

Arquitectura de analítica de datos y storytelling: automatización y visualización estratégica para la gestión comercial en Valveco S.A.S.

Nelson Andrés Jiménez Rojas

Camilo Andrés Salas Parodi

Martha Liliana Torres Londoño

Universidad Ean

Facultad de Ingeniería

Maestría en ciencia de datos

Bogotá, Colombia

08/junio/2026

Arquitectura de analítica de datos y storytelling: automatización y visualización estratégica para la gestión comercial en Valveco S.A.S.

Nelson Andrés Jiménez Rojas

Camilo Andrés Salas Parodi

Martha Liliana Torres Londoño

Trabajo de grado presentado como requisito para optar al título de:

Magister en Ciencia de Datos

Directora:

Carolina María Luque Zabala

Modalidad:

Trabajo Dirigido

Universidad Ean

Facultad de ingeniería

Maestría en ciencia de datos

Bogotá, Colombia

08/junio/2026

Nota de aceptación:

_____

_____

_____

_____

_____

_____
Firma del jurado

_____
Firma del jurado

_____
Firma del director del trabajo de grado

Ciudad, día/mes/año

Con gratitud dedico este trabajo a Dios, por brindarme fortaleza y sabiduría a lo largo de este proceso; a mi familia, por su amor, apoyo incondicional y por ser mi mayor motivación; y a mi pareja, por su compañía, paciencia y comprensión durante este camino académico. Asimismo, agradezco a todas las personas que, de una u otra manera, hicieron parte de este logro y contribuyeron a mi crecimiento personal y profesional.

Camilo

Por los sábados junto a profesores y compañeros que se volvieron amigos e hicieron parte de este camino de maestría, por las enseñanzas, el apoyo y los buenos momentos compartidos durante cada jornada de aprendizaje.

A mi familia, pareja y amigos, por ser esa energía extra a lo largo del camino. ¡Gracias!

Hoy confirmo que "paso a paso se llega lejos".

Andrés

A todos los maestros de vida que me han transformado de manera positiva, especialmente a los que me han acompañado durante este pedacito de camino, todo mi amor y gratitud.

Martha

Agradecimientos

Agradecemos a nuestras familias por el apoyo incondicional, la comprensión y la motivación brindada a lo largo de este proceso académico. Su confianza y acompañamiento fueron importantes para alcanzar esta meta.

A Héctor Galvis gerente general de Valveco S.A.S por su autorización, confianza y apoyo, ya que sin su disposición no habría sido posible desarrollar este proyecto. Asimismo, agradecemos a Stephanie Arroyo, gerente administrativa, con su ayuda se pudo construir y comprender la información necesaria para analizar la situación general de la empresa, facilitando nuestra inmersión en un sector inicialmente desconocido para nosotros.

De igual manera, extendemos nuestro agradecimiento a nuestra directora, Carolina Luque por sus enseñanzas y excelente organización durante el desarrollo del trabajo. Su conocimiento y acompañamiento fueron clave para dar estructura y dirección a este proyecto.

Resumen

La analítica de datos se ha consolidado como una herramienta clave para las organizaciones, en la medida en que posibilita la toma de decisiones informadas basadas en evidencia. En este contexto, la empresa Valveco S.A.S. emplea diversos sistemas para la recolección de información comercial, entre ellos el sistema de gestión empresarial (World Office) y el gestor de relaciones con clientes (DataCRM). No obstante, se evidencia una desconexión entre dichas plataformas, lo que conlleva a la elaboración manual de informes, generando retrasos, inconsistencias y limitada trazabilidad de la información.

El propósito de este trabajo consiste en diseñar una arquitectura de datos ajustada a las necesidades de la organización, que permita a los niveles gerenciales visualizar la información de manera estructurada y extraer mayor valor de la misma. Desde el enfoque metodológico, se realizó una revisión de literatura que aborda conceptos como ciencia de datos, arquitectura analítica, metodología DataOps, integración de datos, ética y gobernanza, así como visualización estratégica.

A partir de un diagnóstico organizacional interno y externo, se determinó que la empresa presenta un nivel de madurez analítica temprano, con oportunidades de mejora en calidad de datos, latencia de la información, esfuerzo operativo de consolidación y automatización de procesos. Como principales resultados, se estructuró una solución que integra la información, automatiza procesos operativos y permite anticipar quiebres de inventario, facilitando su visualización mediante un tablero estratégico orientado al storytelling de datos, con un plan de implementación que permita su adopción.

Palabras clave: analítica de datos, arquitectura de datos, visualización estratégica, gobernanza de datos, integración de datos.

Abstract

Data analytics has become a key tool for organizations, as it enables evidence-based decision-making. In this context, the company Valveco S.A.S. uses various systems to collect commercial information, including the enterprise resource planning system (World Office) and the customer relationship management platform (DataCRM). However, the lack of integration between these platforms has resulted in the manual preparation of reports, leading to delays, inconsistencies, and limited information traceability.

The purpose of this study is to design a data architecture tailored to the organization's needs, enabling management teams to visualize information in a structured manner and extract greater value from it. From a methodological perspective, a literature review was conducted covering concepts such as data science, analytical architecture, the DataOps methodology, data integration, ethics and governance, as well as strategic data visualization.

Based on an internal and external organizational diagnosis, it was determined that the company is at an early stage of analytical maturity, with opportunities for improvement in data quality, information latency, operational consolidation efforts, and process automation. As a result, a solution was designed to integrate information, automate operational processes, and anticipate inventory shortages, while facilitating data visualization through a strategic dashboard focused on data storytelling. Additionally, an implementation plan was proposed to support the adoption of the solution within the organization.

Keywords: data analytics, data architecture, strategic visualization, data governance, data integration.

Tabla de Contenido

Agradecimientos **5**

Resumen **6**

Abstract **7**

Tabla de Contenido **8**

Lista de figuras **12**

Lista de Tablas **14**

1. Introducción **16**

2. Objetivos **19**

2.1 Objetivo general *19*

2.2 Objetivos específicos *19*

3. Justificación **20**

4. Marco Institucional **22**

4.1 Presentación de la empresa *22*

4.2 Referentes estratégicos *22*

4.3 Estructura organizacional *23*

4.4 Productos y servicios ofertados *25*

4.5 Estructura del proceso comercial en Valveco S.A.S *26*

4.6 Análisis y tendencias del sector válvulas industriales *26*

5. Marco Teórico **29**

5.1 Ciencia de datos y gerencia sostenible *29*

5.1.1 Fundamentos de la ciencia de datos en las organizaciones 29

5.1.2 Gerencia sostenible basada en datos 30

5.2. Arquitectura analítica integral para la gestión de datos comerciales *31*

5.2.1. Metodología DataOps 33

5.2.2. Procesos de integración de datos: ERP y CRM 34

5.3 Ética, gobernanza y calidad de datos. *35*
5.3.1 Marco de gobernanza de datos 35
5.3.2 Dimensiones de calidad del dato 36
5.3.3 Ética en el manejo de información comercial 37
5.4 Visualización estratégica y storytelling de datos *38*
5.4.1 Psicología de la percepción visual 38
5.4.2 Narrativa de datos 39
5.5 Analítica en las ventas *39*
6. Diseño Metodológico 41
6.1 Tipo de investigación *41*
6.2 Fases del proyecto *41*
6.3 Análisis externo *44*
6.4 Análisis interno *45*
6.5 Población de la muestra y ficha técnica *46*
6.6 Identificación de variables *47*
6.7 Instrumento de medición *48*
6.8 Validación del instrumento de medición *49*
7. Diagnóstico Organizacional 56
7.1 Análisis externo: Matriz DOFA *56*
7.2 Análisis interno: Herramienta TDWI *58*
7.3 Análisis interno: Encuesta *62*
7.3.1 Calidad de los datos comerciales 62
7.3.2 Latencia de información para la toma de decisiones 66
7.3.3 Esfuerzo operativo de consolidación 69
7.3.4 Automatización y monitoreo de procesos (DataOps) 71
7.3.5 Confiabilidad y alineación de KPIs comerciales 73
7.4 Síntesis de diagnóstico organizacional *76*
8. Diseño de la arquitectura 78

8.1. Selección de herramientas para la arquitectura ... 78

8.2. Componentes de la solución analítica ... 80
8.2.1 Capa de ingesta y orquestación ... 80

8.2.2 Capa de procesamiento y almacenamiento ... 81

8.2.3 Capa de Inteligencia Predictiva ... 85

8.2.4 Capa de consumo y storytelling ... 89

8.3. Marco de gobernanza de datos ... 99
8.3.1 Diccionario de datos ... 99

8.3.2 Política de calidad de datos ... 101

8.3.3 Roles de acceso y permisos diferenciados (Row Level Security) ... 102

8.3.4 Control de gobernanza de datos para la arquitectura ... 103

8.4. Arquitectura analítica orientada a la resolución de interrogantes estratégicos 104

8.5 Plan de criterios de éxito y pruebas de aceptación ... 106

8.6 Plan de casos de prueba técnicos ... 107

8.7 Matriz de gestión de riesgos ... 110

8.8 Tiempo de mantenimiento por componente ... 114

8.9 Análisis financiero: retorno sobre la inversión y costo-beneficio ... 116
8.9.1 Cuantificación del problema actual ... 116

8.9.2 Beneficios cuantificables de la arquitectura propuesta ... 117

8.9.3 Estimación del periodo de retorno de la inversión ... 117

9. Plan de implementación del modelo de analítica de datos y storytelling ... 119

9.1 Plan de adopción ... 120
9.1.1 Nivel estratégico gerencial ... 120

9.1.2 Nivel táctico del área comercial ... 121

9.1.3 Nivel operativo ... 122

9.2 Recursos necesarios ... 122
9.2.1 Recursos Técnicos e Infraestructura ... 122

9.2.2 Roles y Responsabilidades ... 123

9.2.3 Presupuesto 124

9.3 Cronograma 125

10. Implementación en la organización 126

10.1 Retroalimentación de la organización 127

11. Conclusiones y Recomendaciones 130

11.1 Conclusiones 130

11.2 Recomendaciones 133

12. Referencias 136

13. Anexos 143

13.1 Carta de aceptación Valveco S.A.S. 143

13.2 Encuesta de diagnóstico Valveco S.A.S. 144

Lista de figuras

Figura 1 Organigrama Valveco S.A.S. 24

Figura 2 Tamaño del mercado global de válvulas 27

Figura 3 Arquitecturas de inteligencia de negocios para ciencia de datos 33

Figura 4 Resultado TDWI Valveco S.A.S. 59

Figura 5 Puntuación general prueba TDWI 62

Figura 6 Pregunta 1A encuesta 63

Figura 7 Pregunta 2A encuesta 63

Figura 8 Pregunta 3A encuesta 64

Figura 9 Pregunta 4A encuesta 65

Figura 10 Pregunta 1B encuesta 66

Figura 11 Pregunta 2B encuesta 67

Figura 12 Pregunta 3B encuesta 68

Figura 13 Pregunta 1C encuesta 69

Figura 14 Pregunta 2C encuesta 70

Figura 15 Pregunta 1D encuesta 71

Figura 16 Pregunta 2D encuesta 72

Figura 17 Pregunta 1E encuesta 73

Figura 18 Pregunta 2E encuesta 74

Figura 19 Pregunta 3E encuesta 75

Figura 20 Pregunta 4E encuesta 75

Figura 21 Estructura de las fuentes de datos 81

Figura 22 Diseño de arquitectura analítica detallado 85

Figura 23 Arquitectura del modelo de predicción sobre quiebre de inventario 89

Figura 24 Modelo de datos en esquema tipo estrella 90

Figura 25 Secuencia visual del tablero en Power BI 97

Figura 26 Reunión de presentación 127

Lista de tablas

Tabla 1 Revisión bibliográfica por sección 42

Tabla 2 Ficha técnica herramienta TDWI 46

Tabla 3 Ficha técnica encuesta 47

Tabla 4 Variables para diagnóstico interno 48

Tabla 5 Expertos evaluadores de la herramienta encuesta 50

Tabla 6 Resultados validación V de Aiken de la encuesta 51

Tabla 7 Reformulación de preguntas del instrumento 54

Tabla 8 Matriz DOFA 56

Tabla 9 Estrategias FO-FA-DO-DA 58

Tabla 10 Selección de herramientas para la implementación 78

Tabla 11 Procesos de limpieza según la fuente de datos 82

Tabla 12 Métricas de evaluación del modelo sobre el conjunto de prueba 87

Tabla 13 Matriz de confusión del modelo sobre el conjunto de prueba 88

Tabla 14 Análisis de audiencia y decisiones estratégicas 95

Tabla 15 Diccionario de datos 99

Tabla 16 Política de calidad de datos 101

Tabla 17 Roles de accesos y permisos 102

Tabla 18 Control de gobernanza de datos de la arquitectura 103

Tabla 19 . Resolución de interrogantes estratégicos basado en la arquitectura 105

Tabla 20 Criterios de éxito y pruebas de aceptación 106

Tabla 21 Casos de prueba técnicos 108

Tabla 22 Matriz de gestión de riesgos 110

Tabla 23 Tiempos de mantenimiento por componente 114

Tabla 24 Cuantificación del problema 116

Tabla 25 Análisis de sensibilidad del periodo de retorno 119

Tabla 26 Presupuesto plan de implementación 124

Tabla 27 Cronograma de implementación 125

1. Introducción

En el entorno actual de la industria de válvulas y soluciones de conducción, la capacidad de transformar datos en conocimiento estratégico se ha convertido en un diferenciador competitivo clave para empresas con trayectoria como Valveco S.A.S. Al operar en un mercado que proyecta un crecimiento sostenido, la organización se apoya en su misión de satisfacción al cliente y asesoría técnica especializada para mantener su posicionamiento estratégico (Valveco S.A.S., s.f.; Market Data Forecast, 2025). No obstante, para capitalizar plenamente estas oportunidades, resulta necesario evolucionar hacia una cultura organizacional basada en decisiones fundamentadas en datos, que permita un seguimiento riguroso y oportuno de la gestión comercial (Jaggia et al., 2023; Sharda et al., 2018).

Actualmente, la organización presenta una desconexión crítica entre sus principales fuentes de información: el sistema de gestión empresarial (World Office) y el gestor de relaciones con los clientes (DataCRM). Esta fragmentación limita la capacidad analítica del área comercial, ya que, como señalan Bernardo et al. (2024), la analítica de datos pierde efectividad cuando no existen prácticas sólidas de calidad y gobernanza de datos. En este sentido, diversos estudios evidencian que las mejoras en el desempeño financiero y en la competitividad de mercado dependen de que los datos estén integrados, estandarizados y gobernados de manera consistente (Atlas et al., 2025; Solano y Cruz, 2024; Wang et al., 2024).

En la situación actual, la consolidación de los indicadores clave de desempeño (KPI) se realiza de forma manual, lo que demanda entre dos y cuatro horas diarias de trabajo operativo para validar y conciliar información correspondiente a distintos periodos. Esta práctica no solo incrementa el riesgo de errores, sino que también dificulta la obtención de una visión unificada del negocio, afectando la transparencia y la trazabilidad de los procesos comerciales. De acuerdo con Caballero Muñoz-Reja et al. (2019), la ausencia de

mecanismos formales de calidad de datos compromete la confiabilidad de los indicadores utilizados para la toma de decisiones.

El impacto de esta problemática se refleja en latencias significativas en la entrega de reportes y en inconsistencias métricas que no representan con fidelidad el desempeño real del área comercial. Al carecer de una herramienta que consolide y visualice la información de manera integrada, la empresa corre el riesgo de adoptar estrategias reactivas frente a resultados inesperados, en lugar de diseñar acciones preventivas basadas en el análisis sistemático de las ventas.

La literatura especializada evidencia que la automatización de procesos de integración de datos mediante enfoques ETL contribuye a la reducción de errores, la optimización de los ciclos de reporte y la mejora de la trazabilidad de la información, siempre que se acompañe de prácticas de gobernanza de datos y monitoreo continuo (Ain et al., 2019; El-Gharib y Amyot, 2023; Farinha et al., 2024; Ribeiro et al., 2021;).

Para abordar esta problemática, el presente trabajo propone el diseño de una arquitectura de analítica de datos basada en visualización y storytelling (narrativa a partir del análisis de datos) para la toma de decisiones estratégicas en el área comercial de Valveco S.A.S. El proyecto se orienta a la construcción de un flujo automatizado que integre las fuentes de datos existentes y estructure una herramienta de control capaz de narrar el desempeño comercial de Valveco S.A.S. de manera clara, coherente y persuasiva. A través de la aplicación de principios de comunicación visual y análisis de datos, se busca facilitar la identificación de patrones comerciales y operativos, transformando los indicadores en instrumentos basados en evidencia que fortalezcan la competitividad institucional, en línea con los planteamientos de Biemans (2023) y Li et al. (2022). De forma complementaria, se incorpora un modelo de aprendizaje automático, orientado a la predicción de quiebres de stock, el cual permite anticipar la demanda mejorando la respuesta frente a las necesidades de los clientes. Este componente no solo

fortalece la toma de decisiones operativas en el proceso de ventas, sino que se articula con la analítica propuesta, integrando sus resultados en la narrativa de datos para una gestión comercial anticipada y proactiva.

El documento se estructura de la siguiente manera: en la segunda sección, se presentan el objetivo general y los objetivos específicos que orientan la propuesta. Posteriormente, se expone la justificación de la intervención y el marco institucional, con el propósito de contextualizar la organización. La fundamentación teórica se desarrolla en el marco teórico, seguida del diseño metodológico en la sexta sección, donde se detalla la hoja de ruta del trabajo. A partir de un diagnóstico organizacional que evalúa el estado actual de la entidad, se formula el plan de intervención. Finalmente, el documento cierra con las conclusiones, recomendaciones, referencias bibliográficas y anexos.

2. Objetivos

2.1 Objetivo general

Diseñar una arquitectura de analítica de datos basada en visualización y storytelling para la toma de decisiones estratégicas en el área comercial de Valveco S.A.S.

2.2 Objetivos específicos

- Fundamentar teórica y metodológicamente el proyecto mediante la revisión de literatura sobre arquitectura de datos, storytelling estratégico y marcos de visualización aplicados a la gestión comercial.

- Diagnosticar el estado actual de la gestión de información comercial en Valveco S.A.S., mediante un análisis interno y externo que permita identificar problemas técnicos de latencia y trazabilidad, y la madurez de la cultura organizacional para la adopción de decisiones guiadas por datos.

- Proponer una arquitectura de analítica de datos que integre fuentes de información mediante flujos automatizados y un modelo de visualización estratégica basado en storytelling, con el fin de consolidar un ecosistema analítico que transforme los datos comerciales en narrativas accionables para el cumplimiento de las metas estratégicas de Valveco S.A.S.

- Formular una hoja de ruta para la adopción institucional del diseño propuesto, detallando fases de implementación, recursos necesarios, estructura de gobernanza de datos y estrategias de gestión del cambio para favorecer la apropiación de la propuesta dentro de la organización.

3. Justificación

En la actualidad, cada vez más las organizaciones encuentran valor en la calidad, oportunidad y consistencia de sus datos debido a que le permiten la toma de decisiones estratégicas y la generación de ventajas competitivas sostenibles. Diversos estudios evidencian que la analítica de datos impacta la calidad de las decisiones, desempeño financiero y la competitividad (Atlas et al., 2025; Li et al., 2022; Mikalef et al., 2019).

En el caso de Valveco S.A.S., la información comercial relevante para medir el cumplimiento de ventas se encuentra distribuida en sistemas como World Office (gestor empresarial), DataCRM (gestión comercial) y reportes en Excel. En ausencia de una arquitectura analítica de datos, los informes se deben elaborar manualmente, un proceso que puede consumir entre dos y cuatro horas diarias. Este esfuerzo operativo no solo genera retrasos en la disponibilidad de la información, sino también inconsistencias y falta de trazabilidad entre reportes de distintas áreas, lo que afecta directamente la confiabilidad de los indicadores de desempeño (Fahmideh y Beydoun, 2019; Sharda et al., 2018).

Las implicaciones de esta problemática trascienden lo operativo. En contextos comerciales, indicadores clave como el cumplimiento de cuota, la cobertura del pipeline y la optimización del embudo de ventas dependen de datos consistentes y oportunos para habilitar decisiones proactivas y basadas en evidencia (Biemans, 2023; Conde, 2025). Cuando la información es tardía o inconsistente, las organizaciones tienden a adoptar decisiones reactivas, asignar recursos de manera ineficiente y perder oportunidades de mercado, debilitando su capacidad de respuesta frente a la competencia (Solano y Cruz, 2024).

Bajo este escenario, el diseño de una arquitectura de analítica de datos, sustentada en flujos automatizados de integración y transformación de datos que actúen como la base técnica para un modelo de visualización se justifica como una solución estratégica. El valor agregado de esta propuesta radica en la aplicación de un enfoque de gerencia sostenible,

donde la incorporación de procesos analíticos no sólo busca la eficiencia económica, sino la sostenibilidad operativa al liberar al capital humano de tareas repetitivas y propensas al error, permitiendo un enfoque en el análisis de alto valor.

La propuesta se inspira en la unidad de storytelling y visualización de datos e incorpora elementos de la unidad de ética, gobernanza y calidad de los datos, para favorecer su futura adopción dentro de la organización. Se enmarca en el campo institucional de investigación de ciencia, tecnología e innovación, línea de investigación AOPP automatización y optimización de procesos productivos.

En consecuencia, este proyecto es pertinente porque responde a una necesidad real de Valveco S.A.S., aporta una solución con impacto estratégico en la gestión comercial y contribuye al desarrollo de conocimiento práctico y académico en el campo de la ciencia de datos a través de los conocimientos adquiridos en la Maestría en Ciencia de Datos.

4. Marco Institucional

4.1 Presentación de la empresa

Valveco S.A.S fue constituida formalmente el 7 de junio de 1989, en Bogotá D.C., es una compañía colombiana con más de treinta años de trayectoria dedicada a la comercialización y distribución de válvulas industriales y soluciones integrales para sectores estratégicos como hidrocarburos, minería, alimentos, energía y construcción. Se enfoca en brindar soluciones de conducción, con productos de marcas reconocidas a nivel mundial hacía la satisfacción del cliente.

A lo largo de su trayectoria, la empresa ha creado relaciones de confianza con clientes y proveedores, posicionándose como aliado estratégico en el suministro de productos y servicios de alta calidad. Entre sus clientes se destacan empresas de gran importancia en la industria colombiana como Peldar, Ecopetrol, Argos, Corona y Coca-Cola.

4.2 Referentes estratégicos

Valveco S.A.S define su identidad y propósito institucional mediante una misión clara, una visión orientada a la expansión y unos principios estratégicos que guían su operación diaria. Estos referentes permiten las decisiones tácticas y los proyectos estratégicos.

Misión: “Garantizar la satisfacción de las necesidades de nuestros clientes mediante un correcto asesoramiento y suministro de productos de calidad con el apoyo de la estructura organizacional y profesional de la empresa, cuyo compromiso en última instancia será siempre la búsqueda efectiva del desarrollo de valores éticos y económicos de todos los involucrados en la relación comercial.” (Valveco S.A.S. ,s.f)

Visión: “Ser una organización comercial con cubrimiento a nivel nacional e internacional, la cual cuente con una amplia gama de productos que permita atender los sectores industriales y comerciales. Logrando, al mismo tiempo, que la imagen de la empresa sea

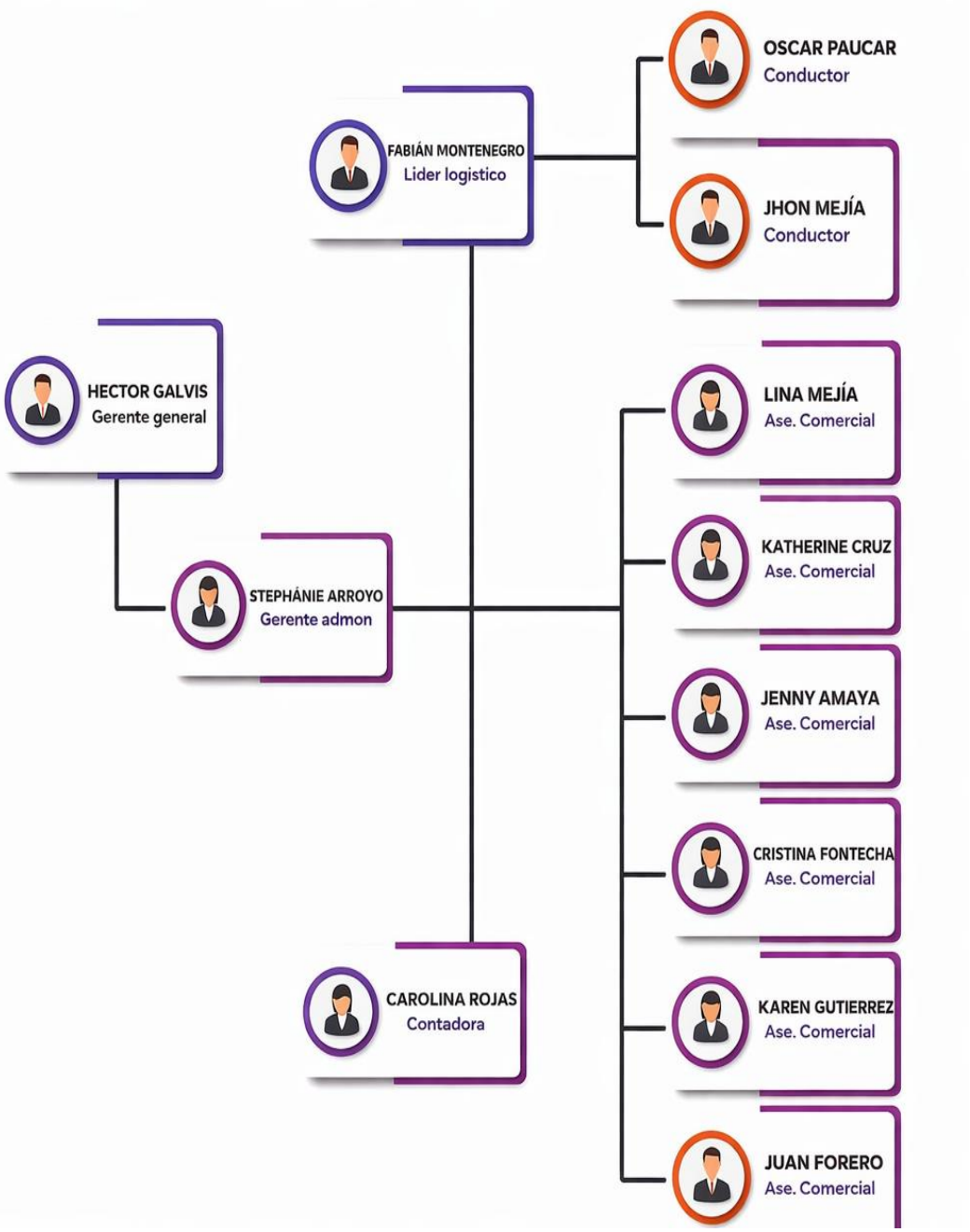
de amplio reconocimiento al estar siempre comprometida con el crecimiento y desarrollo de nuestro país.” (Valveco S.A.S. ,s.f)

Los principios estratégicos están enfocados en la orientación al cliente, responsabilidad social, innovación tecnológica, y la mejora continua. (Valveco S.A.S. ,s.f).

4.3 Estructura organizacional

La estructura organizacional de Valveco S.A.S. se caracteriza por ser de tipo jerárquico en cascada, lo que permite una delimitación clara de funciones y niveles de autoridad. Actualmente está constituida por 10 empleados directos distribuidos en las áreas administrativa, contabilidad, talento humano y comercial y 4 colaboradores externos. Esta organización favorece la comunicación vertical y el control de procesos por tener un tamaño pequeño.

Figura 1 *Organigrama Valveco S.A.S.*



Nota. Elaboración propia

4.4 Productos y servicios ofertados

Valveco S.A.S se especializa en la comercialización y distribución de válvulas industriales y equipos complementarios para la conducción, control y medición de fluidos. Su portafolio está conformado por productos de fabricantes internacionales de alta trayectoria, lo que le permite atender los requerimientos de sectores estratégicos como hidrocarburos, minería, alimentos, energía, construcción y manufactura (Valveco S.A.S., s.f.)

