

## Pronóstico de ventas anuales para ropa de mujer usando regresión lineal

### Taller 2 Individual

María Fernanda Izquierdo Aparicio

#### BUSINESS UNDERSTANDING

##### 1. Background

El sector de la moda para mujeres ha venido experimentado un aumento constante conforme se ha desarrollado mundialmente en los últimos años. Según IMARC Group (2023), el incremento en las ventas de prendas femeninas se atribuye principalmente a las tendencias de moda en ascenso y al creciente impacto de las redes sociales. La industria de la moda satisface a una diversa clientela que abarca distintas edades, estilos de vida, ocasiones, gustos y preferencias. Para el caso colombiano, “datos de Inexmoda revelan que el consumo en moda en el país alcanzó \$27,7 billones en 2021, 21% más que en 2020 y 5% más que en 2019” (Estrada Rudas, 2022) y de acuerdo a Euromonitor International (2022) los canales de distribución más importantes son el de tiendas físicas y *retail online*.

En enero del 2022, el gasto de los hogares destinado a la categoría de moda representó un 3,3%, acercándose al nivel observado antes de la pandemia, que alcanzaba el 3,6% en enero de 2020. Además, en lo que respecta a las tendencias de consumo, las ventas en línea juegan un papel destacado. Según el informe ‘Sector de la moda’ de EAE Business School (2022), el calzado lidera con un 25% de participación en el total de las ventas en línea, experimentando un aumento del 11,2%, seguido de cerca por el sector textil para el hogar con un 19,3% de crecimiento. Por último, el mercado de segunda mano está ganando impulso, y en la actualidad, en Colombia, Bogotá lidera con un 40% de las compras de ropa usada, seguido de Medellín con un 23%, Barranquilla con un 22%, y Cali y Bucaramanga, ambas con un 15%.

En cuanto a canales, Euromonitor Internacional (2022) destaca que la pandemia aceleró la importancia del comercio electrónico minorista en el país, logrando que este canal duplicara su participación en el valor en 2020. Aunque el comercio electrónico perdió terreno entre 2021 y 2022 a medida que los consumidores regresaban lentamente a las tiendas, las ventas totales de comercio electrónico minorista se mantuvieron ligeramente por encima de los niveles anteriores a la pandemia. En contraste, aunque la venta por catálogos no es el canal principal de distribución en el país, las ventas por este medio han crecido un 8.1% en el último año (Redacción PDM, 2023)

En temas económicos, específicamente para los catálogos, se debe destacar que los costos de imprimir y distribuir catálogos han aumentado en un entorno de venta minorista cada vez más en línea. Esto ha llevado a que algunos minoristas de catálogos abandonen la impresión en favor de estrategias en línea (Weston, 2023). A pesar de los mayores costos, los catálogos siguen siendo efectivos para llegar a un grupo más específico de consumidores y ofrecen un alto retorno de inversión (ROI) para algunas marcas (McCoy, 2023). Y para muchas marcas, los catálogos siguen siendo un impulsor efectivo de las ventas en línea y en tiendas físicas (Ruiz, 2015). Pero cabe mencionar que aunque es un canal que da resultados, el aumento de los precios en los servicios postales y de envío puede afectar negativamente a la industria de los catálogos, ya que aumenta los costos de llegar a los consumidores (Davis et al., 2012).

En temas sociales, los resultados de diversos estudios, como el de Zhang en 2020, señalan que los catálogos desempeñan un papel importante en el aumento de las ventas y la participación de los consumidores. Se destacan por su capacidad para destacarse en medio de la saturación de bandejas de entrada digitales y redes sociales. Al tratarse de productos físicos, los catálogos pueden perdurar en los hogares de los consumidores mucho después de que los correos electrónicos hayan sido eliminados, lo que contribuye a mantener su presencia en la mente del consumidor. No obstante, su verdadero poder radica en su capacidad, en el caso de ciertos productos, para incrementar la viveza al permitir que los consumidores visualicen y se imaginen experiencias de uso del producto de una manera más vívida. Además, según Weston (2023), los catálogos actuales representan un punto de contacto adicional con los consumidores y han evolucionado para convertirse en algo más parecido a un libro de mesa de café que brinda a los consumidores la comodidad de explorar productos a su ritmo. Esto fomenta compras más conscientes al permitir que los consumidores se sumerjan por completo en la oferta del minorista. Zagalsky (2023) destaca la eficacia de los catálogos como herramienta de marketing, resaltando el impacto positivo que un hermoso catálogo puede tener en la mente del consumidor en comparación con la

saturación de correos electrónicos y publicidad invasiva. Por otra parte, según Reagan (2018), los catálogos parecen resonar de manera especial con la generación millennial, que valora la experiencia de recibir correo físico en un mundo dominado por lo digital. Además, los compradores que utilizan catálogos suelen ser más leales y dispuestos a gastar más, en contraposición a los compradores en línea, que suelen centrarse en los precios. Sin embargo, es importante tener en cuenta que, como menciona Davis et al. en 2012, los incrementos en los costos de servicios postales tienen un impacto negativo directo en la industria de los catálogos, ya que encarecen la distribución de catálogos a los consumidores. Estos hallazgos demuestran la relevancia continua de los catálogos como una herramienta de marketing efectiva en un entorno digital saturado y en constante cambio.

Teniendo el contexto del mercado se revisan ahora las 5 Fuerzas de Porter:

**Clientes:**

Los compradores en el mercado de ropa femenina en Colombia tienen un poder de negociación relativamente alto ya que existe una amplia variedad de opciones y marcas disponibles para las consumidoras, lo que les da margen para elegir y comparar precios y calidad. Euromonitor International reporta que el mercado colombiano se divide en prendas premium que componen el 3.2% del mercado, prendas de precio moderado componen el 24% del mercado y las prendas de gran consumo componen el restante 72.8% del mercado, por lo que se evidencia que los clientes tienen de donde escoger. Ahora, con la opción que ofrecen los catálogos se ha visto que clientes en zonas periféricas han podido acceder al mercado a través de este canal de venta directa, sobre todo en “zonas que antes no participaban tan activamente en esta industria” (Redacción PDM, 2023).

**Proveedores:**

En el mercado de ropa femenina en Colombia, hay una amplia variedad de proveedores de textiles y confección, tanto locales como internacionales. Esto reduce el poder de negociación de los proveedores, ya que los minoristas pueden elegir entre múltiples fuentes. De hecho Euromonitor International reporta que “según la Confederación Colombiana de Cámaras de Comercio (Confecamaras), la creación de microempresas (relacionadas a confección y comercio de prendas de ropa femenina) en el país creció un 10,6% en 2021 respecto al año anterior” (2022). Además, la globalización y la facilidad de acceso a proveedores extranjeros también disminuyen el poder de los proveedores locales. Sin embargo, los cambios en las tasas de cambio y las políticas comerciales pueden influir en los costos de importación, lo que puede afectar el poder de los proveedores extranjeros.

**Productos Sustitutos:**

La amenaza de productos sustitutos es baja en el mercado de ropa femenina en Colombia, ya que la ropa es una necesidad básica y no tiene sustitutos directos cuando se entiende en un contexto genérico. Aun así, la moda rápida y la ropa de segunda mano son alternativas que algunos consumidores pueden considerar, especialmente en tiempos económicos difíciles.

**Competidores Existentes:**

La rivalidad entre competidores existentes en el mercado de ropa femenina en Colombia es alta. Hay una gran cantidad de marcas nacionales e internacionales compitiendo por la cuota de mercado. Euromonitor Internacional (2022) reporta que los principales competidores en esta industria son Confecciones Leonisa SA, Internacional de Distribuciones de Vestuario de Moda SA, Studio F SA y Inditex, Industria de Diseño Textil SA quienes abarcan el 30% del mercado. El porcentaje restante se divide en más de 40 compañías nacionales como internacionales. Estos competidores se caracterizan por tener diversas estrategias de precios, innovación en diseño y marketing, y la expansión a través de canales en línea y físicos son comunes en la industria. De estas marcas, se destaca que hay 28 compañías que están registradas en la Asociación Colombiana de Venta Directa (ACOVEDI SAS, 2023) quienes ofrecen venta por catálogo.

**Nuevos Competidores:**

La amenaza de nuevos competidores en el mercado de ropa femenina en Colombia es moderada. A pesar de las barreras de entrada, como la inversión inicial y la necesidad de establecer una marca, hay espacio para la entrada de nuevas empresas, especialmente en segmentos de nicho. Las marcas ya establecidas pueden tener ventajas en términos de reconocimiento y economías de escala, lo que dificulta la entrada de competidores a gran escala.

Teniendo en cuenta el contexto macro al que se enfrenta la industria y los componentes de la misma, se hace ahora un análisis de la posición sobre la cual se trabajará este proyecto. Se considerará a Leonisa como la empresa cliente de este proyecto. Leonisa es una marca colombiana de ropa interior, fajas y lencería y que es marca líder en Colombia en su segmento. Leonisa ofrece una amplia gama de productos para mujeres de todas las edades y tallas. Su ropa interior es conocida por su comodidad, calidad y diseño. Sus principales canales de venta son las tiendas físicas y la venta por catálogo, este último es un canal de distribución importante para la marca ya que las ejecutivas de venta de Leonisa ofrecen los productos de la marca a través de catálogos personalizados y distribución permite a la empresa llegar a un público amplio y diverso. A continuación se presenta un análisis DOFA de la entidad:

A continuación se presenta un análisis DOFA de la entidad:

### **Fortalezas**

Leonisa es una marca reconocida y de prestigio, cuenta con más de 60 años de presencia en el país y marca líder en Colombia en ropa interior, fajas y lencería. Ofrece una amplia gama de productos para mujeres de todas las edades y tallas, y en especial su ropa interior es conocida por su comodidad, calidad y diseño. Además de ser reconocida, Leonisa cuenta con canales de distribución diversificados: una red de sucursales físicas, una tienda virtual y un canal de venta por catálogo. Esto le permite llegar a un público amplio y diverso.

### **Debilidades**

La producción y la logística de Leonisa requieren altos costos fijos, lo cual puede representar una desventaja en períodos de recesión económica. También presenta una dependencia de la venta por catálogo ya que la venta por este medio es uno de los canales de distribución más importantes para Leonisa. Sin embargo, este canal puede ser vulnerable a las tendencias cambiantes en el consumo.

### **Oportunidades**

Leonisa, teniendo su canal de venta por catálogo tiene la oportunidad de enfocarse en mercados de periferia y de nicho, pues por un lado y en palabras de Zagalsky (2023) “ las compras en línea consumen cada vez más tiempo, lo que frustra la conveniencia de hacer clic para comprar “ y usar el catálogo como “preludio de un sitio de comercio electrónico, llenando de ideas las mentes de los compradores antes de que naveguen por lo “desconocido” potencial”.

