



UAI
UNIVERSIDAD ADOLFO IBÁÑEZ



Desarrollo de herramienta para pronóstico de demanda de productos NARTD en Coca Cola Chile S.A.

**Proyecto de Pasantía
Facultad de Ingenierías y Ciencias
Universidad Adolfo Ibáñez**

Profesor Guía: Raimundo Sánchez

Sección 4

Santiago, 2023

Rodrigo Rueda

Contenido

Resumen Ejecutivo	3
Abstract	4
Introducción.....	5
a) Contexto de la empresa	5
b) Estructura Organizacional y Área de Interés	7
c) Contexto del problema y oportunidad	9
Objetivos	13
a) Objetivo general	14
b) Objetivos específicos.....	14
Medidas de Desempeño	14
Estado del Arte	16
a) Caso 1: Modelo Abraham y Lodish.....	16
b) Caso 2: Modelos de pronósticos de series temporales.....	19
c) Caso 3: Regresión Lineal Múltiple.....	23
Solución	24
a) Alternativas de Solución	24
b) Solución escogida.....	26
Desarrollo de la solución	26
Metodología.....	26
a) Metodología de Trabajo.....	26
b) Metodología de Solución.....	27
c) Metodología de Validación de Resultados.....	38
Plan de Implementación.....	39
a) Fase de Inicio	39
b) Fase de Desarrollo	39
c) Fase de Aplicación y Ajuste	39
d) Fase de Cierre.....	40
Análisis de Riesgo	40
Evaluación Económica	43
Resultados	45
a) Resultados del Desarrollo de la Solución	45
b) Evaluación de Métricas de Desempeño.....	50
Conclusiones	54
Referencias y Anexos.....	56

Resumen Ejecutivo

En el presente informe se despliega un proyecto que busca implementar una mejora en cómo Coca Cola Chile S.A. hace seguimiento y pronósticos de sus ventas. El problema central radica en los altos errores de pronóstico que ha tenido el modelo actual para los últimos periodos del año 2023 que datan sobre las 5 Millones de Cajas Unitarias con un error porcentual medio absoluto del 12,47% y el hecho de que no exista una herramienta personalizada para Chile actualmente.

Las causas principales de este problema son que el modelo de la compañía únicamente incorpora series temporales, lo que limita su capacidad para capturar la complejidad del entorno y el mercado chileno. Por otro lado, estos pronósticos son armados fuera del país, lo que restringe su alcance, crea una falta de adaptabilidad a las particularidades locales y dificulta la captura de factores específicos que pueden influir en las ventas en el país. Este enfoque genérico y centralizado ha contribuido a errores sustanciales en los pronósticos y destaca la necesidad de una solución más ajustada a las operaciones en Chile.

Para abordar esta problemática, el proyecto propone el armado y la implementación de una herramienta de pronóstico complementaria al modelo actual que incorpore variables relevantes que tengan un impacto en las ventas y logre explicar la variación de los resultados operacionales entre diferentes periodos. La iniciativa busca superar las limitaciones del modelo actual.

La herramienta confeccionada fue un modelo de regresión lineal múltiple que arrojó un MAPE de 6,02% para pronósticos del año 2023, lo que superó en 6,45% al modelo que usa actualmente la compañía y superó en un 2,45% el objetivo general propuesto inicialmente.

Este proyecto tiene el potencial para transformar positivamente el seguimiento de la demanda, la gestión de pronósticos y la toma de decisiones estratégicas en Coca Cola Chile S.A.

Palabras clave: Coca Cola Chile, pronósticos de ventas, modelos estadísticos, modelos de pronóstico, eficiencia operativa, error de pronóstico.

Abstract

The present report unfolds a project aimed at implementing an improvement in how Coca Cola Chile S.A. monitors and forecasts its sales. The central issue lies in the high forecasting errors that the current model has experienced for the last periods of 2023, reaching around 5 million Unit Case errors with an average absolute percentage error of 12.47%, coupled with the absence of a customized tool for Chile at present.

The primary causes of this problem stem from the company's model incorporating solely time series, limiting its ability to capture the complexity of the Chilean environment and market. Additionally, these forecasts are generated outside the country, restricting their scope, creating a lack of adaptability to local peculiarities, and complicating the capture of specific factors that can influence sales in the country. This generic and centralized approach has contributed to substantial errors in forecasts, emphasizing the need for a solution tailored to operations in Chile.

To address this issue, the project proposes the development and implementation of a complementary forecasting tool to the current model, incorporating relevant variables that impact sales and explaining the variation in operational results across different periods. The initiative aims to overcome the limitations of the current model.

The crafted tool was a multiple linear regression model that yielded a MAPE of 6.02% for forecasts in 2023, surpassing the company's current model by 6.45% and exceeding the initially proposed overall goal by 2.45%.

This project has the potential to positively transform demand monitoring, forecasting management, and strategic decision-making at Coca Cola Chile S.A.

Keywords: Coca Cola Chile, sales forecasting, statistical models, forecasting models, operational efficiency, forecasting error.

Introducción

a) Contexto de la empresa

The Coca Cola Company es una empresa líder en la distribución y comercialización de bebidas no alcohólicas en el mercado, con presencia en más de 200 países y una muy amplia gama de productos. La primera Coca Cola fue inventada en 1886 en Atlanta, Georgia, para ser vendida inicialmente como un jarabe refrescante que afirmaba tener propiedades curativas con ingredientes clave: hojas de coca y nuez de cola. Posteriormente, se consolidó como una bebida azucarada que hasta el día de hoy es una de las más consumidas en el mundo

No fue hasta 1892 que se fundó The Coca-Cola Company (TCCC) y se expandió de forma rápida e internacional durante el siglo XX. A medida que la compañía se expandía, se fue convirtiendo en unas de las marcas más reconocidas y consumidas del siglo, lo que abrió apertura para nuevos productos como: Coca Cola Dietética, Coca Cola Light, Fanta, Sprite, entre muchos otros.

La escalada de la compañía ha sido hasta el punto de patrocinar eventos como los Juegos Olímpicos en 1972, la Copa Mundial de la FIFA desde 1966 y la Eurocopa desde 1976 hasta el 2020 aproximadamente. Hoy, se afirma que cualquier individuo en América visualiza al menos tres anuncios de cualquier producto de The Coca Cola Company al día.

En Latino América, The Coca Cola Company se encuentra en: Argentina, Bolivia, Brasil, Chile, Colombia, Costa Rica, República Dominicana, Ecuador, El salvador, Guatemala, Honduras, México, Nicaragua, Panamá, Paraguay, Perú, Uruguay y Venezuela. Cerca del 27% de los volúmenes de venta de toda la compañía provienen de Latino América. Dentro de estos países, los que tienen mayor peso para la empresa son: México, Brasil, Chile y Argentina; debido a la sólida presencia que tiene la marca en estas regiones y el amplio consumo que existe de sus productos.

En Chile, la compañía se introdujo por primera vez en 1940, desde entonces se ha consolidado como una de las marcas más reconocidas y populares del país.

La compañía opera a nivel global con un modelo de negocios basado en franquicias: es decir, Coca Cola otorga contratos de franquicia a empresas embotelladoras que cumplen términos y condiciones bajo los cuales pueden producir y distribuir los productos de la empresa en sus territorios asignados. Aunque son empresas independientes, estas trabajan de la mano con The Coca Cola Company para tomar decisiones estratégicas como: introducir nuevos productos al mercado, realizar campañas de marketing, asegurar que se cumplan los estándares de calidad y sustentabilidad.

En Chile, Coca Cola opera con dos embotelladores:

1. **Coca Cola Andina S.A.**
2. **Coca Cola Embonor S.A.**

Coca Cola Andina está encargada de las operaciones de TCCC en las regiones de Antofagasta, Atacama, Coquimbo, San Antonio, Cachapoal, Aysén, Magallanes y la Región Metropolitana.

Coca Cola Embonor opera en las regiones de Arica y Parinacota, Tarapacá, Valparaíso, Libertador General Bernardo O'Higgins, Maule, Ñuble, Bío Bío, Araucanía, Los Ríos, Los Lagos e Isla de Pascua.

Los volúmenes de venta de Andina corresponden al 63% del total nacional, mientras que Embonor compone cerca del 37% del volumen de TCCC en Chile.

Los embotelladores distribuyen los productos a través de diferentes canales de venta, los cuales son los siguientes:

1. **Canal Tradicional (Traditional Trade):** Son todos los almacenes, tiendas de abarrotes y establecimientos minoristas. Representa cerca del 60% de los volúmenes de venta.
2. **Canal Supermercados (Modern Trade):** Son todos los supermercados grandes del mercado: Cencosud, Walmart, Unimarc. Estos representan cerca del 20% de los volúmenes de venta.
3. **Canal Comer y Beber (On Premise):** Son lugares donde las bebidas se consumen en donde se adquieren, como restaurantes, bares, puntos de comida rápida, lugares de hotelería. Representa aproximadamente el 10% de las ventas.

4. **Otros:** Corresponden a ventas que se hacen a empresas, eventos, máquinas expendedoras en lugares estratégicos y ventas online.

La línea de productos que TCCC produce y comercializa es la de bebidas no alcohólicas como refrescos, aguas embotelladas, jugos y otras bebidas sin alcohol. Esta categoría en la industria tiene un nombre y se le conoce como: NARTD, que significa “Not Alcoholic Ready to Drink” (Listo para beber no alcohólico), es decir, todas las bebidas que están listas para consumir sin necesidad de alguna preparación adicional y que no contienen alcohol. De los productos NARTD que comercializa TCCC, hay dos subcategorías:

1. **Sparkling ó SSDs:** Corresponden a todas las bebidas gaseosas.
2. **Stills:** Corresponden a todas las aguas planas, aguas saborizadas, jugos e isotónicos.

Dentro de Sparkling podemos encontrar marcas conocidas como Coca Cola, Sprite, Fanta. Mientras que en Stills se hayan productos como: Benedictino, Vital, Aquarius, Del Valle, Powerade, entre otros.

Hoy en día Coca Cola Chile S.A. cuenta con más de 140 mil clientes en sus canales y los resultados operacionales del cierre del segundo trimestre de este año 2023 alcanzaron alrededor de 238 millones de cajas unitarias, lo que corresponde aproximadamente a 1.352 millones de litros.

b) Estructura Organizacional y Área de Interés

La casa matriz de Coca Cola Chile, se encuentra en Av. Kennedy 5757 piso 10, Las Condes, Santiago. Las áreas que se encuentran dentro de la empresa están orientadas principalmente a la consolidación comercial de la marca y se pueden encontrar segmentos como: recursos humanos, customer service, marketing, finanzas, franquicias, entre otros.

Para efectos de este informe, se pondrá énfasis en el Área de Franquicias (Franchise), la cual forma parte del equipo comercial de Coca Cola y está encargada de las operaciones de la compañía. El nombre “Área de Franquicias” es debido al modelo de negocios que tiene TCCC con los franquiciados, en este caso, con Andina y Embonor en Chile. El Área de Franchise viene a ser la contra parte de los franquiciados y se encarga que la operación completa que ejecutan los embotelladores asegure que un consumidor final pueda ir a

cualquier punto de venta y adquirir un producto de la marca. Esto significa: asegurar suministros de la cadena productiva, tener discusiones de precios con las franquicias, hacer análisis y reportería de las ventas, generar y hacer estimación de la demanda de los productos, entender el road to market de los productos y tener buena relación con los clientes. Franchise también tiene un rol posicionado en las agendas que la compañía desarrolla con diferentes comunidades, stakeholders, ministerios y el impacto que tiene el negocio con las leyes reguladoras de sustentabilidad (agua, plástico, energía).

En resumen, Franchise maneja toda la relación comercial con Andina y Embonor, para que los embotelladores ejecuten el negocio tal como Coca Cola Chile lo necesite.

A continuación, se adjunta el organigrama del Área de Franquicias para tener una representación gráfica y mejor comprensión de la organización:

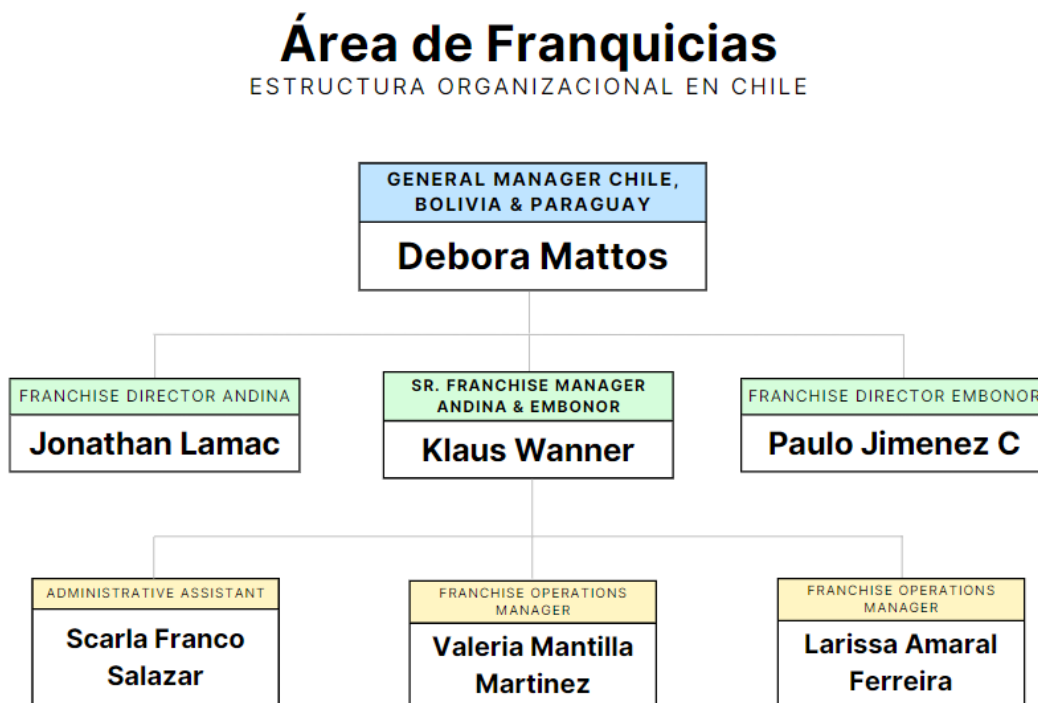


Ilustración 1: Organigrama Franchise Chile.

c) Contexto del problema y oportunidad

La forma de hacer pronósticos de demanda, es decir, la estimación de los volúmenes de venta en la compañía es a través de un proceso llamo “Rolling Estimate Process” (RE), el cual es un proceso periódico y continuo de estimación y actualización de las ventas de la compañía. El RE se hace de forma anual, mensual, semanal y diaria.

