

**UNIVERSIDAD TECNOLÓGICA DE CHIHUAHUA**  
**TECNOLOGÍAS DE LA INFORMACIÓN Y COMUNICACIÓN**



**REPORTE DE LIMPIEZA DE DATOS**

**MATERIA:** Extracción de Conocimiento en Bases de Datos

**MAESTR@:** Enrique Mascote

**ALUMNO:** Carlos Adrián Mata Nevárez

**MATRICULA:** 1123250015

**GRUPO:** IDGS91N

**FECHA:**04/10/2025

# ÍNDICE

CASO DE ESTUDIO.....	1
PROCEDENCIA DE LOS DATOS .....	2
TIPOS Y FUENTES DE DATOS .....	3
TÉCNICAS DE LIMPIEZA DE DATOS.....	5
FUNDAMENTACIÓN Y ESTRUCTURA.....	7
CONCLUSIÓN.....	8
REFERENCIAS .....	9

## CASO DE ESTUDIO

Los sistemas de préstamos digitales dependen en gran medida de la calidad de los datos que procesan para otorgar créditos de manera rápida, segura y confiable.

Sin embargo, si la información proviene de fuentes diversas y no se gestiona correctamente, pueden aparecer inconsistencias, duplicidades y errores que afectan tanto al cliente como a la institución bancaria.

Así que en este caso ficticio se analiza cómo identificar la procedencia, tipos, fuentes y limpieza de los datos en un sistema de préstamos digitales, con el fin de garantizar la confiabilidad en la toma de decisiones crediticias.

## PROCEDENCIA DE LOS DATOS

La procedencia de los datos determina su origen y es crucial para evaluar su confiabilidad, en el caso de los préstamos bancarios digitales, los datos provienen de diversas fuentes tecnológicas y humanas que, al integrarse, permiten un análisis completo del cliente.

- Datos biométricos: huellas digitales y reconocimiento facial utilizados para verificar la identidad del cliente al acceder a la plataforma.
- Datos máquina a máquina (M2M): validaciones automáticas entre el sistema bancario y registros externos (burós de crédito, sistemas gubernamentales de identificación).
- Datos de transacciones: historial de pagos, movimientos en cuentas bancarias y comportamiento crediticio.
- Datos generados por humanos: información proporcionada directamente por el cliente (nombre, dirección, ingresos, referencias personales).
- Datos web: formularios en línea completados en la aplicación bancaria.
- Datos de medios sociales: en algunos modelos experimentales, se utilizan redes sociales como fuente complementaria para analizar patrones de comportamiento y confiabilidad.

## TIPOS Y FUENTES DE DATOS

La clasificación de los datos nos permite entender su naturaleza y como deben de ser gestionados dentro del sistema, en el contexto de los préstamos digitales, se trabaja con información tanto estructurada como no estructurada que requiere diferentes formas de tratamiento (acendes, 2025), (stripe, 2024).

Tipos de datos en préstamos digitales:

### **Cuantitativos:**

- Identificadores digitales como dirección IP.
- Datos financieros: Historial crediticio (score numérico), Transacciones financieras anteriores, Cuentas bancarias (número de cuentas o saldos), Información sobre ingresos, préstamos previos (montos).
- Datos del préstamo solicitado: Monto del préstamo, Plazo, Tasas de interés, Comisiones y términos (si son valores numéricos).
- Datos para evaluación y prevención de fraude: Comparación de geolocalizaciones e IPs.

### **Cualitativos:**

- Nominales (Categorías sin orden):
- Datos personales: Nombre, dirección, número de teléfono, correo electrónico.
- Identificadores: Huella digital del dispositivo, datos biométricos o de identificación electrónica.
- Otros: Tipo de préstamo digital (personal, microcrédito, etc.).

### **Ordinales:**

- Historial crediticio tradicional (si se presenta como "excelente," "bueno," "regular").
- Parámetros personalizados para identificar riesgos (e.g., "bajo," "medio," "alto").

**Estructurados:**

- Generalmente incluye todos los datos que van en campos específicos de un formulario o base de datos: Nombre, dirección, número de teléfono, correo electrónico, dirección IP, score crediticio, montos de transacciones, ingresos, monto y plazo del préstamo.

**No Estructurados:**

- Datos biométricos o de identificación electrónica (imágenes o scans).
- Actividad en línea detectada a través del análisis de múltiples cuentas digitales y dispositivos (logs de actividad, contenido de texto).
- Comportamientos de navegación y patrones de uso de aplicaciones (logs de interacción sin procesar).
- Uso del teléfono, correo, redes sociales (contenido y patrones de actividad detallados).
- Análisis del dispositivo utilizado (información detallada del hardware/software).

# TÉCNICAS DE LIMPIEZA DE DATOS

En la etapa de limpieza de datos es la mas critica ya que busca garantizar la calidad de la información antes de que sea utilizada en procesos analíticos o decisiones de crédito. Su aplicación evita errores que pueden generar riesgos financieros o fraudes (Donnelly, 2023), (Naseem, 2024).

