Maestría en

ADMINISTRACIÓN DE EMPRESAS

Tesis previa a la obtención de título de Magister en Administración de Empresas

AUTORES:

Lic. Pablo Alberto Baca Sosa

Econ. Mario Antonio Garrido Valero

Ing. Christian Andres Martinez de la Vega Boada

TUTOR: José Luis Pérez Galán

Comunicación Digital y Fidelización con Dependientes de Farmacia

PROYECTO DE TITULACIÓN – FIN DE MÁSTER
MAESTRÍA EN ADMINISTRACIÓN DE EMPRESA ENL

REINVENTEMOS

Comunicación Digital y Fidelización con Dependientes de Farmacia

Por

Autores

Lic. Pablo Alberto Baca Sosa

Econ. Mario Antonio Garrido Valero

Ing. Christian Andres Martinez de la Vega Boada

Octubre 2026

Aprobado



Cristian Melo
Presidente(a) del Tribunal
Universidad Internacional del Ecuador

Yo, Cristian Javier Melo González e Ignacio Maroto, declaramos que, personalmente conocemos que los graduandos: Lic. Pablo Alberto Baca Sosa, Econ. Mario Antonio Garrido Valero, y Ing. Christian Andres Martinez de la Vega Boada, son los autores exclusivos de la presente investigación y que ésta es original, auténtica y personal de ellos.



Cristian Melo
Coordinador MBA UIDE



Ignacio Maroto
Provost WBS

Autoría del Trabajo de Titulación

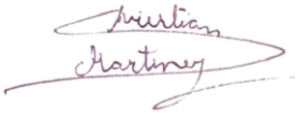
Nosotros, Lic. Pablo Alberto Baca Sosa, Econ. Mario Antonio Garrido Valero, y Ing. Christian Andres Martinez de la Vega Boada, declaramos bajo juramento que el trabajo de titulación titulado **Comunicación Digital y Fidelización con Dependientes de Farmacia** es de nuestra autoría y exclusiva responsabilidad legal y académica; que no ha sido presentado anteriormente para ningún grado o calificación profesional, habiéndose citado las fuentes correspondientes y respetando las disposiciones legales que protegen los derechos de autor vigentes.





Lic. Pablo Alberto Baca Sosa
Correo electrónico: pabacaso@uide.edu.ec

Econ. Mario Antonio Garrido Valero
Correo electrónico: magarridova@uide.edu.ec



Ing. Christian Andres Martinez de la Vega Boada
Correo electrónico: chmartinezdelavegabo@uide.edu.ec

Autorización de Derechos de Propiedad Intelectual

Nosotros, Lic. Pablo Alberto Baca Sosa, Econ. Mario Antonio Garrido Valero, y Ing. Christian Andres Martinez de la Vega Boada, en calidad de autores del trabajo de investigación titulado Comunicación Digital y Fidelización con Dependientes de Farmacia, autorizo a la Universidad Internacional del Ecuador (UIDE) para hacer uso de todos los contenidos que me pertenecen o de parte de los que contiene esta obra, con fines estrictamente académicos o de investigación. Los derechos que como autor me corresponden, lo establecido en los artículos 5, 6, 8, 19 y demás pertinentes de la Ley de Propiedad Intelectual y su Reglamento en Ecuador.

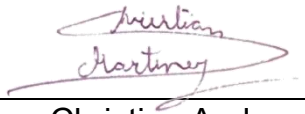
D. M. Quito, octubre de 2026





Lic. Pablo Alberto Baca Sosa
Correo electrónico: pabacaso@uide.edu.ec

Econ. Mario Antonio Garrido Valero
Correo electrónico: magarridova@uide.edu.ec



Ing. Christian Andres Martinez de la Vega Boada
Correo electrónico: chmartinezdelavegabo@uide.edu.ec

Dedicatorias y Agradecimientos

Agradecemos a la Universidad Internacional del Ecuador y a Westfield Business School por la formación recibida a lo largo del programa; al tutor, José Luis Pérez Galán, por sus guías y comentarios durante las diez semanas de trabajo; y a El Laboratorio y a su equipo comercial, por abrir la información que hizo posible este proyecto. También queremos extender nuestro sincero reconocimiento al trabajo y liderazgo de Pablo Baca durante todo el proceso de aprendizaje, es quien con su guía nos ha traído hasta este punto.

ÍNDICE DEL DOCUMENTO

INTRODUCCIÓN Y OBJETO DEL PROYECTO **11**
Contexto de la empresa 11
Problema 11
Objetivo del proyecto 11
Alcance y condiciones del piloto 12
Metodología y estructura del documento 12
PARTE I: BUSINESS INTELLIGENCE **13**
CAPÍTULO 1. Introducción al proyecto de Business Intelligence 14
1. Presentación de la empresa 14
2. Identificación de los objetivos iniciales del proyecto: definición inicial del problema 14
3. Áreas y personas que demandan la información y uso actual 15
4. Objetivos de cada área con la información actual y necesidades detectadas 15
5. Definición de los KPI's relevantes 16
6. Fuentes de conocimiento sobre la empresa y nivel de conocimiento 17
7. Consideraciones adicionales para la ejecución del proyecto 17
CAPÍTULO 2. Fuentes de datos del proyecto 19
Objetivo y alcance del capítulo 19
Proceso de consolidación de información 19
Definición de estructura y gobierno de los datos 20
Tipos de datos relevantes para el proyecto 20
Proceso de transformación de la información 20
Necesidades específicas identificadas para el proyecto 21
Conclusiones del capítulo 21
CAPÍTULO 3. Procesos ETL del proyecto 22
1. Procesos de carga de datos (ETL) 22
2. Información y calidad de la información disponible 22
3. Horarios de carga y accesibilidad 23
4. Variables disponibles y acceso al Data Warehouse 23
5. Sistemas implicados en los procesos 23
6. Recursos de IT o externos empleados 24
7. Necesidad futura: ETL de la base de contacto de dependientes 24
8. Manejo de excepciones y retención de la información 24
Conclusiones del capítulo 25
CAPÍTULO 4. Visualización de la información 26
Premisas de entrega de información 26
1. Análisis de herramientas y selección para el ecosistema del proyecto 26
2. Selección de usuarios y perfiles de uso del tablero 27
3. Necesidades de información y accesos por departamento y perfil 28
Diseño conceptual del tablero: Cobertura y Fidelización de Dependientes 28
4. Revisión y cumplimiento de los objetivos iniciales 29
Conclusiones del capítulo 30
CAPÍTULO 5. Business case y rol de datos 31
Posicionamiento de la función de datos dentro de El Laboratorio 31
Objetivo principal de la función propuesta 31
Objetivo del rol por responsabilidad 31
Impactos esperados en los resultados del proyecto 32
Business case: costos y beneficios 32
Conclusiones del capítulo 34

PARTE II: BIG DATA E INTELIGENCIA ARTIFICIAL **35**
CAPÍTULO 1. Obtención de datos no estructurados 36
Objetivo y alcance del capítulo 36
Compra de datos a entidades externas 36
Datos generados de manera desorganizada por la organización 37
Solución para integrar la información a modelos estructurados 37
Casos de uso para el proyecto 38
Conclusiones del capítulo 38
CAPÍTULO 2. Arquitectura de la información 39
Objetivo y alcance del capítulo 39
1. Análisis de alternativas de proveedores 39
2. Selección de la alternativa de proveedor 39
3. Definición del modelo de cloud computing 40
Zona de aterrizaje ligera para datos no estructurados 40
4. Flujo de integración con el Data Warehouse 41
Conclusiones del capítulo 42
CAPÍTULO 3. Arquitectura open source 43
Objetivo y alcance del capítulo 43
1. Selección crítica del framework de Big Data 43
2. Lenguaje de programación complementario 44
3. Complementos del ecosistema open source 44
Conclusiones del capítulo 45
CAPÍTULO 4. Implementación de inteligencia artificial 46
Objetivo y alcance del capítulo 46
1. Objetivo de aplicación del modelo de Data Science / Machine Learning 46
2. Fuentes de datos utilizadas 47
3. Área de aplicación y departamentos de negocio implicados 47
4. Identificación de la mejora esperada 48
5. Cuantificación de la mejora: business case 48
6. Validación técnica del modelo 48
Conclusiones del capítulo 49
CAPÍTULO 5. Caso de uso en near real time 50
Objetivo y alcance del capítulo 50
1. Objetivo de aplicación del modelo de real time 50
2. Fuentes de datos utilizadas 53
3. Área de aplicación y departamentos de negocio implicados 54
4. Identificación de la mejora esperada 55
5. Cuantificación de la mejora: business case 55
Próximos pasos 58
Conclusiones del capítulo 59
CONCLUSIONES Y APLICACIONES **60**
Síntesis del proyecto 60
Principales hallazgos y decisiones 60
Aplicaciones 61
Limitaciones 61
Reflexión final 61
REFERENCIAS **62**

ÍNDICE DE FIGURAS Y TABLAS

Índice de figuras

Figura 1. Mockup conceptual del tablero propuesto — valores ilustrativos, no corresponden a datos reales del laboratorio. 29
Figura 2. Ubicación de la función de datos dentro de la estructura actual — no representa una nueva área formal. 31
Figura 3. Árbol de decisión del modelo de priorización en su etapa inicial — se ejecuta mensualmente sobre los datos ya existentes (el capítulo 5 de la Parte II lo amplía con fuentes nuevas). 46
Figura 4. Flujo del caso de uso en near real time, de la interacción del dependiente a la acción y al aprendizaje. 52
Figura 5. Tiempo hasta que el laboratorio detecta un problema en un punto de venta, hoy y con el caso de uso. 55

Índice de tablas

Tabla 1. Áreas y personas que demandan la información y su uso actual 15
Tabla 2. Objetivos de cada área y necesidades detectadas 15
Tabla 3. Indicadores de Conocimiento (K): resultado actual, meta y brecha 16
Tabla 4. Indicadores de Organización (O): resultado actual, meta y brecha 16
Tabla 5. Indicadores de Motivación (M): resultado actual, meta y brecha 17
Tabla 6. Tipos de datos relevantes para el proyecto 20
Tabla 7. Necesidades del proceso de transformación de la información 20
Tabla 8. Frecuencia, tipo de ejecución y acceso a la carga de datos 23
Tabla 9. Variables disponibles por origen de datos 23
Tabla 10. Sistemas implicados en los procesos 23
Tabla 11. Conceptos de reporte, indicador, tablero de control y consulta ad-hoc 26
Tabla 12. Herramientas de visualización y su rol en el proyecto 26
Tabla 13. Perfiles con acceso al tablero de control 27
Tabla 14. Necesidades de información y tipo de acceso por perfil 28
Tabla 15. Presupuesto estimado de la función de datos 32
Tabla 16. Ahorro operativo estimado con el tablero 33
Tabla 17. Fuentes externas de datos y su uso en el proyecto 36
Tabla 18. Fuentes internas desorganizadas con valor potencial 37
Tabla 19. Casos de uso de los datos no estructurados 38
Tabla 20. Comparación de proveedores de nube 39
Tabla 21. Capacidades de la zona de aterrizaje y servicios de Azure propuestos 40
Tabla 22. Criterios para seleccionar el framework de Big Data 43
Tabla 23. Justificación de Python como único lenguaje de programación 44
Tabla 24. Complementos del ecosistema open source seleccionados 44
Tabla 25. Fuentes de datos y variables del modelo de priorización 47
Tabla 26. Áreas implicadas y uso del modelo 47
Tabla 27. Clasificación de los flujos de datos según su ritmo 50
Tabla 28. Fases del sistema de alertas y enrutamiento 51
Tabla 29. Fuentes de datos del caso de uso y condiciones de gobierno 53
Tabla 30. Áreas implicadas y rol en el caso de uso 54
Tabla 31. Mejora esperada por indicador 55
Tabla 32. Escenarios de trabajo del piloto 55

Tabla 33. Resultado mensual esperado del piloto (150 puntos de venta) 56
Tabla 34. Regla de decisión al cierre del mes 4..... 57
Tabla 35. VAN a 3 años de la escala nacional..... 58

INTRODUCCIÓN Y OBJETO DEL PROYECTO

Contexto de la empresa

El Laboratorio es una compañía farmacéutica que comercializa sus productos a nivel nacional a través de las principales cadenas de farmacias del país. Su modelo comercial se apoya en dos frentes de relacionamiento: la visita médica, dirigida a profesionales de la salud, y la fuerza de representantes, que atiende 1,079 puntos de venta y a sus dependientes con un equipo de 8 personas. El dependiente de farmacia es un actor clave de la venta responsable de productos farmacéuticos, pero hoy el laboratorio solo llega a él mediante la visita presencial.

Problema

El problema de fondo es de alcance y de tiempo. La visita presencial se programa entre una y cinco veces al mes según la clasificación del punto de venta (A, B, C o D), pero la frecuencia real promedio es de 0.98 visitas por punto frente a una meta ponderada de 2.8, de modo que la capacidad de los representantes cubre apenas el 35% de la demanda de visitas definida por la propia política. Entre una visita y otra no existe divulgación científica continua; la fidelización se reduce a compromisos verbales que no se pueden medir; la rotación de dependientes es alta; y el laboratorio se entera tarde de lo que ocurre en el punto de venta, porque la visita llega, como máximo, una vez al mes en los puntos D y el drenaje que reportan las cadenas llega «a mes caído».

Objetivo del proyecto

El objetivo general es diseñar un canal de comunicación virtual, frecuente y personalizado por cadena, que complemente —sin reemplazar— la labor del representante, y construir sobre datos la capa de información que permita medirlo, priorizar la visita presencial y reaccionar casi en tiempo real a lo que ocurre en el punto de venta. Los objetivos específicos son:

- Diagnosticar las fuentes, los procesos de carga y la calidad de los datos disponibles, y definir los indicadores de Conocimiento, Organización y Motivación que miden la brecha a cerrar (Parte I, capítulos 1 a 3).
- Diseñar el tablero de control y el rol de datos que lo sostiene, con un business case que declare con honestidad sus costos y beneficios (Parte I, capítulos 4 y 5).
- Incorporar las fuentes de datos no estructuradas y definir una arquitectura en la nube y open source proporcionada al tamaño del piloto (Parte II, capítulos 1 a 3).
- Diseñar un modelo de priorización de cobertura con inteligencia artificial (Parte II, capítulo 4).
- Definir un caso de uso en near real time, con escenarios, punto de equilibrio y regla de decisión fijados antes de empezar (Parte II, capítulo 5).