Para cada año, la compañía hace un “Annual Bussines Plan” (ABP), que, en este contexto, es un Rolling madre que se usará para fijar una meta y proyectar las ventas durante todos los periodos del año, poniendo hincapié en periodos de ventas mensuales y trimestrales.

Existen oportunidades de mejora con el modelo de pronóstico actual que la compañía tiene.

La forma en cómo se hace la estimación de los volúmenes de venta es a través de cálculos sencillos basados en la historia, es decir, únicamente incorporando tendencias de ventas, sin necesidad de contar con un soporte estadístico más complejo. En ese contexto, este modelo no incorpora variables que pueden tener un impacto con el resultado operacional, como, por ejemplo: variables macroeconómicas locales como el PIB o el Consumo Privado.

Del mismo modo, el no considerar variables adicionales genera falta de causalidad, es decir, es más difícil identificar con claridad las razones de los cambios repentinos en la demanda.

Además, existen periodos de alta concentración de ventas y alto estrés operativo, como lo son los meses de Septiembre y Diciembre, donde por lo general el modelo tiende a tener mayores márgenes de fallo.

En suma a esto, esta estimación no cuenta con una herramienta customizada para Chile, sino, que se realiza de forma general para todos los mercados de latino américa, lo que, del mismo modo, genera brechas a la hora de su confección debido a la falta de contexto social, político y económico con el mercado local.

Por otro lado, esta estimación se realiza una vez al mes, siendo entregada en ocasiones en la segunda semana del mes en curso, lo que limita el alcance de la herramienta en caso de que se necesite evaluar simulaciones o estrategias previas al mes en porvenir.

Asimismo, este modelo presenta errores de pronóstico significativos, los cuales se pueden visualizar a continuación:

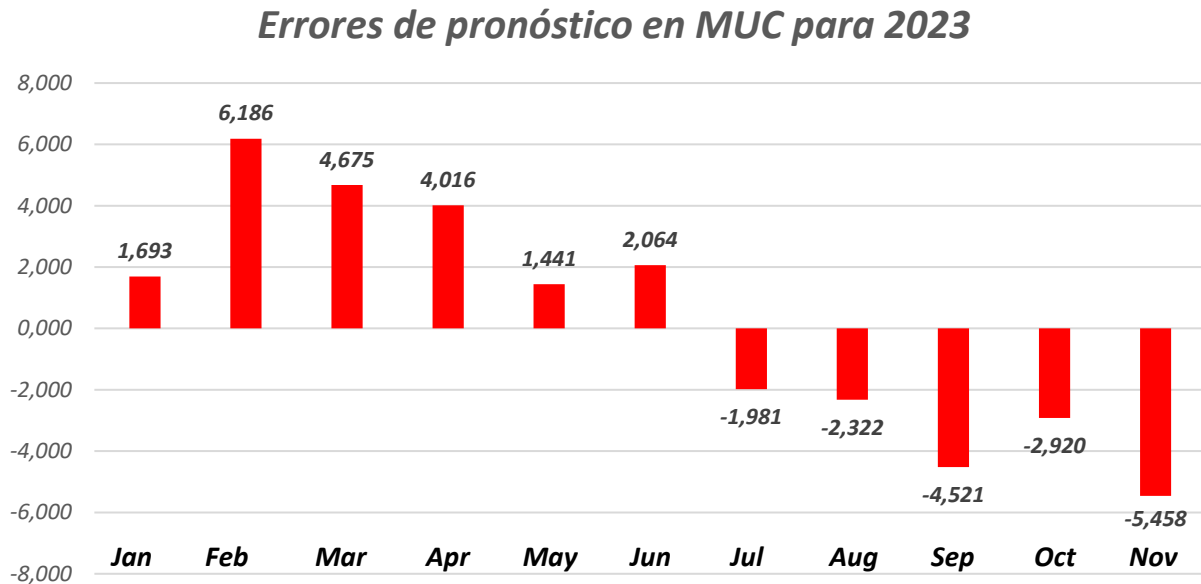


Ilustración 2: Error de pronóstico por periodo en MUC vs RE.

En el gráfico anterior, se pueden visualizar los errores de pronóstico que han existido para los periodos del año 2023 en Millones de Cajas Unitarias. Como se puede apreciar, en algunos periodos se sobreestima la demanda, mientras que en otros se subestima bastante. El hecho de que exista una diferencia de casi 4 a 6 millones de cajas unitarias con el resultado real, es un número que resulta alarmante. Las expectativas de la compañía es que exista una diferencia máxima de 2 MUC para cada periodo.

De forma complementaria, también se pueden examinar los resultados del modelo actual con el **error porcentual medio absoluto**, el cual nos indica el porcentaje promedio de cuanto están fallando las predicciones versus los resultados reales. Esto servirá para evaluar las posibles oportunidades de mejora.

Así, la fórmula del MAPE, es la siguiente:

$$MAPE = \frac{\sum_{i=1}^n \left| \frac{x_i - y_i}{x_i} \right|}{n}$$

Donde:

- x_i es el resultado real en el periodo i
- y_i es el pronóstico para el periodo i
- n es el número de observaciones

Debido a que se maneja información sensible y bajo un acuerdo de confidencialidad con la empresa, se ha procedido a la parametrización de los datos antes de su despliegue. A partir de este punto, para referirnos a los volúmenes de venta de la compañía, utilizaremos el término “Unidades de Transacción Parametrizadas” (UTP), que también se referirá a los volúmenes de venta de los productos NARTD de la compañía.

Para este año 2023 se tienen los siguientes resultados en Millones de Unidades de Transacción Parametrizadas (MUTP):

2023 MUTP	Pronóstico	Resultado Real	Error de Pronóstico Absoluto	Error Porcentual Absoluto
Enero	58,02	61,41	3,39	5,5%
Febrero	62,01	74,38	12,37	16,6%
Marzo	75,58	84,93	9,35	11,0%
Abril	50,88	58,91	8,03	13,6%
Mayo	48,20	51,08	2,88	5,6%
Junio	58,36	62,49	4,13	6,6%
Julio	52,56	48,60	3,96	8,2%
Agosto	56,79	52,15	4,64	8,9%
Septiembre	78,44	69,40	9,04	13,0%
Octubre	60,73	54,89	5,84	10,6%

Tabla 1: Pronósticos vs Resultado Operacional Real.

En la tabla anterior se visualiza el pronóstico del modelo actual para cada periodo, el resultado real de las ventas, el error de pronóstico absoluto para cada observación y el porcentaje que representa ese error del resultado real.

Seguido de esto, se puede graficar el pronóstico de la demanda contra el resultado real para cada periodo, lo cual queda adjunto a continuación:

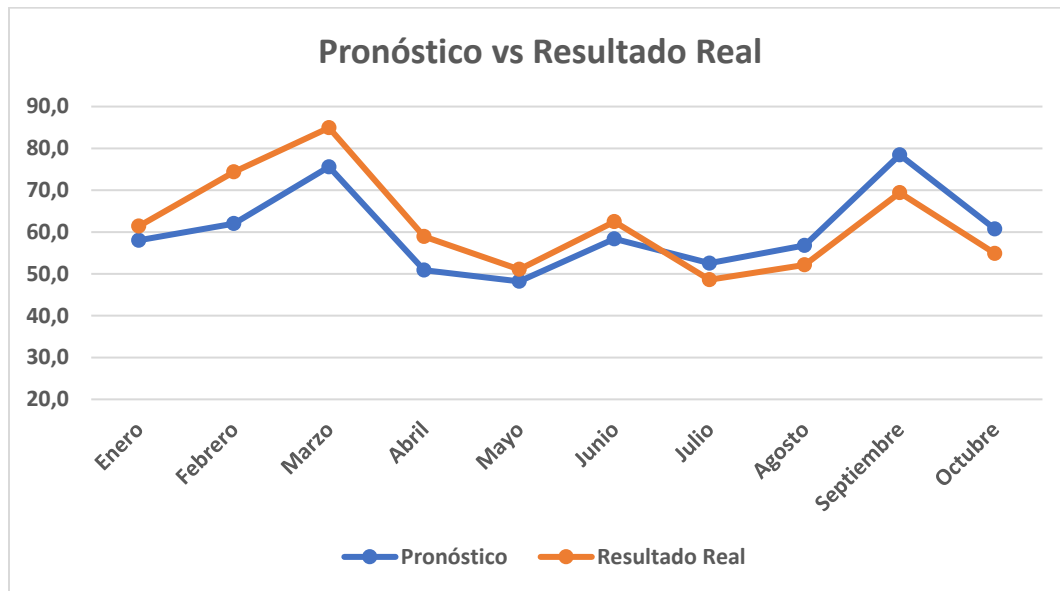


Ilustración 3: Gráfico de Pronóstico vs Resultado Real.

Con esto, se calculó el MAPE, el cual arrojó el siguiente resultado:

MAPE	12,47%
-------------	---------------

Tabla 2: MAPE del modelo actual.

En promedio, las predicciones del modelo están desviadas un 12,13% de los valores reales de las ventas en NARTD. Este porcentaje representa la precisión promedio en términos porcentuales.

De este modo, podemos concluir que, si bien los resultados del modelo no son negativos, existe una posibilidad significativa de mejorar su precisión, rendimiento y que este sea útil para las necesidades de la compañía en Chile.

Para entender y ordenar las causas mencionadas del dolor, se confecciona el siguiente diagrama causa efecto:

Diagrama de Causa-Efecto

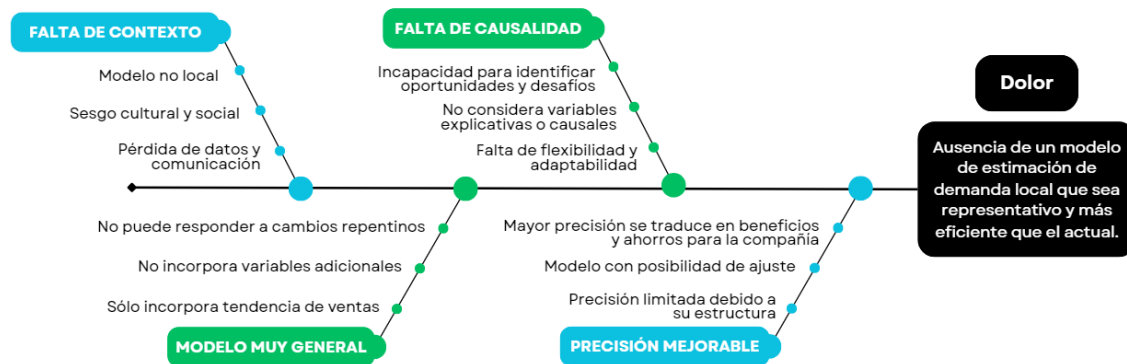


Ilustración 4: Diagrama de causa-efecto.

De este modo, se puede mencionar que la oportunidad del proyecto radica en hacer una mejora en cómo se lleva a cabo el seguimiento y la estimación de demanda de productos NARTD de forma customizada para Chile.

Promover la mejora de este pronóstico es una excelente oportunidad por diferentes razones:

La idea es crear un impacto en la toma de decisiones; un pronóstico mejor previsto y más preciso va a permitir a la compañía adaptabilidad y respuestas eficaces a cambios repentinos en la demanda. Esto se traduce en tomar decisiones más informadas sobre su gestión de inventario, asignación de recursos, planificación de producción, estrategias de marketing y más. Esto, del mismo modo puede llevar a una mejora de eficiencia operativa y ahorro en costos: La reducción de errores en los pronósticos y posible consideración de variables explicativas puede resultar en ahorros significativos a lo largo del tiempo, como, por ejemplo, un impacto en la reducción por inventario obsoleto o excesivo.

Objetivos

El principal deseo de este proyecto es: Crear una herramienta complementaria al modelo de pronóstico actual para los productos NARTD (Not Alcoholic Ready to Drink) de The Coca Cola Company en Chile, que permita explicar el crecimiento de las ventas y logre hacer pronósticos para el futuro. La idea es generar un impacto en la toma de decisiones que la compañía ejecute en diferentes aspectos del negocio.

a) Objetivo general

Desarrollar una herramienta complementaria al modelo de pronóstico actual que incorpore variables adicionales relevantes y logre mantener un error porcentual medio absoluto con una diferencia máxima de 4% con respecto al modelo actual, dentro de un plazo de 6 meses.

b) Objetivos específicos

1. **Recopilar y limpiar datos históricos:** Asegurar de contar con toda la información histórica para procesarla, limpiarla y poder manipularla.
2. **Identificar variables relevantes:** Analizar cuáles son las variables que tienen mayor impacto con la venta de los productos NARTD y determinar cuáles son más relevantes para un buen pronóstico.
3. **Desarrollo y entrenamiento de un modelo más preciso:** Usar herramientas y algoritmos de pronóstico avanzados y ver cual se adapta mejor a este caso.
4. **Validar el modelo:** Aplicar técnicas de validación para verificar el rendimiento del modelo y ajustarlo según sea lo necesario.

Medidas de Desempeño

Primeramente, debemos calcular el error porcentual medio absoluto para el modelo actual y el que se desarrollará. La fórmula de este error es la siguiente:

- **MAPE – Error porcentual medio absoluto**

$$MAPE = \frac{\sum_{i=1}^n \left| \frac{x_i - y_i}{x_i} \right|}{n}$$

Donde:

x_i es el resultado real en el periodo i .

y_i es el pronóstico para el periodo i .

n es el número de observaciones.

En consecuencia, se calculará la diferencia de estos dos errores, la cual será la principal medida de desempeño de este proyecto, ya que corroborará el objetivo propuesto.

- **Diferencia de errores**

$$\Delta MAPE = |MAPE_1 - MAPE_2|$$

Donde:

$MAPE_1$ es el resultado del error absoluto medio del modelo actual.

$MAPE_2$ es el resultado del error absoluto medio del modelo que se construirá.

De forma complementaria, también se añadirán dos medidas de desempeño para calcular un error de pronóstico por periodo; esto nos permitirá visualizar de forma más granular los resultados del modelo actual y el que se desarrollará.

- **Error de pronóstico absoluto**

$$e_t = |Y_t - \hat{Y}_t|$$

Y_t es el valor observado o valor real en el tiempo t .

\hat{Y}_t es el valor pronosticado en el tiempo t .

