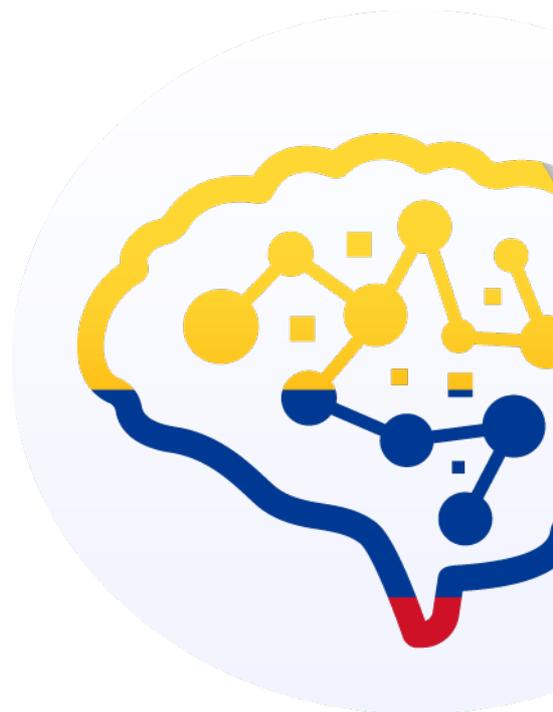


El panorama actual de la calidad de datos en Colombia

Tendencias y estrategias en la gestión de la calidad de datos en Colombia



Índice

1. Ecosistema de datos Colombia
2. Un análisis del estado actual
3. ¿Qué dificulta la implementación de la calidad de datos en Colombia?
4. Calidad de los datos abiertos
5. Casos de éxito en la gestión de calidad de datos.

Ecosistema de datos Colombia

Población	51.609.474 habitantes
No. microempresas	5.597.316
No. pequeñas empresas (≤ 10 empleados)	81.725
No. medianas empresas (10-50 empresas)	19.100
No. grandes empresas (>200 empleados)	6.167
Disponibilidad de Datos Abiertos gubernamentales	https://www.datos.gov.co/ https://www.dane.gov.co/
Disponibilidad de callejero	IDECA: Callejero/portalero de Bogota Alcaldia Medellin: Callejero/portalero Medellin IGAC: Callejero a nivel de Comuna Catastros: Callejeros sin geografía
Niveles administrativos	Departamento/Municipio/Comuna/Barrio
Disponibilidad Información Catastral	Catastro desagregado
Ratio Calidad Datos	Bajo
Fiabilidad Servicio Postal	3/5
Legislación sobre privacidad	Ley Estatutaria 1581 17 de octubre de 2012 de Protección de Datos Personales y su reglamentaria 1377 de 2013
Fichero Oficial Negativo AntiSPAM	No existe

Elaboración propia

Un análisis del estado actual

El 55% de los de datos de las grandes empresas de Colombia tienen problemas de calidad, principalmente debidos a incorrecciones, desactualizaciones y ausencia de datos clave, según un estudio interno de Deyde DataCentric.

Debido a la digitalización y al gran volumen de captura de registros de datos en los últimos años, y teniendo en cuenta que el ciclo de vida de los datos en las empresas hace que rápidamente la información se quede desactualizada, los equipos de datos se enfrentan a ficheros con muchas complicaciones de calidad. Este problema afecta a la contactabilidad con clientes, la toma de decisiones y la propia imagen de las empresas.

La dificultad más común que se produce es la falta de visión única del cliente. Es decir, los empresarios se encuentran con registros duplicados de una misma persona. Esto se produce por tener varias fuentes de entrada de datos, por ejemplo, el sitio web, formularios de landings, publicidad en redes sociales... La mayoría de estos registros están llenos de discrepancias e incoherencias hasta que consigue relacionar y cribar la información correcta con la persona correcta.



