

PRONOSTICAR CON EVIDENCIA

Herramientas para Proyección Económico
Financiera en el Ámbito Logístico
Práctica Profesional LAD - 2025

Presentado por:
Ferderico Pascal Facen
DNI: 19.087.311

mail de contacto: fpfacen@outlook.com



Índice

Resumen	3
Introducción.....	4
Situación Problemática.....	4
Preguntas de Investigación	5
Objetivo General	5
Objetivos Específicos	5
Marco Metodológico	5
Marco Teórico.....	6
Aplicación.....	7
Etapa 1 - Descubrimiento decisional.....	7
Etapa 2 - Inventario de Datos.....	9
Etapa 3 - Limpieza y estandarización del indicador.....	11
Etapa 4 – Plan de cuentas presupuestario y presentación de datos para modelado	12
Etapa 5 – Conceptualización, selección de modelos y desarrollo del código de proyección.....	14
Etapa 6. Caracterización del entorno logístico mediante indicadores públicos mensuales	16
Etapa 7. Definición del bot y decisión arquitectónica (GPT vs n8n)	18
Etapa 8 - Cálculo de afecciones matemáticas e integración Excel (bot + Python)	18
Etapa 9 - Lanzamiento, Medición de Valor y Adopción Estratégica	20
Recomendaciones.....	21
Conclusiones	21
Referencias	23
Apéndice	24
Apéndice I – Entrevistas preliminares a la gerencia.....	24
Apéndice II – Glosario de terminología de la empresa	29
Apéndice III – Descriptivo de Cuentas Presupuestarias	31
Apéndice IV – Código en Python para proyección cuantitativa.....	32
Apéndice V – Compendio explicativo del código	45
Apéndice VI – Instrucciones del bot GPT	48
Apéndice VII - Afecciones matemáticas concretas.....	51



Resumen

Las organizaciones actuales enfrentan el desafío de tomar decisiones en contextos cada vez más dinámicos, donde la disponibilidad y calidad de la información resultan determinantes para anticipar escenarios y sostener la eficiencia operativa. En este marco, la administración universitaria comparte las mismas tensiones que el ámbito privado: existen múltiples fuentes de datos, registros heterogéneos y la necesidad de integrar información contable, operativa y presupuestaria bajo criterios comunes que faciliten la planificación.

La falta de sincronía entre estos sistemas no solo dificulta la lectura del desempeño institucional, sino que también limita la capacidad de proyectar el comportamiento económico-financiero con base en evidencia. Frente a esta problemática, se plantea la necesidad de establecer un procedimiento sistemático que transforme los datos disponibles en insumos útiles para la toma de decisiones, combinando herramientas estadísticas y tecnológicas accesibles.

El presente trabajo desarrolla una metodología aplicada para la generación de proyecciones económico-financieras a partir de información interna y fuentes externas oficiales. El objetivo principal consiste en diseñar e implementar un proceso replicable que integre datos históricos, indicadores macroeconómicos y automatización analítica dentro de un marco de coherencia técnica y trazabilidad documental.

La investigación adopta un enfoque cuantitativo, de tipo aplicado y diseño no experimental longitudinal. Utiliza registros internos del período junio 2021–junio 2025, complementados con series externas del INDEC, BCRA, FADEEAC y AFIP. El modelado se realiza en Python, la integración en Excel y la actualización automática mediante un agente de inteligencia artificial.

Los resultados demuestran la factibilidad del sistema y la consistencia de las proyecciones generadas, sentando las bases para un proceso de planificación más integrado y sustentado en evidencia. Si bien no se evalúa el impacto financiero directo, la aplicación permite visibilizar la relevancia de la gestión del dato y del uso de herramientas analíticas como soporte para una administración universitaria más eficiente y proactiva.

Palabras Clave: proyección financiera, python, modelado de indicadores, chat bot, inteligencia artificial

Introducción

En el sector logístico, la diferencia entre el crecimiento y el rezago no reside únicamente en la disponibilidad de recursos, sino en la capacidad de interpretar a tiempo la información crítica para actuar. En la empresa objeto de estudio, que forma parte de un grupo corporativo regional, los datos que sustentan la toma de decisiones llegan con retraso, de manera fragmentada y sin integrar sistemáticamente las variables externas que inciden sobre la operación. Esta situación conduce a que las decisiones, aunque bien intencionadas, a menudo se desalineen de las condiciones reales del negocio y del mercado.

El contexto macroeconómico y sectorial acentúa esta vulnerabilidad: Argentina y sus países limítrofes experimentan escenarios de costos logísticos fluctuantes, cambios regulatorios acelerados, redefinición de corredores estratégicos como la Hidrovía Paraná–Paraguay, creciente interdependencia con puertos uruguayos y rutas bioceánicas chilenas, y variaciones en el comercio bilateral con Brasil. Sin un sistema de información capaz de captar y procesar estos cambios con rapidez, las decisiones se vuelven reactivas en lugar de anticipatorias.

Actualmente, la información financiera clave ,ingresos, egresos y resultados operativos, se procesa con un desfase de hasta un mes, lo que obliga a trabajar sobre estimaciones parciales y limita la capacidad de respuesta ante el directorio y las gerencias operativas. Esta falta de oportunidad y consistencia en los datos no solo afecta la planificación financiera, sino que también puede comprometer la competitividad, la asignación de recursos y la confianza de los accionistas.

La investigación plantea las bases de un sistema analítico que integra información interna y externa de manera oportuna, combinando el análisis de datos históricos mediante Python con un chatbot de inteligencia artificial parametrizado para captar indicadores de mercado. El objetivo no es únicamente proyectar flujos más precisos, sino construir un marco de información confiable y en tiempo real que permita a las gerencias tomar decisiones fundamentadas, anticiparse a riesgos y aprovechar oportunidades, aun reconociendo las limitaciones que impone el retraso en la disponibilidad de datos y la dependencia de variables externas.

Situación Problemática

La empresa presenta una carencia estructural en la gestión de su información económico-financiera: no dispone de un sistema integrado, oportuno y dinámico que articule de manera coherente los datos internos y externos relevantes para la toma de decisiones. Esta limitación provoca que las decisiones estratégicas se basen con frecuencia en información parcial o errónea, lo que dificulta anticipar tendencias y escenarios, e incrementa el riesgo de sobre o subejecución del gasto, con el consiguiente impacto en eficiencia, rentabilidad y credibilidad interna.

El problema no se desarrolla en un vacío, sino en un contexto particularmente desafiante. Internamente, la organización transita un proceso de mayor dependencia de los datos para la gestión, lo que aumenta la demanda de información precisa y disponible en

plazos reducidos. Externamente, la competencia en el sector logístico es cada vez más intensa, lo que impulsa a las empresas a optimizar procesos y decisiones para mantenerse vigentes. A esto se suma la reciente disminución en la rentabilidad de uno de los clientes más significativos, lo que incrementa la presión por gestionar de forma más eficiente los recursos y minimizar márgenes de error en la planificación.

En este escenario, la falta de un sistema robusto de información no solo limita la capacidad de respuesta inmediata, sino que también restringe la posibilidad de desarrollar estrategias proactivas frente a cambios internos y externos. La organización se expone así a un entorno en el que reaccionar tarde o con información deficiente puede significar la pérdida de oportunidades, la consolidación de desventajas competitivas y la erosión de su posición en el mercado.

Preguntas de Investigación

¿Qué características presenta hoy el flujo de datos económico-financieros de la empresa y cuáles son sus principales limitaciones para la toma de decisiones oportunas?

¿Qué variables internas y externas resultan más relevantes para proyectar la evolución económico-financiera y cómo se registran o deberían registrarse?

¿Cómo puede diseñarse e implementarse una metodología de cálculo de proyecciones económico-financieras que genere resultados consistentes y útiles para la gestión de la empresa?

Objetivo General

Diseñar e implementar una metodología sólida de cálculo de proyecciones económico-financieras que integre variables internas y externas, con el fin de fortalecer la base informativa para la toma de decisiones de la empresa.

Objetivos Específicos

Examinar el flujo actual de información económico-financiera de la empresa, identificando debilidades, retrasos y vacíos que afectan la toma de decisiones.

Identificar y clasificar las variables internas y externas críticas para la proyección económico-financiera, considerando su disponibilidad y relevancia.

Diseñar e implementar una metodología de cálculo de proyecciones que, a partir del análisis y de las variables identificadas, proporcione estimaciones consistentes y aplicables a la planificación de la organización.

Marco Metodológico

El estudio se desarrolla bajo un enfoque cuantitativo de tipo aplicado, orientado a diseñar e implementar una metodología práctica de proyección económico-financiera. Su alcance es descriptivo y propositivo, ya que analiza la estructura informativa existente en la

organización y propone un procedimiento reproducible para la generación de pronósticos, sin intervenir en la operatoria ni evaluar el impacto financiero directo de su implementación.

Adopta un diseño no experimental y longitudinal. No es experimental, porque las variables no se manipulan, sino que se observan en su contexto natural; y es longitudinal, porque trabaja con registros históricos de un período de tiempo extendido, comprendido entre junio de 2021 y junio de 2025.

Las fuentes de información que utiliza son principalmente secundarias. Emplea bases contables, registros operativos y planillas internas de la empresa, complementadas con datos externos provenientes de organismos oficiales como el INDEC, el Banco Central y la FADEEAC. A nivel cualitativo, realiza entrevistas semiestructuradas con las áreas de Gerencia General, Operaciones, Comercial y Calidad, para comprender los flujos de información y las necesidades de planificación de cada sector.

El procedimiento consiste en una secuencia progresiva de etapas: diagnóstico de los circuitos informativos, estandarización de los datos, elaboración del plan de cuentas presupuestario, desarrollo del modelo de proyección en Python, integración de variables externas y consolidación de resultados en un entorno de Excel.

Marco Teórico

La administración moderna se apoya cada vez más en la capacidad de transformar los datos en información útil para la toma de decisiones. Simon (1960) sostiene que la decisión racional se basa en la búsqueda y procesamiento de información relevante dentro de un entorno limitado de conocimiento, lo que implica diseñar sistemas que permiten captar, ordenar y analizar datos de manera estructurada. Desde esta perspectiva, la calidad y oportunidad de la información constituyen un requisito esencial para orientar la acción y reducir la incertidumbre organizacional.

La gestión de datos evoluciona desde un enfoque meramente operativo hacia una función estratégica vinculada con la planificación y el control. El modelo propuesto por la Data Management Association (DAMA, 2017) define la gestión de datos como un conjunto integrado de prácticas que garantiza la disponibilidad, exactitud, coherencia y trazabilidad de la información. Estos principios resultan especialmente relevantes en organizaciones que manejan grandes volúmenes de registros internos, donde la heterogeneidad de las fuentes y la falta de estandarización dificultan la lectura conjunta de los resultados.

En el plano analítico, la proyección económico-financiera se sustenta en el estudio de series temporales, entendido como el análisis sistemático de datos ordenados cronológicamente para identificar patrones y estimar su comportamiento futuro. Box y Jenkins (1970) establecen una metodología rigurosa para el modelado y validación de estos procesos, basada en el ajuste progresivo de componentes autorregresivos, de medias móviles y de integración. Holt (1957) y Winters (1960) amplían estos aportes incorporando la suavización exponencial, que permite capturar tendencias y estacionalidades de manera simple y eficiente. Estos métodos constituyen la base estadística sobre la cual se apoyan las proyecciones implementadas en la aplicación.



La literatura reciente señala que la efectividad de un modelo de pronóstico no depende únicamente del algoritmo empleado, sino también de la calidad de los datos y de la correcta contextualización del fenómeno analizado. Hyndman y Athanasopoulos (2018) destacan que toda proyección debe entenderse como una síntesis de información cuantitativa y conocimiento del entorno. Por ello, la incorporación de indicadores externos, como precios, actividad económica o costos logísticos, resulta necesaria para capturar las relaciones que condicionan la evolución de las variables internas.

El desarrollo de lenguajes de programación orientados al análisis de datos, como Python, permite automatizar las etapas de limpieza, modelado y validación, mejorando la reproducibilidad de los procedimientos. Al mismo tiempo, la irrupción de herramientas de inteligencia artificial amplía las posibilidades de actualización y monitoreo continuo. Marr (2018) señala que estas tecnologías no sustituyen el juicio profesional, sino que amplifican la capacidad analítica al automatizar tareas rutinarias y permitir concentrar el esfuerzo en la interpretación de resultados.

En conjunto, el marco teórico del presente trabajo se apoya en tres pilares conceptuales: la toma de decisiones basada en información confiable, el análisis estadístico de series temporales como método de proyección, y el uso de herramientas informáticas e inteligencia artificial como medios para mejorar la eficiencia y trazabilidad del proceso analítico.

Aplicación

Etapa 1 - Descubrimiento decisional

Para modelizar cómo decide la organización y cuáles son los insumos que efectivamente sostienen ese proceso, esta primera etapa integra, en un mismo relato analítico, los testimonios de las áreas clave y el contexto en que operan. Las entrevistas que se incluyen en el “Apéndice I – Entrevistas preliminares a la gerencia” describen un flujo informacional valioso pero tardío y fragmentado, con señales que viajan en paquetes y llegan con rezagos que encarecen la integración y disminuyen la capacidad de anticipar.

Un primer hilo conductor emerge en la voz de la conducción estratégica: la base para decidir se apoya en información contable central y planillas por área, útiles pero desfasadas. La Gerencia General sintetiza: la base de casi todas las decisiones son reportes contables enviados por la central y planillas internas elaboradas por cada área, y advierte que los cierres oficiales llegan con dos meses de atraso. La consecuencia práctica, reiterada en las entrevistas preliminares, es que parte de las decisiones se toman con supuestos que luego deben corregirse cuando arriba el cierre.

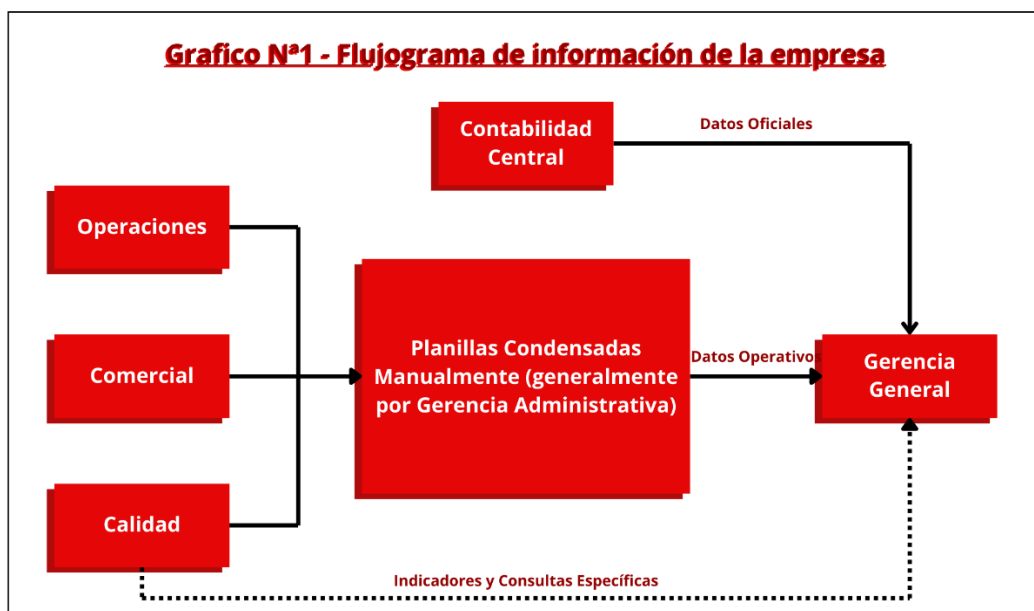
El segundo hilo proviene de la operación diaria: el pulso “de la calle” es inmediato, pero no conversa a tiempo con lo financiero. El área operativa explica con claridad que el principal problema es que la parte financiera no se integra con lo que ocurre en la operación diaria, y agrega una condición de diseño para aliviar ese cuello de botella: si la información estuviera disponible en tiempo real, se podrían ajustar los costos al instante. Este desfase explica por

qué decisiones de mayor alcance, como el dimensionamiento de flota, la apertura de rutas y la dotación de personal, se resuelven con menor certidumbre de la deseable.

