Instituto Superior Tecnológico "Los Andes" ISTLA

CÓDIGO CIENTÍFICO REVISTA DE INVESTIGACIÓN

Vol. 6 – Núm. 1 / Enero – Junio – 2025

Calidad de datos en el Data Lake de una Fintech de Servicios Financieros, basado en las dimensiones de la norma ISO/IEC 25012

Data quality in the Data Lake of a Financial Services Fintech, based on the dimensions of ISO/IEC 25012

Qualidade de dados no Data Lake de uma Fintech de Serviços Financeiros, com base nas dimensões da norma ISO/IEC 25012

Barreno-Pilco, Byron Augusto
Escuela Superior Politécnica de Chimborazo
augusto.barreno@espoch.edu.ec
https://orcid.org/0009-0001-0169-4847

Silva-Peñafiel, Geovanny Euclides
Escuela Superior Politécnica de Chimborazo
geovanny.silva@espoch.edu.ec
https://orcid.org/0000-0002-1069-4574

DOI / URL: https://doi.org/10.55813/gaea/ccri/v6/n1/915

Como citar:
Barreno-Pilco, B. A., & Silva-Peñafiel, G. E. (2025). Calidad de datos en el Data Lake de una Fintech de Servicios Financieros, basado en las dimensiones de la norma ISO/IEC 25012. *Código Científico Revista De Investigación, 6*(1), 731–759. https://doi.org/10.55813/gaea/ccri/v6/n1/915.

Recibido: *18/05/2025* ***Aceptado:*** *27/06/2025* ***Publicado:*** *30/06/2025*

Resumen
En la actualidad, la calidad de los datos es un factor crítico para la toma de decisiones en empresas financieras, incluidas las Fintech. Este estudio evalúa y mejora la calidad de los datos en el dominio "Clientes" de una Fintech ecuatoriana, aplicando un subconjunto de las dimensiones definidas en la norma ISO/IEC 25012, específicamente: unicidad, consistencia, conformidad, y completitud. Para ello, se integró el enfoque Total Data Quality Management (TDQM) con la metodología de la Ciencia del Diseño, estructurada en tres ciclos: relevancia, diseño y rigor. Una vez definidas las reglas de calidad del dominio, se realizó una medición inicial (baseline) a través de su implementación. Posteriormente, los resultados obtenidos permitieron identificar deficiencias y desarrollar planes de remediación. A pesar del corto periodo de evaluación, la implementación de estas estrategias resultó en una mejora del 3.8% en la calidad de los datos. Los hallazgos de este estudio destacan la importancia de integrar la gestión de calidad de datos en la gobernanza organizacional, contribuyendo a mejorar la confianza de los stakeholders y la eficiencia operativa de la Fintech.

Palabras clave: calidad de datos, TDQM, ISO/IEC 25012, gobernanza de datos, lago de datos.

Abstract
Nowadays, data quality is a critical factor for decision-making in financial companies, including Fintechs. This study evaluates and improves data quality in the "Clients" domain of an Ecuadorian Fintech, applying a subset of the dimensions defined in the ISO/IEC 25012 standard, specifically: uniqueness, consistency, conformity and completeness. To achieve this, the Total Data Quality Management (TDQM) approach was integrated with the Design Science methodology, structured into three cycles: relevance, design, and rigor. Once the domain's quality rules were defined, an initial measurement (baseline) was conducted through their implementation. Subsequently, the obtained results allowed for identifying deficiencies and developing remediation plans. Despite the short evaluation period, implementing these strategies resulted in a 3.8% improvement in data quality. The findings of this study highlight the importance of integrating data quality management into organizational governance, contributing to enhanced stakeholder trust and the operational efficiency of the Fintech.
Keywords: data quality, TDQM, ISO/IEC 25012, data governance, datalake.

Resumo
Hoje, a qualidade dos dados é um fator crítico para a tomada de decisões em empresas financeiras, incluindo empresas de Fintech. Este estudo avalia e melhora a qualidade dos dados no domínio "Clientes" de uma empresa fintech equatoriana, aplicando um subconjunto das dimensões definidas na norma ISO/IEC 25012, especificamente: exclusividade, consistência, conformidade e integridade. Para isso, a abordagem Total Data Quality Management (TDQM) foi integrada à metodologia Design Science, estruturada em três ciclos: relevância, design e rigor. Uma vez definidas as regras de qualidade do domínio, foi realizada uma medição inicial (linha de base) por meio de sua implementação. Os resultados obtidos posteriormente permitiram identificar deficiências e desenvolver planos de remediação. Apesar do curto período de avaliação, a implementação dessas estratégias resultou em uma melhoria de 3.8% na qualidade dos dados. Os resultados deste estudo destacam a importância de integrar a gestão da qualidade de dados à governança organizacional, contribuindo para melhorar a confiança das partes interessadas e a eficiência operacional das empresas de fintech.
Palavras-chave: qualidade de dados, TDQM, ISO/IEC 25012, governança de dados, data lake.

Introducción

En la era del Big Data, las organizaciones públicas y privadas enfrentan una creciente necesidad de adoptar un enfoque data-driven para la toma de decisiones, dejando atrás los métodos empíricos tradicionales (Wulff & Finnestrand, 2023). Sin embargo, esta transformación estratégica implica retos significativos, entre los que destacan la inversión en infraestructuras tecnológicas avanzadas, ya sean locales (on-premise) o en la nube (on-cloud) (Vishnu Sakthi et al., 2024) así como la contratación de personal especializado capaz de gestionar, analizar y convertir los datos en información valiosa (insights) para respaldar la toma de decisiones (Araque González et al., 2021).

El crecimiento acelerado en la generación de datos desafía a los equipos de ciencia de datos a garantizar datos de calidad, aunque estos equipos no son quienes provocan una deficiencia en la calidad, son los dolientes directos pues son quienes dan la cara ante los usuarios. Está claro que datos no confiables o de baja calidad deriva en decisiones erróneas, posibles pérdidas económicas, inclusive riesgos o consecuencias legales. Por otro lado, contar con datos de buena calidad brindaran el efecto contrario (Ardagna et al., 2018).

Ante este panorama, surge la interrogante sobre si las organizaciones cuentan con datos de buena o mala calidad. Para responder a esta cuestión, se emplea el enfoque Total Data Quality Management (TDQM), que comprende cuatro fases: definición de reglas de calidad, medición, análisis de resultados y ejecución de acciones correctivas (Bicevskis et al., 2018; Nikiforova, 2020). Como primer paso en este estudio, se ha iniciado la medición de la calidad de los datos en la organización (Ardagna et al., 2018; McMann et al., 2022).

