

UNIVERSIDAD SAN FRANCISCO DE QUITO USFQ

Colegio de Posgrados

**Optimización de la Gobernanza y Calidad de Datos en Marketing:
Un Modelo Estratégico**

Mauricio Alejandro Vega Pérez

**Julián Maya MSc
Director de Trabajo de Titulación**

Trabajo de titulación de posgrado presentado como requisito
para la obtención del título de Máster en Gerencia de Datos y Negocios

Quito, 9 de julio del 2024

Universidad San Francisco de Quito USFQ

Colegio de posgrados

HOJA DE APROBACIÓN DE TRABAJO DE TITULACIÓN

Optimización de la Gobernanza y Calidad de Datos en Marketing: Un Modelo Estratégico

Mauricio Alejandro Vega Pérez

Nombre del Director del Programa: Julián Maya

Título académico: MSc

Director del programa de: Titulación

Nombre del Decano del colegio Académico: Ana María Novillo

Título académico: PhD

Decano del Colegio: Decana CADE

Nombre del Decano del Colegio de Posgrados: Hugo Burgos

Título académico: PhD

Quito, julio 2024

© DERECHOS DE AUTOR

Por medio del presente documento certifico que he leído todas las Políticas y Manuales de la Universidad San Francisco de Quito USFQ, incluyendo la Política de Propiedad Intelectual USFQ, y estoy de acuerdo con su contenido, por lo que los derechos de propiedad intelectual del presente trabajo quedan sujetos a lo dispuesto en esas Políticas.

Asimismo, autorizo a la USFQ para que realice la digitalización y publicación de este trabajo en el repositorio virtual, de conformidad a lo dispuesto en la Ley Orgánica de Educación Superior del Ecuador.

Nombre del estudiante: Mauricio Alejandro Vega Pérez

Código de estudiante: 00337189

C.I. 1716385404

Lugar y fecha:

Quito, 9 de julio de 2024.

ACLARACIÓN PARA PUBLICACIÓN

Nota: El presente trabajo, en su totalidad o cualquiera de sus partes, no debe ser considerado como una publicación, incluso a pesar de estar disponible sin restricciones a través de un repositorio institucional. Esta declaración se alinea con las prácticas y recomendaciones presentadas por el Committee on Publication Ethics COPE descritas por Barbour et al. (2017) Discussion document on best practice for issues around theses publishing, disponible en <http://bit.ly/COPETHeses>.

UNPUBLISHED DOCUMENT

Note: The following graduation project is available through Universidad San Francisco de Quito USFQ institutional repository. Nonetheless, this project – in whole or in part – should not be considered a publication. This statement follows the recommendations presented by the Committee on Publication Ethics COPE described by Barbour et al. (2017) Discussion document on best practice for issues around theses publishing available on <http://bit.ly/COPETHeses>.

DEDICATORIA

To the journey that never ends.

AGRADECIMIENTOS

Quisiera expresar mi más profundo agradecimiento a todas las personas que han contribuido al desarrollo de este trabajo de titulación.

En primer lugar, agradezco a mis padres, cuyo amor y apoyo incondicional han sido fundamentales en mi vida. A mi esposa, por su alegría, paciencia, comprensión y constante apoyo durante todo el proceso. A mi hermano, por su ejemplo, aliento y por siempre impulsarme a dar lo mejor de mí.

También quisiera agradecer a mis profesores de la maestría, cuya orientación y conocimientos han sido invaluable. Su dedicación y pasión han sido una fuente de inspiración constante.

Finalmente, agradezco a mis colegas de la empresa, que me ayudaron a comprender la problemática e importancia de este tema. Su colaboración ha sido esencial para el éxito de este proyecto.

RESUMEN

Este proyecto aborda la mejora de la gobernanza y calidad de los datos en el área de Marketing. A través de un análisis estratégico y la implementación de herramientas para la evaluación, limpieza y monitoreo de los datos, se busca optimizar la toma de decisiones y la eficiencia operativa. La implementación de estos procesos ha demostrado un impacto significativo en la empresa, reduciendo costos y habilitando oportunidades estratégicas.

Palabras clave: gobernanza de datos, calidad de datos, análisis estratégico, automatización, eficiencia operativa.

ABSTRACT

This project addresses the improvement of data governance and quality in the Marketing area. Through strategic analysis and the implementation of tools for evaluation, cleaning, and monitoring of data, it seeks to optimize decision-making and operational efficiency. The implementation of these processes has demonstrated a significant impact on the company, reducing costs and enabling strategic opportunities.

Keywords: data governance, data quality, strategic analysis, automation, operational efficiency.

TABLA DE CONTENIDO

Dedicatoria	6
Agradecimientos	7
Resumen	8
Abstract	9
Introducción	12
Desarrollo del trabajo de titulación	13
CAPÍTULO 1. CONTEXTUALIZACIÓN	14
1.1. Antecedentes y Contexto	14
1.2. Justificación del Proyecto: Calidad de datos	16
1.3. Definición del problema	18
1.4. Objetivos	19
CAPÍTULO 2. ANÁLISIS ESTRATÉGICO de UDT	21
2.1. Valores de la compañía	21
2.2. Objetivos Estratégicos de Marketing en UDT	22
2.4. Análisis de la Estrategia de Analítica de Datos y Transformación Digital	23
CAPÍTULO 3. GENERACIÓN DE VALOR	27
3.1. Estrategias para la Generación de Valor a través de la Analítica de Datos	27
3.2. Identificación de Oportunidades de Valor	29
CAPÍTULO 4. MODELADO DE LA SOLUCIÓN	32
4.1. Investigación cualitativa	32
4.2. Análisis del Flujo de Datos en la Empresa	33
4.3. Características del Modelo Propuesto	37
4.4. Análisis Técnico y Especificaciones	40
4.5. Propuesta de Valor	43
4.6. Implementación y Prueba de Concepto	43
4.7. Análisis de Riesgos	45
4.8. Resultados	46
CAPÍTULO 5. ANÁLISIS FINANCIERO	48
5.1. Cálculo de Proyecciones de Ingresos y Egresos del Proyecto	48
5.2. Métodos de Evaluación	48
5.3. Punto de Equilibrio y Análisis de Sensibilidad	49
5.4. Resultados y Decisión Financiera	51
CAPÍTULO 6. CONCLUSIONES Y RECOMENDACIONES	54
6.1. Conclusiones	54
6.2. Recomendaciones	54
6.3. Impacto del Proyecto en la Empresa	55
Referencias	57

ÍNDICE DE FIGURAS

FIGURA 1. INFRAESTRUCTURA DE DATOS EN UDT.....	34
FIGURA 2. FLUJO DE DATOS ENTRE DEPARTAMENTOS.	35
FIGURA 3. FLUJO DE DATOS EN EL ÁREA DE MARKETING	36
FIGURA 4. INGESTA DE DATOS EN EL ÁREA DE MARKETING.	37
FIGURA 5. SISTEMA DE MEJORA DE CALIDAD DE INGESTA DE DATOS EN EL ÁREA DE MARKETING.....	38
FIGURA 6. TABLERO DE DIAGNÓSTICO INICIAL DE CALIDAD DE DATOS.....	39
FIGURA 7. TABLERO DE MONITOREO DE CALIDAD Y CONSISTENCIA DE DATOS EN EL CRM.....	40
FIGURA 8. TIEMPO DE PROCESAMIENTO Y CALIDAD DE LOS DATOS.....	44
FIGURA 9. ANÁLISIS DE SENSIBILIDAD	50

INTRODUCCIÓN

Este proyecto de maestría se centra en la gobernanza y calidad de los datos, utilizando a la empresa United Data Technologies (UDT) como estudio de caso. El objetivo principal es desarrollar herramientas para la evaluación de la calidad de los datos, la limpieza automatizada, la mejora y el monitoreo.

El primer capítulo detalla los problemas de operaciones ineficientes, decisiones basadas en datos poco fiables, recursos desperdiciados y daños a la reputación debido a la mala calidad de los datos. Se propone un análisis estratégico de la misión, visión, valores y objetivos comerciales de UDT, con un enfoque particular en el departamento de Marketing.

El segundo capítulo se centra en el análisis estratégico de la compañía, incluyendo la misión, visión, valores y los objetivos estratégicos del área de negocio. También se realiza un diagnóstico y análisis de la estrategia de analítica de datos y transformación digital.

El tercer capítulo presenta un análisis exhaustivo de las dimensiones de la calidad de los datos, la gobernanza y las estrategias de modernización para mejorar los procesos de ingesta de datos, particularmente en el área de marketing. Se propone un modelo que se centra en mejorar las dimensiones críticas de la calidad de los datos utilizando scripts automatizados y tableros de control de monitoreo.

El cuarto capítulo destaca la necesidad de procesos estandarizados, integración de inteligencia artificial, colaboración en equipo y automatización continua para optimizar el uso de recursos y mejorar la eficiencia operativa. El análisis financiero sugiere que el proyecto es viable.

El impacto del proyecto en la empresa es significativo, con una reducción de costos y las oportunidades estratégicas al proporcionar un sistema de mejora de la calidad de los datos.

DESARROLLO DEL TRABAJO DE TITULACIÓN

CAPÍTULO 1. CONTEXTUALIZACIÓN

1.1. Antecedentes y Contexto

La tendencia a la baja en la calidad de los datos es motivo de preocupación, especialmente cuando los datos se vuelven cada vez más vitales para la toma de decisiones; a medida que las empresas comienzan a depender más de la analítica predictiva y la inteligencia artificial, las posibles ramificaciones de una mala calidad de los datos se intensifican (Woodie, 2024). Esta disminución en la calidad de los datos a menudo se vincula con la creciente complejidad del panorama tecnológico, que abarca sistemas como la Gestión de Relaciones con el Cliente (CRM), la Gestión de la Cadena de Suministro (SCM) y los sistemas de Planificación de Recursos Empresariales (ERP) (Okoro, 2021).

Para abordar estos problemas de calidad de los datos, es esencial la gobernanza. La gobernanza de los datos no se trata solo de gestionar los datos en sí, sino también de gobernar a las personas y la tecnología que los moldean (Okoro, 2021). Sin esta gobernanza que asegura la calidad, las empresas corren el riesgo de perder oportunidades impulsadas por los datos y de desperdiciar recursos (Petzold, et al. 2020).

El procesamiento y la limpieza de los datos pueden consumir más de la mitad del tiempo de un equipo de análisis, incluyendo el de los científicos de datos altamente remunerados (Petzold, et al. 2020). Esto no solo limita la escalabilidad, sino que también lleva a la frustración de los empleados. Por lo tanto, la gobernanza de los datos efectiva es crucial para garantizar la calidad de los datos y para maximizar su potencial para impulsar las decisiones empresariales (Petzold, et al. 2020).

La gobernanza de datos, que sirve como la base para la toma de decisiones basada en datos, es fundamental para capturar valor a través de análisis, digitalización y otras

oportunidades transformadoras (Petzold, et al. 2020). Al implementar prácticas sólidas, las organizaciones pueden garantizar la calidad de los datos, reducir las inconsistencias y minimizar el riesgo de violaciones/fugas. No se trata solo de cumplir con los requisitos regulatorios como el GDPR (General Data Protection Regulation) y el CCPA (California Consumer Privacy Act) (CastorDoc, 2024), sino también de apoyar la toma de decisiones estratégicas al proporcionar acceso a datos precisos, relevantes, de alta calidad y confiables (Atlan, 2024).

Históricamente, la gobernanza de datos se ha asociado con el cumplimiento, el control y la mitigación de riesgos. Sin embargo, no debería ser una reflexión posterior o simplemente un conjunto de políticas y procedimientos implementados en respuesta a regulaciones o incidentes (Atlan, 2024). En cambio, debería estar integrado dentro de los flujos de trabajo diarios de los profesionales de datos, expandiendo su alcance verticalmente a través de la alineación con las partes interesadas del negocio, y horizontalmente a través del manejo interfuncional de los problemas de datos. Este enfoque mejora la adaptabilidad de las medidas de gobernanza de datos y mejora la comunicación entre los departamentos y los niveles de gestión (Orr, 1998).

