

# PROYECTO DE PASANTÍA

Realizado en el área de Business Intelligence de The Not Company

---

**Mejora de la Planificación Presupuestaria y Proyecciones de Ventas Sell-Out:  
Desarrollo e Implementación de un Sistema Predictivo Avanzado.**

---

Andrés Cerda Waak

Proyecto para optar al título de Ingeniería Civil Industrial con mención en Gerencia Ti de la  
Facultad de Ingeniería y Ciencias de la Universidad Adolfo Ibáñez

Profesor guía:

*Raimundo Sánchez*

Santiago, Chile

2023

## Resumen Ejecutivo

El informe de pasantía realizado en The Not Company se centra en la mejora de la planificación presupuestaria y proyecciones de ventas Sell-Out (Venta a consumidor final). Este desafío se aborda desarrollando un sistema predictivo avanzado, utilizando el modelo estadístico SARIMAX e integrando herramientas de Business Intelligence como Tableau. Esta solución innovadora mejora la precisión en las proyecciones de ventas, lo que es crucial para decisiones estratégicas efectivas.

La metodología empleada incluye un análisis detallado del contexto empresarial, destacando la necesidad de una planificación precisa. La recopilación y preparación de datos, junto con el ajuste del modelo predictivo y su integración en los flujos de trabajo de la empresa, fueron etapas clave. La implementación del modelo final, en particular, demostró ser eficaz en la mejora de las proyecciones de ventas.

El proyecto no solo beneficia a la empresa a nivel local sino que también posee el potencial de ser aplicado en NotCo Global, extendiéndose a los seis países donde opera la compañía. Esta replicabilidad subraya la relevancia del proyecto en un contexto más amplio. Además, se sugiere la integración de externalidades en el modelo SARIMAX para incrementar aún más la precisión de las proyecciones.

Los desafíos enfrentados, como la definición inicial del enfoque y la selección de modelos estadísticos adecuados, son discutidos en el informe, resaltando la importancia de una investigación exhaustiva en las fases iniciales. El proyecto demostró que una preparación cuidadosa puede conducir a un desarrollo más fluido y exitoso.

El proyecto de pasantía ha sido un contribuyente significativo para mejorar la planificación y proyección en The Not Company, reduciendo riesgos y costos asociados con proyecciones inexactas. También logro una considerable reducción de tiempo para la formación de proyecciones Sell-Out, dado a que la nueva solución ayuda a partir de una base más avanzada con inclusiones como las externalidades y estacionalidades. Su éxito no solo beneficia a la empresa localmente, sino que también establece un precedente para futuras mejoras a nivel global, enfatizando la necesidad de adaptabilidad y precisión en el cambiante entorno empresarial.

## **Abstract**

The internship report at The Not Company focuses on optimizing budget planning and Sell-Out sales projections. This challenge is addressed by developing an advanced predictive system using the SARIMAX statistical model and integrating Business Intelligence tools like Tableau. This innovative solution improves accuracy in sales projections, crucial for effective strategic decisions.

The methodology involves a detailed analysis of the business context, emphasizing the need for precise planning. Data collection and preparation, along with the adjustment and integration of the predictive model into the company's workflows, were key stages. Implementing the final model, in particular, proved effective in improving sales projections.

The project benefits the company locally and has the potential to be applied in NotCo Global, extending to the six countries where the company operates. This replicability underscores the project's wider relevance. It suggests integrating externalities into the SARIMAX model to further increase projection accuracy.

Challenges faced, such as initial approach definition and selecting appropriate statistical models, are discussed in the report, highlighting the importance of thorough research in early phases. The project demonstrated that careful preparation can lead to more smooth and successful development.

The internship project has been a significant contributor to improving planning and projection at The Not Company, reducing risks and costs associated with inaccurate projections. It also achieved a considerable reduction in time for the formation of Sell-Out projections, as the new solution provides a more advanced base including externalities and seasonality. Its success benefits the company locally and sets a precedent for future improvements globally, emphasizing the need for adaptability and precision in the changing business environment.

## Índice

Introducción .....	1
Contexto de la Empresa .....	1
Contexto del Problema .....	3
Contexto de la Oportunidad .....	6
Objetivos .....	7
Objetivo General .....	7
Objetivos Específicos .....	7
Medidas de Desempeño .....	7
Estado del Arte .....	8
Solución .....	12
Alternativas de Solución .....	12
Solución escogida .....	12
Desarrollo del Proyecto .....	13
Metodologías .....	13
Desarrollo .....	14
Flujo del Código .....	19
Plan de Implementación .....	22
Análisis de Riesgo .....	24
Evaluación Económica .....	25
Conclusión .....	29
Anexos .....	33
Bibliografía y Referencias .....	41

## Glosario

**Sell-In (SI):** Se refiere al volumen de productos vendidos por un fabricante a un canal de distribución o minorista.

**Sell-Out (SO):** Representa el volumen de productos vendidos por un minorista a los consumidores finales.

**I+D (Investigación y Desarrollo):** Área dedicada a la innovación y desarrollo de nuevos productos o servicios en una empresa.

**AI (Inteligencia Artificial):** Tecnología que simula la capacidad de razonamiento humano, aplicada en diferentes campos como el aprendizaje automático, robótica, y procesamiento del lenguaje natural.

**RMSE (Root Mean Square Error):** Medida estadística de la diferencia entre valores predichos por un modelo y los valores reales.

**ARIMA (Autoregressive Integrated Moving Average):** Modelo estadístico usado para analizar y pronosticar series temporales.

**SARIMAX (Seasonal AutoRegressive Integrated Moving Average with eXogenous variables):** Versión de ARIMA que incluye variables estacionales y exógenas para mejorar la precisión en las predicciones.

**ANN (Artificial Neural Networks):** Redes de neuronas artificiales diseñadas para simular el comportamiento del cerebro humano en el aprendizaje y procesamiento de la información.

**LSTM (Long Short-Term Memory):** Tipo de red neuronal recurrente especializada en aprender dependencias a largo plazo en secuencias de datos.

**SAP:** Empresa de software conocida por sus soluciones de planificación de recursos empresariales (ERP).

**API (Application Programming Interface):** Conjunto de reglas y especificaciones que las aplicaciones pueden seguir para comunicarse entre sí.

**SKU (Stock Keeping Unit):** Código único asignado a un producto en inventario para su identificación y seguimiento.

**B2B (Business to Business):** Transacciones comerciales o interacciones entre empresas.

**EAN (European Article Number):** Sistema de código de barras utilizado para identificar productos a nivel mundial.

**S&OP (Sales and Operations Planning):** Proceso en las empresas para coordinar las operaciones y la planificación de ventas.

## Introducción

### Contexto de la Empresa

The Not Company es una empresa unicornio chilena, en la actualidad está valorada a 1.5 billones llevando solo 7 años en el mercado (nació el 2015), esta se enfoca en la tecnología alimentaria produciendo alternativas de origen vegetal a los productos alimenticios de origen animal. Una de las características más destacadas de la empresa es el uso de inteligencia artificial para poder replicar los sabores, texturas y comportamientos de los alimentos de origen animal usando solo ingredientes vegetales. Esta tecnología la llaman Giuseppe, el cual tiene 4 principales roles, Biagio, Discovery, Toolbox y Flora, estas principalmente focalizadas en reinventar el mundo alimenticio vegetal.

Actualmente, la compañía se sitúa en seis países: Canadá, Estados Unidos, México, Brasil, Argentina y Chile, siendo este el último el que registra la mayor cantidad de participación en ventas, esto teniendo un +20% de participación de mercado corporativo vegetal en Latam, ubicándose como uno de los líderes en el mercado vegetal.



*Figura 1: Distribución NotCo Mundial*

NotCo, una empresa en expansión que opera a través de cuatro canales de venta:

Foodservice (restaurantes), Tradicional (negocios pequeños), Ecommerce (venta online) y Moderno (supermercados), siendo este último el de mayor volumen de ventas. En el canal moderno, sus principales clientes son Cencosud, Walmart, Tottus y Santa Isabel. Las ventas de NotCo se dividen en tres categorías clave: Sell-In (ventas directas a clientes como Walmart, tiendas pequeñas, restaurantes o ecommerce), Sell-Out (ventas de clientes a consumidores finales, como los productos NotCo en supermercados) y Market Share (participación en el mercado, tanto general como en el segmento vegetal), siendo esta última esencial para análisis.

En Chile, NotCo se compone de equipos especializados, incluyendo I+D y AI, logística y producción (mayormente externalizada), ventas, marketing, finanzas y un reciente equipo de Business Intelligence (BI) que apoya a los anteriores mediante el análisis de datos para facilitar la toma de decisiones.

Las asociaciones estratégicas o "partnerships" con empresas como Papa John's, Burger King y Starbucks son fundamentales para impulsar las ventas de NotCo, ayudando a expandir la presencia de sus productos en el mercado.

Mirando hacia el futuro, NotCo planea diversificar su cartera de productos y explorar nuevas categorías más allá de los productos vegetales, como las "Not Snacks". Este enfoque permitirá a la empresa expandirse en diferentes categorías mientras mantiene su compromiso con la alimentación basada en plantas, compitiendo en mercados adicionales e innovando en el ámbito de los productos plant-based.

## Contexto del Problema

El proyecto se centra exclusivamente en la sede de Chile de Notco, con planes de replicabilidad en otras sedes. Se enfoca en el canal moderno, cuya justificación se detallará más adelante. El contexto del problema se origina en el área de Business Intelligence (BI) de Notco, que se formalizó tan solo en marzo de 2023. Anteriormente, la empresa, en todas sus sedes, no empleaba tecnología avanzada para la gestión de datos, confiando principalmente en informes de Excel para la toma de decisiones y la planificación. Con la creación del área de BI, se empezó a usar softwares como Tableau para analizar datos, colaborando con diversas empresas para obtener la información necesaria. Notco dispone de datos desde 2021 en adelante para Sell-In, Sell-Out y Market Share, lo que significa que los informes del equipo de BI son recientes. Se han ido añadiendo gradualmente nuevos temas que impactan directamente en los departamentos de ventas, marketing y finanzas. Desde junio de este año, se han consolidado numerosos informes sobre Sell-In y Sell-Out.

Notco se enfrenta a desafíos significativos que necesitan atención inmediata en su proceso de planificación y toma de decisiones. El equipo de marketing tarda tres semanas en generar el presupuesto de Sell-In basándose en una proyección de Sell-Out, dicha proyección consume tiempo valioso y actualmente es inexacta, lo que genera un desequilibrio en la planificación y toma de decisiones de la empresa. Dado que el Sell-Out es el último punto de la cadena de valor de Notco (visualizado en la Figura 3), no tener una comprensión clara de las futuras ventas puede afectar negativamente la capacidad de la empresa para alcanzar sus objetivos comerciales. La falta de un plan detallado y estructurado para el Sell-Out limita el proceso de toma de decisiones de Notco, impidiendo la optimización de sus estrategias de ventas y operaciones. Este desafío requiere una solución para mejorar la precisión y eficiencia en la generación de proyecciones de ventas y, en consecuencia, en la formulación de estrategias comerciales efectivas.



Total Por Categoría	
Septiembre SO	
Categoría	Accuracy Cat
Not Apanados	75%
Not Burger	69%
Not Cheese	28%
Not Dairy	55%
Not Ice Cream	75%
Not Mayo	56%
Not Meat	54%
Not Milk	76%
Not Salchicha	✓
Not Sancks	0%
<b>Total general</b>	<b>64%</b>

Tabla 1: Proyección Sell-Out

Total Por Categoría	
Septiembre SI	
Categoría	Accuracy Cat
Not Apanados	85%
Not Burger	70%
Not Cheese	80%
Not Dairy	35%
Not Ice Cream	89%
Not Mayo	63%
Not Meat	41%
Not Milk	63%
Not Salchicha	✓
Not Sancks	26%
<b>Total general</b>	<b>61%</b>

Tabla 2: Proyección Sell-In

En estas dos figuras se observa la precisión calculada por la empresa

$$\left( \frac{\text{Target Ventas}}{\text{Target Ventas} - \text{Ventas Reales}} - 1 \right)$$

que tienen actualmente en septiembre tanto para Sell-out como Sell-in por marca. Se destacan dos cosas importantes, lo primero es que las proyecciones actuales están entregando resultados deficientes, lo que dificulta la toma de decisiones como también el ensamble de los presupuestos y la producción que tendrá que generar NotCo. Lo segundo es la influencia que se puede notar entre lo que se genera para la proyección de SO y lo que se produce después para SI, lo que nos reafirma lo importante que es asegurar buenos resultados en la primera proyección.

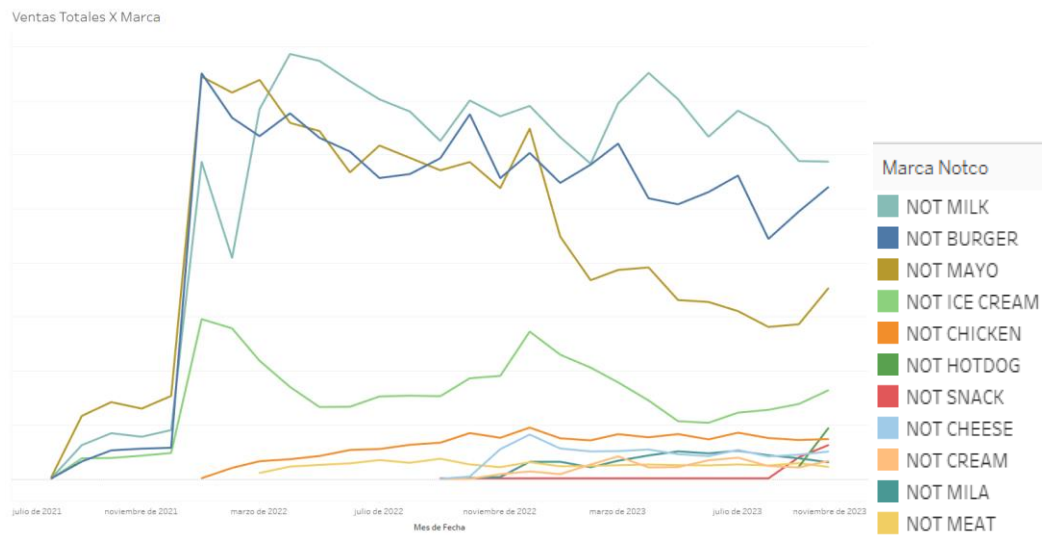


Figura 2: Gráfico Ventas Totales por Marcas

También es importante notar que generar proyecciones inexactas para las ventas de productos nuevos como Not Meat o Not Mila, que representan un 1,1% y 1,5% de las ventas totales respectivamente, no influye tanto como los productos más vendidos como Not Milk, Not Burger y Not Mayo, las cuales constituyen un 73% de las ventas totales (se observa en la Figura 6), por lo que es importante asegurar a lo menos una buena proyección y plan para este tipo de marca.