## **Eliminación de duplicados**

- En muchas bases de datos es común encontrar registros repetidos, ya sea por errores al capturar información o al integrarla desde distintas fuentes. Para evitar cálculos equivocados o sesgos en los resultados, se eliminan esas entradas duplicadas.

## **Relleno o imputación de datos faltantes**

- Cuando se detectan valores importantes que no fueron registrados, se aplican técnicas de imputación, ya sea con métodos estadísticos o modelos predictivos, con el fin de completar la información y evitar que los análisis pierdan fiabilidad.

## **Corrección de datos erróneos o inconsistentes**

- Se revisan y corrigen errores como faltas de captura, tipografías equivocadas, diferencias en el formato (por ejemplo, fechas escritas de distintas maneras) o valores que no coinciden con lo que sería válido en un rango.

## **Estandarización de formatos**

- Se busca unificar la forma en que se presentan los datos fechas, números de identificación, direcciones, entre otros para que sean consistentes y más fáciles de procesar.

## **Eliminación de datos irrelevantes o redundantes**

- También se descartan aquellos campos que no aportan información útil al análisis. Esto ayuda a reducir ruido y a que los modelos trabajen de manera más eficiente.

## **Validación y verificación de datos**

- Finalmente, se aplican reglas de validación para comprobar que la información cumpla con los criterios esperados: que los valores estén dentro de los rangos correctos, que las relaciones entre campos sean coherentes y que los identificadores realmente sean únicos.

## **Herramientas y enfoques utilizados**

Se emplean diversas herramientas como Excel, Python (con librerías como Pandas), R, OpenRefine o plataformas especializadas en limpieza de datos, que permiten automatizar procesos y corregir grandes volúmenes de información de manera eficiente.

Un dato importante es que implementan validaciones automáticas en los sistemas digitales de préstamos para identificar de inmediato datos inválidos o sospechosos al momento de capturarlos.

Y también se aplican técnicas de detección de fraudes mediante el análisis de patrones y consistencias, con el objetivo de limpiar datos que puedan haber sido manipulados o que presenten irregularidades.



## FUNDAMENTACIÓN Y ESTRUCTURA

La fundamentación conceptual permite comprender la importancia de mantener datos de alta calidad en los sistemas digitales.

Tomando en cuenta el caso de estudio, una correcta gestión de la información no solo garantiza la confiabilidad de los procesos, sino que también contribuye a reducir riesgos financieros, mejorar la eficiencia operativa y fortalecer la transparencia en la toma de decisiones.

Los datos precisos y confiables se convierten en un recurso estratégico que respalda tanto la gestión interna como la experiencia del cliente ya que es primordial si lo que queremos es brindar un servicio de calidad.

Y la calidad de los datos en los sistemas bancarios digitales depende de un manejo adecuado de la información, asegurando que sea precisa, completa, consistente y actualizada.

Para lograrlo, se implementan procesos de limpieza, validación, estandarización y control de integridad, que permiten que los datos estén listos para ser utilizados en análisis de riesgo crediticio, aprobación de préstamos y detección de fraudes.

Una estructura de datos bien definida facilita la integración de información proveniente de múltiples fuentes, como datos biométricos, transacciones financieras, registros humanos, portales web y redes sociales, permitiendo consolidar información de manera organizada y coherente.

Esto no solo optimiza la eficiencia operativa, sino que también fortalece la capacidad de la entidad para realizar análisis predictivos, segmentación de clientes y toma de decisiones más acertadas.

Además, un sistema estructurado de gestión de datos contribuye a la trazabilidad y auditoría de la información, lo que resulta clave para cumplir con normativas legales y regulatorias, así como para garantizar la seguridad y privacidad de los datos de los clientes ya que es información muy sensible.

La fundamentación y la estructura adecuada de la información se convierten, por tanto, en pilares esenciales para un sistema bancario digital confiable, eficiente y orientado al cliente.

## CONCLUSIÓN

La conclusión del análisis nos da a entender que la procedencia de los datos, sus tipos, las técnicas de limpieza aplicadas y la fundamentación conceptual son elementos fundamentales para asegurar la calidad de la información en los sistemas digitales de préstamos bancarios.

Contar con una estrategia integral de gestión de datos no solo optimiza la eficiencia operativa, sino que también disminuye los riesgos de fraude, mejora la exactitud en la toma de decisiones financieras y refuerza la confianza y la reputación de la institución ante sus clientes y reguladores.

Además, la calidad de los datos permite realizar análisis más profundos y oportunos, apoyando la innovación en productos y servicios financieros.

## REFERENCIAS

*acendes*. (4 de Oct de 2025). Obtenido de <https://acendes.com/prestamos-digitales-la-guia-completa-para-implementarlos-en-tu-financiera>

Donnelly, S. (2 de Junio de 2023). *financealliance*. Obtenido de <https://www.financealliance.io/data-cleaning-techniques>

Naseem, O. (19 de Julio de 2024). *datadriveninvestor*. Obtenido de <https://www.datadriveninvestor.com/2024/07/19/data-cleaning-strategies-for-financial-data>

*stripe*. (27 de Junio de 2024). Obtenido de <https://stripe.com/mx/resources/more/digital-lending-101-how-its-changing-the-banking-and-loan-servicing-industry>