

Taller 002 individual

Conociendo el catálogo.

Laura Juliana Mora Páez^{a,c},

Ana María Beltrán Cortés^{b,c}

^aEstudiante de Maestría en Ingeniería de Sistemas y Computación

^bProfesor, Departamento de Ingeniería Industrial

^cPontificia Universidad Javeriana, Bogotá, Colombia

1. ENTENDIMIENTO DEL NEGOCIO

Actualmente uno de los métodos con los que cuentan diferentes compañías, enfocadas en la venta de sus productos, es a través de las ventas por catálogo; esto debido a que esta es una estrategia de marketing funcional, donde se le brinda a un distribuidor un catálogo, y este comercializa los productos con su círculo social, de esta forma la compañía puede ofrecer algunos productos a menores costos, al no tener asociados gastos de establecimientos físicos, y la publicidad se maneja a través del voz a voz.ⁱ

Leonisa, es una de las varias compañías en Colombia que se dedica a la venta por catálogo, donde también destacan compañías como Avon, Yanbal, Cyson y Natura, entre muchas otras. Leonisa nace en la ciudad de Armenia (Colombia), como emprendimiento de los hermanos Urrea, ya que, en Colombia, para ese momento, no existía ninguna empresa dedicada a la producción de ropa interior femeninaⁱⁱ. Actualmente Leonisa cuenta como estrategia competitiva, el brindar un total look, es decir sus clientes pueden conseguir todas las prendas que requieran para organizar sus diferentes atuendos, desde la ropa interior, hasta el maquillaje y accesoriosⁱⁱⁱ, esto gracias a que en la actualidad Leonisa es su principal productor, y cuenta con diferentes alianzas empresariales, como MaxFactor y Vogue^{iv} con el fin de ampliar su catálogo de productos.

Sin embargo, con la aparición de la pandemia, se ha tenido que replantear la venta por catálogo, no solo por la disminución del contacto con los círculos sociales de las personas, sino a su vez con la aparición de diferentes emprendimientos y retailers de ropa, a través de redes sociales, compañías como Leonisa, debieron no solo pasar los catálogos físicos a digitales, sino a su vez abrir nuevos canales de ventas en línea, con el fin de seguir llegando a sus compradores y mantener el negocio^v.

Objetivo de negocio:

Tras la disminución de ventas a causa de la pandemia, se busca conocer predicciones de ventas para ropa de mujer, en las diferentes zonas manejadas.

- Definir posibles aspectos que influyen en las compras de los clientes.

Objetivo de minería:

A través del uso de diferentes modelos de regresión lineal, determinar las predicciones en compras.

- Encontrar un RMSE menor al determinado por la plataforma Kaggle.

2. ENTENDIMIENTO DE LOS DATOS

Se cuenta con una base de datos de 1.600 registros de compra en Leonisa, y una base con los datos de 400 registros, en las cuales se debe realizar un pronóstico de venta en la ropa para mujer. Si bien se desconoce el proceso de recolección de los datos, se conoce que las bases de datos fueron entregadas en un archivo Excel “Datos Taller individual 2 - 2210 (Estudiantes).xls”, el cual cuenta con dos hojas, una con los 1.600 registros y los 12 atributos, y la otra con los 400 registros, donde el atributo 12 (ropamujer) debe ser predicho. Cabe resaltar que los sets de datos cuentan con una estructura de tabla clásica, adicionalmente no se cuenta con datos referentes a información geográfica, sin embargo, se observa que los datos se encuentran organizados por zonas de venta.

A continuación, se presenta el diccionario de variables en la Tabla 1, con la descripción de cada uno de los atributos con los que se cuenta en ambas bases de datos, con el fin de dar mayor claridad sobre estos.

Tabla 1. Diccionario de variables.

Diccionario de variables			
Variable	Explicación	Tipo de dato	Valores posibles
Ropamujer (variable a predecir)	Ventas de prendas para mujer en millones de pesos durante el año.	Numérico/ Cuantitativa	1289- 117709
correo	Número de catálogos enviados durante el año.	Numérico/ Cuantitativa	1147- 15259
páginas	Número de páginas del catálogo.	Numérico/ Cuantitativa	51-114
teléfono	Número promedio de líneas para llamada abiertas para pedidos.	Numérico/ Cuantitativa	17-59
impresa	Cantidad gastada en publicidad impresa.	Numérico/ Cuantitativa	18061- 40028
servicio	Número de representantes del servicio al cliente.	Numérico/ Cuantitativa	15-68
idmercado	Tipo de mercado. Solo se listan números de clases o tipos de mercado que se desea atacar. No se proveerá información adicional.	Numérico / Cualitativa	1-9

tamañomer	Tamaño potencial del mercado, proyectado de acuerdo a cifras del área de marketing.	Char/ Cualitativa	"Median", "Grande", "Pequeño"
Idloc	ID de la tienda.	Numérico/ Cualitativa	1-1998
Edadloc	Años transcurridos desde la llegada a la zona.	Numérico/ Cuantitativa	1-25
promo (promoción)	Tipo de promoción que se llevó a cabo durante el año. De nuevo, información descriptiva sobre estas promociones es reservada.	Numérico/ Cualitativa	1-3
nomina	Valor total de la nómina durante el año.	Numérico	901-135882