El 55% de los de datos de las grandes empresas de Colombia tienen problemas de calidad

Otra problemática está relacionada con las direcciones de correo electrónico. Muchos emails que se introducen en formularios contienen errores o no existen. Es muy común que al escribir las direcciones se comenten errores de escritura, entren caracteres no permitidos o estén mal formateadas (ausencia de arrobas, espacios, etcétera).

Por último, debemos hablar de las direcciones postales. La nomenclatura urbana en Colombia se basa en un modelo numérico. A nivel administrativo, existen callejeros y catastros, pero utilizan diversas fuentes de datos. Además, no fue hasta 2013 que se empezó a utilizar el código postal como iniciativa para facilitar y automatizar el encaminamiento de los envíos postales. Todo esto contribuye a una baja calidad de datos postales.

A estas alturas en la transformación digital del país, las empresas colombianas no se pueden conformar con solo capturar información. Como tal, se centran en prácticas innovadoras de automatización de procesos mediante, por ejemplo, la gestión de inteligencia artificial.



¿Qué dificulta la implementación de la calidad de datos en Colombia?

Sin duda, existe un gran interés entre las empresas colombianas en convertirse en organizaciones guiadas por datos y sacar el mayor provecho de los mismos. Pero el camino para llegar a ello no es sencillo.

Según los CEOs de la región, las principales limitantes para la aplicación de una estrategia basada en datos de calidad son menos de origen tecnológico y sí más humanas. En primer lugar, aunque muchas compañías están informatizadas, hay una falta de "alfabetización" analítica de los cuerpos directivos.

Los datos de la encuesta de IDC de agosto de 2021, Global Future of Intelligence, muestran que solo el 21% de los ejecutivos, el 20.7% de los gerentes y el 14.7% de los trabajadores fueron calificados como completamente alfabetizados en datos.

La única manera de franquear esta barrera es proporcionar una formación a los directivos para que ellos entiendan qué significa tomar decisiones en base a datos.

Más aún, que conozcan la diferencia entre trabajar con y sin datos y cuáles son los beneficios esperables en el primer caso. Si no logramos este cambio, lo más probable es que los directivos no entiendan la propuesta de la consultora. O que terminen comprometiendo un presupuesto para, por ejemplo, una solución de Inteligencia Artificial, que se convierta en una pesada carga dada la falta de objetivos concretos de negocio.

Hasta 2018, en Colombia era mínimo el número de líderes empresariales entendía y había implementado soluciones de analítica, según lo establecido por el Observatorio de las TIC. Hoy el panorama es muy diferente, pues ya se ha multiplicado por seis los líderes que tienen incorporado modelos analíticos en sus organizaciones (según la más reciente Encuesta de Transformación Digital de la ANDI).

1 Falta de comunicación y desconocimiento sobre los datos

Por otra parte, también es perceptible la falta de comunicación de las estrategias comerciales en algunas empresas. A modo de ilustración, los científicos de datos no saben cómo puede ganar dinero la compañía con los datos y los líderes de las empresas tampoco. Esto se debe tanto a la falta de formación de la que hablamos, pero también al poco interés en comunicar las estrategias comerciales. Razón por la que los científicos de datos pueden quedar en una suerte de limbo y no podrán aportar suficiente valor.

El objetivo de convertir a las empresas en **datacentric** es que todos sus integrantes tomen decisiones basadas en evidencia. Sin embargo, los colaboradores de las empresas no usan lo que no conocen. Y se preguntan frecuentemente: ¿qué datos tenemos?, ¿dónde están?, ¿qué **calidad** tienen estos datos?, ¿podemos usarlos o no?

Según el estudio [The data-powered enterprise de Capgemini](#), solo el 20 % de los responsables de toma de decisiones confían ciegamente en su data. Los datos confiables son fundamentales para la agilidad organizacional y la capacidad de una organización para monetizar los datos. Mientras que el 56% de los ejecutivos técnicos creen que sus datos/la estrategia de análisis está alineada con la estrategia comercial, solo el 38% de los usuarios comerciales comparten esa opinión.