En el grupo corporativo al que pertenece la Fintech, se han llevado a cabo análisis internos previos que han permitido identificar las dimensiones más críticas aplicables a datos de banca. Dichos análisis previos llevaron a adoptar el estándar ISO/IEC 25012 para la medición de calidad, el estándar contiene un total de 15 dimensiones entre 5 inherentes y 10

dependientes (Calabrese et al., 2019). Además, existen otras metodologías, como las desarrolladas por DAMA e IBM, para la gestión de la calidad de datos. En este estudio, se han priorizado cuatro dimensiones clave: unicidad, consistencia, conformidad, y completitud, que no solo permiten medir el nivel de calidad, sino también facilitan la implementación de planes de remediación para su mejora continua (Andrew Black (Van Nederpelt & Black) & Peter van Nederpelt (Van Nederpelt & Black), 2020).

Las empresas financieras en Ecuador, incluidas las Fintech, no están exentas de los problemas derivados de la baja calidad de los datos. El presente estudio tiene como objetivo principal implementar un artefacto que permita medir la calidad de los datos, que apalanque planes de remediación y mejora en la calidad de datos para el dominio: Clientes.

Para ello, se plantean tres objetivos específicos:

Primero: Diagnosticar el estado actual de la calidad de los datos en el Data Lake de la Fintech, analizando las dimensiones de unicidad, precisión, consistencia, conformidad, completitud y actualidad.

Segundo: Implementar reglas y mecanismos de control para gestionar y mejorar la calidad de los datos en el Data Lake sobre la plataforma Databricks, para cada dimensión de calidad.

Tercero: Evaluar el impacto de la implementación del artefacto, sobre las dimensiones de calidad de datos, comparando indicadores antes y después del proceso de implementación.

Como medida de mitigación ante los riesgos asociados a la baja calidad de los datos, se resalta la necesidad de integrar la disciplina de calidad de datos dentro de la gobernanza de datos, buscando fortalecer la confianza de los usuarios de los datos para lograr ser una compañía guiada por datos (Herrera & Kapur, 2007).

Los resultados de este estudio sin duda contribuirán en el sector financiero ecuatoriano y de la región, para la adopción de la gestión de calidad de datos, siempre con la finalidad de promover las decisiones estratégicas de las compañías basado en datos confiables y precisos.

Esta investigación tiene su fundamento teórico llevado a cabo mediante una revisión sistemática de libros y revistas sobre calidad de datos, incluyendo el estándar ISO/IEC 25012, el libro DMBOOK2, y varios artículos científicos publicados en revistas indexadas en bases de datos como Scopus, Science Direct, Springer y Google Scholar.

Esta investigación se ha llevado a cabo en una Fintech de servicios financieros que basa sus decisiones y estrategia en datos, por ende, prioriza la calidad de los datos para una adecuada generación de insights. La pregunta central que orienta este trabajo es:

¿La aplicación de una metodología para la medición de calidad de datos y la ejecución de planes de remediación puede mejorar la calidad de los datos en la organización?

Para garantizar una mejora continua, es fundamental implementar planes de remediación efectivos que cubran tanto la corrección de datos de baja calidad existentes, así como los ajustes necesarios en los procesos o sistemas para evitar que los mismos problemas que causaron baja calidad sigan ocurriendo; es mandatorio una sinergia adecuada entre las personas que poseen los roles de Product Owner, Data Steward y Data Engineer. La colaboración entre estos roles permitirá por un lado la remediación en deficiencias de calidad de datos pasadas y actuales, y por otro lado, viendo hacia el futuro, el establecimiento de mecanismos de control y gobernanza que aseguren la sostenibilidad de la calidad de los datos (DAMA International, 2017).

Metodología

El proyecto pretende aplicar la disciplina de calidad de datos, esencial dentro de la gestión y gobernanza de los datos (DAMA International, 2017), con el objetivo de garantizar que los datos de la Fintech, sean aptos para el propósito de guiar su estrategia por datos. Con ese antecedente claro, se adopta la metodología de la ciencia del diseño, estructurada en tres fases: 1. El ciclo de la relevancia, 2. El ciclo del diseño y 3. El ciclo del rigor, esta metodología realza la validez científica y la aplicabilidad práctica de los resultados obtenidos.

Figura 1

Ciencia del diseño (DSR) aplicada a la gestión de calidad de datos

Nota: (Autores, 2025).

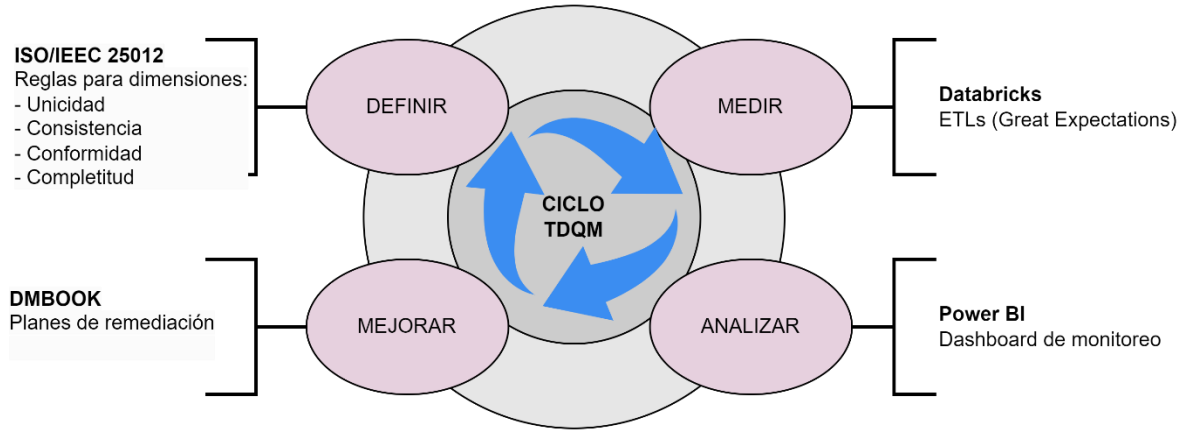
La figura representa el ciclo de investigación aplicado para mejorar la calidad de datos, partiendo de la identificación del problema, el diseño del artefacto con base en la metodología TDQM, y la evaluación comparativa del índice de calidad del dominio en el tiempo.

Ciclo de la Relevancia

El ciclo de la relevancia dentro de este estudio está orientado a comprender y abordar el problema del desconocimiento del nivel de calidad de los datos en el dominio "Clientes" de la Fintech. Este ciclo vincula las necesidades del entorno organizacional con el desarrollo de soluciones tecnológicas y metodológicas basadas en Total Data Quality Management (TDQM) y el estándar ISO/IEC 25012.

Figura 2

Ciclo TDQM aplicado a la gestión de calidad de datos



Nota: (Autores, 2025).

La figura muestra el ciclo de la metodología Total Data Quality Management (TDQM), compuesto por las fases de definir, medir, analizar y mejorar. Este proceso se complementa con las dimensiones de calidad de datos del estándar ISO/IEC 25012 y los planes de remediación propuestos en el marco del DMBOOK.

Para ello, se han llevado a cabo cinco actividades principales que establecen la base del estudio, cabe mencionar que la primera actividad mencionada no forma parte del TDQM, sin embargo, fue necesaria para llevar a cabo la experimentación de esta investigación:

1. Selección de herramienta para la medición de calidad de datos:

Evaluación comparativa de tres posibles soluciones tecnológicas para la medición y monitoreo de la calidad de los datos: 1. Desarrollo de un motor de calidad a medida. 2. Librería Great Expectations (GX) y 3. Delta Live Tables (DLT) con expectations de Databricks.