Con el fin de explorar los datos, y tener una primera impresión de cuales datos podían influenciar más a la hora de generar las predicciones, se inició por una descripción de los atributos de carácter cuantitativo, en donde como se puede observar en la ilustración 1, la gran mayoría presenta kurtosis cercanas a 0, al igual que las desviaciones estándar en su mayoría son muy cercanas a 0, a excepción de los atributos como correo, impresa y nómina.

vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
edadloc	1	1600	10.47	7.39	9.00	9.96	8.90	1.0	25.00	24.00	0.47	-1.06 0.18
correo	2	1600	10598.68	2400.67	10480.00	10732.08	1427.74	1147.0	15259.00	14112.00	-0.97	2.73 60.02
paginas	3	1600	85.04	15.94	84.00	85.30	16.31	51.0	114.00	63.00	-0.05	-0.73 0.40
telefono	4	1600	37.64	9.75	36.00	37.15	8.90	17.0	59.00	42.00	0.42	-0.40 0.24
impresa	5	1600	28834.62	6251.17	28734.79	28782.09	7737.01	18061.2	40027.78	21966.58	0.06	-1.12 156.28
servicio	6	1600	43.65	13.24	43.00	43.66	14.83	15.0	68.00	53.00	0.01	-0.83 0.33
nomina	7	1600	24104.13	18217.35	18886.50	21349.06	14045.41	901.0	135882.00	134981.00	1.55	2.93 455.43

Ilustración 1. Medidas base atributos cuantitativos.

Adicionalmente, con el fin de observar un poco más el comportamiento de los datos, se generan diferentes boxplot (ilustración 2), donde resaltan atributos como el correo y nomina, por su cantidad de datos atípicos.

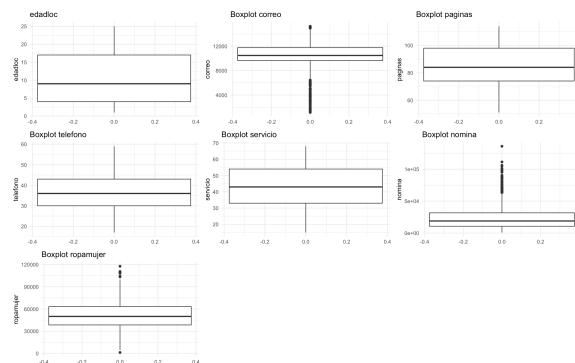


Ilustración 2. Boxplot atributos cuantitativos

Posteriormente se pasa a observar el comportamiento que tienen estas variables cuantitativas, mezcladas con alguna de las variables categóricas, y su relación con el atributo a predecir ropa mujer, como se muestra en las gráficas a continuación.

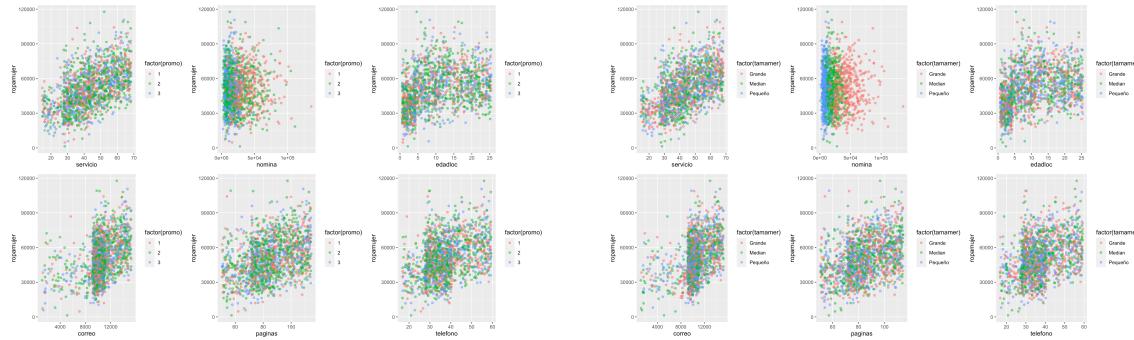


Ilustración 3. Relación ropamujer, promo, atributos cuantitativos *Ilustración 4. Relación ropamujer, tamamer, atributos cuantitativos*