- **Error de pronóstico porcentual**

$$ea_t(\%) = \frac{|Y_t - \hat{Y}_t|}{Y_t} * 100$$

Y_t es el valor observado o valor real en el tiempo t .

\hat{Y}_t es el valor pronosticado en el tiempo t .

Estado del Arte

Para llevar a cabo un desarrollo e implementación efectiva de este proyecto, se realizó una investigación sobre casos en los que se abordó una problemática similar a la existente y planteada en este informe. Dado lo anterior, a continuación, se expondrán 3 casos con sus respectivas soluciones.

a) Caso 1: Modelo Abraham y Lodish

El primer caso es sobre una empresa perteneciente al rubro de *retail* en Chile y consta en averiguar cuál es el impacto que tienen las promociones en la industria del *retail*, ya que estas afectan en la demanda a corto plazo y otros factores como la rentabilidad. Para evaluar este impacto, se utilizó un modelo llamado “Abraham y Lodish” que se centra en evaluar las promociones a corto plazo y analizando los productos a nivel categoría. Para poder evaluar este impacto, se usa una metodología que calcula una línea base de ventas utilizando regresiones lineales.

Primeramente, se hace un ajuste por estacionalidad natural de las ventas independiente de la presencia de promociones. Se crea un promedio móvil de 52 semanas; “Tendencia”: $T(t)$

Luego, se utiliza un modelo de regresión para desestacionalizar y remover las promociones. Matemáticamente el modelo es el siguiente:

$$\log\left(\frac{S(t)}{T(t)}\right) = \alpha + \beta P(t) + \sum_{i=1}^m a_i X_i(t) + \sum_{j=1}^{12} r_j D_j(t) + \sum_{l=1}^L h_l H_l(t)$$

Donde:

$S(t)$: Ventas semanales de la categoría en la semana t .

$T(t)$: Tendencia de las ventas en la semana t .

$P(t)$: Precio promedio categoría en la semana t .

$X_i(t)$: Porcentaje de las ventas realizadas por la promoción i en la semana t .

$D_j(t)$: Variable binaria, la cual toma valor 1 cuando la semana de estudio pertenece al j .

$H_i(t)$: Variable binaria, la cual toma valor 1 cuando la semana de estudio presenta un feriado.

m : Número de promociones.

L : Número de semanas con promoción.

Además, se necesita calcular el coeficiente de estacionalidad normalizado I_j para el periodo j . La fórmula es:

$$I_j = \frac{12e^{rj}}{\sum_{k=1}^{12} e^{rk}}$$

Luego, se hacen cero los parámetros correspondientes a las promociones para eliminar su efecto y se recalcula la tendencia inicial, ya que incluía el efecto de las promociones. El procedimiento anterior se repite 4 veces

Posterior a esto, se eliminan los datos de las semanas en las que hay promociones: Para cada serie semanal desestacionalizada, se divide por la tendencia y por el coeficiente de estacionalidad.

Luego, se eliminan *outliers*: eliminando datos que queden fuera de intervalos de confianza.

Se suavizan los datos, se re-estacionalizan y se hace una reinserción de la tendencia: esto se hace rellenando datos borrados usando promedios ponderados de las semanas anteriores y se multiplican los datos por la tendencia y por el coeficiente de estacionalización.

Se ajustan los quiebres de *stock*, estos datos se distinguen donde no existieron ventas.

Finalmente, se hace un ajuste de la línea base por factores específicos de mercado, es decir, se comparan las ventas del local con otros locales que no tuvieron promoción en esa fecha.

Sean:

$A(t)$: factor de ajuste a calcular.

$B_i(t)$: línea base final para la tienda i en el tiempo i .

$b_i(t)$: línea base desestacionalizada para la tienda i en el tiempo t .

$I(t)$: índice de estacionalidad para el tiempo t .

N : número de tiendas sin promoción en el tiempo t .

El ajuste requiere que todas las tiendas que no promocionaron en la semana de estudio, las sumas de sus ventas sean igual a la suma de las estimaciones de la línea base.

$$\sum_{i \in N} B_i(t) = \sum_{i \in N} S_i(t)$$

Por definición, $B_i(t) = A(t)I(t)b_i(t)$, luego:

$$A(t) = \frac{\sum_{i \in N} S_i(t)}{\sum_{i \in N} I(t)b_i(t)}$$

Una vez calculada la línea base, el volumen incremental de las ventas es fácilmente calculable restando la línea base a las ventas reales para cada semana en la que no hubo promoción.

De este modo, se adjunta el diagrama de flujo para cada uno de los pasos:



Ilustración 5: Diagrama de Flujo del Modelo de Abraham-Lodish.

En este caso se concluye que este modelo posee grandes ventajas para el *retail* y para las distintas marcas ya que entrega información útil para medir la eficacia y la rentabilidad de las promociones. Del mismo modo, se puede evaluar si la empresa estima de manera certera o no las ventas incrementales de sus productos.

b) Caso 2: Modelos de pronósticos de series temporales

En este caso se hace un estudio para un operador logístico que posee 23 puntos de dispensación zonal para sus puntos de venta, donde sus compras surgen de la revisión mensual del inventario y la consolidación de los pedidos en la bodega, los cuales son despachados posteriormente hacia estos puntos. El problema: elevada cantidad de producto que no rota por razones internas y externas del proceso. El objetivo es usar diferentes modelos de pronósticos que utilizan series temporales para encontrar cual es el que se adecua mejor a la necesidad del operador.

Este estudio aborda 4 modelos de pronóstico distintos, los cuales serán descritos a continuación:

1. **Pronóstico por Media Móvil:** Este método es una técnica que implica calcular el promedio de un número fijo de observaciones anteriores en un periodo de tiempo específico para prever valores futuros.

La fórmula matemática es la siguiente:

$$M_T = \frac{x_T + x_{T-1} + x_{T-2} + \dots + x_{T-N+1}}{N}$$

Esto se puede escribir cómo:

$$M_T = \frac{1}{N} \sum_{i=0}^{N-1} X_{T-i}$$

Donde:

M_T : es el valor pronosticado en el periodo T .

N : es el número de periodos que se están promediando.

X_{T-i} : son los valores observados en los N periodos anteriores.

2. **Suavización Exponencial Simple:** Es una técnica que asigna pesos exponenciales decrecientes a observaciones pasadas. Es adecuada para datos con patrones estacionarios y sin tendencias significativas.

La fórmula matemática es la siguiente:

$$S_T = \alpha x_T + (1 - \alpha) S_{T-1}$$

Donde:

S_T : es el pronóstico al final del periodo T .

S_{T-1} : es el pronóstico del periodo anterior.

x_T : es la demanda real en el periodo T.

α : constante de suavización exponencial que debe ser ajustado.

Para la optimización de la constante de suavización α se sugieren valores entre 0,01 y 0,30 o entre 0,05 y 0,20 debido al contexto de la demanda intermitente, eligiendo el que minimice los errores de pronóstico.

3. Método Holt-Winters: Es una técnica de pronóstico que considera tres componentes principales en los datos de series temporales: tendencia, estacionalidad y error. Es útil especialmente cuando los datos tienen patrones de tendencia y estacionalidad.

Para esta técnica se deben calcular 3 factores:

- **Factor de nivel para el periodo t:**

$$S_t = \alpha \frac{x_t}{C_{t-N}} + (1 - \alpha)(S_{t-1} + G_{t-1})$$

- **Factor de tendencia para el periodo t:**

$$G_t = \beta(S_t - S_{t-1}) + (1 - \beta)G_{t-1}$$

- **Factor estacionario para el periodo t:**

$$C_t = \gamma \frac{x_t}{S_t} + (1 - \gamma)C_{t-N}$$

Donde:

α : primera constante de suavización con $0 < \alpha < 1$.

β : segunda constante de suavización con $0 < \beta < 1$.

γ : tercera constante de suavización con $0 < \gamma < 1$.

N : duración de la estación.

x_t : demanda en el periodo t.

Finalmente, se pronostica la demanda de cualquier periodo futuro $t + \tau$ del siguiente modo:

$$F_{t,t+\tau} = (S_t + \tau G_t + C_{t-N})$$

- 4. Método de Croston:** Se utiliza para pronosticar la demanda intermitente o esporádica, donde las observaciones de demanda son erráticas o irregulares con grandes fluctuaciones en el tiempo.

Cuando se usa este modelo, se usa una variable binaria para definir la existencia u ausencia de la demanda.

Esto se define:

$$x_t = y_t z_t$$

Donde x_t es la demanda observada en el periodo t , z_t es el tamaño de la transacción en el periodo t , y y_t es la variable binaria definida como:

$$y_t \begin{cases} 1 & \text{si existe transacción.} \\ 0 & \text{si no existe transacción.} \end{cases}$$

Si $y_t = 0$, no hay demanda y si $y_t = 1$ hay transacción, luego se definen las ecuaciones:

$$\hat{Z}_t = \alpha x_t + (1 - \alpha)\hat{Z}_{t-1}$$

$$\hat{n}_t = \alpha n_t + (1 - \alpha)\hat{n}_{t-1}$$

Donde:

n_t : periodos desde la última transacción.

\hat{n}_t : valor de n al final del periodo t .

\hat{Z}_t : valor estimado de z al final del periodo t .

Los valores \hat{Z}_t y \hat{n}_t al inicio de la simulación se pueden asumir como el promedio de los datos históricos.

Luego, el pronóstico para el periodo $t + 1$ se calcula cómo:

$$\hat{x}_{t+1} = \frac{\hat{Z}_t}{\hat{n}_t}$$

En este caso, para obtener resultados interpretables, se redistribuyeron productos en 4 categorías distintas según sus porcentajes de rotación.

Luego de ejercicios experimentales, se concluye que generalizar un modelo de pronóstico para todos los productos ocasiona un distanciamiento con la demanda real, en la gran mayoría de los casos. Ocasionalmente, al actualizar la información, la elección del método varía con respecto a un periodo de tiempo pasado, por lo tanto, el analista debe hacer constantes actualización de los datos y corroborar que tipo de pronóstico arroja un error menor.

c) Caso 3: Regresión Lineal Múltiple

El tercer y último caso, manifiesta que hay una carencia práctica en el pronóstico de *stock* de piezas de repuesto para equipos médicos en Cuba, haciendo hincapié en que las soluciones actuales no logran mejorar la exactitud de estos pronósticos y en consecuencia se entorpece o atrasa el uso de estos.

De este modo, proponen un modelo llamado “MPREDSTOCK” para el proceso de predicción del *stock* de estas piezas de repuesto mediante la Regresión Lineal Múltiple como método de solución.

El primer paso para su confección fue hacer una selección de variables explicativas del *stock* anual, esto a partir de un análisis de correlación bivariada. De este análisis se seleccionaron las siguientes variables: frecuencia de fallas de la pieza (X_{1i}), confiabilidad operacional (X_{2i}) y disponibilidad técnica del equipo (X_{3i}).

El algoritmo parte de $k + 1$ variables cuantitativas, siendo Y , la variable de respuesta (*stock* anual) y $X_{1i} + X_{2i} + \dots + X_{ki}$ las variables explicativas, implementando el método de mínimos cuadrados ordinarios (MCO) para estimar parámetros poblacionales en la Regresión Lineal Múltiple ($\beta_0, \beta_1, \beta_2, \beta_3$).

De este modo, la fórmula matemática del modelo es la siguiente:

$$Y = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i}$$

Donde:

Y : Corresponde a la variable de respuesta, *stock* anual.

X_{1i}, X_{2i}, X_{3i} : Corresponden a las variables explicativas.

$\beta_0, \beta_1, \beta_2, \beta_3$: Son los coeficientes por estimar.

La interacción de cada uno de los componentes que conforman el modelo a partir del método Regresión Lineal Múltiple permite generar una salida que es plan de *stock* de piezas al año consecutivo al actual, con el objetivo de satisfacer los reportes de piezas de repuesto de equipos médicos, ya sea para mantenimiento, reparación o cambio.

Se concluye que la experimentación realizada permitió comprobar la contribución del modelo a la mejora de la exactitud de las predicciones del *stock* de piezas de repuesto para equipos médicos de Cuba.

Solución

a) Alternativas de Solución

Tomando en cuenta todas las investigaciones mencionadas se pueden desprender ventajas y desventajas de cada una de estas para abordar la problemática en Coca Cola Chile.

El primer caso, una investigación realizada por la Pontificia Universidad Católica de Chile, propone armar una regresión lineal a través del modelo Abraham Y Lodish para ver cómo impactan las promociones en las ventas incrementales de la empresa. Este modelo incorpora una serie de pasos para ajustarse a la realidad del mercado, desestacionalizar datos, considerar el *stock* y adaptarse a otros mercados. La ventaja de este modelo es que permite evaluar las promociones en los productos de una manera cuantitativa para tomar una mejor decisión estratégica y hacer evaluación periódica sin grandes costos asociados. La principal desventaja de este modelo es que se necesita una gran cantidad de información para llevarlo a cabo, es un modelo complejo de armar, tiene una serie de pasos reiterados a seguir y además se utilizan softwares especializados como lo es Stata y Matlab; estos softwares son caros y resulta difícil que las empresas los adquieran únicamente con este propósito. Por último, este modelo no se ajusta de lleno a la problemática de este informe: hacer una predicción explicativa de las ventas, sin considerar promociones.

El segundo caso, investigación realizada por la revista Biotecnología en el Sector Agropecuario y Agroindustrial de la Universidad de Cauca, propone hacer una evaluación

con diferentes modelos de pronósticos que utilizan principalmente técnicas estadísticas de suavización y descomposición de series temporales para hacer pronósticos. La principal ventaja de estos modelos es que sólo se necesitan series temporales para confeccionar su armado, es decir, la recopilación de la información a manipular es mucho más expedita en el caso que no existan otros componentes a considerar. Otras ventajas de estos modelos es que son simples de entender e implementar, tienen flexibilidad para capturar tendencias, estacionalidades y demandas intermitentes. La desventaja es que cada modelo presenta resultados que pueden ser coherentes para un producto y no para otro, lo que obliga al analista a ejecutar cada modelo con cada producto y entender cuál se ajusta mejor según lo certera que sean las predicciones. Otra desventaja que tienen estos modelos es que no consideran variables adicionales que pueden tener relación y afectar el resultado real de los valores observados.