Alcance y condiciones del piloto

El proyecto se concibe como un piloto o prueba de concepto, sin aprobación formal a gran escala ni presupuesto definitivo asignado. Tres condiciones externas determinan su diseño: la restricción ecuatoriana sobre la divulgación de productos que no son de venta libre (OTC) en redes sociales y canales abiertos; la Ley Orgánica de Protección de Datos Personales (LOPDP) y su Reglamento General, que exigen autorización previa, expresa e informada para tratar los datos de los dependientes; y la política de mensajería de WhatsApp Business, que prohíbe los mensajes sobre medicamentos con receta y trata los OTC como una vertical regulada. Por eso el diseño es independiente del canal y contempla una validación con el área Legal antes de construir.

Se aplica además un principio de honestidad: todas las cifras económicas son estimaciones con supuestos declarados, y el documento evita proyectar ventas sin evidencia que las respalde.

Metodología y estructura del documento

El trabajo se desarrolló durante diez semanas, con un entregable semanal que se apoyó en el anterior; cada capítulo corresponde a una de esas semanas. La Parte I (Business Intelligence, capítulos 1 a 5) parte del problema, las fuentes y los procesos de carga, y llega a la visualización y al business case. La Parte II (Big Data e Inteligencia Artificial, capítulos 1 a 5) incorpora los datos no estructurados, la arquitectura en la nube y open source, el modelo de priorización y el caso de uso en near real time. Cierran el documento las conclusiones y aplicaciones y las referencias.

PARTE I: BUSINESS INTELLIGENCE

Corresponde al Bloque 1 del programa. Cubre las semanas 1 a 5 del programa y recorre la cadena de valor de la información: el problema de negocio y los indicadores que lo miden, las fuentes de datos, los procesos de carga, la visualización y el caso de negocio con el rol de datos que sostiene el proyecto.

CAPÍTULO 1. Introducción al proyecto de Business Intelligence

1. Presentación de la empresa

El Laboratorio es una compañía farmacéutica que comercializa productos a nivel nacional a través de las principales cadenas de farmacias del país. Su modelo comercial se apoya en dos frentes de relacionamiento: la visita médica, orientada a profesionales de la salud y especialidades clínicas, y la fuerza de representantes, orientada a los puntos de venta y sus dependientes.

El proyecto propuesto busca establecer comunicación e interacción frecuente con los dependientes de farmacias de las cadenas más importantes a nivel nacional, mediante mensajes de fidelización y divulgación científica personalizados por cadena. Se trata de la primera iniciativa del laboratorio orientada a ampliar su alcance hacia los dependientes de punto de venta mediante un canal virtual.

2. Identificación de los objetivos iniciales del proyecto: definición inicial del problema

Problema

Actualmente no existe ningún canal virtual de comunicación entre El Laboratorio y los dependientes de farmacia. Todo el relacionamiento depende exclusivamente de la visita presencial del representante comercial, cuya frecuencia varía según la clasificación del punto de venta (A, B, C o D) y puede llegar a ser de una sola visita al mes en los puntos de menor categoría. Esto limita la divulgación científica continua y no permite sostener un vínculo de fidelización medible con el dependiente, actor clave en la venta responsable de productos farmacéuticos.

Procesos de negocio relacionados

- Proceso de visita comercial a puntos de venta (representantes): único canal de contacto actual, con frecuencia limitada por la clasificación del punto de venta.
- Proceso de divulgación científica en el punto de venta: se realiza únicamente durante la visita presencial, sin continuidad entre una visita y otra.
- Proceso de fidelización con dependientes: no existe un programa formal; solo compromisos verbales no cuantificables.
- Proceso de gestión comercial corporativa (Key Account): negocia condiciones y grillas promocionales por cadena, información que debe alimentar los mensajes personalizados del proyecto.

Objetivo del proyecto

Diseñar e implementar un canal de comunicación virtual que permita al Laboratorio interactuar de forma frecuente y personalizada por cadena con los dependientes de farmacia, complementando —sin reemplazar— la labor del representante, para

fortalecer la divulgación científica y sentar las bases de un programa de fidelización medible.

3. Áreas y personas que demandan la información y uso actual

Tabla 1. Áreas y personas que demandan la información y su uso actual

Área	Rol frente al proyecto	Uso actual de la información
Dirección Comercial / Ventas	Patrocinador del proyecto; gestiona representantes y Key Account	Reportes de ventas y visitas en Excel; sin área de BI formal
Representantes	Ejecutan la visita y divulgación científica en el punto de venta	Información de grilla promocional por cadena; registro manual de visitas
Key Account	Negocia condiciones comerciales y grillas por cadena	Información de condiciones, exhibición y portafolio por cadena
Marketing	Sin injerencia actual sobre el canal de dependientes (solo visita médica)	Material científico únicamente para médicos
TI / Sistemas	Socio tecnológico para el desarrollo del canal virtual	No participa actualmente en la comunicación con dependientes

No existe actualmente un área de Analítica de Datos / BI; la información comercial se gestiona íntegramente desde la Dirección Comercial mediante hojas de cálculo.

4. Objetivos de cada área con la información actual y necesidades detectadas

Tabla 2. Objetivos de cada área y necesidades detectadas

Área	Objetivo frente al proyecto	Necesidad detectada
Dirección Comercial	Ampliar el alcance de la comunicación con dependientes sin depender solo de la visita física	Un canal virtual propio, y estructurar la información de ventas/visitas más allá de Excel
Representantes	Sostener la divulgación científica entre visitas presenciales	Contenido digital reutilizable, personalizado por cadena
Key Account	Alinear los mensajes digitales con las condiciones y grillas vigentes por cadena	Integración de su información de portafolio por cadena al canal virtual
TI / Sistemas	Habilitar la infraestructura tecnológica del canal virtual	Requerimientos claros de alcance, volumen de usuarios y restricciones regulatorias (ej. productos OTC)

5. Definición de los KPI's relevantes

Siguiendo el marco Conocimiento (K) – Organización (O) – Motivación (M), se definen los indicadores clave para evidenciar la brecha (gap) que el proyecto busca cerrar. Los resultados actuales se calculan con base en los datos disponibles (1,079 puntos de venta cubiertos por 8 representantes a nivel nacional); los supuestos usados se detallan al pie de cada tabla y deben validarse con datos operativos reales antes de la ejecución del proyecto.

I. Indicadores de Conocimiento (K)

Miden el alcance real de la divulgación científica hacia el dependiente.

Tabla 3. Indicadores de Conocimiento (K): resultado actual, meta y brecha

Indicador	Resultado actual	Meta empresa	Gap
Frecuencia real de contacto por punto de venta (visitas/mes)	0.98 visitas/mes	2.8 visitas/mes (según política A/B/C/D)	1.82 visitas/mes
% de puntos de venta con acceso a un canal virtual de divulgación científica	0%	100%	100 p.p.

Supuesto: capacidad estimada de 8 representantes × 22 días hábiles × 6 visitas efectivas/día = 1,056 visitas posibles al mes, frente a 1,079 puntos de venta. La meta de 2.8 visitas/mes surge de ponderar la política de frecuencia por clasificación de punto de venta (A/B: 5x, C: 3x, D: 1–2x).

II. Indicadores de Organización (O)

Miden la capacidad operativa actual del equipo comercial frente a la demanda de cobertura.

Tabla 4. Indicadores de Organización (O): resultado actual, meta y brecha

Indicador	Resultado actual	Meta empresa	Gap
Cobertura de visitas: capacidad mensual real / demanda según política	35%	100%	65 p.p.
Puntos de venta por representante	135 puntos/representante	A definir vía estudio de carga (referencial: 80–100)	35–55 puntos

Demanda mensual estimada: 1,079 puntos × 2.8 visitas = 3,021 visitas requeridas; capacidad real: 1,056 visitas. Cobertura = 1,056 / 3,021 ≈ 35%.

III. Indicadores de Motivación (M)

Miden el estado actual de la fidelización y el engagement con el dependiente.

Tabla 5. Indicadores de Motivación (M): resultado actual, meta y brecha

Indicador	Resultado actual	Meta empresa	Gap
% de dependientes en un programa de fidelización formal y medible	0%	100%	100 p.p.
Nivel de interacción/engagement posterior al contacto	No medible (sin canal de seguimiento)	A definir con el canal virtual (ej. 60% de apertura de mensajes)	No cuantificable aún

El programa de fidelización actual consiste en compromisos verbales del representante (ej. combos), no formalizados ni cuantificables; para efectos de este análisis se considera 0%.

6. Fuentes de conocimiento sobre la empresa y nivel de conocimiento

Fuentes de conocimiento disponibles

- Reportes comerciales y de ventas gestionados por la Dirección Comercial (Excel): frecuencia de visitas, clasificación de puntos de venta, grillas promocionales por cadena.
- Documentación interna formal: manuales de procesos comerciales, políticas de visita y materiales de divulgación científica.
- Conocimiento directo de la operación diaria, por la interacción constante con el equipo de representantes y con Key Account.
- Acceso directo a otras áreas internas (TI, Marketing) mediante consulta, al formar parte de la organización.

Nivel de conocimiento: 8 / 10

La posición del equipo autor dentro de la Dirección Comercial —desde donde se supervisa y representa al equipo de representantes— da acceso directo y amplio a la información operativa y comercial relevante para el proyecto. No se asigna la puntuación máxima porque existen aspectos que no se manejan de forma nativa, como el detalle técnico de las capacidades de TI/Sistemas para el canal virtual, o la información financiera y estratégica de la Gerencia General, que requerirán entrevistas o coordinación con las áreas correspondientes.

7. Consideraciones adicionales para la ejecución del proyecto

Restricción regulatoria sobre divulgación de productos

La normativa vigente en Ecuador restringe la divulgación pública de productos que no sean de venta libre (OTC) en redes sociales y canales digitales abiertos. Esto condiciona el diseño del canal: se recomienda un mecanismo cerrado o de acceso restringido a dependientes verificados, diferenciando el tratamiento de productos OTC frente a los de prescripción, y validando el alcance de contenidos con el área legal/regulatoria antes de definir el proyecto final.

Naturaleza piloto del proyecto

Esta iniciativa se concibe como un proyecto piloto o prueba de concepto, sin aprobación formal a gran escala ni presupuesto definitivo asignado. Se recomienda un alcance inicial acotado (por ejemplo, cadenas o puntos de venta A/B en una ciudad) con criterios de éxito medibles que sustenten una eventual expansión.

Contexto favorable de respaldo interno

No se han identificado riesgos significativos de resistencia interna. La Dirección Comercial, que patrocina el proyecto, y el área de TI, como socio tecnológico, no presentan conflictos de prioridades conocidos, lo cual constituye un entorno favorable para la ejecución.

CAPÍTULO 2. Fuentes de datos del proyecto

Objetivo y alcance del capítulo

La información en El Laboratorio tiene diversos propósitos; su uso y estructura dependen de las necesidades de quien la requiere, de las herramientas que la soportan y del público al cual va dirigida. Para efectos de este proyecto, estos propósitos se clasifican de la siguiente manera:

- Operativos / Tácticos: reportes que responden a una pregunta específica de la operación diaria, como el listado de visitas realizadas por representante o el detalle de puntos de venta por clasificación (A, B, C, D).
- Corporativos: datos agregados que apoyan la toma de decisiones de la Dirección Comercial, como tableros de cobertura de visitas o cumplimiento de frecuencia por cadena.
- Comerciales: información desde la perspectiva del cliente (la cadena de farmacias), como el desempeño de ventas y rotación por cadena.
- De procesos: reportes con el detalle de la ejecución del proceso de visita (fecha, representante, punto de venta, productos expuestos).
- Analítica y minería: para este proyecto, aplicaría a futuro para identificar patrones de receptividad de los dependientes y personalizar la frecuencia y el contenido de los mensajes del canal virtual.

En El Laboratorio aún no existen reportes de tipo Legal asociados a este proceso, dado que la comunicación con dependientes es una iniciativa nueva.

Proceso de consolidación de información

El Laboratorio cuenta con un Data Warehouse corporativo administrado por el área de TI/Sistemas, responsable de establecer las políticas y definiciones sobre el uso y la estructura de la información. Este repositorio centraliza los datos generados por los distintos sistemas de la compañía, entre ellos el sistema de gestión de visitas de representantes (CRM), que es la fuente principal de información para este proyecto.

Data Warehouse de El Laboratorio

El proceso de carga de información relacionada con la gestión comercial presenta la siguiente rutina:

- El CRM de gestión de visitas de representantes consolida la información de cada visita realizada en el punto de venta (fecha, representante, punto de venta, productos expuestos).
- Esta información se actualiza hacia el Data Warehouse con una periodicidad mensual.
- Desde el Data Warehouse, las áreas acceden a la información mediante herramientas de Excel y Power BI para la generación de reportes y análisis.

Oportunidad de mejora identificada: la actualización mensual limita la capacidad de reacción del proyecto. Para un canal de comunicación virtual que busca ser frecuente, sería deseable contar con una

actualización diaria de la información de visitas, lo cual debe evaluarse junto con TI/Sistemas como parte del alcance del proyecto.

Definición de estructura y gobierno de los datos

El gobierno de la estructura de la información está a cargo del área de TI/Sistemas, responsable de autorizar los accesos a las tablas del Data Warehouse y de validar que la información cumpla con los criterios definidos de calidad, fuente y frecuencia de actualización. Actualmente no se han identificado problemas graves de calidad de datos; la principal oportunidad de mejora es la frecuencia de actualización (mensual vs. diaria), más que la integridad o consistencia de la información.

Tipos de datos relevantes para el proyecto

Se relacionan los tipos de datos disponibles o requeridos para el proyecto. La variable en común que permitiría relacionar la información de cada punto de venta es el código único de punto de venta (o de cadena).

Tabla 6. Tipos de datos relevantes para el proyecto

Tipo de dato	Descripción
Visitas comerciales	Fecha de visita, representante asignado, punto de venta, productos expuestos según grilla promocional (fuente: CRM de representantes, actualización mensual).
Clasificación de puntos de venta	Categoría A, B, C o D y ubicación (ciudad / fuera de ciudad), que determinan la frecuencia de visita establecida por política.
Grilla promocional por cadena	Portafolio de 5-6 productos vigentes por cadena, gestionado por el equipo de Key Account.
Ventas y rotación por cadena	Información comercial consolidada por cadena, gestionada actualmente por la Dirección Comercial en Excel.
Datos de contacto de dependientes	Actualmente inexistente. Sería un nuevo tipo de dato a construir (nombre, punto de venta, cadena, canal de contacto) como insumo indispensable para el canal virtual.