### **Amenazas**

Por un lado la competencia en el mercado de ropa es cada vez más intensa en Colombia, por lo que encontrar nuevos mercados es clave para la marca. Y esto va acompañado de los cambios en los hábitos de consumo, como la creciente preferencia por la ropa cómoda y casual, ya que este factor afecta directamente la demanda de ropa femenina.

## **2. Business goal**

El objetivo de negocio es incrementar la tasa de retorno a la inversión (*Return of Investment - ROI*) a partir del pronóstico de ventas anuales para la ropa de mujer basado en la información de la sucursal.

## **3. Business success criteria**

Incrementar la tasa de retorno a la inversión anual entre 5 y 10 puntos porcentuales.

## **4. Determine Data mining goals**

Desarrollar un modelo de regresión lineal adecuado que permita pronosticar las ventas anuales para la ropa de mujer que cumpla con los estándares de capacidad predictiva.

## **5. Data mining success criteria:**

El criterio de éxito para la regresión lineal es que tenga el *menor* error cuadrático medio.

## DATA UNDERSTANDING

### 1. Describe data

Se tiene una base de datos con información de “1600 zonas ordenadas de manera contigua durante un año de un detallista de ropa y joyería que vende por catálogo” (Beltrán et al, 2023). La base de datos presenta información acerca de las ventas de prendas para mujer en millones de pesos, información sobre los catálogos enviados (número de catálogos enviados y número de páginas del catálogo) y gastos de la empresa (cantidad gastada en publicidad impresa y valor total de la nómina). También presenta información respecto al servicio del negocio, como el número promedio de líneas para llamada abiertas para pedidos y número de representantes del servicio al cliente. Finalmente la base cuenta con información acerca del mercado (tipo y tamaño potencial del mercado; años transcurridos desde la llegada a la zona y tipo de promoción que se llevó a cabo durante el año)

### 2. Explore data

Tanto los análisis univariado como multivariado se realizaron sobre la base de datos entera. De éste análisis se debe mencionar que la data contiene 12 variables en total, 11 variables escalares y una categórica. Solamente se analizan 11 variables escalares ya que *Idloc* es un identificador único para cada zona.

A partir de un análisis univariado de cada una de las variables se puede hacer el siguiente resumen:

- **ropamujer:** Es el valor de ventas de prendas para mujer en millones de pesos durante el año y la variable a predecir. Los valores van desde \$ 1.289,24 hasta \$117.709,26 millones de pesos. En el año de estudio, en promedio se vendieron \$ 50,978.38 millones de pesos. La desviación estándar es de aproximadamente \$17.558,26, la cual una desviación relativamente baja (comparándola con la media), y el coeficiente de variación es de 0.34, métricas que indican que no hay una cantidad significativa de variabilidad en los datos de venta de prendas de mujer. En cuanto a medidas de forma, la el valor de asimetría es 0.31, lo que informa sobre una ligera asimetría positiva en los datos, y el valor de curtosis es 0.00, lo que indica una distribución casi mesocúrtica, en otras palabras, ésta variable tiene una forma muy similar a la distribución normal.
- **correo:** Esta variable indica el número de catálogos enviados durante el año. Durante el año estudiado, las zonas enviaron entre 1.147 y 15.259 y catálogos, con un promedio de 10.598,68 catálogos enviados por zona. En cuanto a medidas de dispersión, la variable de correo tiene una desviación estándar 2.400,67 y un coeficiente de variación es 0.23, lo que indica que la variabilidad en el número de catálogos enviados, en relación con la media, es relativamente baja. Y en cuanto a forma, esta variable tiene un coeficiente de asimetría es -0,97, y una curtosis de 2,74. Estas métricas indican 1) una asimetría negativa, siendo la cola izquierda de la distribución es más larga que la derecha, y evidenciando una concentración de valores en la parte inferior del rango; y 2) que los datos tienen una mayor concentración de valores alrededor de la media y colas más pesadas en comparación con una distribución normal, y sugiere la presencia de valores atípicos en los datos.
- **páginas:** Indica el número de páginas del catálogo. Los catálogos podían tener entre 51 hasta 114 páginas, con un promedio de 85,04 por catálogo. Esta distribución presenta una desviación estándar de 15,94 páginas, y un coeficiente de variación de 0.19, lo que indica que los catálogos no variaban por un número muy alto de páginas, por lo cual se puede pensar que la distribución no presenta una variabilidad significativa en los datos. Por último se tiene que la distribución tiene un coeficiente de asimetría -0,05 y curtosis igual a -0,73, medidas que indican una distribución casi simétrica la cola izquierda un poco más larga que la cola derecha, también tiene una concentración de datos concentración alrededor del promedio y, en comparación con una distribución normal, ésta variable tiene una forma relativamente más plana.
- **teléfono:** Esta variable indica el número promedio de líneas para llamada abiertas para pedidos. La dispersión de la variable indica que habían entre 17 y 59 líneas abiertas para pedidos, y en promedio una zona podía tener 37,64 líneas abiertas. La desviación estándar de 9,75 y el coeficiente de variación de 0.26 sugieren una moderada variabilidad en relación con la media y la asimetría positiva de 0,42 indica una ligera cola hacia la derecha en la distribución, lo que sugiere una concentración de valores más bajos. La curtosis negativa de -0,40 señala una distribución relativamente achatada en comparación con una distribución normal.
- **impresa:** Es la cantidad gastada en publicidad impresa. Su dispersión muestra que las zonas gastaron entre \$18.061,20 y \$40.027,78 unidades monetarias, con un valor medio de \$28.834,62 gastado por zona. La desviación estándar de 6,251.17 y el coeficiente de variación de 0.22, métricas que evidencian una variabilidad moderada en relación con la media. Esta variable tiene una asimetría cercana a cero (0,06) lo

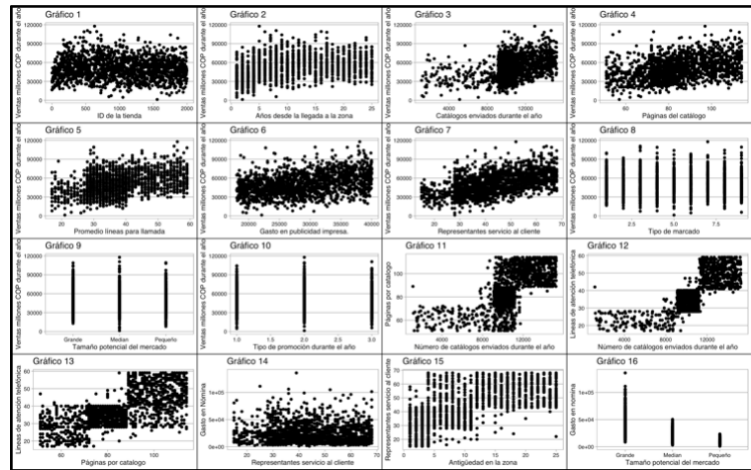
cual indica que es una distribución casi simétrica, con una ligera inclinación hacia la derecha, y tiene una curtosis negativa de -1,12, lo cual señala que la distribución es relativamente achatada en comparación con una distribución normal, y muestra menos concentración de valores alrededor de la media y colas más ligeras.

- **servicio:** Es el número de representantes del servicio al cliente. Presenta un rango que va desde 15 hasta 68 representantes, con un valor medio de 43,65 representantes por zona. Tiene una desviación estándar de 13,24 y coeficiente de variación de 0,30, medidas que indican una moderada variabilidad en relación con la media. Y en cuanto a forma, la variable tiene una asimetría cercana a cero, siendo el coeficiente de asimetría 0,01 y una curtosis negativa de -0,82. Esto señala que la distribución es casi que simétrica y es relativamente plana en comparación con una distribución normal.
- **edadloc:** Es el número de años transcurridos desde la llegada a la zona. Esta variable indica que el negocio puede tener una antigüedad de 1 a 25 años en la zona correspondiente, siendo el promedio 10,47 años de antigüedad y la mediana de 9 años, lo que sugiere que el negocio ha estado en la mayoría de las zonas por alrededor de 9 a 10 años. En cuanto a distribución, esta variable tiene una desviación estándar de 7,39 y el coeficiente de variación de 0,71, medidas que evidencian una variabilidad moderada en relación con la media. Y en cuanto a la forma, la variable tiene coeficiente de asimetría de 0,47 lo que indica una asimetría positiva, lo que indica que hay algunas zonas con mayor antigüedad que la mayoría. También tiene una curtosis negativa de -1,06 lo que significa que la distribución es relativamente plana, tiene colas más ligeras y una menor concentración de valores en torno a la media en comparación con una distribución normal.
- **nomina:** Esta variable indica el valor total de la nómina durante el año. La nómina durante el año en estudio costó entre \$901 hasta \$135.882 unidades monetarias para el negocio, con un valor promedio de \$24.104,13 por zona. La distribución tiene una desviación estándar de \$18.217,35 y un coeficiente de variación de 0,76, medidas que muestran que la variabilidad de la nómina es alta en relación con su media. Esta distribución tiene una asimetría positiva pues su coeficiente de asimetría es de 1,55, lo que señala una cola pronunciada hacia la derecha en la distribución, lo que implica la presencia de unos pocos valores extremadamente altos que elevan el promedio. Y tiene una curtosis de 2,94, medida que evidencia una alta concentración de valores alrededor de la media y una cola izquierda más pesada en comparación con una distribución normal.
- **idmercado:** Es la variable correspondiente a tipo de mercado, aunque solamente se listan números de clases o tipos de mercado que se desea atacar mas no se tiene contexto adicional de que indican los diversos niveles de la variable. Si esta variable se toma como una variable escalar se puede decir que los valores oscilan de 1 a 9. El promedio es de 4,93 y su media es de 5, lo que indica que una buena parte de las zonas tienen mercados de tipo 4 o 5. Esta variable tiene una desviación estándar de 2,61 y coeficiente de variación de 0,53, lo que sugiere una variabilidad moderada en relación con la media. La asimetría cercana a cero ( de 0,07) señala una distribución casi simétrica, y la curtosis negativa de -1,24 indica que en comparación con una distribución normal, ésta variable tiene una forma relativamente más plana. En resumen, los números de clases o tipos de mercado tienen una variabilidad moderada, pero sin más contexto, no se puede proporcionar una interpretación concreta sobre su significado en sí mismos.
- **promo:** Indica el tipo de promoción que se llevó a cabo durante el año. Al igual que con la variable anterior, no se tiene contexto adicional de que indican los diversos niveles de la variable. Si la variable se toma como una variable escalar se puede decir que hay tres niveles de promoción dado que los valores oscilan entre 1, 2 y 3, pero no hay interpretación precisa de que tipo de promoción significa cada valor. La media se encuentra en 1,98 y la mediana en 2, lo que sugiere una tendencia hacia el tipo de promoción etiquetado como "2". La desviación estándar de 0,75 y el coeficiente de variación de 0,38. Estas medidas señalan una variabilidad moderada en relación con la media. La variable tiene una asimetría de 0,04, muy cercana a cero lo que sugiere una distribución casi simétrica. Y tiene una curtosis de -1,22 que evidencia que la distribución es casi que plana en comparación con una distribución normal. Si esta variable se toma como una categórica, cabe resaltar que el nivel "1" tiene el 29,25% de las zonas, el nivel "2" tiene el 43,81% de las zonas y el nivel "3" tiene el 26,94% de las zonas registradas.
- **tamamer:** Esta variable es categórica e indica el tamaño potencial del mercado, proyectado de acuerdo a cifras del área de marketing. De las zonas registradas, se tiene que el 30,75% tienen mercados potencialmente pequeños, el 35,56% de las zonas tienen mercados potencialmente medianos y el restante 33,69% fueron categorizadas como mercados potencialmente grandes. Esto indica que el tamaño potencial de los mercados a través de las zonas está relativamente balanceado en las tres categorías.