El tercero caso, investigación realizada por la Revista Cubana de Ciencias Informáticas de la Universidad de las Ciencias Informáticas, propone un modelo de Regresión Lineal Múltiple como método de solución para el pronóstico del *stock* de piezas de repuesto de equipos médicos. Este modelo incorpora un análisis de correlación bivariada para seleccionar indicadores que tengan mayor influencia con el *stock* anual de las piezas. Estas variables seleccionadas se utilizaron en el modelo para generar la variable de respuesta “*stock* anual” y satisfacer los reportes de piezas de repuesto de equipos médicos. La principal ventaja de este modelo es su índole explicativa, es decir, permite modelar la relación entre la variable de respuesta y múltiples variables explicativas, lo que puede proporcionar información sobre cómo estas variables afectan la cantidad de stock. Además, es un modelo flexible, puede manejar múltiples variables explicativas, lo que permite capturar la complejidad de las relaciones entre el stock y factores relevantes. Por último, es altamente mejorable, si el modelo no se ajusta adecuadamente a los datos, se pueden incluir o excluir interacciones entre las variables de tal modo que se llegue a una mejor predicción. Como desventaja principal, se puede mencionar que este modelo asume linealidad entre las variables, si la verdadera relación es no lineal, el modelo puede no ajustarse bien a los datos. Además, el modelo puede ser sensible a valores atípicos (*outliers*) en la estimación de los coeficientes, afectando la validez del modelo.

b) Solución escogida

Teniendo en consideración los puntos a favor y en contra que presentan las alternativas planteadas, se escoge proceder con la tercera alternativa, la cual implica confeccionar un modelo de Regresión Lineal Múltiple con variables explicativas asociadas a las ventas de la compañía. Esto, debido a su facilidad, versatilidad y viabilidad económica: Se cuenta con las herramientas, los recursos, la información, el conocimiento y el software necesario para llevar a cabo el armado de este modelo. Adicional a esto, es una solución que aborda un punto importante de la problemática: incluir variables explicativas que pueden tener impacto en el resultado de las ventas.

Desarrollo de la solución

A continuación, se despliega todo lo que es el desarrollo de la solución escogida para abordar la problemática inicial de este proyecto.

Metodología

a) Metodología de Trabajo

El problema por abordar tiene como foco desarrollar una herramienta complementaria de pronóstico de demanda en Chile que sirva de apoyo y agilice las tomas de decisiones.

Para esto, se ha realizado un entendimiento de los procesos y factores que pueden influir en las ventas de la compañía con el propósito de diseñar un modelo que se adapte a estas necesidades.

Así mismo, se confeccionaron los siguientes pasos a seguir para implementar la solución:

- 1. Levantamiento de la información:** Recopilación de datos disponibles en empresa y variables que influyen el impacto de las ventas.
- 2. Preparación de los datos:** Seleccionar, limpiar y transformar toda la información en un formato adecuado para su modelado.

3. **Armado del modelo:** Aplicar la técnica de modelado seleccionada, Regresión Lineal Múltiple, con los datos extraídos.
4. **Entrenamiento:** Calibrar y ajustar parámetros del modelo para mejorar su rendimiento. Se realizarán ajustes según sea necesario para mejorar los resultados.
5. **Validación del modelo y análisis de confiabilidad:** Aplicación de las medidas de desempeño seleccionadas para validar el modelo de pronóstico y corroborar que está alineado con el objetivo general.
6. **Implementación:** Implementar el modelo en el entorno operativo y monitorear su rendimiento.

b) Metodología de Solución

Para hacer el armado del modelo, se hizo selección de 3 tipos de variables:

1. **Variable por Modelar:** Volúmenes de venta de la compañía en cajas unitarias (UC).
2. **Variables Externas:** Variables macroeconómicas y variables asociadas al clima.
3. **Variables de Ejecución:** Días efectivos de venta y métricas de desempeño en el mercado de NARTD.

Esta selección de variables se debe a que el equipo de trabajos las considera influyentes en el resultado operacional.

La extracción de la **variable a modelar** se hizo de la base de datos de TCC, la cual opera a nivel global y es llamada NSR. Esta base de datos registra todas las transacciones de la compañía a nivel país y embotellador. Desde aquí se puede extraer información como volúmenes físicos de venta, métricas de ingresos netos y brutos, métricas que comparan volúmenes de periodos pasados, información de venta por producto: categoría, segmento y formato del producto, ventas por canal, ventas por embotellador y por país.

El proceso de cómo se actualiza la información en NSR se muestra a continuación:

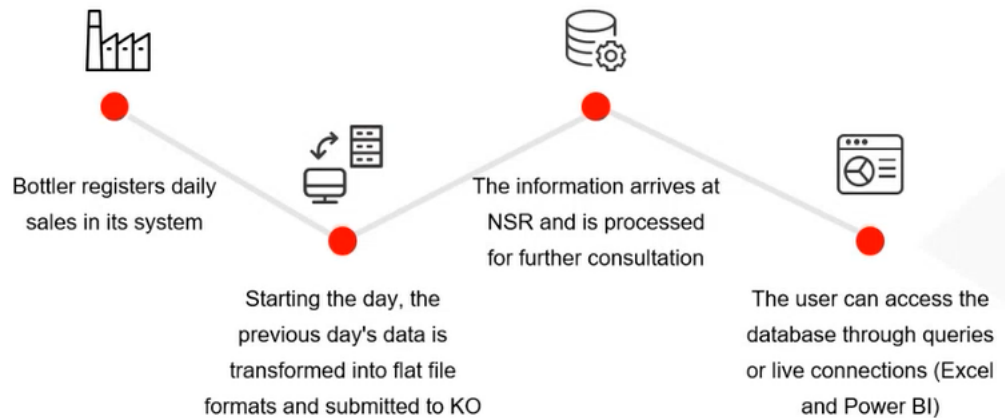


Ilustración 6: NSR Overview.

Desde aquí, se hizo consulta a la base de datos para hacer extracción de los volúmenes de venta mensuales de la compañía en UC (cajas unitarias) para el total de NARTD en Chile a partir del año 2013, es decir, el total de cajas unitarias vendidas por Andina y Embonor en todos los canales de ventas a partir del 2013 con una apertura mensual.

Para las **variables externas**, se hizo extracción de datos macroeconómicos a nivel país desde la página del Banco Central de Chile y el INE (Instituto Nacional de Estadísticas). Estos datos se extrajeron con una apertura mensual y desde el año 2013. Las variables extraídas fueron las siguientes:

- IMACEC.
- Variación anual del IMACEC.
- PIB.
- Variación anual del PIB.
- Consumo Privado.
- Variación anual del Consumo Privado.
- Variación anual de la Inflación.
- Variación mensual de la Inflación.
- Tasa de Desempleo.
- Tipo de Cambio.
- Variación anual del Tipo de Cambio.

- Variación mensual del Tipo de Cambio.
- Salario Nominal.
- Variación anual del Salario Nominal.
- Variación mensual del Salario Nominal.
- Tasa de Política Monetaria.
- Índice de Precios al Consumidor.
- Importaciones en millones de USD.
- Exportaciones en millones de USD.
- Fuerza de trabajo.
- Índice de Remuneraciones.
- Índice del costo de la mano de obra.

Además, se incluyeron dos variables asociadas al clima, las cuales fueron extraídas del modelo de pronóstico “*WeatherTrends360*” (<https://www.weatherTrends360.com>). Este modelo, ofrece servicios de pronóstico meteorológico con un algoritmo estadístico para diferentes empresas en el mundo que tengan necesidades de planificación en función del clima. TCC adquiere este servicio y se puede hacer monitoreo y extracción de diferentes datos o variables. Para este caso en particular, se hizo extracción de dos variables con apertura mensual desde el año 2013:

- Temperatura máxima en grados °C.
- Precipitación climática total.

Para las **variables ejecución**, se consideraron los días efectivos de venta y métricas de desempeño en la industria de NARTD. Los días efectivos de venta se calcularon haciendo un recuento de las salidas de ventas diarias en NSR para cada mes desde el año 2013.

Para las métricas de desempeño, se utilizó una base de datos que provee Nielsen Company. Nielsen, es un proveedor de información, datos y tendencias en mercados. La forma de operar de Nielsen es auditando en muestras representativas que ocurre con las tendencias de consumo; analizando el desempeño de marcas, productos y la competencia presente para dar una visión completa y comparable de las ventas. Desde Nielsen, se extrajeron **4 variables de ejecución**:

- **Distribución ponderada manejante:** Indica que porcentaje del 100% de las ventas en los almacenes corresponde al producto seleccionado.
- **Distribución numérica manejante:** Indica en que porcentaje del 100% de los almacenes el producto está presente.
- **Precio promedio por unidad (en cajas unitarias)**
- **Inventario activo frio (en cajas unitarias):** Indica cuantas cajas unitarias del inventario frio presente en los almacenes corresponden al producto seleccionado.
- **Inventario activo no frio (en cajas unitarias):** Indica cuantas cajas unitarias del inventario no frio presente en los almacenes corresponden al producto seleccionado.

Todas estas métricas fueron extraídas para los productos: Coca Cola y Pepsi. Se consideraron estas dos ya que son competencia directa y ambas tienen un protagonismo alto en el mercado. Estas métricas entregan datos mensuales a partir de diciembre de 2016.

El total de variables extraídas, contando la variable a predecir, asciende a 36 variables. Todas estas fueron seleccionadas debido a que se considera que tienen un impacto o una relación directa con las ventas de la compañía; esta selección fue objeto de discusión con el equipo de trabajo. Además, el objetivo de seleccionar datos de naturaleza distinta, como macroeconómicos, climáticos y de ejecución, es debido a que se busca modelar fenómenos distintos en una misma herramienta con el propósito de predecir y explicar las ventas.

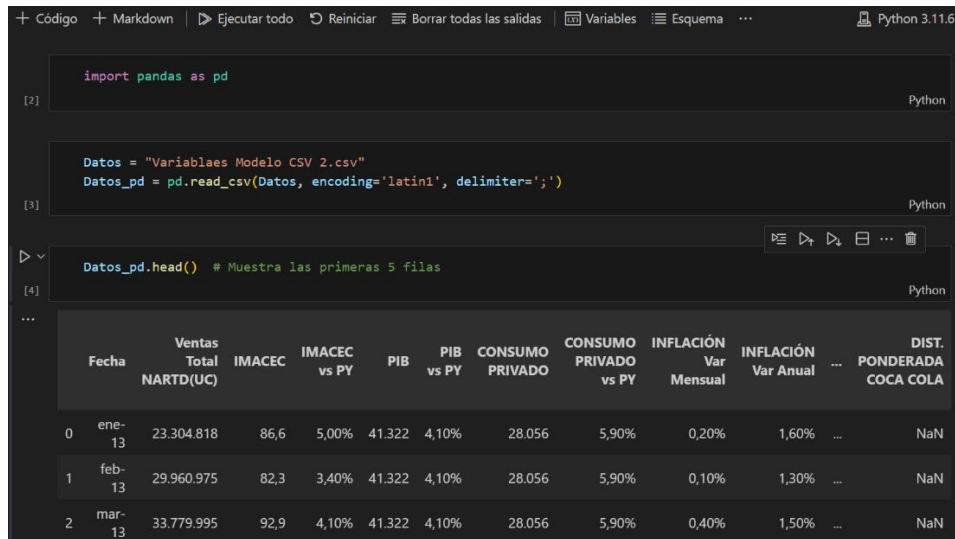
Con todo este levantamiento de información, se construyó un archivo de tipo CSV mediante la herramienta Excel que permitirá manipular y preparar los datos para su modelado.

Fecha	Ventas Total NARTD(UC)	IMACEC	IMACEC vs PY	PIB	PIB vs PY	CONSUMO PRIVADO	CONSUMO PRIVADO vs PY	INFLACIÓN Var Mensual	INFLACIÓN Var Anual	DESEMPLEO	TIPO DE CAMBIO	TC vs PY
ene-13	23.304.818	86,6	5,00%	41.322	4,10%	28.056	5,90%	0,20%	1,60%	6,10%	472,67	-5,70%
feb-13	29.960.975	82,3	3,40%	41.322	4,10%	28.056	5,90%	0,10%	1,30%	6,40%	472,3	-1,90%
mar-13	33.779.995	92,9	4,10%	41.322	4,10%	28.056	5,90%	0,40%	1,50%	6,40%	472,5	-2,70%
abr-13	24.803.998	91,4	4,30%	42.736	3,40%	31.040	5,00%	-0,50%	1,00%	6,50%	472,1	-2,90%
may-13	22.034.732	90,7	3,00%	42.736	3,40%	31.040	5,00%	0,00%	0,90%	6,50%	479,6	-3,50%
jun-13	26.826.884	88,6	3,00%	42.736	3,40%	31.040	5,00%	0,60%	1,90%	6,30%	502,9	-0,50%
jul-13	20.094.805	85,6	3,10%	41.003	2,90%	30.633	4,90%	0,30%	2,20%	5,90%	505	2,60%
ago-13	22.425.231	88,9	2,60%	41.003	2,90%	30.633	4,90%	0,20%	2,20%	5,80%	512,6	6,60%
sept-13	33.449.395	85,3	3,10%	41.003	2,90%	30.633	4,90%	0,50%	2,00%	5,90%	504,6	6,20%
oct-13	23.085.403	92,4	2,80%	44.804	2,80%	33.540	4,00%	0,10%	1,50%	6,00%	500,8	5,40%
nov-13	26.444.496	94,5	3,10%	44.804	2,80%	33.540	4,00%	0,40%	2,40%	5,90%	519,3	8,00%
dic-13	49.495.775	97	2,40%	44.804	2,80%	33.540	4,00%	0,60%	3,00%	5,80%	529,5	11,00%
ene-14	23.926.048	88,1	1,70%	42.332	2,40%	29.249	4,30%	0,20%	2,80%	6,20%	537	13,60%
feb-14	30.213.476	84,8	3,10%	42.332	2,40%	29.249	4,30%	0,50%	3,20%	6,20%	554,4	17,40%
mar-14	35.968.920	95,3	2,50%	42.332	2,40%	29.249	4,30%	0,80%	3,50%	6,50%	563,8	19,30%
abr-14	23.846.311	92,8	1,60%	43.460	1,70%	31.897	2,80%	0,60%	4,30%	6,10%	554,6	17,50%
may-14	22.947.110	92,5	2,00%	43.460	1,70%	31.897	2,80%	0,30%	4,70%	6,30%	555,4	15,80%
jun-14	26.102.152	89,9	1,50%	43.460	1,70%	31.897	2,80%	0,10%	4,30%	6,50%	553,1	10,00%
jul-14	19.800.924	86,2	0,70%	41.541	1,30%	31.203	1,90%	0,20%	4,50%	6,60%	558,2	10,50%

Ilustración 7: Extracto de archivo csv final con todas las variables incorporadas.

