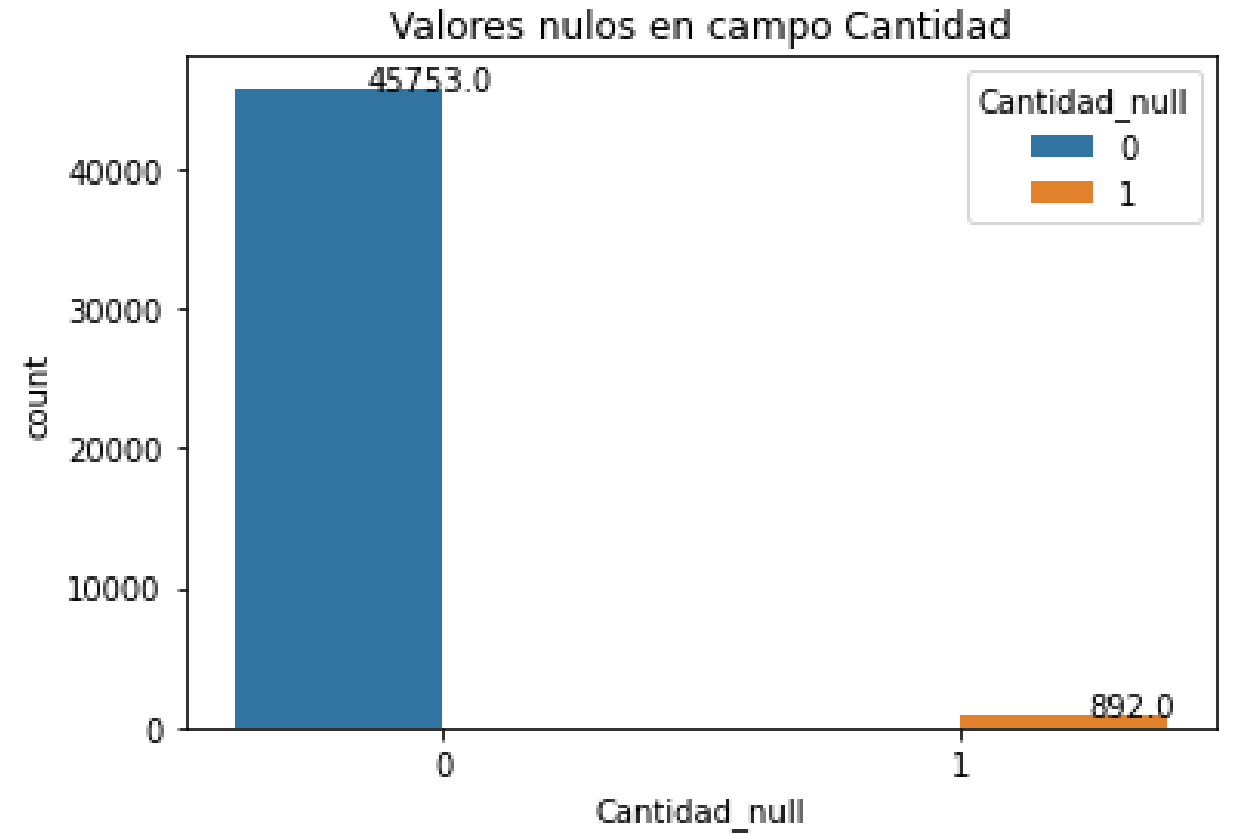
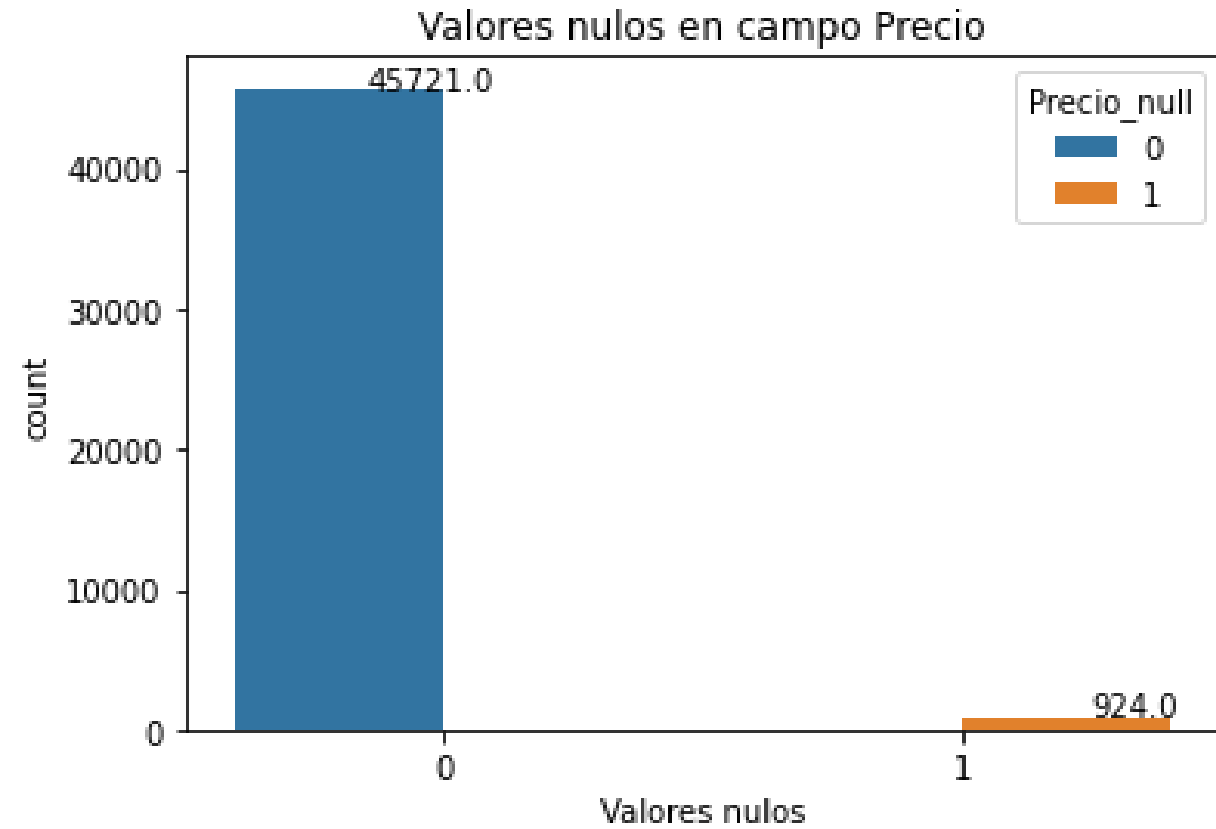