Dentro de los principales productos que ofrece la compañía se destacan:

- Válvulas industriales: Válvulas de mariposa, bola inoxidable, compuerta, globo, retención, válvulas Watson, utilizadas para procesos de control y de flujo y presión.
- Válvulas solenoides y actuadores: Elementos que permiten la automatización en la operación de válvulas industriales.
- Instrumentación y medición: Medidores de flujo, presión, temperatura y nivel, así como soluciones en tubing, racores y conectores para asegurar la confiabilidad de las operaciones.
- Componentes de control avanzado: Equipos y accesorios de más de 20 proveedores internacionales, entre los que destacan marcas líderes como Danfoss, Amstrong, Flir, Golden Partner, Rockwell y Emerson.

En cuanto a servicios, Valveco S.A.S ofrece un acompañamiento de inicio a fin, no se limitan solo a la venta de productos, también incluyen:

- Asesoría técnica especializada para la selección de válvulas y equipos según la necesidad que tenga el cliente.
- Soporte en montaje, mantenimiento y reemplazo de válvulas e instrumentación.
- Servicios de postventa y respaldo técnico, si surge algún inconveniente, bajo parámetros establecidos se generan respaldos con la marca oficial para tener un servicio de calidad.
- Gestión de proyectos industriales, con experiencia en los sectores previamente mencionados.

4.5 Estructura del proceso comercial en Valveco S.A.S

En Valveco S.A.S la asignación de metas comerciales se realiza con base en el comportamiento de ventas del mismo período del año anterior. Las cuentas más grandes son asignadas a la asesora comercial con mayor experiencia, ella asume un mayor porcentaje de la meta comercial, el resto de los asesores se distribuyen según el tamaño y nivel de sus clientes.

Las metas y resultados se registran mensualmente en el CRM y semanalmente se realizan seguimientos del cumplimiento, se analiza el avance frente a la meta, se revisan las desviaciones, se verifican las cotizaciones realizadas, los negocios perdidos y las estrategias comerciales como ofertas y contraofertas.

Adicionalmente, se realiza la verificación de la información entre el CRM y el World Office para validar que las ventas reportadas como ganadas correspondan a las facturas electrónicas, esto se debe a que Valveco S.A.S considera una venta ganada cuando está factura en el World Office. En los casos donde hay devoluciones, se generan notas de crédito que se registran en el sistema y se ven en los informes descargados como valores negativos.

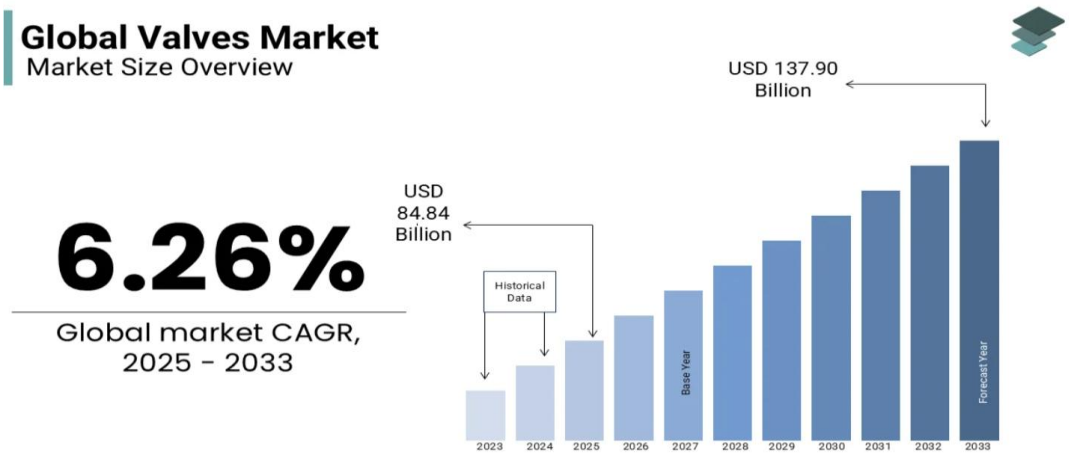
Cada bimestre la empresa realiza un análisis de ventas, utilidad, inventario y desviación frente a las metas con el objetivo de evaluar el desempeño comercial e intervenir con acciones necesarias.

4.6 Análisis y tendencias del sector válvulas industriales

El mercado global de válvulas industriales presenta un crecimiento que ha sido impulsado por la industrialización de los últimos 5 años, el desarrollo de nuevas aplicaciones tecnológicas y la transición hacía procesos más sostenibles. Se estima que para el 2030-2033 el mercado alcance valores entre 100 y 146 mil millones de dólares, con una tasa compuesta anual entre el 4% y el 6% (Market Research Report, 2025). En

América Latina, aunque el crecimiento es moderado respecto al promedio global, la demanda continúa en crecimiento, gracias a la expansión de los sectores como energía, alimentos, minería y construcción.

Figura 2 *Tamaño del mercado global de válvulas*



Nota: Adaptado de The Observatory of Economic Complexity. (2023).

La implementación de tecnologías digitales se identifica como un factor diferenciador. La inteligencia artificial aplicada al mantenimiento predictivo, la digitalización de procesos de control de fluidos y la incorporación de sistemas de monitoreo inteligente están transformando la industria. Estas tendencias permiten a las compañías mejorar su productividad, reduciendo fallas, realizando controles de riesgo y calidad y tomar decisiones basadas en datos confiables.

A nivel país, el sector muestra un aumento en la importación de válvulas y equipos relacionados, con un crecimiento de 16.9 millones de dólares en 2023 en la categoría “válvulas de seguridad o alivio” reportado por the observatory of economic complexity, (2023). Valveco S.A.S. se consolida como un ente importante, ya que cuenta con representación de marcas internacionales y un portafolio sólido y especializado. En el

último año, la empresa registró un crecimiento del 16% en sus ventas y alcanzó ingresos de 4.800 millones de pesos, esto demuestra su capacidad para aprovechar la dinámica del mercado y posicionarse entre sus competidores.

La integración de una arquitectura de inteligencia de negocios se hace relevante al observar este panorama. La implementación de procesos ETL y tableros de control permitirá a Valveco S.A.S. mejorar la trazabilidad de sus indicadores comerciales, optimizar la proyección de ventas y aprovechar con mayor precisión las oportunidades del mercado. Es necesario estructurar unas metas claras y confiables que mejoren la gestión comercial, con acceso inmediato y fortaleciendo con herramientas de inteligencia de negocios la competitividad de la empresa en el mercado.

5. Marco Teórico

5.1 Ciencia de datos y gerencia sostenible

La intersección entre la ciencia de datos y la gestión organizacional ha dejado de ser una periferia técnica para convertirse en el núcleo de la estrategia competitiva moderna (Mikalef et al., 2019; Li et al., 2022). Las organizaciones han enfrentado una aceleración sin precedentes en la adopción de tecnologías inteligentes, impulsada por la necesidad de una búsqueda de eficiencia operativa (Jaggia et al., 2023; Sharda et al., 2018). En este apartado se consideran los fundamentos de la ciencia de datos en el ámbito corporativo y la transición hacia una gerencia sostenible que utiliza la automatización y la analítica avanzada como pilares para la optimización del capital humano y la escalabilidad del conocimiento.

5.1.1 Fundamentos de la ciencia de datos en las organizaciones

En el contexto de la transformación digital contemporánea, la ciencia de datos se constituye como un paradigma interdisciplinario que mezcla métodos científicos, procesos de gobernanza de la información y algoritmos de aprendizaje automático para la extracción de conocimiento estratégico. Teóricamente, este campo se sustenta en la Visión Basada en Recursos (RBV), la cual postula que las capacidades analíticas actúan como recursos estratégicos valiosos, raros, inimitables y no sustituibles (VRIN), catalizando una ventaja competitiva sostenible a través de la reconfiguración proactiva de activos (Sultana et al., 2024).

Para que estas capacidades analíticas se traduzcan efectivamente en una ventaja competitiva, la organización debe evolucionar en su madurez de procesamiento, pues la calidad de la toma de decisiones está intrínsecamente ligada al nivel de uso de la analítica de datos (Li et al., 2022). Bajo esta premisa, la arquitectura de analítica de datos no solo debe almacenar información, sino facilitar una transición lógica entre niveles de

conocimiento. En este sentido, la literatura científica establece una distinción crítica entre el análisis descriptivo tradicional y los enfoques prospectivos modernos. El análisis descriptivo se caracteriza por su naturaleza retrospectiva, orientada al diagnóstico de eventos históricos mediante técnicas de agregación y visualización de datos (Anifowose, 2023). Si bien este enfoque es indispensable para establecer líneas base operativas y comprender el desempeño pasado, su utilidad es limitada en entornos con cambios estructurales acelerados, donde se requiere de inteligencia de decisión aumentada para mitigar la incertidumbre estratégica (Anifowose, 2023; Mandaliya y Kulkarni, 2025).

Por otro lado, el enfoque predictivo se fundamenta en la previsión estratégica. Este utiliza modelos de machine learning (ML) y algoritmos estadísticos para proyectar escenarios futuros y tendencias probables, permitiendo a las firmas reducir la incertidumbre. Investigaciones empíricas sugieren que las firmas que integran inteligencia de decisión aumentada logran mejoras de entre el 20% y el 30% en la precisión de sus pronósticos y una reducción en la latencia de respuesta (Anifowose, 2023).

5.1.2 Gerencia sostenible basada en datos

La gerencia sostenible en el ámbito comercial se define como la capacidad de articular la eficiencia operativa de corto plazo con la resiliencia estratégica de mercado a largo plazo. Bajo esta visión, la automatización analítica no es solo una herramienta técnica, sino el eje central para la creación de valor responsable y la obtención de una ventaja competitiva sostenible en entornos industriales (Gupta et al., 2022). Para Valveco S.A.S., esto implica que la arquitectura de datos debe garantizar que la toma de decisiones comerciales no dependa de esfuerzos aislados, sino de un ecosistema analítico escalable que reduzca el error humano y la latencia en la trazabilidad (Akpe et al., 2021).

En este contexto, la sostenibilidad estratégica se apoya en la "capacidad analítica del talento" (Big Data Talent Analytics Capability). A diferencia del enfoque administrativo

tradicional, esta perspectiva postula que el impacto financiero y de mercado de una firma depende de cómo el talento comercial utiliza la analítica para transformar el conocimiento en impacto de mercado (Atlas et al., 2025). Por tanto, más allá de la asignación de cargos, el enfoque se centra en optimizar las interacciones entre las herramientas digitales y los equipos de ventas para mejorar la percepción de valor y la efectividad de la gestión (Biemans, 2023).

Finalmente, la consolidación de este modelo se logra al transformar la arquitectura técnica en una capacidad de innovación guiada por datos. En lugar de sistemas de información pasivos, se requiere una infraestructura que potencie directamente la calidad de la toma de decisiones, reduciendo la ambigüedad en la interpretación de los indicadores comerciales (Li et al., 2022). Esta sinergia asegura que los datos automatizados se conviertan en un activo estratégico que permita a Valveco S.A.S. adaptarse proactivamente a las dinámicas del mercado, garantizando una ventaja competitiva persistente y alineada con sus objetivos de crecimiento (Market Data Forecast, 2025; Sultana et al., 2024).

5.2. Arquitectura analítica integral para la gestión de datos comerciales

Para el fortalecimiento de la inteligencia comercial en Valveco S.A.S., la literatura científica contemporánea valida la adopción de una arquitectura centrada en bodega de datos (Data Warehouse o DW) que, en su diseño moderno, integra la flexibilidad de los lagos de datos (Nambiar y Mundra, 2022). A diferencia de estructuras puramente descentralizadas, la arquitectura presentada en la Figura 3 constituye el estándar para garantizar la integridad, consistencia y gobernanza requeridas en el análisis de datos estructurados provenientes de sistemas ERP y CRM (Nambiar y Mundra, 2022). Esta arquitectura se descompone en cinco fases críticas que transforman los datos operacionales en inteligencia estratégica:

Fuentes de datos: La capa inicial identifica los sistemas de registro de la organización, que para el caso de estudio incluyen el ERP (World Office) para la gestión de recursos y el CRM (DataCRM) para la gestión de relaciones comerciales (Valveco S.A.S., s.f.). La heterogeneidad de estas fuentes exige procesos de extracción robustos que no afecten el rendimiento transaccional ni la disponibilidad de la información operativa (Saddad et al., 2020).

Proceso de extracción, transformación y carga (ETL): Según Bollineni (2023), este componente actúa como el motor de integración ya que mientras que los procesos ETL tradicionales suelen ser manuales y propensos a generar silos de datos, el ETL moderno potenciado por DataOps automatiza la validación para asegurar que solo la información con el formato correcto ingrese al sistema. En esta fase se limpian, estandarizan y unifican las definiciones de "Cliente" y "Venta", garantizando que la conversión de datos ocurra de manera idéntica en todos los entornos.

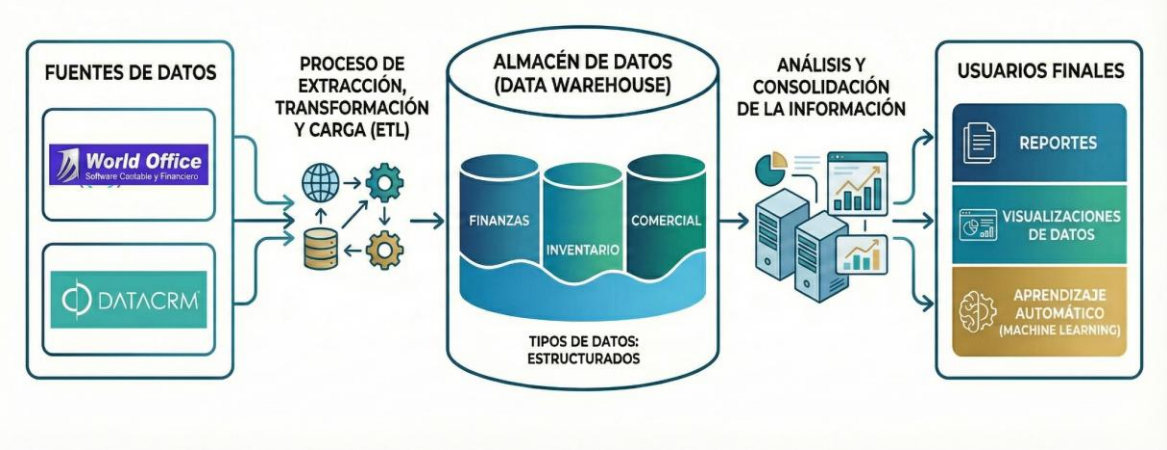
Almacén de datos (Enfoque Lake-DW): El núcleo de la arquitectura es un repositorio centralizado organizado en áreas funcionales como Finanzas, Inventario y Comercial (Nambiar y Mundra, 2022). Como se ilustra en la base de los cilindros de la Figura 3, el sistema integra niveles lógicos que permiten la persistencia de datos crudos, lo que otorga a cada silo la flexibilidad de un lago de datos para fines de auditoría, junto con tablas pre-agregadas que optimizan las consultas frecuentes (Saddad et al., 2020; Nambiar y Mundra, 2022). Esta convergencia, conocida como Lakehouse, unifica el almacenamiento masivo con el análisis estratégico en una plataforma abierta (Armbrust et al., 2021). Todo el sistema está regido por la Metadata, es decir, el diccionario que garantiza la gobernanza al definir la procedencia y el significado de los datos (Nambiar y Mundra, 2022).

Análisis y consolidación de la información: Esta etapa prepara los activos de datos para su consumo mediante marcos de procesamiento (Cuzzocrea, 2024). En este punto, la eficiencia se gestiona mediante la selección automática de índices que sustituyen

estructuras rígidas, permitiendo un procesamiento ágil de grandes volúmenes de datos comerciales (Yahyaoui et al., 2025).

Usuarios finales: Finalmente, la última etapa facilita el consumo estratégico de la información procesada a través de tres ejes principales: reportes estáticos, visualizaciones de datos para exploración dinámica y aprendizaje automático para la detección de patrones. Esta estructura asegura que la toma de decisiones comerciales esté fundamentada estrictamente en evidencia analítica (Cuzzocrea, 2024).

Figura 3 *Arquitecturas de inteligencia de negocios para ciencia de datos*



Nota: Adaptado de Nambiar y Mundra (2022)

5.2.1. Metodología DataOps

La implementación operativa de la arquitectura propuesta se fundamenta en DataOps, una práctica de gestión de datos colaborativa que se centra en mejorar la comunicación, integración y automatización de los flujos de información (Bollineni, 2023). Esta metodología adapta los principios de DevOps, la cual se define como una cultura de ingeniería que combina el desarrollo de software y las operaciones de TI para proporcionar entregas continuas de alta calidad, aplicándolos específicamente a la ingeniería de datos

para reducir la latencia y minimizar el error humano en los procesos analíticos (Bollineni, 2023).

Para Valveco S.A.S., la implementación de DataOps se estructurará en tres etapas clave que garantizan la agilidad y fiabilidad del ecosistema:

- Automatización de flujos (Pipelines): Consiste en la reducción o eliminación de intervenciones manuales en la extracción y carga de datos para reducir los silos de información.
- Integración y despliegue continuo (CI/CD): Permite realizar mejoras constantes en los modelos de datos y visualizaciones sin interrumpir la disponibilidad del sistema para el área comercial.
- Validación y Pruebas Automáticas: Se introducen controles de calidad en intervalos de tiempo definidos para detectar e identificar inconsistencias en los datos antes de que estos afecten los reportes estratégicos (Bollineni, 2023; Fannouch et al., 2025).

La adopción de estos principios permite mitigar riesgos operativos, optimizando la calidad de los activos de información y acelerando la velocidad de entrega de nuevas soluciones analíticas (Bollineni, 2023).

5.2.2. Procesos de integración de datos: ERP y CRM

La integración efectiva de World Office y DataCRM es el desafío central para obtener una visión 360° del cliente. Rahman et al. (2025) destacan que la integración de estas plataformas no es meramente técnica, sino estratégica para la optimización de procesos de negocio.

El proceso de integración transforma datos dispersos en "activos de información". Esto implica la resolución de entidades (identificar que el cliente "Valveco" en el CRM es el mismo "Valveco S.A.S." en el ERP) y la normalización de modelos de datos dispares en un esquema dimensional unificado (Estrella o copo de nieve) dentro del DW.

Para asegurar una toma de decisiones basada en evidencia y manejar grandes volúmenes de datos históricos, la arquitectura se orienta hacia el alto rendimiento y la eficiencia en las consultas. Se emplean marcos de trabajo de alto rendimiento que permiten realizar búsquedas analíticas complejas de manera eficiente mediante estructuras de datos optimizadas (Cuzzocrea, 2024). Además, la integración de un modelo Lakehouse facilita que el procesamiento de datos sea ágil, permitiendo consultas rápidas sobre la "única fuente de verdad" de la organización (Armbrust et al., 2021; Cuzzocrea, 2024).

Esta estructura permite una trazabilidad comercial de Valveco S.A.S. que sea consistente, admitiendo que la interacción con el cliente, registrada en el CRM se traduzca de forma precisa en indicadores de gestión financiera dentro del ERP (Rahman et al., 2025).

5.3 Ética, gobernanza y calidad de datos.

5.3.1 Marco de gobernanza de datos

La gobernanza de datos define los roles, políticas, estándares y mecanismos de control que garantizan que los datos se gestionen a favor de la organización y que los productos analíticos como lo son los indicadores, los tableros y los modelos sean confiables, auditables y sostenibles en el tiempo. En este contexto, la gobernanza de datos define responsabilidades sobre la creación, registro, actualización y uso de la información, así como normas para la calidad de los datos, la estandarización de definiciones y control de acceso de la información. Esto permite que los reportes comerciales y financieros de la empresa sean consistentes entre los diferentes sistemas de información y que se pueda realizar seguimiento de metas, análisis de ventas y toma de decisiones estratégicas.

En una organización como Valveco S.A.S., donde las métricas comerciales se obtienen y consolidan desde diferentes fuentes (World Office, DataCRM y Excel), la ausencia de

gobernanza se evidencia en métricas inconsistentes entre áreas y procesos manuales difíciles de auditar. Por esta razón la gobernanza que se propone busca sustentarse desde tres pilares:

1. **Estandarización semántica:**

Se incluyen las definiciones y el diccionario de datos, se formaliza qué significa cada KPI y sus componentes. Incluyendo la periodicidad, la fórmula y las fuentes de registro.

2. **Gestión de responsabilidades**

Se propone otorgar unas responsabilidades mínimas de función; donde se encuentran tres principalmente, definición y validación semántica de los indicadores, control de calidad y conciliación de la información, operación y monitoreo del flujo automatizando datos.

Esto permite pasar del proceso manual a un proceso controlado por reglas y calidad, mitigando reprocesos y asegurando continuidad.

3. **Control y trazabilidad**

Son un principio básico de la analítica de datos, orientado a asegurar que los indicadores comerciales sean interpretables y comparables. Para cada métrica se explica su fuente de origen, la fórmula del cálculo y el período de análisis.

5.3.2 Dimensiones de calidad del dato

La calidad de las decisiones y los productos analíticos está condicionada por la calidad de los datos (Li et al.,2022; Wang et al., 2024) En el caso que se está realizando se reportan latencias e inconsistencias, es necesario evaluar la calidad por medio de dimensiones operativas y medibles.

Exactitud: El dato representa correctamente el evento real. En la empresa se presentan problemas cuando la información registrada en el CRM no coincide con la plataforma World

Office, por ejemplo, cuando una venta se marca como ganada pero no tiene factura, o cuando existen diferencias en los valores o devoluciones no registradas en ambos sistemas.

Completitud: La completitud se refiere a que los registros contengan toda la información necesaria. En la empresa se presentan casos con campos obligatorios incompletos, como tipo de contacto, nit del cliente, celular del cliente, motivo de no venta.

Consistencia: Se refiere a que no existan diferencias entre la información registrada en los sistemas. En la empresa se identifica la mayor inconsistencia en este ítem, hay datos duplicados, reprocesos y diferencias entre las ventas registradas en CRM y las facturadas en World Office.

Oportunidad / Latencia: La demora o el desfase entre el momento en que ocurre la venta o el registro a cuando se tiene la disponibilidad de información para analizar. Actualmente los reportes realizados se cierran con retrasos e inconsistencias entre los sistemas, afectando el tiempo para reaccionar proactivamente. Estas dimensiones se deben validar con reglas de calidad de datos *(Data Quality Rules)*, las cuales están integradas por: validaciones de unicidad; que no se identifiquen duplicados, Integridad referencial; los datos de vendedor y cliente existen ,rangos aceptables; que estén los datos requeridos como fechas y que los valores no sean negativos, consistencia temporal; que las ventas estén en el período adecuado y las métricas de latencia; el tiempo que se tiene entre el momento de registrar y su disponibilidad para el análisis.

5.3.3 Ética en el manejo de información comercial

La ética en analítica de datos se refiere al uso responsable de los datos y de los indicadores para la toma de decisiones, evitando interpretaciones erróneas o decisiones sin justificación dentro de la organización (Provost y Fawcett, 2013). En el área comercial los principales riesgos están relacionados con los datos privados de los clientes, la

exposición de información sensible y la falta de claridad en los indicadores que pueden afectar la evaluación del desempeño.

La forma en la que se registran los datos en las plataformas puede generar sesgos en los indicadores. Por ejemplo, si un asesor registra todas las gestiones que realiza y otro no lo hace diligenciando todos los campos solicitados, esto puede afectar el desempeño en la calidad del registro y no se debe a la gestión comercial. Es importante que las métricas no se construyan a partir de datos incompletos o sin controles de calidad.

Por eso la ética se integra con la gobernanza de datos como un principio que busca garantizar el uso responsable de la información, la confidencialidad de los datos comerciales y lograr interpretar de forma adecuada los indicadores para tomar decisiones.

5.4 Visualización estratégica y storytelling de datos

5.4.1 Psicología de la percepción visual

La visualización con estrategia busca reducir la carga cognitiva y acelerar la interpretación en lo posible en tiempo real de la información clave con la que se toman decisiones. En los tableros comerciales el problema no está en mostrar más sino en permitir que el usuario detecte señales o patrones donde se pueda identificar caídas por estacionalidad, comportamiento variado o anormal, desviaciones de cuota o incumplimiento, con un mínimo de esfuerzo. Desde la teoría de percepción visual, el diseño debe contener jerarquía preatencional, en este concepto se realiza correctamente el uso de posición, tamaño, alineación, contraste y colores, esto permite identificar patrones antes de realizar un análisis detallado (Ware, 2012; Few 2013)

En el ejercicio se prioriza, la jerarquía visual que guíe la atención a indicadores críticos, la reducción de ruido gráfico y la consistencia en ejes, períodos y escala. El tablero

adicional para comunicar los resultados permite detectar anomalías, inconsistencias y tendencias, apoyando la interpretación analítica más allá de un reporte descriptivo.

5.4.2 Narrativa de datos

La narrativa de datos consiste en estructurar una secuencia para hilar el contexto del negocio con la narrativa que se quiere exponer en la gestión comercial. Es importante generar esta dinámica porque no se pueden basar los resultados en un indicador aislado, sino en una constructiva desarrollada por análisis creados. Esta narrativa debe iniciar evaluando el cumplimiento de metas entre lo que se presupuestos, el resultado de la cuota de cada asesor, la variación y la tendencia temporal; posteriormente profundizar los factores que intervienen en estas métricas, por ejemplo, las líneas de negocio, los clientes, el producto. Finalmente se deben analizar las oportunidades comerciales y la evolución donde se pueda identificar riesgos o posibilidades en diferentes segmentos para tomar líneas de acción.

El valor principal de la narrativa de datos no solo debe describir datos también debe estructurar la información donde se logren identificar patrones, desviaciones importantes y posibles cuello de botella. Este enfoque fortalece la capacidad preventiva de la organización y se alinea con la literatura en analítica en ventas, que destaca la importancia de analizar las dinámicas de proceso comercial más allá de los promedios agregados (Biemans, 2023; Conde, 2025)

5.5 Analítica en las ventas

La analítica de ventas consiste en la aplicación de técnicas cuantitativas, estadísticas y de aprendizaje automático sobre los datos comerciales de las organizaciones (Sultana et al., 2024). Actualmente debido a los ciclos más largos, múltiples variables para la toma de

decisiones, y altas inversiones, se hace necesario en materia de competitividad la aplicación de estas (Conde, 2025).

El uso de técnicas de analítica se ha empleado recientemente para optimizar operaciones clave como la gestión de inventarios, pronóstico de demanda, personalización de la experiencia del cliente, y en general la integración para la toma de decisiones. Se reportan mejoras entre 15 y 30% en costos de inventarios, 2 y 5% en ingresos por disponibilidad de productos. (Velásquez, 2025)

El monitoreo inteligente, se comprende como la observación constante de los datos de ventas, el comportamiento de los compradores, y las condiciones del mercado en tiempo real, lo que ofrece una visión sobre las tendencias operativas y comerciales clave, para la toma de decisiones estratégicas.

Para conocer el estado de la organización, y poder identificar los cambios que suponen la toma de decisiones, existen indicadores clave, que la gerencia debe conocer y seguir, y le permiten evaluar la efectividad de su fuerza de ventas y la sostenibilidad. Algunos de los principales indicadores son: tasa de conversión (proporción de oportunidades que llegan a ser ventas), valor del pipeline, el valor del cliente en el tiempo, y cumplimiento de cuotas.

Este último indicador, refleja la capacidad de la fuerza de ventas para alcanzar los objetivos propuestos por la organización en ingresos percibidos. Desde el punto de vista de la analítica, es una variable que puede ser gestionada gracias a la aplicación de diferentes técnicas, optimizando la priorización de oportunidades y la asignación en la labor comercial. Se puede fortalecer la precisión, para la evaluación más objetiva del rendimiento individual y colectivo del equipo de ventas.