La herramienta que se usará para confeccionar el modelo es el Software Visual Studio Code (VSC) de Microsoft. Este programa es un editor de código fuente que permite usar múltiples lenguajes de programación, para este caso se usará el lenguaje de Python. El modelo será construido a través de las herramientas estadísticas de Python “*sklearn*” y “*statsmodels*”, que permiten realizar análisis de regresión, estimar los valores que se necesitan y hacer entrenamiento del conjunto de datos. En suma, se usarán otras bibliotecas como “*pandas*” que permiten la manipulación inicial de los datos antes de construir y entrenar el modelo.

La imagen que se adjunta a continuación muestra la importación del archivo CSV que se manipulará en el Software VSC:



```

import pandas as pd

Datos = "Variables Modelo CSV 2.csv"
Datos_pd = pd.read_csv(Datos, encoding='latin1', delimiter=';')

Datos_pd.head() # Muestra las primeras 5 filas
  
```

	Fecha	Ventas Total NARTD(UC)	IMACEC	IMACEC vs PY	PIB	PIB vs PY	CONSUMO PRIVADO	CONSUMO PRIVADO vs PY	INFLACIÓN Var Mensual	INFLACIÓN Var Anual	...	DIST. PONDERADA COCA COLA
0	ene-13	23.304.818	86,6	5,00%	41.322	4,10%	28.056	5,90%	0,20%	1,60%	...	NaN
1	feb-13	29.960.975	82,3	3,40%	41.322	4,10%	28.056	5,90%	0,10%	1,30%	...	NaN
2	mar-13	33.779.995	92,9	4,10%	41.322	4,10%	28.056	5,90%	0,40%	1,50%	...	NaN

Ilustración 8: VSC con importación del archivo CSV a través de Python.

En adelante se comienza a trabajar los datos y transformar las variables para su modelado. A continuación, podemos visualizar la naturaleza inicial de los datos:

```
Datos_pd.info()
✓ 0.0s

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 133 entries, 0 to 132
Data columns (total 38 columns):
#   Column                                     Non-Null Count  Dtype
---  -
0   Fecha                                     133 non-null    object
1   Ventas Total NARTD(UC)                   130 non-null    object
2   IMACEC                                    132 non-null    object
3   IMACEC vs PY                             132 non-null    object
4   PIB                                       132 non-null    float64
5   PIB vs PY                                132 non-null    object
6   CONSUMO PRIVADO                          132 non-null    float64
7   CONSUMO PRIVADO vs PY                    132 non-null    object
8   INFLACIÓN Var Mensual                    129 non-null    object
9   INFLACIÓN Var Anual                      129 non-null    object
10  DESEMPLEO                                128 non-null    object
11  TIPO DE CAMBIO                            132 non-null    object
12  TC vs PY                                  132 non-null    object
13  TC vs PM                                  132 non-null    object
14  SALARIO NOMIAL                            132 non-null    object
15  SALARIO NOMINAL vs PY                    132 non-null    object
16  SALARIO NOMINAL vs PM                    132 non-null    object
17  TEMPERATURA°C                             133 non-null    object
18  PRECIPITACIÓN CLIMÁTICA                  133 non-null    object
19  IPC vs PY                                 133 non-null    object
...
36  INVENTARIO ACTIVO FRIO PEPSI              82 non-null    object
37  INVENTARIO ACTIVO NO FRIO PEPSI           82 non-null    object
```

Ilustración 9: Naturaleza inicial de las variables.

En la naturaleza de los datos, se puede apreciar que la gran mayoría tienen formato de “*object*”. Este es un tipo de dato más general que puede contener cualquier tipo de objeto, pero para nuestro caso necesitamos manejar datos con valores numéricos, es decir, de formato: “*float64*”.

Particularmente, la variable “Fecha”, se necesita manejar en tipo “*datetime64*”, esto para que pueda ser manipulada propiamente tal como una fecha.

A continuación, se muestra el formato final de todas las variables:


```

RangeIndex: 133 entries, 0 to 132
Data columns (total 38 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   Fecha                                133 non-null    datetime64[ns]
1   Ventas Total NARTD(UC)               130 non-null    float64
2   IMACEC                               132 non-null    float64
3   IMACEC vs PY                         132 non-null    float64
4   PIB                                  132 non-null    float64
5   PIB vs PY                           132 non-null    float64
6   CONSUMO PRIVADO                     132 non-null    float64
7   CONSUMO PRIVADO vs PY               132 non-null    float64
8   INFLACIÓN Var Mensual                129 non-null    float64
9   INFLACIÓN Var Anual                 129 non-null    float64
10  DESEMPLEO                           128 non-null    float64
11  TIPO DE CAMBIO                      132 non-null    float64
12  TC vs PY                            132 non-null    float64
13  TC vs PM                            132 non-null    float64
14  SALARIO NOMIAL                      132 non-null    float64
15  SALARIO NOMINAL vs PY               132 non-null    float64
16  SALARIO NOMINAL vs PM               132 non-null    float64
17  TEMPERATURA°C                       133 non-null    float64
18  PRECIPITACIÓN CLIMÁTICA             133 non-null    float64
19  IPC vs PY                           133 non-null    float64
...
36  INVENTARIO ACTIVO FRIO PEPSI         82 non-null     float64
37  INVENTARIO ACTIVO NO FRIO PEPSI     82 non-null     float64
dtypes: datetime64[ns](1), float64(37)

```

Ilustración 10: Formato final de variables para su modelado.

Luego de esto, se procedió a hacer una descripción de las variables para corroborar escalabilidad de estas:

	count	mean	min	25%	50%	75%	max	std
Fecha	133	2018-07-01 18:46:00.902255616	2013-01-01 00:00:00	2015-10-01 00:00:00	2018-07-01 00:00:00	2021-04-01 00:00:00	2024-01-01 00:00:00	NaN
Ventas Total NARTD(UC)	130.0	29087163.484615	18674024.0	23516352.5	26971090.0	32643239.5	51081749.0	7308799.542407
IMACEC	132.0	98.383333	82.3	92.475	97.35	104.1	121.2	8.193538
IMACEC vs PY	132.0	0.021515	-0.16	0.0	0.02	0.03	0.21	0.054918
PIB	132.0	46599.022727	41003.0	43970.75	46563.5	48972.5	54840.0	3523.52394
PIB vs PY	132.0	0.021364	-0.15	0.0	0.02	0.03	0.19	0.053099
CONSUMO PRIVADO	132.0	35182.454545	28056.0	32093.5	34403.5	37633.5	45558.0	4076.161466
CONSUMO PRIVADO vs PY	132.0	0.033258	-0.2	0.02	0.035	0.0425	0.34	0.079775
INFLACIÓN Var Mensual	129.0	0.003256	-0.01	0.0	0.0	0.01	0.02	0.005025
INFLACIÓN Var Anual	129.0	0.044574	0.01	0.02	0.03	0.05	0.14	0.031125
DESEMPLEO	128.0	0.075469	0.06	0.07	0.07	0.08	0.13	0.015156
TIPO DE CAMBIO	132.0	694.3025	472.1	625.95	681.5	785.615	955.9	111.576893
TC vs PY	132.0	0.054848	-0.17	-0.03	0.07	0.14	0.28	0.100149
TC vs PM	132.0	0.004697	-0.06	-0.01	0.0	0.02	0.11	0.027503
SALARIO NOMIAL	132.0	112.928788	83.3	97.175	110.15	125.075	156.0	19.724962
SALARIO NOMINAL vs PY	132.0	0.059621	0.0	0.05	0.06	0.07	0.11	0.020874
SALARIO NOMINAL vs PM	132.0	0.005	-0.01	0.0	0.0	0.01	0.02	0.005456
TEMPERATURA°C	133.0	22.924812	14.8	18.5	23.2	27.5	31.0	4.82943
PRECIPITACIÓN CLIMÁTICA	133.0	22.98797	3.2	10.0	16.0	29.0	115.0	19.919916
IPC vs PY	133.0	0.043233	0.0	0.02	0.03	0.05	0.14	0.031588
IPC vs PM	129.0	0.003256	-0.01	0.0	0.0	0.01	0.02	0.005025
TPM	130.0	0.041308	0.01	0.03	0.03	0.05	0.11	0.028102
IMPORTACIONES	129.0	6231.124574	4093.24	5343.34	5969.22	6914.29	9908.79	1281.930095
EXPORTACIONES	129.0	6421.29876	4606.0	5564.55	6206.7	7271.73	9728.41	1191.561952
FUERZA DE TRABAJO	128.0	0.015469	-0.16	0.0175	0.02	0.03	0.1	0.044595

Ilustración 11: Descripción y escalabilidad de variables

Se aprecia que tienen medias con escalabilidad distinta, eso puede significar una dificultad a la hora de modelar e interpretar los resultados finales.

Para complementar este análisis, se ha hecho una matriz bivariada que permite visualizar la correlación de las variables. Esta se adjunta a continuación:

Correlaciones Variables

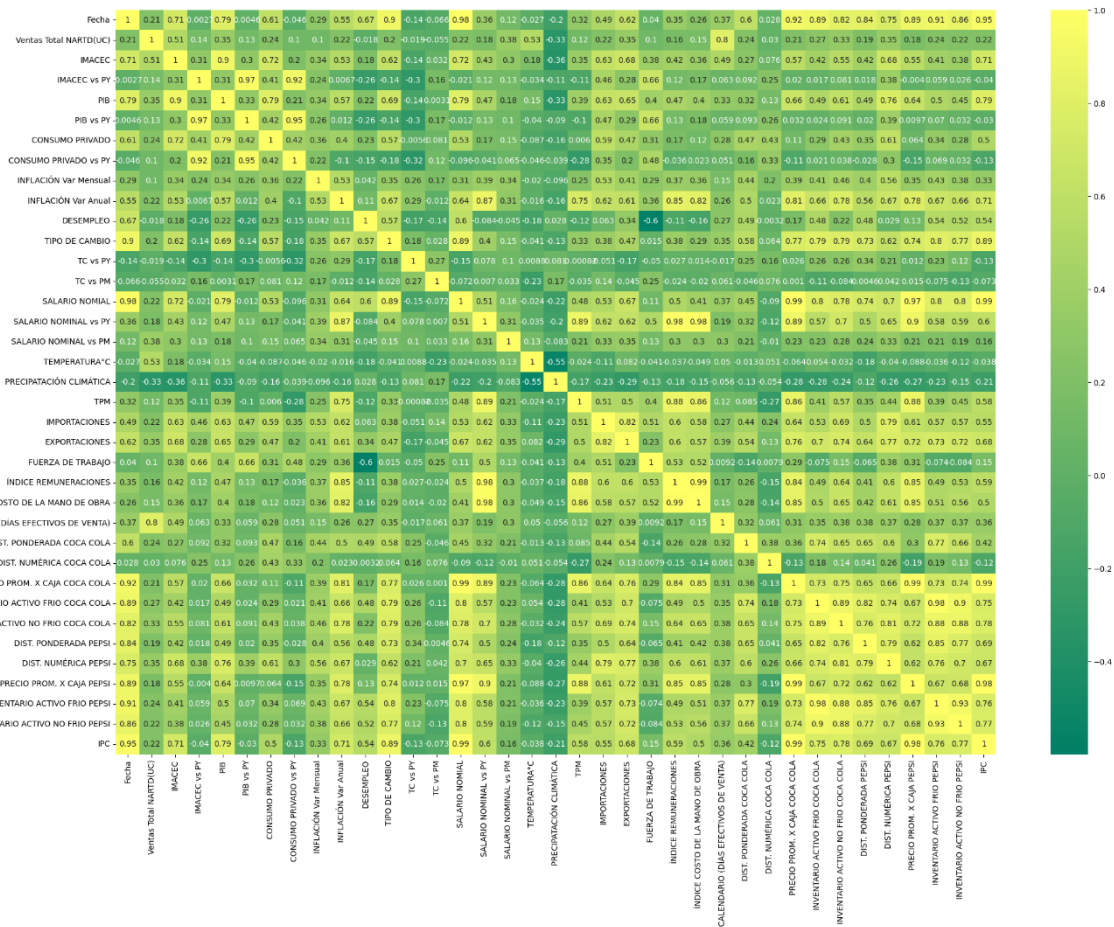


Ilustración 12: Matriz de correlación de las variables.

El color amarillo representa una alta correlación y a medida que se pone más verde, quiere decir que hay una menor correlación. A simple vista, se observa una alta variabilidad de los datos, lo que al igual que la escalabilidad de las medias, indica que estos deben ser transformados a una escala similar para ser trabajados y que el modelo tenga resultados coherentes. Por lo tanto, se ha utilizado la herramienta de Python “*StandardScaler*” para estandarizar todas las características del conjunto de datos.

Posterior a esto, se ha realizado una visualización de los datos a través de “Boxplots” o Diagramas de Cajas. Esto se hace para visualizar la distribución de los datos y tratar de detectar cuales pueden tener potenciales “outliers”.