Se incentiva a las empresas a pasar de una "mentalidad defensiva" a una "mentalidad de oportunidad", viendo la gobernanza de datos como una herramienta estratégica para mejorar el valor empresarial. Abordar los problemas de gobernanza de datos puede llevar a análisis más precisos y a una mejor toma de decisiones (Segner, 2023). Vale la pena señalar que la gobernanza de datos es una de las diferencias entre las empresas que capturan valor y las que no. Las empresas que han invertido poco en gobernanza exponen a sus organizaciones a un riesgo regulatorio real, que puede ser costoso (Petzold, et al. 2020). Por lo tanto, cada empresa

puede tener éxito al integrar estratégicamente la gobernanza de datos en la forma en que la organización trabaja todos los días (Petzold, et al. 2020).

1.2. Justificación del Proyecto: Calidad de datos

La mala calidad de los datos tiene implicaciones de gran alcance, un estudio de Gartner de 2021 estimó que les cuesta a las organizaciones un promedio de \$12.9 millones por año (Woodie, 2024). Además, la investigación de Gartner encontró que las organizaciones atribuyen pérdidas de ingresos a la mala calidad de los datos (Moore, 2018), con estimaciones que sugieren que impacta el 26% de los ingresos de sus empresas (Monte Carlo, 2023).

La incapacidad de los datos de mala calidad para proporcionar información basada en análisis sin un ajuste manual sustancial es una preocupación importante. La encuesta de MDM reveló que el 82% de los encuestados pasó uno o más días a la semana resolviendo problemas de calidad de los datos, con el 66% confiando en la revisión manual para evaluar, monitorear y gestionar sus datos maestros (Shaikh, et al. 2024). Esto tiene un impacto significativo en la productividad, con los empleados gastando un promedio del 30% de su tiempo total en la empresa en tareas que no añaden valor debido a la mala calidad y disponibilidad de los datos (Petzold, et al. 2020); mientras que otros pasan el 40% de su tiempo evaluando o comprobando la calidad de los datos (Monte Carlo, 2023).

Este problema se extiende a las grandes organizaciones multidivisionales, que pueden tener dificultades para generar eficientemente KPIs u otras métricas. Los problemas más prevalentes en la calidad de los datos de las organizaciones fueron la incompletitud, la inconsistencia y la inexactitud (Shaikh, et al. 2024). Muchas empresas también luchan con problemas de unicidad, o información duplicada, a través de sus sistemas (Woodie, 2024).

Las preocupaciones sobre la calidad de los datos son multifacéticas. Incluyen la dispersión de los datos a través de varios sistemas dentro de las organizaciones, la recopilación

y uso de los datos por varios niveles en la organización, y la falta de garantía de calidad de los datos en las metodologías de desarrollo de sistemas (Okoro, 2021). Sin embargo, la calidad de los datos no es la única preocupación. Los profesionales de los datos también se preocupan por la ambigüedad en la propiedad de los datos, la mala alfabetización de los datos, la integración de múltiples fuentes de datos y la documentación de los productos de datos (Woodie, 2024).

Los silos organizacionales complican aún más las cosas. El 80% de las organizaciones informaron que algunas de sus divisiones operan en silos, cada una con sus propios requisitos de gestión de datos, prácticas, sistemas de origen y comportamientos de consumo. Por ejemplo, un equipo de ventas puede mantener los datos del cliente en un sistema de gestión de relaciones con el cliente (CRM), mientras que un equipo de marketing puede usar una plataforma de datos del cliente (CDP) para crear perfiles de clientes e informar sobre las campañas publicitarias. Estos silos pueden llevar a inconsistencias y errores, dificultando la toma de decisiones relacionadas con el negocio, los datos y la tecnología (Shaikh, et al. 2024).

En resumen, la mala calidad de los datos es un problema costoso que afecta a la productividad y a los ingresos de las organizaciones. Los problemas comunes incluyen la incompletitud, la inconsistencia y la inexactitud de los datos, así como la información duplicada. Además, la dispersión de los datos a través de varios sistemas y la falta de garantía de calidad de los datos en las metodologías de desarrollo de sistemas son preocupaciones adicionales. Los silos organizacionales, que resultan en requisitos de gestión de datos divergentes y comportamientos de consumo, complican aún más la situación. Estos desafíos subrayan la necesidad de una gestión de datos eficaz y de estrategias de mejora de la calidad de los datos.

1.3. Definición del problema

En la era de la transformación digital, los datos han surgido como la nueva moneda. No se trata solo de recopilar datos, sino de desbloquear su poder para impulsar mejores resultados empresariales, mejorar la toma de decisiones y proporcionar nuevos conocimientos (Microsoft, 2023). Los datos se utilizan para mejorar los productos y servicios organizacionales, fortalecer el servicio al cliente y respaldar decisiones estratégicas e inteligencia empresarial (Abraham, et al. 2019).

Los datos se utilizan para mejorar los productos y servicios organizacionales, fortalecer el servicio al cliente y respaldar decisiones estratégicas e inteligencia empresarial (Abraham, et al. 2019). El valor comercial de los datos es directamente proporcional a su integridad. Los altos niveles de integridad de los datos comienzan con una buena calidad de datos, riqueza contextual e integración, todos los cuales están respaldados por herramientas y procesos sólidos de gobernanza de datos (Levy, 2024).

El valor de los activos de datos depende de la calidad de los datos de origen. Un aumento en la productividad de los datos es el resultado de una inteligencia empresarial efectiva (Okoro, 2021), que conduce a la toma de decisiones informada. Sin embargo, una organización no puede ser impulsada por datos, y un equipo de datos no puede agregar valor si los datos no son confiables. Esta falta de confianza puede tensar las relaciones con los interesados internos y, en casos donde los datos son parte del producto, puede impactar la satisfacción del cliente y los resultados finales. Por lo tanto, mantener la confianza en los datos es primordial para cualquier organización (Monte Carlo, 2023).

La empresa por analizar, en este estudio, enfrenta problemas significativos relacionados con la calidad y consistencia de los datos. Estos problemas están afectando negativamente las

operaciones y decisiones estratégicas de la empresa. A continuación, se detallan los problemas específicos:

Operaciones ineficientes: Los datos incorrectos y duplicados están causando ineficiencias en los procesos operativos de ingesta de datos.

Decisiones basadas en datos no confiables: Las decisiones estratégicas de un área de negocio pueden ser tardías y erróneas debido a la falta de confianza en los datos.

Recursos y tiempo malgastados: Se está destinando una gran cantidad de tiempo y recursos a la corrección manual de datos.

Reputación afectada: La mala calidad de los datos puede dañar la reputación de la empresa frente a clientes y stakeholders.

1.4. Objetivos

- **Objetivo General**

El objetivo general es desarrollar herramientas para evaluar la calidad de los datos, automatizar su limpieza, mejorarlos y monitorear su calidad. El proyecto se centrará en la calidad de los datos y la automatización de la ingesta de datos, utilizando como estudio de caso los datos que el equipo de marketing ingresa en el CRM.

- **Objetivos Específicos**

Limpieza y Mejora Automática de los Datos: Desarrollar y aplicar flujos de trabajo automatizados para la limpieza y transformación de datos basados en reglas y estándares predefinidos, capaces de corregir automáticamente los problemas de calidad de datos identificados durante la evaluación.

Visualización de la Calidad de los Datos: Diseñar y crear un tablero que visualice las métricas clave de calidad y consistencia de los datos, capaz de demostrar las mejoras logradas

en la calidad de los datos como resultado de la implementación del sistema integral de gestión de la calidad de los datos.

CAPÍTULO 2. ANÁLISIS ESTRATÉGICO DE UDT

2.1. Valores de la compañía

Como proveedor de soluciones tecnológicas, United Data Technologies (UDT) moderniza, asegura y gestiona sistemas de Tecnología de la Información (TI) completos para empresas comerciales, gobiernos estatales y locales, y organizaciones educativas.

La oferta de la empresa incluye servicios administrados de tecnologías de la información (TI) integrados y servicios de ciclo de vida, soluciones de ciberseguridad, informática y conectividad. Todo con un enfoque singular: empoderar a sus clientes para que tomen decisiones de TI basadas en información que impulsan sus prioridades digitales más importantes.

Reconocida como una de las 50 empresas de TI de más rápido crecimiento en los Estados Unidos, UDT es un proveedor líder de soluciones tecnológicas de alto rendimiento en el sureste y medio oeste de los Estados Unidos.

Dedicada a ayudar a sus clientes a maximizar sus inversiones en TI, la empresa reduce los riesgos, costos y complejidad de desplegar tecnología de próxima generación. Para proporcionar los productos más recientes, desde computación personal, infraestructura audiovisual hasta soluciones personalizadas, UDT se asocia con los principales proveedores de tecnología y marcas de la industria.

Establecida en 1995 y con sede en Florida, la empresa cuenta con oficinas principales y equipos de soluciones en cuatro estados. Con más de 300 empleados, genera más de 300 millones de dólares en ingresos anuales. Posee dos centros logísticos y ha desarrollado subproductos como Havrion y Conquest (recientemente adquirida por Blueboyant).

La misión y el propósito de la empresa se resumen en: “Proteger a nuestros clientes, sus sistemas, sus usuarios y las comunidades a su alrededor. Empoderar a nuestros clientes con soluciones tecnológicas e insights para lograr sus prioridades más importantes”.

2.2. Objetivos Estratégicos de Marketing en UDT

El área de negocio, dentro de la empresa, que va a ser el foco del presente trabajo, es Marketing. Dentro de esta área, las prioridades, u objetivos estratégicos, planteados para el 2024 son los siguientes:

1. Optimizar la Generación de Demanda de Negocios por Territorio: Entregar resultados de generación de demanda que apoyen los objetivos de ventas regionales, enfocando la estrategia y ejecución de Marketing al modelo de ventas por Territorio.
2. Demostrar un ROI de Marketing Medible: Para alcanzar sus objetivos de crecimiento, UDT necesita aprovechar el poder del marketing como un multiplicador de fuerza. Sin embargo, esto requiere un aumento en la inversión, que a su vez exige evidencia de un retorno de inversión (ROI) predecible a través de informes bien documentados. Esto implica la planificación estratégica de programas y campañas dentro del sistema CRM para vincular sistemáticamente los esfuerzos con los resultados. Además, es esencial seguir el progreso de estos programas y campañas desde el punto de generación de leads hasta el cierre de negocios.
3. Activar la Iniciativa de Crecimiento Estratégico: Ejecutar una estrategia de marketing que no solo promueva los productos o servicios de UDT, sino que también fomente una conexión más profunda con la audiencia. Aprovechando la demanda estratégica de nuestros proveedores y marcas tecnológicas, mientras se distingue la marca UDT.

4. Reposicionar UDT como un Proveedor de Soluciones Tecnológicas: Establecer la plataforma de marketing y los programas que lanzarán con éxito un negocio de proveedor de servicios de “telecomunicaciones” de \$100 millones.
5. Lanzar Soluciones de Telecomunicaciones: Establecer la posición de UDT como un Proveedor de Soluciones Tecnológicas, implementando una estrategia de marca que refuerce lo que es exclusivo de UDT y lo posicione en las mentes de los clientes.