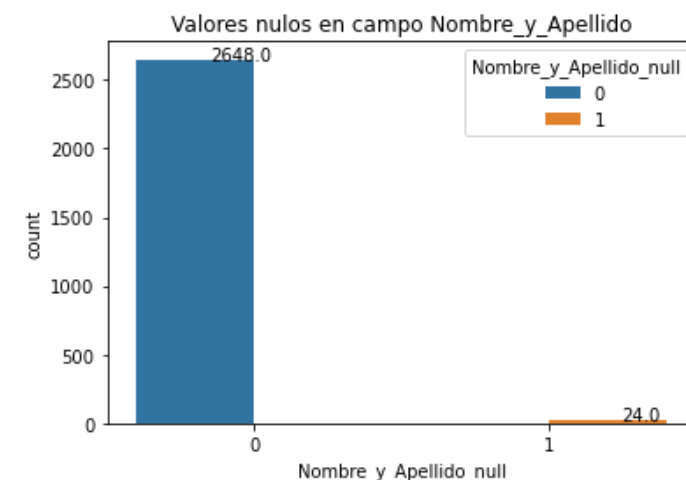
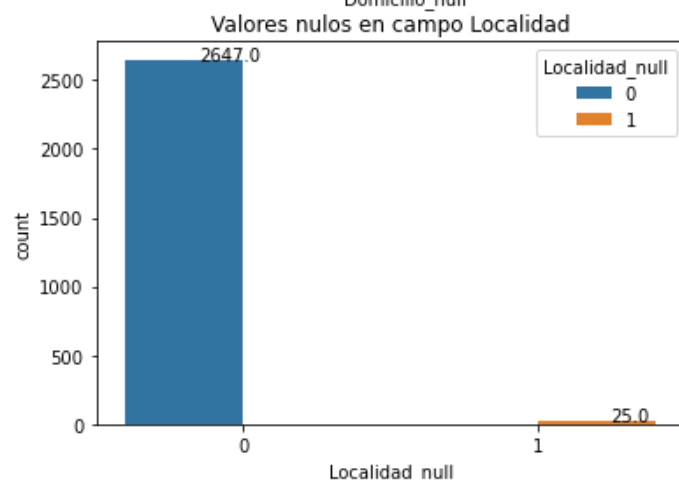
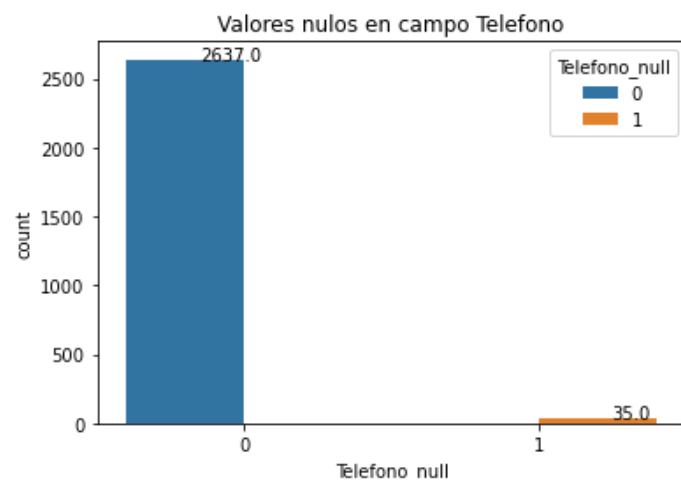
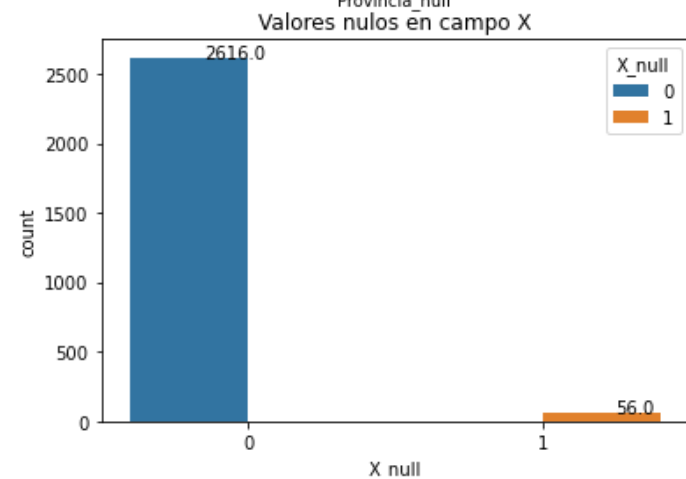
