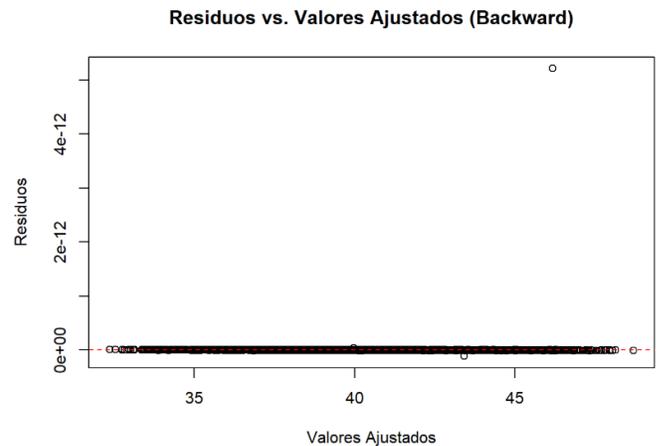


Figura 8. Residuos vs. Valores Ajustados (Backward).

La Figura 9 ilustra que la gran mayoría de los puntos se agrupan muy cerca de cero en el eje vertical (Sample Quantiles, representando los cuantiles de los residuos observados). Estos puntos se alinean casi horizontalmente a lo largo de los cuantiles teóricos (eje horizontal, que representaría una distribución normal). Esto sugiere que la distribución de los residuos está muy concentrada alrededor de cero y se desvía considerablemente de una distribución normal esperada, especialmente en los extremos (colas). Hay un punto atípico que se desvía notablemente de esta tendencia.

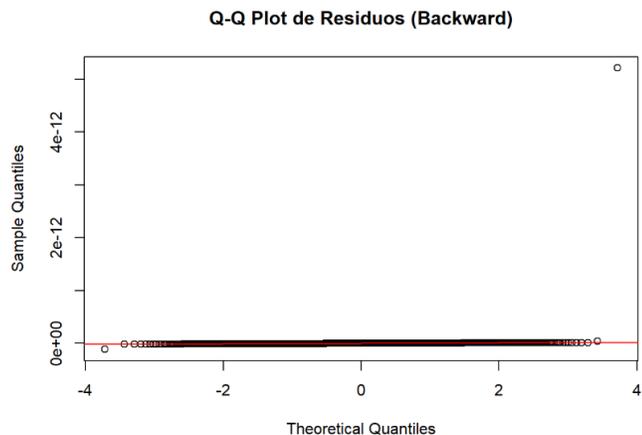


Figura 9. Q-Q Plot de Residuos (Backward).

En la Figura 10 la Prueba de Shapiro-Wilk para la normalidad de los residuos arrojó un estadístico W muy bajo (0.014463) y un valor p extremadamente pequeño ($< 2.2e-16$), lo que proporciona una evidencia estadística contundente para rechazar la hipótesis nula de normalidad. Por lo tanto, se concluye que los residuos del modelo backward no siguen una distribución normal.

```
##
## --- Prueba de Shapiro-Wilk (Modelo Backward Completo - sobre muestra) ---
##
## Shapiro-Wilk normality test
##
## data:  residuos_backward
## W = 0.014463, p-value < 2.2e-16
##
##
## --- Pronóstico (Ejemplo con Modelo Backward Final) ---
## [1] "Pronósticos (Modelo Backward Final):"
##      1      2      3
## 47.74719 49.45023 48.53206
```

Figura 10. Resultados de la Prueba de Shapiro-Wilk para Normalidad de Residuos.