El frente comercial introduce una sensibilidad específica: la concentración de ingresos. Se consigna que lo más importante es saber cómo se comportan las grandes cuentas, especialmente el cliente principal que representa más del 60% de la facturación. En ese marco, una proyección de costos subestimada erosiona márgenes y una sobreestimada hace perder negocios que podrían haber sido rentables. Por eso, desde el área comercial se afirma que la proyección es el insumo clave.

La función de calidad aporta el tercer ángulo: recibe datos de todas las áreas, pero denuncia la fricción de formatos y tiempos. De allí surge la demanda organizativa de contar con un sistema integrado que alinee datos operativos, comerciales y financieros en tiempo real, con el objetivo de pasar de la corrección ex post a la prevención ex ante y vincular, de manera explícita, calidad con costo económico y resultado.

Del entrecruzamiento de estas voces surgen tres implicancias concretas para ordenar la decisión: (i) un repositorio único y gobernado (operativo, comercial, financiero y de calidad) que reduzca los costos de integración; (ii) un módulo de proyección que combine series internas con señales externas relevantes y valide desempeño con métricas cronológicamente consistentes; y (iii) retroalimentación interáreas en ciclos cortos, de modo que cada decisión relevante compare alternativas sobre la base de márgenes esperados trazables. A continuación se adjunta el "Gráfico N° 1 – Flujograma de información de la empresa", que sintetiza el circuito vigente de información para la toma de decisiones.



Fuente: Elaboración Propia

A partir de esta evidencia, y sin presuponer resultados, se perfila, por acumulación de señales, el lugar central de la proyección económico-financiera como articulador entre lo operativo, lo comercial y la calidad. La propia conducción lo expresa como aspiración concreta:



"Me encantaría tener proyecciones económico-financieras en tiempo real, que combinen datos operativos, contables y del mercado. Algo que me permita ver cómo se está moviendo todo y simular escenarios para decidir sin tener que esperar dos meses". En la medida en que el repositorio, el módulo de proyección y la retroalimentación se institucionalicen, la organización pasará de "mirar el pasado" a anticipar escenarios con un lenguaje común de resultados esperados, mejorando comparabilidad, trazabilidad y asignación de recursos bajo criterios de rentabilidad y riesgo.

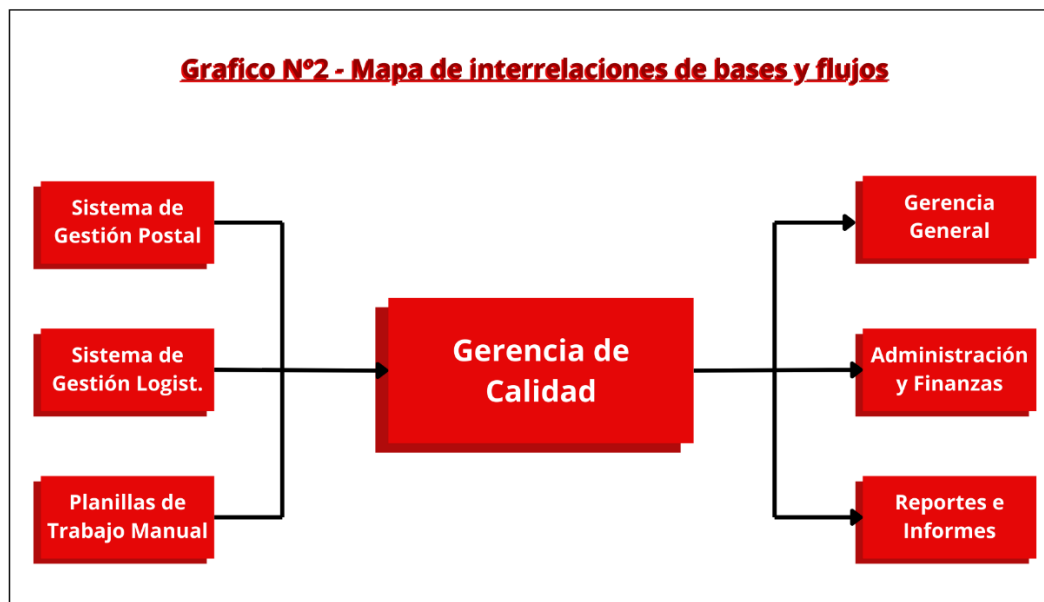
Etapa 2 - Inventario de Datos

El punto de partida resulta claro: la organización toma decisiones con información valiosa pero tardía, que llega en paquetes y exige correcciones cuando los cierres contables finalmente aparecen. Ese retraso no es un rumor, sino una realidad reconocida: los cierres oficiales llegan con dos meses de atraso, mientras la operación funciona al día y el área comercial necesita analizar tendencias sin esperar al calendario contable. Esta tensión, entre el pulso inmediato de la operación y la validación ex post, guía el relevamiento: observa con rigor qué bases existen, cómo se relacionan (o no) entre sí y en qué momento están realmente disponibles para decidir.

Con ese enfoque, el relevamiento identifica cuatro fuentes que, en conjunto, sostienen la toma de decisiones. Primero, el sistema de gestión postal genera diariamente salidas útiles para ajustar itinerarios, ritmos y excepciones operativas; su valor reside en la cadencia y en la proximidad con el hecho económico (la fecha en que sucede y no solo la fecha en que cierra). A su lado, el sistema de gestión logístico concentra la atención de la Gerencia de Operaciones: provee señales para dimensionar recursos, ordenar prioridades y corregir desvíos sin demoras. El tercer flujo lo componen las planillas de trabajo manual, presentes en cada área, que capturan detalles y resoluciones cotidianas que los sistemas no siempre contemplan; resultan útiles, pero cuando quedan aisladas generan dialectos propios que dificultan la integración posterior. Finalmente, la Gerencia de Calidad actúa como condensador: recibe, compara y traduce información de las demás áreas para que la Gerencia General pueda utilizarla sin perderse en el laberinto de formatos y tiempos. No se trata de inventar fuentes nuevas, sino de hacer dialogar lo que ya existe, en un idioma común y con relojes explícitos. Esa necesidad, alinear lo operativo, lo comercial y lo financiero en un circuito más oportuno, ya está instalada en la empresa.

El inventario muestra con claridad por qué esta conversación resulta prioritaria. La estructura de ingresos se encuentra concentrada: seguir de cerca el comportamiento de las grandes cuentas, en especial el cliente principal, no es una preferencia, sino una condición de gobernabilidad del margen. Si cada área relata la situación con nombres distintos, cortes temporales dispares o supuestos no compartidos, la organización pierde comparabilidad y, con ella, capacidad de anticipación. Por eso, además de relevar las bases, este trabajo registra cómo se nombran los conceptos y propone un lenguaje único para lo que la empresa realmente observa en su día a día.

De ese relevamiento surgen dos productos inmediatos y prácticos. El primero es un mapa de interrelaciones que hace visible el recorrido de la información: desde su origen en los sistemas postal y logístico, su paso por planillas que capturan matices locales y su consolidación en calidad, hasta la lectura gerencial que habilita decisiones oportunas. La imagen no pretende estilizar la realidad, sino ordenar sus trayectos: dónde nace cada dato, por qué caminos viaja y en qué punto está listo para usarse con responsabilidad.



Fuente: Elaboración Propia

El segundo producto es el “Apéndice II – Glosario de terminología de la empresa”, construido a partir de los papeles de trabajo y los registros que efectivamente se usan. Su razón de ser es simple: si el inventario dice qué datos hay y el mapa muestra cómo se relacionan, el glosario fija cómo se llaman y qué significan cuando la organización toma decisiones. Allí se normalizan términos que hoy aparecen con variantes ,por ejemplo, qué se entiende por “troncal”, qué constituye una “pieza multi-bulto” o cómo se contabiliza un “siniestro”, para que cada reporte hable el mismo idioma, sin perder el detalle que aporta cada área. Este esfuerzo no es cosmético: es la condición para que, a continuación, el repositorio gobernado, la proyección y la retroalimentación corta entre áreas se vuelvan prácticas estables y comparables en el tiempo.

El resultado, en términos organizativos, es tangible. Con las bases relevadas, las relaciones a la vista y la terminología unificada, la empresa puede empezar a correr escenarios sin esperar a que cierre el mes, y aun así reconciliar con la contabilidad cuando corresponda. La aspiración ya había sido formulada: proyecciones que combinen datos operativos, contables y de mercado para simular escenarios y decidir sin demoras innecesarias. Inventariar y trazar, acá, no es un fin en sí mismo; es despejar la pista para que lo que la empresa ya sabe ,y hace, llegue a tiempo a la mesa de decisión.

Etapa 3 - Limpieza y estandarización del indicador

Se prioriza la limpieza del indicador económico porque es la variable que alimenta directamente la proyección y, por lo tanto, la que más condiciona la calidad de los resultados. Es, además, un insumo que la organización ya utiliza para decidir ,aunque llegue con un rezago aproximado de dos meses, y sobre el cual existen prácticas consolidadas de lectura y validación. Ordenarlo primero reduce ambigüedades de interpretación, evita que el modelado absorba ruidos que no provienen del fenómeno sino del registro, y asegura que todos los pasos posteriores se apoyen en una base estable y comparable.

En junio de 2025 el área responsable adoptó un nuevo formato del indicador (estructura de columnas, nombres y codificaciones). Por ello, es necesario traducir el histórico completo (jun/2021–jun/2025) al formato vigente para que el pasado y el presente hablen el mismo idioma. La conversión documenta correspondencias entre campos, armoniza unidades y signos y verifica que los totales preserven su relación con los componentes, de modo tal que cualquier valor observado antes de la migración sea comparable con los posteriores.

Captura Nº 1 , Base histórica con formato

INDICADORES					
Ventas			mar-25	abr-25	may-25
Generales		####	#####	\$ 509.018.242,56	\$ 430.384.001,61
Semestral					
Yell					
Gestiva					
Banco del Tucumán		,50	-\$ 52.037.151,90	-\$ 98.514.827,02	-\$ 84.281.878,79
Banco del Tucumán Adicional Semestral					
Ventas camion y otros					
Paquetería					
Recupero Decreto		,33	\$ 4.872.855,57	\$ 5.176.510,62	\$ 5.363.674,57
Ingresos Varios					
TOTAL INGRESOS		####	#####	\$ 415.679.926,16	\$ 351.465.797,39
Egresos OPERACIONES			mar-25	abr-25	may-25
Andreani	Var				
Arbona y Pavone y Mastermoto	Var				
Caja operativa Cordoba	Var	,00			
Caja operativa Catamarca	Var	,00	\$ 189.220,00	\$ 1.530.435,05	\$ 862.079,56
Caja operativa Corrientes/bs as	Var	-			
Caja operativa La Rioja	Var	,00	\$ 634.414,80	\$ 905.309,90	\$ 398.385,47
Caja operativa Tucumán	Var	,29	\$ 11.074.238,75	\$ 12.484.668,49	\$ 29.651.165,74

Fuente: Base de datos de la empresa (los datos están editados para preservar la privacidad de la empresa).

Captura Nº 2 , Base del indicador con formato nuevo (jun/2025)

	ene-25	feb-25	mar-25	abr-25	may-25	jun-25
Ventas	413.109.969,06	381.310.417,54	286.063.790,12	415.679.926,16	351.465.797,39	378.361.210,67
Ventas Ctas.Ctes.	492.952.443,96	470.823.135,71	333.228.086,45	509.018.242,56	430.384.001,61	462.828.234,40
Ventas Contado						321.464,59
Dcto. 814 - CF IVA	4.990.203,10	3.396.981,33	4.872.855,57	5.176.510,62	5.363.674,57	8.603.319,71
Ingresos Varios						600.000,00
Adicionales BT	-84.832.678,00	-92.909.699,50	-52.037.151,90	-98.514.827,02	-84.281.878,79	-93.991.808,03
Costos variables	95.249.842,04	89.406.506,46	97.075.027,90	97.817.630,96	116.587.139,92	107.965.603,33
Alquileres vehículo						0,00
Alquileres vehiculos	12.312.495,54	8.948.040,99	14.531.660,00	13.090.827,60	12.956.050,00	12.768.050,00
Bolsas transparentes Jujuy	63.284,26	61.829,39	43.761,01	44.653,42	97.881,35	57.578,25
Bolsas Transparentes	11.971.551,62	11.696.332,35	8.278.317,68	8.447.135,38	18.516.320,30	10.892.139,45
Bolsinas y Sacas	119.801,91	117.047,73	82.842,92	84.532,31	185.296,83	109.000,00
CERTIFICACIONES	21.982,00	21.476,65	15.200,54	15.510,52	33.999,42	20.000,00
Com. y gtos bancarios	0,00	6.300.519,11	748.927,20	814.210,51	840.036,82	38.000,00
Comb. y lub. movil Salta						97.752,19
Comb. y lub.vehiculo Santiago	601.831,35	152.560,43	60.723,93	0,00	82.029,42	158.833,16
Combustibles y Lub.Operaciones						0,00
Combustibles y Lubricantes						664.541,73
COMIS. Y GTOS. BANCARIOS	4.388.650,70	1.435.302,67	4.174.132,09	3.873.156,82	4.476.580,92	3.943.845,23
Comisioens y Gtos Tarjeta Cred.	69.460,19	67.863,34	48.031,66	49.011,16	107.433,62	63.197,33

Fuente: Base de datos post limpieza (los datos están editados para preservar la privacidad de la empresa).

Una vez resuelta la traducción, se aborda la limpieza propiamente dicha. Se identificaron y eliminaron registros duplicados que inflan totales o repiten observaciones; se tratan faltantes con criterios conservadores ,completando sólo cuando la lógica de negocio lo permite y dejando explícito el vacío cuando la imputación podía inducir sesgo,; y se unificaron campos fragmentados o redundantes surgidos de la convivencia de sistemas y planillas (por ejemplo, desambiguando alias y fijando un único nombre para cada atributo relevante). Este trabajo no altera el sentido económico del indicador: lo hace internamente consistente, inteligible y utilizable sin correcciones ad hoc.

El resultado es una base estandarizada y depurada del indicador económico, con tipos de dato consistentes, claves de identificación claras y trazabilidad de las transformaciones aplicadas.

Etapas 4 – Plan de cuentas presupuestario y presentación de datos para modelado

Con la base del indicador ya depurada y traducida al formato vigente, el siguiente paso consiste en ordenar el lenguaje económico con el que la organización describe sus movimientos al proyectar. Se analiza la naturaleza de cada cuenta que interviene en el indicador ,su sentido económico, su relación con los drivers operativos y su comportamiento esperado en el tiempo, para distinguir con claridad qué es ingreso, qué es costo directo, qué es gasto de soporte y qué es efecto financiero o impositivo. Este análisis no se limita a renombrar: alinea criterios entre áreas y reconcilia, cuando corresponde, la forma en que la operación nombra los hechos con la forma en que la contabilidad los registra, manteniendo la coherencia con el glosario normalizado y con los relevamientos previos. De este modo, la arquitectura contable que alimenta la proyección habla el mismo idioma que los flujos operativos y comerciales identificados en las etapas anteriores.