Solo el 20 % de los responsables de toma de decisiones confían ciegamente en su data

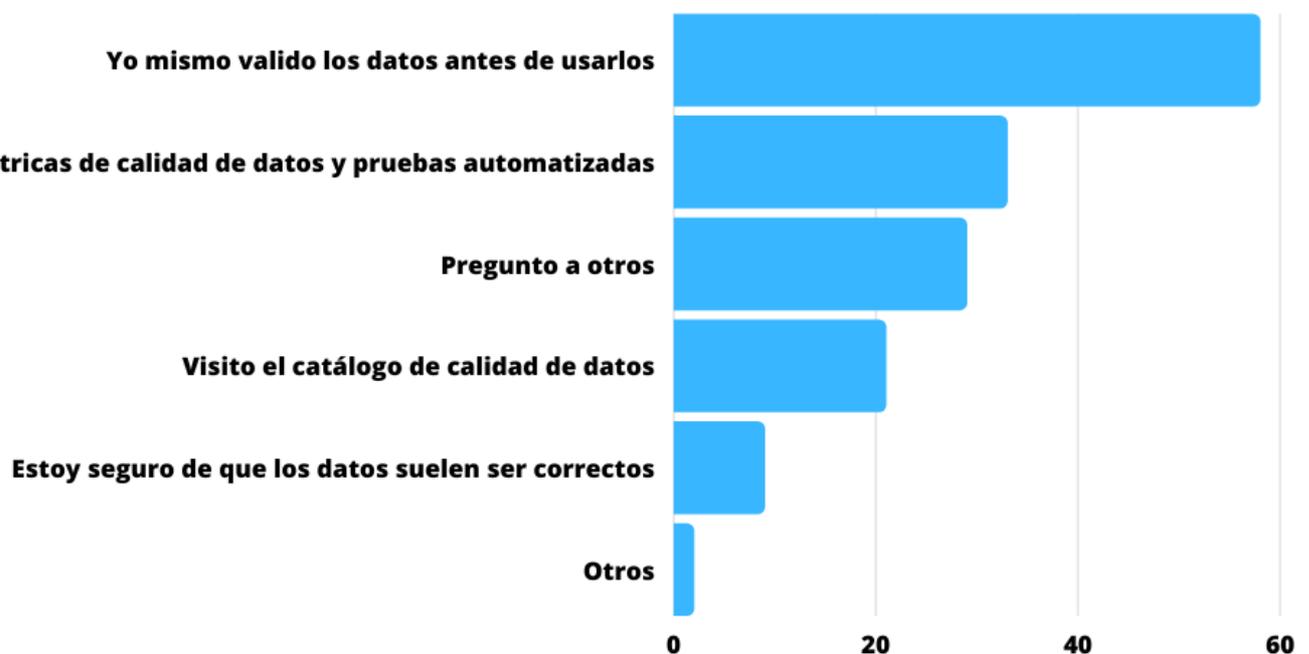
No basta con crear un repositorio de datos centralizado. A esto debemos sumar una capa de visibilidad y entendimiento que proporcione respuestas a las preguntas expuestas. Esto solo se logra con **un programa de data quality sólido** y consistente.

la calidad y la confiabilidad de los datos es el principal KPI (45%) para los equipos que trabajan directamente con datos, seguido de la democratización de los datos (43%).

A pesar de estos resultados, parece que cuesta tomar cartas en el asunto y solucionar el problema. En una encuesta realizada sobre el estado de la calidad de los datos en 2021 por [Datafold](#) se revela que la mayoría de los equipos aún confían en las verificaciones manuales de datos o en preguntar a otros antes de usar los datos para su trabajo. Las pruebas automatizadas y los catálogos de datos son utilizados actualmente por aproximadamente el 30 % y el 20 % de los equipos, respectivamente, como fuente de confianza para la calidad de los datos.