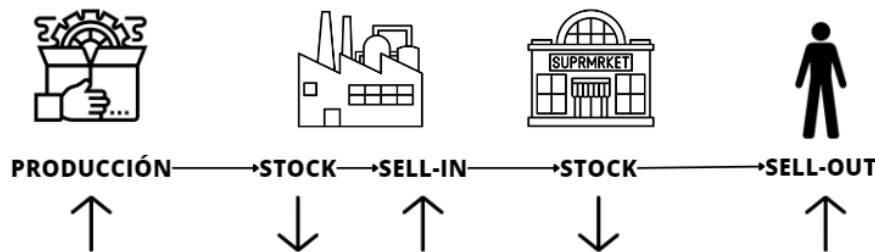
Marca Notco	Oct 2023
NOT MILK	29,14%
NOT BURGER	26,81%
NOT MAYO	17,52%
NOT ICE CREAM	8,16%
NOT CHICKEN	3,68%
NOT HOTDOG	4,68%
NOT SNACK	3,13%
NOT CHEESE	2,53%
NOT CREAM	1,68%
NOT MILA	1,55%
NOT MEAT	1,13%

*Tabla 3: Peso Marcas NotCo Octubre*

Notco enfrenta otro aspecto crítico relacionado con la falta de un plan de acción efectivo basado en el Sell-Out, que es la tendencia a sufrir quiebres de stock en sus productos debido a estimaciones inexactas de ventas en las cadenas de distribución. Esta situación podría llevar a pérdidas monetarias considerables, un tema que será analizado más detalladamente en la evaluación económica.

Es crucial entender que este problema no es exclusivo de la sede en Chile, sino que afecta a todas las sedes de la empresa. Mientras que Chile lidera en el uso de tecnologías avanzadas para el análisis de datos, seguido por Argentina y Brasil, otras sedes como Canadá, Estados Unidos y México están menos avanzadas en este aspecto. Esto destaca la necesidad de una solución global para optimizar las operaciones y alcanzar consistentemente los objetivos de ventas.

Además, Notco no está cumpliendo con las metas de ventas de Sell-Out, en parte debido a que las proyecciones se basan en métodos subjetivos y anticuados. Se requiere con urgencia un enfoque más preciso y actualizado para lograr los objetivos de ventas de manera efectiva y sostenible. Conocer de antemano un aumento estimado (por proyección) de Sell-Out permitiría tomar medidas preventivas contra la disminución del inventario en manos de los clientes, lo que a su vez impulsaría un incremento en las ventas a los clientes y, como resultado, una reducción del stock de la empresa. Este proceso conlleva un aumento en la producción para satisfacer la demanda, lo que podría ser anticipado y gestionado eficientemente con una adecuada planificación de ventas y acciones, evitando posibles complicaciones futuras.



*Figura 3: Esquema cadena de valor*

## Contexto de la Oportunidad

El desafío enfrentado en el área de Business Intelligence (BI) en The Not Company, específicamente en Chile, tiene repercusiones significativas a nivel global, dado que la empresa está en una etapa temprana de desarrollo en BI a nivel internacional. Por lo tanto, la implementación exitosa de una nueva solución para este problema cobra una importancia crítica.

Encontrar una solución eficaz para este desafío abre un abanico de oportunidades para la empresa. Mejorar la eficiencia operativa y aumentar las ventas son solo el comienzo. La optimización de la producción y la gestión de inventarios, junto con la personalización de estrategias de marketing, son aspectos clave que pueden beneficiarse significativamente. Además, una solución robusta en BI permite a la empresa adaptarse con mayor rapidez a los cambios en el mercado, expandir su alcance y obtener una ventaja competitiva sólida.

## Objetivos

### Objetivo General

Desarrollar un sistema para apoyar la planificación de presupuesto, relacionada con el Sell-Out y así disminuir el error de predicción del mes actual en un 15%.

### Objetivos Específicos

1. Planificación Estructurada del Sell-Out: Desarrollar un plan estructurado para el Sell-Out, detallado por cadena y marca, que presente presupuestos claros y precisos. Esto permitirá la toma de decisiones estratégicas basadas en datos, con el objetivo de mejorar la alineación de las proyecciones con los resultados reales.
2. Proyecciones de Ventas Modernizadas: Implementar técnicas estadísticas avanzadas para mejorar la fiabilidad de las proyecciones de ventas en un 15%, priorizando los datos recientes y eliminando la dependencia de métodos obsoletos. Esto se reflejará en una menor variación entre las cifras proyectadas y las ventas reales.
3. Integración de Herramientas de Business Intelligence: Integrar herramientas de Business Intelligence como Tableau para mejorar la visualización y comprensión de los datos de ventas Sell-Out. Se espera que este análisis mejorado conduzca a una planificación presupuestaria un 10% más rápida y una toma de decisiones más efectiva.
4. Optimización del Proceso de Generación de Informes: Optimizar el proceso de generación de informes para que el equipo de marketing reduzca el tiempo dedicado a esta tarea en un 25%. Este cambio permitirá que el equipo invierta más tiempo en estrategias de marketing y toma de decisiones operativas.

### Medidas de Desempeño

-Precisión Forecast: Esta medida evalúa la precisión de las proyecciones al comparar la raíz del error cuadrático medio con las previstas. Relacionado con el objetivo de "Proyecciones de Ventas Modernizadas". Para esto primero se calcularán los dos RMSE (Bi y actual) eso a través de esta fórmula:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (Real_i - Proyectado_i)^2}$$

Una vez obtenidos los resultados se utilizara la siguiente fórmula para sacarla diferencia % entre los RMSE, así se podrá visualizar con mayor facilidad el cambio de error que estaremos mejorando:

$$Diff\% = \left( \frac{RMSE1 - RMSE2}{RMSE2} \right) \times 100$$

- Disminución en el Tiempo de Generación de Información: Esta métrica directamente refleja el éxito del objetivo de "Optimización del Proceso de Generación de Informes", mostrando la eficiencia ganada en la generación de informes post-implementación. Sera calculada analizando el tiempo tomado en generar las proyecciones actuales contra la generación de las nuevas proyecciones de SO.

## Estado del Arte

Este documento analiza diversos modelos de pronóstico de ventas, destacando técnicas avanzadas y comparando diferentes metodologías.

Para este estudio se enfatiza la importancia de las variables predictoras en modelos de series temporales y se examinan enfoques que incluyen patrones temporales y modelos mixtos. La elección del modelo adecuado depende de factores como recursos disponibles, nivel de precisión deseada y el propósito específico en el contexto de su aplicación, con un enfoque especial en metodologías para series temporales que ajustan los datos para reflejar patrones periódicos y facilitar el análisis predictivo.

El modelo de serie temporal es esencial en el análisis de datos temporales (Mostrado en la siguiente ecuación del libro de Hyndman & Athanasopoulos, 2021), proporcionando una base sólida para comprender cómo evolucionan las variables a lo largo del tiempo. Este modelo se destaca por incorporar tanto externalidades como variables dependientes, incluyendo el término  $t+1$  y el componente de error. Su importancia radica en su capacidad para establecer patrones y predecir tendencias futuras, lo que lo convierte en un punto de

partida crucial para el análisis de series temporales antes de adentrarse en metodologías más complejas.

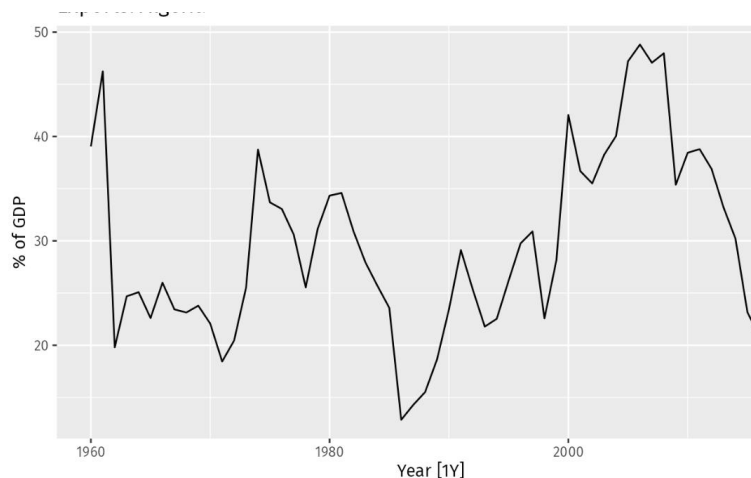
$$ED_{t+1} = f(ED_t, ED_{t-1}, ED_{t-2}, ED_{t-3}, \dots, error)$$

### Modelo de Suavizado Exponencial

Los modelos de suavizado exponencial son técnicas de pronóstico que utilizan promedios ponderados de datos históricos, otorgando mayor peso a las observaciones más recientes. Esta metodología es particularmente efectiva y eficiente en una variedad de series temporales, siendo altamente valorada en entornos industriales donde la rapidez y adaptabilidad son esenciales.

En el suavizado exponencial, hay varios enfoques, entre ellos:

- Suavizado Exponencial Simple: Adecuado para series sin tendencia o patrón estacional (ejemplo en Figura 4). Ajusta los pesos de las observaciones pasadas, dándole mayor relevancia a las más recientes. Se basa en un parámetro de suavizado, alpha, que regula el decaimiento de los pesos para observaciones antiguas, enfocándose en la previsión del tiempo  $t+1$ .



*Figura 4: Serie temporal sin tendencia ni patrón estacional (Hyndman & Athanasopoulos, 2021)*

- Modelo de Suavizado Exponencial Doble (o de Holt): Agrega un componente de tendencia al suavizado exponencial simple, ideal para datos con una tendencia clara, pero sin estacionalidad.

- Modelo de Suavizado Exponencial Triple (o de Holt-Winters): Incorpora tanto la tendencia como la estacionalidad, ajustándose a series temporales con patrones estacionales y tendencias, ya sean crecientes o decrecientes.

$$\hat{y}_{t+1|t} = \alpha y_t + \alpha(1 - \alpha)y_{t-1} + \alpha(1 - \alpha)^2 y_{t-2} + \dots ,$$

El parámetro alpha en los modelos de suavizado exponencial mostrado en la ecuación anterior (Hyndman & Athanasopoulos, 2021), es fundamental para equilibrar la influencia de los datos históricos y recientes: un alpha bajo enfatiza observaciones antiguas, mientras que un alto da más peso a los datos recientes, siendo útil en situaciones que requieren adaptación rápida a cambios. La selección del modelo y sus parámetros específicos depende de las características de la serie temporal y los objetivos del análisis de pronóstico.

### **Modelos ARIMA**

“Los modelos ARIMA proporcionan otro enfoque para el pronóstico de series temporales. El suavizado exponencial y los modelos ARIMA son los dos enfoques más utilizados para el pronóstico de series temporales y proporcionan enfoques complementarios al problema. Mientras que los modelos de suavizado exponencial se basan en una descripción de la tendencia y la estacionalidad de los datos, los modelos ARIMA pretenden describir las autocorrelaciones de los datos” (Hyndman & Athanasopoulos, 2021, cap. 9).

El modelo ARIMA es un enfoque de pronóstico que utiliza datos estadísticos para identificar patrones y predecir tendencias futuras. Se compone de tres elementos: el Autorregresivo (AR), que relaciona valores actuales con pasados; la Media Móvil (MA), que asume el error de predicción como una combinación lineal de errores previos; y el Integrado (I), que usa la diferencia entre valores consecutivos. Representados por los parámetros p, d y q, ARIMA es ideal para series temporales no estacionales. Para series con estacionalidad, se extiende al modelo SARIMAX, que añade componentes de estacionalidad (S) y factores exógenos (X), como promociones o cambios de precio, mejorando así la precisión y efectividad en la proyección de ventas. Es decir, SARIMAX será útil en los casos donde los residuos puedan exhibir una tendencia estacional o patrón. La sintaxis de este modelo se conforma por esta siguiente ecuación:

La ecuación del modelo en conjunto con su sintaxis es la siguiente:

$$SARIMAX([\beta], \mu, \sigma, d, [\varphi], [\theta], s, sd, [s\varphi], [s\theta])$$

- $[\beta]$  (Opcional). Son los coeficientes de la matriz de los factores exógenos.
- $\mu$  (Opcional). Es la media del modelo ARMA (Ej.  $\mu$ ). Si falta, la media es asumida como cero.
- $\sigma$  (Obligatorio). Es el valor de la desviación estándar del modelo residual/innovaciones.
- $d$  (Obligatorio). Es el orden no diferencial no estacional.
- $[p]$  (Opcional). Son los parámetros del modelo componente AR(p):  $[\varphi_1, \varphi_2 \dots \varphi_p]$  (comenzando con el retraso menor (lag)).
- $[\theta]$  (Opcional). Son los parámetros del modelo componente MA(q):  $[\theta_1, \theta_2 \dots \theta_q]$  (comenzando con el retraso menor (lag)).
- $s$  (Opcional). Es el número de observaciones por un periodo (Ej. 12 = Anual, 4 = Trimestral).
- $sd$  (Opcional). Es el orden diferencial estacional.
- $[s\varphi]$  (Opcional). Son los parámetros del componente del modelo AR estacional AR(PP):  $[s\varphi_1, s\varphi_2 \dots s\varphi_{pp}]$  (comenzando con el retraso menor (lag)).
- $[s\theta]$  (Opcional). Son los parámetros del componente del modelo MA estacional MA(QQ):  $[s\theta_1, s\theta_2 \dots s\theta_{qq}]$  (comenzando con el retraso menor (lag)).