Proceso de transformación de la información

El proceso de transformación de la información estaría orientado a atender dos necesidades del proyecto:

Tabla 7. Necesidades del proceso de transformación de la información

Necesidad	Descripción
Reportería y Análisis	Consolidación de la información de visitas, cobertura por clasificación de punto de venta y desempeño por cadena, en tableros de control que apoyen a la Dirección Comercial en el seguimiento del proyecto

Necesidad	Descripción
	piloto.
Analítica y minería	A futuro, uso de la información del canal virtual (aperturas de mensajes, interacción, frecuencia de contacto efectiva) para identificar patrones de receptividad de los dependientes y personalizar la frecuencia y el contenido de los mensajes por cadena.

Necesidades específicas identificadas para el proyecto

Necesidad de una base de contacto de dependientes

A diferencia de la información de visitas o ventas, hoy no existe ninguna base de datos que registre el contacto directo de los dependientes (nombre, punto de venta, cadena, canal de comunicación). Esta es la necesidad de información más crítica para poder ejecutar el proyecto, y debe construirse desde cero, posiblemente apoyándose en el registro que realizan los representantes durante sus visitas.

Necesidad de actualización más frecuente del Data Warehouse

Para que el canal virtual complemente de forma oportuna la visita presencial (por ejemplo, reforzando contenido después de una visita reciente), se requeriría una actualización diaria de la información de visitas, en lugar de la periodicidad mensual actual.

Necesidad de segmentación por clasificación y cadena

La personalización de mensajes por cadena, mencionada en el objetivo del proyecto, requiere que la información de grilla promocional y clasificación de punto de venta esté disponible y actualizada de forma que pueda cruzarse con la futura base de contacto de dependientes.

Conclusiones del capítulo

El Laboratorio cuenta con una base tecnológica aprovechable —un Data Warehouse gobernado por TI/Sistemas y un CRM de visitas— que puede servir de punto de partida para el proyecto. Sin embargo, existen dos brechas clave que deben resolverse antes de la implementación: la periodicidad de actualización de la información (mensual, no diaria) y, sobre todo, la inexistencia de una base de datos de contacto de los dependientes, que es el insumo indispensable para activar cualquier canal de comunicación virtual y personalizada por cadena.

CAPÍTULO 3. Procesos ETL del proyecto

1. Procesos de carga de datos (ETL)

La información relevante para el proyecto llega al Data Warehouse de El Laboratorio a través de dos flujos claramente distintos, con orígenes y dinámicas muy diferentes entre sí:

Flujo 1: CRM de visitas de representantes (fuente interna)

- Extracción: el CRM registra automáticamente cada visita (fecha, representante, punto de venta, productos expuestos).
- Transformación: el equipo de TI/Sistemas procesa la información mediante consultas SQL desarrolladas internamente, aplicando validaciones de formato, campos obligatorios y control de duplicados.
- Carga: el proceso se ejecuta de forma automática mediante un batch programado, sin intervención manual.

Flujo 2: Drenaje por punto de venta (fuente externa)

- Extracción: la información de ventas/rotación (drenaje) es reportada por las propias cadenas de farmacias, no generada por un sistema interno del laboratorio.
- Esta información se recibe "a mes caído", es decir, una vez que el mes ya cerró, por lo que su disponibilidad depende de terceros externos y no del laboratorio.
- Una vez recibida, se integra al Data Warehouse mediante el mismo proceso SQL gestionado por TI.

> **Hallazgo clave: aunque los dos flujos se cargan hoy con la misma periodicidad mensual, sus restricciones son de naturaleza distinta. El CRM de visitas es 100% interno y automatizado — técnicamente podría actualizarse a diario sin depender de nadie más. El drenaje, en cambio, tiene una limitación real e inevitable: depende de que las cadenas reporten la información, y eso solo ocurre una vez al mes. Sincronizar ambos flujos al ritmo del más lento no es una necesidad técnica, es una decisión de diseño que conviene cuestionar.**

2. Información y calidad de la información disponible

El proceso cuenta con validaciones formales de calidad antes de que los datos entren al Data Warehouse: se verifica el formato correcto de cada campo, la presencia de campos obligatorios y se controla la existencia de registros duplicados. Esto es una fortaleza real del proceso actual: la información que llega al repositorio es estructuralmente confiable.

Sin embargo, esta calidad estructural no resuelve el problema de oportunidad. Un dato puede estar perfectamente validado en formato y, aun así, llegar sistemáticamente tarde para el propósito del proyecto. La calidad de un dato no se mide solo por su exactitud, sino también por si está disponible en el momento en que se necesita tomar una decisión o generar una interacción — y en ese segundo criterio, el proceso actual sí presenta una debilidad real para un proyecto que busca comunicación frecuente.

3. Horarios de carga y accesibilidad

Tabla 8. Frecuencia, tipo de ejecución y acceso a la carga de datos

Aspecto	Situación actual
Frecuencia de carga (ambos flujos)	Mensual
Tipo de ejecución	Automática (batch programado por TI)
Acceso a la información	Abierto: cualquier persona de la Dirección Comercial puede consultar libremente, sin necesidad de solicitud formal a TI

La accesibilidad abierta hacia el equipo Comercial es positiva porque democratiza el uso de la información sin fricciones burocráticas. No obstante, el acceso libre a un dato no compensa que ese dato tenga hasta 30 días de desfase: dar acceso instantáneo a información desactualizada resuelve un problema de gobierno de datos, pero no resuelve el problema de fondo que el proyecto necesita resolver, que es la frecuencia de contacto con el dependiente.

4. Variables disponibles y acceso al Data Warehouse

Tabla 9. Variables disponibles por origen de datos

Origen	Variables disponibles
CRM de visitas (interno)	Fecha de visita, representante, punto de venta, clasificación (A/B/C/D), productos expuestos según grilla promocional
Drenaje por punto de venta (externo, cadenas)	Ventas / rotación de producto por punto de venta, reportado mensualmente por cada cadena

El acceso a estas tablas está gobernado por el área de TI/Sistemas, que definió la estructura de datos y habilitó el acceso directo del equipo Comercial vía Power BI, sin requerir solicitudes individuales de acceso.

5. Sistemas implicados en los procesos

Tabla 10. Sistemas implicados en los procesos

Sistema	Rol en el proceso
CRM de visitas de representantes	Fuente interna; genera el dato de origen de forma automática con cada visita registrada
Motor de procesamiento SQL (TI)	Transforma, valida y carga la información de ambos flujos hacia el Data Warehouse
Reporte de Drenaje (proporcionado por las cadenas)	Fuente externa; su disponibilidad depende del calendario de cierre de cada cadena, no del laboratorio

Sistema	Rol en el proceso
Power BI	Herramienta de consulta y explotación de la información para el equipo Comercial

6. Recursos de IT o externos empleados

El proceso de ETL es gestionado en su totalidad por el equipo interno de TI/Sistemas, utilizando SQL como herramienta principal de transformación y carga. No se emplean consultores o proveedores externos para este proceso. La única dependencia externa real del proceso no es tecnológica, sino de información: el dato de drenaje que las cadenas de farmacias entregan al laboratorio.

7. Necesidad futura: ETL de la base de contacto de dependientes

A diferencia de los dos flujos descritos anteriormente, la información de contacto de los dependientes (nombre, punto de venta, canal de comunicación) no existe hoy en ningún sistema. Su incorporación futura al Data Warehouse requeriría definir un tercer flujo de carga, con características muy distintas a los dos actuales:

- Extracción: se propone que la captura inicial la realice el representante durante la visita, registrándola en el mismo CRM que ya usa hoy para las visitas, evitando construir una herramienta nueva desde cero para el piloto.
- Transformación: requeriría validaciones adicionales a las ya existentes (formato, obligatoriedad, duplicados), específicas para datos personales: validación de formato de número de WhatsApp y control de duplicados por dependiente (no solo por registro de visita).
- Carga y actualización: al depender del contacto humano en cada visita, esta base solo podría actualizarse con la misma frecuencia de visita del representante (hasta 30 días en puntos tipo D), lo cual es una limitación relevante frente a la alta rotación de dependientes ya identificada.

> **Limitación a validar: delegar la captura y actualización de esta base en el representante añade una tarea administrativa a un recurso que, según el capítulo 1 de la Parte I, ya cubre solo ~35% de la demanda de visitas. Existe una alternativa — un modelo híbrido donde el representante invita y el propio dependiente confirma su registro (ej. vía WhatsApp)— que reduciría esta carga y resolvería mejor el consentimiento explícito del dependiente, pero su diseño excede el alcance de este capítulo y se propone como línea de trabajo para una fase posterior del proyecto.**

8. Manejo de excepciones y retención de la información

Manejo de excepciones

El proceso actual no tiene definido qué ocurre cuando una cadena no entrega su reporte de drenaje a tiempo o lo entrega incompleto. Esta es una brecha a resolver antes de escalar el proyecto: se recomienda definir una regla explícita (por ejemplo, mantener el último dato disponible marcado como "desactualizado" en los tableros, en lugar de dejar el campo vacío) para que el equipo Comercial sepa siempre si está viendo un dato vigente o uno rezagado.

Tiempo de retención de la información

Tampoco está definido cuánto tiempo se conserva el histórico de visitas y drenaje en el Data Warehouse. Para el proyecto esto es relevante porque medir la efectividad de la fidelización requerirá comparar el comportamiento de un mismo punto de venta a lo largo de varios meses; se recomienda validar con TI/Sistemas la política de retención vigente y, de ser necesario, gestionar un histórico dedicado para los datos de este proyecto.

Conclusiones del capítulo

El proceso ETL actual de El Laboratorio tiene una base técnica sólida: es automático, cuenta con validaciones de calidad, y su acceso está bien gobernado por TI sin fricciones para el usuario de negocio. El verdadero cuello de botella para este proyecto no es la capacidad técnica de TI, sino una decisión de diseño: acoplar la actualización del CRM de visitas (que podría ser prácticamente inmediata) al ritmo del dato de drenaje (que depende de terceros y solo puede ser mensual).

Para que el proyecto de comunicación virtual con dependientes sea viable, se recomienda desacoplar ambos flujos: mantener el drenaje con su periodicidad mensual natural (dado que no puede cambiarse, depende de las cadenas), pero llevar la actualización del CRM de visitas a una frecuencia semanal o incluso diaria, ya que no existe ninguna limitación técnica real que lo impida. Esto permitiría que el canal virtual reaccione con datos recientes de visita, sin necesidad de esperar el cierre mensual del drenaje.

Finalmente, el mayor reto de datos del proyecto no está en los flujos que ya existen, sino en el que aún no se ha construido: la base de contacto de los dependientes. Su diseño debe considerarse una prioridad temprana del proyecto, dado que ningún canal virtual puede activarse sin ella.

CAPÍTULO 4. Visualización de la información

Premisas de entrega de información

Para la entrega de información de tipo Corporativo y Comercial —relevante para el seguimiento del proyecto por parte de la Dirección Comercial y Key Account— la estrategia se basa en el autoservicio: los usuarios de negocio consultan directamente la información sin depender de un desarrollo a medida de TI. Para información de tipo Operativo o de Proceso (por ejemplo, el detalle diario de una visita específica), la entrega seguirá dependiendo de los reportes propios del CRM, gestionados por TI/Sistemas.

Para evitar ambigüedad, se definen los siguientes conceptos aplicados al contexto de El Laboratorio:

Tabla 11. Conceptos de reporte, indicador, tablero de control y consulta ad-hoc

Término	Descripción aplicada al proyecto
Reporte	Resultado de una consulta dirigido a un público específico (ej. reporte mensual de visitas por representante), cuyo objetivo es informar para la toma de decisiones.
Indicador	Magnitud que permite medir una característica del proyecto en el tiempo, como la frecuencia real de visita o el nivel de fidelización, y compararla contra la meta establecida.
Tablero de control	Conjunto de indicadores del proyecto (cobertura, fidelización, desempeño por cadena) cuyo seguimiento periódico apoya la toma de decisiones de la Dirección Comercial.
Consulta ad-hoc	Consulta no programada que un usuario de Comercial puede realizar directamente sobre el Data Warehouse a través de Power BI, sin depender de TI.

1. Análisis de herramientas y selección para el ecosistema del proyecto

El Laboratorio ya cuenta con un estándar de visualización establecido: Power BI, con licenciamiento básico (Pro), utilizado hoy para consultar la información del Data Warehouse. Esta es la herramienta natural para el proyecto, ya que no implica adoptar una tecnología nueva ni negociar licenciamiento adicional con TI.

Tabla 12. Herramientas de visualización y su rol en el proyecto

Herramienta	Rol para el proyecto
Power BI (licencia básica/Pro)	Herramienta principal de visualización y autoservicio para Dirección Comercial y Key Account; ya es el estándar de la empresa.

Herramienta	Rol para el proyecto
Excel	Uso complementario para análisis puntuales o exportaciones rápidas, tal como se usa hoy junto a Power BI.
Reportes nativos del CRM de visitas	Cubren la necesidad operativa/de proceso (detalle de una visita puntual), gestionados por TI, fuera del alcance del tablero de autoservicio.

Limitación a considerar: al tratarse de licenciamiento básico (Pro) y no premium, Power BI impone límites de frecuencia de actualización automática (típicamente hasta 8 actualizaciones programadas por día). Esto no representa un problema hoy, dado que la fuente de datos se actualiza mensualmente; pero si se implementa la recomendación del capítulo 3 de la Parte I de llevar el CRM de visitas a una actualización diaria o semanal, deberá validarse con TI si la licencia básica es suficiente o si el proyecto requerirá una capacidad premium a futuro.

2. Selección de usuarios y perfiles de uso del tablero

Se definieron dos perfiles con acceso al tablero de control del proyecto: Dirección Comercial y Key Account. A diferencia del acceso general al Data Warehouse descrito en el capítulo 3 de la Parte I —abierto a cualquier persona de Comercial—, el tablero específico de este proyecto se plantea con acceso más acotado.

