



UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA INDUSTRIAL

**ESTIMACIÓN DEL POTENCIAL DE COMPRA DE LOS CLIENTES DE UN
PROVEEDOR PERUANO DE PRODUCTOS DE CONSUMO MASIVO**

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERA CIVIL INDUSTRIAL

IGNACIA BELÉN RIQUELME PICART

PROFESOR GUÍA:

ALEJANDRA PUENTE CHANDÍA

MIEMBROS DE LA COMISIÓN:

LUIS ABURTO LAFOURCADE

CLAUDIO PIZARRO TORRES

SANTIAGO DE CHILE

2016

RESUMEN DE LA MEMORIA PARA
OPTAR AL TÍTULO DE INGENIERA
CIVIL INDUSTRIAL
POR: Ignacia Belén Riquelme Picart
FECHA: 07/07/2016
PROFESOR GUÍA: Sra. Alejandra
Puente Chandía

ESTIMACIÓN DEL POTENCIAL DE COMPRA DE LOS CLIENTES DE UN PROVEEDOR PERUANO DE PRODUCTOS DE CONSUMO MASIVO

El presente trabajo se realiza en el contexto de un proveedor peruano de productos de consumo masivo en la industria del *retail* en Perú, que tiene por principal dificultad identificar patrones de comportamiento de sus clientes para determinar la capacidad máxima de compra, lo que se define como el potencial de compra de un cliente.

El principal objetivo de este trabajo es determinar el potencial de compra en clientes de un proveedor peruano de productos de consumo masivo y proponer líneas de acción con el fin de aumentar las ventas y asignar de mejor manera los esfuerzos de marketing y fuerza de venta. Lo anterior, se realiza con una metodología que permita identificar las variables que influyen en el comportamiento de compra, y se desarrolla un modelo que incluye los efectos de las variables transaccionales del cliente, de contexto de la región a la que pertenece y del tipo de vendedor que lo atiende y un modelo de las principales categorías del proveedor. Ambos modelos se comparan con regresiones lineales y regresiones de fronteras estocásticas, calculando el potencial de compra de los clientes.

Los resultados principales de esta memoria, arrojan un aumento potencial de las ventas por casi un 11% en el Modelo (I) de 3 efectos, 12% en el Modelo (II) de regresiones lineales y 77% en el de regresiones de fronteras estocásticas, lo que significa un aumento de 4.8, 3.6 y 20.7 MMUSD respectivamente. Se selecciona el modelo de categorías principales, respaldado con estadísticos significativos, considerando que los modelos tienen distintos enfoques. Posteriormente, se clasifican los clientes, priorizando a los con baja venta real y alto potencial de compra para enfocar las propuestas de acción como aumentar el mejor tipo de vendedores en aquel segmento y creando estrategias tácticas para Lima.

Como conclusión, se proyecta que se pueden aumentar las ventas, diseñando e implementando líneas de acción que consideren las estimaciones de potencial de compra por cada cliente. Finalmente, se propone como trabajos futuros, profundizar en métodos para estimar el potencial y gestión de categorías para optimizar los esfuerzos de fuerza de venta y marketing para la empresa.

*A mis 3 sobrinos,
Francisco, Josefina y a quien viene en camino,
que cada día sea un presente,
de una vida plena, feliz y trascendente.*

*En memoria de un hombre y amigo excepcional,
Oscar Grasset,
quien estará en nuestras memorias por siempre.*

AGRADECIMIENTOS

En primer lugar, quiero agradecer sinceramente a mi papá, mis hermanos, mi tía favorita y mi primo Rafa, por inspirarme en continuar y darme la oportunidad de finalizar esta etapa universitaria, además de complementar mi vida con su maravillosa e incondicional compañía. También, agradezco a mi segunda familia, los Canales y los Martínez Peña y Lillo, por aceptarme como una más en sus hogares.

A mis adorados amigos, Andrés, Cata C., Cata F., Gonzalo, Carlota, Javier, Diego, Moya, Nico, Joaquín, Jetro, Dani H., Dani Y., Marcelo, Juanjo, Pancho, Pauli y Matías. Faltan palabras para expresar mi gratitud por el amor y apoyo que me han brindado desde que los conocí. Gracias por nunca dejar que me rindiera y ayudar a levantarme al caer. Son lo más grande que tengo, junto con Rocky y Bart. Todo esto no tendría sentido sin ustedes y juntos podemos lograr cualquier meta – *we can work it out*.

Gracias a Gabriel y Sergio, por ayudarme y tener siempre la mejor disposición al hacerlo. Espero nunca cambien, ya que siempre estarán guiados por su luz interior hacia el éxito y el bien.

Agradezco a los profesores de la facultad y a mi comisión, Alejandra, Luis y Claudio, por brindarme las herramientas, tiempo y orientación que necesitaba para sacar adelante esta memoria. Destaco especialmente a mi profesora guía por su buena voluntad, paciencia, comprensión y gran aporte en el ámbito docente y profesional.

Finalmente, gracias a todo el equipo de P. A., particularmente, Ariel, Camila y Erick, por colaborar en mi formación profesional y darme la oportunidad de desarrollar este tema de memoria.

TABLA DE CONTENIDO

| | |
|--|----|
| CAPÍTULO 1. INTRODUCCIÓN | 1 |
| 1.1 Antecedentes generales | 1 |
| 1.2 Antecedentes de la empresa | 3 |
| CAPÍTULO 2. DESCRIPCIÓN DEL PROYECTO Y JUSTIFICACIÓN | 5 |
| CAPÍTULO 3. OBJETIVOS | 9 |
| 3.1 Objetivo General | 9 |
| 3.2 Objetivos Específicos..... | 9 |
| CAPÍTULO 4. ALCANCES..... | 10 |
| CAPÍTULO 5. MARCO CONCEPTUAL | 12 |
| 5.1 Metodología <i>Knowledge Discovery in Databases</i> | 12 |
| 5.2 Modelos de Regresión | 12 |
| 5.2.1 Modelo de Regresión Lineal..... | 12 |
| 5.2.1.1 Algoritmo <i>Stepwise</i> | 13 |
| 5.2.1.2 Evaluación del modelo de Regresión Lineal | 14 |
| 5.2.2 Modelo de Regresión de Fronteras | 14 |
| 5.2.2.1 Modelo de frontera de producción estocástica..... | 15 |
| 5.3 Método para estimar <i>Customer Lifetime Value</i> | 17 |
| 5.3.1 Método RFM..... | 17 |
| CAPÍTULO 6. METODOLOGÍA | 18 |
| CAPÍTULO 7. DESARROLLO METODOLÓGICO | 19 |
| 7.1 Análisis y selección de datos | 19 |
| 7.1.1 Recopilación datos | 19 |
| 7.1.2 Pre procesamiento de datos..... | 20 |
| 7.1.3 Análisis descriptivo | 22 |
| 7.1.4 Creación de variables..... | 24 |
| 7.1.5 Selección de variables..... | 27 |
| 7.1.6 Selección muestra de datos | 30 |
| 7.2 Modelo predictivo de potencial de compra del cliente | 31 |
| 7.2.1 Modelo simple: promedio general..... | 31 |

| | | |
|--------------|--|----|
| 7.2.1.1 | Resultados por región | 31 |
| 7.2.1.2 | Resultados por zona | 33 |
| 7.2.1.3 | Resultados categorías principales | 33 |
| 7.2.2 | Modelo de regresión lineal | 34 |
| 7.2.2.1 | Modelo (I): Efectos de las ventas 2014 | 34 |
| 7.2.2.2 | Modelo (II): Categorías principales | 43 |
| 7.2.3 | Modelo de Regresión de Fronteras | 49 |
| 7.2.3.1 | Modelo (I): Efectos de las ventas 2014 | 49 |
| 7.2.3.2 | Modelo (II): Categorías principales | 49 |
| 7.2.3.3 | Cálculo de potencial de compra | 53 |
| 7.2.4 | Resultados complementarios | 54 |
| 7.2.5 | Elección del modelo de potencial de compra | 54 |
| 7.3 | Ranking de clientes | 58 |
| 7.4 | Propuestas de líneas de acción | 63 |
| CAPÍTULO 8. | CONCLUSIONES | 66 |
| CAPÍTULO 9. | TRABAJOS PROPUESTOS | 69 |
| BIBLIOGRAFÍA | | 70 |
| ANEXOS | | 73 |

ÍNDICE DE GRÁFICOS

| | |
|---|----|
| Gráfico 1: Crecimiento del sector <i>retail</i> | 1 |
| Gráfico 2: Número de clientes y ventas según tipo de minorista 2014..... | 10 |
| Gráfico 3: Modelo de regresión de frontera estocástica..... | 16 |
| Gráfico 4: Histograma ventas 2014 de bodegas | 21 |
| Gráfico 5: <i>Box plot con y sin outliers</i> | 21 |
| Gráfico 6: Ventas totales y bodegas 2013 | 23 |
| Gráfico 7: Ejemplo – Bodegas similares | 23 |
| Gráfico 8: Ventas totales 2013 y 2014 por zona | 26 |
| Gráfico 9: Ventas propias vs. socios comerciales según categoría 2013 | 29 |
| Gráfico 10: Modelo (I) – Efecto cliente: comparación pesos de los parámetros betas en la regresión lineal 2 | 36 |
| Gráfico 11: Modelo (I) – Efecto contexto: Comparación pesos de los parámetros betas en la regresión lineal 2 | 37 |
| Gráfico 12: Modelo (I) – Efecto tipo de vendedor: comparación pesos de los parámetros betas en la regresión lineal 2..... | 39 |
| Gráfico 13: Modelo (I) – 3 Efectos: Comparación de parámetros betas de la regresión lineal 2..... | 40 |
| Gráfico 14: Categoría: Aceites – Estadísticos principales de los modelos de regresión de fronteras según número de variables incluidas | 51 |
| Gráfico 15: Categoría aceites – Venta real vs. Venta potencial 2013 | 58 |
| Gráfico 16: Número de clientes por zonas y cuadrantes..... | 61 |
| Gráfico 17: Número de vendedores por tipo y cuadrantes | 62 |
| Gráfico 18: Categoría aceites – Venta real vs. Venta estimada por el modelo de frontera 2013 | 63 |
| Gráfico 19: Frontera de posibilidades de producción | 78 |
| Gráfico 20: Comparación de los modelos de regresión de fronteras | 81 |
| Gráfico 21: Categoría: Detergentes – Estadísticos principales de los modelos de regresión de fronteras según número de variables incluidas | 95 |
| Gráfico 22: Categoría: Galletas – Estadísticos principales de los modelos de regresión de fronteras según número de variables incluidas | 96 |
| Gráfico 23: Categoría: Pastas – Estadísticos principales de los modelos de regresión de fronteras según número de variables incluidas | 97 |

| | |
|---|-----|
| Gráfico 24: Porcentaje de clientes por zona y cuadrantes | 102 |
| Gráfico 25: Porcentaje de vendedores por tipo y cuadrantes | 103 |

ÍNDICE DE TABLAS

| | |
|--|----|
| Tabla 1: Análisis descriptivo de variables transaccionales | 22 |
| Tabla 2: Clasificación del tipo de vendedor según ventas 2013 (S/.) | 24 |
| Tabla 3: Datos según región..... | 25 |
| Tabla 4: Análisis descriptivo de variables RFM | 27 |
| Tabla 5: Matriz de correlación de Pearson de las variables transaccionales principales de los clientes..... | 28 |
| Tabla 6: Matriz de correlación de Pearson de las variables de región..... | 28 |
| Tabla 7: Análisis descriptivo de la venta 2013 (S/.) de las categorías propias | 30 |
| Tabla 8: Resultados preliminares de potencial de compra por región | 32 |
| Tabla 9: Resultados preliminares de potencial de compra por zonas..... | 33 |
| Tabla 10: Resultados preliminares de potencial de compra por categorías | 34 |
| Tabla 11: Modelo (I) – Efecto cliente: Resultados regresión lineal 1 (parámetros betas) | 35 |
| Tabla 12: Modelo (I) – 3 Efectos: Resultados regresión lineal 2..... | 41 |
| Tabla 13: Estadísticos principales de las regresiones lineales de los efectos | 41 |
| Tabla 14: Estadísticos del modelo de regresión lineal de los 3 efectos en conjunto | 42 |
| Tabla 15: Estimación del potencial de compra por zonas basado en el modelo de regresión lineal de los 3 efectos en conjunto | 43 |
| Tabla 16: Parámetros betas estimados por categoría de las regresiones lineales | 46 |
| Tabla 17: Estadísticos de las regresiones lineales de cada categoría..... | 47 |
| Tabla 18: Estimación potencial de compra según categorías..... | 48 |
| Tabla 19: Categoría: Aceites – Variables incluidas en los modelos de regresión de fronteras | 50 |
| Tabla 20: Parámetros betas estimados por categoría del modelo de fronteras estocásticos..... | 52 |
| Tabla 21: Eficiencia técnica de cada categoría del modelo de fronteras estocásticos .. | 53 |
| Tabla 22: Cálculo de potencial de compra según modelos de regresión de fronteras de cada categoría..... | 53 |
| Tabla 23: Comparación de los modelos de regresión lineal | 55 |
| Tabla 24: Comparación de los modelos de regresión de fronteras estocásticos..... | 56 |
| Tabla 25: Resumen de ventas de los cuadrantes de la categoría aceites..... | 59 |

| | |
|---|----|
| Tabla 26: Promedios de las variables transaccionales de los clientes de cada cuadrante | 60 |
| Tabla 27: Categorías a potenciar en orden de prioridad..... | 65 |
| Tabla 28: Modelo (I) – Efecto cliente: Resultados regresión lineal 1 (parámetros betas) | 86 |
| Tabla 29: Modelo (I) – Efecto cliente: Resultados regresión lineal 1 (estadísticos) | 86 |
| Tabla 30: Modelo (I) – Efecto cliente: Resultados regresión lineal 2 (parámetros betas) | 86 |
| Tabla 31: Modelo (I) – Efecto cliente: Resultados regresión lineal 2 (estadísticos) | 87 |
| Tabla 32: Modelo (I) – Efecto contexto: Resultados regresión lineal 1 (parámetros betas) | 88 |
| Tabla 33: Modelo (I) – Efecto contexto: Resultados regresión lineal 1 (estadísticos).... | 89 |
| Tabla 34: Modelo (I) – Efecto contexto: Resultados regresión lineal 2 (parámetros betas) | 89 |
| Tabla 35: Modelo (I) – Efecto contexto: Resultados regresión lineal 2 (estadísticos).... | 89 |
| Tabla 36: Modelo (I) – Efecto tipo de vendedor: Resultados regresión lineal 1 y 2 | 90 |
| Tabla 37: Modelo (I) – Efecto tipo de vendedor: Resultados regresión lineal 1 y 2 (estadísticos) | 90 |
| Tabla 38: Resultados regresión lineal categoría aceites | 91 |
| Tabla 39: Resultados regresión lineal categoría detergentes | 91 |
| Tabla 40: Resultados regresión lineal categoría galletas | 92 |
| Tabla 41: Resultados regresión lineal categoría pastas | 92 |
| Tabla 42: Resultados regresión lineal categoría salsas..... | 93 |
| Tabla 43: Residuos de las regresiones lineales de cada categoría | 93 |
| Tabla 44: Modelo (I): 3 efectos – Resultados regresión de frontera | 94 |
| Tabla 45: Categoría: Detergentes – Variables incluidas en los modelos de regresión de fronteras | 95 |
| Tabla 46: Categoría: Galletas – Variables incluidas en los modelos de regresión de fronteras | 96 |
| Tabla 47: Categoría: Pastas – Variables incluidas en los modelos de regresión de fronteras | 97 |
| Tabla 48: Ineficiencia estocástica en cada Modelo (II) de fronteras | 98 |

| | |
|---|-----|
| Tabla 49: Análisis descriptivo de las principales variables transaccionales, incluyendo la superficie de la bodega | 99 |
| Tabla 50: Resultados modelo de regresión lineal 1 con variables de la encuesta..... | 100 |
| Tabla 51: Resultados modelo de regresión lineal 2 con variables de la encuesta..... | 101 |
| Tabla 52: Porcentaje de clientes por zona y cuadrantes | 102 |
| Tabla 53: Resultados Test Chi-Cuadrado de clientes por zona y cuadrantes | 102 |
| Tabla 54: Porcentaje de vendedores por tipo y cuadrantes..... | 103 |
| Tabla 55: Resultados Test Chi-Cuadrado de clientes por zona y cuadrantes | 103 |

ÍNDICE DE ANEXOS

| | |
|--|-----|
| Anexo 1: Marco conceptual – KDD | 73 |
| Anexo 2: Marco conceptual – Regresión Lineal | 74 |
| Anexo 3: Marco conceptual – Evaluación de modelos | 75 |
| Anexo 4: Marco conceptual – Modelo de Regresión de Fronteras..... | 78 |
| Anexo 5: Marco conceptual – Modelo de frontera de producción determinística | 80 |
| Anexo 6: Marco conceptual – Comparación de modelos de fronteras de producción.... | 81 |
| Anexo 7: Marco conceptual – Customer Lifetime Value | 82 |
| Anexo 8: Mapa político de Perú | 84 |
| Anexo 9: Matriz de correlación de Pearson entre ventas 2013 de cada categoría..... | 85 |
| Anexo 10: Modelo (I) - Regresión lineal efecto cliente | 86 |
| Anexo 11: Modelo (I) - Regresión lineal efecto contexto | 88 |
| Anexo 12: Modelo (I) - Regresión lineal efecto tipo de vendedor | 90 |
| Anexo 13: Resultados Modelo (II): Regresión lineal categorías principales | 91 |
| Anexo 14: Modelo (I) - Resultados regresión de fronteras estocásticas | 94 |
| Anexo 15: Selección de modelos de regresión de frontera para detergentes, galletas y pastas..... | 95 |
| Anexo 16: Ineficiencia estocástica en cada categoría del modelo de fronteras | 98 |
| Anexo 17: Resultados complementarios | 99 |
| Anexo 18: Test Chi-Cuadrado de clientes por zona y cuadrantes..... | 102 |
| Anexo 19: Test Chi-Cuadrado de vendedores por tipo y cuadrantes..... | 103 |

CAPÍTULO 1. INTRODUCCIÓN

1.1 Antecedentes generales

El *retail* es un sector industrial compuesto por empresas que comercializan al por menor o al detalle. Este rubro en Perú, sólo en el 2012 la actividad comercial creció 6,7% en cifras reales (ver **Gráfico 1: Crecimiento del sector *retail***), mientras que en términos acumulados del PIB (Producto Bruto Interno) del sector comercio alcanzó una tasa de crecimiento de 102,9% respecto al 2002, convirtiendo a este país en el más atractivo y dinámico de Latinoamérica en cuanto a potencial de expansión del comercio moderno por encima de otros países como Brasil (90,3%), Chile (66,4%), Colombia (63,4%) y México (44,1%) (CCL y IEDEP Agosto 2013).

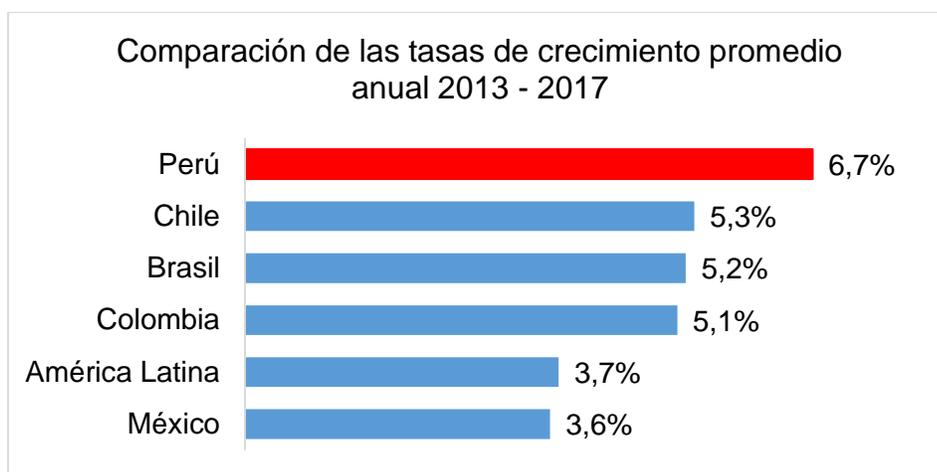


Gráfico 1: Crecimiento del sector *retail*

Fuente: Elaboración propia con datos de Latin Consensus Forecast, y IEDEP-CCL

De acuerdo al Índice de Desarrollo Global del Comercio Retail de A.T. Kearney 2013, algunos de los factores como el crecimiento significativo del PIB per cápita, un ambiente de negocios que va mejorando y el nivel riesgo país, han posibilitado que Perú alcance el puesto 12 de 30 de economías emergentes que son atractivas para el desarrollo de inversiones en el sector del *retail*. Por lo tanto, este buen rendimiento de la economía y actividad comercial en Perú, junto con una excelente oportunidad y enorme potencial es aprovechado por inversionistas locales y extranjeros que están en la búsqueda de nuevos nichos de mercado.

Lo anterior se traduce en mayor competencia en esta industria, por lo que surge la necesidad de buscar estrategias para crear valor al cliente (Garrido 2009), ya que se ha reconocido que el verdadero valor no está en los productos, sino en los clientes

(Naumann 1994). Para ello, se debe considerar que hay una escasez de clientes, provocado por la situación actual, donde los clientes son más exigentes, están condicionados por el desarrollo tecnológico, y, en particular, las empresas enfocan sus esfuerzos en aumentar su participación de mercado, elevando los costos de atraer a nuevos clientes (Webster 2000).

Captar nuevos clientes corresponde a una actividad de vital importancia para el crecimiento de las empresas en *retail*, pero, hoy en día, también se deberían concentrar en retener a los clientes actuales, creando relaciones rentables y sostenibles en el tiempo, ya que, en términos económicos, resulta más conveniente priorizar conservar a los clientes versus atraer a nuevos (Rasmussen 1999). Esto exige un cambio en la gestión de la relación con el cliente, pasando de un marketing transaccional (lanzamiento de los productos al mercado esperando que este sea el que capte a los clientes), a un marketing más relacional (Cava 2004). Aquello, espera maximizar las oportunidades de ventas cruzadas (*cross-selling*) y de ventas de productos de mayor valor agregado (*up-selling*), intensificando la satisfacción y lealtad del cliente. Además, los costos operativos asociados al servicio del cliente se reducen a lo largo del tiempo, debido a que se adquiere aprendizaje de los procesos de la empresa (Reichled y Sasser Jr. 1990).

Dado lo anterior, surge el concepto de *Customer Relationship Management* (CRM), el cual permite a las empresas identificar, caracterizar, atraer y retener a sus clientes, incrementando su satisfacción y, de este modo, optimizar la rentabilidad de la empresa (Kotler y Armstrong s.f.).

Dentro de los beneficios que logra una estrategia CRM, se encuentran (Renart Cavas 2004):

- Mayor conocimiento del cliente.
- Aumento de la satisfacción y lealtad del cliente.
- Aumento de ventas.
- Reducción de costos.

Por otro lado, la empresa tiene que estar dispuesta a tratar a sus clientes de manera distinta entre ellos, es decir, realizar discriminación positiva con el objetivo de crear planes de acción diferenciados. A través de CRM analítico es posible analizar datos de información transaccional relacionados con el cliente que no se están explotando al máximo en las empresas (Buttle 2009).

Actualmente, el mercado admite llevar registro de las transacciones de las compras que se realizan, quedando ligadas a un identificador único, lo cual permite conocer a cada comprador de forma personalizada, y mediante un análisis transaccional y estudio analítico se puede mejorar la gestión de clientes. Este proceso de recolección de datos y análisis de la información es conocido como Inteligencia de Clientes, y permite diseñar campañas de marketing más efectivas y focalizadas (Kotler y Armstrong s.f.). Se pueden reconocer grupos de clientes más homogéneos y parecidos en su comportamiento de compra, además de identificar necesidades específicas para cada tipo de cliente.

La importancia de lo anterior, es que los *retailers* deben ser cuidadosos con la fuga de clientes que representan mayor rentabilidad en el largo plazo, por ende, se recomienda seleccionar a los clientes más valiosos y potenciar una buena estrategia hacia el futuro que maximice su *Customer Lifetime Value* (CLV), la cual es una herramienta para poder identificar a los clientes más rentables en el tiempo (Forbes 2007).

Por lo tanto, las preguntas que surgen en las empresas son ¿Quiénes son y serán los clientes más valiosos?, ¿Se está aprovechando toda la capacidad de compra que tienen los clientes? Si no es así, ¿Cuánto más podrían comprar estos clientes? ¿A qué segmento de clientes se deben enfocar los esfuerzos? Actualmente, ¿Es eficiente la manera de asignar los recursos, tanto de marketing como fuerza de venta, para atender a los clientes?

1.2 Antecedentes de la empresa

La empresa donde se basa el caso de estudio es un proveedor peruano, la cual es una Sociedad Anónima Abierta que se dedica a la exportación, importación, distribución y comercialización de productos de consumo masivo, principalmente alimenticios y de limpieza, en sus más variadas formas y con diferentes marcas. Actualmente, opera a nivel nacional y en 16 países del mundo, principalmente en Latinoamérica.

En Perú, la empresa está posicionada dentro de los primeros proveedores de bienes de consumo masivo con ventas netas anuales equivalentes a más de 1,8 millones de dólares americanos (USD) y una utilidad bruta de USD 500 mil con un aumento de 8,3% con respecto al año 2013.

La empresa cuenta con dos tipos de negocios, productos de consumo masivo (principalmente detergentes, pastas, salsas, galletas y aceites domésticos) y productos industriales (margarinas y mantequillas, harinas y premezclas, entre otros), los cuales

representan un 66% y 34% de la venta total, respectivamente. Los productos son distribuidos a través del canal moderno (hipermercados, supermercados y grandes almacenes) y el canal tradicional (mayoristas y minoristas).

La visión de la compañía es triplicar el valor de su compañía para el 2021, con lo cual ha formado una gran red de distribución, una posición líder en el mercado, alto poder de negociación con los clientes, junto con un crecimiento sostenido.

Por lo tanto, la empresa peruana presentada anteriormente, cae dentro del rubro del *retail* y, como se mencionó previamente, el constante aumento del nivel de competitividad y complejidad presente en el área del *retail*, hace que se genere la necesidad de evaluar diferentes opciones para mejorar el margen de ganancias (Freemium Economics: Leveraging Analytics and User Segmentation to Drive Revenue s.f.). Existen opciones como estimación del valor del cliente, potencial de compra, tasa de fuga de clientes, segmentaciones, gestión de categorías, productos, trabajadores, entre otros, acordes a las necesidades de la empresa (Seufert 2014).

CAPÍTULO 2. DESCRIPCIÓN DEL PROYECTO Y JUSTIFICACIÓN

En el contexto descrito inicialmente, para tener un liderazgo de mercado, se hace fundamental conocer a los clientes en profundidad con las características propias de cada uno, tales como sus patrones de compra, su dinamismo en el tiempo, particularidades y tipos de clientes, para así generar relaciones de largo plazo basadas en la lealtad de los clientes (Zapata 2006).

En adelante, se consideran clientes a aquellos que compran directamente al proveedor peruano, es decir, se refiere a tiendas que venden al por menor y no al usuario final que consume y/o utiliza los productos comprados.

Entonces, se hace esencial para aumentar las ventas, buscar nuevas herramientas para implementar e innovar en la empresa que generen una ventaja competitiva en el mercado. Por lo tanto, se proponen modelos más sofisticados que permitan tener una mejor estimación del comportamiento del cliente y de su rol en la empresa y, de este modo, optimizar los recursos disponibles en función del tiempo.

El valor y capacidad de compra que pueda tener un cliente es una métrica esencial en marketing, dado que el cálculo de ello, proporciona un umbral máximo para el gasto de adquisición en el desarrollo de campañas de *Return On Investment* (ROI) positivo. También, provee una medida para diseñar nuevos proyectos, optimizando la asignación de recursos, con el fin de tomar decisiones estratégicas más acertadas (Seufert 2014). Lo anterior, debe incluir las variables significativas para la empresa o definiciones de su negocio, basado en el seguimiento de toda la relación entre el cliente y la entidad en base a factores como su nivel de facturación, número de productos consumidos, la probabilidad de abandono que se le puede asignar, el costo de retención estimado, entre otros, tanto como para su comportamiento presente como para el potencial recorrido que podría alcanzar.

Actualmente, en la empresa se tiene una gran base de datos con información transaccional de los clientes, pero no se conocen sus patrones de comportamiento, por lo tanto, no se tiene una estimación de la capacidad de compra ni el valor del cliente para la empresa.

El valor de un cliente se puede medir a partir del volumen de compra, sin embargo, lo que no es evidente es cuánto no le está comprando al proveedor peruano. Entonces, se define el **potencial de compra** de un cliente como la capacidad máxima de compra que tiene,

es decir, el total que le es posible comprar de acuerdo a sus características propias. Además, la identificación del potencial de compra del cliente y lo que compra actualmente constituye el **diferencial de potencial**, que se calcula a partir de la diferencia entre aquellos montos, y mide el nivel de eficiencia del cliente.

Una forma simple de estimar el potencial de compra sería imponerle al cliente que en cada transacción futura que realice, ésta sea igual o mayor a su monto máximo histórico transaccional. Entonces, el potencial de compra anual de un cliente, sería el número de transacciones del año multiplicado por el monto máximo transaccional del año correspondiente. Lo anterior, sobrestima la capacidad de compra total que tiene el cliente, dado que su monto máximo histórico puede diferir a su promedio general, lo cual se puede deber a una ocasión de compra de características únicas o puede estar condicionado por decisiones o errores del cliente como sobre *stock* o *stock piling* intencional u orden de compra mal ejecutada, fallas en la distribución de los productos por parte del proveedor, mala estimación de la demanda, entre otras razones.

Entonces, para calcular el potencial de compra, se pueden agregar más variables y analizar las relaciones entre ellas como, por ejemplo, a través de las características que definen al cliente y, según los datos transaccionales se puede pronosticar cuánto podría llegar a comprar en un periodo de tiempo, e incluso identificar qué tipo de cliente será uno nuevo, comparándolo con los patrones de compra y atributos de los clientes antiguos.

Complementando lo anterior, es menos costoso invertir en un cliente actual que en adquirir un cliente nuevo (Bhattacharya 1998). Entonces, es importante identificar a los clientes más “descuidados” o con un gran diferencial de potencial, es decir, los que están comprando menos de su venta potencial o capacidad, y que podrían necesitar una atención especial para aprovechar su capacidad total de compra, ya que son los que en la actualidad le generan ingresos a la empresa.

Por lo tanto, el problema corresponde a la falta de acciones comerciales por parte de la empresa para capturar la parte del mercado potencial que a la fecha no está siendo aprovechada.

Dado que el proveedor está presente en regiones de Perú muy diferentes entre sí, se deben considerar factores exógenos como desigualdad económica en la que está inserto el cliente y usuario final, lo cual hace que varíe su poder de adquisición y presupuesto. Complementando lo anterior, las ventas pueden estar condicionadas de acuerdo a las particularidades del tipo de vendedor, junto con las características propias del cliente, tanto transaccionales como preferencias de compra.

Entonces, como no se está aprovechando toda la capacidad de compra de los clientes se proponen las siguientes hipótesis:

- La ubicación del cliente determina el potencial de compra, tanto por características propias del sector y su nivel de competencia.
- La distribución de los vendedores geográficamente de acuerdo a su tipo, afecta directamente el potencial de compra de los clientes.
- La distribución de los vendedores dado la cantidad de clientes que atienden, afecta directamente el potencial de compra de los clientes.
- Los clientes con baja frecuencia de compra tienen mayor potencial, en comparación a los que tienen mayor frecuencia.
- Los clientes que compran la categoría más vendida tienen menor potencial de compra.

Con ello, se pueden formular propuestas de acciones de marketing focalizadas para clientes con un alto diferencial de potencial, con énfasis en el uso eficiente tanto de los recursos humanos como de los datos, aumentando la calidad de servicio y asignando de mejor manera la fuerza de venta, de forma de lograr coherencia entre las acciones comerciales y los objetivos estratégicos del proveedor peruano. Así, se tomarán acciones comerciales más acertadas con el foco de aumentar las ventas en sus categorías junto con un mejor servicio de atención para los clientes con mayor potencial.

El problema a resolver es complejo, pues no es trivial estimar el potencial de compra de los clientes por las siguientes razones:

- Información incompleta de mercado:
La información disponible está sesgada a la base de datos propia de la empresa, es decir, sólo se tienen los datos relacionados a las transacciones del proveedor peruano de productos de consumo masivo y no se conoce lo que compran los clientes a la competencia, dado que hay más marcas en el mercado dentro de las categorías de productos que maneja la empresa.
- Efectos que influyen en el consumo de los clientes:
La empresa no conoce el comportamiento de compra del cliente ni los factores que condicionan una transacción, tanto modificables a nivel táctico como la distribución de la fuerza de venta o a nivel operacional, como la atención de los vendedores, o, incluso, de efectos exógenos como la ubicación y tamaño del negocio del cliente, entre otros.

- **Grandes volúmenes de información:**
La gran cantidad de información existente, a nivel transaccional del proveedor peruano, requiere ser trabajada con modelos de consultas a la base de datos y herramientas más sofisticadas que Microsoft Excel. Además, se debe tener un entendimiento previo de los datos para relacionar las distintas bases de datos disponibles.

CAPÍTULO 3. OBJETIVOS

3.1 Objetivo General

Determinar el potencial de compra en clientes de una empresa proveedora peruana y proponer líneas de acción con el fin de aumentar las ventas.

3.2 Objetivos Específicos

Se presentan a continuación los objetivos específicos que apoyan el desarrollo y logro del objetivo general descrito previamente:

- Identificar las variables que influyen en el comportamiento de compra.
- Desarrollar modelos de predicción de estimación de potencial.
- Evaluar y seleccionar un modelo de estimación de potencial.
- Rankear clientes en función del potencial de compra calculado.
- Diseñar líneas de acción de marketing para los clientes con mayor potencial y valor para la empresa.

CAPÍTULO 4. ALCANCES

El principal alcance esperable es que la metodología propuesta pueda ser usada para todo tipo de clientes minoristas de la empresa de consumo masivo, pero para el caso aplicado sólo se trabajará con los siguientes datos disponibles (selección de la muestra):

- **Años:** 2013 y 2014.
- **Nivel de agregación de los productos de consumo masivo:** categorías a nivel individual.
- **Canal:** tradicional.
- **Clientes:** minoristas activos.
- **Tipo de cliente minorista:** sólo tipo bodegas peruanas (equivalente almacén de barrio en Chile en surtido y categorías), ya que son el tipo de cliente minorista con un nivel de ventas acumulado mayor que el resto de las clasificaciones (ver **Gráfico 2**). Entonces, los clientes se limitan a ser sólo bodegas para este trabajo.

