

Propuesta de un modelo de predicción de canibalizaciones de las referencias de línea en presencia de ofertas para una empresa manufacturera de alimentos de la ciudad de Medellín

Daniela Flórez Jaramillo

Trabajo de profundización para optar al título de Magíster en Ingeniería

Asesor:

Carlos Alberto Castro Zuluaga

Máster en Ingeniería Industrial

Escuela de Ingeniería

Universidad EAFIT

Medellín, 2020

Tabla de contenido

	Pág.
Introducción	7
1. Descripción y formulación del problema.....	9
2. Objetivos	11
2.1 Objetivo General.....	11
2.2 Objetivos específicos.....	11
3. Marco Teórico	12
3.1 Antecedentes de investigaciones referidas a la aplicación del modelo de canibalizaciones	12
3.2 Aproximación teórica al concepto de canibalización.....	16
3.2.1 Tipos de promociones que sientan la base del proceso de canibalización.	19
3.3 Gestión de la demanda	19
3.3.1 Naturaleza de los datos en modelos cuantitativos para la gestión de demanda	21
3.4 Aprendizaje automático o Machine Learning.....	23
4. Metodología	25
4.1 Paradigma de investigación	25
4.2 Enfoque de investigación.....	25
4.3 Tipo de estudio.....	26
4.4 Fuentes de información.....	26
5. Modelo de proyección de canibalización.....	28
5.1 Definición de variables.....	29
5.2 Obtención y limpieza de la información – variable canibalización	31

5.2.1 Cálculo del patrón de comportamiento normal de la referencia de línea.....	31
5.2.1.1 Detección y limpieza de datos atípicos.....	34
5.2.1.2 Pronóstico con análisis de series de tiempo.....	37
5.4 Evaluación de modelos para la predicción de canibalización.....	47
5.4.1 Regresión Lineal Múltiple	49
5.4.2 Árbol de Decisión Regresión.....	50
5.5 Conexión del modelo de canibalización con el proceso planear ventas y operaciones (PVO) de la compañía	54
5.6 Validación de la eficiencia del modelo de canibalización propuesto.....	57
5.7 Percepción de expertos frente al modelo de canibalización propuesto	60
6. Conclusiones.....	64
7. Recomendaciones	66
Referencias bibliográficas.....	67

Lista de figuras

	Pág.
Figura 1. Etapas del modelo de canibalización propuesto.....	28
Figura 2. Ejemplo a) venta referencia de línea en presencia de oferta y b) estimación del patrón de comportamiento normal de la referencia de línea.....	32
Figura 3. Combinaciones referencia, cliente, cedi.....	33
Figura 4. Coeficiente de variación ventas referencias de línea sin eventos promocionales a detalle cliente, cedi.	34
Figura 5. Datos normalizados referencia, cliente, cedi, semana.....	35
Figura 6. Detección de atípicos para una combinación referencia, cliente, cedi.....	36
Figura 7. Atípicos suavizados para una combinación referencia, cliente, cedi.	36
Figura 8. Coeficiente de canibalización por evento.	42
Figura 9. Mapa de calor entre variables evaluadas del modelo de canibalización.....	43
Figura 10. Variables con mayor correlación con la variable canibalización.	44
Figura 11. Coeficiente de canibalización según el “Cliente”.....	46
Figura 12. Coeficiente de canibalización según el “Tipo de Oferta”.....	46
Figura 13. Coeficiente de canibalización según el “Tipo de evento”.....	47
Figura 14. Procesos ejecutados para cargue de modelos en Jupyter.....	48
Figura 15. Medidas de rendimiento Regresión Lineal Múltiple para la predicción de la canibalización.	50
Figura 16. Árbol de decisión regresión para la predicción de la canibalización.	51
Figura 17. Importancias variables independientes en modelo de Árbol de decisión regresión.....	52

Figura 18. Medidas de rendimiento Árbol de Decisión Regresión para la predicción de la canibalización.	53
Figura 19. Rangos de planeaciones del ciclo PVO.....	54

Lista de tablas

	Pág.
Tabla 1. Tipos de canibalización	17
Tabla 2. Variables iniciales a incluir para la estimación de la canibalización	30
Tabla 3. Unidades dejadas de vender referencia de línea en presencia de una oferta	38
Tabla 4. Coeficiente de canibalización.....	38
Tabla 5. Ejemplo coeficiente de canibalización calculado por semana.	41
Tabla 6. Ejemplo coeficiente de canibalización calculado por evento.....	42
Tabla 7. Resultados medidas de desempeño modelos de predicción de la canibalización	53
Tabla 8. Coeficiente de canibalización estimado por equipo comercial	57
Tabla 9. Coeficiente de canibalización estimado por el modelo	58
Tabla 10. Coeficiente de canibalización real	59
Tabla 11. Comparación del MAD de los coeficientes de canibalización estimados vs los reales	59
Tabla 12. Comparación del MAD de los coeficientes de canibalización estimados vs los reales	60

Introducción

En los últimos años ha cobrado gran importancia la ciencia de datos, en la medida que, esta posibilita que las empresas accedan, recopilen, limpien y almacenen información clave y relevante para el desarrollo de procesos de toma de decisiones a niveles estratégicos, tácticos y operativos, Una de las aplicaciones actuales de la ciencia de datos se sitúa en la predicción de la demanda, elemento que puede estimarse a través de diversos métodos estadísticos y técnicas enfocadas en el tiempo. Sin embargo, este proceso no es lineal ni absoluto y depende de las características de los productos que brinda la empresa. Por otra parte, la forma en la cual se manejan las promociones, puede acarrear variaciones en la cantidad de ventas de artículos de referencia, proceso que se conoce bajo el nombre de canibalización.

En consecuencia, para evitar la ocurrencia de errores demasiado elevados en la estimación de la demanda de los artículos de línea ocasionados por las promociones, las compañías deben contar con herramientas que les permitan generar el mejor estimado de la canibalización, es decir, de la disminución de las ventas de la referencia de línea cada que en simultánea se presentan eventos promocionales. En consonancia con esta premisa, el presente trabajo se desarrolla con el objetivo de proponer un modelo de predicción de canibalizaciones que permita estimar de manera más asertiva las ventas de las referencias de línea cuando se tiene presencia de una oferta, tomando como punto de partida la realidad y contexto de una empresa manufacturera de alimentos de la ciudad de Medellín.

Para la realización de este estudio se utilizarán métodos de aprendizaje automático con algoritmos supervisados mediante la aplicación de modelos de regresión, siendo esta una metodología pertinente para el manejo y tratamiento de la cantidad de datos proporcionados por la

empresa. De igual modo, el desarrollo del modelo se apoya en el lenguaje de programación Python, el cual tiene un uso frecuente en Machine Learning.

Para el diseño del modelo de predicción de canibalizaciones, en primer lugar, se identifican las variables que afectan la demanda de los artículos de línea cuando existen ofertas o promociones. Seguidamente, se obtiene, limpia, estructura y analiza la información existente de las variables definidas para así evaluar algunos métodos de predicción de acuerdo con medidas de rendimiento adecuadas para este propósito. Finalmente, se integra el resultado del modelo de predicción con la planeación de ventas y operaciones de la empresa, para lo cual se valida el resultado del modelo en comparación con lo propuesto por el equipo comercial.

Se espera que con los resultados de la investigación se logren obtener proyecciones que permitan mejorar el coeficiente de canibalización, lo que, por ende, permita a la empresa hacer estimaciones más acertadas de la demanda, para de este modo evitar pérdidas en los productos y tener una mejor gestión de inventarios que reduzca el número de agotados en los pedidos.

1. Descripción y formulación del problema

El dinamismo actual del mercado lleva a las organizaciones a flexibilizar sus procesos buscando atender oportunamente los cambios en los gustos y preferencias del consumidor. (Esmailikia, et. al, 2016). Para ello es importante que exista una excelente planeación y gestión de la cadena de suministro donde operen de forma sincronizada las ventas y las operaciones con una mirada enfocada en el cliente (Madhani, 2017).

El insumo de partida para llevar a cabo dicha planeación es la estimación de la demanda, la cual se convierte en un desafío debido a la incertidumbre y variabilidad que en ocasiones se presenta en los consumos (Vereecke, et. al, 2018).

Hoy la compañía objeto de estudio lleva a cabo el ejercicio de planeación de la demanda bajo métodos estadísticos y series de tiempo para todo el portafolio de referencias de línea, pero carece de una herramienta que le permita generar el mejor estimado de la canibalización, es decir, de la disminución de las ventas de la referencia de línea cada que en simultánea se presentan eventos promocionales.

Las diferencias entre lo que se estima vender de la referencia de línea y lo que realmente se vende en presencia de evento promocional han traído quiebres importantes en la cadena de suministro dejando de producir más de una referencia de lo que el mercado estaba requiriendo o por el contrario produciendo excesos que no son solicitados.

Lo anterior ha llevado a afectar los niveles de servicio de las referencias de línea impactas por la entrada de las ofertas en un 20% aproximadamente durante el evento y ha traído destrucciones de producto que no logran ser vendidos. Por un lado, las sobrestimaciones en la información entregada han traído en promedio vencimientos de cerca de 3.500 millones anuales de producto

que por su vida útil no alcanzan a ser comercializables. Por su parte, las subestimaciones en dicha información han generado aumentos en el número de agotados de la referencia de línea en 1.200 millones anuales aproximadamente. Las anteriores cifras muestran la relevancia del problema para la compañía.

Hoy para el ejercicio de predicción de la canibalización en la compañía se usan métodos cualitativos, es decir, se dan bajo supuestos intuitivos en este caso del área de ventas, marketing y Trade, los cuales al estar basados principalmente en el juicio humano pueden verse afectados por factores externos (Hyndman y Athanasopoulos, 2013).

Este proyecto busca proponer un modelo que permita estimar de una manera más asertiva los efectos que la canibalización proveniente de eventos promocionales tendrá sobre las referencias de línea, y así poder mejorar la planeación de las operaciones en la compañía, que permita preparar la cadena de suministro para atender las disminuciones de consumo, para así retomar la producción y el abastecimiento una vez finalicen los eventos. Se busca que lo anterior permita traer beneficios económicos para la compañía tales como reducción en el número de agotados de las referencias de línea durante los eventos y reducción en los vencimientos de producto terminado.

El modelo propuesto requiere determinar cuáles son aquellas posibles variables que puedan tener una mayor incidencia en la canibalización, para así determinar el efecto de dichas variables sobre las ventas. Para desarrollar el modelo se realizará un piloto con cuatro (4) de las referencias principales de línea, de cuatro (4) marcas diferentes, con las nueve (9) ofertas del segmento con mayores ventas, que incluyen nueve (9) centros de distribución y a tres (3) de los mayores clientes de la empresa, usando para ello técnicas de aprendizaje de máquina con el fin de minimizar el uso de la intuición en el proceso de toma de decisiones.

2. Objetivos

2.1 Objetivo General

Proponer un modelo de predicción de canibalizaciones que permita estimar de manera más asertiva las ventas de las referencias de línea cuando se tiene presencia de una oferta.

2.2 Objetivos específicos

- Identificar las variables que afectan la demanda de los artículos de línea cuando existen ofertas o promociones.
- Obtener, limpiar, estructurar y analizar la información existente de las variables definidas.
- Evaluar modelos multivariados de predicción de la canibalización.
- Integrar el resultado del modelo de predicción con la planeación de ventas y operaciones.
- Validar el resultado del modelo en comparación con lo propuesto por el equipo comercial.

3. Marco Teórico

Todo proceso de investigación requiere de la exploración sistemática de teorías y conceptos que permitan ampliar el espectro de conocimiento con relación al fenómeno que se desea estudiar. En consonancia con lo anterior, Farías (2009), plantea la importancia de identificar las posiciones epistemológicas que fundamentan las construcciones en torno a un tema, en la medida que estos precedentes permiten verificar si las preposiciones establecidas en el planteamiento del problema de la investigación son claras o, por el contrario, deben ser replanteadas a causa de su inviabilidad.

A partir de lo anterior, en la presente sección se explora un conjunto de antecedentes, avances teóricos y conceptuales, en aras de fundamentar el presente trabajo y, al tiempo, sentar un insumo para la estructura metodológica y la interpretación de los resultados de la investigación. Es menester resaltar que no se encontró un proceso de investigación documentado referido propiamente al objeto de estudio y al tipo de compañía, sin embargo, se identificaron algunos estudios que se han desarrollado en torno al tema de la predicción de canibalizaciones.

3.1 Antecedentes de investigaciones referidas a la aplicación del modelo de canibalizaciones

La revisión inicial de la literatura permitió identificar un estudio realizado por Lomax et. al (1997), en el cual los autores examinan tres alternativas técnicas para estimar la canibalización en una empresa de manufactura de productos para el hogar. Para el desarrollo del estudio los autores emplean un análisis de pérdidas-ganancias, donde se efectúa un proceso de duplicación de tablas y compra, utilizando métodos de pronóstico cuantitativo, mediante las pruebas de Farnum y Stanton (1989), referidas a las desviaciones y pruebas no paramétricas de estacionalidad. En

conclusión, los autores establecen que las magnitudes esperadas de la canibalización están en proporción directa a la participación que posee cada marca antes del lanzamiento de un nuevo producto, sin embargo, que se debe seguir desarrollando estudios para de esta manera validar la hipótesis presentada.

Este antecedente resulta oportuno para la presente investigación, en tanto posibilita identificar la manera en la cual se han efectuado medidas de canibalizaciones, los métodos elegidos y sus resultados, aspectos que sientan la base o se presentan como insumos para la generación de un modelo para la predicción de canibalizaciones.

Por su parte Van et. al (2010), propone un modelo que permite estimar los efectos de la canibalización, y calcular el porcentaje de ventas del nuevo producto en relación con la demanda total de partes fabricadas para el sector automotriz. El modelo implementado se basa en el modelo de Corrección de Errores Vectoriales, el cual permite descomponer las ventas base de un nuevo producto y determinar cuánto de su la nueva demanda se debe a la canibalización de los demás productos de la empresa. El modelo fue aplicado para analizar la introducción de un nuevo auto, tomando como punto de partida los datos detallados de las transacciones de los demás autos de la empresa. El estudio permite concluir que la introducción de la nueva referencia de auto provocaría que el 15% de los propietarios del segmento de sedanes de lujo perdieran el interés por consumir la categoría de SUV. Estos cálculos ilustran que la canibalización puede ser un “mal necesario”, ya que estimula y motiva la adquisición de nuevos productos e incluso la generación de nuevos nichos de mercado.