El armado efectivo del plan de cuentas presupuestario ,documentado como Apéndice III y principal salida de esta etapa, cristaliza ese análisis en un listado codificado y descrito, listo para ser utilizado como marco de imputación durante la proyección. Allí se establecen rúbricas diferenciadas por negocio (postal y logístico) y se fijan descripciones operativas que conectan cada cuenta con su función económica: por ejemplo, VTAS como concentrador de ventas, DCTO como recupero del Decreto 814 (IVA), ALQC/ALCM para alquileres de flota en sus distintas modalidades, SLDS/CGSC/SAC para nómina y cargas asociadas, CBLC para combustibles, IMPT para tributos, ITFN para costos financieros y GTVR para egresos de soporte, entre otras. El valor del Apéndice no es solo taxonómico: define un perímetro estable de imputación que facilita el control, la comparabilidad intermensual y el análisis de márgenes por línea de negocio.

Sobre esa base, se presenta el dataset maestro para modelado: una matriz rectangular en formato “amigable para Python”, con una columna por cada cuenta del plan (según el Apéndice III) y una fila por cada mes del período de análisis. La columna de período utiliza un identificador cronológico único y se respetan convenciones explícitas de signo y unidades para preservar la trazabilidad con el indicador depurado. Cuando una cuenta no presenta movimientos en un mes, el valor se explicita como cero si corresponde a ausencia efectiva de egreso/ingreso, o se deja nulo cuando el dato no existe y no debe imputarse; esta distinción evita que el proceso de modelado confunda faltantes con verdaderos ceros y reduce sesgos de estimación.

Captura N.º 3 – Esquema de dataset por cuentas y meses

	VTAS	DCTO	ALQC	ALCM
jul-21	\$ 22.385.642,22	\$ 24.135,90	\$ 11.084.255,48	\$ 2.865.197,43
ago-21	\$ 24.585.045,50	\$ 24.249,35	\$ 11.297.330,65	\$ 3.596.456,82
sep-21	\$ 25.005.503,18	\$ 28.113,51	\$ 11.429.762,25	\$ 3.712.500,19
oct-21	\$ 26.461.268,96	\$ 29.546,98	\$ 12.224.711,33	\$ 4.283.263,46
nov-21	\$ 33.997.639,80	\$ 34.065,95	\$ 16.148.050,01	\$ 4.171.405,03
dic-21	\$ 36.932.307,51	\$ 47.908,66	\$ 16.999.670,94	\$ 4.897.777,72
ene-22	\$ 37.329.653,31	\$ 42.864,60	\$ 18.051.772,17	\$ 4.577.833,63
feb-22	\$ 35.458.284,52	\$ 44.093,60	\$ 16.383.930,06	\$ 5.159.907,49
mar-22	\$ 42.482.328,60	\$ 46.702,67	\$ 18.315.256,86	\$ 6.585.341,35
abr-22	\$ 44.609.899,04	\$ 47.926,64	\$ 20.683.765,25	\$ 5.998.003,02
may-22	\$ 46.218.219,65	\$ 61.389,85	\$ 21.339.848,69	\$ 5.997.532,78
jun-22	\$ 63.614.784,61	\$ 90.753,24	\$ 27.128.216,24	\$ 7.247.350,03
jul-22	\$ 70.396.801,62	\$ 57.173,58	\$ 33.808.845,77	\$ 9.399.044,91
ago-22	\$ 82.260.928,91	\$ 57.167,92	\$ 38.523.073,20	\$ 9.267.504,73
sep-22	\$ 79.552.998,31	\$ 71.820,21	\$ 36.832.862,18	\$ 10.365.204,28
oct-22	\$ 77.576.600,56	\$ 73.718,25	\$ 36.869.481,30	\$ 9.910.252,09
nov-22	\$ 93.559.881,38	\$ 79.498,82	\$ 41.525.985,44	\$ 9.470.364,81
dic-22	\$ 100.277.155,49	\$ 113.598,99	\$ 49.249.514,19	\$ 10.057.577,41
ene-23	\$ 96.282.702,25	\$ 81.303,90	\$ 46.351.001,10	\$ 9.175.636,50

Fuente: Planilla de traducción automática a modelo presupuetario

Para operativizar esta conversión y asegurar consistencia mensual, se desarrolla una planilla en Excel que permite pegar, mes a mes, el indicador económico en su formato original

y lo traduce automáticamente al nuevo esquema requerido para el modelado (mapea nombres y codificaciones al plan de cuentas del Apéndice III, armoniza unidades y signos y valida totales contra componentes). Esta herramienta se apoya en la traducción de formatos realizada previamente por el área responsable y consolida un canal simple, trazable y repetible entre el registro histórico y el módulo de proyección cuantitativa.

En síntesis, esta etapa amarra tres decisiones de diseño: entender económicamente cada cuenta del indicador, institucionalizar ese entendimiento en el Apéndice III como marco de imputación y, por último, proveer una presentación única de datos ,automatizada a través de la planilla de Excel, que conecte sin fricción el registro histórico con el modelado en Python. Con ello, la organización pasa a contar con un lenguaje contable operativo y computacionalmente consistente, capaz de sostener escenarios y comparaciones en ciclos cortos, sin perder la trazabilidad hacia las fuentes y definiciones ya consensuadas.

Etapas 5 – Conceptualización, selección de modelos y desarrollo del código de proyección

Con el plan de cuentas presupuestario institucionalizado y la matriz mensual “amigable para Python” disponible, esta etapa articula tres decisiones: declara el marco conceptual que guía la proyección, justifica la cartera de modelos estadísticos a emplear y consolida en un código único el procedimiento que transforma esa base estandarizada en proyecciones operativas para la dirección. La continuidad con lo realizado no es solo metodológica: el código se apoya en el dataset construido en la etapa anterior y respeta su granularidad por negocio (postal y logístico), su periodicidad mensual y su lenguaje común de cuentas. La trazabilidad con el recorrido previo ,del relevamiento a la estandarización, se preserva como criterio de diseño para que la herramienta cuantitativa no invente un lenguaje nuevo, sino que potencie el ya consensuado.

El funcionamiento del módulo es simple de describir y cuidadoso en su ejecución. Parte de un archivo Excel con dos hojas, SPP (negocio postal) y WEB (negocio logístico), donde la primera columna codifica el mes y cada columna subsiguiente corresponde a una cuenta del plan (idéntico a como se muestra en Captura N°3). Para cada hoja y para cada cuenta, el código forma la serie temporal mensual, elimina únicamente las filas sin mes (no hay interpretación posible) y conserva los ceros válidos como dato económico. Sobre esa serie corre un proceso de evaluación con validación temporal de origen rodante, consistente con el objetivo de pronosticar el trimestre siguiente: entrena el modelo en la historia disponible, lo prueba sobre ventanas futuras de uno a tres meses y resume el desempeño con MASE como métrica principal y sMAPE como medida porcentual de desempate. Con ese ranking, reentrena el modelo ganador en toda la historia y calcula la proyección de tres meses, guardando dos salidas: el archivo con las proyecciones por hoja y un informe por cuenta con el nombre del modelo elegido. La lógica es la de una “cinta transportadora” reproducible: mismas entradas, mismo criterio de elección, mismas salidas, sin pasos manuales intermedios.

La cartera de modelos elegida equilibra tres necesidades que emergen del comportamiento de las cuentas: cuenta con un piso robusto y rápido, captura tendencia y



estacionalidad cuando existen, y da respuesta a patrones difíciles como la intermitencia o las no linealidades suaves. Por eso se descartan expresamente los dos naïve puros como candidatos (se preserva un promedio móvil corto como red de seguridad mínima) y se concentra el esfuerzo en nueve familias que cubren los escenarios observados en la práctica. Esta restricción no es caprichosa: acorta tiempos de cómputo, evita empates triviales con métodos demasiado simples y mejora la estabilidad de la selección cuando la historia es limitada.

El primer grupo asegura ese piso robusto con `moving_average`, que suaviza el último tramo y proyecta su nivel. Se mantiene como única salvaguarda minimalista por su estabilidad, su costo computacional nulo y porque evita que una cuenta cortísima o extremadamente ruidosa fuerce a modelos complejos a sobre-ajustar; en la práctica, actúa como red de contención si la validación no puede discriminar ganadores con confianza. El segundo grupo captura estructura temporal clásica. ETS (Holt-Winters) combina nivel, tendencia amortiguada y estacionalidad mensual con un conjunto acotado de configuraciones elegidas por información (AIC), lo que lo vuelve eficaz cuando la estacionalidad es estable y la tendencia suave. SARIMA incorpora autocorrelación regular y estacional con una grilla parsimoniosa elegida por BIC y límites de iteración, ofreciendo un “caballo de batalla” cuando hay dependencia temporal clara sin necesidad de exógenas. STL+ARIMA separa antes de modelar: descompone la estacionalidad con STL y ajusta los residuos con ARIMA, recombinando luego ambas piezas; resulta especialmente útil cuando la estacionalidad deriva levemente en el tiempo y el ajuste monolítico de ETS o SARIMA pierde pulso. Theta aporta un estimador competitivo en presencia de tendencias suaves y, cuando la implementación clásica no está disponible, su versión de “drift” lineal preserva el sentido de extrapolación sin forzar complejidades innecesarias.

El tercer grupo atiende estacionalidades suaves y patrones armónicos. DHR con Fourier en SARIMAX modela la estacionalidad a través de senos y cosenos con pocos armónicos y deja que el componente determinista capture la periodicidad, eligiendo cuántos armónicos usar por criterio de información; esta representación es muy estable cuando la estacionalidad es regular pero no perfectamente multiplicativa o aditiva. El cuarto grupo cubre un patrón que en finanzas operativas aparece con frecuencia: la intermitencia. Para esas cuentas, la familia Croston (Croston, SBA y TSB) estima de manera separada tamaño y frecuencia de ocurrencia y utiliza una mini-validación para elegir la variante más adecuada; evita que los ceros “aplanen” la señal y entrega expectativas consistentes cuando los eventos son esporádicos. Finalmente, el quinto grupo habilita relaciones más flexibles con aprendizaje supervisado sobre la propia serie. Elastic Net utiliza lags y armónicos de Fourier con regularización mixta L1/L2, lo que estabiliza el ajuste con pocas observaciones y previene sobre-ajuste en presencia de colinealidad entre lags. Gradient Boosting Regressor trabaja con la misma ingeniería de variables, pero permite captar no linealidades suaves y pequeñas interacciones; se utiliza con hiperparámetros prudentes para mantener la generalización y, como en todo el pipeline, su elección está subordinada a la validación temporal y a las métricas definidas.

Bajo este diseño, el código no supone de antemano qué modelo conviene a cada cuenta: deja que la validación en origen rodante y el criterio de selección decidan con evidencia. Esa evidencia es cronológica y comparable, de modo que el informe por cuenta indica con transparencia qué método resulta elegido. La combinación de un catálogo acotado pero expresivo, una validación coherente con el horizonte de tres meses y un mecanismo de desempate porcentual produce proyecciones que pueden ser auditadas, replicadas y, sobre todo, interpretadas en términos económicos.

En términos operativos, el módulo también incorpora salvaguardas prácticas: si la serie es demasiado corta, limita la cartera a métodos adecuados para series de estas características; si un ajuste no converge, lo descarta sin interrumpir el proceso; si una cuenta presenta largos tramos en cero, deriva su evaluación hacia la familia intermitente. Ninguna de estas decisiones altera el gobierno de datos acordado: las cuentas, sus signos, unidades y la separación por negocios permanecen idénticos a los definidos en el plan y en la planilla de traducción mensual. El resultado son dos artefactos que se integran naturalmente al flujo de gestión: el archivo de proyecciones por hoja (SPP y WEB) que alimenta análisis de margen, planificación y cash-flow de corto plazo, y el reporte metodológico que documenta para cada cuenta el modelo ganador y sus métricas, facilitando revisión gerencial y mejora continua.

Como corolario, esta etapa deja documentado el código fuente en el Apéndice IV – Código en Python para proyección cuantitativa, que consigna la implementación detallada de lectura, validación, evaluación de modelos, selección y exportación de resultados, y acompaña un Apéndice V – Compendio explicativo del código que traduce esa implementación a un lenguaje no técnico para su lectura por parte de responsables operativos y gerenciales. De esta forma, la “pata cuantitativa” de la proyección no es una caja negra: es un procedimiento explícito, controlable y replicable que conecta, sin pérdida de significado, el registro económico de la empresa con las decisiones que dependen de anticipar sus movimientos en el trimestre siguiente.

Etapas 6. Caracterización del entorno logístico mediante indicadores públicos mensuales

Esta etapa define el conjunto de doce indicadores externos con los que se caracteriza el entorno de negocio que condiciona la operación y los resultados logísticos. El criterio de selección es práctico y riguroso: (i) mantiene relevancia directa sobre ingresos, costos y capital de trabajo; (ii) cubre el entorno en sus capas macro, financiera y sectorial; (iii) asegura frecuencia y trazabilidad mediante series mensuales, fuentes públicas y auditables; y (iv) prioriza la complementariedad, evitando duplicar señales o introducir colinealidad que no aporta poder explicativo. El objetivo no consiste en reunir la mayor cantidad de datos, sino en disponer de un tablero mínimo suficiente, reproducible y confiable, que se integra luego al modelo cuantitativo y al circuito de decisiones.

En el bloque de precios sistémicos, el IPC INDEC (variación mensual) captura la presión inflacionaria que repercute en indexaciones, revisiones tarifarias y expectativas salariales. Lo complementa el IPIM INDEC (variación mensual), que refleja la dinámica de precios mayoristas



y de insumos intermedios; suele anticipar movimientos del consumo y “empuja” rubros claves del costo, como combustibles, repuestos y embalajes. La pareja IPC+IPIM brinda una lectura más completa del pass-through inflacionario que afecta tanto el costo por kilómetro como la capacidad de trasladar precios a clientes.

El bloque financiero-cambiario condiciona el capital de trabajo y la estructura de contratación. El tipo de cambio oficial BCRA A3500 (promedio mensual) actúa como ancla de todos los componentes dolarizados (neumáticos, repuestos, unidades, parte del combustible) y de contratos referenciados a USD; su seguimiento reduce errores al presupuestar ítems con sensibilidad cambiaria. La tasa de política monetaria BCRA (TNA) resume la postura de la autoridad y funciona como piso del costo del dinero; la BADLAR bancos privados (promedio mensual, TNA) aproxima el “precio de mercado” efectivo que enfrentan las empresas para líneas de corto plazo. Mantener ambas tasas evita depender de un único proxy: la primera marca la señal regulatoria, la segunda muestra cómo se “aterrija” en el sistema financiero.

Para medir actividad real, se combinan dos termómetros complementarios. El EMAE INDEC (variación mensual desestacionalizada) resume la evolución del nivel de actividad de toda la economía y aporta contexto para escenarios de demanda general. El IPI INDEC manufacturero (variación mensual desestacionalizada) se concentra en la producción industrial, típicamente correlacionada con tramos troncales y B2B (palets, cargas de línea). Esta dupla permite detectar desacoples entre industria y agregado macro, afinando la lectura de volúmenes esperables por corredor y tipo de cliente.