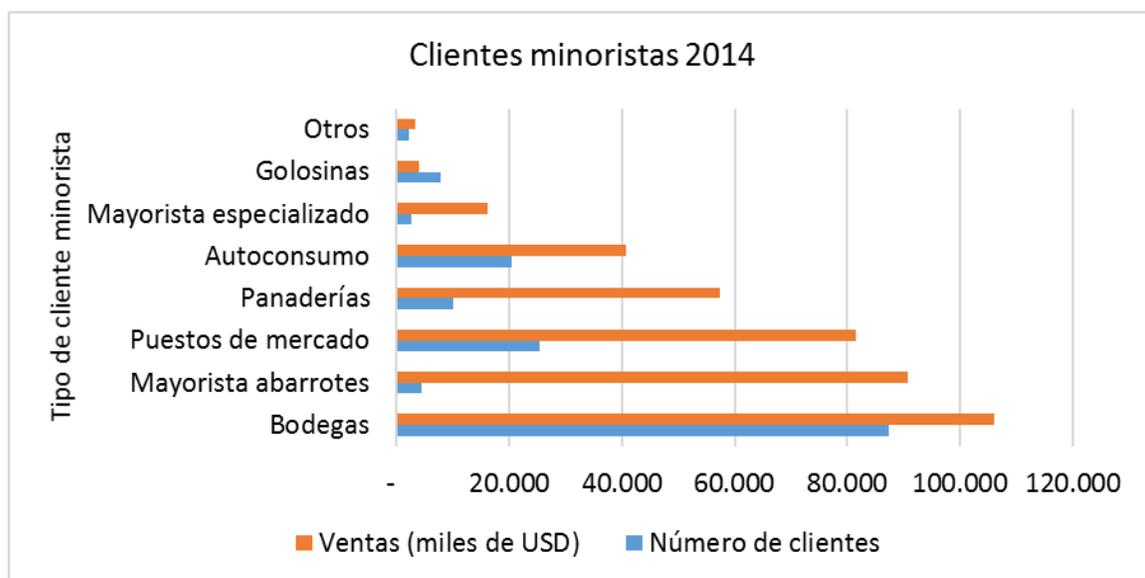


Gráfico 2: Número de clientes y ventas según tipo de minorista 2014
Fuente: Elaboración propia en base a información de la empresa

- **Nivel de venta:** se considerarán los clientes que tengan en promedio una venta mensual mayor a S/.100, alrededor de USD 30 (dado que son clientes que no son visitados por un vendedor, sino que son atendidos de manera telefónica) y con una venta mensual menor a S/.6.000, equivalentes a USD 1.780 (pues son clientes que en la empresa se consideran como mayoristas, por su nivel de compra, a pesar que estén en la base de datos de los clientes minoristas, pues la venta es al detalle de todas maneras).

Además, sólo se diseñarán las propuestas de acciones de marketing, ya que el proyecto no abarca implementación de la metodología debido al tiempo que implicaría.

Es importante destacar que hay requerimientos mínimos esperados por la empresa como variables que las consideran relevantes en su negocio, lo cual no puede ser alterado en la selección de variables.

CAPÍTULO 5. MARCO CONCEPTUAL

Los modelos y conceptos más relevantes para la estimación del potencial de compra obtenidos de la revisión bibliográfica son:

5.1 Metodología *Knowledge Discovery in Databases*

La metodología KDD es un proceso iterativo e interactivo con el usuario y entre etapas, las cuales se resumen a continuación (Fayyad, Piatetsky-Shapiro y Smyth 1996) (ver en Anexo 1 más detalles):

- I. **Selección de datos:** determinación de objetivos del proyecto y selección de datos.
- II. **Pre procesamiento:** limpieza de datos (detección y manejo de *outliers* e información faltante) y con formato común.
- III. **Transformación:** base de datos estructurada según el modelo lógico de información (reducción de datos relevantes al proyecto).
- IV. **Minería de datos:** aplicación de algoritmos para encontrar patrones en los datos.
- V. **Interpretación y evaluación:** validación de los patrones encontrados por el experto en el negocio y usuario.
- VI. **Generación de conocimiento:** documentación del conocimiento adquirido e incorporado al proceso KDD, acorde con los objetivos del proyecto.

5.2 Modelos de Regresión

Dependiendo de la incidencia que tiene cada variable para explicar el comportamiento de los datos disponibles en una base de datos, existen varios modelos econométricos y, a continuación, se mencionan dos de los métodos aplicables para la metodología descrita en el siguiente capítulo, regresiones lineales y regresiones de fronteras.

5.2.1 Modelo de Regresión Lineal

Método matemático que representa la relación entre una variable dependiente y variables independientes que explican su comportamiento. Esta relación queda representada por la siguiente ecuación lineal [8]:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \dots + \beta_N X_{Ni} + \varepsilon_i$$

Los X_{1i} corresponden al valor que toma la variable explicativa (independiente) i -ésima, y los β_i son los parámetros que representan la incidencia que cada variable i tiene para explicar el valor de la dependiente Y_i . La sumatoria del producto entre la variable independiente con su respectivo parámetro, más un intercepto de estimación β_0 (constante), logra predecir la variable de interés Y_i . El término de error queda representado por ε_i (ver propiedades del modelo y más detalles en Anexo 2).

Los supuestos que asume un modelo de regresión lineal son los siguientes (Chmabers 1992):

- La variable dependiente es función lineal de las independientes más un término de error.
- Las variables independientes son independientes del término de error.
- El error se distribuye normal, tiene media nula y la misma varianza para todas las observaciones (homocedasticidad), es decir, $\varepsilon_i \sim N(0, \sigma^2)$.
- El error es independiente entre observaciones (datos disponibles).

5.2.1.1 Algoritmo *Stepwise*

El algoritmo *Stepwise* se utiliza para incluir y eliminar variables, de manera paralela, en los modelos de regresión lineal para poder encontrar variables que son significativas para la estimación de la variable dependiente (Venables y Ripley 2002).

Stepwise combina los siguientes procedimientos (Hastie y Pregibon 1992):

- *Backward Stepwise Regression*: procedimiento que parte del modelo de regresión con todas las variables explicativas y en cada etapa se elimina la variable menos influyente según el contraste individual de t y $p - valor$.
- *Forward Stepwise Regression*: algoritmo que funciona de forma inversa que el anterior, inicia con el modelo sin ninguna variable explicativa (solo una constante) y en cada etapa se introduce la más significativa.

Stepwise comienza con el algoritmo *Backward Stepwise Regression*, pero en cada etapa se plantea si todas las variables introducidas deben de permanecer de acuerdo al procedimiento *Forward Stepwise Regression*. Finalizan las iteraciones cuando ninguna variable entra o sale del modelo.

5.2.1.2 Evaluación del modelo de Regresión Lineal

Para evaluar un modelo, se debe medir cualitativamente, la capacidad de interpretación respecto al comportamiento de clientes, y cuantitativamente, el ajuste de la data de calibración y que pueda predecir más allá de los datos de calibración.

Una posibilidad para contrastar diferentes modelos y determinar si son suficientemente buenos, es comparar las siguientes métricas de ajuste resumidas a continuación y para mayor detalle ver Anexo 3.

- **Coeficiente de Determinación (R^2):** mide la proporción de variación total de las observaciones con respecto a su media que puede ser atribuida a la recta de regresión estimada, es decir, cuánto de la variabilidad de la variable dependiente es explicada por el modelo.
- **Mean Absolute Percentage Error (MAPE):** indicador del porcentaje de error de la media absoluta que mide la precisión de la predicción de un método de pronóstico, es decir, cuánto se equivoca porcentualmente el modelo.
- **Criterio de información de Akaike (AIC):** es una medida de la calidad relativa de un modelo estadístico para un conjunto dado de dato. AIC penaliza el exceso de parámetros ajustados, estimando la pérdida de información. Puede demostrarse que, al aumentar el tamaño muestral, el AIC es eficiente, escogiendo el modelo que proporciona (en promedio) mejores predicciones. Sin embargo, AIC tiende a seleccionar modelos con más parámetros de los necesarios (Peña 2002).
- **Criterio de información Bayesiano (BIC):** criterio de información alternativo a partir de un enfoque bayesiano, como una aproximación a una transformación de la probabilidad posterior de un modelo candidato, que penaliza el número de parámetros con $\ln(n)$ (Sakamoto, Ishiguro y Kitagawa 1986).

5.2.2 Modelo de Regresión de Fronteras

Los modelos paramétricos que se estudiarán serán de ecuación única, de los cuales se estiman los parámetros que describen la estructura de la frontera y las estimaciones de la eficiencia técnica de cada productor se obtienen como subproducto de lo anterior (ver Anexo 4). El término asociado a la eficiencia técnica puede ser determinado a través de dos puntos de vista: del determinístico (ver Anexo 5), que sólo consideran el término

como una componente de ineficiencia unilateral; y el estocástico, que agrega el ruido aleatorio tradicional simétrico (ver en Anexo 6 la comparación de los modelos) (González 2015).

5.2.2.1 Modelo de frontera de producción estocástica

Para atribuir variaciones en la frontera producción no asociadas con variaciones en las entradas, se combinan efectos aleatorios e ineficiencia que permitan desarrollar un modelo más realista. Para ello, se requiere especificar una frontera de producción estocástica, la cual puede ser escrita como (D. Aigner 1977) (Kumbhakar S.C 2000):

$$y_i = f(\vec{x}_i, \vec{\beta}) * \exp\{v_i\} * TE_i$$

Donde, $f(\vec{x}_i, \vec{\beta}) * \exp\{v_i\}$ es la frontera de producción estocástica, la cual está formada por una parte determinística, $f(\vec{x}_i, \vec{\beta})$, común a todos los productores y una parte específica de cada productor, $\exp\{v_i\}$, que recoge el efecto aleatorio. En este caso, la eficiencia técnica se calcula como:

$$TE_i = \frac{y_i}{f(\vec{x}_i, \vec{\beta}) * \exp\{v_i\}}$$

De esta forma, la eficiencia técnica puede ser estimada utilizando el modelo de frontera de producción estocástica, el cual es preferido por incluir el efecto aleatorio en la producción, también conocido como el modelo de error compuesto.

Considerando que $TE_i = \exp\{-u\}$, suponiendo $f(\vec{x}_i, \vec{\beta}) = \prod_n x_{ni}^{\beta_n}$, entonces la ecuación queda:

$$y_i = \prod_n x_{ni}^{\beta_n} * \exp\{v_i\} * \exp\{-u_i\}$$

Finalmente, tomando logaritmos en la ecuación anterior, resulta el modelo de frontera de producción estocástica, y se escribe similar a una regresión lineal con una perturbación negativa de la siguiente manera, con el fin de estimar los betas de la ecuación:

$$\ln y_i = \beta_0 + \sum_n \beta_n \ln x_{ni} + v_i - u_i$$

Donde, $\exp\{v_i\}$ es la componente ruido y $\exp\{-u_i\}$ mide la ineficiencia técnica y u_i es la componente no negativa. Juntas, es decir, $\epsilon_i = v_i - u_i$, forman el término de error. La componente $v_i \sim N(0, \sigma_v^2)$ que es simétrica, idénticamente distribuida e independiente de u_i , el cual tiene una distribución norma no negativa $u_i \sim N^+(0, \sigma_u^2)$. Como $u_i \geq 0$, el término de error, ϵ_i , es asimétrico respecto al cero.

Por lo tanto, el objetivo es obtener estimaciones de los parámetros de producción $\vec{\beta}$ en $f(\vec{x}_i, \vec{\beta})$ como de la eficiencia técnica de cada productor. Para conseguir la segunda parte se requiere de estimaciones separadas del ruido estadístico, v_i , y de la eficiencia técnica, u_i , sean extraídas de la estimación de ϵ_i para cada productor, para lo que hay que asumir distribuciones en las dos componentes de error.

En el **Gráfico 3**, se observan los niveles de producción de una firma al modelarlos con la frontera de producción estocástica, el output puede ir por sobre la frontera de producción, como es el caso de x_i , en que el error estocástico o ruido es positivo. Caso contrario corresponde al de x_j , ya que el ruido es negativo, por lo que el output del modelo se posiciona bajo la frontera de producción.

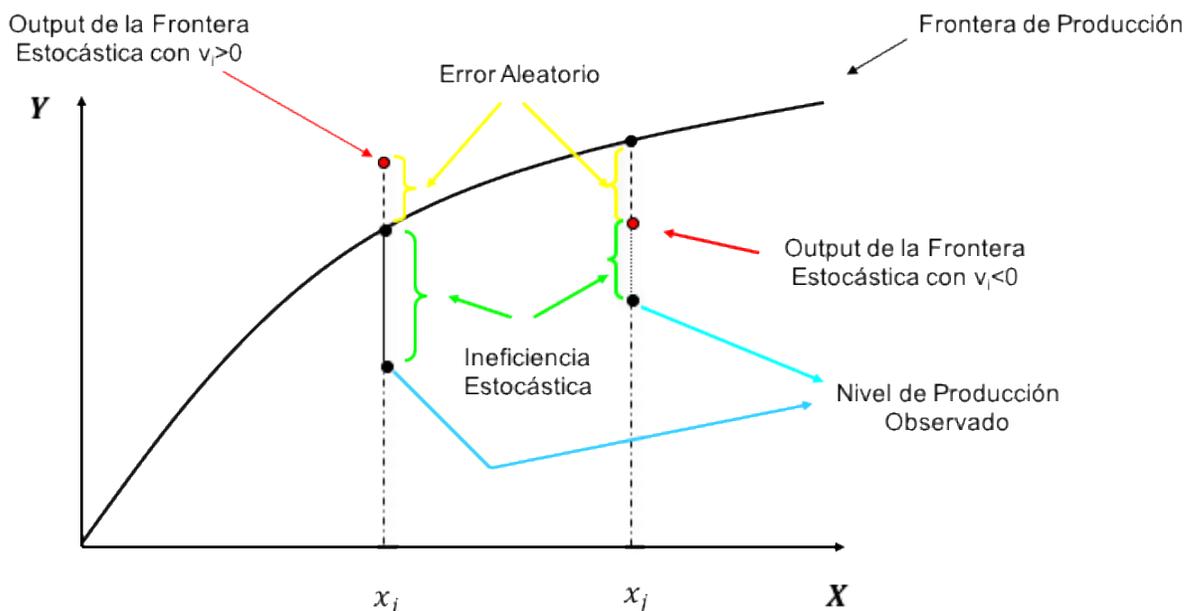


Gráfico 3: Modelo de regresión de frontera estocástica
Fuente: Elaboración propia en base a (University of New England 2014)

5.3 Método para estimar *Customer Lifetime Value*

Existen distintos enfoques para estimar CLV (ver en Anexo 7), a continuación, se mencionará uno de los métodos más simples.

5.3.1 Método RFM

Se utiliza para tasas de respuesta a estrategias de marketing cuantitativo, creando grupos de clientes en función de 3 variables (S. Gupta 2006):

- I. *Recency* (R): tiempo desde la última compra hasta momento de evaluación.
- II. *Frequency* (F): esta variable se puede estimar de 2 maneras:
 - a. Cantidad de transacciones en un determinado período.
 - b. Tiempo entre transacciones
- III. *Monetary Value* (M): monto promedio de las transacciones.

A partir de cada grupo de cliente, se puede analizar la tasa de respuesta dentro de la clase, y definir el CLV para cada cliente según donde pertenece.

CAPÍTULO 6. METODOLOGÍA

El diseño metodológico de este proyecto, contempla una muestra seleccionada de la base de datos de la empresa, de la cual se busca extraer todos los conceptos e información relevantes que sirvan para desarrollar el tema de estimar el potencial de compra de los clientes del proveedor peruano de productos de consumo masivo. Lo anterior se desarrolla en paralelo de la definición de los alcances del estudio, en cuanto a tipos de clientes y períodos de análisis con el fin de acotarlo y utilizar una muestra relevante que permita llegar a resultados concluyentes.

En particular, la aplicación de esta metodología, permitirá entender mejor al cliente a partir de información transaccional, identificando sus características, efectos que influyen en los patrones de compra y cuantificando el potencial de compra de los clientes de la empresa, con el fin de tomar decisiones comerciales más acertadas.

A continuación, en el **Diagrama 1** se resumen los principales pasos de la metodología que se desarrollan en el siguiente capítulo 7 del proyecto:

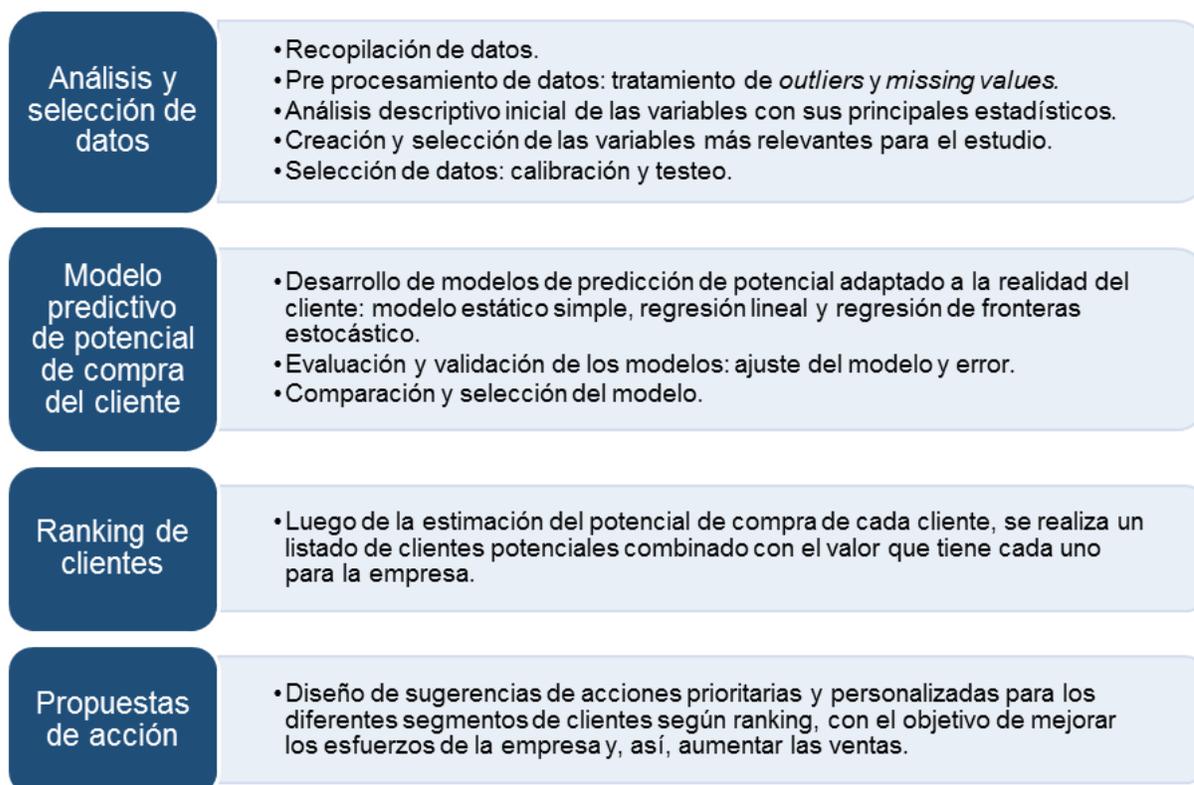


Diagrama 1: Resumen metodología

Fuente: Elaboración propia

CAPÍTULO 7. DESARROLLO METODOLÓGICO

Se presentan en esta sección el detalle del trabajo realizado clasificado según los principales hitos logrados.

7.1 Análisis y selección de datos

7.1.1 Recopilación datos

Se comenzó con la recopilación de toda la información disponible en la base de datos, la cual tiene sus primeros registros en el año 2011, basado en el proceso KDD. Es necesario destacar que sólo se tomarán en cuenta los datos para el caso aplicado, donde el tipo de cliente son las bodegas en los años 2013 y 2014, definidos en el capítulo 4: Alcances.

El detalle de cada transacción fue sacado de la base de datos de la empresa con consultas utilizando SQL (*Structured Query Language*). Generando el siguiente listado de variables por cliente:

- Tipo de clientes: Bodegas
- Nombre del cliente
- Antigüedad: 2 años activos.
- Número de meses con compra.
- Ubicación: dirección, distrito, ciudad, región y zona del país.
- Nombre y código de la sede de la empresa
- Número de clientes por región (competencia).
- Periodo: anual
- Monto preventa: orden de compra
- Monto facturado en el año
- Monto bruto (volumen de compra en kg) en el año
- Número de transacciones en el año
- Días entre compras
- Ticket promedio: ventas/transacciones
- Número de productos distintos comprados
- Número de marcas distintos comprados
- Número de categorías distintos comprados
- Venta de cada categoría en el año
- Transacciones que tiene cada categoría en el año

Desde de una encuesta realizada por la empresa a 1.512 bodegas de Lima, se agregaron los m^2 de la tienda de los clientes.

7.1.2 Pre procesamiento de datos

Los datos de las bodegas se limpiaron antes de utilizarse con el fin de realizar un estudio más confiable y con datos que no produzcan ruido en los resultados de los modelos de estimación, pues existen registros con valores nulos (*missing values*) o extremos (*outliers*).

El criterio para tratar *missing values* fue no utilizar las variables incompletas y, dado que el caso son 0,01% del total de los datos, todas estas observaciones fueron omitidas para los cálculos de estadísticos descriptivos básicos y modelos desarrollados, exceptuando la información disponible sobre los m^2 de la tienda del cliente, ya que, intuitivamente, se considera un factor importante y el volumen de registros que contiene esta variable es muy reducido para permitir su eliminación.

Para identificar *outliers* (Han, Kamber y J.Peí 2006) en las variables, se utilizaron gráficos *box plots*, ya que es una manera simple de tener una noción de la dispersión de los datos, ya que los valores extremos, es decir, aquellos elementos de los datos que difieren en gran medida del resto, puede ser producto de un comportamiento de un cliente muy distinto al normal, o debido a que se registraron de manera incorrecta los datos.

En el caso de las ventas del año 2014, el **Gráfico 4** muestra la distribución de las ventas y el **Gráfico 5** detecta que hay una gran cantidad de clientes (bodegas) que se comportan como *outliers*.

Por lo tanto, para tener una distinción más robusta de estos valores extremos, se analizaron, los *outliers* con el criterio *maximum residual test*, detectando distribución normal con el test *Grubb*, y se define z como: $z = \frac{|x - \bar{x}|}{\sigma}$, donde \bar{x} es el promedio, σ es la desviación estándar y un valor x es *outlier* si:

$$z \geq \frac{N - 1}{\sqrt{N}} \sqrt{\frac{t_{\frac{\alpha}{2N}, N-2}^2}{N - 2 + t_{\frac{\alpha}{2N}, N-2}^2}}$$

Donde $t_{\frac{\alpha}{2N}, N-2}^2$ es el valor que toma una t -distribución a un nivel de significancia de $\frac{\alpha}{2N}$, y N es el número de datos en la variable elegida.

Así, al eliminar aquellos datos, la distribución de la variable de las ventas del año 2014 es representada por el **Gráfico 5**.

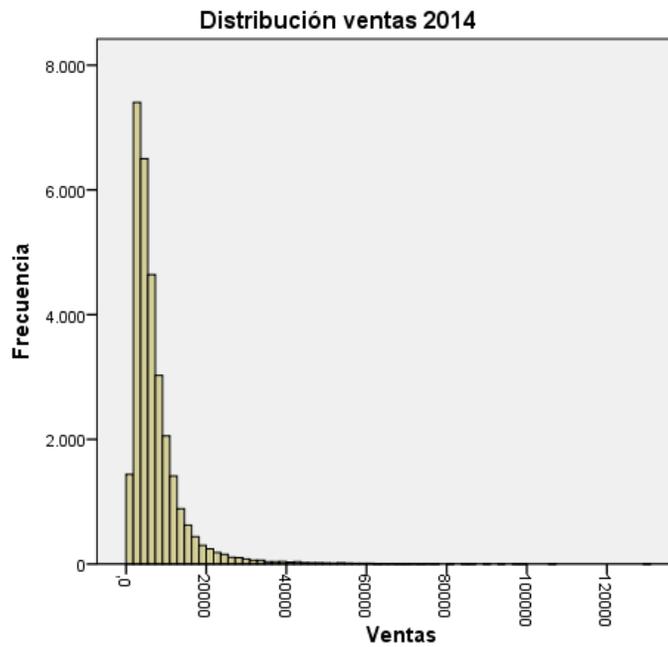


Gráfico 4: Histograma ventas 2014 de bodegas
Fuente: Elaboración propia en base a información de la empresa

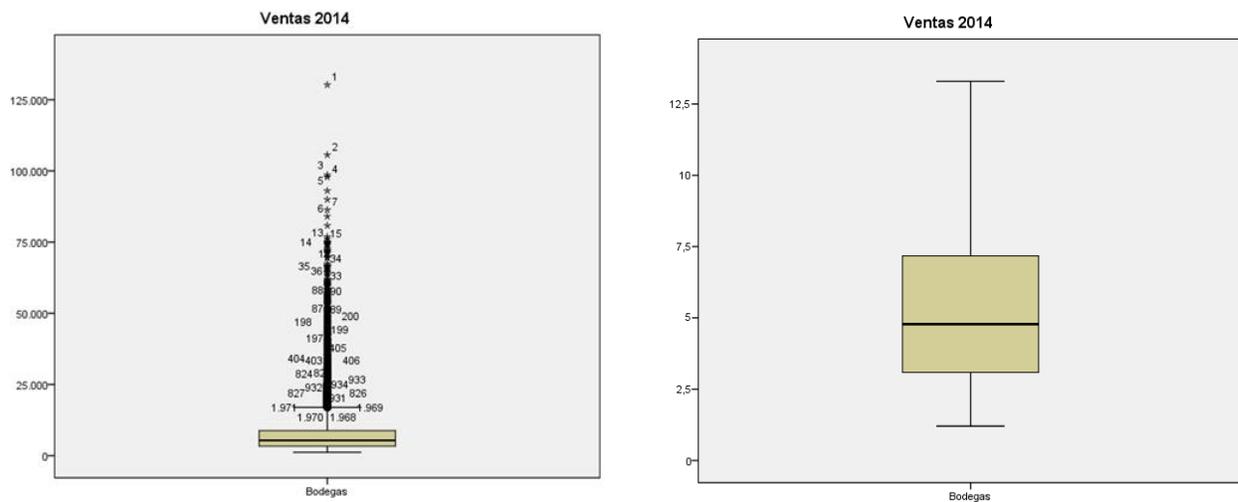


Gráfico 5: Box plot con y sin outliers
Fuente: Elaboración propia en base a información de la empresa

Entonces, luego del tratamiento de *outliers* y *missing values*, queda un total de 26.156 clientes, correspondientes a un 84,43% del total inicial (30.980 clientes activos filtrados).

De este modo, se consolidaron los datos con todas las variables necesarias en una planilla .txt.

7.1.3 Análisis descriptivo

Luego, se realiza un estudio descriptivo inicial de la situación de las variables para entender cómo se comportan, con el fin de tener un conocimiento preliminar de los datos disponibles. La siguiente **Tabla 1** resume los principales estadísticos que describen a las variables transaccionales más relevantes (obtenidos utilizando el *software* SPSS¹):

| Variable | Año | Mínimo | Máximo | Rango | Media | Desv. típ. | Varianza |
|---------------------------|------|--------|--------|--------|--------|------------|----------|
| Venta (miles S/.) | 2014 | 1,20 | 13,44 | 12,24 | 5,38 | 2,85 | 8,15 |
| Venta (miles S/.) | 2013 | 1,20 | 14,00 | 12,80 | 5,42 | 2,91 | 8,49 |
| Transacciones | 2013 | 3,00 | 344,00 | 341,00 | 86,89 | 45,70 | 2.088,48 |
| Días distintos con compra | 2013 | 2,00 | 129,00 | 127,00 | 50,63 | 23,43 | 548,81 |
| Días entre compras | 2013 | 2,84 | 287,00 | 284,16 | 9,00 | 5,89 | 34,70 |
| Recencia | 2013 | - | 309,00 | 309,00 | 7,40 | 12,63 | 159,48 |
| Frecuencia mensual | 2013 | 0,17 | 10,75 | 10,58 | 4,22 | 1,95 | 3,81 |
| Frecuencia diaria | 2013 | 0,01 | 0,35 | 0,35 | 0,14 | 0,06 | 0,00 |
| R/F | 2013 | - | 40,85 | 40,85 | 0,82 | 1,25 | 1,56 |
| # categorías | 2013 | 2,00 | 28,00 | 26,00 | 16,23 | 3,15 | 9,90 |
| # productos | 2013 | 4,00 | 302,00 | 298,00 | 127,68 | 43,35 | 1.879,60 |
| # marcas | 2013 | 3,00 | 46,00 | 43,00 | 27,10 | 5,33 | 28,44 |

Tabla 1: Análisis descriptivo de variables transaccionales
Fuente: Elaboración propia en base a información de la empresa

Es importante destacar que, hay diferencias entre las ventas y número de las bodegas según la región del país (ver **Gráfico 6**), en donde se destaca Lima, capital de Perú, como la región con mayor venta y número de clientes, seguida de las principales ciudades del país, Arequipa (sur), Cusco (centro) y Piura (norte).

¹ *Statistical Package for the Social Sciences*

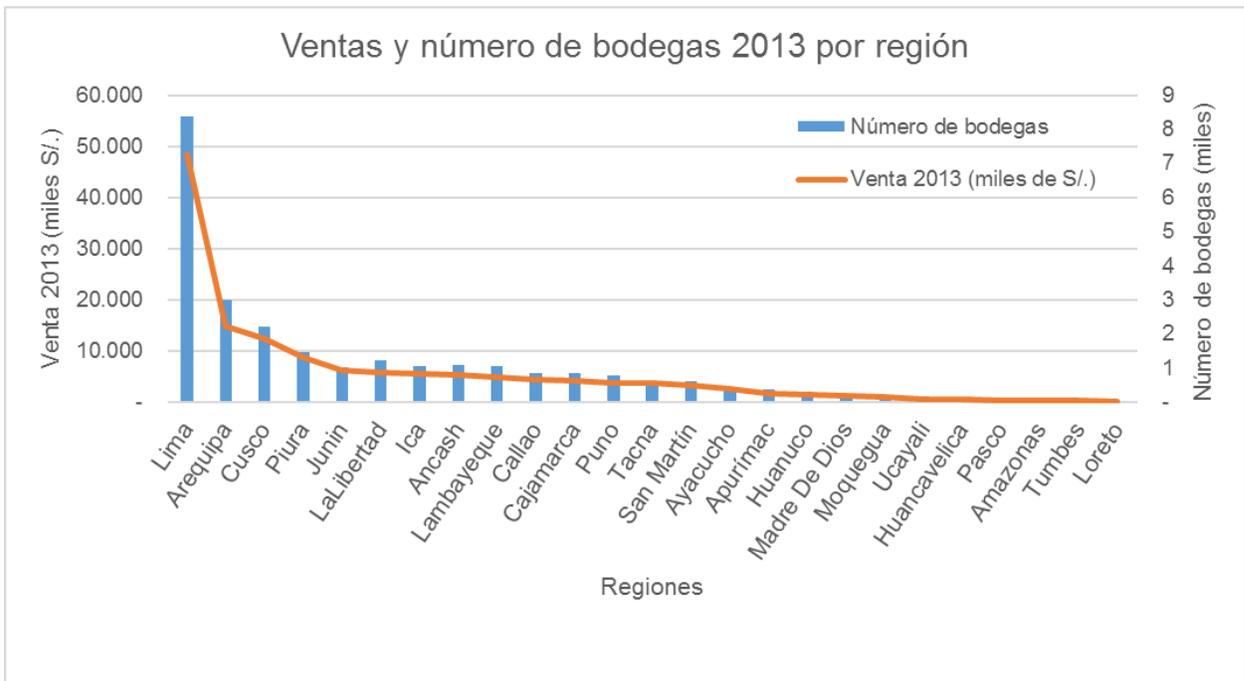


Gráfico 6: Ventas totales y bodegas 2013

Fuente: Elaboración propia en base a información de la empresa

Complementando lo anterior, a medida que se observan más y más los datos, a modo de ejemplo, se encuentran 3 clientes que, a pesar de pertenecer a la misma región (Lima) y distrito (Miraflores), e incluso sólo a una calle de distancia y, además, que compran la misma cantidad de diferentes categorías (25) y productos (340) al año, tienen un nivel de venta anual distinto (ver **Gráfico 7**), por lo tanto, se debe generar un modelo que permita entender estas diferencias, y más aún, que pueda explicar cuanto más podrían comprar los clientes de acuerdo al set de variables que se consideren en el modelo.