Se definieron criterios de selección basados en efectividad, escalabilidad, facilidad de implementación y presupuesto dentro de la infraestructura de la Fintech, resultado de esta evaluación se optó por trabajar con Great Expectations (GX).

Tabla 1

Comparación de herramientas para la medición de calidad de datos

Criterio	Delta Live Tables (DLT)	Great Expectations (GX)	Motor de Calidad Propio
Definición	Servicio gestionado de Databricks para crear y monitorear pipelines de datos confiables.	Framework de validación de datos open source para crear pruebas y monitorear la calidad.	Solución personalizada desarrollada para validar y mejorar la calidad de datos según necesidades específicas.
Facilidad de uso	Alta, interfaz intuitiva, integración con Databricks y SQL/PySpark.	Moderada, requiere configuración manual de las expectativas.	Depende del diseño, inicialmente puede ser más complejo.
Integración	Se integra nativamente con Databricks, Spark y Delta Lake.	Compatible con múltiples fuentes de datos (SQL, Pandas, Spark).	Desarrollo específico para Spark
Automatización	Automático con mantenimiento gestionado, incluye auto-healing.	Parcial, se puede automatizar, pero requiere scripts adicionales.	Totalmente personalizable, pero requiere desarrollo y monitoreo.
Personalización	Limitada a las funcionalidades ofrecidas por Databricks.	Alta, permite personalizar validaciones con código Python.	Alta, adaptable a las necesidades del negocio.
Escalabilidad	Alta, diseñado para grandes volúmenes de datos.	Alta, soporta grandes volúmenes en entornos distribuidos.	Alta, dependiendo del entorno.
Costos	Altos, requiere licencia de Databricks.	Es open source.	Desarrollo propio., sin embargo requiere de muchas horas hombre
Tiempo de Implementación	Rápido, configuración simple en entornos existentes de Databricks.	Moderado, depende de la complejidad de las reglas.	Lento, desarrollo desde cero según las necesidades.
Documentación y Soporte	Amplia, soporte oficial de Databricks.	Amplia, comunidad activa y soporte comercial opcional.	Depende del equipo interno.

Flexibilidad en Reglas	Reglas predefinidas y pipelines declarativos.	Reglas completamente definidas por el usuario.	Totalmente flexible, según los requerimientos definidos.
Monitoreo	Integrado, dashboards en Databricks para monitorear pipelines.	Integrado, dashboards de validación y reportes automatizados.	Depende del desarrollo, puede integrarse con herramientas BI.
Mantenimiento	Bajo, gestionado por Databricks.	Moderado, requiere actualizaciones manuales.	Alto, responsabilidad completa del equipo.

Nota: (Autores, 2025).

La tabla muestra detalles de la comparativa realizada para elegir la herramienta a usar/implementar, proporcionando información valiosa a favor de DLT y GX, sin embargo el costo dio un aporte adicional en la elección final por Great Expectations GX.

2. Definición de estándares de calidad:

Identificación de las dimensiones de calidad y establecimiento de reglas de calidad aplicables al dominio de estudio “Clientes”.

3. Medición de la calidad de los datos:

Implementación de procesos automáticos ETL de medición, mismos que serán de utilidad tanto para la medición inicial, así como las mediciones recurrentes sobre las dimensiones de calidad priorizadas.

4. Análisis de la calidad de los datos:

Análisis mediante monitoreo continuo de los datos para identificar tendencias y puntos críticos de mejora. Para este fin se ha diseñado un dashboard que permite el análisis de una forma amigable.

5. Limpieza y remediación de datos:

Gestión de planes de remediación con un enfoque iterativo.

Implementación directamente en los procesos ETL de transformaciones y limpieza de datos basados en las reglas que deben cumplir los campos para corregir inconsistencias y prevenir futuros problemas de calidad de datos.

El estudio sigue un enfoque cuantitativo y experimental, con el objetivo de medir el índice de calidad de datos en el dominio "Clientes" mediante el análisis de datos reales obtenidos de las bases de datos de la Fintech.

Población y Muestra

- Población: Clientes del grupo corporativo al que pertenece la Fintech.
- Muestra: Clientes que han instalado la aplicación desde 2023 hasta la actualidad (2025).
- Criterios de inclusión: Clientes con onboarding completo.
- Criterios de exclusión: No se excluyen datos.

Instrumentos de Recolección de Datos

Para evaluar la calidad de los datos, se utilizarán los datos almacenados en el Datalake de las capas bronze y gold de la arquitectura medallion, considerando los siguientes campos:

- Datos de identificación: Número de identificación, tipo de documento, nombre del cliente.
- Datos de contacto: Teléfono, correo electrónico, dirección principal.
- Información financiera y de afiliación: Agencia de anclaje, código interno, estado del cliente.
- Datos demográficos: Estado civil, fecha de nacimiento, género, nacionalidad.
- Datos de ubicación: País, provincia, cantón y parroquia de domicilio.
- Datos operativos: Fecha de creación del registro.

Ciclo del Diseño

En esta primera fase del DSR, se ha diseñado el artefacto que permitirá abordar la problemática de calidad de datos en el dominio "Clientes". La solución se basa en la metodología Total Data Quality Management (TDQM) y el estándar ISO/IEC 25012, asegurando una evaluación sistemática de la calidad de los datos mediante reglas predefinidas y procesos de monitoreo continuo.

Cabe aclarar algo adicional que podría causar confusión. DSR es la metodología usada en esta investigación que permite diseñar el artefacto de gestión de calidad de datos, TDQM por su parte es la metodología que ha sido elegida luego de la investigación para llevar a cabo de inicio a fin la gestión como tal de la calidad de datos, y finalmente es estándar ISO/IEC 25012 es el que proporciona las dimensiones a evaluar para medir la calidad de datos, de estas dimensiones a su vez se derivan reglas de calidad aplicables a los campos de las tablas de las bases de datos.

Dado que gran parte de los problemas de calidad se originan desde las fuentes de datos, la remediación no será automática en esta etapa, sino que dependerá de los planes de acción y la colaboración con otros equipos técnicos dentro de la organización. No obstante, este proyecto establece las bases para medir, monitorear y visualizar la calidad de los datos, proporcionando insumos clave para los planes de remediación.

Solución propuesta

La solución consiste en el diseño e implementación de un conjunto de procesos automatizados ELTs de medición de calidad de datos, con la posibilidad de expandirse en el futuro a otros dominios de información. Para la medición, se usó la librería Great Expectations.

El artefacto opera a tres niveles de evaluación de datos:

- Histórica: Evalúa la calidad de todos los datos almacenados en las tablas del dominio.

- Diaria: Analiza solo los nuevos datos ingresados en el sistema en un día específico.
- Near Real-Time (NRT): Realiza validaciones cada hora sobre los datos más recientes.