En el presente trabajo se centrará en el objetivo empresarial dentro del área de Marketing de demostrar un Retorno de Inversión (ROI) de marketing medible. Este enfoque se alinea con el objetivo estratégico de la empresa, que requiere de una planificación estratégica de programas y campañas dentro del sistema CRM y una vinculación sistemática de los esfuerzos con los resultados. Al mejorar la calidad de los datos en el CRM, este proyecto contribuirá a la planificación y ejecución de programas y campañas de marketing más efectivos. Además, la creación de tableros para la generación de informes proporcionará a la empresa la evidencia documentada que necesita para demostrar un ROI predecible.

2.4. Análisis de la Estrategia de Analítica de Datos y Transformación Digital

Actualmente, los objetivos de la gobernanza de datos en UDT se centran en establecer métodos, responsabilidades y procesos para estandarizar, integrar, proteger y almacenar datos corporativos. La meta es proporcionar información confiable a las personas adecuadas, en el momento correcto y en un formato fácilmente consumible. Esto implica minimizar riesgos, establecer reglas internas para el uso de datos, cumplir con los requisitos normativos, mejorar la comunicación tanto interna como externa, incrementar el valor de los datos, facilitar su administración, reducir costos y garantizar la continuidad de la empresa mediante la gestión de riesgos y optimización.

El plan actual en marcha en UDT incluye objetivos estratégicos alineados con las iniciativas corporativas, OKR (Objetivos y Resultados Clave), y los KTLO (Mantener Las Luces Encendidas) del equipo EDP (Plataforma de Datos Empresariales). Han establecido un Azure Data Warehouse que alberga el 99% de sus datos. La información está segura y el acceso a los datos se proporciona únicamente a los miembros del equipo que presentan solicitudes formales y tienen los derechos correspondientes. Han implementado las Mejores Prácticas de Gobernanza de Datos, que incluyen flujos de trabajo y SOPs (Procedimientos Operativos Estándar), gestión de datos, cumplimiento de datos y soporte para todas las áreas de la empresa. El 95% de los datos utilizados para los negocios se almacenan en la nube, mientras que el 5% restante se mantiene en servidores de datos locales. La información analítica se gestiona con permisos, y la Gobernanza de Datos MDM (Gestión de Datos Maestros) está en funcionamiento.

Los objetivos futuros de UDT incluyen la integración segura de Azure AD (Active Directory) en todos los servidores, datos gobernados federados y accesibles a todas las Líneas de Negocio, mejora de la ingestión de datos de terceros, y la implementación de una Calificación de Veracidad de Datos para clasificar la calidad y confiabilidad de los activos de datos utilizados. Este sistema de calificación de veracidad de datos clasifica los datos en función de su calidad y confiabilidad, proporcionando una medida cuantitativa de la utilidad de los datos.

La hoja de ruta de UDT incluye entender los activos de datos, federar datos para mejorar la usabilidad, implementar Procedimientos Operativos Estándar, utilizar Inteligencia Artificial para analizar datos e implementar Microsoft Purview. Su plan de análisis incluye depreciar los informes redundantes o no utilizados, mejorar las métricas de uso, evangelizar los informes por

Línea de Negocio, e integrar análisis predictivos y prescriptivos, así como Inteligencia Artificial generativa, como ChatGPT.

La evangelización de los informes en UDT busca utilizar los activos de datos para complementar las iniciativas estratégicas de toda la empresa, ya sean objetivos inmediatos o a largo plazo. Su objetivo es proporcionar datos curados, bien definidos y administrados que se consideren como “Registro de Oro” y que estén disponibles a través de una plataforma empresarial. Esto proporcionará una fuente común y única de verdad que ofrece accesibilidad a los datos para las soluciones empresariales y los sistemas de informes, impulsando los ingresos y la rentabilidad, reduciendo costos, permitiendo que los activos de datos vitales sean medidos, construyendo un léxico o diccionario de datos, empoderando decisiones clave y habilitando oportunidades estratégicas para oportunidades inmediatas y futuras.

Los impulsores estratégicos de UDT incluyen un enfoque unificado para recopilar datos en todas las plataformas y entregar una instancia curada y normalizada que pueda ser federada para la empresa. Esto permite la gobernanza y administración de todos los activos de datos, estableciendo confianza en todos los datos y acelerando la precisión de los informes y el tiempo de comercialización para la entrega de informes.

Basándose en la información proporcionada, UDT parece tener un enfoque estructurado y avanzado en varios aspectos de la gobernanza de datos. El establecimiento de métodos y procesos para estandarizar, integrar, proteger y almacenar datos indica un marco organizativo en marcha, lo cual es un signo de madurez con respecto a la gobernanza de datos.

El objetivo de proporcionar información confiable a las personas correctas en el momento adecuado y en un formato fácilmente consumible sugiere que en UDT se han establecido mecanismos para asegurar la calidad y accesibilidad de los datos. Además, el enfoque en minimizar riesgos, establecer reglas internas para el uso de datos e implementar

requisitos de cumplimiento muestra una preocupación por la seguridad y la conformidad normativa.

La intención de mejorar la comunicación interna y externa y aumentar el valor de los datos indica que UDT entiende la importancia de los datos como activos estratégicos. Los objetivos de facilitar la administración y reducir costos reflejan una intención de optimizar la eficiencia operativa y gestionar los recursos de manera efectiva. Finalmente, el enfoque en la gestión de riesgos y la optimización para la existencia continua sugiere un compromiso con la sostenibilidad y la resiliencia de la empresa.

Dado estos puntos, UDT parece estar en un nivel de madurez intermedio a avanzado en la gobernanza de datos. Sin embargo, una evaluación detallada con una herramienta específica sería necesaria para calificar su nivel exacto.

CAPÍTULO 3. GENERACIÓN DE VALOR

3.1. Estrategias para la Generación de Valor a través de la Analítica de Datos

La gobernanza de datos es la gestión de los datos como un activo estratégico, que implica controles en torno a su contenido, estructura, uso y seguridad. Requiere conocimiento de los datos existentes, su calidad, usabilidad, acceso y patrones de uso, asegurando que todos los casos de uso sean seguros, cumplan con las normativas y están gobernados (Microsoft, 2023). Esta gobernanza se logra a través de un conjunto de roles, procesos, políticas y herramientas que garantizan la calidad de los datos a lo largo del ciclo de vida de los datos y el uso adecuado de los datos en toda la organización. Esto permite a los usuarios encontrar, preparar, usar y compartir conjuntos de datos de confianza de manera independiente (Qlik, 2024).

El beneficio principal de la gobernanza de datos es la provisión de datos de alta calidad necesarios para las herramientas de análisis de datos e inteligencia empresarial. Estas herramientas ofrecen perspectivas que conducen a mejores decisiones empresariales y mejor rendimiento. Los beneficios adicionales incluyen la mejora de la precisión, la integridad y la consistencia de los datos (Qlik, 2024).

La gestión de la calidad de los datos, según la define Asociación Internacional de Gestión de Datos (DAMA) Internacional, es una función de medición, evaluación, mejora y garantía de la aptitud de los datos para su uso. Es un subnivel de la gestión de datos que incluye la planificación, el control y la provisión de activos de datos de alta calidad. La responsabilidad de manejar y mejorar la calidad de los datos a menudo recae en los departamentos de TI, aunque algunas empresas implementan estrategias de gestión de datos más amplias. En el contexto de

la gobernanza de datos, la gestión de la calidad de los datos involucra tanto a profesionales de negocios como de TI (Okoro, 2021).

La calidad de los datos es un aspecto vital de la gobernanza de datos. Se refiere a la integridad, corrección, unicidad, integridad referencial y consistencia en todos los dominios de datos (Okoro, 2021). La aptitud de los datos para su uso se alinea directamente con las seis dimensiones de la calidad de los datos: precisión, completitud, consistencia, oportunidad, unicidad y validez (DAMA, 2017). Las dimensiones de calidad de los datos son características medibles que ayudan a definir los requisitos de calidad de los datos y se utilizan para evaluar la calidad de los datos. DAMA ha identificado seis dimensiones principales: Precisión (los valores de los datos deben estar lo más cerca posible de los valores reales), Completitud (todos los valores de datos requeridos deben estar presentes), Consistencia (los valores de datos dentro de una columna deben cumplir con una regla), Oportunidad (los datos deben representar la realidad desde un punto en el tiempo requerido), Unicidad (los valores distintos deben aparecer solo una vez), y Validez (los datos deben ajustarse al formato, tipo o rango de su definición) (IBM, 2024).

Estas dimensiones son atributos medibles de los datos que pueden ser evaluados, interpretados y mejorados individualmente. Las puntuaciones agregadas de múltiples dimensiones representan la calidad de los datos en un contexto específico e indican la aptitud de los datos para su uso (Collibra, 2023). Estas dimensiones ayudan a definir los requisitos de calidad de los datos y a determinar los resultados esperados de las evaluaciones de calidad de los datos, ya sean iniciales o continuas (IBM, 2024).

Medir las dimensiones de la calidad de los datos ayuda a identificar oportunidades para mejorar la calidad de los datos. Las reglas de calidad de los datos pueden ser utilizadas para determinar si tus datos son aptos para su uso y dónde necesitan mejoras. Estas reglas aseguran

que los datos representen de manera precisa, completa y consistente la entidad del mundo real. Las reglas automatizadas ayudan a identificar rápidamente los errores de los datos y proporcionan una actualización constante sobre el estado de la salud de los datos (Collibra, 2023).

3.2. Identificación de Oportunidades de Valor

El volumen de datos está expandiéndose a un ritmo que supera la capacidad humana para su gestión manual (Atlan, 2023). Por ejemplo, entre 2020 y 2022, el volumen total de datos aumentó de 1 petabyte a más de 2 petabytes. Además, una empresa promedio gestiona alrededor de 400 fuentes de datos (Atlan, 2023).

Históricamente, la gobernanza se ha centrado en la gestión de datos 'finalizados', como métricas financieras, presentaciones regulatorias e indicadores clave de rendimiento, que requieren definiciones formales y alta calidad de datos. Sin embargo, la ciencia y análisis de datos modernos a menudo utilizan datos brutos y semiprocesados. Este cambio ha llevado a una tensión entre los proveedores de datos, que se esfuerzan por proporcionar datos de manera responsable sin poner en peligro el negocio, y los consumidores de datos, que demandan acceso inmediato a los datos para sus proyectos (Qlik, 2024).

La gobernanza de datos moderna tiene como objetivo hacer que los flujos de datos sean rastreables y los procesos relacionados con los datos sean transparentes, lo que permite una mejor comprensión de las operaciones, la mejora del rendimiento y el logro de objetivos (Atlan, 2024). Sin embargo, el proceso ETL (Extract, Transform, Load) puede ser lento, particularmente la fase de transformación que implica la limpieza y estandarización de datos, lo que puede crear cuellos de botella. Para superar estos desafíos, se están empleando soluciones como flujos de trabajo automatizados y herramientas de inteligencia artificial (IA) para acelerar el proceso ETL y minimizar los errores humanos. Otra estrategia es la técnica ELT, donde se

retrasa la fase de transformación, permitiendo que los datos brutos se carguen en un destino y se transformen más tarde (Atlan, 2022).

Establecer un proceso de ingesta de datos robusto es crucial para una fuente de datos efectiva. Las organizaciones están pasando de procesos manuales laboriosos a herramientas de alta tecnología para operaciones más rápidas y con menos errores. Un adecuado catálogo y gobernanza de la ingesta de datos pueden prevenir o mitigar los cuellos de botella en la limpieza, mejorando la eficiencia (Atlan, 2022).

La automatización en la gobernanza de datos reemplaza los procesos manuales, que son difíciles de escalar y propensos a errores, con procesos automatizados, sostenibles y reproducibles. Esto es esencial para escalar la gobernanza de datos para satisfacer las crecientes demandas de datos (Atlan, 2023). Por ejemplo, Tide, un banco digital con sede en el Reino Unido redujo el tiempo requerido para identificar toda la data sensible en su sistema de 50 días laborables a solo cinco horas utilizando la gobernanza de datos automatizada (Atlan, 2023).