6. Diseño Metodológico

6.1 Tipo de investigación

Esta investigación tiene un enfoque aplicado, por lo que busca el diseño de una solución práctica para la empresa Valveco S.A.S., que se orienta en la mejora de la medición de las metas comerciales, y la agilidad en la toma de decisiones estratégicas. Así mismo, el razonamiento se define entonces como deductivo porque parte del diagnóstico específico para una solución concreta. El grado de profundidad es descriptivo, con una fase exploratoria de diagnóstico, pero que se centra en la caracterización de procesos para diseñar una arquitectura específica a partir de conocimientos ya existentes.

Las mediciones se llevarán a cabo en un solo momento de tiempo por lo que el enfoque es transversal. Las fuentes de información son mixtas, dado que se requiere información cualitativa (revisión de literatura e identificación de los factores organizacionales) y cuantitativa (Indicadores, latencia de datos y rendimiento del sistema).

6.2 Fases del proyecto

El desarrollo de este trabajo se articula a través de cinco fases secuenciales que permiten transitar desde la conceptualización teórica hasta la formulación de una estrategia de adopción empresarial para Valveco S.A.S.

Fase 1: Fundamentación teórica y revisión de referentes

Esta fase establece la base sólida conceptual del proyecto mediante la revisión de fuentes bibliográficas de alto impacto. A continuación, se presenta la tabla de literatura seleccionada que sustenta el diseño de la arquitectura y el modelo de storytelling:

Tabla 1 *Revisión bibliográfica por sección*

Sección	Autor	Título	Año	Utilidad
Arquitectura y Analítica	Armbrust, M., et al.	Lakehouse: A New Generation of Open Platforms that Unify Data Warehousing and Advanced Analytics. Proceedings of CIDR	2021	Fundamenta la unificación de almacenamiento y analítica avanzada en plataformas abiertas.
	Nambiar, A., & Mundra, D.	An Overview of Data Warehouse and Data Lake in Modern Enterprise Data Management	2022	Proporciona una visión integral sobre la gestión de datos moderna en entornos empresariales.
Inteligencia de Negocios	Ain, N., et al.	Two decades of research on business intelligence system adoption, utilization and success – A systematic literature review	2019	Ofrece una perspectiva global sobre la adopción y éxito de los sistemas de BI.
Narrativa Visual	Ware, C.	Information visualization: Perception for design	2012	Sustenta los principios de diseño basados en la percepción humana para la visualización de datos.
Calidad de Datos	Wang, J., et al.	Overview of data quality: Examining the dimensions, antecedents, and impacts of data quality	2024	Define los estándares de calidad necesarios para garantizar la integridad de la información.

Nota: Elaboración propia

Fase 2: Diagnóstico de la gestión de información y madurez analítica

El diagnóstico se orienta a identificar el estado actual de la gestión comercial en Valveco S.A.S. mediante un análisis de factores internos y externos. En primera instancia, se aplica la matriz DOFA como herramienta de administración estratégica para el análisis externo, permitiendo identificar las oportunidades y amenazas derivadas de tendencias tecnológicas y acontecimientos del mercado sobre los cuales la organización no tiene control. De igual manera, se analizan las fortalezas y debilidades internas, entendidas como aquellas actividades controlables que impactan directamente en el manejo de la información y el desempeño del negocio.

En segunda instancia, el análisis interno se profundiza mediante el instrumento *TDWI BI and AI Maturity Model Assessment*, el cual permite medir el estado actual de los sistemas de información y la cultura analítica de la empresa. Además, se utiliza una encuesta como un segundo instrumento de medición basada en cinco variables críticas: la calidad de los datos comerciales, la latencia de la información para la toma de decisiones, el nivel de automatización y monitoreo bajo principios DataOps, el esfuerzo operativo requerido para la consolidación de datos y la confiabilidad y alineación de los KPIs comerciales. Este enfoque garantiza una comprensión técnica y estratégica de las brechas de madurez antes del diseño de la solución.

Fase 3: Diseño de la Arquitectura de Ciencia de Datos

En esta fase se materializa el diseño técnico de la arquitectura, integrando las fuentes de información relevantes identificadas en las bases de datos de la organización, específicamente World Office y DataCRM. Se establece un proceso de Extracción, Transformación y Carga (ETL) mediante principios de DataOps para asegurar la agilidad y la limpieza de los datos, corrigiendo posibles inconsistencias o valores nulos. La arquitectura propuesta garantiza la integración de datos en un repositorio centralizado

(Data Warehouse), permitiendo una vista única y consolidada de la información de finanzas, inventarios y ventas para homogeneizar el formato y la estructura de los datos institucionales.

Fase 4: Modelo de visualización estratégica y storytelling

En esta fase se construye el modelo de visualización estratégica con el objetivo de brindar a la gerencia herramientas visuales y métricas clave para el seguimiento efectivo del rendimiento comercial. El diseño se centra en la creación de narrativas visuales accionables que transformen los datos en historias coherentes mediante el uso de dashboards y storytelling. Este enfoque busca potenciar la eficiencia operativa y reforzar la validez de los resultados obtenidos, asegurando que la información sea procesada de manera eficiente para el cumplimiento de las metas estratégicas de Valveco S.A.S.

Fase 5: Propuesta de implementación y hoja de ruta

La fase final define el camino para la adopción institucional de la arquitectura diseñada, enmarcada en un cronograma detallado de actividades y recursos requeridos para su ejecución. Se establecen los procesos y usuarios intervinientes, describiendo perfiles y responsabilidades para asegurar una distribución eficiente de las tareas analíticas y de gobernanza de datos. Finalmente, se integran planes de capacitación y gestión del cambio para garantizar que el personal de Valveco S.A.S. se adapte efectivamente a las nuevas herramientas de ciencia de datos y procesos storytelling.

6.3 Análisis externo

Para este proyecto, se emplea la matriz DOFA como instrumento para el análisis externo, debido a que nos permite contextualizar a Valveco S.A.S., identificando factores, como la disponibilidad y uso de nuevas tecnologías, y desde un enfoque estratégico y

organizacional, determinar cuáles de esos factores pueden influir de forma positiva o negativa sobre los objetivos propuestos.

La matriz DOFA es una herramienta ampliamente aplicada en administración estratégica, debido a que permite tener una visión amplia de una organización. Según David, et al. (2023), los factores externos que se analizan son las amenazas y oportunidades, que abarca las tendencias y acontecimientos que pueden impactar la organización, pero sobre los cuales no existe control.

“Las fortalezas y debilidades internas son las actividades que una organización puede controlar y cuyo desempeño es especialmente bueno o deficiente” (David et al, 2023). Es decir, las fortalezas y debilidades se relacionan directamente con la administración del negocio, y uno de los aspectos clave es el manejo de la información.

6.4 Análisis interno

Para el análisis interno de Valveco S.A.S., se requiere comprender el estado actual de la organización en varios aspectos que comprenden una parte técnica sobre la gobernanza de los datos, el manejo de ellos y las herramientas que emplean para convertirlos en información útil, y también una parte referente a la cultura organizacional encaminada a la adopción del diseño propuesto.

Se emplearán dos herramientas para ese propósito, una es el instrumento para evaluar el modelo de madurez de BI e IA, desarrollada por the data warehouse institute (TDWI), que es una organización que se dedica a la educación e investigación en el área analítica y gestión de datos desde 1995.

La otra herramienta, consiste en una encuesta dirigida a todos los colaboradores del área de ventas de la empresa, que sirve para recoger información cualitativa de las variables seleccionadas.

6.5 Población de la muestra y ficha técnica

A continuación, presentamos la ficha técnica para la aplicación del instrumento de TDWI para evaluar el modelo de madurez de BI e IA.

Tabla 2 *Ficha técnica herramienta TDWI*

Características	Descripción
Objetivo	Evaluar las variables relacionadas con el modelo de madurez de BI e IA: organización, recursos, infraestructura de datos, analítica y gobernanza.
Período de recolección de datos	Marzo de 2026
Lugar de aplicación	Virtual
Población	2 colaboradores clave: gerente general y gerente administrativo
Muestra	1 participante
Modo de recolección	A través del enlace: https://tdwi.org/research/2025/02/bi-all-tdwi-bi-and-ai-maturity-model-assessment.aspx

Nota: Elaboración propia

Tabla 3 *Ficha técnica encuesta*

Características	Descripción
Objetivo	Recopilar información cualitativa sobre el estado actual de los procesos de integración, calidad y trazabilidad de datos comerciales en Valveco S.A.S.
Período de recolección de datos	Marzo 2026
Lugar de aplicación	Bogotá D.C
Población	9 colaboradores
Censo	9 participantes
Modo de recolección	Encuesta recolectada a través de la herramienta Google forms: https://docs.google.com/forms/d/e/1FAIpQLSdi30MFCq3nOWqTXyey8KESMKssJvtMZAL4rVZ_j_NGGUX3KQ/viewform?usp=dialog

Nota: Elaboración propia

6.6 Identificación de variables

Para la aplicación de la encuesta como uno de los instrumentos de medición en el área comercial de Valveco S.A.S., se definen las siguientes variables clave para el análisis interno, las cuales se pueden observar en la tabla 4.

Tabla 4 *Variables para diagnóstico interno*

Variable	Descripción
Calidad de los datos comerciales	Esta variable busca evaluar el estado de las dimensiones de calidad (exactitud, completitud, consistencia, oportunidad y trazabilidad) en los datos que alimentan los KPIs de ventas, con el fin de asegurar decisiones confiables y tableros válidos. (Wang et al., 2024; Li et al., 2022; Bernardo et al., 2024)
Latencia de información para la toma de decisiones	Esta variable busca cuantificar el desfase entre la generación del dato comercial y su disponibilidad validada en reportes y tableros, con el fin de reducir decisiones reactivas y mejorar la oportunidad gerencial. (Conde, 2025; Biemans, 2023)
Automatización y monitoreo de procesos (DataOps)	Esta variable busca evaluar la automatización de los pipelines de integración y su monitoreo (bitácoras, alertas, pruebas), con el fin de reducir errores, acortar ciclos y fortalecer la trazabilidad. (Ribeiro et al., 2021; Farinha et al., 2024; Fannouch et al., 2025)
Esfuerzo operativo de consolidación	Esta variable busca medir la carga manual dedicada a descargar, depurar y conciliar información desde World Office, DataCRM y Excel, con el fin de identificar reprocesos, cuellos de botella y oportunidades de automatización. (Solano y Cruz, 2024)
Confiabilidad y alineación de KPIs comerciales	Esta variable busca asegurar que los indicadores clave (como cuota y cobertura del pipeline) están definidos con fórmulas y fuentes claras, sean consistentes entre áreas y cuenten con trazabilidad. El objetivo es garantizar métricas comparables y útiles para la gestión, facilitando decisiones ágiles y basadas en datos. (Ain et al., 2019; Bernardo et al., 2024)

Nota: Elaboración propia

6.7 Instrumento de medición

Para obtener información que permita conocer el estado de la organización referente a las variables definidas, se emplean dos herramientas de análisis. La primera, es un instrumento proporcionado por the data warehouse institute (TDWI) para evaluar el modelo

de madurez de BI e IA. Nos permite tener un diagnóstico de la organización en las siguientes dimensiones: organización, recursos, infraestructura, analítica y gobernanza de datos, a través de un cuestionario de 50 preguntas de selección múltiple.

La segunda herramienta para el análisis interno consiste en una encuesta dirigida a los colaboradores implicados en el proceso comercial, en ella se plantean preguntas con base en las variables calidad de los datos comerciales, latencia de información para la toma de decisiones, automatización y monitoreo de procesos (DataOps), esfuerzo operativo de consolidación, confiabilidad y alineación de KPIs comerciales. Se encuentra en el anexo 1. Se usa una escala de Likert para responder a las preguntas.

6.8 Validación del instrumento de medición

La encuesta propuesta para Valveco S.A.S. fue sometida a una validación de contenido con un panel de cuatro especialistas externos a la organización. Esta revisión tuvo como objetivo confirmar que cada pregunta resulta clara, pertinente y relevante para evaluar las variables del estudio (calidad de datos, latencia, automatización/monitoreo, esfuerzo operativo y confiabilidad de KPIs), además de garantizar claridad, coherencia conceptual y redacción adecuada. Para la valoración se empleó el modelo de V de Aiken, a partir del cual se estimó el grado de acuerdo de los expertos sobre la idoneidad de las afirmaciones incluidas en el instrumento.

Tabla 5 *Expertos evaluadores de la herramienta encuesta*

Nombre	Institución y cargo	Título	Experiencia	Tiempo
Lina Marcela Torres Londoño	Visa. Consulting and analytics senior manager.	Ingeniero de sistemas. Msc. Ciencias de la computación. Especialista en dirección de proyectos.	Arquitecto de soluciones y gerente de operaciones de cliente. Proyectos de sistemas de información (arquitectura, diseño, implementación, migración).	25 años
Roberto Adolfo Díaz Díaz	Universidad del Tolima. Director del programa de economía.	Economista. Msc en economía. Aspirante al título de doctor en ciencias administrativas.	Docencia e investigación.	20 años
Jaime Alberto Ochoa Durán	Universidad EAN. Docente.	Msc. Ingeniería analítica de datos, especialista en IA, Ingeniero de petróleos, antropólogo social.	Docencia e investigación. Experto en IA, analítica avanzada y proyectos I+D+i.	9 años
Santiago Alzate Cote	Universidad Sergio Arboleda. Decano del programa de finanzas y comercio exterior.	Doctor en administración estratégica, Msc Administración de empresas, especialista en matemáticas aplicadas, Ingeniero industrial.	Docente, experto en gestión estratégica e integración de tecnología.	22 años

Nota: elaboración propia

En la tabla 6 se muestran los resultados de la validación. La V-Aiken con valor 1 indica completo acuerdo entre los expertos. Valores superiores a 0.8 indican validez aceptable de las preguntas, y valores inferiores débil validez por lo que la pregunta debe ser reformulada.

Para calcular la V de Aiken, se promedian los valores de calidad, pertinencia y relevancia que cada uno de los evaluadores le dio a la pregunta. Después se calcula el coeficiente con la fórmula a continuación, donde S es la suma de los valores positivos, n el número de validadores, y c el número de valores en la escala.

$$v = \frac{S}{\big(n(c-1)\big)} \qquad (1)$$

Tabla 6 *Resultados validación V de Aiken de la encuesta*

A. Calidad de los datos comerciales		EVALUADOR 1	EVALUADOR 2	EVALUADOR 3	EVALUADOR 4	V DE AIKEN
1	La información comercial disponible en los sistemas de Valveco S.A.S. contiene los datos necesarios y suficientes para realizar adecuadamente mi trabajo	1,00	1,00	1,00	1,00	**1,00**
2	Cuando debo comparar cifras comerciales en diferentes reportes o sistemas (por ejemplo, CRM, ERP o reportes financieros), estas coinciden entre sí.	1,00	1,00	1,00	1,00	**1,00**
3	Los datos comerciales que debo utilizar reflejan de manera precisa lo que ocurre en la operación semanal de ventas.	1,00	1,00	1,00	1,00	**1,00**

4	La estructuración, organización y estandarización de los datos comerciales, facilitaría el análisis y la toma de decisiones en mi trabajo.	1,00	1,00	0,67	0,00	**0,67**
B. Latencia de información para la toma de decisiones		EVALUADOR 1	EVALUADOR 2	EVALUADOR 3	EVALUADOR 4	V DE AIKEN
1	La información comercial está disponible en el momento en que la necesito para hacer seguimiento a mis resultados y metas.	1,00	1,00	1,00	1,00	**1,00**
2	Los reportes disponibles me permiten conocer el avance de mis resultados en periodos específicos (diarios, semanales, mensuales).	1,00	1,00	0,67	1,00	**0,92**
3	La información comercial que utilizo se encuentra actualizada cuando la consulto.	1,00	1,00	1,00	1,00	**1,00**
4	Con frecuencia debo consolidar información manualmente desde diferentes fuentes (archivos, sistemas o reportes).	1,00	1,00	1,00	1,00	**1,00**
C. Esfuerzo operativo de consolidación		EVALUADOR 1	EVALUADOR 2	EVALUADOR 3	EVALUADOR 4	V DE AIKEN
1	La revisión o verificación de cifras comerciales requiere tiempo adicional dentro de mis actividades habituales	1,00	1,00	1,00	1,00	**1,00**
2	La forma en que se organiza y gestiona la información comercial genera dificultades en el desarrollo de las actividades del área.	0,67	1,00	1,00	1,00	**0,92**

D. Automatización y monitoreo de procesos (DataOps)		EVALUADOR 1	EVALUADOR 2	EVALUADOR 3	EVALUADOR 4	V DE AIKEN
1	Con frecuencia se deben realizar reprocesos o se repiten tareas al revisar información o seguimiento a las cifras.	1,00	1,00	1,00	1,00	**1,00**
2	Mejorar la forma en que se organizan, integran y consolidan los datos comerciales reduciría la carga operativa actual del área comercial.	1,00	1,00	1,00	1,00	**1,00**
E. Confiabilidad y alineación de KPIs comerciales		EVALUADOR 1	EVALUADOR 2	EVALUADOR 3	EVALUADOR 4	V DE AIKEN
1	Los indicadores comerciales que utilizo están claramente definidos y documentados.	1,00	1,00	1,00	1,00	**1,00**
2	Entiende cómo se calcula el cumplimiento de metas en el área comercial.	1,00	1,00	1,00	1,00	**1,00**
3	Contar con indicadores definidos aumenta mi confianza en los resultados reportados	1,00	0,67	1,00	1,00	**0,92**
4	Tengo claridad sobre los criterios con los cuales se evalúa mi desempeño en el área comercial.	1,00	1,00	1,00	1,00	**1,00**

Nota: Elaboración propia

Los resultados de V de Aiken muestran que todas las preguntas, excepto una, cumplen con el criterio de calidad, pertinencia y relevancia, el resultado del coeficiente es superior a 0.80, sin embargo, hay algunos comentarios de los evaluadores que son tenidos en

cuenta para hacer cambios en la redacción de algunas preguntas. El instrumento definitivo se encuentra en el anexo 1. La pregunta “La forma en que se organiza y gestiona la información comercial genera dificultades en el desarrollo de las actividades del área” se elimina porque en las observaciones a pesar de cumplir con los criterios, fue considerada repetitiva por los evaluadores.

A continuación, en la tabla 7 se visualizan las observaciones de los evaluadores, y el cambio en las preguntas.

Tabla 7 *Reformulación de preguntas del instrumento*

	Pregunta	Observación	Pregunta reformulada
A3	Los datos comerciales que debo utilizar reflejan de manera precisa lo que ocurre en la operación semanal de ventas.	¿Solo semanal?	Los datos comerciales reflejan con precisión lo que ocurre en la operación real de ventas
A4	La estructuración, organización y estandarización de los datos comerciales, facilitaría el análisis y la toma de decisiones en mi trabajo.	Cambiaría "facilitaría" simplemente por "facilita" No se entiende el alcance de la pregunta, se debe definir bien qué es estructuración, estandarización	La forma en que se estructuran, organizan y estandarizan los datos comerciales en Valveco S.A.S. (formatos, nombres y reglas comunes para registrar la información) facilita el análisis y la toma de decisiones en mi trabajo.
B4	Con frecuencia debo consolidar información manualmente desde diferentes fuentes (archivos, sistemas o reportes).	¿No debería estar en el siguiente grupo?	Observación: Se pasa a la variable esfuerzo operativo de consolidación.

C2	Mejorar la forma en que se organizan, integran y consolidan los datos comerciales reduciría la carga operativa actual del área comercial.	Esta pregunta induce a la respuesta afirmativa en el encuestado, lo que puede sesgar la investigación.	La forma en que se organizan, integran y consolidan los datos comerciales influye en la carga operativa del área comercial.
E2	Entiende cómo se calcula el cumplimiento de metas en el área comercial.	Para ser consistentes con la redacción debería decir "entiendo"	Entiendo cómo se calcula el cumplimiento de metas en el área comercial.
E3	Contar con indicadores definidos aumenta mi confianza en los resultados reportados	La pregunta induce a una respuesta afirmativa	Contar con indicadores definidos influye en mi nivel de confianza en los resultados reportados.

Nota: elaboración propia

7. Diagnóstico Organizacional

Para realizar el diagnóstico organizacional, seguimos la metodología descrita en el capítulo 6. Para el análisis externo elaboramos una matriz DOFA en equipo con el gerente administrativo y gerente general de Valveco S.A.S. Para el análisis interno la gerente administrativa diligenció el instrumento desarrollado por TDWI y se hizo un censo entre los colaboradores del área comercial para conocer la percepción de ellos frente a las variables seleccionadas por medio del instrumento encuesta. Los resultados se presentan a continuación.

7.1 Análisis externo: Matriz DOFA

Tabla 8 *Matriz DOFA*

FORTALEZAS	**DEBILIDADES**
Equipo de ventas con amplio conocimiento de los productos y actitud de aprendizaje. Aliados estratégicos reconocidos en el sector (Danfoss y asco). Son importadores directos de algunos productos. Buena reputación y satisfacción del cliente. Entregas rápidas y servicio ágil. Base de clientes con seguimiento. Trayectoria con empresas grandes de varios sectores. Se lleva un CRM con seguimiento de las actividades comerciales. Cuentan con datos históricos de ventas desde el año 2020. Hay análisis de inventario y conexión con comercial.	El crecimiento en los datos puede superar la capacidad tecnológica actual. Información de clientes puede estar duplicada o incompleta en el CRM. Sólo se mide el desempeño comercial por cumplimiento de cuotas de ventas. Falta integración entre ventas, marketing y atención al cliente. No existe trazabilidad de PQR. Dependencia de los tiempos de proveedores y de logística. Datos de ventas en dos sistemas diferentes. No existe segmentación avanzada de clientes. Existe un conocimiento limitado sobre el comportamiento de los clientes que no finalizan el proceso de compra.

OPORTUNIDADES	AMENAZAS
Interés en implementar modelos de analítica de datos. Integrar datos por los nuevos canales digitales de ventas (TikTok, Linkedin, Whatsapp). Posibilidad de analizar comportamiento de clientes y estacionalidad de demanda. Expansión hacia sectores como gas y petróleo. Predicción de la demanda haciendo uso de los datos históricos. Implementación de visualización de datos. Los productos se requieren de forma inmediata para no causar interrupción en los procesos productivos del cliente. Integración entre los datos de ventas, inventario y atención a cliente para mejorar la segmentación y la toma de decisiones.	En el mercado externo para ser competitivo hay que mantener volúmenes de pedido altos. Competidores más grandes en el mercado. Ingreso de productos chinos de menor precio. Variaciones de la TRM que afectan las importaciones, lo que exige mayor capacidad de análisis. Competidores usando métricas y analítica comercial. Cambios normativos o certificaciones pueden dejar productos obsoletos. Productos con precio muy alto que no se pueden mantener en almacén.

Nota: Elaboración propia

A partir de la matriz, se pueden formular algunas estrategias cruzadas, combinando los factores evaluados. Fortalezas con oportunidades para mejorar la competitividad comercial, mitigar las debilidades y enfrentar mejor los cambios en el entorno.

Tabla 9 *Estrategias FO-FA-DO-DA*

Estrategias	Fortalezas	Debilidades
Oportunidades	Emplear la base de clientes para ampliar las oportunidades de negocio implementando modelos de analítica de datos. Emplear herramientas de visualización de datos y analítica comercial a partir de los datos históricos Aprovechar la capacidad de hacer entregas rápidas y servicio ágil para posicionarse.	Implementar una arquitectura que conecte las fuentes de datos del CMR, inventarios y ventas. Aplicar segmentación de clientes para entender las razones por las que no se concreta la venta. Incorporar indicadores comerciales más completos (tasa de conversión valor del pipeline y el valor del cliente en el tiempo)
Amenazas	Optimizar la gestión de inventarios para conseguir la compra con mejor TRM. Aprovechar el conocimiento técnico y la gestión de los inventarios para evitar que los productos se vuelvan obsoletos. Diferenciarse de la competencia por la rapidez del servicio y entrega y el respaldo de proveedores reconocidos, mejorando la disponibilidad de productos.	Implementar el monitoreo de indicadores y la analítica en las ventas para competir con empresas que ya lo hacen. Mejorar la gestión y calidad de los datos para no perder oportunidades de negocio frente a competidores. Mejorar la integración de los sistemas de información para reducir la dependencia de procesos manuales.

Nota: Elaboración propia

7.2 Análisis interno: Herramienta TDWI

El director administrativo diligenció la prueba de acuerdo con lo descrito en el punto 6.4 del diseño metodológico, los resultados se presentan a continuación. En cada una de las cinco dimensiones se obtuvo un puntaje que determina en qué etapa se encuentra la organización, en una escala de 1 a 20 de la siguiente forma:

Etapa 1 - Naciente: 0 a 6

Etapa 2 - Temprano: 7 a 10

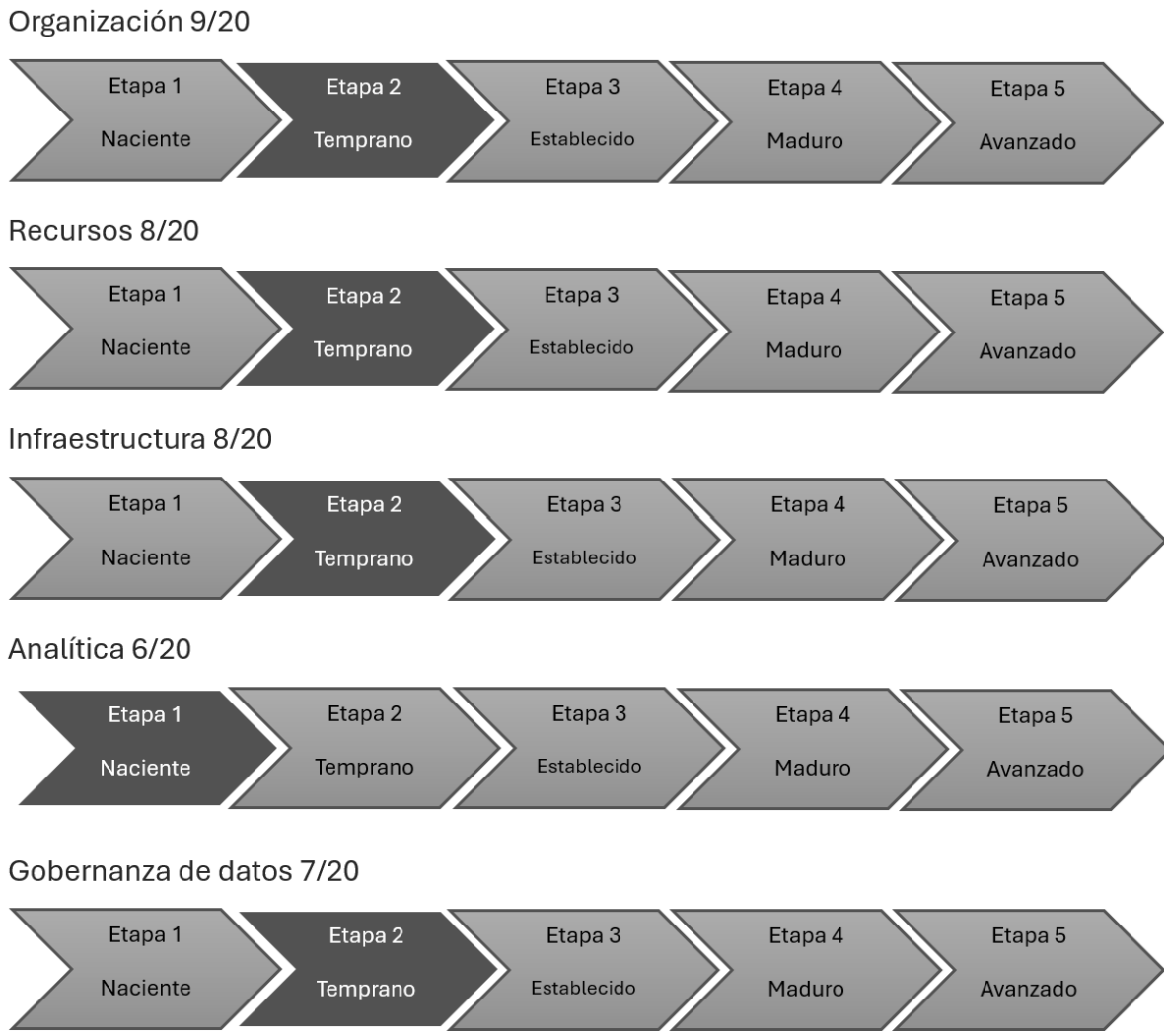
Etapa 3 - Establecido: 11 a 14

Etapa 4 - Maduro: 15 - 18

Etapa 5 - Avanzado/Visionario: 19 a 20

A continuación, se hace el análisis de cada una de las dimensiones incluidas:

Figura 4 *Resultado TDWI Valveco S.A.S.*



Nota: Elaboración propia

Organización: En la dimensión organización se obtuvo la mejor puntuación, lo que evidencia que Valveco S.A.S. comprende la importancia de los datos para tomar decisiones y muestra apertura hacia la implementación de iniciativas de analítica de datos.