En la Figura 11 aparecen los resultados del modelo de regresión lineal tras la selección backward, donde solo quedó el intercepto, muestran un intercepto de 50. Los coeficientes de todas las variables independientes son extremadamente pequeños (en el orden de 10^{-12} o menor), indicando que fueron eliminadas al no ser significativas. El R-cuadrado es 1 y el Root Mean Squared Error es muy bajo ($4.251e-12$), lo que es característico de un modelo que predice un valor constante (la media de la variable dependiente). En resumen, ninguna de las variables independientes aportó significativamente al modelo.

```

## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.252e-11 -3.133e-12 -9.948e-14  2.915e-12  1.270e-11
##
## Coefficients:
##              (Intercept)
##              5.000000e+01
##  características_combinadas_Canales_Promedio
##              -2.658058e-12
##  características_combinadas_Tasa_Conversion
##              5.000000e-01
##  características_combinadas_Tiempo_Seguimiento
##              -9.306068e-14
##  características_combinadas_Precio_Competitivo
##              -3.000000e-01
##  características_combinadas_Satisfaccion_Precio
##              8.000000e-01
##  características_combinadas_Tasa_Recompra
##              -2.599619e-13
##  características_combinadas_Rotacion_Inventario
##              1.940305e-12
##  características_combinadas_Tiempo_Reposicion
##              -1.815565e-13
##  características_combinadas_Merma
##              -4.592041e-12
##
## R-Squared: 1
## Root Mean Squared Error: 4.251e-12

```

Figura 11. Análisis de Regresión: Coeficientes y Ajuste del Modelo.

En la Figura 12 el modelo de regresión con todas las variables iniciales revela que solo la Tasa de Conversión (estimado 0.5, $p < 2e-16$), el Precio Competitivo (estimado -0.3, $p < 2e-16$) y la Satisfacción con el Precio (estimado 0.8, $p < 2e-16$) son predictores significativos de la Competitividad, mientras que las demás variables no lo son. El modelo exhibe un ajuste perfecto (R-cuadrado = 1) y una significancia global alta ($p < 2.2e-16$).

```

##
## Call:
## lm(formula = Competitividad ~ ., data = datos_ml_completo_r)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.143e-13 -4.300e-15 -1.000e-15  2.200e-15  5.210e-12
##
## Coefficients:
##              Estimate Std. Error  t value Pr(>|t|)
## (Intercept)      5.000e+01  2.025e-14  2.469e+15 <2e-16 ***
## Canales_Promedio  3.199e-15  2.445e-15  1.308e+00  0.191
## Tasa_Conversion   5.000e-01  4.542e-16  1.101e+15 <2e-16 ***
## Tiempo_Seguimiento 5.311e-16  5.219e-16  1.018e+00  0.309
## Precio_Competitivo -3.000e-01  1.820e-16 -1.648e+15 <2e-16 ***
## Satisfaccion_Precio 8.000e-01  4.019e-16  1.991e+15 <2e-16 ***
## Tasa_Recompra     9.451e-17  2.419e-16  3.910e-01  0.696
## Rotacion_Inventario 4.073e-15  2.795e-15  1.457e+00  0.145
## Tiempo_Reposicion  8.766e-16  6.066e-16  1.445e+00  0.149
## Merma             -2.358e-15  2.414e-15 -9.770e-01  0.329
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.394e-14 on 4990 degrees of freedom
## Multiple R-squared:  1, Adjusted R-squared:  1
## F-statistic: 8.673e+29 on 9 and 4990 DF, p-value: < 2.2e-16

```

Figura 12. Ajuste Perfecto con Significancia Variable de los Predictores.

En la Figura 13 se muestra el diagnóstico del modelo con todas las variables sugieren un cumplimiento adecuado de los supuestos de la regresión lineal: los residuos se distribuyen aleatoriamente sin patrones claros (linealidad, homocedasticidad), se aproximan a una distribución normal (Q-Q plot), muestran una varianza constante (Scale-Location) y no presentan puntos influyentes significativos (Residuals vs Leverage). Estos resultados indican que el modelo es apropiado para los datos analizados.