Mientras que muchas organizaciones adoptan un enfoque holístico para la gobernanza de datos, considerando todos los activos de datos simultáneamente, este amplio alcance puede ralentizar el progreso y correr el riesgo de desconectar los esfuerzos de las necesidades del negocio. Para tener éxito, los activos de datos deben ser priorizados por dominios y por datos dentro de cada dominio (Petzold, et al. 2020). Las organizaciones pueden comenzar a integrar herramientas de gestión de datos probando en un dominio, lo que les permite validar el diseño, el modelo de gobernanza y los flujos de trabajo en un entorno controlado. Esto también proporciona una oportunidad para recopilar comentarios de usuarios y partes interesadas, identificar áreas de mejora y optimizar la herramienta de gestión de datos y los flujos de trabajo (Shaikh, et al. 2024).

Es importante habilitar casos de uso prioritarios rápidamente, incluso si la solución no es perfecta. El desarrollo a largo plazo para hacer que los casos de uso estén listos para producción puede proceder una vez que se ha demostrado el valor. Por ejemplo, mejorar las campañas de clientes puede no requerir un conjunto de datos completamente integrado en toda la empresa, sino más bien un enfoque personalizado en una plataforma dedicada. La gobernanza de datos debe apoyar y acelerar este enfoque personalizado, centrándose en resolver problemas de disponibilidad y calidad, además de establecer una sólida gestión de datos maestros (Petzold, et al. 2020).

CAPÍTULO 4. MODELADO DE LA SOLUCIÓN

4.1. Investigación cualitativa

Se llevaron a cabo cuatro entrevistas clave en la empresa, el Chief Digital Information Officer, el Director de Datos, el VP de Marketing, y el Gerente de Contenido de Marketing, para entender los desafíos prevalentes relacionados con la gestión de datos y los procesos comerciales en UDT. Las entrevistas revelaron temas comunes que destacaron la necesidad de una gestión de datos efectiva, incluyendo la mejora del intercambio de conocimientos, la estandarización de procesos, la integración de herramientas, y la calidad de los datos. Se enfatizó la consistencia y la creación de una plataforma unificada de datos empresariales, así como la necesidad de procesos estandarizados y un repositorio compartido de información. La integración de varios flujos de datos y herramientas utilizadas por el equipo surgió como una preocupación común. Además, se resaltó el potencial de la IA para mejorar la eficiencia operativa, la productividad y la toma de decisiones estratégicas.

En cuanto a las ideas clave compartidas entre los entrevistados, se enfatizó que mejorar el intercambio de conocimientos entre los equipos puede asegurar la consistencia en la marca y la terminología. Se destacó la importancia de una cultura orientada a los datos y el papel de la IA en las operaciones. La colaboración dentro y entre los equipos es crucial para mejorar los procesos de datos y crear tableros de control efectivos. La responsabilidad de los datos dentro de la organización surgió como un punto clave. Además, las entrevistas también subrayaron la necesidad de asignación de recursos y eficiencia, con discusiones en torno a la necesidad de asignar recursos de manera efectiva y mejorar la eficiencia operativa a través de la automatización y mejores prácticas de gestión.

Las posibles causas de los problemas actuales son que los datos provienen de diferentes formatos y herramientas dispares sin estandarización adecuada para su ingreso. El uso de múltiples herramientas no integradas totalmente dificulta la gestión y limpieza de datos. Las políticas y procedimientos de gobernanza de datos son insuficientes y no están socializados entre toda la organización.

Las conclusiones obtenidas son que abordar los problemas de calidad y consistencia de los datos es crítico para asegurar información confiable y operaciones eficientes. Integrar las herramientas y sistemas existentes para crear un ambiente de trabajo cohesivo puede mejorar significativamente la productividad. Establecer mecanismos robustos de intercambio de conocimientos puede prevenir inconsistencias y mejorar la alineación del equipo. Implementar procesos estandarizados puede agilizar las operaciones y proporcionar una dirección clara a los equipos. Ampliar el uso de la IA puede llevar a una mejora de la productividad y una ventaja estratégica. Mejorar la colaboración en equipo y reducir los silos puede llevar a una mejor alineación y un trabajo en equipo más efectivo. Buscar continuamente oportunidades de automatización puede eliminar ineficiencias y optimizar el uso de recursos.

4.2. Análisis del Flujo de Datos en la Empresa

Los datos son una mercancía indispensable para las empresas. Los ingenieros de datos construyen procesos para transportar datos desde su origen hasta su destino, como un data warehouse o un data lake. Estos datos se utilizan para alimentar tableros de análisis, impulsar la automatización con IA, entrenar modelos de aprendizaje automático y desarrollar productos de datos (Woodie, 2023).

Un data warehouse es un sistema de almacenamiento diseñado para contener grandes volúmenes de datos estructurados de diversas fuentes. Por otro lado, un data lake es un sistema de almacenamiento diseñado para contener grandes volúmenes de datos crudos, ya sean

estructurados, no estructurados o semiestructurados. A pesar de ser rentables, pueden desorganizarse rápidamente debido a la gran cantidad de datos que pueden contener. Un data mart es un repositorio de datos altamente estructurado. Sin embargo, a diferencia de un data warehouse que podría almacenar todos los datos de una organización, un data mart almacena datos específicos de un solo tema, como datos financieros, de ventas o de marketing (Atlas, 2022).

A nivel general (Figura 1), la infraestructura para el manejo de datos desarrollada dentro de UDT consta de un data lake para el almacenamiento de datos no transformados, un data warehouse para el almacenamiento y un data mart para el desarrollo de APIs (Application Programming Interfaces) y la analítica e inteligencia de negocios. El flujo de datos va desde la ingesta de diferentes fuentes hacia la generación de insights o el desarrollo de herramientas propias.

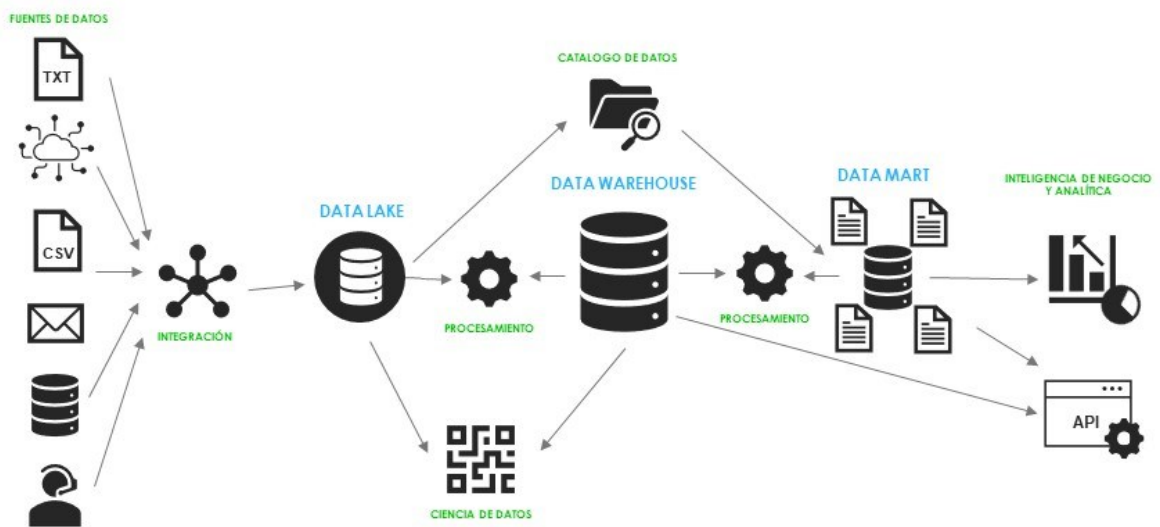


Figura 1. Infraestructura de datos en UDT

La Figura 2 ilustra el flujo de datos dentro de los diferentes departamentos de la empresa, como Marketing, Ventas, Experiencia del Cliente (CX) y Tecnología de la Información (IT). Los datos viajan desde diferentes fuentes de marketing hacia el CRM,

mostrando una integración entre la automatización de marketing y la gestión de relaciones con los clientes. Posteriormente, los datos fluyen a través de la Plataforma de Datos Empresarial (EDP) bajo la gestión de IT, para finalmente ser utilizados en la analítica e inteligencia de negocios empresarial. De manera paralela, Marketing maneja sus propias analíticas para obtener insights valiosos para mejorar la toma de decisiones en sus campañas.

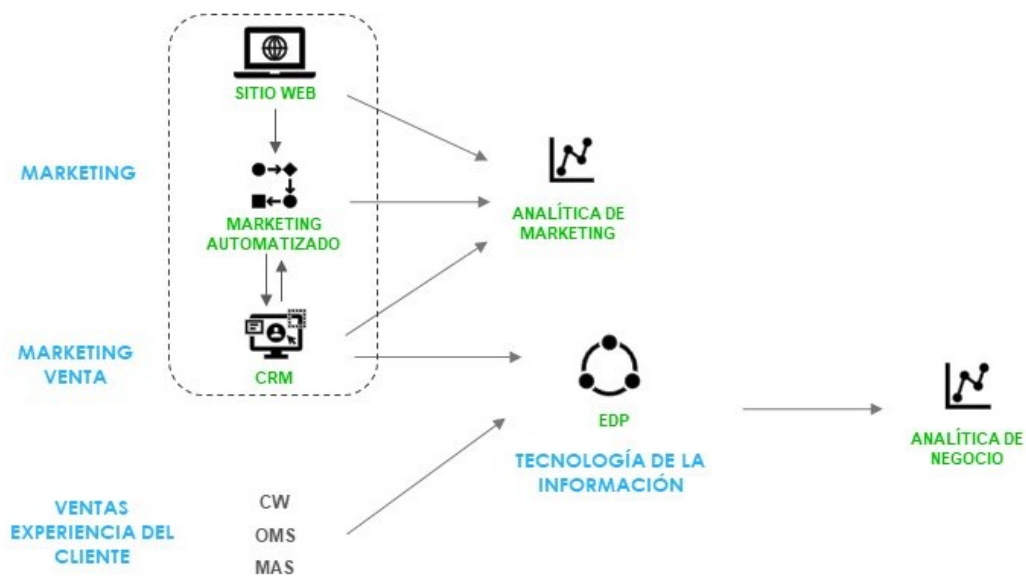


Figura 2. Flujo de datos entre departamentos.

En el caso del área de Marketing (Figura 3), los datos son recolectados de eventos y campañas de marketing, procesados e ingresados al CRM. Luego, estos datos se utilizan en varias estrategias de marketing como email marketing y anuncios pagados, así como en la Analítica de Marketing y en la generación de reportes.



Figura 3. Flujo de datos en el área de Marketing

La ingesta de datos es el proceso de mover datos brutos desde una fuente a un destino donde se pueden almacenar o utilizar para consultas y análisis (Atlan, 2022). El proceso actual para la ingesta de datos en el área de marketing se puede visualizar en la Figura 4. La información que se recibe se limpia manualmente para eliminar cualquier inconsistencia y comprobar su validez. Posteriormente, se utiliza una herramienta para completar la información faltante en los registros. La información se vuelve a limpiar y se la coloca en el formato para la ingesta en el CRM. Dependiendo de la situación, se añaden pasos previos para evitar información duplicada comparándola con la ya existe en el CRM. En términos generales, este proceso requiere 3 horas al día para su completitud, donde la limpieza y el completar la información representan el 62% de ese tiempo. Sin embargo, este tiempo puede variar significativamente dependiendo de la cantidad de registros que se necesiten ingresar al sistema, lo cual puede comprometer la calidad de los datos en función del tiempo disponible para su limpieza y revisión.