Esto es coherente con lo abordado en el proyecto, ya que la empresa cuenta con un sistema centralizado para el área comercial (CRM) y otro para la gestión contable (World Office). En la actualidad es posible realizar auto seguimiento de las ventas efectuadas para el área comercial.

Identificar este aspecto es fundamental para la ejecución del proyecto, porque nos demuestra que la empresa está dispuesta a avanzar en términos de analítica de datos. Sin esta apertura organizacional, cualquier propuesta de implementación tendría un alto riesgo de fracasar.

Recursos: Las capacidades del personal, presupuesto y técnicas aún son limitadas. Se cuenta con el uso de software para recolectar datos, pero la información no se analiza. En nuestro proyecto no debemos asumir que la empresa cuenta con competencias avanzadas para analítica. Por el contrario, se debe considerar la necesidad de capacitación para facilitar la comprensión y el aprovechamiento de las soluciones que se implementen.

Infraestructura de datos: Existe una base importante para la gestión de datos, sin embargo, aún no dispone de una arquitectura de analítica integrada. Desde el año 2020 con la implementación del CRM, se recolectan datos, hay una conexión de este sistema con los movimientos de inventario, lo que facilita la identificación de la disponibilidad de stock al momento de realizar una negociación comercial.

Como desventaja el sistema World Office y el CRM están desalineados, lo que genera dificultad para seguimiento a las ventas en dinero y en ventas efectivas, falta una integración de las plataformas usadas.

Analítica: Esta es la dimensión más baja y es el hallazgo más importante de la encuesta realizada. La empresa actualmente tiene un comportamiento reactivo y no proactivo, el

análisis de información se realiza de forma descriptiva con seguimiento comercial y no es tan cercano a una estructura analítica de datos.

A pesar de que se realizan seguimientos comerciales, se compara el período con el año anterior y el crecimiento con el presupuesto del año, se debe hacer una explotación avanzada con una segmentación marcada, predicciones confiables, centralización de información y automatización.

Gobernanza de datos: Esta es la dimensión que contempla la estrategia de gobernanza, acceso, uso correcto de datos, catálogos, privacidad, así como el seguimiento y control de modelos.

En Valveco S.A.S. actualmente no hay una trazabilidad de las quejas, ni de las razones de pérdida de ventas. Además, se identifican registros de clientes duplicados, campos diligenciados de forma incompleta y ausencia de conexión entre los sistemas de información.

Los resultados del TDWI BI and AI Maturity Model demuestran que Valveco S.A.S. se encuentra en una etapa temprana de madurez analítica, con una puntuación global de 38/100, Aunque la organización presenta avances iniciales en estructura organizacional, uso de CRM, seguimiento comercial y disponibilidad de información histórica, persisten brechas relevantes en infraestructura integrada, recursos especializados, gobernanza de datos y capacidades analíticas.

La propuesta de implementación se debe orientar hacia la realización de un plan progresivo, priorizando la calidad, integración, estandarización y visualización de datos. Tiene prioridad frente a la implementación de soluciones más avanzadas de IA.

Figura 5 *Puntuación general prueba TDWI*



Nota: Elaboración propia basados en el resultado de la prueba TDWI aplicada a Valveco S.A.S.

Valveco S.A.S. ha superado la etapa naciente y ha realizado inversiones para el manejo de los datos. Cuenta con algunas herramientas, sobre las cuales se puede obtener un mayor valor a partir de la información. La empresa tiene bases para usar BI y la capacidad para evolucionar hacia el uso de la ciencia de datos, aunque aún no cuenta con una capacidad analítica consolidada.

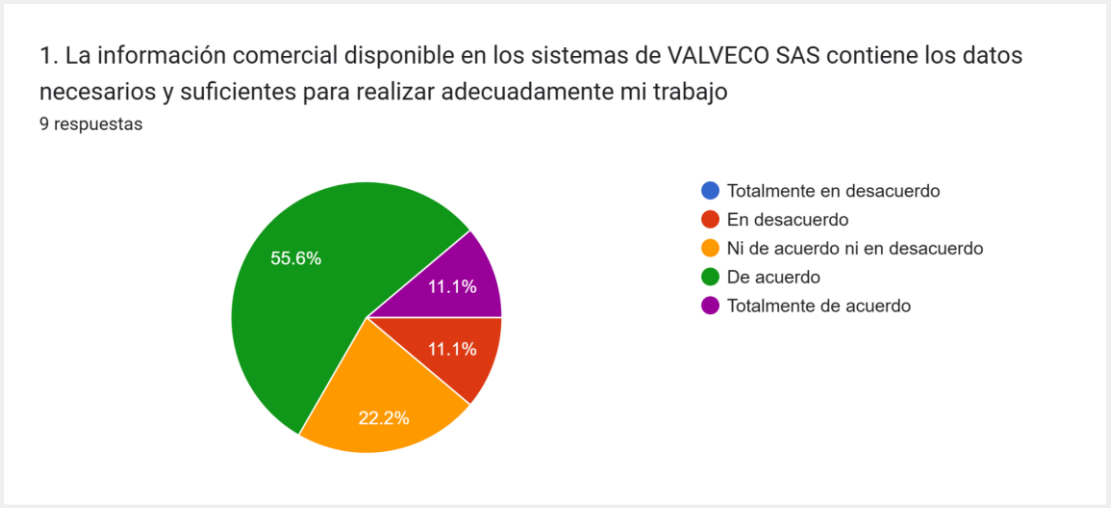
7.3 Análisis interno: Encuesta

A continuación, se presentan los resultados de la herramienta diseñada por nuestro grupo de trabajo, que fue aplicada como un censo a los colaboradores del área comercial, con las variables identificadas.

7.3.1 Calidad de los datos comerciales

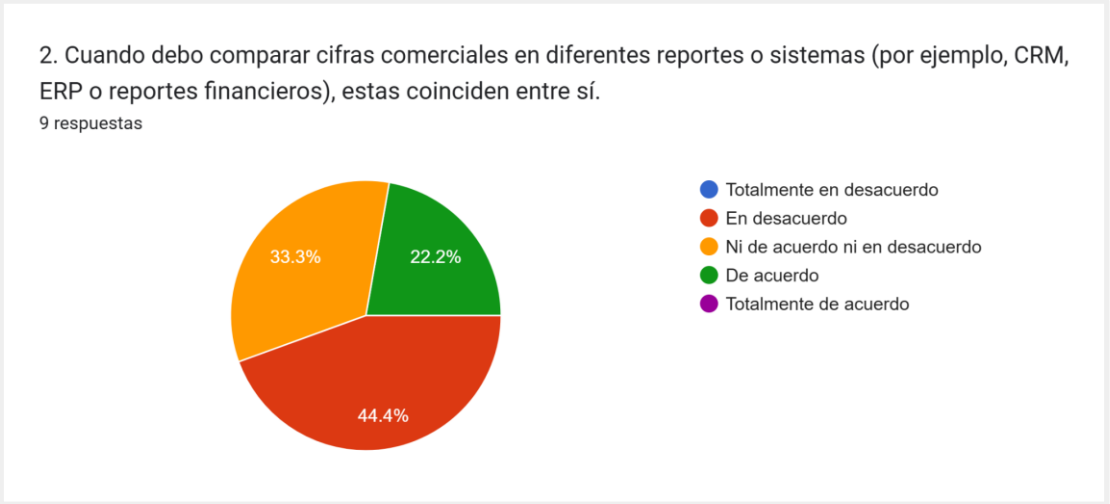
La mayoría de los colaboradores del área comercial considera favorable la suficiencia de la información comercial disponible, el 66.7% de los encuestados considera que los sistemas, aunque estén separados contienen los datos necesarios para ejecutar sus actividades diarias. Sin embargo, existe un 22.2% que tiene respuesta neutral y un 11.1% para los que no es suficiente la información para su ejecución diaria. Hay información funcional, pero puede existir una mejora.

Figura 6 *Pregunta 1A encuesta*



Nota: Elaboración propia

Figura 7 *Pregunta 2A encuesta*



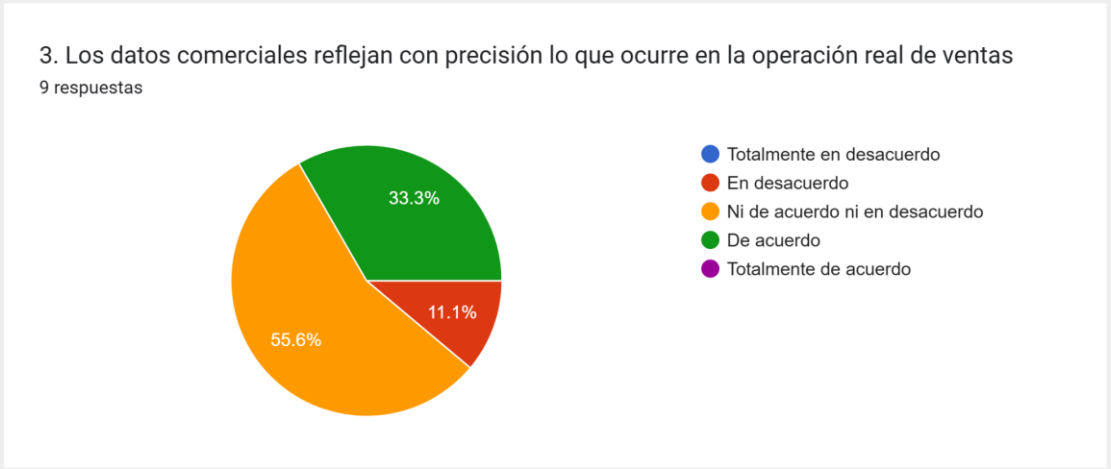
Nota: Elaboración propia

La percepción que se tiene sobre la consistencia de la información comercial entre los sistemas que se usan en Valveco S.A.S. es desfavorable, el 44.4% de los encuestados están en desacuerdo con que las cifras coincidan entre el World Office y el CRM, el 33.3%

elige posición neutral y el 22.2% está de acuerdo que sí coinciden los sistemas. Este resultado evidencia diferencia entre las cifras reportadas por los sistemas usados.

Este hallazgo soporta la debilidad en la confiabilidad y coherencia de datos, lo que puede afectar la confianza en los reportes y dificultar la toma de decisiones. Se debe fortalecer la integración y el control de la información entre los sistemas usados.

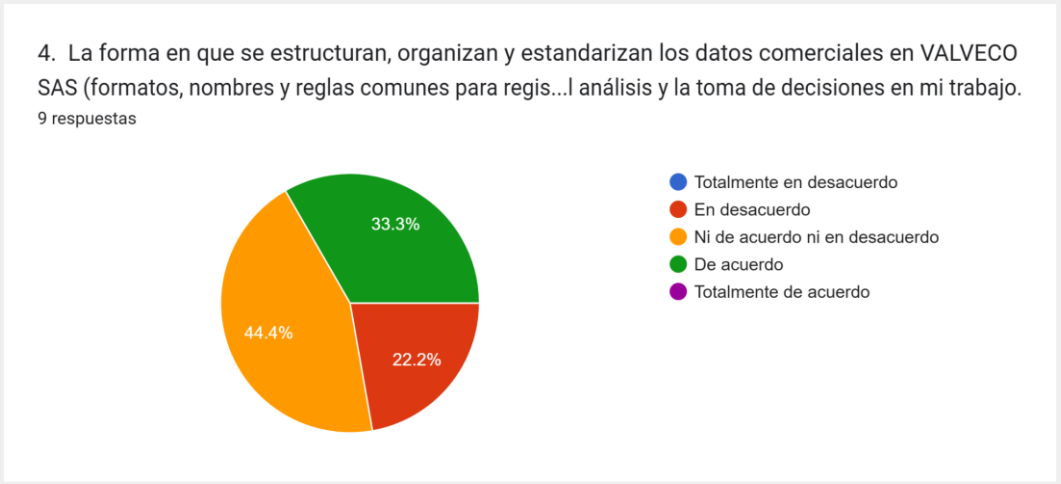
Figura 8 *Pregunta 3A encuesta*



Nota: Elaboración propia.

Los resultados muestran una percepción intermedia en cuanto a la noción que tienen de la precisión de los datos comerciales frente a la operación semanal de ventas. El 33.3% de los encuestados están de acuerdo con la afirmación, el 56.6% se mantiene en una posición neutral y el 11.1% está en desacuerdo. Esto indica que, aunque la tercera parte del personal considera que la información si muestra la realidad operativa, la mayoría tiene duda en si es clara o no. Esto sugiere que los datos comerciales son útiles para realizar seguimientos, pero todavía no generan la suficiente confianza en precisión.

Figura 9 *Pregunta 4A encuesta*

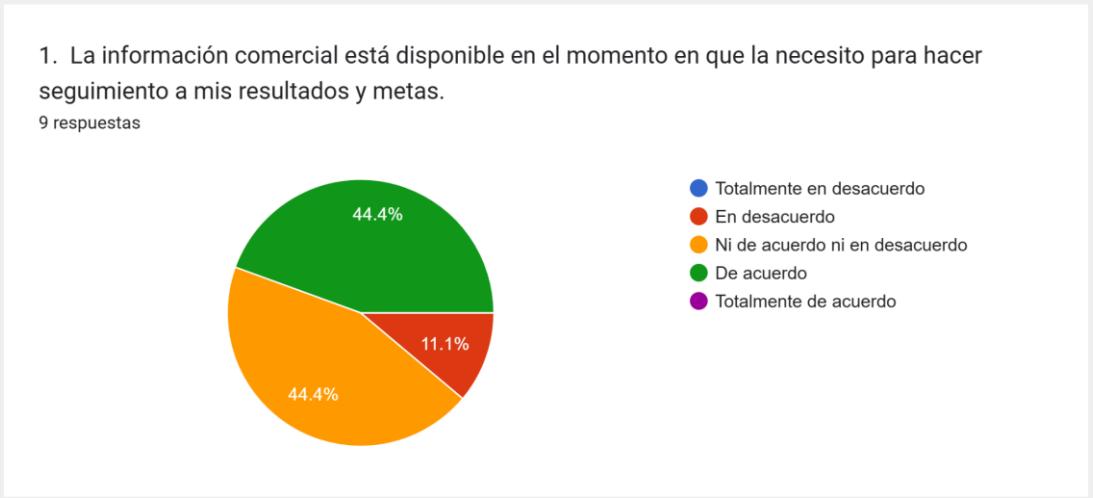


Nota: Elaboración propia

Se evidencia una percepción de desconocimiento en la forma en la que actualmente la empresa organiza, estructura y estandariza los datos comerciales. El 44.4% de los encuestados adopta una posición neutral, lo que demuestra que no hay conocimiento sobre cómo se realizan dichos procesos o que no intervienen en esta organización, el 22.2% está en desacuerdo y el 33.3% está de acuerdo con que la manera que existe actualmente. Por no tener una arquitectura de analítica de datos no se tiene una consistencia en la información.

7.3.2 Latencia de información para la toma de decisiones

Figura 10 *Pregunta 1B encuesta*

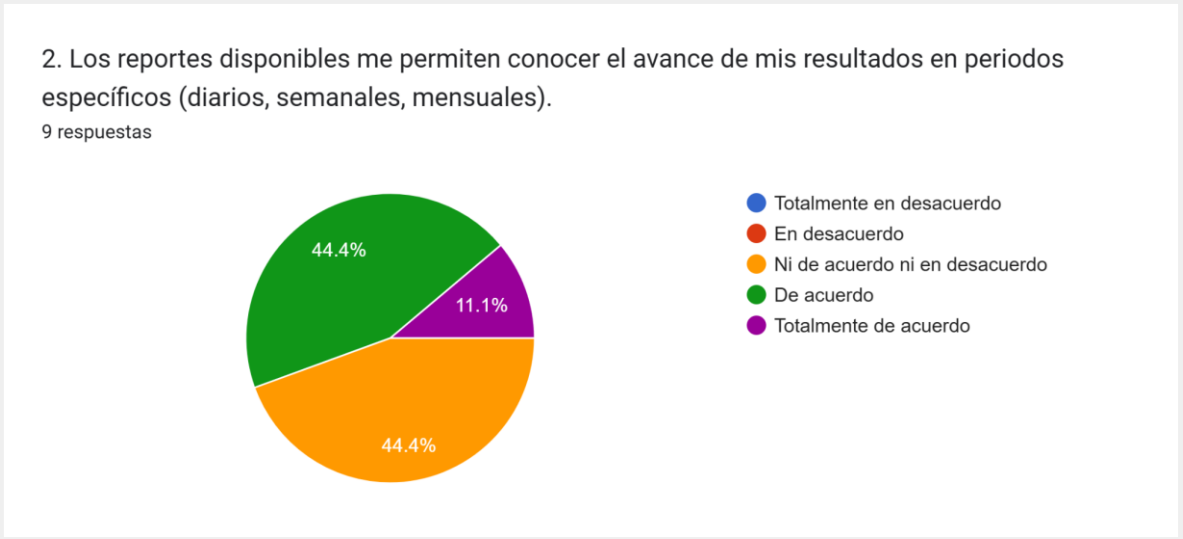


Nota: Elaboración propia

El 44.4% de los encuestados está de acuerdo con que la información se encuentra disponible cuando se necesita, mientras que el otro 44.4% elige una posición neutral y un 11.1% está en desacuerdo. A pesar de que existe un gran porcentaje que considera que la información está para el momento en el que se requiere todavía otro gran porcentaje no tiene la percepción plena que esté de manera oportuna

Esto sugiere que, aunque la empresa disponga de información útil para el seguimiento comercial, aún presenta limitaciones en la disponibilidad inmediata o en la percepción de acceso oportuno para todos los usuarios.

Figura 11 *Pregunta 2B encuesta*



Nota: Elaboración propia

Los resultados muestran que la utilidad de reportes para conocer el avance de los resultados en períodos específicos es útil. El 44.4% de los encuestados está de acuerdo y el 11.1% totalmente de acuerdo, esto en su totalidad representa un 55.5% de valoración positiva, sin embargo, el 44.4% adopta una posición neutral. Esto indica que los reportes que existen y están disponibles en este momento, si logran ser útiles para el tema de seguimiento en el desempeño comercial, aunque no todos los encuestados consideran la misma utilidad o claridad en la información. Esto nos refuerza el tema de una buena iniciativa para intervenir, pero se debe mejorar en control de resultados para aprovecharlo para todos los usuarios.

Figura 12 *Pregunta 3B encuesta*



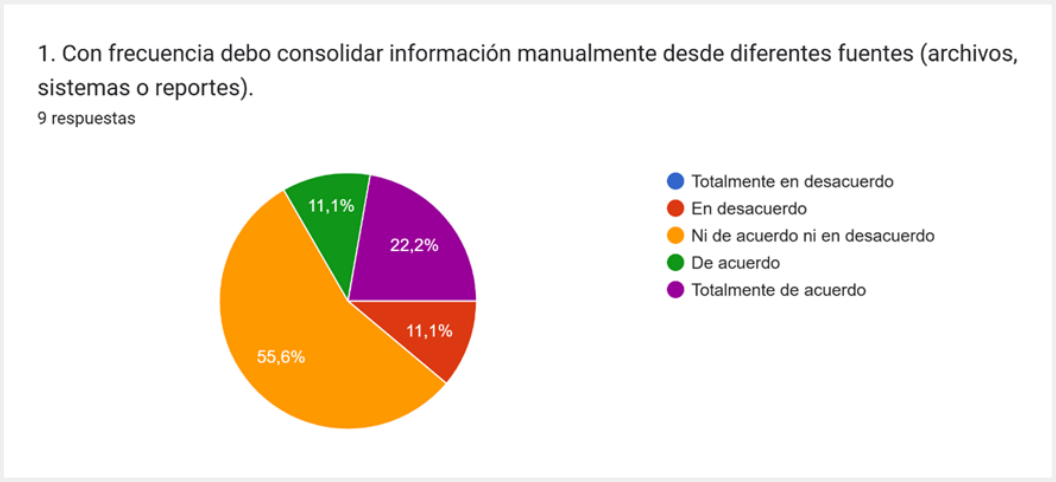
Nota: Elaboración propia

Los resultados muestran una baja favorabilidad sobre la actualización de la información comercial al momento que es consultada. El 33.3% de los encuestados están en desacuerdo o totalmente en desacuerdo, el 44.4% adopta una posición neutral y el 22.2% está de acuerdo. Esto evidencia que no existe una certeza que la información siempre esté actualizada al ser requerida

Esto nos sugiere la oportunidad para intervenir en tema de sincronizar los datos para intervenir el seguimiento comercial tomando decisiones en tiempo real

7.3.3 Esfuerzo operativo de consolidación

Figura 13 *Pregunta 1C encuesta*

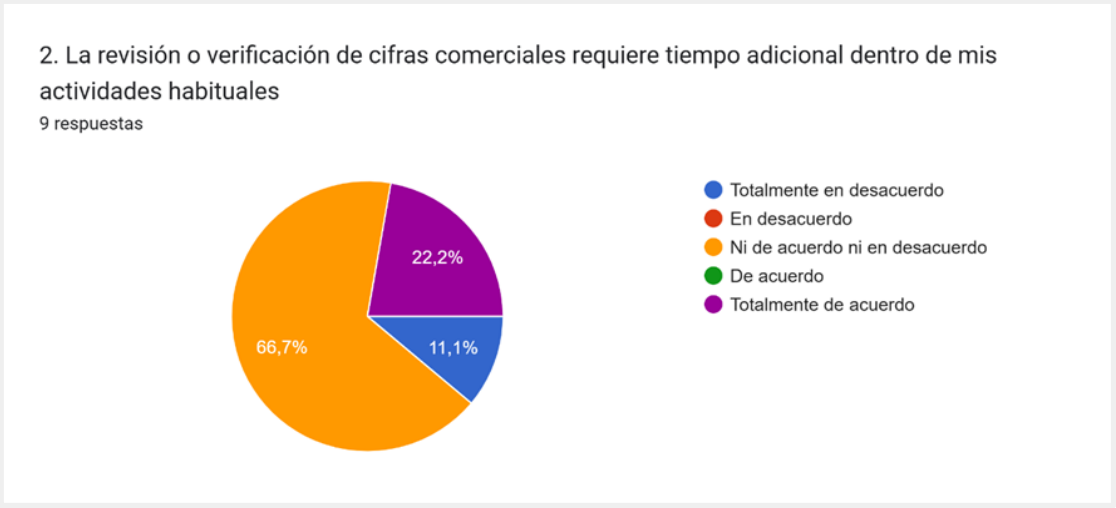


Nota: Elaboración propia

Los resultados evidencian una consolidación manual de información desde diferentes fuentes de datos y es una práctica frecuente. El 33.3% de los encuestados están de acuerdo o totalmente de acuerdo con esta afirmación, mientras que el 55.6% adopta una posición neutral y el 11.1% están en desacuerdo. Se requiere una intervención manual para recolectar los datos y descargar reportes en las plataformas existentes

Esto sugiere una oportunidad de mejora en la integración y automatización de los datos, ya que hacerlo manualmente incrementa el esfuerzo operativo.

Figura 14 *Pregunta 2C encuesta*

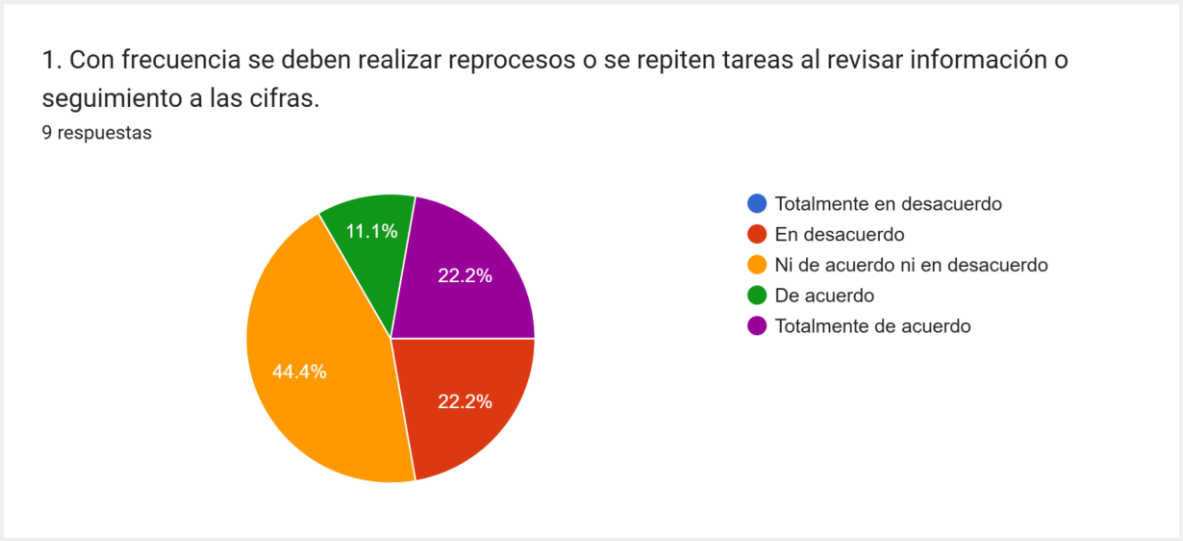


Nota: Elaboración propia

Los resultados evidencian que la revisión o verificación de cifras comerciales representa una carga adicional dentro de las actividades habituales. El 22.2% de los encuestados está totalmente de acuerdo con esta afirmación, mientras que el 66.7% mantiene una postura neutra y el 11.1% está totalmente en desacuerdo. Aunque la mayoría de los encuestados expresa una postura de no invertir tiempo adicional para revisar y verificar los datos, si sugiere una validación de cifras que demanda tiempo extra para algunos colaboradores.

7.3.4 Automatización y monitoreo de procesos (DataOps)

Figura 15 *Pregunta 1D encuesta*



Nota: Elaboración propia

Los resultados evidencian que la repetición de tareas o reprocesos al revisar la información y dar seguimiento a las cifras es una situación común en la empresa. El 33.3% de los encuestados está de acuerdo o totalmente de acuerdo con esta afirmación, el 44.4% adopta una posición neutral y el 22.2% está en desacuerdo. Esto sugiere que no todos los encuestados perciben el problema con la misma intensidad, pero demuestra señales de ineficiencia en el manejo para validar la información comercial

Esto permite identificar una oportunidad de mejora en automatización y monitoreo de procesos, ya que existen reprocesos por fallas de actualización, integración o control de datos.

Figura 16 *Pregunta 2D encuesta*

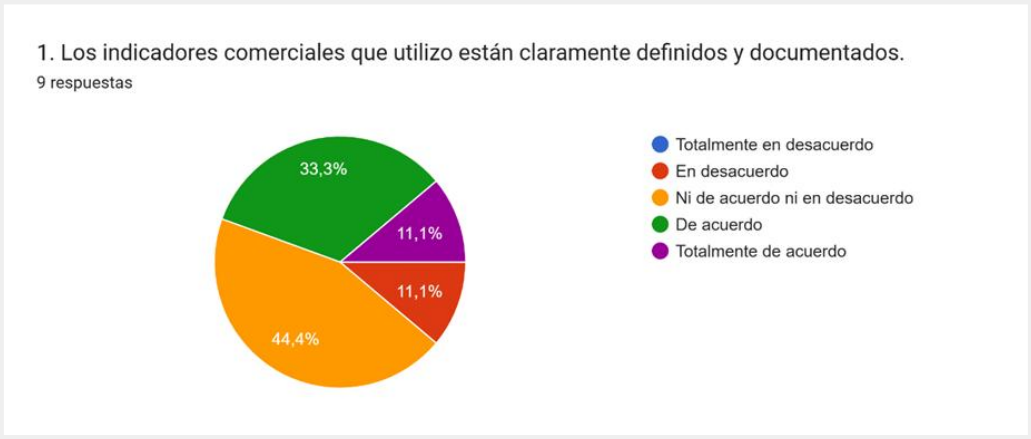


Nota: Elaboración propia

Estos resultados evidencian que la organización, integración y consolidación de los datos comerciales tienen una incidencia en la carga operativa del área comercial. El 33.3% de los encuestados está de acuerdo o totalmente de acuerdo con esta afirmación, el 44.4% adopta una posición neutral y el 22.2% está en desacuerdo o totalmente en desacuerdo. Esto muestra que la forma en la que se gestionan los datos impacta sus actividades diarias. Con esta información se sugiere que hay deficiencia en estructuración e integración de la información y esto requiere de mayor esfuerzo operativo.