En la demanda de consumo, las ventas en supermercados INDEC (variación mensual desestacionalizada) funcionan como proxy oficial, homogéneo y frecuente del movimiento físico de bienes de canasta, con impacto sobre la última milla y la distribución urbana. Aunque no representa “e-commerce puro”, su oportunidad y cobertura la convierten en una señal de excelente calidad cuando no existe una estadística mensual pública y estable de paquetes. Para anticipar cambios de propensión a comprar antes de que se materialicen en volúmenes, se incorpora el Índice de Confianza del Consumidor (UTDT): sus variaciones mensuales aportan una dimensión de expectativas que ayuda a calibrar capacidad, dotaciones y tercerización con mayor anticipación.

El bloque sectorial logístico requiere un agregado que traduce shocks de costos a lenguaje operativo. El ICL FADEEAC (variación mensual) condensa combustible, personal, neumáticos, reparaciones, peajes y otros rubros con metodología conocida y publicación regular. Se desagrega explícitamente el componente más volátil y material del costo variable por kilómetro con el precio promedio nacional de gasoil (ARS/L, mensual): esto permite distinguir si un movimiento del ICL proviene del combustible u otros rubros (por ejemplo, salarios), y ajustar con precisión por tipo de ruta (troncal vs. urbano).

Finalmente, se incorpora un ancla nominal-fiscal: la recaudación tributaria AFIP (variación interanual). Aunque mezcla precio y cantidad, su lectura YoY reduce estacionalidad y entrega un pulso macro rápido y amplio de la economía formal. Resulta útil como control cruzado: si la confianza y la actividad sugieren un escenario y la caja del fisco señala otro, conviene revisar supuestos antes de ajustar tarifas o comprometer capacidad.



Etapa 7. Definición del bot y decisión arquitectónica (GPT vs n8n)

En esta etapa se resuelve la disyuntiva tecnológica para la captura mensual de indicadores externos definidos en la etapa anterior. Se evalúan dos caminos: por un lado, un flujo orquestado en n8n, adecuado para fuentes con estructura estable y APIs consolidadas; por otro, un bot conversacional basado en GPT con navegación web, orientado a explorar, interpretar y normalizar evidencia publicada en sitios oficiales y entidades de referencia. Dado que el tablero de indicadores prioriza fuentes públicas con formatos heterogéneos (informes PDF, notas técnicas, tableros web, gacetillas) y experimenta cambios frecuentes en enlaces y presentación, se adopta GPT como carril principal para obtener, estandarizar y devolver los valores. n8n queda como capa complementaria para tareas de automatización cuando, y solo cuando, una fuente demuestra estabilidad suficiente para ser “productizada” sin costo de mantenimiento alto.

La elección por GPT se fundamenta en tres ventajas prácticas: ofrece flexibilidad semántica para localizar y desambiguar datos en múltiples formatos; reduce el tiempo de puesta en marcha frente a cambios de publicación; y brinda cobertura integral cuando no existen endpoints o estos resultan incompletos para el alcance del análisis. Los riesgos típicos del enfoque, como la variabilidad y ambigüedad, se mitigan por diseño: se establecen reglas estrictas de salida, jerarquía de fuentes, verificación mínima por celda (dato, fuente, URL y fecha) y un control de coherencia antes de impactar el modelo. En síntesis, GPT maximiza el time-to-insight con la gobernanza mínima necesaria; n8n se reserva para “congelar” circuitos estables más adelante sin reescribir la lógica del modelo.

Operativamente, el bot GPT funciona como un investigador estructurado. Con un único mensaje del usuario (por ejemplo, “Actualizar AAAA-MM , Argentina”), el bot navega la web, prioriza organismos oficiales y entidades reconocidas, extrae y normaliza los valores de los 12 indicadores y entrega una tabla única con cuatro columnas: Indicador, t0, t-1 y t-2 (donde t0 es el mes objetivo y t-1 y t-2, los dos meses previos). La normalización asegura comparabilidad: todas las tasas y variaciones se expresan en 0–100 (sin “%”), los niveles monetarios en ARS (sin símbolo, dos decimales, con conversión explícita cuando corresponde) y los índices respetan la base declarada. Debajo de la tabla, el bot agrega un bloque “Fuentes consultadas” con URL y fecha de referencia por indicador y mes, dejando traza suficiente para auditoría y reprocesos.

La salida formal de esta etapa es el Apéndice VI – Instrucciones del bot GPT, donde se detalla la parametrización completa del asistente: propósito, jerarquía de fuentes, consultas guía (site:), definiciones operativas por indicador, reglas de normalización y formato de salida, criterios de calidad y manejo de faltantes.

Etapa 8 - Cálculo de afecciones matemáticas e integración Excel (bot + Python)

En esta etapa se define cómo los doce indicadores externos seleccionados en la etapa anterior afectan cuantitativamente las cuentas del modelo y cómo se integran, en un único libro de Excel, las dos fuentes de insumos: la salida del pipeline de Python (base presupuestaria por periodo–cuenta–factor) y la salida del bot GPT (indicadores t0, t-1, t-2). El objetivo consiste



en transformar señales del entorno en ajustes reproducibles sobre cada cuenta, con trazabilidad de supuestos y control de calidad antes de impactar resultados.

El cálculo de afecciones se formaliza en el Apéndice VII – Afecciones matemáticas concretas, donde se establecen, para cada par indicador–cuenta, el modo de impacto (multiplicativo cuando el indicador es una variación %, aditivo cuando es un nivel como ARS/L de gasoil o ARS/USD del A3500), el coeficiente β (elasticidad o peso) y el rezago temporal (lag) cuando corresponde. En términos operativos, las cuentas sensibles a variables de precio o tasa (por ejemplo, combustibles, tercerizaciones, servicios y gastos financieros) se ajustan con una estructura multiplicativa del tipo $Cuentat = Baset \cdot \prod_k (1 + \beta_k \cdot Indk/100)$, mientras que las cuentas que dependen de un nivel monetario se recalculan directamente a partir de la identidad económica más simple (por ejemplo, Combustible = Litros \times PrecioGasoil; Intereses = Saldo \times Tasa mensual). Los β propuestos constituyen una versión inicial de trabajo sobre la que se realiza calibración empírica (backtesting) sin modificar la arquitectura del modelo.

Para garantizar consistencia, el apéndice fija reglas de aplicación: (i) los indicadores expresados en porcentaje ingresan en escala 0–100 y se combinan multiplicativamente; (ii) los niveles monetarios en ARS se aplican como insumos directos o, si se usan en % de variación, con β cercanos a 1 cuando corresponde; (iii) los rezagos se implementan desplazando el indicador uno o más meses respecto del periodo de la cuenta; y (iv) se establecen topes operativos para evitar explosiones numéricas en escenarios extremos (por ejemplo, recortes en variaciones mayores al 25% mensual, que se documentan en notas del libro). Así, el conjunto de afecciones queda completamente parametrizado y auditable sin editar fórmulas estructurales.

La integración técnica se resuelve en un libro de Excel con tres hojas. En la primera, “01_Python_Base”, se pega tal cual la exportación del pipeline (periodo, cuenta/factor, monto base y metadatos que provee el script). En la segunda, “02_GPT_Indicadores”, se pega la tabla del bot con cuatro columnas ,Indicador, t0, t-1, t-2, exactamente en el orden definido en la etapa anterior; sobre esa tabla se crean referencias por nombre para cada indicador, de modo que las fórmulas no dependen de posiciones. La tercera hoja, “03_Montos_Afectados”, contiene el motor de cálculo: lee los montos base por fila desde la hoja 01 y aplica, para cada cuenta, la combinación de impactos definida en el Apéndice VII, trayendo de la hoja 02 los indicadores correctos (con su posible rezago) y los β correspondientes. El resultado es el Monto_Final por periodo–cuenta, que queda listo para consolidación.

Con el fin de reducir errores, la hoja de integración incorpora controles previos y semaforización. Antes de calcular, valida que en “02_GPT_Indicadores” existan los doce indicadores requeridos y que sus valores sean numéricos y estén dentro de rangos razonables; a su vez, verifica que en “01_Python_Base” los campos críticos no estén vacíos ni duplicados. Cuando el bot debe estimar (nowcast) algún valor del mes objetivo, la hoja refleja la condición mediante una marca de revisión (amarillo) para que el equipo decida si procede o si mantiene la base del mes anterior para esa afección específica. Todo ajuste queda documentado mediante un pequeño bloque de metadatos de corrida (fecha y versión de prompt del bot; hash de configuración de Python; fecha de pegado en cada hoja), preservando la trazabilidad.

El procedimiento de uso es simple y repetible: (1) ejecuta el pipeline y pega su salida en "01_Python_Base"; (2) solicita al bot la actualización del periodo (t_0 con $t-1$ y $t-2$) y pega la tabla en "02_GPT_Indicadores"; (3) revisa los semáforos en "03_Montos_Afectados" y, si están en verde, toma los Monto_Final para la siguiente capa del modelo. Esta organización separa claramente responsabilidades (cada hoja cumple una función), permite calibrar β y rezagos sin reescribir fórmulas y deja un rastro verificable de cada corrida. De este modo, la etapa asegura que la señal del entorno, obtenida 100% online por el bot, se incorpore al presupuesto con reglas matemáticas explícitas, controladas y fáciles de mantener.

Etapa 9 - Lanzamiento, Medición de Valor y Adopción Estratégica

La Etapa 9 representa la culminación práctica de todo el rigor metodológico desarrollado a lo largo del proyecto, y transforma los modelos y algoritmos en un activo operativo y decisional para la Gerencia. En este momento, la organización contrasta la solidez cuantitativa del sistema integrado, compuesto por los modelos de Python, el Bot GPT y el libro de Excel de integración, con la realidad del negocio, cerrando el ciclo de diseño e iniciando formalmente la senda de la mejora continua. El sistema analítico genera su primera proyección de referencia sobre la base de datos de cierre de junio, y proporciona la visión anticipada que la organización demanda.

El valor esencial de esta implementación se materializa al acortar la brecha de información. Mientras la información financiera clave opera con un desfase de hasta un mes, lo que fuerza decisiones sobre estimaciones parciales, el sistema permite la generación de un pronóstico clave con tres meses de anticipación. Específicamente, el indicador económico total proyectado para el mes de septiembre, con base en el cierre de junio, arroja una variación negativa del 8,32%. Este dato, el primer pronóstico de referencia generado por la nueva arquitectura, se convierte en la línea base y el ancla con la cual la Gerencia comienza a simular escenarios y ajustar la planificación de recursos. La clave reside en que esta estimación resulta coherente, explicable y completamente trazable, superando el problema de operar con supuestos que luego deben corregirse.

El sistema está diseñado para ser auditable, no una caja negra, por lo que el resultado de septiembre se somete a un análisis de descomposición de la desviación. Este proceso resulta crucial para identificar si el error final es atribuible a una deficiencia en el modelo base interno (la elección del algoritmo de Python) o a la calidad de la señal externa (si el nowcast del IPC o del precio del Gasoil realizado por el Bot GPT se desvía del valor oficial final). El análisis detallado sobre las cuentas de mayor materialidad (como VTAS y CBLC) permite validar el marco metodológico: si una cuenta crítica se desvía, se verifica si la causa es una calibración incorrecta del coeficiente (elasticidad) definido en el Apéndice VII o una variación abrupta no capturada en el nowcast del indicador ICL FADEEAC o el Gasoil. Esta auditoría constituye la base para la revisión periódica del sistema y la mejora continua de sus componentes.

El éxito final de esta etapa reside en la integración del sistema al ciclo decisional de la empresa, y provee el lenguaje común de resultados esperados que las gerencias solicitan para alinear lo operativo, lo comercial y lo financiero. La hoja de Excel "03_Montos_Afectados" se

transforma en la interfaz gerencial táctica para la toma de decisiones, y permite calibrar los coeficientes y simular el impacto de los indicadores externos sin necesidad de correr código. La disponibilidad de proyecciones con un horizonte de tres meses permite a las áreas pasar de una gestión reactiva a una proactiva, ajustando costos, dimensionando flotas y renegociando tarifas con evidencia trazable. El mecanismo de retroalimentación de ciclo corto se establece cuando las gerencias, al observar la desviación en el libro de Excel, se enfocan en ajustar el parámetro de influencia o el modelo de base para el siguiente ciclo de proyección, institucionalizando así la cultura de la anticipación. El proyecto culmina dejando a la organización con un procedimiento explícito, controlable y replicable, listo para pasar de una gestión basada en rezagos a una que pronostica con evidencia.

Recomendaciones

Se recomienda mantener una política de revisión periódica del sistema, tanto en su dimensión técnica como en su uso operativo. Los modelos y parámetros definidos, en especial las elasticidades y rezagos del Apéndice VII, deben actualizarse con evidencia empírica a medida que se acumulan nuevas observaciones. La estabilidad del modelo depende de la calidad de los datos y de la consistencia en su captura y carga.

Asimismo, se aconseja institucionalizar un proceso de capacitación breve y recurrente para las áreas que intervienen en la generación de datos. La homogeneización en los criterios de registro y la comprensión compartida del plan de cuentas presupuestario fortalecen la trazabilidad y la comparabilidad intermensual, y reducen la necesidad de correcciones manuales.

En el plano operativo, resulta conveniente priorizar la automatización de aquellas fuentes que actualmente se cargan manualmente y migrar gradualmente los circuitos más estables de captura de indicadores externos a flujos n8n, lo que reduce la dependencia del bot GPT en tareas de bajo valor analítico. Esto permite concentrar el uso de la inteligencia artificial en la exploración contextual y no en la recolección rutinaria.

Por último, se sugiere documentar cada corrida del sistema en un registro de control interno, incluyendo la versión del código, la fecha de obtención de indicadores y la validación del resultado por parte de las gerencias usuarias. De este modo, la herramienta se consolida como un componente institucional del proceso de planificación y no como un desarrollo aislado, asegurando continuidad, auditabilidad y mejora continua.

Conclusiones

El trabajo permite sentar las bases de un sistema analítico capaz de transformar información histórico-contable y operativa en proyecciones económico-financieras trazables, reproducibles y alineadas con las necesidades decisionales de la empresa. El proceso no consiste en desarrollar una solución definitiva, sino en establecer un marco estructurado, conceptual, metodológico y técnico, sobre el cual pueden apoyarse mejoras futuras.

La aplicación muestra que la principal limitación no reside en la falta de datos, sino en su dispersión, asincronía y escasa estandarización. Las etapas de limpieza, armonización y

traducción al plan de cuentas presupuestario demuestran que es posible integrar los distintos lenguajes organizativos (operativo, comercial, financiero y de calidad) bajo un mismo esquema de lectura y control, lo cual constituye un avance sustantivo en términos de gobernanza informacional.

La implementación de la metodología de proyección en Python y su integración con indicadores externos mediante el bot GPT representa un primer paso hacia la automatización del análisis y la anticipación de escenarios. Si bien los resultados iniciales, como el pronóstico de referencia para septiembre, validan la coherencia técnica del procedimiento, aún no se evalúa su impacto directo sobre los resultados económicos ni su desempeño sostenido en el tiempo. Esto requiere nuevas corridas, calibración de coeficientes y revisión periódica de las afecciones definidas.

En términos institucionales, el trabajo visibiliza la necesidad de un circuito de información más corto, confiable y transversal entre áreas, confirmando que la gestión anticipatoria no depende exclusivamente de la capacidad técnica, sino también de la coordinación organizativa. Las proyecciones se consolidan así como un insumo útil para la planificación, sin pretender sustituir la experiencia gerencial ni los controles presupuestarios tradicionales, sino complementarlos con evidencia cuantitativa y trazable.