Este estudio se considera relevante para la investigación en marcha, en la medida que, presenta y desarrolla un modelo de Corrección de Errores Vectoriales, el cual varía con el tiempo y es un modelo matemáticamente consistente. Este modelo al ser empleado a partir de un estudio

empírico expone elementos metodológicos y procedimentales que pueden ser tomados en cuenta como base para el modelo de predicción de canibalizaciones que se desea desarrollar a partir del presente proceso de investigación.

Hvolby y Steger (2015), en su estudio buscan proyectar los posibles escenarios de canibalización y cómo estos desafían las operaciones de una empresa de productos alimentarios perecederos. Para ello, los autores se centran especialmente en la canibalización de un solo producto en términos de calidad y precio y, a partir de allí, presentan escenarios de estrategias de calidad y precio, para de este modo desarrollar un modelo conceptual de reposición de productos, el cual consiste en determinar los periodos en los cuales se deben reponer los productos antes de que estos alcancen su fecha de vencimiento. Los resultados del estudio indican que, los dos desafíos principales para los minoristas son optimizar las ganancias y reducir el desperdicio. Una posible estrategia para realizar esto puede ser introducir precios dinámicos y decisiones de reposición de productos, lo cual es posible toda vez se conozca la dinámica de canibalización entre productos.

Otra investigación evalúa la canibalización entre productos y promociones minoristas (Herrala, 2018), para lo cual, se propone un método de aprendizaje automático mediante el uso de la herramienta Python, a partir del cual se pueda identificar el alcance de la canibalización entre artículos individuales a partir de datos de ventas de series temporales. La metodología usada por el autor para determinar la canibalización se centra en la identificación de los datos de ventas, a partir de los cuales se valida la relación entre la caída de volumen del producto canibalizado y el aumento de volumen del producto promocionado. Para ello, se identifica el volumen de los productos en lugar de la rotación debido a la conexión más clara con la elección del consumidor y la sustitución de la demanda.

Los resultados permiten concluir que, al probar el método en datos simulados durante tres años, este se reconoce como estable, ya que converge en la misma solución independientemente de la suposición o hipótesis inicial. De igual modo, se identifica que la precisión del modelo disminuye a medida que aumenta el número de productos o la distorsión en los datos. Es importante señalar que el ruido de los datos afecta a la bondad de los resultados de forma lineal, lo cual se explica porque las líneas de base son cada vez más débiles para predecir los volúmenes a medida que aumenta el ruido. A partir del estudio se puede concluir que la planificación de promociones tiene un gran potencial para aumentar los márgenes de promoción y brindar a las empresas ventajas competitivas.

Dentro de las investigaciones desarrolladas en el contexto nacional se identifica el estudio de Sánchez (2019), cuyo objetivo se centra en determinar la importancia del desarrollo de un modelo que permita medir la canibalización de la puesta en marcha de las alianzas estratégicas en cada marca, a través del comportamiento de las órdenes de cada marca en el tiempo. La metodología del estudio se basa principalmente en el análisis de datos, el cual permite encontrar un método de proyección para capturar la tendencia o comportamiento de la venta. Con base a lo anterior, se desarrolla un modelo de canibalización, el cual permite llegar a la conclusión de que, este tipo de modelos permite optimizar procesos y hacer una verdadera planeación estratégica, ya que a partir del mismo se logra evidenciar la verdadera venta incremental (o cantidad de órdenes) que, en este caso, la alianza estaba generando, “o lo que es lo mismo, un análisis profundo de su eficiencia, permitiendo priorizar las alianzas y orientar los esfuerzos de forma más acertada” (Sánchez, 2019, p.2).

De acuerdo con la revisión de literatura, se logra evidencia que existe un interés por parte de la comunidad académica y directamente afectados en desarrollar modelos predictivos con el ánimo

de identificar y medir el nivel de canibalización, así como, determinar la afectación económica que este fenómeno implica sobre la empresa. Sin embargo, es menester resaltar que no existen modelos específicos que puedan ser aplicados al problema que se aborda en este proyecto, ya que, las características de la empresa manufacturera de alimentos objeto de estudio, el tipo de productos y las necesidades específicas relacionadas con la gestión de la demanda, distan considerablemente de los antecedentes analizados.

A continuación, se presentan las teorías transversales al estudio. En primer lugar, se explora el concepto de canibalización, seguidamente, los procesos de gestión de la demanda, la naturaleza de los datos en modelos cuantitativos y finalmente se hace una aproximación teórica al aprendizaje automático o Machine Learning.

3.2 Aproximación teórica al concepto de canibalización

Al efectuar la revisión de la literatura se identifica que la canibalización es un objeto de estudio económico que ha llamado la atención de los investigadores en los últimos años, elemento que permite entender la manera en la cual una situación diferencial o específica desarrollada para potenciar a la empresa, puede canibalizar o afectar otra situación de la misma.

Ahora bien, no existe una definición generalmente aceptada de canibalización (Lomax et. al, 1997). Ésta en términos generales puede entenderse como el proceso por el cual un nuevo producto gana ventas desviándolos de un producto existente (Kerin et. al, 1978). Así mismo, otros autores indican que la canibalización es la medida en que los clientes de un producto se interesan por otros productos ofrecidos por la misma firma (Mason y Milne, 1994).

En la actualidad se identifican siete tipos de canibalización (Ivanov, 2008), a saber: canibalización de producto, base de clientes, canal de distribución, corporativo, intra-empresarial, de recursos e imagen. La Tabla 1, muestra una breve descripción de los tipos de canibalización.

Tabla 1. *Tipos de canibalización*

Tipo de canibalización	Descripción
Canibalización de producto	Las promociones de determinado producto o servicio pueden canibalizar o afectar las ventas de otros productos o servicios no promovidos de igual forma, lo cual puede generar ingresos de bajo margen.
Canibalización de base de clientes	Se refiere a la sustitución de clientes de alto valor por clientes de bajo valor
Canibalización de canal de distribución	La estimulación de un canal de distribución particular exprime las ventas de otros canales.
Canibalización corporativa	Las fusiones y adquisiciones son demasiado optimistas y estas no producen efectos sinérgicos, por tanto, el valor de mercado de la empresa combinada es inferior a la suma de los valores de mercado de las empresas fusionadas
Canibalización intra-empresarial	Las oficinas de la empresa compiten entre sí por las ventas al mismo cliente
Canibalización de recursos	La introducción de nuevos productos requiere el desvío de recursos del proyecto existente actualmente más rentable
Canibalización de imagen	Erosión de la imagen de la empresa, a partir de las percepciones de los clientes y de los empleados sobre la empresa

Fuente: elaboración propia basada en Ivanov (2008).

En el presente caso de estudio se tendrá en cuenta la canibalización de los productos, el cual se da cuando la empresa crea un nuevo producto u oferta con el objetivo de atraer nuevos clientes o aumentar sus ventas. Esta acción también puede atraer a una gran proporción de su base actual de clientes, los cuales pueden preferir y adquirir la oferta o el nuevo producto, dejando de consumir el producto que generalmente adquirirían. En este punto es cuando se produce el proceso de canibalización.

La canibalización es importante, en la medida que se constituye como una estrategia para que las empresas crezcan y expandan sus operaciones, por medio de la retención de sus clientes actuales y la atracción de clientes nuevos. Sin embargo, es necesario que la compañía que desea desarrollar esta estrategia tenga un control sobre la misma y haga las estimaciones pertinentes para que no se perjudiquen el volumen de las ventas y, por consiguiente, sus ingresos.

El Instituto de Finanzas Corporativas (CFI) de Estados Unidos (2019), indica que existen dos formas para que los procesos de canibalización no afecten a las compañías: (1) identificación de los mercados objetivo y (2) evaluación de la demanda. El primer proceso indica que las empresas deben identificar los mercados específicos para cada uno de los productos, para de esta manera determinar qué brecha cubre el producto existente y los consumidores específicos a los que sirve el mismo. El segundo proceso, sugiere que las empresas determinen la cantidad de ingresos netos que probablemente generará el nuevo producto o la oferta, por lo que es necesario que se evalúen elementos como los costos de producción incurridos frente a los beneficios, estos últimos en forma de nuevos ingresos.

Es importante resaltar que los nuevos productos no siempre generan mayores ingresos. Por el contrario, si bien estos pueden aumentar el volumen de ventas a corto plazo, pueden provocar que los ingresos disminuyan a largo plazo. Si ese es el caso, es mejor que una empresa se quede

con su producto original. No obstante, cuando una compañía solo ofrece un producto debido a su fuerte cuota de mercado, corre el riesgo de desatender las necesidades de sus clientes. De allí la importancia de contar con un modelo que permita a la empresa la predicción de canibalizaciones y tomar decisiones estratégicas.

3.2.1 Tipos de promociones que sientan la base del proceso de canibalización.

La detección oportuna del porcentaje de canibalización es sin duda necesario para preparar los eslabones y la información relevante para el flujo de la cadena de suministro. Para este caso el foco estará dirigido en las ofertas centradas en los productos en promoción que afectan directamente la gestión de demanda de los productos de línea.

En este caso se identifican tres tipos de promoción: promociones de precio, promoción de regalos y promociones personales. Las primeras buscan usar al precio como estímulo de compra, entre las cuales se encuentran las promociones de rebaja, reembolso, 2x1, sistema de puntos o lanzamientos. Las segundas, se centran en añadir un porcentaje de contenido adicional al producto o se incluye un producto adicional gratis que añada valor al consumidor. Las terceras, se dirigen a interactuar específicamente con el consumidor, como por ejemplo, la distribución de muestras gratis o las degustaciones.

3.3 Gestión de la demanda

Uno de los conocimientos de mayor importancia y con mayor valor estratégico para la gestión de una compañía es la identificación de las características y las variables de la demanda.

La comprensión de la demanda implica que la empresa reconozca la percepción de valor de los productos que ofrece, las motivaciones y restricciones de los consumidores y, en general, todo lo que puede condicionar las decisiones de compra (Salazar, 2011).

Una vez la compañía reconozca los aspectos mencionados, puede desarrollar procesos de previsión de demanda, en los cuales puede identificar la cantidad de productos que los consumidores están dispuestos a adquirir por unidad de tiempo (López 2006). Es importante resaltar que las características de la demanda de un determinado producto definen el tipo de proceso productivo más adecuado para su fabricación. Entre las características más importantes se encuentran: tamaño y frecuencia de los pedidos, posibilidad de atender los pedidos no satisfechos y posibilidad de prever la demanda entre otras.

Con relación al tamaño y frecuencia de los pedidos, es importante tener en cuenta que si los clientes de la empresa hacen pedidos de gran volumen es más fácil que la empresa se quede sin existencias para atender su demanda, lo cual sugiere un sistema de producción flexible que le posibilite dar una respuesta rápida a los requerimientos y necesidades de sus clientes (López 2006).

En lo que respecta a la posibilidad de atender los pedidos no satisfechos, específicamente en los productos de uso común, los clientes esperan encontrarlos disponibles siempre que los desean adquirir. Bajo esta premisa, el tiempo que el consumidor sea capaz de esperar para obtener un producto debe determinar el tipo de producción que se debe emplear para cada producto. Así mismo, la característica referida a la posibilidad de prever la demanda, indica que, si la demanda de un producto es uniforme y predecible, se pueden emplear procesos de producción menos flexibles.

Ahora bien, tomando como base la compañía objeto de estudio, es importante indicar que, una de las formas de gestionar la demanda se centra en efectuar un proceso de segmentación de

los materiales. Según la variabilidad en el consumo y frecuencia en los pedidos se genera una clasificación entre demanda ordinaria y demanda extraordinaria. De forma periódica (generalmente cada mes) se toma como punto de partida del ciclo de Planear Ventas y Operaciones (PVO) de la compañía, la generación de predicciones para cada tipo de demanda.

Para la predicción de los materiales de demanda ordinaria (referencias de línea) se aplican métodos cuantitativos, de los cuales se usan modelos de pronósticos de series de tiempo dado el tipo de datos con los que se cuenta. Dependiendo el patrón identificado en las series de tiempo (estacionario, tendencial, estacional y/o cíclico) se usa el modelo de pronóstico. Para la predicción de la demanda extraordinaria se aplican métodos cualitativos, en donde el equipo comercial de la compañía entrega los planes de las ofertas y la canibalización que estas generan en las referencias de demanda ordinaria.

Una vez es recibida la anterior información, se ajusta el pronóstico de la referencia de línea que será canibalizada, se consolida toda la información y se genera el plan de ventas el cual posteriormente es entregado a la red de operaciones. Precisamente a partir del presente estudio se busca proponer un modelo de predicción de canibalizaciones que permita estimar de manera más asertiva las ventas de las referencias de línea cuando se tiene presencia de una oferta.

3.3.1 Naturaleza de los datos en modelos cuantitativos para la gestión de demanda

La previsión de la canibalización se puede realizar mediante métodos cualitativos o subjetivos o métodos cuantitativos que buscan tener un mayor grado de objetividad. Para utilizar un enfoque de predicción cuantitativo, como se busca en el presente proyecto, es necesario y

fundamental conocer la naturaleza de los datos ya que, dependiendo de ello, se utilizan para su tratamiento unas técnicas u otras.

La naturaleza de los datos puede ser de series de tiempo, de corte transversal, o de panel. La primera categoría incorpora información de variables individuales medidas en puntos sucesivos en el tiempo o a lo largo de periodos regulares (semanas, meses, etc.). El objetivo es pronosticar el comportamiento para la secuencia de las observaciones en el futuro.

Los datos de corte transversal no incorporan el aspecto temporal, sino que representa el análisis de información para las unidades individuales de estudio en un momento determinado de tiempo. Con esta información se quiere predecir algo que no ha sido observado, a partir de aquellos datos que si han sido observados.

Finalmente, los datos de panel combinan información tanto de series de tiempo como de corte transversal, incluyendo una muestra de observaciones para un periodo de tiempo determinado.

Teniendo en cuenta que para el caso de estudio lo que se busca es generar un modelo de predicción en el cual se indague sobre distintas variables que puedan explicar el comportamiento de la canibalización, se usarán modelos de corte transversal, también llamados explicativos o causales, pues son estos los modelos que se usan cuando se percibe una relación entre la variable que se va a predecir con una o más variables predictoras (Hyndman y Athanasopoulos, 2017).

En la ciencia de datos, específicamente en el aprendizaje automático o Machine Learning muchas de las técnicas usadas tratan con este tipo de modelos, siendo esta una de las principales razones para usar dichas técnicas en el presente proyecto.

3.4 Aprendizaje automático o Machine Learning

El aprendizaje automático o Machine Learning busca crear sistemas que sean capaces de aprender automáticamente, esto es, sistemas capaces de encontrar patrones completos en grandes conjuntos de datos por sí solos (Herrera y Mazón, 2016). Dentro de los algoritmos utilizados en el aprendizaje automático existen técnicas de aprendizaje supervisado y aprendizaje no supervisado. En las primeras se aplican lo que se ha aprendido con los datos históricos, es decir encuentran patrones para sacar conclusiones sobre nuevos datos, mientras en las segundas se pueden extraer inferencias de conjuntos de datos, las cuales se usan para realizar análisis exploratorio de datos y son útiles cuando no se está seguro de qué buscar.