Tabla 13. Perfiles con acceso al tablero de control

Perfil	Objetivo de uso	Nivel de detalle
Dirección Comercial	Seguimiento ejecutivo del piloto: cobertura de visitas, avance de fidelización y desempeño general por cadena	Vista agregada, con filtros por cadena y clasificación de punto de venta
Key Account	Validar que los mensajes y el desempeño del canal estén alineados con las condiciones negociadas por cadena	Vista filtrada por cadena específica

Punto a resolver: los representantes —quienes generan el dato base y ejecutan la visita— no tienen acceso propuesto al tablero en esta fase. Esto es coherente con mantenerlos enfocados en la ejecución más que en el análisis, dado que ya operan al límite de su capacidad (~35% de cobertura, según el capítulo 1 de la Parte I); sin embargo, sin visibilidad de su propio desempeño, es más difícil que se apropien de las metas de cobertura. Se recomienda evaluar en una fase posterior una vista simplificada para representantes (solo sus propios puntos de venta), sin necesidad de ampliar el acceso a la información de otras cadenas o representantes. Esta vista

se concreta en el capítulo 4 de la Parte II mediante una salida filtrada del modelo de priorización.

3. Necesidades de información y accesos por departamento y perfil

Tabla 14. Necesidades de información y tipo de acceso por perfil

Perfil	Necesidad de información	Tipo de acceso
Dirección Comercial	Cobertura de visitas vs. meta, avance de fidelización, desempeño de ventas por cadena (drenaje)	Autoservicio vía Power BI, sin restricción de cadena
Key Account	Desempeño y cobertura de su(s) cadena(s) asignada(s)	Autoservicio vía Power BI, filtrado por cadena
TI / Sistemas	Administración de la estructura de datos y del modelo del tablero	Acceso técnico completo, vía SQL y configuración de Power BI

Un punto que deberá resolverse antes de incorporar la futura base de contacto de dependientes (capítulo 3 de la Parte I, sección 7) es que ese dato es de carácter personal, a diferencia de la información de visitas y ventas que hoy se comparte sin restricciones dentro de Comercial. Cuando ese flujo se integre al Data Warehouse, el tablero deberá diferenciar entre información operativa (abierta) e información de contacto personal (con acceso restringido y trazable), siguiendo un protocolo de confidencialidad específico para ese tipo de dato.

Diseño conceptual del tablero: Cobertura y Fidelización de Dependientes

Se propone un tablero de control en Power BI, estructurado en cinco campos de información, orientado a operacionalizar el seguimiento de los indicadores definidos en el capítulo 1 de la Parte I (Conocimiento, Organización y Motivación):

Campo 1: Cuadro de mando (filtros)

- Cadena: permite aislar el desempeño de una cadena de farmacias específica.
- Clasificación de punto de venta (A/B/C/D): permite comparar cobertura según la frecuencia de visita esperada por categoría.
- Mes: permite evaluar la evolución mensual, en línea con la periodicidad actual de actualización del drenaje.

Campo 2: Indicadores de cobertura

Muestra la frecuencia real de visita frente a la meta por clasificación (indicador definido en el capítulo 1 de la Parte I), permitiendo visualizar el gap de cobertura por cadena o punto de venta.

Campo 3: Indicador de fidelización

Una vez implementado el canal virtual, este campo mostrará el nivel de interacción/engagement de los dependientes contactados. Mientras no exista el canal, el campo reflejará 0%, como referencia del punto de partida del proyecto.

Campo 4: Desempeño por cadena (drenaje)

Gráfica de ventas/rotación por cadena, para correlacionar la cobertura de visitas y el futuro canal virtual con el desempeño comercial real, respondiendo a la pregunta de fondo del proyecto: si efectivamente impacta en ventas, no solo en volumen de mensajes.

Campo 5: Tendencia temporal

Evolución mes a mes de la cobertura y (a futuro) de la fidelización, para identificar si el gap se está cerrando conforme avanza el piloto.



Figura 1. Mockup conceptual del tablero propuesto — valores ilustrativos, no corresponden a datos reales del laboratorio.

4. Revisión y cumplimiento de los objetivos iniciales

El tablero propuesto no introduce indicadores nuevos: opera directamente sobre los KPI's de Conocimiento, Organización y Motivación definidos en el capítulo 1 de la Parte I, dándoles una capa de visualización periódica y accesible para la Dirección Comercial. Esto permite verificar de forma continua si el proyecto se está acercando a las metas planteadas (100% de cobertura de canal virtual, cierre del gap de frecuencia de visita, fidelización medible) o si es necesario ajustar el alcance del piloto.

Conclusiones del capítulo

El Laboratorio no requiere adoptar una herramienta nueva para visualizar la información del proyecto: Power BI, ya estandarizado en la empresa, es suficiente para dar autoservicio a Dirección Comercial y Key Account. La decisión relevante no es tecnológica sino de gobierno: qué perfiles acceden a qué nivel de detalle, y cómo se tratará de forma diferenciada la información personal de los dependientes una vez se incorpore al Data Warehouse. El tablero conceptual propuesto conecta directamente con los KPI's ya definidos, evitando construir un nuevo marco de medición y permitiendo verificar, de forma continua, si el proyecto avanza hacia sus objetivos iniciales.

CAPÍTULO 5. Business case y rol de datos

Posicionamiento de la función de datos dentro de El Laboratorio

A diferencia de organizaciones con una gerencia de analítica dedicada, El Laboratorio no cuenta hoy con un área de datos formal: la información se gestiona con Excel y Power BI directamente desde la Dirección Comercial (capítulo 2 de la Parte I). Crear una gerencia completa no es viable ni necesario para un proyecto en fase piloto; en su lugar, se propone asignar una responsabilidad de datos a un perfil ya existente dentro de Comercial, evitando el costo y la complejidad de contratar una estructura nueva.

Organigrama propuesto



Figura 2. Ubicación de la función de datos dentro de la estructura actual — no representa una nueva área formal.

Objetivo principal de la función propuesta

Administrar el tablero de control del proyecto (capítulo 4 de la Parte I), mantener actualizada la relación entre las necesidades de información de Comercial y Key Account, y servir de punto de contacto con TI/Sistemas para la evolución del proyecto (por ejemplo, la eventual migración a actualización diaria del CRM de visitas, discutida en el capítulo 3 de la Parte I), sin necesidad de crear una nueva área dentro de la compañía.

Objetivo del rol por responsabilidad

Responsable de Datos del Proyecto (Analista Comercial existente, dedicación parcial)

Administrar el tablero de control, coordinar con TI la actualización y calidad de los datos, y actuar como punto de escalamiento para consultas ad-hoc de otros usuarios de Comercial.

> **Las consultas ad-hoc de autoservicio en Power BI pueden resultar complejas para un usuario de negocio sin formación técnica. En lugar de asumir que cualquier persona de Comercial puede construir sus propias consultas, se designa un responsable que absorbe esa complejidad y da soporte de primer nivel, antes de escalar a TI.**

Gobernanza de accesos al tablero

Queda pendiente definir si los representantes deben tener acceso al tablero (capítulo 4 de la Parte I). Se trata de una decisión de negocio, no técnica: no la toma TI ni se resuelve por defecto. La Dirección Comercial, con recomendación del Responsable de Datos del Proyecto, deberá evaluarla formalmente una vez el piloto tenga resultados iniciales, sopesando el valor de la visibilidad para el representante frente al riesgo de sobrecarga ya identificado en el capítulo 1 de la Parte I. Mientras tanto, el capítulo 4 de la Parte II propone una alternativa acotada: una salida filtrada del modelo de priorización, limitada a los puntos de venta de cada representante.

Impactos esperados en los resultados del proyecto

La función propuesta no genera valor por sí misma: su impacto se mide por su efecto sobre los indicadores ya definidos en el capítulo 1 de la Parte I (cobertura de visitas, fidelización) y sobre la calidad y oportunidad de la información que alimenta esas decisiones (capítulo 3 de la Parte I). Concretamente, se espera que reduzca el tiempo que hoy toma consolidar manualmente la información en Excel, y que disminuya la dependencia de TI para resolver solicitudes puntuales de información.

Business case: costos y beneficios

Se investigaron rangos salariales de mercado en Ecuador para perfiles de analista de datos/comercial (fuentes: Paylab, sueldo.ec, Talently, Glassdoor Ecuador), que oscilan entre USD 600 y USD 1,200 mensuales para un perfil junior-medio. Se toma como supuesto un salario de mercado de referencia de USD 900/mes para dimensionar el costo de la dedicación parcial.

Presupuesto estimado

Tabla 15. Presupuesto estimado de la función de datos

Concepto	Supuesto	Costo mensual	Costo anual
Dedicación parcial del Responsable de Datos (~30% de su tiempo)	Salario de mercado USD 900/mes × 30%	USD 270	USD 3,240
Licencias Power BI Premium Per User (opcional, si se adopta la recomendación del capítulo 3 de	2 licencias PPU × USD 24/mes	USD 48	USD 576

Concepto	Supuesto	Costo mensual	Costo anual
la Parte I de actualización diaria)			
Capacitación inicial en Power BI para el equipo Comercial (única vez)	Estimado de mercado para taller interno/externo	—	USD 300 (único pago)

Cifras estimadas a partir de rangos de mercado consultados en julio de 2026 (Paylab Ecuador, sueldo.ec, Talently, Glassdoor); deben validarse con el área de Recursos Humanos y con TI antes de aprobar el presupuesto real. La licencia Premium es un costo contingente: solo aplica si se decide avanzar con la recomendación de actualización diaria del CRM.

Beneficio conservador: ahorro de tiempo operativo

Se identifican dos fuentes de ahorro directamente cuantificables, siguiendo el mismo tipo de supuestos declarados usados en capítulos anteriores:

Tabla 16. Ahorro operativo estimado con el tablero

Concepto	Hoy (sin tablero)	Con tablero	Ahorro estimado
Horas/mes consolidando manualmente visitas y drenaje en Excel	8 horas	1 hora	7 horas/mes (~USD 472/año)
Horas/mes de TI atendiendo solicitudes puntuales de reportes de Comercial	10 horas	2 horas	8 horas/mes (~USD 900/año)

> **Hallazgo crítico: el ahorro de tiempo combinado (~USD 1,372/año) no cubre el costo estimado de la propuesta (~USD 3,240 a 4,116 en el primer año, según se adopten la licencia opcional y la capacitación). Presentar este business case únicamente como un caso de ahorro operativo sería, honestamente, insuficiente para justificarlo. El valor real de esta inversión no está en el ahorro de horas, sino en habilitar el proyecto de fidelización en sí — algo que no se puede monetizar todavía con datos reales.**

Beneficio estratégico: potencial de impacto en ventas (no garantizado)

El verdadero argumento de negocio no es el ahorro operativo, sino el valor de contar con información confiable y visible para tomar decisiones sobre el piloto. Si la cobertura de visitas mejora del 35% actual hacia la meta (capítulo 1 de la Parte I), y el canal virtual logra activarse, es razonable esperar algún efecto positivo en la rotación de producto por punto de venta — pero, a diferencia de los ahorros de tiempo, este efecto no puede cuantificarse con datos reales hasta que el piloto esté en marcha, y no se debe forzar una cifra de venta proyectada sin evidencia. Se recomienda tratar este beneficio como la

justificación estratégica de la inversión (bajo costo, aprendizaje temprano), no como una cifra financiera comprometida.

Conclusiones del capítulo

El Laboratorio no necesita crear una nueva gerencia de analítica para sostener este proyecto: basta con asignar una responsabilidad de datos a un perfil existente dentro de Comercial, con una dedicación parcial y un costo acotado (entre USD 3,240 y 4,116 en el primer año). Sin embargo, es importante ser honestos: bajo un análisis estrictamente de ahorro de tiempo, esta inversión no se paga sola. Su justificación real es estratégica — sin esta función, el proyecto de fidelización no tiene quién administre el tablero, resuelva las consultas ad-hoc que preocupan al usuario de negocio, ni sirva de puente formal con TI. El siguiente paso es que el piloto genere datos reales que permitan reemplazar esta justificación estratégica por una cifra financiera concreta.

PARTE II: BIG DATA E INTELIGENCIA ARTIFICIAL

Corresponde al Bloque 2 del programa. Cubre las semanas 6 a 10 del programa e incorpora los datos no estructurados, define la arquitectura en la nube y open source, diseña un modelo de inteligencia artificial de priorización y un caso de uso en near real time con su business case.

CAPÍTULO 1. Obtención de datos no estructurados

Objetivo y alcance del capítulo

Además de la información ya centralizada en el Data Warehouse (visitas y drenaje, capítulo 3 de la Parte I), existen dos grandes fuentes de datos adicionales relevantes para el proyecto, que hoy no están estructuradas ni integradas:

- Compra de datos a entidades públicas y privadas (paneles de mercado farmacéutico).
- Datos generados de manera desorganizada a través de canales propios de la operación (notas de visita y chats informales).

Compra de datos a entidades externas

Para este proyecto, la compra de datos externos no debería orientarse a obtener información personal de los dependientes. Comprar datos de contacto de personas naturales sin una base de consentimiento válida sería, además de éticamente cuestionable, contrario a los principios de la Ley Orgánica de Protección de Datos Personales de Ecuador (LOPDP, 2021) y su Reglamento General (Decreto Ejecutivo 904, 2023), que exigen que el tratamiento de datos personales cuente con autorización previa, expresa e informada del titular.

El valor real de la compra de datos externos para este proyecto está en información de mercado agregada, no en datos personales de individuos:

Tabla 17. Fuentes externas de datos y su uso en el proyecto

Fuente externa	Tipo de dato	Uso para el proyecto
Paneles de mercado farmacéutico (ej. IQVIA u otros proveedores especializados)	Volúmenes de venta y rotación de producto por cadena/región, a nivel agregado	Contrastar el drenaje reportado por las cadenas (capítulo 3 de la Parte I) contra una fuente independiente, para detectar reportes incompletos o subestimados
ARCSA (Agencia Nacional de Regulación, Control y Vigilancia Sanitaria)	Información pública de registro sanitario y clasificación de productos (OTC vs. venta bajo receta)	Validar qué productos del portafolio pueden divulgarse libremente en el canal virtual, dada la restricción regulatoria ya identificada en el capítulo 1 de la Parte I

Nota de gobernanza: cualquier compra de datos de mercado debe evaluarse junto con el equipo legal antes de contratarse, para confirmar que la fuente externa cumple con la LOPDP y que el dato adquirido es efectivamente agregado y no rastreable a personas naturales.