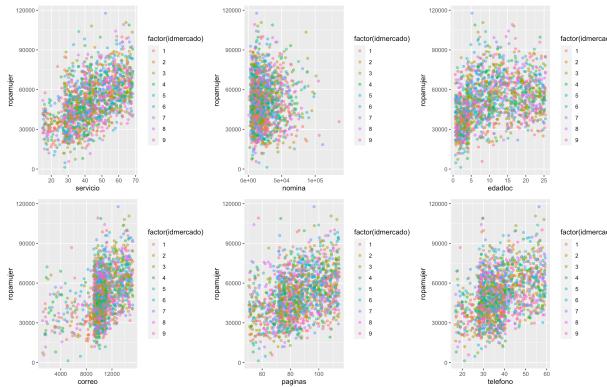


Ilustración 5. Relación ropamujer, idmercado, atributos cuantitativos

Como se puede observar en las ilustraciones anteriores, la gran mayoría de las relaciones presentan comportamientos bastante similares, exceptuando la relación de nómina con ropamujer según el atributo tamamer, ya que aquí se alcanza a evidenciar casi una sectorización de los datos, el primer cuadrante representado por el mercado pequeño, seguido del mediano y finalmente en el cuadrante más grande el mercado grande.

3. PREPARACIÓN DE LOS DATOS

Se realizo una preparación de datos diferente, según el Nº Prueba (ver tabla 2). Para el caso de los modelos del Nº Prueba V0, se dumifica el atributo tamamer debido a su naturaleza de carácter, y con el fin de evitar la multicolinealidad, se remueve una de las variables dummy (tamamerPequeño). Posteriormente, para las diferentes versiones planteadas, se tiene en cuenta la matriz de correlaciones (Ilustración 6), tras haber realizado una dummificación de los atributos tamamer y promo.

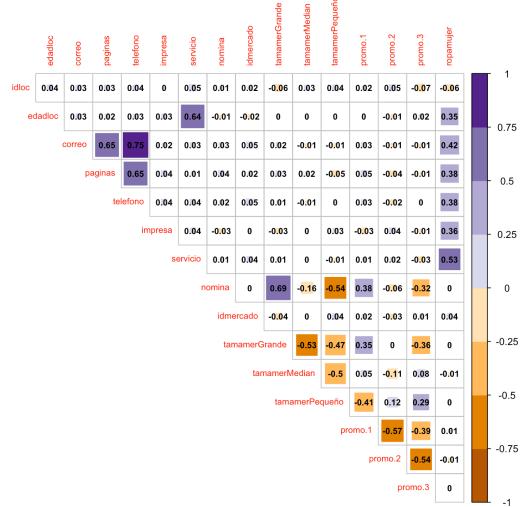


Ilustración 6. Matriz correlación

Para el caso del Nº Prueba V1, se retiran los atributos idloc, nomina, idmercado, promo, debido a su baja correlación con ropamujer, sin embargo, como se mencionó anteriormente, se realiza la dumificación de tamamer y se retira la variable tamamerPequeño. Continuando con Nº Prueba V2, se mantienen las dumificaciones anteriores, de estas se retiran los atributos tamamerPequeño, promo.3, y adicionalmente se crea un nuevo atributo, basado en la multiplicación de los atributos teléfono, paginas, correo, debido a la alta relación que presentan entre ellas y con el fin de no perder la relación de estas con ropamujer, adicionalmente, se retiran los atributos idloc, nomina, idmercado, teléfono, edadloc, paginas, correo, en algunos casos debido a su poca relación con el atributo a predecir, y otros con el fin de evitar alta correlación entre las variables predictoras.

En el caso de NºPrueba V3, se decide descartar todas las variables debido a su poca relación con ropamujer, adicionalmente, se mantiene la multiplicación de los atributos mencionados para la V2, en total son descartadas las variables idloc, nomina, idmercado, teléfono, edadloc, paginas, correo, promo, tamamer. Con el fin de realizar pruebas diferentes en el caso de la Nº Prueba V4, se decide retirar la variable con mayor correlación, es decir servicio, adicionalmente se realiza una estandarización de todos los atributos cuantitativos, con el fin de eliminar los outliers, y se mantiene la dummificación del atributo tamamer y se elimina tamamerPequeño. Finalmente se realiza un pequeño cambio en los atributos de NºPrueba V2, donde para darle aún más peso al atributo servicio, se eleva este al cuadrado, adicionalmente, debido a la alta relación que presentan el atributo nómina y tamamerGrande, se realiza una multiplicación con estos atributos, y se descartan de dataset final para NºPrueba V2^2; igualmente se realiza la aplicación de un logaritmo sobre la variable nomina para NºPrueba V1Log, conservando el resto de atributos como en la V1 y para se aplica un logaritmo a nómina y correo, conservando el resto de atributos igual para el caso V2^2 y a este NºPrueba se le llama V2^2Log.