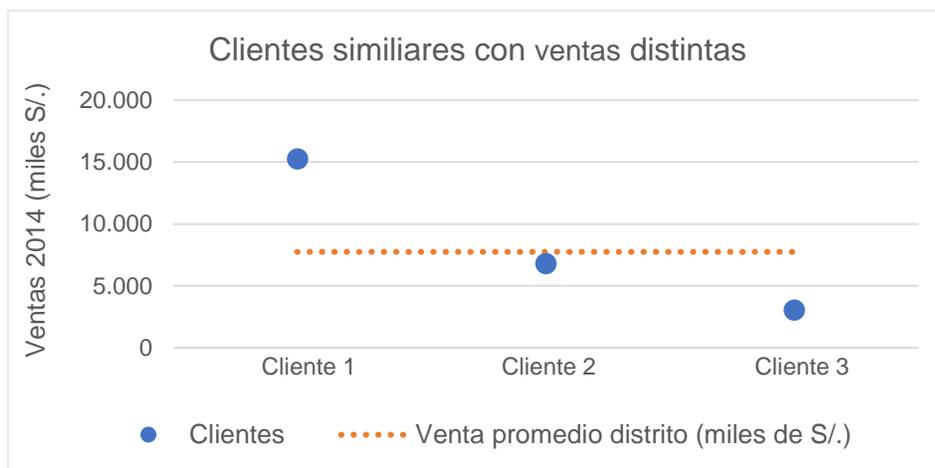


Gráfico 7: Ejemplo – Bodegas similares

Fuente: Elaboración propia en base a información de la empresa

7.1.4 Creación de variables

A partir del análisis anterior, se levantaron las siguientes variables:

- i. **Tipo de vendedor:** dado que un vendedor, en promedio, atiende a 132 clientes y, aproximadamente, un cliente es atendido por 3 vendedores en el año, se consideran las ventas anuales totales de cada vendedor, ya que representa la contribución económica que realizan los vendedores a la empresa. Para ello, se sumaron todas las ventas del año 2013 según vendedor y se clasificaron de acuerdo a cuartiles para formar los tipos de distintos vendedores, donde A es el tipo que corresponde al mejor vendedor (el que más vendió) y D es el peor (el que menos vendió). En la **Tabla 2** se observa cual fue la clasificación según los cuartiles de sus ventas totales del año 2013.

| Tipo de vendedor | Rango (S/.) | Número de vendedores | Venta total (millones S/.) | Número de clientes | Venta (millones S./) /número de vendedores | Venta (millones S./) /número de clientes |
|------------------|-----------------------|----------------------|----------------------------|--------------------|--|--|
| A | (418.725; 21.979.132] | 166 | 411.362 | 38.746 | 2.478,08 | 10,62 |
| B | (253.791; 418.724] | 165 | 6.027 | 17.866 | 36,53 | 0,34 |
| C | (117.807; 253.790] | 166 | 2.872 | 16.228 | 17,30 | 0,18 |
| D | [51; 117.806] | 166 | 1.039 | 14.432 | 6,26 | 0,07 |
| Total | | 663 | 421.301 | 87.272 | 635,45 | 4,83 |

Tabla 2: Clasificación del tipo de vendedor según ventas 2013 (S/.)

Fuente: Elaboración propia en base a información de la empresa

- ii. A nivel de **región**, se añadieron variables con datos secundarios obtenidos por el Instituto Nacional de Estadística e Informática de Perú (INEI), tales como:
- Superficie (km^2).
 - Número de personas (miles).
 - Densidad poblacional ($personas/km^2$).
 - Crecimiento poblacional (%).
 - Ingreso promedio mensual del hogar (S/.).
 - Valor Bruto Agregado (VAB).
 - Producto Interno Bruto (PIB) per cápita.

En la siguiente **Tabla 3**, se pueden observar los valores de los datos secundarios, agregando el número de clientes del proveedor peruano según región, donde se destaca la magnitud del valor con diferentes colores en cada columna (bajo valor es rojo, alto valor es verde y las escalas de colores amarillo/verde claro representan los valores intermedios):

| Nombre región | Número de clientes | Número de personas (miles) | Superficie (km ²) | Densidad (personas/km ²) | Crecimiento poblacional | Ingreso prom. mensual del hogar (S/.) | VAB comercio | PIB per cápita |
|---------------|--------------------|----------------------------|-------------------------------|--------------------------------------|-------------------------|---------------------------------------|--------------|----------------|
| Amazonas | 79 | 423 | 39.249 | 10,77 | 1,0 | 859 | 7,8 | 7.531 |
| Ancash | 1.086 | 1.149 | 35.915 | 31,98 | 0,5 | 973 | 8,7 | 18.014 |
| Apurímac | 372 | 459 | 20.896 | 21,96 | 0,5 | 620 | 7,2 | 5.156 |
| Arequipa | 3.011 | 1.287 | 63.345 | 20,32 | 1,1 | 1.300 | 8,0 | 23.212 |
| Ayacucho | 458 | 689 | 43.815 | 15,72 | 1,1 | 750 | 7,7 | 8.637 |
| Cajamarca | 851 | 1.530 | 33.318 | 45,91 | 0,3 | 798 | 7,7 | 9.564 |
| Callao | 861 | 1.010 | 147 | 6.872,89 | 1,4 | 1.306 | - | - |
| Cusco | 2.226 | 1.317 | 71.987 | 18,29 | 0,6 | 974 | 8,3 | 13.961 |
| Huancavelica | 64 | 495 | 22.132 | 22,36 | 0,8 | 642 | 7,0 | 7.840 |
| Huanuco | 306 | 861 | 37.022 | 23,24 | 0,7 | 813 | 6,9 | 5.425 |
| Ica | 1.038 | 787 | 21.328 | 36,91 | 1,0 | 1.023 | 8,7 | 20.893 |
| Junín | 1.025 | 1.351 | 44.327 | 30,47 | 0,7 | 976 | 7,1 | 11.319 |
| La Libertad | 1.241 | 1.860 | 25.500 | 72,93 | 1,2 | 1.001 | 9,0 | 14.150 |
| Lambayeque | 1.069 | 1.261 | 14.231 | 88,58 | 0,8 | 843 | 8,9 | 10.870 |
| Lima | 8.380 | 9.838 | 34.802 | 282,70 | 1,5 | 1.487 | 9,0 | 25.748 |
| Loreto | 16 | 1.039 | 368.852 | 2,82 | 1,0 | 963 | 8,0 | 9.879 |
| Madre de Dios | 222 | 137 | 85.301 | 1,61 | 2,4 | 1.822 | 7,5 | 17.288 |
| Moquegua | 161 | 180 | 15.734 | 11,47 | 1,0 | 1.780 | 7,9 | 43.205 |
| Pasco | 71 | 304 | 25.028 | 12,15 | 0,7 | 835 | 6,1 | 18.241 |
| Piura | 1.448 | 1.844 | 35.893 | 51,38 | 0,8 | 886 | 10,1 | 13.631 |
| Puno | 767 | 1.416 | 71.999 | 19,66 | 0,9 | 721 | 7,2 | 8.249 |
| San Martín | 600 | 841 | 51.253 | 16,40 | 1,4 | 983 | 8,8 | 7.491 |
| Tacna | 606 | 342 | 16.076 | 21,26 | 1,3 | 1.244 | 7,7 | 21.463 |
| Tumbes | 45 | 238 | 4.669 | 50,90 | 1,3 | 1.118 | 8,1 | 11.050 |
| Ucayali | 153 | 496 | 102.400 | 4,84 | 1,2 | 1.015 | 8,4 | 10.151 |

Tabla 3: Datos según región

Fuente: Elaboración propia en base a información de la empresa y valores de los datos secundarios según el INEI² de Perú

Se destaca el alto número de habitantes que hay en Lima, que representa cerca de un 32% del total de personas, al igual que el porcentaje de clientes con respecto al total. Se observa que hay grandes diferencias en el ingreso promedio mensual del hogar y PIB per cápita en las regiones. Además, Perú tiene un crecimiento poblacional en torno a un 1,01%, y hay diferencias notorias en la densidad de las regiones.

² Instituto Nacional de Estadística e Informática de Perú.

iii. **Zona:** se clasificó a cada región según la zona del donde pertenece como se detalla a continuación país (ver Anexo 8: Mapa político de Perú):

- Centro: Ancash, Ayacucho, Callao, Huancavelica, Huánuco, Ica, Junín, Pasco y Ucayali.
- Norte: Amazonas, Cajamarca, La Libertad, Lambayeque, Loreto, Piura, San Martín y Tumbes.
- Sur: Apurímac, Arequipa, Cusco, Madre de Dios, Moquegua, Puno y Tacna.
- Lima

Además, como se puede observar en el **Gráfico 8**, el único sector donde aumentan las ventas el 2014 en un 1,63% con respecto al año 2013, es la zona centro.

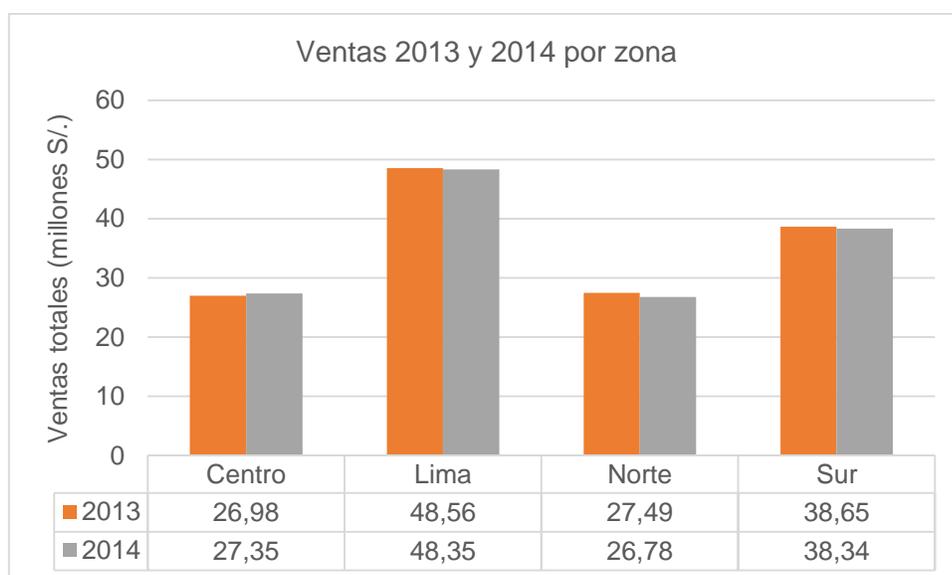


Gráfico 8: Ventas totales 2013 y 2014 por zona

Fuente: Elaboración propia en base a información de la empresa

ii. **Puntaje RFM** (S. Gupta 2006): tomando en cuenta variables como número de transacciones, último día con compra del año, días entre las compras, entre otras y con el fin de analizar el valor del cliente, se define:

- R: Recencia (qué tan reciente fue su última compra el cliente).
- F: Frecuencia (qué tan seguido compra el cliente).
- M: Valor monetario (suma del valor de la venta).

Un método es asignar una escala del 1 al 5 donde 5 es el máximo valor que pueden tomar las variables R, F y M, luego se concatenan y se forma un puntaje en donde se ordena para crear un ranking. Además, se agregó la variable R/F, que es recencia dividido en frecuencia para medir fuga de clientes.

Los principales estadísticos de estas variables creadas se pueden ver en la siguiente **Tabla 4**.

| Variable | Año | Mínimo | Máximo | Rango | Media | Desv. típ. | Varianza |
|----------|------|--------|--------|-------|--------|------------|-----------|
| Score R | 2013 | 1 | 5 | 4 | 3,54 | 1,23 | 1,52 |
| Score F | 2013 | 1 | 5 | 4 | 2,76 | 1,08 | 1,16 |
| Score M | 2013 | 1 | 5 | 4 | 2,47 | 1,11 | 1,24 |
| RFM | 2013 | 111 | 554 | 443 | 384,23 | 128,51 | 16.515,25 |
| R/F | 2013 | - | 40,85 | 40,85 | 0,82 | 1,25 | 1,56 |

Tabla 4: Análisis descriptivo de variables RFM

Fuente: Elaboración propia en base a información de la empresa

- iii. **Variables *dummies*:** Para incorporar en los modelos de estimación de potencial de compra, se crean variables *dummies* para categorizar las variables no escalares, tales como zona, región, distrito y sede de distribución, de los productos de consumo masivo del proveedor peruano, a la que pertenece el cliente.

7.1.5 Selección de variables

Luego de analizar las variables de forma univariada, midiendo la variabilidad interna de cada variable y verificando la completitud de las variables como se hizo anteriormente, se observa la matriz de correlación de las variables y utilizando un criterio de eliminación de variables correlacionadas, se procedió a elegir las variables explicativas de los modelos:

1. **Modelo (I): Estimación ventas 2014:** Se seleccionan las variables de tal manera de minimizar la correlación entre las variables de acuerdo a la siguiente **Tabla 5** (nula correlación es rojo, correlación perfecta es verde y las escalas de colores amarillo/verde claro representa los valores intermedios):

| Variable / Correlación de Pearson | Venta (miles S/.) 2014 | Venta (miles S/.) 2013 | Transacciones 2013 | Días distintos con compra 2013 | Días entre compras 2013 | Recencia 2013 | Frecuencia mensual 2013 | Frecuencia diaria 2013 | R/F 2013 | Score R 2013 | Score F 2013 | Score M 2013 | RFM 2013 | # categorías 2013 | # productos 2013 | # marcas 2013 |
|-----------------------------------|------------------------|------------------------|--------------------|--------------------------------|-------------------------|---------------|-------------------------|------------------------|----------|--------------|--------------|--------------|----------|-------------------|------------------|---------------|
| Venta (miles S/.) 2014 | 1 | 0,80 | 0,53 | 0,39 | -0,24 | -0,14 | 0,39 | 0,39 | -0,08 | 0,19 | 0,32 | 0,76 | 0,22 | 0,39 | 0,51 | 0,41 |
| Venta (miles S/.) 2013 | 0,80 | 1 | 0,61 | 0,45 | -0,29 | -0,16 | 0,45 | 0,45 | -0,08 | 0,20 | 0,38 | 0,93 | 0,23 | 0,47 | 0,59 | 0,50 |
| Transacciones 2013 | 0,53 | 0,61 | 1 | 0,90 | -0,63 | -0,30 | 0,90 | 0,90 | -0,10 | 0,40 | 0,80 | 0,61 | 0,45 | 0,60 | 0,79 | 0,64 |
| Días distintos con compra 2013 | 0,39 | 0,45 | 0,90 | 1 | -0,73 | -0,34 | 1 | 1 | -0,10 | 0,43 | 0,92 | 0,46 | 0,50 | 0,50 | 0,64 | 0,53 |
| Días entre compras 2013 | -0,24 | -0,29 | -0,63 | -0,73 | 1 | 0,30 | -0,73 | -0,73 | 0,01 | -0,32 | -0,75 | -0,30 | -0,38 | -0,41 | -0,50 | -0,44 |
| Recencia 2013 | -0,14 | -0,16 | -0,30 | -0,34 | 0,30 | 1 | -0,34 | -0,34 | 0,84 | -0,64 | -0,34 | -0,17 | -0,64 | -0,20 | -0,25 | -0,22 |
| Frecuencia mensual 2013 | 0,39 | 0,45 | 0,90 | 1 | -0,73 | -0,34 | 1 | 1 | -0,10 | 0,43 | 0,92 | 0,46 | 0,50 | 0,50 | 0,64 | 0,53 |
| Frecuencia diaria 2013 | 0,39 | 0,45 | 0,90 | 1 | -0,73 | -0,34 | 1 | 1 | -0,10 | 0,43 | 0,92 | 0,46 | 0,50 | 0,50 | 0,64 | 0,53 |
| R/F 2013 | -0,08 | -0,08 | -0,10 | -0,10 | 0,01 | 0,84 | -0,10 | -0,10 | 1 | -0,60 | -0,08 | -0,08 | -0,58 | -0,06 | -0,08 | -0,06 |
| Score R 2013 | 0,19 | 0,20 | 0,40 | 0,43 | -0,32 | -0,64 | 0,43 | 0,43 | -0,60 | 1 | 0,41 | 0,21 | 1 | 0,24 | 0,30 | 0,24 |
| Score F 2013 | 0,32 | 0,38 | 0,80 | 0,92 | -0,75 | -0,34 | 0,92 | 0,92 | -0,08 | 0,41 | 1 | 0,40 | 0,48 | 0,48 | 0,60 | 0,51 |
| Score M 2013 | 0,76 | 0,93 | 0,61 | 0,46 | -0,30 | -0,17 | 0,46 | 0,46 | -0,08 | 0,21 | 0,40 | 1 | 0,24 | 0,48 | 0,60 | 0,51 |
| RFM 2013 | 0,22 | 0,23 | 0,45 | 0,50 | -0,38 | -0,64 | 0,50 | 0,50 | -0,58 | 1 | 0,48 | 0,24 | 1 | 0,27 | 0,34 | 0,28 |
| # categorías 2013 | 0,39 | 0,47 | 0,60 | 0,50 | -0,41 | -0,20 | 0,50 | 0,50 | -0,06 | 0,24 | 0,48 | 0,48 | 0,27 | 1 | 0,78 | 0,88 |
| # productos 2013 | 0,51 | 0,59 | 0,79 | 0,64 | -0,50 | -0,25 | 0,64 | 0,64 | -0,08 | 0,30 | 0,60 | 0,60 | 0,34 | 0,78 | 1 | 0,85 |
| # marcas 2013 | 0,41 | 0,50 | 0,64 | 0,53 | -0,44 | -0,22 | 0,53 | 0,53 | -0,06 | 0,24 | 0,51 | 0,51 | 0,28 | 0,88 | 0,85 | 1 |

Tabla 5: Matriz de correlación de Pearson de las variables transaccionales principales de los clientes

Fuente: Elaboración propia en base a información de la empresa

Además, se eliminan variables si son una combinación entre ellas como R/F con recencia y frecuencia, pero se conservan las variables que intuitivamente puedan tener un poder explicativo en el modelo. Por lo tanto, las variables seleccionadas por cliente son, venta (miles S/.) 2014, venta (miles S/.) 2013, días entre compras 2013, R/F 2013 y número de categorías distintas con compra.

Aplicando el mismo análisis para las variables de las regiones, se elimina superficie (km^2) y número de habitantes para incluir densidad poblacional ($personas/km^2$) como combinación de las variables anteriores, junto con la eliminación de variables por correlaciones de acuerdo a la siguiente **Tabla 6:**

| Variable/ Correlación de Pearson | Número de clientes | Densidad | Crecimiento poblacional | Ingreso prom. mensual del hogar | VAB comercio | PIB per cápita |
|----------------------------------|--------------------|----------|-------------------------|---------------------------------|--------------|----------------|
| Número de clientes | 1 | -0,06 | 0,69 | 0,80 | 0,33 | 0,75 |
| Densidad | -0,06 | 1 | 0,21 | 0,16 | -0,88 | -0,38 |
| Crecimiento poblacional | 0,69 | 0,21 | 1 | 0,83 | 0,02 | 0,53 |
| Ingreso prom. mensual del hogar | 0,80 | 0,16 | 0,83 | 1 | 0,06 | 0,79 |
| VAB comercio | 0,33 | -0,88 | 0,02 | 0,06 | 1 | 0,53 |
| PIB per cápita | 0,75 | -0,38 | 0,53 | 0,79 | 0,53 | 1 |

Tabla 6: Matriz de correlación de Pearson de las variables de región

Fuente: Elaboración propia en base a información de la empresa

Dado que existe un alto nivel de correlación entre la mayoría de las variables (color verde claro en la tabla), se seleccionan las siguientes variables a estudiar posteriormente en los modelos: nombre de la región, densidad poblacional (*personas/km²*), número de clientes de la región y zona a la que pertenece.

Por último, de acuerdo a las variables del tipo de vendedor (A, B, C y D) 2013, se selecciona el número de vendedores de cada tipo que atienden a cada cliente.

2. **Modelo (II): categorías principales:** Algunas de las categorías que provee la empresa son de origen propio y otras son de socios comerciales. En el siguiente **Gráfico 9**, se observa cual es la distribución de las ventas de cada categoría según su origen, junto con el número de clientes que compran en la categoría correspondiente:

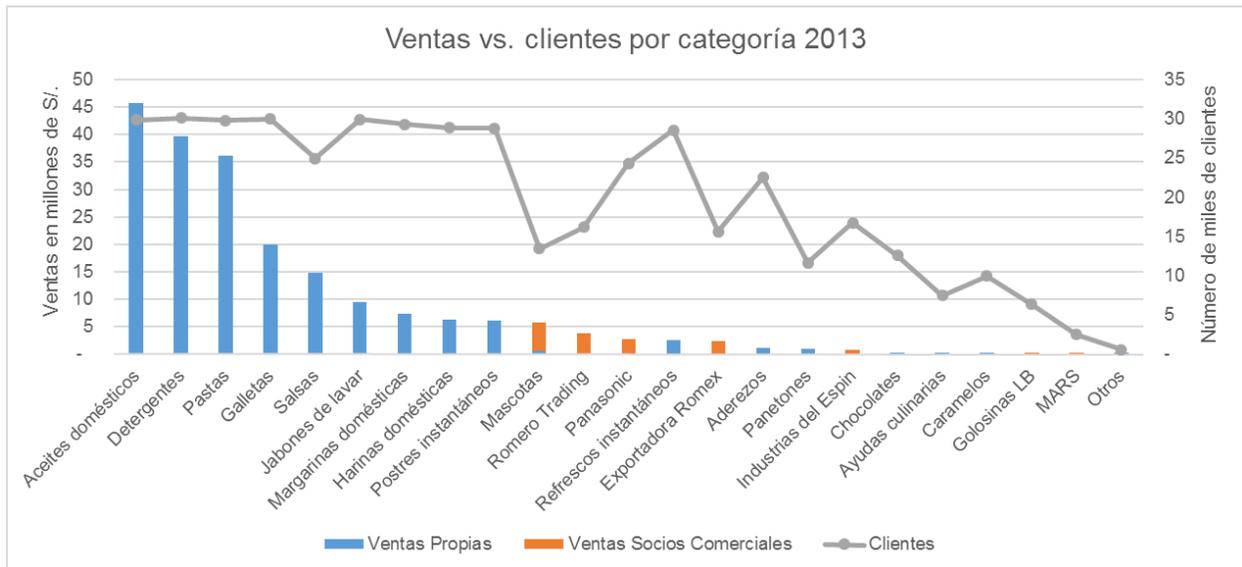


Gráfico 9: Ventas propias vs. socios comerciales según categoría 2013

Fuente: Elaboración propia en base a información de la empresa

En primer lugar, se descarta utilizar las ventas de las categorías de socios comerciales de la empresa. Luego, se analizó la matriz de correlaciones entre las ventas 2013 de cada categoría (ver Anexo 9), sin resultados concluyentes dada la baja correlación entre las variables, implica que las categorías seleccionadas, para estimar en los modelos, son las 5 más importantes que hacen el 81,44% de las ventas totales propias del 2013 de la empresa (ver **Tabla 7**), las cuales son aceites domésticos, detergentes, galletas, pastas y salsas (variables a estudiar). Sin embargo, se incorporan las otras categorías como variables independientes en los modelos de categorías por la gran cantidad de clientes que las han comprado.

Complementando lo anterior, se realizó un análisis descriptivo de las ventas (S/.) del año 2013 de cada categoría, resumiendo los principales estadísticos en la **Tabla 7**, junto con el número de clientes que compran y el porcentaje de la venta que representa en comparación al resto.

| Categorías | Mínimo | Máximo | Media | Mediana | Desv.típ | Varianza | Clientes | Venta (%) |
|------------------------|--------|-----------|----------|---------|---------------|------------|----------|---------------|
| Aceites domésticos | 0 | 9.052,41 | 1.079,26 | 899,72 | 771,45 | 595.145,04 | 25.906 | 23,06% |
| Detergentes | 0 | 11.797,28 | 1.008,09 | 436,89 | 705,68 | 497.993,16 | 25.977 | 21,54% |
| Pastas | 0 | 9.052,47 | 806,69 | 109,47 | 712,00 | 506.956,54 | 25.773 | 17,24% |
| Galletas | 0 | 9.625,89 | 546,00 | 106,49 | 437,80 | 191.675,94 | 26.020 | 11,67% |
| Salsas | 0 | 7.405,75 | 371,53 | - | 364,89 | 133.141,68 | 21.330 | 7,94% |
| Jabones de lavar | 0 | 6.054,07 | 222,38 | 158,23 | 222,50 | 49.505,16 | 25.741 | 4,75% |
| Margarinas domésticas | 0 | 5.421,03 | 205,92 | - | 188,00 | 35.346,17 | 25.280 | 4,40% |
| Postres instantáneos | 0 | 2.825,82 | 153,18 | 52,62 | 156,08 | 24.361,91 | 24.770 | 3,27% |
| Harinas domésticas | 0 | 8.033,92 | 146,73 | 171,86 | 180,71 | 32.655,90 | 24.932 | 3,13% |
| Refrescos instantáneos | 0 | 1.746,35 | 71,07 | 290,94 | 72,85 | 5.306,86 | 24.696 | 1,52% |
| Panetones | 0 | 2.439,60 | 27,60 | 611,94 | 73,34 | 5.379,64 | 9.672 | 0,59% |
| Aderezos | 0 | 1.481,80 | 25,77 | - | 42,95 | 1.844,43 | 19.092 | 0,55% |
| Chocolates | 0 | 347,31 | 7,09 | - | 14,86 | 220,83 | 10.404 | 0,15% |
| Ayudas culinarias | 0 | 401,54 | 5,19 | - | 15,34 | 235,34 | 6.050 | 0,11% |
| Caramelos | 0 | 500,45 | 3,97 | - | 11,67 | 136,24 | 8.557 | 0,08% |
| Cuidado del cabello | 0 | 123,55 | 0,02 | 848,45 | 1,08 | 1,17 | 45 | 0,00% |
| Suavizante | 0 | 8,41 | 0,00 | - | 0,08 | 0,01 | 3 | 0,00% |

Tabla 7: Análisis descriptivo de la venta 2013 (S/.) de las categorías propias

Fuente: Elaboración propia en base a información de la empresa

Se puede observar que existe una desviación estándar mayor al promedio en la mayoría de las categorías, excepto con aceites, salsas y margarinas, además de una media muy diferente entre todas las categorías propias del proveedor. Lo cual, significa que existe una alta dispersión de las ventas entre los clientes, por lo que se induce que los clientes tienen un comportamiento de compra diferente, a pesar que muchos compren las mismas categorías. Cabe destacar que suavizante y cuidado del cabello, tienen una cantidad despreciable de clientes que las compran y promedio de ventas en comparación al resto de las categorías, lo que hace que no sean relevantes en análisis posteriores.

7.1.6 Selección muestra de datos

Para el cálculo del error y validación de resultados en los modelos, es necesario seleccionar 2 muestras de datos para calibrar el modelo y testarlo. Aleatoriamente, se toma 80% de los datos para ajustar el modelo y el 20% restante de los datos para medir

la capacidad predictiva del modelo, lo cual se realiza comparando si son similares los coeficientes de determinación.

7.2 Modelo predictivo de potencial de compra del cliente

7.2.1 Modelo simple: promedio general

Previo a los siguientes modelos más complejos, sólo en esta parte para el cálculo de potencial de compra se considera que el potencial de compra de los clientes será el promedio de venta del sector al que pertenecen en cada caso a analizar. Por lo tanto, los clientes con diferencial de potencial son las bodegas bajo el promedio de ventas, de acuerdo a las ventas de las regiones/zonas a las que pertenecen y las 5 categorías principales del proveedor.

7.2.1.1 Resultados por región

Considerando la venta total y el número de clientes por región del año 2014, se calcula la venta promedio de cada región con el fin de obtener la cantidad de clientes que están bajo ese promedio de ventas, los cuales serán los clientes con venta potencial. Luego de identificar a aquellos clientes potenciales, la suma total de la diferencia entre el promedio de venta por región y la venta real de cada bodega, constituye el diferencial de potencial total de cada región. Además, el porcentaje de aumento de ventas entre la venta real y la venta potencial, indica en la **Tabla 8** que las primeras cinco regiones aumentarían en más de un 23% son Tumbes, La Libertad, Pasco, Ancash y Apurímac, las cuales pertenecen a distintas zonas de Perú, por lo que no hay una clara relación entre las zonas o cercanía entre las regiones en términos de potencial. Dado lo anterior, hace sentido agregar más variables para llegar a conclusiones más certeras.

| Región | Venta 2014 (miles S./) | Número de clientes | Venta promedio (miles S./) | Venta potencial (miles S./) | # clientes potenciales | Diferencial de potencial (miles S./) | Aumento en ventas (%) |
|----------------------|---------------------------|--------------------------|----------------------------------|-----------------------------------|---------------------------|---|-----------------------------|
| Lima | 48.345,41 | 8.380 | 5,77 | 58.618,58 | 4.684 | 10.273,17 | 21,2% |
| Arequipa | 15.273,99 | 3.011 | 5,07 | 18.559,10 | 1.778 | 3.285,11 | 21,5% |
| Cusco | 11.959,84 | 2.227 | 5,37 | 14.439,55 | 1.302 | 2.479,72 | 20,7% |
| Piura | 7.761,65 | 1.448 | 5,36 | 9.377,06 | 808 | 1.615,41 | 20,8% |
| Junin | 6.104,65 | 1.025 | 5,96 | 7.300,52 | 563 | 1.195,87 | 19,6% |
| Ancash | 5.609,68 | 1.086 | 5,17 | 6.916,72 | 642 | 1.307,05 | 23,3% |
| Ica | 5.568,46 | 1.038 | 5,36 | 6.765,45 | 615 | 1.196,99 | 21,5% |
| LaLibertad | 5.375,09 | 1.241 | 4,33 | 6.642,18 | 749 | 1.267,09 | 23,6% |
| Lambayeque | 4.846,99 | 1.069 | 4,53 | 5.944,90 | 637 | 1.097,91 | 22,7% |
| Cajamarca | 4.545,61 | 851 | 5,34 | 5.536,89 | 492 | 991,28 | 21,8% |
| Callao | 4.456,90 | 861 | 5,18 | 5.489,87 | 497 | 1.032,97 | 23,2% |
| Puno | 3.668,44 | 767 | 4,78 | 4.458,60 | 453 | 790,16 | 21,5% |
| Tacna | 3.529,37 | 606 | 5,82 | 4.229,67 | 338 | 700,31 | 19,8% |
| SanMartin | 3.478,20 | 600 | 5,80 | 4.194,95 | 338 | 716,75 | 20,6% |
| Ayacucho | 2.463,40 | 458 | 5,38 | 2.971,31 | 264 | 507,91 | 20,6% |
| Apurímac | 1.809,47 | 372 | 4,86 | 2.230,48 | 223 | 421,01 | 23,3% |
| Huanuco | 1.616,34 | 307 | 5,26 | 1.989,61 | 177 | 373,27 | 23,1% |
| MadredeDios | 1.124,92 | 222 | 5,07 | 1.367,48 | 126 | 242,55 | 21,6% |
| Moquegua | 980,05 | 161 | 6,09 | 1.164,30 | 88 | 184,25 | 18,8% |
| Ucayali | 688,61 | 153 | 4,50 | 844,22 | 99 | 155,61 | 22,6% |
| Amazonas | 444,19 | 79 | 5,62 | 536,10 | 42 | 91,91 | 20,7% |
| Pasco | 420,51 | 71 | 5,92 | 518,68 | 41 | 98,17 | 23,3% |
| Huancavelica | 435,65 | 64 | 6,81 | 509,41 | 30 | 73,76 | 16,9% |
| Tumbes | 217,71 | 45 | 4,84 | 269,65 | 29 | 51,94 | 23,9% |
| Loreto | 109,15 | 16 | 6,82 | 130,91 | 10 | 21,76 | 19,9% |
| Total general | 140.834,28 | 26.158 | 5,38 | 171.006,19 | 15.025 | 30.171,92 | 21,4% |

Tabla 8: Resultados preliminares de potencial de compra por región

Fuente: Elaboración propia en base a información de la empresa

En la misma **Tabla 8**, se puede observar que las principales regiones de Perú, Lima, Arequipa, Cusco y Piura son las que tienen una venta potencial mayor, junto con un número de clientes potenciales superior al resto de las regiones, por lo tanto, poseen un diferencial de potencial mayor, es decir, existe más dispersión en la distribución de los clientes de esas regiones dado que los clientes que menos compran están más alejados del promedio que las otras regiones.

7.2.1.2 Resultados por zona

Se aplica el mismo procedimiento explicado en la sección anterior (resultados por regiones) pero con las zonas de Perú, con lo que se corrobora que Lima tiene mayor venta potencial y número de clientes potenciales y, por lo tanto, un alto diferencial de potencial, pero con el menor porcentaje de aumento en ventas que el resto de las zonas.

| Zona | Venta 2014 (miles S/.) | Número de clientes | Venta promedio (miles S/.) | Venta potencial (miles S/.) | # clientes potenciales | Diferencial de potencial (miles S/.) | Aumento en ventas (%) |
|----------------------|---------------------------|-----------------------|----------------------------------|-----------------------------------|---------------------------|--|-----------------------------|
| Lima | 48.345,41 | 8.380 | 5,77 | 58.618,58 | 4.684 | 10.273,17 | 21,25% |
| Sur | 38.346,08 | 7.366 | 5,21 | 46.509,61 | 4.316 | 8.163,53 | 21,29% |
| Centro | 27.364,19 | 5.063 | 5,40 | 33.383,20 | 2.916 | 6.019,01 | 22,00% |
| Norte | 26.778,60 | 5.349 | 5,01 | 32.777,15 | 3.091 | 5.998,55 | 22,40% |
| Total general | 140.834,28 | 26.158 | 5,38 | 171.288,54 | 15.007 | 30.454,27 | 21,62% |

Tabla 9: Resultados preliminares de potencial de compra por zonas
Fuente: Elaboración propia

7.2.1.3 Resultados categorías principales

En el cálculo de potencial en las categorías principales del año 2013, se toman en cuenta las cinco categorías con mayor venta del proveedor y se realizan cálculos semejantes a los resultados anteriores por región/zona, es decir, se suma la venta total de cada categoría para calcular la venta promedio con el número de clientes que compran la categoría respectiva, con el objetivo de determinar la cantidad de clientes que están bajo ese promedio. De un total de 26.156 clientes, como muestra la **Tabla 10** el número de clientes potenciales es muy alto dada las grandes diferencias de venta en las categorías lo cual es respaldado por los estadísticos del análisis descriptivo anterior de categorías (ver **Tabla 7:** Análisis descriptivo de la venta 2013 (S/.) de las categorías propias).

Así, las ventas aumentarían en un 31% aproximadamente en las categorías seleccionadas, lo cual es considerable para el proveedor dado que implica que las bodegas que están bajo el promedio varía entre las categorías y, además, una bodega compra más de una categoría. Cabe destacar que, con este método no se pueden diseñar propuestas de acción para aumentar las ventas si se utilizara aquel cálculo de potencial, por lo que, igual a los casos anteriores, se debiese agregar más variables, no obstante, si entrega un valor de referencia que permite la evaluación de los modelos matemáticos que se utilizan en las secciones posteriores de este trabajo.

| Categorías | Venta 2013 (miles S/.) | Número de clientes | Venta promedio (miles S/.) | Venta potencial (miles S/.) | # clientes potenciales | Diferencial de potencial (miles S/.) | Aumento en ventas (%) |
|--------------|---------------------------|-----------------------|----------------------------------|-----------------------------------|---------------------------|--|-----------------------------|
| Aceites | 28.228,77 | 25.906 | 1,08 | 35.799,39 | 15.715 | 7.570,62 | 26,82% |
| Detergentes | 26.366,98 | 25.977 | 1,02 | 33.361,54 | 15.554 | 6.994,56 | 26,53% |
| Pastas | 21.099,81 | 25.773 | 0,82 | 27.851,58 | 16.446 | 6.751,77 | 32,00% |
| Galletas | 14.280,94 | 26.020 | 0,55 | 18.374,32 | 16.167 | 4.093,39 | 28,66% |
| Salsas | 9.717,73 | 21.330 | 0,46 | 13.269,26 | 15.530 | 3.551,53 | 36,55% |
| Total | 99.694,24 | 125.006 | 0,77 | 128.656,10 | 79.412 | 28.961,86 | 30,11% |

Tabla 10: Resultados preliminares de potencial de compra por categorías
Fuente: Elaboración propia en base a información de la empresa

Finalmente, se concluye lo mismo que en el caso anterior que, el gran problema de este enfoque es que no se pueden incorporar ni relacionar más variables, incluso se tratan a todas las bodegas como si fueran iguales y tuvieran la misma capacidad de compra, por lo que se hace necesario generar otros modelos para explicar el comportamiento con mayor precisión.