Del análisis multivariado cabe mencionar que:



- Existen correlaciones positivas importantes (coeficientes de correlación mayores que 0,5) entre: 1) el número de años transcurridos desde la llegada a la zona y el número de representantes del servicio al cliente, 2) número de catálogos enviados durante el año, el número de páginas del catálogo y número promedio de líneas para llamada abiertas para pedidos, 3) el valor total de la nómina durante el año y el tamaño de mercado potencialmente grande.
- Hay correlaciones negativas importantes (coeficientes de correlación menores que -0,5) entre: 1) el valor total de la nómina durante el año y el tamaño de mercado potencialmente pequeño y 2) el tamaño de mercado potencialmente, el tamaño de mercado potencialmente mediano y el tamaño de mercado potencialmente pequeño.
- En un análisis preliminar visual, a partir de la graficación de las diversas variables explicativas contra la variable a predecir, como se muestra en el cuadro a continuación gráficos 1 al 10, no es posible identificar una relación directa que explique el comportamiento de la variable a predecir a partir de las variables explicativas, originado entre otras cosas, por la dispersión de los datos.
- En contraste, el análisis visual, a partir de la graficación de algunas variables explicativas, como se muestra en los gráficos 11 al 16, se puede observar que: 1) a mayor número de catálogos enviados el número de páginas también aumenta, 2) a mayor número de catálogos enviados hay mayor número de líneas de atención telefónica, 3) el número de representantes de servicio al cliente no tiene una clara incidencia en el gasto de nómina, 4) a mayor antigüedad en la zona hay mayor número de representantes de servicio al cliente y 5) hay una mayor variabilidad del gasto de nómina mercados potencialmente grandes, que en mercados potencialmente pequeños.



### 3. Verify Data Quality

Al revisar la data no detectaron valores nulos ni información faltante tanto para observaciones como para variables, por lo que se puede concluir que la calidad de los datos es sólida. La revisión de las variables escalares indica que no hay presencia de ceros, pero si se identifica que de las 1600 observaciones hay 90 con un valor de gasto en nómina mayor a \$60.000 unidades monetarias y hay 4 observaciones con gasto en publicidad impresa mayor a \$40.000.

## DATA PREPARATION

### 1. Clean and Construct data

Debido a que la revisión de la data indica que la calidad de datos es sólida no es necesario realizar ningún tipo de limpieza a la data. Aun así, debido a que este proyecto se enfoca en un modelo de aprendizaje supervisado es necesario dividir la base de datos en dos, una parte para entrenar el modelo y la otra para realizar la evaluación de este. Y dentro del marco del modelo supervisado, en este caso usando una regresión lineal, es necesario transformar variables con el fin de identificar una relación que explique el comportamiento de la variable a predecir a partir de las variables explicativas.

- Transformación de variables

Se realizan las siguientes transformaciones que van a ser evaluadas en diversas iteraciones de la regresión lineal.

- Variables dummies para categóricas
- Variable de interacción entre correo y páginas ( $\text{correo\_páginas} = \text{correo} * \text{páginas}$ )
- Variable de interacción de impresa y IDMercado ( $\text{impresa\_IDMercado} = \text{impresa} * \text{IDMercado}$ )
- Cuadrado de variables numéricas (variables con sufijo “\_sqr”)
- Raíz cuadrada de variables numéricas (variables con prefijo “sqrt\_”)
- Inversa de variables numéricas (variables con sufijo “\_inversa”)

- Inversa del cuadrado de variables numéricas (variables con sufijo “\_inversa\_sqr”)
  - Logaritmo natural de variables numéricas (variables con prefijo “ln\_”)
  - Exponencial de variables numéricas (variables con prefijo “e\_”)
- Creación de *dataset train* y *dataset test*  
Una vez creadas las nuevas posibles variables y teniendo las variables originales, se procede a dividir la base de datos en una porción para entrenamiento y otra para evaluación. Se decide, dada la cantidad de datos, hacer una separación aleatoria 80%-20%, siendo 80% de las observaciones usadas para entrenamiento y el restante 20% para la evaluación del modelo.

## 2. Dataset description

De acuerdo a lo mencionado se tendrán dos *datasets*, el primero con el 80% (1.280) de las observaciones escogidas al azar para fines de entrenamiento del modelo y el segundo con el 20% (320) de las observaciones para evaluar el modelo. Ambos *datasets* cuentan con las mismas variables originales y con las mismas variables transformadas mencionadas en el punto anterior, lo que constituye que ambos datasets tienen 64 variables.

## MODELING

### 1. Select modeling techniques:

Para este proyecto se elige usar una regresión lineal porque este tipo de modelo supervisado permite una interpretación relativamente sencilla de la relación entre variable dependiente y variables explicativas. Además, es un modelo que no impone restricciones a los coeficientes, lo cual resulta útil en este caso, ya que se parte del supuesto de que todas las variables explicativas tienen cierta relevancia para la predicción de la variable dependiente. Finalmente la regresión lineal, en este caso, también sirve porque puede ser usada como modelo de referencia para comparar rendimiento de modelos un poco más avanzados.

Con el fin obtener las variables independientes que van a ir en las diversas iteraciones de la regresión, y teniendo en cuenta que se realizaron diversas transformaciones las variables originales, se aplicará el método *stepwise*. Este método, usando la heurística del vecino más cercano, indica qué variables independientes deberán incluirse en el modelo de regresión y cuáles se deberán excluir. Esto es importante porque automatiza el proceso de selección de características, se puede ir mejorando el rendimiento de la regresión puede mitigar problemas de multicolinealidad que pueden existir en nuestros datasets.

En conjunto con la regresión lineal, se aplicarán los métodos de regularización, Ridge y Lasso, con el fin abordar el problema del sobreajuste (overfitting) y mejorar el rendimiento de algunos de los modelos de regresión planteados. Se evaluarán las regresiones utilizando Ridge porque con este método se puede reducir el impacto de la multicolinealidad entre variables predictoras, puede prevenir el overfitting al reducir los coeficientes y no realiza *feature selection* sino que reduce los coeficientes a cero. Por otro lado, se evaluarán las regresiones utilizando Lasso porque éste método selecciona automáticamente los predictores más relevantes, descartando el resto y haciendo una *feature selection*.

### 2. Generate test design:

Para el diseño y evaluación de las regresiones de datos, se comenzó por revisar las variables originales y se propusieron modelos de regresión lineal utilizando solo estas variables, considerando el conocimiento del negocio. Luego, se llevaron a cabo transformaciones graduales en las variables originales, generando iteraciones adicionales de regresión lineal. Para seleccionar las variables más relevantes, se aplicó el método de *stepwise*. Posteriormente, se aplicaron técnicas de regularización mencionadas anteriormente Ridge, Lasso y redes elásticas a las regresiones con el menor error cuadrático medio (RMSE). Se compararon los modelos de regresión en función del RMSE del *dataset* de evaluación y se seleccionaron las tres mejores regresiones para una evaluación más detallada. Este proceso permitió optimizar y refinar el modelo de regresión para mejorar su capacidad predictiva. En el Anexo 1 se pueden ver cada una de las 35 iteraciones realizadas con sus métricas correspondientes.

### 3. Build model:

De acuerdo al proceso mencionado se evaluaron 35 iteraciones de una regresión lineal usando las variables explicativas originales así como transformaciones de las mismas. Del proceso se seleccionaron los siguientes tres modelos conforme al criterio de selección del menor error cuadrático medio (RMSE).

Modelo	Iteración	r2	r2_Adj	AIC	DWTest	DWTest pvalue	RMSE
1	31	0.6557148	0.6474454	27360.60	2.060.512	0.8591231	9.487.465
2	20	0.6542957	0.6468404	27359.87	2.061.168	0.8630361	9.631.867
3	21	0.6466272	0.6398694	27381.95	2.053.620	0.8310342	9.761.693

#### Resultados Modelo 1

```
Call:
lm(formula = ropamujer ~ idloc + edadloc + correo + telefono +
  impresa + servicio + correo_paginas + correo_sqr + telefono_sqr +
  impresa_sqr + sqrt_edadloc + sqrt_correo + sqrt_telefono +
  sqrt_impresa + sqrt_servicio + edadloc_inversa + correo_inversa +
  paginas_inversa + telefono_inversa + impresa_inversa + servicio_inversa +
  edadloc_inversa_sqr + correo_inversa_sqr + servicio_inversa_sqr +
  ln_edadloc + ln_correo + ln_telefono + ln_servicio + e_adadloc +
  ln_nomina, data = data_train)
```

Residuals:

Min	1Q	Median	3Q	Max
-32140	-7364	-820	6395	47952

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	3.879e+07	2.583e+07	1.192	0.2336
idloc	-2.654e+00	5.122e-01	-5.181	2.58e-07 ***
edadloc	5.245e+04	8.003e+03	6.554	8.17e-11 ***
correo	-6.624e+02	3.859e+02	-1.716	0.0883
telefono	6.136e+05	3.655e+05	1.679	0.0934
impresa	2.597e+02	1.389e+02	1.870	0.0617
servicio	-6.505e+04	4.188e+04	-1.556	0.1199
correo_paginas	2.693e+02	6.120e+03	4.480	1.18e-05 ***
correo_sqr	4.696e+03	2.757e+03	1.703	0.0888
telefono_sqr	-9.568e+02	5.861e+02	-1.632	0.1028
impresa_sqr	-1.836e+03	5.518e+04	-1.880	0.0604
sqrt_edadloc	-8.921e+05	1.287e+05	-6.930	6.70e-12 ***
sqrt_correo	2.840e+05	1.195e+05	2.370	0.0180
sqrt_telefono	-1.280e+07	7.553e+06	-1.695	0.0904
sqrt_impresa	-7.612e+04	4.182e+04	-1.856	0.0637
sqrt_servicio	2.453e+06	1.576e+06	1.556	0.1199
edadloc_inversa	1.534e+06	2.161e+05	7.097	2.13e-12 ***
correo_inversa	-4.533e+09	2.818e+09	-1.602	0.1095
paginas_inversa	6.115e+05	4.840e+05	1.513	0.1304
telefono_inversa	7.507e+07	4.394e+07	1.709	0.0878
impresa_inversa	-2.091e+10	1.141e+10	-1.833	0.0671
servicio_inversa	-7.005e+07	4.634e+07	-1.512	0.1309
edadloc_inversa_sqr	-5.044e+05	7.418e+04	-6.800	6.2e-11 ***
correo_inversa_sqr	1.144e+12	7.754e+11	1.475	0.1405
servicio_inversa_sqr	2.140e+08	1.418e+08	1.510	0.1313
ln_edadloc	1.821e+06	1.424e+05	7.172	1.27e-12 ***
ln_correo	-4.948e+06	2.936e+06	-1.685	0.0922
ln_telefono	2.087e+07	1.224e+07	1.705	0.0885
ln_servicio	-6.611e+06	4.293e+06	-1.540	0.1238
e_adadloc	-8.040e-08	4.281e-08	-1.914	0.0599
ln_nomina	-5.584e+02	3.907e+02	-1.429	0.1532

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 10470 on 1249 degrees of freedom  
Multiple R-squared: 0.6557, Adjusted R-squared: 0.6474  
F-statistic: 79.29 on 38 and 1249 DF, p-value: < 2.2e-16

#### Resultados Modelo 2

```
Call:
lm(formula = ropamujer ~ idloc + edadloc + correo + impresa +
  servicio + nomina + idmercado + correo_paginas + impresa_idmercado +
  correo_sqr + correo_sqr + nomina_sqr + ln_edadloc + ln_correo +
  ln_servicio + ln_nomina + sqrt_nomina + sqrt_edadloc + sqrt_correo +
  sqrt_impresa + sqrt_servicio + edadloc_inversa + correo_inversa +
  paginas_inversa + impresa_inversa + edadloc_inversa_sqr +
  correo_inversa_sqr, data = data_train)
```

Residuals:

Min	1Q	Median	3Q	Max
-32046	-7124	-960	6405	47372

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	4.002e+07	1.759e+07	2.275	0.0231 *
idloc	-2.661e+00	5.124e-01	-5.193	2.41e-07 ***
edadloc	4.293e+04	6.225e+03	6.896	8.44e-12 ***
correo	-6.947e+02	3.503e+02	-1.983	0.0476 *
impresa	2.282e+02	1.386e+02	1.646	0.1000
servicio	-2.581e+03	1.844e+03	-1.616	0.1063
nomina	1.457e+09	9.382e+01	1.567	0.1175
idmercado	-7.768e+02	5.502e+02	-1.412	0.1583
correo_paginas	2.699e+02	5.998e+03	4.501	7.40e-06 ***
impresa_idmercado	2.974e+02	1.855e+02	1.604	0.1091
impresa_sqr	-9.876e+04	5.498e+04	-1.651	0.0990
correo_sqr	4.830e+03	2.588e+03	1.926	0.0544
nomina_sqr	-3.725e+06	2.524e+06	-1.476	0.1403
ln_edadloc	8.673e+05	1.167e+05	7.430	2.01e-13 ***
ln_correo	-5.283e+06	2.672e+06	-1.977	0.0483 *
ln_servicio	-1.590e+05	6.680e+04	-2.381	0.0174 *
ln_nomina	1.539e+04	1.060e+04	1.451	0.1470
sqrt_nomina	-6.876e+02	3.902e+02	-1.757	0.1196
sqrt_edadloc	-7.462e+05	1.029e+05	-7.248	7.35e-13 ***
sqrt_correo	2.160e+05	1.085e+05	1.990	0.0468 *
sqrt_impresa	-6.702e+04	4.095e+04	-1.636	0.1020
sqrt_servicio	9.694e+04	4.493e+04	2.158	0.0312 *
edadloc_inversa	1.321e+06	1.844e+05	7.163	1.34e-12 ***
correo_inversa	-4.890e+09	2.583e+09	-1.893	0.0586
paginas_inversa	6.850e+05	4.800e+05	1.515	0.1301
impresa_inversa	-1.853e+10	1.139e+10	-1.626	0.1041
edadloc_inversa_sqr	-4.361e+05	6.486e+04	-6.725	2.67e-11 ***
correo_inversa_sqr	1.256e+12	6.172e+11	1.752	0.0801

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 10480 on 1252 degrees of freedom  
Multiple R-squared: 0.6543, Adjusted R-squared: 0.6468  
F-statistic: 87.76 on 27 and 1252 DF, p-value: < 2.2e-16

#### Resultados Modelo 3

```
Call:
lm(formula = ropamujer ~ edadloc + correo + impresa + servicio +
  idmercado + correo_paginas + impresa_idmercado + impresa_sqr +
  correo_sqr + ln_edadloc + ln_correo + ln_servicio + sqrt_edadloc +
  sqrt_correo + sqrt_impresa + sqrt_servicio + edadloc_inversa +
  correo_inversa + paginas_inversa + impresa_inversa + servicio_inversa +
  edadloc_inversa_sqr + correo_inversa_sqr + servicio_inversa_sqr,
  data = data_train2)
```

Residuals:

Min	1Q	Median	3Q	Max
-30277	-7225	-587	6471	48820

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	4.691e+07	2.003e+07	2.342	0.0193 *
edadloc	4.310e+04	6.305e+03	6.836	1.27e-11 ***
correo	-5.742e+02	3.524e+02	-1.629	0.1035
impresa	2.173e+02	1.399e+02	1.553	0.1206
servicio	-6.624e+04	4.204e+04	-1.575	0.1154
idmercado	-7.843e+02	5.543e+02	-1.415	0.1573
correo_paginas	2.701e+02	6.043e+03	4.469	8.58e-06 ***
impresa_idmercado	2.961e+02	1.869e+02	1.584	0.1134
impresa_sqr	-8.702e+04	5.547e+04	-1.569	0.1170
correo_sqr	3.961e+03	2.522e+03	1.570	0.1166
ln_edadloc	8.691e+05	1.181e+05	7.358	3.35e-13 ***
ln_correo	-4.376e+06	2.689e+06	-1.627	0.1040
ln_servicio	-6.643e+06	4.319e+06	-1.538	0.1243
sqrt_edadloc	-7.486e+05	1.042e+05	-7.183	1.17e-12 ***
sqrt_correo	1.789e+05	1.092e+05	1.638	0.1017
sqrt_impresa	-6.361e+04	4.132e+04	-1.539	0.1240
sqrt_servicio	2.482e+06	1.585e+06	1.565	0.1177
edadloc_inversa	1.319e+06	1.863e+05	7.077	2.45e-12 ***
correo_inversa	-4.038e+09	2.601e+09	-1.553	0.1207
paginas_inversa	6.114e+05	4.028e+05	1.518	0.1293
impresa_inversa	-1.749e+10	1.150e+10	-1.521	0.1284
servicio_inversa	-6.928e+07	4.664e+07	-1.485	0.1377
edadloc_inversa_sqr	-4.340e+05	6.552e+04	-6.624	5.18e-11 ***
correo_inversa_sqr	1.034e+12	7.224e+11	1.431	0.1526
servicio_inversa_sqr	2.083e+08	1.428e+08	1.459	0.1449

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 10580 on 1255 degrees of freedom  
Multiple R-squared: 0.6466, Adjusted R-squared: 0.6399  
F-statistic: 95.69 on 24 and 1255 DF, p-value: < 2.2e-16

## EVALUATION:

### 1. Assess model:

Teniendo en cuenta el objetivo de *data mining* del proyecto el cual es escoger aquella regresión lineal con el *menor error cuadrático medio*, el modelo elegido es el modelo 1. Este modelo tiene la forma de:

$$Y = \beta_0 + \beta_1 idloc + \beta_2 edadloc + \beta_3 correo + \beta_4 telefono + \beta_5 impresa + \beta_6 servicio + \beta_7 (correo \text{ paginas}) + \beta_8 (correo)^2 + \beta_9 (telefono)^2 + \beta_{10} (impresa)^2 + \beta_{11} \sqrt{edadloc} + \beta_{12} \sqrt{correo} + \beta_{13} \sqrt{telefono} + \beta_{14} \sqrt{impresa} + \beta_{15} \sqrt{servicio} + \beta_{16} \left( \frac{1}{edadloc} \right) + \beta_{17} \left( \frac{1}{correo} \right) + \beta_{18} \left( \frac{1}{paginas} \right) + \beta_{19} \left( \frac{1}{telefono} \right) + \beta_{20} \left( \frac{1}{impresa} \right) + \beta_{21} \left( \frac{1}{servicio} \right) + \beta_{22} \left( \frac{1}{edadloc^2} \right) + \beta_{23} \left( \frac{1}{correo^2} \right) + \beta_{24} \left( \frac{1}{servicio^2} \right) + \beta_{25} \ln(edadloc) + \beta_{26} \ln(correo) + \beta_{27} \ln(telefono) + \beta_{28} \ln(servicio) + \beta_{29} e^{edadloc} + \beta_{30} \ln(nomina) + \epsilon$$

Siendo los betas iguales a los coeficientes que aparecen en el resultado mostrado en el anterior punto.

De este modelo, se puede concluir que las relaciones más significativas, de acuerdo a la magnitud del coeficiente son:

- Relaciones de los inversos de variables explicativas con la variable dependiente:
  - Coefficientes negativos: Aumentos en el valor del gasto en publicidad impresa, catálogos enviados, y representantes de servicio conllevan a aumentos pequeños en el valor de ventas de prendas para mujer.