7.3.5 Confiabilidad y alineación de KPIs comerciales

Figura 17 *Pregunta 1E encuesta*

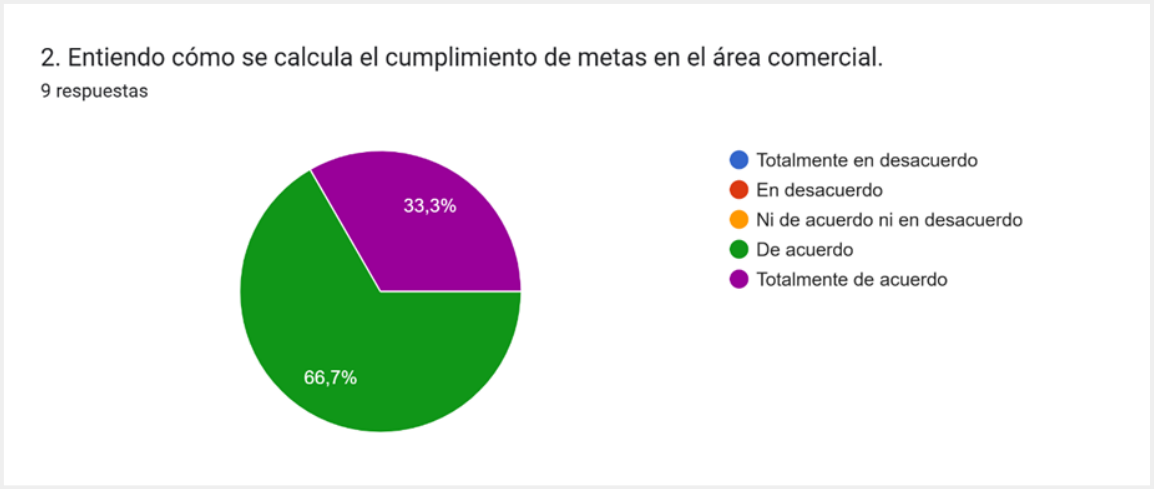


Nota: Elaboración propia

Los resultados muestran favorabilidad sobre la definición y documentación de los indicadores comerciales utilizados en el área comercial. El 33.3% de los encuestados está de acuerdo y el 11.1% totalmente de acuerdo, lo que suma un 44.4% de valoración positiva, por el otro lado el 44.4% adopta una posición neutral y el 11.1% está en desacuerdo. Esto indica que a pesar de que exista una base de indicadores reconocida por los colaboradores, todavía no todos tienen la claridad de medición en sus métricas.

Esto sugiere la necesidad de fortalecer y estandarizar los indicadores comerciales para que todos los comprendan, sea útil en sus seguimientos y se puedan tomar decisiones de manera proactiva.

Figura 18 *Pregunta 2E encuesta*



Nota: Elaboración propia

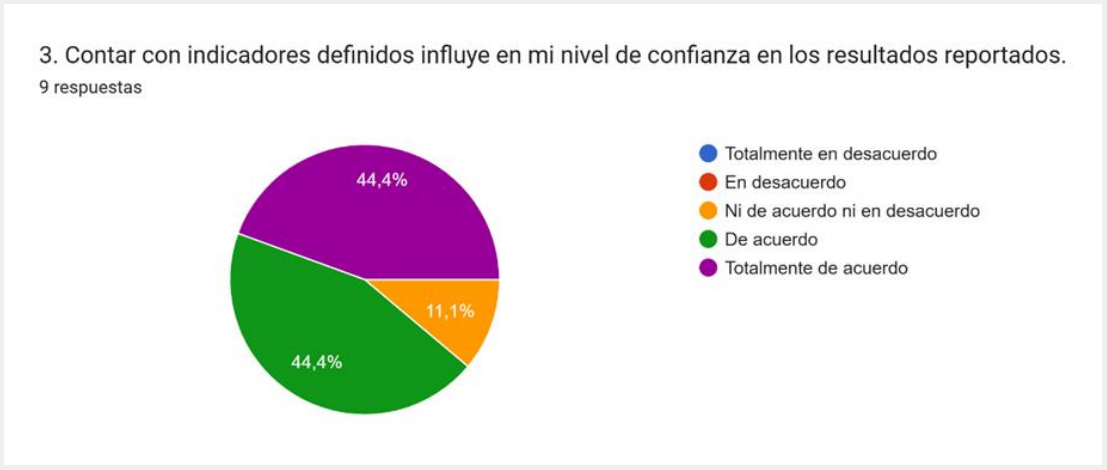
Los resultados muestran contundentemente que la compresión del cálculo del cumplimiento de metas en el área comercial es clara. El 66.7% de los encuestados está de acuerdo y el 33.3% totalmente de acuerdo, esto indica que todos los encuestados son conscientes de la forma en la que se miden los indicadores en el área comercial.

Esto es una fortaleza que tiene la empresa, en la que se puede llegar a profundizar y que también denota un comportamiento de seguimiento en el área que ayuda a la cultura cuando se intervenga con el proyecto de ciencia de datos.

Los resultados evidencian favorabilidad en contar con indicadores definidos en la confianza de los resultados reportados. El 44.4% de los encuestados está de acuerdo y el 44.4% totalmente de acuerdo, mientras que el 11.1% actúa de forma neutral. Para la mayoría la claridad en los indicadores fortalece la credibilidad y seguridad de la información utilizada.

Con este hallazgo se evidencia la importancia de mantener indicadores definidos y estandarizados.

Figura 19 *Pregunta 3E encuesta*



Nota: Elaboración propia

Figura 20 *Pregunta 4E encuesta*



Nota: Elaboración propia

Los resultados muestran claridad de los criterios con los cuales se evalúa el desempeño en el área comercial. El 77.8% de los encuestados está de acuerdo y el 11.1% totalmente de acuerdo, el 11.1% restante actúa de forma neutral. La mayoría de los encuestados reconoce con claridad los criterios de evaluación que son aplicados para la gestión comercial.

Por el lado de la encuesta realizada a los asesores comerciales, se percibe un 33% de respuestas neutrales y un 44% de personas que sí evidencian bases de uso tecnológicas y de uso de información para su operatividad diaria. Estos hallazgos evidencian que los colaboradores no presentan resistencia a la inclusión de herramientas de analítica. De la mano con los resultados obtenidos ya mencionados, se tiene la misma percepción desde el equipo comercial, hay que fortalecer la latencia, la confiabilidad, la automatización y la visualización de la información.

7.4 Síntesis de diagnóstico organizacional

Gracias al diagnóstico organizacional, se evidenció que Valveco S.A.S. cuenta con una base funcional en la gestión de los datos y una disposición organizacional favorable para avanzar en procesos de analítica. Dentro de la propuesta resulta fundamental considerar el factor humano como un elemento determinante para la adopción del modelo que se busca implementar, manteniendo la confianza del equipo en los indicadores actuales, los cuales, aunque son limitados en número, reflejan un nivel adecuado de transparencia en la gestión de información. Se identifican oportunidades de mejora mediante la incorporación de nuevos indicadores que puedan integrarse en los sistemas de visualización, fortaleciendo la toma de decisiones estratégicas.

Contar con información y registros constituye una ventaja organizacional que debe ser aprovechada; sin embargo, para obtener un mayor beneficio es necesario asegurar la integración de las fuentes de datos, con el fin de reducir la dependencia de procesos manuales, fortaleciendo la precisión de la información y disminuir reprocesos y tiempos operativos.

Se identifican brechas estructurales sobre las cuales es necesario intervenir en el marco del proyecto, tales como la latencia de la información, la integración de las plataformas, la

calidad y gobernanza de los datos, la automatización de procesos y el desarrollo de capacidades analíticas.

En términos generales, la organización presenta un nivel de analítica temprano, caracterizado por un enfoque descriptivo y reactivo en el uso de los datos. No obstante, existen oportunidades claras para fortalecer la estandarización de la información, la sincronización de los datos, y la implementación progresiva de prácticas de analítica y visualización que permitan generar mayor valor.

8. Diseño de la arquitectura

Para transformar el modelo operativo de Valveco S.A.S., se propone una arquitectura de datos basada en el paradigma de interoperabilidad y automatización. El objetivo es mitigar la latencia de procesamiento (actualmente de 2 a 4 horas diarias) mediante la integración de fuentes heterogéneas (World Office y DataCRM) en un ecosistema unificado.

8.1. Selección de herramientas para la arquitectura

Para el desarrollo de la arquitectura analítica en Valveco S.A.S., se opta por un componente tecnológico que equilibra la potencia del código abierto con la robustez de las herramientas empresariales líderes. Esta selección busca mitigar los problemas de latencia y fragmentación de datos identificados en el diagnóstico inicial.

La siguiente tabla presenta el análisis de idoneidad de las herramientas seleccionadas frente a las alternativas más comunes del mercado, evaluadas bajo los criterios de costo de implementación, flexibilidad técnica y capacidad de integración.

Tabla 10 *Selección de herramientas para la implementación*

Categoría	**Herramienta Seleccionada**	**Competencia Directa**	**Ventajas estratégicas para Valveco S.A.S.**
Orquestación y Automatización	**n8n** (Auto hosteado)	Zapier / Make	**Soberanía de datos:** Al ser auto-hospedado, los datos sensibles no salen del entorno de la empresa. **Costo:** No tiene cobro por volumen de ejecuciones, permitiendo procesar miles de registros sin sobrecostos.

Extracción y Procesamiento	**Python**	Excel VBA / Alteryx	**Versatilidad:** Permite realizar limpieza de datos a través de diferentes librerías para grandes volúmenes de datos y la automatización de interfaz gráfica (GUI) en portales que no tienen API. **Escalabilidad:** Capacidad de procesar grandes volúmenes de datos con bibliotecas como Pandas y Numpy.
Visualización y Storytelling	**Power BI**	Tableau / QlikView	**Integración:** Compatibilidad nativa con el ecosistema de Microsoft, posibilidad de extraer información de diferentes tipos de fuentes de datos y APIs de terceros. **Curva de aprendizaje:** Facilidad de adopción para los usuarios finales y costos de licenciamiento significativamente menores (Pro vs. Tableau Server).

Nota: Elaboración propia

La elección de estas tres herramientas no es aislada, sino que responde a una sinergia técnica específica para los objetivos de este proyecto:

n8n como núcleo de interoperabilidad, que a diferencia de las herramientas *no-code* tradicionales, permite una manipulación técnica profunda de la API de DataCRM. Su capacidad de ser instalado en un servidor propio (Docker) garantiza que Valveco S.A.S. mantenga el control total sobre su infraestructura de automatización.

Python contribuye al cierre de brechas, en este escenario donde algunos portales externos de gestión comercial no cuentan con APIs abiertas para la integración de datos. Mediante scripts de automatización, esta herramienta permite extraer información de forma desatendida y sistemática, ofreciendo un nivel de flexibilidad e interoperabilidad que otras plataformas comerciales no proporcionan.

La capacidad de Power BI para manejar modelos de datos complejos y su lenguaje DAX facilitan el diseño de Storytelling basado en datos, permitiendo transformar la visualización de información en una narrativa clara y coherente. Esto asegura que la gerencia disponga de una única fuente que explique de manera integrada el desempeño comercial y financiero de la organización.

8.2. Componentes de la solución analítica

8.2.1 Capa de ingesta y orquestación

Para la implementación de flujos de trabajo para la orquestación y automatización de procesos dentro de la arquitectura del proyecto, la herramienta n8n gestionará la lógica de integración entre las APIs de los diferentes aplicativos, mientras que Python se emplea para ejecutar tareas de extracción de datos por interfaz y procesos de limpieza y transformación de datos en aquellos entornos donde no existan interfaces de programación abiertas.

Esta capa es el punto de entrada de la arquitectura y está diseñada para manejar la heterogeneidad de los sistemas principales de la empresa, resumido en la figura 21.

Figura 21 *Estructura de las fuentes de datos*



Nota: Elaboración propia

El software de n8n coordina la ejecución de estos scripts de Python, asegurando que la información de ambas fuentes se sincronice para mitigar la latencia actual (de 2 a 4 horas) y garantizar una base de datos actualizada para el resto de la cadena.

8.2.2 Capa de procesamiento y almacenamiento

En esta fase, los datos capturados se someten a un pipeline de ETL (Extracción, Transformación y Carga) desarrollado en Python. Esta capa es crítica para transformar datos administrativos en activos analíticos, garantizando que la fuente de referencia principal cumpla con los estándares de rigor técnico y validez estadística requeridos.

Para garantizar la interoperabilidad, se aplican protocolos de limpieza diferenciados según la naturaleza y el origen del archivo, tal como se describe en la tabla 11.

Tabla 11 *Procesos de limpieza según la fuente de datos*

Acción / Proceso	Descripción Técnica
Fuente: Comercial (DataCRM)	
Codificación de Negocios	Codificación de estados de venta como variables categóricas.
Normalización de Valor	Limpieza de símbolos ($) y puntos; conversión a tipo float.
Tratamiento de Fechas	Conversión de Est_Cierre de formato String a Datetime.
Normalizar NIT	Eliminación de guiones y del dígito de verificación.
Fuente: Comercial (World Office)	
Estandarización de ID	Remoción de puntos, guiones y dígito de verificación en CC/NIT.
Limpieza de Cliente	Aplicación de Trim y eliminación de sufijos legales (S.A.S., S.A., etc.).
Separación de Documento	Uso de funciones de string para eliminar redundancias (ej. "DMC DMC136").
Tratamiento de Fechas	Conversión de campo de fecha de tipo String a Datetime.
Depuración de Columnas	Eliminación de la columna "Otros Imp." por ausencia de datos (valores nulos).

Fuente: Inventario (World Office)	
Codificación de Referencias	Aplicación de codificación categórica a las descripciones existentes.
Extracción de Código	Aislamiento del identificador único para cada referencia del sistema.
Orden Cronológico	Organización temporal de las referencias y sus niveles de existencias.
Conversión de Unidades	Unificación de medidas (ej. Pulgadas a mm) según el catálogo.
Tratamiento de Fechas	Conversión de campo de fecha de tipo String a Datetime.

Fuente: Elaboración propia

La consolidación de silos e integridad de los datos se realiza una vez los datos están limpios, y se integran mediante operaciones de unión (Joins) lógicos: vínculo Comercial-Operativo donde se cruzan los negocios del CRM con los registros de World Office para validar las actividades comerciales que fueron realmente consolidadas. Y cálculo de KPIs descriptivos, donde se generan métricas que permiten validar el desempeño de las actividades comerciales realizadas y el consumo de inventario.

La salida de esta capa es un conjunto de datos estructurados y "limpio" almacenado en un repositorio intermedio, en este caso se usa google sheets ya que la empresa cuenta con acceso a las herramientas de google y con conexiones directas desde esta plataforma a PowerBi que se utiliza para la visualización de datos. Este repositorio actúa como el insumo

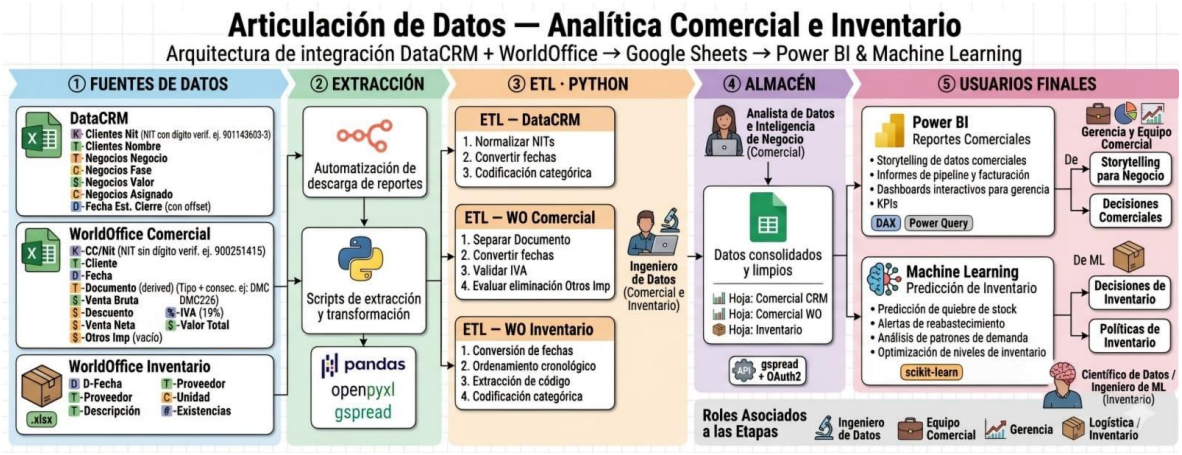
base para las capas superiores, garantizando que el científico de datos y el analista de BI trabajen sobre una base de datos sin errores de forma, duplicidad o latencia.

Es necesario aclarar el alcance de este componente. En términos conceptuales, cumple la función de un repositorio analítico centralizado, ya que reúne en un solo lugar la información depurada de distintas fuentes y la pone a disposición de las capas superiores. Sin embargo, su implementación no corresponde a un *Data Warehouse* en sentido estricto, sino a una hoja de cálculo colaborativa (Google Sheets). Esta elección responde a la necesidad de contar con una solución inicial práctica y de bajo costo (*producto mínimo viable*), alineada con el nivel de madurez analítica temprana identificado en Valveco S.A.S. y con el uso previo de herramientas de Google dentro de la organización. De esta manera, se evita realizar inversiones en infraestructura que aún no serían justificables en esta primera etapa.

Ahora bien, es importante reconocer que este enfoque tiene limitaciones en cuanto a escalabilidad. Google Sheets restringe el volumen de registros, el rendimiento de las consultas y la concurrencia de accesos, además de que el pipeline actual reprocesa todos los datos en cada ejecución. Por ello, a medida que el volumen de información aumente, se recomienda migrar hacia una solución más robusta (como una base de datos relacional o un *Data Warehouse* en la nube) e incorporar mecanismos de procesamiento incremental basados en deltas y llaves compuestas, que eviten el reprocesamiento completo en cada corrida. Esta ruta de evolución coincide con la retroalimentación recibida por la organización durante la socialización de la propuesta (véase la sección 10.1)

La imagen que se presenta a continuación en la figura 22 presenta una versión más detallada de la arquitectura planteada en la figura 3, pero orientada a las capas mencionadas en esta sección.

Figura 22 *Diseño de arquitectura analítica detallado*



Nota: Elaboración propia

8.2.3 Capa de Inteligencia Predictiva

Esta capa implementa modelos de analítica predictiva que analizan el comportamiento histórico de los inventarios. Utiliza algoritmos de aprendizaje supervisado para detectar patrones de consumo y generar alertas tempranas sobre posibles quiebres de stock con 60 días de antelación. Este componente transforma el dato histórico en una herramienta de prevención, permitiendo una gestión de abastecimiento basada en la probabilidad y no solo en la intuición.

El modelo aborda un problema de clasificación binaria para responder si un producto específico llegará a cero existencias en un plazo de 60 días. La elección de un horizonte de 60 días responde directamente a los tiempos de reposición del portafolio de Valveco S.A.S. Una parte significativa de las referencias corresponde a productos importados de proveedores internacionales. El ciclo completo de reposición, que incluye la colocación del pedido, el tránsito internacional y el proceso de nacionalización, suele superar los 30 días, plazo que resultaría insuficiente para reaccionar ante una alerta de quiebre.

Un horizonte de 60 días otorga al equipo de compras el margen operativo necesario para gestionar la reposición antes de que la existencia llegue a cero, evitando así pérdidas

comerciales derivadas del desabastecimiento. Esta decisión se encuentra alineada con las debilidades y amenazas identificadas en la matriz DOFA, particularmente la dependencia de los tiempos de proveedores y logística, así como las variaciones de la TRM que inciden en los procesos de importación.

8.2.3.1 Ingeniería de características

Se transformaron los registros diarios de inventario (más de 7,600 registros) en un conjunto de 23 variables predictivas agrupadas en:

1. Estado actual: Niveles de stock y razón de capacidad máxima.
2. Tendencia: Velocidad de consumo diario en ventanas de 7 y 14 días, junto con la pendiente de regresión lineal para identificar aceleración en la demanda.
3. Variabilidad: Desviaciones estándar y coeficientes de variación para capturar la irregularidad del consumo.
4. Contexto: Codificación de proveedores y categorías mediante Label Encoding.

8.2.3.2 Selección y configuración del algoritmo

Tras un análisis comparativo, se selecciona Clasificación de bosques aleatorios (Random Forest Classifier, en inglés) por su robustez frente a datos no lineales y su resistencia al sobreajuste en comparación con modelos como Gradient Boosting. La configuración del modelo incluye:

1. Ensemble: 200 árboles de decisión para estabilizar las predicciones.
2. Control de complejidad: Profundidad máxima de 12 niveles para asegurar la generalización del modelo.
3. Balanceo de clases: Uso de class_weight='balanced' para compensar la distribución de eventos de quiebre en el historial.

8.2.3.3 Estrategia de validación y métricas

Dada la naturaleza secuencial de los datos, se aplica una división temporal (entrenamiento con datos históricos y evaluación con datos futuros) para evitar la fuga de información. Para el modelo se prioriza la métrica de Recall (Sensibilidad, definida en la Ecuación 2), asumiendo que para la operación de Valveco S.A.S. es más costoso un "falso negativo" (un quiebre no detectado que detiene la operación) que una "falsa alarma" (un pedido preventivo).

$$Recall = \frac{TP}{TP + FN}, \qquad (2)$$

$$\begin{cases} TP\ (Verdaderos\ Positivos): Casos\ positivos\ que\ el\ modelo\ predijo\ correctamente\ como\ positivos. \\ FN\ (Falsos\ Negativos): Casos\ positivos\ que\ el\ modelo\ predijo\ erróneamente\ como\ negativos. \end{cases}$$

Los resultados obtenidos sobre el conjunto de prueba confirman el desempeño del modelo bajo la métrica priorizada. La Tabla 12 presenta las métricas de evaluación y la Tabla 13 la matriz de confusión correspondiente.

Tabla 12 *Métricas de evaluación del modelo sobre el conjunto de prueba*

Métrica	Valor
Exactitud (Accuracy)	82,8 %
Precisión (Precision)	72,4 %
Sensibilidad (Recall)	99,5 %
F1-Score	83,8 %
AUC-ROC	91,7 %

Nota: Elaboración propia

Tabla 13 *Matriz de confusión del modelo sobre el conjunto de prueba*

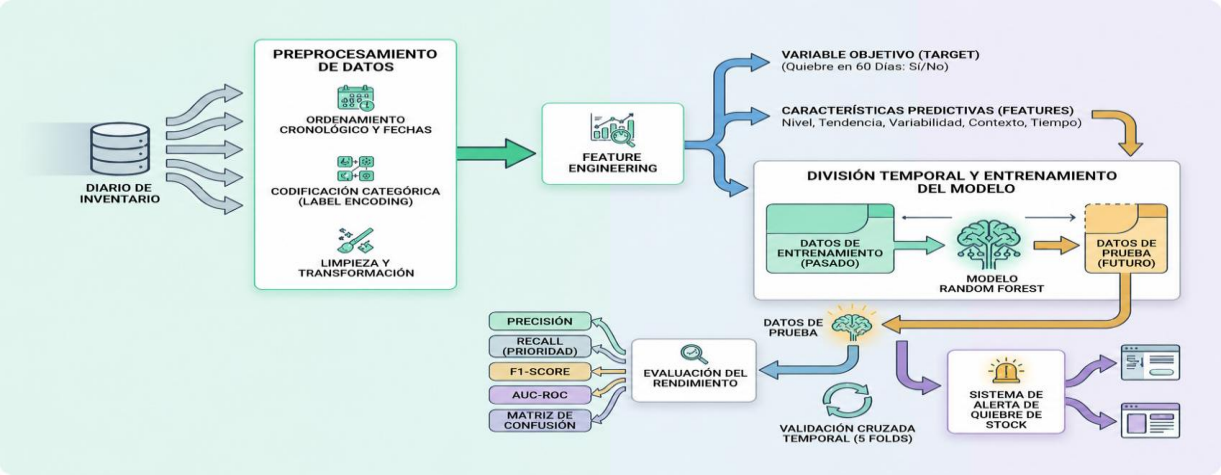
	Predicho: No-Quiebre	**Predicho: Quiebre**
Real: No-Quiebre	365 *(VN)*	162 *(FP)*
Real: Quiebre	2 *(FN)*	424 *(VP)*

Nota: Elaboración propia. VN = verdaderos negativos; FP = falsos positivos; FN = falsos negativos; VP = verdaderos positivos.

El modelo alcanza una sensibilidad (recall) del 99,5 %, detectando 424 de los 426 quiebres reales del conjunto de prueba y dejando escapar únicamente dos falsos negativos, lo cual es coherente con la priorización de esta métrica frente al costo operativo de un quiebre no detectado. El costo de este umbral conservador se refleja en 162 falsas alarmas (falsos positivos), equivalentes a pedidos preventivos que resultan asumibles para la operación de Valveco S.A.S. dada la asimetría de costos descrita. El AUC-ROC de 91,7 % evidencia una adecuada capacidad discriminativa entre referencias en riesgo y seguras. Para verificar la estabilidad temporal del modelo, se aplicó una validación cruzada con TimeSeriesSplit de cinco particiones, obteniendo un F1-Score promedio de 0,915 (± 0,089) y un AUC-ROC promedio de 0,970 (± 0,033), lo que confirma que el desempeño no depende de un periodo particular del histórico. Cabe precisar que la prevalencia de eventos de quiebre en los datos históricos es elevada, dado que un porcentaje considerable de las referencias permanece en niveles de existencia cero durante periodos prolongados; este comportamiento respalda el uso de la ponderación de clases y contextualiza los valores altos de sensibilidad obtenidos.

La figura 23 muestra de manera gráfica el diseño de la solución que se plantea para el modelo de machine learning considerando los ítems anteriores.

Figura 23 *Arquitectura del modelo de predicción sobre quiebre de inventario*



Nota: Elaboración propia

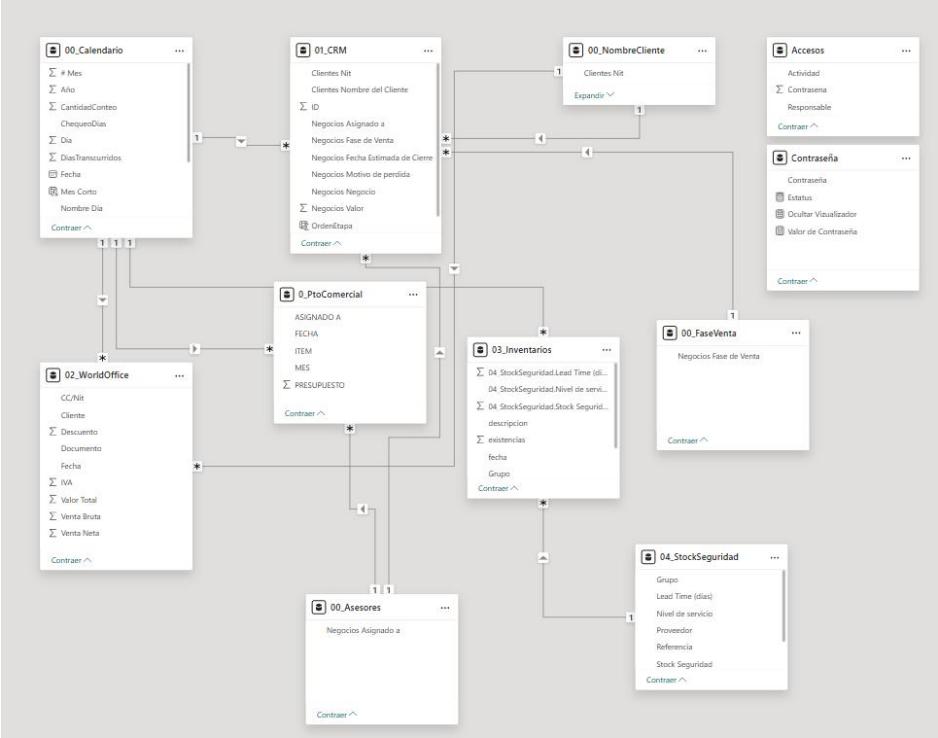
8.2.4 Capa de consumo y storytelling

Aplicación de técnicas de storytelling con datos. No se busca solo la visualización, sino la construcción de una narrativa que guíe al usuario desde el "qué sucedió" (análisis descriptivo) hasta el "por qué sucedió" (análisis diagnóstico).

