

Informe de Análisis de la Calidad de Datos - Ironhack Payments

En el análisis exploratorio realizado para IronHack Payments, se llevaron a cabo varias modificaciones para asegurar la coherencia y la integridad de los datos, garantizando relaciones lógicas y consistentes entre las variables.

1. Valores Nulos

- **Identificación y Tratamiento:** Se identificaron valores nulos en columnas críticas y se trataron adecuadamente.
- **Mantenimiento de Valores Nulos:** En casos donde los valores nulos eran lógicos y representativos del estado de los datos (por ejemplo, valores nulos en la columna `incident_rate` que indican ausencia de incidencias), se mantuvieron para preservar información valiosa.
- **Imputación de `user_id`:** Los valores faltantes en `user_id` se imputaron utilizando `deleted_account_id` cuando `user_id` estaba vacío, asegurando la integridad de las identificaciones de usuarios.

2. Duplicados

- **Verificación:** No se encontraron duplicados en los conjuntos de datos, asegurando la unicidad de cada registro.

3. Valores Atípicos

- **Análisis y Decisión:** Se analizaron los valores atípicos y se decidió mantenerlos para representar la diversidad de las transacciones, reflejando así un panorama más completo del comportamiento de los usuarios.

4. Unión de Datos

- **Integración de Conjuntos de Datos:** Se unieron los conjuntos de datos de solicitudes de efectivo y tarifas utilizando el identificador `id`, facilitando un análisis más integrado y coherente.

5. Eliminación de Columnas Innecesarias

- **Optimización de Datos:** Se eliminaron columnas innecesarias tanto en el conjunto de datos de comisiones como en el de solicitudes de efectivo (`cash request`), simplificando los datos para un análisis más claro y enfocado.

6. Creación de Nuevas Variables

- **Reducción y Simplificación:**
 - **Tiempo de Cobro:** Se crearon nuevas variables que redujeron las columnas relativas a fechas a una única columna indicando el tiempo de cobro.
 - **Tasa de Incidentes:** Se creó una columna que resumiera los datos de incidentes, indicando únicamente si hubo una incidencia o no, simplificando el análisis de incidencias.

Conclusión

Este análisis de calidad de datos proporciona una base sólida para el análisis posterior, asegurando que los datos utilizados sean fiables y representativos de las operaciones de IronHack Payments. Las modificaciones realizadas han mejorado la coherencia y la integridad de los datos, permitiendo un análisis más preciso y significativo de las tendencias y comportamientos de los usuarios.