- b. Coeficientes positivos: Aumentos en el número promedio de líneas para llamada abiertas para pedidos, número de años de antigüedad de la zona y número de páginas en el catálogo conllevan pequeñas disminuciones en el valor de ventas de prendas para mujer.
- Relaciones de las raíces de variables explicativas con la variable independiente:
  - a. Coeficientes negativos: Aumento en el número promedio de líneas para llamada abiertas para pedidos, número de años de antigüedad de la zona o en el valor del gasto en publicidad impresa están asociados con un disminuciones en la variable de ventas de prendas para mujer.
  - b. Coeficientes positivos: Aumento en catálogos enviados, y representantes de servicio están asociados aumentos en la variable de ventas de prendas para mujer.
- Relaciones de logaritmos naturales de variables explicativas con la variable independiente:
  - a. Coeficientes negativos: Aumentos en catálogos enviados, y representantes de servicio, en términos porcentuales, se asocian con una disminución en la variable de ventas de prendas para mujer
  - b. Coeficientes positivos: Aumento en el número promedio de líneas para llamada abiertas para pedidos, número de años de antigüedad de la zona, en términos porcentuales, indican una disminución en la variable de ventas de prendas para mujeres.

## 2. Produce final report:

El análisis de la regresión lineal muestra que los recursos que tienen mayor impacto en la demanda son el número promedio de líneas para llamada abiertas para pedidos, el número de años de antigüedad de la zona y el número de páginas en el catálogo. Por lo tanto, Leonisa debe optimizar la asignación de estos recursos para maximizar la demanda. En concreto, la empresa puede aumentar el número promedio de líneas para llamada abiertas para pedidos en las zonas con mayor potencial de demanda y priorizar la asignación de recursos (representantes y catálogos) en las zonas con mayor antigüedad.

### I. Estrategia y análisis 1

- **Estrategia:** Aumentar el número promedio de líneas para llamada abiertas para pedidos en las zonas con mayor potencial de demanda
- **Costo:** Aumentar el número de líneas, Leonisa necesitará expandir su infraestructura actual y aumentar el número de representantes que atiendan las líneas.
- **Beneficio:** Se identifica que si hay aumentos en las líneas abiertas donde el mercado es potencialmente grande, Leonisa puede tener un incremento en las ventas, esto a razón de que, al tener más líneas disponibles para atender pedidos, Leonisa podrá atender a un mayor número de clientes, lo que podría traducirse en un aumento en las ventas. Así mismo se puede mejorar la en la experiencia del cliente ya que tener más líneas abiertas se puede reducir el tiempo de espera para atender un pedido, y se ofrecerá una mejor experiencia al cliente, lo que podría traducirse en una mayor satisfacción y lealtad de los clientes.
- **Mecanismo de evaluación:** Leonisa puede medir el aumento en las ventas en las zonas donde se implementó la estrategia, medir la satisfacción del cliente por medio de la realización de encuestas a los clientes para medir su satisfacción y probabilidad de volver a comprar y finalmente Leonisa puede medir el tiempo de espera promedio para atender un pedido y ver si esto aumenta la conversión.

### II. Estrategia y análisis 2

- **Estrategia:** Priorizar la asignación de recursos (representantes y catálogos) en las zonas con mayor antigüedad.
- **Costos:** Al centrarse en las zonas con mayor antigüedad, Leonisa puede perder oportunidades de expansión en áreas nuevas y potencialmente rentables. Así mismo puede haber un punto de saturación en las zonas más antiguas donde los consumidores ya tienen sus proveedores preferidos y, por ende, la competencia puede ser más intensa en estas áreas.
- **Beneficios:** De los posibles beneficios se puede destacar que al priorizar la oferta y atención en las zonas con mayor antigüedad (es decir donde probablemente hay una base de clientes más establecida) se puede fortalecer la relación y la lealtad del cliente a lo largo del tiempo. De la misma forma, enfocarse en estas zonas, es probable que se logren mayores ventas y una mejor retención de clientes, ya que estos consumidores pueden tener una mayor probabilidad de realizar compras repetidas. Finalmente, al concentrar recursos en áreas con historial de ventas sólido, la empresa puede optimizar sus recursos y reducir costos al evitar la expansión en áreas menos probables de tener éxito.

- **Mecanismo de evaluación:** Primeramente, se debe realizar un seguimiento continuo de las ventas en las zonas con mayor antigüedad para evaluar si la estrategia está generando un aumento en las ventas y la retención de clientes. De la misma forma hay que obtener la opinión de los clientes en estas áreas, ya que ellos pueden proporcionar información valiosa sobre su satisfacción y la efectividad de la estrategia. Además, es importante comparar los costos asociados con la implementación de la estrategia con los beneficios en términos de aumento de ventas y lealtad del cliente. Finalmente hay que evaluar cómo la competencia responde a la estrategia y si están ganando terreno en nuevas áreas.

### ***III. Potenciales siguientes pasos***

Para ambas estrategias es importante considerar los siguientes pasos:

- **Análisis de comportamiento de transacciones:** Realizar un análisis más detallado del comportamiento de transacciones de los clientes. Esto permitirá reconocer si hay aumentos de venta, la rentabilidad de las estrategias y calcular posibles retornos a la inversión.
- **Análisis competitivo:** Realizar un análisis comparativo de otros catálogos en el mercado que compiten con los propios. Evaluar cómo se posicionan éstos en relación con los catálogos propios y su acogida por zonas.
- **Inclusión de data adicional:** Es importante que para futuros análisis se incluyan otras variables aparte de las que ya se tienen para predecir demanda. Por ejemplo, el tipo de artículo es importante porque se ha identificado que el interés por ciertos tipos de prendas impulsan el crecimiento del mercado y de la misma forma se puede segmentar de manera más sencilla.

## ANEXO 1

Iteración	Modelo	r <sup>2</sup>	r <sup>2</sup> _Adj	AIC	DWTest Estadístico	DWTest Pvalue	RMSE
1	$Y = \beta_0 + \beta_1 \text{idloc} + \beta_2 \text{edadloc} + \beta_3 \text{correo} + \beta_4 \text{paginas} + \beta_5 \text{teléfono} + \beta_6 \text{impresa} + \beta_7 \text{servicio} + \beta_8 \text{nomina} + \beta_9 \text{idmercado} + \beta_{10} \text{tamamer}_{\text{Median}} + \beta_{11} \text{tamamer}_{\text{Pequeño}} + \beta_{12} \text{promo} + \varepsilon$	0,5832767	0,5793298	27569,02	2.023.499	0,6633444	10.814,1
2	$Y = \beta_0 + \beta_1 \text{edadloc} + \beta_2 \text{correo} + \beta_3 \text{impresa} + \beta_4 \text{servicio} + \beta_5 \text{nomina} + \varepsilon$	0,5537848	0,5520335	27642,55	2.009.638	0,5680701	11.097,2
3	$Y = \beta_0 + \beta_1 \text{edadloc} + \beta_2 \text{correo} + \beta_3 \text{impresa} + \beta_4 \text{servicio} + \beta_5 \text{nomina} + \beta_6 \text{promo} + \varepsilon$	0,5540719	0,5519702	27643,72	2.004.838	0,5347359	11.112,4
4	$Y = \beta_0 + \beta_1 \text{edadloc} + \beta_2 \text{correo} + \beta_3 \text{paginas} + \beta_4 \text{teléfono} + \beta_5 \text{impresa} + \beta_6 \text{servicio} + \beta_7 \text{nomina} + \beta_8 \text{idmercado} + \beta_9 \text{promo} + \varepsilon$	0,5746876	0,5716736	27589,14	2.019.670	0,6385938	11.017,8
5	$Y = \beta_0 + \beta_1 \text{edadloc} + \beta_2 \text{correo} + \beta_3 \text{paginas} + \beta_4 \text{teléfono} + \beta_5 \text{impresa} + \beta_6 \text{servicio} + \beta_7 \text{nomina} + \beta_8 \text{idmercado} + \beta_9 \text{promo} + \beta_{10} \text{tamamer} + \varepsilon$	0,5750332	0,5713466	27592,1	2.019.589	0,6377261	11.045,6
6	$Y = \beta_0 + \beta_1 \text{idloc} + \beta_2 \text{correo} + \beta_3 \text{paginas} + \beta_4 \text{impresa} + \beta_5 \text{servicio} + \varepsilon$	0,581949	0,5803083	27559,09	2.019.535	0,6368852	10.724,2
7	$Y = \beta_0 + \beta_1 \text{idloc} + \beta_2 \text{correo} + \beta_3 \text{paginas} + \beta_4 \text{impresa} + \beta_5 \text{servicio} + \varepsilon$	0,581949	0,5803083	27559,09	2.019.535	0,6368852	10.724,2
8	$Y = \beta_0 + \beta_1 \text{edadloc} + \beta_2 \text{correo} + \beta_3 \text{paginas} + \beta_4 \text{teléfono} + \beta_5 \text{impresa} + \beta_6 \text{servicio} + \beta_7 \text{nomina} + \beta_8 \text{idmercado}_1 + \beta_9 \text{idmercado}_2 + \beta_{10} \text{idmercado}_3 + \beta_{11} \text{idmercado}_4 + \beta_{12} \text{idmercado}_5 + \beta_{13} \text{idmercado}_6 + \beta_{14} \text{idmercado}_7 + \beta_{15} \text{idmercado}_8 + \beta_{16} \text{promo}_1 + \beta_{17} \text{promo}_2 + \beta_{18} \text{tamamer}_\text{Grande} + \beta_{19} \text{tamamer}_\text{Median} + \varepsilon$	0,5772756	0,5709011	27601,32	2.020.524	0,6447635	11.027,8
9	$Y = \beta_0 + \beta_1 \text{correo} + \beta_2 \text{paginas} + \beta_3 \text{impresa} + \beta_4 \text{servicio} + \beta_5 \text{id\_mercado}_7 + \varepsilon$	0,575533	0,5731971	27582,59	2.020.142	0,6416398	10.942,7
10	$Y = \beta_0 + \beta_1 \text{correo} + \beta_2 \text{paginas} + \beta_3 \text{impresa} + \beta_4 \text{servicio} + \varepsilon$	0,5733623	0,5720238	27583,12	2.018.381	0,629385	10.936,5
11	$Y = \beta_0 + \beta_1 \text{correo} + \beta_2 \text{impresa} + \beta_3 \text{servicio} + \beta_4 \text{idmercado}_7 + \beta_5 \text{correo\_paginas} + \varepsilon$	0,5764694	0,5741387	27579,76	2.019.405	0,6369186	10.820,9
12	$Y = \beta_0 + \beta_1 \text{correo} + \beta_2 \text{impresa} + \beta_3 \text{servicio} + \beta_4 \text{idmercado} + \beta_5 \text{correo\_paginas} + \beta_6 \text{impresa\_idmercado} + \varepsilon$	0,5759021	0,5735682	27581,48	2.012.416	0,5892474	10.859,1
13	$Y = \beta_0 + \beta_1 \text{edadloc} + \beta_2 \text{correo} + \beta_3 \text{paginas} + \beta_4 \text{impresa} + \beta_5 \text{servicio} + \beta_6 \text{nomina} + \beta_7 \text{idmercado} + \beta_8 \text{correo\_paginas} + \beta_9 \text{impresa\_idmercado} + \beta_{10} \text{impresa}^2 + \beta_{11} \text{correo}^2 + \beta_{12} \text{nomina}^2 + \beta_{13} \text{promo}_1 + \beta_{14} \text{promo}_2 + \beta_{15} \text{tamamer}_\text{Grande} + \beta_{16} \text{tamamer}_\text{Median} + \varepsilon$	0,5781418	0,5724591	27594,7	2.023.207	0,6617146	11.029,5