De acuerdo con lo anterior, para el diseño del modelo de predicción de canibalización se usará aprendizaje automático con algoritmos supervisados dentro de los cuales se encuentran diferentes modelos de regresión y de clasificación. Con los modelos de regresión se tiene un número de variables predictoras y una variable respuesta y se busca encontrar una relación entre dichas variables que proporcionen un resultado (Novales, 2010). Por su parte, a partir de los modelos de clasificación lo que se pretende es detallar y predecir una categoría posible teniendo como punto de partida los datos históricos, en este caso del producto.

En el presente proyecto se utilizarán métodos de aprendizaje automático con algoritmos supervisados mediante la aplicación de modelos de regresión ya que es una metodología útil para el manejo y tratamiento de la cantidad de datos que se tomará como punto de partida para el análisis y la generación del modelo. De igual forma, los métodos que se emplearán permiten una mejor comprensión de las variables que afectan la demanda de los artículos de línea (cuando existen

ofertas o promociones) de la empresa objeto de estudio y permite efectuar procesos de predicción para apoyar los procesos de planeación de ventas y operaciones.

Es importante resaltar que para el desarrollo del presente proyecto se utilizará como lenguaje de programación Python, el cual tiene un uso frecuente en Machine Learning (Ozgun, et. al, 2017).

4. Metodología

4.1 Paradigma de investigación

Luego de explorar la literatura académica y contrastarla con el problema de investigación identificado, se determina que el presente estudio se asocia al paradigma de investigación positivista (empírico- analítico), el cual indica que la realidad es aprehensible y está impulsada por leyes y mecanismos naturales inmutables explicados a partir del binomio “causa-efecto” (Guba y Lincoln, 2001). Para los autores, este paradigma supone que el investigador y el fenómeno estudiado sean entidades independientes, es decir, que las concepciones, perspectivas, opiniones y nociones previas del investigador no influyan en el objeto estudiado, ni este último influya en el investigador. A nivel metodológico este paradigma presenta preguntas o hipótesis que son presentadas en forma de proposiciones, las cuales están sujetas a verificación empírica.

En este caso se parte del supuesto de la asociación del efecto directo de una oferta sobre una referencia de línea ya detectada, es decir, se explica el movimiento en volumen de la referencia de línea usando el movimiento inverso en la referencia promocional que se encuentra ya definida por el área de mercadeo y ventas.

4.2 Enfoque de investigación

Un enfoque cuantitativo busca descubrir leyes o principios generales, a partir de sistemas estadísticos con los cuales se analizan hechos conocibles y recuperables, que se pueden contar y medir (Galeano, 2004). Este enfoque se dirige a entender los fenómenos con independencia de los estados subjetivos de los participantes, de allí que su base sea la observación empírica y la

medición de lo observado (con su correspondiente verificación objetiva). En el enfoque cuantitativo los datos son rigurosos, exactos y confiables, ya que este usa la validez y confiabilidad de los modelos matemáticos.

4.3 Tipo de estudio

El estudio es de tipo descriptivo, el cual, permite caracterizar fenómenos, a partir especificación de sus propiedades, características, procesos y los elementos que en él convergen (Mousalli, 2015). En este tipo de estudio sólo se pretende medir o recoger información sobre las variables de manera independiente, más no identificar cómo se relacionan entre sí.

4.4 Fuentes de información

Para este estudio se utilizaron dos tipos de fuentes de información: primaria y secundaria. Con relación a la información primaria el estudio contó con la participación de dos expertos (jefe de planeación de demanda con foco a los planes de venta y sensibilidad comercial, y jefe de monitoreo de la cadena de suministro con foco en la ciencia de datos), quienes mediante el análisis de los resultados planteados proporcionaron su opinión como usuarios expertos con relación al modelo de predicción de canibalizaciones, al validar si este permite estimar de manera más asertiva las ventas de las referencias de línea cuando se tiene presencia de una oferta.

Con relación a las fuentes secundarias, en primer lugar, se exploraron artículos, tesis y en general investigaciones sobre el tema en cuestión, los cuales fueron recabados a partir de la revisión de la literatura tomando como punto de partida repositorios y bases de datos

especializadas. A partir de los elementos teóricos y conceptuales identificados en esta revisión de la literatura se identificaron las variables y procedimientos para el diseño de un modelo de corte transversal donde se usa técnicas supervisadas de Machine Learning, ya que esta es una metodología útil para el manejo y tratamiento de la cantidad de datos que se tomará como punto de partida para el análisis y la generación del modelo.

De igual modo, dentro de las fuentes secundarias de información se encuentran los datos proporcionados por la compañía, específicamente por las áreas de ventas, mercadeo y Trade, los cuales permitieron efectuar el proceso de simulación para el diseño, desarrollo y validación del modelo propuesto.

5. Modelo de proyección de canibalización

Con el objetivo de concretar el modelo de proyección de canibalización, se desarrollaron un conjunto de procesos, los cuales se describen, a continuación, en la figura 1.

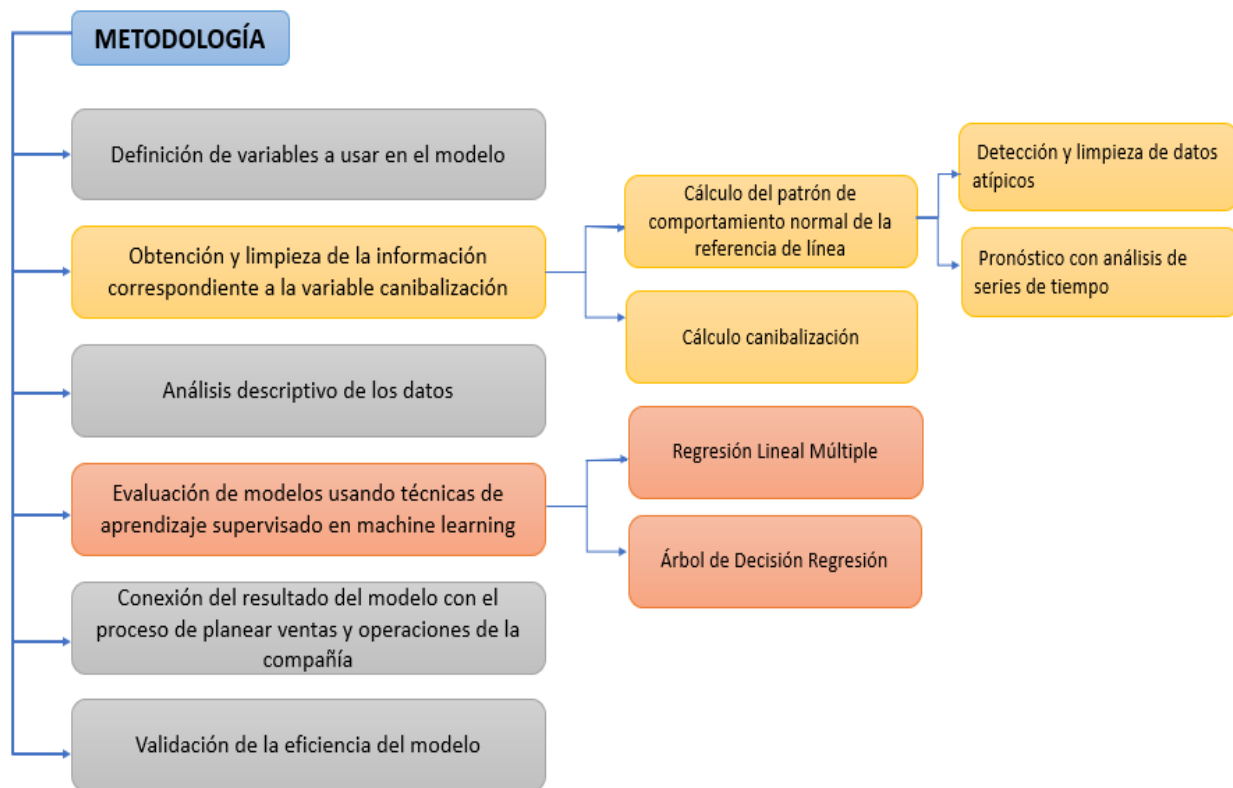


Figura 1. Etapas del modelo de canibalización propuesto. Elaboración Propia.

Como puede observarse en la figura 1, para la definición de las variables a usar para el modelo, se tuvo en cuenta inicialmente la revisión bibliográfica y se terminaron de concretar en compañía del equipo comercial teniendo en cuenta como restricción la disponibilidad de información histórica. Después fue necesario obtener la información histórica con la cual se modelaría la variable canibalización, para lo cual se calculó primero el patrón de comportamiento

normal de la referencia de línea, es decir se pronosticó la referencia de línea mediante series de tiempo asumiendo que no hubo presencia de oferta, luego se calcularon las unidades dejadas de vender en aquellas semanas donde si hubo oferta y así se obtuvo para cada combinación de referencia, cliente, cedi, semana presentes en el piloto el factor de canibalización, dato que junto a las demás variables servirían de insumo para la incorporación del modelo.

Seguido de este paso, se procede con el análisis descriptivo de toda la información con el fin de encontrar patrones de comportamiento de los datos y hallazgos importantes para la investigación. A continuación, se incorporan dos modelos usando técnicas de aprendizaje supervisado en Machine Learning los cuales se comparan en términos de coeficiente de determinación y exactitud para la elección de uno de ellos.

Se propone un proceso de integración del resultado del modelo con el proceso de planear ventas y operaciones de la compañía haciendo énfasis en cada uno de sus nodos (abastecimiento, producción, distribución) y finalmente se valida un evento reciente con una de las referencias y clientes estudiadas en el proyecto, mediante una comparación del resultado real de canibalización con los estimados entregados por el equipo comercial y por el modelo.

Cada uno de los bloques de la figura 1 se irán desarrollando en esta sección.

5.1 Definición de variables

Con la finalidad de desarrollar el modelo para estimación de la canibalización se estableció de manera preliminar un conjunto de variables, las cuales se definieron en conjunto con un equipo interdisciplinario de la compañía en el cual participaron áreas de mercado, ventas, trade y planeación. Desde la experiencia de cada área se citaron una serie de variables, las cuales sentaron

la base para la búsqueda de la información que fue filtrada y extraída teniendo en cuenta la disponibilidad de los datos históricos.

Adicionalmente, el proceso de identificación de variables también se apoyó en la literatura académica, específicamente en estudios relacionados con temas como referencia de línea o referencia base; promociones de venta, clientes; Centros de Distribución y Canibalización entre otros (Hazbun, 2014; Roman, 2016; Diaz, Consuegra y Esteban, 2014; Wheelen y Hunger, 2007).

A continuación, se presenta una tabla con las variables finalmente seleccionadas y que serán usadas como entrada para el modelo.

Tabla 2. *Variables iniciales a incluir para la estimación de la canibalización*

Variable	Descripción
Referencia de línea	Es la referencia impactada, es decir, la que sufre la canibalización
Tipo de Oferta	Amarre, extra-contenido o combo virtual
Cliente	Es el almacén de cadena al cual se le vende la oferta y la referencia de línea
Cedi	Es el centro de distribución desde donde se efectúa la venta
Periodo	Es el periodo en que se lleva a cabo la oferta, (semana – año)
Marca	Es la marca de la referencia ofertada
Tipo de evento	Se especifica si el evento es triple A (eventos publicados en boletín generalmente en aniversarios de los clientes) o si se trata de un evento convencional (activados en su mayoría para incrementar flujo en los puntos de venta, o por estrategia de bloqueo a la competencia).
Descuento	Es el descuento aplicado a la oferta
Canibalización	Son las unidades dejadas de vender de la referencia de línea por cada unidad vendida de la oferta.

Fuente: elaboración propia.

Tomando como punto de partida las variables definidas en la Tabla 2, se da paso a la obtención y limpieza de la información correspondiente a la variable canibalización ya que esta variable debe primero calcularse para generar la información histórica y posteriormente modelarse y pronosticarse.

5.2 Obtención y limpieza de la información – variable canibalización

Para obtener la información histórica de la variable canibalización, la cual es nuestra variable objetivo es necesario llevar a cabo los siguientes pasos.

5.2.1 Cálculo del patrón de comportamiento normal de la referencia de línea

En primer lugar, se debe establecer de manera hipotética cuál podría haber sido la demanda del producto de línea si no se hubiera realizado la promoción. En la Figura 2 se ejemplifica el proceso propuesto.



Figura 2. Ejemplo a) venta referencia de línea en presencia de oferta y b) estimación del patrón de comportamiento normal de la referencia de línea. Fuente elaboración propia.

La línea azul en la Figura 2 muestra las ventas reales de la referencia de línea en cada periodo con y sin presencia de ofertas. Por su parte la línea roja en la Figura 2b) representa las ventas estimadas que se hubieran dado de la referencia de línea si no se hubiera activado la oferta.

Para realizar la estimación del patrón de comportamiento normal de la referencia de línea se analizan 4 años de historia llevados a las semanas donde no hubo evento promocional para las 4 referencias de línea que más ventas aportan a la compañía (aproximadamente el 15% de las ventas) y que fueron incluidas en el piloto, en 3 de los clientes principales y para los 9 centros de distribución como se muestra en la Figura 3, lo cual representa 80 combinaciones de referencia, cliente, cedi que se utilizaran en el modelo propuesto.

El análisis a las ventas se lleva a cabo con el fin de detectar el patrón de los datos y así pronosticar las semanas donde la venta de la referencia de línea presenta una caída en el consumo debido a la presencia de una oferta en el mercado.