Datos generados de manera desorganizada por la organización

En la operación diaria del proyecto ya se generan datos con valor potencial, pero que hoy no están estructurados ni conectados al Data Warehouse:

Tabla 18. Fuentes internas desorganizadas con valor potencial

Fuente interna desorganizada	Qué contiene	Por qué genera valor
Campo de notas/observaciones libres del CRM de visitas	Comentarios en texto libre del representante sobre lo ocurrido en cada visita (dudas del dependiente, objeciones sobre un producto, quejas)	Permite identificar patrones de objeciones o dudas recurrentes que hoy no se capturan en ningún campo estructurado del CRM
Chats informales de WhatsApp entre representantes y dependientes/cadenas	Conversaciones sueltas, sin formato ni estructura, que ya ocurren de manera informal	Es la evidencia más temprana de cómo reaccionarían los dependientes a un canal virtual formal, antes incluso de que este exista

> **Punto a resolver antes de usar los chats de WhatsApp: son conversaciones entre personas, no un canal oficial de la empresa. Antes de analizar su contenido de forma sistemática, debe evaluarse si eso requiere informar o pedir consentimiento a los dependientes involucrados, dado que ya se identificó como brecha en el capítulo 3 de la Parte I que la futura base de contacto de dependientes deberá tratarse como dato personal, con reglas de acceso distintas a las de la información operativa.**

Solución para integrar la información a modelos estructurados

Text Analytics sobre las notas de visita

Aplicar técnicas de minería de texto sobre el campo de observaciones del CRM permitiría identificar palabras clave y temas recurrentes (por ejemplo, qué producto genera más dudas, o qué objeción se repite en una cadena específica), sin necesidad de rediseñar el CRM ni pedirle al representante que llene campos adicionales.

Protocolo de captura para los chats informales

A diferencia de las notas del CRM (que ya están en un sistema de la empresa), los chats de WhatsApp de representantes están hoy en dispositivos personales, sin ningún proceso ETL que los traiga al Data Warehouse. Se requiere primero una definición de gobierno (qué se puede capturar y con qué autorización) antes de construir cualquier proceso técnico de integración.

Casos de uso para el proyecto

Tabla 19. Casos de uso de los datos no estructurados

Caso de uso	Fuente de datos	Beneficio esperado
Detección de objeciones recurrentes sobre productos	Notas de visita (Text Analytics)	Ajustar el contenido de divulgación científica del canal virtual a las dudas reales del dependiente, no solo a la ficha técnica del producto
Validación cruzada del drenaje reportado por las cadenas	Panel de mercado farmacéutico externo	Detectar si una cadena está subreportando ventas, complementando el manejo de excepciones ya identificado en el capítulo 3 de la Parte I
Señal temprana de aceptación del futuro canal virtual	Chats informales (una vez resuelto el consentimiento)	Anticipar el nivel de fidelización esperado antes de tener datos formales del canal

Conclusiones del capítulo

El proyecto ya genera datos de valor que hoy se pierden por falta de estructura: las notas de visita y los chats informales de los representantes. Para El Laboratorio, el valor de las fuentes externas está en la información de mercado agregada, no en datos de individuos. La prioridad inmediata no es tecnológica sino de gobierno: definir qué se puede capturar y bajo qué autorización, antes de construir cualquier proceso de Text Analytics o ETL sobre estas fuentes no estructuradas.

CAPÍTULO 2. Arquitectura de la información

Objetivo y alcance del capítulo

En el capítulo 1 de la Parte II se identificaron fuentes de datos no estructuradas con valor para el proyecto (notas de visita en texto libre, chats informales de representantes, paneles de mercado externos) que hoy no tienen dónde almacenarse ni procesarse antes de llegar al Data Warehouse. Este capítulo define la arquitectura en la nube necesaria para capturar, almacenar y preparar esa información, evitando construir infraestructura sobredimensionada para un proyecto que aún está en fase piloto.

1. Análisis de alternativas de proveedores

El Data Warehouse de El Laboratorio ya opera sobre Microsoft Azure, y la empresa utiliza Microsoft 365 como suite corporativa. Esto condiciona fuertemente el análisis: adoptar un proveedor distinto implicaría duplicar conocimiento técnico, credenciales y procesos de integración, sin un beneficio claro para un proyecto piloto.

Tabla 20. Comparación de proveedores de nube

Proveedor	Ventaja para el proyecto	Desventaja
Microsoft Azure	Ya aloja el Data Warehouse; se integra de forma nativa con Power BI (capítulo 4 de la Parte I) y con Microsoft 365; el equipo de TI ya tiene conocimiento operativo de la plataforma	Ninguna relevante dado el contexto ya existente
AWS	Amplio catálogo de servicios de datos y IA	Requeriría una integración nueva con el Data Warehouse actual y curva de aprendizaje adicional para TI, sin necesidad real que lo justifique
Google Cloud	Buenas herramientas de procesamiento de lenguaje natural	Mismo problema que AWS: introduce un segundo proveedor sin sinergia con lo que ya existe

El Laboratorio ya tomó esta decisión de forma implícita al elegir Azure para su Data Warehouse. Evaluar otros proveedores sin una razón de peso sería ignorar una decisión ya tomada y encarecer el proyecto sin necesidad.

2. Selección de la alternativa de proveedor

Se selecciona Microsoft Azure como proveedor único, aprovechando la inversión y el conocimiento ya existente en TI/Sistemas, y la integración nativa con Power BI y Microsoft 365.

3. Definición del modelo de cloud computing

Se recomienda un modelo PaaS (Platform as a Service), en lugar de IaaS o SaaS puro, por las siguientes razones:

- IaaS exigiría que alguien administre servidores virtuales, sistemas operativos y parches — una carga operativa que no tiene sentido para un equipo sin infraestructura dedicada (recordar que la responsabilidad de datos del proyecto recae en un perfil de Comercial con dedicación parcial, capítulo 5 de la Parte I).
- SaaS puro (aplicaciones ya armadas) no ofrece la flexibilidad necesaria para procesar texto libre con reglas propias del negocio farmacéutico.
- PaaS entrega servicios ya administrados por Azure (almacenamiento, cómputo, procesamiento de lenguaje) sin que el equipo del proyecto tenga que operar infraestructura, mientras conserva la flexibilidad de configurar el procesamiento según las necesidades del proyecto.

Zona de aterrizaje ligera para datos no estructurados

Se propone una zona de aterrizaje ("landing zone") acotada al tamaño real del proyecto, evitando replicar una arquitectura de nivel corporativo con capacidades que este piloto no necesita. De las capacidades típicas de una Landing Zone, solo se seleccionan las que resuelven una necesidad concreta ya identificada en capítulos anteriores:

Tabla 21. Capacidades de la zona de aterrizaje y servicios de Azure propuestos

Capacidad	Servicio Azure propuesto	Necesidad que resuelve
Almacenamiento distribuido de bajo costo	Azure Blob Storage / Data Lake Storage Gen2	Guardar las notas de visita exportadas, los chats (una vez resuelto el consentimiento) y los archivos de paneles de mercado comprados
Procesamiento de lenguaje natural / Text Analytics	Azure AI Language (Cognitive Services)	Ejecutar la minería de texto sobre notas de visita definida en el capítulo 1 de la Parte II, sin construir un modelo propio desde cero
Integración de archivos planos	Azure Data Factory (conector de archivos)	Cargar de forma programada los archivos del panel de mercado externo (ej. IQVIA) hacia el Data Warehouse
Autenticación y autorización	Microsoft Entra ID (Azure AD)	Aplicar el control de acceso diferenciado que se identificó como pendiente en el capítulo 4 de la Parte I, especialmente relevante una vez se incorpore la base de contacto de dependientes
Auditoría y cumplimiento	Azure Monitor / registros de actividad	Dejar trazabilidad de quién accede a los datos, requisito derivado de la

Capacidad	Servicio Azure propuesto	Necesidad que resuelve
		LOPDP (capítulo 1 de la Parte II)

> **Capacidades explícitamente descartadas para este proyecto: motores NoSQL, plataformas de procesamiento de streams a gran escala (el caso de uso de near real time del capítulo 5 de la Parte II se resuelve con Azure Functions y una cola de mensajes, sin plataforma de streaming), y aceleradores especializados tipo SAS. Estas capacidades son propias de arquitecturas de datos a gran escala y no del volumen de datos que maneja un proyecto piloto de una sola línea de negocio. Incluirlas encarecería la solución sin aportar valor real en esta etapa.**

Estructura y codificación de los datos

Al igual que en el Data Warehouse actual, no se imponen restricciones estrictas sobre el formato de los archivos que ingresen a esta zona de aterrizaje (texto plano, CSV, o exportes del CRM); se recomienda, sin ser obligatorio, que todo archivo tabular incluya encabezado, para facilitar su interpretación por parte de Azure Data Factory y de quien administre el proceso.

4. Flujo de integración con el Data Warehouse

El flujo propuesto conecta las fuentes no estructuradas del capítulo 1 de la Parte II con el Data Warehouse ya existente (capítulo 3 de la Parte I), sin duplicar la arquitectura actual:

- 1) Fuentes no estructuradas: notas de visita del CRM, chats informales (sujetos a definición de consentimiento) y archivos de paneles de mercado.
- 2) Zona de aterrizaje (Azure Blob Storage): recepción y almacenamiento temprano de estos archivos, sin necesidad de transformarlos de inmediato.
- 3) Procesamiento (Azure AI Language + Azure Data Factory): extracción de temas/palabras clave de las notas de texto libre, y estructuración de los archivos de panel de mercado en formato tabular.
- 4) Carga hacia el Data Warehouse: los resultados ya estructurados se integran a las tablas existentes mediante el mismo proceso SQL que administra TI (capítulo 3 de la Parte I), evitando crear un segundo repositorio paralelo.
- 5) Visualización: los nuevos indicadores (por ejemplo, objeciones recurrentes por cadena) se incorporan al tablero de Power BI ya definido (capítulo 4 de la Parte I).

Consideración de costo

Los servicios propuestos (Blob Storage y Azure AI Language) se cobran por consumo, no por infraestructura fija. Para el volumen de datos de un proyecto piloto (notas de unas cuantas centenas de visitas al mes), el costo estimado es marginal —del orden de decenas de dólares mensuales—, por lo que se puede absorber dentro del presupuesto ya estimado en el capítulo 5 de la Parte I, sin representar una nueva partida significativa.

Conclusiones del capítulo

El Laboratorio no necesita una arquitectura de datos nueva ni un segundo proveedor de nube: la decisión correcta es extender el Azure que ya sostiene el Data Warehouse, usando un modelo PaaS que no exige operar infraestructura. La zona de aterrizaje propuesta se limita deliberadamente a las capacidades que resuelven necesidades ya identificadas en capítulos anteriores —almacenamiento, procesamiento de texto, integración de archivos y control de acceso—, evitando construir una plataforma de nivel corporativo que ni el tamaño del proyecto ni su presupuesto justifican.

CAPÍTULO 3. Arquitectura open source

Objetivo y alcance del capítulo

Una vez definida la zona de aterrizaje en Azure (capítulo 2 de la Parte II) para las fuentes no estructuradas del proyecto (notas de visita, chats, paneles de mercado), corresponde seleccionar de forma crítica qué herramientas open source de procesamiento de Big Data y qué lenguajes de programación complementan esa arquitectura — evitando adoptar por defecto herramientas pensadas para volúmenes que este proyecto no tiene.

1. Selección crítica del framework de Big Data

Frameworks como Apache Spark o Hadoop existen para resolver un problema específico: procesar volúmenes de datos que superan la capacidad de una sola máquina, distribuyéndolos en un clúster de múltiples nodos tolerante a fallos. Antes de adoptar uno de estos frameworks por ser el estándar de la industria, es necesario preguntarse si El Laboratorio tiene ese problema.

Tabla 22. Criterios para seleccionar el framework de Big Data

Criterio	Situación de El Laboratorio
Volumen de datos	Cientos de notas de visita al mes, drenaje mensual de ~1,079 puntos de venta, y paneles de mercado periódicos (capítulos 3 de la Parte I y 1 y 2 de la Parte II) — órdenes de magnitud por debajo de lo que justifica un clúster distribuido
Plan de escalamiento	Piloto acotado a Ecuador, sin planes confirmados de expansión a otros países o líneas de producto
Capacidad del equipo	El Responsable de Datos del Proyecto (capítulo 5 de la Parte I) no tiene hoy conocimientos de programación; solo maneja Excel y Power BI

> **Conclusión de la selección: no se adopta Apache Spark ni Hadoop para este proyecto. Ambos frameworks resuelven un problema de escala que El Laboratorio no tiene, y su adopción implicaría una complejidad operativa (gestión de clústeres, nodos, tolerancia a fallos) desproporcionada frente al volumen real de datos y a la capacidad actual del equipo. Adoptarlos solo porque son el estándar de la industria sería replicar una decisión de arquitectura sin replicar el problema que la justifica.**

En su lugar, se propone un procesamiento en un solo nodo, usando un lenguaje de programación de propósito general sobre los servicios ya definidos en el capítulo 2 de la Parte II (Azure Blob Storage, Azure AI Language), sin necesidad de un motor de clúster.

Criterio de reevaluación futura

Esta decisión no es permanente: si el proyecto escalara a múltiples países o líneas de producto, y el volumen de notas, conversaciones y transacciones creciera en uno o dos órdenes de magnitud, sí sería razonable reevaluar la adopción de Spark. La arquitectura propuesta en este capítulo no cierra esa puerta, simplemente no fuerza su apertura anticipadamente.

2. Lenguaje de programación complementario

Se selecciona Python como único lenguaje de programación del proyecto, descartando el uso combinado con R.

Tabla 23. Justificación de Python como único lenguaje de programación

Razón	Detalle
Preferencia y familiaridad	Es el lenguaje con el que el equipo ya tiene mayor afinidad, aun sin experiencia previa de programación
Soporte nativo en Azure	Los servicios ya seleccionados en el capítulo 2 de la Parte II (Azure AI Language, Azure Data Factory, Azure Functions) tienen SDKs y ejemplos de referencia en Python de primera clase
Un solo lenguaje para un equipo pequeño	Con una sola persona a cargo de los datos (dedicación parcial, capítulo 5 de la Parte I), sostener dos lenguajes (R y Python) duplicaría el esfuerzo de aprendizaje sin un beneficio claro

Decisión deliberada: R se descarta porque su ventaja (rigor estadístico) se justifica en equipos de analistas especializados, que El Laboratorio no tiene. Dividir el esfuerzo entre dos lenguajes sería replicar una división de trabajo que no existe en la estructura real del proyecto (organigrama propuesto en el capítulo 5 de la Parte I).