8.2.4.1 Modelos de datos semánticos en Power BI

El modelo semántico se diseña con un esquema tipo estrella, se realizan tablas de dimensiones claves que son: cliente, fecha, asesor y producto, para poder integrar las tablas de hechos obtenidas de las bases de datos de DataCRM y el World Office, asegurando consistencia, trazabilidad y eficiencia en el análisis.

Figura 24 *Modelo de datos en esquema tipo estrella*



Desde: tabla (columna)	Relación	A: tabla (columna)	Estado
01_CRM (Clientes Nit)	* — 1	00_NombreCliente (Clientes Nit)	Activo
01_CRM (Negocios Asignado a)	* — 1	00_Asesores (Negocios Asigna...	Activo
01_CRM (Negocios Fase de Ve...	* — 1	00_FaseVenta (Negocios Fase ...	Activo
01_CRM (Negocios Fecha Esti...	* — 1	00_Calendario (Fecha)	Activo
02_WorldOffice (CC/Nit)	* — 1	00_NombreCliente (Clientes Nit)	Activo
02_WorldOffice (Fecha)	* — 1	00_Calendario (Fecha)	Activo
03_Inventarios (descripcion)	* — 1	04_StockSeguridad (Referencia)	Activo
03_Inventarios (fecha)	* — 1	00_Calendario (Fecha)	Activo
0_PtoComercial (ASIGNADO A)	* — 1	00_Asesores (Negocios Asigna...	Activo
0_PtoComercial (FECHA)	* — 1	00_Calendario (Fecha)	Activo

Nota: Elaboración propia

La integración entre las bases de datos no se realiza de forma directa entre tablas, sino mediante dimensiones comunes. En especial, la dimensión de cliente se construye por el NIT normalizado, que permite vincular las oportunidades comerciales con lo que se factura. Asimismo, la dimensión fecha permite analizar el movimiento entre los dos sistemas por períodos de tiempo.

Con este enfoque se logran comparar los registros de ambos sistemas para resolver la falta de coherencia entre los sistemas de información que tiene Valveco S.A.S.

8.2.4.2 Definición de KPIS comerciales

A partir de lo que actualmente realiza la empresa en sus dinámicas de seguimiento comercial y la unificación de los sistemas que se realiza, se definen los siguientes indicadores clave:

Cumplimiento de cuota: Indicador porcentual que mide el nivel de ventas que se han facturado respecto a la meta comercial establecida. Se calcula dividiendo las ventas registradas en el World Office entre la cuota asignada en el período.

$$Cumplimiento\ cuota = \frac{Total\ ventas\ realizadas}{Cuota\ de\ ventas} * 100 \qquad (3)$$

Variación de ventas: Indicador porcentual que muestra el incremento o la disminución de las ventas en el tiempo. Se comparan las ventas actuales con las del período anterior.

$$Variación\ de\ ventas = \frac{ventas\ actuales - ventas\ del\ periodo\ anterior}{ventas\ del\ periodo\ anterior} * 100 \qquad (4)$$

Tasa de conversión: Indicador porcentual que mide la eficiencia comercial, relacionando las ventas facturadas con las oportunidades generadas en el CRM. Esto permite evaluar la capacidad de cierre del equipo o se puede identificar por qué razón se están perdiendo las negociaciones.

$$Tasa\ de\ conversión = \frac{ventas\ facturadas}{total\ oportunidades\ en\ el\ CRM} * 100 \quad (5)$$

Valor del pipeline: Indicador absoluto que representa el valor total de las oportunidades abiertas en el CRM. Permite estimar el ingreso potencial en proceso de negociación.

$$Valor\ del\ pipeline = \sum_{n=0}^{n} Valor\ de\ cada\ oportunidad \quad (6)$$

Tasa de pérdida de oportunidades: Indicador porcentual que mide la proporción de oportunidades comerciales que no se logró concretar en la venta. Esto sirve para realizar intervenciones o recuperación de clientes.

$$Tasa\ de\ pérdida = \frac{Oportunidades\ perdidas}{Total\ de\ oportunidades} * 100 \quad (7)$$

Valor del cliente en el tiempo: Indicador absoluto que mide el valor total generado por cada cliente desde el momento en el que se inició la relación Valveco S.A.S. - cliente. Esto sirve para priorizar o fidelizar cuentas claves.

$$CLV = valor\ medio\ compra * frecuencia\ compra * duracion\ de\ la\ relación \quad (8)$$

Ticket promedio Indicador absoluto que mide el promedio de las ventas realizadas. Con este KPI se puede identificar el comportamiento de compra de los clientes y el promedio de venta por asesor.

$$Ticket\ promedio = \frac{Ventas\ totales}{Número\ de\ transacciones} * 100 \quad (9)$$

8.2.4.3 Capa de visualización estratégica y toma de decisiones

La capa de visualización estratégica se implementa en Power BI mediante un tablero estructurado que guía usuario a través de una secuencia lógica de análisis, desde el acceso al sistema hasta la identificación de riesgos y oportunidades para la toma de decisiones. Cada página cumple una función específica dentro del proceso analítico, permitiendo comprender el comportamiento del área comercial.

La primera página corresponde a la validación de credenciales, donde se controla el acceso a la información mediante autenticación de usuarios. Esta funcionalidad garantiza que cada perfil visualice únicamente la información autorizada, fortaleciendo la seguridad y confidencialidad de los datos comerciales.

La segunda página presenta el menú principal. Desde esta vista el usuario puede acceder a cada uno de los módulos analíticos: Resumen ejecutivo, embudo comercial, análisis por asesor e inventario y alertas, organizando la consulta de la información de forma intuitiva.

En la tercera página, resumen ejecutivo, se muestra una visión general del estado comercial de la organización. En esta página se consolidan indicadores estratégicos como el valor del pipeline, las ventas realizadas, las ventas perdidas, el número de negocios y el inventario en riesgo. Adicionalmente, se presenta el comportamiento de las ventas ganadas

y perdidas por asesor, proporcionando un panorama inicial del desempeño del negocio durante el período seleccionado.

La cuarta página corresponde al embudo comercial, donde se analiza el comportamiento del proceso de ventas a lo largo de sus etapas. Esta vista incorpora indicadores como la tasa de conversión, la tasa de pérdida, el ticket promedio y el número de negocios, complementados con el embudo de oportunidades y la comparación de la conversión por asesor. Asimismo, se destacan alertas sobre oportunidades que permanecen un tiempo prolongado en negociación o propuesta, facilitando la identificación de posibles retrasos en el proceso comercial.

La quinta página desarrolla el análisis por asesor, se orienta a evaluar el desempeño individual del equipo comercial. Se incorporan indicadores relacionados con el avance proyectado, ventas alcanzadas, porcentaje de cumplimiento de la meta, brecha comercial, ticket promedio y número de negocios. De igual forma, se consideran las variables de contexto como los días hábiles, los días transcurridos y el porcentaje de avance del período, permitiendo interpretar los resultados en función del tiempo efectivo disponible para el cumplimiento de los objetivos comerciales.

Finalmente, la sexta página presenta el módulo de inventarios y alertas, cuyo propósito es apoyar las decisiones de abastecimiento mediante la identificación de productos que requieren atención. En esta vista se monitorea el estado del inventario, identificar referencias críticas y evaluar el nivel de riesgo asociado a posibles quiebres de inventario. Como parte de la propuesta desarrollada, este módulo está diseñado para consumir las predicciones generadas por le modelo de aprendizaje automático, permitiendo recomendar oportunamente los productos que deben solicitarse considerando el lead time de abastecimiento y reduciendo así el riesgo de perder oportunidades comerciales por falta de disponibilidad.

En conjunto, esta secuencia de visualización establece una narrativa analítica que conduce al usuario desde una visión general del negocio hasta análisis específicos sobre el desempeño comercial y los riesgos operacionales. De esta manera, el tablero facilita responder preguntas como ¿cuál es el estado del negocio?, ¿cómo se comporta el proceso comercial?, ¿quién está generando los resultados? y ¿qué acciones deben tomarse para mejorar el desempeño comercial y minimizar los riesgos asociados al inventario?, apoyando la toma de decisiones estratégicas basada en datos.

8.2.4.4 Análisis de audiencia y mapa de decisiones estratégicas

La eficacia en la toma de decisiones estratégicas depende de la alineación entre el nivel de granularidad de los datos y el rol jerárquico del interlocutor. Bajo el principio de storytelling estratégico, cada interfaz de visualización en la arquitectura propuesta ha sido diseñada para una audiencia específica, respondiendo a interrogantes críticos predefinidos. La tabla 12 formaliza el despliegue de los tableros en Power BI, vinculando cada vista con su perfil de usuario, las dimensiones de análisis habilitadas y su contribución directa a los objetivos tácticos y estratégicos de Valveco S.A.S.

Tabla 14 *Análisis de audiencia y decisiones estratégicas*

Vista en Power BI	Preguntas que responde	Decisión que habilita	Quien la toma	Objetivo estratégico asociado	Frecuencia de uso
Embudo de ventas interactivo	¿En qué fase se pierden más oportunidades? ¿Qué vendedor tiene mayor tasa de conversión?	Reasignar oportunidades estancadas / Reforzar capacitación en la fase de mayor abandono	Gerente Comercial	Aumentar la tasa de cierre global de la empresa	Semanal

Tablero de cumplimiento de cuota	¿Cuánto lleva cada vendedor vs. su meta? ¿Quién puede alcanzar la cuota antes del cierre del mes?	Aplicar incentivos, redistribuir cargas comerciales o ajustar la cuota del próximo periodo	Gerente Comercial + Gerencia General	Cumplimiento del presupuesto de ventas	Semanal / Mensual
Alerta de quiebre de inventario + cruce CRM	¿Qué referencias están en riesgo de quiebre en 60 días?	Priorizar orden de compra de reposición	Gerente Comercial + Gerencia	Eficiencia operativa	Tiempo real / Diario
Pipeline de oportunidades estancadas	¿Cuántos días llevan sin actividad estas oportunidades ? ¿Cuánto vale el pipeline paralizado?	Acción comercial directa: llamada de seguimiento, escalada o descarte de oportunidad	Gerente Comercial	Reducir el ciclo de venta promedio	Semanal
Análisis de ticket promedio y ciclo de venta por vendedor	¿Quién vende más rápido? ¿Quién vende negocios de mayor valor?	Identificar buenas prácticas del vendedor con mejor desempeño y replicarlas / Asignar cuentas estratégicas al vendedor con mejor perfil de cierre	Gerente Comercial	Optimizar el desempeño del equipo comercial	Mensual

Nota: Elaboración propia

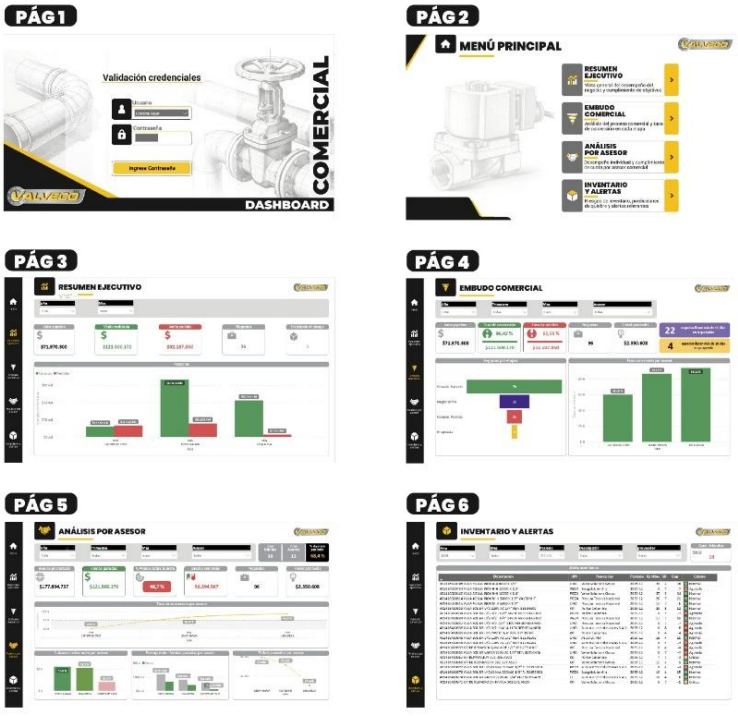
8.2.4.5 Ciclo de decisión y relación con el diagnóstico

El flujo inicia con la integración de datos desde DataCRM y World Office, continúa con su transformación y análisis predictivo, y finaliza con la visualización en Power BI. En la figura 25 se puede observar la secuencia visual del tablero con mayor detalle.

A partir del tablero, los usuarios identifican desviaciones, riesgos y oportunidades, logrando tomar decisiones como priorizar ventas, analizar comportamiento de ventas o planificar inventarios. Este ciclo responde a los problemas identificados en el diagnóstico, como la falta de integración, reprocesos y baja capacidad de anticipación. Esto hace que se tomen decisiones oportunas y basadas en datos.

8.2.4.6 Secuencia visual del tablero

Figura 25 *Secuencia visual del tablero en Power BI*



Nota: Elaboración propia

La Figura 25 presenta la secuencia visual diseñada para el tablero de control, cuya finalidad es crearle un camino al usuario a través de un proceso de análisis progresivo. El tablero se organiza para seguir una narrativa que facilita la compresión del comportamiento comercial y orienta la toma de decisiones estratégicas.

La secuencia inicia con la autenticación del usuario y el acceso al menú principal, desde donde se selecciona el módulo de análisis requerido. Posteriormente, el recorrido comienza con una visión general del negocio mediante el resumen ejecutivo, permitiendo identificar rápidamente el estado de los principales indicadores comerciales. A continuación, el análisis profundiza en el proceso de ventas a través del embudo comercial, donde es posible comprender cómo evolucionan las oportunidades comerciales y en qué etapas se presentan las principales pérdidas o retrasos.

Una vez identificado el comportamiento general del proceso comercial, la secuencia dirige el análisis hacia el desempeño del equipo de ventas mediante la vista de análisis por asesor, relacionando los resultados obtenidos con el cumplimiento de las metas establecidas y el avance del período evaluado. Finalmente, el recorrido concluye con la vista de inventario y alertas, la cual incorpora información proveniente de analítica avanzada para anticipar riesgos de abastecimiento que puedan afectar futuras oportunidades comerciales.

Esta organización responde a una estructura de análisis de lo general a lo particular, donde cada vista aporta el contexto necesario para interpretar la siguiente. De esta manera, el usuario transita desde una compresión global del estado del negocio hasta la identificación de causas específicas y riesgos potenciales, favoreciendo un proceso de análisis coherente y una toma de decisiones fundamentada en la información presentada.

8.3. Marco de gobernanza de datos

La gobernanza de datos es el conjunto de políticas, roles y estándares que garantizan que los datos de la organización sean confiables, seguros, comprensibles y utilizados de manera consistente. Sin gobernanza, la misma métrica puede ser interpretada de formas distintas por diferentes personas, generando decisiones contradictorias. Esta sección define el marco de gobernanza aplicado a la arquitectura de Valveco S.A.S.

8.3.1 Diccionario de datos

El diccionario de datos que se detalla en la tabla 12 actúa como el estándar técnico de verdad para el modelo analítico, normalizando la estructura y el origen de cada atributo. A través de este instrumento, se documenta la propiedad del dato, los dominios permitidos y las lógicas de cálculo subyacentes, eliminando discrepancias en la interpretación de los KPIs. Esta formalización asegura que métricas críticas, tales como 'Venta Neta' o 'Margen Bruto', operen bajo una definición técnica uniforme y auditable en todas las unidades de Valveco S.A.S.

Tabla 15 *Diccionario de datos*

Fuente de Datos	Nombre de la Columna	Tipo de Dato	Descripción
World Office_Inventario	fecha	Texto	Fecha del registro de inventario
World Office_Inventario	proveedor	Texto	Nombre del proveedor del ítem
World Office_Inventario	descripción	Texto	Descripción técnica del producto
World Office_Inventario	unidad de medida	Texto	Unidad de medida del ítem (p. ej., UND)
World Office_Inventario	existencias	Entero	Cantidad disponible en inventario

World Office_Comercial	CC/Nit	Texto	Número de identificación tributaria del cliente
World Office_Comercial	Cliente	Texto	Nombre o razón social del cliente
World Office_Comercial	Fecha	Fecha	Fecha de registro de la transacción
World Office_Comercial	Documento	Texto	Referencia del documento de venta
World Office_Comercial	Venta Bruta	Numérico	Valor bruto de la venta en pesos
World Office_Comercial	Descuento	Numérico	Monto del descuento aplicado
World Office_Comercial	Venta Neta	Numérico	Valor de venta tras aplicar descuentos
World Office_Comercial	Otros Imp	Numérico	Otros impuestos distintos al IVA
World Office_Comercial	IVA	Numérico	Impuesto al valor agregado
World Office_Comercial	Valor Total	Numérico	Valor total de la transacción
DataCRM_Comercial	Clientes Nit	Texto	NIT o identificación del cliente
DataCRM_Comercial	Clientes Nombre del Cliente	Texto	Nombre o razón social del cliente
DataCRM_Comercial	Negocios Negocio	Texto	Categoría o tipo de negocio registrado
DataCRM_Comercial	Negocios Fase de Venta	Texto	Etapa actual del proceso de venta (Ej: Cerrada-ganada, Cerrada-perdida)
DataCRM_Comercial	Negocios Valor	Numérico	Valor estimado del negocio en pesos
DataCRM_Comercial	Negocios Asignado a	Texto	Asesor comercial responsable del negocio
DataCRM_Comercial	Negocios Fecha Estimada de Cierre	Texto	Fecha proyectada de cierre del negocio

Nota: Elaboración propia.

8.3.2 Política de calidad de datos

Se establecen umbrales cuantitativos de calidad para los campos más críticos del modelo. El pipeline de n8n evalúa estos umbrales en cada ejecución antes de continuar hacia la capa de transformación. Si se supera el umbral de parada, el proceso se detiene automáticamente y se genera una alerta al analista de datos, los detalles se especifican en la tabla 14.

Tabla 16 *Política de calidad de datos*

Campo	Umbral de alerta (warning)	Umbral de parada (stop)	Acción automática del pipeline
NIT_Cliente nulo	>2% del total de registros	>5% del total de registros	Pipeline detenido + alerta por correo al analista con detalle de registros afectados
Fase_Oportunidad con valor no controlado	>1% de registros	>3% de registros	Registros excluidos de KPIs + log de excepción en Google Sheets
Stock_Disponible negativo	Cualquier registro	---	Registro marcado como "revisar" en columna de estado + alerta
Fecha_Creacion futura (error de registro)	Cualquier registro	---	Registro excluido del cálculo + log de excepción
Venta_Bruta = 0 en oportunidad Cerrado-Ganado	>0.5% de registros	>2% de registros	Alerta: posible error de registro. Pipeline continua, pero marca los registros.
Duplicados por NIT + Fecha + Referencia	>0.1% del total	>1% del total	De duplicación automática (conserva registro más reciente) + log del proceso

Nota: Elaboración propia.

8.3.3 Roles de acceso y permisos diferenciados (Row Level Security)

No todos los usuarios deben tener acceso a todos los datos. La arquitectura implementa Row Level Security (RLS) en Power BI para filtrar automáticamente los datos según el usuario autenticado, garantizando confidencialidad sin requerir tableros separados, los detalles de los diferentes roles de acceso se definen en la tabla 15.

Tabla 17 *Roles de accesos y permisos*

Rol	Usuario típico	Que ve en Power BI	Que NO ve	Nivel de acceso
Gerente General	Gerencia	Todos los KPIs, todos los vendedores, márgenes totales, alertas ML, análisis financiero consolidado	---	Acceso completo sin filtros
Gerente Comercial	Dirección comercial	KPIs por vendedor, embudo de ventas, alertas de quiebre, ciclo de venta, oportunidades estancadas	Márgenes individuales por producto (solo accede a totales)	Acceso filtrado por rol comercial
Vendedor Individual	Cada vendedor del equipo	Solo sus propias oportunidades, su cuota personal, su ciclo de venta y su tasa de conversión	Datos de otros vendedores, márgenes, pipeline completo	Acceso filtrado por identidad (RLS por usuario en Power BI)
Finanzas	Área financiera	Ventas netas, márgenes consolidados, cumplimiento de cuota agregado, análisis de rentabilidad por referencia	Detalle individual de oportunidades en el CRM, datos de contacto de clientes	Acceso a capa financiera

Analista de datos	Responsable del ecosistema	Todas las capas incluyendo datos crudos, logs del pipeline, tablero de calidad de datos	---	Acceso administrativo completo

Nota: Elaboración propia. La implementación técnica se realiza mediante Row Level Security (RLS) nativo de Power BI, configurado por el analista de datos.

8.3.4 Control de gobernanza de datos para la arquitectura

Dentro de la arquitectura propuesta, Google Sheets se posiciona como el repositorio de persistencia temporal (Staging Area). Si bien esta elección responde a la necesidad de interoperabilidad nativa con el ecosistema de Power BI y a la optimización de la curva de adopción tecnológica por parte de los usuarios, su naturaleza de edición concurrente introduce riesgos de consistencia. En consecuencia, se han diseñado mecanismos de gobernanza de datos para salvaguardar la integridad y trazabilidad de la información, los cuales se describen en la tabla 16.

Tabla 18 *Control de gobernanza de datos de la arquitectura*

Control	Implementación técnica	Responsable
Control de versiones	Activar historial de versiones nativo de Google Sheets (retención de 100 versiones). Complementado con backup semanal automatizado a Google Drive en formato CSV con timestamp en el nombre del archivo.	n8n (automatizado)
Protección de hojas de datos crudos	Las hojas alimentadas por el pipeline de n8n se configuran en modo Solo lectura para todos los usuarios, excepto la cuenta de servicio del pipeline. Los usuarios finales acceden únicamente a través de Power BI.	Analista de datos

Política de tránsito	Google Sheets actúa como capa de staging, no como sistema de registro definitivo. El archivo histórico se consolida semanalmente en Google Drive como backup inmutable con acceso restringido.	Analista de datos
Validación de integridad post-escritura	Tras cada carga del pipeline, un nodo de n8n ejecuta una consulta de validación que compara el número de registros cargados vs. los registros en la fuente. Si hay discrepancia superior al 1%, se genera alerta automática al analista.	n8n
Auditoría de ediciones manuales	Semanalmente, un script de Python verifica si existen modificaciones no autorizadas en celdas controladas por el pipeline. Si detecta cambios, genera alerta al analista con el detalle de las celdas modificadas.	Script Python (cron semanal)

Nota: Elaboración propia

8.4. Arquitectura analítica orientada a la resolución de interrogantes estratégicos

El diseño de una arquitectura analítica de orientación estratégica trasciende la mera disponibilidad de los activos de información, fundamentándose en los interrogantes críticos que la organización busca resolver. Bajo esta premisa, la estructura propuesta establece una trazabilidad directa entre cada capa técnica y las dimensiones de negocio que habilita. A continuación, la tabla 17 detalla la vinculación de los componentes de la arquitectura con las necesidades operativas y estratégicas de Valveco S.A.S., transformando el flujo técnico en un motor de valor comercial tangible.

Tabla 19 *. Resolución de interrogantes estratégicos basado en la arquitectura*

Capa analítica	Pregunta de negocio que responde	Dato fuente	Herramienta
Descriptiva	¿En qué cantidad vendió cada vendedor este mes vs. el mes anterior?	DataCRM	Power BI
Descriptiva	¿Cuál es la distribución del pipeline comercial por fase y por valor total?	DataCRM	Power BI
Descriptiva	¿Qué referencias de inventario tienen mayor rotación este trimestre?	World Office	Power BI
Descriptiva	¿Cuál es el ticket promedio por vendedor y por cliente?	DataCRM + World Office	Power BI
Diagnóstica	¿En qué fase del embudo de ventas se pierden más oportunidades y cuál es el patrón de abandono?	DataCRM	Power BI
Diagnóstica	¿Por qué el ciclo de venta de un vendedor es significativamente más largo que el promedio del equipo?	DataCRM	Power BI
Diagnóstica	¿Qué clientes o referencias concentran la mayor parte del margen bruto?	DataCRM + World Office	Power BI
Predictiva	¿Qué referencias tienen riesgo de quiebre de stock en los próximos 60 días dado el ritmo actual de consumo?	World Office + histórico	Modelo ML (Python)
Prescriptiva (fase futura)	¿A qué vendedor debería asignarse esta oportunidad estratégica según su historial de cierre con clientes similares?	DataCRM	Power BI + análisis

Nota: Elaboración propia

8.5 Plan de criterios de éxito y pruebas de aceptación

Los criterios de éxito y el plan de pruebas se definen antes del inicio de la implementación. Estos criterios constituyen la condición de cierre del proyecto: la arquitectura se considera implementada exitosamente únicamente cuando todos los criterios sean verificados y aprobados. Los criterios definidos para el proyecto se muestran en la tabla 18.

Tabla 20 *Criterios de éxito y pruebas de aceptación*

Criterio	Indicador	Meta	Forma de medición	Fecha de evaluación
CE-01 Latencia del pipeline	Tiempo extremo a extremo: desde extracción hasta disponibilidad en Power BI	Menor a 15 minutos	Log de n8n: timestamp de inicio vs. timestamp de carga en Sheets	Al finalizar fase de implementación (mes 4)
CE-02 Calidad predictiva del modelo ML	Recall y Precisión en conjunto de prueba con datos de evaluación temporal	Recall >= 85% y Precisión >= 70%	Reporte de evaluación del modelo (scikit-learn classification_report)	Antes del despliegue en producción
CE-03 Adopción del tablero Power BI	Frecuencia de acceso de los 3 usuarios durante las primeras 4 semanas de operación	Mínimo 3 accesos por usuario por semana durante 4 semanas consecutivas	Registro de acceso en Power BI Service (Usage Metrics)	Semanas 1 a 4 post-lanzamiento
CE-04 Reducción de trabajo manual	Reducción del tiempo dedicado a consolidación manual de reportes comerciales	Reducción >= 70% del tiempo semanal vs. línea base pre-implementación	Encuesta de tiempo pre vs. post implementación con los 3 usuarios (estimación en horas/semana)	Mes 2 post-lanzamiento

CE-05 Disponibilidad del pipeline	Tasa de ejecuciones exitosas del pipeline en ventana de 30 días	Mayor o igual al 95% de ejecuciones exitosas	Dashboard de monitoreo de n8n (executions history)	Mensual desde mes 1 de operación
CE-06 Integridad de datos	Porcentaje de registros con campos críticos nulos (NIT, Fecha, Referencia)	Menor al 5% de registros con campos críticos nulos	Script de auditoría en n8n ejecutado en cada ciclo	Semanal desde inicio de operación
CE-07 Satisfacción de usuarios	Puntuación de usabilidad percibida del tablero por los usuarios finales	Puntuación promedio >= 4 sobre 5 en encuesta SUS simplificada	Encuesta de 5 preguntas enviada a los 3 usuarios	6 semanas post-lanzamiento
CE-08 Precisión de KPIs en tablero	Concordancia entre los KPIs mostrados en Power BI y los registros en las fuentes originales	0% de divergencia en muestra auditada de 100 registros	Auditoría cruzada manual: valor en Power BI vs. valor en DataCRM/World Office	Al finalizar implementación y mensualmente

Nota: Elaboración propia

8.6 Plan de casos de prueba técnicos

Se define un plan de casos de prueba técnicos (tabla 19) que verifican el funcionamiento correcto de cada componente de la arquitectura de manera individual (pruebas unitarias) y en conjunto (pruebas de integración).