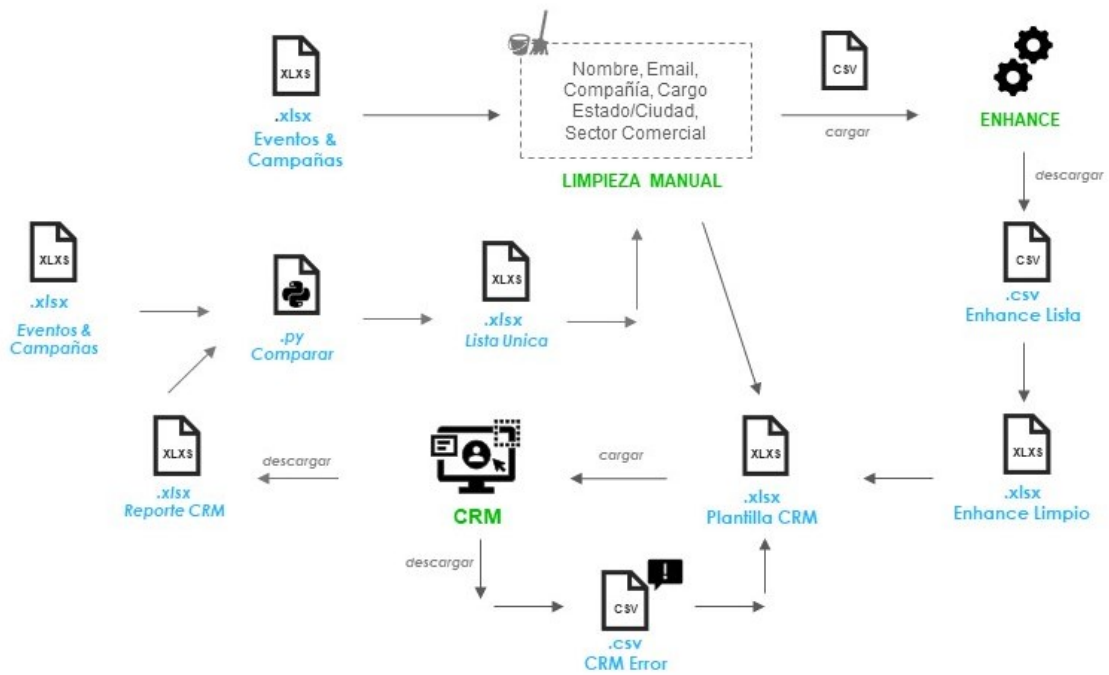


Figura 4. Ingesta de datos en el área de Marketing.

4.3. Características del Modelo Propuesto

Tomando como referencia el proceso desarrollado por Dropbox para su proceso de validación de datos (Sanko, 2023), con la idea de cubrir la mayor parte de los problemas de calidad de datos con un par de mejoras en el proceso. Este modelo se enfoca en mejorar las dimensiones de calidad de datos más importantes para este proceso: completitud (la cantidad de datos que están presentes), precisión (la corrección de los datos), consistencia (la coherencia de los datos) y validez (la conformidad de los datos con el formato correcto). En la Figura 5 se muestra el sistema de mejora de calidad de ingesta de datos, que consta de dos pasos.

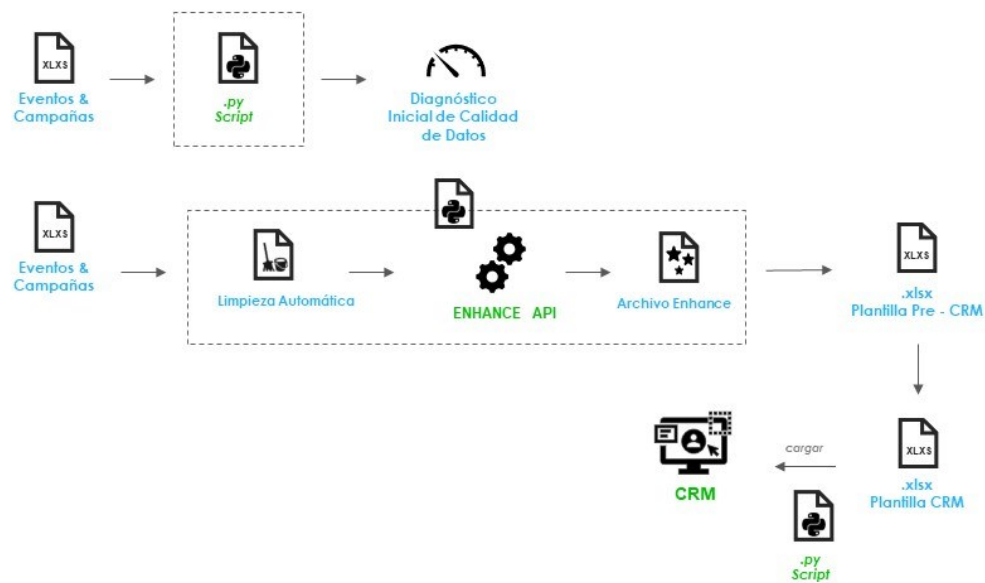


Figura 5. Sistema de mejora de calidad de ingesta de datos en el área de Marketing.

En el primer paso, la información que se recibe se analiza en un tablero con métricas sobre la calidad de los datos para valorar su estado y generar un reporte inicial con recomendaciones para su mejora (Figura 6). En el segundo paso, la información se limpia de manera automatizada y luego, mediante una API, se completa la información faltante. Se genera un archivo compatible con la plantilla para la ingesta en el CRM.

Finalmente, la información se transfiere a la plantilla de ingesta del CRM. Esta información es entonces revisada por una persona que busca errores y problemas de formato. Paralelamente, un código automatizado verifica la completitud de la información y la ausencia de errores técnicos. Una vez que la información ha sido revisada y validada tanto manual como automáticamente, se procede a cargarla en el CRM.

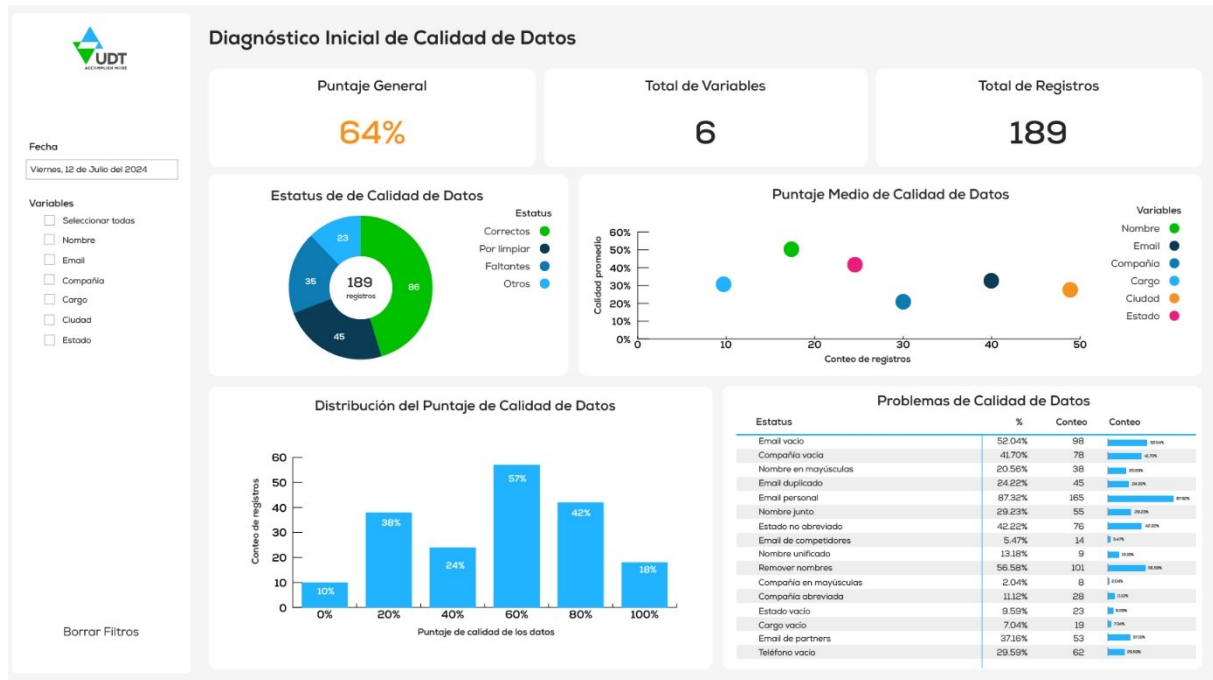


Figura 6. Tablero de Diagnóstico Inicial de Calidad de Datos.

Adicionalmente, otro tablero monitoriza la calidad y consistencia de los datos en el CRM (Figura 7), permitiendo generar reportes periódicos que proporcionan una visión general de la calidad de los datos y destacan cualquier problema de consistencia.

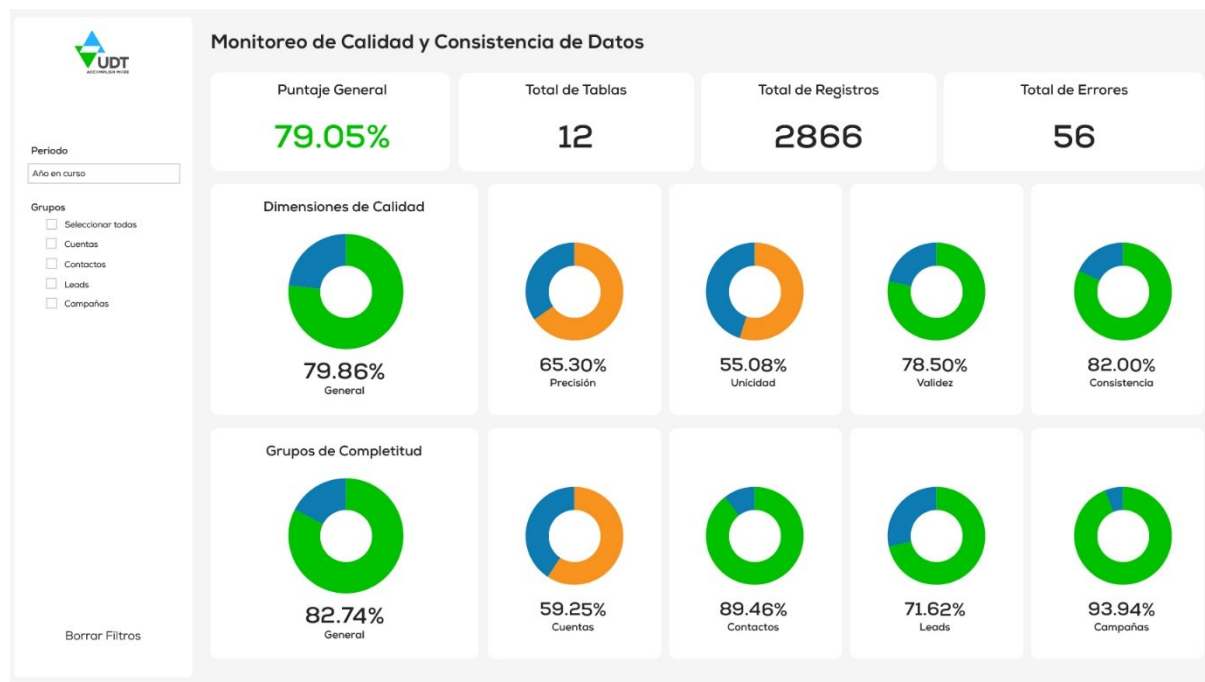


Figura 7. Tablero de Monitoreo de Calidad y Consistencia de Datos en el CRM.

El sistema propuesto proporciona una perspectiva clara de cómo abordar de manera sistemática la disminución de tiempos y el aumento de la calidad de los datos, lo cual es esencial para la toma de decisiones basadas en información confiable dentro del área de Marketing.