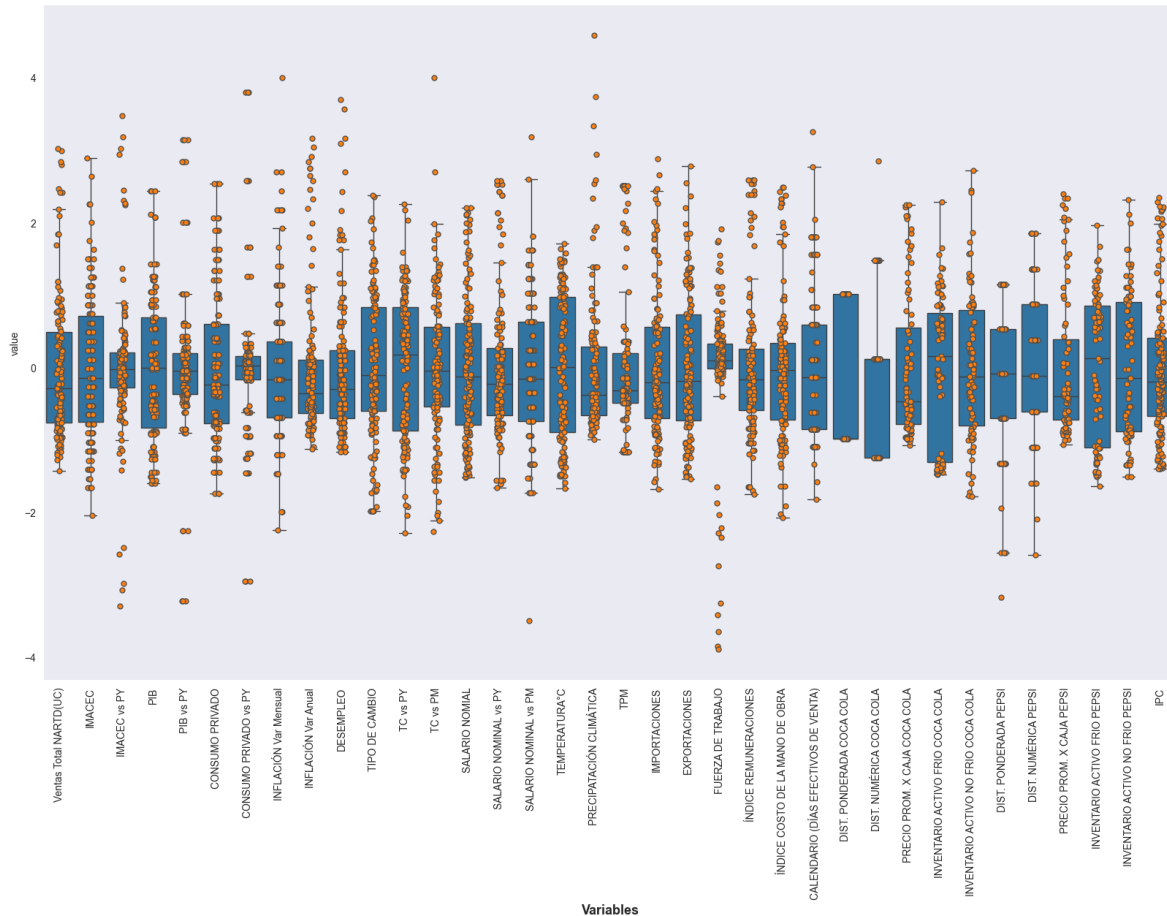


Ilustración 13: Diagrama de cajas para todas las variables extraídas.

Se pueden detectar una serie de datos atípicos en la gran mayoría de las variables, pero dado el contexto en el que se está trabajando la información, estos no serán eliminados, ya que contienen información de valor. La gran mayoría muestran como cambiaron los datos debido al contexto sociopolítico vivido en Chile el año 2019 y el periodo de Pandemia a partir del año 2020.

Luego, para construir el modelo, es necesario escoger el mejor set de variables. Esto se hizo a través de criterios que permitieron seleccionar aquellas variables que ofrecían un rendimiento superior. Los criterios utilizados fueron los siguientes:

1. **Eliminación Recursiva de Características (RFE):** Es una técnica de selección de características para identificar y seleccionar variables relevantes. En palabras sencillas, lo que esta técnica realiza es eliminar las características menos relevantes del conjunto de variables, proporcionando aquellas que se relacionan mejor o tienen un impacto significativo en la variable a predecir.
2. **Orden de Importancia de Características:** Está técnica también es para seleccionar variables relevantes. La diferencia es que este método evalúa la importancia de cada característica de manera individual, estableciendo un orden jerárquico para una cada una sin eliminarlas.
3. **Factor de Inflación de Varianza (VIF):** Es una medida estadística utilizada para evaluar la multicolinealidad en un modelo de regresión. La multicolinealidad se refiere a la alta correlación que puede existir entre dos o más variables. Esto puede causar problemas, ya que hace que sea difícil determinar la contribución individual de cada variable al modelo. Lo que hace es calcular cuanto aumenta la varianza de un coeficiente en el modelo en función de su multicolinealidad, lo que se busca hacer es encontrar valores VIF bajos: inferiores a 10. Si los valores ya son superiores a 10, estamos en presencia de multicolinealidad.
4. **Criterio Comercial:** Se refiere a la relevancia práctica y comercial que tiene cada variable en función de las necesidades que la empresa espera modelar. Es decir, es asegurarse que los datos seleccionados para el modelo tienen un sentido lógico y práctico para la toma de decisiones.

Luego de aplicar todos estos criterios las variables seleccionadas fueron las siguientes:

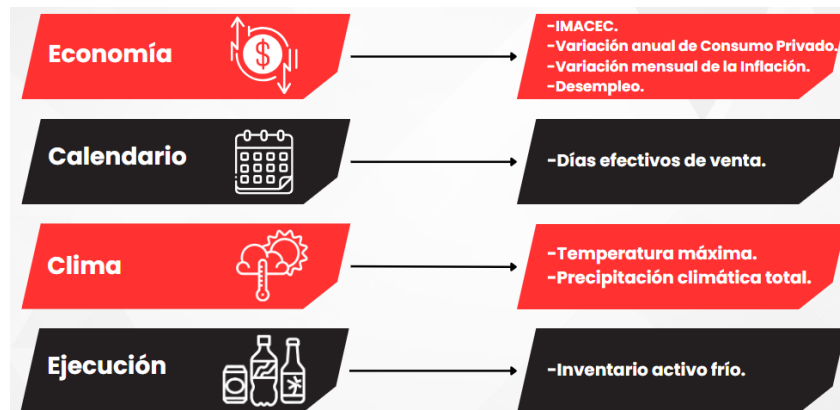


Ilustración 14: Variables seleccionadas para su modelado.

Posterior a la limpieza, escalabilidad de los datos y selección de variables finales se procedió con el armado del modelo, el cual debe cumplir con la fórmula de una regresión lineal múltiple, la cual para este caso es la siguiente:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n$$

Donde:

Y : Corresponde a la variable de respuesta, en este caso es el pronóstico de ventas NARTD.

X_1, X_2, \dots, X_n : Corresponden a las variables predictoras.

β_0 : Constante del modelo, también conocido como intercepto.

$\beta_1, \beta_2, \dots, \beta_n$: Son los coeficientes asociados a las variables predictoras.

Este modelo fue construido con la herramienta estadística “*statsmodels*”. Luego de su confección, se empleó el 80% de los datos importados para hacer un **entrenamiento del modelo**. Esto se llevó a cabo mediante la función “*fit*” de la biblioteca “*sklearn*” que realiza un entrenamiento supervisado, es decir, un entrenamiento a partir de datos entregados. En este proceso, el modelo ajusta sus parámetros internos para hacer predicciones más precisas en función de las características y las etiquetas de los datos de entrenamiento.

Finalmente, para reafirmar que las variables seleccionadas fueron las más robustas, se utilizaron 4 criterios adicionales que nos ayudan a evaluar la solidez y efectividad del modelo. Estos criterios son:

5. **Criterio Coherente:** Se refiere a la necesidad de que los coeficientes asignados a cada variable tengan sentido con la realidad que se busca modelar. Es decir, se espera que magnitud de los coeficientes sea coherente con la relación esperada entre las variables. Por ejemplo, si estamos modelando el PIB, se esperaría que su coeficiente sea positivo, ya que un coeficiente negativo estaría indicando una disminución de la variable a predecir, lo cual no tiene sentido según este contexto.
6. **P-value:** El valor p en el contexto de una regresión lineal, es una métrica estadística que evalúa la significancia predictora de los coeficientes en el modelo. Un valor p bajo indica que la variable tiene una importancia alta en las predicciones.

7. **R-Squared:** El R cuadrado es una medida que explica la calidad y robustez de ajuste de un modelo de regresión. En estricto rigor, representa cuanta variabilidad en la variable de respuesta es explicada por el modelo. Un R cuadrado cercano a 1 indica un modulo bien ajustado, mientras que un valor cercano a 0 sugiere que el modelo no está captando bien la variabilidad de los datos.
8. **Validación Cruzada:** Previo a la aplicación de las medidas de desempeño, se utilizó el 20% de la información no usada en el entrenamiento para validar las predicciones del modelo. Este proceso asegura que las variables seleccionadas sean relevantes más allá del conjunto de entrenamiento inicial.

c) Metodología de Validación de Resultados

Como iniciativa para validar el modelo construido se utilizará información de los resultados reales de las ventas correspondientes al año 2023, abarcando los periodos mensuales desde Enero hasta Octubre. Este proceso de validación tiene como objetivo principal contrastar las proyecciones generadas con el modelo nuevo, el modelo actual que usa la compañía y los resultados reales de las ventas.

De este modo, se hará un cálculo de todas las medidas de desempeño propuestas para ambos modelos de pronóstico. En consecuencia, estas serán comparadas para corroborar los resultados, los pronósticos por periodo y verificar que se está alcanzando el objetivo general del proyecto.

Recordemos que las medidas de desempeño fueron explicadas con detalle anteriormente y son las siguientes:

- **Error absoluto medio:** $MAPE = \frac{\sum_{i=1}^n \left| \frac{x_i - y_i}{x_i} \right|}{n}$
- **Diferencia de errores:** $\Delta MAPE = |MAPE_1 - MAPE_2|$
- **Error de pronóstico absoluto:** $e_t = |Y_t - \hat{Y}_t|$
- **Error de pronóstico porcentual:** $ea_t(\%) = \frac{|Y_t - \hat{Y}_t|}{Y_t} * 100$

Los cálculos se harán de la siguiente manera:

- El MAPE se evaluará de forma individual para cada modelo.
- La diferencia de errores será la diferencia entre el MAPE del modelo actual y el MAPE del modelo nuevo.
- Los errores de pronóstico, tanto el absoluto como el porcentual, se calcularán para cada periodo observado, es decir, se considerarán errores para cada mes desde Enero hasta Octubre del modelo nuevo y del modelo actual.

Con los resultados obtenidos, se espera brindar comprensión completa del rendimiento del modelo que será implementado y el modelo actual, como también se espera alcanzar el objetivo propuesto para este proyecto.

Plan de Implementación

A continuación, se adjunta cual ha sido el plan de implementación del proyecto, detallando las fases y actividades llevadas a cabo para alcanzar los objetivos propuestos.

a) Fase de Inicio

- Planteamiento del dolor y la brecha al equipo de trabajo para llevar a cabo la mejora.
- Planificación y propuesta de la estrategia para llevar a cabo el proyecto.
- Identificar fuentes de extracción de la información.

b) Fase de Desarrollo

- Reuniones semanales con el supervisor y reuniones periódicas con el equipo de trabajo para hacer seguimientos de los avances y posibles ajustes.
- Levantamiento de la información y adquisición de los datos.
- Limpieza de los datos, análisis de estos y selección de variables.
- Elección y armado del modelo econométrico.

c) Fase de Aplicación y Ajuste

- Capacitación al equipo de trabajo para un correcto entendimiento de la herramienta.

En consecuencia, se ha diseñado una matriz de riesgo y estrategias de mitigación que proporcionen una visión integral de los posibles obstáculos que pueden surgir. Esta matriz se adjunta a continuación:

	Insignificante	Menor	Moderado	Mayor	Catastrófico
Improbable	1	2	3	4	5
Posible	2	4	6	8	10
Moderado	3	6	9	12	15
Probable	4	8	12	16	20
Muy probable	5	10	15	20	25

Tabla 3: Matriz de Riesgo.

Las columnas de la izquierda representan la probabilidad de que un hecho ocurra y las filas superiores indican el impacto que puede tener ese hecho en el proyecto. A modo de complemento, también se adjunta una tabla que indica que tan tolerable puede ser un problema según su puntaje:

Riesgo	Puntaje
Aceptable	1 a 4
Adecuado	5 a 9
Tolerable	10 a 16
Inaceptable	17 a 25

Tabla 4: Tolerabilidad del riesgo.

Dado lo anterior y entendiendo como opera la matriz, para este proyecto en particular se pueden reconocer los siguientes riesgos:

- Falta de entendimiento del modelo:** Poco entendimiento de cómo opera el modelo que puede afectar la interpretación en su ejecución.

Puntuación de riesgo: 4, corresponde a un riesgo aceptable.

Mitigación: Realizar capacitación y documentación adecuada al equipo de trabajo.

- 2. Riesgo en los datos:** Datos incompletos, inexactos o sesgados; esto puede afectar la precisión del modelo.

Puntuación de riesgo: 9, corresponde a un riesgo adecuado.

Mitigación: Realizar limpieza minuciosa de los datos, identificar y corregir datos faltantes, diversificar fuentes de extracción en caso de que sea necesario.

- 3. Riesgo del modelo:** Modelo no lo suficientemente robusto o ajustado, esto podría ser un problema en resultados que se esperan de la herramienta.

Puntuación de riesgo: 12, corresponde a un riesgo tolerable.

Mitigación: Aplicar validación cruzada, es decir, evaluar el modelo con diferentes variables y para diferentes productos de la compañía. Mantener simplicidad, evitar que el modelo sea demasiado complejo para que pueda capturar lo que se espera.

- 4. Cambios en el entorno:** Condiciones políticas, económicas y sociales, cambios en la estructura de la empresa.

Puntuación de riesgo: 15, corresponde a un riesgo tolerable.

Mitigación: Monitoreo continuo del entorno, reuniones con el equipo para adoptar estrategias, realizar escenarios de sensibilidad para ver cómo los cambios en el entorno afectan al modelo, actualizar el modelo periódicamente.

Dicho lo anterior y para una mejor comprensión de este análisis, se adjunta una tabla resumen con los riesgos, su puntuación y las respectivas mitigaciones:

Riesgo	Puntaje	Mitigaciones
Falta de entendimiento del modelo	4	Hacer introducción y capacitación al equipo de trabajo.
Riesgo de datos	9	Realizar limpieza minuciosa de los datos, identificar y corregir datos faltantes, diversificar fuentes de extracción.
Riesgo del modelo	12	Validar el modelo constantemente, mantener simplicidad, ejecutar validación cruzada.
Cambios en el entorno	15	Monitoreo continuo del entorno, reuniones con el equipo, realizar escenarios de sensibilidad, actualizar el modelo periódicamente.

Tabla 5: Riesgos y mitigaciones.

Evaluación Económica

Para realizar la evaluación económica, se llevó a cabo la extracción de datos de inversión en activos e insumos de Coca Cola Andina S.A. correspondientes al año 2023. El propósito principal de esto es utilizar esta información como referencia clave para evaluaciones futuras de inversión, permitiendo estimar el potencial ahorro que podría materializarse en periodos venideros en caso de que el proyecto se implemente de manera exitosa.

Se han planteado 3 tipos de escenarios:

- Escenario pesimista, que corresponde a un ahorro del 2% de la inversión
- Escenario realista, que corresponde a un ahorro del 5% de la inversión

- Escenario optimista, que corresponde a un ahorro del 10% de la inversión.