4. MODELACIÓN

Para la realización del modelo final, se plantearon 5 diferentes versiones, las cuales se combinaron con las diferentes preparaciones de datos mencionadas en la sección anterior. Se planteo una red elástica, utilizando el lamda más pequeño que encontrara, y otra versión de la red elástica según la desviación estándar; por otro lado, se tuvieron en cuenta los modelos de ridge y lasso, al igual que con la red elástica, por cada uno hay una versión con el lamda más pequeño y otro según la desviación estándar, adicionalmente se plantea un modelo usando step. A continuación, en la Tabla 2 se muestran los RMSE obtenidos según el modelo utilizado, resaltando en verde los que obtuvieron un RMSE menor.

Tabla 2. Modelos con sus RMSE

Nº Prueba	Versión del modelo	RMSE	Nº Prueba	Versión del modelo	RMSE
V0	pred_red	11371.51	V4	pred_red	13628.36
	pred_red_se	11640.04		pred_red_se	14054.91
	pred_ridge	11421.14		pred_ridge	13638.79
	pred_ridge_se	11772.34		pred_ridge_se	14012.98
	pred_lasso	11371.02		pred_lasso	13654.49
	pred_lasso_se	11664.79		pred_lasso_se	13985.3
	pred_step	11353.44		pred_step	13619.49
V1	pred_red	11501.26	V2^2	pred_red	11550.04
	pred_red_se	11807.38		pred_red_se	11824.61
	pred_ridge	11541.32		pred_ridge	11594.24
	pred_ridge_se	11900.53		pred_ridge_se	11946.23
	pred_lasso	11497.12		pred_lasso	11548.25
	pred_lasso_se	11828.07		pred_lasso_se	11829.63
	pred_step	11482.34		pred_step	11528.21
V2	pred_red	11584.16	V1Log	pred_red	11501.26
	pred_red_se	11848.38		pred_red_se	11807.38
	pred_ridge	11618.99		pred_ridge	11543.07
	pred_ridge_se	11958.74		pred_ridge_se	11946.38
	pred_lasso	11582.66		pred_lasso	11515.78
	pred_lasso_se	11853.82		pred_lasso_se	11828.07
	pred_step	11575.49		pred_step	11482.34
V3	pred_red	11570.31	V2^2Log	pred_red	11696.12
	pred_red_se	11848.38		pred_red_se	11987.95
	pred_ridge	11606.61		pred_ridge	11735.34
	pred_ridge_se	11930.94		pred_ridge_se	12062.38
	pred_lasso	11570.28		pred_lasso	11694.48

	pred_lasso_se	11853.82		pred_lasso_se	11946.91
	pred_step	11566.83		pred_step	11674.69

Con el fin de dejar públicos todos los scripts de los diferentes modelos, se dispuso la siguiente carpeta online, https://livejaveriana.edu-my.sharepoint.com/:f/g/personal/mora_laura_javeriana_edu_co/EmYwK0TSqwlAg7pX0hV5x8wB0pYmd4VX05ajVV10f1DiA?e=ExdSvC, adicionalmente el modelo que mejor resultado tuvo en el sistema de Kaggle es el adjunto al documento.

5. EVALUACIÓN

Con el objetivo de evaluar los diferentes modelos planteados se subieron algunos de los de los resultados con los RMSE más bajos. Iniciando con el modelo pred_step para el V1, el cual a nivel local obtuvo un RMSE de 11482.34, sin embargo en Kaggle obtuvo un 11419.51190; para el caso del pred_step de V0, a nivel local obtuvo un RMSE de 11353.44 y en Kaggle un 11434.37747; continuando con el pred_step del V2, se obtuvo a nivel local un RMSE de 11575.49 y en Kaggle uno de 11880.63904; con el pred_step de V2^2, se obtuvo un RMSE de 11528.21, mientras en Kaggle se obtuvo de 11891.85236. Finalmente, para pred_red de V4, se obtiene un RMSE de 13628.36 y en Kaggle se obtiene el valor más alto de todas las pruebas, siendo este de 221167049.66664 y para el pred_step de V1Log, se obtiene a nivel local un RMSE de 11674.69, mientras que a nivel de Kaggle se obtiene un 11419.51, como con el caso del V1.