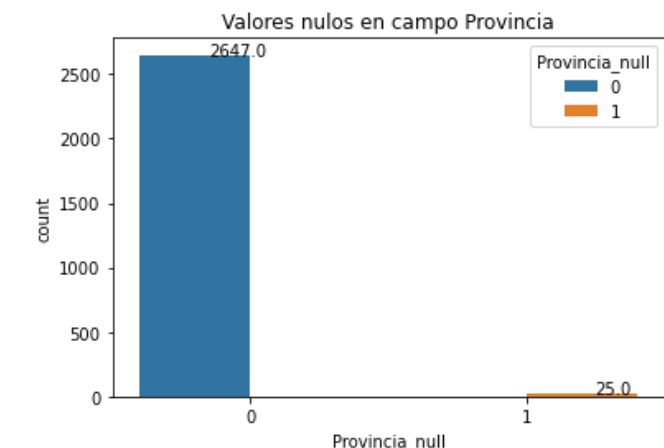
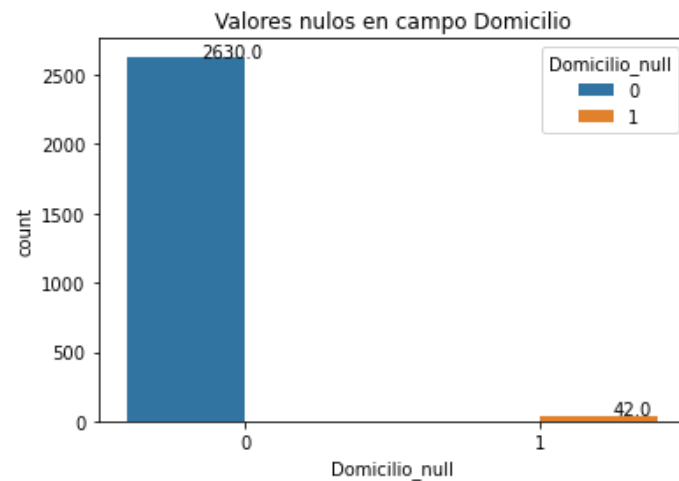
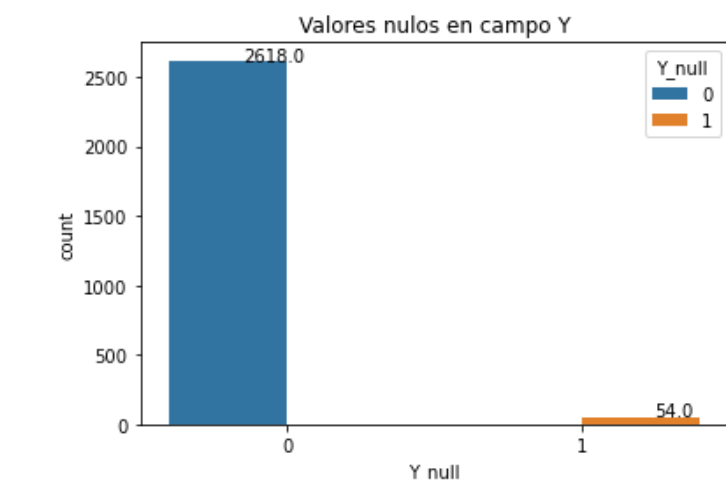
Informe – Calidad de los datos