14	$Y = \beta_0 + \beta_1 \text{idloc} + \beta_2 \text{correo} + \beta_3 \text{impresa} + \beta_4 \text{servicio} + \beta_5 \text{idmercado} + \beta_6 \text{correo\_paginas} + \beta_7 \text{impresa\_idmercado} + \epsilon$	0,58403 43	0,58174 52	27556,6 9	2.017.57 4	0,62402 9	10.600,3
15	$Y = \beta_0 + \beta_1 \text{idloc} + \beta_2 \text{edadloc} + \beta_3 \text{correo} + \beta_4 \text{paginas} + \beta_5 \text{impresa} + \beta_6 \text{servicio} + \beta_7 \text{idmercado} + \beta_8 \text{impresa\_idmercado} + \beta_9 \ln\_edadloc + \beta_{10} \ln\_paginas + \beta_{11} \ln\_servicio + \epsilon$	0,62318 51	0,61991 62	27438,1 7	2.050.35 7	0,81660 54	10.112,1
16	$Y = \beta_0 + \beta_1 \text{edadloc} + \beta_2 \text{correo} + \beta_3 \text{paginas} + \beta_4 \text{impresa} + \beta_5 \text{servicio} + \beta_6 \text{idmercado} + \beta_7 \text{impresa\_idmercado} + \beta_8 \ln\_edadloc + \beta_9 \ln\_paginas + \beta_{10} \ln\_servicio + \epsilon$	0,61526 16	0,61222 98	27462,8	2.047.23 7	0,80162 49	10.277,2
17	$Y = \beta_0 + \beta_1 \text{idloc} + \beta_2 \text{edadloc} + \beta_3 \text{correo} + \beta_4 \text{impresa} + \beta_5 \text{servicio} + \beta_6 \text{idmercado} + \beta_7 \text{impresa\_idmercado} + \beta_8 \ln\_edadloc + \beta_9 \ln\_paginas + \beta_{10} \ln\_servicio + \beta_{11} \sqrt{\text{edadloc}} + \beta_{12} \sqrt{\text{paginas}} + \beta_{13} \sqrt{\text{servicio}} + \epsilon$	0,63324 43	0,62947 82	27407,5 3	2.068.70 7	0,89074 32	9.997,1
18	$Y = \beta_0 + \beta_1 \text{idloc} + \beta_2 \text{edadloc} + \beta_3 \text{telefono} + \beta_4 \text{impresa} + \beta_5 \text{impresa\_sqr} + \beta_6 \ln\_edadloc + \beta_7 \ln\_correo + \beta_8 \ln\_paginas + \beta_9 \ln\_telefono + \beta_{10} \ln\_servicio + \beta_{11} \sqrt{\text{edadloc}} + \beta_{12} \sqrt{\text{correo}} + \beta_{13} \sqrt{\text{telefono}} + \beta_{14} \sqrt{\text{impresa}} + \beta_{15} \text{edadloc\_inversa} + \beta_{16} \text{correo\_inversa} + \beta_{17} \text{paginas\_inversa} + \beta_{18} \text{telefono\_inversa} + \beta_{19} \text{impresa\_inversa} + \beta_{20} \text{servicio\_inversa} + \epsilon$	0,64055 55	0,63484 55	27395,7 6	2.054.35 1	0,83391 12	9.912,4
19	$Y = \beta_0 + \beta_1 \text{edadloc} + \beta_2 \text{impresa} + \beta_3 \text{servicio} + \beta_4 \text{idmercado} + \beta_5 \text{impresa\_idmercado} + \beta_6 \text{impresa\_sqr} + \beta_7 \ln\_edadloc + \beta_8 \ln\_correo + \beta_9 \ln\_paginas + \beta_{10} \ln\_servicio + \beta_{11} \sqrt{\text{edadloc}} + \beta_{12} \sqrt{\text{correo}} + \beta_{13} \sqrt{\text{impresa}} + \beta_{14} \sqrt{\text{servicio}} + \beta_{15} \text{edadloc\_inversa} + \beta_{16} \text{paginas\_inversa} + \beta_{17} \text{impresa\_inversa} + \epsilon$	0,63304 21	0,62809 89	27416,2 4	2.058.14 2	0,85099 88	10.140,5
20	$Y = \beta_0 + \beta_1 \text{idloc} + \beta_2 \text{edadloc} + \beta_3 \text{correo} + \beta_4 \text{impresa} + \beta_5 \text{servicio} + \beta_6 \text{nomina} + \beta_7 \text{idmercado} + \beta_8 \text{correo\_paginas} + \beta_9 \text{impresa\_idmercado} + \beta_{10} \text{impresa\_sqr} + \beta_{11} \text{correo\_sqr} + \beta_{12} \text{nomina\_sqr} + \beta_{13} \ln\_edadloc + \beta_{14} \ln\_correo + \beta_{15} \ln\_servicio + \beta_{16} \ln\_nomina + \beta_{17} \sqrt{\text{nomina}} + \beta_{18} \sqrt{\text{edadloc}} + \beta_{19} \sqrt{\text{correo}} + \beta_{20} \sqrt{\text{impresa}} + \beta_{21} \sqrt{\text{servicio}} + \beta_{22} \text{edadloc\_inversa} + \beta_{23} \text{correo\_inversa} + \beta_{24} \text{paginas\_inversa} + \beta_{25} \text{impresa\_inversa} + \beta_{26} \text{edadloc\_inversa\_sqr} + \beta_{27} \text{correo\_inversa\_sqr} + \epsilon$	0,65429 57	0,64684 04	27359,8 7	2.061.16 8	0,86303 61	9.631,9
21	$Y = \beta_0 + \beta_1 \text{edadloc} + \beta_2 \text{correo} + \beta_3 \text{impresa} + \beta_4 \text{servicio} + \beta_5 \text{idmercado} + \beta_6 \text{correo\_paginas} + \beta_7 \text{impresa\_idmercado} + \beta_8 \text{impresa\_sqr} + \beta_9 \text{correo\_sqr} + \beta_{10} \ln\_edadloc + \beta_{11} \ln\_correo + \beta_{12} \ln\_servicio + \beta_{13} \sqrt{\text{edadloc}} + \beta_{14} \sqrt{\text{correo}} + \beta_{15} \sqrt{\text{impresa}} + \beta_{16} \sqrt{\text{servicio}} + \beta_{17} \text{edadloc\_inversa} + \beta_{18} \text{correo\_inversa} + \beta_{19} \text{paginas\_inversa} + \beta_{20} \text{impresa\_inversa} + \beta_{21} \text{servicio\_inversa} + \beta_{22} \text{edadloc\_inversa\_sqr} + \beta_{23} \text{correo\_inversa\_sqr} + \beta_{24} \text{servicio\_inversa\_sqr} + \epsilon$	0,64662 72	0,63986 94	27381,9 5	2.053.62 0	0,83103 42	9.761,7
22	Regresión lineal usando data modelo 20 y aplicando regularización Ridge  $Y = \beta_0 + \beta_1 \text{idloc} + \beta_2 \text{edadloc} + \beta_3 \text{correo} + \beta_4 \text{paginas} + \beta_5 \text{telefono} + \beta_6 \text{impresa} + \beta_7 \text{servicio} + \beta_8 \text{nomina} + \beta_9 \text{idmercado} + \beta_{10} \text{correo\_paginas} + \beta_{11} \text{impresa\_idmercado} + \beta_{12} \text{impresa\_sar} + \beta_{13} \text{correo\_sgr} + \beta_{14} \text{nomina\_sar} + \beta_{15} \text{promo\_1} + \beta_{16} \text{promo\_2} + \beta_{17} \text{tamamer\_Grande} + \beta_{18} \text{tamamer\_Median} + \beta_{19} \ln\_edadloc + \beta_{20} \ln\_correo + \beta_{21}$	NA	NA	NA	NA	NA	10.255,7