En el libro “Definición de un Modelo SARIMAX” de Mohamad mencionan como los modelos ARIMA y SARIMAX son ampliamente usados en diversos sectores. En economía, se aplican para predecir indicadores como el PIB, la inflación y el desempleo. En el sector empresarial, son esenciales para pronosticar ventas, inventario y demanda de clientes. Por ejemplo, las empresas pueden usar este modelo para predecir ventas diarias considerando factores estacionales y tendencias de consumo, mientras que ARIMA puede ayudar a prever la demanda de productos para optimizar la producción e inventario. Estos modelos facilitan la toma de decisiones estratégicas y mejoran la eficiencia y rentabilidad de las organizaciones.



## Redes Neuronales

Las redes neuronales artificiales (ANN) y su variante para series temporales, las redes LSTM (Long Short Term Memory), representan avances significativos en el análisis de información y tecnología. Estas redes, estructuradas inspirándose en el cerebro humano, forman parte esencial del Machine learning y los algoritmos de Deep learning. Una ANN consta de varias capas de nodos, incluyendo capas de entrada, ocultas y de salida, con cada nodo enviando datos a la siguiente capa si supera un umbral específico.

Las ANN son cruciales para tareas exigentes como el reconocimiento de voz o de figuras y se utilizan ampliamente en plataformas como Google y en inteligencias artificiales avanzadas como Bing y GPT. Sin embargo, su implementación puede ser compleja, y en el caso de las series temporales, las redes LSTM, aunque capaces de almacenar cierta memoria de instancias anteriores, tienen limitaciones, como la dificultad para integrar efectos externos en las proyecciones.

## Solución

### Alternativas de Solución

Tras un análisis detallado, se identificaron tres metodologías clave para el pronóstico de ventas: Redes Neuronales (LSTM), Modelos ARIMA/SARIMAX y Suavizado Exponencial. Las redes LSTM son efectivas para series temporales complejas, pero limitadas en interpretación y factores externos. ARIMA es adecuado para tendencias estables, mientras que SARIMAX, al incluir componentes estacionales y exógenos, es más versátil para patrones estacionales y eventos externos. El Suavizado Exponencial, siendo más simple y adaptable, es ideal para datos con patrones estacionales o tendencias, enfatizando la simplicidad y rapidez en adaptarse a cambios de tendencia.

### Solución escogida

Se ha seleccionado el modelo SARIMAX como herramienta analítica para el proyecto en The Not Company, con los resultados presentados en Tableau. Se eligió por su capacidad para modelar estacionalidad y tendencias en series temporales, incluir factores externos, ofrecer una interpretación clara y requerir menos recursos computacionales. Este modelo es particularmente adecuado para analizar el impacto de variables exógenas en las ventas.

y proporciona transparencia en la toma de decisiones. Tableau, utilizado para la visualización, complementa al modelo al permitir una presentación efectiva y comprensible de las tendencias y predicciones, facilitando el análisis y la comunicación con los equipos. Juntos, SARIMAX y Tableau ofrecen una solución eficiente y precisa para mejorar las proyecciones de ventas y la toma de decisiones basada en datos.

## Desarrollo del Proyecto

### Metodologías

La metodología se basa en principios avanzados para la proyección de ventas y visualización de datos, enfocándose en el modelo SARIMAX y el uso de Tableau. Esta metodología combina un modelo estadístico avanzado con herramientas de visualización de vanguardia para mejorar la precisión en las proyecciones de ventas y optimizar los procesos de toma de decisiones.

#### **Metodología para desarrollar la solución**

**Recolección y Preparación de Datos:** El primer paso consiste en la recolección exhaustiva de datos de ventas históricos y variables exógenas que puedan influir en las proyecciones. Esto incluye la recopilación de datos internos de The Not Company y cualquier información relevante del mercado.

**Análisis Exploratorio de Datos:** Se realizará un análisis exploratorio para comprender las características fundamentales de los datos, identificar patrones, tendencias y anomalías, y preparar los datos para su posterior análisis con el modelo SARIMAX.

**Implementación del Modelo SARIMAX:** El modelo SARIMAX se implementará para realizar las proyecciones de ventas. Este proceso incluirá la selección de parámetros óptimos, la prueba de diferentes configuraciones del modelo y la validación de los resultados para garantizar su precisión y fiabilidad.

**Desarrollo de Informes con Tableau:** Se utilizará Tableau para crear informes y paneles de control que visualicen los resultados del modelo SARIMAX. Esto facilitará la interpretación de las proyecciones de ventas y permitirá una comunicación efectiva de los hallazgos a los equipos clave.

**Validación y Ajustes:** Las proyecciones generadas por el modelo SARIMAX se validarán continuamente comparándolas con los resultados de ventas reales. Se realizarán ajustes en el modelo y en los informes de Tableau según sea necesario para mejorar la precisión de las proyecciones.

**Documentación y Capacitación:** Se documentará todo el proceso de desarrollo y se proporcionará formación al personal relevante de The Not Company sobre cómo utilizar y entender los informes generados en Tableau. Esto asegurará que el proyecto tenga un impacto duradero y sostenible.

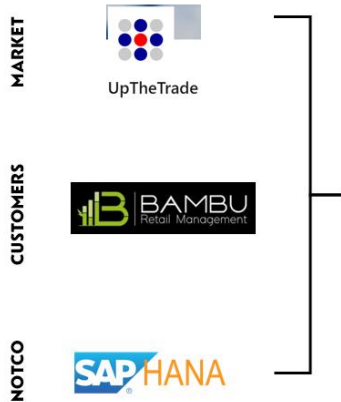
## Desarrollo

Durante la pasantía en el área de Business Intelligence de The Not Company, en Chile, se inició un desafío clave relacionado con la proyección de ventas de Sell-Out, que afectaba directamente la planificación del Replan y, por ende, las decisiones estratégicas de la empresa. Este proyecto significó un cambio notable desde la dependencia de informes de Excel hacia la adopción de herramientas más avanzadas como Tableau y Power BI.

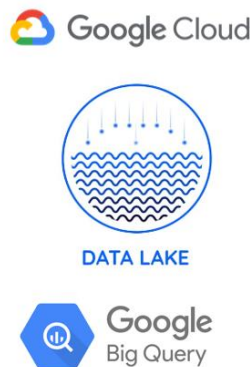
El proyecto comenzó con un enfoque en adquirir un profundo entendimiento de la dinámica de los datos en la industria alimenticia y en aprender a utilizar eficazmente las herramientas de BI para una mejor visualización de los datos. Estas herramientas se utilizaron no solo para generar reportes de sell-in, sell-out, market share, fill rate y stock, sino también para crear visualizaciones claras que ayudaran a comprender el proyecto de proyección.

Por otro lado, una parte crucial del trabajo fue familiarizarse con el origen y manejo de las bases de datos. Para contextualizar, los datos de sell-in provenían de SAP y eran organizados por Cortex, mientras que la información del market share se obtenía de Upthetrade. Estos datos se añadían al datalake de la empresa en BigQuery, lo que implicó un aprendizaje adicional en el manejo de esta plataforma. Además, la información sobre sell-out y stock, así como los datos de clientes del canal moderno, eran proporcionados por Bambú a través de una API. Esta API fue esencial, ya que integraba datos de múltiples B2B, utilizando los SKU como puntos de conexión para generar una base de datos coherente y precisa. Esta información se puede visualizar en la siguiente Figura extraída de una presentación de introducción a lo que hace el equipo BI.

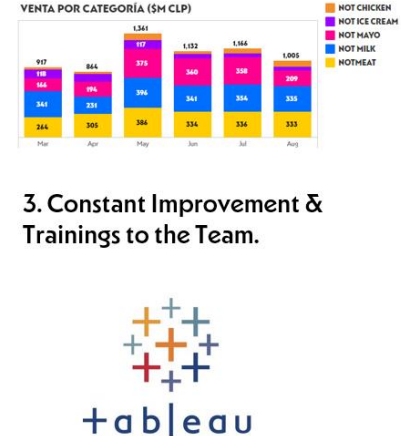
**1. Data Management (real time)**



**2. DDBB Architecture Design**



**3. Visual Deployment**



*Figura 5: Flujo de trabajo bases de datos NotCo (Presentaciones BI)*

El proyecto tomó un enfoque particular en el sell-out, área donde la empresa Bambú jugó un rol crucial. Bambú proporcionaba datos esenciales a través de una API, recopilando información de stock, sell-out y los locales de los clientes B2B. La complejidad de esta tarea residía en la diversidad y variabilidad de las bases de datos de cada B2B. Para afrontar este desafío, se establecieron puntos de conexión a través de los SKU, generando una base de datos maestra (Generada por el área BI) que enlazaba los SKU de cada B2B con los códigos EAN de Notco.

Upc	Categoría	Marca Notco	Descri Item	skub2b_cencc	skub2b_walm	skub2b_tottus	skub2b_smu
7804659650370	BURGER	NOT BURGER	NotBurger Pack 5x100gr 500gr	1846119	-	20538485	-
7804659650387	BURGER	NOT BURGER	NotBurger Flowpack 100gr	1846118	743635	20538484	-
7804659650745	BURGER	NOT BURGER	NotBurger Flowpack XL 150gr	1897160	-	20873368	-
7804659650790	BURGER	NOT BURGER	NotChicken Burger Naked 100gr	1897161	857753	20873369	661658
7804659650943	APANADOS	NOT CHICKEN	NotChicken Burger Crispy 95gr	1930828	111691	20934239	-

*Tabla 4: Maestra Upc/Ean Notco vs SKU Clientes (Acortada)*

Esta conexión aseguraba la coherencia y exactitud de la información. Posteriormente, se creó otra conexión entre esta base de datos maestra y la API de Bambú, enfocándose en la unión de los códigos EAN, lo que permitió formar una base de datos organizada y fiable. El entendimiento profundo de estos sistemas y la integración efectiva de las bases de datos fueron pasos preliminares cruciales en el proyecto.

La identificación del problema central en el proyecto de Business Intelligence de The Not Company se inició con una serie de entrevistas con líderes de diversas áreas de la empresa, incluyendo ventas, marketing, finanzas y producción. Fue especialmente relevante para el área de producción, encargada de generar el plan S&OP para la planificación de

producción. Estas conversaciones revelaron la necesidad de una planificación más detallada, destacando que las proyecciones inexactas del sell-out impactaban negativamente en toda la cadena de valor.

Un análisis detallado de las proyecciones existentes descubrió numerosas inexactitudes. Por ejemplo, en el mes de noviembre, la venta real fue de 1063 millones, mientras que la proyección sell-out actual fue de 1370 millones, mostrando una discrepancia significativa que se repetía en varias ocasiones. Este hallazgo condujo a una investigación más profunda del proceso de proyección de ventas, evidenciando que la metodología empleada era demasiado manual y propensa a errores.

Para abordar este problema, se exploraron diferentes modelos de proyección de ventas. Inicialmente, como novato en el área, se comenzó con modelos más complicados sin una comprensión clara de las necesidades específicas del proyecto. Este enfoque resultó en una pérdida de tiempo valioso. Sin embargo, tras una investigación más detallada (indicada en el estado del arte) y una mejor comprensión de los requisitos del proyecto, se seleccionó finalmente el modelo SARIMAX por su eficiencia y precisión. La elección de este modelo marcó un punto de inflexión en el proyecto, permitiendo un avance más estructurado y enfocado en la solución de los problemas identificados.

### **Exploración de Datos**

En el proyecto de The Not Company, se tomó una decisión crucial durante el análisis de ventas, especialmente en lo que respecta a los productos nuevos. Dada la limitada cantidad de datos disponibles para estos productos (aproximadamente un mes), se optó por iniciar las proyecciones con modelos lineales para estos productos, dado que el modelo entrega mejores resultados si se entrena con data que muestre estacionalidad, por lo que pocos meses de información no funcionaría para una buena proyección. Esta estrategia era fundamental para asegurar que la escasez de datos no afectara negativamente la precisión de los resultados deseados.

Paralelamente, se enfrentó el reto de manejar la irregularidad en los datos diarios. Algunos días presentaban una gran cantidad de datos, mientras que otros tenían muy pocos. Esta variabilidad planteó un desafío en términos de cómo ajustar mejor el modelo de proyección de ventas. Tras un análisis detallado, se determinó que la mejor solución era la agrupación diaria de datos. Aunque esta elección implicaba un tiempo de entrenamiento más largo para el modelo, se justificaba por la precisión de los resultados obtenidos.