Francisco Daniel Lugo Reyes

Valores faltantes (2%) – Tabla Venta

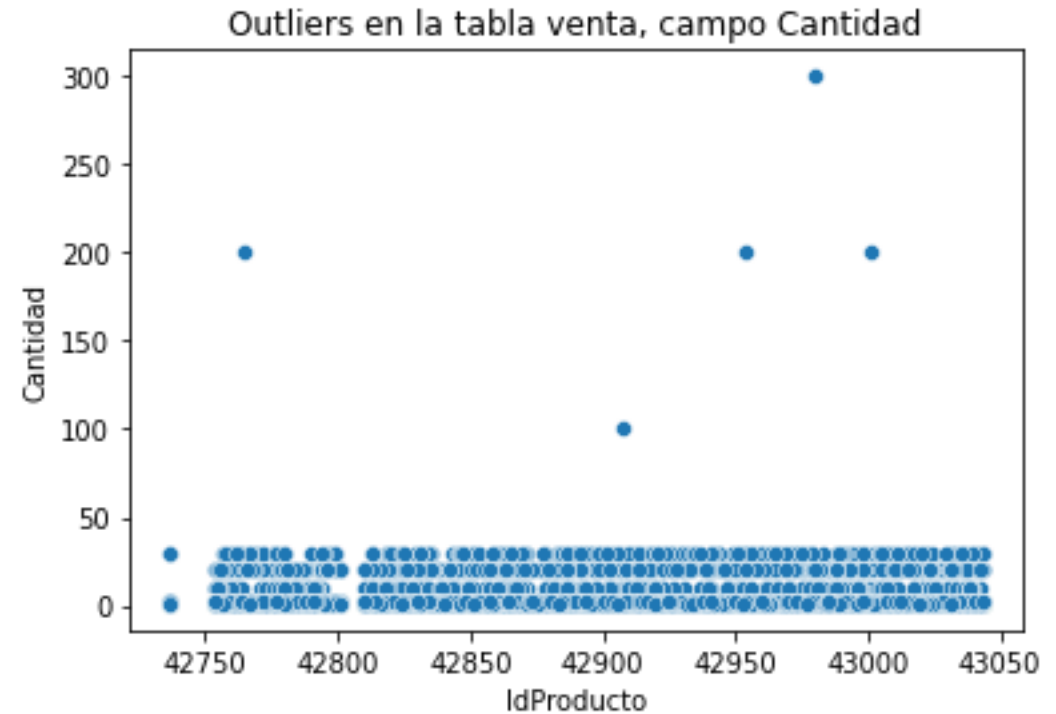
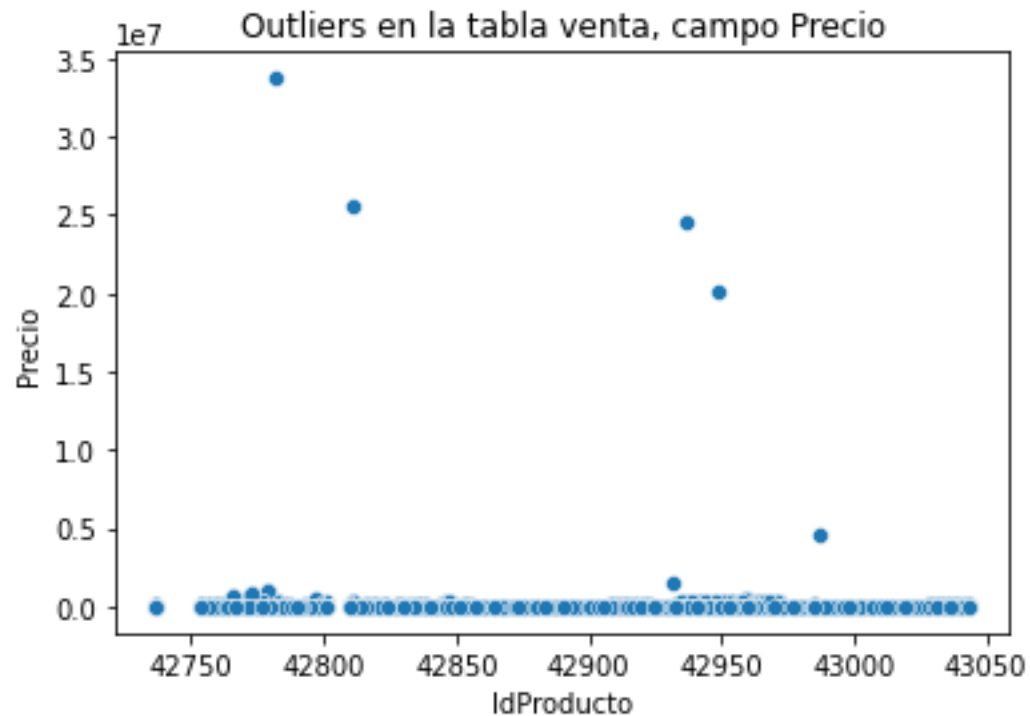