	$\text{In\_paginas} + \beta_{22} \text{In\_telefono} + \beta_{23} \text{In\_servicio} + \beta_{24} \text{In\_nomina} + \beta_{25} \text{sart\_nomina} + \beta_{26} \text{sart\_edadloc} + \beta_{27} \text{sqrt\_correo} + \beta_{28} \text{sqrt\_paginas} + \beta_{29} \text{sqrt\_telefono} + \beta_{30} \text{sqrt\_impresa} + \beta_{31} \text{sqrt\_servicio} + \beta_{32} \text{edadloc\_inversa} + \beta_{33} \text{correo\_inversa} + \beta_{34} \text{paginas\_inversa} + \beta_{35} \text{telefono\_inversa} + \beta_{36} \text{impresa\_inversa} + \beta_{37} \text{servicio\_inversa} + \beta_{38} \text{edadloc\_inversa\_sqr} + \beta_{39} \text{correo\_inversa\_sqr} + \beta_{40} \text{paginas\_inversa\_sqr} + \beta_{41} \text{telefono\_inversa\_sqr} + \beta_{42} \text{impresa\_inversa\_sqr} + \beta_{43} \text{servicio\_inversa\_sqr} + \epsilon$						
23	<p>Regresión lineal usando data modelo 20 y aplicando regularización Lasso</p> $Y = \beta_0 + \beta_1 \text{idloc} + \beta_2 \text{edadloc} + \beta_3 \text{correo} + \beta_4 \text{paginas} + \beta_5 \text{servicio} + \beta_6 \text{nomina} + \beta_7 \text{idmercado} + \beta_8 \text{correo\_paginas} + \beta_9 \text{impresa\_idmercado} + \beta_{10} \text{impresa\_sqr} + \beta_{11} \text{correo\_sqr} + \beta_{12} \text{nomina\_sqr} + \beta_{13} \ln\_edadloc + \beta_{14} \ln\_correo + \beta_{15} \ln\_servicio + \beta_{16} \ln\_nomina + \beta_{17} \text{sqrt\_nomina} + \beta_{18} \text{sqrt\_edadloc} + \beta_{19} \text{sqrt\_correo} + \beta_{20} \text{sqrt\_impresa} + \beta_{21} \text{sqrt\_servicio} + \beta_{22} \text{edadloc\_inversa} + \beta_{23} \text{correo\_inversa} + \beta_{24} \text{paginas\_inversa} + \beta_{25} \text{impresa\_inversa} + \beta_{26} \text{edadloc\_inversa\_sqr} + \beta_{27} \text{correo\_inversa\_sqr} + \beta_{28} \text{impresa\_sqr} + \epsilon$	NA	NA	NA	NA	NA	9.840,4
24	<p>Regresión lineal usando data modelo 20 y aplicando regularización redes elasticas con <math>\alpha = 0.5</math></p> $Y = \beta_0 + \beta_1 \text{idloc} + \beta_2 \text{edadloc} + \beta_3 \text{correo} + \beta_4 \text{paginas} + \beta_5 \text{impresa} + \beta_6 \text{servicio} + \beta_7 \text{correo\_paginas} + \beta_8 \text{impresa\_idmercado} + \beta_9 \text{impresa\_sqr} + \beta_{10} \text{correo\_sqr} + \beta_{11} \text{promo\_2} + \beta_{12} \ln\_edadloc + \beta_{13} \ln\_nomina + \beta_{14} \text{sqrt\_correo} + \beta_{15} \text{sqrt\_impresa} + \beta_{16} \text{sqrt\_servicio} + \beta_{17} \text{edadloc\_inversa} + \beta_{18} \text{impresa\_inversa} + \beta_{19} \text{telefono\_inversa\_sqr} + \beta_{20} \text{impresa\_inversa\_sqr} + \beta_{21} \text{servicio\_inversa\_sqr} + \epsilon$	NA	NA	NA	NA	NA	10.986,1
25	<p>Regresión lineal usando data modelo 20 y aplicando regularización Ridge excluyendo variable IDLOC</p> $\beta_1 \text{edadloc} + \beta_2 \text{correo} + \beta_3 \text{paginas} + \beta_4 \text{telefono} + \beta_5 \text{impresa} + \beta_6 \text{servicio} + \beta_7 \text{nomina} + \beta_8 \text{idmercado} + \beta_9 \text{correo\_paginas} + \beta_{10} \text{impresa\_idmercado} + \beta_{11} \text{impresa\_sqr} + \beta_{12} \text{correo\_sqr} + \beta_{13} \text{nomina\_sqr} + \beta_{14} \text{promo\_1} + \beta_{15} \text{promo\_2} + \beta_{16} \text{tamamer\_Grande} + \beta_{17} \text{tamamer\_Median} + \beta_{18} \ln\_edadloc + \beta_{19} \ln\_correo + \beta_{20} \ln\_paginas + \beta_{21} \ln\_telefono + \beta_{22} \ln\_servicio + \beta_{23} \ln\_nomina + \beta_{24} \text{sqrt\_nomina} + \beta_{25} \text{sqrt\_edadloc} + \beta_{26} \text{sqrt\_correo} + \beta_{27} \text{sqrt\_paginas} + \beta_{28} \text{sqrt\_telefono} + \beta_{29} \text{sqrt\_impresa} + \beta_{30} \text{sqrt\_servicio} + \beta_{31} \text{edadloc\_inversa} + \beta_{32} \text{correo\_inversa} + \beta_{33} \text{paginas\_inversa} + \beta_{34} \text{telefono\_inversa} + \beta_{35} \text{impresa\_inversa} + \beta_{36} \text{servicio\_inversa} + \beta_{37} \text{edadloc\_inversa\_sqr} + \beta_{38} \text{correo\_inversa\_sqr} + \beta_{39} \text{paginas\_inversa\_sqr} + \beta_{40} \text{telefono\_inversa\_sqr} + \beta_{41} \text{impresa\_inversa\_sqr} + \beta_{42} \text{servicio\_inversa\_sqr} + \epsilon$	NA	NA	NA	NA	NA	10.455,7



26	<p>Regresión lineal usando data modelo 20 y aplicando regularización Lasso excluyendo variable IDLOC</p> $Y = \beta_0 + \beta_1 \text{edadloc} + \beta_2 \text{correo} + \beta_3 \text{paginas} + \beta_4 \text{servicio} + \beta_5 \text{idmercado} + \beta_6 \text{correo\_paginas} + \beta_7 \text{impres\_idmercado} + \beta_8 \text{nomina\_sqr} + \beta_9 \text{promo\_1} + \beta_{10} \text{promo\_2} + \beta_{11} \text{tamamer\_Median} + \beta_{12} \ln\_edadloc + \beta_{13} \ln\_nomina + \beta_{14} \text{sqrt\_correo} + \beta_{15} \text{sqrt\_impres\_a} + \beta_{16} \text{sqrt\_servicio} + \beta_{17} \text{impres\_inversa} + \beta_{18} \text{edadloc\_inversa\_sqr} + \beta_{19} \text{paginas\_inversa\_sqr} + \beta_{20} \text{telefono\_inversa\_sqr} + \beta_{21} \text{servicio\_inversa\_sqr} + \epsilon$	NA	NA	NA	NA	NA	10.039,8
27	<p>Modelo 20 data sin idloc con regularización redes elasticas con alpha = 0.1 y excluyendo variable IDLOC</p> $Y = \beta_0 + \beta_1 \text{edadloc} + \beta_2 \text{correo} + \beta_3 \text{paginas} + \beta_4 \text{impres\_a} + \beta_5 \text{servicio} + \beta_6 \text{idmercado} + \beta_7 \text{correo\_paginas} + \beta_8 \text{impres\_idmercado} + \beta_9 \text{impres\_sqr} + \beta_{10} \text{correo\_sqr} + \beta_{11} \text{nomina\_sqr} + \beta_{12} \text{promo\_1} + \beta_{13} \text{promo\_2} + \beta_{14} \text{tamamer\_Median} + \beta_{15} \ln\_edadloc + \beta_{16} \ln\_correo + \beta_{17} \ln\_servicio + \beta_{18} \ln\_nomina + \beta_{19} \text{sqrt\_correo} + \beta_{20} \text{sqrt\_paginas} + \beta_{21} \text{sqrt\_impres\_a} + \beta_{22} \text{sqrt\_servicio} + \beta_{23} \text{edadloc\_inversa} + \beta_{24} \text{impres\_inversa} + \beta_{25} \text{edadloc\_inversa\_sqr} + \beta_{26} \text{paginas\_inversa\_sqr} + \beta_{27} \text{telefono\_inversa\_sqr} + \beta_{28} \text{impres\_inversa\_sqr} + \beta_{29} \text{servicio\_inversa\_sqr} + \epsilon$	NA	NA	NA	NA	NA	11.399,3
28	<p>Modelo 20 (27 vars seleccionadas a partir del stepwise) con regularización Ridge</p> $Y = \beta_0 + \beta_1 \text{idloc} + \beta_2 \text{edadloc} + \beta_3 \text{correo} + \beta_4 \text{impres\_a} + \beta_5 \text{servicio} + \beta_6 \text{nomina} + \beta_7 \text{idmercado} + \beta_8 \text{correo\_paginas} + \beta_9 \text{impres\_idmercado} + \beta_{10} \text{impres\_sqr} + \beta_{11} \text{correo\_sqr} + \beta_{12} \text{nomina\_sqr} + \beta_{13} \ln\_edadloc + \beta_{14} \ln\_correo + \beta_{15} \ln\_servicio + \beta_{16} \ln\_nomina + \beta_{17} \text{sqrt\_nomina} + \beta_{18} \text{sqrt\_edadloc} + \beta_{19} \text{sqrt\_correo} + \beta_{20} \text{sqrt\_impres\_a} + \beta_{21} \text{sqrt\_servicio} + \beta_{22} \text{edadloc\_inversa} + \beta_{23} \text{correo\_inversa} + \beta_{24} \text{paginas\_inversa} + \beta_{25} \text{impres\_inversa} + \beta_{26} \text{edadloc\_inversa\_sqr} + \beta_{27} \text{correo\_inversa\_sqr} + \epsilon$	NA	NA	NA	NA	NA	10.265,8
29	<p>Modelo 20 (28 vars seleccionadas a partir del stepwise) con regularización Lasso</p> $Y = \beta_0 + \beta_1 \text{idloc} + \beta_2 \text{edadloc} + \beta_3 \text{correo} + \beta_4 \text{impres\_a} + \beta_5 \text{servicio} + \beta_6 \text{nomina} + \beta_7 \text{idmercado} + \beta_8 \text{correo\_paginas} + \beta_9 \text{impres\_idmercado} + \beta_{10} \text{correo\_sqr} + \beta_{11} \ln\_edadloc + \beta_{12} \ln\_servicio + \beta_{13} \ln\_nomina + \beta_{14} \text{sqrt\_correo} + \beta_{15} \text{sqrt\_impres\_a} + \beta_{16} \text{correo\_inversa} + \beta_{17} \text{paginas\_inversa} + \beta_{18} \text{impres\_inversa} + \beta_{19} \text{edadloc\_inversa\_sqr} + \beta_{20} \text{correo\_inversa\_sqr} + \epsilon$	NA	NA	NA	NA	NA	9.829,8
30	<p>Modelo 20 (28 vars seleccionadas a partir del stepwise) con regularización redes elasticas con alpha = 0.6</p>	NA	NA	NA	NA	NA	10.945,9