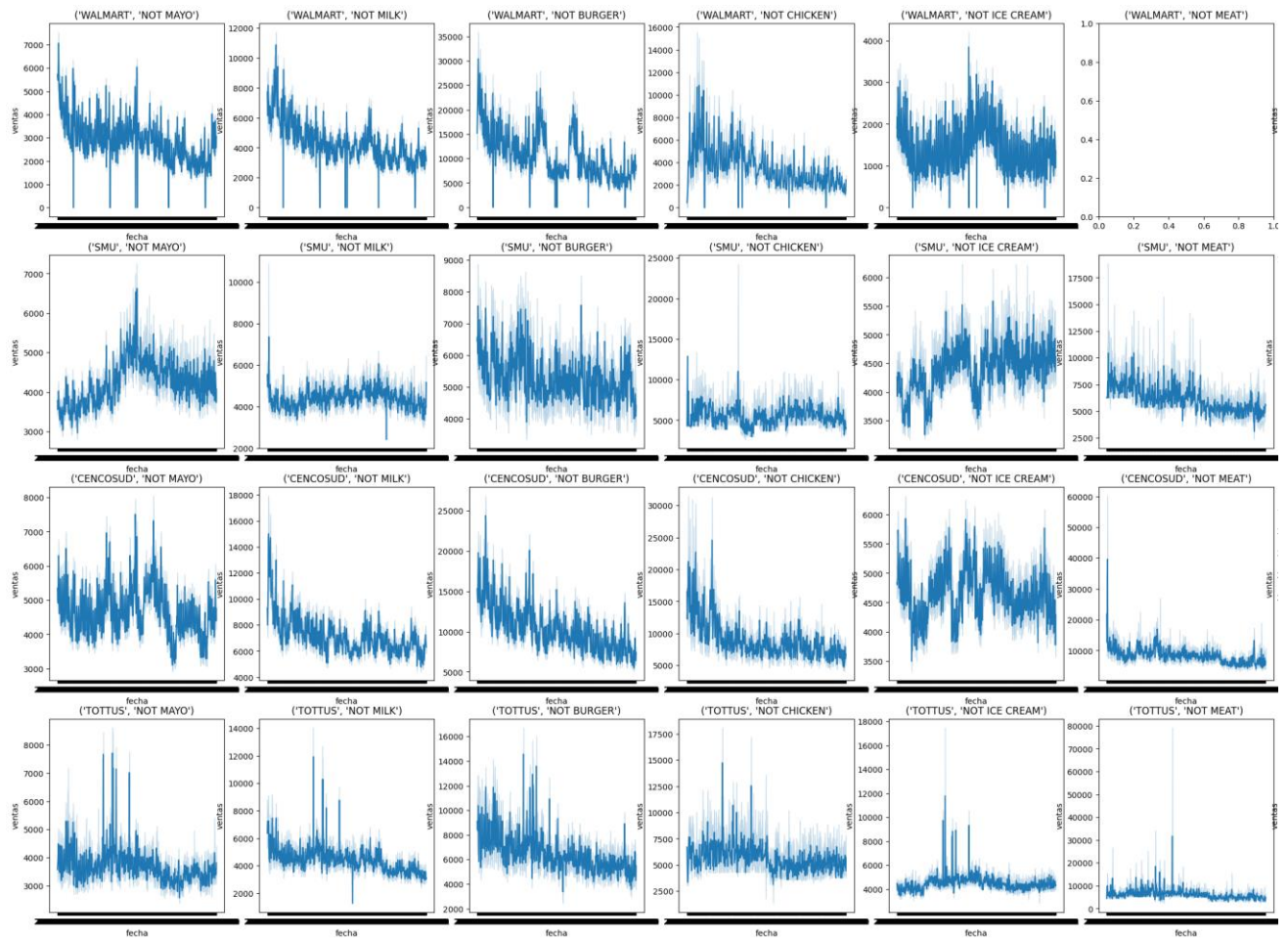
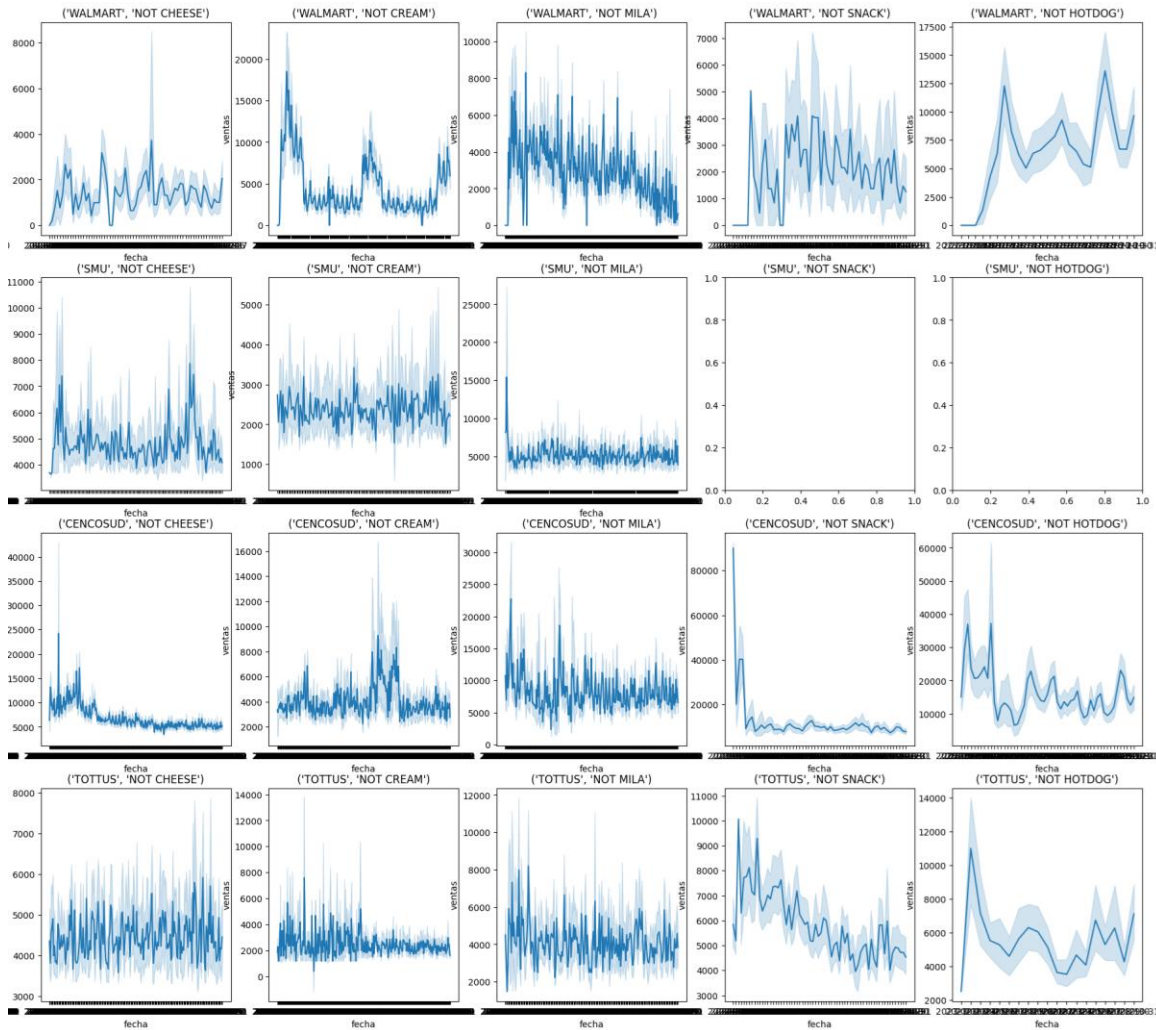


Figura 6: Lista de Dataframes graficados Cliente x Marca (Código elaboración propia)





*Figura 7: Lista de Dataframes graficados Cliente x Marca (Código elaboración propia)*

Este enfoque de agrupación diaria se complementó con el análisis realizado a través de dos figuras clave (figuras 6 y 7), que mostraban las tendencias de ventas de diversas marcas y productos en su estado original, no agrupado periódicamente, pero sí evidenciando tendencias. Estas figuras fueron fundamentales para entender la dinámica del mercado y los patrones de venta.

La combinación de estos enfoques permitió abordar eficientemente los desafíos presentados tanto por los productos establecidos como por los nuevos lanzamientos, formando una estrategia integral que mejoró significativamente la capacidad de proyección y análisis de ventas de la empresa.

## Flujo del Código

Con un enfoque dirigido hacia la base del proyecto, se explicará el flujo del código para la generación de las proyecciones. Este inicia con la importación de bibliotecas clave en Python donde se utilizan `datetime` y `calendar` para el manejo eficiente de fechas, mientras que `numpy` y `pandas` son esenciales para el análisis de datos. La biblioteca `optuna` se integra para la optimización de hiperparámetros en el modelo SARIMAX, y `statsmodels` se emplea para su construcción. Las herramientas de `sklearn.metrics` facilitan la evaluación del modelo mediante métricas como MSE y RMSE.

El proyecto avanza con la extracción de datos a través de la API de Bambu, obteniendo tablas fundamentales como Locales, EAN y Sell-Out, que aportan datos cruciales sobre ventas y productos. Estos datos se transforman en dataframes de `pandas`, preparándolos para el análisis. La fusión de estas tablas en `pandas` crea un conjunto de datos integrado que refleja distintos aspectos del negocio.

Posteriormente, los datos se desagregan por holding y marca, generando dataframes específicos para un análisis más detallado y adaptado a segmentos particulares del negocio, lo que podemos visualizar directamente en las figuras 6 y 7. Este nivel de detalle es crucial para predicciones y análisis precisos.

El núcleo del análisis es el modelo SARIMAX. Aquí, los datos se preparan, se define el modelo y se optimiza con `optuna`, incluyendo la conversión de datos a series temporales diarias y la normalización de las predicciones. Esto asegura consistencia y comparabilidad con los datos históricos.

La función `ELECTRONIC_mensual` es clave en el proceso, donde se procesan los datos para generar predicciones de ventas. Se implementa un bucle que recorre los dataframes desagregados, y se utiliza un manejo de excepciones para garantizar la robustez del proceso. Los resultados, almacenados en diccionarios, se preparan para la exportación.

Este flujo lo podemos ver representado en detalle en la figura 8.

Finalmente, las predicciones se exportan a Excel, facilitando su uso en decisiones empresariales y su aplicación en plataformas como Tableau.



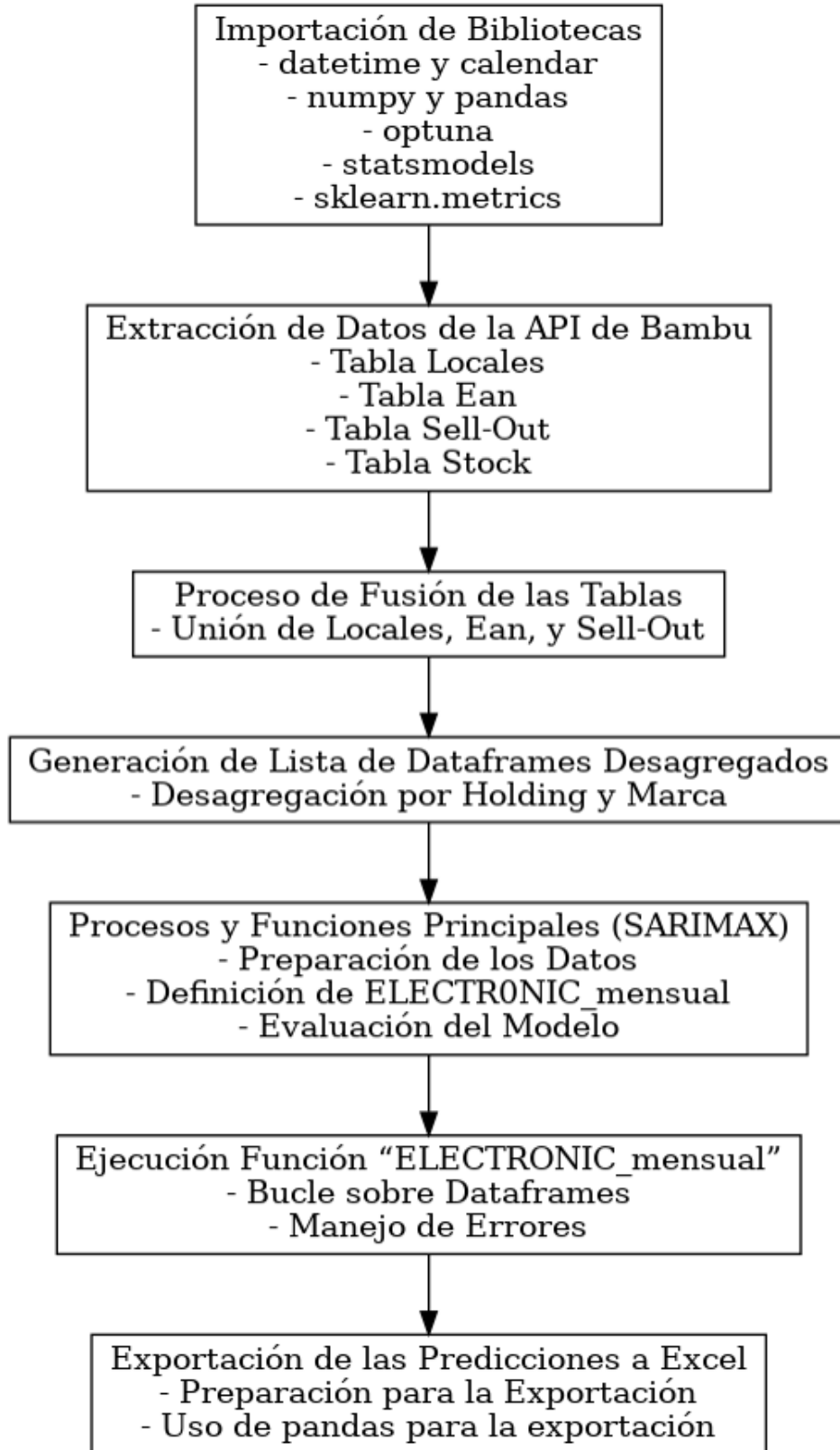


Figura 8: Flujo del Código (Generación Propia)

Cabe destacar que este mismo flujo y el proceso detallado del código puede ser revisado en el anexo, donde se muestran figuras en orden para visualizar cada parte del flujo detallado anteriormente.