Para los periodos de estudio, se han utilizados datos de inversión por trimestre, específicamente el primer, segundo y tercer trimestre del año 2023.

Dicho lo anterior, se adjunta una tabla que muestra los resultados del análisis realizado en millones de pesos chilenos:

Trimestre	Mes	Inversión	Pesimista (2%)	Realista (5%)	Optimista (10%)
T1	Enero	\$40.682	\$39.868	\$38.648	\$36.614
	Febrero				
	Marzo				
T2	Abril	\$85.339	\$83.632	\$81.072	\$76.805
	Mayo				
	Junio				
T3	Julio	\$141.389	\$138.561	\$134.320	\$127.250
	Agosto				
	Septiembre				

Total	\$267.410	\$262.062	\$254.040	\$240.669
Ahorro		\$5.348	\$13.371	\$26.741

Promedio	\$252.257
Ahorro del Promedio	\$15.153

Tabla 6: Posibles escenarios de ahorro en Mill. CH\$

Adicionalmente, se consideró un ahorro promedio de los tres escenarios propuestos, que sirve como un indicador complementario al estudio efectuado.

Resultados

En el siguiente punto se abordará todo lo que fueron los resultados del proyecto realizado

a) Resultados del Desarrollo de la Solución

Recordemos que, para efectos de este informe, algunos valores desplegados serán de UTP, es decir, Unidades de Transacción Parametrizadas, esto debido a que no se puede revelar información explícita de los resultados operacionales del 2023. Recordemos también que la variable que se busca predecir son las ventas totales de NARTD en MUC (Millones de Cajas Unitarias).

Como se mencionó en la metodología, las variables seleccionadas para el armado del modelo fueron:

- IMACEC.
- Variación anual del Consumo Privado.
- Variación mensual de la Inflación.
- Desempleo.
- Días efectivos de venta.
- Temperatura máxima en grados °C.
- Precipitación climática total.

Dicho lo anterior, el modelo que se confeccionó arrojó un R cuadrado de 0,9. Esto quiere decir que el ajuste y la calidad del modelo son notablemente aceptables, destacando la eficiencia con la que el modelo explica la variabilidad de los datos.

Así mismo, para evaluar el desempeño de forma preliminar, se hicieron predicciones con el modelo para compararlo con los resultados reales desde el año 2017 al año 2023.

Los resultados de estas predicciones pueden verse a continuación:

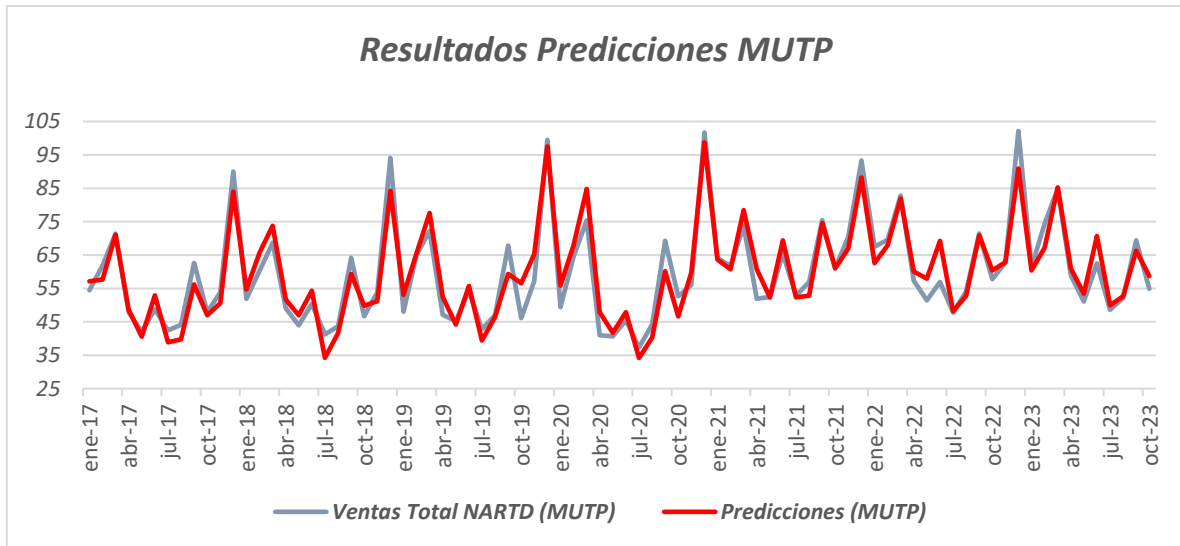


Ilustración 16: Predicciones del modelo confeccionado desde 2017 a 2023.

A simple vista, se puede observar que las predicciones se alinean bastante bien con los resultados reales de las ventas. Estas observaciones arrojaron un MAPE de 6,61%, es decir, que en promedio las predicciones fallan en un 6,61%. Este bajo porcentaje indica una precisión considerable en el modelo.

Así, podemos ver de forma granular los resultados de errores de pronóstico por periodo en MUC (Millones de Cajas Unitarias). Se adjuntan los errores en MUC para que se tenga referencia con los errores de pronóstico del modelo que usa actualmente la compañía.

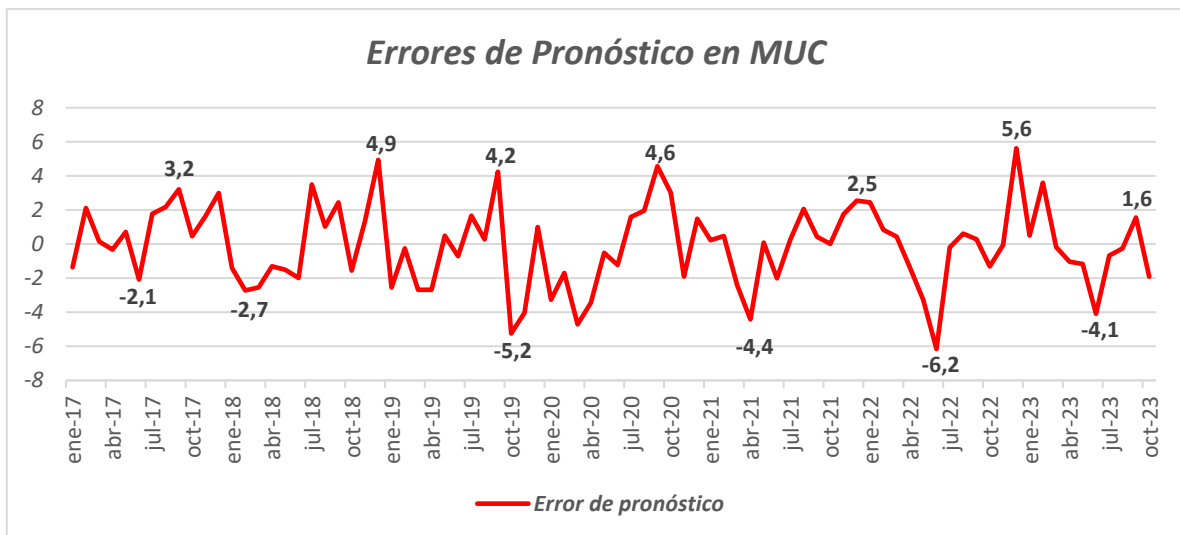


Ilustración 17: Errores de pronóstico para evaluación desde 2017 a 2023.

Si bien, hay resultados donde los errores ascienden de 5 a 6,2 MUC, el promedio de los errores es de 1,9 MUC, con una desviación de 1,4 MUC con respecto al promedio. Considerando que estas evaluaciones abarcan desde año 2017 hasta octubre del 2023, este resultado se encuentra alineado con los márgenes de error esperados que la compañía establece para sus pronósticos, que es de 2 MUC.

De forma complementaria, también podemos ver los errores de pronóstico anuales en MUC que el modelo arrojó para los periodos cerrados de 2017 a 2022.

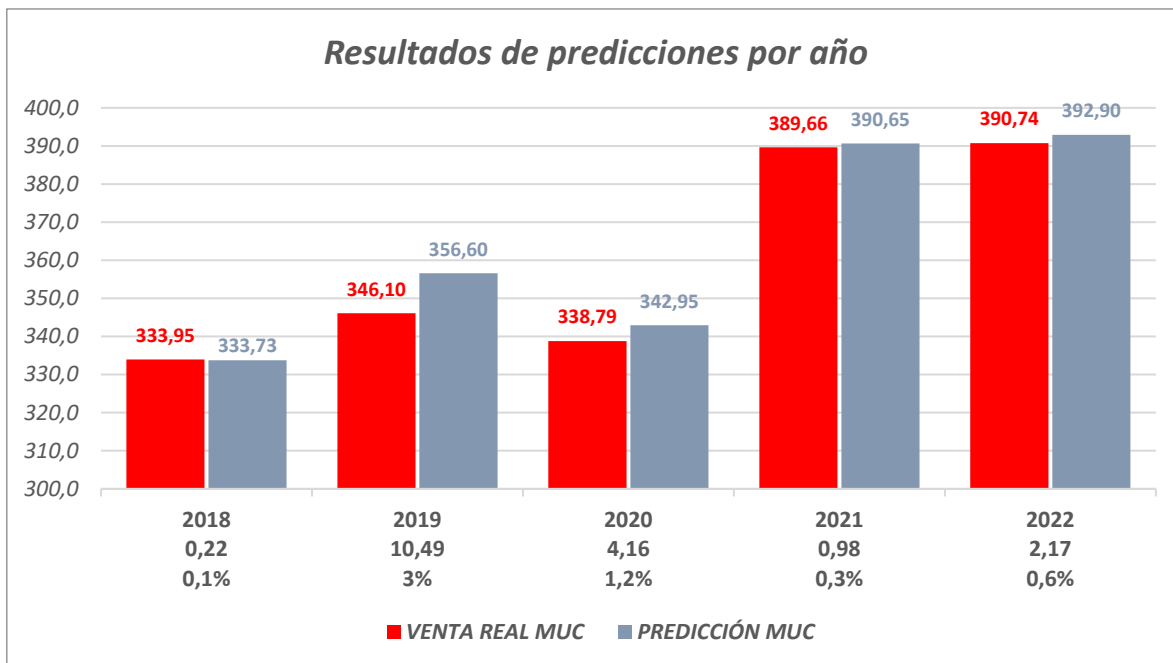


Ilustración 18: Errores de pronóstico anuales desde 2017 a 2023.

Cada observación presenta en el eje horizontal el año, seguido del error de pronóstico para el año en MUC y el porcentaje de cuanto falló el pronóstico anual al resultado real. Se destaca que el error máximo fue de un 3% al resultado operacional y el error mínimo de un 0,1%, lo que indica un buen desempeño en general de estas predicciones.

Una forma de validar las predicciones hechas entre 2017 y 2023 es realizando visualización de la distribución que siguen los residuos de los pronósticos. Se le llama residuo a la diferencia que existe entre el valor real y valor pronosticado, en otras palabras, es un error de pronóstico. Se sostiene que, si estos residuos siguen una distribución normal, el modelo estadístico se ajusta de manera adecuada a los datos, lo que fortalece la confianza en la

precisión y fiabilidad del modelo. A continuación, se adjunta la densidad que siguieron los residuos de los pronósticos realizados desde el año 2017 al 2023:

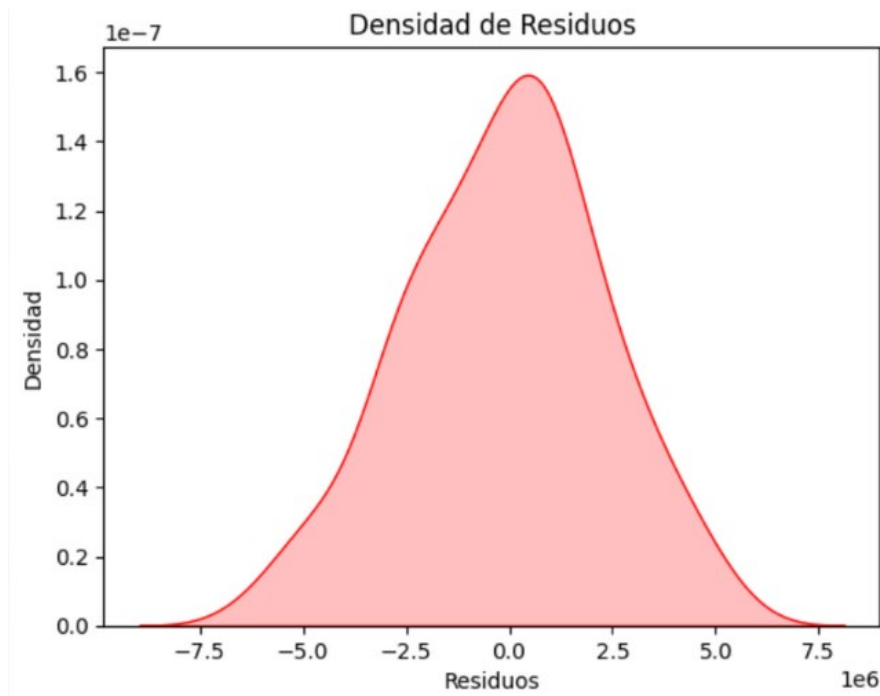


Ilustración 19: Distribución de residuos para pronósticos realizados desde el año 2017 al 2023.

En vista de lo anterior y corroborando que los residuos siguen una distribución muy cercana a la normal, se puede concluir que el modelo exhibido emplea un buen ajuste de los datos. Esta observación aumenta la credibilidad en la capacidad del modelo para realizar pronósticos precisos y respalda su utilidad en futuras necesidades.

Ahora bien, una de las razones por las que se armó este modelo fue con el propósito de incorporar variables explicativas que tuvieran relación con las ventas. De este modo, incorporar estas características permite explicar cuál es la contribución de cada variable a la evolución de los volúmenes de ventas.

Así, podemos obtener las elasticidades de cada una de las variables incorporadas en función de las ventas. Este análisis de elasticidades contribuye significativamente a la comprensión de cómo diferentes factores impulsan variaciones en las ventas de NARTD y permite tomar decisiones estratégicas en el negocio.