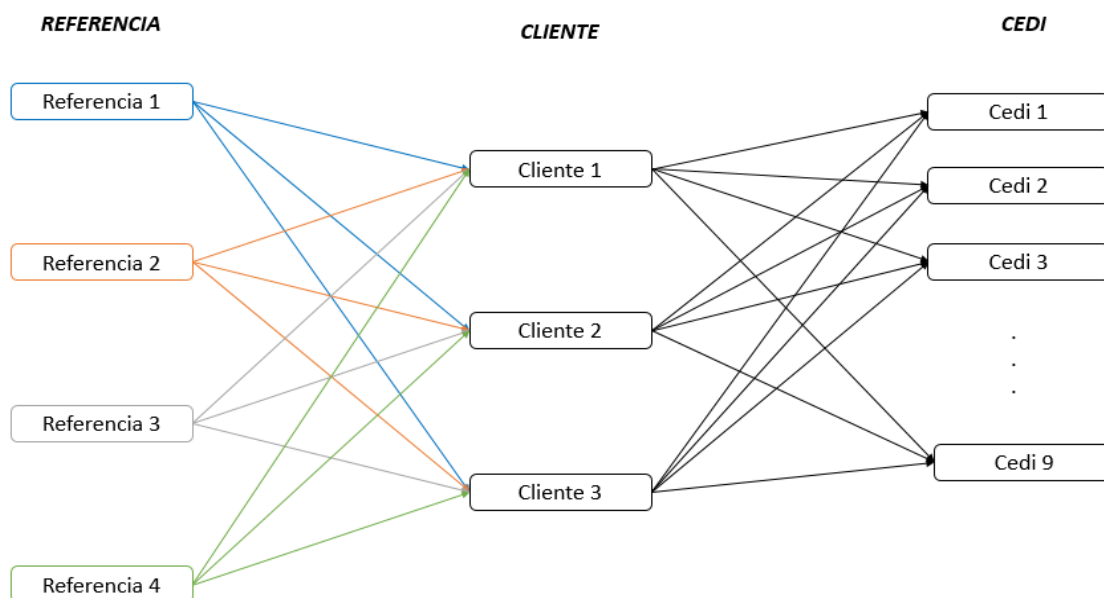


Figura 3. Combinaciones referencia, cliente, cedi. Fuente elaboración propia.

Se evidencia un patrón en las ventas aproximadamente estable para todas las combinaciones de referencias, clientes, cedís, a partir de la obtención de sus coeficientes de variación, los cuales se calculan como la relación entre la desviación estándar y la demanda promedio, dando para la mayoría de estos, valores por debajo de 0,25. La figura 4 muestra los coeficiente de variación de las ventas de las diferentes combinaciones mencionada anteriormente, en donde se observa que cerca del 80% de las combinaciones referencia, cliente, cedi, tienen coeficientes de variación que oscilan entre 0,02 y 0,25, los cuales pueden considerarse como artículos con demanda aproximadamente estable (Sipper y Bulfin, 1998, Silver et, al, 2018).

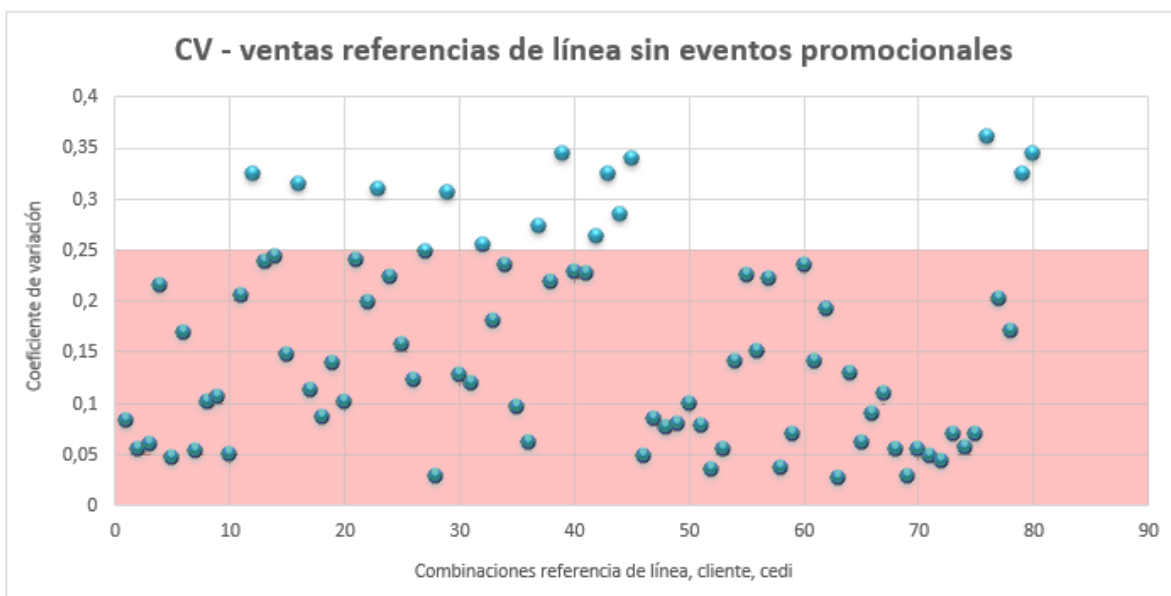


Figura 4. Coeficiente de variación ventas referencias de línea sin eventos promocionales a detalle cliente, cedi.

Fuente elaboración propia.

5.2.1.1 Detección y limpieza de datos atípicos

Para la detección de datos atípicos presente en las ventas por semana de cada una de las combinaciones referencia, cliente, cedi, sin presencia de evento promocional se aplicó normalización estadística a los datos con el objetivo de hacer comparaciones respecto a cada combinación y su media, para así poder determinar el grado de desviación de los datos. En la figura 5 se grafican los datos normalizados para los cuales se dividió cada registro (referencia, cliente, cedi, semana) por el promedio de su combinación referencia, cliente, cedi.

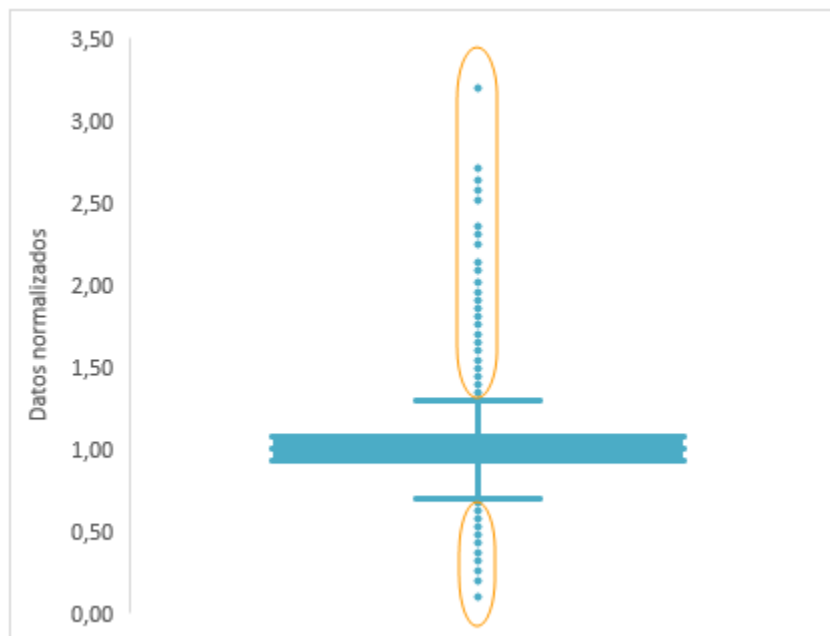


Figura 5. Datos normalizados referencia, cliente, cedi, semana.

En la figura 5, es posible distinguir puntos notablemente alejados del resto de las observaciones (encerrados en un ovalo), estos corresponden a los datos que se encuentran por encima o por debajo de los bigotes, esto es, por encima del cuartil 3 + 1,5 RIC (rango intercuartil que se calcula como la diferente entre el cuartil 3 y el cuartil 1) y por debajo del cuartil 1 - 1,5 RIC, lo anterior corresponde a los valores por encima de 1,29 y por debajo de 0,70, los cuales se consideran que están fuera del patrón general en la distribución, es decir, que son datos atípicos.

Junto al equipo comercial de la compañía, se analizan esos puntos y se determinan las razones que puedan explicar dicho comportamiento, entre las cuales se encuentran semanas de agotados debido a averías en equipos de producción, semanas de ventas superiores a la tendencia normal por descuentos puntuales efectuados directamente por el cliente, entre otras. Se decide ajustar dichos datos aplicando la siguiente regla:

Se define un límite superior y un límite inferior por cada año de historia para cada combinación de datos referencia, cliente, cedi. Teniendo en cuenta la naturaleza y la dispersión

de los datos los cuales se encuentran en su mayoría agrupados, los límites fueron establecidos en $\pm 1,5$ desviaciones estándar (Galvanize, 2020), esto buscando mayor sensibilidad para la detección y suavización de datos atípicos. Para los datos que se encuentren más dispersos con un rango más amplio se recomienda valores de desviación estándar para los límites a partir de 3.

Posteriormente, se evalúa la posición de cada punto con base a un gráfico de control, en donde, aquellos puntos que estén por encima del límite superior o por debajo del límite inferior son reemplazados por el promedio de ventas del año y del grupo de datos que se esté evaluando, dado que se trata de referencias de línea con comportamiento estacionario como se evidenció en la figura 4. Las Figuras 6 y 7 se muestran el proceso realizado para una de las referencias del piloto en un cliente y centro de distribución (CEDI) específico.

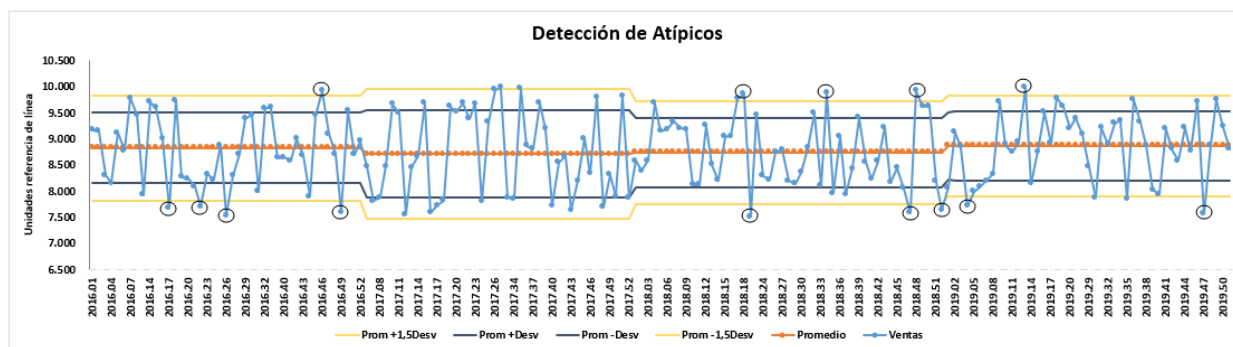


Figura 6. Detección de atípicos para una combinación referencia, cliente, cedi. Fuente elaboración propia.

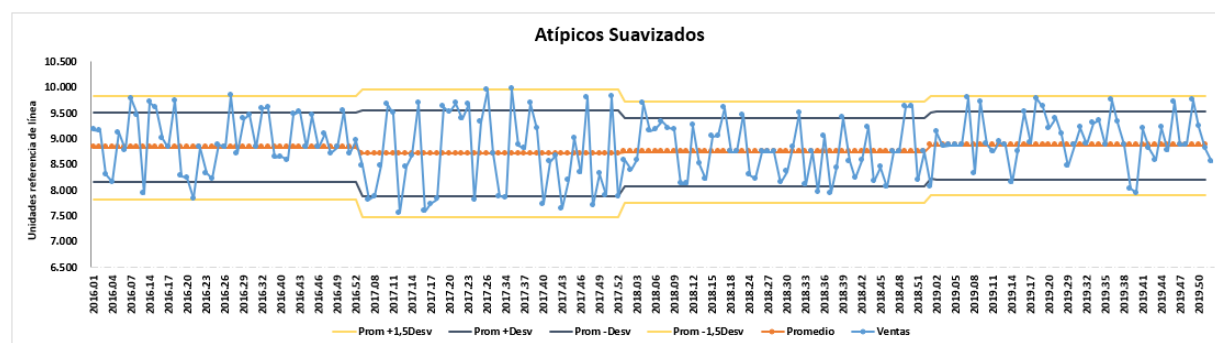


Figura 7. Atípicos suavizados para una combinación referencia, cliente, cedi. Fuente elaboración propia.

5.2.1.2 Pronóstico con análisis de series de tiempo

El siguiente paso consiste en generar un **pronóstico estadístico** con series de tiempo para las referencias de línea presentes en el piloto al detalle referencia, cliente, cedi, semana. Teniendo en cuenta que el patrón identificado en los datos fue estacionario ya que se detectaron series relativamente estables, el modelo de pronóstico usado fue suavización exponencial simple. Dicho modelo fue seleccionado por su simplicidad y por su flexibilidad para asignar diferentes valores del coeficiente de suavización a las series que hacen parte del análisis. Para el modelo de pronóstico se usaron diferentes valores para el parámetro alpha (α) para cada combinación referencia, cliente, cedi, cuyos valores se obtuvieron por medio optimización usando Solver de Excel para minimizar la desviación media absoluta (MAD). (Berry y Bliemel 1974; Rasmussen, 2003, Billah et. al, 2006).

Una vez corrido el modelo se obtuvo un resultado del error porcentual absoluto medio (MAPE) para el total de las combinaciones referencia, cliente, cedi, semana de 7,6%, resultado que genera confianza en la predicción realizada. Ya con el patrón de comportamiento normal de las 4 referencias de línea que hacen parte del piloto se da paso al cálculo de la canibalización.

5.2.2 Cálculo canibalización

El cálculo de la canibalización histórica que servirá de insumo para la generación del modelo de predicción se realiza en dos pasos. En primer lugar, se calculan las unidades estimadas dejadas de vender de la referencia de línea en las semanas donde hubo oferta ya que estas son las semanas en las cuales se desea conocer el impacto en la disminución de las ventas de la referencia de línea. Para ello se toma el dato del patrón de comportamiento normal hallado en el pronóstico a la

referencia de línea y se le restan las unidades reales vendidas de la referencia de línea, tal como se muestra en la Tabla 3 donde se expone a manera de ejemplo como se obtiene el estimado de las unidades que se dejan de vender en una semana específica para una de las referencias de línea en presencia de una oferta.

Tabla 3. Unidades dejadas de vender referencia de línea en presencia de una oferta

Semana	Unidades vendidas de la oferta	Venta Real referencia de línea	Pronóstico de venta referencia de línea	Unidades estimadas dejadas de vender referencia de línea
2017.08	1.000	400	1.300	1.300 - 400
				900

Fuente: elaboración propia.

Seguido de este paso, se calcula el coeficiente de canibalización el cual indica cuántas unidades se estima se dejaron de vender de la referencia de línea por cada unidad de oferta vendida.

$$\text{Coeficiente de canibalización} = \frac{\text{Unidades dejadas de vender de la ref de línea}}{\text{Unidades vendidas de la oferta}}$$

Tabla 4. Coeficiente de canibalización

Semana	Unidades vendidas de la oferta	Venta Real referencia de línea	Pronóstico de venta referencia de línea	Unidades dejadas de vender referencia de línea	Coeficiente de canibalización (CoefCan)
2017.08	1.000	400	1.300	1.300 - 400	900 / 1.000
				900	0,9

Fuente: elaboración propia.