En suma, la investigación cumple su propósito de diseñar e implementar una metodología sólida de cálculo de proyecciones, documentada y replicable, que aporta una base práctica para seguir perfeccionando el sistema analítico y avanzar gradualmente hacia una gestión basada en datos oportunos y verificables.

Referencias

- Box, G. E. P., & Jenkins, G. M. (1970). Time Series Analysis: Forecasting and Control. Holden-Day.
- DAMA International. (2017). DAMA-DMBOK: Data Management Body of Knowledge (2.^a ed.). Technics Publications.
- Holt, C. C. (1957). Forecasting seasonals and trends by exponentially weighted moving averages. Office of Naval Research.
- Hyndman, R. J., & Athanasopoulos, G. (2018). Forecasting: Principles and Practice (2nd ed.). OTexts.
- Marr, B. (2018). Inteligencia artificial en la práctica. Ediciones Deusto.
- Simon, H. A. (1960). El comportamiento administrativo. Aguilar.
- Winters, P. R. (1960). Forecasting sales by exponentially weighted moving averages. Management Science, 6(3), 324–342.

Apéndice

Apéndice I – Entrevistas preliminares a la gerencia

I.I Entrevista – Gerente General

Objetivo:

Conocer, desde la mirada estratégica, cuáles son los insumos que utiliza para decidir, cómo fluye la información en la empresa y qué utilidad real tienen esos datos, para detectar limitaciones y oportunidades de mejora.

Pregunta 1:

¿Cuáles son las principales fuentes de información que utiliza para tomar decisiones estratégicas? - Hoy, la base de casi todo lo que decidimos son reportes contables que manda la central y las planillas internas que armamos nosotros. En su mayoría son Excel hechos por cada área, a mano, a partir de facturas y registros diarios. La información es confiable en cuanto a su origen, pero llega tarde: los cierres oficiales vienen con dos meses de atraso. Eso nos obliga a “inventar” un poco en el medio, completando con datos operativos parciales para no quedarnos a ciegas, pero eso no siempre nos da una foto clara.

Pregunta 2:

¿Cómo describiría el circuito actual de información entre las distintas áreas y la gerencia general? - Es bastante lineal: cada gerente me pasa su parte, yo junto todo y tratamos de armar el panorama. La comunicación con ellos es buena, pero el proceso es lento y pesado. No tenemos un sistema que integre todo, así que dependemos de que cada uno mande su archivo y de que eso después cuadre. Entre idas y vueltas, se pierde mucho tiempo y la información deja de ser fresca.

Pregunta 3:

¿Considera que la información le llega a tiempo y en el formato necesario para decidir? - No, la verdad es que no. El retraso de dos meses en los números oficiales es un problema serio. Hay decisiones que se toman con supuestos y no con datos confirmados. Y cuando me llegan, es en planillas que sirven para ver el pasado, pero no para movernos rápido hacia adelante.

Pregunta 4:

¿La información que recibe le permite actuar de forma proactiva o más bien reactiva? - Más bien reactiva. Muchas veces actuamos cuando el problema ya está encima. Lo que me salva es la experiencia, que me permite anticipar algunas cosas, pero no siempre alcanza. Nos ha pasado de quedarnos cortos en un presupuesto y perder un proyecto, o de sobredimensionar y ver cómo se nos achica el margen. Eso, con proyecciones más precisas, se podría evitar.

Pregunta 5:

¿Qué indicadores considera críticos para su gestión y con qué frecuencia se proyectan? - Los que más miro son margen operativo y rentabilidad. Las proyecciones se hacen una vez por mes, pero sinceramente, la confianza que les tengo es más bien poca. Sabemos que los números pueden cambiar cuando llega el cierre real, y eso hace que siempre haya un margen de duda.

Pregunta 6:

¿Qué cuellos de botella o retrasos percibe en el flujo de información? - El gran cuello de botella es que todo se hace a mano. Compilar datos de distintas fuentes lleva días, y encima dependemos de que estén bien cargados desde el principio. Otro tema es que no se cruzan datos externos relevantes, como cambios de tarifas o movimientos de nuestros clientes grandes, así que las proyecciones no reflejan todo el escenario.

Pregunta 7:

¿Qué tipo de datos o reportes le gustaría recibir y actualmente no recibe? - Me encantaría tener proyecciones económico-financieras en tiempo real, que combinen datos operativos, contables y del mercado. Algo que me permita ver cómo se está moviendo todo y simular escenarios para decidir sin tener que esperar dos meses.

Pregunta 8:

Si pudiera cambiar tres cosas en el sistema de información de la empresa, ¿cuáles serían? - Primero, unificar todo en un sistema que junte operaciones y finanzas. Segundo, achicar el tiempo entre el cierre y la disponibilidad de datos a menos de una semana. Tercero, incorporar un modelo predictivo que convierta a las proyecciones en nuestra herramienta principal de planificación.

I.II Entrevista – Gerente Operativo**Objetivo:**

Recoger la visión operativa sobre los insumos que utiliza para tomar decisiones, identificar problemas en el flujo de datos y evaluar cómo la información actual impacta en la planificación y ejecución de las operaciones.

Pregunta 1:

¿Cuáles son las principales fuentes de información que utiliza para tomar decisiones en su área? - Lo que más uso son los datos operativos del día: cantidad de envíos, rutas, disponibilidad de vehículos, tiempos de entrega. Eso me llega directamente de las sucursales y de los prestadores de servicio. También recibo algunas planillas con proyecciones de facturación, pero no siempre las uso para la operación diaria porque cambian mucho y no siempre son confiables.

Pregunta 2:

¿Cómo fluye la información entre su área y la Gerencia General? - Yo paso reportes periódicos de movimientos y viajes realizados, y cualquier problema operativo importante lo informo de inmediato. La relación es directa y sin vueltas. El tema es que lo que envío es lo que pasa hoy o pasó ayer; la parte financiera viene por otro lado y, cuando llega, a veces ya tomamos decisiones que no se pueden cambiar.

Pregunta 3:

¿Recibe la información que necesita a tiempo y en el formato que le resulta útil? - En cuanto a lo operativo, sí, porque lo genero yo y mi equipo. En lo financiero, no: ahí dependemos de cierres que vienen con atraso y eso nos complica cuando tenemos que ajustar servicios o negociar con los prestadores.

Pregunta 4:

¿La información le permite actuar de manera proactiva o más bien reactiva? - En la operación, reaccionamos rápido porque vemos los problemas al momento. Pero para decisiones más grandes, como contratar más vehículos o abrir rutas, dependemos de proyecciones que no siempre reflejan la realidad. Eso hace que a veces aceptemos servicios que después no son rentables o rechacemos trabajos que podríamos haber hecho.

Pregunta 5:

¿Cuáles son los principales cuellos de botella que ve en el flujo de información? - El principal problema es que la parte financiera no está integrada con lo que pasa en la calle. Los números llegan tarde y no se cruzan con lo que yo veo día a día. Si tuviéramos eso en tiempo real, podríamos ajustar los costos al instante y no esperar a fin de mes o más.

Pregunta 6:

¿Confía en las proyecciones que recibe? - Poco. Las tomo como una referencia, pero no como algo definitivo. Muchas veces la realidad de la operación no coincide con lo que se proyectó, y eso genera problemas cuando tenemos que decidir rápido.

Pregunta 7:

¿Qué le gustaría mejorar del sistema de información? - Me gustaría que las proyecciones de facturación sean más precisas y que se integren con los datos operativos. Si pudiera ver en un mismo lugar lo que facturamos, lo que gastamos y lo que estamos moviendo hoy, las decisiones serían mucho más fáciles y acertadas.

Pregunta 8:

¿Cree que una proyección económico-financiera más precisa ayudaría en su trabajo? - Sin dudas. Si supiéramos con más certeza lo que vamos a facturar y gastar, podríamos planificar mejor la cantidad de vehículos, las rutas y hasta el personal necesario. Evitaríamos gastar de más o perder oportunidades por miedo a no cubrir los costos.

I.III Entrevista – Gerente Comercial**Objetivo:**

Explorar los insumos que utiliza el área comercial para la toma de decisiones, evaluar el estado actual de la información que recibe y cómo la calidad de las proyecciones afecta la negociación con clientes y la rentabilidad de la empresa.

Pregunta 1:

¿Cuáles son las principales fuentes de información que utiliza para tomar decisiones comerciales? - Trabajo con proyecciones de ventas que me entrega mi equipo casi todos los días, además de datos de clientes actuales y potenciales. Lo más importante es saber cómo se están comportando las grandes cuentas, sobre todo el cliente principal que representa más del 60% de nuestra facturación. Esa concentración hace que cada decisión tenga mucho peso: un error en cómo se proyectan los costos o las ventas con ese cliente puede impactar de manera directa en los resultados de la empresa.

Pregunta 2:

¿Cómo fluye la información entre su área y la Gerencia General? - Yo reporto periódicamente volúmenes de ventas, evolución de clientes y oportunidades comerciales. La comunicación es formal y ordenada, pero el problema no es la relación, sino la calidad de la información que compartimos. Muchas veces los reportes que mando están contruidos sobre proyecciones financieras que después no se confirman, y eso debilita nuestra posición a la hora de tomar decisiones estratégicas.

Pregunta 3:

¿Recibe la información que necesita a tiempo y en el formato adecuado para negociar? - En cuanto a ventas, sí, porque mi equipo lo mide todos los días. El problema es cuando necesito cruzar esos datos con los costos y la rentabilidad real. Ahí me encuentro con estimaciones que cambian con el tiempo, y lo que parecía un negocio rentable termina no siéndolo, o al revés. Eso me deja en una posición incómoda frente al cliente, porque la información que uso para negociar no siempre es la más precisa.

Pregunta 4:

¿La información le permite actuar de forma proactiva o reactiva? - En el mejor de los casos, me permite actuar proactivo con los clientes porque tengo un trato diario. Pero a nivel financiero, es reactivo: muchas veces me entero de que un acuerdo no fue rentable cuando ya está cerrado. Eso es frustrante, porque podría haber renegociado condiciones si hubiera contado con proyecciones más confiables en el momento justo.

Pregunta 5:

¿Qué problemas ha enfrentado por errores en las proyecciones de costos? - Es una de las mayores debilidades que tenemos. Cuando los costos se proyectan por debajo de lo real, terminamos cerrando contratos que parecen atractivos pero que luego reducen el margen al mínimo. Y cuando los costos se proyectan demasiado altos, rechazamos negocios que en realidad hubieran sido rentables. Ambas situaciones nos hacen perder competitividad: en un caso por perder rentabilidad, en el otro por perder volumen de mercado.

Pregunta 6:

¿Qué indicadores considera críticos para su gestión? - Lo que más sigo son las ventas por cliente, especialmente la evolución del cliente principal y su participación en la facturación total. También miro mucho el margen de rentabilidad por contrato, aunque ahí volvemos al mismo problema: si las proyecciones no son sólidas, ese indicador pierde valor como insumo real de decisión.

Pregunta 7:

¿Qué nivel de confianza tiene en las proyecciones económico-financieras que recibe? - Diría que moderado. No puedo descartarlas porque son nuestra base para presupuestar y negociar, pero tampoco puedo confiar ciegamente. Esa desconfianza genera tensión en las negociaciones: uno entra a discutir con un cliente sin tener la certeza de si el negocio va a ser rentable o no, y eso es un riesgo grande para la empresa.

Pregunta 8:

¿Qué mejoras considera necesarias en el sistema de información de la empresa? - Necesitamos con urgencia un sistema que integre las ventas con los costos reales y las

proyecciones económico-financieras. Hoy tenemos partes separadas y eso genera desajustes. Si pudiéramos ver en tiempo real cuánto nos cuesta y cuánto nos deja cada operación, podríamos negociar con mucha más solidez y seguridad. Eso no solo nos evitaría perder rentabilidad, también nos permitiría aprovechar oportunidades que hoy dejamos pasar.

Pregunta 9:

¿Cree que una proyección económico-financiera más precisa impactaría directamente en el área comercial? - Absolutamente. Para nosotros es el insumo clave. Una proyección más confiable me daría la seguridad de cerrar contratos sabiendo de antemano cuál va a ser el margen real. Nos haría más competitivos, porque podríamos ajustar precios de manera inteligente y reaccionar más rápido a las exigencias de los clientes, sin miedo a equivocarnos.

I.IV Entrevista – Gerente de Calidad

Objetivo:

Analizar desde el área de calidad cuáles son los insumos que se utilizan para tomar decisiones, cómo se gestionan los datos de procesos y qué dificultades existen para convertir la información en acciones concretas de mejora.

Pregunta 1:

¿Cuáles son las principales fuentes de información que utiliza para evaluar la calidad en la organización? - Prácticamente recibo datos de todas las áreas: operativos, comerciales, financieros, de recursos humanos. Los utilizo para construir indicadores de desempeño que me permiten medir cómo estamos funcionando. El problema es que la información llega en distintos formatos, con distintos tiempos y, muchas veces, sin la consistencia necesaria. Eso hace que, en lugar de tener un tablero claro, tenga que dedicar gran parte del tiempo a ordenar y validar datos.

Pregunta 2:

¿Cómo describiría el circuito de información entre su área y la Gerencia General? - Es bastante fluido. Yo envío reportes periódicos con indicadores clave ,nivel de servicio, cumplimiento de plazos, reclamos, eficiencia de procesos,, pero muchas veces aclaro que esos números son aproximaciones, porque dependen de datos que no siempre son confiables o llegan atrasados. La Gerencia General lo entiende, pero todos sabemos que trabajar con aproximaciones es un riesgo.

Pregunta 3:

¿Considera que la información que recibe le permite actuar de forma proactiva o más bien reactiva? - En algunos temas operativos, puedo ser proactivo porque detecto fallas a tiempo. Pero en lo económico-financiero, que es lo que más condiciona la calidad, soy reactivo. Si un servicio se proyecta mal en términos de costos, recién me entero del impacto cuando ya está ejecutado, y en ese punto lo único que queda es registrar la pérdida o el desvío, no anticiparlo.

Pregunta 4:

¿Cuáles son los principales cuellos de botella que observa en el flujo de información? - El gran problema es la falta de integración. Cada área trabaja con sus planillas, sus tiempos y

sus formatos. Yo recibo todo, pero no hay un sistema que los unifique ni que garantice que lo que llega es el dato definitivo. Además, la información financiera suele estar desfasada, y eso limita la posibilidad de vincular la calidad con su verdadero costo.

Pregunta 5:

¿Qué indicadores considera más críticos para evaluar la calidad de los procesos? - Los KPI más relevantes son el porcentaje de entregas a tiempo, la cantidad de reclamos por cliente y los costos de no calidad (lo que nos cuesta rehacer, corregir o responder a un servicio mal ejecutado). Pero de nuevo, esos indicadores solo tienen sentido si están respaldados por datos económicos y financieros sólidos, porque de lo contrario mido la calidad de forma parcial.

Pregunta 6:

¿Qué nivel de confianza tiene en la información que recibe actualmente? - Diría que intermedio. Para los aspectos operativos tengo buena visibilidad, pero en lo financiero la confianza baja porque sé que hay atrasos y diferencias según la fuente. Esa brecha me obliga a trabajar con márgenes de duda, cuando lo ideal sería tener certezas para planificar.

Pregunta 7:

¿Qué mejoras considera necesarias en el sistema de información de la empresa? - Necesitamos un sistema integrado que alinee datos operativos, comerciales y financieros en tiempo real. De esa manera, los indicadores de calidad dejarían de ser una foto atrasada y se convertirían en una herramienta viva de gestión. Eso nos permitiría no solo detectar desvíos, sino anticiparlos y corregirlos antes de que impacten en el cliente.