¿Cómo compruebas la calidad de los datos que utilizas?

3 mejores respuestas permitidas



Fuente: Datafold

2 Cambio cultural en las organizaciones

Se ha podido percibir que muchas organizaciones van directamente a implementar las soluciones más avanzadas: IA, machine learning, big data y analítica. Pero esto es apenas la punta del iceberg. Esto lo podemos ver en una encuesta llevada a cabo por [KPMG](#). Para el 85% de los encuestados y el 71% del grupo global, su estrategia de inversión digital se define como “agresiva”. Sin embargo, para el 60% es cierto afirmar que uno de los problemas está en “decidir sobre la tecnología adecuada”.

En la base, deben estar las disciplinas fundamentales como son: el gobierno de datos, la gestión de metadatos, datos maestros y datos de referencia. Igualmente, debemos considerar, la calidad de datos y su seguridad.

Esto no quiere decir que las organizaciones no puedan introducir las soluciones digitales que mencionamos sin tener una calidad de datos cimentada.

Lo que sí es cierto es que gestionar las mismas les llevará más tiempo, les resultará más costoso y asumirán más riesgos. Al final, aportarán menos valor. Desde esta perspectiva, hay que entender que contar con datos de buena calidad comprensibles y plasmables es un habilitador de la democratización de los datos en las organizaciones. De hecho, al no aplicarlo, lo más probable es que la capacidad de trabajar con datos sea excluyente y se reduzca a los científicos de datos o al equipo de data.

Para corregir lo anterior, es necesario hacer énfasis en el cambio cultural dentro de las empresas. ¡Es imprescindible atender el factor humano de los datos! En efecto, las personas y los procesos siguen siendo los principales obstáculos en la conversión de las empresas en **data driven**. Por tanto, es indispensable desmitificar una gran cantidad de paradigmas y vicios que tienen las compañías. En particular, aquellas más tradicionales que están transitando hacia la **transformación digital**. Esto hará posible aplicar las buenas prácticas de datos en las organizaciones.

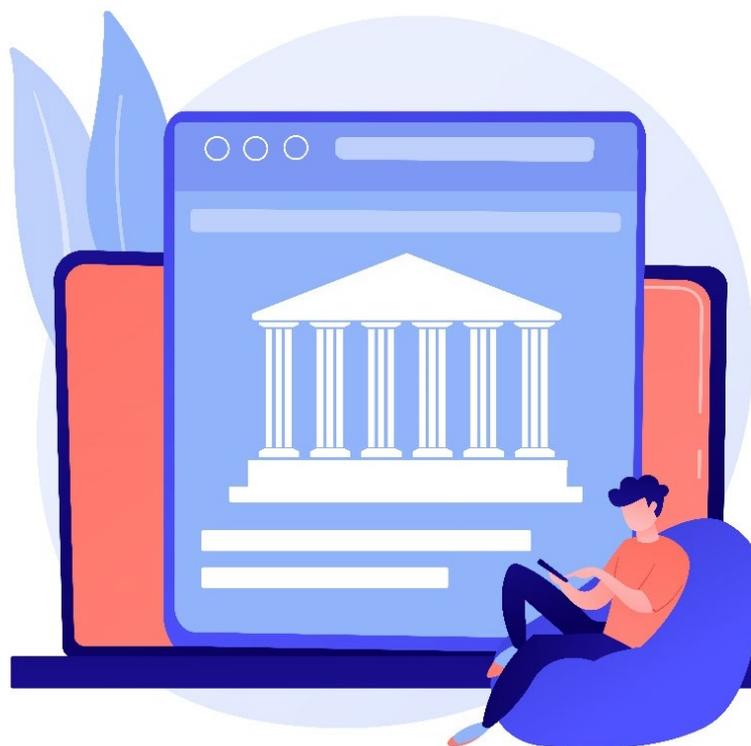


Calidad de los datos abiertos

La Gobernanza de la Iniciativa Datos Abiertos en Colombia está respaldada por el nuevo enfoque de la política de Gobierno Digital, que busca la reconstrucción de los lazos de confianza entre el ciudadano y el Estado.