Para el ejemplo en la Tabla 4 se concluye que por cada 100 unidades de oferta vendidas se dejan de vender 90 unidades de la referencia de línea. Finalmente, para el cálculo de la canibalización la cual se da en función de las unidades que se oferten, se multiplica el coeficiente de canibalización por las unidades vendidas de la oferta.

$$\text{Canibalización} = \text{Unidades estimadas a vender de la oferta} * \text{CoefCan}$$

El valor de la canibalización es aquel que debe tenerse en cuenta para afectar el plan de ventas de la referencia de línea y así plasmar el efecto que trae la activación de la oferta. Para el ejemplo expuesto en la Tabla 4, serían 900 unidades las que se deben restar y ajustar al plan de demanda que se tenga de la referencia de línea asociada a la oferta y así dar la visual a las operaciones de la disminución en el consumo que se tendrá de la referencia de línea.

Después de obtener el coeficiente de canibalización para las referencias del piloto a detalle cliente, cedi, semana, se hace el análisis descriptivo de la información de las variables presentadas en la Tabla 2, con el fin de describir las tendencias de los datos y encontrar relaciones entre cada una de las variables que permita tomar decisiones para llevar a cabo la ejecución del modelo.

5.3 Análisis descriptivo de la información

En este apartado, se hace uso de estadísticas descriptivas y técnicas de visualización que permiten comprender el contenido de los datos, evaluar la calidad de éstos y descubrir información relevante que aporte al desarrollo del modelo para la predicción de la canibalización.

Para esta parte del proyecto se utiliza el lenguaje de programación Python, el cual tiene un uso frecuente en Machine Learning (Muenchen, 2017). La herramienta utilizada para el estudio de datos y la modelación de estos fue Jupyter Notebook.

Para el proceso de cargue de la información a la herramienta es necesario dummificar las variables, es decir sustituir las columnas categóricas por varias columnas booleanas con el fin de llevar todos los datos de categóricos a numéricos.

Por ejemplo, en los datos iniciales se contaba con una columna llamada “Cliente” en la que existían 3 valores posibles (“Cliente1”, “Cliente2” y “Cliente3”). Esta columna se dummificó y se sustituyó por tres variables: “ Cliente1”, “ Cliente2”, y “Cliente3”, cuyos valores son 1 si la fila pertenece a esa categoría o 0 si no pertenece.

En la exploración de los datos, donde se cuenta con 2.223 combinaciones de referencia-cliente-credi-semana, inicialmente se analiza el comportamiento de la variable a predecir (canibalización) para la cual se obtuvo la información histórica llevando a cabo los pasos descritos en el apartado 5.2. El coeficiente de variación para esta variable fue de 1,90, resultado que denota una alta variación en los datos.

Para entender la relación entre las variables descritas en la Tabla 2, es decir, las variables con las que se cuenta para el desarrollo del modelo, se llevó a cabo la técnica de análisis de componentes principales PCA (por sus siglas en inglés), la cual permite agrupar las variables correlacionadas entre sí separándolas de las que no, obteniendo de esta forma un nuevo espacio reducido (Quiroga. C, Villalobos. A, 2015) que permita comprender y explicar el comportamiento de la variable a predecir.

Al aplicar la técnica PCA no se encontraron correlaciones superiores al 20% de las variables independientes (cliente, credi, semana, marca, tipo de evento, descuento) con la variable

dependiente (canibalización), lo cual indica que no existen correlaciones fuertes que permitan explicar la variable objetivo. Con lo anterior se infiere que el modelo no tendría un buen desempeño y estaría sub ajustado dada la poca información relevante con la que se cuenta.

Dados los resultados anteriormente descritos para la variable objetivo que es la que se busca predecir y además entender qué variables influyen en su comportamiento, se indaga con los expertos del área comercial y se concluye que el análisis y estudio a la variable canibalización debe hacerse por evento y no por semana ya que el comportamiento de cada semana del evento depende de la estrategia utilizada por cada cliente para el surtido de sus góndolas, lo anterior significa que para el cálculo histórico de la variable canibalización debe ajustarse el proceso descrito en el punto 5.2 y realizar el proceso sumando las semanas para cada evento promocional.

Para ejemplificar el impacto del hallazgo de realizar el proceso de estimación de la canibalización por eventos y no por semanas, en las Tablas 5 y 6 se muestra un ejemplo detallado de la activación de una oferta con duración de 4 semanas para un cliente y cedi específico.

Tabla 5. Ejemplo coeficiente de canibalización calculado por semana.

Semana	Análisis por semana				
	Unidades vendidas de la oferta	Venta Real referencia de línea	Pronóstico de venta referencia de línea	Unidades dejadas de vender referencia de línea	Coeficiente de canibalización (CoefCan)
2018.02	5.000	-	3.835	3.835	0,77
2018.03	3.100	500	3.818	3.318	1,07
2018.04	331	2.860	3.818	958	2,90
2018.05	100	3.780	3.865	85	0,85

Fuente: elaboración propia.

Tabla 6. Ejemplo coeficiente de canibalización calculado por evento

Duración evento	Análisis por evento				
	Unidades vendidas de la oferta	Venta Real referencia de línea	Pronóstico de venta referencia de línea	Unidades dejadas de vender referencia de línea	Coeficiente de canibalización (CoefCan)
4 semanas	8.531	7.140	15.336	8.196	0,96

Fuente: elaboración propia.

El realizar el análisis de la canibalización por evento promocional y no por semana, implica que la base de datos original compuesta por 2.223 combinaciones, pasara a 597 combinaciones, agrupados ahora por eventos.

Se lleva a cabo el análisis descriptivo de la nueva información por evento y se encuentra que el coeficiente de dispersión de los datos de canibalización por evento se reduce significativamente con respecto a la canibalización por semana pasando de 1,90 a 0,29, lo cual indica la presencia de datos más estables y menos desviados con respecto a su promedio. En la figura 8 se expone un gráfico de caja de bigotes para la variable canibalización por evento en el cual se identifica una media de 0,72, un valor mínimo de 0,19, un valor máximo de 1,19 y no se detectan datos por fuera de los bigotes, lo que indica que no hay presencia de atípicos en esta serie de datos.

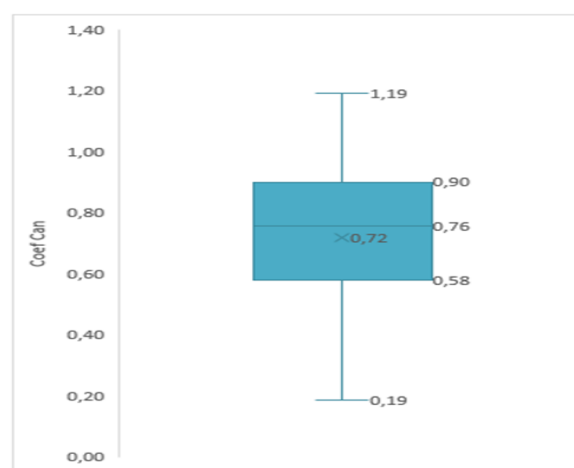


Figura 8. Coeficiente de canibalización por evento. Elaboración propia

La Figura 9 corresponde a un mapa de calor en el que se observa el grado de correlación entre todas las variables incluidas en el modelo. Mediante este mapa se logra identificar la existencia de colinealidad entre algunas variables independientes, lo que indica una correlación entre estas, permitiendo descartar alguna de ellas y simplificar de esta forma el modelo. En el mapa de calor entre más rojo o azul sea el recuadro mayor correlación existe, en donde el color rojo indica que la correlación es positiva, mientras el color azul señala que la correlación negativa, es decir, una correlación en sentido inverso.

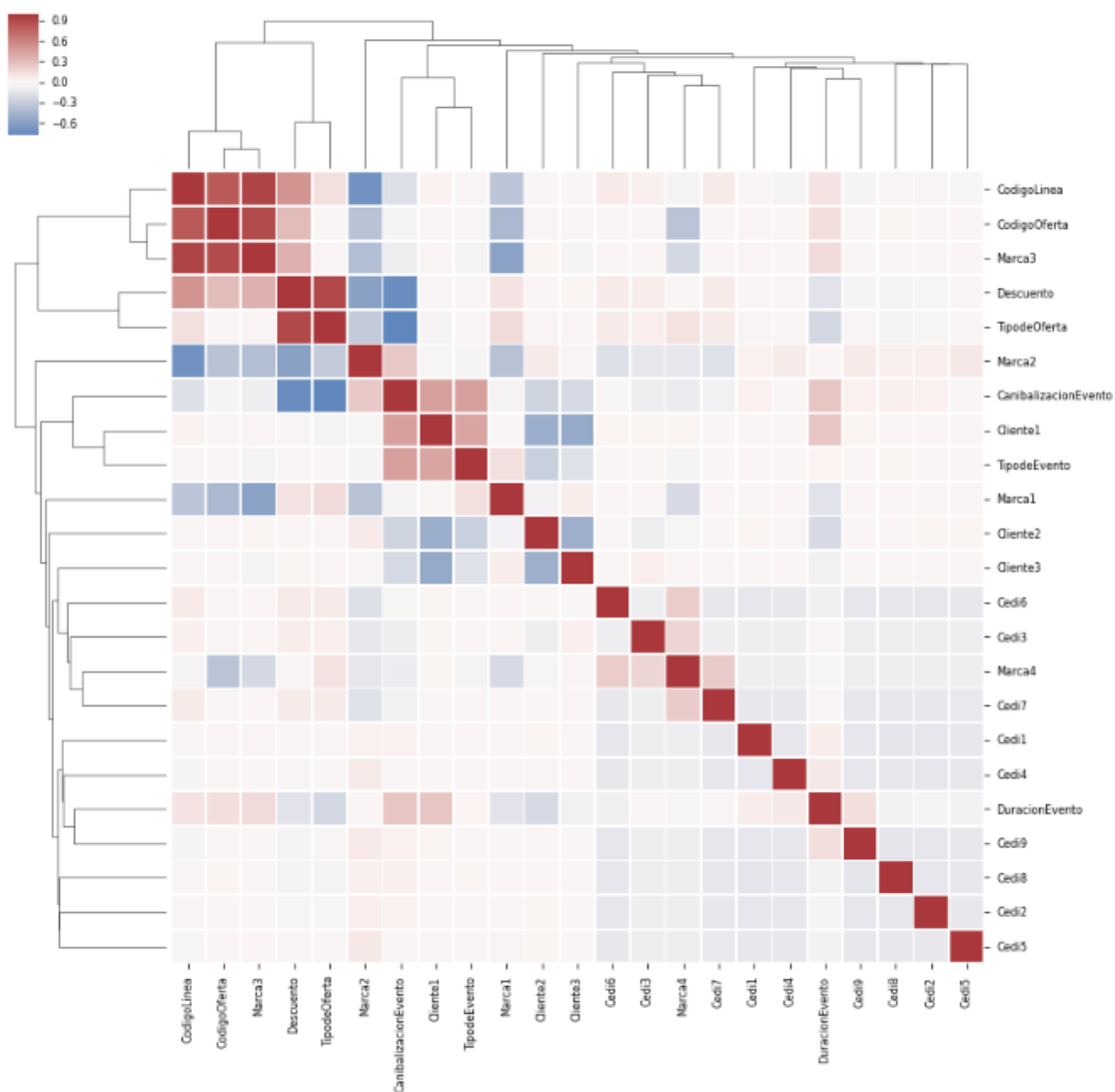


Figura 9. Mapa de calor entre variables evaluadas del modelo de canibalización. Fuente: elaboración propia

Se logra evidenciar una fuerte correlación entre algunas variables independientes, como por ejemplo “Descuento” y “Tipo de oferta”. La correlación existente entre ambas variables tiene explicación ya que el tipo de descuento que se brinda a las ofertas depende del tipo de oferta, es decir, una depende de la otra. Por esta razón debe excluirse una de las dos, y esta debe ser la que menor correlación tenga con la variable dependiente. En la Figura 10 se evidencian las variables con mayor correlación con la variable dependiente (encerradas en círculo blanco), en donde los valores positivos más altos y los valores negativos más altos son aquellos que presentan mayor incidencia en la explicación del comportamiento de la variable canibalización. Para el caso de las variables que presentaron colinealidad “Descuento” y “Tipo de oferta”, identificadas en la Figura 9, la variable “**Tipo de oferta**” es la que presenta mayor correlación con la variable objetivo, por lo cual se excluirá del modelo la variable “Descuento”, sin que esto afecte el modelo debido a la colinealidad anteriormente descrita.

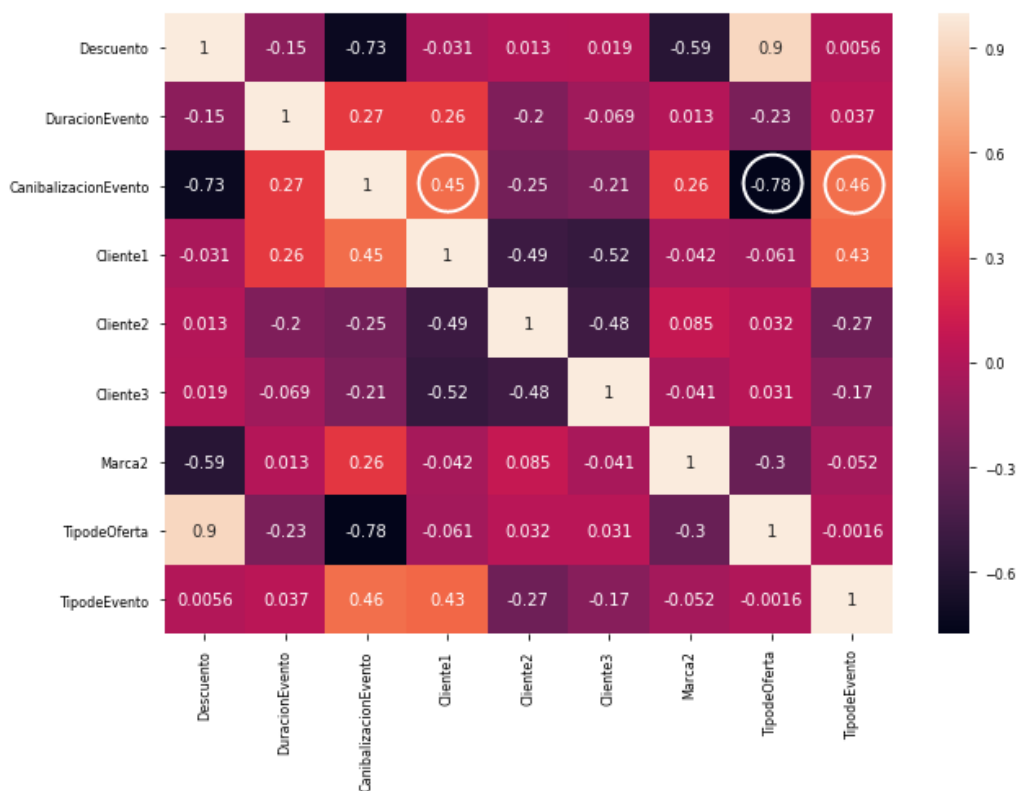


Figura 10. Variables con mayor correlación con la variable canibalización. Fuente: elaboración propia

En la figura 10 se resaltan las correlaciones superiores a +/- 0,30 con la variable a predecir, La variable **“Cliente1”**, hace parte de la conversión a variable numérica de la variable Cliente descrita al inicio de esta sección. Se observa una mayor relación de la variable **“Cliente1”** con la canibalización, que las variables **“Cliente2”** y **“Cliente3”**, (estas dos últimas presentan una correlación muy similar), por lo que en el modelo se usará la variable **“Tipo de Cliente”**, donde una categoría será **“Cliente1”** y la otra será **“Cliente2 o Cliente3”**, pues no hay diferencia significativa en la relación existente de estas dos últimas variables con la variable respuesta.