3. Complementos del ecosistema open source

Además del lenguaje, se seleccionan librerías y herramientas open source ya maduras, evitando construir nada desde cero:

Tabla 24. Complementos del ecosistema open source seleccionados

Componente	Herramienta open source	Uso en el proyecto
Manipulación de datos	pandas	Limpieza y estructuración de los archivos de panel de mercado y de las notas de visita exportadas del CRM
Procesamiento de lenguaje natural	spaCy / NLTK	Alternativa open source y sin costo a Azure AI Language para la minería de texto de las notas de visita (capítulo 1 de la Parte II), útil como respaldo o para pruebas antes de comprometerse con el servicio administrado

Componente	Herramienta open source	Uso en el proyecto
Control de versiones	Git	Trazabilidad de los scripts de procesamiento, aun con un equipo de una sola persona, para no perder el historial de cambios
Documentación y aprendizaje	Jupyter Notebooks	Entorno accesible para que el Responsable de Datos del Proyecto empiece a familiarizarse con Python de forma incremental, en línea con la capacitación ya presupuestada en el capítulo 5 de la Parte I
Modelado y aprendizaje automático	scikit-learn	Entrenamiento del árbol de decisión de priorización y del clasificador de texto de interacciones (capítulos 4 y 5 de la Parte II), sin costo de licenciamiento

Para el almacenamiento estructurado, se mantiene el motor SQL ya utilizado por TI (capítulo 3 de la Parte I), sin introducir una base de datos NoSQL adicional: el volumen y la naturaleza tabular de los datos del proyecto no lo requieren.

Conclusiones del capítulo

La arquitectura open source correcta para El Laboratorio no es la que usaría una organización con millones de registros, sino la que resuelve el problema real de este piloto con la menor complejidad operativa posible. Se descarta Apache Spark y Hadoop por ausencia de un problema de volumen que los justifique, se selecciona Python como único lenguaje para no fragmentar el esfuerzo de un equipo pequeño, y se complementa con librerías open source maduras (pandas, spaCy/NLTK) que no representan costo de licenciamiento y son coherentes con el presupuesto ya definido en el capítulo 5 de la Parte I.

CAPÍTULO 4. Implementación de inteligencia artificial

Objetivo y alcance del capítulo

Un modelo predictivo de aceptación o fidelización requiere, para entrenarse, años de historial de interacciones ya aceptadas o rechazadas. El proyecto de El Laboratorio no tiene ese tipo de historial: el canal virtual con dependientes todavía no existe, por lo que no hay datos de «aceptación» o «fidelización» que un modelo pueda aprender hoy. Construir ese modelo ahora sería un ejercicio sin datos reales para entrenarlo.

En cambio, sí existen entre 1 y 2 años de historial de visitas y drenaje en el Data Warehouse (capítulo 3 de la Parte I), suficientes para entrenar un modelo distinto pero igual de útil en esta etapa: priorizar, dentro de la capacidad limitada de representantes (capítulo 1 de la Parte I), qué puntos de venta necesitan refuerzo con mayor urgencia.

1. Objetivo de aplicación del modelo de Data Science / Machine Learning

Modelo de Priorización de Cobertura y Riesgo de Bajo Desempeño: predice, para cada punto de venta, la probabilidad de que su desempeño (rotación/drenaje) se deteriore dado el nivel de cobertura de visitas que está recibiendo actualmente, con el fin de priorizar dónde reforzar la visita presencial o, a futuro, el canal virtual.

> **Este modelo no aumenta la capacidad total de los representantes (siguen siendo 8, con ~35% de cobertura posible según el capítulo 1 de la Parte I). Su valor está en decidir mejor a quién visitar primero dentro de esa misma capacidad restringida, en lugar de reforzar el reparto fijo por clasificación A/B/C/D.**

Diagrama de aplicación del modelo

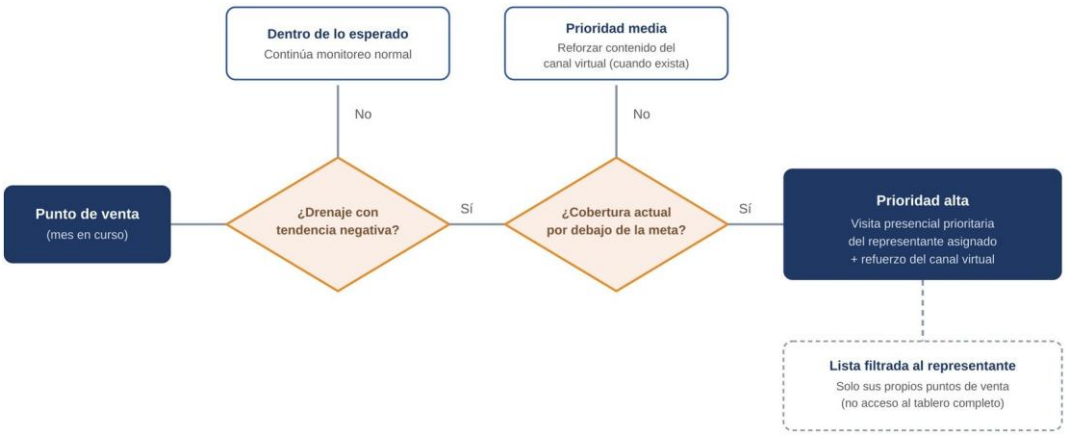


Figura 3. Árbol de decisión del modelo de priorización en su etapa inicial — se ejecuta mensualmente sobre los datos ya existentes (el capítulo 5 de la Parte II lo amplía con fuentes nuevas).

2. Fuentes de datos utilizadas

El modelo se construye en dos etapas. En la etapa inicial usa únicamente datos que ya existen en el Data Warehouse (capítulo 3 de la Parte I), sin requerir ninguna fuente nueva. En la segunda etapa incorpora las fuentes no estructuradas y del canal virtual (notas de visita, notas de voz, app y web), que se detallan en el capítulo 5 de la Parte II, donde además se define el escenario de trabajo del piloto.

Tabla 25. Fuentes de datos y variables del modelo de priorización

Fuente	Variables utilizadas
CRM de visitas (interno)	Frecuencia real de visita vs. meta por clasificación (A/B/C/D), representante asignado, historial de 1-2 años
Drenaje por punto de venta (cadenas)	Tendencia de ventas/rotación mensual del punto de venta, historial de 1-2 años
Notas de visita del CRM, texto libre (etapa 2)	Temas y objeciones recurrentes por punto de venta, obtenidos con minería de texto (capítulo 1 de la Parte II)
Notas de voz, app y web del canal virtual (etapa 2)	Número de interacciones y alertas por punto, tiempo de respuesta, última interacción y activación del dependiente (capítulo 5 de la Parte II)

La etapa inicial es deliberadamente simple de construir, porque no requiere ningún dato adicional al que ya circula por el Data Warehouse; las fuentes de la segunda etapa se incorporan solo después de resolver las condiciones de consentimiento y gobierno.

3. Área de aplicación y departamentos de negocio implicados

Tabla 26. Áreas implicadas y uso del modelo

Área	Uso del modelo
Dirección Comercial	Visualiza la priorización agregada por cadena y clasificación en el tablero (capítulo 4 de la Parte I), para ajustar la estrategia comercial del piloto
Representantes	Reciben una lista filtrada, limitada a sus propios puntos de venta asignados, con el orden de prioridad sugerido para el mes

Esto resuelve, de forma acotada, el punto que quedó pendiente en el capítulo 4 de la Parte I: si los representantes debían tener acceso al tablero. La respuesta no es darles acceso al tablero completo (que expone información de otras cadenas y representantes), sino una salida específica y filtrada del modelo, limitada a la información que necesitan para actuar. Esta distinción —dato operativo puntual vs. acceso a la herramienta de análisis completa— es la misma que se usó en el capítulo 2 de la Parte II para diferenciar accesos.

4. Identificación de la mejora esperada

Hoy, la frecuencia de visita se asigna únicamente según la clasificación estática del punto de venta (A, B, C o D), sin importar si su desempeño real está mejorando o deteriorándose. El modelo permite pasar de una asignación fija por categoría a una asignación dinámica dentro de esa misma categoría, dando prioridad a los puntos con mayor riesgo detectado.

5. Cuantificación de la mejora: business case

Costo

El modelo se construye con los recursos ya presupuestados en capítulos anteriores, sin costo incremental significativo:

- Lenguaje y librerías: Python + pandas + scikit-learn (capítulo 3 de la Parte II), sin costo de licenciamiento.
- Responsable: el mismo Responsable de Datos del Proyecto (capítulo 5 de la Parte I), con un estimado de 20-30 horas de dedicación única para el desarrollo y validación inicial del modelo, dentro de su dedicación parcial ya presupuestada.

Beneficio: por qué no se cuantifica en ventas todavía

Sería posible estimar una cifra de "ventas recuperadas" atribuible al modelo, pero hacerlo sin evidencia sería repetir el mismo error que se evitó deliberadamente en el capítulo 5 de la Parte I: forzar una cifra financiera sin datos reales que la respalden. El beneficio real y verificable en esta etapa es la mejora en la asignación de un recurso ya escaso (35% de cobertura posible), no un incremento de ventas todavía no observado.

> **Recomendación de validación: antes de escalar el modelo a todos los puntos de venta, se sugiere una prueba controlada de 2-3 meses comparando un grupo de puntos priorizados por el modelo contra un grupo con la asignación tradicional por clasificación, midiendo la diferencia real en el drenaje. Solo con ese resultado se podrá construir un business case financiero defendible, en lugar de uno especulativo. El escenario de trabajo, el punto de equilibrio y la regla de decisión de esa prueba se definen en el capítulo 5 de la Parte II.**

6. Validación técnica del modelo

Variable objetivo (a definir con la Dirección Comercial): se propone etiquetar como «deterioro» a un punto de venta cuyo drenaje del mes siguiente cae por debajo de su promedio de los tres meses previos en un umbral a fijar (por ejemplo, 15%). Variables explicativas: frecuencia real de visita frente a la meta de su clasificación, tendencia del drenaje, clasificación A/B/C/D y ciudad; en la segunda etapa, los temas de las notas de visita y los indicadores de interacción.

Validación: división temporal (se entrena con los primeros meses del historial y se prueba con los últimos 3 a 6, nunca al azar, para no usar información del futuro). El árbol se compara contra una línea base —la asignación actual por clasificación A/B/C/D— y se mide el recall de los puntos que efectivamente se deterioran dentro de los N puntos que los representantes pueden cubrir en el mes, junto con la precisión de esa lista. Se limita la profundidad a 3 o 4 niveles para que cada recomendación sea explicable al representante.

Riesgo y criterio de descarte: con 1,079 puntos y entre 12 y 24 meses de historial hay entre unas 13,000 y 26,000 observaciones, pero los eventos de deterioro pueden ser pocos y desbalanceados; por eso la métrica principal es el recall de los puntos que se deterioran y no la exactitud global. Si el modelo no supera a la línea base en la prueba temporal, se mantiene la asignación por clasificación.

Conclusiones del capítulo

El Laboratorio puede empezar a usar inteligencia artificial en este proyecto sin esperar a que el canal virtual con dependientes exista: los datos de visitas y drenaje que ya tiene son suficientes para un modelo de priorización que mejora la asignación de un recurso limitado. Este modelo, además, resuelve de forma concreta una decisión de gobierno que había quedado abierta desde el capítulo 4 de la Parte I sobre el acceso de los representantes a la información. El paso pendiente no es tecnológico sino de validación: correr una prueba controlada antes de convertir la mejora esperada en una cifra financiera del business case.

CAPÍTULO 5. Caso de uso en near real time

Objetivo y alcance del capítulo

El problema que atraviesa todo el proyecto es de tiempo: el laboratorio se entera de lo que pasa en un punto de venta tarde. La visita presencial llega, como máximo, una vez al mes en los puntos D (capítulo 1 de la Parte I), y el drenaje que reportan las cadenas llega «a mes caído» (capítulo 3 de la Parte I). Este capítulo propone el único componente del proyecto que genera información nueva con una latencia de minutos, en lugar de semanas: un sistema que recibe las interacciones de los dependientes, las clasifica y alerta a la persona que debe actuar.

Como base conceptual se tomaron tres ideas del curso: la toma de decisiones en tiempo real como una asignación que combina datos de contexto con una regla de decisión; el uso de las interacciones para mejorar la atención al cliente, con la pregunta de fondo sobre el uso de los datos con consentimiento; y la distinción entre el análisis en tiempo real, el de tendencias y el de patrones, junto con el criterio de evaluación de inversiones mediante el Valor Actual Neto (VAN).

1. Objetivo de aplicación del modelo de real time

Sistema de alertas y enrutamiento de interacciones de dependientes. Cada vez que un dependiente escribe, envía una nota de voz o interactúa con la app o la web del laboratorio, el sistema identifica quién es, en qué punto de venta trabaja y de qué trata su mensaje (desabastecimiento, objeción sobre un producto, queja, consulta, solicitud de muestra). Según esa clasificación, avisa al representante o al Key Account responsable, responde con contenido ya aprobado, o deriva el caso al área que corresponde.

Por qué near real time y no real time

En un detector de fraude, la decisión debe tomarse en milisegundos porque el daño ocurre en el instante de la transacción. Aquí la decisión final siempre es humana: un representante que visita un punto o un Key Account que llama a una cadena. Que la alerta llegue en menos de cinco minutos es suficiente; reducirla a milisegundos no cambiaría ningún resultado de negocio y encarecería la solución. La clasificación correcta de cada flujo de datos del proyecto es la siguiente:

Tabla 27. Clasificación de los flujos de datos según su ritmo

Flujo de datos	Origen	Ritmo actual → objetivo	Clasificación
Drenaje por punto de venta	Cadenas de farmacias	Mensual, «a mes caído» → mensual (no depende del laboratorio)	Batch mensual
Visitas y clasificación A/B/C/D	CRM interno	Mensual → diario o semanal (recomendación del capítulo 3 de la Parte I)	Batch frecuente

Flujo de datos	Origen	Ritmo actual → objetivo	Clasificación
Notas de visita en texto libre	CRM interno	Mensual → diario	Batch frecuente
Interacciones de dependientes	Canal virtual (nuevo)	No existe → ≤ 5 minutos hasta la alerta	Near real time
Tablero de control	Power BI (capítulo 4 de la Parte I)	Actualización diaria; la licencia Pro permite hasta 8 por día	Batch diario

Las alertas no pasan por el tablero: llegan directamente al representante, por lo que el límite de actualizaciones de la licencia Power BI Pro (capítulo 4 de la Parte I) no condiciona este caso de uso.