La Gobernanza de la Iniciativa Datos Abiertos en Colombia está respaldada por el nuevo enfoque de la política de Gobierno Digital, que busca la reconstrucción de los lazos de confianza entre el ciudadano y el Estado. El Ministerio de Tecnologías de la Información y las Comunicaciones (TIC) ha dispuesto el Portal de Datos del Estado Colombiano <http://www.datos.gov.co/>, como el espacio digital para la disposición de la información pública estructurada del país y el punto de encuentro del ecosistema para su uso y aprovechamiento. Éste es operado y administrado por la Dirección de Gobierno Digital y reúne las herramientas que se proveen a los ciudadanos y entidades para publicar, acceder y usar los datos abiertos de Gobierno.

Desde 2021 el Ministerio TIC ha estado trabajando en la mejora de los conjuntos de datos en el portal. Estas mejoras se incluyen en el proyecto encaminado a Diseñar e Implementar la estrategia de calidad de datos abiertos.



Open Data, para todos y de calidad

A partir de la hoja de ruta y el análisis de la calidad de los datos del portal datos.gov.co, se incluyó la migración, puesta en operación y el acompañamiento técnico especializado sobre la herramienta de calidad de datos y la mejora al material. También se brindan talleres de formación y capacitación que permita el avance en la apertura, calidad, uso y explotación de los datos públicos del estado. Todo este proyecto se respalda en labores de analítica, monitoreo y mejora continua de la calidad de los conjuntos de datos publicados en el portal de datos abiertos.

¿Cómo mejoran la calidad de dichos datos? [La Guía estándares de Calidad e Interoperabilidad de Datos Abiertos](#) explica que la segunda fase en la estrategia de apertura se basa en “Monitorear la calidad y el uso”. Es decir, estandarizar la calidad y mejorar la interoperabilidad de los datos abiertos publicados o a publicar en el portal de datos abiertos del gobierno colombiano utilizando dieciséis criterios:



Ministerio TIC Colombia

Según estos criterios, el gobierno de Colombia ha publicado el [diagnóstico de calidad en el portal](#). En 2022 cuenta con 5.439 conjuntos de datos y 102 mil registros. La calidad de sus datos se sitúa en 8.71 puntos sobre 10. En comparación con el año pasado, el puntaje general es algo menor. En 2021 contaba con 3638 conjuntos de datos, 171 mil registros, con un puntaje medio de 9.18 puntos.

Sello de Excelencia Gobierno Digital

Además, el Gobierno colombiano cuenta con el [Sello de Excelencia de Gobierno Digital](#). Es un modelo de madurez que garantiza y certifica la alta calidad de los productos y servicios digitales de las entidades públicas del Estado Colombiano.

Para crear los más altos estándares de calidad, se plantea y se desarrolla el modelo del Sello de Excelencia del Gobierno Digital, el cual fue creado a través del decreto 2573 del año 2014 y adoptado en la resolución 2405 de 2016, en donde se define como un modelo de certificación que busca garantizar la alta calidad de los servicios digitales del Estado colombiano.

El Sello de Excelencia certifica productos y servicios de entidades públicas que han liberado sus datos en el portal de datos abiertos del gobierno colombiano. Las tres entidades con mayor puntuación según el Sello de Excelencia son el Departamento Administrativo de la Función Pública, el Ministerio de Relaciones Exteriores y el Ministerio de Trabajo.

Con todo esto, Colombia se convierte en el tercer país de América Latina y el Caribe con mejor puntuación - 60.47 sobre 100 - en [la IV edición Barómetro de Datos Abiertos](#).