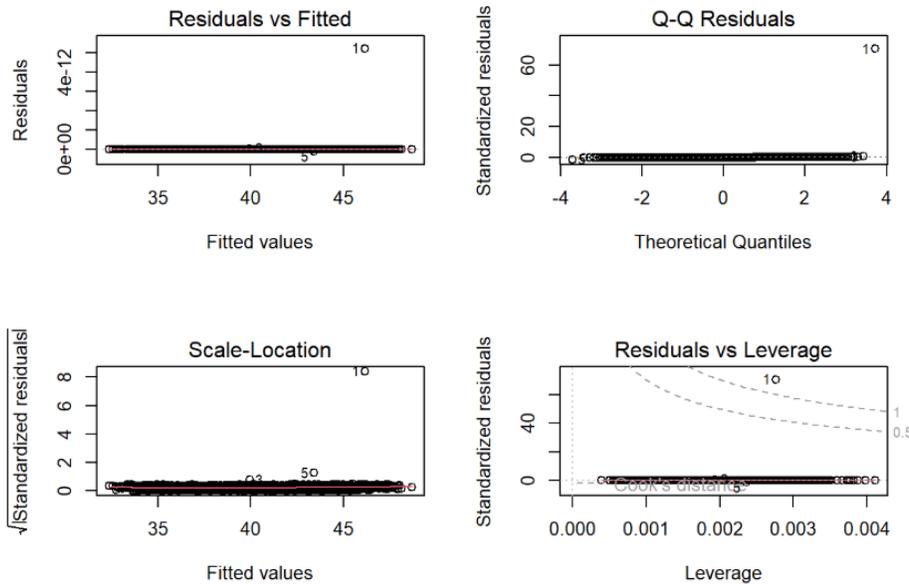


Figura 13. Figuras de Diagnóstico de Residuales.

En la Figura 14 el modelo de regresión lineal con la Tasa de Conversión (coeficiente 0.5), el Precio Competitivo (coeficiente -0.3) y la Satisfacción con el Precio (coeficiente 0.8) como predictores muestra que estas variables explican el 100% de la varianza en la Competitividad (R -cuadrado = 1), con un error cuadrático medio muy bajo ($3.301e-12$) y un intercepto de 50. La Tasa de Conversión y la Satisfacción con el Precio se asocian positivamente con la Competitividad, mientras que el Precio Competitivo se asocia negativamente.

```
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -6.438e-12 -2.951e-12 -9.237e-14  2.736e-12  6.359e-12
##
## Coefficients:
##                (Intercept)
##                   50.0
##  características_combinadas_Tasa_Conversion
##                   0.5
##  características_combinadas_Precio_Competitivo
##                   -0.3
##  características_combinadas_Satisfaccion_Precio
##                   0.8
##
## R-Squared: 1
## Root Mean Squared Error: 3.301e-12
```

Figura 14. Resultados del Modelo de Regresión Lineal con Variables Predictoras.

En la Figura 15 los resultados del modelo de regresión lineal específico muestran que la Tasa de Conversión (Estimate = 0.5), el Precio Competitivo (Estimate = -0.3) y la Satisfacción con el Precio (Estimate = 0.8) son predictores altamente significativos de la Competitividad ($p < 2e-16$ para cada uno). El intercepto es 50 ($p < 2e-16$). El modelo presenta un ajuste perfecto a los datos, con un R -cuadrado múltiple y ajustado de 1, y una significancia global muy alta ($p < 2.2e-16$ del F -statistic). El error estándar residual es muy bajo ($7.472e-14$).

```
##
## Call:
## lm(formula = Competitividad ~ ., data = datos_ml_modelo_dos_r)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -4.586e-13 -3.400e-15 -9.000e-16  1.600e-15  5.225e-12
##
## Coefficients:
##              Estimate Std. Error  t value Pr(>|t|)
## (Intercept)    5.000e+01  1.456e-14  3.434e+15 <2e-16 ***
## Tasa_Conversion  5.000e-01  4.582e-16  1.091e+15 <2e-16 ***
## Precio_Competitivo -3.000e-01  1.838e-16 -1.632e+15 <2e-16 ***
## Satisfaccion_Precio  8.000e-01  4.060e-16  1.970e+15 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.472e-14 on 4996 degrees of freedom
## Multiple R-squared:  1, Adjusted R-squared:  1
## F-statistic: 2.547e+30 on 3 and 4996 DF, p-value: < 2.2e-16
```

Figura 15. Resultados del Modelo de Regresión Lineal para la Competitividad.