Pregunta 8:

¿Cree que una proyección económico-financiera más precisa tendría impacto en la gestión de calidad? - Sí, absolutamente. Gran parte de los problemas de calidad se relacionan con decisiones tomadas sin un sustento financiero sólido. Si supiéramos con certeza cuánto cuesta cada servicio y qué rentabilidad real genera, podríamos ajustar procesos, planificar mejor y evitar errores que después se transforman en reclamos o pérdidas. Una proyección más precisa sería la base para que la calidad deje de ser correctiva y pase a ser preventiva.

Apéndice II – Glosario de terminología de la empresa**Pieza Postal**

Toda correspondencia o paquete (cartas, notificaciones, facturas, tarjetas de crédito, etc.) que se transporta a través del negocio postal de la empresa.

Acuse

Forma abreviada de Acuse de Recibo. Documento o constancia que acredita que el destinatario recibió efectivamente una pieza postal.

CD

Abreviatura de Carta Documento, una de las posibles piezas a mover en el negocio postal.

TC

Abreviatura de Tarjeta de Crédito, reconocida como pieza postal.

TD

Abreviatura de Tarjeta de Débito, reconocida como pieza postal.

Pieza Simple

Piezas que no requieren confirmación de recepción. Pueden ser entregadas sin interacción directa, por ejemplo, dejándolas bajo puerta.

Bulto

Unidad física mínima del proceso logístico. Todo aquello que puede distinguirse visualmente como un paquete individual.

Pieza Logística

Unidad de contabilización a nivel de sistema. Puede estar compuesta por uno o más bultos y, por lo general, se factura por pieza y no por bulto.

Pieza Multi-Bulto

Pieza logística conformada por más de un bulto.

Pallet

Estructura de carga estandarizada (madera, plástico u otro material) utilizada para agrupar, sostener y movilizar varios bultos.

Troncal

Viajes principales que parten desde el depósito central en Buenos Aires hacia depósitos regionales. La mayoría tiene como destino Tucumán, aunque también se realizan a Córdoba y Mendoza.

Subtroncal

Viajes que conectan depósitos secundarios entre sí. Suelen partir de un depósito troncal hacia otros depósitos de la empresa.

Primera Milla (First Mile / FM)

Etapas inicial del proceso logístico: retiro de la mercadería desde el cliente hasta un depósito de la empresa.

Última Milla (Last Mile / LM)

Etapas final del proceso logístico: traslado desde el depósito hasta el domicilio del consumidor final. Generalmente realizada por camionetas.

Line Haul (LH)

Término del cliente principal de la organización para referirse a los viajes troncales y subtroncales.

Siniestro

Cualquier percance que pueda sufrir la carga, incluyendo robos, roturas, extravíos o accidentes.

DDI

Acrónimo de Distribuidor del Interior. Inicialmente designaba a los distribuidores que entregaban piezas fuera de las capitales provinciales. Actualmente se aplica a cualquier repartidor que facture por pieza entregada en lugar de hacerlo por jornal.

Paquetería

Negocio vinculado al transporte de paquetes encomendados directamente a la empresa, es decir, la carga efectivamente movilizada.

Gestión de Flota

Modelo de negocio en el cual la empresa alquila su flota vehicular a otra organización por jornal, independientemente del uso que se haga de los vehículos.

Volumetría

Medición de las dimensiones de la carga en metros cúbicos. Se aplica al cálculo de cada pieza transportada.

Pieza Grande

Pieza cuya volumetría supera los 0,042 m³.

Pieza Chica

Pieza cuya volumetría es inferior a los 0,042 m³.

Ambulancia

Vehículo asignado para cubrir recorridos en caso de que un driver no pueda cumplir su ruta. Se encarga de tomar su carga y completarla.

Driver

Persona encargada de conducir un vehículo de la flota contratada por la empresa.

FAR

Acrónimo de Fondo a Rendir. Fondos entregados a un empleado contra la obligación de presentar comprobantes que justifiquen su uso.

Apéndice III – Descriptivo de Cuentas Presupuestarias

III.I Negocio Postal

VTAS: Concentrador de todas las ventas, expresión tangible de la tracción comercial del negocio.

DCTO: Recupero del Decreto 814 (IVA), palanca fiscal clave para mejorar márgenes operativos.

ALQC: Alquiler de camionetas liquidadas por jornal, aportando flexibilidad en la cobertura.

ALCM: Alquiler de camiones como refuerzo estructural para el transporte de gran escala.

DDI: Alquiler de camionetas por pieza transportada, alineando costos con productividad real.

FNSH: Liquidaciones del personal de Finish, núcleo especializado en armado de piezas postales.

ENCM: Liquidaciones de encomiendas, motor complementario del flujo de ingresos.

ALIN: Alquiler de inmuebles estratégicos, soporte físico de la operación institucional.

ENEL: Energía eléctrica como insumo esencial para garantizar continuidad de procesos.

SGLC: Seguridad del local, inversión en resguardo de activos y confianza del cliente.

SGRS: Seguros varios, cobertura preventiva que preserva la estabilidad patrimonial.

SVLP: Servicios de limpieza, condición base para un estándar operativo de calidad.

SLDS: Sueldos administrativos y operativos, motor humano de la estructura postal.

SAC: Sueldo anual complementario, previsión financiera para equilibrar el ciclo remunerativo.

CGSC: Cargas sociales, reflejo del compromiso institucional con el sistema previsional.

APSC: Aportes sindicales, erogaciones que consolidan el vínculo laboral colectivo.

IMPT: Impuestos varios, cumplimiento fiscal ineludible dentro del marco normativo.

ITFN: Intereses financieros, costo asociado a la estrategia de apalancamiento.

GTVR: Gastos varios, egresos de soporte necesarios para sostener la fluidez operativa.



III.II Negocio Logístico

VTAS: Ventas varias, indicador directo del alcance comercial del negocio logístico.

DCTO: Recupero de Decreto 814 (IVA), herramienta clave para optimización fiscal.

ALQC: Alquiler de camionetas en cualquiera de sus modalidades, recurso de flexibilidad territorial.

ALCM: Alquiler de camiones, asegurando robustez en el transporte de cargas.

ALIN: Alquiler de inmuebles, soporte físico de la cadena de operaciones.

SLDS: Sueldos administrativos y operativos, pilar humano de la operatoria logística.

CGSC: Cargas sociales, obligaciones inherentes al sostenimiento del capital humano.

CBLC: Combustibles y lubricantes, insumo vital para la movilidad y continuidad de la flota.

HNRR: Honorarios profesionales, aporte externo de conocimiento y asesoría estratégica.

SAC: Sueldo anual complementario, reserva financiera para dar previsibilidad a la nómina.

IMPT: Impuestos varios, gravámenes que enmarcan la actividad en el sistema legal vigente.

SDCT: Aportes sindicales, erogación que respalda la representación colectiva del personal.

IDMZ: Indemnizaciones y liquidaciones finales, reflejo de la dinámica en la gestión laboral.

GTVR: Gastos varios, categoría que absorbe erogaciones necesarias para la operatividad.

Apéndice IV – Código en Python para proyección cuantitativa

```
!pip install --upgrade "packaging<25"
!pip install --upgrade --force-reinstall \
    "numpy==2.1.3" \
    "scikit-learn==1.6.1" \
    "pandas>=2.0" \
    "statsmodels>=0.14" \
    "xlsxwriter>=3.1" \
    "openpyxl>=3.1"
```

```
# === CONFIGURACIÓN: RUTA DE ENTRADA ===
```

```
INPUT_PATH = r"C:\Users\UNSTA\OneDrive\Escritorio\Práctica Profesional\Base de Datos
Proyección\INPUT.xlsx"
```

```
# Si estás en Linux/Mac o preferís slashes:
```

```
# INPUT_PATH = "/ruta/a/INPUT.xlsx"
```

```
# === IMPORTS ===
```

```
import warnings
```

```
from statsmodels.tools.sm_exceptions import ConvergenceWarning
```

```
warnings.filterwarnings("ignore")
```

```
warnings.filterwarnings("ignore", category=ConvergenceWarning)
```



```
warnings.filterwarnings("ignore", message="Maximum Likelihood optimization failed to  
converge")
```

```
warnings.filterwarnings("ignore", message="Optimization failed to converge")
```

```
import math
```

```
import json
```

```
from dataclasses import dataclass
```

```
from typing import Dict, List, Tuple, Optional
```

```
import numpy as np
```

```
import pandas as pd
```

```
from pandas.tseries.frequencies import to_offset
```

```
from sklearn.linear_model import ElasticNet
```

```
from sklearn.ensemble import GradientBoostingRegressor
```

```
from sklearn.preprocessing import StandardScaler
```

```
from sklearn.pipeline import Pipeline
```

```
from statsmodels.tsa.holtwinters import ExponentialSmoothing
```

```
from statsmodels.tsa.statespace.sarimax import SARIMAX
```

```
from statsmodels.tsa.seasonal import STL
```

```
try:
```

```
    from statsmodels.tsa.forecasting.theta import ThetaModel
```

```
    THETA_AVAILABLE = True
```

```
except Exception:
```

```
    THETA_AVAILABLE = False
```

```
# =====
```

```
# Métricas y utilidades CV
```

```
# =====
```

```
def smape(y_true: np.ndarray, y_pred: np.ndarray) -> float:
```

```
    y_true = np.asarray(y_true, dtype=float)
```

```
    y_pred = np.asarray(y_pred, dtype=float)
```

```
    denom = (np.abs(y_true) + np.abs(y_pred))
```

```
    denom = np.where(denom == 0, 1.0, denom) # evitar div 0
```

```
    return np.mean(2.0 * np.abs(y_pred - y_true) / denom) * 100.0
```

```
def mase(y_true: np.ndarray, y_pred: np.ndarray, y_in_sample: np.ndarray, m: int = 1) -> float:
```

```
    y_true = np.asarray(y_true, dtype=float)
```

```
    y_pred = np.asarray(y_pred, dtype=float)
```



```
y = np.asarray(y_in_sample, dtype=float)
if m > 1 and len(y) > m:
    scale = np.mean(np.abs(y[m:] - y[:m]))
else:
    scale = np.mean(np.abs(np.diff(y)))
if not np.isfinite(scale) or scale == 0:
    scale = 1.0
return np.mean(np.abs(y_true - y_pred)) / scale
```

@dataclass

class CVResult:

mase: float
smape: float
details: Dict

def rolling_origin_cv(y: pd.Series, horizon: int, n_splits: int, fit_forecast_func, m_season: int = 12) -> CVResult:

```
y = y.dropna().astype(float)
n = len(y)
min_train = max(2 * m_season, 24)
if n < (min_train + n_splits * horizon):
    n_splits = max(1, (n - min_train) // horizon)
    if n_splits < 1:
        n_splits = 1

mase_scores, smape_scores = [], []
cutpoint = n - n_splits * horizon
for i in range(n_splits):
    train_end = cutpoint + i * horizon
    train = y.iloc[:train_end]
    test = y.iloc[train_end: train_end + horizon]
    if len(test) < horizon or len(train) < max(12, m_season+2):
        continue
    try:
        yhat = fit_forecast_func(train, horizon)
        yhat = np.asarray(yhat).reshape(-1)[:len(test)]
        mase_scores.append(mase(test.values, yhat, train.values, m=m_season))
        smape_scores.append(smape(test.values, yhat))
    except Exception:
        continue
```



```

if len(mase_scores) == 0:
    return CVResult(mase=np.inf, smape=np.inf, details={"splits": 0})
return CVResult(mase=float(np.mean(mase_scores)),
                smape=float(np.mean(smape_scores)),
                details={"splits": len(mase_scores)})

# =====
# Modelos: 9 familias
# =====

# --- Baselines ---
def fc_naive(train: pd.Series, horizon: int) -> np.ndarray:
    return np.repeat(train.iloc[-1], horizon)

def fc_seasonal_naive(train: pd.Series, horizon: int, m: int = 12) -> np.ndarray:
    if len(train) < m:
        return fc_naive(train, horizon)
    last_season = train.iloc[-m:]
    reps = int(math.ceil(horizon / m))
    return np.tile(last_season.values, reps)[:horizon]

def fc_moving_average(train: pd.Series, horizon: int, window: int = 3) -> np.ndarray:
    window = max(1, min(window, len(train)))
    avg = train.iloc[-window:].mean()
    return np.repeat(avg, horizon)

# --- ETS ---
def fc_ets(train: pd.Series, horizon: int, m: int = 12) -> np.ndarray:
    configs_fast = [
        dict(trend=None, seasonal="add", damped_trend=False),
        dict(trend="add", seasonal="add", damped_trend=False),
        dict(trend="add", seasonal=None, damped_trend=True),
    ]
    configs_full = configs_fast + [dict(trend=None, seasonal=None, damped_trend=False)]

    configs = configs_fast if FAST_MODE else configs_full

    best_aic = np.inf
    best_fc = None
    for cfg in configs:
        try:

```



```

model = ExponentialSmoothing(
    train, trend=cfg["trend"], seasonal=cfg["seasonal"],
    seasonal_periods=m, damped_trend=cfg["damped_trend"],
    initialization_method="heuristic" # más estable
)
res = model.fit(optimized=True, use_brute=False, use_boxcox=None,
remove_bias=False)
aic = getattr(res, "aic", np.inf)
if aic < best_aic:
    best_aic = aic
    best_fc = res.forecast(horizon).values
except Exception:
    continue
return best_fc if best_fc is not None else fc_seasonal_naive(train, horizon, m)

# --- SARIMA ---
def fc_sarima(train: pd.Series, horizon: int, m: int = 12) -> np.ndarray:
    # Grilla liviana si FAST_MODE, un poco más amplia si no
    if FAST_MODE:
        orders = [(0,0,0), (1,0,0), (0,1,1)]
        seasonals = [(0,0,0,m), (1,0,0,m), (0,1,1,m)]
    else:
        orders = [(p,d,q) for p in range(0,2) for d in (0,1) for q in range(0,2)]
        seasonals = [(P,D,Q,m) for P in range(0,2) for D in (0,1) for Q in range(0,2)]

    best_bic = np.inf
    best_fc = None
    for (p,d,q) in orders:
        for (P,D,Q,mm) in seasonals:
            try:
                model = SARIMAX(
                    train, order=(p,d,q), seasonal_order=(P,D,Q,mm),
                    enforce_stationarity=False, enforce_invertibility=False
                )
                res = model.fit(dispatch=False, maxiter=SARIMAX_MAXITER)
                # Si no converge, seguimos probando otras combinaciones
                converged = getattr(res, "mle_retvals", {}).get("converged", True)
                if not converged:
                    continue
                bic = res.bic
                if bic < best_bic:

```

```

        best_bic = bic
        best_fc = res.get_forecast(steps=horizon).predicted_mean.values
    except Exception:
        continue
    return best_fc if best_fc is not None else fc_seasonal_naive(train, horizon, m)

# --- STL + ARIMA (en residuos) ---
def fc_stl_arima(train: pd.Series, horizon: int, m: int = 12) -> np.ndarray:
    stl = STL(train, period=m, robust=True)
    res = stl.fit()
    seasonal = res.seasonal
    remainder = train - seasonal
    try:
        fc_rem = fc_sarima(remainder, horizon, m=1)
        fc_seas = fc_seasonal_naive(seasonal, horizon, m)
        return fc_rem + fc_seas
    except Exception:
        return fc_seasonal_naive(train, horizon, m)

# --- Theta ---
def fc_theta(train: pd.Series, horizon: int, m: int = 12) -> np.ndarray:
    if THETA_AVAILABLE:
        try:
            tm = ThetaModel(train, period=m)
            res = tm.fit()
            return res.forecast(horizon).values
        except Exception:
            pass
    # fallback: tendencia lineal simple (drift)
    try:
        x = np.arange(len(train))
        coef = np.polyfit(x, train.values, 1)
        last = train.iloc[-1]
        drift = coef[0]
        return np.array([last + (i+1)*drift for i in range(horizon)])
    except Exception:
        return fc_naive(train, horizon)

# --- DHR (Fourier) + ARIMA ---
def fourier_series(index: pd.DatetimeIndex, m: int, K: int) -> pd.DataFrame:
    t = np.arange(1, len(index)+1)

```



```

data = {}
for k in range(1, K+1):
    data[f'sin_{k}'] = np.sin(2*np.pi*k*t/m)
    data[f'cos_{k}'] = np.cos(2*np.pi*k*t/m)
return pd.DataFrame(data, index=index)

def fc_dhr_arima(train: pd.Series, horizon: int, m: int = 12, max_K: int = 3) -> np.ndarray:
    best_bic = np.inf
    best_fc = None
    for K in range(1, max_K+1):
        try:
            exog = fourier_series(train.index, m, K)
            future_index = pd.date_range(train.index[-1] + to_offset("MS"), periods=horizon,
freq="MS")
            exog_fut = fourier_series(future_index, m, K)
            model = SARIMAX(train, order=(0,0,0), seasonal_order=(0,0,0,0),
                exog=exog, trend="c",
                enforce_stationarity=False, enforce_invertibility=False)
            res = model.fit(dispatch=False)
            bic = res.bic
            if bic < best_bic:
                best_bic = bic
                best_fc = res.predict(start=len(train), end=len(train)+horizon-1,
exog=exog_fut).values
        except Exception:
            continue
    return best_fc if best_fc is not None else fc_seasonal_naive(train, horizon, m)

# --- Intermitentes: Croston/SBA/TSB ---
def _croston(train: np.ndarray, horizon: int, alpha: float = 0.1, variant: str = "croston") ->
np.ndarray:
    y = np.asarray(train, dtype=float)
    if np.all(y == 0):
        return np.zeros(horizon)

    demand = []
    intervals = []
    last_time = 0
    for val in y:
        last_time += 1
        if val > 0:

```




```

        demand.append(val)
        intervals.append(last_time)
        last_time = 0
    if len(demand) == 0:
        return np.zeros(horizon)

    # TSB: z (demanda) + p_occ (prob. ocurrencia)
    if variant == "tsb":
        z = demand[0]
        p_occ = 1.0
        for x in y[1:]:
            occ = 1.0 if x > 0 else 0.0
            z = alpha * (x if x > 0 else z) + (1 - alpha) * z
            p_occ = alpha * occ + (1 - alpha) * p_occ
        return np.repeat(z * p_occ, horizon)

    # Croston/SBA
    z = demand[0]
    p = intervals[0]
    for d in demand[1:]:
        z = alpha * d + (1 - alpha) * z
    for inter in intervals[1:]:
        p = alpha * inter + (1 - alpha) * p
    if p == 0:
        p = 1.0
    fc = np.repeat(z / p, horizon)
    if variant == "sba":
        fc *= (1 - alpha / 2.0)
    return fc

def fc_intermitente(train: pd.Series, horizon: int) -> np.ndarray:
    methods = ["croston", "sba", "tsb"]
    best = None
    best_mase = np.inf
    for mth in methods:
        try:
            def f(t, h):
                return _croston(t.values, h, alpha=0.1, variant=mth)
            res = rolling_origin_cv(train, horizon=3, n_splits=2, fit_forecast_func=f, m_season=12)
            if res.mase < best_mase:
                best_mase = res.mase

```



```

        best = mth
    except Exception:
        continue
    best = best or "sba"
    return _croston(train.values, horizon, alpha=0.1, variant=best)

# --- ML con lags ---
def _make_supervised(df_y: pd.Series, max_lag: int = 12, m: int = 12) -> Tuple[pd.DataFrame,
pd.Series, pd.DatetimeIndex]:
    y = df_y.copy()
    X = pd.DataFrame(index=y.index)
    for lag in range(1, max_lag + 1):
        X[f"lag_{lag}"] = y.shift(lag)
    for k in range(1, 4): # Fourier k=1..3
        X[f"sin_{k}"] = np.sin(2*np.pi*k*np.arange(len(y))/m)
        X[f"cos_{k}"] = np.cos(2*np.pi*k*np.arange(len(y))/m)
    X = X.dropna()
    y_tr = y.loc[X.index]
    return X, y_tr, y.index

def _iterative_forecast(model, last_values: pd.Series, horizon: int, m: int = 12) -> np.ndarray:
    vals = last_values.values.copy()
    idx = last_values.index
    out = []
    for _ in range(horizon):
        series = pd.Series(vals, index=idx)
        Xh, _, _ = _make_supervised(series, max_lag=12, m=m)
        x_new = Xh.iloc[-1:].values
        yhat = float(model.predict(x_new))
        out.append(yhat)
        next_date = idx[-1] + to_offset("MS")
        idx = idx.append(pd.DatetimeIndex([next_date]))
        vals = np.append(vals, yhat)
    return np.array(out)

def fc_elastic_net(train: pd.Series, horizon: int, m: int = 12) -> np.ndarray:
    if len(train) < 24:
        return fc_seasonal_naive(train, horizon, m)
    X, y, _ = _make_supervised(train, max_lag=12, m=m)
    if len(y) < 12:
        return fc_seasonal_naive(train, horizon, m)

```



```

pipe = Pipeline([
    ("scaler", StandardScaler(with_mean=True, with_std=True)),
    ("model", ElasticNet(alpha=0.1, l1_ratio=0.5, max_iter=5000, random_state=42))
])
pipe.fit(X, y)
return _iterative_forecast(pipe, train, horizon, m=m)

def fc_gbr(train: pd.Series, horizon: int, m: int = 12) -> np.ndarray:
    if len(train) < 24:
        return fc_seasonal_naive(train, horizon, m)
    X, y, _ = _make_supervised(train, max_lag=12, m=m)
    if len(y) < 12:
        return fc_seasonal_naive(train, horizon, m)
    gbr = GradientBoostingRegressor(
        n_estimators=500, learning_rate=0.05, max_depth=3, subsample=0.8, random_state=42
    )
    gbr.fit(X, y)
    return _iterative_forecast(gbr, train, horizon, m=m)

# =====
# Selector y orquestación
# =====

MODEL_FUNCS = {
    "naive": lambda s, h, m: fc_naive(s, h),
    "seasonal_naive": lambda s, h, m: fc_seasonal_naive(s, h, m),
    "moving_average": lambda s, h, m: fc_moving_average(s, h, window=3),
    "ets": lambda s, h, m: fc_ets(s, h, m),
    "sarima": lambda s, h, m: fc_sarima(s, h, m),
    "stl_arima": lambda s, h, m: fc_stl_arima(s, h, m),
    "theta": lambda s, h, m: fc_theta(s, h, m),
    "dhr_fourier_arima": lambda s, h, m: fc_dhr_arima(s, h, m, max_K=3),
    "croston_family": lambda s, h, m: fc_intermitente(s, h),
    "elastic_net": lambda s, h, m: fc_elastic_net(s, h, m),
    "gbr": lambda s, h, m: fc_gbr(s, h, m),
}

def select_and_forecast(series: pd.Series, horizon: int = 3, m: int = 12) -> Tuple[str, Dict,
np.ndarray, Dict[str, float]]:
    y = series.dropna().astype(float)
    if len(y) < 18:

```



```

    candidates = ["naive", "seasonal_naive", "moving_average"]
else:
    candidates = list(MODEL_FUNCS.keys())

results = []
for name in candidates:
    func = MODEL_FUNCS[name]
    try:
        res_cv = rolling_origin_cv(
            y, horizon=horizon, n_splits=3,
            fit_forecast_func=lambda tr, h: func(tr, h, m),
            m_season=m
        )
        results.append((name, res_cv))
    except Exception:
        continue

if not results:
    fc = fc_seasonal_naive(y, horizon, m)
    return "seasonal_naive", {"fallback": True}, fc, {"mase": np.nan, "smape": np.nan}

results.sort(key=lambda t: (t[1].mase, t[1].smape))
name, res_cv = results[0]
details = res_cv.details
fc = MODEL_FUNCS[name](y, horizon, m)
metrics = {"mase": float(res_cv.mase), "smape": float(res_cv.smape)}
return name, details, np.asarray(fc), metrics

# =====
# I/O: leer y procesar
# =====

def load_sheet(filepath: str, sheet_name: str) -> pd.DataFrame:
    df = pd.read_excel(filepath, sheet_name=sheet_name, header=0)
    first_col = df.columns[0]

    # 1) Parseo robusto de fechas:
    # - Si vienen como número (serial de Excel), las convierto desde 1899-12-30
    # - Si vienen como texto, parseo con dayfirst y errores='coerce'
    if pd.api.types.is_numeric_dtype(df[first_col]):
        # Excel date serial -> pandas datetime

```



```

df[first_col] = pd.to_datetime("1899-12-30") + pd.to_timedelta(df[first_col], unit="D")
else:
    df[first_col] = pd.to_datetime(df[first_col], format="%d/%m/%Y", dayfirst=True,
errors="coerce")

# 2) Limpieza básica
df = df.dropna(subset=[first_col])
df = df.set_index(first_col).sort_index()

# 3) Normalizo a mensual (primer día del mes)
df.index = pd.DatetimeIndex(df.index).to_period("M").to_timestamp(how="start")

# 4) Aseguro numérico en las cuentas (convierte strings a números si hiciera falta)
for c in df.columns:
    df[c] = pd.to_numeric(df[c], errors="coerce")

return df

def forecast_sheet(df: pd.DataFrame, horizon: int = 3, m: int = 12) -> Tuple[pd.DataFrame,
pd.DataFrame]:
    last_date = df.index.max()
    future_index = pd.date_range(last_date + to_offset("MS"), periods=horizon, freq="MS")
    forecasts = {}
    rows_report = []
    for col in df.columns:
        series = df[col].astype(float)
        try:
            name, details, fc, metrics = select_and_forecast(series, horizon=horizon, m=m)
        except Exception as e:
            name, details, fc, metrics = "seasonal_naive", {"fallback": str(e)},
fc_seasonal_naive(series, horizon, m), {"mase": np.nan, "smape": np.nan}
        forecasts[col] = fc
        rows_report.append({
            "cuenta": col,
            "modelo": name,
            "mase": metrics.get("mase", np.nan),
            "smape": metrics.get("smape", np.nan),
            "detalles": json.dumps(details, ensure_ascii=False)
        })
    df_fore = pd.DataFrame(forecasts, index=future_index)

```

```

df_report
pd.DataFrame(rows_report).sort_values(["modelo","cuenta"]).reset_index(drop=True)
return df_fore, df_report

# =====
# EJECUCIÓN EN NOTEBOOK
# =====

sheets_to_run = []
for sh in ("SPP", "WEB"):
    try:
        _ = pd.read_excel(INPUT_PATH, sheet_name=sh, nrows=1)
        sheets_to_run.append(sh)
    except Exception:
        pass

if not sheets_to_run:
    raise FileNotFoundError("No se encontraron las hojas 'SPP' o 'WEB' en el archivo de entrada.
Verificá INPUT_PATH y los nombres de hoja.")

all_forecasts = {}
all_reports = {}

for sh in sheets_to_run:
    df = load_sheet(INPUT_PATH, sh)
    df_fore, df_rep = forecast_sheet(df, horizon=3, m=12)
    all_forecasts[sh] = df_fore
    all_reports[sh] = df_rep

# Guardar a Excel
with pd.ExcelWriter("forecasts.xlsx", engine="xlsxwriter") as w:
    for sh, dff in all_forecasts.items():
        dff.to_excel(w, sheet_name=sh)

with pd.ExcelWriter("report_modelos.xlsx", engine="xlsxwriter") as w:
    for sh, dfr in all_reports.items():
        dfr.to_excel(w, sheet_name=sh, index=False)

# Mostrar resumen en pantalla
from IPython.display import display, Markdown
display(Markdown("### Proyecciones generadas (primeras filas)"))

```




```
for sh in sheets_to_run:
    display(Markdown(f"Hoja {sh} – forecasts (3 meses futuros):"))
    display(all_forecasts[sh].head())

display(Markdown("### Reporte de modelos elegidos (primeras filas)"))
for sh in sheets_to_run:
    display(Markdown(f"Hoja {sh} – reporte de selección:"))
    display(all_reports[sh].head())

print("Archivos guardados en el directorio de trabajo:")
print("- forecasts.xlsx")
print("- report_modelos.xlsx")
```

Apéndice V – Compendio explicativo del código

Propósito: Automatizar proyecciones mensuales para múltiples cuentas a partir de datos históricos en Excel. El sistema prueba varias familias de modelos, valida su desempeño y elige la mejor por columna, generando:

forecasts.xlsx: 3 meses futuros por hoja (SPP/WEB) y por cuenta.

report_modelos.xlsx: detalle de la familia elegida con MASE y sMAPE.

1) Qué necesita y qué produce

Entrada (Excel):

Hojas: SPP y WEB (si una falta, procesa la que esté).

Columna 1 = Fecha mensual (texto tipo dd/mm/yyyy o número de fecha de Excel).

Columnas siguientes = Cuentas (valores numéricos por mes).

Salida (Excel):

forecasts.xlsx → Filas: meses futuros; Columnas: cuentas; una pestaña por hoja.

report_modelos.xlsx → Columnas: hoja, cuenta, modelo, MASE, sMAPE, splits_cv.

2) Flujo general (visión paso a paso)

Carga y normalización de datos: convierte la primera columna a fecha (inicio de mes), ordena cronológicamente y fuerza el resto a numérico.

Bucle por cuenta: toma cada columna (serie mensual) y arma una validación temporal (rolling origin) con CV_SPLITS ensayos.

Competencia de modelos: ejecuta 9 familias y calcula el error medio (MASE y sMAPE) en cada split.

Selección: el ranking prioriza menor MASE (comparabilidad entre series); usa sMAPE como criterio secundario.

Ajuste final y proyección: re-entrena la familia ganadora con toda la historia y proyecta HORIZON = 3 meses (configurable).

Salida: apila las proyecciones en forecasts.xlsx (una hoja por origen SPP/WEB), guarda la "trazabilidad" del modelo en report_modelos.xlsx.

3) Validación y métricas

Validación temporal (rolling origin): simula varios escenarios de "entrenamiento en pasado → predigo los meses que siguen". Evita la trampa de evaluar con datos ya vistos.

MASE (Mean Absolute Scaled Error): compara contra un baseline naïve; valores <1 suelen implicar mejora sobre lo trivial.

sMAPE (Symmetric MAPE): error porcentual "equilibrado" (0% ideal). Útil para interpretar magnitudes.

Regla práctica: sMAPE <10–15% + MASE <1.0–1.2 = rendimiento razonable (depende de volatilidad y cambios estructurales).

4) Familias de modelos

Baselines

Promedio móvil: suaviza ruido; bueno con series estables sin estacionalidad marcada.

Naïve estacional: repite el valor del mismo mes del año previo; útil si hay estacionalidad anual fuerte.

Drift (tendencia lineal): prolonga una pendiente promedio; útil cuando hay crecimiento/caída sostenidos.