Fases del sistema

Tabla 28. Fases del sistema de alertas y enrutamiento

Fase	Qué hace	Herramienta (coherente con capítulos 2 y 3 de la Parte II)
1. Recepción	Recibe cada interacción y transcribe las notas de voz a texto	Azure Functions + cola de mensajes; Azure AI Speech para audio
2. Filtro	Verifica que el dependiente esté registrado y con consentimiento vigente; descarta duplicados y mensajes no válidos. Si no está identificado, solo se almacena y no se analiza	Python; consulta a la base de dependientes
3. Contexto	Cruza al dependiente con su punto de venta, cadena, clasificación A/B/C/D y representante asignado	SQL sobre el Data Warehouse existente
4. Clasificación	Asigna tema, producto mencionado, urgencia y grupo regulatorio (OTC / no OTC)	Reglas + modelo de clasificación de texto (spaCy o Azure AI Language)
5. Acción	Alerta al representante o al Key Account; responde con contenido aprobado; o deriva a Legal / Información médica	Notificación al representante; plantillas aprobadas
6. Aprendizaje	Guarda la interacción, la acción tomada y su resultado; reentrena el clasificador cada mes	Data Warehouse; Power BI; modelo del capítulo 4 de la Parte II



Figura 4. Flujo del caso de uso en near real time, de la interacción del dependiente a la acción y al aprendizaje.

Diseño de la solución y del algoritmo

Escala: con unos 135 dependientes activos en el piloto y alrededor de 4 interacciones al mes cada uno, el flujo sería de unas 25 interacciones diarias; a escala nacional serían unas 180 por día. Es una corriente de datos muy pequeña. Por la misma razón expuesta en el capítulo 3 de la Parte II, no se justifican plataformas de streaming como TIBCO o Spark Streaming, diseñadas para flujos de miles de eventos por segundo.

Cada interacción se describe con atributos simples, equivalentes a los que usa un detector de anomalías: canal (texto, voz, clic), tema, producto mencionado, categoría regulatoria (OTC o no OTC), urgencia o tono, cadena, clasificación del punto de venta, representante asignado y hora del mensaje.

El algoritmo es un clasificador de texto, no un detector estadístico de anomalías. Se arranca con reglas y un diccionario de los 5 o 6 productos de la grilla de cada cadena, y se incorpora un modelo supervisado entrenado con unas 400 notas de visita históricas etiquetadas manualmente (unas 15 horas entre el Responsable de Datos y un representante). Como en el caso de fraude, las clases están desbalanceadas: la mayoría de las interacciones serán consultas simples y pocas serán desabastecimientos o menciones reguladas. Por eso se fijan dos metas de calidad distintas: recall mínimo de 90% en la detección de menciones de productos no OTC (un falso negativo es un riesgo regulatorio) y precisión mínima de 70% en las alertas al representante (para evitar la fatiga de alertas).

No se usan modelos generativos para responder consultas científicas. En un laboratorio, una respuesta inventada o mal formulada sobre un medicamento es un riesgo mayor que cualquier ahorro de tiempo; las respuestas automáticas se limitan a contenido OTC previamente aprobado.

Riesgo de canal que condiciona el diseño: la política de mensajería de WhatsApp Business (actualizada el 23 de septiembre de 2026) incluye entre los usos prohibidos las «drogas, sean de prescripción, recreativas o de otro tipo», y la documentación de proveedores autorizados de la plataforma indica que los mensajes sobre medicamentos con receta están prohibidos y que los OTC se tratan como una vertical regulada. A esto se suma la restricción ecuatoriana sobre divulgar productos no OTC en redes sociales (capítulo 1 de la Parte I). Por eso el sistema se diseña independiente del canal: las fases 2 a 6 no cambian si la entrada es WhatsApp, un portal web con registro por QR o notificaciones por correo o SMS. Antes de construir se plantea una compuerta de validación (Legal y el proveedor de WhatsApp Business) para confirmar por escrito qué contenido puede circular por cada canal.

2. Fuentes de datos utilizadas

Este caso de uso amplía las fuentes del modelo de priorización del capítulo 4 de la Parte II, que en su etapa inicial se alimenta solo de datos mensuales del CRM y del drenaje. Las fuentes se incorporan por fases, según su disponibilidad y las condiciones de uso:

Tabla 29. Fuentes de datos del caso de uso y condiciones de gobierno

Fuente	Estado hoy	Uso en el caso de uso	Condición de gobierno / legal
Notas de visita del CRM (texto libre)	Existe	Entrenar y validar el clasificador; derivar temas por punto de venta	Quitar nombres propios antes de procesar (minimización)
Visitas, clasificación A/B/C/D y drenaje	Existe en el Data Warehouse	Contexto de cada interacción; variable de resultado para verificar casos	Ya gobernado por TI (capítulo 3 de la Parte I)
Chats informales de WhatsApp entre representantes y dependientes	Existe, disperso en dispositivos personales	Solo como muestra para etiquetar, si hay autorización	No usar sin informar y obtener consentimiento (LOPDP, capítulo 1 de la Parte II)
Mensajes de texto del canal virtual	No existe	Núcleo del caso de uso	Opt-in expreso y registro previo del dependiente
Notas de voz	No existe	Se transcriben con Azure AI Speech; se guarda el texto y se descarta el audio original	Consentimiento específico para voz; minimización del dato
Eventos de app y web (registro por QR, aperturas, clics, formularios)	No existe	Medir activación y engagement; señal de abandono	Aviso de privacidad y consentimiento en el registro
Panel de mercado farmacéutico externo	A evaluar (capítulo 1 de la Parte II)	Validar cruzadamente el drenaje reportado por las cadenas	Solo datos agregados

Cómo estas fuentes alimentan el modelo de priorización del capítulo 4 de la Parte II

- Temas recurrentes en las notas de visita por punto de venta (por ejemplo, objeciones repetidas sobre un producto): variable de riesgo que el drenaje mensual solo mostraría semanas después.
- Número de alertas por punto, tiempo de respuesta y última interacción: indican si un punto está «frío» aunque su cobertura de visitas sea la esperada.
- Activación del dependiente (registrado, activo, inactivo): permite distinguir un punto bien cubierto de un punto con contacto real.

Con estas variables, el árbol de decisión del capítulo 4 de la Parte II deja de apoyarse únicamente en datos con un mes de rezago y puede recalcular la prioridad semanalmente.

3. Área de aplicación y departamentos de negocio implicados

Tabla 30. Áreas implicadas y rol en el caso de uso

Área	Rol en el caso de uso
Dirección Comercial	Patrocinadora. Decide si se escala, se ajusta o se detiene el piloto según la regla del business case
Representantes	Reciben las alertas de sus propios puntos de venta (y no el tablero completo, según lo definido en el capítulo 4 de la Parte II) y registran el resultado de cada acción
Key Account	Recibe las alertas que se repiten dentro de una misma cadena (por ejemplo, un desabastecimiento en varios puntos)
Responsable de Datos del Proyecto	Opera el clasificador, revisa la calidad de las alertas, audita por muestreo los casos recuperados y reentrena el modelo
TI / Sistemas	Provee la infraestructura en Azure e integra el flujo con el Data Warehouse
Legal / Regulatorio (nuevo actor)	Valida la política de cada canal, el texto de consentimiento y la lista de contenido OTC aprobado
Marketing / Visita médica	Aprueba el contenido científico de las respuestas automáticas y atiende las consultas no OTC derivadas. Hoy Marketing solo cubre visita médica (capítulo 1 de la Parte I), por lo que esta es una coordinación nueva

4. Identificación de la mejora esperada

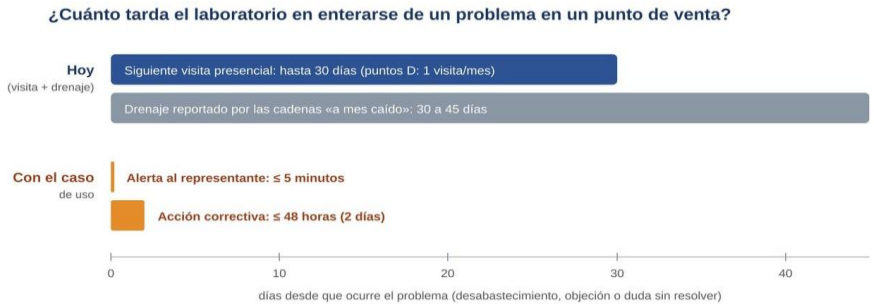


Figura 5. Tiempo hasta que el laboratorio detecta un problema en un punto de venta, hoy y con el caso de uso.

Tabla 31. Mejora esperada por indicador

Indicador	Hoy	Con el caso de uso	Cómo se mide
Tiempo hasta detectar un desabastecimiento u objeción	Hasta 30 días (siguiente visita) o 30 a 45 días (drenaje)	≤ 5 min hasta la alerta; acción en ≤ 48 h	Marca de tiempo del mensaje y de la acción registrada
Tiempo de respuesta a una duda del dependiente	Hasta la siguiente visita	Misma jornada laboral para consultas OTC	Registro del canal
Contacto efectivo en puntos C y D	1 a 3 visitas al mes	Visitas más contacto digital continuo	Dependientes activos por punto de venta
Control de cumplimiento sobre productos no OTC	No existe un control sistemático	Toda mención marcada y derivada (recall ≥ 90%)	Auditoría mensual de una muestra

5. Cuantificación de la mejora: business case

Sin un escenario de trabajo definido, es muy difícil saber si un piloto tuvo éxito. Este business case lo define de forma explícita: tres escenarios con supuestos declarados, un punto de equilibrio, una métrica de éxito y una regla de decisión fijada antes de empezar. Todas las cifras son estimaciones calculadas con los supuestos de la tabla siguiente y deben validarse con la Dirección Comercial; no son resultados observados.

5.1 Escenarios de trabajo

El piloto dura 4 meses y cubre 150 puntos de venta, con un grupo de control de otros 150 puntos equivalentes en cadena y clasificación.

Tabla 32. Escenarios de trabajo del piloto

Variable	Conservador	Base	Optimista	Fundamento
Dependientes por punto de venta	2	2	2	Dependientes de turno; alta rotación

Variable	Conservador	Base	Optimista	Fundamento
Activación (con consentimiento y activos)	25%	45%	60%	A validar en las primeras semanas del piloto
Señales accionables por dependiente activo al mes	0.25	0.40	0.50	Desabastecimiento, objeción, queja o solicitud
Alertas atendidas en ≤ 48 h	50%	70%	80%	Capacidad del representante (capítulo 1 de la Parte I)
Alertas atendidas que recuperan una venta	20%	35%	45%	Verificado con el drenaje siguiente
Margen recuperado por caso (USD)	20	30	35	≈ USD 70 de ventas en riesgo por caso × margen de contribución ≈ 43%

Supuestos de base a validar con la Dirección Comercial: ventas de productos del laboratorio de unos USD 600 al mes por punto de venta (≈ USD 100 por producto de la grilla) y un problema detectado unas 3 semanas antes de lo que ocurre hoy.

5.2 Resultado mensual esperado del piloto (150 puntos)

Tabla 33. Resultado mensual esperado del piloto (150 puntos de venta)

Concepto	Conservador	Base	Optimista
Dependientes activos	75	135	180
Alertas al mes	19	54	90
Alertas atendidas en ≤ 48 h	9	38	72
Casos con venta recuperada	2	13	32
Margen recuperado (USD/mes)	38	397	1,134
Costo mensual de operación (USD)	192	254	321
Resultado neto mensual (USD)	−154	+143	+813
Casos de equilibrio (casos/mes que cubren el costo)	≈ 10	≈ 9	≈ 9

El costo mensual del piloto en el escenario base se compone de USD 90 de dedicación adicional del Responsable de Datos (cerca de un 10% de su salario de referencia de USD 900, sobre el 30% ya presupuestado en el capítulo 5 de la Parte I), USD 59 de tiempo de atención de alertas, USD 50 de plataforma de mensajería, USD 40 de mensajes salientes y USD 15 de servicios en la nube (la capa gratuita de Azure AI Language cubre 5,000 registros de texto al mes). Los mensajes se valoran a la tarifa de marketing de WhatsApp para el resto de Latinoamérica (≈ USD 0.074 por mensaje), por lo que representan una cota superior: los de servicio y los de otros canales cuestan menos.

A este costo se suma una inversión inicial de unos USD 1,200: desarrollo del flujo (≈ USD 374), etiquetado de notas (≈ USD 84), revisión legal y textos de consentimiento (≈ USD 500) y una contingencia cercana al 25%. En cuatro meses, el resultado acumulado del piloto sería de −USD 1,817 (conservador), −USD 629 (base) y +USD 2,053 (optimista); es decir, la pérdida máxima que se arriesga para obtener información real es de unos USD 1,800.

5.3 Métrica de éxito y regla de decisión

La métrica principal es el número de casos con venta recuperada y verificada por mes. Un caso se considera verificado cuando existe una alerta, una acción correctiva registrada por el representante y una confirmación independiente: que el producto en ese punto supere su tendencia de los 3 meses previos en el siguiente reporte de drenaje, o que la confirme el Key Account. El Responsable de Datos audita el 20% de los casos para evitar que el registro propio de los representantes infle los resultados. El punto de equilibrio está entre 9 y 10 casos verificados por mes.

Tabla 34. Regla de decisión al cierre del mes 4

Decisión al cierre del mes 4	Condición
Escalar a más puntos de venta	≥ 13 casos verificados por mes durante 2 meses consecutivos (escenario base), activación ≥ 45%, ≥ 70% de alertas atendidas en 48 h, recall regulatorio ≥ 90% y ningún incumplimiento de política de canal o normativa
Ajustar y extender 3 meses	Entre 9 y 12 casos verificados por mes, o activación entre 25% y 45% con tendencia creciente
Detener o cambiar de canal	Menos de 9 casos verificados por mes, activación menor a 25%, o cualquier bloqueo de cumplimiento (política de la plataforma o normativa)

5.4 Cómo se mide: por qué no basta comparar ventas

La comparación de ventas entre los 150 puntos del piloto y los 150 de control es un control de seguridad, pero no puede ser el criterio de decisión. Suponiendo una variabilidad mensual de ventas por punto del 20% y promediando 3 meses, con 150 puntos por grupo solo se distinguirían diferencias de aproximadamente 3.7% o más. El efecto esperado equivale a cerca de 0.1% de las ventas del piloto en el escenario conservador, 1.0% en el base y 2.9% en el optimista: ninguno superaría ese umbral. Por eso la decisión se apoya en los casos verificados, y la prueba comparativa sobre ventas recién sería concluyente en una segunda fase con unos 400 puntos por grupo (diferencia detectable de aproximadamente 2.3%).