Valores faltantes (~2%) – Tabla Cliente



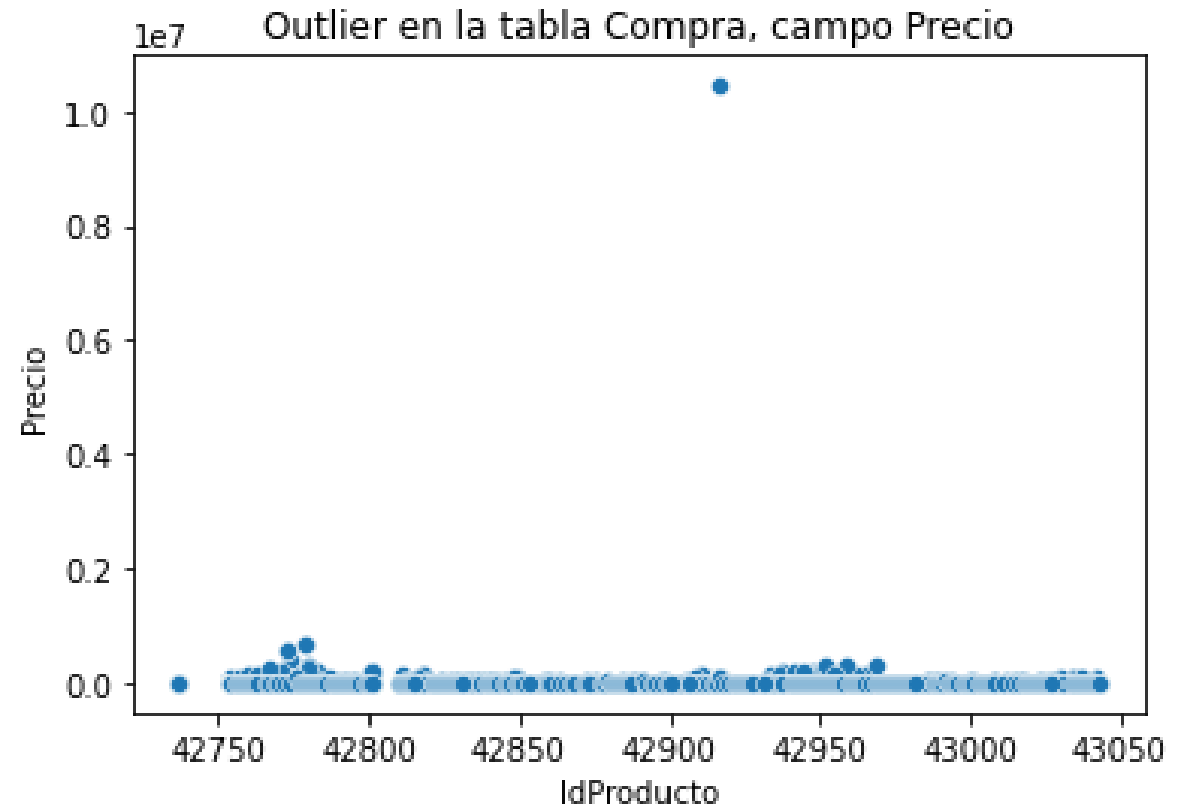
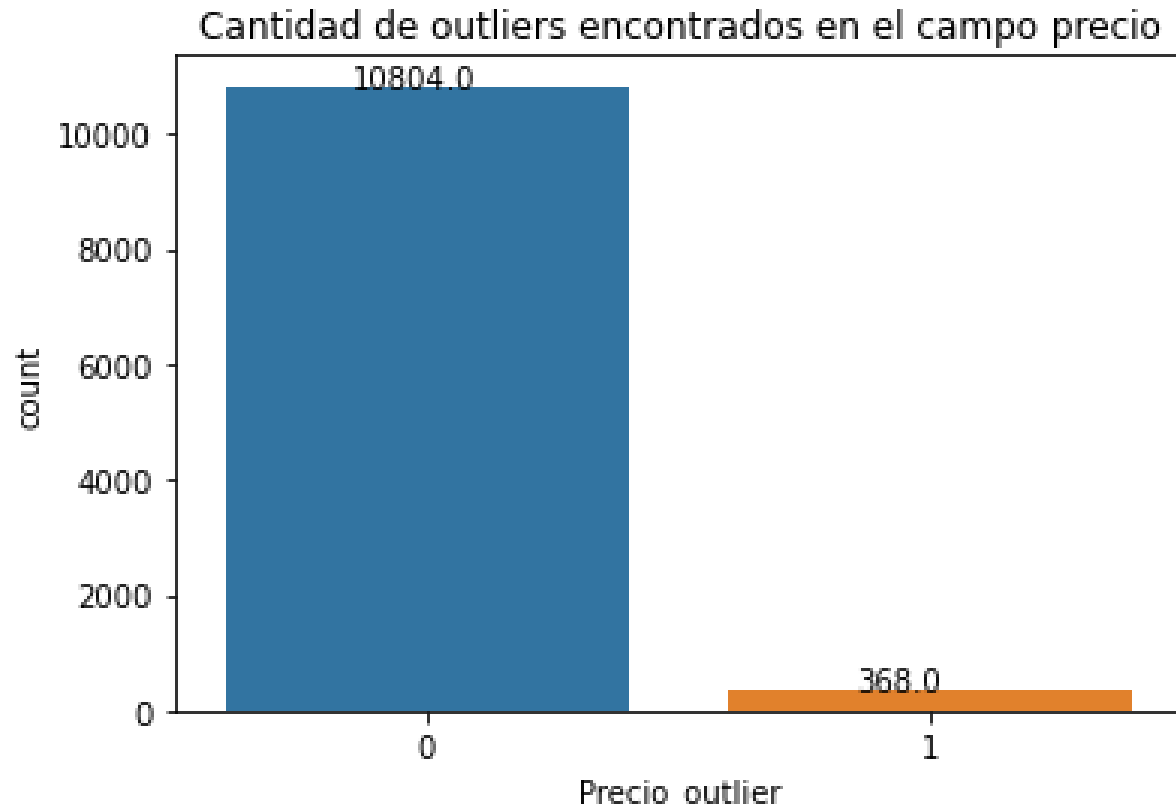
Outliers