Otra de las variables con mayor correlación con la variable canibalización es **“Tipo de Evento”**, la cual puede ser Amarre o Extracontenido y finalmente se tiene la variable **“Tipo de Oferta”** la cual puede corresponder a evento convencional o evento triple A., tal cómo fueron descritas en la sección 5.1

A continuación, se analiza por separado el comportamiento de cada una de las variables seleccionadas para el modelo y su relación con la variable canibalización.

En la Figura 11 se observa la variable Cliente con sus categorías, **“Cliente1”**, **“Cliente2”** y **“Cliente3”**, donde se logra evidenciar una notable diferencia en el comportamiento de los datos del coeficiente de canibalización entre **“Cliente1”** y los demás clientes, siendo mayor el coeficiente para **“Cliente1”**. Dicho comportamiento está relacionado con los valores de correlación estudiados en la figura 10 donde **“Cliente2”** y **“Cliente3”** presentaban valores muy similares, de 0,25 y 0,21 respectivamente con la variable canibalización.

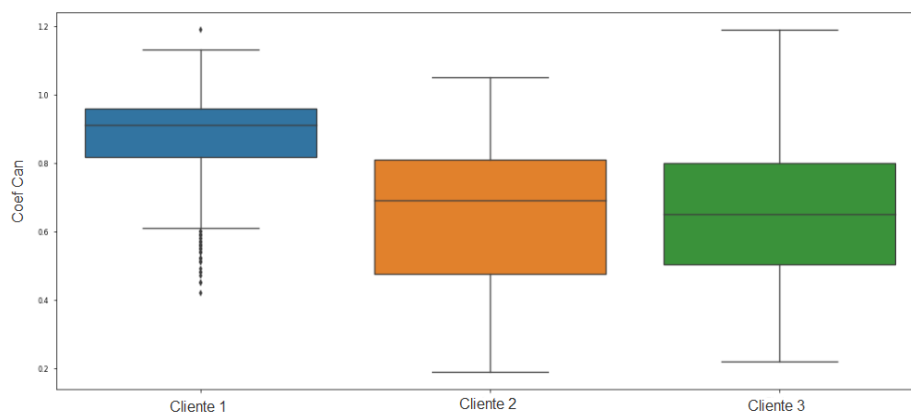


Figura 11. Coeficiente de canibalización según el “Cliente”. Fuente: elaboración propia

En las cajas de bigotes de la Figura 12 se observa la variable “Tipo de Oferta” con sus dos categorías “Amarre” y “Extracontenido”, se logran evidenciar valores más altos del coeficiente de canibalización si la oferta es un Extracontenido, siendo su media de 0,82, mientras que para el Amarre está en 0,44, lo que indica que las ventas de la referencia de línea se ven más impactadas si la oferta activada es un Extracontenido.

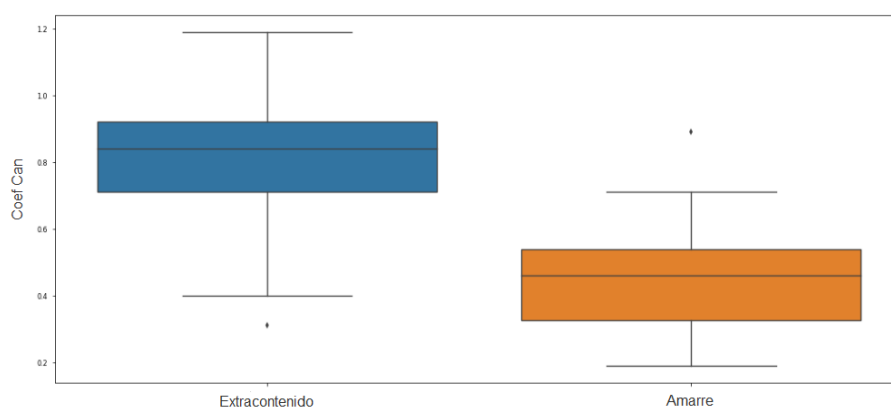


Figura 12. Coeficiente de canibalización según el “Tipo de Oferta”. Fuente: elaboración propia

Dependiendo el tipo de evento, si es “Convencional”(evento que busca incentivar la compra en segmentos focalizados sin presencia de publicidad) o “Triple A” (evento de alto impacto

para el cliente con presencia de publicidad) los valores del coeficiente de canibalización también presentan una diferencia como se observa en la Figura 13. La media para el evento convencional se encuentra en 0,78 mientras que para el evento triple A se encuentra en 0,55 lo que significa que se dejan de vender más unidades de línea cuando el evento del cliente es convencional.

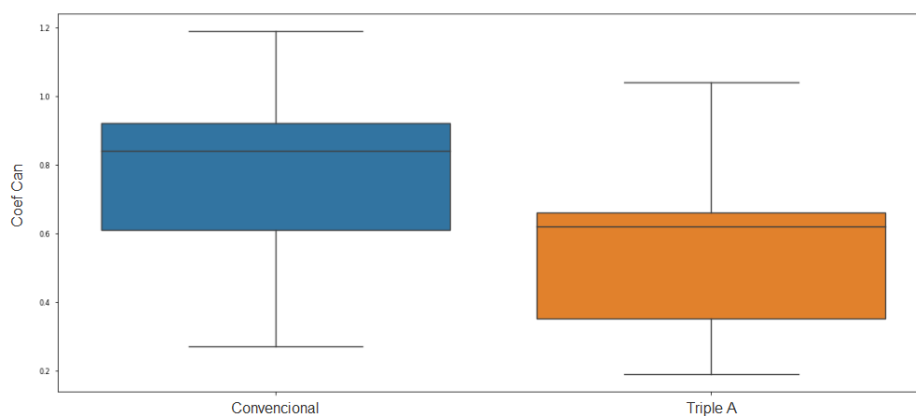


Figura 13. Coeficiente de canibalización según el “Tipo de evento”. Fuente: elaboración propia

5.4 Evaluación de modelos para la predicción de canibalización

Los datos de entrada para los modelos se obtuvieron de la base de datos CanibEvento.xlsx. la cual se encuentra disponible en el siguiente enlace <https://bit.ly/2BcAJPP>.

Teniendo en cuenta que la naturaleza de los datos es de corte transversal, y que se usarán técnicas de aprendizaje supervisado con el objetivo de hallar patrones de la información del pasado para retornar conclusiones sobre nuevos datos, el proyecto se centrará en análisis de modelos de regresión, dado que se desea determinar la relación entre la variable dependiente (canibalización) y las demás variables explicativas (tipo de cliente, tipo de oferta, tipo de evento).

Para la etapa de entrenamiento de los modelos de regresión a evaluar los cuales serán regresión lineal múltiple y árbol de decisión regresión, se tomó el 80% del total de registros

equivalente a 478 datos, y los datos de clasificación o prueba correspondieron al 20% restante, un porcentaje usual en el tratamiento de datos (Hyndman y Athanasopoulos, 2013). A cada uno de los modelos se le evaluará el rendimiento y con base en este se definirá cuál utilizar para finalmente llevar a cabo la predicción de la canibalización. En la figura 14 se muestran las librerías usadas para correr los dos modelos propuestos.

Modelos de Predicción - Canibalización evento

Regresión Lineal Múltiple

```
In [242]: ##Separación de Features (X1, X2,...,Xk) y Target (y)
X = train.loc[:,['TipodeCliente','TipodeOferta','TipodeEvento']].values
y = train.loc[:,['CanibalizacionEvento']].values
```

```
In [243]: # Particionamiento de datos Entrenamiento y Prueba
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=7)
```

```
In [244]: import matplotlib.pyplot as plt
```

```
In [245]: #Entrenamiento de modelo
from sklearn import linear_model
regLineal = linear_model.LinearRegression()
regLineal.fit(X_train, y_train)
```

```
Out[245]: LinearRegression(copy_X=True, fit_intercept=True, n_jobs=None,
normalize=False)
```

Árbol de Decisión Regresión

```
In [271]: pip install graphviz

Requirement already satisfied: graphviz in c:\users\acdflorez\appdata\local\continuum\anaconda3\lib\site-packages (0.13.2)
Note: you may need to restart the kernel to use updated packages.
```

```
In [272]: pip install pydotplus

Requirement already satisfied: pydotplus in c:\users\acdflorez\appdata\local\continuum\anaconda3\lib\site-packages (2.0.2)
Requirement already satisfied: pyparsing>=2.0.1 in c:\users\acdflorez\appdata\local\continuum\anaconda3\lib\site-packages (from pydotplus) (2.3.1)
Note: you may need to restart the kernel to use updated packages.
```

```
In [273]: ## Separación de Features (X1, X2,...,Xk) y Label (y)
X=train.loc[:,['TipodeCliente','TipodeOferta','TipodeEvento']].values
y = train.loc[:,['CanibalizacionEvento']].values
```

```
In [274]: ## Particionamiento de datos
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=0)
```

```
In [275]: #Entrenar modelo
from sklearn import tree
import matplotlib.pyplot as plt
arboldecisionreg = tree.DecisionTreeRegressor()
arboldecisionreg.fit(X_train , y = y_train)
predict = arboldecisionreg.predict(X_test)
accuracy_test = round(arboldecisionreg.score(X_test,y_test)*100,2)
accuracy_train = round(arboldecisionreg.score(X_train,y_train)*100,2)

print('train accuracy of decision tree classifier',accuracy_train)
print('test accuracy of decision tree classifier',accuracy_test)
```

Figura 14. Procesos ejecutados para cargue de modelos en Jupyter Fuente: elaboración propia

A continuación, se exponen los resultados para cada uno de los modelos.

5.4.1 Regresión Lineal Múltiple

El primer modelo planteado para la predicción de la canibalización fue **Regresión Lineal Múltiple**, el cual se representa matemáticamente con la expresión:

$$y_t = \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_3x_3 + e$$

Donde:

$y_t =$ *Canibalización proyectada*

$\beta_0 =$ *Intercepto*

$\beta_i =$ *Peso de la variable x_i*

$x_1 =$ *Tipo de Cliente*

$x_2 =$ *Tipo de Oferta*

$x_3 =$ *Tipo de Evento*

Los parámetros arrojados por el modelo fueron los siguientes:

$$y_t = 0.651 + 0.112x_1 - 0.373x_2 + 0.169x_3 + e$$

La variable independiente que mayor explica la canibalización de las referencias de línea es el Tipo de oferta (si es amarre o extracontenido) esta es X_2 , seguido de esta variable está X_3 , tipo de Evento (Convencional o Triple A) y finalmente la variable que menos aporta a la explicación de la canibalización es X_1 , Tipo de Cliente.

Las medidas de rendimiento obtenidas para este modelo se muestran en la figura 15.

```
In [246]: ## Coeficiente de Determinación
          regLineal.score(X_train, y_train)
Out[246]: 0.8749748798133731

In [247]: ## Exactitud
          print("Accuracy --> ", regLineal.score(X_test, y_test)*100)
          Accuracy --> 83.59777542556944
```

Figura 15. Medidas de rendimiento Regresión Lineal Múltiple para la predicción de la canibalización. Fuente: elaboración propia

Con estas medidas se puede identificar que las variables independientes explican el comportamiento de la variable dependiente en un 87.49%, presentado una exactitud en la predicción del 83.59%, esta última medida de rendimiento se calcula teniendo en cuenta el número de eventos que tuvieron una asertividad con un error admisible sobre el total de eventos.

5.4.2 Árbol de Decisión Regresión

El segundo modelo planteado para la predicción de la canibalización fue **árbol de decisión regresión**. Estos árboles se construyen usando algoritmos que buscan elegir la opción óptima en cada paso, con la intención de llegar a una solución general óptima.

El nodo raíz del árbol (nodo superior) corresponde a la variable con mayor peso en el modelo, es decir aquella que permite predecir mejor la variable respuesta, que para este caso es el tipo de

Oferta lo cual coincide con lo encontrado en el anterior modelo. Desde este nodo el árbol se divide en dos ramas donde cada nodo de ahí en adelante actúa como un caso de prueba para algunas de las variables usando un conjunto de reglas sucesivas que ayudan a tomar una decisión haciendo preguntas de tipo $X_k \leq C$? Donde X_k es la variable independiente y C el valor que ésta puede tomar.