Según los resultados obtenidos con los diferentes modelos, se procedió a observar más a fondo el modelo con el mejor RMSE en kaggle, es decir el pred_step de V1, a continuación en la ilustración 7 se pueden observar las gráficas de dicho modelo.

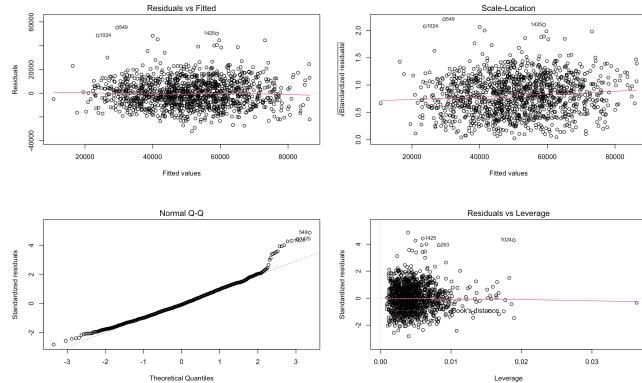


Ilustración 7. Graficas mejor modelo

Adicionalmente, en la ilustración 8, se pueden observar los resultados del modelo mencionado anteriormente, donde se aprecia que el modelo tiene un error residual estándar de 11390, R^2 de 0.5791, es decir que el modelo tiene un 57.91% de la variabilidad de logra captar o recoger del conjunto de variables , por su parte cuenta con un R^2 ajustado de 0.5775, la estadística F nos es de 350 y se cuenta con un p-value: < 2.2e-16; adicionalmente se puede destacar que las variables más influyentes en este son correo, paginas, impresa y servicio.

```

Call:
lm(formula = ropamujer ~ correo + paginas + telefono + impresa +
    servicio, data = ropa.train)

Residuals:
    Min      1Q Median      3Q     Max 
-32054   -7823   -840    7190   55232 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) -4.264e+04 2.476e+03 -17.222 < 2e-16 ***
correo       1.927e+00 2.096e-01  9.193 < 2e-16 ***
paginas      1.659e+02 2.806e+01  5.914 4.29e-09 ***
telefono     8.163e+01 5.292e+01  1.542  0.123  
impresa      9.236e-01 5.100e-02 18.109 < 2e-16 ***
servicio     6.732e+02 2.423e+01 27.784 < 2e-16 *** 
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 11390 on 1274 degrees of freedom
Multiple R-squared:  0.5791,    Adjusted R-squared:  0.5775 
F-statistic: 350.6 on 5 and 1274 DF,  p-value: < 2.2e-16

```

Ilustración 8. Resultados del modelo

Con esta información en mente, se puede generar como recomendación al negocio, prestar suma atención como se está prestando el servicio que se presenta, al igual que la presentación del catálogo mostrando mayor calidad y atractivo para el cliente.

BIBLIOGRAFÍA

ⁱ “Ventas por Catálogo ▷ Qué es, Ejemplos 【 2022 】 .” Guía del Empresario, <https://guiadeempresario.com/ventas/ventas-por-catalogo/>. Accessed 5 Apr. 2022.

ⁱⁱ Ramírez R., Carlos Enrique. “INTERNACIONALIZACIÓN DE LEONISA: UNA EMPRESA COLOMBIANA DE ROPA INTERIOR.” Estudios Gerenciales, vol. 22, no. 98, 2006, pp. 25–55, http://www.scielo.org.co/scielo.php?script=sci_abstract&pid=S0123-59232006000100002&lng=en&nrm=iso&tlng=es.

ⁱⁱⁱ ALZATE FERNÁNDE, FELIPE. CARACTERIZACIÓN DEL CENTRO CORPORATIVO DE LEONISA. 2016, https://repository.eafit.edu.co/bitstream/handle/10784/11625/AlzateFernandez_Felipe_2016.pdf?sequenc e=2#:~:text=La%20compa%C3%B1a%C3%ADa%20ha%20establecido%20negociaciones,Coordinadora%20y%20Enviexpress%2C%20entre%20otros.

^{iv} Leonisa Marcas Aliadas | Leonisa Colombia. <https://www.leonisa.com/col/marcas-aliadas/>. Accessed 5 Apr. 2022.

^v Semana. “El cambio radical que trajo la digitalización a las empresas de venta directa.”

Semana.com Últimas Noticias de Colombia y el Mundo, 29 Oct. 2020, <https://www.semana.com/empresas/articulo/como-estan-operando-las-empresas-de-venta-directa-en-la-pandemia/305000/>.