Great Expectations proporciona los resultados en formatos JSON, estos a su vez en los ETL son llevados hacia una tabla, la cual se constituye como la fuente para el dashboard de monitoreo en Power BI, que permite la visualización de la calidad de los datos desde una jerarquía de: dominio → tabla → campo. El dashboard facilita la identificación de problemas de calidad y sirve como insumo para coordinar acciones de remediación con los equipos técnicos.

Para evaluar la calidad conforme a cada dimensión, a nivel de campo se aplicaron las siguientes fórmulas:

Unicidad. - Evaluación de registros únicos, sin duplicidad conforme a los campos clave o que dan unicidad a los registros de las tablas a evaluar.

$$Unicidad = \left(\frac{Total\ de\ registros\ únicos}{Total\ de\ registros}\right)$$

En donde un valor de Unicidad al 100% es interpretado como que no existen duplicados.

Completitud. – Medición del porcentaje de valores faltantes siendo estos valores nulos o blancos.

$$Completitud = \left(1 - \left(\frac{Total\ valores\ nulos + Total\ valores\ en\ blanco}{Total\ de\ registros\ evaluados}\right)\right)$$

En donde un valor de Completitud al 100% indica que no hay datos faltantes. Como aclaración acá se toma como denominador “Total de registros evaluados”, ya que existen casos en los que, si es correcto que el campo no posea valor, por lo tanto, la evaluación se hará sobre los casos de registro que obligatoriamente deben poseer un valor válido, distinto de nulo o blanco.

Conformidad. - Verificación de que los datos cumplen con formatos y estructuras definidas, por ejemplo, un campo con correos electrónicos debe contener un formato válido.

$$Conformidad = \left(\frac{Total\ de\ registros\ que\ cumplen\ con\ el\ formato\ definido}{Total\ de\ registros\ evaluados}\right)$$

Consistencia (Lógica). – Evaluación si los datos mantienen una lógica coherente entre sí, ya sea dentro de una misma tabla o con otras incluso de otras fuentes. Un dato es consistente cuando no presenta contradicciones lógicas respecto a otros datos relacionados.

$$Consistencia = \left(\frac{Total\ de\ registros\ \text{sin}\ inconsistencias\ detectadas}{Total\ de\ registros\ evaluados}\right)$$

Mientras que para obtener el porcentaje general de cada dimensión se promedian los resultados individuales de los campos.

$$Calidad\ general\ por\ dimensión = \frac{\sum_{i=1}^{n} Ci}{n}$$

Donde:

Ci representa el porcentaje de valores correctos obtenido para el “campo i” según su respectiva regla de calidad.

n es el número total de campos evaluados bajo la dimensión de validez.

El numerador representa la suma de todos los valores individuales de la dimensión, y el denominador simplemente normaliza el resultado dividiéndolo por el número total de campos.

Por último, para obtener el índice de calidad del dominio, se suman los resultados generales de cada dimensión de acuerdo a su ponderación:

$$ICD = \sum_{i=1}^{m} (Di * wi)$$

Donde:

m es el número de dimensiones usadas para la medición.

Di representa el porcentaje de calidad de cada dimensión

wi representa la ponderación o peso que se le da a cada dimensión, para este estudio se da la misma ponderación para cada dimensión, al ser 4 les corresponde el 25% a cada una.

Diseño del artefacto

El diseño de la solución se compone de los siguientes módulos:

1. ETLs con Great Expectations (GX): Encargados de ejecuta reglas de validación sobre los datos en la capa cruda y master. Realiza validaciones en tres niveles: histórico para evaluar todos los datos almacenados en la tabla, diario para evaluar los datos generados en un día y NRT para evaluar los datos generados durante la última hora.

2. Dashboard de monitoreo: Desarrollado en Power BI para visualizar indicadores clave y tendencias de calidad de datos, mismo que permite la identificación de campos específicos con problemas.

3. Mecanismo de remediación: Actualmente, la remediación de datos no es automática y depende de los equipos responsables de las fuentes de datos para el caso de zona bronze, mientras que para zonas silver y gold, depende del equipo de ingeniería de datos. Se identifican y documentan problemas de calidad en un libro Excel, desde donde se coordinan los planes de remediación.

Proceso de desarrollo e iteración

La implementación sigue un enfoque iterativo, pasando por varias fases de prueba y ajuste:

- Implementación inicial: Desarrollo del primer ETL con reglas de validación en Great Expectations (GX). Se prueba con datos históricos sobre una tabla de clientes.
- Ajustes y optimización: Modificación de las reglas y ajustes en la arquitectura de los ETLs según los primeros resultados obtenidos.
- Expansión de la solución: Desarrollo de 8 ETLs adicionales para cubrir todas las tablas del dominio (capa cruda y master).

- Validación con usuarios internos: Data Stewards, Data Architects y Data Engineers para evaluar la efectividad del artefacto.
- Automatización: Programación o calendarización de los ETLs para que se ejecuten de forma automática de acuerdo a la periodicidad.

Herramientas y tecnologías utilizadas

La solución está construida utilizando las siguientes herramientas.

Tabla 2

Herramientas usadas para la gestión de calidad de datos

Componente	Herramienta
Plataforma y procesamiento	Databricks
ETLs y validación de datos	Python, PySpark, Spark SQL, Great Expectations (GX)
Almacenamiento de resultados	Delta Lake (Hive)
Visualización y monitoreo	Power BI
Gestión de planes de remediación	Excel (manual, en este primer release)

Nota: (Autores, 2025).

Validación del diseño

Para asegurar que el sistema cumple con sus objetivos, se implementa una validación en dos niveles:

Nivel 1: Evaluación cuantitativa en la que se comparan métricas de calidad antes y después de implementar el artefacto. Además se analiza la evolución de indicadores como porcentaje de registros duplicados, valores nulos o inconsistentes.

Nivel 2: Evaluación cualitativa en donde se recoge feedback de los Data Stewards y Data Engineers sobre la usabilidad y efectividad del artefacto. Se evalúa que tan fácil es interpretar los resultados y tomar decisiones basadas en el dashboard.

Limitaciones de la solución

El artefacto representa un avance significativo en la gestión de calidad de datos, con oportunidad de mejora debido a ciertas limitaciones de esta versión inicial:

1. Ausencia de alertas y notificaciones automáticas

Para el monitoreo de la calidad se cuenta con el dashboard, cuya interpretación de resultados o indicadores lo deben realizar personas. En una fase futura se considera la incorporación de alertas automáticas para notificar problemas críticos.

2. Falta de automatización en los planes de remediación

Aunque el sistema identifica problemas de calidad, la remediación sigue siendo un proceso manual. La corrección de datos depende de los equipos responsables de las fuentes de datos, lo que puede generar demoras en la resolución.

3. Dependencia de la colaboración interdepartamental

La mejora en la calidad de datos requiere la cooperación de diferentes áreas de la organización. La efectividad del artefacto dependerá del grado de compromiso de los equipos de desarrollo y operaciones.