- Se detectaron outliers en la tabla venta, en los campos “Precio” y “Cantidad”.



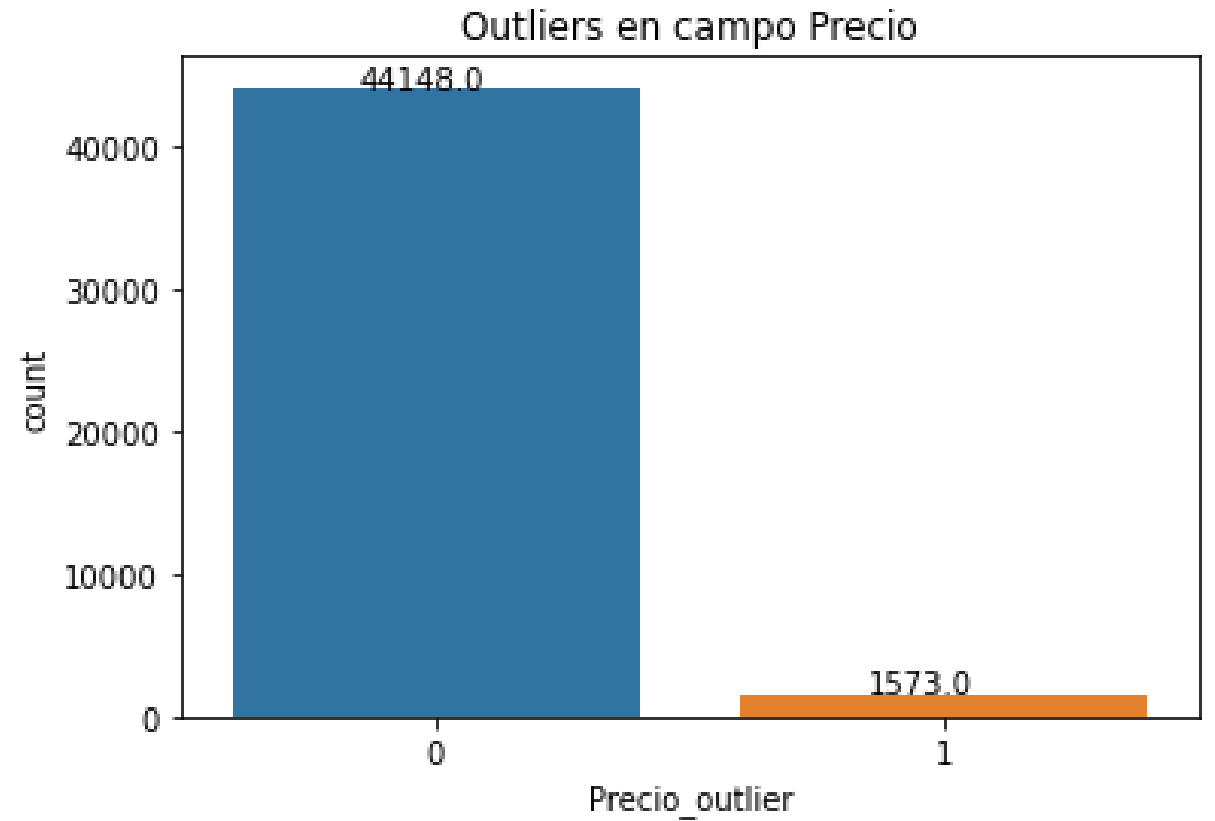
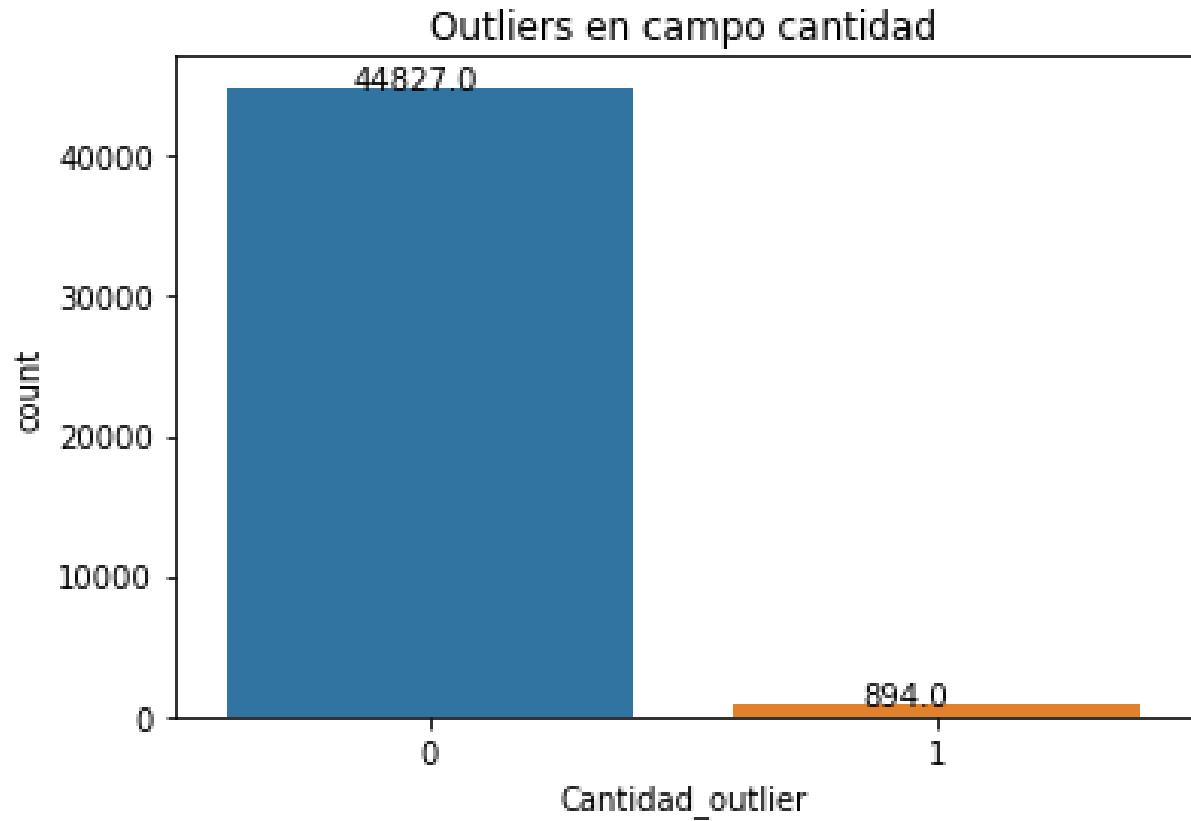
A la izquierda, en la tabla venta se muestran precios por encima de los \$30,000,000, además, a la derecha en el campo “Cantidad” tambien se muestran outliers con cantidades que ascienden a las 300 piezas.

Outliers en la tabla compra (~3.5%)