7.2.2 Modelo de regresión lineal

Para estudiar la relación existente entre más variables y la importancia de estas, se realizan regresiones y se comparan con los resultados del algoritmo *Stepwise* del paquete *step* del *software* RStudio, y se desarrollan los siguientes modelos de regresión lineal con variables transaccionales y de las categorías principales del proveedor con el fin de agregar más variables que el caso anterior a la estimación del potencial de compra de los clientes.

7.2.2.1 Modelo (I): Efectos de las ventas 2014

En este Modelo (I) se desea comprobar la influencia que tienen las variables transaccionales, demográficas y del tipo de vendedor del 2013 en las ventas del siguiente año (2014).

Para identificar cuáles son los efectos que influyen en el comportamiento de compra de los clientes, se selecciona como variable dependiente las ventas del año 2014 y las siguientes variables independientes según efecto.

Para poder cuantificar dicha relación, se debe representar la recta de la regresión lineal que subyace en el modelo matemático que relaciona las variables señaladas anteriormente, donde se minimiza la suma de los cuadrados de los residuos.

7.2.2.1.1 Resultados regresión lineal según efecto (individual)

En primer lugar, se realizó por efecto individual cada regresión lineal según la selección de variables anterior. En segundo lugar, aquellas regresiones se comparan con los resultados de la misma regresión lineal, pero utilizando el algoritmo *Stepwise* para ver la mejor combinación de las variables mencionadas en cada efecto.

7.2.2.1.1.1 Efecto cliente

Las variables del año 2013 que se consideran en el efecto cliente son ventas, días entre compras, R/F, número de categorías distintas compradas. Por lo tanto, la ecuación de la regresión lineal es:

$$Ventas_{2014,i} = \beta_0 + \beta_1 Ventas_{2013,i} + \beta_2 Días\ entre\ compras_{2013,i} + \beta_3 R/F_{2013,i} + \beta_4 Número\ de\ categorías\ distintas\ compradas_{2013,i} + \varepsilon_i$$

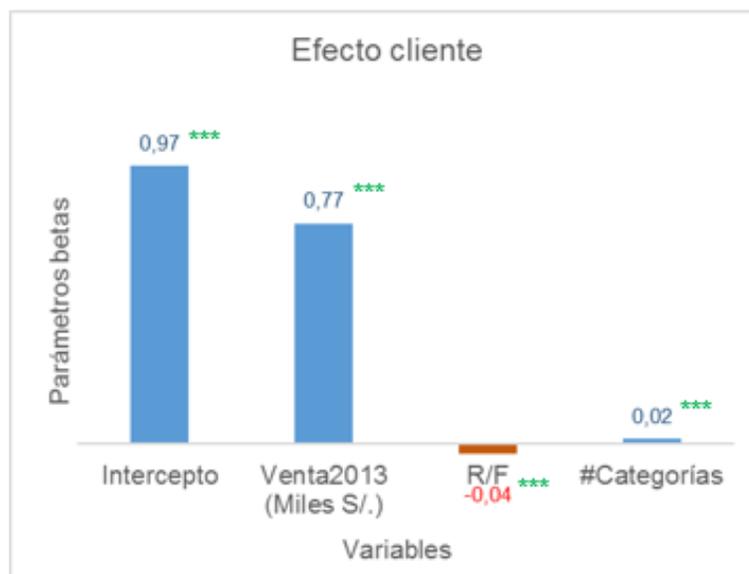
En el efecto del cliente, se ha eliminado la variable días entre compras (2013) dado que no es significativa dentro del conjunto de variables que se tomó para analizar el nivel de ventas del año 2014, lo cual se puede contrastar en los resultados de la regresión lineal 1 de la **Tabla 11** sin utilizar el algoritmo *Stepwise*, junto con sus estadísticos principales en el Anexo 10.

| Variables | Coeficientes | Error std. | t valor | Pr(> t) | Signif. |
|------------------------|--------------|------------|---------|----------|---------|
| Intercepto | 1,01 | 0,08 | 12,88 | < 2e-16 | *** |
| Venta 2013 (Miles S/.) | 0,77 | 0,00 | 167,02 | < 2e-16 | *** |
| Días entre compra 2013 | 0,00 | 0,00 | -0,81 | 0,42 | |
| R/F | -0,04 | 0,01 | -3,95 | 0,00 | *** |
| #Categorías | 0,02 | 0,00 | 3,37 | 0,00 | *** |

Códigos de significancia: 0 '***' 0,001 '**' 0,01 '*' 0,05 '.' 0,1 ' ' 1

Tabla 11: Modelo (I) – Efecto cliente: Resultados regresión lineal 1 (parámetros betas)
Fuente: Elaboración propia en base a información de la empresa

Luego, los resultados de la regresión lineal 2 del efecto cliente, utilizando el algoritmo *Stepwise*, mostrados en el **Gráfico 10** muestran que todas las variables son significativas sin incorporar la variable días entre compras 2013.



Códigos de significancia: 0 '***' 0,001 '**' 0,01 '*' 0,05 '.' 0,1 '.' 1

Gráfico 10: Modelo (I) – Efecto cliente: comparación pesos de los parámetros betas en la regresión lineal 2

Fuente: Elaboración propia en base a información de la empresa

Comparando los parámetros betas presentados, y sin incluir el intercepto, la variable con mayor relevancia para el efecto cliente es la venta del año anterior (2013), el beta de la variable R/F afecta negativamente a la venta, puesto esta variable indica que mientras más grande sea el R/F más tiempo ha pasado desde su último periodo de compra, es decir, ha pasado más de un ciclo de compra desde la última vez que compró, por lo tanto, es coherente el signo del beta. Por otro lado, el beta positivo de la variable de número de categorías distintas que compra el cliente, implica que mientras más categorías compre el cliente, afecta positivamente a la venta del siguiente año (2014).

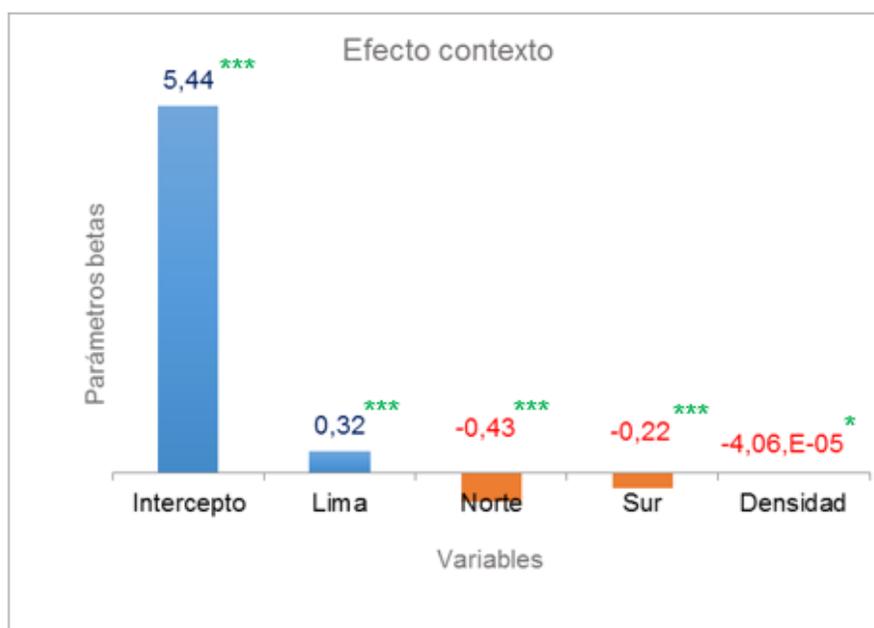
7.2.2.1.1.2 Efecto contexto

Para este efecto se incluyen las variables de las regiones, tales como el nombre de la región, densidad poblacional (*personas/km²*), número de clientes de la región y zona a la que pertenece. La ecuación de la regresión lineal es:

$$Ventas_{2014,i} = \beta_{02} + \beta_1 Regi\acute{o}n_i + \beta_2 Densidad + \beta_3 N\acute{u}mero\ de\ clientes + \beta_4 Zona + \varepsilon_i$$

Al comparar las regresiones en el efecto del contexto con y sin el algoritmo *Stepwise* (ver Anexo 11) el principal cambio es que el algoritmo *Stepwise* elimina la variable dummy nombre de la regi3n (utiliza como base la regi3n Amazonas), junto con el n3mero de clientes de la regi3n y mantiene la variable densidad poblacional y la dummy zonas, dado que esa combinaci3n minimiza el AIC del modelo, adem3s algunas de las variables resultaron no ser significativas.

Los principales resultados de la regresi3n lineal 2, utilizando algoritmo *Stepwise*, se presentan en el **Gr3fico 11** y en Anexo 11, a continuaci3n:



C3digos de significancia: 0 '***' 0,001 '**' 0,01 '*' 0,05 '.' 0,1 '.' 1

Gr3fico 11: Modelo (I) – Efecto contexto: Comparaci3n pesos de los par3metros betas en la regresi3n lineal 2

Fuente: Elaboraci3n propia en base a informaci3n de la empresa

Aquellos resultados, se basan en la zona centro como dummy (puesto que se debe eliminar una de las variables dummies para realizar la regresi3n lineal) y los betas de las zonas Norte y Sur afectan negativamente a la venta, esto se puede respaldar con la siguiente hip3tesis basada en el **Gr3fico 8**: Ventas totales 2013 y 2014 por zona, dado que el centro es la 3nica zona que aumenta sus ventas el a3o 2014 con respecto al a3o

2013. A la vez, se observa que el coeficiente de Lima es positivo y esto se debe a que el tamaño (en ventas) del sector duplica las ventas del Centro, además de presentar una disminución menor porcentualmente respecto de las otras (0.4% versus 2.6% y 0.8%), lo que explica el efecto positivo.

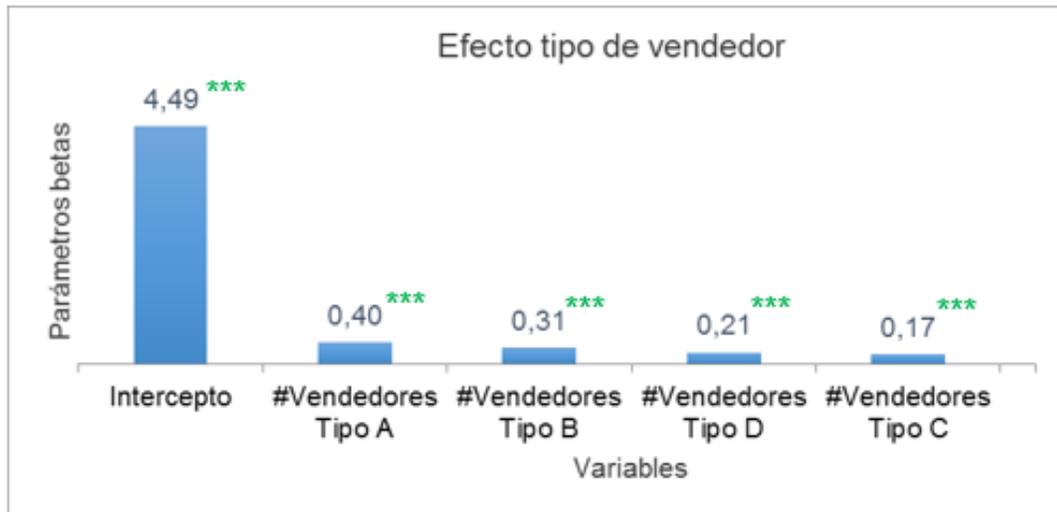
Complementando lo anterior, el coeficiente de la densidad poblacional a pesar de ser negativo, no se considera tan relevante puesto que su valor es cercano a 0 y mucho menor que los coeficientes de las otras variables consideradas, puesto que intuitivamente es contradictorio que mientras más personas por km^2 (superficie de la región), influya negativamente en las ventas.

7.2.2.1.1.3 Efecto tipo de vendedor

De acuerdo a la clasificación anterior del tipo de vendedor y considerando que los vendedores pueden atender a más de un cliente y un cliente puede ser atendido por más de un vendedor, se agregó el número de vendedores del tipo A, B, C y D que tiene cada cliente. La ecuación de la regresión lineal es:

$$Ventas_{2014,i} = \beta_0 + \beta_1 Tipo A_i + \beta_2 Tipo B_i + \beta_3 Tipo C_i + \beta_4 Tipo D_i + \varepsilon_i$$

Las variables de las regresiones lineales del efecto del tipo de vendedor tienen los mismos betas en ambos resultados al aplicar el algoritmo *Stepwise*, puesto que en la regresión simple (sin utilizar *Stepwise*) todos los tipos de vendedores son significativos como se muestra en el **Gráfico 12** (ver Anexo 12).



Códigos de significancia: 0 '***' 0,001 '**' 0,01 '*' 0,05 '.' 0,1 '.' 1

Gráfico 12: Modelo (I) – Efecto tipo de vendedor: comparación pesos de los parámetros betas en la regresión lineal 2

Fuente: Elaboración propia en base a información de la empresa

A pesar de tener un alto valor el intercepto, se puede notar que en general la magnitud del beta relacionado al tipo de vendedor es decreciente, es decir, el beta del tipo de vendedor A es el mayor y luego sigue el beta del tipo B, pero genera ruido al comparar los betas del tipo C y D puesto que intuitivamente deberían ser decrecientes las ventas mientras disminuye el nivel del vendedor.

7.2.2.1.2 Resultados regresión lineal de los 3 efectos

A pesar de tener un análisis individual de los efectos, se desean ver los efectos en conjunto de forma tal de evaluar cómo se comportan de manera integrada en la venta del año 2014. De este modo, la ecuación del Modelo (I) de regresión lineal de los 3 efectos es:

$$Ventas_{2014} = \beta_0 + \beta_1 Cliente + \beta_2 Contexto + \beta_3 TipoDeVendedor + \varepsilon_t$$

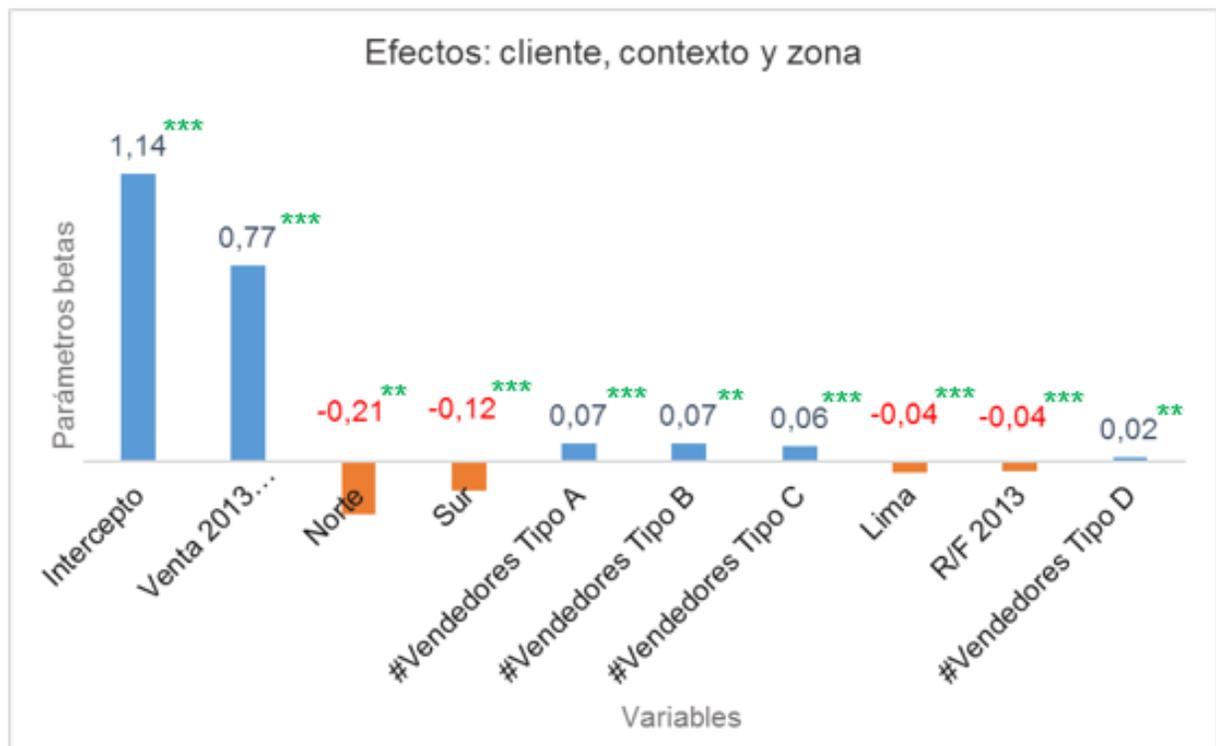
La cual se puede reescribir con todas las variables importantes detalladas como:

$$\begin{aligned}
 Ventas_{2014,i} = & \beta_0 + \beta_1 Ventas_{2013,i} + \beta_2 \cdot R/F_{2013,i} \\
 & + \beta_3 \text{Número de categorías distintas compradas}_{2013,i} + \beta_4 \text{Zona} \\
 & + \beta_5 \text{Densidad} + \beta_6 \text{Tipo } A_i + \beta_7 \text{Tipo } B_i + \beta_8 \text{Tipo } C_i + \beta_9 \text{Tipo } D_i + \varepsilon_i
 \end{aligned}$$

Donde *Tipo A, B, C y D*, es la cuenta del número de vendedores del tipo A, B C y D que tiene cada cliente, respectivamente.

7.2.2.1.3 Resultados regresión lineal de los 3 efectos

Al realizar el mismo procedimiento anterior para los 3 efectos, los resultados de los betas de las variables del modelo se presentan en el **Gráfico 13** y en la **Tabla 12**:



Códigos de significancia: 0 '***' 0,001 '**' 0,01 '*' 0,05 '.' 0,1 '' 1

Gráfico 13: Modelo (I) – 3 Efectos: Comparación de parámetros betas de la regresión lineal 2

Fuente: Elaboración propia en base a información de la empresa

| Variables | Coefficientes | Error std. | t valor | Pr(> t) | Signif. |
|-----------------------|---------------|------------|---------|-----------|---------|
| Intercepto | 1,138 | 0,04 | 25,96 | < 2e-16 | *** |
| Venta 2013(Miles S/.) | 0,774 | 4,11,E-03 | 188,50 | < 2e-16 | *** |
| R/F 2013 | -0,038 | 0,04 | -5,40 | 6,84,E-08 | *** |
| Norte | -0,209 | 0,04 | -3,22 | 1,27,E-03 | ** |
| Sur | -0,117 | 0,02 | 4,30 | 1,71,E-05 | *** |
| Lima | -0,044 | 0,02 | 4,08 | 4,57,E-05 | *** |
| #Vendedores Tipo A | 0,068 | 0,02 | 3,62 | 2,99,E-04 | *** |
| #Vendedores Tipo B | 0,068 | 0,03 | -1,27 | 1,37,E-03 | ** |
| #Vendedores Tipo C | 0,059 | 0,01 | -4,00 | 6,42,E-05 | *** |
| #Vendedores Tipo D | 0,016 | 0,01 | 1,16 | 1,58,E-03 | ** |

Códigos de significancia: 0 '***' 0,001 '**' 0,01 '*' 0,05 '.' 0,1 ' ' 1

Tabla 12: Modelo (I) – 3 Efectos: Resultados regresión lineal 2

Fuente: Elaboración propia en base a información de la empresa

En primer lugar, se destaca que se eliminan las variables número de distintas categorías que compra cada cliente y la densidad poblacional. Al ver el efecto del cliente en la regresión, se obtienen las mismas conclusiones que el análisis individual del efecto, donde el beta de la venta 2013 es el más importante después del intercepto, y el beta del R/F sigue siendo negativo lo cual es intuitivamente correcto. Incluso, se mantiene el análisis de las zonas, ya que como la zona centro es la dummy base, Lima, Norte y Sur afectan negativamente a la venta en conjunto, también basado en la información del **Gráfico 8:** Ventas totales 2013 y 2014 por zona. Al incluir las variables del efecto del tipo de vendedor en conjunto con el resto de las variables, en este caso, sus betas si son decrecientes en magnitud cuando disminuye la calidad del vendedor.

A continuación, en la **Tabla 13** se comparan los principales estadísticos de las regresiones lineales presentadas anteriormente:

| Estadísticos | Cliente | Contexto | Tipo de vendedor | 3 efectos |
|-------------------------|-----------|-----------|------------------|-----------|
| R2 | 0,64 | 0,01 | 0,20 | 0,64 |
| R2 ajustado | 0,64 | 0,01 | 0,02 | 0,64 |
| R2: grupo control | 0,59 | 0,02 | 0,15 | 0,62 |
| F-estadístico | 12.330,00 | 53,58 | 107,90 | 4.138,00 |
| P-valor | < 2,2e-16 | < 2,2E-16 | <2,2E-16 | <2,2E-16 |
| Error residual estándar | 1,71 | 2,83 | 2,82 | 1,71 |

Tabla 13: Estadísticos principales de las regresiones lineales de los efectos

Fuente: Elaboración propia en base a información de la empresa

El coeficiente de determinación (R^2) en el efecto contexto y tipo de vendedor son muy bajos, pero los valores del efecto cliente y de los 3 efectos en conjunto son parecidos por lo que puede deberse a que la variable 2013, presente en ambas regresiones, tiene un alto poder explicativo de las ventas del 2014. El R^2 ajustado, que toma en cuenta la cantidad de datos y variables explicativas en las regresiones, en todas ellas es parecido a su R^2 , lo que implica que es suficiente con mirar el R^2 múltiple. Además, dado que se trabajan los modelos con grupo de tratamiento para estimar los betas y con grupo control para analizar la capacidad predictiva del modelo, se comparan los R^2 de ambos grupos, los cuales son similares entre ellos.

Todos los p-valores son pequeños por lo que se rechaza la hipótesis nula (todos los coeficientes son 0) y agregando el estadístico F, se complementa que el modelo de 3 efectos es significativo.

Complementando lo anterior, al calcular aquellas métricas de ajuste es importante mirar tanto la capacidad de explicar bien la data que se observa como la capacidad de pronóstico de los modelos, por lo que en la **Tabla 14** se muestran más estadísticos del modelo de regresión lineal que incluye los 3 efectos.

| Estadísticos | 3 efectos |
|--------------|------------|
| Skewness | 0,37 |
| AIC | 81.808,41 |
| BIC | 81.895,84 |
| LL | -40.893,00 |
| MAPE | 28,72% |

Tabla 14: Estadísticos del modelo de regresión lineal de los 3 efectos en conjunto
Fuente: Elaboración propia en base a información de la empresa

Se destaca favorablemente, el valor del MAPE como indicador del porcentaje de error de la media absoluta que mide la precisión de la predicción de un método de pronóstico, es decir, el modelo se equivoca porcentualmente en un 28,72%. También es importante que, el valor de *skewness* (falta de simetría) es positivo lo que indica que los residuos de la regresión están inclinados hacia el lado positivo, con respecto al signo y valor de los residuos.

7.2.2.1.4 Cálculo de potencial de compra por zona

De acuerdo a lo anterior, el modelo con mejor rendimiento y poder explicativo es el modelo de regresión lineal que considera los 3 efectos en conjunto, por lo que, se procede a realizar el cálculo del potencial de compra de los clientes.

Dado que el modelo de los 3 efectos incluye las zonas, se realiza un cálculo en torno a esa variable en la siguiente **Tabla 15**, donde se puede ver que las ventas 2014 comparadas con la venta potencial aumentan proporcionalmente, en torno a un 11,33%.

| Modelo (I): 3 efectos | Total | Zonas | | | |
|--------------------------------------|-------------------|-----------|-----------|-----------|-----------|
| | | Lima | Sur | Norte | Centro |
| Venta 2014 (miles S/.) | 140.812,66 | 47.979,99 | 38.385,37 | 27.116,72 | 27.330,59 |
| Número de clientes | 26.156 | 8.378 | 7.366 | 5.349 | 5.063 |
| Venta potencial (miles S/.) | 156.760,73 | 53.019,12 | 42.769,16 | 30.541,82 | 30.430,64 |
| # clientes potenciales | 11.786 | 3.813 | 3.283 | 2.424 | 2.266 |
| Diferencial de potencial (miles S/.) | 16.143,41 | 5.087,39 | 4.497,53 | 3.456,79 | 3.101,70 |
| Venta promedio (miles S/.) | 5,38 | 5,73 | 5,21 | 5,07 | 5,40 |
| Venta potencial promedio (miles S/.) | 5,99 | 6,33 | 5,81 | 5,71 | 6,01 |
| Aumento en ventas (%) | 11,33% | 10,50% | 11,42% | 12,63% | 11,34% |

Tabla 15: Estimación del potencial de compra por zonas basado en el modelo de regresión lineal de los 3 efectos en conjunto

Fuente: Elaboración propia en base a información de la empresa

Los clientes potenciales representan un 45% del total de clientes en cada zona y total, lo que implica que casi la mitad de los clientes no están comprando su venta potencial individual y, si se consideran a todos los clientes, el diferencial de potencial asciende a más de 16 millones de soles, lo cual excede el 11% de las ventas totales reales.

7.2.2.2 Modelo (II): Categorías principales

El Modelo (II) se define como la estimación de la compra de una de las categorías más importantes en función al resto de las categorías y, para ello, se utiliza el mismo algoritmo a las regresiones anteriores, *Stepwise*.

7.2.2.2.1 Regresión lineal categoría aceites domésticos

La ecuación de la regresión lineal de las ventas del proveedor en la categoría aceites domésticos queda definida como:

$$\begin{aligned} \text{LogAceites}_i &= \beta_0 \\ &+ \sum_i \beta_i (\text{LogAderezos}_i + \text{LogAyudasCulinarias}_i + \text{LogCaramelos}_i \\ &+ \text{LogChocolates}_i + \text{LogDetergentes}_i + \text{LogGalletas}_i + \text{LogHarinas}_i \\ &+ \text{LogJabones}_i + \text{LogMargarinas}_i + \text{LogPanetones}_i + \text{LogPastas}_i \\ &+ \text{LogPostres}_i + \text{LogRefrescos}_i + \text{LogSalsas}_i) + \varepsilon_i \end{aligned}$$

7.2.2.2.2 Regresión lineal categoría detergentes

La ecuación de la regresión lineal de las ventas del proveedor en la categoría detergentes queda definida como:

$$\begin{aligned} \text{LogDetergentes}_i &= \beta_0 \\ &+ \sum_i \beta_i (\text{LogAceites}_i + \text{LogAderezos}_i + \text{LogAyudasCulinarias}_i \\ &+ \text{LogCaramelos}_i + \text{LogChocolates}_i + \text{LogGalletas}_i + \text{LogHarinas}_i \\ &+ \text{LogJabones}_i + \text{LogMargarinas}_i + \text{LogPanetones}_i + \text{LogPastas}_i \\ &+ \text{LogPostres}_i + \text{LogRefrescos}_i + \text{LogSalsas}_i) + \varepsilon_i \end{aligned}$$

7.2.2.2.3 Regresión lineal categoría galletas

La ecuación de la regresión lineal de las ventas del proveedor en la categoría galletas queda definida como:

$$\begin{aligned} \text{LogGalletas}_i &= \beta_0 \\ &+ \sum_i \beta_i (\text{LogAceites}_i + \text{LogAderezos}_i + \text{LogAyudasCulinarias}_i \\ &+ \text{LogCaramelos}_i + \text{LogChocolates}_i + \text{LogDetergentes}_i + \text{LogHarinas}_i \\ &+ \text{LogJabones}_i + \text{LogMargarinas}_i + \text{LogPanetones}_i + \text{LogPastas}_i \\ &+ \text{LogPostres}_i + \text{LogRefrescos}_i + \text{LogSalsas}_i) + \varepsilon_i \end{aligned}$$

7.2.2.2.4 Regresión lineal categoría pastas

La ecuación de la regresión lineal de las ventas del proveedor en la categoría pastas queda definida como:

$$\begin{aligned} \text{LogPastas}_i = & \beta_0 \\ & + \sum_i \beta_i (\text{LogAceites}_i + \text{LogAderezos}_i + \text{LogAyudasCulinarias}_i \\ & + \text{LogCaramelos}_i + \text{LogChocolates}_i + \text{LogDetergentes}_i + \text{LogGalletas}_i \\ & + \text{LogHarinas}_i + \text{LogJabones}_i + \text{LogMargarinas}_i + \text{LogPanetones}_i \\ & + \text{LogPostres}_i + \text{LogRefrescos}_i + \text{LogSalsas}_i) + \varepsilon_i \end{aligned}$$

7.2.2.2.5 Regresión lineal categoría salsas

La ecuación de la regresión lineal de las ventas del proveedor en la categoría salsas queda definida como:

$$\begin{aligned} \text{LogSalsas}_i = & \beta_0 \\ & + \sum_i \beta_i (\text{LogAceites}_i + \text{LogAderezos}_i + \text{LogAyudasCulinarias}_i \\ & + \text{LogCaramelos}_i + \text{LogChocolates}_i + \text{LogDetergentes}_i + \text{LogGalletas}_i \\ & + \text{LogHarinas}_i + \text{LogJabones}_i + \text{LogMargarinas}_i + \text{LogPanetones}_i \\ & + \text{LogPastas}_i + \text{LogPostres}_i + \text{LogRefrescos}_i) + \varepsilon_i \end{aligned}$$

7.2.2.2.6 Resultados regresiones lineales de cada categoría

Los principales resultados de las regresiones lineales de cada categoría, utilizando el algoritmo *Stepwise*, se resumen en las siguientes tablas. Los resultados de los betas se detallan en el Anexo 13, y en particular, la **Tabla 16**, muestra el resumen de los betas significativos de las regresiones de cada categoría (sin incluir el intercepto, el color verde destaca los betas con mayor magnitud, color rojo los de menor magnitud y las escalas de colores amarillo/verde claro representa los valores intermedios).

| Variables | Aceites | Deterg. | Galletas | Pastas | Salsas |
|----------------------------|-----------|-----------|----------|-----------|-----------|
| Intercepto | 1,88 *** | 2,30 *** | 4,55 *** | 1,12 *** | 0,75 *** |
| Log Aceites | - | 0,19 *** | -0,02 * | 0,43 *** | 0,08 *** |
| Log Aderezos | 0,01 * | -0,01 *** | 0,10 *** | -0,02 *** | 0,04 *** |
| Log Ayudas culinarias | -0,03 *** | 0,03 *** | 0,01 ** | 0,02 *** | 0,10 *** |
| Log Caramelos | -0,02 *** | - | 0,14 *** | - | - |
| Log Chocolates | 0,01 * | - | 0,05 *** | -0,02 *** | - |
| Log Detergentes | 0,23 *** | - | -0,03 ** | 0,12 *** | 0,19 *** |
| Log Galletas | -0,01 * | -0,01 ** | - | - | 0,08 *** |
| Log Harinas domésticas | 0,06 *** | 0,02 *** | 0,02 ** | 0,18 *** | 0,10 *** |
| Log Jabones de lavar | 0,13 *** | 0,35 *** | 0,07 *** | 0,06 *** | -0,05 * |
| Log Margarinas domésticas | 0,12 *** | 0,10 *** | 0,06 *** | -0,05 *** | 0,14 *** |
| Log Panetones | - | 0,01 *** | 0,02 *** | 0,02 *** | -0,03 *** |
| Log Pastas | 0,23 *** | 0,05 *** | - | - | 0,09 *** |
| Log Postres instantáneos | 0,08 *** | 0,05 *** | 0,05 *** | 0,09 *** | - |
| Log Refrescos instantáneos | 0,02 *** | 0,06 *** | 0,08 *** | 0,04 *** | 0,04 ** |
| Log Salsas | 0,01 *** | 0,02 *** | 0,01 *** | 0,02 *** | - |

Códigos de significancia: 0 '***' 0,001 '**' 0,01 '*' 0,05 '.' 0,1 ' ' 1

Tabla 16: Parámetros betas estimados por categoría de las regresiones lineales
Fuente: Elaboración propia en base a información de la empresa

Se puede inferir, de los resultados de la regresión lineal de aceites que los betas con mayor peso son las categorías de detergentes y pastas, estas pertenecen a las categorías principales del proveedor, lo cual se debe a que, junto con aceites, son las 3 categorías más compradas por los clientes.

De los betas de la regresión lineal de detergentes, el de la categoría jabones de lavar tiene un mayor peso, lo cual se debe a que ambos son productos de limpieza, por lo tanto, hace sentido que correspondan a una misma ocasión de compra, lo que explica la correlación que genera el valor. Además, aceites también tiene un peso significativo puesto que es la categoría más comprada por los clientes.

En la categoría galletas, el beta de caramelos es el con mayor impacto, lo cual se puede deber a que ambas son “golosinas” del tipo snacks y las categorías que menos peso tienen son aceites, ayudas culinarias, harinas, panetones y salsas que no son compras de tipo “al paso” para consumir en el momento.

Los resultados de la regresión lineal de pastas muestran que el beta con mayor magnitud es el de la categoría de aceites, lo cual se puede deber a que son categorías para el hogar que al utilizarlas se necesitan en conjunto. Después de aceites, los betas con mayor magnitud son harinas y detergentes que se pueden complementar con pastas y aceites como una canasta del hogar, pues son productos básicos de uso.

En la regresión lineal de salsas, el beta con mayor magnitud en valor es detergentes, no son categorías que se complementen pero pueden tener relación con la frecuencia de compra dado que hay distintos formatos de salsas y, esto puede ser visto como una oportunidad de vender productos “unidos”, tipo estrategia de *bundling*.