Estas limitaciones destacan posibles áreas de mejora y futuras líneas de investigación, con el objetivo de evolucionar hacia una solución más automatizada e integrada con los procesos de negocio.

Ciclo del rigor

El modelo de evaluación de calidad de datos utilizado en este estudio se fundamenta en la norma ISO/IEC 25012, la cual define un conjunto de 15 dimensiones para la gestión de calidad de datos. Sin embargo, estudios previos han demostrado que no todas las dimensiones son igualmente relevantes en todos los dominios de datos (Calabrese et al., 2019). Para este estudio, se han seleccionado cuatro dimensiones clave (unicidad, consistencia, conformidad y completitud) debido a su impacto en la calidad de los datos en el sector financiero, alineándose con investigaciones previas en entornos de datos estructurados (Andrew Black (Van Nederpelt & Black) & Peter van Nederpelt (Van Nederpelt & Black), 2020; Zibak et al., 2022).

Otros estudios recientes, han usado métricas similares para evaluar la integridad y confiabilidad de los datos (McMann et al., 2022). Si bien dentro de la revisión sistemática no se hallaron estudios que hayan usado Great Expectations (GX), diversos foros, páginas, la misma documentación de la librería y los resultados de la comparativa realizada, dieron el aval de que sería la mejor decisión para usar sobre la plataforma analítica que utiliza la Fintech. Las fórmulas utilizadas para evaluar cada dimensión han sido adoptadas en diversas metodologías de calidad de datos y han demostrado su validez en escenarios similares.

Por ejemplo: La métrica de unicidad ha sido ampliamente utilizada en modelos de detección de duplicados en sistemas financieros (Bena et al., 2024). La métrica de completitud es una de las más utilizadas en Data Quality Management y ha sido aplicada en estudios sobre la confiabilidad de los datos de clientes (Ardagna et al., 2018). Por lo tanto, la combinación de estas métricas proporciona un marco sólido y científicamente validado para la evaluación de calidad de datos en la Fintech, asegurando que los resultados obtenidos sean comparables con otros estudios y puedan ser utilizados como referencia en futuros proyectos de calidad de datos.

La revisión sistemática de la literatura proporcionó información interesante entre ello se identificó que la mayoría de las investigaciones relacionadas con calidad de datos se enfocan principalmente en la etapa de medición o diagnóstico, lo cual es tan solo una etapa en la gestión de calidad de datos. Como punto a favor, el presente estudio incorpora una etapa adicional en cuanto a la gestión de planes de remediación, fortaleciendo el vínculo entre diagnóstico y mejora efectiva. Esta integración responde a la necesidad de generar valor real a partir del análisis de calidad, alineándose con los principios del enfoque TDQM. Para cubrir con esta validación se realiza una comparación de las métricas de calidad de datos por dimensión, antes (línea base) y después (actualidad) de modo que quede la evidencia de la gestión realizada para la mejora del índice de calidad de datos.

Resultados

Baseline

Como se mencionó, el punto de partida sería la medición inicial, que permite identificar el punto de partida en términos de calidad de datos, de aquí nacen los planes de remediación. A continuación, se presentarán muestras de los resultados de la medición inicial, se presentaran las tablas 3, 4, 5 y 6 que contienen los resultados por cada dimensión, a nivel de tabla, regla de calidad aplicada, cantidad total de registros evaluados, cantidad de registros correctos y el porcentaje de calidad de cada campo.

Tabla 3

Resultados de la evaluación de dimensión de completitud.

Tabla	Regla	Registros analizados	Registros correctos	Porcentaje calidad
Client	Agencia anclaje no puede ser nulo ni blanco	1.780.453	1.744.166	97,96%
Client	Cantón no puede ser nulo ni blanco	1.780.453	1.647.648	92,54%
Client	País no puede ser nulo ni blanco	1.780.453	-	0,00%
Customers	Agencia anclaje no puede ser nulo ni blanco	1.810.305	1.740.186	96,13%
Customers	Código interno no puede ser nulo ni blanco	1.810.305	1.777.142	98,17%
Customers	Fecha creación no puede ser nulo ni blanco	1.810.305	1.777.230	98,17%
Users	Cantón no puede ser nulo ni blanco	1.810.305	1.654.357	91,39%
Users	Celular no puede ser nulo ni blanco	1.810.305	1.783.231	98,50%
Users	Estado civil no puede ser nulo ni blanco	1.810.305	1.787.098	98,72%
Clients	Tipo de cliente no puede ser nulo ni blanco	2.688.432	2.688.432	100,00%
Clients	Descripción de usuario no puede ser nulo ni blanco	2.688.432	2.688.432	100,00%

Clients	Ds_churn_category no puede ser nulo ni blanco	2.688.432	2.688.432	100,00%
Users_pilot	Numero de cuentas no debe estar vacía	746.926	746.601	99,96%
Users_pilot	Apps no debe estar vacía	746.926	746.926	100,00%
Users_pilot	Tipo de cliente no puede ser nulo ni blanco	746.926	746.926	100,00%

Nota: (Autores, 2025).

Después de promediar los porcentajes de calidad para cada campo de cada tabla se obtiene un 90.58% de calidad en completitud, en términos generales el resultado de la métrica de calidad para la dimensión de completitud es bueno, sin embargo, se ve impactado por campos como los de ubicación, latitud y longitud que en su mayoría son datos faltantes, para el resto de campos se tiene un porcentaje mayor al 90%.

Dimensión de Unicidad

Tabla 4

Resultados de la evaluación de dimensión de unicidad

Tabla	Regla	Registros analizados	Registros correctos	Porcentaje calidad
Client	Celular no debe estar repetidos	1.780.453	1.621.720	91,08%
Client	Identificación no puede estar duplicado	1.780.453	1.621.722	91,08%
Client	Correo electrónico no debe estar repetido los correos	1.780.453	1.621.720	91,08%
Users	Celular no debe estar repetidos	1.810.305	1.623.590	89,69%
Users	Identificación no puede estar duplicado	1.810.305	1.649.746	91,13%
Users	Correo electrI³nico no debe estar	1.810.305	1.650.629	91,18%

	repetidos los correos			
Clients	Identificación no puede estar duplicado	2.688.432	2.688.432	100,00%
Clients	Celular no debe estar repetidos	2.688.432	2.678.287	99,62%
Users_pilot	Identificación no puede estar duplicado	746.926	746.926	100,00%

Nota: (Autores, 2025).

A pesar de que el porcentaje de calidad en unicidad es bueno 93.66% (promedio de los porcentajes por campo y tabla), la existencia de duplicidad causa problemas, obligando a realizar depuraciones de la duplicidad en todos los procesos en donde se use esta data.

Dimensión de Validez

Tabla 5

Resultados de la evaluación de dimensión de validez.