En la fase final del proyecto en The Not Company, se realizó un paso crucial: la integración de los datos de ventas reales con las predicciones del modelo SARIMAX. Antes de esta integración, fue necesario esperar a obtener los datos reales de ventas para poder realizar una comparación efectiva. Esta integración, detallada en figura 9, permitió comparar directamente las ventas reales con las proyecciones del modelo y analizar el error cuadrático medio (RMSE) de estas predicciones. Este análisis fue esencial para evaluar la precisión del modelo y para informar las decisiones estratégicas futuras de la empresa.

## FORECAST SELL-OUT NOVIEMBRE

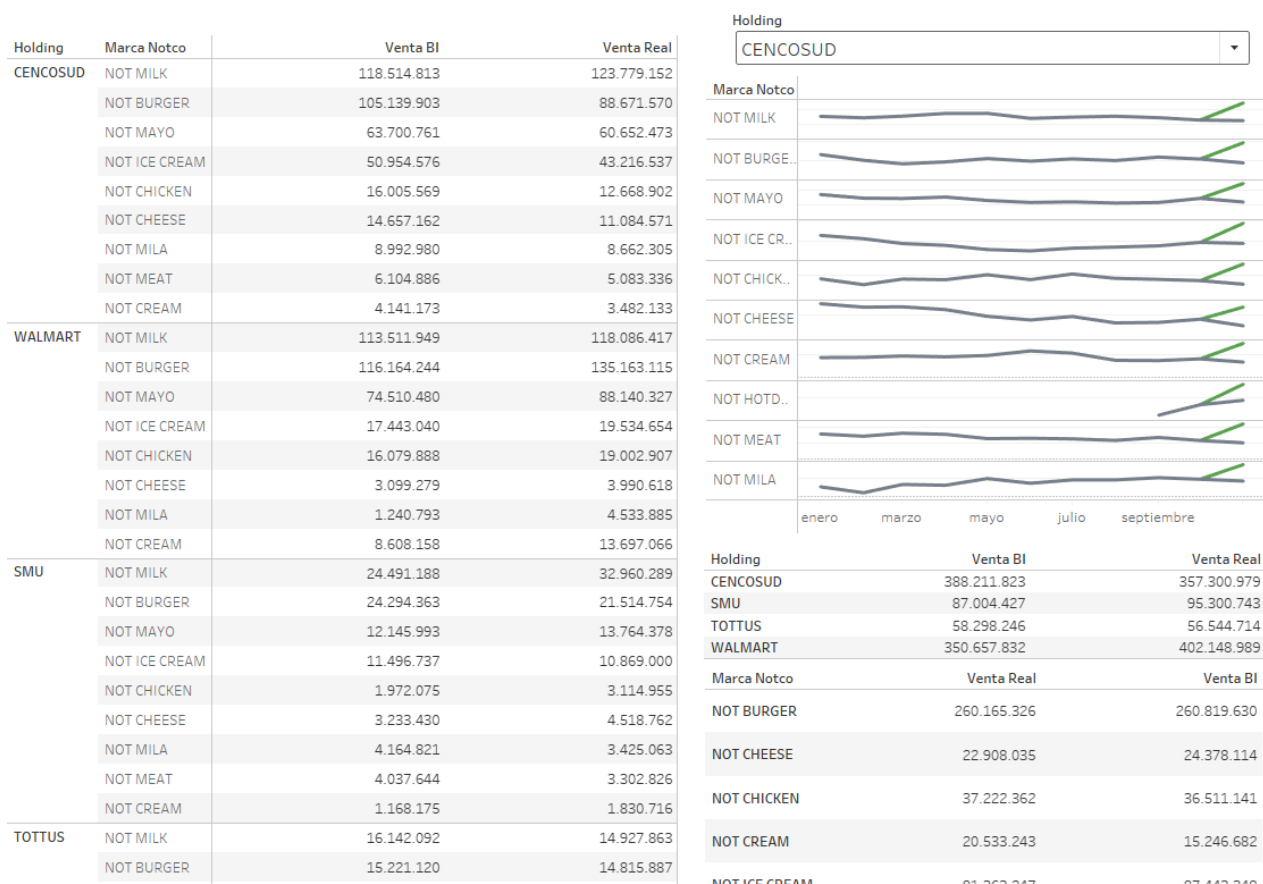


Figura 9: Visualización resultados en Tableau (Reporte elaboración propia)

## Plan de Implementación

En el plan de implementación del proyecto de Business Intelligence en la compañía, se estableció una secuencia lógica y orientada a objetivos, enfocándose en maximizar el impacto y la utilidad del modelo de proyección de ventas. El proceso se desarrolló en varias etapas clave:

### 1. Validación del Modelo:

La primera etapa consistió en validar la eficacia del modelo SARIMAX, comprobando si los resultados obtenidos cumplían con los objetivos planteados. Esta validación era crucial para asegurar que el modelo proporcionara proyecciones fiables y útiles para la planificación de ventas y producción.

### 2. Presentación de Resultados:

Una vez confirmada la validez del modelo, se buscó la mejor manera de presentar los resultados a los equipos clave de The Not Company. El objetivo principal era aportar una proyección de ventas más precisa, lo que a su vez facilitaría una mejor planificación del sell-in, mejorando el S&OP y el replan. Esto contribuiría significativamente a reducir los costes asociados a una mala planificación. Esta visualización se muestra en la figura 9.

### 3. Integración en Reportes:

El segundo objetivo era integrar las proyecciones en los reportes existentes o crear nuevos reportes para que los equipos de ventas (como el mostrado en la figura 9), marketing y finanzas tuvieran acceso a estas proyecciones mejoradas. Esta integración permitiría a los distintos equipos tomar decisiones más informadas, como planificar promociones o establecer metas de ventas. Una idea para esta etapa de implementación es poder agregar las nuevas proyecciones a un reporte de ventas SO (Mostrado en la figura 10), el cual se envía vía email todos los días. Mostrando dentro una comparación entre las ventas reales de este mes 2023, del mismo mes, pero 2022 y de las proyecciones generadas para ese mes. Así puede existir un acceso fácil a la información de predicción aportando a decisiones diarias de los equipos.

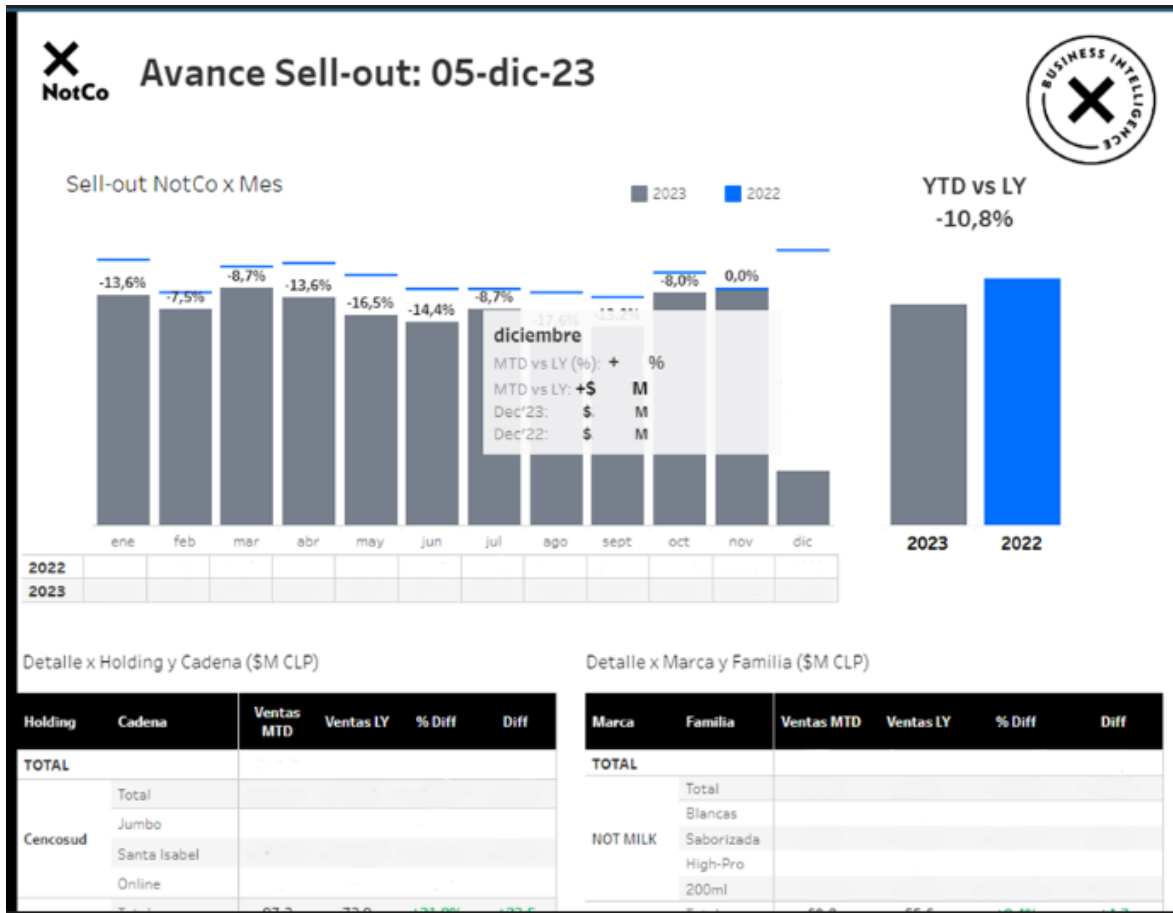


Figura 10: Reporte diario Sell-Out (Reportes Bi) (Enmascarado por confidencialidad)

#### 4. Colaboración con Líderes de Equipo:

Se programaron reuniones con líderes de equipo para discutir cómo los resultados del modelo podrían beneficiar a cada área. Se buscó feedback sobre qué tipo de visualizaciones y formatos de reporte serían más útiles para ellos. Aunque la decisión final sobre la presentación de los resultados recaería en el equipo de Business Intelligence, estas reuniones eran fundamentales para garantizar que el informe final atendiera las necesidades de los distintos equipos.

#### 5. Implementación y Distribución Regular:

Se decidió ejecutar el modelo cada dos semanas, coincidiendo con el principio y la mitad del mes, para asistir en el replan y realizar los ajustes necesarios. Además, se planificó la implementación de la solución en un reporte accesible para los distintos equipos, señalado

anteriormente para la figura 10. Este enfoque garantizaría que todos los equipos estuvieran informados sobre las tendencias actuales de las ventas y pudieran ajustar sus estrategias y operaciones en consecuencia.

Este plan de implementación se diseñó para asegurar que los resultados del modelo fueran utilizados de manera efectiva y eficiente, impactando positivamente en la toma de decisiones estratégicas y operativas de la empresa.

## Análisis de Riesgo

El presente análisis de riesgo se centra en la implementación de un nuevo sistema de apoyo de decisiones para la proyección de ventas en The Not Company. Este sistema busca mejorar la precisión de las proyecciones actuales y proporcionar una guía más clara para las decisiones estratégicas, especialmente en el contexto de las revisiones del S&OP (Sales and Operations Planning) y el ajuste mensual del replan. A pesar de sus beneficios potenciales, es crucial identificar y abordar los posibles riesgos asociados con la implementación y el funcionamiento de este sistema.

### Identificación de Riesgos

Riesgo	Probabilidad	Impacto	Gravedad	Plan de Mitigación
Técnico	Media	Alto	Medio-Alto	Realizar pruebas exhaustivas de integración y compatibilidad antes de la implementación completa.
Precisión de Datos	Alta	Muy Alto	Muy Alto	Establecer protocolos para la actualización y verificación regular de los datos ingresados al sistema.
Operativo	Alta	Medio	Medio-Alto	Proporcionar capacitación y soporte adecuado al personal durante la transición al nuevo sistema.
Dependencia del Sistema	Media	Muy Alto	Alto	Mantener un enfoque equilibrado, utilizando el sistema como una herramienta de apoyo y no como la única fuente para la toma de decisiones.

*Tabla 5: Matriz de Riesgo (Elaboración Propia)*

**Riesgo Técnico:** Existe la posibilidad de que el nuevo sistema no se integre adecuadamente con las plataformas y procesos existentes, lo que podría llevar a errores en los datos o incompatibilidades técnicas.

**Riesgo de Precisión de Datos:** Aunque el sistema busca mejorar la precisión en las proyecciones, existe el riesgo de que los datos alimentados al sistema no sean lo suficientemente precisos o estén desactualizados, afectando la calidad de las proyecciones generadas.

**Riesgo Operativo:** La adopción de un nuevo sistema puede enfrentar resistencia por parte del personal o requerir una curva de aprendizaje que afecte temporalmente la eficiencia operativa.

**Riesgo de Dependencia del Sistema:** Una dependencia excesiva en las proyecciones del sistema podría llevar a una menor atención a los factores de mercado cambiantes que el sistema no puede prever.

La implementación del nuevo sistema de apoyo de decisiones para la proyección de ventas en The Not Company conlleva varios riesgos que deben ser gestionados cuidadosamente. Sin embargo, con planes de mitigación adecuados y una constante evaluación de riesgos, el proyecto tiene el potencial de proporcionar un valor significativo y mejorar las decisiones de ventas y operaciones de la empresa.