El resultado de este modelo se muestra en la figura 16 en donde es necesario tener en cuenta lo siguiente:

- mse: error cuadrático medio (coef can real – coef can estimado)²
- samples: número de muestras que cumplen las características del nodo
- value: predicción para el coeficiente de canibalización

Out[279]:

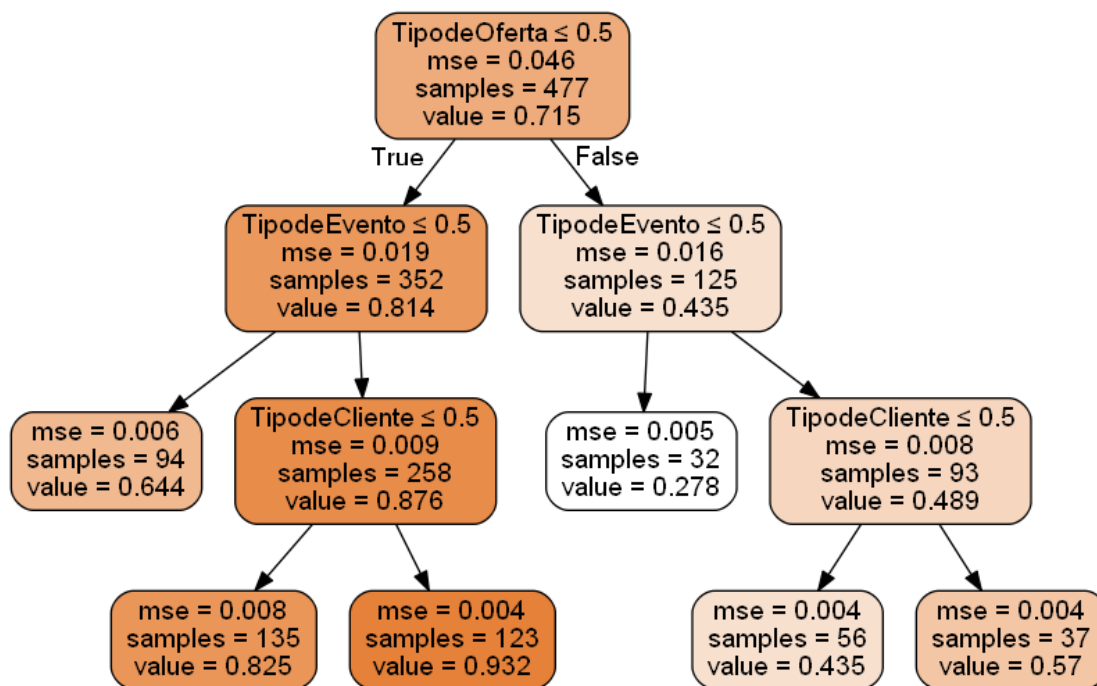


Figura 16. Árbol de decisión regresión para la predicción de la canibalización. Fuente: elaboración propia

Para la interpretación de la figura 16 se especifican los valores de cada variable.

Tipo de Oferta = (1 = *Amarre*), (0 = *Extracontenido*)

Tipo de Evento = (1 = *Convencional*), (0 = *Triple A*)

Tipo de Cliente = (1 = *Cliente1*), (0 = *Cliente2, Cliente3*)

A manera de ejemplo, para el caso donde el tipo de oferta sea un Extracontenido (tipo de oferta = 0) y este sea activado en un evento convencional (tipo de evento = 1) y para el tipo de cliente 1 (tipo de cliente = 1) de acuerdo al árbol de la figura 16, la predicción para el coeficiente de canibalización es 0,932.

El orden de importancia de las variables independientes que aportan al modelo de árbol de regresión se presenta en la Figura 17.

```
In [285]: importances = arboldecisionreg.feature_importances_

In [299]: df = pd.DataFrame({'Caracteristicas': ['TipodeEvento', 'TipodeOferta', 'TipodeCliente'],
                             'Importancia de Caracteristicas': importances})
ax = df.plot.barh(x='Caracteristicas', y='Importancia de Caracteristicas')
```

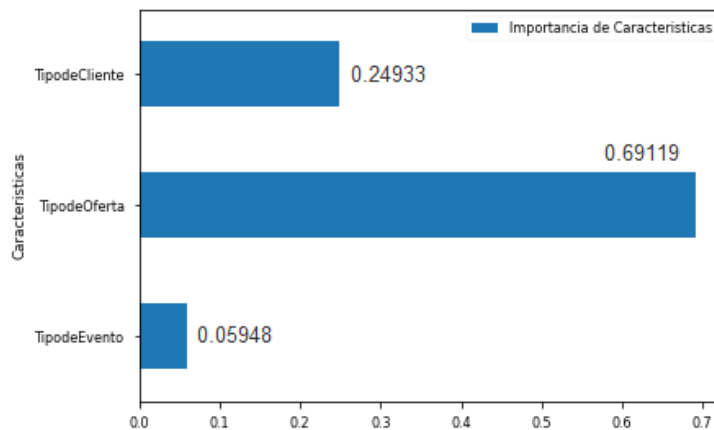


Figura 17. Importancias variables independientes en modelo de Árbol de decisión regresión. Fuente: elaboración propia

En la figura 18 se muestran las medidas de rendimiento obtenidas para este modelo.

```
In [276]: ## Coeficiente de Determinación
          arboldecisionreg.score(X_train, y_train)

Out[276]: 0.8775663602114226

In [277]: ## Exactitud
          print("Accuracy --> ", arboldecisionreg.score(X_test, y_test)*100)

Accuracy --> 82.85323833232765
```

Figura 18. Medidas de rendimiento Árbol de Decisión Regresión para la predicción de la canibalización. Fuente: elaboración propia

Estas medidas indican que las variables independientes están explicando el comportamiento de la variable dependiente en un 87.75%, presentado una exactitud en la predicción entregada del 82.85%, es decir, cerca del 83% de las predicciones estuvieron correctas.

Comparando las medidas de desempeño para cada modelo se encuentra un resultado muy similar de los coeficientes de determinación, al utilizar el modelo de regresión lineal múltiple o el árbol de decisión regresión, pero el indicador de exactitud es un poco mejor la predicción mediante la regresión lineal múltiple.

Tabla 7. Resultados medidas de desempeño modelos de predicción de la canibalización

	<i>Regresión Lineal Múltiple</i>	<i>Árbol de Decisión Regresión</i>
R²	0,874	0,877
Exactitud	83,597	82,853

Fuente: elaboración propia.

Ambos modelos de predicción generan buenos resultados de acuerdo con las medidas de rendimiento. Para el cumplimiento del objetivo de este proyecto cobra mayor relevancia la exactitud que se logra con los datos de prueba, ya que entre más cercano a la realidad sea el estimado entregado a la cadena de suministro, mejor esta se preparará para atender la demanda de los productos en oferta y de la referencia de línea impactada, minimizando el número de agotados y destrucciones. Por lo anterior, el modelo que se propone utilizar inicialmente por la compañía es la regresión lineal múltiple.

5.5 Conexión del modelo de canibalización con el proceso planear ventas y operaciones (PVO) de la compañía

La compañía objeto de estudio cuenta con un proceso integral, formal y de ejecución mensual que involucra niveles estratégicos del negocio para balancear la demanda del mercado y la oferta de la red de operaciones en diferentes horizontes de tiempo, lo cual facilita la toma de decisiones acerca de cuál podría ser la mejor combinación de mercados, productos y clientes a servir acorde con el plan de negocio. De acuerdo con el tipo de decisiones y su alcance, el negocio cuenta con los rangos de planeaciones que se muestran en la Figura 19 en su ciclo de PVO:

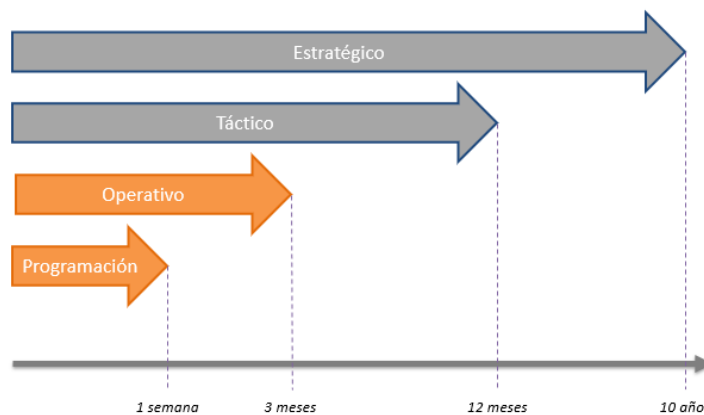


Figura 19. Rangos de planeaciones del ciclo PVO. Fuente: elaboración propia

La base para cada una de estas planeaciones es el plan de ventas, el cual es entregado en diferentes niveles de agregación de acuerdo con el tipo de planeación, de manera que entre más horizonte de tiempo sea entregado, mayor será la agregación en los datos.

Es en la planeación operativa y en la programación donde la información de **canibalización** cobra mayor relevancia , aquí los planes se entregan en un mayor detalle llegando a semana, material, cliente y cedi y es donde más novedades y reprocesos se han presentado debido a información errada entregada a las operaciones frente al impacto de las referencias de línea por presencia de ofertas.

Con el plan de ventas entregado se genera un plan maestro de operaciones que contempla los tiempos de reposición de cada uno de los nodos (abastecimiento, producción y distribución) y las restricciones existentes en términos de lotes mínimos.

En la planeación **Operativa** se usa dicho plan, en el cual se detallan las necesidades de recursos, de manera que en cada nodo se evalúa disponibilidad y se confirman o dan señales oportunas a los procesos en caso de no poder responder a alguna necesidad.

En el nodo de *abastecimiento* se visualiza cuál será el consumo estimado de cada material. Para el caso de las referencias canibalizadas que se integrarán al plan de ventas cada que se corra el modelo de canibalización, se evidencia una disminución en el consumo de las materias primas asociadas, pero también se encuentra la necesidad de contar con los materiales para la producción de las ofertas. por lo que se requiere revisar los lead times del proveedor y dar las señales pertinentes de aumentos y disminución en las órdenes de compra.

En el nodo de *producción* se revisa la necesidad de turnos, y se dan señales frente a los niveles de ocupación de los recursos restrictivos de la planta. Así, ara las semanas donde hay oferta

y canibalización, se debe tener claridad de cuántas unidades realmente se estima disminuir de la referencia de línea para hacer la respectiva nivelación de recursos.

En el nodo de *distribución* se validan necesidades de vehículos y niveles de ocupación en centros de distribución.

En la **Programación** se definen las operaciones de abastecimiento, producción y distribución a ejecutar día a día, es decir, es el lanzamiento de las órdenes y allí se ejecuta el qué, cuánto, cuándo y dónde de cada material.

Para este horizonte de la programación debe existir una excelente sincronización entre el lanzamiento de órdenes de producción y distribución de las referencias de línea y de las ofertas, de tal forma que la disminución de inventarios de la referencia de línea y el aumento en el inventario de la oferta se haga de acuerdo con las señales entregadas por el plan y el modelo combinado con las señales del consumo real. Así mismo cuando esté finalizando la oferta se debe contar con suficiente inventario de la referencia de línea para retomar inventarios en los cedís y no generar desabastecimientos. Para ello es necesario realizar diariamente un monitoreo a las ventas para validar que no hayan cambiado los supuestos iniciales plasmados en el plan maestro de operaciones. En caso de requerirse algún cambio o ajuste, se actualiza el plan y se montan nuevas órdenes que respondan a los cambios dados.

Los resultados del modelo de canibalización y su integración oportuna al plan de ventas de la compañía es fundamental para el proceso y traerá ventajas notables en términos de disminución de destrucciones de producto terminado y mejoras en el nivel de servicio de las referencias de línea ya que se espera una mejor estimación en los coeficientes de canibalización lo cual permita una mejor preparación y planeación como se plasmó anteriormente para los nodos de abastecimiento, producción y distribución.

5.6 Validación de la eficiencia del modelo de canibalización propuesto

Como una primera validación a la eficiencia del modelo, se evaluó un evento reciente de 2 semanas de uno de los clientes que hizo parte del piloto de este proyecto. Se compara la canibalización entregada por el equipo comercial, la canibalización arrojada por el modelo y el resultado real de la canibalización. Las estimaciones y el resultado final se muestran en las tablas 8, 9 y 10.

Tabla 8. Coeficiente de canibalización estimado por equipo comercial

Cliente	Tipo de Oferta	Tipo de Evento	Unidades planeadas vender de la oferta	Pronóstico de venta referencia de línea	Venta estimada por el comercial para la referencia de línea	Unidades estimadas por el comercial que se dejarán de vender de la referencia de línea	Coeficiente de canibalización estimado por el comercial (<i>CoefCan</i>)
Cliente 2	Extracontenido	Triple A	2.000	2.000	-	2.000	1,00
Cliente 2	Amarre	Triple A	8.450	5.300	3.610	1.690	0,20

Fuente: elaboración propia.

Para este evento triple A el equipo comercial por cada 100 unidades que se proyectaban vender del extracontenido, estimaba dejar de vender la misma cantidad de la referencia de línea. A diferencia del amarre, que por cada 100 unidades que estimaba vender de la oferta, estimaba dejar de vender 20 unidades de la referencia de línea. En la tabla 8 se muestran las estimaciones proporcionadas por el equipo comercial de la compañía.

Por su parte, el modelo de regresión lineal múltiple propuesto predijo para este evento un menor coeficiente de canibalización con la venta del extracontenido que lo estimado por el equipo comercial. Con el modelo se estima que, por cada 100 unidades vendidas de la oferta, se dejarán de vender 65 unidades de la referencia de línea. Para el amarre la predicción del modelo estima 0,07 puntos por encima del coeficiente que lo estimado por el equipo comercial. La tabla 9 muestra los resultados arrojados por el modelo de canibalización propuesto.

Tabla 9. Coeficiente de canibalización estimado por el modelo

Cliente	Tipo de Oferta	Tipo de Evento	Unidades planeadas vender de la oferta	Pronóstico de venta referencia de línea	Venta estimada por el modelo para la referencia de línea	Unidades estimadas por el modelo que se dejarán de vender de la referencia de línea	Coeficiente de canibalización estimado por el modelo (CoefCan)
Cliente 2	Extracontenido	Triple A	2.000	2.000	700	1.300	0,65
Cliente 2	Amarre	Triple A	8.450	5.300	3.019	2.282	0,27

Fuente: elaboración propia.

Finalmente, en los resultados reales del evento realizado durante 2 semanas en el cliente 2 se encontró que la referencia de línea se dejó de consumir en 52 unidades por cada 100 unidades consumidas del extracontenido y para el caso del amarre por cada 100 unidades vendidas de este, se dejaron de vender 33 unidades de la referencia regular, lo que significa que generó mayor disminución en las ventas de la referencia de línea la activación del extracontenido. La tabla 10 muestra las ventas reales de la referencia de línea con el coeficiente de canibalización real presentado en el evento.

Tabla 10. Coeficiente de canibalización real

Cliente	Tipo de Oferta	Tipo de Evento	Unidades vendidas de la oferta	Pronóstico de venta referencia de línea	Venta real referencia de línea	Unidades reales dejadas de vender referencia de línea	Coeficiente de canibalización real (CoefCan)
Cliente 2	Extracontenido	Triple A	2.000	2.000	960	1.040	0,52
Cliente 2	Amarre	Triple A	8.450	5.300	2.512	2.789	0,33

Fuente: elaboración propia.

A continuación, se muestra en la tabla 11 el MAD (Desviación media absoluta) el cual mide el tamaño del error de los coeficientes de canibalización estimados por el equipo comercial versus los arrojados por el modelo con respecto al resultado real.

Tabla 11. Comparación del MAD de los coeficientes de canibalización estimados vs los reales

Cliente	Tipo de Oferta	Tipo de Evento	MAD del coeficiente de canibalización estimado por el equipo comercial	MAD del coeficiente de canibalización estimado por el modelo
Cliente 2	Extracontenido	Triple A	0,48	0,13
Cliente 2	Amarre	Triple A	0,13	0,06

Fuente: elaboración propia.