En la **Tabla 17**, se muestran los principales estadísticos de las regresiones lineales de cada categoría:

| Estadístico | Aceites | Detergentes | Galletas | Pastas | Salsas |
|---------------------|-----------|-------------|-----------|-----------|-----------|
| R2 Múltiple | 0,52 | 0,54 | 0,20 | 0,39 | 0,07 |
| R2 Ajustado | 0,52 | 0,54 | 0,20 | 0,39 | 0,07 |
| R2 grupo control | 0,47 | 0,52 | 0,21 | 0,37 | 0,07 |
| Error std. Residual | 0,70 | 0,64 | 0,85 | 0,97 | 2,30 |
| Grados de libertad | 20.911 | 20.912 | 20.911 | 20.912 | 20.913 |
| F-Estadístico | 1.714,00 | 2.017,00 | 405,00 | 1.108,00 | 143,30 |
| p-valor | < 2,2e-16 | <2,2e-16 | <2,2e-16 | <2,2e-16 | <2,2e-16 |
| Skewness | -1,96 | -1,19 | -1,35 | -2,30 | -1,51 |
| AIC | 44.638,31 | 40.753,08 | 52.492,04 | 57.934,23 | 94.186,75 |
| BIC | 44.757,54 | 40.864,36 | 52.611,27 | 58.045,51 | 94.282,14 |
| LL | -22.304 | -20.363 | -26.231 | -28.953 | -47.081 |
| MAPE | 0,50 | 0,57 | 0,94 | 1,49 | 0,54 |

Tabla 17: Estadísticos de las regresiones lineales de cada categoría
Fuente: Elaboración propia en base a información de la empresa

El coeficiente de determinación (R^2) en salsas y galletas son muy bajos por lo que las variables, lo cual significa que es poca la variabilidad de la variable dependiente que es explicada por la presencia de las variables de predicción en la ecuación del modelo y puede se puede entender como que las variables son independientes. El R^2 ajustado, que toma en cuenta la cantidad de datos y variables explicativas en las regresiones, en todas ellas es parecido a su R^2 , lo que implica que es suficiente con mirar el R^2 múltiple. Además, dado que se trabajan los modelos con grupo de tratamiento para estimar los betas y con grupo control para analizar la capacidad predictiva del modelo, se comparan los R^2 de ambos grupos, los cuales son similares entre ellos.

Todos los p-valores son pequeños por lo que se rechaza la hipótesis nula (todos los coeficientes son 0) y agregando el estadístico F, se complementa que el modelo de 3 efectos es significativo.

Complementando lo anterior, al calcular aquellas métricas de ajuste es importante mirar tanto la capacidad de explicar bien la data que se observa como la capacidad de pronóstico de los modelos, por lo que, se destaca desfavorablemente, el valor del MAPE como indicador del porcentaje de error de la media absoluta que mide la precisión de la predicción de un método de pronóstico, es decir, los modelos de aceites, detergentes y salsas se equivocan casi en un 54% en promedio, galletas en un 94% y pastas en más de 149%, lo cual es un valor muy alto.

También es importante que, el valor de *skewness* (falta de simetría) es negativo en todas las categorías (aceites, detergentes, galletas, pastas y salsas) lo que indica que los residuos de la regresión están inclinados hacia el lado negativo, con respecto al signo y valor de los residuos.

7.2.2.2.7 Cálculo de potencial de compra en categorías principales

A continuación, en la **Tabla 18**, se calcula el potencial de compra de los modelos de cada categoría, con lo cual se podrían aumentar las ventas en un 12,09% en promedio. De todas las categorías, el modelo de salsas es donde hay menos margen de aumento en ventas, lo cual se puede deber a que es la categoría con menos ventas en comparación a aceites, detergentes, galletas y pastas. Se observa que el número de clientes potenciales es cercano a la mitad del total de clientes dado que en el cálculo del potencial de compra ha sido utilizado los resultados de un modelo de regresión lineal.

| Categorías | Aceites | Deterg. | Galletas | Pastas | Salsas | Total |
|---|---------|---------|----------|--------|--------|---------------|
| Venta 2013 (miles S/.) | 28.229 | 26.368 | 14.281 | 21.100 | 9.718 | 99.695 |
| Número de clientes | 25.906 | 25.977 | 26.020 | 25.773 | 21.330 | 125.006 |
| Venta potencial (miles S/.) | 31.754 | 29.711 | 16.217 | 23.695 | 10.374 | 111.751 |
| # clientes con diferencial de potencial | 12.717 | 12.175 | 12.239 | 11.054 | 5.740 | 53.925 |
| Diferencial de potencial (miles S/.) | 3.525 | 3.344 | 1.936 | 2.596 | 656 | 12.056 |
| Venta promedio (miles S/.) | 1,09 | 1,02 | 0,55 | 0,82 | 0,46 | 3,93 |
| Venta potencial promedio (miles S/.) | 1,09 | 1,02 | 0,55 | 0,82 | 0,46 | 3,93 |
| Aumento en ventas (%) | 12,49% | 12,68% | 13,55% | 12,30% | 6,75% | 12,09% |

Tabla 18: Estimación potencial de compra según categorías
Fuente: Elaboración propia en base a información de la empresa

7.2.3 Modelo de Regresión de Fronteras

Para desarrollar un análisis del modelo de regresión de fronteras estocástico, el *software* que se utilizará es RStudio, a partir del paquete *frontier*.

Dado que el modelo de regresión de fronteras estocásticas es un método paramétrico, se elige que la función determinada que seguirán las ventas de las categorías, será una función de Cobb Douglas para tratar de ver como los valores de producción se acercan a esa función que determina la frontera eficiente, ya que la función de producción representa la máxima producción alcanzable de un producto a partir de los inputs utilizados en la producción (ventas del resto de las categorías). Como la estimación econométrica requiere linealidad para estimar los parámetros beta, normalmente las funciones se linealizan aplicando logaritmos en base a Cobb Douglas.

7.2.3.1 Modelo (I): Efectos de las ventas 2014

Considerando los resultados del modelo previo de regresión lineal de los 3 efectos, se toman las mismas variables para realizar el modelo de regresión de fronteras.

Los resultados de la regresión de fronteras estocásticas son inconclusos puesto que el valor de *skewness* de la regresión lineal es positivo, por lo tanto, no es probable que los resultados reflejen una buena estimación y significa que puede haber clientes que no sean ineficientes o que el modelo está mal especificado para este tipo de regresión de fronteras estocástica como se puede ver en el Anexo 14.

El parámetro *gamma*, que mide la proporción de la varianza explicada por las ineficiencias de producción, está cerca del límite del espacio de parámetros [0,1] lo que puede causar problemas de convergencia; esto implica que, puede afectar negativamente a la validez y la fiabilidad de las pruebas estadísticas. Además, la proporción de la varianza total que se debe a la ineficiencia es 2,65e-05, lo cual es un valor casi nulo y toda la varianza del error se debería a efectos aleatorios del modelo.

7.2.3.2 Modelo (II): Categorías principales

De acuerdo a los resultados de la **Tabla 17**: Estadísticos de las regresiones lineales de cada categoría, el valor del *skewness* es negativo para todas las categorías, lo que significa que los residuos están desviados y se concentran hacia el lado izquierdo de su distribución, lo que implica que es probable que no todos los clientes son completamente

eficientes técnicamente, en otras palabras, se concluye que existe potencial de compra de estos clientes que no está siendo explotado por la empresa. Por lo tanto, tiene sentido realizar un análisis de fronteras estocásticas para ver la ineficiencia de los clientes con el fin de calcular el potencial de compra por categoría.

Entonces, tomando como base las variables de todas las categorías, se realizaron distintas combinaciones de ellas con diferente número de variables a incluir, con el fin de minimizar la cantidad de variables agregadas en los modelos de regresión de fronteras estocásticas para que la interpretación del modelo sea más simple. Para comenzar, se agregan variables al modelo en orden de venta total decreciente hasta que la variable que sea agregada no sea significativa. Luego, se comparan los diferentes modelos con estadísticos como log-verosimilitud (LL), AIC y BIC.

A modo de ejemplo, para el caso de la categoría de aceites, se desarrollaron varios modelos de fronteras estocásticas con distinto número de variables. Los modelos que tienen mayor significancia en las variables de acuerdo a la importancia en ventas totales de categorías son el modelo con 6, 7, 8 y 13 categorías y las variables incluidas en cada uno están en la **Tabla 19**:

| # | 6 categorías | 7 categorías | 8 categorías | 13 categorías |
|----|---------------|---------------|---------------|-------------------|
| 1 | Pastas | Pastas | Pastas | Pastas |
| 2 | Deterg. | Deterg. | Deterg. | Deterg. |
| 3 | Jabones | Jabones | Jabones | Jabones |
| 4 | Margarinas | Margarinas | Margarinas | Margarinas |
| 5 | Postres inst. | Postres inst. | Postres inst. | Postres inst. |
| 6 | Harinas | Harinas | Harinas | Harinas |
| 7 | | Galletas | Galletas | Galletas |
| 8 | | | Salsas | Salsas |
| 9 | | | | Ayudas culinarias |
| 10 | | | | Refrescos |
| 11 | | | | Caramelos |
| 12 | | | | Aderezos |
| 13 | | | | Chocolates |

Tabla 19: Categoría: Aceites – Variables incluidas en los modelos de regresión de fronteras

Fuente: Elaboración propia en base a información de la empresa

Para seleccionar el mejor modelo, se compararon los estadísticos AIC, BIC y LL, presentados en el **Gráfico 14**, donde se puede ver que el modelo que minimiza el AIC y BIC y maximiza la log-verosimilitud es el modelo con 13 variables.

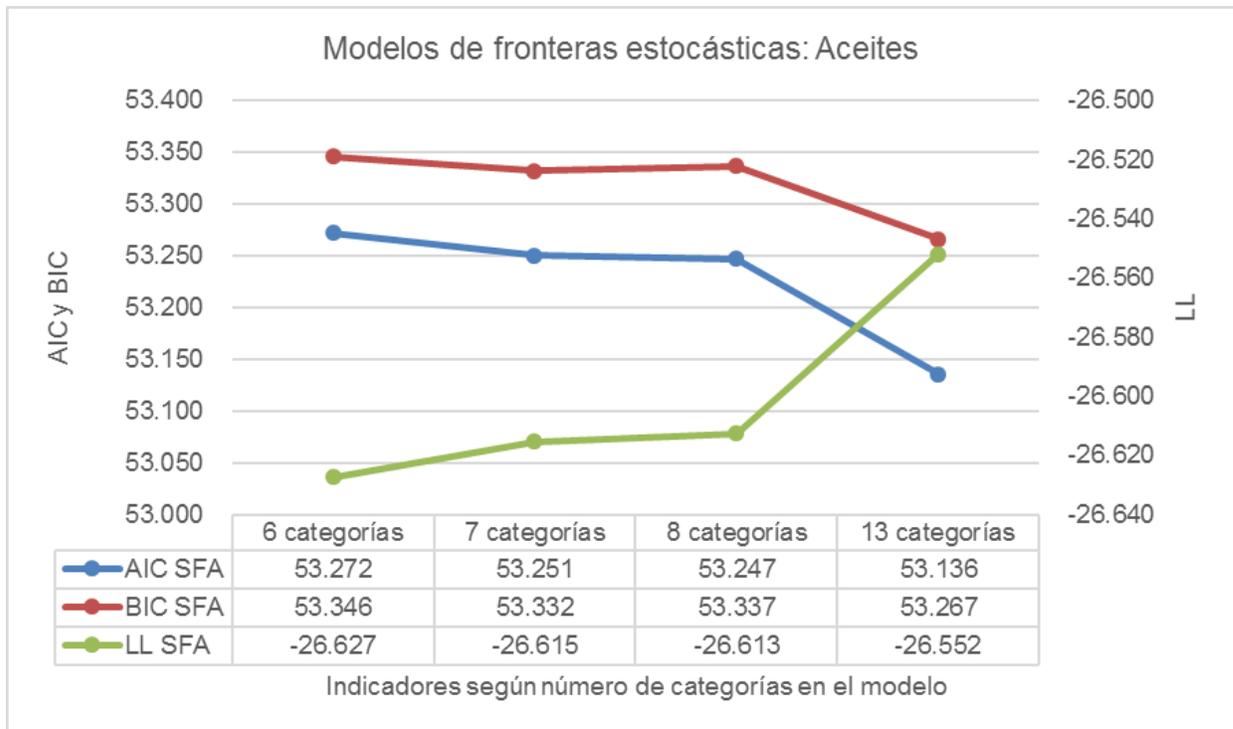


Gráfico 14: Categoría: Aceites – Estadísticos principales de los modelos de regresión de fronteras según número de variables incluidas

Fuente: Elaboración propia en base a información de la empresa

En el Anexo 15, están los análisis de las categorías detergentes, galletas y pastas para seleccionar el mejor modelo para cada una y estos, se resumen en la **Tabla 20** (color verde representa los valores absolutos más altos de los parámetros betas, mientras que el color rojo, los menores) y en el Anexo 16, están los valores del error que se deben a la ineficiencia estocástica en cada categoría del modelo de fronteras.

| Coeficientes | Aceites | Deterg. | Galletas | Pastas |
|----------------------------|-------------|----------|-----------|-----------|
| Intercepto | 3,45 *** | 3,58 *** | 6,14 *** | 2,95 *** |
| Log Aceites | - | 0,13 *** | -0,06 *** | 0,34 *** |
| Log Aderezos | 0,01 *** | - | 0,09 *** | -0,02 *** |
| Log Ayudas Culinarias | -0,02 *** | 0,03 *** | 0,03 *** | 0,03 *** |
| Log Caramelos | -0,02 *** | - | 0,13 *** | - |
| Log Chocolates | 0,01 *** | - | 0,04 *** | -0,02 *** |
| Log Detergentes | 0,19 *** | - | -0,03 *** | 0,15 *** |
| Log Galletas | -0,02 *** | -0,01 * | - | - |
| Log Harinas Domésticas | 0,06 *** | 0,02 *** | - | 0,14 *** |
| Log Jabones de Lavar | 0,14 *** | 0,32 *** | 0,08 *** | 0,06 *** |
| Log Margarinas Domésticas | 0,06 *** | 0,07 *** | 0,04 *** | -0,09 *** |
| Log Panetones | - | 0,01 *** | 0,01 *** | 0,01 *** |
| Log Pastas | 0,18 *** | 0,05 *** | - | - |
| Log Postres Instantáneos | 0,06 *** | 0,04 *** | 0,03 *** | 0,08 *** |
| Log Refrescos Instantáneos | 0,02 *** | 0,05 *** | 0,05 *** | 0,04 *** |
| Log Salsas | 4,28,E-03 * | 0,01 *** | 0,01 *** | 0,01 *** |

Códigos de significancia: 0 '***' 0,001 '**' 0,01 '*' 0,05 '.' 0,1 ' ' 1

Tabla 20: Parámetros betas estimados por categoría del modelo de fronteras estocásticas

Fuente: Elaboración propia en base a información de la empresa

Los resultados de los betas estimados para cada categoría se interpretan como la elasticidad de la variable respectivamente; por lo tanto, si aumenta/disminuye un beta en una unidad la variable dependiente también lo hace en la misma medida. De este modo, la función de producción estimada aumenta/disminuye monótonamente con todos los inputs, respectivamente.

Se pueden obtener análisis parecidos a los resultados del Modelo (II) de regresiones lineales de las categorías principales, comparando a nivel de magnitudes y signos de los betas, dado que los coeficientes son similares.

Con los modelos estocásticos que se desarrollaron en aceites, detergentes, galletas y pastas, en la **Tabla 21**, se muestran los principales indicadores de la distribución de la eficiencia técnica de los clientes en cada modelo.

| Eficiencia Técnica | Aceites | Detergent | Galletas | Pastas |
|--------------------|-----------|-----------|-----------|-----------|
| Min. | 1,89,E-03 | 3,03,E-03 | 2,74,E-03 | 8,80,E-04 |
| 1er Q | 0,49 | 0,51 | 0,37 | 0,36 |
| Mediana | 0,59 | 0,63 | 0,52 | 0,50 |
| Promedio | 0,57 | 0,60 | 0,50 | 0,48 |
| 3er Q | 0,68 | 0,71 | 0,64 | 0,62 |
| Max. | 0,95 | 0,96 | 0,93 | 0,96 |

Tabla 21: Eficiencia técnica de cada categoría del modelo de fronteras estocásticos
Fuente: Elaboración propia en base a información de la empresa

Por lo tanto, al observar el valor máximo de la eficiencia técnica de cada categoría, se puede concluir que ninguna alcanza a ser 100% técnicamente eficiente, ya que todos los valores son en torno al 95%. Además, el mínimo de la eficiencia técnica es casi nulo para todas las categorías, y la mediana no alcanza a ser un valor tan alto, lo que implica que hay muchos clientes que son poco eficientes en su capacidad de compra, por lo tanto, hay espacio para mejorar en las categorías.

7.2.3.3 Cálculo de potencial de compra

A partir de las predicciones de los modelos de fronteras estocásticos, se calcula el potencial de compra de cada categoría en la **Tabla 22**, donde el principal análisis es que las ventas aumentan en un alto porcentaje (77,27% en general), dado que la venta potencial es superior a la real, incluso comparado con los otros resultados de cálculo de potencial basados en los modelos de regresión lineal. Lo anterior, indica que el modelo de regresión de fronteras apunta a que el cliente compre su capacidad total, es decir, la mejor situación en que el cliente podría vender, maximizando su eficiencia y, por ende, sus ventas.

| Categorías | Aceites | Deterg. | Galletas | Pastas | Total |
|---|---------|---------|----------|--------|---------|
| Venta 2013 (miles S/.) | 28.229 | 26.368 | 14.281 | 21.100 | 89.978 |
| Número de clientes | 25.906 | 25.977 | 26.020 | 25.773 | 103.676 |
| Venta potencial (miles S/.) | 47.900 | 42.856 | 27.085 | 41.662 | 159.504 |
| # clientes con diferencial de potencial | 23.490 | 23.228 | 23.367 | 24.278 | 94.363 |
| Diferencial de potencial (miles S/.) | 19.671 | 16.488 | 12.804 | 20.563 | 69.526 |
| Venta promedio (miles S/.) | 1,09 | 1,02 | 0,55 | 0,82 | 3,47 |
| Venta potencial promedio (miles S/.) | 1,85 | 1,60 | 1,56 | 1,00 | 6,01 |
| Aumento en ventas (%) | 69,68% | 62,53% | 89,66% | 97,45% | 77,27% |

Tabla 22: Cálculo de potencial de compra según modelos de regresión de fronteras de cada categoría

Fuente: Elaboración propia en base a información de la empresa

7.2.4 Resultados complementarios

En búsqueda de agregar las variables obtenidas a través de la encuesta, realizada por la empresa a bodegas de Lima, se realizaron modelos que incluyeron la superficie de la bodega (m^2) y si esta está en ubicada en una avenida principal o si pertenece a una zona residencial o comercial/industrial. Dado que los resultados no son concluyentes (ver Anexo 17), se recomienda continuar con esta investigación como trabajo futuro cuando el número de los encuestados sea mucho mayor al 5,9% actual y la muestra sea más representativa, pues son variables que, intuitivamente, influyen en las ventas del año 2014.

7.2.5 Elección del modelo de potencial de compra

El modelo de potencial de compra que sea seleccionado debe ser evaluado de acuerdo a criterios como simplicidad para interpretar e implementarlo y que esté adaptado a la realidad del proveedor peruano y de sus clientes, además de ver el ajuste de los datos, capacidad predictiva, robustez, número de variables incluidas en los modelos, entre otros.

Previo al análisis de los modelos (I) y (II), es clave tener en cuenta que no son comparables en cuanto a lo que miden y a como estiman o predicen, pues utilizan diferentes variables, tanto dependientes como independientes, y son modelos que buscan la mejor forma, de diferentes maneras, de estimar el potencial de compra de los clientes del proveedor peruano.

La **Tabla 23** presenta los estadísticos que permiten comparar los modelos de regresión lineal realizados de manera analítica. Al respecto se observa que el Modelo (I) presenta una mayor robustez que (II) de las categorías, ya que comparando el R^2 de (I), teniendo una menor cantidad de variables incluidas en su construcción, explica una mayor proporción de la varianza del modelo; esto es válido para las tres versiones del estadístico (múltiple, ajustado y grupo de control) debido a que tiene el mismo valor por modelo en todos los modelos.

| Estadísticos | Modelo I | Modelo II | | | | |
|---------------------|-----------|-----------|-------------|-----------|-----------|-----------|
| | 3 efectos | Aceites | Detergentes | Galletas | Pastas | Salsas |
| R2 Múltiple | 0,64 | 0,52 | 0,54 | 0,20 | 0,39 | 0,07 |
| R2 Adjustado | 0,64 | 0,52 | 0,54 | 0,20 | 0,39 | 0,07 |
| R2 grupo control | 0,62 | 0,47 | 0,52 | 0,21 | 0,37 | 0,07 |
| Error std. Residual | 1,71 | 0,70 | 0,64 | 0,85 | 0,97 | 2,30 |
| Grados de libertad | 20.915 | 20.911 | 20.912 | 20.911 | 20.912 | 20.913 |
| p-valor | <2,2e-16 | < 2,2e-16 | <2,2e-16 | <2,2e-16 | <2,2e-16 | <2,2e-16 |
| Skewness | 0,37 | -1,96 | -1,19 | -1,35 | -2,30 | -1,51 |
| AIC | 81.808,41 | 44.638,31 | 40.753,08 | 52.492,04 | 57.934,23 | 94.186,75 |
| BIC | 81.895,84 | 44.757,54 | 40.864,36 | 52.611,27 | 58.045,51 | 94.282,14 |
| LL | -40.893 | -22.304 | -20.363 | -26.231 | -28.953 | -47.081 |
| MAPE | 28,72% | 49,80% | 57,32% | 94,44% | 149,49% | 54,11% |

Tabla 23: Comparación de los modelos de regresión lineal
Fuente: Elaboración propia en base a información de la empresa

Por otro lado, se observa que el error estándar residual del Modelo (I) presenta, en general, un valor más elevado que los modelos de categorías, exceptuando el modelo de salsas. Debido a esto, los resultados entregados por este modelo son empíricamente menos acertados que los del Modelo (II), no obstante, el error estándar es pequeño en contraste con los valores que se introducen en la construcción del modelo, por lo que se puede considerar como un estadístico que entrega poca información relevante para decidir la calidad de los modelos por sí solo; de la misma manera se observa que los grados de libertad entre modelos son virtualmente iguales.

El análisis del estadístico *skewness* indica que la distribución de los residuos del Modelo (II) están inclinados hacia el lado izquierdo con respecto al cero, lo que significa que es probable que todos los clientes no sean eficientes, lo que se ve más claramente en los modelos estocásticos de frontera, mientras que el mismo análisis para el Modelo (I) indica que los residuos están centrados positivamente, lo cual no es concluyente para evaluar si son eficientes los clientes, pero se ha comprobado en las secciones anteriores de este trabajo que existen clientes con potencial de compra sin explotar.

Así mismo, se observa que el estadístico de Akaike y Bayesiano (AIC y BIC) resulta ser menor en el caso del Modelo (II), lo que indica que este último es un mejor modelo que el primero.

Finalmente, los estadísticos de log-verosimilitud y MAPE, en línea con la idea anterior, resultan favorables para el Modelo (II) respecto del (I), razón por la cual se concluye que el modelo de categorías es, analíticamente, un mejor modelo predictivo del potencial de

compra de los clientes, a pesar de tener un menor poder explicativo de la varianza del efecto estudiado.

Al comparar entre los modelos de categorías, por regresiones lineales y de fronteras estocásticas, vale la pena complejizar el modelo de acuerdo al ratio de LR, que se utiliza para ver si los parámetros adicionales o complejidad ayudan a mejorar el ajuste, pero esto no es lo único que se debe evaluar al seleccionar un modelo como se mencionó en un principio.

Con respecto a los modelos de frontera realizados, la **Tabla 24** resume los estadísticos principales que permiten analizarlos:

| Estadísticos | | Modelo II | | | |
|---------------------------|----------|------------|-------------|------------|------------|
| | | Aceites | Detergentes | Galletas | Pastas |
| AIC | | 53.135,77 | 49.477,24 | 63.396,44 | 67.187,17 |
| BIC | | 53.266,52 | 49.591,64 | 63.519,02 | 49.586,58 |
| LL | | -26.551,89 | -24.724,62 | -31.683,22 | -33.578,59 |
| Varianza por ineficiencia | | 0,60 | 0,57 | 0,66 | 0,72 |
| Skewness | | -1,96 | -1,19 | -1,35 | -2,30 |
| Eficiencia técnica | Mínimo | 0,002 | 0,003 | 0,003 | 0,001 |
| | Promedio | 0,574 | 0,597 | 0,500 | 0,478 |
| | Máximo | 0,955 | 0,955 | 0,931 | 0,957 |

Tabla 24: Comparación de los modelos de regresión de fronteras estocásticas
Fuente: Elaboración propia en base a información de la empresa

Se destaca el valor negativo del *skewness* para todos los modelos corroborando la existencia de potencial de compra de los clientes, específico por categoría de productos, lo cual representa que el modelo tiene margen para mejorar la eficiencia de las bodegas, pues podrían estar mejor de lo que actualmente están, a diferencia de lo que busca el modelo lineal.

A pesar de esto, se observa que los valores de AIC y BIC resultan ser mayores en todos los modelos respecto de su contraparte lineal, a la vez que la log-verosimilitud es menor. Ambos criterios muestran que los modelos lineales son mejores estimadores del comportamiento de los clientes que los modelos de frontera.

En general, los modelos lineales son mucho más fáciles de interpretar e implementar, aunque con menor poder de estimación del potencial de compra, puesto que la regresión

lineal tiende a subestimar la capacidad efectiva de compra del cliente, dejando poco espacio para crecer, en comparación a la regresión de fronteras que estima la mejor situación en que podría estar la bodega. Si bien el modelo de fronteras entrega un valor superior de aumento en las ventas al modelo lineal, este puede estar sobre estimando la capacidad de compra del cliente que tiene con el proveedor peruano, puesto que la base de datos está sesgada a las transacciones que tiene la empresa y no se incorporó la información de lo que las bodegas le pueden estar comprando a la competencia.

Pero la realidad es que la necesidad de la empresa, de obtener un modelo sencillo y útil, favorece finalmente al modelo lineal. Es importante aclarar que todos los modelos construidos cumplen con el objetivo del trabajo, pues permiten identificar nichos de acción, por medio de medidas tácticas o estratégicas, que permiten a la empresa explotar el potencial de compra que en la actualidad no está siendo aprovechado; desde este punto de vista los modelos de 3 efectos permiten tomar medidas sobre políticas de precio, localización y distribución de vendedores que no son posibles definir utilizando los modelos por categorías, mientras que estos últimos permiten el diseño de estrategias de marketing enfocado a productos con el objetivo de aprovechar el potencial de compra disponible.

Finalmente, aunque los modelos pueden ser utilizados de forma paralela y complementaria, se decide tomando en cuenta todos los hechos expuestos, elegir el modelo lineal de categorías como recomendación para la empresa, aunque el modelo de frontera de categorías entrega mayor información, pero con una mayor dificultad de implementación.

7.3 Ranking de clientes

Para clasificar a las bodegas, se asigna un score de forma tal de obtener una distribución de los clientes según la venta real y potencial calculada durante el año 2013 para la categoría de aceites, ya que representan el 23% aproximadamente de las ventas totales del proveedor peruano.

Así, los clientes fueron clasificados en 4 tipos de acuerdo la venta real y el potencial de compra de aceites, de acuerdo al siguiente criterio basado en el promedio de ventas real y promedio de ventas potencial de la categoría, mostrado en el **Gráfico 15**:

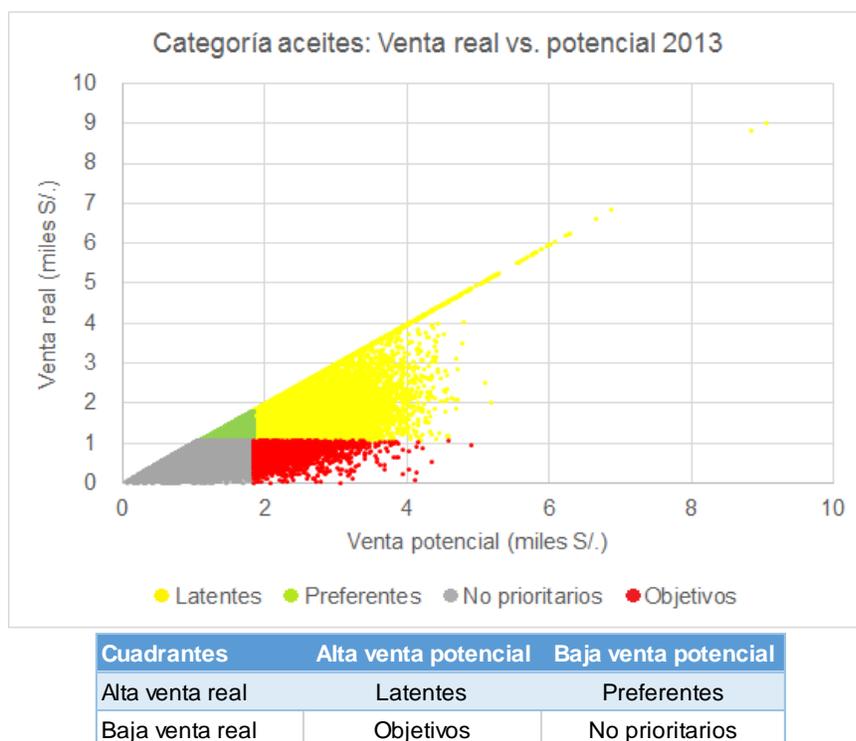


Gráfico 15: Categoría aceites – Venta real vs. Venta potencial 2013

Fuente: Elaboración propia en base a información de la empresa

Los cuadrantes del gráfico presentado han sido separados utilizando como criterio la siguiente meta, conseguir el potencial de aquellos clientes con elevado potencial de compra que en la actualidad están comprando la mitad o menos de la mitad de su potencial. Utilizando el promedio de venta real (S/. 1.079) y el promedio de venta potencial (S/. 1.831) en la categoría aceites, se cumple el criterio mencionado.

El cuadrante “Latentes” corresponde a aquellos clientes con un alto potencial de compra y elevado o moderado valor en ventas reales; por lo tanto, corresponde a clientes que pueden tener un margen de explotación que vale la pena considerar.

El cuadrante “Preferentes” enmarca a clientes que tienen bajo potencial de venta y a los que se les está vendiendo la mayor parte o todo su potencial, razón por la que no hay espacio de acción posible, salvo cuidarlos para mantener la relación con estos clientes.

Los clientes “No Prioritarios” corresponden a clientes con bajo nivel de potencial y bajo nivel de compras, en los que poner esfuerzo significa invertir recursos por un bajo nivel de retornos.

El cuadrante “Objetivos” corresponde a los clientes que están comprando a un nivel muy por debajo de su potencial disponible (mitad o menos). Son identificados como los clientes en los que se debe invertir pues tienen la mayor tasa de retorno.

El detalle de la venta real, potencial y estimada por el modelo de frontera, se resume en la **Tabla 25**:

| Categoría | Aceites | Latentes | Preferentes | No prioritarios | Objetivo | Total |
|---|------------|----------|-------------|-----------------|----------|--------|
| Clientes | Total | 8.484 | 1.958 | 12.461 | 3.253 | 26.156 |
| | Porcentaje | 32,44% | 7,49% | 47,64% | 12,44% | 100% |
| Venta real (miles S/.) | Total | 16.213 | 2.626 | 6.792 | 2.598 | 28.229 |
| | Porcentaje | 57,43% | 9,30% | 24,06% | 9,20% | 100% |
| Venta potencial (miles S/.) | Total | 23.125 | 3.019 | 14.492 | 7.264 | 47.900 |
| | Porcentaje | 48,28% | 6,30% | 30,25% | 15,16% | 100% |
| Venta estimada (miles S/.) | Total | 21.739 | 2.943 | 13.653 | 8.008 | 46.343 |
| | Porcentaje | 46,91% | 6,35% | 29,46% | 17,28% | 100% |
| Diferencial de potencial (miles S/.) | Total | 6.912 | 393 | 7.700 | 4.666 | 19.671 |
| | Porcentaje | 35,14% | 2,00% | 39,14% | 23,72% | 100% |

Tabla 25: Resumen de ventas de los cuadrantes de la categoría aceites

Fuente: Elaboración propia en base a información de la empresa

Se puede concluir que los clientes que pertenecen a “Latentes” (alta venta real y alta venta potencial) es el que más contribuye a la venta total de la empresa, pues está compuesto por los mejores clientes. Sin embargo, es importante observar la contribución del cuadrante de clientes “No prioritarios” (baja venta real y bajo potencial), que sigue a “Latentes” con un 23% de la venta total debido a que es el segmento que está conformado por la mayor cantidad de clientes.

Para determinar los patrones de compra o factores que influyan en el nivel de compra de los clientes de la empresa, se realizó un estudio descriptivo de los promedios de las variables transaccionales de los cuadrantes. En la **Tabla 26**, se pueden observar las principales diferencias:

| Promedios de los cuadrantes | Latentes | Preferentes | No prioritarios | Objetivos | Total |
|--------------------------------------|----------|-------------|-----------------|-----------|--------|
| Venta 2014 (miles S/.) | 7,86 | 4,97 | 3,61 | 5,96 | 5,60 |
| Venta 2013 (miles S/.) | 8,44 | 5,00 | 3,24 | 6,13 | 5,70 |
| Transacciones 2013 | 116,11 | 77,36 | 65,11 | 99,83 | 89,60 |
| Días entre compras 2013 | 7,09 | 9,50 | 10,48 | 7,98 | 8,76 |
| Frecuencia mensual 2013 | 5,18 | 3,97 | 3,48 | 4,67 | 4,33 |
| R/F 2013 | 0,71 | 0,81 | 0,90 | 0,77 | 0,80 |
| Número de categorías 2013 | 17,63 | 15,63 | 15,10 | 17,26 | 16,41 |
| Número de productos 2013 | 154,84 | 113,56 | 106,53 | 146,34 | 130,32 |
| Número de marcas 2013 | 29,92 | 26,16 | 24,82 | 29,03 | 27,48 |
| Aceites: ventas 2013 (miles S/.) | 1,91 | 1,34 | 0,55 | 0,80 | 1,15 |
| Detergentes: ventas 2013 (miles S/.) | 1,53 | 0,80 | 0,61 | 1,32 | 1,06 |
| Galletas: ventas 2013 (miles S/.) | 0,70 | 0,50 | 0,43 | 0,62 | 0,56 |
| Pastas: ventas 2013 (miles S/.) | 1,33 | 0,60 | 0,42 | 1,02 | 0,84 |
| Salsas: ventas 2013 (miles S/.) | 0,55 | 0,31 | 0,24 | 0,45 | 0,39 |

Tabla 26: Promedios de las variables transaccionales de los clientes de cada cuadrante
Fuente: Elaboración propia en base a información de la empresa

Los clientes que tienen alta venta potencial (“Latentes” y “Objetivos”), en promedio, son similares en los días entre compras, R/F, número de categorías, productos y marcas diferentes con compra, en comparación a los clientes de los cuadrantes de baja venta potencial (“Preferentes” y “No prioritarios”).