## Evaluación Económica

El actual proyecto de mejora contiene una característica importante, este no requiere inversiones o costos adicionales, por lo cual la evaluación se enfocará en la disminución del error del costo de oportunidad mediante la mejora en la precisión de las proyecciones de ventas. Al alinear mejor la producción con las ventas reales, se busca reducir estos costos de inventario y almacenamiento y los riesgos de pérdidas de ventas o daño a la reputación. Siguiendo el objetivo SMART del proyecto de desarrollar un sistema para la planificación del Sell-Out y reducir los errores de predicción del mes actual en un 15%, la utilización del modelo SARIMAX para mejorar la precisión de las proyecciones promete mejorar la rentabilidad y eficiencia operativa, alineándose con los objetivos establecidos.

### Error de Baja y Alta Producción

La metodología del proyecto en The Not Company permite calcular directamente las pérdidas monetarias por baja producción, presentando los quiebres por falta de stock en términos monetarios y ofreciendo una clara estimación de las pérdidas económicas por no satisfacer la demanda del mercado. En la tabla 6, se muestra el quiebre monetario por marca debido a ventas menores a las órdenes de caja. El análisis de la siguiente Figura se

enfoca en los datos de octubre, proporcionando una visión detallada de las consecuencias financieras de no alcanzar las expectativas de venta, y resalta la necesidad de una planificación y producción precisas para reducir pérdidas y mejorar la eficiencia económica de la empresa. Las marcas se han enmascarado por privacidad.

Marca	Fill-rate	Cantidad Orden Caja	Cantidad Facturada	Cantidad Corte	Quiebre Valorizado
Grand Total	96,8%	95.621	92.542	-2.860	103,00M
	100,0%	2.285	2.285	0	0,00M
	96,5%	27.264	26.323	-940	34,23M
	93,1%	3.667	3.413	-254	15,94M
	100,0%	1.600	1.600	0	0,00M
	100,0%	132	132	0	0,00M
	97,0%	10.445	10.128	-316	6,41M
	95,2%	10.554	10.047	-510	14,33M
	96,6%	1.060	1.024	-36	1,43M
	99,2%	32.429	32.184	-26	0,52M
	93,6%	2.576	2.410	-165	7,68M
	85,1%	4.109	3.496	-613	22,46M

*Tabla 6: Reporte Fill-Rate Notco: Apertura Marca (Reportes Bi) (Elaboración propia)*

En las figuras siguientes, se refleja la pérdida debida a la mala planificación en The Not Company en los meses de septiembre a noviembre, abierto por motivos de pérdida. Esta visualización muestra cómo la planificación es responsable en promedio de casi el 50% del quiebre valorizado. Este dato es revelador de la necesidad crítica de mejorar las proyecciones de ventas, ya que una proyección más precisa es fundamental para lograr una planificación más efectiva. La mejora en las proyecciones puede contribuir significativamente a la reducción de pérdidas asociadas con errores en la planificación, enfatizando la importancia de una precisión y análisis mejorados en las estrategias de planificación de la empresa.

## Septiembre

Motivo Rechazo Not..	Cantidad Corte	%	Quiebre Valorizado
<b>Total</b>	<b>-2.310</b>	<b>100,0%</b>	<b>82,69M</b>
Error de planificación	-984	42,6%	35,17M
Error del Copacker	-613	26,5%	21,20M
Liberación de Calidad	-268	11,6%	8,95M
Dropsiz	-187	8,1%	6,98M
Producto defectuoso	-139	6,0%	4,36M
Quiebre por materia ..	-119	5,2%	6,03M

## Octubre

Motivo Rechazo Not..	Cantidad Corte	%	Quiebre Valorizado
<b>Total</b>	<b>-2.930</b>	<b>100,0%</b>	<b>105,51M</b>
Error del Copacker	-1.429	48,8%	51,52M
Error de planificación	-1.410	48,1%	50,97M
Dropsiz	-65	2,2%	2,49M
Decision del Negocio	-26	0,9%	0,52M

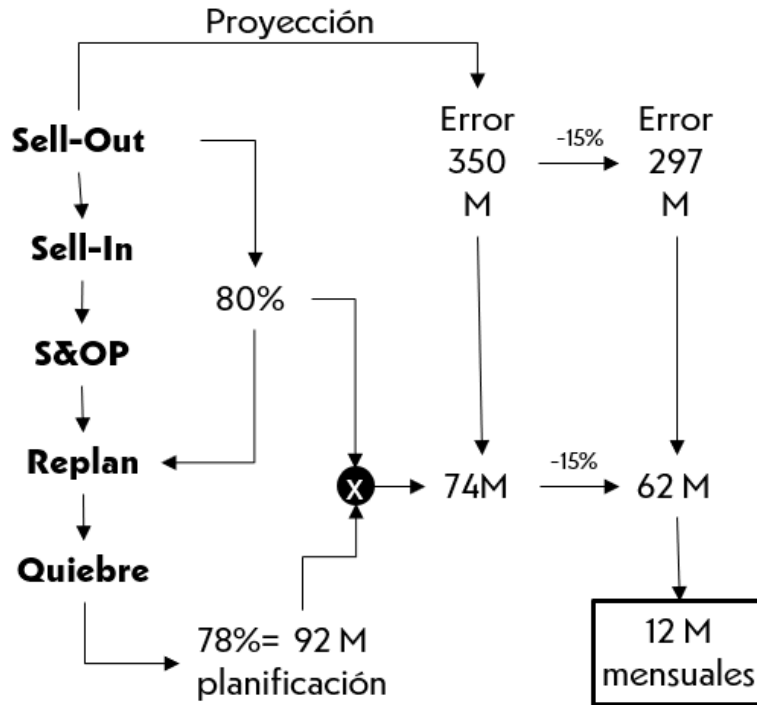
## Noviembre

Motivo Rechazo Not..	Cantidad Corte	%	Quiebre Valorizado
<b>Total</b>	<b>-4.825</b>	<b>100,0%</b>	<b>120,83M</b>
Error de planificación	-3.756	77,8%	92,39M
Mal solicitado por Cli..	-362	7,5%	7,95M
Back Order	-339	7,0%	8,63M
Error del Copacker	-187	3,9%	6,44M
Dropsiz	-181	3,8%	5,42M

*Tabla 7: Reporte Fill-Rate Notco: Motivo de perdida Sep-Nov (Reportes Bi) (Elaboración propia)*

En The Not Company, un error en la planificación en noviembre condujo a un quiebre valorado en 92,39 millones de CLP. Mejorar la precisión de la planificación en un 15%, conforme al objetivo del proyecto, podría teóricamente reducir esta pérdida en un monto proporcional, aproximadamente 12 millones de CLP al mes. Esto se puede visualizar en el siguiente diagrama:





*Figura 11: Diagrama ejemplo de mejora presupuestaria (Elaboración propia)*

La proyección del Sell-In y el proceso de Sales and Operations Planning (S&OP) están completamente determinados por la proyección de sell-out. Durante la generación del replan, el S&OP es sometido a un proceso de validación por parte de los gerentes de producción. Esta etapa suele incluir una inflación deliberada de las cifras para establecer metas de producción más altas, lo que se traduce en un aumento en las ventas proyectadas. Esta inflación representa aproximadamente un 20% sobre las cifras reales. Por lo tanto, se puede concluir que el Replan está influenciado en un 80% por el sell-out.

Además, dado que los errores en la planificación se derivan principalmente del replan y que los quiebres están directamente relacionados con este, se estima que el 80% de los errores de planificación se pueden atribuir a la proyección de sell-out. En este contexto, se observa que el 80% del quiebre, equivalente a 74 millones, está directamente relacionado con la proyección de sell-out. Por lo tanto, una reducción del 15% en el error de proyección del sell-out se traduciría, como indica el diagrama, en un ahorro mensual de 12 millones.

Estas cifras destacan la importancia y el impacto positivo de optimizar la planificación y proyección, lo cual puede mejorar significativamente la eficiencia económica y la salud financiera de la empresa a largo plazo.

## Conclusión

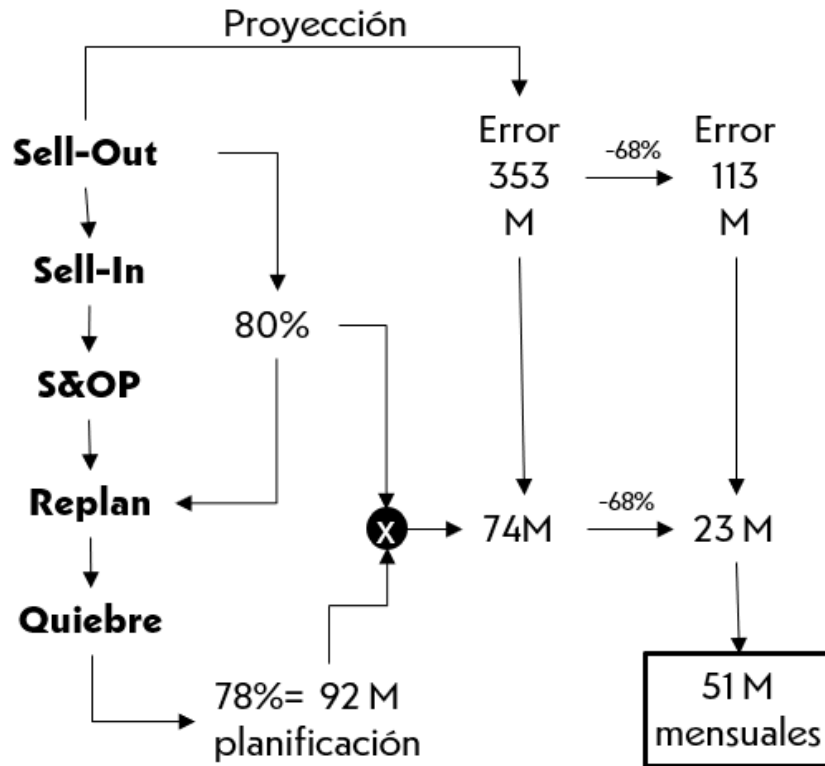
La implementación de un sistema predictivo avanzado para la optimización de la planificación presupuestaria y proyecciones de ventas Sell-Out en The Not Company ha demostrado ser un cambio significativo en la eficiencia y precisión de las proyecciones. A través del uso del modelo SARIMAX y la integración con herramientas de Business Intelligence como Tableau, se logró una reducción significativa del error en las proyecciones de ventas.

Una tabla comparativa (Tabla 8), desglosando los resultados por holding y marca, ilustra claramente la superioridad de las proyecciones generadas con este nuevo sistema frente a las proyecciones actuales y las ventas reales. Los datos indican una disminución notable en el error de proyección, evidenciada tanto en la comparación de los valores RMSE como en el porcentaje de mejora en la precisión de las proyecciones. La fórmula utilizada para medir el desempeño muestra una disminución del 68% en el error de las proyecciones actuales, destacando el impacto sustancial de este nuevo enfoque en la precisión de la planificación.