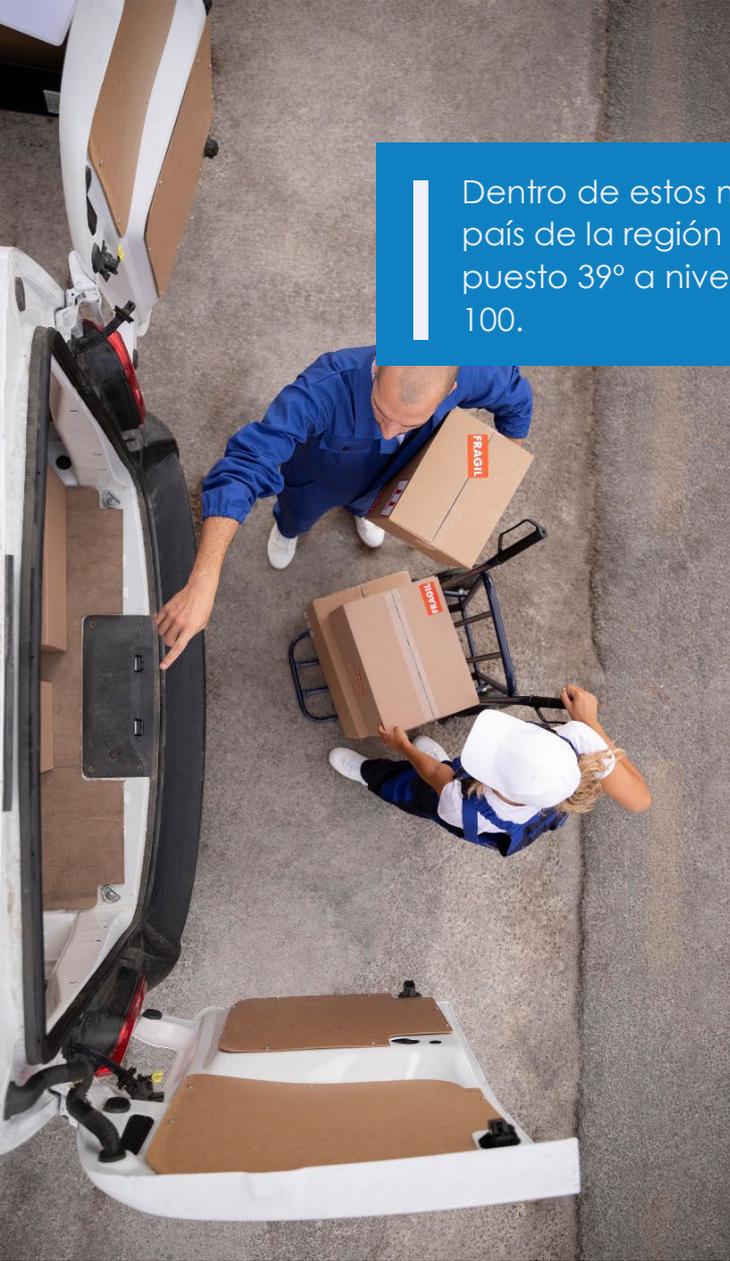
Colombia se convierte en el tercer país de América Latina y el Caribe con mejor puntuación en el Barómetro de Datos Abiertos

Desafío del servicio postal

La industria de servicios postales es uno de los sectores claves de la economía colombiana, y a su vez se pueden ver resultados inferiores comparado a otros sectores de la infraestructura nacional. La fiabilidad del servicio postal en América Latina y el Caribe es la segunda más baja del mundo, solo por encima de África. Además, su accesibilidad está por debajo de la media mundial, 8.1 sobre 100.

La causa de estas calificaciones se basa en un pilar fundamental: la baja calidad de los datos postales. Colombia cuenta con un callejero y un sistema de numeración complejos, donde no se utiliza el código postal y existe mucho ámbito rural.