Tabla	Regla	Registros analizados	Registros correctos	Porcentaje calidad
Client	Identificación solo debe tener números	1.780.453	1.780.453	100,00%
Client	Fecha de nacimiento debe estar en formato 'YYYY-MM-DD'	1.780.453	1.780.397	100,00%
Client	Genero debe ser MALE o FEMALE	1.780.453	1.780.397	100,00%
Client	Correo electrónico no debe contener doble arroba, puntos seguidos ni ser uno de los correos por defecto	1.780.453	1.780.109	99,98%
Users	Nombre completo debe tener al menos un nombre y apellido	1.810.305	1.612.042	89,05%
Users	País debe pertenecer al catalogo	1.810.305	1.787.116	98,72%
Users	Parroquia debe pertenecer al catalogo	1.810.305	669.729	37,00%

Users	Correo electrónico no debe contener doble arroba, puntos seguidos ni ser uno de los correos por defecto	1.810.305	1.809.906	99,98%
Users	Nacionalidad debe ser ECUATORIANA	1.810.305	1.667.263	92,10%
Clients	Nombre completo debe tener al menos un nombre y apellido	2.688.432	2.388.547	88,85%
Clients	Latitud debe estar entre -5 y 2 (Ecuador)	2.688.432	209.318	7,79%
Clients	Longitud debe estar entre -81 y -75 (Ecuador)	2.688.432	207.345	7,71%
Users_pilot	Identificación solo debe tener números	746.926	744.359	99,66%
Users_pilot	Identificación debe tener la longitud de 10	746.926	744.330	99,65%
Users_pilot	Tipo de usuario debe ser User	746.926	746.926	100,00%

Nota: (Autores, 2025).

En el caso de esta dimensión, se evidencia un porcentaje de calidad menor: 87.50%, sin embargo, se ratifican problemas de calidad en campos de ubicación: parroquia, cantón, provincia, latitud y longitud.

Dimensión de conformidad

Tabla 6

Resultados de la evaluación de dimensión de conformidad

Tabla	Regla	Registros analizados	Registros correctos	Porcentaje calidad
Client	Identificación debe cumplir con algoritmo Cedula	1.780.453	1.780.453	100,00%
Client	Fecha de nacimiento de estar entre 18 y 110 años	1.780.453	1.780.140	99,98%
Client	Fecha creación debe estar entre las fechas definidas	1.780.453	1.780.453	100,00%

Users	Estado cliente es 'NEWLY_REGISTERED', 'ACTIVE_ACCOUNT', 'ACTIVE' y fecha deceso no puede tener valor	1.810.305	693.277	38,30%
Users	Estado civil si es CASADO o UNIÓN LIBRE debe tener pareja, caso de SOLTERO, DIVORCIADO o VIUDO no debe tener pareja	1.810.305	820.630	45,33%
Customers	Fecha creación debe estar entre las fechas definidas	1.810.305	1.777.230	98,17%
Users	Identificación debe cumplir con algoritmo Cedula	1.810.305	1.810.305	100,00%
Users	Fecha de nacimiento de estar entre 18 y 110 años	1.810.305	1.786.819	98,70%
Users	Celular debe tener 9 dígitos y ser numérico y debe empezar con 09 o +593	1.810.305	1.782.815	98,48%
Clients	Fecha creación debe estar entre las fechas definidas	2.688.432	2.688.432	100,00%
Clients	Celular debe tener 9 dígitos y ser numérico y debe empezar con 09 o +593	2.688.432	2.686.525	99,93%
Users_pilot	Fecha creación debe estar entre las fechas definidas	746.926	746.926	100,00%

Nota: (Autores, 2025).

En cuanto a la conformidad, las reglas aplicables para el dominio de estudio son pocas, dando como resultado un 90.68% de calidad, este resultado se ve impactado por dos campos “estado civil” y “estado del cliente”, los mismos que también serán como parte de los planes de remediación.

Una vez que se cuenta con los resultados detallados por cada dimensión, queda obtener el porcentaje de calidad de partida, para lo cual según la fórmula previamente descrita se suman los resultados individuales de cada dimensión según su ponderación.

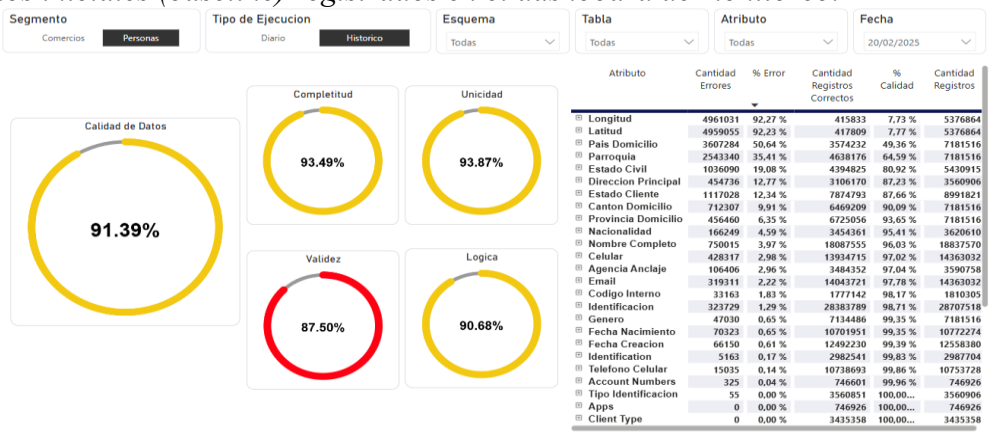
$$ICD = \sum_{i=1}^{m} (Di * wi)$$

$$ICD\ Clientes = (90.68 * 0.25) + (93.49 * 0.25) + (87.50 * 0.25) + (93.87 * 0.25)$$

$$ICD\ Clientes = 91.3\%$$

Figura 3

Resultados iniciales (baseline) registrados en el dashboard de monitoreo.



Atributo	Cantidad Errores	% Error	Cantidad Registros Correctos	% Calidad	Cantidad Registros
Longitud	4961031	92,27 %	415833	7,73 %	5376864
Latitud	4959055	92,23 %	417809	7,77 %	5376864
Pais Domicilio	3607284	50,64 %	3574232	49,36 %	7181516
Parroquia	2543340	35,41 %	4638176	64,59 %	7181516
Estado Civil	1036090	19,08 %	4394825	80,92 %	5430915
Direccion Principal	454736	12,77 %	3106170	87,23 %	3560906
Estado Cliente	1117028	12,34 %	7874793	87,66 %	8991821
Canton Domicilio	712307	9,91 %	6469209	90,09 %	7181516
Provincia Domicilio	456460	6,35 %	6725056	93,65 %	7181516
Nacionalidad	166249	4,59 %	3454361	95,41 %	3620610
Nombre Completo	750015	3,97 %	18087555	96,03 %	18837570
Celular	428317	2,98 %	13934715	97,02 %	14363032
Agencia Anclaje	106406	2,96 %	3484352	97,04 %	3590758
Email	319311	2,22 %	14043721	97,78 %	14363032
Codigo Interno	33163	1,83 %	1777142	98,17 %	1810305
Identificacion	323729	1,29 %	28383789	98,71 %	28707518
Genero	47030	0,65 %	7134486	99,35 %	7181516
Fecha Nacimiento	70323	0,65 %	10701951	99,35 %	10772274
Fecha Creacion	66150	0,61 %	12492230	99,39 %	12558380
Identification	5163	0,17 %	2982541	99,83 %	2987704
Telefono Celular	15035	0,14 %	10738693	99,86 %	10753728
Account Numbers	325	0,04 %	746601	99,96 %	746926
Tipo Identificacion	55	0,00 %	3560851	100,00...	3560906
Apps	0	0,00 %	746926	100,00...	746926
Client Type	0	0,00 %	3435358	100,00...	3435358

Nota: (Autores, 2025).

