# Informe del Proceso de Transformación de Datos por Tabla

Cambios en todas las tablas

* Renombramos todas las columnas a fin de tener un acceso normalizado a cada una de ellas.

Tabla Clientes

* Eliminamos la columna ‘col10’ ya que no aporta nada al dataset.
* Capitalizamos las columnas 'nom\_y\_ape', 'provincia', 'localidad' y 'domicilio'.
* Llenamos los nulos de edad con el promedio de todas las edades, a fin de no dejar nulos esos campos y poder transformar esa columna a tipo entero.
* Reemplazamos en latitud y longitud las comas por vacíos.
* Convertimos Latitud y Longitud a float. Los nulos y cadenas de texto dejamos igual ya que no podemos deducir con que completar.
* Normalizamos las provincias.
* Por último, eliminamos los registros que tienen más de dos datos faltantes, debido a que no nos brindan demasiada información.

Tabla Compras

* Completamos los nulos en precio por la media de valores en dicha columna, ya que representan solo el 3% (aproximadamente) de todos los registros.
* Columna 'precio' convertida a float.
* Columnas 'idCompra', 'año', 'mes', 'periodo', 'idProducto', 'cantidad', 'idProveedor' a entero.
* Marcamos los Outliers tanto en precio como en cantidad con una columna destinada a ello. En precio, reemplazamos el valor outlier por su media. En cantidad, reemplazamos el valor por su división por 10, considerando que fue un error de tipeo en el cuál se agregó un cero de más al dato.

Tabla Ventas

* Completamos los nulos en precio con la media de valores de este campo. Esta decisión se debe a que los registros que presentan nulos en este campo no llegan al 2% del total de registros, esto hace que no genere inconvenientes.
* Completamos los nulos en cantidad con la media de valores de este campo. Lo hacemos debido a que los registros con esta característica no llegan al 2% del total de registros, por lo que no afecta nuestro análisis.
* Marcamos los Outliers tanto en precio como en cantidad con una columna destinada a ello. En precio, reemplazamos el valor outlier por su media. En cantidad, reemplazamos el valor por su división por 10, considerando que fue un error de tipeo en el cuál se agregó un cero de más al dato.

Tabla Localidades

* Transformamos Latitud y Longitud en string para reemplazar los puntos por comas y luego devolverlos a float nuevamente.
* Capitalizamos la columna nombreLocalidad.
* Normalizamos los nombres de provincias con el dataset que provee el gobierno.

Tabla Proveedores

* Capitalizamos las tablas dirección, ciudad, estado, país y departamento.

Tabla Sucursales

* Capitalizamos la columna dirección.
* Normalizamos las columnas de localidades y provincias utilizando los datasets normalizados que provee el gobierno.