4.4. Análisis Técnico y Especificaciones

- **Diseño del Modelo y flujo de trabajo**

Inicialmente, se identificaron las variables obligatorias para que la información recibida tenga validez en el negocio y pueda ser ingresada al CRM. Se determinaron qué variables son complementarias, lo que permite entender y segmentar cada registro de manera más efectiva. Finalmente, se establecieron qué variables son específicas para el CRM y deben completarse para tener sentido y relación dentro de los sistemas del CRM.

El modelo se divide en cinco entregables que, en conjunto, forman el sistema de mejora de la calidad de la ingesta de datos:

- 1. Tablero de Diagnóstico Inicial de Calidad de Datos:** Este tablero evalúa la calidad de los datos recibidos mediante métricas sobre la completitud y precisión de las variables obligatorias y complementarias. Se muestra la distribución de los registros por su calidad de datos para cada variable, se presentan recuentos de registros que están limpios y se enumeran tipos de errores junto con sus respectivos porcentajes y recuentos. Esto permite generar un informe inicial sobre el estado de los datos y las recomendaciones que se deben seguir para mejorarlos.
- 2. Script de Depuración Automática de Datos:** Este código realiza una limpieza de datos en un archivo Excel, eliminando filas con correos electrónicos de dominios específicos y filas duplicadas, separando los nombres en diferentes columnas en nombre y apellido, capitalizando ciertas columnas, formateando números de teléfono y abreviando nombres de estados. Finalmente, se guardan los datos limpios en un nuevo archivo Excel.
- 3. Script de Autocompletado de datos:** Este código analiza un archivo Excel y extrae ciertas columnas basándose en un mapeo cargado desde un archivo de configuración, lo que permite flexibilidad para diferentes conjuntos de datos. Los datos extraídos se convierten en una cadena JSON y se envían en lotes a la Enrich API de Zoominfo para completarlos. Tras la solicitud a la API, se procesa la respuesta, se aplanan los datos anidados, se combinan los datos de entrada con los enriquecidos y se guardan en un nuevo archivo Excel. Si la respuesta de la API no contiene datos, se imprime un mensaje de error. Este proceso se realiza en lotes de 25 registros para manejar eficientemente grandes conjuntos de datos.
- 4. Script de Validación de Datos:** Este código realiza un análisis de un archivo Excel en el directorio actual. Se busca el archivo, se carga en un DataFrame de pandas y

luego se verifica cada columna en busca de celdas vacías y fórmulas. Al final, se imprimen los nombres de las columnas vacías y las que contienen fórmulas, si las hay, y se muestra el número de columnas y filas llenas en el DataFrame. Las columnas y filas completas se definen como aquellas que no están completamente vacías.

- 5. Tablero de Monitoreo de Calidad y Consistencia de Datos en el CRM:** Este tablero monitorea la calidad y consistencia de los datos en el CRM, utilizando las dimensiones de la calidad de datos. Se muestra un puntaje general de las dimensiones, así como de los diferentes sistemas dentro del CRM para el periodo seleccionado. La parte superior presenta algunos KPIs para entender la procedencia de la información. La parte central del tablero presenta indicadores que representan diferentes dimensiones de calidad como Precisión, Completitud, Consistencia, Oportunidad, Unicidad y Validez con porcentajes que indican sus respectivas puntuaciones. El esquema de colores representa diferentes niveles o umbrales de calidad de datos.

- **Herramientas y Tecnologías Utilizadas**

Los scripts fueron creados en Python 3.8, utilizando las librerías de pandas (para la manipulación y análisis de datos), openpyxl (para leer y escribir archivos Excel), json (para trabajar con datos en formato JSON), http.client (para hacer solicitudes HTTP), zi_api_auth_client (para autenticar las solicitudes a la API de Zoominfo) y os (para interactuar con el sistema operativo). El script de Autocompletado de datos utiliza el Enhance API de Zoominfo (una API que proporciona datos enriquecidos sobre empresas y contactos) mediante una autenticación con una API key. Los tableros fueron construidos en Power BI usando como fuente de datos un archivo Excel (para visualizar y analizar los datos) o una conexión directa al

data warehouse de UDT (para acceder a grandes volúmenes de datos almacenados de manera eficiente). Estas herramientas y tecnologías fueron elegidas por su eficacia y eficiencia en la manipulación y análisis de datos, así como por su capacidad para interactuar con otros sistemas y APIs.

4.5. Propuesta de Valor

La propuesta de valor se enfoca en la reducción del tiempo de limpieza en un 20% y aumentar la calidad de los datos en un 10%. Esto se alinea con el objetivo del proyecto de priorizar la calidad de los datos y la automatización de su ingesta, con especial atención a los datos que el equipo de marketing introduce en el CRM. Se desarrollarán herramientas para evaluar y mejorar la calidad de los datos, automatizar su limpieza y monitorear su calidad. En el contexto empresarial, esta propuesta tiene un impacto significativo al incrementar la eficiencia y la calidad de los datos, lo cual, a su vez, potencia la efectividad de las estrategias de marketing y la toma de decisiones basada en datos.

4.6. Implementación y Prueba de Concepto

Para llevar a cabo la prueba de concepto, se consideraron exclusivamente los scripts propuestos anteriormente, dado que estos permiten evaluar la viabilidad del proyecto. Se emplearon dos conjuntos de datos, de los cuales se tiene información sobre el tiempo invertido en su limpieza y en la mejora de la información. Tomando el promedio como referencia, el procesamiento de 160 registros con una calidad inicial del 23% requirió originalmente 1.86 horas persona, logrando una calidad final del 67%.

El diseño de cada script comenzó con la creación de un pseudocódigo para establecer la secuencia de tareas y su lógica. Posteriormente, se tradujo al lenguaje Python y cada tarea se

implementó como función independiente. Cada script fue validado con un conjunto de prueba simple para verificar la integridad del resultado, realizando mejoras en cada iteración.

Una vez finalizados los scripts, se aplicaron a los conjuntos de datos seleccionados. Se registraron los tiempos de ejecución y de revisión manual, así como la calidad final de los datos.

- **Pruebas y Validación**

Con la implementación de los scripts propuestos, se logró una mejora significativa en la eficiencia del procesamiento de datos. Al aplicar estos scripts a los mismos 160 registros, que inicialmente presentaban una calidad del 23%, se observó una reducción en el tiempo de procesamiento de 1.86 horas a tan solo 0.25 horas. Además, se logró un incremento en la calidad de los datos, alcanzando un 72% de calidad en los registros procesados. Estos resultados confirman la efectividad de los scripts propuestos en la mejora de la eficiencia y calidad del procesamiento de datos (Figura 8).

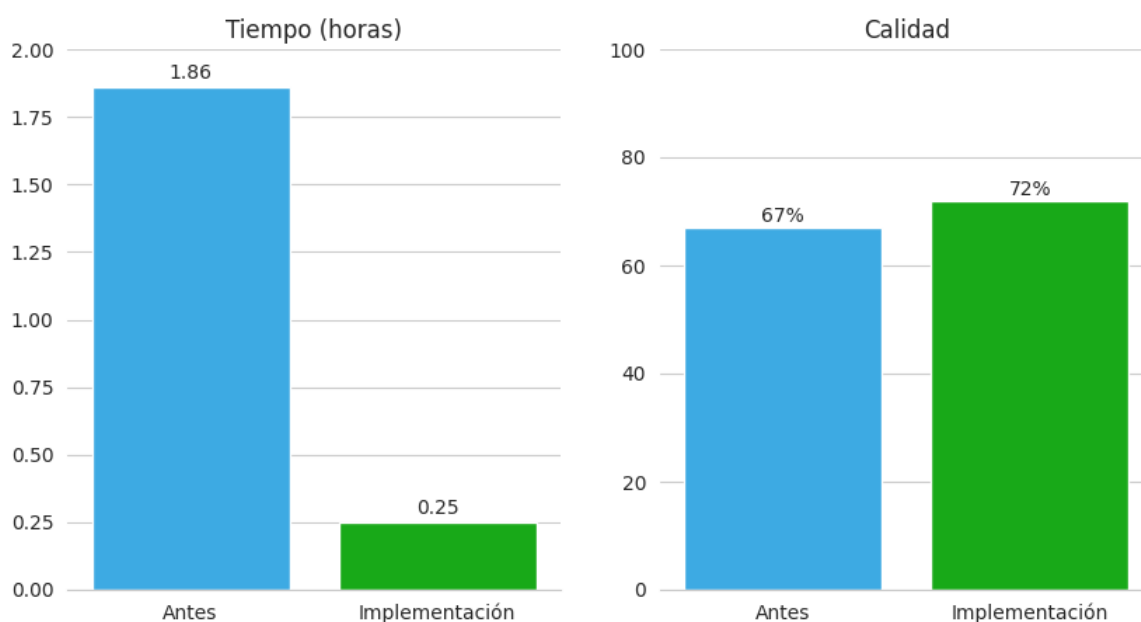


Figura 8. Tiempo de procesamiento y calidad de los datos.

4.7. Análisis de Riesgos

- **Identificación de Riesgos**

Cambios en las variables específicas al CRM: Cualquier cambio en las variables específicas al CRM puede afectar la eficacia del tablero y los scripts de limpieza y autocompletado de datos.

Cambios en las API utilizadas: Las actualizaciones o cambios en las API utilizadas, como la API de Zoominfo, pueden requerir modificaciones en los scripts.

Cambios en las fuentes de datos: Si las fuentes de datos cambian o se actualizan, puede ser necesario ajustar los scripts para adaptarse a los nuevos formatos o estructuras de datos.

Problemas de compatibilidad de software: Las diferencias en las versiones de las librerías de Python utilizadas, pueden causar problemas de compatibilidad que afecten la funcionalidad de los scripts.

- **Estrategias de Mitigación**

Monitoreo constante de las variables del CRM: Mantener una comunicación constante con el equipo de CRM para estar al tanto de cualquier cambio en las variables específicas.

Actualización regular de los scripts: Mantener los scripts actualizados para adaptarse a cualquier cambio en las API utilizadas.

Validación de los datos de entrada: Implementar controles de validación de datos para asegurar que los datos de entrada cumplen con los formatos y estructuras esperados.

Mantenimiento de la compatibilidad del software: Asegurar que todas las herramientas y tecnologías utilizadas estén actualizadas y sean compatibles entre sí.

4.8. Resultados

Los resultados obtenidos a partir de la implementación del sistema propuesto para mejorar la ingesta de datos en el área de Marketing son notables y significativos.

- **Mejora en el Tiempo de Procesamiento y calidad de los Datos**

La implementación de los scripts propuestos ha permitido una mejora significativa en la eficiencia del procesamiento de datos. Al aplicar estos scripts a los mismos 160 registros, que inicialmente presentaban una calidad del 23%, se observó una reducción en el tiempo de procesamiento de 1.86 horas a tan solo 0.25 horas. Esto representa una mejora del 86.5% en el tiempo de procesamiento, liberando recursos valiosos que pueden ser redirigidos a otras tareas importantes. Además de la mejora en la eficiencia, también se logró un incremento en la calidad de los datos. Los registros procesados mostraron un aumento en la calidad de los datos del 23% inicial a un 72% después de la implementación de los scripts. Esto representa una mejora del 49% en la calidad de los datos, lo que significa que los datos ahora son más confiables y útiles para la toma de decisiones.

- **Cambios en el Proceso de Ingesta de Datos**

La implementación de los scripts ha cambiado significativamente el proceso de ingesta de datos. Anteriormente, este proceso requería una considerable cantidad de tiempo y esfuerzo manual para limpiar y verificar los datos. Ahora, con la implementación de los scripts, gran parte de este proceso se ha automatizado, liberando tiempo y recursos valiosos. Además, la implementación de los scripts ha mejorado la consistencia y la precisión de los datos, ya que reduce la posibilidad de errores humanos.