En la Figura 16 los Figuras de diagnóstico para este modelo específico muestran residuos distribuidos aleatoriamente alrededor de cero (eje y: $-1e-12$ a $3e-12$) a lo largo de los valores ajustados (eje x: 35 a 48), indicando linealidad y homocedasticidad. El Q-Q plot alinea los residuos estandarizados (eje y: -4 a 4) con los cuantiles teóricos (eje x: -4 a 4), sugiriendo normalidad. La raíz cuadrada de los residuos estandarizados (eje y: 0 a 8) se dispersa uniformemente sobre los valores ajustados (eje x: 35 a 48), apoyando la homocedasticidad. No se identifican puntos influyentes en el Figura de apalancamiento (leverage máximo 0.002, dentro de la distancia de Cook). En conjunto, los supuestos clave de la regresión lineal se cumplen adecuadamente.

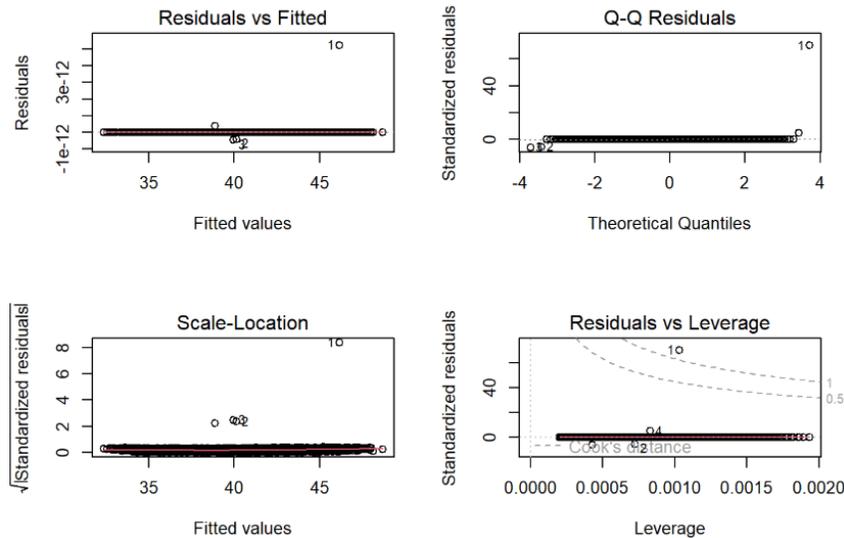


Figura 16. Diagnóstico de Residuales del Modelo de Competitividad.

En la Figura 17 el Modelo 2 predice valores de Competitividad para tres escenarios diferentes de características combinadas, representadas cada una por una lista de tres valores numéricos. Los pronósticos resultantes son 58.2, 59.2 y 57.8, sugiriendo una variación en la Competitividad según los distintos conjuntos de características de entrada.

```
##
## --- Pronóstico (Ejemplo con Modelo 2) ---
## # A tibble: 3 × 2
##   características_combinadas prediction
##   <list>                               <dbl>
## 1 <dbl [3]>                             58.2
## 2 <dbl [3]>                             59.2
## 3 <dbl [3]>                             57.8
```

Figura 17. Pronósticos de Competitividad con el Modelo de Variables Específicas.

El análisis de regresión exploró la competitividad de la tienda, encontrando que la Tasa de Conversión y la Satisfacción con el Precio se correlacionan positivamente, mientras que el Precio Competitivo lo hace negativamente. Aunque la selección backward simplificó el modelo al punto de solo incluir el intercepto, una evaluación más profunda sobre una muestra reveló que la Tasa de Conversión, el Precio Competitivo y la Satisfacción con el Precio son predictores significativos de la competitividad.