De acuerdo con los resultados expuestos en la tabla 11 se evidencia una mayor aproximación al dato real aplicando el sugerido del modelo obteniendo un menor error. Lo anterior confirma que el modelo puede estar arrojando buenos resultados.

Finalmente, para el monitoreo de los resultados del modelo se propone la incorporación de un nuevo indicador el cual se revise con expertos del área comercial cada que termine un evento.

El indicador propuesto es asertividad de la canibalización y su fórmula será la siguiente.

$$\text{Asertividad Canibalización} = 1 - \frac{\text{ABS (Plan de venta estimado referencia línea - venta real referencia línea)}}{\text{Venta real referencia línea}}$$

Para la evaluación a la prueba realizada con el evento reciente y con base a los datos de las tablas 9 y 10, el resultado de asertividad a la canibalización se muestra en la tabla 12.

Tabla 12. Comparación del MAD de los coeficientes de canibalización estimados vs los reales

Cliente	Tipo de Oferta	Tipo de Evento	Plan de venta estimado referencia de línea	Venta real referencia de línea	Asertividad de la estimación de la canibalización
Cliente 2	Extracontenido	Triple A	700	960	0,73
Cliente 2	Amarre	Triple A	3.019	2.512	0,80

Fuente: elaboración propia.

5.7 Percepción de expertos frente al modelo de canibalización propuesto

Con la finalidad de validar la eficiencia del modelo de canibalización se desarrolló una entrevista semiestructurada aplicada a dos (2) expertos con roles afines al tema desarrollado en este proyecto. Uno de los expertos es jefe de planeación de demanda con foco a los planes de venta y a la sensibilidad comercial, el segundo experto es el jefe de monitoreo de la cadena de suministro con foco ciencia de datos.

Se preguntó al experto 1 acerca de sus observaciones frente a los resultados encontrados en el análisis de canibalización por semana y canibalización por evento, al respecto, indicó que al analizar los datos de canibalización por semana se encuentra un rango demasiado amplio en los porcentajes que no permite tener una conclusión relevante que ayude a la toma de decisiones, por tanto, el experto 1 manifestó que la propuesta de canibalización por evento supone una solución interesante a un problema que se tiene continuamente en la empresa, en la medida que “al tener estabilidad y confianza en el porcentaje de canibalización del evento se pueden concentrar las acciones del equipo de ventas sobre la repartición de la venta y de la oferta por semana y, con base a ello, sacar conclusiones sobre el comportamiento de las referencias de línea.

De igual modo, se preguntó a experto 1 sobre si las conclusiones del modelo y del comportamiento de los datos son coherentes con lo que sucede en la realidad de la empresa, al respecto indicó que este modelo demuestra con datos claros que algunas variables que se consideraban tenían una influencia en la canibalización, como lo son la marca o la región donde se vende, en realidad no son relevantes. A partir de esta afirmación, el experto 1, expresó que se debía continuar con la recolección de datos para corroborar esta hipótesis, afinar los resultados y ampliar las conclusiones a otros segmentos de mercado.

Finalmente, se preguntó al Experto 1 acerca de las recomendaciones para la réplica y escalamiento futuro del modelo, a lo cual indicó que “los resultados son un gran aporte a la actividad de la planeación de la demanda, la cual busca articular las operaciones de la cadena de suministro con el consumo real, por tanto, una mejor asertividad en el modelo de canibalización implicará una mejor anticipación en las disminuciones e incrementos del consumo de las referencias de línea y así, un mejor uso del capital de trabajo y una disminución en el riesgo de destrucción de producto no comercializable por vida útil corta.

De acuerdo con el experto 1: “la idea debe seguir evolucionando con el hallazgo de nuevas variables que tengan influencia en la magnitud de la canibalización. Algunas de ellas podrían ser el momento del mes en que se realiza (quincena), la presencia de asesoras que apoyen el evento o el hecho de que la oferta esté presente en un catálogo físico o virtual. Asimismo, se debe ampliar el uso de esta herramienta a la planeación de acciones estrategias e innovadoras” (Comunicación personal, experto 1, 2020).

Por otra parte, se preguntó al experto 2 acerca de su percepción sobre el desarrollo planteado y la aplicación de ciencia de datos para resolver el problema, al respecto indicó que “es relevante el uso de la metodología de análisis de componentes principales, ya que se empezó con paradigmas asociados a la cantidad de variables a incluir encontrando que pocas variables independientes explicaban la variable de respuesta”. El experto 2, resaltó que se usaron modelos no complejos obteniendo resultados de asertividad superiores al 70% lo cual, según él, es bueno para este primer análisis exploratorio.

En esta misma línea se preguntó al Experto 2, sobre las recomendaciones para la replicación futura del modelo, a lo cual indicó que, para la implementación masiva del modelo se debe hacer un foco en el método de cálculo del patrón de comportamiento normal de la referencia de línea sobre la cual se calcula la venta perdida. Para el Experto “este es el punto más sensible en los resultados y cuando se amplíe a otras líneas de negocio es probable que se encuentre la necesidad de utilizar modelos diferentes para su determinación” (Comunicación personal, Experto 2, 2020).

De acuerdo con los resultados obtenidos del modelo frente a las proyecciones realizadas con el proceso actual de la compañía y, teniendo en cuenta los análisis de los expertos se observa una mejora significativa del coeficiente de canibalización, y aunque es apresurado concluir que este es

el modelo definitivo, es claro que al incluir procesos estructurados para tomar decisiones en relación a variables complejas de predecir se robustece el proceso, y se le da mayor credibilidad al plan de ventas entregado. Se deben seguir monitoreando y calibrando los resultados y sin duda se abre el panorama para seguir indagando en alternativas y herramientas ofrecidas por la ciencia de datos.

6. Conclusiones

La técnica de análisis de componentes principales permitió evaluar el aporte de cada variable independiente al modelo mediante su correlación con la variable objetivo, además, de detectar colinealidad entre algunas de las variables, pasando de 9 variables inicialmente planteadas como posibles variables explicativas a 3, lo cual simplificó la complejidad del espacio muestral. Puede concluirse que la aplicación de esta técnica en el análisis exploratorio de datos fue de gran importancia para el proceso ya que desde la construcción del modelo predictivo se tenía certeza del impacto de las variables independientes en el resultado final de la canibalización.

De igual modo, se estructuró la forma de analizar el impacto de las ofertas en las referencias de línea pasando de semana a evento. Este hallazgo fue muy significativo para el desarrollo del proyecto ya que le dio norte y coherencia a la investigación pues la variación de los datos disminuyó en un 85% y los enfoques se trasladaron en entender el evento completo resaltando que la distribución del evento por semana estaba dada de acuerdo con la negociación y acuerdos que se hicieran con cada cliente.

Otra de las conclusiones a las cuales se llega, es que los modelos multivariados de predicción de la canibalización corridos y evaluados presentaron medidas de rendimiento favorables, coeficientes de determinación de 0,87 y resultados de exactitud de 82% y 83%. La empresa continuará sus análisis y tomará decisiones con el modelo que obtuvo mejor medida de exactitud, es decir, mayor número de aciertos en la predicción, hecho que genera tranquilidad ya que se trata de un modelo estructurado y pensado que reúne diferentes variables, el cual presenta una mejoría con relación al proceso de estimación actual de un 69%. Este es sin duda un buen

acercamiento de la herramienta que se desea tener cuando se migre masivamente con el proceso en los demás segmentos del negocio.

Así pues, contar con una herramienta de predicción de canibalización para el negocio es importante, pero más importante aún es incorporar la información entregada por este al proceso formal de planeación de ventas y operaciones, empezar a generar confianza con los equipos de operaciones frente a la información entregada en el plan de ventas y generar sinergias que faciliten el flujo de los procesos y permitan cumplir con todos los objetivos transversales de la compañía, mejorar el nivel de servicio y disminuir el nivel de destrucciones.

Se propone el indicador de asertividad de la canibalización con el fin de medir y monitorear la eficiencia del modelo. Con el resultado obtenido en la primera evaluación a un evento frente al resultado de la canibalización se resalta el buen resultado entregado por el modelo, que, aunque presenta oportunidades de mejora en su asertividad, muestra tener un mejor resultado que el proceso que se está haciendo actualmente para la entrega de la información.

Finalmente, tras el diseño del modelo propuesto para la predicción de la canibalización la compañía contará con un proceso robusto y estructurado, tendrá esta primer aproximación que le permitirá disminuir la brecha en los errores de la predicción de la canibalización lo cual era una preocupación debido a la frecuencia en las desviaciones de la información que se entregaba, se deja a la compañía un modelo que debe seguir nutriéndose y alimentándose de información para que el sistema continúe aprendiendo de los patrones, tendencias y relación de los datos y así mejore cada vez más las predicciones, aprovechando de este modo el potencial que demuestra la ciencia de datos y pueda lograr aplicarse en otras áreas.

7. Recomendaciones

Para la aplicación masiva del modelo es necesario retar la disponibilidad de variables, iniciar con la documentación y seguimiento de nuevos datos que de acuerdo con lo conversado con el equipo interdisciplinario que hace parte de este proyecto, puedan influir en el resultado de la canibalización, buscando de esta manera incluir más variables en el modelo que puedan ser analizadas y modeladas usando técnicas de Machine Learning.

Es necesario sistematizar el proceso del cálculo del patrón de comportamiento normal de la referencia de línea, ya que al replicar el modelo a referencias de otros segmentos y marcas del negocio dependiendo del comportamiento identificado en las series de tiempo se deben aplicar diferentes modelos de pronóstico.

Un futuro trabajo podría estar enfocado en el desarrollo de un modelo que prediga la canibalización generada debido a la introducción de productos nuevos, a entender cómo y de qué forma la llegada de productos nuevos afecta la venta de las referencias de línea, por cuánto tiempo y si es permanente o no en el tiempo dicho impacto.

Referencias bibliográficas

- Berry, W. Bliemel, F. (1974). Selecting exponential smoothing constants: an application of pattern search. *Nt. J. Prod. Res.* 12(4). 483-499
- Billah, B. King, M. Snyder, R. Koehler, A. (2006). Exponential smoothing model selection for forecasting. *International Journal of Forecasting* 22 (). 239 – 247
- Brueggen, A. Grabner, I. Sedatole, K. L. (2014). The Folly of Forecasting: The Effects of Sales Forecast Accuracy and Bias on Inventory and Production Decisions under Aggregated and Disaggregated Forecasting Regimes.
- Díaz, E. Consuegra, D. Esteban, A. (2014). Canibalización del servicio: ¿influye la distribución multicanal en la prestación incorrecta del servicio? *DOCFRADIS.* 4(1). 1-27.
- Esmailikia, M. Fahimnia, B. Sarkis, J. Govindan, K. (2016). A tactical supply chain planning model with multiple flexibility options: An empirical evaluation. *Annals of Operations Research.* DOI: 10.1007 / s10479-013-1513-2
- Fariás, F. 2009. La epistemología de las ciencias sociales en la formación por competencias *Cinta moebio* 34: 58-66.
- Galvanize. (2020). Identifying outliers. Recuperado de: https://help.highbond.com/helpdocs/analytics/142/user-guide/en-us/Content/analytics/analyzing_data/identifying_outliers.htm
- Hazbun, C. (2014). Estudio del efecto de canibalización en una tienda de retail al incorporar una nueva tienda de la misma cadena utilizando información transaccional. [Tesis Maestría]. Universidad De Chile

- Herrera, C. Mazón, B. (2016). Implementación de un módulo de análisis estadístico y predictivo para agricultura utilizando bigdata y machine learning, integrado al sistema iotmach. [Tesis]. UTMACHALA, Machala, Ecuador
- Hvolby, H. Steger, K. (2015). Managing Cannibalization of Perishable Food Products in the Retail Sector. *Procedia Computer Science*. 64. 1051-1056.
- Hyndman, R.J. Athanasopoulos, G. (2013) Forecasting: principles and practice. Recuperado de <http://otexts.org/fpp/2/3>.
- Ivanov, S. (2008). Conceptualizing Cannibalization: The Case of Tourist Companies. *Tourism Today*. 8(1). 7-32
- Kerin, R.A., Harvey, M.G. and Rothe, J.T. (1978). Cannibalism and new product development. *Business Horizons*. 25-31
- Lomax, W. Hammond, K. East, R. Clemente, M. (1997). The measurement of cannibalization. *Journal of product & brand management*. 6(1). 27-39 33
- López, R. (2006). Operaciones de almacenaje. Madrid: Paraninfo.
- Madhani, P. (2017). Logistics and Marketing Integration: Enhancing Competitive Advantages. *The IUP Journal of Management Research*. 16(3). 7-29
- Mason, C.H. Milne, G.R. (1994). An approach for identifying cannibalization within product line extensions and multi brand strategies”, *Journal of Business Research*. 31(). 63-70
- Novalés, A. (2010). Análisis de Regresión. [Tesis]. Universidad Complutense, Madrid, España.
- Ozgun, C. Colliau, T. Rogers, G. Hughes, Z. Bennie, E (2017). MatLab vs. Python vs. R. *Journal of data science*: DOI. 15. 355-372.

- Ragharan Srinivasan, S. , Ramakrishnan, S. y Grasman, S. (2005), "Incorporando modelos de canibalización en el pronóstico de la demanda", *Marketing Intelligence & Planning*, 23(5). 470-485
- Rasmussen, R. (2003). On time series data and optimal parameters. *Omega* 32 (). 111 – 120
- Roman, D. (2016). Un análisis de las promociones de ventas al consumidor: los cambios que supone la introducción de los smartphones en esta táctica de marketing. [Tesis doctoral]. Universitat Central De Catalunya.
- Salazar, G. (2011). Gestión estratégica de la demanda. Bogotá: Colegio de Estudios Superiores de Administración CESA.
- Sánchez, W. (2019). Desarrollo de modelo de “Canibalización” de órdenes en las Alianzas estratégicas de Alsea Colombia. [Tesis]. Universidad Santo Tomás. Bogotá, Colombia.
- Sipper, D. Bulfin, R. (1998). Planeación y Control de La Producción. McGRAW-HILL: México D.F
- Van, H. Srinivasan, S. Dekimpe, M. (2010). Estimating Cannibalization Rates for Pioneering Innovations. *Marketing Science*. 29(6). 1024-1039
- Vereecke A, Vanderheyden K, Baecke P, Van Steendam T. (2018). Mind the gap : assessing maturity of demand planning, a cornerstone of S&OP. *International Journal of Operations & Production Management. Emerald*. 38(8):1618–39.
- Wheelen, T. Hunger, D. (2007). Administración Estratégica y Política de Negocios 10ed. Pearson Educación. Ciudad de México.