- **Evaluación de la Implementación**

La simplicidad del sistema propuesto no solo facilita su mantenimiento, sino que también facilita la modificación y ampliación de sus capacidades. Esto significa que el sistema puede adaptarse y evolucionar con el tiempo para satisfacer las necesidades cambiantes dentro del área de Marketing. Los resultados obtenidos a partir de la implementación del sistema propuesto han demostrado que es posible mejorar significativamente la eficiencia y la calidad de la ingesta de datos en el área de Marketing. Estos resultados confirman la efectividad de los scripts propuestos y destacan el valor de la automatización y la mejora de la calidad de los datos en la toma de decisiones basada en datos.

- **Comparación con los Objetivos**

Los resultados obtenidos con la implementación de los scripts propuestos cumplen con los objetivos establecidos para el proyecto. Se logró una reducción significativa en el tiempo de limpieza de los datos y una mejora en la calidad de los mismos. Esto se alinea con el objetivo general del proyecto de desarrollar herramientas para evaluar la calidad de los datos, automatizar su limpieza, mejorarlos y crear tableros para monitorear su calidad. En términos de los objetivos específicos, se logró desarrollar y aplicar flujos de trabajo automatizados para la limpieza y transformación de datos, así como diseñar y crear un tablero que visualiza las métricas clave de calidad y consistencia de los datos. Estos resultados demuestran que se han cumplido los objetivos del proyecto.

CAPÍTULO 5. ANÁLISIS FINANCIERO

5.1. Cálculo de Proyecciones de Ingresos y Egresos del Proyecto

Esta sección presenta las proyecciones de ingresos y egresos del proyecto durante un período de 24 meses. Los ahorros mensuales se derivan de la reducción del tiempo de procesamiento multiplicado por el número de procesos y el costo por hora de la persona en Marketing. Por otro lado, los egresos se originan del costo anual del API, los costos de implementación (que incluyen el costo por hora de un Programador y un Analista) y los costos de mantenimiento trimestral (también basados en el costo por hora de un Programador y un Analista).

Los ingresos, representados como ahorros, se mantienen constantes en \$2,664.71 mensuales. Por otro lado, los egresos incluyen un costo inicial significativo de \$20,042.00 en el primer mes (costo anual del API más el costo de implementación), seguido de costos trimestrales recurrentes de \$691.20 (mantenimiento). En el mes 12, se presenta un aumento en los costos, alcanzando \$15,691.20 (costo anual del API).

El flujo neto, que es la diferencia entre los ingresos y los egresos, varía a lo largo del tiempo. En el primer mes, el flujo neto es negativo con -\$17,377.29, debido al alto costo inicial. En los meses posteriores, el flujo neto es positivo en \$2,664.71, excepto en algunos meses donde los costos adicionales lo reducen a \$1,973.51. En el mes 12, el flujo neto vuelve a ser negativo con -\$13,026.49, reflejando el incremento en los costos de ese mes.

5.2. Métodos de Evaluación

Para evaluar la viabilidad financiera del proyecto, se han utilizado tres índices clave: el Valor Actual Neto (VAN), el Retorno sobre la Inversión (ROI) y la Tasa Interna de Retorno (TIR).

El VAN del proyecto es de \$18,271.66, calculado con una tasa de descuento del 15%. Esta tasa es acorde para una empresa de tecnología en Estados Unidos, ya que refleja el costo de oportunidad del capital en este sector. En la industria tecnológica, es común ver tasas de descuento que oscilan entre el 10% y el 20% para proyectos de inversión y evaluaciones internas, dependiendo de la madurez de la empresa y el perfil de riesgo del proyecto. El VAN positivo indica que el proyecto generará más ingresos de los que cuesta, considerando el valor temporal del dinero, y por lo tanto, es financieramente viable.

El ROI del proyecto es del 48%. Este índice mide la rentabilidad de la inversión al comparar las ganancias netas obtenidas con el costo total de la inversión. Un ROI del 48% significa que, por cada dólar invertido, se espera un retorno adicional de 0.48 dólares, lo cual es una señal positiva de la rentabilidad del proyecto.

La TIR del proyecto es del 117%. La TIR es la tasa de descuento que hace que el VAN del proyecto sea igual a cero. Una TIR del 117% es significativamente alta, lo que sugiere que el proyecto es altamente rentable y que los ingresos esperados superan ampliamente los costos.

En resumen, con un VAN de \$18,271.66, un ROI del 48% y una TIR del 117%, el proyecto muestra ser financieramente viable, rentable y atractivo, especialmente dentro del sector de tecnología en Estados Unidos, donde la tasa de descuento del 15% es apropiada para evaluar inversiones.

5.3. Punto de Equilibrio y Análisis de Sensibilidad

El punto de equilibrio es el nivel de reducción del tiempo de procesamiento necesario para que el proyecto no incurra en pérdidas ni ganancias, es decir, el punto en el cual el Valor Actual Neto (VAN) es igual a cero. En este caso, analizamos cómo las variaciones en la reducción del tiempo de procesamiento impactan el VAN del proyecto.

Basándonos en la Figura 9, el punto de equilibrio se encuentra entre el 50% y el 60% de reducción del tiempo de procesamiento. Esto significa que el proyecto empezará a ser rentable cuando se logre una reducción en el tiempo de procesamiento superior al 57%.

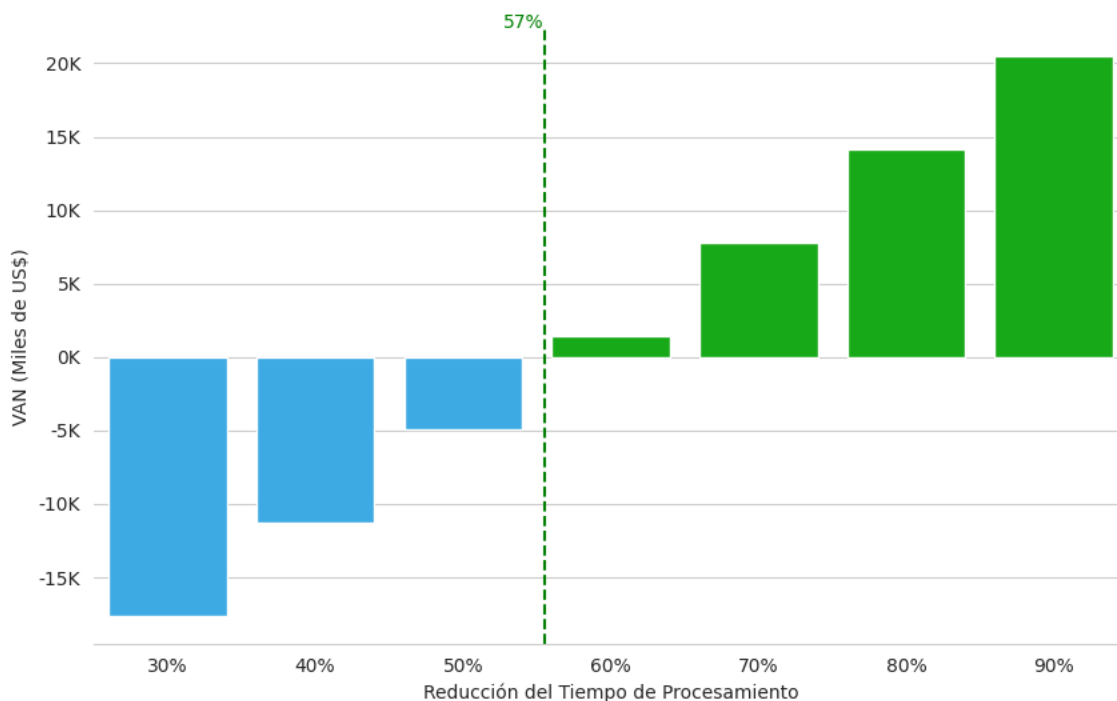


Figura 9. Análisis de sensibilidad

El análisis de sensibilidad examina cómo el VAN del proyecto cambia en respuesta a variaciones en la reducción en el tiempo de procesamiento, revelando que el proyecto es altamente sensible a las variaciones en la reducción del tiempo de procesamiento. La rentabilidad comienza a partir de una reducción del 60%, con aumentos significativos en el VAN a medida que se incrementa la eficiencia. Esto sugiere que incluso pequeñas mejoras adicionales más allá del punto de equilibrio pueden llevar a un aumento considerable en la rentabilidad del proyecto.

La clave para maximizar el valor del proyecto radica en lograr y superar la reducción del tiempo de procesamiento en al menos un 60%, lo cual asegura que los ingresos netos superen los costos, proporcionando un retorno positivo.

5.4. Resultados y Decisión Financiera

El análisis financiero del proyecto ofrece una comprensión clara de su viabilidad y rentabilidad. A partir de las proyecciones de ingresos y egresos, y los métodos de evaluación utilizados, como el Valor Actual Neto (VAN), el Retorno sobre la Inversión (ROI) y la Tasa Interna de Retorno (TIR), se pueden extraer varias conclusiones importantes.

Primero, las proyecciones de ingresos y egresos indican que los ingresos, representados como ahorros constantes de \$2,664.71 mensuales, son cruciales para contrarrestar los costos. Los egresos incluyen un costo inicial significativo de \$20,042.00 en el primer mes y costos mensuales recurrentes de \$691.20. En el mes 12, los costos aumentan drásticamente a \$15,691.20, lo que afecta el flujo de caja de ese mes. A pesar del alto costo inicial, el flujo neto se mantiene positivo después del primer mes y refleja la capacidad del proyecto para generar ingresos superiores a sus costos operativos a largo plazo. Aunque el flujo neto es negativo en el primer mes y nuevamente en el mes 12 debido a los altos costos, la tendencia general es hacia la estabilidad y la rentabilidad.

En cuanto al Valor Actual Neto (VAN), con una tasa de descuento del 15%, el proyecto muestra un VAN de \$18,271.66. Este valor positivo sugiere que el proyecto es financieramente viable y que generará más ingresos de los que cuesta, considerando el valor temporal del dinero. El análisis de sensibilidad revela que el VAN empieza a ser positivo a partir de una reducción del 60% en el tiempo de procesamiento. A medida que se incrementa esta reducción, el VAN aumenta significativamente, subrayando la importancia de optimizar los procesos para maximizar la rentabilidad.

El Retorno sobre la Inversión (ROI) del proyecto es del 48%, lo que indica que por cada dólar invertido se espera un retorno adicional de 0.48 dólares. Este índice positivo resalta la rentabilidad del proyecto, mostrando que generará beneficios significativos. Además, la Tasa Interna de Retorno (TIR) del 117% es excepcionalmente alta, lo que sugiere que el proyecto no solo es rentable sino altamente atractivo. Esta tasa indica que los ingresos esperados superan ampliamente los costos, haciendo que el proyecto sea una inversión muy rentable.

El análisis de sensibilidad enfatiza la importancia de la reducción del tiempo de procesamiento. El proyecto comienza a ser rentable con una reducción del 60% en el tiempo de procesamiento, y la rentabilidad mejora sustancialmente con mayores reducciones. Esto indica que el éxito financiero del proyecto depende en gran medida de alcanzar y mantener niveles altos de eficiencia operativa.

- **Recomendaciones**

Para maximizar los beneficios del proyecto, es crucial enfocarse en optimizar la reducción del tiempo de procesamiento. Dado que la rentabilidad del proyecto está directamente vinculada a esta métrica, se recomienda implementar estrategias y tecnologías que permitan alcanzar una reducción del 60% o más. Mantener un control riguroso sobre los costos es también esencial.