Con respecto al promedio de las ventas de las categorías principales, se encuentra un patrón de compra en detergentes, galletas, pastas y salsas en cuanto a proporciones de venta en los clientes de alta venta potencial, pero en aceites, se diferencian los clientes “Objetivos” por tener cercano a la mitad de venta promedio que los “Latentes”, lo cual explica que sean grupos diferentes, por su baja venta real en aceites.

Se puede replicar el análisis anterior para los clientes que tienen baja venta potencial (“Preferentes” y “No prioritarios”), pero se diferencian en menor medida en las ventas de las categorías, excepto en aceites, que nuevamente es la categoría que separa a los cuadrantes.

Además, en el **Gráfico 16** se muestra la distribución por zona de los clientes de cada cuadrante:

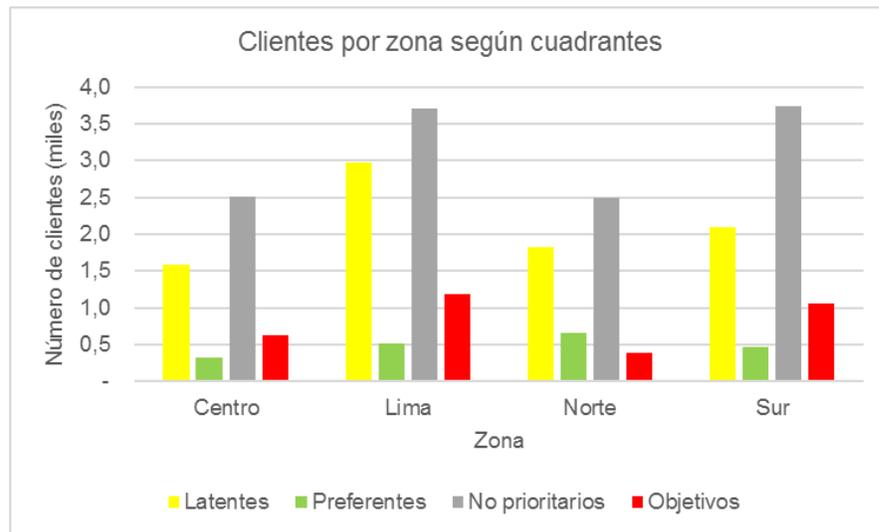


Gráfico 16: Número de clientes por zonas y cuadrantes
Fuente: Elaboración propia en base a información de la empresa

Los clientes “Objetivos” y “No prioritarios” se concentran en la zona sur y Lima, mientras que los clientes “Preferentes” están distribuidos en todas las zonas de manera más equitativa.

En el **Gráfico 17**, se puede observar la distribución de los tipos de vendedores que atienden a los clientes de cada cuadrante, en donde hay un mayor número de vendedores tipo A (los mejores vendedores) en los cuadrantes de los clientes “Latentes” y “No prioritarios”. Se destaca que los clientes “Preferentes” tienen un menor número de vendedores de todos los tipos, lo cual se puede justificar con que es el cuadrante con menor número de clientes (ver **Tabla 25**), pero proporcionalmente son los que tienen un mayor porcentaje de vendedores tipo A (ver gráfico en Anexo 19).

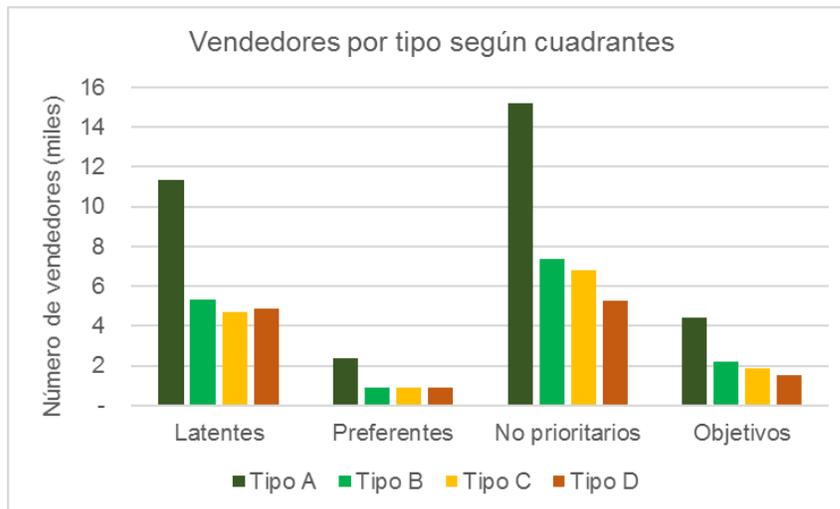


Gráfico 17: Número de vendedores por tipo y cuadrantes
Fuente: Elaboración propia en base a información de la empresa

Para evaluar si las diferencias de los **Gráfico 16** y **Gráfico 17** son significativas, se realizó un test Chi-Cuadrado (χ^2) sobre las proporciones de clientes según zona y tipo de vendedores en cada cuadrante (ver Anexo 18 y Anexo 19, respectivamente). Se puede concluir del test que, para ambos casos, las diferencias son significativas entre los distintos cuadrantes, dado que a un nivel de confianza del 95%, se rechaza la hipótesis nula que plantea que son iguales, ya que la significancia del test es menor a 0,05.

Finalmente, como complemento de lo anterior, se realiza el mismo ejercicio de los cuadrantes de venta real vs. venta potencial, con los resultados de la venta estimada en el modelo de fronteras de categorías en el **Gráfico 18** para ver el diferencial de potencial de los clientes al comparar la venta real con la venta estimada por el modelo de frontera.

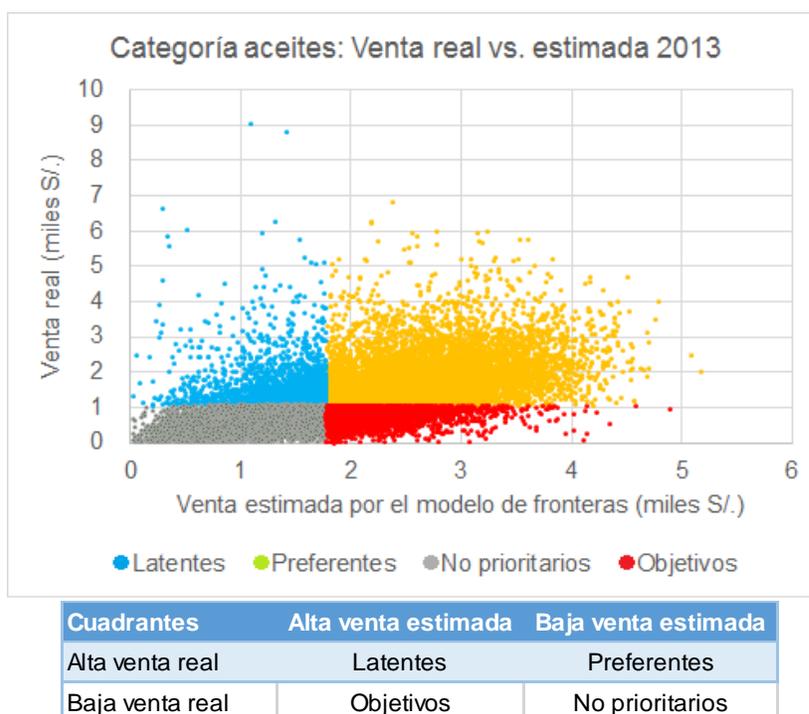


Gráfico 18: Categoría aceites – Venta real vs. Venta estimada por el modelo de frontera 2013

Fuente: Elaboración propia en base a información de la empresa

En el **Gráfico 18**, se puede apreciar la asimetría que existe entre los cuadrantes con respecto al diferencial de potencial de compra de los clientes. Se reafirman las conclusiones anteriores que los clientes que pertenecen a “Objetivos” tienen mayor espacio para aumentar sus ventas si el proveedor peruano toma alguna acción sobre ellos.

7.4 Propuestas de líneas de acción

Para este trabajo, las propuestas que se recomiendan son para fomentar la relación entre los clientes con mayor potencial de compra y la empresa, de acuerdo al ranking de clientes de la sección anterior.

Se puede observar que, para aumentar las ventas de la empresa, puede ser de gran utilidad enfocar los esfuerzos en incrementar el nivel de ventas en el segmento “Objetivos”, ya que los clientes tienen baja venta real y alta venta potencial. Aquel cuadrante de clientes tiene más oportunidades para crecer que los demás.

Se infiere que, los clientes que tienen mayor venta real y menor potencial de compra (“Preferentes”) al estar bien abordados por la empresa, de acuerdo al modelo de frontera, están comprando cercano al total de su capacidad, se pueden imitar las buenas prácticas de ventas que hay dentro de este segmento, que hace que tengan poco potencial. Por lo que, se recomienda realizar una acción para hacer más eficaz la ocasión de compra de las bodegas, ya que los “Preferentes” a pesar de no tener un alto número de transacciones, en comparación al resto, son los clientes que están más próximos a la frontera de ventas, por ende, son más eficientes en cada compra, tal vez porque gestionan mejor sus pedidos.

Además, se propone aumentar los vendedores tipo A en los clientes “Objetivos”, ya que los clientes “Preferentes” en proporción tienen un mayor número de vendedores de este tipo.

Por otro lado, tras analizar el Modelo (I) construido para la estimación del potencial de compra se proponen las siguientes líneas de acción concretas:

- Manipular la política de precios, presentando ofertas atractivas para los clientes de manera que permita aumentar la frecuencia de compra de los clientes, aprovechando el impulso para mejorar el nivel de ventas y con ello aprovechar el potencial de compra disponible.
- Proponer nueva política de recaptura de clientes cuyo R/F excede 2, de esta manera, si un cliente supera esa cifra, se puede ofrecer facilidades de compra de productos para conseguir que éstos vuelvan a comprar a la empresa, por ejemplo: por medio de compra en cuotas, descuentos, microcréditos, etc.
- Generar una cultura de capacitación de vendedores que apunte a la replicación de prácticas que se identifiquen como comunes entre los vendedores de tipo A.
- Estudiar el efecto de la ubicación de la cartera de clientes de los vendedores sobre su desempeño, buscando identificar la existencia de espacios de mejora en los resultados reubicando los vendedores disponibles.
- Y como recomendaciones generales:
 - Diseñar un plan de marketing enfocado en la captura de clientes potenciales que en la actualidad están comprando a la competencia.
 - Desarrollo de una política de empresa que apoye a microempresarios (bodegas) emergentes, generando una ampliación del mercado de los clientes, por ejemplo, política preferencial de precios a bodegas con menos de 5 años de antigüedad.

Se recomienda generar estrategias tácticas, en precios y promociones, en torno a las variables más significativas de los modelos de regresión lineal y de frontera (Modelo (II) – Categorías principales) para cada categoría.

En particular, para aceites, se sugiere potenciar pastas y detergentes para los clientes “Objetivos”, ya que el promedio de ventas en la categoría aceites es mucho menor al resto de clientes, aunque, las ventas en las otras categorías principales sean incluso mayor, por ejemplo, el promedio de venta en detergentes y pastas es mayor que el de los clientes “Preferentes” y “No prioritarios”.

En la misma directriz, se resume en la **Tabla 26** el orden, según ventas totales de las categorías base, para aumentar las ventas de acuerdo a etapas de prioridad para la gestión de las categorías:

| Categoría base | | Categorías a potenciar | | |
|----------------|-------------|------------------------|---------------|-----------------------------|
| Orden | Prioridad | Primera etapa | Segunda etapa | Tercera etapa |
| 1 | Aceites | Detergentes | Pastas | Jabones de lavar |
| 2 | Detergentes | Jabones de lavar | Aceites | Margarinas |
| 4 | Pastas | Aceites | Harinas | Detergentes |
| 3 | Galletas | Caramelos | Aderezos | Refrescos instantáneos |
| 5 | Salsas | Detergentes | Margarinas | Harinas y ayudas culinarias |

Tabla 27: Categorías a potenciar en orden de prioridad

Fuente: Elaboración propia en base a información de la empresa

Se sugiere no dar énfasis a salsas para aumentar las ventas, a pesar de ser una de las categorías principales del proveedor peruano con casi un 8% de las ventas totales, se contradice con los resultados de todos los modelos realizados, pues no influye en las ventas significativamente como el resto de las categorías. Lo mismo ocurre con la influencia que tiene panetones en las ventas, pero esta categoría representa menos de un 1% de las ventas totales.

Se recomienda priorizar los clientes “Objetivos” de Lima y, en segundo lugar, la zona sur, ya que tienen un mayor número de clientes de ese segmento y por simplicidad a nivel logístico y de gestión de control de las recomendaciones propuestas. Además, existe un alto nivel de habitantes en Lima que son los clientes finales de las bodegas.

CAPÍTULO 8. CONCLUSIONES

El mercado del *retail* enfrenta un entorno competitivo basado en la diferenciación, productividad y calidad de servicio, con el factor común los recursos humanos disponibles, con una mirada en el largo plazo. Por lo tanto, es necesario conocer a los clientes, identificando a los clientes potenciales, de manera de formular acciones focalizadas que permitan rentabilizar y estrechar la relación entre clientes y empresa. En el caso del proveedor peruano, tiene implicancias más significativas aún dado los grandes volúmenes de dinero que mueve la empresa y, trae consigo, un aumento en las ventas, lo cual está alineado con la visión y definiciones de negocio de la empresa.

La principal forma de aumentar las ventas es vender más, vender mejor y anticiparse a la competencia. Lo que se puede lograr a través de propuestas interesantes de categorías preferentes para maximizar oportunidades de compra o crear acciones de *up selling* (incentivar la compra de productos de mayor calidad) y/o *cross selling* (aumentar el número de categorías compradas), que incluyan reglas de asociación e interacción entre productos y categorías, lo cual intensifica la satisfacción y lealtad del cliente.

Además, se pueden formular propuestas de acciones de marketing focalizadas para estos clientes que fueron mal atendidos, enfocadas en el uso eficiente tanto de los recursos humanos como de los datos, aumentando la calidad de servicio y asignando de mejor modo las fuerzas de venta, de manera de lograr coherencia entre las acciones comerciales y los objetivos estratégicos del proveedor peruano.

Se concluye que las etapas de la metodología fueron desarrolladas satisfactoriamente, lo que reafirma el problema que no se conocen los patrones de comportamiento de los clientes y da una mejor visión de cómo abordarlo, identificando quiebres a lo largo del proceso.

El manejo, orden y estandarización de las bases de datos, es un proceso lento y difícil, pero es el paso base para realizar modelos con un alto nivel de datos transaccionales, por lo que la metodología basada en KDD tuvo un buen desempeño, que otras técnicas que se pudieron utilizar.

Con respecto a los modelos simples, estos sirven para tener una visión de parte del problema y resolverlo de manera rápida y fácil, por ejemplo, asumir que los clientes se comportarán de acuerdo al promedio general o promedio del sector donde pertenecen, o se mantendrán igual, es decir, que en el siguiente periodo analizado gastarán lo mismo.

Mediante la creación y selección de variables, fue posible identificar patrones y simplificar el reconocimiento de características esenciales de los clientes. En particular, la elección de las categorías principales juega un rol importante para el proveedor peruano, ya que generan más de un 77% de las ventas y fue una buena base para generar los modelos.

Luego de analizar los resultados, se puede concluir que todos los modelos pueden estimar el potencial de compra de los clientes de diferentes maneras, identificando las variables que influyen en las ventas del proveedor peruano. A pesar de que los resultados no son iguales en los modelos, se puede indicar que son robustos, ya que entregan confianza en sus pronósticos de aumento de ventas, al no presentar grandes cambios frente a pequeñas variaciones. Sin embargo, no se debe dejar de lado el juicio de un experto, ya que es útil para identificar problemas y validar los resultados, puesto que las reglas de negocio y características de la industria son importantes para la formulación de los modelos, pero no se explicitaron a lo largo de la memoria, puesto que se trabajó con un solo tipo de cliente minorista.

Los resultados principales de esta memoria, arrojan un aumento potencial de las ventas por casi un 11% en el Modelo (I) de 3 efectos, 12% en el Modelo (II) de regresiones lineales y 77% en el de regresiones de fronteras estocásticas. Cifras no despreciables para la empresa considerando que el *retail* es un sector muy competitivo, pues representa en torno a S/. 16.012.972, S./ 12.053.178 y S/. 69.525.773 en el caso de modelo de fronteras, respectivamente.

Lo anterior, no constituye la ganancia absoluta sin márgenes de error, lo que no sólo dependerá de los valores entregados por los modelos, sino de la implementación y coherencia en las decisiones de la empresa para aumentar las ventas.

También se debe tener cuidado de atribuir a una disminución en las ventas del proveedor a la efectividad del modelo, puesto a que la base de datos esta sesgada a la información disponible de la misma empresa y no se ha tomado en cuenta a lo que compran las bodegas a la competencia. A pesar que el proveedor peruano está creciendo, en promedio, más rápido que el *retail* en Perú, dejando más espacio para aumentar las ventas, y aunque tenga una gran participación de mercado, la evidencia no es clara y se debe tratar de superar esta complejidad, lo cual hace que el porcentaje de aumento de ventas del modelo de regresión de frontera esté sobre estimado por esa razón.

Como conclusión general, los objetivos del trabajo fueron logrados, pero queda espacio para avanzar en profundidad en la gestión de categorías del proveedor peruano, reconociendo problemas, de manera clara, en la fuerza de venta y patrones de

comportamiento, en relación al potencial de compra de los clientes. Esto no solo supe las necesidades del proveedor peruano, si no también, de otras empresas similares en conjunto.

Se destaca que no se validaron hipótesis como alta rotación de los vendedores, que podría explicar la cantidad de vendedores que atienden a un cliente, lo que dificultó la elección de que variable incorporar a los modelos para estudiar el efecto que tiene el tipo de vendedor en las ventas. Pero se concluye que no es trivial optimizar la distribución de la fuerza de ventas de manera de asignar eficientemente los recursos de la empresa para captar el potencial de compra total de los clientes.

En cuanto a los modelos utilizados para resolver el problema de estimación del potencial de compra de los clientes, es que el inconveniente de utilizar otros métodos como programación matemática es que los parámetros son calculados en vez de estimados, en los mínimos cuadrados corregidos es que la frontera estimada es paralela a la obtenida por la regresión de los mínimos cuadrados ordinarios, puesto que solo se corrige el término independiente. También, un inconveniente del método de mínimos cuadrados modificados es que no garantiza que todas las producciones queden limitadas superiormente por la frontera de producción, es decir, si un cliente tiene un residuo de mínimos cuadrados ordinarios grande, es posible que la eficiencia técnica sea mayor que la unidad, y, además, la frontera de producción igual es paralela a la de regresión con mínimos cuadrados ordinarios y no permite que haya diferencia entre la mejor producción y la producción media de todos los productores.

Complementando lo anterior, las técnicas descritas anteriormente, tienen el inconveniente común de que cada una de ellas mide la eficiencia técnica con una frontera de producción determinística, es decir, toda la variación en la producción, no asociada con la variación de entradas, se atribuye a la ineficiencia técnica, y ninguna de las técnicas mencionadas permite el efecto de perturbaciones aleatorias, las cuales también podrían contribuir a la variación de las salidas. Por lo tanto, lo que se necesita es un modelo que atribuya las variaciones en las producciones, no asociadas con variaciones en las entradas, como combinación de perturbaciones aleatorias e ineficiencia técnica. Este tipo de modelo de fronteras de producción estocástica es más complejo, pero también más realista.

CAPÍTULO 9. TRABAJOS PROPUESTOS

Se presentan algunas sugerencias sobre los trabajos futuros que se pueden desarrollar tomando como base el trabajo presentado anteriormente, de esta forma se puede indagar con mayor profundidad en otros métodos y variables para estimar el potencial de compra de los clientes y compararlos con los modelos propuestos en este documento.

Continuando en la misma línea de lo anteriormente planteado, se plantea la integración de un sistema de metas para la fuerza de venta en base al potencial de compra, es decir, implementar un mecanismo que establezca objetivos a los vendedores, junto con planes de redistribución, que sea modelado tomando como base el comportamiento del potencial de clientes dentro de la cartera de cada uno de los vendedores.

Se recomienda profundizar con la investigación de las variables de la encuesta que, intuitivamente, son relevantes al estimar las ventas de las bodegas, pero en los modelos desarrollados no mostraron resultados concluyentes.

Otra posibilidad de realizar el análisis, sería disminuir el período de tiempo de estudio, por ejemplo, el equivalente a la frecuencia de compra promedio. Sin embargo, este período no puede ser muy reducido, ya que para calcular el potencial de cliente se deben evitar presencias sobre stock o problemas externos a las ventas, como por ejemplo que los vendedores no asistan por razones de fuerza mayor, o que el dueño de bodega no esté presente para realizar el pedido, entre otras.

Además, es importante destacar que se debe tener un conocimiento de cuánto vale el cliente hoy y a largo plazo, ya que este dato es relevante para generar propuestas de acción para estos mismos. Por esta misma razón se recomienda calcular la valoración del cliente incluyendo estimaciones de los montos futuros de las transacciones a las cuales estará ligado, junto con un análisis de sensibilidad de la tasa de descuento, tasa de retención, entre otros supuestos de *Customer Lifetime Value*.

Finalmente, se propone aplicar la misma metodología anteriormente desarrollada, en conjunto con una metodología de evaluación, para aquellos clientes que no fueron considerados, es decir, otros tipos de clientes minoritas de la empresa. No se tiene certeza de ellos, pero podrían tener sorpresivamente un alto potencial como clientes.

BIBLIOGRAFÍA

Bhattacharya, C. B. 1998. "When customers are members: Customer retention in paid." *Journal of the Academy of Marketing Science* 26 (1): 31-44.

Buttle, Francis. 2009. *Customer Relationship Management: Concepts and Technologies*. Estados Unidos: Elsevier.

Cava, L. G. R. 2004. *IESE Business School: University of Navarra*. Accessed 11 06, 2015. http://www.iese.edu/en/files/6_13439.pdf.

CCL, and IEDEP. Agosto 2013. *Informe Sectorial: Comercio*. Lima, Perú: Cámara de Comercio – Lima; Instituto de Economía y Desarrollo Empresarial.

Chmabers, J.M. 1992. *Statistical Models in S. Capítulo 4: Linear models*. Edited by J.M. Chmabers and T. J. Hastie. Chapman and Hall/CRC.

D. Aigner, C.A. Lovell, P. Schmidt. 1977. "Formulation and Estimation of Stochastic Frontier Production Function Models." *Journal of Econometrics* 6 (1): 21-37.

Dufour, Jean-Marie. 1983. "Coefficients of determination." McGill University.

Fayyad, U., G Piatetsky-Shapiro, and P. Smyth. 1996. *From Data Mining to Knowledge Discovery in Databases* (Communications of the ACM) (39): 24-26.

Forbes, T. 2007. "Valuing Customers." *Journal of Database Marketing & Customer Strategy* 4-10.

n.d. *Freemium Economics: Leveraging Analytics and User Segmentation to Drive Revenue*.

Garrido, Marcel. 2009. *Estimación del Valor de Clientes de un Club de Fidelización de una Tienda de Retail. Memoria para optar al título de Ingeniero Civil Industrial*. Santiago, Chile: Universidad de Chile, Facultad de Ciencias Físicas y Matemáticas, Departamento de Ingeniería Industrial.

González, Angela Corte. 2015. "Frontera de producción." Investigación matemática, estadística y computación, Universidad de Oviedo.

Han, J., M. Kamber, and J.Pei. 2006. *Data Mining: Concepts and Techniques*. Morgan Kaufmann Publishers.

Hastie, T. J., and D. Pregibon. 1992. *Generalized linear models: Chapter 6 of Statistical Models in S*. Edited by J. M. Chambers and T. J. Hastie. Wadsworth & Brooks/Cole.

J.S. Armstrong, F. Collopy. 1992. "Error Measures for Generalizing About Forecasting Methods: Empirical Comparisons." *International Journal of Forecasting* 8 (1): 69-80.

Kotler, Philip, and Gary Armstrong. n.d. *Fundamentos de marketing. Sexta Edición*. Pearson.

Kumbhakar S.C, Lovell C.A.K. 2000. *Stochastic Frontier Analysis*. Binghamton, New York: Iberoamericana.

Memoria Anual Empresa, 2014. n.d. "Bolsa de Valores de Lima." Accessed 11 06, 2015. <http://www.bvl.com.pe/>.

Naumann, Earl. 1994. *Creating Customer Value: The Path to Sustainable Competitive Advantage*. Suecia: South-Western Pub; 1st edition.

Núñez, V. 2012. *Estimación del efecto de atención incidental en las ventas de un supermercado. Memoria para optar al título de Ingeniero Civil Industrial*. Santiago de Chile: Universidad de Chile, Facultad de Ciencias Físicas y Matemáticas, Departamento de Ingeniería Industrial.

Pardoe, Dr. Iain. n.d. *STAT 501 Online course materials website*. Dr. Laura Simon and Dr. Derek Young. <https://onlinecourses.science.psu.edu/stat501/print/book/export/html/266>.

Peña, D. 2002. *Análisis de datos multivariantes*. Madrid: McGraw-Hill.

Rasmussen, E. 1999. "Complaints can build relationships." *Sales and Marketing Management* 9 (151): 89-90.

Reichled, F, and W.E. Sasser Jr. 1990. "Zero Defections: Quality Comes to Services." *Harvard Business Review* 68 (5): 105-111.

Renart Cavas, Luis G. 2004. *CRM: tres estrategias de éxito*. IESE. Accessed 11 06, 2015. <http://www.iese.edu/research/pdfs/ESTUDIO-15.pdf>.

S. Gupta, D. Hanssens, B. Hardie, W. Kahn, B. Kumar, N. Lin and N.R.S. Sriram. 2006. "Modeling Customer Lifetime Value." *Journal of Service Research* 9 (2): 139-155.

Sakamoto, Y., M. Ishiguro, and G. Kitagawa. 1986. *Akaike Information Criterion Statistics*. D. Reidel Publishing Company.

Schwarz, G. 1978. *Estimating the dimension of a model*. Vol. 6.

Seufert, Eric Benjamin. 2014. *Freemium Economics*. Massachusetts.

University of New England. 2014. "Topic 4.1 Stochastic Frontier Analysis Part 1."

Vallejos, S. 2006. *Minería de Datos*. Corrientes: Universidad Nacional del Nordeste.

Venables, W. N., and B. D. Ripley. 2002. *Modern Applied Statistics with S*. 4. Nueva York: Springer.

Webster, Allen. 2000. *Estadística aplicada a los negocios y la economía. Capítulo 10*. Madrid: Irwin.

Wikipedia. 2010. *Likelihood-ratio test*. November. Accessed Enero 2016.
https://en.wikipedia.org/wiki/Likelihood-ratio_test.

Wilkinson, G. N., and C. E. Rogers. n.d. *Symbolic Description of Factorial Models for Analysis of Variance. Applied Statistics*. Vol. 22.

Zapata, W. 2006. *Un Nuevo enfoque de la fidelización del cliente: CRM*.

ANEXOS

Anexo 1: Marco conceptual – KDD

El proceso de “Descubrimiento de Conocimiento en Bases de Datos” (KDD por sus iniciales en inglés), propuesta metodológica de M.J. Berry, se define como la “extracción no trivial de patrones en los datos válidos, novedosos, potencialmente útiles y entendibles a partir de una base de datos” (Vallejos 2006).

KDD aparece como solución a la necesidad de nuevas técnicas y metodologías para el trabajo con grandes bases de datos, y este proceso, se hace cargo y da solución, conceptualmente, a un número de problemas que surgen del constante crecimiento de las bases de datos, tales como (Fayyad, Piatetsky-Shapiro y Smyth 1996):

- Manejo de gran cantidad de datos.
- Integración de datos de distintas fuentes (en distinto formato).
- Manejo de datos dañados, errados o faltantes.
- Identificación de patrones donde resulta muy complejo hacer hipótesis a priori dado el tamaño de la muestra.

De esta forma, el proceso KDD presenta una metodología que permite encontrar información y conocimientos relevantes de manera estructurada, asegurando la calidad de los datos a utilizar, se enfoca sólo en los datos relevantes al problema y se encarga de la clasificación de los métodos para el descubrimiento, entendimiento y validación de patrones, alineados con los objetivos del proyecto (Fayyad, Piatetsky-Shapiro y Smyth 1996).

Anexo 2: Marco conceptual – Regresión Lineal

La técnica utilizada para estimar los parámetros es Mínimos Cuadrados (MLS), $\hat{\beta} = (X'X)^{-1}X'Y$, que busca minimizar el error cuadrático medio. Se debe considerar que, al ocupar este método para ajustar la regresión, el valor del coeficiente de determinación (R^2) aumenta mientras mayor es el número de variables independientes (explicado en mayor profundidad en la sección 6.2.1.2 Evaluación del modelo de regresión lineal). Sin embargo, se debe medir la bondad del ajuste, corroborando que el coeficiente de determinación también aumente al agregar más variables X_i .

Algunas propiedades de la regresión lineal son (Wilkinson y Rogers s.f.):

- El estimador de mínimos cuadrados es insesgado.
- El estimador de mínimos cuadrados es eficiente en el sentido que tiene la menor varianza posible dada la información disponible.
- El estimador máximo verosímil maximiza la probabilidad de la observación. Este estimador es asintóticamente insesgado.
- El estimador máximo verosímil, para el caso lineal con errores normales coincide con el estimador de mínimos cuadrados, es decir, minimiza el error de ajuste.

Anexo 3: Marco conceptual – Evaluación de modelos

- **Coefficiente de Determinación (R^2):** es definido (Dufour 1983) como:

$$R^2 = \frac{SCR}{SCT} = 1 - \frac{SCE}{SCT} = 1 - \frac{\sum_i(\hat{y}_i - \bar{y})^2}{\sum_i(y_i - \bar{y})^2}$$

Donde, SCT (suma de cuadrados totales) representa la variación total con respecto a la media, SCR (suma de cuadrados de la regresión) la porción de variación que es atribuible a un efecto lineal de las variables de predicción sobre las variables de respuesta y SCE (suma de los cuadrados del error) interpreta a la variación del error.

Por lo tanto, este coeficiente es una medida estandarizada que toma valores entre 0, cuando las variables son independientes, y 1 cuando existe una relación perfecta entre las variables. Si $R^2 = 1$, puede afirmarse que toda variación presente en las observaciones es explicada por la presencia de las variables de predicción en la ecuación de regresión.

- **Prueba F de Fisher (F – test):** se realiza para probar la hipótesis nula sobre los parámetros, $H_0: \beta_j = 0, \forall j$ utilizando el siguiente estadístico (Pardoe s.f.):

$$F = \frac{SCR/(m - 1)}{SCE/(n - 1)}$$

Donde, el estadístico $F_{(m-1, n-1)}$ con $m - 1$ grados de libertad en el numerador y $n - 1$ grados de libertad en el denominador, m es el número de parámetros y n el número de observaciones.

Si F es grande, la mayor proporción de variación en los datos es debida a la regresión, por lo tanto, se rechaza la hipótesis si $F > F_\alpha$ para un α de confianza determinado.

F – test es una prueba de idoneidad general del modelo, que indica si los datos proporcionan o no pruebas suficientes que muestren que el modelo global contribuye con la información a la predicción.

- **Mean Absolute Percentage Error (MAPE):** se calcula con la siguiente fórmula (J.S. Armstrong 1992):

$$MAPE = \frac{1}{n} \sum_i \left| \frac{y_i - \hat{y}_i}{y_i} \right|$$

- **Índices de Ratios de Verosimilitud:**

$$\rho = 1 - \frac{LL(\hat{\beta})}{LL(0)}$$

Asociado con el método de máxima verosimilitud, para un modelo de k parámetros estimados sobre n observaciones, se construyen criterios de información como estimadores aproximadamente insesgados de la log-verosimilitud esperada:

- **Criterio de información de Akaike (AIC):** propuesto por Akaike (1974) como estimador insesgado asintótico de la información de Kullback-Leiber (K-L) esperada, entre un modelo candidato ajustado y el modelo verdadero, es decir, es una medida de la calidad relativa de un modelo estadístico para un conjunto dado de dato.

La idea clave de AIC es penalizar un exceso de parámetros ajustados, lo cual no lo resuelve el test asintótico de chi-cuadrado (χ^2). AIC es un estimador muestral de la esperanza de la log-verosimilitud, $E[\ln f(X | \theta)]$, que viene dado por la expresión (Sakamoto, Ishiguro y Kitagawa 1986):

$$AIC(k) = -2 \ln LL(\hat{\beta}) + 2k = -2 \ln \mathcal{L}[\hat{\theta}(k)] + 2k$$

Donde, $\mathcal{L}[\hat{\theta}(k)]$ es la función de verosimilitud de las observaciones, $\hat{\theta}(k)$ es la estimación máximo verosímil del vector de parámetros θ y k es el número de parámetros independientes estimados dentro del modelo. El primer término del estimador puede ser interpretado como una medida de la bondad del ajuste, mientras que el segundo término es una penalización creciente conforme aumenta el número de parámetros de acuerdo al Principio de Parsimonia, independiente del tamaño de muestra.

Algunas de las ventajas del AIC que lo hacen tan utilizado en la práctica, son su simplicidad y facilidad para ser implementado, y el hecho de que no existe el problema de especificar subjetivamente un nivel de significación arbitrario para contrastar dos modelos.

El criterio AIC no pretende seleccionar el modelo correcto, ya que admite que el modelo verdadero puede no estar entre los estimados, y trata de obtener el modelo que proporcione mejores predicciones entre los existentes.

Suponiendo que el modelo verdadero puede aproximarse arbitrariamente bien con los estimados, puede demostrarse que, al aumentar el tamaño muestral, el AIC es

eficiente, escogiendo el modelo que proporciona (en promedio) mejores predicciones. Sin embargo, AIC tiende a seleccionar modelos con más parámetros de los necesarios (Peña 2002).