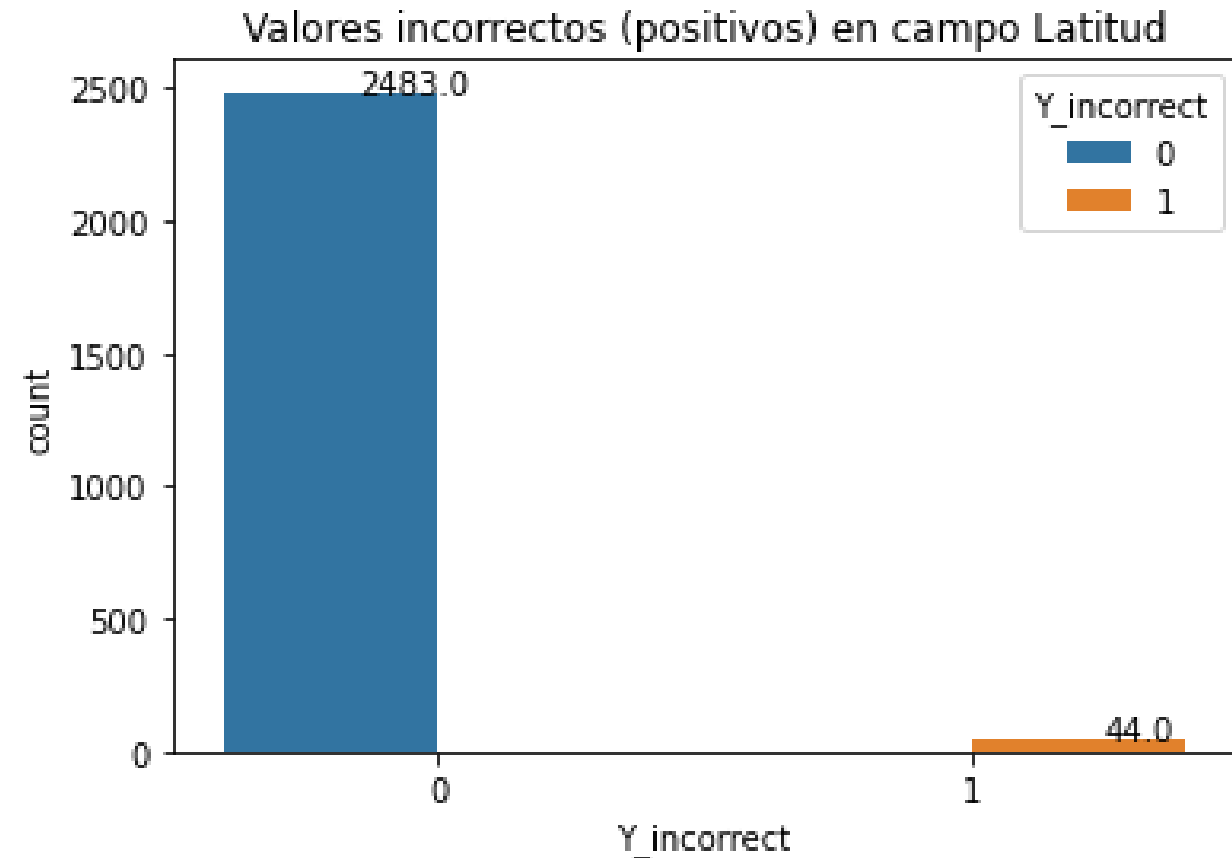
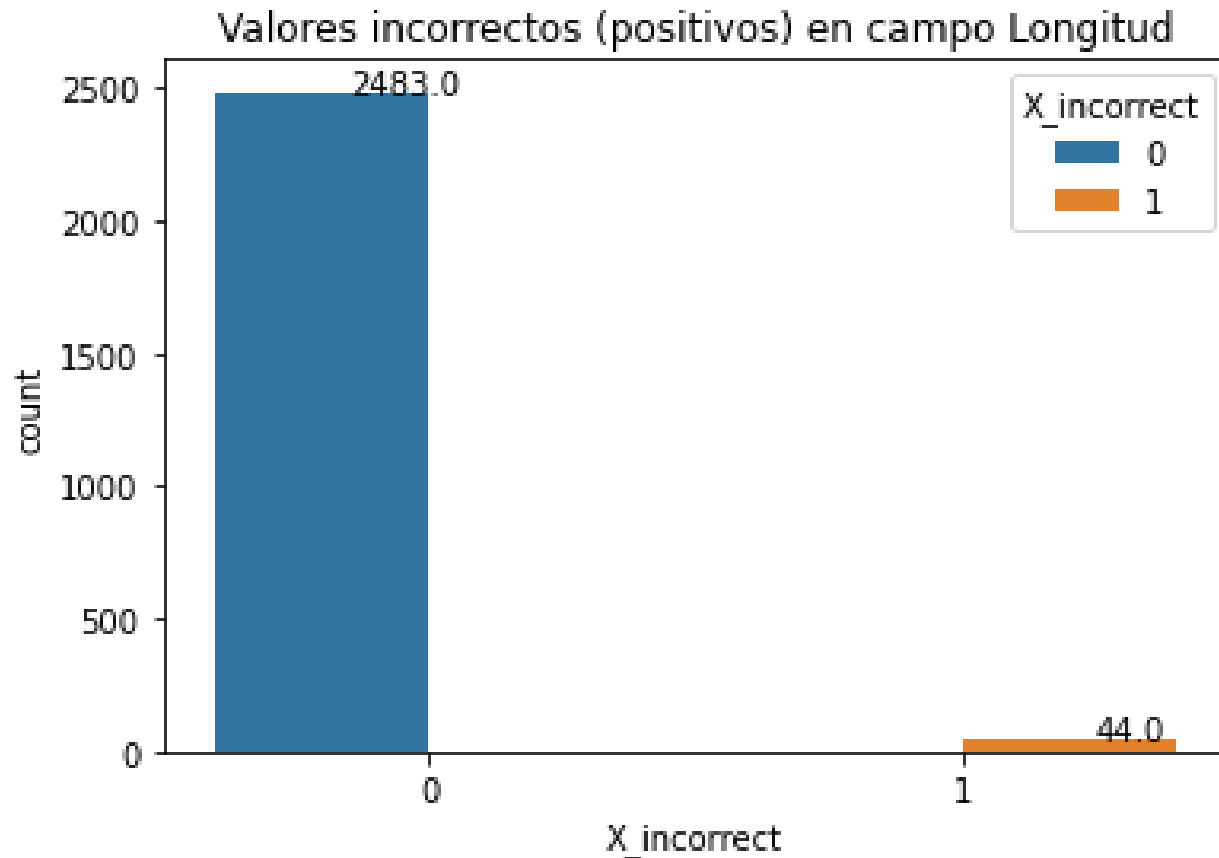
Para descartar estos outliers se utilizó $3 \times \text{IQR}$ (Método de los valores atípicos mayores) debido a la variación del precio a lo largo del tiempo.