La figura muestra una primera visualización del dashboard desarrollado, del cual se puede revisar los resultados históricos, además se puede aplicar varios filtros, por último, proporciona una tabla con el detalle por dimensión, tabla y campo de los resultados individuales de la medición.

Plan de remediación

El objetivo del siguiente plan de remediación, es elevar la calidad de datos del dominio Clientes que parte desde un baseline del 91.3%, priorizando los campos que afectan procesos clave entre ellos la analítica de datos para las dimensiones: completitud, unicidad, validez y conformidad. Para este fin aplicaremos dos principios clave mencionados en DMBOOK (DAMA International, 2017), como son: Remediación correctiva: Arreglar los datos que ya están mal, es decir limpieza y Remediación preventiva: Corregir procesos para que los errores no vuelvan a ocurrir.

Tabla 7

Campos críticos identificados

Campo / Regla	Dimensión	Problema	Criticidad
Latitud y Longitud	Completitud / Validez	Altísimo nivel de nulos y valores fuera de rango	Alta
País, Parroquia, Cantón	Completitud / Validez	Datos faltantes o no pertenecen al catálogo	Media
Estado Civil / Estado del cliente	Conformidad	Reglas de negocio no cumplidas	Media
Duplicidad de Celular, Correo, ID	Unicidad	Registros duplicados con impacto operativo	Media
Dirección principal	Validez	Registros con datos incompletos o inválidos	Media

Nota: (Autores, 2025).

Accionables de remediación

Remediación correctiva:

Se delimita el alcance a nivel de capa analítica (datalake) en las capas silver y gold, realizando los correctivos mediante la obtención de data correcta de otras fuentes oficiales. Esta estrategia se justifica debido a la naturaleza de la capa bronze, en la cual no se realiza limpieza ni transformación y para la remediación correctiva de estos datos históricos en las fuentes, se comunica formalmente los hallazgos de baja calidad al Data Steward, para que los coordine oportunamente con las áreas de negocio, después de esto automáticamente se remediarán los problemas de calidad desde las fuentes.

Tabla 8

Actividades para remediación de variables críticas

Actividad	Detalle	Ámbito de aplicación
Georreferenciar coordenadas	Basado en las coordenadas de inicio de sesión obtener la ubicación de mayor permanencia basado en franjas horarios.	Capa gold
Validar ubicación vs. catálogos	Comparar y corregir parroquia, cantón, provincia, país usando catálogos oficiales.	Capa silver
Limpieza de duplicados	Usar técnicas de deduplicación exacta y fuzzy matching sobre celular, ID y correo.	Capa silver
Corrección de valores de Estado Civil / Estado Cliente	Reglas programadas para establecer relaciones válidas.	Capa silver
Revisión de formatos	Corregir longitud de cédula, formato de correo, número de celular, etc.	Capas silver, gold

Nota: (Autores, 2025).

Remediación preventiva:

Para cubrir la remediación preventiva se requiere de la intervención de las áreas de producto, quienes, con los hallazgos encontrados deben establecer los mecanismos adecuados dentro de los desarrollos, para garantizar que no vuelvan a ocurrir. En esta investigación el alcance llega hasta: 1. La comunicación formal de los hallazgos al Data Steward quien es el encargado por parte de gobierno de datos de escalar los errores de calidad y velar por el cumplimiento de las acciones correctivas preventivas y 2. El monitoreo continuo de la calidad desde las fuentes de datos. Por tal razón no se plantean actividades dentro de este apartado de remediación preventiva.

Roles y responsabilidades.

Cabe recalcar que toda remediación es dirigida por el Data Steward, quien para los casos de remediación a nivel de capa analítica dispondrá de Data Engineers para su ejecución, mientras que, para remediación a nivel de fuentes, coordina con los Data Owner.

Tabla 9

Roles y responsabilidades para la gestión de calidad de datos

Rol	Responsabilidad
Data Steward	Coordina y supervisar ejecución de remediación, validar reglas.
Product Owner	Aprobar reglas de negocio y priorización de focos.
Data Engineer	Implementar validaciones, limpieza y pipelines.
Data Qualty Engineer	Medir KPIs, generar alertas y reportes de seguimiento.
Software Developer	Implementar las reglas a nivel de aplicación para ambos tipos de remediación, correctiva y preventiva.
Chief Data Officer	Asegurar recursos y alinear con objetivos del negocio.

Nota: (Autores, 2025).

Resultados actuales

Posterior a la ejecución de gran parte de los planes de remediación se llega en un corto plazo al 95.1%, iniciando como se mencionó anteriormente por los accionables dependientes del equipo de datos es decir a nivel del datalake, en las capas silver y gold. Se realiza una nueva medición y comparando con los resultados anteriores se tiene:

Tabla 10

Comparativa de los resultados por dimensión al inicio (Baseline) y en la actualidad (Después de remediación)

Dimensión	Baseline	Después de remediación
Completitud	93,49%	93,56%
Unicidad	93,87%	99.80%
Validez	87,5%	88,13%
Conformidad	90,68%	98,95%

Nota: (Autores, 2025).

$$IDC\ Clientes\ (actual) =$$

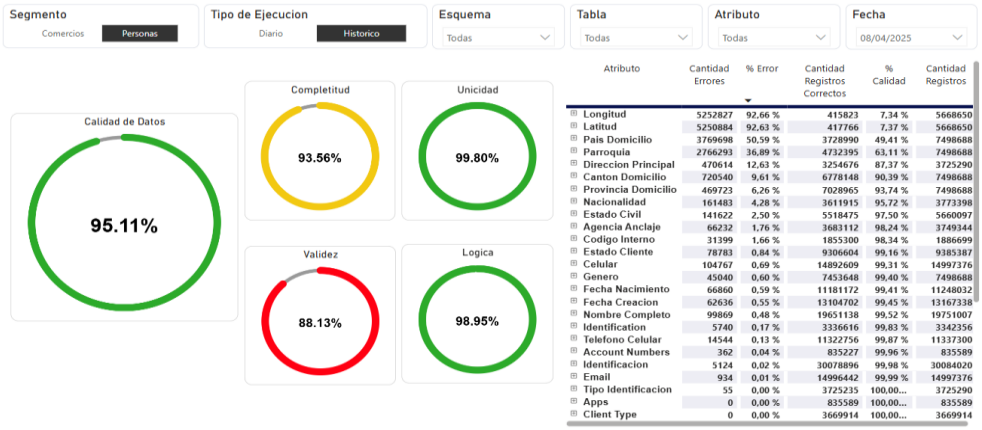
$$(93.56 * 0.25) + (99.80 * 0.25)$$

$$+(88.13 * 0.25) + (98.95 * 0.25)$$

$$IDC\ Clientes\ (actual) = 95.11\%$$

Figura 4

Resultados actuales registrados en el dashboard de monitoreo



Nota: (Autores, 2025).