- **Criterio de información Bayesiano (BIC):** en el contexto de procedimientos basados en la verosimilitud, Schwarz (1978) sugiere que el AIC podría no ser asintóticamente justificable, y presentó un criterio de información alternativo a partir de un enfoque bayesiano, como una aproximación a una transformación de la probabilidad posterior de un modelo candidato, que penaliza el número de parámetros con $\ln(n)$, en lugar de 2, así (Sakamoto, Ishiguro y Kitagawa 1986):

$$BIC(k) = -2\ln LL(\hat{\beta}) + k * \ln(n) = -2 \ln \mathcal{L}[\hat{\theta}(k)] + k * \ln(n)$$

Donde, $\mathcal{L}[\hat{\theta}(k)]$ es la función de verosimilitud de las observaciones, $\hat{\theta}(k)$ es la estimación máximo verosímil del vector de parámetros θ y k es el número de parámetros independientes estimados dentro del modelo.

Por lo tanto, el BIC trata de seleccionar el modelo correcto, con máxima probabilidad a posteriori, y puede demostrarse que es un criterio consistente, de manera que la probabilidad de seleccionar un modelo correcto tiende a 1 cuando aumenta el tamaño muestral (Schwarz 1978).

- **Test de Ratio de Verosimilitud** (Wikipedia 2010): Considerando 2 modelos, A y B, si se puede derivar el modelo B imponiendo restricciones sobre los parámetros del modelo A, se dice que el modelo B está anidado en el modelo A. Para determinar si el modelo A, que tiene más parámetros que B, se ajusta mejor a los datos que el modelo B, se puede usar el test de ratios de verosimilitud (LR), calculando el estadístico: $LR = 2(LL_A - LL_B)$. Siendo k el número de parámetros adicionales del modelo A, si: $LR > \chi^2(0,05; k)$, entonces el modelo A es mejor que B, es decir, los parámetros adicionales ayudan a mejorar el ajuste.

Anexo 4: Marco conceptual – Modelo de Regresión de Fronteras

En la práctica, las fronteras de producción no son más que una regresión que se ajusta a los datos con la restricción que todas las observaciones deben encontrarse debajo de la frontera y al menos una de ella debe estar sobre ella. De esta forma, se puede interpretar la frontera de producción como el límite máximo de producción que sirve de referencia para el cálculo de la ineficiencia del resto de las unidades observadas, tal como se puede ver en el **Gráfico 19** (González 2015):

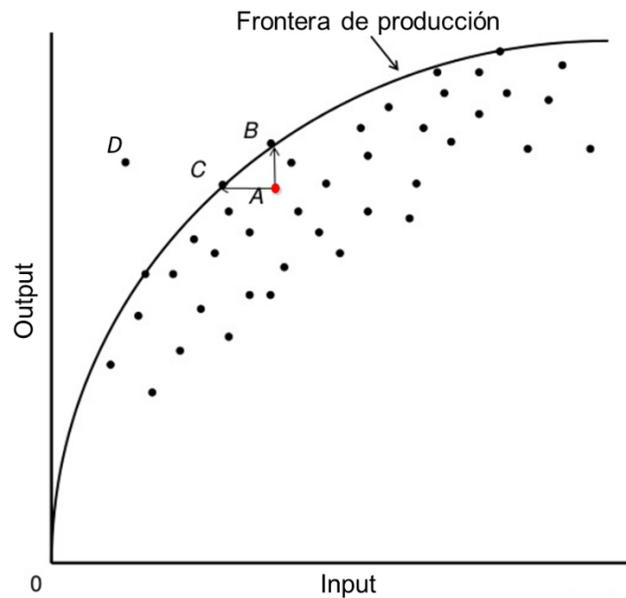


Gráfico 19: Frontera de posibilidades de producción

Fuente: Elaboración propia en base a (University of New England 2014)

El **Gráfico 19** muestra el conjunto de las combinaciones entre input (recursos) y output (producción) en este ejemplo, pero el subconjunto de puntos en la frontera de producción representa la eficiencia de producción posible de acuerdo a aquellos datos. En el gráfico, el punto A representa una firma ineficiente, mientras que el punto B y C son eficientes. Para que A sea eficiente, debe llegar al punto B (aumentando su producción) sin requerir más recursos y, también, puede reducir el uso de recursos, pero conservando su nivel de producción (punto C). El punto D muestra una producción no factible técnicamente según la asignación de recursos, ya que representan volúmenes de producción que no son alcanzables a partir de los recursos con los que cuenta.

Por lo tanto, los productores que operan en su frontera de producción se dirá que son técnicamente eficientes, puesto que son los niveles máximos de producción posibles dados unos recursos determinados. Por otro lado, los productores que están por debajo de ella serán técnicamente ineficientes.

La productividad de una firma es el ratio entre los outputs que produce con los recursos de los inputs que se utilizan, es decir: $Productividad = Outputs/Inputs$.

El primer autor que motivó la utilización de las fronteras de producción para analizar la eficiencia fue Farrell en 1957. Para él, la forma correcta de medir la eficiencia era mediante la comparación de cada observación con la mejor práctica observada. Luego en 1977, Aigner, Lovell y Schmidt propusieron el modelo de frontera estocástica.

Los modelos paramétricos que se estudiarán serán de ecuación única, de los cuales se estiman los parámetros que describen la estructura de la frontera y las estimaciones de la eficiencia técnica de cada productor se obtienen como subproducto de lo anterior. El término asociado a la eficiencia técnica puede ser determinado a través de dos puntos de vista: del determinístico, que sólo consideran el término como una componente de ineficiencia unilateral; y el estocástico, que agrega el ruido aleatorio tradicional simétrico.

Anexo 5: Marco conceptual – Modelo de frontera de producción determinística

Bajo el supuesto que el vector de recursos es no-negativo y se denota por $\vec{x} = (x_1, x_2, \dots, x_n) \in (\mathbb{R}^+)^n$. La producción es un vector no-negativo de salidas y . Entonces, modelo de frontera de producción se escribe como (González 2015):

$$y_i = f(\vec{x}_i, \vec{\beta}) * TE_i$$

Donde, y_i es el escalar salida del productor i , $i = 1, \dots, I$, \vec{x}_i es un vector de N entradas del productor i , $f(\vec{x}_i, \vec{\beta})$ es la frontera de producción, y se supondrá que la función de producción es la siguiente $f(\vec{x}_i, \vec{\beta}) = \prod_n x_{ni}^{\beta_n}$, y $\vec{\beta}$ es el vector de parámetros a estimar.

Como TE_i representa la eficiencia técnica del productor i , despejando de la ecuación anterior, se tiene que:

$$TE_i = \frac{y_i}{f(\vec{x}_i, \vec{\beta})}$$

Lo anterior, define la eficiencia técnica como una proporción entre la salida observada y el máximo valor factible de la producción. De modo que y_i alcanza el máximo valor factible de $f(\vec{x}_i, \vec{\beta})$, si y sólo si, $TE = 1$, es decir, implica que está en la frontera de producción. En otro caso, la eficiencia técnica menor que 1 y proporciona una medida de déficit observado en la producción sobre el máximo de la producción posible.

En la ecuación anterior de y_i , la frontera de producción $f(\vec{x}_i, \vec{\beta})$ es determinística. Como consecuencia, todo el déficit de producción observado, y_i , sobre el máximo valor alcanzable, $f(\vec{x}_i, \vec{\beta})$, se atribuye a la ineficiencia técnica, ya que no se recoge el efecto que pueden ejercer perturbaciones aleatorias sobre la producción no controlables por el productor.

Anexo 6: Marco conceptual – Comparación de modelos de fronteras de producción

Para comparar las diferencias en los resultados de los modelos de fronteras de producción determinístico y estocástico, se presenta el siguiente **Gráfico 20**, donde se puede observar la frontera de producción de una firma y el nivel de producción, interpretado por los dos análisis anteriormente nombrados.

El error técnico del modelo determinístico está sólo expresado por la ineficiencia determinística, mientras que el error técnico del modelo estocástico lo indican en conjunto la ineficiencia estocástica más el error aleatorio o ruido (que en este caso es negativo).

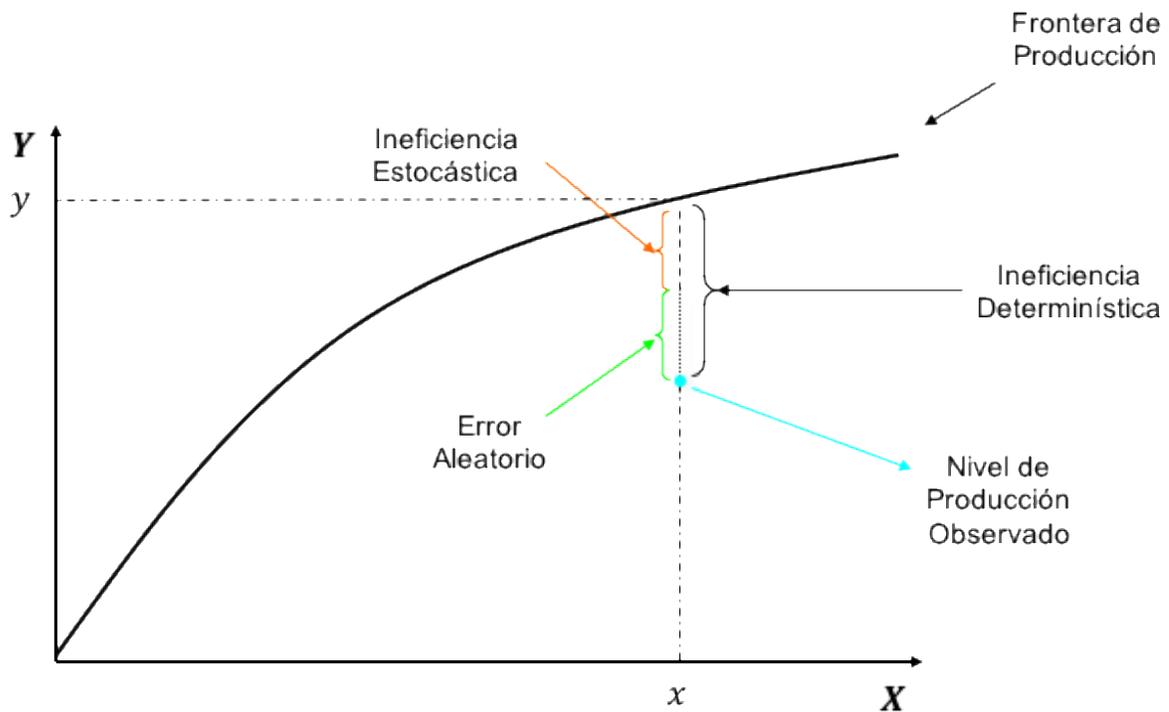


Gráfico 20: Comparación de los modelos de regresión de fronteras

Fuente: Elaboración propia en base a (University of New England 2014)

Anexo 7: Marco conceptual – Customer Lifetime Value

Para cuantificar el valor del cliente asociado a este trabajo se usa el CLV, el cual fue definido hace aproximadamente 30 años por Kepler (1974), como “el valor actual de los beneficios que se espera obtener a lo largo del tiempo que vaya a durar la relación con el cliente”.

En términos más modernos se define como la suma de los flujos descontados que un consumidor, o un segmento de consumidores, genera durante su relación con la empresa (Berger y Nasr, 1998), o, dicho de otro modo, “el valor presente neto de los beneficios ligados a cada consumidor, una vez que este ha sido adquirido, menos los costes asociados a cada consumidor a lo largo de la vida de dicho consumidor” (Dywer, 1997; Blattberg, 2008).

La necesidad asociada a negocios específicos, generó que se fuera complejizando, por ejemplo, Haenlein (2007), utiliza un modelo basado en una combinación de una cadena de Markov de primer orden y un árbol de clasificación y regresión (CART) para definir los posibles estados probabilísticos en que se puede encontrar un cliente.

Donkers (2007) realiza una comparación de los modelos formulados para el cálculo del CLV. En términos simples compara la capacidad predictiva de los modelos cuando son aplicados al *retail*. El trabajo de Donkers es uno de los más complejos en el cálculo del CLV en este mercado. Donkers concluye que los modelos sencillos proporcionan buenas predicciones del CLV y que los más complejos consiguen mejoras marginales.

El indicador CLV es fundamental para las empresas que deseen desarrollar una mejor estrategia de CRM, dado que al conocer el valor de los clientes se puede determinar cuántos recursos utilizar en ellos, en cuanto a esfuerzos de marketing y tiempo.

Matemáticamente, el CLV de un cliente es:

$$CLV = \sum_{t=0}^T \frac{(p_t - c_t) r_t}{(1 + i)^t} - CA$$

Donde:

- p_t : monto gastado por el cliente en el periodo t .
- c_t : costo asociado a los servicios del cliente en el periodo t .
- i : tasa de descuento.
- r_t : tasa de retención del periodo t .
- CA : costo de adquisición del cliente.
- T : horizonte de tiempo para la estimación del CLV.

Anexo 8: Mapa político de Perú



Fuente: <http://www.veomapas.com/mapa-politico-del-peru-m265.html>

Anexo 9: Matriz de correlación de Pearson entre ventas 2013 de cada categoría

| Categorías | Aceites domésticos | Ade- rezos | Ayudas culinarias | Carame- los | Choco- lates | Cuidado del cabello | Deter- gentes | Exporta- Romex | Galle- tas | Golo- sinas LB | Harinas domésticas | Indus- trias del Espin | Jabo- nes de lavar | Margarinas domésticas | MARS | Mas- cotas | Pana- sonic | Pane- tones | Postres instantáneos | Resfrescos instantáneos | Romero Trading | Sal- sas | Suavizante | Otros | |
|-------------------------|--------------------|------------|-------------------|-------------|--------------|---------------------|---------------|----------------|------------|----------------|--------------------|------------------------|--------------------|-----------------------|-------|------------|-------------|-------------|----------------------|-------------------------|----------------|----------|------------|-------|-------|
| Acetiles domésticos | 1 | 0.22 | 0.09 | 0.05 | 0.13 | 0.00 | 0.46 | 0.12 | 0.20 | 0.06 | 0.31 | 0.11 | 0.34 | 0.35 | -0.01 | 0.15 | 0.28 | 0.08 | 0.50 | 0.39 | 0.31 | 0.08 | 0.31 | 0.01 | -0.01 |
| Aderezos | 0.22 | 1 | 0.31 | 0.04 | 0.03 | -0.01 | 0.21 | 0.34 | 0.26 | 0.05 | 0.28 | 0.24 | 0.13 | 0.26 | 0.18 | 0.20 | 0.22 | 0.00 | 0.14 | 0.17 | 0.16 | 0.16 | 0.36 | 0.01 | 0.02 |
| Ayudas culinarias | 0.09 | 0.31 | 1 | 0.01 | 0.00 | -0.01 | 0.13 | 0.24 | 0.13 | 0.03 | 0.15 | 0.20 | 0.00 | 0.19 | 0.16 | 0.19 | 0.05 | 0.03 | 0.11 | 0.09 | 0.14 | 0.24 | 0.28 | 0.00 | 0.00 |
| Carameles | 0.05 | 0.04 | 0.01 | 1 | 0.22 | 0.02 | 0.06 | 0.01 | 0.21 | 0.18 | 0.04 | 0.05 | 0.06 | 0.04 | 0.06 | 0.04 | 0.11 | 0.14 | 0.04 | 0.06 | 0.06 | 0.02 | 0.03 | 0.02 | 0.02 |
| Chocolates | 0.13 | 0.03 | 0.00 | 0.22 | 1 | 0.00 | 0.15 | 0.01 | 0.15 | 0.10 | 0.06 | 0.03 | 0.14 | 0.11 | 0.02 | 0.05 | 0.14 | 0.17 | 0.10 | 0.09 | 0.10 | 0.03 | 0.05 | 0.01 | 0.00 |
| Cuidado del cabello | 0.00 | -0.01 | -0.01 | 0.02 | 0.00 | 1 | 0.00 | -0.01 | -0.01 | 0.00 | 0.01 | 0.05 | -0.01 | 0.00 | 0.00 | 0.03 | -0.01 | 0.00 | 0.01 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Detergentes | 0.46 | 0.21 | 0.13 | 0.06 | 0.15 | 0.00 | 1 | 0.21 | 0.28 | 0.08 | 0.26 | 0.21 | 0.44 | 0.45 | 0.06 | 0.24 | 0.28 | 0.10 | 0.37 | 0.33 | 0.32 | 0.15 | 0.35 | 0.00 | 0.04 |
| Exportadora | 0.12 | 0.34 | 0.24 | 0.01 | 0.01 | -0.01 | 0.21 | 1 | 0.27 | 0.06 | 0.18 | 0.30 | 0.08 | 0.39 | 0.29 | 0.23 | 0.22 | -0.07 | 0.08 | 0.00 | 0.07 | 0.09 | 0.36 | 0.00 | 0.07 |
| Romex | 0.20 | 0.26 | 0.13 | 0.21 | 0.15 | -0.01 | 0.28 | 0.27 | 1 | 0.21 | 0.16 | 0.23 | 0.20 | 0.27 | 0.26 | 0.14 | 0.33 | 0.10 | 0.15 | 0.17 | 0.19 | 0.10 | 0.29 | 0.02 | 0.13 |
| Galletas | 0.06 | 0.05 | 0.03 | 0.18 | 0.10 | 0.00 | 0.08 | 0.06 | 0.21 | 1 | 0.03 | 0.07 | 0.04 | 0.07 | 0.08 | 0.05 | 0.11 | 0.05 | 0.05 | 0.04 | 0.06 | 0.04 | 0.05 | 0.00 | 0.04 |
| Golosinas LB | 0.31 | 0.28 | 0.15 | 0.04 | 0.06 | 0.01 | 0.26 | 0.18 | 0.16 | 0.03 | 1 | 0.12 | 0.14 | 0.31 | 0.05 | 0.17 | 0.23 | 0.06 | 0.30 | 0.26 | 0.20 | 0.12 | 0.30 | 0.00 | -0.02 |
| Harinas domésticas | 0.11 | 0.24 | 0.20 | 0.05 | 0.03 | 0.05 | 0.21 | 0.30 | 0.23 | 0.07 | 0.12 | 1 | 0.07 | 0.27 | 0.18 | 0.16 | 0.18 | -0.01 | 0.07 | 0.04 | 0.13 | 0.08 | 0.23 | 0.01 | 0.06 |
| Industrias del Espin | 0.34 | 0.13 | 0.00 | 0.06 | 0.14 | -0.01 | 0.44 | 0.08 | 0.20 | 0.04 | 0.14 | 0.07 | 1 | 0.19 | -0.01 | 0.05 | 0.20 | 0.06 | 0.24 | 0.17 | 0.12 | 0.02 | 0.17 | 0.01 | 0.00 |
| Jabones de lavar | 0.35 | 0.26 | 0.19 | 0.04 | 0.11 | 0.00 | 0.45 | 0.39 | 0.27 | 0.07 | 0.31 | 0.27 | 0.19 | 1 | 0.11 | 0.28 | 0.31 | 0.01 | 0.21 | 0.22 | 0.30 | 0.16 | 0.45 | 0.01 | 0.08 |
| Margarinas domésticas | -0.01 | 0.18 | 0.16 | 0.06 | 0.02 | 0.00 | 0.06 | 0.29 | 0.26 | 0.08 | 0.05 | 0.18 | -0.01 | 0.11 | 1 | 0.12 | 0.09 | -0.03 | -0.01 | -0.04 | 0.02 | 0.07 | 0.17 | 0.00 | 0.10 |
| MARS | 0.15 | 0.20 | 0.19 | 0.04 | 0.05 | 0.03 | 0.24 | 0.23 | 0.14 | 0.05 | 0.17 | 0.16 | 0.05 | 0.28 | 0.12 | 1 | 0.14 | 0.01 | 0.09 | 0.11 | 0.16 | 0.18 | 0.24 | 0.00 | 0.05 |
| Mascolitas | 0.28 | 0.22 | 0.05 | 0.11 | 0.14 | -0.01 | 0.28 | 0.22 | 0.33 | 0.11 | 0.23 | 0.18 | 0.20 | 0.31 | 0.09 | 0.14 | 1 | 0.04 | 0.16 | 0.15 | 0.19 | 0.02 | 0.24 | 0.02 | 0.01 |
| Panetonics | 0.08 | 0.00 | 0.03 | 0.14 | 0.17 | 0.00 | 0.10 | -0.07 | 0.10 | 0.05 | 0.06 | -0.01 | 0.06 | 0.01 | -0.03 | 0.01 | 0.04 | 1 | 0.12 | 0.16 | 0.09 | 0.06 | 0.01 | 0.00 | -0.02 |
| Pastas | 0.50 | 0.14 | 0.11 | 0.04 | 0.10 | 0.01 | 0.37 | 0.08 | 0.15 | 0.05 | 0.30 | 0.07 | 0.24 | 0.21 | -0.01 | 0.09 | 0.16 | 0.12 | 1 | 0.38 | 0.28 | 0.04 | 0.24 | 0.00 | -0.04 |
| Postres instantáneos | 0.39 | 0.17 | 0.09 | 0.06 | 0.09 | 0.01 | 0.33 | 0.00 | 0.17 | 0.04 | 0.26 | 0.04 | 0.17 | 0.22 | -0.04 | 0.11 | 0.15 | 0.16 | 0.38 | 1 | 0.38 | 0.11 | 0.24 | 0.00 | -0.03 |
| Resfrescos instantáneos | 0.31 | 0.16 | 0.14 | 0.06 | 0.10 | 0.00 | 0.32 | 0.07 | 0.19 | 0.06 | 0.20 | 0.13 | 0.12 | 0.30 | 0.02 | 0.16 | 0.19 | 0.09 | 0.28 | 0.38 | 1 | 0.18 | 0.28 | 0.01 | -0.01 |
| Romero Trading | 0.08 | 0.16 | 0.24 | 0.02 | 0.03 | 0.00 | 0.15 | 0.09 | 0.10 | 0.04 | 0.12 | 0.08 | 0.02 | 0.16 | 0.07 | 0.18 | 0.02 | 0.06 | 0.04 | 0.11 | 0.18 | 1 | 0.20 | 0.00 | 0.00 |
| Salidas | 0.31 | 0.36 | 0.28 | 0.03 | 0.05 | 0.00 | 0.35 | 0.36 | 0.29 | 0.05 | 0.30 | 0.23 | 0.17 | 0.45 | 0.17 | 0.24 | 0.24 | 0.01 | 0.24 | 0.24 | 0.28 | 0.20 | 1 | -0.01 | 0.03 |
| Suavizante | 0.01 | 0.01 | 0.00 | 0.02 | 0.01 | 0.00 | 0.00 | 0.00 | 0.02 | 0.00 | 0.00 | 0.01 | 0.01 | 0.01 | 0.00 | 0.00 | 0.02 | 0.00 | 0.00 | 0.00 | 0.01 | 0.00 | -0.01 | 1 | 0.00 |
| Otros | -0.01 | 0.02 | 0.00 | 0.02 | 0.00 | 0.00 | 0.04 | 0.07 | 0.13 | 0.04 | -0.02 | 0.06 | 0.00 | 0.08 | 0.10 | 0.05 | 0.01 | -0.02 | -0.04 | -0.03 | -0.01 | 0.00 | 0.03 | 0.00 | 1 |

Anexo 10: Modelo (I) - Regresión lineal efecto cliente

- **Resultados regresión lineal 1**, sin utilizar algoritmo *Stepwise*:

| Variables | Coefficientes | Error std. | t valor | Pr(> t) | Signif. |
|------------------------|---------------|------------|---------|----------|---------|
| Intercepto | 1,01 | 0,08 | 12,88 | < 2e-16 | *** |
| Venta 2013 (Miles S./) | 0,77 | 0,00 | 167,02 | < 2e-16 | *** |
| Días entre compra 2013 | 0,00 | 0,00 | -0,81 | 0,42 | |
| R/F | -0,04 | 0,01 | -3,95 | 0,00 | *** |
| #Categorías | 0,02 | 0,00 | 3,37 | 0,00 | *** |

Códigos de significancia: 0 '***' 0,001 '**' 0,01 '*' 0,05 '.' 0,1 ' ' 1

Tabla 28: Modelo (I) – Efecto cliente: Resultados regresión lineal 1 (parámetros betas)
Fuente: Elaboración propia en base a información de la empresa

| Estadísticos | Efecto cliente |
|-------------------------|----------------|
| R2 | 0,64 |
| R2 ajustado | 0,64 |
| F-estadístico | 1,71 |
| P-valor | < 2,2e-16 |
| Error residual estándar | 9251,00 |

Códigos de significancia: 0 '***' 0,001 '**' 0,01 '*' 0,05 '.' 0,1 ' ' 1

Tabla 29: Modelo (I) – Efecto cliente: Resultados regresión lineal 1 (estadísticos)
Fuente: Elaboración propia en base a información de la empresa

- **Resultados regresión lineal 2**, utilizando algoritmo *Stepwise*:

| Variables | Coefficientes | Error std. | t valor | Pr(> t) | Signif. |
|----------------------|---------------|------------|---------|----------|---------|
| Intercepto | 0,97 | 0,06 | 15,27 | < 2e-16 | *** |
| Venta2013(Miles S./) | 0,77 | 0,00 | 168,37 | < 2e-16 | *** |
| R/F | -0,04 | 0,01 | -3,93 | 8,5E-05 | *** |
| #Categorías | 0,02 | 0,00 | 3,82 | 0,00013 | *** |

Códigos de significancia: 0 '***' 0,001 '**' 0,01 '*' 0,05 '.' 0,1 ' ' 1

Tabla 30: Modelo (I) – Efecto cliente: Resultados regresión lineal 2 (parámetros betas)
Fuente: Elaboración propia en base a información de la empresa

| Estadísticos | Cliente |
|-------------------------|-----------|
| R2 | 0,64 |
| R2 ajustado | 0,64 |
| R2: grupo control | 0,59 |
| F-estadístico | 12.330,00 |
| P-valor | < 2,2e-16 |
| Error residual estándar | 1,71 |

Tabla 31: Modelo (I) – Efecto cliente: Resultados regresión lineal 2 (estadísticos)
Fuente: Elaboración propia en base a información de la empresa

Anexo 11: Modelo (I) - Regresión lineal efecto contexto

- **Resultados regresión lineal 1**, sin utilizar algoritmo *Stepwise*:

| Variables | Coefficientes | Error std. | t valor | Pr(> t) | Signif. |
|------------------|---------------|------------|---------|-----------|---------|
| Intercepto | 4,63 | 0,26 | 17,65 | <2,00E-16 | *** |
| #Cliente (miles) | -0,44 | 0,10 | -4,32 | 1,57E-05 | *** |
| Densidad | -0,02 | 0,01 | -1,50 | 0,13 | |
| Lima | 11,58 | 4,24 | 2,73 | 0,01 | ** |
| Norte | 1,36 | 0,48 | 2,85 | 4,41,E-03 | ** |
| Sur | 1,97 | 0,38 | 5,22 | 1,81E-07 | *** |
| Ancash | 1,56 | 0,49 | 3,20 | 1,39,E-03 | ** |
| Apurímac | -1,13 | 0,21 | -5,34 | 9,53E-08 | *** |
| Arequipa | 0,41 | 0,30 | 1,37 | 0,17 | |
| Ayacucho | 1,31 | 0,34 | 3,89 | 9,92E-05 | *** |
| Cajamarca | 0,88 | 0,45 | 1,96 | 0,05 | * |
| Callao | 153,20 | 101,67 | 1,51 | 0,13 | |
| Cusco | 0,08 | 0,21 | 0,40 | 0,69 | |
| Huancavelica | 2,71 | 0,53 | 5,13 | 3,00E-07 | *** |
| Huanuco | 1,38 | 0,41 | 3,34 | 8,32,E-04 | *** |
| Ica | 2,03 | 0,56 | 3,65 | 2,63,E-04 | *** |
| Junin | 2,54 | 0,48 | 5,30 | 1,17E-07 | *** |
| La Libertad | 0,51 | 0,78 | 0,65 | 0,52 | |
| Lambayeque | 0,98 | 1,00 | 0,98 | 0,33 | |
| Lima | NA | NA | NA | NA | |
| Loreto | 1,71 | 0,89 | 1,92 | 0,05 | , |
| Madre de Dios | -1,32 | 0,38 | -3,44 | 5,81,E-04 | *** |
| Moquegua | -0,13 | 0,32 | -0,40 | 0,69 | |
| Pasco | 1,62 | 0,46 | 3,55 | 3,94,E-04 | *** |
| Piura | 1,28 | 0,53 | 2,44 | 0,01 | * |
| Puno | -1,21 | 0,17 | -6,94 | 4,14E-12 | *** |
| San Martín | 0,55 | 0,34 | 1,64 | 0,10 | |
| Tacna | NA | NA | NA | NA | |
| Tumbres | NA | NA | NA | NA | |
| Ucayali | NA | NA | NA | NA | |

Códigos de significancia: 0 '***' 0,001 '**' 0,01 '*' 0,05 '.' 0,1 ' ' 1

Tabla 32: Modelo (I) – Efecto contexto: Resultados regresión lineal 1 (parámetros betas)

Fuente: Elaboración propia en base a información de la empresa

| Estadísticos | Contexto |
|-------------------------|-----------|
| R2 | 0,02 |
| R2 ajustado | 0,02 |
| F-estadístico | 21,34 |
| P-valor | < 2,2E-16 |
| Error residual estándar | 2,81 |

Tabla 33: Modelo (I) – Efecto contexto: Resultados regresión lineal 1 (estadísticos)
Fuente: Elaboración propia en base a información de la empresa

- **Resultados regresión lineal 2**, utilizando algoritmo *Stepwise*:

| Variables | Coefficientes | Error std. | t valor | Pr(> t) | Signif. |
|------------|---------------|------------|---------|-----------|---------|
| Intercepto | 5,44 | 0,05 | 110,98 | <2,00E-16 | *** |
| Lima | 0,32 | 0,06 | 5,49 | 4,14,E-08 | *** |
| Norte | -0,43 | 0,07 | -6,64 | 3,19,E-11 | *** |
| Sur | -0,22 | 0,06 | -3,66 | 2,51,E-04 | *** |
| Densidad | -4,06,E-05 | 1,71,E-05 | -2,37 | 0,02 | * |

Códigos de significancia: 0 '***' 0,001 '**' 0,01 '*' 0,05 '.' 0,1 ' ' 1

Tabla 34: Modelo (I) – Efecto contexto: Resultados regresión lineal 2 (parámetros betas)

Fuente: Elaboración propia en base a información de la empresa

| Estadísticos | Contexto |
|-------------------------|-----------|
| R2 | 0,01 |
| R2 ajustado | 0,01 |
| R2: grupo control | 0,02 |
| F-estadístico | 53,58 |
| P-valor | < 2,2E-16 |
| Error residual estándar | 2,83 |

Tabla 35: Modelo (I) – Efecto contexto: Resultados regresión lineal 2 (estadísticos)
Fuente: Elaboración propia en base a información de la empresa

Anexo 12: Modelo (I) - Regresión lineal efecto tipo de vendedor

- Resultados regresión lineal, sin y con utilizar el algoritmo *Stepwise* son iguales dado que todas las variables son significativas:

| Variables | Coefficientes | Error std. | t valor | Pr(> t) | Signif. |
|--------------------|---------------|------------|---------|------------|---------|
| Intercepto | 4,49 | 0,05 | 92,66 | < 2,00E-16 | *** |
| #Vendedores Tipo A | 0,40 | 0,03 | 15,90 | < 2,00E-16 | *** |
| #Vendedores Tipo B | 0,31 | 0,03 | 11,96 | < 2,00E-16 | *** |
| #Vendedores Tipo D | 0,21 | 0,02 | 9,43 | < 2,00E-16 | *** |
| #Vendedores Tipo C | 0,17 | 0,03 | 6,87 | 6,55E-12 | *** |

Códigos de significancia: 0 '***' 0,001 '**' 0,01 '*' 0,05 '.' 0,1 '.' 1

Tabla 36: Modelo (I) – Efecto tipo de vendedor: Resultados regresión lineal 1 y 2
Fuente: Elaboración propia en base a información de la empresa

| Estadísticos | Tipo de vendedor |
|-------------------------|------------------|
| R2 | 0,20 |
| R2 ajustado | 0,02 |
| R2: grupo control | 0,15 |
| F-estadístico | 107,90 |
| P-valor | <2,2E-16 |
| Error residual estándar | 2,82 |

Tabla 37: Modelo (I) – Efecto tipo de vendedor: Resultados regresión lineal 1 y 2
(estadísticos)

Fuente: Elaboración propia en base a información de la empresa

Anexo 13: Resultados Modelo (II): Regresión lineal categorías principales

| Variables | Coefficientes | Error std. | t valor | Pr(> t) | Signif. |
|----------------------------|---------------|------------|---------|-----------|---------|
| Intercepto | 1,88 | 0,04 | 41,85 | <2,00E-16 | *** |
| Log Detergentes | 0,23 | 0,01 | 30,59 | <2,00E-16 | *** |
| Log Pastas | 0,23 | 4,77,E-03 | 48,15 | <2,00E-16 | *** |
| Log Margarinas domésticas | 0,12 | 4,86,E-03 | 24,37 | <2,00E-16 | *** |
| Log Postres instantáneos | 0,08 | 4,02,E-03 | 19,49 | <2,00E-16 | *** |
| Log Jabones de lavar | 0,13 | 0,01 | 21,76 | <2,00E-16 | *** |
| Log Harinas domésticas | 0,06 | 4,64,E-03 | 13,46 | <2,00E-16 | *** |
| Log Ayudas culinarias | -0,03 | 4,15,E-03 | -7,01 | 2,47E-12 | *** |
| Log Refrescos instantáneos | 0,02 | 4,58,E-03 | 4,48 | 7,46E-06 | *** |
| Log Caramelos | -0,02 | 4,63,E-03 | -4,31 | 1,62E-05 | *** |
| Log Salsas | 0,01 | 2,12,E-03 | 3,62 | 2,91,E-04 | *** |
| Log Aderezos | 0,01 | 3,53,E-03 | 2,52 | 0,01 | * |
| Log Galletas | -0,01 | 0,01 | -2,52 | 0,01 | * |
| Log Chocolates | 0,01 | 3,78,E-03 | 2,23 | 0,03 | * |