Tabla 21 *Casos de prueba técnicos*

ID	Componente	Caso de prueba	Métrica	Umbral de aceptación	Herramienta de verificación
CP-01	Pipeline n8n	Latencia extremo a extremo (extracción -> carga en Google Sheets)	Tiempo en minutos	<= 15 minutos	Log de ejecución n8n + timestamp en Google Sheets
CP-02	Pipeline n8n	Tasa de éxito de ejecuciones automáticas en ventana de 30 días	% de ejecuciones sin error	>= 95%	Dashboard de monitoreo n8n (Executions)
CP-03	Modelo ML	Precisión del modelo de alerta de quiebre de inventario	Recall en conjunto de prueba	>= 85%	Reporte scikit-learn classification_report
CP-04	Modelo ML	Tasa de falsos positivos en alertas de quiebre	Precisión	>= 70%	Reporte scikit-learn classification_report
CP-05	Power BI	Concordancia de KPIs entre tablero y fuente original	Diferencia absoluta en muestra de 100 registros	0% de divergencia	Auditoría manual cruzada mensual
CP-06	Google Sheets	Integridad referencial entre pestañas del repositorio (registros sin NIT válido)	Registros huérfanos	< 1% del total	Script Python de validación post-carga

CP-07	Power BI	Tiempo de carga del tablero principal en condiciones de uso estándar	Segundos hasta carga completa	<= 8 segundos	Medición directa en entorno de usuario
CP-08	Pipeline n8n	Comportamiento ante error de API de DataCRM (inyección de fallo simulado)	¿Se genera alerta automática?	Si: notificación en < 5 minutos al analista	Prueba de inyección de fallo en nodo HTTP de n8n
CP-09	Extracción World Office	Integridad de datos de inventario post-extracción por script Python	Registros perdidos vs. fuente en World Office	0% de pérdida de registros	Conteo comparativo Python vs. World Office
CP-10	Gobernanza	Activación de alerta cuando campos críticos superan el umbral de nulos	¿Se detiene el pipeline y se genera alerta?	Si, en < 5 minutos	Prueba con dataset sintético con >5% de NITs nulos

Nota: Elaboración propia.

Casos de prueba complementarios:

- Pruebas de integración: Verificar que DataCRM -> n8n -> Google Sheets fluye sin pérdida de registros en cada ejecución completa del pipeline.

- Pruebas de regresión del modelo ML: Cada vez que se re-entrenan los datos, el Recall no debe caer más de 3 puntos porcentuales respecto a la línea base del modelo en producción.

- Pruebas de aceptación de usuario (UAT): Los 3 usuarios deben poder responder correctamente las preguntas de negocio de la sección 8.4 usando únicamente el tablero Power BI, sin consultar los sistemas fuente.

8.7 Matriz de gestión de riesgos

La viabilidad y sostenibilidad de la implementación tecnológica dependen de una gestión proactiva de las vulnerabilidades inherentes al despliegue. Por ello, se ha estructurado una matriz de criticidad de riesgos en la tabla 20, diseñada para identificar, cuantificar y establecer protocolos de mitigación ante factores que puedan comprometer la continuidad del proyecto. La evaluación se fundamenta en el cálculo de la Exposición al Riesgo (E = P*I), donde la Probabilidad (P) y el Impacto (I) se ponderan en una escala de 1 a 5. Bajo este modelo, se categorizan los eventos en niveles de riesgo Bajo (1-5), Medio (6-11) y Alto (> 12), permitiendo priorizar las respuestas estratégicas conforme a su severidad.

Tabla 22 *Matriz de gestión de riesgos*

ID	Riesgo	Categoría	P	I	Exp.	Nivel	Plan de mitigación	Responsable
R-01	DataCRM cambia su API y rompe los flujos de n8n	Técnico	0.6	4	12	ALTO	1. Documentar la versión de API utilizada. 2. Diseñar flujos n8n con adaptadores desacoplados. 3. Suscribirse a changelog de DataCRM. 4. Contrato de soporte para corrección en < 2 días hábiles.	Científico de datos

R-02	World Office actualiza su interfaz y quiebra scripts de Python	Técnico	0.6	4	12	ALTO	1. Evaluar si World Office ofrece API oficial (priorizar sobre GUI). 2. Documentar selectores utilizados. 3. Prueba de regresión automática semanal del script de extracción.	Desarrollador externo
R-03	Resistencia del equipo comercial a adoptar Power BI	Organizacional	0.8	4	16	ALTO	1. Involucrar usuarios clave en diseño de KPIs desde el inicio. 2. Capacitación práctica de 4 horas por usuario. 3. Designar un champion interno. 4. Medir adopción en semanas 2, 4 y 8.	Gerencia comercial
R-04	Pérdida o corrupción de datos en Google Sheets por edición accidental	Datos	0.8	5	20	ALTO	1. Proteger hojas de datos crudos (solo lectura para usuarios). 2. Activar historial de versiones. 3. Backup semanal automatizado a Google Drive en CSV. 4. Validación de integridad post-carga en n8n.	Analista de datos

R-05	Baja calidad de datos en DataCRM (campos nulos, duplicados, formatos inconsistentes)	Datos	0.8	4	16	ALTO	1. Implementar reglas de calidad en n8n. 2. Capacitar al equipo comercial en el registro correcto en el CRM. 3. Dashboard de calidad visible para el analista.	Analista de datos
R-06	Modelo ML genera alertas con alta tasa de falsos positivos (baja Precisión)	Técnico-Analítico	0.6	3	9	MEDIO	1. Umbral mínimo de Precisión en 70% antes de producción (CP-04). 2. Periodo de observación de 4 semanas con validación manual paralela. 3. Protocolo de feedback del equipo comercial para marcar alertas incorrectas.	Científico de datos
R-07	Cambio o salida del analista interno capacitado en el ecosistema	Organizacional	0.4	4	8	MEDIO	1. Documentar todos los procedimientos operativos en manual interno. 2. Capacitar a un segundo usuario como respaldo. 3. Mantener contrato de soporte externo activo.	Gerencia

R-08	Incumplimiento de la Ley 1581 (Habeas Data) por manejo de datos de clientes	Legal	0.4	5	10	MEDIO	1. Inventariar qué datos personales se procesan. 2. Implementar política de retención (sección 8.8.4). 3. Anonimizar datos en ambientes de desarrollo y prueba.	Gerencia + Asesor legal
R-09	Costos de herramientas o licencias superan el presupuesto estimado	Financiero	0.4	3	6	MEDIO	1. Priorizar herramientas open source (n8n self-hosted, Python). 2. Revisar costos de Power BI Pro mensualmente. 3. Incluir contingencia del 15% en el presupuesto total.	Gerencia
R-10	Histórico de datos insuficiente para entrenar el modelo ML con alta confianza	Técnico-Analítico	0.6	4	12	ALTO	1. Auditar el histórico disponible en DataCRM y World Office en la fase de diagnóstico. 2. Si es menor a 12 meses, aplicar reglas heurísticas inicialmente y reentrenar al acumular más datos. 3. Plan de acumulación de datos para reentrenamiento a los 6 meses.	Científico de datos

Nota: Elaboración propia. P = Probabilidad (0-1), I =Impacto (1-5), Exp. = Exposición (PxI*)*

8.8 Tiempo de mantenimiento por componente

En este contexto, el presupuesto de construcción del proyecto cubre los cinco meses de implementación; sin embargo, la sostenibilidad del ecosistema analítico depende de la existencia de un plan de mantenimiento formal que establezca con claridad los responsables de las actualizaciones, la periodicidad y las condiciones bajo las cuales deben ejecutarse. La ausencia de este plan genera un riesgo operativo, ya que, aunque el científico de datos entregue el sistema funcional, no se define explícitamente quién asume su sostenimiento cuando se presentan cambios en el comportamiento del inventario o cuando plataformas externas, como DataCRM, modifican sus interfaces de programación de aplicaciones (API). Por ende, la tabla 21 detalla los tiempos de mantenimiento requeridos por componente para el sostenimiento operativo.

Tabla 23 *Tiempos de mantenimiento por componente*

Componente	Evento de mantenimiento	Frecuencia	Responsable	Tiempo estimado
Modelo ML	Reentrenamiento con nuevos datos de inventario y ventas acumulados	Mensual	Científico de datos (contrato de soporte)	4 horas/mes
Modelo ML	Evaluación de deriva del modelo (data drift): verificar que el Recall no haya caído más de 3 puntos vs. línea base	Quincenal	Científico de datos	1 hora
n8n	Revisión de logs de errores y ejecuciones fallidas del periodo	Semanal	Analista de datos interno (capacitado)	30 minutos

n8n	Ajuste de flujos si DataCRM cambia su API o estructura de respuesta (Riesgo R-01)	Por evento	Científico de datos / Desarrollador externo	4-16 horas por cambio
Google Sheets	Auditoría de integridad: verificar que no haya celdas editadas manualmente en hojas protegidas	Semanal	Analista de datos interno	1 hora
Power BI	Actualización de medidas DAX si cambian los KPIs o las definiciones de negocio	Trimestral o por evento	Analista de datos interno	2-4 horas
Power BI	Revisión del modelo de datos si World Office cambia su estructura de exportación (Riesgo R-02)	Por evento	Analista de datos + Científico de datos	4-8 horas
Scripts Python (World Office)	Actualización de selectores o lógica si World Office cambia su interfaz (Riesgo R-02)	Por evento	Desarrollador externo / Científico de datos	4-16 horas
Gobernanza	Revisión y actualización del diccionario de datos y políticas de calidad	Semestral	Gerencia + Analista de datos	2 horas
Modelo de datos Power BI	Backup del archivo .pbix y verificación de conexiones activas	Mensual	Analista de datos	30 minutos

Nota: Elaboración propia

8.9 Análisis financiero: retorno sobre la inversión y costo-beneficio

Un proyecto de analítica de datos genera valor cuando los beneficios que produce superan su costo total (ver tabla 24), incluyendo los costos de construcción, los costos ocultos frecuentemente omitidos y los costos de operación post-implementación. Esta sección cuantifica el impacto financiero de la arquitectura propuesta para Valveco S.A.S.

8.9.1 Cuantificación del problema actual

Antes de calcular el retorno de la inversión, se cuantifica el costo del problema que la arquitectura resuelve de acuerdo con la tabla 22.

Tabla 24 *Cuantificación del problema*

Problema actual	Estimación cuantitativa mensual	Base del cálculo
Latencia de reportes: 2-4 horas de consolidación manual diaria por tres personas	~$900.000 COP/mes en horas-hombre	Supuesto: 3 horas/semana x 3 personas x valor hora $25.000 x 4 semanas. A validar con encuesta interna al inicio del proyecto.
Quiebres de stock no detectados oportunamente que detienen o retrasan ventas	~$5.000.000 COP/mes en ventas en riesgo	Supuesto: 1 quiebre de stock/mes promedio x venta promedio afectada. A validar con histórico de World Office.
Decisiones tardías por falta de datos en tiempo real	Variable - estimar con gerencia	Costo de oportunidad: oportunidades perdidas o retrasadas x ticket promedio de Valveco S.A.S.

Nota: Los valores son estimaciones de referencia que deben validarse durante la fase de diagnóstico con datos reales de Valveco S.A.S. Elaboración propia.

8.9.2 Beneficios cuantificables de la arquitectura propuesta

La automatización de los flujos de información no solo optimiza la integridad de los datos, sino que representa un beneficio monetario directo para la organización. A continuación, la tabla 23 presenta la estimación del valor económico generado por la nueva infraestructura.

Tabla 25 *Análisis del impacto económico usando la arquitectura de datos propuesta*

Beneficio	Estimación mensual (COP)	Base del cálculo
Eliminación de la consolidación manual de reportes (CE-04: reducción >= 70%)	~$630.000/mes en ahorro de horas-hombre	70% de reducción sobre línea base de $900.000/mes
Reducción de quiebres de stock: el modelo ML detecta con 60 días de anticipación y permite reposición oportuna	~$2.500.000/mes (estimado 50% de reducción de impacto)	Asumiendo que el modelo evita al menos 1 de cada 2 quiebres de stock mensual
Mejora en tasa de cierre comercial por visibilidad del pipeline en tiempo real (estimado +2%)	Depende del ticket promedio y volumen de oportunidades de Valveco S.A.S.	A calcular con gerencia comercial usando datos históricos de DataCRM
Total beneficio mensual estimado (conservador)	~$3.130.000/mes	Suma de los dos primeros rubros. No incluye mejora en tasa de cierre por ser variable.

Nota: Estimaciones de referencia sujetas a validación con datos reales. Elaboración propia.

8.9.3 Estimación del periodo de retorno de la inversión

Con base en las estimaciones anteriores, el periodo de retorno aproximado de la inversión se calcula con las siguientes fórmulas:

$$\text{ROI} = \frac{Beneficios\ del\ periodo - costo\ total\ del\ proyecto}{Costo\ total\ del\ proyecto} * 100 \quad (10)$$

$$Periodo\ de\ retorno = \frac{costo\ total\ del\ proyecto}{beneficio\ mensual\ neto\ estimado} \quad (11)$$

Con base en el costo total de implementación de $73.752.463 COP (Tabla 24) y el beneficio mensual neto conservador de $3.130.000 COP (Tabla 23), el periodo de retorno de la inversión se calcula de acuerdo con la Ecuación (11):

$$Periodo\ de\ retorno = \frac{\$73.752.463}{\$3.130.000/mes} \approx 24\ meses$$

En el escenario base, la inversión se recupera en aproximadamente 24 meses (dos años). A partir de ese punto, el beneficio neto mensual pasa a ser retorno positivo para la organización. Aplicando la fórmula de ROI a distintos horizontes de tiempo se obtiene:

Tabla 26 *Retorno de la inversión (ROI) por horizonte de evaluación*

Horizonte	Beneficio Acumulado	ROI
12 meses	$37.560.000	−49,1 %
24 meses	$75.120.000	+1,9 %
36 meses	$112.680.000	+52,8 %

Nota: ROI = (Beneficios acumulados − Costo total) / Costo total × 100. Elaboración propia.

Dado que los beneficios cuantificados constituyen estimaciones de referencia sujetas a validación con datos reales de Valveco S.A.S., el análisis se presenta en tres escenarios que reflejan distintos niveles de materialización del beneficio

Tabla 25 *Análisis de sensibilidad del periodo de retorno*

Escenario	Beneficio Mensual Neto	Periodo de Retorno
Pesimista (30 % del beneficio)	$939.000	~79 meses
Base (100 %)	$3.130.000	~24 meses
Optimista (150 %)	$4.695.000	~16 meses

Nota: Elaboración propia.

El análisis muestra que la rentabilidad del proyecto depende en gran medida de la materialización de los beneficios, especialmente de la reducción efectiva de quiebres de stock, que constituye el factor de mayor impacto. En el escenario base, un periodo de retorno cercano a dos años resulta razonable para un proyecto de infraestructura analítica en una organización con un nivel de madurez temprana, considerando además que la arquitectura habilita beneficios adicionales no incluidos en este cálculo (como la mejora en la tasa de cierre comercial y la disminución de decisiones tardías) y que la inversión en capital humano especializado, que concentra la mayor parte del presupuesto, fortalece capacidades analíticas sostenibles más allá del horizonte de recuperación. Se recomienda, por tanto, validar los supuestos de beneficio con el histórico real de World Office y DataCRM durante la fase inicial del proyecto, con el fin de precisar el escenario efectivo.

9. Plan de implementación del modelo de analítica de datos y storytelling

9.1 Plan de adopción

El plan de adopción busca garantizar la incorporación del modelo de analítica de datos y storytelling propuesto, dentro de la organización. Como parte del proceso, se busca que los colaboradores del área comercial y los directivos de la empresa comprendan, utilicen y confíen en las herramientas desarrolladas.

Debido a que se requiere una transformación organizacional, con cambios en la cultura de datos, en los procesos de gestión y las competencias de todos los implicados, se proponen las siguientes actividades.

9.1.1 Nivel estratégico gerencial

Debido a que la gerencia es el actor principal para la adopción de la arquitectura analítica, se ha planteado una sustitución progresiva de los informes que se recogían de forma manual por el dashboard corporativo diseñado en esta propuesta. Con este fin, se realiza una reunión para la presentación del proyecto liderada por el analista de datos e inteligencia de negocios, con el propósito de socializar los resultados del diagnóstico, las necesidades del negocio, y la estructura de indicadores alineados con los objetivos estratégicos de Valveco S.A.S.

A partir de la retroalimentación gerencial, se efectúan los ajustes necesarios al dashboard, garantizando su pertinencia y utilidad para la toma de decisiones.

Antes de dar inicio la fase de implementación, se incluye un proceso de capacitación enfocado en tres aspectos: (i) lectura e interpretación de visualizaciones, (ii) análisis de tendencias temporales y (iii) toma de decisiones basadas en datos.

Con el fin de asegurar la adopción del dashboard, la gerencia debe integrarlo como herramienta principal en los comités gerenciales, reemplazando los reportes tradicionales.

De manera complementaria, se establecen mecanismos de monitoreo del uso, a través del seguimiento de accesos por usuario y frecuencia de consulta en la herramienta de visualización.

Finalmente, tras seis semanas de uso, se aplica un instrumento de medición de satisfacción y adopción, orientado a evaluar la utilidad percibida, la facilidad de uso y el impacto en la toma de decisiones. Los resultados obtenidos permiten identificar oportunidades de mejora y ajustar la estrategia de adopción. A partir de ese momento se continúa con el plan de mantenimiento periódico.

9.1.2 Nivel táctico del área comercial

Los asesores comerciales deben conocer de manera clara los indicadores a partir de los cuales será medido su desempeño. En este sentido, la gerencia debe determinar un método de medición alineado con los indicadores considerados estratégicos y socializarlo con el equipo comercial, con el fin de preservar la confianza identificada en el diagnóstico organizacional y evitar que el proceso de automatización genere percepciones de incertidumbre o resistencia.

El rol del nivel táctico resulta fundamental, en la medida que son responsables de la recolección de una parte significativa de los datos que alimentan el dashboard. Por ello el analista de datos debe diseñar una sensibilización orientada a fortalecer la comprensión sobre (i) la importancia del dato, (ii) las buenas prácticas de ingreso de datos y (iii) evidenciar cómo los errores en la captura afectan la calidad de la información presentada. Lo anterior permite mejorar la calidad de los datos sin implementar controles excesivos. En este espacio es posible identificar barreras de adopción, aunque el software del CRM continúa siendo el mismo.

9.1.3 Nivel operativo

En el nivel operativo la adopción de la arquitectura se orienta a garantizar la calidad, oportunidad y consistencia de los datos desde su origen. Los actores de este nivel capturan información de inventarios y movimiento de productos, insumos críticos para la ingesta del modelo. La estandarización de los procedimientos de registro y los lineamientos para el diligenciamiento de la información con el fin de minimizar errores, omisiones o inconsistencias que afectan el desempeño del modelo. Se implementan validaciones de inventarios para asegurar la sostenibilidad del sistema.

9.2 Recursos necesarios

A continuación, se presentan los recursos técnicos, de infraestructura y presupuesto (ver la Tabla 24) requeridos para el proyecto.

9.2.1 Recursos Técnicos e Infraestructura

- **Hardware:** Servidor local o instancia en la nube (VPS) para el despliegue de Docker, donde residirá la instancia de n8n.
- **Software y SaaS:**
    - **n8n (Self-hosted):** Para la orquestación de flujos sin costos por ejecución.
    - **Python 3.x:** Para procesos de extracción de datos por interfaz gráfica y transformación de datos.
    - **Power BI Pro:** Para la publicación y compartición de informes en la organización.
    - **Conectores API:** Acceso a los endpoints de World Office y DataCRM.

9.2.2 Roles y Responsabilidades

1. **Analista de datos e inteligencia de negocios:** Es el puente entre la gerencia y la técnica. Se encarga de definir los KPIs críticos, diseñar la narrativa de los tableros en Power BI y asegurar que la herramienta responda a las necesidades de toma de decisiones de Valveco S.A.S.

2. **Ingeniero de datos (Especialista en automatización):** Perfil técnico encargado de las actividades relacionadas al uso de datos. Desarrolla los flujos en n8n, escribe los scripts de Python para la extracción y garantiza que la información fluya sin errores entre los sistemas.

3. **Científico de datos:** Perfil técnico encargado de los modelos de machine learning para la predicción de quiebre de stock en el inventario de las referencias que maneja Valveco S.A.S. Debe realizar mantenimiento del modelo y garantizar las mejores métricas para que las alertas de stock y decisiones de gerencia sean adecuadas.

4. **Técnico en infraestructura y seguridad TI:** Responsable de la puesta a punto del servidor (Docker), la gestión de credenciales y permisos de las APIs, y la implementación de protocolos de respaldo (backups) para proteger la integridad de la información.

5. **Alta dirección:** Es responsable de la adopción del modelo.

9.2.3 Presupuesto

Este presupuesto ha sido calculado específicamente para una implementación que impacta a 3 usuarios líderes (Gerencia, comercial y finanzas). Los costos reflejan tanto la inversión tecnológica como el capital humano especializado necesario para sustituir el trabajo manual por procesos automatizados.

Tabla 26 *Presupuesto plan de implementación*

Ítem	Descripción del Recurso	Costo Total (5 Meses)	Costo por Mes	Costo por Usuario (3)
Licencias Power BI Pro	Suscripción para visualización y publicación de reportes.	$ 777,963.45	$ 155,592.69	$ 51,864.23
Analista de Datos	Definición de KPIs, diseño de Dashboards y Storytelling.	$ 20,000,000.00	$ 4,000,000.00	$ 4,000,000.00
Ingeniero de Datos	Desarrollo de flujos n8n, scripts Python y ETL.	$ 17,500,000.00	$ 3,500,000.00	$ 3,500,000.00
Científico de Datos	Desarrollo y mantenimiento de modelos de ML para predicción de quiebre de Stock	$ 20,000,000.00	$ 4,000,000.00	$ 4,000,000.00
Infraestructura Cloud / VPS	Servidor para ejecución de n8n, Python y base de datos.	$ 324,500.00	$ 64,900.00	$ 64,900.00
Seguridad y Soporte TI	Gestión de servidores, backups y ciberseguridad.	$ 15,000,000.00	$ 3,000,000.00	$ 3,000,000.00
Capacitación y Cierre	Transferencia de conocimiento y manuales técnicos.	$ 150,000.00	$ 30,000.00	$ 10,000.00
TOTAL ESTIMADO		$ 73,752,463.45	$ 14,750,492.69	$ 14,626,764.23

Nota: Elaboración propia

9.3 Cronograma

El proyecto se ejecutará en un periodo de 20 semanas, garantizando una transición suave desde los procesos manuales hacia la automatización final.

Tabla 27 *Cronograma de implementación*

Actividad / Semana	Agosto				Septiembre				Octubre				Noviembre				Diciembre			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Establecer plan de trabajo y acta de inicio																				
Asignar recursos técnicos y entornos (Docker/n8n)																				
Recopilar datos mediante entrevistas (Ventas/Contable)																				
Analizar requerimientos y documentar procesos actuales																				
Realizar análisis de brechas																				
Desarrollar esquema inicial de recopilación (n8n/API)																				
Ejecución de procesos ETL y limpieza con Python																				
Modelado de datos relacionales para Business Intelligence																				
Ajustes de diseño de dashboards y storytelling																				
Implementación inicial del modelo en producción																				
Pruebas de validación y calidad de datos																				
Capacitación de usuarios finales																				
Seguimiento a la transición del nuevo proceso																				
Evaluación de resultados con KPIs																				
Presentación final y documento de cierre																				

Nota: Elaboración propia

10. Implementación en la organización

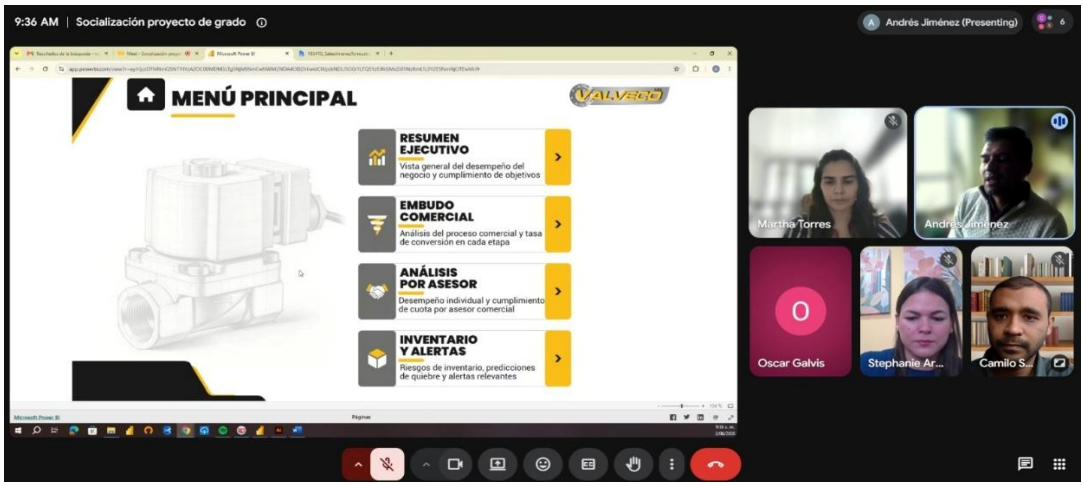
El 3 de junio de 2026 se llevó a cabo una reunión virtual con el propósito de presentar la propuesta desarrollada en el marco de este trabajo de investigación. Por parte de Valveco S.A.S. se contó con la participaron de Stephanie Arroyo, gerente administrativa, y Óscar Galvis, socio de la organización, ingeniero y experto en analítica de datos y Big Data. Si bien Héctor Galvis, gerente general de la empresa, no pudo estar presente en la reunión por razones de salud, su participación fue fundamental durante todo el desarrollo del proyecto, dado que realizó seguimiento permanente a las actividades ejecutadas y brindó apoyo en la atención de las inquietudes que se tuvieron durante la ejecución del proyecto. Aunque los lineamientos del trabajo de grado dirigido establecen un alcance que llega únicamente hasta el diseño del plan de implementación, se tomó la iniciativa de generar este espacio con los directivos de la organización con el propósito de conocer su percepción frente a la propuesta desarrollada e identificar los aspectos clave para su futura materialización. De esta manera, este encuentro se constituyó como el primer acercamiento y el puente estratégico para facilitar la eventual implementación de la solución dentro de la empresa.

La presentación inició con la exposición de los resultados obtenidos en el diagnóstico organizacional, abordando tanto los factores internos como externos identificados durante el análisis. Posteriormente, se presentó el pipeline propuesto del modelo, explicando de manera detallada cada una de sus etapas, desde la integración y transformación de los datos hasta el diseño del tablero.

A continuación, se mostró un prototipo funcional del tablero de control, construido a partir de los datos históricos suministrados por la empresa correspondientes al año 2025, complementado con datos sintéticos para el año 2026, con el fin de ilustrar el comportamiento esperado de la solución. Hay que aclarar que ese prototipo contiene datos

reales de la empresa por lo que no es posible anexar el enlace en este documento. Finalmente se expuso el plan de implementación propuesto, incluyendo las actividades, recursos y consideraciones necesarias para su puesta en marcha dentro de la organización.

Figura 26 *Reunión de presentación*



Nota: Captura de pantalla

10.1 Retroalimentación de la organización

Posteriormente a la presentación de la propuesta, se obtuvo retroalimentación por parte de Oscar Galvis y Stephanie Arroyo, sus aportes fueron significativos para la consolidación del proyecto. Sus observaciones permitieron validar la pertinencia de la solución propuesta frente a problemáticas asociadas a la latencia de la información comercial, la trazabilidad de los datos y la integración de la información proveniente de los distintos sistemas utilizados. Asimismo, se identificaron oportunidades de mejora orientadas a fortalecer la escalabilidad, robustez y sostenibilidad de la solución en escenarios futuros.