	$Y = \beta_0 + \beta_1 \text{idloc} + \beta_2 \text{edadloc} + \beta_3 \text{correo} + \beta_4 \text{impresa} + \beta_5 \text{servicio} + \beta_6 \text{correo\_paginas} + \beta_7 \text{impresa\_idmercado} + \beta_8 \text{impresa\_sqr} + \beta_9 \ln\_edadloc + \beta_{10} \ln\_nomina + \beta_{11} \sqrt{\text{correo}} + \beta_{12} \sqrt{\text{impresa}} + \beta_{13} \text{impresa\_inversa} + \beta_{14} \text{edadloc\_inversa\_sqr} + \varepsilon$						
31	Stepwise de modelo con todo  $Y = \beta_0 + \beta_1 \text{idloc} + \beta_2 \text{edadloc} + \beta_3 \text{correo} + \beta_4 \text{telefono} + \beta_5 \text{impresa} + \beta_6 \text{servicio} + \beta_7 \text{correo\_paginas} + \beta_8 \text{correo\_sqr} + \beta_9 \text{telefono\_sqr} + \beta_{10} \text{impresa\_sqr} + \beta_{11} \sqrt{\text{edadloc}} + \beta_{12} \sqrt{\text{correo}} + \beta_{13} \sqrt{\text{telefono}} + \beta_{14} \sqrt{\text{impresa}} + \beta_{15} \sqrt{\text{servicio}} + \beta_{16} \text{edadloc\_inversa} + \beta_{17} \text{correo\_inversa} + \beta_{18} \text{paginas\_inversa} + \beta_{19} \text{telefono\_inversa} + \beta_{20} \text{impresa\_inversa} + \beta_{21} \text{servicio\_inversa} + \beta_{22} \text{edadloc\_inversa\_sqr} + \beta_{23} \text{correo\_inversa\_sqr} + \beta_{24} \text{servicio\_inversa\_sqr} + \beta_{25} \ln\_edadloc + \beta_{26} \ln\_correo + \varepsilon$	0,65571 48	0,64744 54	27360,6	2.060.51 2	0,85912 31	9.487,5
32	Modelo con todo sin idloc y stepwise  $Y = \beta_0 + \beta_1 \text{edadloc} + \beta_2 \text{telefono} + \beta_3 \text{impresa} + \beta_4 \text{servicio} + \beta_5 \text{idmercado} + \beta_6 \text{impresa\_idmercado} + \beta_7 \text{telefono\_sqr} + \beta_8 \text{impresa\_sqr} + \beta_9 \sqrt{\text{edadloc}} + \beta_{10} \sqrt{\text{correo}} + \beta_{11} \sqrt{\text{paginas}} + \beta_{12} \sqrt{\text{telefono}} + \beta_{13} \sqrt{\text{impresa}} + \beta_{14} \sqrt{\text{servicio}} + \beta_{15} \text{edadloc\_inversa} + \beta_{16} \text{paginas\_inversa} + \beta_{17} \text{telefono\_inversa} + \beta_{18} \text{impresa\_inversa} + \beta_{19} \text{servicio\_inversa} + \beta_{20} \text{edadloc\_inversa\_sqr} + \beta_{21} \text{servicio\_inversa\_sqr} + \beta_{22} \ln\_edadloc + \beta_{23} \ln\_correo + \beta_{24} \ln\_telefono + \beta_{25} \ln\_servicio + \beta_{26} \ln\_edadloc + \varepsilon$	0,64751 3	0,64019 88	27382,7 4	2.047.62 5	0,80169 21	10.859,1
33	Modelo 32 con regularización Ridge  $Y = \beta_0 + \beta_1 \text{edadloc} + \beta_2 \text{telefono} + \beta_3 \text{impresa} + \beta_4 \text{servicio} + \beta_5 \text{idmercado} + \beta_6 \text{impresa\_idmercado} + \beta_7 \text{telefono\_sqr} + \beta_8 \text{impresa\_sqr} + \beta_9 \sqrt{\text{edadloc}} + \beta_{10} \sqrt{\text{correo}} + \beta_{11} \sqrt{\text{paginas}} + \beta_{12} \sqrt{\text{telefono}} + \beta_{13} \sqrt{\text{impresa}} + \beta_{14} \sqrt{\text{servicio}} + \beta_{15} \text{edadloc\_inversa} + \beta_{16} \text{paginas\_inversa} + \beta_{17} \text{telefono\_inversa} + \beta_{18} \text{impresa\_inversa} + \beta_{19} \text{servicio\_inversa} + \beta_{20} \text{edadloc\_inversa\_sqr} + \beta_{21} \text{servicio\_inversa\_sqr} + \beta_{22} \ln\_edadloc + \beta_{23} \ln\_correo + \beta_{24} \ln\_telefono + \beta_{25} \ln\_servicio + \beta_{26} \ln\_edadloc + \varepsilon$	NA	NA	NA	NA	NA	10.488,1
34	Modelo 32 con regularización con regularización Lasso  $Y = \beta_0 + \beta_1 \text{edadloc} + \beta_2 \text{servicio} + \beta_3 \text{idmercado} + \beta_4 \text{impresa\_idmercado} + \beta_5 \text{telefono\_sqr} + \beta_6 \sqrt{\text{correo}} + \beta_7 \sqrt{\text{paginas}} + \beta_8 \sqrt{\text{telefono}} + \beta_9 \sqrt{\text{impresa}} + \beta_{10} \sqrt{\text{servicio}} + \beta_{11} \text{edadloc\_inversa} + \beta_{12} \text{paginas\_inversa} + \beta_{13} \text{impresa\_inversa}$	NA	NA	NA	NA	NA	10.069,7

	$+ \beta_{14} \text{edadloc\_inversa\_sqr} + \beta_{15} \ln\_edadloc + \beta_{16} \ln\_correo + \beta_{17} \ln\_telefono + \beta_{18} \ln\_servicio + \beta_{19} e\_edadloc + \varepsilon$						
35	<p>Modelo con todo pero sin idloc con regularización redes elasticas con <math>\alpha = 0.5</math></p> $Y = \beta_0 + \beta_1 \text{edadloc} + \beta_2 \text{impresa} + \beta_3 \text{servicio} + \beta_4 \text{impresa\_idmercado} + \beta_5 \text{telefono\_sqr} + \beta_6 \text{impresa\_sqr} + \beta_7 \text{sqr\_correo} + \beta_8 \text{sqr\_paginas} + \beta_9 \text{sqr\_impresa} + \beta_{10} \text{sqr\_servicio} + \beta_{11} \text{edadloc\_inversa} + \beta_{12} \text{impresa\_inversa} + \beta_{13} \text{edadloc\_inversa\_sqr} + \beta_{14} \text{servicio\_inversa\_sqr} + \beta_{15} \ln\_edadloc + \varepsilon$	NA	NA	NA	NA	NA	11.207,3

## REFERENCIAS

- ACOVEDI SAS. (2023). *Inicio - Acovedi*. ACOVEDI - INICIO. Recuperado 13 de octubre de 2023, de <https://acovedi.org.co/>
- Beltran, A. M., Alvarado, J., Garcia, J., Gonzalez Sanmiguel, J., Cuervo, O. A., & Jones Segura, P. (2023). *Principios de Machine Learning: regresión*. Kaggle. Recuperado 13 de octubre de 2023, de <https://kaggle.com/competitions/principios-de-machine-learning-regresion-2330>
- Davis, K., Fire, D., & Davison, H. (2012). Catalogs: The Consumers' Point of View. En *FGI Research*. Recuperado 13 de octubre de 2023, de <https://www.memberize.net/clubportal/clubdocs/2129/ACMA%20Consumer%20Survey%20Final%20release.pdf>
- EAE Business School. (2022, 11 febrero). *En el primer semestre de 2021 el consumo de moda en España creció un 23,5%, aunque se mantuvo un 25% por debajo de 2019* | EAE. EAE Business School. Recuperado 16 de octubre de 2023, de <https://www.eae.es/actualidad/noticias/en-el-primer-semestre-de-2021-el-consumo-de-moda-en-espana-crecio-un-235-aunque-se-mantuvo-un-25-por-debajo-de-2019#:~:text=EAE%20Business%20School%20ha%20publicado,facturaci%C3%B3n%20muy%20cerano%20al%202019>
- Estrada Rudas, C. (2022, 4 marzo). Consumo de moda en Colombia cerró ventas por \$27,7 billones el año pasado. *Diario La República*. Recuperado 14 de octubre de 2023, de <https://www.larepublica.co/empresas/consumo-de-moda-en-colombia-cerro-ventas-por-27-7-billones-el-ano-pasado-3316367>
- Euromonitor International. (2022). Womenswear in Colombia. En *Euromonitor International*. Recuperado 12 de octubre de 2023, de <https://www.portal.euromonitor.com/?wSe5shkCR395IzRO%2b7J8fAC9nTXrL96FqDU2RyJOfY6JnNbVzO9NGA%3d%3d>
- IMARC Group. (2023). *Women Apparel Market Size, Share, Industry Trends 2023-2028*. IMARC Groupe. Recuperado 16 de octubre de 2023, de <https://www.imarcgroup.com/women-apparel-market>
- McCoy, K. (2023, 21 agosto). Why brands like Nordstrom and Zappos are revisiting print marketing. *Digiday*. Recuperado 14 de octubre de 2023, de <https://digiday.com/marketing/why-brands-like-nordstrom-and-zappos-are-revisiting-print-marketing/>
- Portafolio. (s. f.). *Ropa cómoda impulsa el crecimiento de la industria textil en Colombia*. Portafolio.co. <https://www.portafolio.co/innovacion/ropa-comoda-impulsa-el-crecimiento-de-la-industria-textil-en-colombia-586798>
- Reagan, C. (2018, 13 septiembre). Millennials are more interested in catalogs than your grandmother is. *CNBC*. Recuperado 14 de octubre de 2023, de <https://www.cnn.com/2017/12/21/millennials-are-more-interested-in-catalogs-than-your-grandmother-is.html>
- Redacción PDM. (2023, 9 junio). *Redacción PDM*. Periódico del Meta. Recuperado 16 de octubre de 2023, de <https://periodicodelmeta.com/en-colombia-las-ventas-por-catalogo-han-crecido-cerca-de-un-8-1/>
- Ruiz, R. R. (2015, 26 enero). Catalogs, after years of decline, are revamped for changing times. *The New York Times*. Recuperado 14 de octubre de 2023, de <https://www.nytimes.com/2015/01/26/business/media/catalogs-after-years-of-decline-are-revamped-for-changing-times.html>
- Weston, S. (2023, 2 agosto). *How retailers are harnessing the power of catalogues*. Drapers. Recuperado 13 de octubre de 2023, de <https://www.drapersonline.com/insight/analysis/how-retailers-are-harnessing-the-power-of-catalogues>
- Zagalsky, A. (2023, 12 septiembre). Why fashion has reinvented the mail-order catalogue. *The Telegraph*. Recuperado 13 de octubre de 2023, de <https://www.telegraph.co.uk/fashion/shopping/why-fashion-has-reinvented-the-mail-order-catalogue/>
- Zhang, J. Z. (2020, 11 febrero). *Why catalogs are making a comeback*. Harvard Business Review. Recuperado 11 de octubre de 2023, <https://hbr.org/2020/02/why-catalogs-are-making-a-comeback>