Discusión

Los resultados arrojados por el artefacto en su etapa de evaluación inicial (baseline) permitió identificar claramente los principales puntos de dolor en cuanto a la calidad de datos del dominio “Clientes”. Un índice de calidad inicial del 91.3% no parecería malo al contrario es bueno con oportunidades de mejora, principalmente en atributos relacionados con ubicación, estado del cliente y estado del cliente.

Como se mencionó con anterioridad un diferencial importante que representa uno de los aportes más relevantes del presente estudio fue la incorporación de la gestión de planes de

remediación. Siendo clave la sinergia entre los actores involucrados permitió iniciar un proceso de mejora continua. Por tanto, se puede responder afirmativamente a la pregunta de investigación, pues se verificó que la aplicación de una metodología para la medición de calidad de datos y la ejecución de planes de remediación sí permite mejorar la calidad de datos. Este caso de éxito aplicado al dominio de "Clientes" ahora mismo ya se está extendiendo hacia otros dominios de información.

La mejora observada de 91,3% a 95,1% en el índice de calidad, sumada a la identificación detallada de problemas específicos y la articulación de equipos técnicos para su corrección, confirman que una gestión activa y sistemática genera un impacto positivo y medible.

El incremento de un 3.8% respecto a la evaluación inicial, es pequeño y modesto, pero de gran importancia si se considera que fue logrado en un entorno real de producción, con datos activos y las limitantes de participación interdepartamental. Además, las metodologías adoptadas Ciencia del Diseño (DSR) y TDQM demostraron ser eficaces para construir un artefacto replicable, escalable y con valor práctico.

Otro aspecto para destacar es la utilidad del dashboard de monitoreo, el cual ha permitido visualizar la calidad de los datos de forma granular, facilitando la toma de decisiones para los responsables técnicos y de negocio. Sin embargo, también se evidenciaron limitaciones importantes: la ausencia de alertas automáticas y la dependencia de procesos manuales tanto para el seguimiento como para la remediación.

Conclusión

Este estudio ha demostrado que la aplicación de una metodología sistemática y práctica para la gestión de calidad de datos en una Fintech permite no solo diagnosticar, sino también mejorar la calidad de la información crítica para el negocio. La integración del enfoque TDQM

con la Ciencia del Diseño (DSR) permitió construir una solución orientada a resultados concretos, validada empíricamente y apoyada en herramientas modernas como Great Expectations y Power BI.

Uno de los principales logros fue evidenciar que los planes de remediación son una pieza clave para cerrar el ciclo de calidad, logrando pasar de un diagnóstico del 91,3% a un 95,1% de calidad global en el dominio "Clientes". Además, se reforzó la importancia de involucrar distintos roles organizacionales Data Stewards, Data Engineers, Product Owners para sostener las mejoras alcanzadas.

No obstante, el sistema aún tiene oportunidades de mejora, como la automatización de alertas y la remediación directa desde las fuentes. Estos elementos se contemplan como siguientes pasos, así como la extensión de este modelo a otros dominios de datos dentro de la organización.

Referencias bibliográficas

Andrew Black (Van Nederpelt & Black), & Peter van Nederpelt (Van Nederpelt & Black). (2020). Dimensions of Data Quality (DDQ). 1–113. https://www.dama-nl.org/wp-content/uploads/2020/09/DDQ-Dimensions-of-Data-Quality-Research-Paper-version-1.2-d.d.-3-Sept-2020.pdf

Araque González, G. A., Gómez Vásquez, M., Vélez Uribe, J. P., & Suárez Hernández, A. H. (2021). Big Data y las implicaciones en la cuarta revolución industrial - Retos, oportunidades y tendencias futuras. Revista Venezolana de Gerencia, 26(93), 33–47. https://doi.org/10.52080/rvg93.04

Ardagna, D., Cappiello, C., Samá, W., & Vitali, M. (2018). Context-aware data quality assessment for big data. Future Generation Computer Systems, 89, 548–562. https://doi.org/10.1016/J.FUTURE.2018.07.014

Bena, Y. A., Ibrahim, R., & Mahmood, J. (2024). Current Challenges of Big Data Quality Management in Big Data Governance: A Literature Review. Lecture Notes on Data Engineering and Communications Technologies, 210, 160–172. https://doi.org/10.1007/978-3-031-59711-4_15

Bicevskis, J., Bicevska, Z., Nikiforova, A., & Oditis, I. (2018). An Approach to Data Quality Evaluation. 2018 5th International Conference on Social Networks Analysis, Management and Security, SNAMS 2018, 196–201. https://doi.org/10.1109/SNAMS.2018.8554915

Calabrese, J., Esponda, S., Pasini, A. C., Boracchia, M., & Pesado, P. M. (2019). Guía para evaluar calidad de datos basada en ISO/IEC 25012. XXV Congreso Argentino de Ciencias de La Computación, 1288–1296. http://sedici.unlp.edu.ar/handle/10915/91086

DAMA International. (2017). DAMA-DMBOK: Data Management Body of Knowledge (2nd ed.).

Herrera, Y. M., & Kapur, D. (2007). Improving data quality: Actors, incentives, and capabilities. Political Analysis, 15(4), 365–386. https://doi.org/10.1093/pan/mpm007

McMann, K., Pemstein, D., Seim, B., Teorell, J., & Lindberg, S. (2022). Assessing Data Quality: An Approach and An Application. Political Analysis, 30(3), 426–449. https://doi.org/10.1017/pan.2021.27

Nikiforova, A. (2020). Definition and Evaluation of Data Quality: User-Oriented Data Object-Driven Approach to Data Quality Assessment. BALTIC JOURNAL OF MODERN COMPUTING, 8(3), 391–432. https://doi.org/10.22364/bjmc.2020.8.3.02

Vishnu Sakthi, D., Valarmathi, V., Surya, V., Karthikeyan, A., & Malathi, E. (2024). Bigdata clustering and classification with improved fuzzy based deep architecture under MapReduce framework. Intelligent Decision Technologies, 18(2), 1511–1540. https://doi.org/10.3233/IDT-230537

Wulff, K., & Finnestrand, H. (2023). Data-driven information for action. Gruppe. Interaktion. Organisation. Zeitschrift Fur Angewandte Organisationspsychologie, 54(1), 65–77. https://doi.org/10.1007/s11612-023-00666-9

Zibak, A., Sauerwein, C., & Simpson, A. C. (2022). Threat Intelligence Quality Dimensions for Research and Practice. Digital Threats: Research and Practice, 3(4). https://doi.org/10.1145/3484202