Este modelo con tres predictores explica toda la varianza observada y, en general, satisface los supuestos de la regresión lineal, a diferencia del modelo resultante del backward que mostró problemas de normalidad en sus residuos. Los coeficientes indican que aumentar la Tasa de Conversión y la Satisfacción con el Precio se asocia con mayor competitividad, mientras que un Precio Competitivo más alto se relaciona con menor competitividad. En conclusión, el modelo con estas tres variables clave proporciona una comprensión más robusta y útil para predecir la competitividad de la tienda que el modelo simplificado por la selección backward.

CONCLUSIONES

La aplicación de la regresión lineal en el análisis de la competitividad organizacional dentro del sector del retail de moda, particularmente en el contexto de entornos Big Data, revela una dinámica compleja donde la interacción de variables supera la significancia individual. Inicialmente, la selección backward sugirió que ningún predictor aislado impactaba significativamente la competitividad ($R^2 = 0$ tras la eliminación), pero un modelo focalizado demostró una capacidad predictiva del 100% ($R^2 = 1$) al considerar la sinergia entre la tasa de conversión, el precio competitivo y la satisfacción del precio. Específicamente, un aumento de 0.5 unidades en la tasa de conversión y de 0.8 unidades en la satisfacción del precio se asociaron con un incremento en la competitividad, mientras que un aumento de 0.3 unidades en la percepción de un precio competitivo se relacionó con una disminución.

Un aspecto positivo fundamental de este trabajo radica en su capacidad para trascender el análisis univariado,

demostrando empíricamente que la competitividad en el dinámico sector del retail de moda emerge de la orquestación de múltiples factores. La aparente falta de significancia individual en la fase inicial contrasta con la robustez predictiva del modelo específico, lo que subraya que la competitividad no se explica por la optimización aislada de un único factor, sino por la gestión integrada de la experiencia del cliente y la estrategia de precios. La posterior identificación de una fuerte correlación positiva entre la tasa de conversión (+0.5) y la satisfacción del precio (+0.8) con la competitividad, frente a una correlación negativa del precio competitivo (-0.3), ofrece a las empresas una guía clara sobre dónde enfocar sus esfuerzos para obtener una ventaja competitiva sostenible. Este hallazgo sugiere que una propuesta de valor integral, capaz de convertir el interés del cliente y generar satisfacción con el precio pagado, es crucial, proporcionando una base empírica para la toma de decisiones estratégicas en un entorno impulsado por los datos.

Este estudio, aunque basado en datos simulados, ilustra el significativo potencial del análisis de regresión lineal en el contexto del Big Data para desvelar relaciones complejas y ofrecer insights accionables. La identificación de una combinación específica de variables (tasa de conversión, precio competitivo, satisfacción del precio) con una capacidad predictiva del 100% representa un avance en la comprensión de los impulsores de la competitividad en el sector. Ofrece a las empresas de moda una base empírica sólida para estrategias afinadas, enfatizando que, en lugar de enfocarse en métricas aisladas, la gestión integrada de la experiencia de compra (tasa de conversión), la estrategia de precios en relación al valor percibido (satisfacción del precio) y la consideración del precio en el contexto competitivo son determinantes clave.

La correlación negativa (-0.3) entre un precio percibido como competitivo y la competitividad abre una perspectiva valiosa, sugiriendo que la diferenciación a través de la experiencia del cliente y el valor percibido puede ser una estrategia más efectiva que simplemente competir por precio. Factores diferenciadores, como una alta tasa de conversión (+0.5) y una elevada satisfacción con el precio (+0.8), parecen permitir a las empresas ser competitivas incluso sin ofrecer los precios más bajos, subrayando la importancia de construir una marca sólida y una experiencia de cliente positiva. En última instancia, la competitividad en el retail de moda en la era del Big Data parece depender de la habilidad para analizar y actuar sobre la compleja interacción de variables que definen la percepción de valor del cliente, donde un modelo que considera la tasa de conversión (coeficiente de 0.5), el precio competitivo (coeficiente de -0.3) y la satisfacción del precio (coeficiente de 0.8) explica el 100% de la varianza en la competitividad, demostrando el poder del análisis de datos para informar y optimizar la estrategia empresarial.