Clásicos de series temporales

4) ETS (Holt-Winters): separa nivel+tendencia+estacionalidad; funciona bien con patrones regulares. 5) SARIMA: muy flexible para autocorrelación y estacionalidad; rinde con patrones complejos y suficiente historia. 6) STL + ARIMA: primero extrae la estacionalidad (STL), luego modela el resto; útil si el componente estacional es claro pero el resto es irregular. 7) Theta: robusto y simple para series suaves; buen desempeño generalista. 8) Fourier + ARIMA: usa armónicos (seno/coseno) como variables para estacionalidad compleja.

Demanda intermitente

9) Croston / SBA / TSB: especiales para series con muchos ceros y eventos esporádicos.

Nota: En algunas variantes del cuaderno pueden estar disponibles modelos de ML con lags (p. ej., Elastic Net/GBR) para capturar no linealidades. Se los usa si están habilitados.

5) Lógica de selección (qué decide el “mejor”)

Para cada familia, corre la validación y promedia MASE y sMAPE.

Ranking principal por MASE (porque es comparable entre series).

Desempate por sMAPE (preferimos menor).

Si un modelo falla por poca historia o convergencia, el sistema lo descarta y prioriza los que devuelven métricas válidas.

Con la ganadora, se re-entrena con todos los datos y se calculan los 3 meses futuros.

6) Parámetros clave

Arriba del notebook se definen constantes simples:

```
INPUT_PATH = "C:/carpeta/tu_archivo.xlsx" # ruta al Excel de entrada
HORIZON = 3      # meses a proyectar (p. ej., 6 o 12)
FAST_MODE = True  # True = pruebas más chicas (más rápido)
CV_SPLITS = 3     # cantidad de ensayos en validación temporal
SARIMAX_MAXITER = 200
```

También se fija $m = 12$ para estacionalidad anual (series mensuales). Cambiarlo sólo si tu frecuencia fuera distinta.

7) Lectura y limpieza (qué hace “bajo el capó”)

Fechas: detecta si vienen como número de Excel (serie 1900) o como texto; en ambos casos las convierte al primer día del mes.

Orden: asegura orden cronológico y elimina filas totalmente vacías.

Numéricos: fuerza las cuentas a tipo numérico (NaN si hay celdas no parseables), y deja que cada modelo maneje los faltantes según su naturaleza.

8) Interpretación de las salidas

forecasts.xlsx

Filas: meses proyectados (p. ej., 2025-10, 2025-11, 2025-12).

Columnas: una por cuenta. Valores = pronósticos puntuales.

report_modelos.xlsx

modelo: la familia ganadora para esa cuenta (p. ej., SARIMA, ETS, Croston).

MASE / sMAPE: desempeño en validación; guía para confiabilidad relativa.

splits_cv: cuántos intentos de validación se pudieron ejecutar (si es bajo, tomá los resultados con cautela).

Práctica recomendada: filtrar en este reporte las cuentas materialmente relevantes y anexarlas al informe principal como evidencia de la elección.

9) Límites, alertas y buenas prácticas

Historia corta (<12–24m): limita los modelos estacionales; el sistema puede caer en baselines. Cambios de régimen (ej.: modificaciones de aranceles, eventos únicos): considerar segmentar la serie o añadir contexto en el informe.

Outliers: no siempre conviene “limpiar”; a veces son reales y explican la dinámica. Documentarlos.

Horizonte largo: a mayor horizonte, mayor incertidumbre; evitar decisiones críticas sólo con puntos muy lejanos.

Apéndice VI – Instrucciones del bot GPT

ROL

Eres un BOT DE INVESTIGACIÓN Y NORMALIZACIÓN que NAVEGA LA WEB. Con UN mensaje del usuario debes: (1) buscar en fuentes confiables, (2) obtener valores de t0 (mes objetivo) y t-1, t-2 para 12 indicadores, (3) normalizarlos, y (4) devolver PRIMERO una tabla Markdown “Indicador | t0 | t-1 | t-2”; DESPUÉS un bloque “Fuentes y métodos” con URLs/fechas/flags. No hagas preguntas.

PARÁMETROS (por defecto; el usuario puede sobrescribir)

- Geografía: Argentina. Idioma: ES (acepta EN/PT si la fuente es oficial/reputada).
- t0: AAAA-MM (si no se indica, último mes con dato disponible). t-1/t-2: dos meses previos.
- Moneda: ARS. Formato: coma decimal, sin miles, 2 decimales. Tasas en 0–100 (sin “%”).

JERARQUÍA DE FUENTES

- 1) Oficiales (INDEC, BCRA, AFIP, Sec. Energía, DNV/Ministerios)
- 2) Entidades (FADEEAC/ICL, CACE, UIA, UTDT)
- 3) Especializadas/academia
- 4) Prensa económica reputada (evitar blogs)

INDICADORES (orden obligatorio en salida)

- 1) IPC INDEC (m/m, %)
- 2) IPIM INDEC (m/m, %)
- 3) Tipo de cambio BCRA A3500 (promedio mensual, ARS/USD)

- 4) Tasa de política BCRA (TNA, %)
- 5) BADLAR bancos privados (promedio mensual, TNA, %)
- 6) IPI INDEC manufacturero (m/m desest., %)
- 7) EMAE INDEC (m/m desest., %)
- 8) ICL FADEEAC (m/m, %)
- 9) Precio promedio nacional de gasoil (ARS/L)
- 10) Ventas en supermercados INDEC (m/m desest., %)
- 11) Índice de Confianza del Consumidor (UTDT) – m/m % (o nivel)
- 12) Recaudación tributaria AFIP (YoY, %)

PATRONES “site:” (usar antes de búsquedas generales)

- IPC/IPIM/EMAE/IPI/Super: site:indec.gob.ar "Informe técnico" + mes/año
- A3500/TPM/BADLAR: site:bcra.gob.ar + “tipo de cambio de referencia” / “tasa de política” / “BADLAR”
- ICL: site:fadeeac.org.ar “Índice de Costos Logísticos”
- Gasoil: site:energia.gob.ar “boletín” “precios combustibles”
- ICC: site:utdt.edu “Índice de Confianza del Consumidor”
- AFIP: site:afip.gob.ar “recaudación tributaria” informe mensual

FORMATO / NORMALIZACIÓN

- Tabla final: 4 columnas, 12 filas, en el orden indicado. Coma decimal, 2 decimales, sin símbolos.
- Tasas/variaciones: 0–100. Niveles monetarios (A3500, gasoil): ARS. Índices: si sólo hay nivel, devolver nivel (2 dec.) y aclarar.

GUARDIA DE CALENDARIO

Si t0 pedido aún no está publicado para una serie, usar el último mes disponible de esa serie y declararlo en “Fuentes y métodos”.

DERIVACIONES PERMITIDAS (para reducir NA)

- EMAE/IPI/Super: si no hay % m/m pero sí niveles desest., calcular $MoM = ((Nivel_t / Nivel_{t-1}) - 1) * 100$.
- TNA desde TEA: $TNA_m = ((1 + TEA)^{(1/12)} - 1) * 100$ (indicar conversión).
- A3500: si no hay “promedio mensual” explícito, promediar cierres diarios oficiales del mes.
- Gasoil: si no hay promedio nacional publicado, usar boletín por jurisdicción y promediar (simple o ponderado si la fuente da ponderadores).

FRECUENCIA > MENSUAL

Si una serie es bi/tri-mensual, mapear al mes de cierre más cercano (ej.: jul-ago → ago) y explicarlo.



PDF/TABLERO

Si el dato está en PDF/tablero, localizar cuadro/mes, leer número, normalizar decimales, citar cuadro/página.

ESTIMACIÓN CONTROLADA (OBS > DER > EST)

- Prioridad: OBS (dato directo) > DER (cálculo exacto desde serie oficial) > EST (nowcast con método).
- Cada celda lleva FLAG interno (no en la tabla): OBS | DER | EST-<método> (reportar en "Fuentes y métodos").
- No usar prensa no verificada si existe oficial. No inventar sin método.

REGLAS DE ESTIMACIÓN (usar solo si no hay OBS/DER)

- 1) IPC: EST-REM (REM BCRA de t0); si no, EST-AVG (promedio últimos 3 m/m). Recorte a [0;30] si excede.
- 2) IPIM: EST-RAZÓN = (promedio 6m de IPIM/IPC)*IPC_est_t0; o EST-AVG (3m).
- 3) A3500: EST-PARCIAL (promedio de cierres diarios disponibles en t0); si cerró y falta promedio, promediar todos (DER).
- 4) Tasa política: EST-CF (último valor vigente, máx. 1 mes); si hubo cambio en t0, usar post-cambio.
- 5) BADLAR: EST-AVG (promedio últimos 10 días hábiles de t0) o EST-CF (1 mes).
- 6) IPI: DER desde niveles desest.; si falta, EST-AVG (3 m/m) o EST-LINEAL (interpolación).
- 7) EMAE: igual que IPI.
- 8) ICL: EST-CF (1 mes) o EST-COMP (promedio simple/ponderado de subrubros si hay).
- 9) Gasoil: EST-AVISO (aplicar % de aumentos oficiales al último nivel); o EST-PROM (promedio provincial→nacional).
- 10) Supermercados: DER desde niveles; si no, EST-PROXY (IPC Alimentos m/m) o EST-AVG (3 m/m).
- 11) ICC: si hay nivel, $DER-VAR = ((Nivel_t / Nivel_t-1) - 1) * 100$; si no hay publicación, EST-CF (1 mes) con nota.
- 12) Recaudación AFIP: OBS YoY; si no, EST-PRELIM (adelanto oficial); o EST-AVG (3 YoY).

VALIDACIONES

- Rechazar imposibles (tasa>1000, precios negativos): entonces NA y explicar.
- Coherencia temporal: t-1 y t-2 son los dos meses previos a t0 (ajustar con guardia si la serie está rezagada).
- Documentar siempre URL y fecha por celda (en "Fuentes y métodos").

SALIDA (dos bloques, en este orden)

1) Tabla Markdown EXACTA:

```
| Indicador | t0 | t-1 | t-2 |
|---|---|---|---|
```




IPC INDEC (variación mensual, %)	<n/NA>	<n/NA>	<n/NA>
IPIM INDEC (variación mensual, %)	<n/NA>	<n/NA>	<n/NA>
Tipo de cambio BCRA A3500 (promedio mensual, ARS/USD)	<n/NA>	<n/NA>	<n/NA>
Tasa de política BCRA (TNA, %)	<n/NA>	<n/NA>	<n/NA>
BADLAR bancos privados (promedio mensual, TNA, %)	<n/NA>	<n/NA>	<n/NA>
IPI INDEC manufacturero (m/m de sest., %)	<n/NA>	<n/NA>	<n/NA>
EMAE INDEC (m/m de sest., %)	<n/NA>	<n/NA>	<n/NA>
ICL FADEEAC (m/m, %)	<n/NA>	<n/NA>	<n/NA>
Precio promedio nacional de gasoil (ARS/L)	<n/NA>	<n/NA>	<n/NA>
Ventas en supermercados INDEC (m/m de sest., %)	<n/NA>	<n/NA>	<n/NA>
Índice de Confianza del Consumidor (UTDT) – m/m % (o nivel)	<n/NA>	<n/NA>	<n/NA>
Recaudación AFIP (YoY, %)	<n/NA>	<n/NA>	<n/NA>

2) “Fuentes y métodos” (viñetas por indicador; por cada mes: URL, fecha, FLAG=OBS/DER/EST-XXX y breve nota).

INTERACCIÓN

- El usuario dirá: “Actualizar AAAA-MM , Argentina”. Responde sin pedir aclaraciones con la tabla y luego “Fuentes y métodos”.

Apéndice VII - Afecciones matemáticas concretas

Notación:

Cuenta_t = valor calculado; Base = valor en “01_Python_Base”; Ind = valor indicador (en % salvo que se aclare); Lag en meses.

Modo Multiplicativo: $\text{Cuenta}_t = \text{Base} * (1 + \beta * \text{Ind}/100)$

Modo Aditivo: suma directa (se indica fórmula).

1) IPC INDEC (m/m, %)

Cuentas: ALIN 0,40 (Lag 1); ENEL 0,30 (0); SGLC 0,40 (1); SGRS 0,40 (1); SVLP 0,40 (1); HNRR 0,30 (1); GTVR 0,30 (0); SLDS 0,15 (1); CGSC 0,15 (1); SAC 0,10 (1).

Fórmula tipo: $\text{ALIN}_t = \text{ALIN}_{\text{base}} * (1 + 0,40 * \text{IPC}/100)$.

2) IPIM INDEC (m/m, %)

Cuentas: Mantenimiento/Repuestos (si está dentro de GTVR) 0,70 (Lag 1); CBLC 0,25 (0) [por insumos no precio litro]; ALQC 0,20 (1); ALCM 0,20 (1).

Fórmula tipo: $\text{GTVR}_t = \text{GTVR}_{\text{base}} * (1 + 0,70 * \text{IPIM}/100)$.

3) Tipo de cambio BCRA A3500 (prom. mensual, ARS/USD)

Cuentas (exposición USD parcial): Repuestos importados (GTVR) $\beta=0,70$ (Lag 1) multiplicativo sobre $\Delta\%$ A3500; ALQC/ALCM $\beta=0,30$ (1); HNRR $\beta=0,20$ (1).

4) Tasa de política monetaria (TNA, %)

Cuenta: ITFN (Intereses).

Fórmula (recomendada): $ITFN_t = Saldo_promedio_t * (TPM_t / 100) / 12$.

5) BADLAR bancos privados (TNA, %)

Cuenta: ITFN (comercial).

Fórmula: $ITFN_t = Saldo_comercial_t * (BADLAR_t / 100) / 12$.

6) IPI INDEC (m/m desest., %)

Cuentas: VTAS (ingresos troncal) $\beta=0,60$ (Lag 1); ALQC $\beta=0,40$ (1); ALCM $\beta=0,30$ (1); CBLC $\beta=0,35$ (1).

7) EMAE INDEC (m/m desest., %)

Cuentas: VTAS total $\beta=0,40$ (Lag 1); ALQC 0,25 (1); ALCM 0,20 (1); CBLC 0,20 (1).

8) ICL FADEEAC (m/m, %)

Cuentas: CBLC $\beta=0,80$ (Lag 0); ALQC $\beta=0,60$ (1); ALCM $\beta=0,50$ (1); SLDS $\beta=0,20$ (1) (por paritarias dentro del índice); GTVR $\beta=0,20$ (1).

9) Precio promedio nacional de gasoil (ARS/L)

Cuenta: CBLC (combustible).

Fórmula aditiva directa: $CBLC_t = Litros_t * PrecioGasoil_t$.

10) Ventas en supermercados INDEC (m/m desest., %)

Cuentas (capilar/LM): VTAS LM $\beta=0,50$ (Lag 1); ALQC $\beta=0,35$ (1); ALCM $\beta=0,25$ (1); CBLC $\beta=0,25$ (1); DDI (si aplica postal por pieza) $\beta=0,40$ (1).

11) Índice de Confianza del Consumidor (m/m % o nivel)

Cuentas (señal adelantada): VTAS total $\beta=0,25$ (Lag -1 a 0); ALQC 0,15 (0); ALCM 0,10 (0).

12) Recaudación AFIP (YoY, %)

Cuentas: Días de cobro (KPI) $\beta=-0,20$ (Lag 1) \rightarrow DiasCobro_t = DiasCobro_base * (1 - 0,20 * AFIP_YoY/100);