Códigos de significancia: 0 '***' 0,001 '**' 0,01 '*' 0,05 '.' 0,1 ' ' 1

Tabla 38: Resultados regresión lineal categoría aceites
Fuente: Elaboración propia en base a información de la empresa

| Variables | Coefficientes | Error std. | t valor | Pr(> t) | Signif. |
|----------------------------|---------------|------------|---------|-----------|---------|
| Intercepto | 2,30 | 0,04 | 58,31 | < 2e-16 | *** |
| Log Jabones de lavar | 0,35 | 4,85,E-03 | 71,45 | < 2e-16 | *** |
| Log Aceites | 0,19 | 0,01 | 30,58 | < 2e-16 | *** |
| Log Margarinas domésticas | 0,10 | 4,45,E-03 | 21,95 | < 2e-16 | *** |
| Log Refrescos instantáneos | 0,06 | 4,16,E-03 | 13,56 | < 2e-16 | *** |
| Log Pastas | 0,05 | 4,57,E-03 | 11,93 | < 2e-16 | *** |
| Log Postres instantáneos | 0,05 | 3,70,E-03 | 13,00 | < 2e-16 | *** |
| Log Ayudas culinarias | 0,03 | 3,78,E-03 | 8,42 | < 2e-16 | *** |
| Log Harinas domésticas | 0,02 | 4,25,E-03 | 5,30 | 1,15E-07 | *** |
| Log Salsas | 0,02 | 1,93,E-03 | 7,93 | 2,28E-15 | *** |
| Log Galletas | -0,01 | 0,01 | -2,84 | 4,57,E-03 | ** |
| Log Panetones | 0,01 | 2,35,E-03 | 4,17 | 3,01,E-05 | *** |
| Log Aderezos | -0,01 | 3,22,E-03 | -1,78 | 0,08 | , |

Códigos de significancia: 0 '***' 0,001 '**' 0,01 '*' 0,05 '.' 0,1 ' ' 1

Tabla 39: Resultados regresión lineal categoría detergentes
Fuente: Elaboración propia en base a información de la empresa

| Variables | Coefficientes | Error std. | t valor | Pr(> t) | Signif. |
|----------------------------|---------------|------------|---------|----------|---------|
| Intercepto | 4,5536 | 0,0464 | 98,1550 | < 2e-16 | *** |
| Log Caramelos | 0,1440 | 0,0055 | 26,0130 | < 2e-16 | *** |
| Log Aderezos | 0,0982 | 0,0042 | 23,3500 | < 2e-16 | *** |
| Log Refrescos instantáneos | 0,0784 | 0,0055 | 14,2390 | < 2e-16 | *** |
| Log Jabones de lavar | 0,0727 | 0,0071 | 10,1720 | < 2e-16 | *** |
| Log Margarinas domésticas | 0,0618 | 0,0059 | 10,4200 | < 2e-16 | *** |
| Log Postres instantáneos | 0,0522 | 0,0049 | 10,7090 | < 2e-16 | *** |
| Log Chocolates | 0,0463 | 0,0046 | 9,9730 | < 2e-16 | *** |
| Log Detergentes | -0,0266 | 0,0091 | -2,9110 | 0,0036 | ** |
| Log Panetones | 0,0181 | 0,0032 | 5,5950 | 0,0000 | *** |
| Log Aceites | -0,0179 | 0,0079 | -2,2580 | 0,0239 | * |
| Log Harinas domésticas | 0,0161 | 0,0055 | 2,9160 | 0,0035 | ** |
| Log Ayudas culinarias | 0,0147 | 0,0050 | 2,9410 | 0,0033 | ** |
| Log Salsas | 0,0112 | 0,0026 | 4,3840 | 0,0000 | *** |

Códigos de significancia: 0 '***' 0,001 '**' 0,01 '*' 0,05 '.' 0,1 ' ' 1

Tabla 40: Resultados regresión lineal categoría galletas
Fuente: Elaboración propia en base a información de la empresa

| Variables | Coefficientes | Error std. | t valor | Pr(> t) | Signif. |
|----------------------------|---------------|------------|---------|-----------|---------|
| Intercepto | 1,12 | 0,05 | 21,13 | < 2e-16 | *** |
| Log Aceites | 0,43 | 0,01 | 48,18 | < 2e-16 | *** |
| Log Harinas domésticas | 0,18 | 0,01 | 28,29 | < 2e-16 | *** |
| Log Detergentes | 0,12 | 0,01 | 11,94 | < 2e-16 | *** |
| Log Postres instantáneos | 0,09 | 0,01 | 16,54 | < 2e-16 | *** |
| Log Jabones de lavar | 0,06 | 0,01 | 7,06 | 1,67,E-12 | *** |
| Log Margarinas domésticas | -0,05 | 0,01 | -7,54 | 5,07,E-14 | *** |
| Log Refrescos instantáneos | 0,04 | 0,01 | 5,81 | 6,24,E-09 | *** |
| Log Chocolates | -0,02 | 0,01 | -4,68 | 2,89,E-06 | *** |
| Log Ayudas culinarias | 0,02 | 0,01 | 3,77 | 1,67,E-04 | *** |
| Log Aderezos | -0,02 | 4,79,E-03 | -3,91 | 9,23,E-05 | *** |
| Log Panetones | 0,02 | 3,64,E-03 | 4,44 | 8,87,E-06 | *** |
| Log Salsas | 0,02 | 2,91,E-03 | 5,34 | 9,52,E-08 | *** |

Códigos de significancia: 0 '***' 0,001 '**' 0,01 '*' 0,05 '.' 0,1 ' ' 1

Tabla 41: Resultados regresión lineal categoría pastas
Fuente: Elaboración propia en base a información de la empresa

| Variables | Coefficientes | Error std. | t valor | Pr(> t) | Signf. |
|----------------------------|---------------|------------|---------|-----------|--------|
| Intercepto | 0,75 | 0,15 | 4,96 | 7,30,E-07 | *** |
| Log Detergentes | 0,19 | 0,02 | 7,89 | 3,23,E-15 | *** |
| Log Margarinas domésticas | 0,14 | 0,02 | 9,01 | < 2e-16 | *** |
| Log Harinas domésticas | 0,10 | 0,02 | 6,71 | 2,06,E-11 | *** |
| Log Ayudas culinarias | 0,10 | 0,01 | 7,35 | 2,05,E-13 | *** |
| Log Pastas | 0,09 | 0,02 | 5,29 | 1,23,E-07 | *** |
| Log Aceites | 0,08 | 0,02 | 3,55 | 3,93,E-04 | *** |
| Log Galletas | 0,08 | 0,02 | 4,18 | 2,88,E-05 | *** |
| Log Jabones de lavar | -0,05 | 0,02 | -2,35 | 0,02 | * |
| Log Aderezos | 0,04 | 0,01 | 3,70 | 2,18,E-04 | *** |
| Log Refrescos instantáneos | 0,04 | 0,01 | 2,79 | 0,01 | ** |
| Log Panetones | -0,03 | 0,01 | -3,34 | 8,31,E-04 | *** |

Códigos de significancia: 0 '***' 0,001 '**' 0,01 '*' 0,05 '.' 0,1 '.' 1

Tabla 42: Resultados regresión lineal categoría salsas
Fuente: Elaboración propia en base a información de la empresa

| Categorías | Aceites | Detergentes | Galletas | Pastas | Salsas |
|------------|-----------|-------------|----------|-----------|---------|
| Min | 4,3 | 7,8 | 71,0 | 1,4 | 1,2 |
| Max | 3.342,7 | 3.405,5 | 1.692,8 | 2.781,7 | 561,9 |
| Media | 958,4 | 901,3 | 430,0 | 633,4 | 136,3 |
| Desv. Típ. | 526,9 | 472,1 | 181,7 | 364,2 | 73,8 |
| Varianza | 277.644,7 | 222.845,4 | 33.016,9 | 132.666,7 | 5.449,7 |

Tabla 43: Residuos de las regresiones lineales de cada categoría
Fuente: Elaboración propia en base a información de la empresa

Anexo 14: Modelo (I) - Resultados regresión de fronteras estocásticas

| Variables | Coefficientes | Std Error | z value | Pr(> z) | Signif. |
|------------------------|---------------|-----------|---------|-----------|---------|
| Intercepto | 1,16 | 0,14 | 8,08 | 6,68,E-16 | *** |
| Venta 2013 (miles S/.) | 0,78 | 0,00 | 211,81 | <2,20E-16 | *** |
| R/F 2013 | -0,04 | 0,01 | -4,81 | 1,54,E-06 | *** |
| Lima | -0,03 | 0,03 | -1,00 | 0,32 | |
| Norte | -0,21 | 0,04 | -5,74 | 9,59,E-09 | *** |
| Sur | -0,13 | 0,03 | -3,92 | 8,92,E-05 | *** |
| #Vendedores Tipo A | 0,06 | 0,01 | 4,25 | 2,09,E-05 | *** |
| #Vendedores Tipo B | 0,07 | 0,02 | 4,34 | 1,45,E-05 | *** |
| #Vendedores Tipo C | 0,06 | 0,01 | 3,92 | 8,84,E-05 | *** |
| #Vendedores Tipo D | 4,49,E-03 | 0,01 | 0,37 | 0,71 | *** |
| gammaVar | 0,00003 | NA | NA | NA | |

Códigos de significancia: 0 '***' 0,001 '**' 0,01 '*' 0,05 '.' 0,1 '.' 1

Tabla 44: Modelo (I): 3 efectos – Resultados regresión de frontera
Fuente: Elaboración propia en base a información de la empresa

Anexo 15: Selección de modelos de regresión de frontera para detergentes, galletas y pastas

- **Categoría detergentes:** el modelo seleccionado es el de 11 variables:

| # | 9 categorías | 10 categorías | 11 categorías | 12 categorías |
|----|-------------------|-------------------|-------------------|-------------------|
| 1 | Jabones | Jabones | Jabones | Jabones |
| 2 | Aceites | Aceites | Aceites | Aceites |
| 3 | Margarinas | Margarinas | Margarinas | Margarinas |
| 4 | Pastas | Pastas | Pastas | Pastas |
| 5 | Postres inst. | Postres inst. | Postres inst. | Postres inst. |
| 6 | Harinas | Harinas | Harinas | Harinas |
| 7 | Salsas | Salsas | Salsas | Salsas |
| 8 | Refrescos | Refrescos | Refrescos | Refrescos |
| 9 | Ayudas culinarias | Ayudas culinarias | Ayudas culinarias | Ayudas culinarias |
| 10 | | Panetones | Panetones | Panetones |
| 11 | | | Galletas | Galletas |
| 12 | | | | Aderezos |

Tabla 45: Categoría: Detergentes – Variables incluidas en los modelos de regresión de fronteras

Fuente: Elaboración propia en base a información de la empresa

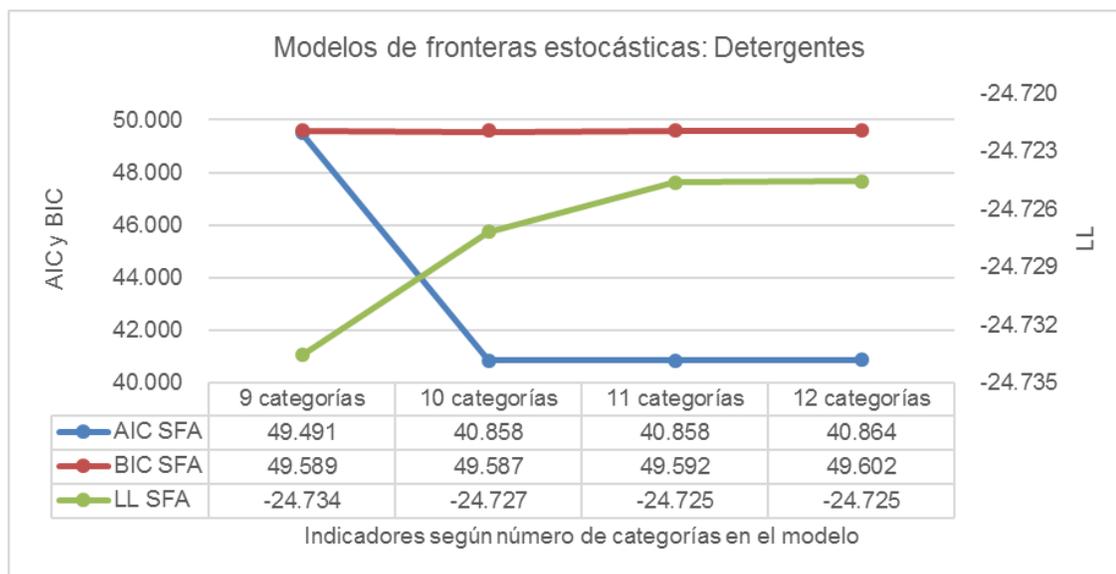


Gráfico 21: Categoría: Detergentes – Estadísticos principales de los modelos de regresión de fronteras según número de variables incluidas

Fuente: Elaboración propia en base a información de la empresa

- **Categoría galletas:** el modelo seleccionado es el de 12.2 variables:

| # | 11 categorías | 12.1 categorías | 12.2 categorías | 13 categorías |
|----|-----------------|-----------------|-------------------|-------------------|
| 1 | Jabones | Jabones | Jabones | Jabones |
| 2 | Margarinas | Margarinas | Margarinas | Margarinas |
| 3 | Postres inst. | Postres inst. | Postres inst. | Postres inst. |
| 4 | Detergentes | Detergentes | Detergentes | Detergentes |
| 5 | Aceites | Aceites | Aceites | Aceites |
| 6 | Salsas | Harinas | Salsas | Harinas |
| 7 | Caramelos | Salsas | Caramelos | Salsas |
| 8 | Aderezos | Caramelos | Aderezos | Caramelos |
| 9 | Refrescos inst. | Aderezos | Refrescos inst. | Aderezos |
| 10 | Chocolates | Refrescos inst. | Chocolates | Refrescos inst. |
| 11 | Panetones | Chocolates | Panetones | Chocolates |
| 12 | | Panetones | Ayudas culinarias | Panetones |
| 13 | | | | Ayudas culinarias |

Tabla 46: Categoría: Galletas – Variables incluidas en los modelos de regresión de fronteras de fronteras

Fuente: Elaboración propia en base a información de la empresa

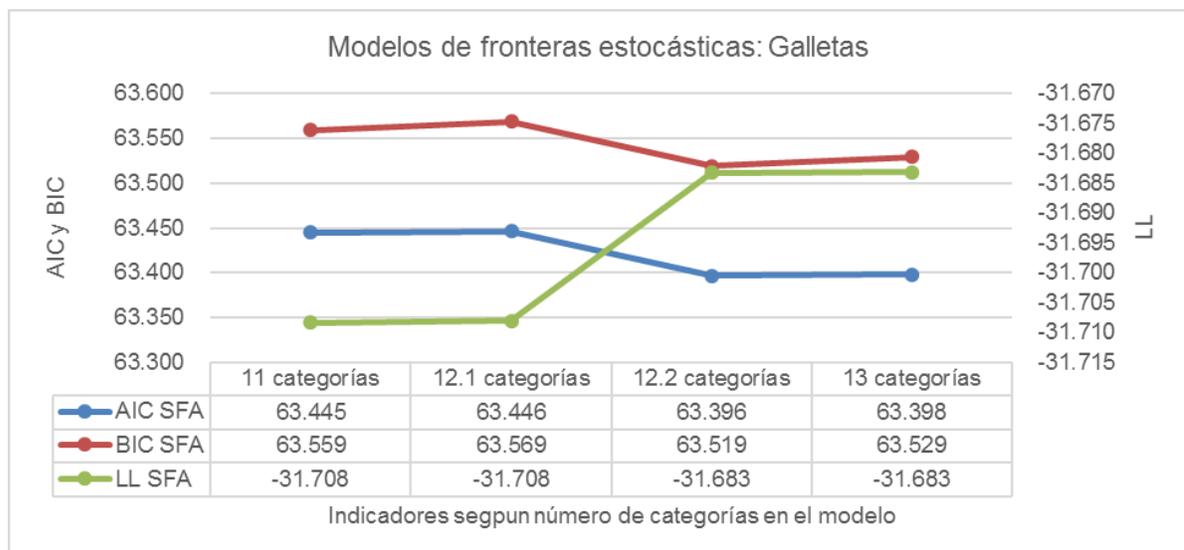


Gráfico 22: Categoría: Galletas – Estadísticos principales de los modelos de regresión de fronteras según número de variables incluidas

Fuente: Elaboración propia en base a información de la empresa

- **Categoría pastas:** el modelo seleccionado es el de 12 variables:

| # | 7 categorías | 11 categorías | 12 categorías |
|----|---------------|-------------------|-------------------|
| 1 | Aceites | Aceites | Aceites |
| 2 | Harinas | Harinas | Harinas |
| 3 | Detergentes | Detergentes | Detergentes |
| 4 | Postres inst. | Postres inst. | Postres inst. |
| 5 | Jabones | Jabones | Jabones |
| 6 | Margarinas | Margarinas | Margarinas |
| 7 | Salsas | Salsas | Salsas |
| 8 | | Refrescos inst. | Refrescos inst. |
| 9 | | Chocolates | Chocolates |
| 10 | | Ayudas culinarias | Ayudas culinarias |
| 11 | | Aderezos | Aderezos |
| 12 | | | Panetones |

Tabla 47: Categoría: Pastas – Variables incluidas en los modelos de regresión de fronteras

Fuente: Elaboración propia en base a información de la empresa

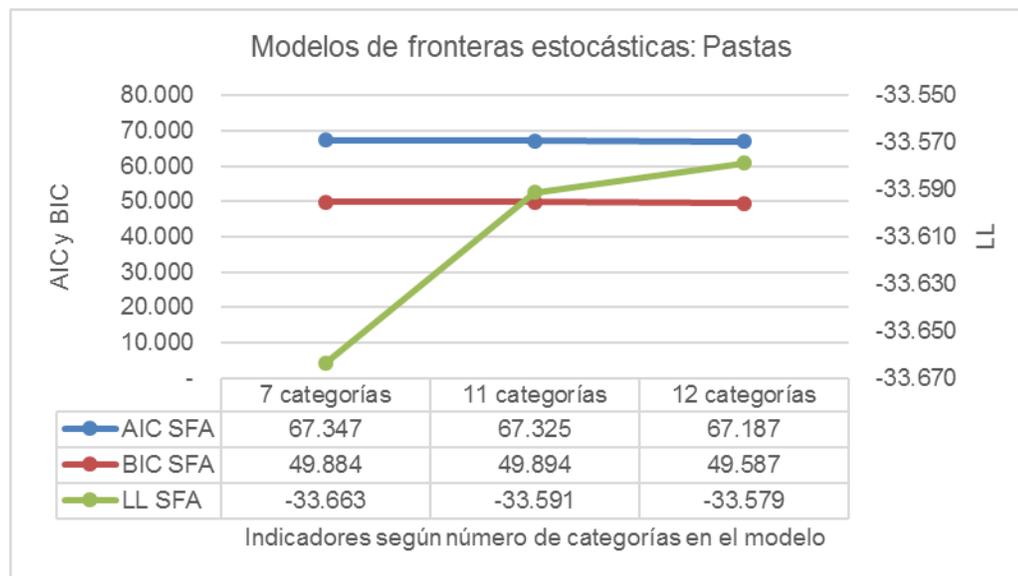


Gráfico 23: Categoría: Pastas – Estadísticos principales de los modelos de regresión de fronteras según número de variables incluidas

Fuente: Elaboración propia en base a información de la empresa

Anexo 16: Ineficiencia estocástica en cada categoría del modelo de fronteras

| Ineficiencia estocástica | Aceites | Detergent | Galletas | Pastas |
|--------------------------|---------|-----------|----------|--------|
| Min. | 0,05 | 0,05 | 0,07 | 0,04 |
| 1er Q | 0,42 | 0,38 | 0,49 | 0,54 |
| Mediana | 0,58 | 0,51 | 0,73 | 0,77 |
| Promedio | 0,67 | 0,61 | 0,87 | 0,96 |
| 3er Q | 0,78 | 0,73 | 1,07 | 1,12 |
| Max. | 6,35 | 5,87 | 6,02 | 7,14 |

Tabla 48: Ineficiencia estocástica en cada Modelo (II) de fronteras
Fuente: Elaboración propia en base a información de la empresa

Anexo 17: Resultados complementarios

Hasta el momento, no se había incluido el efecto de las variables obtenidas de la encuesta realizada por la empresa en los modelos realizados anteriormente, pues se cuenta con aquella información para 1.438 clientes, menos del 5,92% del total. Entonces, como resultado complementario, se evaluarán las variables seleccionadas en el modelo (I) de los tres efectos (cliente, contexto y tipo de vendedor), incluyendo superficie disponible de la bodega (m^2) y las siguientes variables *dummys*:

- Si la bodega está ubicada en una avenida principal (valor 1 y 0 en caso contrario).
- Si la bodega pertenece a una zona comercial/industrial (valor 1) o residencial (valor 0).

En primer lugar, se realiza un análisis descriptivo de las variables en la **Tabla 49** con la muestra de clientes reducida:

| Variable | Mínimo | Máximo | Rango | Media | Desv. típ. | Varianza |
|--------------------------------|-------------|------------|---------------|--------------|--------------|---------------|
| Venta (miles S/.) 2014 | 1,23 | 13,31 | 12,07 | 6,62 | 2,91 | 8,50 |
| Venta (miles S/.) 2013 | 1,21 | 13,97 | 12,76 | 6,49 | 2,94 | 8,63 |
| Transacciones 2013 | 9 | 330 | 321 | 119,35 | 49,33 | 2.433,80 |
| Días distintos con compra 2013 | 3 | 113 | 110 | 64,77 | 22,26 | 495,73 |
| Recencia 2013 | - | 262 | 262 | 5,33 | 10,97 | 120,35 |
| Frecuencia mensual 2013 | 0,25 | 9,42 | 9,17 | 5,40 | 1,86 | 3,44 |
| Frecuencia diaria 2013 | 0,01 | 0,31 | 0,30 | 0,18 | 0,06 | 0,00 |
| R/F 2013 | - | 24,61 | 24,61 | 0,80 | 1,25 | 1,56 |
| # categorías 2013 | 3 | 26 | 23 | 18,28 | 2,90 | 8,42 |
| # productos 2013 | 16 | 270 | 254 | 160,46 | 42,54 | 1.809,47 |
| # marcas 2013 | 4 | 44 | 40 | 30,58 | 4,87 | 23,72 |
| Tamaño bodega (m2) | 1,25 | 208 | 206,75 | 29,65 | 26,62 | 708,75 |
| #Vendedores Tipo A | - | 3 | 3 | 1,26 | 0,72 | 0,51 |
| #Vendedores Tipo B | - | 3 | 3 | 0,76 | 0,78 | 0,60 |
| #Vendedores Tipo C | - | 3 | 3 | 0,88 | 0,87 | 0,75 |
| #Vendedores Tipo D | - | 7 | 7 | 0,65 | 1,10 | 1,22 |

Tabla 49: Análisis descriptivo de las principales variables transaccionales, incluyendo la superficie de la bodega

Fuente: Elaboración propia en base a información de la empresa

Lo anterior, se puede comparar con la **Tabla 1** que resume los estadísticos descriptivos de las mismas variables de la muestra completa de clientes. La principal diferencia es que, en la muestra de clientes encuestados, las medias de las variables son mayores en

ventas, transacciones, días con compra, frecuencia de compra, número de categorías, productos y marcas compradas, sin embargo, la media de recencia y R/F disminuyen, lo que implica que son clientes seleccionados intencionalmente por la empresa para realizar una encuesta de prueba, lo cual no es representativo para el total de los clientes que tienen un rendimiento menor.

Luego, se realiza la regresión lineal con las variables mencionadas de la encuesta, con el fin de entender el impacto que tienen en las ventas del año 2014. Cabe mencionar que, se elimina la variable zona dado que todas las bodegas encuestadas son de Lima y que se aplica el mismo procedimiento utilizado para las regresiones lineales anteriores. Por lo tanto, la ecuación del modelo con las variables que se incluyen en esta sección es:

$$\begin{aligned}
 Ventas_{2014,i} = & \beta_0 + \beta_1 Ventas_{2013,i} + \beta_2 R/F_{2013,i} \\
 & + \beta_3 \text{Número de categorías distintas compradas}_{2013,i} \\
 & + \beta_6 \text{Tipo } A_i + \beta_7 \text{Tipo } B_i + \beta_8 \text{Tipo } C_i + \beta_9 \text{Tipo } D_i \\
 & + \beta_{10} \text{Superficie} + \beta_{11} \text{ZonaResidencial} + \beta_{12} \text{AvenidaPrincipal} + \varepsilon_i
 \end{aligned}$$

Los resultados de la regresión lineal se presentan en la **Tabla 50**, que no emplea el algoritmo *Stepwise*, y muestra que no es significativo el efecto del tipo de vendedor, ni si pertenece a una zona residencial o si está en una avenida principal, incluso afectarían negativamente a las ventas 2014 si la bodega está en una avenida principal, pero positivamente si está en una zona comercial o industrial.

| Variables | Coefficientes | Error std. | t valor | Pr(> t) | Signif. |
|------------------------|---------------|------------|----------|-----------|---------|
| Intercepto | 1,210 | 0,17 | 6951,00 | 6,10,E-12 | *** |
| Venta 2013 (miles S/.) | 0,819 | 0,02 | 47414,00 | < 2e-16 | *** |
| R/F 2013 | -0,078 | 0,04 | -2167,00 | 3,04,E-02 | * |
| Superficie (m2) | 0,003 | 0,00 | 1709,00 | 0,09 | , |
| Avenida Principal | -0,001 | 0,10 | -0,01 | 9,95,E-01 | |
| Zona residencial | 0,188 | 0,12 | 1510,00 | 1,31,E-01 | |
| #Vendedores Tipo A | -0,043 | 0,07 | -0,58 | 5,62,E-01 | |
| #Vendedores Tipo B | 0,029 | 0,07 | 0,41 | 6,82,E-01 | |
| #Vendedores Tipo C | 0,043 | 0,06 | 0,71 | 4,75,E-01 | |
| #Vendedores Tipo D | 3,43,E-02 | 0,05 | 0,74 | 0,46 | |

Códigos de significancia: 0 '***' 0,001 '**' 0,01 '*' 0,05 ',' 0,1 ' ' 1

Tabla 50: Resultados modelo de regresión lineal 1 con variables de la encuesta
Fuente: Elaboración propia en base a información de la empresa

Se destaca que la superficie de la bodega es significativa a un 90% de confianza, pero la magnitud del coeficiente es muy pequeña en comparación al parámetro beta de las ventas del año 2013 y el intercepto, lo cuales son los valores más grandes. Esto, se respalda con los resultados de la regresión lineal de los 3 efectos, donde las ventas del año 2013 es la variable con mayor importancia para estimar las ventas del siguiente año.

Dado que el resultado de la regresión anterior tiene muchas variables no significativas para el modelo, se utiliza el algoritmo *Stepwise*. Estos resultados, se muestran en la **Tabla 51**:

| Variables | Coeficientes | Error std. | t valor | Pr(> t) | Signif. |
|------------------------|--------------|------------|----------|-----------|---------|
| Intercepto | 1,22 | 0,13 | 9577,00 | <2e-16 | *** |
| Venta 2013 (miles S/.) | 0,82 | 0,02 | 50574,00 | <2e-16 | *** |
| R/F 2013 | -0,08 | 0,04 | -2230,00 | 2,59,E-02 | * |
| Superficie (m2) | 0,00 | 0,00 | 1646,00 | 0,10 | |
| Zona residencial | 0,20 | 0,12 | 1616,00 | 1,07,E-01 | |

Códigos de significancia: 0 '***' 0,001 '**' 0,01 '*' 0,05 ',' 0,1 '' 1

Tabla 51: Resultados modelo de regresión lineal 2 con variables de la encuesta

Fuente: Elaboración propia en base a información de la empresa

De acuerdo a los resultados y que el *skewness* del modelo es 0,59 para realizar una regresión de fronteras estocástica, se procede a no considerar, por el momento, profundizar en el análisis con las variables de la encuesta, dado que se elimina la variable de ubicarse en una avenida principal y que los coeficientes de superficie y zona residencial o comercial/industrial no son significativos para el modelo. Intuitivamente, las variables de la encuesta son relevantes para estimar las ventas de la bodega por lo que se recomienda continuar con esta investigación como trabajo futuro.

Anexo 18: Test Chi-Cuadrado de clientes por zona y cuadrantes

| Zona | Latentes | Preferentes | No prioritarios | Objetivos |
|--------|----------|-------------|-----------------|-----------|
| Centro | 18,72% | 16,80% | 20,20% | 19,31% |
| Lima | 35,08% | 26,00% | 29,78% | 36,40% |
| Norte | 21,43% | 33,61% | 19,98% | 11,77% |
| Sur | 24,78% | 23,60% | 30,04% | 32,52% |

Tabla 52: Porcentaje de clientes por zona y cuadrantes
Fuente: Elaboración propia en base a información de la empresa

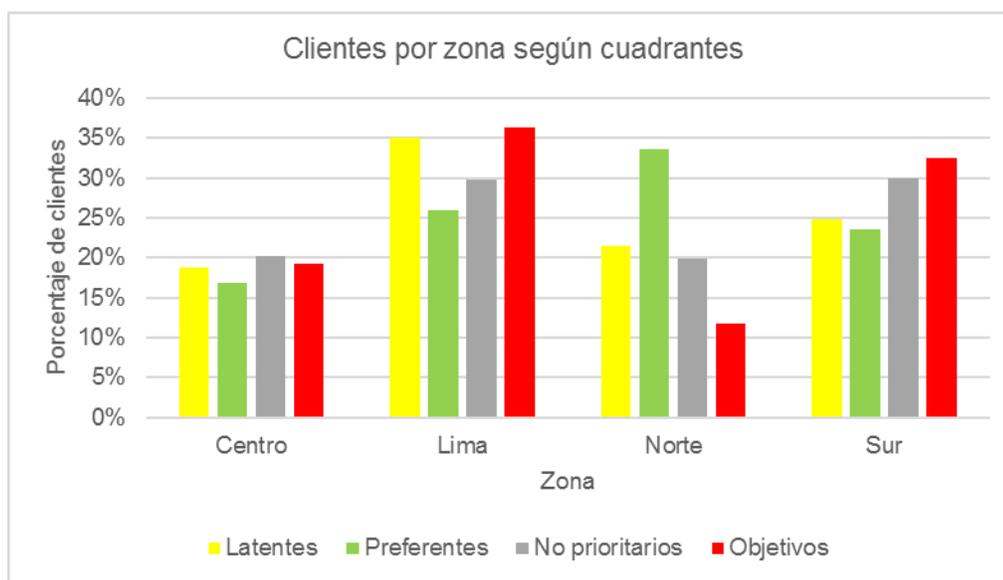


Gráfico 24: Porcentaje de clientes por zona y cuadrantes
Fuente: Elaboración propia en base a información de la empresa

| Pruebas de chi-cuadrado | Valor | Grados de libertad | Sig. asintótica (bilateral) |
|--------------------------|----------|--------------------|-----------------------------|
| Chi-cuadrado de Pearson | 476,197* | 9 | 0,00 |
| Razón de verosimilitudes | 472,42 | 9 | 0,00 |
| Número de casos válidos | 26.156 | | |

* 0 casillas (0,0%) tienen una frecuencia esperada inferior a 5.
La frecuencia mínima esperada es 378,93.

Tabla 53: Resultados Test Chi-Cuadrado de clientes por zona y cuadrantes
Fuente: Elaboración propia en base a información de la empresa

Anexo 19: Test Chi-Cuadrado de vendedores por tipo y cuadrantes

| Tipo de vendedor | Latentes | Preferentes | No prioritarios | Objetivos |
|------------------|----------|-------------|-----------------|-----------|
| Tipo A | 43,31% | 46,22% | 43,90% | 44,01% |
| Tipo B | 20,34% | 17,70% | 21,26% | 21,84% |
| Tipo C | 17,83% | 17,95% | 19,64% | 18,64% |
| Tipo D | 18,52% | 18,13% | 15,20% | 15,51% |

Tabla 54: Porcentaje de vendedores por tipo y cuadrantes
Fuente: Elaboración propia en base a información de la empresa

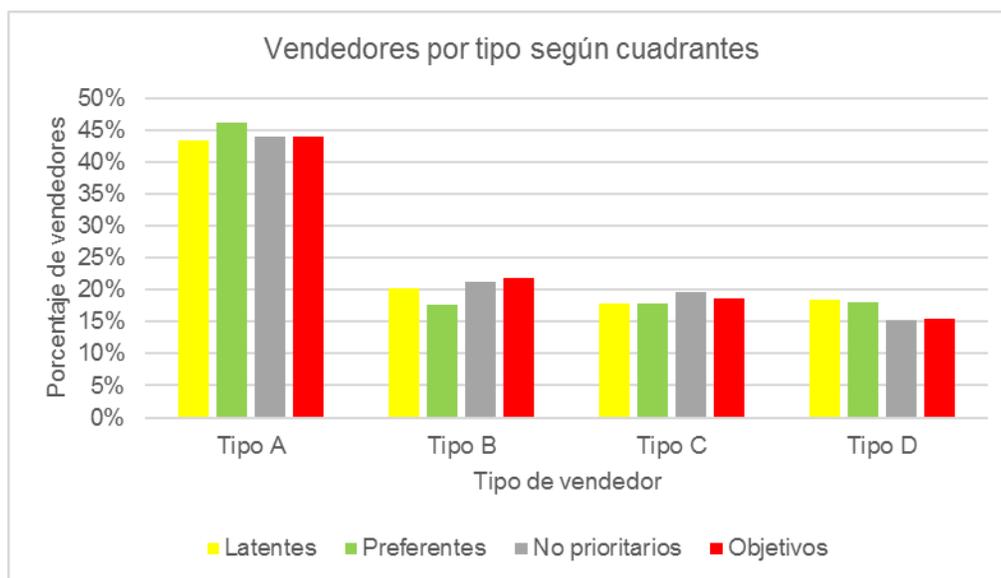


Gráfico 25: Porcentaje de vendedores por tipo y cuadrantes
Fuente: Elaboración propia en base a información de la empresa

| Pruebas de chi-cuadrado | Valor | Grados de libertad | Sig. asintótica (bilateral) |
|--------------------------|---------|--------------------|-----------------------------|
| Chi-cuadrado de Pearson | 95,524* | 9 | 0,00 |
| Razón de verosimilitudes | 97,75 | 9 | 0,00 |
| Número de casos válidos | 26.156 | | |

* 0 casillas (0,0%) tienen una frecuencia esperada inferior a 5.
La frecuencia mínima esperada es 147,62.

Tabla 55: Resultados Test Chi-Cuadrado de clientes por zona y cuadrantes
Fuente: Elaboración propia en base a información de la empresa