5.5 Evaluación de la inversión a escala nacional

Siguiendo el criterio de evaluación de inversiones mediante VAN, se evalúa la decisión de escalar a los 1,079 puntos de venta con VAN a 3 años y una tasa de descuento del 12%

(supuesto). Se considera una inversión de despliegue de USD 2,500 (principalmente el tiempo de los representantes para registrar a unos 2,158 dependientes), un primer año con 50% de la operación y dos años a régimen.

Supuestos de escala: el margen recuperado y los costos variables del piloto (tiempo de atención de alertas y mensajes: USD 99 de los USD 254 mensuales del escenario base) se multiplican por 7.19, la razón entre 1,079 y 150 puntos de venta; los costos fijos (dedicación del Responsable de Datos, plataforma y servicios en la nube: USD 155 al mes) no se escalan. En el primer año opera el 50% del volumen de casos, con los costos fijos completos. Resultado neto mensual a régimen: −USD 148 (conservador), +USD 1,989 (base) y +USD 6,808 (optimista). Son estimaciones simplificadas: no incluyen soporte adicional ni rotación de dependientes, y deben validarse con TI y la Dirección Comercial.

Tabla 35. VAN a 3 años de la escala nacional

Escenario	VAN a 3 años (USD)	Lectura
Conservador	−6,800	No escalar: los costos de operación superan el beneficio
Base	+43,300	Escalar si el piloto confirma al menos este escenario; la inversión se recupera en menos de 3 meses de operación (con el 50% de la operación prevista para el primer año)
Optimista	+156,400	Cota superior; no se usa como criterio de aprobación

No se reporta la TIR porque la inversión inicial es muy pequeña frente a los flujos anuales y el indicador pierde utilidad.

Sensibilidad: en el escenario base, el piloto sigue cubriendo su costo mensual mientras la tasa de recuperación se mantenga sobre 22% (el base supone 35%) o la activación sobre 23% (el base supone 45%). El caso soporta una caída de la tasa de recuperación de cerca de un tercio (de 35% a 22%) o de la activación de cerca de la mitad (de 45% a 23%), pero no ambas a la vez.

Próximos pasos

- Semana 1: compuerta de canal. Validar por escrito con Legal y el proveedor de mensajería qué contenido puede circular por WhatsApp; si no es viable, arrancar con el portal web con registro por QR.
- Semanas 1 a 3: etiquetar unas 400 notas de visita, construir el flujo y probar el clasificador contra la meta de recall y precisión.
- Validar con la Dirección Comercial los supuestos de ventas por punto, margen y dependientes por punto antes de iniciar.
- Meses 1 a 4: ejecutar el piloto con grupo de control, medir los casos verificados cada mes y decidir según la regla del apartado 5.3.

Conclusiones del capítulo

El caso de uso en near real time es la parte del proyecto que resuelve de raíz el problema de latencia que el drenaje mensual nunca podrá resolver, porque depende de las cadenas: crea una señal nueva, propia del laboratorio, que llega en minutos. No requiere tecnología de streaming de gran escala, sino un flujo simple sobre los servicios de Azure y el Python ya definidos en capítulos anteriores, con una persona siempre a cargo de la acción.

Su viabilidad no depende de la tecnología sino de dos condiciones que hoy están abiertas: que el canal elegido sea compatible con las políticas de plataforma y con la normativa sobre productos no OTC, y que los dependientes acepten registrarse. El business case lo reconoce: bajo el escenario conservador no vale la pena escalar, bajo el base sí, y la diferencia entre ambos se decide con datos del piloto, con una regla fijada antes de empezar. Esa regla es lo que permite afirmar, al cierre del mes 4, si el proyecto tuvo éxito o no.

CONCLUSIONES Y APLICACIONES

Síntesis del proyecto

El proyecto parte de una constatación simple: El Laboratorio solo conoce a los dependientes de farmacia a través de una visita presencial que, con 8 representantes y 1,079 puntos de venta, cubre alrededor del 35% de la demanda de visitas definida por su propia política (0.98 visitas por punto y mes frente a una meta ponderada de 2.8). A lo largo de los diez capítulos se construyó, sobre los datos y la infraestructura que la compañía ya tiene, una respuesta proporcionada al tamaño de un piloto: un canal virtual medible, un tablero de control, una arquitectura ligera, un modelo de priorización y un sistema de alertas en near real time.

Principales hallazgos y decisiones

- **Datos y oportunidad.** El proceso ETL actual es automático, validado y bien gobernado por TI, pero llega tarde: el CRM de visitas y el drenaje se cargan mensualmente. Solo el drenaje depende de terceros, por lo que se recomienda desacoplar ambos flujos y llevar el CRM a una frecuencia semanal o diaria (Parte I, capítulo 3).
- **Visualización y rol de datos.** Power BI, ya estandarizado en la empresa, es suficiente. Basta con asignar una responsabilidad de datos a un perfil existente de Comercial, con una dedicación parcial de cerca del 30% y un costo acotado de entre USD 3,240 y 4,116 en el primer año. El ahorro de tiempo (~USD 1,372 al año) no cubre ese costo, así que la inversión se justifica por razones estratégicas y no por ahorro operativo (Parte I, capítulos 4 y 5).
- **Datos no estructurados y arquitectura.** El valor está en las notas de visita, en los chats informales (con consentimiento) y en paneles de mercado agregados, no en comprar datos personales. La arquitectura se limita a Azure en modelo PaaS y a Python con librerías abiertas; se descartan Spark y Hadoop porque el volumen no los justifica (Parte II, capítulos 1 a 3).
- **Inteligencia artificial.** Sin historial del canal virtual, el modelo viable hoy es un árbol de decisión de priorización de cobertura entrenado con uno a dos años de visitas y drenaje. No aumenta la capacidad de los representantes: mejora a quién visitar primero dentro de ella (Parte II, capítulo 4).
- **Near real time.** El sistema de alertas reduce el tiempo hasta detectar un problema en un punto de venta, de hasta 30 a 45 días, a menos de cinco minutos hasta la alerta. El piloto de 4 meses, con 150 puntos de venta y un grupo de control de otros 150, cuesta entre USD 192 y 321 al mes; su punto de equilibrio está entre 9 y 10 casos con venta recuperada y verificada por mes; el escenario base arroja un resultado neto de +USD 143 mensuales y la pérdida máxima que se arriesga para obtener información real es de unos USD 1,800 (Parte II, capítulo 5).

Aplicaciones

Los resultados se traducen en una ruta de ejecución con puntos de control claros:

- Semana 1: validar por escrito con Legal y el proveedor de mensajería qué contenido puede circular por cada canal; si WhatsApp no es viable, iniciar con un portal web con registro por QR.
- Semanas 1 a 3: etiquetar unas 400 notas de visita, construir el flujo en Azure y Python, y probar el clasificador contra las metas de recall (≥ 90% en menciones de productos no OTC) y de precisión (≥ 70% en las alertas al representante).
- Meses 1 a 4: ejecutar el piloto con grupo de control y decidir al cierre del mes 4 según una regla fijada de antemano: escalar (≥ 13 casos verificados por mes durante dos meses consecutivos, activación ≥ 45% y cumplimiento normativo), ajustar y extender tres meses (entre 9 y 12 casos) o detener (menos de 9 casos, activación menor a 25% o bloqueo de cumplimiento).
- Si el piloto confirma al menos el escenario base, evaluar la escala a los 1,079 puntos de venta: con una tasa de descuento del 12%, el VAN a tres años es de +USD 43,300 en el escenario base y de −USD 6,800 en el conservador.

Limitaciones

Las cifras económicas son estimaciones basadas en supuestos declarados (ventas de unos USD 600 al mes por punto de venta, margen de contribución cercano al 43% y activación de dependientes entre 25% y 60%) que deben validarse con la Dirección Comercial antes de iniciar. Con 150 puntos por grupo, una comparación de ventas solo detectaría diferencias de aproximadamente 3.7% o más; por eso la decisión del piloto se apoya en casos verificados y no en una prueba de ventas, que sería concluyente recién en una segunda fase con unos 400 puntos por grupo. La viabilidad depende, además, de dos condiciones todavía abiertas: que el canal elegido sea compatible con las políticas de la plataforma y con la normativa sobre productos no OTC, y que los dependientes acepten registrarse con consentimiento expreso.

Reflexión final

El principal aprendizaje es de método. En un proyecto de datos de tamaño piloto, la decisión correcta suele ser la más simple que resuelve el problema real —Power BI antes que una plataforma nueva, un solo nodo con Python antes que un clúster, un árbol de decisión antes que un modelo complejo— y declarar con honestidad lo que todavía no se puede medir. Sobre esa base, la compañía puede aprender con un riesgo acotado y decidir con evidencia si escala, ajusta o detiene el canal.

REFERENCIAS

Agencia Nacional de Regulación, Control y Vigilancia Sanitaria [ARCSA]. (s. f.). *Sitio institucional*. https://www.controlsanitario.gob.ec/

Asamblea Nacional Constituyente del Ecuador. (2008). *Constitución de la República del Ecuador* (Registro Oficial 449, 20 de octubre de 2008).

Asamblea Nacional del Ecuador. (2021). *Ley Orgánica de Protección de Datos Personales* (Registro Oficial Suplemento 459, 26 de mayo de 2021).

Bird, S., Klein, E., & Loper, E. (2009). *Natural language processing with Python*. O'Reilly Media.

Brealey, R. A., Myers, S. C., & Allen, F. (2020). *Principles of corporate finance* (13.ª ed.). McGraw-Hill Education.

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and regression trees*. Wadsworth.

Chacon, S., & Straub, B. (2014). *Pro Git* (2.ª ed.). Apress. https://git-scm.com/book

Chen, H., Chiang, R. H. L., & Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. *MIS Quarterly*, *36*(4), 1165–1188. https://doi.org/10.2307/41703503

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2.ª ed.). Lawrence Erlbaum Associates.

Davenport, T. H., & Harris, J. G. (2007). *Competing on analytics: The new science of winning*. Harvard Business School Press.

Davenport, T. H., & Patil, D. J. (2012). Data scientist: The sexiest job of the 21st century. *Harvard Business Review*, *90*(10), 70–76.

Few, S. (2013). *Information dashboard design: Displaying data for at-a-glance monitoring* (2.ª ed.). Analytics Press.

Honnibal, M., Montani, I., Van Landeghem, S., & Boyd, A. (2020). *spaCy: Industrial-strength natural language processing in Python* [Software]. Zenodo. https://doi.org/10.5281/zenodo.1212303

Inmon, W. H. (2005). *Building the data warehouse* (4.ª ed.). Wiley.

Kaplan, R. S., & Norton, D. P. (1996). *The balanced scorecard: Translating strategy into action*. Harvard Business School Press.

Kimball, R., & Ross, M. (2013). *The data warehouse toolkit: The definitive guide to dimensional modeling* (3.ª ed.). Wiley.

Kleppmann, M. (2017). *Designing data-intensive applications*. O'Reilly Media.

Kohavi, R., Tang, D., & Xu, Y. (2020). *Trustworthy online controlled experiments: A practical guide to A/B testing*. Cambridge University Press.

Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to information retrieval*. Cambridge University Press.

McKinney, W. (2010). Data structures for statistical computing in Python. En S. van der Walt & J. Millman (Eds.), *Proceedings of the 9th Python in Science Conference* (pp. 56–61). https://doi.org/10.25080/Majora-92bf1922-00a

Mell, P., & Grance, T. (2011). *The NIST definition of cloud computing* (NIST Special Publication 800-145). National Institute of Standards and Technology. https://doi.org/10.6028/NIST.SP.800-145

Meta Platforms. (2026). *Pricing on the WhatsApp Business PlaVorm*. https://developers.facebook.com/docs/whatsapp/pricing

Meta Platforms. (2026). *WhatsApp Business messaging policy*. https://whatsappbusiness.com/policy/

Microsoft. (2026). *Azure AI Language*. https://azure.microsoft.com/products/ai-services/ai-language

Microsoft. (2026). *Azure Functions overview*. Microsoft Learn. https://learn.microsoft.com/azure/azure-functions/functions-overview

Microsoft. (2026). *Azure landing zones*. Microsoft Learn. https://learn.microsoft.com/azure/cloud-adoption-framework/ready/landing-zone/

Microsoft. (2026). *Power BI pricing*. https://www.microsoft.com/en/power-platform/products/power-bi/pricing

Microsoft. (2026). *What is Azure Data Factory?* Microsoft Learn. https://learn.microsoft.com/azure/data-factory/introduction

Microsoft. (2026). *What is Microsok Entra ID?* Microsoft Learn. https://learn.microsoft.com/entra/fundamentals/whatis

Oliver, R. L. (1999). Whence consumer loyalty? *Journal of Marketing, 63*(4_suppl1), 33–44. https://doi.org/10.1177/00222429990634s105

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research, 12*, 2825–2830.

Presidencia de la República del Ecuador. (2023). *Reglamento General a la Ley Orgánica de Protección de Datos Personales* (Decreto Ejecutivo 904; Registro Oficial 435, 13 de noviembre de 2023). https://www.gob.ec/regulaciones/reglamento-general-ley-organica-proteccion-datos-personales-0

Provost, F., & Fawcett, T. (2013). *Data science for business: What you need to know about data mining and data-analytic thinking*. O'Reilly Media.

Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning, 1*(1), 81–106. https://doi.org/10.1007/BF00116251

Reichheld, F. F., & Sasser, W. E. (1990). Zero defections: Quality comes to services. *Harvard Business Review, 68*(5), 105–111.

Reis, J., & Housley, M. (2022). *Fundamentals of data engineering*. O'Reilly Media.

White, T. (2015). *Hadoop: The definitive guide* (4.ª ed.). O'Reilly Media.

Zaharia, M., Xin, R. S., Wendell, P., Das, T., Armbrust, M., Dave, A., Meng, X., Rosen, J., Venkataraman, S., Franklin, M. J., Ghodsi, A., Gonzalez, J., Shenker, S., & Stoica, I. (2016). Apache Spark: A unified engine for big data processing. *Communications of the ACM, 59*(11), 56–65. https://doi.org/10.1145/2934664