Holding	Marca Notco	Proyección BI	Venta Real	Proyección Act	Rmse BI	Rmse Act	% Rmse
CENCOSUCO	NOT MILK	\$ 118.514.813	\$ 123.779.152	\$ 168.655.338	\$ 5.264.339	\$ 44.876.186	88%
	NOT BURGER	\$ 105.139.903	\$ 88.671.570	\$ 154.431.751	\$ 16.468.333	\$ 65.760.181	75%
	NOT MAYO	\$ 63.700.761	\$ 60.652.473	\$ 100.423.513	\$ 3.048.288	\$ 39.771.040	92%
	NOT ICE CREAM	\$ 50.954.576	\$ 43.216.537	\$ 43.197.797	\$ 7.738.039	\$ 18.740	-41191%
	NOT CHICKEN	\$ 16.005.569	\$ 12.668.902	\$ 21.639.662	\$ 3.336.667	\$ 8.970.760	63%
	NOT CHEESE	\$ 14.657.162	\$ 11.084.571	\$ 14.778.325	\$ 3.572.591	\$ 3.693.754	3%
	NOT MILA	\$ 8.992.980	\$ 8.662.305	\$ 9.783.529	\$ 330.675	\$ 1.121.224	71%
	NOT MEAT	\$ 6.104.886	\$ 5.083.336	\$ 5.218.487	\$ 1.021.550	\$ 135.151	-656%
	NOT CREAM	\$ 4.141.173	\$ 3.482.133	\$ 4.394.874	\$ 659.040	\$ 912.741	28%
	<b>TOTAL</b>	<b>\$ 388.211.823</b>	<b>\$ 357.300.979</b>	<b>\$ 522.523.277</b>	<b>\$ 41.439.522</b>	<b>\$ 165.259.779</b>	<b>75%</b>
WALMART	NOT MILK	\$ 113.511.949	\$ 118.086.417	\$ 112.997.407	\$ 4.574.468	\$ 5.089.010	10%
	NOT BURGER	\$ 116.164.244	\$ 135.163.115	\$ 233.302.443	\$ 18.998.871	\$ 98.139.328	81%
	NOT MAYO	\$ 74.510.480	\$ 88.140.327	\$ 75.242.546	\$ 13.629.847	\$ 12.897.781	-6%
	NOT ICE CREAM	\$ 17.443.040	\$ 19.534.654	\$ 23.040.332	\$ 2.091.614	\$ 3.505.678	40%
	NOT CHICKEN	\$ 16.079.888	\$ 19.002.907	\$ 16.970.281	\$ 2.923.019	\$ 2.032.626	-44%
	NOT CHEESE	\$ 3.099.279	\$ 3.990.618	\$ 8.886.857	\$ 891.339	\$ 4.896.239	82%
	NOT MILA	\$ 1.240.793	\$ 4.533.885	\$ 9.616.511	\$ 3.293.092	\$ 5.082.626	35%
	NOT CREAM	\$ 8.608.158	\$ 13.697.066	\$ 13.071.031	\$ 5.088.908	\$ 626.035	-713%
	<b>TOTAL</b>	<b>\$ 350.657.832</b>	<b>\$ 402.148.989</b>	<b>\$ 493.127.408</b>	<b>\$ 51.491.157</b>	<b>\$ 132.269.323</b>	<b>61%</b>
SMU	NOT MILK	\$ 24.491.188	\$ 32.960.289	\$ 39.287.541	\$ 8.469.101	\$ 6.327.252	-34%
	NOT BURGER	\$ 24.294.363	\$ 21.514.754	\$ 30.904.315	\$ 2.779.609	\$ 9.389.561	70%
	NOT MAYO	\$ 12.145.993	\$ 13.764.378	\$ 26.034.702	\$ 1.618.385	\$ 12.270.324	87%
	NOT ICE CREAM	\$ 11.496.737	\$ 10.869.000	\$ 13.192.836	\$ 627.737	\$ 2.323.836	73%
	NOT CHICKEN	\$ 1.972.075	\$ 3.114.955	\$ 2.720.753	\$ 1.142.880	\$ 394.202	-190%
	NOT CHEESE	\$ 3.233.430	\$ 4.518.762	\$ 2.221.714	\$ 1.285.332	\$ 2.297.048	44%
	NOT MILA	\$ 4.164.821	\$ 3.425.063	\$ 5.111.395	\$ 739.758	\$ 1.686.332	56%
	NOT MEAT	\$ 4.037.644	\$ 3.302.826	\$ 3.289.734	\$ 734.818	\$ 13.092	-5513%
	NOT CREAM	\$ 1.168.175	\$ 1.830.716	\$ 1.553.825	\$ 662.541	\$ 276.831	-139%
	<b>TOTAL</b>	<b>\$ 87.004.427</b>	<b>\$ 95.300.743</b>	<b>\$ 124.316.815</b>	<b>\$ 18.060.160</b>	<b>\$ 34.978.537</b>	<b>48%</b>
TOTTUS	NOT MILK	\$ 16.142.092	\$ 14.927.863	\$ 21.916.830	\$ 1.214.229	\$ 6.988.967	83%
	NOT BURGER	\$ 15.221.120	\$ 14.815.887	\$ 19.782.656	\$ 405.233	\$ 4.966.770	92%
	NOT MAYO	\$ 8.182.836	\$ 8.086.935	\$ 14.128.195	\$ 95.901	\$ 6.041.260	98%
	NOT ICE CREAM	\$ 7.548.994	\$ 7.743.056	\$ 6.905.796	\$ 194.062	\$ 837.261	77%
	NOT CHICKEN	\$ 2.453.608	\$ 2.435.598	\$ 3.315.918	\$ 18.010	\$ 880.320	98%
	NOT CHEESE	\$ 3.388.243	\$ 3.314.084	\$ 3.422.366	\$ 74.159	\$ 108.282	32%
	NOT MILA	\$ 2.067.405	\$ 2.144.106	\$ 2.561.983	\$ 76.701	\$ 417.878	82%
	NOT MEAT	\$ 1.964.773	\$ 1.553.857	\$ 1.696.269	\$ 410.916	\$ 142.413	-189%
	NOT CREAM	\$ 1.329.176	\$ 1.523.328	\$ 1.703.042	\$ 194.152	\$ 179.714	-8%
	<b>TOTAL</b>	<b>\$ 58.298.246</b>	<b>\$ 56.544.714</b>	<b>\$ 75.433.054</b>	<b>\$ 2.683.363</b>	<b>\$ 20.562.863</b>	<b>87%</b>
	<b>GRAND TOTAL</b>	<b>\$ 884.172.328</b>	<b>\$ 911.295.425</b>	<b>\$ 1.215.400.553</b>	<b>\$ 113.674.203</b>	<b>\$ 353.070.501</b>	<b>68%</b>

Tabla 8: Tabla de resultados (Comparación proyecciones y RMSE) (Elaboración propia)

En la siguiente Figura se presentará un diagrama el cual sigue la misma lógica explicada en el ejemplo de la evaluación económica (Figura 11), pero actualizado con los valores reales de noviembre. Este diagrama será una herramienta visual clave para comprender cómo una reducción del 68% en el error de las proyecciones actuales de ventas podría impactar en la disminución del quiebre monetario de la empresa. Este análisis no solo mostrará la mejora en la precisión de las proyecciones, sino que también ilustrará de manera concreta el potencial de disminución de error monetario y la eficiencia operativa resultante de una planificación más ajustada a la realidad.



*Figura 12: Diagrama de posible mejora por implementación de nuevas proyecciones (Elaboración propia).*

Tras presentar el diagrama, es crucial contextualizarlo, reconociendo que las proyecciones actuales pueden estar infladas por efectos externos que se representan en la formación del replan. Sin embargo, a pesar de estos factores, la magnitud de las pérdidas monetarias que actualmente producen las proyecciones inexactas sugiere una necesidad urgente de mejorar este proceso. Utilizando las nuevas proyecciones como una guía mensual realista, la empresa podría reducir significativamente los costos de oportunidad monetarios. Como se discutió anteriormente, esto podría impactar directamente al error de planificación que explica un 50% del quiebre monetario total, lo cual es un cambio significativo y beneficioso para la empresa.

Por otro lado, la implementación del reporte sugerido en el plan de implementación extiende sus beneficios más allá de la mejora en la planificación de la producción. Este avance no solo aborda el problema principal del proyecto, que es generar proyecciones más precisas y realistas, sino que también ofrece un valioso soporte para otras áreas de la empresa. Facilitará la toma de decisiones informadas en diferentes departamentos, contribuyendo a una gestión más eficiente y alineada con los objetivos estratégicos de la organización.

Por otro lado, un logro notable del proyecto fue la optimización del tiempo para generar proyecciones. Anteriormente, este proceso requería dos semanas para considerar tendencias y estacionalidades. Con las mejoras implementadas, se logró reducir este período a solo una semana. Esta agilización ha permitido que la empresa utilice las nuevas proyecciones como base y agregue externalidades de manera más rápida y efectiva, facilitando una toma de decisiones más ágil y alineada con las dinámicas del mercado actual.

Por último, para mejorar el proyecto, se propone su expansión a NotCo Global, abarcando los seis países donde opera la empresa. Esto ofrecería una perspectiva más acertada de las ventas a nivel internacional. Además, sería beneficioso integrar factores externos al modelo SARIMAX para aumentar la precisión de las proyecciones, aunque esta tarea no se realizó debido a la complejidad y el tiempo necesario para adaptar el formato de estas externalidades.

En cuanto a los aspectos a mejorar, el principal desafío inicial fue determinar el enfoque correcto del proyecto, dado que identificar el problema principal resultó complicado. Una vez establecido el problema, una investigación más profunda y un mejor análisis exploratorio habría facilitado un desarrollo más eficiente del proyecto, evitando invertir tiempo en modelos que no eran adecuados para los objetivos planteados. La investigación exhaustiva es esencial para un desarrollo más ágil y efectivo de cualquier proyecto.

## Anexos

### Explicación Código:

#### Importación de Bibliotecas

`datetime` y `calendar`: Estas bibliotecas se usan para manejar fechas y tiempos. `datetime` permite trabajar con fechas y horas de manera eficiente, mientras que `calendar` es útil para obtener información relacionada con el calendario, como el número de días en un mes.

`numpy` y `pandas`: Son esenciales para el análisis de datos. `numpy` proporciona soporte para arreglos y matrices grandes y funciones matemáticas para trabajar con estos. `pandas` es crucial para la manipulación y limpieza de datos, ofreciendo estructuras de datos como `DataFrame`, que facilitan la manipulación de tablas de datos.

`optuna`: Una biblioteca de optimización de hiperparámetros, utilizada para automatizar la selección de los mejores parámetros para el modelo. Esta biblioteca es fundamental para mejorar el rendimiento del modelo SARIMAX.

`statsmodels`: Utilizada para la construcción y análisis de modelos estadísticos. Es clave para implementar el modelo SARIMAX, que es central en este análisis.

`sklearn.metrics`: Proporciona herramientas para calcular métricas de rendimiento del modelo, como el error cuadrático medio (MSE) y su raíz (RMSE), que son críticas para evaluar la precisión de las predicciones.

#### Extracción de Datos de la API de Bambu

El análisis comenzó con la importación de datos cruciales para comprender las ventas y su distribución. Estos datos fueron extraídos utilizando consultas específicas a una API proporcionada por la empresa Bambu, que ofrecía acceso a cuatro tablas fundamentales: Locales, Ean, Sell-Out y Stock.

**Tabla Locales:** Contenía información vital sobre los holdings, crucial para entender la distribución y el rendimiento de las ventas en diversas agrupaciones empresariales. La identificación de los holdings era esencial para realizar análisis segmentados y comprender las dinámicas de ventas en diferentes áreas de la empresa.

**Tabla Ean:** Proveía detalles importantes para identificar las marcas de los productos, permitiendo un análisis de ventas desagregado por producto y ofreciendo una visión más profunda de los productos que impulsaban las tendencias de ventas.

Tabla Sell-Out: Incluía datos sobre las ventas y las fechas correspondientes, siendo el núcleo del análisis de rendimiento de ventas.

Tabla Stock: Aunque estaba disponible, esta tabla no se utilizó en el análisis y, por tanto, no se extrajo.

Las consultas utilizadas en el proceso de importación permitieron extraer selectivamente los datos necesarios de la API, manipulándolos y convirtiéndolos en dataframes de pandas para análisis posteriores. Esta técnica aseguró una extracción eficiente y precisa de los datos relevantes lo cual podemos revisar en la Figura siguiente:

### Query Sell-out

```
1 import datetime
2 #Query Eans Bambu
3 QUERY = 'SELECT * FROM `notco-data-analytics.api_bambu.sell-out`'
4 query_job = client.query(QUERY) # API request
5 rows = query_job.result()
6 Big_Query_Sell_Out = rows.to_dataframe()
7 Big_Query_Sell_Out = pd.DataFrame(Big_Query_Sell_Out)
8 Big_Query_Sell_Out
9 #Query Eans Bambu
10 QUERY = 'SELECT * FROM `notco-data-analytics.api_bambu.Ean`'
11 query_job = client.query(QUERY) # API request
12 rows = query_job.result()
13 Maestra_eans = rows.to_dataframe()
14 Maestra_eans = pd.DataFrame(Maestra_eans)
15 #Pasamos el Upc a tipo objeto
16 Maestra_eans['Upc'] = Maestra_eans['Upc'].astype('object')
17
18 #DataFrame salas
19 api_salas = requests.get('https://thenot.bambub2b.com/api/locales.php', headers = headers)
20 df_salas = api_salas.json()
21 df_salas = pd.DataFrame(df_salas)
```

*Figura Anexo 1: Query tablas Sell-out/Ean/Locales*

## Proceso de Fusión de las Tablas

Una vez importadas, las tablas fueron combinadas para crear un conjunto de datos integral y diverso.

Unión de Locales, Ean y Sell-Out: La fusión de estas tablas se realizó utilizando operaciones de combinación en pandas, empleando identificadores comunes como los códigos de locales y productos (EAN). Esta fusión meticulosa integró múltiples aspectos del negocio en un único conjunto de datos.

Creación de un Dataframe Integral: El resultado fue un dataframe exhaustivo que amalgamó la información de los holdings, las marcas de los productos y las ventas. Este dataframe proporcionó una visión comprensiva de las ventas, enriquecida con contextos de ubicación y producto, y formó la base para análisis más profundos y contextualizados. Este proceso lo podemos ver en la siguiente Figura:

**Creacion Dataframe Sell\_Out**

```

1 #Creamos una copia del dataframe para no modificar el original y unirlo con la maestra de eans y la tabla de salas
2 df_sell_out = Big_Query_Sell_Out.copy()
3 #df_sell_out = df_sell_out.replace( 'SIN EAN', 0 )
4 df_sell_out['ean'] = df_sell_out['ean'].astype(float)
5 df_sell_out = pd.merge(df_sell_out, Maestra_eans, left_on='ean', right_on='Upc', how='left')
6
7 df_sell_out = pd.merge(df_sell_out, df_salas, left_on='cod_local', right_on='cod_local', how='left')
8 #Ahora eliminamos la columna CADENA
9 #df_sell_out = df_sell_out.drop(['CADENA'], axis=1)
10 #Y pondremos cadena como la primera columna del dataframe
11 df_sell_out = df_sell_out[['cadena_x', 'holding', 'Categoria', 'ean', 'local', 'cod_local', 'fecha', 'Marca_Notco', 'Descri_Item', 'ventas', 'unidades', 'costo']]
12 #Ahora ordenamos respecto a la cadena y fecha
13 df_sell_out = df_sell_out.sort_values(by=['cadena_x', 'fecha', 'Marca_Notco'])
14 #Ahora veamos la ventas por mes de cada holding y marca
15 df_sell_out['fecha'] = pd.to_datetime(df_sell_out['fecha'], errors='coerce')
16 #Pasamos las variables ventas y costo a numericas
17 df_sell_out['ventas'] = pd.to_numeric(df_sell_out['ventas'], errors='coerce')
18 df_sell_out['costo'] = pd.to_numeric(df_sell_out['costo'], errors='coerce')
19 df_sell_out['unidades'] = pd.to_numeric(df_sell_out['unidades'], errors='coerce')
20
21 #Variables a utilizar
22 df_sell_out = df_sell_out[['fecha', 'holding', 'Marca_Notco', 'ventas']]
23 exploratory_data_analysis(df_sell_out)
24 #Usaremos solo los valores del marzo en Adelante
25 # Get the current date
26 today = datetime.date.today()
27 # Calculate the first day of the current month
28 first_day_of_current_month = today.replace(day=1)
29 # Calculate the last day of the last month
30 last_day_of_last_month = first_day_of_current_month - datetime.timedelta(days=1)
31 df_sell_out = df_sell_out[(df_sell_out['fecha'] >= '2022-03-01') & (df_sell_out['fecha'] <= pd.to_datetime(last_day_of_last_month))] #
32 #Pasamos el FechaFactura a tipo str
33 df_sell_out['fecha'] = df_sell_out['fecha'].astype(str)
34 df_sell_out['Año'] = df_sell_out['fecha'].str.slice(0, 4)
35 df_sell_out['Mes'] = df_sell_out['fecha'].str.slice(5, 7)
36
37 #Eliminamos todos los valores de unidades = 0
38 df_sell_out = df_sell_out.sort_values(by=['Año', 'Mes'])
39 df_sell_out

```

*Figura Anexo 2: Creación Dataframe Ventas (Unión entre las 3 tablas principales)*

## Generación de Lista de Dataframes Desagregados

Desagregación por Holding y Marca: Para realizar un análisis más granular y obtener insights específicos, los datos se desagregan por holding y marca. Esto implica la creación de dataframes separados para cada holding y cada marca dentro del conjunto de datos. Este nivel de desagregación permite realizar predicciones y análisis más específicos, adaptados a segmentos particulares del negocio.