Reevaluar periódicamente la viabilidad del proyecto es otra recomendación clave. Continuar evaluando el VAN, ROI y TIR en diferentes escenarios permitirá ajustar las estrategias según sea necesario y asegurar que el proyecto siga siendo rentable a lo largo del tiempo.

Considerar opciones de financiamiento puede ser una estrategia útil para mitigar el impacto del alto costo inicial en el flujo de caja durante los primeros meses. Explorar alianzas

con otras empresas puede ofrecer una solución viable para gestionar el capital de trabajo de manera más eficiente.

- **Decisión Financiera**

En base a los análisis y resultados obtenidos, se recomienda proceder con la implementación del proyecto. Los indicadores financieros, como el VAN positivo, el ROI del 48% y la TIR del 117%, sugieren que el proyecto no solo es viable sino también altamente rentable. El enfoque debe centrarse en alcanzar y mantener al menos un 60% de reducción en el tiempo de procesamiento para asegurar la rentabilidad y maximizar las oportunidades de ahorro y eficiencia que el proyecto ofrece.

En resumen, el proyecto presenta una oportunidad financiera sólida y justifica su implementación, siempre que se mantenga una vigilancia continua sobre los costos y se optimicen los procesos operativos para mejorar la eficiencia y la rentabilidad.

CAPÍTULO 6. CONCLUSIONES Y RECOMENDACIONES

6.1. Conclusiones

La implementación del sistema propuesto para mejorar la calidad de la ingesta de datos en el área de Marketing de UDT ha demostrado ser altamente eficaz. Los resultados obtenidos indican una mejora significativa en la calidad de los datos, que se incrementó del 23% inicial al 72% tras la implementación de los scripts automatizados. Además, el tiempo necesario para la limpieza y procesamiento de los datos se redujo en un 86.5%, lo que ha liberado recursos valiosos y aumentado la eficiencia operativa.

La automatización de la limpieza y validación de datos ha reducido la posibilidad de errores humanos, mejorando la consistencia y precisión de los datos ingresados en el CRM. La implementación de tableros para evaluar y monitorear la calidad de los datos proporciona una visión clara del estado de los datos en tiempo real, facilitando la toma de decisiones informadas.

El análisis financiero muestra un Valor Actual Neto (VAN) positivo de \$18,271.66, un Retorno sobre la Inversión (ROI) del 48%, y una Tasa Interna de Retorno (TIR) del 117%, lo que indica que el proyecto es financieramente viable y rentable. Estos resultados reflejan no solo la eficiencia del sistema, sino también su potencial para generar un retorno significativo sobre la inversión realizada.

6.2. Recomendaciones

Para maximizar los beneficios del proyecto, se proponen varias recomendaciones. En primer lugar, es fundamental proporcionar un vocabulario básico y establecer criterios a través de los cuales se evalúa y se mide la calidad de los datos. Desde una perspectiva práctica, si las dimensiones de la calidad de los datos son confusas y ambiguas, será más difícil para el negocio alinear roles y responsabilidades. Por lo tanto, se recomienda simplificar las dimensiones de la

calidad de los datos a las más esenciales. Lo que realmente importa es la completitud, es decir, ¿en qué medida los datos representan el panorama completo? Y la precisión, es decir, ¿cuánto podemos confiar en los datos? Al final del día, los datos están destinados a respaldar las decisiones del negocio.

En segundo lugar, continuar optimizando los procesos de ingesta de datos para mantener y mejorar la reducción del tiempo de procesamiento. Además, mantener un monitoreo constante de las variables del CRM y las API utilizadas para asegurar que los scripts y tableros se mantengan actualizados y funcionen correctamente.

En tercer lugar, es crucial capacitar al personal en el uso de las nuevas herramientas y procesos, y socializar las políticas de gobernanza de datos a lo largo de toda la organización para asegurar una adopción efectiva. La creación de una cultura orientada a los datos es esencial para lograr una integración completa y efectiva de estas nuevas prácticas.

Además, se recomienda realizar evaluaciones periódicas del sistema para identificar áreas de mejora y ajustar las estrategias según sea necesario, asegurando la sostenibilidad y efectividad del proyecto a largo plazo. Finalmente, investigar y adoptar nuevas tecnologías y herramientas que puedan complementar y mejorar aún más el sistema de ingesta de datos, incluyendo el uso de Inteligencia Artificial y análisis predictivos. La inclusión de herramientas avanzadas puede ofrecer oportunidades adicionales para mejorar la calidad y la eficiencia de los datos.

6.3. Impacto del Proyecto en la Empresa

El impacto del proyecto en UDT es significativo y multifacético. Al mejorar la calidad de los datos, proporcionamos información más confiable y precisa, facilitando una mejor toma de decisiones estratégicas y operativas. Esta mejora en la calidad de los datos, junto con la reducción en el tiempo de procesamiento, libera recursos que pueden ser redirigidos a otras

áreas críticas, mejorando así la eficiencia general de la empresa. Además, la automatización de procesos reduce la carga de trabajo manual y minimiza errores, contribuyendo a un aumento en la productividad del equipo de marketing.

Este proyecto también ha demostrado el valor de tener un modelo y flujo de trabajo bien diseñado para la ingesta de datos. Al identificar las variables obligatorias y complementarias para la validez de la información en el negocio, hemos establecido un sistema de mejora de la calidad de la ingesta de datos, lo que ha llevado a una mejora significativa en la calidad y consistencia de los datos.

En cuanto a la gobernanza de datos, la implementación de prácticas sólidas asegura el cumplimiento con las normativas regulatorias, reduciendo el riesgo de sanciones y mejorando la reputación de la empresa. Este aspecto es particularmente importante para UDT, que opera en un entorno altamente regulado. Finalmente, la capacidad de demostrar un Retorno de Inversión (ROI) medible en las actividades de marketing fortalece la posición de UDT como un proveedor de soluciones tecnológicas confiable y orientado a datos, lo que puede atraer nuevos socios estratégicos.

Además, el proyecto ha fomentado una mayor colaboración entre los departamentos, ya que la mejora en la calidad de los datos beneficia a múltiples áreas dentro de la empresa. Esta colaboración interdepartamental es clave para el éxito continuo de las iniciativas de gobernanza de datos y para la creación de un entorno de trabajo más cohesivo y eficiente.

REFERENCIAS

- Abraham, R., Schneider, J., & vom Brocke, J. (2019). Data governance: A conceptual framework, structured review, and research agenda. *International Journal of Information Management*. Recuperado en junio de 2024, de <https://www.cairn.info/revue-projectique-2018-2-page-155.htm?contenu=citepar>
- Atlan. (2022). Data Ingestion 101: Using Big Data Starts Here. Recuperado en junio de 2024, de <https://atlan.com/data-ingestion-101/?ref=/improve-data-quality/>
- Atlan. (2023). 7 Best Practices for Data Governance. Recuperado en junio de 2024, de <https://atlan.com/data-governance-best-practices/?ref=/data-governance-framework/>
- Atlan. (2024). Data Governance Framework: Examples, Templates, Best practices. Recuperado en junio de 2024, de <https://atlan.com/data-governance-framework/>
- Atlan. (2024b). Data Governance Maturity Model: A Roadmap to Optimizing Your Data Initiatives and Driving Business Value. Recuperado en junio de 2024, de <https://atlan.com/data-governance-maturity-model/>
- CastorDoc. (2024). McKinsey Data Governance Framework: 5 Steps to Use It. Recuperado en junio de 2024, de <https://www.castordoc.com/data-strategy/mckinsey-data-governance-framework-5-steps-to-use-it>
- Collibra. (2023). The 6 data quality dimensions with examples. Recuperado en junio de 2024, de <https://www.collibra.com/us/en/blog/the-6-dimensions-of-data-quality>
- DAMA. (2017). *The DAMA guide to the data management body of knowledge (DAMA-DMBOK2) (2nd ed.)*. Technics Publications.
- IBM. (2024). Data quality dimensions (IBM Knowledge Catalog). Recuperado en junio de 2024, de <https://www.ibm.com/docs/en/cloud-paks/cp-data/4.8.x?topic=quality-data-dimensions>
- Kayuchenko, A. (2024). How DAMA Data Quality Dimensions could harm your business? LinkedIn. Recuperado en junio de 2024, de <https://www.linkedin.com/pulse/how-dama-data-quality-dimensions-could-harm-your-kayuchenko-phd-vtfee>
- Levy, R. (2024). Data Quality Dimensions: How Do You Measure Up? (+ Downloadable Scorecard). Precisely. Recuperado en junio de 2024, de <https://www.precisely.com/blog/data-quality/data-quality-dimensions-measure>

- Marchildon, P., Bourdeau, S., Hadaya, P., & Labissière, A. (2018). Data governance maturity assessment tool: A design science approach. *Projectics / Proyéctica / Projectique*, 20, 155-193. doi:10.3917/proj.020.0155
- Microsoft. (2023). Creating a modern data governance strategy to accelerate digital transformation at Microsoft. Recuperado en junio de 2024, de <https://www.microsoft.com/insidetrack/blog/driving-effective-data-governance-for-improved-quality-and-analytics/>
- Monte Carlo. (2023). Data Pipeline Monitoring 101. Recuperado en junio de 2024, de https://resources.montecarlo.com/ebooks/data-pipeline-monitoring-101?lx=LPgDLW&_gl=11oixo4u_gcl_auMzQ0MDY5MDU5LjE3MTcxMjg1MjU._ga_aMjAwNjA1ODQ5NS4xNzE3MTI4NTI1_ga_SZGJ8KW5Z8*MTcxOTQ1ODI0MS4yLjEuMTcxOTQ1ODM2Ny42MC4wLjA.&_ga=2.250954945.672193606.1719458242-2006058495.1717128525
- Moore, S. (2018). How to Create a Business Case for Data Quality Improvement. Gartner. Recuperado en junio de 2024, de <https://www.gartner.com/smarterwithgartner/how-to-create-a-business-case-for-data-quality-improvement>
- Okoro, R. (2021). Proposed Data Governance Framework for SMEs. Minnesota State University.
- Orr, K. (1998). Data quality and systems theory. *Communications of the ACM*, 41(2), 66–71. doi:10.1145/269012.269023.
- Petzold, B., Roggendorf, M., & Rowshnakish, K. (2020). Designing data governance that delivers value. McKinsey Digital. Recuperado en junio de 2024, de <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/designing-data-governance-that-delivers-value>
- Qlik. (2024). Data Governance. Recuperado en junio de 2024, de <https://www.qlik.com/us/data-governance>
- Sanko, A. (2023). Balancing quality and coverage with our data validation framework. Dropbox. Recuperado en junio de 2024, de <https://dropbox.tech/infrastructure/balancing-quality-and-coverage-with-our-data-validation-framework>
- Segner, M. (2023). Data Governance Frameworks: 5 Lessons from McKinsey, Microsoft & More. MonteCarlo. Recuperado en junio de 2024, de <https://www.montecarlo.com/blog-data-governance-frameworks/>

Shaikh, A., Harreis, H., Machado, J., & Rowshankish, K. (2024). Master data management: The key to getting more from your data. McKinsey Digital. Recuperado en junio de 2024, de <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/master-data-management-the-key-to-getting-more-from-your-data>

Tallon, P., Ramirez, R., & Short, J. (2013). The Information Artifact in IT Governance: Toward a Theory of Information Governance. *Journal of Management Information Systems*, 30, 141-178. doi:10.2753/MIS0742-1222300306.

Woodie, A. (2023). Data Quality Is Getting Worse, Monte Carlo Says. Datanami. Recuperado en junio de 2024, de <https://www.datanami.com/2023/05/02/data-quality-is-getting-worse-monte-carlo-says/>

Woodie, A. (2024). Data Quality Getting Worse, Report Says. Datanami. Recuperado en junio de 2024, de <https://www.datanami.com/2024/04/05/data-quality-getting-worse-report-says/>