

Envío a Proveedores

En primer lugar, se genera el envío de casos que se desean enriquecer a los proveedores externos. Una de las tareas es generar un envío semanal con todos los casos en **Estado: No Visto** que existan, para los clientes de Mora Tardía que se indiquen.

En todos los casos (para cualquier proveedor) se deberá enviar un listado de DNIs/CUITs (según corresponda) y Nombre.

1. Generar una query que contenga IDMOROSO, ESTADO, IDCLIENTE y FILTROA, de la tabla CARTERA. Esta query será necesaria para el preprocesamiento posterior de los datos. (Referencia: ver query Envío_semanal_a_enriquecimiento.sql a modo de ejemplo).
2. Limpiar el archivo .csv que se exporte de forma tal que queden solamente las columnas IDMOROSO y NOMBRE. Este será el archivo que se envíe al proveedor.

Si el envío incluye a AMEX:

1. Generar una query que contenga IDMOROSO, ESTADO, IDCLIENTE y FILTROA, de la tabla CARTERA, y DATOADIC de la tabla MOROSO. Esta query será necesaria para el preprocesamiento posterior de los datos.
 - a. El DNI/CUIT de los casos de AMEX se encuentra ubicado en el campo DATOADIC, de la tabla MOROSO.
 - b. Ejecutar el notebook Envío a Proveedores - Preprocesamiento, que mapea los CUITs de personas físicas de la columna DATOADIC a DNIs (por una cuestión de eficiencia en el enriquecimiento, los morosos tienen mejor devolución cuando enviamos DNIs).
 - c. La salida de este notebook será lo que se envíe al proveedor externo. Validar que incluye DNI y NOMBRE.

Data Cleaning

Una vez recibida la devolución del proveedor, se procede con la limpieza de los datos. Para esto, se ejecutan ciertas jupyter notebooks. Cada notebook tiene un código al final de su nombre, perteneciente a la fecha de ejecución; aquella versión de la notebook con la fecha más reciente, corresponderá a la inserción más reciente realizada. En general, no hay demasiadas variaciones entre fechas (aunque en ciertas ocasiones los proveedores suman o restan columnas, lo que ocasiona que haya que realizar ciertas modificaciones a los pipelines).

Para **todos** los proveedores se utilizan los siguientes pipelines:

- Data Cleaning - Telefonos: Limpieza de teléfonos (*).
- Data Cleaning - Mails: Limpieza de Mails (*).

- DAM_LOG: Limpieza de datos para insertar en la tabla DAM_LOG, dentro de Active_testing (que posteriormente impacta en el tablero de enriquecimiento).

Para el proveedor Informes Digitales, los pipelines de teléfonos están separados en diferentes notebooks (fijos, celulares y vecinos).

Para los proveedores VIS/PyP e Insiders (AD) se agregan además las siguientes notebooks:

- Data Cleaning - Inmuebles
- Data Cleaning - Vehículos
- Data Cleaning - Deudas
- Data Cleaning - Scoring

(*) Duplicados

En algunos casos, algunos proveedores nos devuelven duplicados ciertos teléfonos o mails de la siguiente forma:

- El teléfono 1111-1111 aparece en el campo “Teléfono Fijo Personal” pero también en el campo “Teléfono Fijo Laboral”. Por lo tanto, no tenemos un propietario al que podamos atribuirle con certeza ese dato.
- Es por esto, que si al final de los pipelines Data Cleaning - Teléfonos o Data Cleaning - Mails encontramos estos duplicados, los mismos deberán pasar a un pipeline llamado DAM_CONTACTABILIDAD - Teléfonos - Duplicados. Los pasos están indicados en el mismo notebook.
- En el notebook DAM_CONTACTABILIDAD - Teléfonos - Duplicados, simplemente se generan puntajes en función de ciertas reglas definidas, para elegir a quién finalmente atribuimos ese dato.
- El output de este pipeline es el que finalmente se inserta en la tabla DAM_CONTACTABILIDAD.

Inserción de Datos

Una vez realizada la limpieza en cada pipeline, es necesario generar las inserciones correspondientes. Para esto, cada pipeline exporta un archivo .csv que contiene los datos ya listos para insertar en la base de datos (con esto referimos a que el output de cada notebook tiene ya las columnas ordenadas de forma tal que la inserción a la database sea lo más limpia posible).

Dentro del módulo de Datos Adicionales del Moroso (DAM) se insertan datos en las siguientes tablas.

- DAM_CONTACTABILIDAD: Se insertan datos adicionales de contactabilidad para cada moroso. Esta tabla contempla tanto teléfonos (fijos, celulares) como mails.

- DAM_INMUEBLES: Se insertan datos referidos a la cantidad de inmuebles que tiene ese moroso a su nombre.
- DAM_VEHICULOS: Se insertan datos referidos a los vehículos que tiene ese moroso a su nombre.
- DAM_DEUDAS: Se insertan datos que muestran las entidades con las que opera ese moroso, junto con la situación que remite a la misma, informada por el BCRA.
- DAM_DATOS_PERSONALES: Por el momento, simplemente se insertan datos de Scoring Patrimonial de cada moroso.
- DAM_LOG: Es un log de los morosos que fueron enriquecidos históricamente por los diferentes proveedores externos. El tablero de enriquecimiento toma información de esta tabla en particular.

Post-inserción

Una vez que los datos sean insertados, es necesario notificar a los supervisores correspondientes que poseen nuevos datos para gestionar. Para esto, se generan dos queries:

- Una que muestre la totalidad de los casos insertados.
- Otra que muestre la totalidad de casos nuevos insertados.