A continuación, se adjuntan las elasticidades estadísticas que arrojó cada variable para el modelo confeccionado:

- Si el IMACEC cambia en +1%, las ventas de NARTD cambian en +1,9%.
- Si el Consumo Privado cambia en +1, las ventas de NARTD cambian en +0,8%.
- Si el Desempleo cambia en +1%, las ventas de NARTD cambian en -0,22%.
- Si la Inflación cambia en +1%, las ventas de NARTD cambian en -0,45%.
- Si la Temperatura máxima cambia en +1%, las ventas de NARTD cambian en +2,1%.
- Si la Precipitación climática cambia en +1%, las ventas de NARTD cambian en -0,3%.
- Si los Días Efectivos de Venta cambian en +1%, las ventas de NARTD cambian en 3,5%.
- Si el Inventario Activo Frío cambia en +1%, las ventas de NARTD cambian en +2,8%.

De este modo, se puede desglosar cual es el aporte de cada variable a la evolución de las ventas entre un periodo y otro. Por ejemplo, el delta de ventas entre el año 2021 y 2022 es de 1,08 MUC. Si se hace una apertura de cómo contribuyó cada variable en este delta, se obtiene lo siguiente:

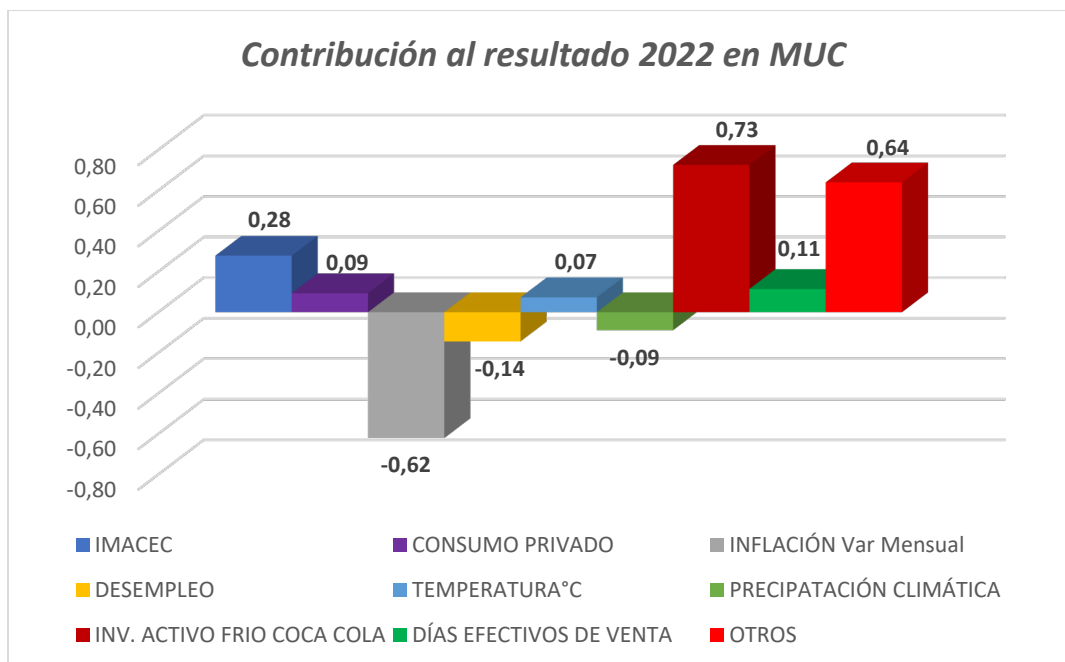


Ilustración 20: Contribución de cada variable al resultado 2022.

Se puede ver que algunas variables aportan valores muy elevados mientras que otras contribuyen de manera más modesta. Esto se debe a las variaciones en las variables entre

2021 y 2022. Por ejemplo, la variación que existió del inventario activo frío del año 2021 al año 2022 fue bastante significativa, es por eso que 0,73 MUC son explicadas por esta variable en estos periodos.

Ahora bien, hay que tener en cuenta que este modelo es una representación estadística y, por ende, la realidad puede diferir en varios aspectos. Aunque las contribuciones proporcionan información valiosa sobre la influencia relativa de cada variable al resultado de las ventas de NARTD, es importante entender que el entorno está sujeto a factores, cambios inesperados y complejidades que un modelo estadístico no podrá capturar completamente. En este contexto, los resultados explicativos para cada periodo se deben interpretar minuciosamente para usarse como un complemento en la toma de decisiones estratégicas.

b) Evaluación de Métricas de Desempeño

En el siguiente punto se calcularán las medidas de desempeño propuestas para este proyecto y del modo que fue explicado en la metodología de validación de resultados.

Hay que recordar que para hacer estas observaciones se usarán los datos del año 2023, por lo tanto, todo será expresado en MUTP (Millones de Unidades de Transacción Parametrizadas).

Recapitular también que, el modo de hacer esta evaluación es contrastando los resultados del modelo armado en este proyecto contra el modelo que usa actualmente la compañía.

Para referirnos al modelo actual que usa la compañía hablaremos de “Pronósticos CC” y para referirnos al modelo armado en este proyecto, se hablará de “Pronósticos RR”.

A continuación, se adjuntan los resultados de los errores de pronósticos absolutos (e_t) y los errores de pronóstico porcentuales ($ea_t(\%)$) para cada periodo del año 2023:

2023 MUTP	Pronóstico CC	Pronóstico RR	Resultado Real	Error Absoluto CC	Error Porcentual CC	Error Absoluto RR	Error Porcentual RR
Enero	58,02	60,40	61,41	3,39	5,5%	1,00	1,6%
Febrero	62,01	67,19	74,38	12,37	16,6%	7,20	9,7%
Marzo	75,58	85,29	84,93	9,35	11,0%	0,36	0,4%
Abril	50,88	60,97	58,91	8,03	13,6%	2,06	3,5%
Mayo	48,20	53,43	51,08	2,88	5,6%	2,34	4,6%
Junio	58,36	70,67	62,49	4,13	6,6%	8,19	13,1%
Julio	52,56	49,95	48,60	3,96	8,2%	1,36	2,8%
Agosto	56,79	52,66	52,15	4,64	8,9%	0,51	1,0%
Septiembre	78,44	66,29	69,40	9,04	13,0%	3,12	4,5%
Octubre	60,73	58,71	54,89	5,84	10,6%	3,82	7,0%

Tabla 7: Resultados errores de pronóstico para Pronósticos CC y Pronósticos RR.

Se puede observar que todos los errores de pronósticos RR son menores a los errores de pronósticos CC. De forma complementaria, se pueden visualizar los promedios de estos errores:

Promedio Error Absoluto CC	6,36
Promedio Error Absoluto RR	3,00
Promedio Error Porcentual CC	9,98%
Promedio Error Porcentual RR	4,81%

Tabla 8: Promedio de los errores de pronóstico para Pronósticos CC y Pronósticos RR.

En el costado izquierdo de la tabla vemos el nombre del dato que corresponde y al lado derecho el resultado del promedio indicado. Estos resultados sugieren que para los pronósticos del año 2023 el modelo RR superó en términos de exactitud al modelo CC, lo cual es una señal de que el proyecto realizado ha logrado resultados positivos, proporcionado a la compañía una herramienta que crea valor y es confiable para su uso corporativo.

En suma, podemos graficar estos pronósticos y visualizar como se alinean con las ventas reales:

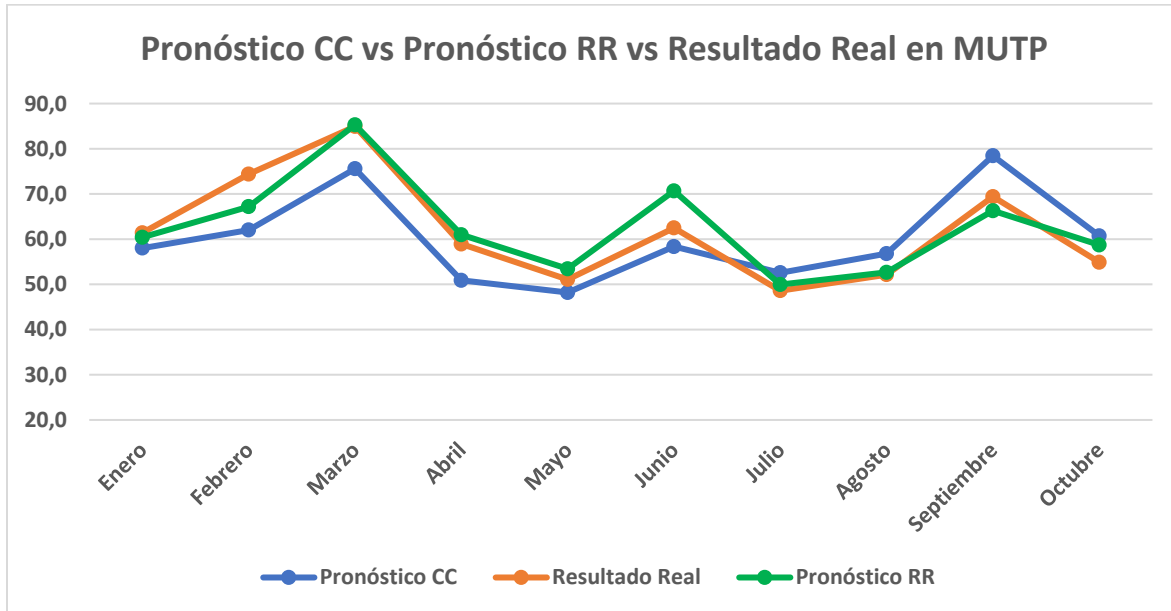


Ilustración 21: Gráfico de pronósticos CC vs pronósticos RR y los resultados de las ventas.

El gráfico anterior facilita una comprensión visual de los resultados de las predicciones, evidenciando que los pronósticos RR están más cerca de los valores reales que los pronósticos CC, respaldando la robustez del modelo confeccionado.

Por último, se calcularon los MAPE de cada modelo y la diferencia de estos dos errores para corroborar si el objetivo general del proyecto fue alcanzado o no. Estos valores se adjuntan a continuación:

MAPE CC	12,47%
MAPE RR	6,02%
DELTA MAPE	6,45%

Tabla 9: Resultado de medidas de desempeño propuestas.

Estos números nos están diciendo lo siguiente:

- En promedio los pronósticos CC fallaron un 12,47%.
- En promedios los pronósticos RR fallaron un 6,02%.
- La diferencia entre estos errores fue de 6,45%.

Si bien, esta diferencia se sitúa fuera del objetivo general planteado inicialmente para este proyecto, es crucial destacar que esta diferencia indica que se está superando de manera positiva el objetivo propuesto. Concretamente, se está un 2,45% por encima de la diferencia inicialmente establecida, que era de un 4%.

Estos resultados sugieren que el modelo armado se alinea de manera efectiva con los desafíos abordados en la problemática identificada y que los pronósticos RR aportan una mejora sustancial en cómo se hace el seguimiento y los pronósticos de las ventas para productos NARTD en Coca Cola Chile S.A.

Conclusiones

Para el desarrollo de este proyecto, se han logrado consolidar todos los objetivos propuestos, tanto el general como los específicos. La recopilación de la información fue la adecuada para abordar la problemática y las necesidades de la empresa, basándose en análisis que consideraban factores internos y externos que podían influir en las ventas. La selección de variables para el modelo se hizo a través estudios que permitieron escoger aquellas que tuvieran un mayor impacto, mejor resultado estadístico y coherencia práctica. El diseño y construcción del modelo de pronóstico RR, ha demostrado ser efectivo en la mejora de los pronósticos en comparación con el modelo que usa actualmente la compañía, entregando una herramienta útil que podrá ser usada como complemento para prever dificultades o demoras que presenten los pronósticos CC. Además, el modelo construido para este proyecto ayuda a desglosar cual es la contribución relativa de cada variable al resultado operacional en cada periodo, brindando una capacidad analítica que sirve de apoyo para futuros trabajos de gestión operativa y estratégica que se ejecuten. Esta herramienta entrega valor y utilidad a la compañía en diferentes aspectos del negocio.

Para la evaluación que se hizo con las medidas de desempeño en los pronósticos del año 2023, el modelo RR mostró un error porcentual medio absoluto (MAPE) de 6,02%, mientras que el modelo CC un error de 12,47%. Estos resultados indican que el modelo RR fue superior en pronósticos y superó las expectativas del proyecto.

La superación de los objetivos establecidos inicialmente es una buena señal de cómo opera el modelo RR. Esta superación de 2,45% por sobre el objetivo general supone que el trabajo a lo largo del proyecto fue llevado a cabo meticulosamente comprendiendo los desafíos y oportunidades que existían.

Los resultados del modelo tienen sentido con la realidad, aunque bien, en algunos cambios relativos hay brechas que podrían estar sobreestimadas. Esto sugiere que el modelo podría beneficiarse de ajustes y perfecciones adicionales que le permitan estar más cerca a la realidad que se busca modelar. La identificación de estas brechas proporciona una valiosa oportunidad para la refinación de la herramienta. Un enfoque más ajustado garantizaría que el modelo capte de forma más precisa las complejidades y variaciones en los datos.

Aún así, hay que tener en cuenta que este trabajo sigue siendo una herramienta estadística. La realidad de la industria es compleja, dinámica y muchas veces imprevista. Aunque el modelo entregue valiosas percepciones y predicciones del entorno, es difícil que alcance completamente todos los matices y variaciones que existen en la realidad. En este contexto, la interpretación de los resultados del modelo debe hacerse con cuidado y siempre reconocer sus limitaciones inherentes. La implementación de este tipo de herramientas debe ir de la mano de una mentalidad flexible, que logre comprender como funciona la naturaleza predictiva y del mismo modo, logre entender las realidades cambiantes del negocio.

Referencias y Anexos

1. Coca Cola Andina S.A. <https://www.koandina.com/>
2. Coca Cola Embonor S.A. <https://www.embonor.cl/>
3. WeatherTrends360 <https://www.weathertrends360.com/>
4. Díez, B. C. (2014). *Desarrollo e implementación de un método para la evaluación de las promociones en la industria del retail*. Pontificia Universidad Católica de Chile (Chile).
5. Pérez, R. A., Sánchez, S. A. M., & Bastidas, J. J. B. (2012). *Aplicación de modelos de pronósticos en productos de consumo masivo*. *Biotecnología en el Sector Agropecuario y Agroindustrial: BSAA*, 10(2).
6. Morales Tabares, Z. E., Cabrera Campos, A., Vázquez Silva, E., & Caballero Mota, Y. (2016). *MPREDSTOCK: Modelo multivariado de predicción del stock de piezas de repuesto para equipos médicos*. *Revista Cubana de Ciencias Informáticas*, 10(3).