Además, el callejero y el catastro cuentan con una base de datos única. Son muchas fuentes de información distintas, por lo que son servicios muy desagregados. Por ejemplo, Colombia cuenta con IDECA, el callejero/portalero de Bogotá; la Alcaldía de Medellín cuenta con el callejero/portalero de Medellín, y IGAC, el callejero a nivel comuna.



Dentro de estos malos datos, Colombia se posiciona como el país de la región con mejor desempeño y desarrollo postal, el puesto 39° a nivel mundial con una puntuación de 53.2 sobre 100.

Estos buenos datos dentro del contexto de la región son gracias a las medidas del Gobierno Nacional. Una de estas iniciativas es el “Pacto por la Transformación Digital de Colombia”, incluido en el “Plan Nacional de Desarrollo 2018 - 2022 Pacto por Colombia, Pacto por la Equidad” (PND). pretende utilizar las tecnologías de la información y las comunicaciones como catalizadores que mejoren la calidad de vida de los colombianos. Dentro de los objetivos del pacto se encuentra el de “Crear las condiciones habilitantes para la masificación de las TIC”, haciendo especial énfasis en la transformación del sector postal, en el contexto de la digitalización y las tendencias del comercio electrónico.

Casos de éxito en la gestión de calidad de datos

Liberty Seguros, agiliza las gestiones comerciales mediante una visión precisa de los datos del cliente

Consciente de la actual tendencia de las empresas colombianas en asumir o continuar su transformación digital, en **Deyde DataCentric** aportamos al mercado de la región nuestras soluciones en **calidad de datos**. En especial, **MyDataQ**, una herramienta versátil de desarrollo propio para el tratamiento automatizado en la especialidad de normalización, deduplicación y **enriquecimiento de bases de datos**.

Como una de las principales empresas del sector asegurador en el mundo, Liberty Seguros le ofrece al mercado un amplio portafolio de productos y servicios dentro de dos líneas de productos: seguros para personas y seguros para bienes. La compañía tiene presencia desde hace 20 años en 19 ciudades de Colombia.

Liberty Seguros necesitaba disponer de un sistema de normalización y deduplicación de direcciones postales rápido y fiable para verificar las visitas realizadas por los inspectores de riesgo a los precios asegurados.

MyDataQ con los módulos de normalización de direcciones postales y deduplicación para obtener una visión única del cliente en modo Servicio en Oficinas.

Liberty Seguros obtuvo una base de datos en óptimas condiciones para agilizar las gestiones comerciales y logró tener una visión única de cliente evitando duplicidades.



Liberty
Seguros®

Fincomercio, proceso automático con información completa, confiable y precisa

Fincomercio es una cooperativa de ahorro y crédito con más de 63 años en el mercado, que presta servicios financieros a estudiantes, pensionados, independientes y empleados. Cuenta con convenios con más de 900 empresas para que sus empleados puedan acceder fácilmente a sus beneficios. Tienen más de 210.000 asociados en toda Colombia.

En la delegación de operaciones y mesa de ayuda nace el proyecto de higienización de datos de contacto, donde se requiere optimizar los recursos y el tiempo dedicado a realizar procesos manuales de identificación de datos incorrectos de los asociados.

MyDataQ, se hace imprescindible con los módulos de Names, Addresses, Email y Phones en licencia de uso Batch. Deyde DataCentric con MyDataQ cambia la perspectiva de un proceso manual a un proceso automático generando valor a la higienización de datos y optimización del tiempo en actividades de fidelización de asociados.

Fincomercio consiguió identificar correos electrónicos ficticios y/o vacíos para campañas de e-commerce, y unificar datos de contacto de asociados naturales y jurídicos.

Además, logró obtener una base de datos correcta mediante el estandarizado de direcciones para cumplimiento de estándar de reportes a la DIAN, así como optimizar recursos dedicados a procesos manuales y segmentar asociados de Créditos, Seguros, Ahorro e inversión, logrando una visión única.



deyde  datacentric by Accumin

www.deyde.com

www.datacentric.es

 info@deyde.com