Estructuración de Datos para Análisis Específicos: La estructuración de estos dataframes desagregados es esencial para realizar análisis que consideren las características únicas de cada holding o marca. Esto podría incluir diferencias en tendencias de ventas, estacionalidad y otros factores específicos del segmento.

La siguiente Figura muestra el proceso de generación de la lista de dataframes:



```
1 lista_df={}
2 i=1
3 for holding in df_sell_out["holding"].unique():
4     print(holding)
5     for marca in df_sell_out["Marca_Notco"].unique():
6         print(f'    ', marca)
7         df=df_sell_out[(df_sell_out["holding"]==holding) & (df_sell_out["Marca_Notco"]==marca)]
8         df=df[['fecha', 'ventas']]
9         print(df)
10        lista_df[holding,marca]=df
11        i+=1
12
13
✓ 16.7s
```

*Figura Anexo 3: Generación Lista Dataframes*

### **Procesos y funciones principales (SARIMAX)**

**Preparación de los Datos:** Antes de construir el modelo, se realiza un preprocesamiento de los datos. Esto incluye la conversión de fechas al formato estándar de datetime y la reestructuración de los datos para su análisis.

**Definición de ELECTRONIC\_mensual:** Esta función es el núcleo del análisis. Incluye varios pasos clave:

**Configuración del Modelo SARIMAX:** Dentro de ELECTRONIC\_mensual, se define una subfunción sarimax\_trial para configurar el modelo SARIMAX con diferentes combinaciones de parámetros. Los parámetros q, d, p, y s son seleccionados y optimizados para modelar la estacionalidad y las tendencias autorregresivas en los datos de ventas.

**Conversión a Series Temporales Diarias:** Un paso crucial en el análisis de series temporales es la conversión de los datos en una forma que represente adecuadamente la temporalidad. En este caso, los datos se agrupan a nivel diario, proporcionando una base detallada para el análisis de tendencias y patrones a lo largo del tiempo.

**Optimización con optuna:** Se utiliza optuna para encontrar la mejor combinación de parámetros que minimice el RMSE, lo que implica ajustar repetidamente el modelo con diferentes parámetros y seleccionar el mejor conjunto.

**Normalización y Estandarización de las Predicciones:** Las predicciones generadas por el modelo SARIMAX se ajustan en relación con la desviación estándar y la media del conjunto de datos. Esta normalización es crucial para que las predicciones sean consistentes y comparables con los datos históricos.

Generación de Predicciones Futuras: Utilizando el modelo optimizado, se generan predicciones para el próximo mes. Estas predicciones son esenciales para la planificación empresarial y la toma de decisiones.

Evaluación del Modelo: Se calcula el RMSE de las predicciones para evaluar la precisión del modelo. Un RMSE más bajo indica una mayor precisión del modelo.

Este proceso lo podemos visualizar en las siguientes figuras (4 y 5):

```
def ELECTRONIC_mensual(df):
    optuna.logging.set_verbosity(optuna.logging.ERROR)
    warnings.filterwarnings("ignore", module="statsmodels")
    def sarimax_trial(trial):
        q = trial.suggest_int('q', 0, 4)
        d = trial.suggest_int('d', 0, 4)
        p = trial.suggest_int('p', 0, 4)
        s = int(trial.suggest_categorical('s', ['4', '12', '31']))
        trend = trial.suggest_categorical('trend', ['n', 'c', 't', 'ct'])
        model = sm.tsa.statespace.SARIMAX(data_train['Sales'], order=(q, d, p), seasonal_order=(q, d, p, s), trend=trend, enforce_stationarity=False)
        results = model.fit()
        data_train['forecast'] = results.predict(start=len(data_train)-number_of_days, end=len(data_train), dynamic=True)
        future_dates = [data_train.index[-1] + DateOffset(days=x) for x in range(0, number_of_days+1)]
        future_data_test_train = pd.DataFrame(index=future_dates[1:], columns=data_train.columns)
        future_data_train = pd.concat([data_train, future_data_test_train])
        future_data_train['forecast'] = results.predict(start=len(data_train), end=len(data_train) + number_of_days+1, dynamic=True)
        y_pred = pd.DataFrame(future_data_train['forecast'][len(future_data_train)-number_of_days:])
        y_pred.columns = ['Sales']

        rmse = sqrt(mean_squared_error(data_test, y_pred))
        return(rmse)

    prueb = df.copy()
    prueb['fecha'] = pd.to_datetime(prueb['fecha'])

    daily_df = prueb.resample('D', on='fecha').sum()
    # print(f'ceros:', daily_df['ventas'][daily_df['ventas']==0].count())
    daily_df.columns = ['Sales']
    daily_df_mean = daily_df.mean()
    daily_df_std = daily_df.std()
    daily_df = (daily_df - daily_df_mean) / daily_df_std
    daily_df = daily_df.replace(-np.inf, np.nan).dropna()

    # last_month_date = datetime.datetime.now() - pd.DateOffset(months=1)

    today = datetime.datetime.now().date()

    # Calculate the start date (1st of last month)
    start_date = (today - pd.DateOffset(months=1)).replace(day=1).date()

    # Generate the list of dates from the start_date to today
    # date_range = pd.date_range(start_date, today)
    number_of_days = (today - start_date).days + 1
```

*Figura Anexo 4: Código de implementación SARIMAX*

```
data_test = pd.DataFrame(daily_df['Sales'][:len(daily_df)-number_of_days])
data_train = pd.DataFrame(daily_df['Sales'][:len(daily_df)-number_of_days])

# study_sarimax = optuna.create_study(study_name="my_study_mes", storage="sqlite:///my_study_mes.db", direction="minimize", load_if_exists=True)
study_sarimax = optuna.create_study(direction="minimize")
try:
    study_sarimax.optimize(sarimax_trial, n_trials=20, catch=(IndexError, ValueError))
except Exception as e:
    print("Error during optimization:", e)
    return None, None

print(study_sarimax.best_params, study_sarimax.best_value)
q, d, p, s, trend = study_sarimax.best_params['q'], study_sarimax.best_params['d'], study_sarimax.best_params['p'], int(study_sarimax.best_params['s']), study_sarimax.best_params['trend']

sarimax = sm.tsa.statespace.SARIMAX(daily_df['Sales'], order=(q, d, p), seasonal_order=(q, d, p, s), trend=trend, enforce_stationarity=False)
results_sarimax = sarimax.fit()
# Get the current year and month
current_year = datetime.datetime.now().year
current_month = datetime.datetime.now().month

# Get the number of days in the current month
_, days_in_current_month = monthrange(current_year, current_month)

future_dates = [daily_df.index[-1] + DateOffset(days=x) for x in range(0, days_in_current_month+1)]
future_datest_daily_df = pd.DataFrame(index=future_dates[1:], columns=daily_df.columns)
future_daily_df = pd.concat([daily_df, future_datest_daily_df])

future_daily_df['forecast'] = results_sarimax.predict(start=len(daily_df), end=len(daily_df) + days_in_current_month+1, dynamic=True)
pred_prox_mes = pd.DataFrame(future_daily_df['forecast'][:len(future_daily_df)-days_in_current_month])
pred_prox_mes['forecast'] = pred_prox_mes['forecast'] * daily_df_std['Sales'] + daily_df_mean['Sales']
pred_prox_mes.columns = ['Sales']

mean_forecast = pred_prox_mes['Sales'].mean()
pred_prox_mes[pred_prox_mes <= 0] = mean_forecast

error_orig = study_sarimax.best_value * daily_df_std

return pred_prox_mes, error_orig['Sales']
```

*Figura Anexo 5: Código de implementación SARIMAX*

## Ejecución Función “ELECTRONIC\_mensual”

Tras la preparación y fusión de los datos, el proyecto avanzó a una etapa crítica: la ejecución de la función ELECTRONIC\_mensual. Esta función se centró en el modelado de series temporales y la generación de predicciones de ventas.

Preparación para la Ejecución: Se crearon dos diccionarios, df\_predicciones y diccionario de errores, para almacenar las predicciones y los errores asociados respectivamente. Estos diccionarios se inicializaron vacíos: {}.

Bucle sobre Dataframes: Se implementó un bucle que recorría lista\_df, una colección de dataframes. Cada dataframe en lista\_df representa un segmento de datos para el cual se necesitaba generar predicciones. En cada iteración del bucle, se llamó a ELECTRONIC\_mensual con el dataframe actual (lista\_df[df]) como argumento.

La función ELECTRONIC\_mensual procesó los datos y devolvió las predicciones de ventas junto con el error de predicción correspondiente.

Se implementó manejo de excepciones para capturar y manejar cualquier error durante la ejecución de la función. En caso de una excepción, se imprimía el error y se asignaban valores None a las predicciones y errores para ese dataframe específico.

Almacenamiento de Resultados: Los resultados de cada llamada a la función, incluyendo tanto las predicciones (res) como los errores (error), se almacenaron en los diccionarios `df_prediccionessss` y `dicssaionariderrores`, indexados por el nombre del dataframe.

Esto se puede visualizar en la siguiente Figura:

```
1 df_prediccionessss, dicssaionariderrores = {}, {}
2
3 for df in lista_df:
4     print(df)
5     try:
6         res, error = ELECTRONIC_mensual(lista_df[df])
7         print(res, error)
8         df_prediccionessss[df], dicssaionariderrores[df] = res, error
9     except Exception as e:
10        print('error:', e)
11        res, error = None, None
12        df_prediccionessss[df], dicssaionariderrores[df] = res, error
```

*Figura Anexo 6: Llamado ELECTRONIC\_MENSUAL (Bucle para recorrer la lista)*

### Exportación de las Predicciones a Excel

El paso final en el análisis fue la exportación de las predicciones generadas a un archivo Excel. Este proceso es vital para la utilización práctica de los resultados del análisis.

Preparación para la Exportación: Las predicciones obtenidas se organizaron en un formato adecuado para su exportación, asegurando que la información fuera clara y accesible para su uso en decisiones empresariales.

Exportación Utilizando pandas: Utilizando las capacidades de pandas, una herramienta versátil para el manejo de datos en Python, las predicciones se exportaron a un archivo Excel. Este formato es ampliamente utilizado en el ámbito empresarial, lo que facilita la interpretación y aplicación de los resultados a la plataforma Tableau,. Este proceso se puede ver en la siguiente Figura:

```
1 import pandas as pd
2
3 # Lista para almacenar las sumas totales de cada DataFrame, sus fechas y errores
4 rows = []
5
6 # Procesa df_predicciones para obtener las ventas
7 for key, df in df_predicciones.items():
8     if df is not None:
9         holding, marca = key # Asume que key es una tupla ('holding', 'marca')
10        error = diccionario_errores.get(key, None) # Obtiene el error del diccionario_errores si existe
11
12        for index, row in df.iterrows():
13            fecha = index.date().strftime('%d-%m-%Y') # Obtiene solo la fecha en formato 'dd-mm-yyyy'
14            sales = row['Sales']
15
16            rows.append({
17                'holding': holding,
18                'Marca_Notco': marca,
19                'Fecha': fecha,
20                'Sales': sales,
21                'Error': error # Asigna el error
22            })
23
24 # Combina todas las sumas totales y errores en un DataFrame
25 df_total = pd.DataFrame(rows)
26
27 # Exporta a Excel
28 with pd.ExcelWriter('prediccion_total_Noviembre_D2.xlsx') as writer:
29     df_total.to_excel(writer, index=False)
```

*Figura Anexo 7: Exportación predicciones a Excel*

## Bibliografía y Referencias

- Mohamad. (2016). SARIMAX – Definición de un Modelo SARIMAX. Numxl. [SARIMAX - Definición de un Modelo SARIMAX – Centro de ayuda \(numxl.com\)](https://numxl.com/sarimax/)
- Yvan Tijoux. (2022). Distribución Notco Mundial. Presentaciones The Not Company
- Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting: Principles and Practice* (3ra ed.). Monash University. [Forecasting: Principles and Practice \(3rd ed\) \(otexts.com\)](https://otexts.com/fpp3/)
- Ricardo Montezillos. (2023). Reporte Fill-Rate Logística
- Caterina Vargas. (2023). Reporte Forecasting SO/SI The Not Company
- IBM. ¿Qué son las redes neuronales? IBM. [¿Qué son las redes neuronales? | IBM](https://www.ibm.com/latam/think/insights/stories/que-son-las-redes-neuronales)
- Figura Anexo 1-7. Código Visual Studio. Elaboración Propia
- Figuras 1, 7 y 13. Presentaciones equipo Bi.
- Figuras 2,3,4,5,6,19, 21 y 22. Reportes equipo Bi.