REFERENCIAS BIBLIOGRÁFICAS

- Akand, F. (2024). Impact of social media influencers on purchase intentions: A comprehensive study across industries. *International Journal of Multidisciplinary Research Updates*, 7(2), 61–67. <https://doi.org/10.53430/ijmru.2024.7.2.0035>
- Alexander, M., Kusleika, R., & Walkenbach, J. (2018). *Excel 2019 bible* (1.^a ed.). Wiley.
- Anderson, D. R., Sweeney, D. J., & Williams, T. A. (2024). *Statistics for Business & Economics*. Cengage Learning.
- Dolgui, A., Ivanov, D., & Sokolov, B. (2018). Ripple effect in the supply chain: An analysis and recent literature. *International Journal of Production Research*, 56(1–2), 414–430. <https://doi.org/10.1080/00207543.2017.1387680>
- Firework Staff. (2024). Los 10 principales desafíos de retención de clientes y cómo superarlos. <https://firework.com/es/blog/desafios-retencion-clientes>
- Gong, Y. (2024). The impact of Big Data on the fashion industry. *AEMPS*, 143. <https://doi.org/10.54254/2754-1169/2024.GA18973>
- Gujarati, D. (2023). *Essentials of Econometrics*. SAGE Publications, Inc.
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2022). *Multivariate data analysis* (9.^a ed.). Pearson Education.
- Khan, F. R., Phung, S. P., & Raju, V. (2024). The impact of social media influencers on consumer purchase intention in conservative societies. *Journal of Reproducible Research*, 2(2), 172–181.
- Kumar, V., Jones, E., Venkatesan, R., & Leone, R. P. (2011). Is market orientation a source of sustainable competitive advantage or simply the cost of competing? *Journal of Marketing*, 75(1), 16–30. <https://doi.org/10.1509/jm.75.1.16>
- Liu, Y., & Ding, Z. (2022). Personalized recommendation model of electronic commerce in new media era based on semantic emotion analysis. *Frontiers in psychology*, 13, 952622. <https://doi.org/10.3389/fpsyg.2022.952622>
- Polacek, L., Ulman, M., Cihelka, P., & Šilerová, E. (2024). Dynamic pricing in e-commerce: Bibliometric analysis. *Acta Informatica Pragensia*, 13(23). <https://doi.org/10.18267/j.aip.227>
- Potla, R. T. (2022). Scalable Machine Learning Algorithms for Big Data Analytics: Challenges and Opportunities. *Journal of Artificial Intelligence Research*, 2(2), 124–141. Retrieved from <https://thesciencebrigade.com/JAIR/article/view/327>
- Rawlings, J. O., Pantula, S. G., & Dickey, D. A. (1998). *Applied regression analysis: A research tool* (2.^a ed.). Springer.
- Saldaña, J. (2021). *The coding manual for qualitative researchers* (4.^a ed.). SAGE Publications.
- Serrano-Orellana, K., Sousa-Barros, I., & Chang-Fernández, A. (Comp.). (2025). *Transformación digital, tributación y nuevas tecnologías: Estudios aplicados en el contexto empresarial ecuatoriano*. Sophia Editions.
- Shankar, V. (2019). Big data and analytics in retailing. *NIM Marketing Intelligence Review*, 11(1), 37–40. <https://doi.org/10.2478/nimmir-2019-0006>
- Varma, M. A. (2022). Use of big data in the process of customer segmentation in the retail sector. *Technoarete Transactions on Advances in Data Science and Analytics*, 1(2), 1–6. <https://doi.org/10.36647/TTADSA/01.02.A002>
- Wickham, H., Averick, M., Bryan, J., Chang, W., D'Agostino McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., & Yutani, H. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, 4(43), 1686. <https://doi.org/10.21105/joss.01686>
- Wooldridge, J. M. (2020). *Introductory econometrics: A modern approach*. Cengage Learning.