Actualmente, el proyecto avanza de manera favorable a una fase de análisis y evaluación de alternativas para su eventual implementación dentro de la empresa. A

continuación, se presenta en detalle la retroalimentación recibida durante la sesión de socialización.

Durante la reunión, tanto Oscar Galvis como Stephanie Arroyo destacaron el valor de la solución desarrollada, especialmente en lo relacionado con el proceso de integración, limpieza, estandarización de los datos. Señalaron que este componente permite abordar problemáticas actuales de la organización, tales como la latencia de la información comercial, la falta de trazabilidad entre las diferentes fuentes de datos y las dificultades para contrastar la información registrada en los distintos sistemas corporativos.

En relación con la arquitectura de datos propuesta, se resaltó la importancia de considerar criterios de escalabilidad desde las etapas iniciales, con el fin de soportar futuros incrementos en el volumen de información procesada. Como una posible evolución de la solución, Oscar Galvis sugirió evaluar la implementación de mecanismos de procesamiento incremental basados en deltas y llaves compuestas. Esta estrategia, permitiría evitar el reprocesamiento completo de los datos en cada ejecución, optimizando así el desempeño y la eficiencia de los procesos ETL y facilitando la escalabilidad de la plataforma a largo plazo.

Respecto al componente analítico, se valoró positivamente la incorporación de modelos predictivos orientados a la gestión de inventarios. Como oportunidad de evolución, se propuso enriquecer el modelo de machine learning mediante la inclusión de variables temporales asociadas a la estacionalidad de las ventas, así como factores propios del negocio, entre ellos las márgenes de utilidad de las referencias, los tiempos de reposición y los patrones comerciales asociados a las especializaciones de los asesores comerciales. A partir de esta observación, se planteó la posibilidad de evolucionar hacia un motor de recomendaciones capaz de sugerir cantidades óptimas de reposición para cada referencia, alineando las decisiones de abastecimiento con los objetivos de rentabilidad y generación de valor para la organización.

Asimismo, por parte de la empresa se destacó el potencial de la solución para fortalecer los procesos de análisis y control de la calidad de los datos. En particular, identificaron como un desarrollo futuro la incorporación de mecanismos que permitan visualizar, explicar y monitorear las diferencias existentes entre la información registrada en el CRM y la proveniente del sistema World Office.

Finalmente, Stephanie Arroyo, señaló que las políticas de captura de información en el CRM han evolucionado a lo largo del tiempo. Actualmente, la organización registra la totalidad de los negocios, incluidas las cotizaciones telefónicas, situación que genera variaciones en los indicadores históricos de conversión y que deberá ser considerada en futuras fases del proyecto para contextualizar adecuadamente los análisis y mejorar la interpretación de los resultados. Los directivos coincidieron en que los proyectos basados en datos representan una inversión estratégica para las organizaciones cuyo valor se materializa cuando los resultados pueden demostrarse de manera objetiva, medible y alineada con los objetivos del negocio. Desde esta perspectiva, consideraron que la propuesta desarrollada representa una base sólida para avanzar hacia una gestión empresarial soportada en datos, analítica y procesos de toma de decisiones informadas.

11. Conclusiones y Recomendaciones

11.1 Conclusiones

A partir del desarrollo del presente trabajo se evidenció la necesidad de Valveco S.A.S. de evolucionar hacia una cultura organizacional basada en datos, que permita superar la fragmentación de sus sistemas de información y transformar la gestión comercial mediante herramientas analíticas que habiliten decisiones estratégicas oportunas, confiables y fundamentadas en evidencia. El proyecto respondió a esta necesidad a través de un proceso estructurado que integró fundamentación teórica, diagnóstico organizacional, diseño de arquitectura y formulación de una hoja de ruta de adopción institucional.

El primer paso para llevar a cabo este proyecto fue el cumplimiento del primer objetivo específico, orientado a fundamentar teórica y metodológicamente la propuesta mediante la revisión de literatura especializada. La revisión de la literatura permitió comprender que la adopción de una arquitectura de analítica de datos no se reduce a la selección de herramientas tecnológicas, sino que exige un enfoque integrado que articule procesos de integración entre sistemas heterogéneos como el ERP y el CRM, principios de calidad y gobernanza del dato, y técnicas de comunicación visual orientadas al storytelling estratégico. Esta revisión también proporcionó una perspectiva más amplia sobre los marcos metodológicos disponibles, permitiendo una selección informada de enfoques adaptados a las necesidades específicas de Valveco S.A.S. y garantizando que la propuesta resultante estuviera fundamentada de forma coherente con la literatura vigente del campo.

El diagnóstico interno y externo, correspondiente al segundo objetivo específico, proporcionó una comprensión completa del estado actual de la gestión de información comercial en la organización, identificando con precisión las oportunidades de mejora y las brechas que debían ser resueltas por la arquitectura propuesta. Los resultados del

instrumento TDWI evidenciaron que Valveco S.A.S. se ubica en una etapa de madurez analítica temprana con una puntuación global de 38 sobre 100, siendo la analítica y la gobernanza de datos las dimensiones de mayor déficit. Este análisis estratégico permitió identificar no solo las fuentes de datos disponibles, sino también las limitaciones más críticas en su gestión: inconsistencias percibidas entre World Office y DataCRM, una carga operativa de consolidación manual de entre dos y cuatro horas diarias, y una latencia de información que comprometía la capacidad de decisión del área comercial. Paralelamente, la Matriz DOFA evidenció que la organización cuenta con una apertura institucional favorable hacia la analítica de datos.

En cuanto al tercer objetivo específico, el diseño de la arquitectura de analítica de datos aseguró la disponibilidad de información integrada, estandarizada y confiable para el análisis comercial de Valveco S.A.S., sentando las bases técnicas para una toma de decisiones proactiva y basada en evidencia. La arquitectura propuesta, estructurada en cuatro capas funcionales, resuelve de manera articulada los problemas identificados en el diagnóstico: la capa de ingesta y orquestación con n8n y Python elimina la fragmentación entre fuentes y automatiza la consolidación de datos; la capa de procesamiento y almacenamiento construye mediante un pipeline ETL un repositorio analítico centralizado que sirve como fuente única de consulta para toda la organización, implementado inicialmente en Google Sheets como una solución práctica y de bajo costo que, en fases posteriores, podrá escalar hacia un Data Warehouse más robusto; la capa de inteligencia predictiva incorpora un modelo de Random Forest Classifier capaz de anticipar quiebres de inventario con sesenta días de antelación; y la capa de consumo y storytelling en Power BI traduce los datos en narrativas accionables para los perfiles de gerencia, dirección comercial y área administrativa. Al centralizar los datos de ventas e inventarios bajo un esquema semántico de estrella con Row Level Security, se facilita el acceso a información

relevante para cada nivel jerárquico de la organización, promoviendo una toma de decisiones más ágil, transparente y diferenciada.

El cuarto objetivo específico se llevó a cabo mediante la formulación de una hoja de ruta de adopción institucional estructurada en veinte semanas, organizada en tres niveles de intervención: estratégico gerencial, táctico del área comercial y operativo. Este plan define los roles especializados requeridos, el presupuesto estimado de implementación e incorpora un marco de gobernanza que incluye diccionario de datos, política de calidad con umbrales automatizados y controles de acceso diferenciados. Reconociendo la importancia del factor humano, el plan contempla estrategias de gestión del cambio orientadas a construir comprensión sobre el valor del dato en todos los niveles de la organización, dado que la sostenibilidad de cualquier ecosistema analítico depende tanto de la solidez técnica de su diseño como de la apropiación cultural que los equipos desarrollen frente a la herramienta.

La implementación del diseño propuesto en el área comercial de Valveco S.A.S. ofrece a la gerencia y a los asesores comerciales la oportunidad de transformar significativamente la manera en que gestionan la información: al liberar recursos de tareas operativas repetitivas, se potencia la capacidad estratégica del equipo para concentrarse en el análisis de alto valor y en la generación de acciones que impulsen el cumplimiento de las metas comerciales. El objetivo general del proyecto (“Diseñar una arquitectura de analítica de datos basada en visualización y storytelling para la toma de decisiones estratégicas en Valveco S.A.S.”) se cumple a través de la integración coherente de sus cuatro objetivos específicos, demostrando que es posible, incluso para una organización en etapa analítica temprana, construir un ecosistema de datos robusto y viable que responda a necesidades reales de negocio, sin requerir infraestructuras complejas de partida, siempre que la solución se sustente en evidencia, sea técnicamente sólida y esté diseñada para ser adoptada por las personas que la utilizarán.

11.2 Recomendaciones

Valveco S.A.S. cuenta con condiciones favorables para avanzar significativamente en su nivel de madurez analítica, y la implementación de la arquitectura propuesta en este proyecto puede tener un impacto positivo y sostenido en la estrategia comercial y en la capacidad de toma de decisiones de la organización. Por tanto, con el propósito de maximizar el valor generado por el diseño desarrollado, se presentan a continuación las recomendaciones para las etapas subsiguientes del proyecto.

Como primera recomendación, se sugiere priorizar la implementación de las capas de ingesta y procesamiento como fase productiva inicial, dado que la automatización de la consolidación de datos representa el beneficio de mayor impacto operativo inmediato para el equipo comercial. Esta decisión permitirá generar resultados tangibles en el corto plazo, construyendo confianza institucional en el sistema antes de incorporar las capas de mayor complejidad técnica. Para garantizar la calidad de esta transición, se recomienda ejecutar los casos de prueba técnicos definidos en el plan, verificando la integridad de los flujos entre World Office, DataCRM y Power BI, y validar los umbrales de calidad automatizados antes de habilitar el acceso de los usuarios finales.

En cuanto a la sostenibilidad del ecosistema analítico, se recomienda institucionalizar las prácticas de gobernanza de datos formuladas en este proyecto (diccionario de datos, política de calidad y controles de acceso mediante Row Level Security) como procedimientos ordinarios del área, con revisiones periódicas que garanticen su vigencia frente a cambios en los sistemas fuente o en los indicadores estratégicos definidos por la gerencia. La figura del analista de datos debe consolidarse como custodio permanente de esta gobernanza, siendo el responsable de preservar la coherencia del modelo semántico en Power BI y de actuar como puente entre las necesidades de negocio y la arquitectura técnica que las sustenta.

Adicionalmente, se recomienda que, una vez estabilizado el modelo predictivo de quiebre de stock, se evalúe su enriquecimiento con variables contextuales del entorno de negocios, tales como la estacionalidad de la demanda por sector industrial, los tiempos históricos de despacho por proveedor y las variaciones de la TRM que inciden en los costos de importación. En paralelo, se recomienda profundizar en la integración entre los datos comerciales y los de inventario, que si bien corresponden a dimensiones distintas de la operación, su articulación representa un valor estratégico significativo para el equipo de ventas: contar con visibilidad en tiempo real sobre las referencias disponibles en stock permitirá a los asesores comerciales orientar sus esfuerzos hacia los productos con mayor disponibilidad inmediata, establecer compromisos más precisos con los clientes en cuanto a tiempos de despacho y reducir fricciones en el proceso de cierre de negocios. Esta evolución conjunta permitirá avanzar desde la analítica predictiva actual hacia un enfoque prescriptivo, donde el sistema no solo anticipe riesgos de desabastecimiento, sino que recomiende acciones específicas tanto al equipo de compras como al comercial, fortaleciendo la capacidad de Valveco S.A.S. para proteger sus ingresos, mejorar la experiencia del cliente y consolidar su reputación de servicio ágil y confiable.

Se recomienda también evaluar la extensión de la arquitectura diseñada hacia otras áreas de la organización como finanzas, logística y servicio postventa, aprovechando que la infraestructura tecnológica seleccionada —n8n, Python y Power BI— ofrece condiciones favorables de escalabilidad sin requerir inversiones adicionales de gran magnitud. Esta expansión permitiría construir una visión integrada del negocio que conecte el desempeño comercial con la trazabilidad operativa y la gestión financiera, habilitando un nivel de análisis organizacional más completo, y contribuyendo al avance progresivo de Valveco S.A.S. en el modelo de madurez analítica TDWI más allá de la etapa temprana en la que actualmente se encuentra.

Finalmente, las recomendaciones para la adopción exitosa de esta solución de analítica de datos convergen en un elemento transversal: la necesidad de construir una cultura organizacional orientada al dato. Se sugiere diseñar un programa de formación continua dirigido tanto a los colaboradores del área comercial como a la alta dirección, que vaya más allá del entrenamiento técnico en el uso de Power BI y apunte a transformar la manera en que la organización lee e interpreta su propio desempeño. La actualización automática de los datos desde las fuentes, los controles de calidad integrados al pipeline y la narrativa estructurada del tablero proporcionan las condiciones técnicas para esta transformación; sin embargo, serán la gestión del cambio y el compromiso institucional los factores determinantes para que la toma de decisiones basada en datos se consolide como una práctica natural, transversal e indispensable dentro de Valveco S.A.S.

12. Referencias

Ain, N., Vaia, G., DeLone, W. H., & Waheed, M. (2019). Two decades of research on business intelligence system adoption, utilization and success: A systematic literature review. *Decision Support Systems, 127*, 113113. https://doi.org/10.1016/j.dss.2019.113113

Akpe, O.-e. E., Owoade, S., Ubanadu, B. C., Daraojimba, A. I., & Gbenle, T. P. (2021). *A conceptual framework for data-driven decision making in enterprise IT management*. *Iconic Research and Engineering Journals, 5*(3), 318–339. https://www.irejournals.com/formatedpaper/1708518.pdf

Anifowose, O. (2023). Augmented Decision Intelligence (ADI): The convergence of AI, predictive analytics, and executive cognition as a transformative paradigm for strategic decision-making. Journal of Computational Analysis and Applications, 31(3),750-777. https://eudoxuspress.com/index.php/pub/article/download/4136/3018/8232

Armbrust, M., Ghodsi, A., Xin, R., & Zaharia, M. (2021). *Lakehouse: A new generation of open platforms that unify data warehousing and advanced analytics*. En *Proceedings of the 11th Conference on Innovative Data Systems Research (CIDR 2021)*. https://www.cidrdb.org/cidr2021/papers/cidr2021_paper17.pdf

Atlas, F., Yitong, Y., & Khan, K. U. (2025). *The financial and market impact of big data analytics and big data talent analytics capability: A knowledge management*

perspective. *Humanities and Social Sciences Communications, 12*, Article 919. https://doi.org/10.1057/s41599-025-05206-y

Benoliel, D., & Feldman, D. K. D. (2025). *The social cost of artificial knowledge*. *Journal of Law, Technology and Policy, 2025*(1), 331–378. https://illinoisjltp.com/current-issue/article/the-social-cost-of-artificial-knowledge

Bernardo, B. M. V., Mamede, H. S., Barroso, J. M. P., & Santos, V. M. P. D. (2024). *Data governance & quality management—Innovation and breakthroughs across different fields*. *Journal of Innovation & Knowledge, 9*(4), Article 100598. https://doi.org/10.1016/j.jik.2024.100598

Biemans, W. (2023). *The impact of digital tools on sales-marketing interactions and perceptions*. *Industrial Marketing Management, 115*, 395–407. https://doi.org/10.1016/j.indmarman.2023.10.015

Bollineni, S. (2023). *DataOps: Applying DevOps principles to data engineering for improved data management*. *Journal of Artificial Intelligence, Machine Learning and Data Science, 1*(4), 1285–1288. https://doi.org/10.51219/JAIMLD/satyadeepak-bollineni/293

Caballero Muñoz-Reja, I., Gómez Carretero, A. I., Gualo Cejudo, F., Merino García, J., Rivas García, B., & Piattini Velthuis, M. (2019). *Calidad de datos*. Ediciones de la U.

Conde, R. (2025). *Necessary condition analysis for sales funnel optimization*. *Journal of Marketing Analytics*, 1–15. https://doi.org/10.1057/s41270-025-00388-5

Cuzzocrea, A. (2024). *Evolving OLAP and BI towards complex, high-performance Big-OLAP-Data-Cube-Processing analytics frameworks: How to speed-up large-scale, high-dimensional queries over clouds*. *Procedia Computer Science, 246*, 4169–4175. https://doi.org/10.1016/j.procs.2024.09.256

David, F. R., David, F. R., & David, M. E. (2023). *Conceptos de administración estratégica*. Pearson Educación. https://ebooks7-24-com.bdbiblioteca.universidadean.edu.co/?il=32928

El-Gharib, N. M., & Amyot, D. (2023). *Robotic process automation using process mining: A systematic literature review*. *Data & Knowledge Engineering, 148*, Article 102229. https://doi.org/10.1016/j.datak.2023.102229

Fahmideh, M., & Beydoun, G. (2019). *Big data analytics architecture design: An application in manufacturing systems*. *Computers & Industrial Engineering, 128*, 948–963. https://doi.org/10.1016/j.cie.2018.08.004

Fannouch, A., Gharib, J., & Gahi, Y. (2025). *Enhancing DataOps practices through innovative collaborative models: A systematic review*. *International Journal of Information Management Data Insights, 5*(1), Article 100321. https://doi.org/10.1016/j.jjimei.2025.100321

Farinha, D., Pereira, R., & Almeida, R. (2024). *A framework to support robotic process automation*. *Journal of Information Technology, 39*(1), 149–166. https://doi.org/10.1177/02683962231165066

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

Gupta, A., Kr Singh, R., Kamble, S., & Mishra, R. (2022). *Knowledge management in Industry 4.0 environment for sustainable competitive advantage: A strategic framework*. *Knowledge Management Research & Practice, 20*(6), 878–892. https://doi.org/10.1080/14778238.2022.2144512

Jaggia, S., Kelly, A., Lertwachara, K., & Chen, L. (2023). *Analítica de Negocios: Comunicación con números*. McGraw-Hill Interamericana

Li, L., Lin, J., Ouyang, Y., & Luo, X. (2022). Evaluating the impact of big data analytics usage on the decision-making quality of organizations. *Technological Forecasting and Social Change, 175*, 121355. https://doi.org/10.1016/j.techfore.2021.121355

Mandaliya, S., & Kulkarni, P. (2025). Leveraging AI and ML for enhanced Business Intelligence systems. *International Journal of Advanced Computer Science and Applications, 16*(11). https://thesai.org/Downloads/Volume16No11/Paper_33-Leveraging_AI_and_ML_for_Enhanced_Business_Intelligence_Systems.pdf

Market Data Forecast. (Abril 2025). Valves market size, share, trends & growth forecast report, 2024–2033. Recuperado el 31 agosto de 2025. https://www.marketdataforecast.com/market-reports/valves-market

Mikalef, P., Boura, M., Lekakos, G., & Krogstie, J. (2019). Exploring the relationship between big data analytics capability and competitive performance: The mediating roles of dynamic and operational capabilities. *Information & Management, 57*(2), 103169. https://doi.org/10.1016/j.im.2019.103169

Nambiar, A., & Mundra, D. (2022). An overview of data warehouse and data lake in modern enterprise data management. *Big Data and Cognitive Computing, 6*(4), 132. https://doi.org/10.3390/bdcc6040132

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

Rahman, M. A., Zamil, M. H., & Hoque, A. B. (2025). A meta-analysis of ERP and CRM integration tools in business process optimization. *ASRC Procedia: Global Perspectives in Science and Scholarship, 1*(1), 278–312. https://doi.org/10.63125/yah70173

Ribeiro, J., Lima, R., Eckhardt, T., & Paiva, S. (2021). Robotic process automation and artificial intelligence in Industry 4.0 – A literature review. *Procedia Computer Science, 181*, 51–58. https://doi.org/10.1016/j.procs.2021.01.104

Saddad, E., El-Bastawissy, A., Mokhtar, H., & Hazman, M. (2020). Lake Data Warehouse Architecture for Big Data Solutions. *International Journal of Advanced Computer Science and Applications*, 11(8). https://thesai.org/Downloads/Volume11No8/Paper_54-Lake_Data_Warehouse_Architecture.pdf

Sharda, R., Turban, E., & Delen, D. (2018). *Business intelligence, analytics, and data science: A managerial perspective*. Pearson Educación.

Solano, M. C., & Cruz, J. C. (2024). Integrating analytics in enterprise systems: A systematic literature review of impacts and innovations. Administrative Sciences, 14(7), 138. https://doi.org/10.3390/admsci14070138

Sultana, S., Akter, S., & Kyriazis, E. (2024). Theorising data-driven innovation capabilities to survive and thrive in the digital economy. Journal of Strategic Marketing, 32(7), 864-890. https://doi.org/10.1080/0965254X.2021.2013934

The Observatory of Economic Complexity. (2023). *Safety or relief valves (HS 848140): Colombia trade data*. https://oec.world/

Valveco S.A.S. (n.d.). Misión, visión y objetivos. Recuperado el 1 de septiembre de 2025, de https://valveco.com.co/pages/acerca-de-valveco

Velásquez, J. (2025). Emerging trends in retail analytics: A bibliometric analysis of the last decade / Tendencias emergentes en la analítica del retail: Un análisis bibliométrico

de la última década. *DYNA, 92*(237), 16–29. https://doi.org/10.15446/dyna.v92n237.116880

Wang, J., Liu, Y., Li, P., Lin, Z., Sindakis, S., & Aggarwal, S. (2024). Overview of data quality: Examining the dimensions, antecedents, and impacts of data quality. *Journal of the Knowledge Economy, 15*(1), 1159–1178. https://doi.org/10.1007/s13132-022-01096-6

Ware, C. (2012). Information visualization: Perception for design (3rd ed.). Morgan Kaufmann

Yahyaoui, M., Aknin, N., Amjad, S., & Benameur, L. (2025). Automatic selection of bitmap join indexes in data warehouses using CFPGrowth++ algorithm. *Informatica, 49*(27), 365–378. https://doi.org/10.31449/inf.v49i27.7807

13. Anexos

13.1 Carta de aceptación Valveco S.A.S.

Bogotá 25-agosto-25

SEÑORES
SUBPROCESO DE TRABAJOS DE GRADO
Universidad Ean
Bogotá

Respetados señores,

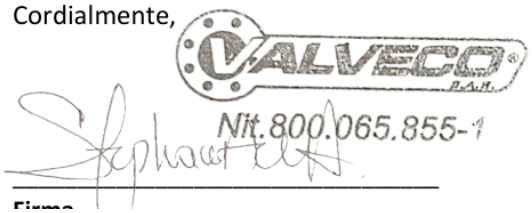
Por medio de la presente nos permitimos autorizar a **Andrés Jiménez, Camilo Salas y Martha Torres,** identificados con cedula de ciudadanía 1026279113, 1140904926, 41954513 estudiantes del programa de maestría **Ciencia de datos** de la Universidad Ean, para que realicen en nuestra organización **VALVECO S.A.S** su trabajo de grado titulado: **Implementación de tablero para integración y estandarización de bases de datos para medición de metas y resultados** bajo la modalidad de trabajo dirigido.

Como empresa nos comprometemos a brindar la información requerida para adelantar este proyecto académico, así mismo, autorizamos que el documento resultado de este trabajo sea publicado en el repositorio documental Minerva de la Universidad Ean.

A continuación, relacionamos los datos de la persona que será el contacto designado por la empresa.

NOMBRE DEL CONTACTO: Stephanie Arroyo A.
CARGO QUE OCUPA: Gerente Administrativa
TELÉFONO: 3132008006
CORREO ELECTRÓNICO: administrativa@valveco.com.co

Cordialmente,



Firma

13.2 Encuesta de diagnóstico Valveco S.A.S.

Esta encuesta se realiza en el marco del proyecto **"Arquitectura de ciencia de datos y storytelling: automatización y visualización estratégica para la gestión comercial en Valveco S.A.S"**.

Su objetivo es conocer la percepción de los colaboradores sobre diferentes aspectos relacionados con la gestión de la información comercial, tales como la calidad de los datos, latencia de la información para la toma de decisiones, automatización y monitoreo de procesos (DataOps), esfuerzo operativo de consolidación, confiabilidad y alineación de KPIs comerciales.

La información recolectada será utilizada únicamente con fines académicos. Agradecemos su participación y sinceridad en las respuestas.

Nombre _____

Cargo _____

A. Calidad de los datos comerciales

1. La información comercial disponible en los sistemas de Valveco S.A.S. contiene los datos necesarios y suficientes para realizar adecuadamente mi trabajo

- Totalmente en desacuerdo
- En desacuerdo
- Ni de acuerdo ni en desacuerdo
- De acuerdo
- Totalmente de acuerdo

2. Cuando debo comparar cifras comerciales en diferentes reportes o sistemas (por ejemplo, CRM, ERP o reportes financieros), estas coinciden entre sí.

- Totalmente en desacuerdo
- En desacuerdo
- Ni de acuerdo ni en desacuerdo
- De acuerdo
- Totalmente de acuerdo

3. Los datos comerciales reflejan con precisión lo que ocurre en la operación real de ventas

- Totalmente en desacuerdo
- En desacuerdo
- Ni de acuerdo ni en desacuerdo
- De acuerdo
- Totalmente de acuerdo

4. La forma en que se estructuran, organizan y estandarizan los datos comerciales en Valveco S.A.S. (formatos, nombres y reglas comunes para registrar la información) facilita el análisis y la toma de decisiones en mi trabajo.

- Totalmente en desacuerdo
- En desacuerdo
- Ni de acuerdo ni en desacuerdo
- De acuerdo
- Totalmente de acuerdo

B. Latencia de información para la toma de decisiones

1. La información comercial está disponible en el momento en que la necesito para hacer seguimiento a mis resultados y metas.

- Totalmente en desacuerdo

- En desacuerdo
- Ni de acuerdo ni en desacuerdo
- De acuerdo
- Totalmente de acuerdo

2. Los reportes disponibles me permiten conocer el avance de mis resultados en periodos específicos (diarios, semanales, mensuales).

- Totalmente en desacuerdo
- En desacuerdo
- Ni de acuerdo ni en desacuerdo
- De acuerdo
- Totalmente de acuerdo

3. La información comercial que utilizo se encuentra actualizada cuando la consulto.

- Totalmente en desacuerdo
- En desacuerdo
- Ni de acuerdo ni en desacuerdo
- De acuerdo
- Totalmente de acuerdo

C. Esfuerzo operativo de consolidación

1. Con frecuencia debo consolidar información manualmente desde diferentes fuentes (archivos, sistemas o reportes).

- Totalmente en desacuerdo
- En desacuerdo

- Ni de acuerdo ni en desacuerdo
- De acuerdo
- Totalmente de acuerdo

2. La revisión o verificación de cifras comerciales requiere tiempo adicional dentro de mis actividades habituales

- Totalmente en desacuerdo
- En desacuerdo
- Ni de acuerdo ni en desacuerdo
- De acuerdo
- Totalmente de acuerdo

D. Automatización y monitoreo de procesos (DataOps)

1. Con frecuencia se deben realizar reprocesos o se repiten tareas al revisar información o seguimiento a las cifras.

- Totalmente en desacuerdo
- En desacuerdo
- Ni de acuerdo ni en desacuerdo
- De acuerdo
- Totalmente de acuerdo

2. Mejorar la forma en que se organizan, integran y consolidan los datos comerciales reduciría la carga operativa actual del área comercial

- Totalmente en desacuerdo
- En desacuerdo

- Ni de acuerdo ni en desacuerdo
- De acuerdo
- Totalmente de acuerdo

E. Confiabilidad y alineación de KPIs comerciales

1. Los indicadores comerciales que utilizo están claramente definidos y documentados.

- Totalmente en desacuerdo
- En desacuerdo
- Ni de acuerdo ni en desacuerdo
- De acuerdo
- Totalmente de acuerdo

2. Entiendo cómo se calcula el cumplimiento de metas en el área comercial.

- Totalmente en desacuerdo
- En desacuerdo
- Ni de acuerdo ni en desacuerdo
- De acuerdo
- Totalmente de acuerdo

3. Contar con indicadores definidos influye en mi nivel de confianza en los resultados reportados.

- Totalmente en desacuerdo
- En desacuerdo
- Ni de acuerdo ni en desacuerdo
- De acuerdo

- Totalmente de acuerdo

4. Tengo claridad sobre los criterios con los cuales se evalúa mi desempeño en el área comercial.

   - Totalmente en desacuerdo
   - En desacuerdo
   - Ni de acuerdo ni en desacuerdo
   - De acuerdo
   - Totalmente de acuerdo