Outliers (~ 3.5%) – Tabla Venta



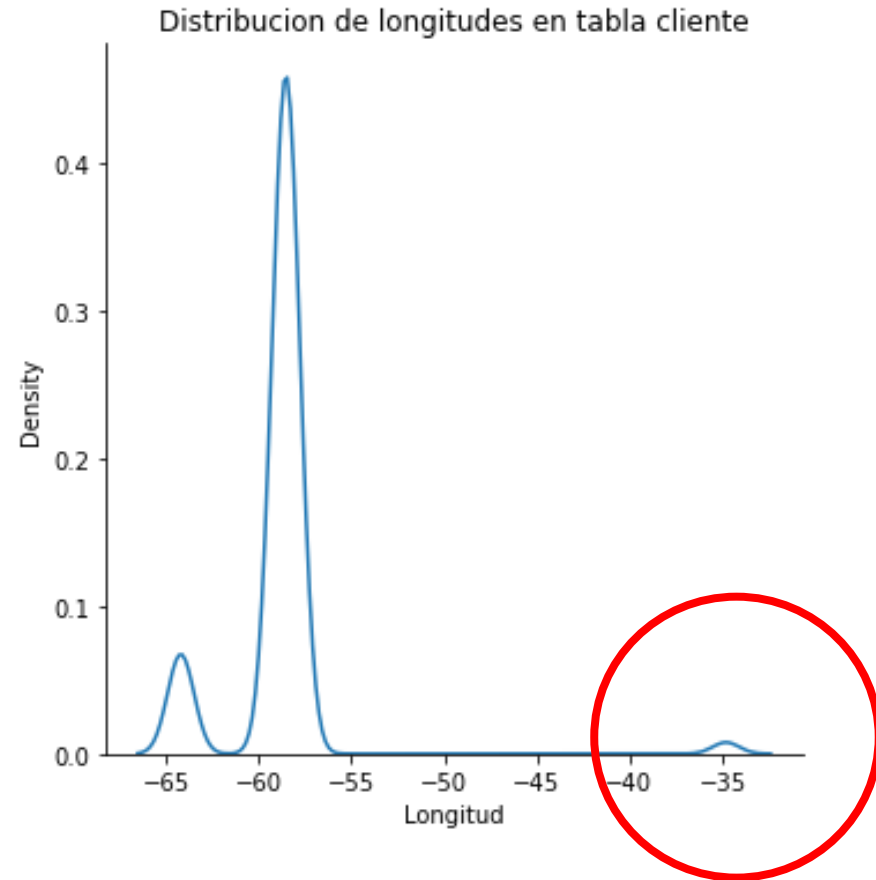
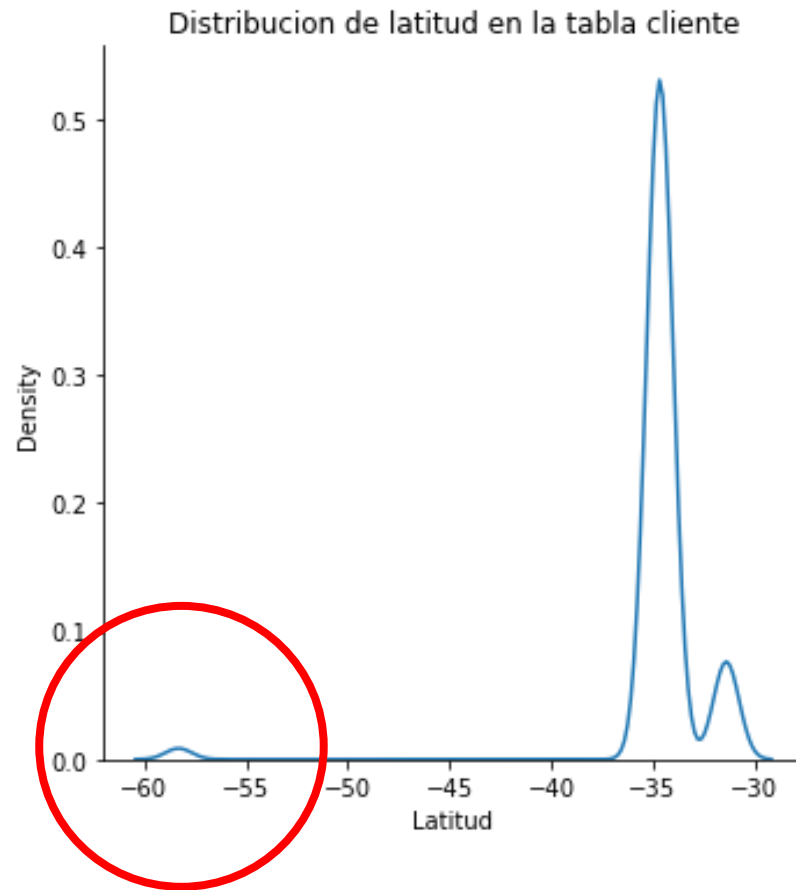
Valores incorrectos (~2%)

- Se encontraron valores positivos en las coordenadas de la ubicación de los clientes, es decir, ubicados fuera del país.



Valores incorrectos – Tabla clientes

Además, se encontraron valores de latitud y longitud capturados en orden contrario:



- Los valores mínimos y máximos en Argentina para la latitud son: -55.055, -21.781 y para la longitud: -73.566, -53.637, respectivamente.

Normalización

- Se encontraron datos de tipo string en la tabla Cliente, Sucursal y proveedores que no estaban normalizados, es decir, los nombres no eran uniformes a lo largo de las diferentes columnas.
- Es importante capturar los datos correctamente y en la medida de lo posible, evitar abreviaturas para que al momento de normalizar, el algoritmo reconozca correctamente cada string.

```
... {'Ciudad de Buenos Aires': 'Buenos Aires',  
    'Buenos Aires': 'Buenos Aires',  
    'C deBuenos Aires': 'Buenos Aires',  
    'Bs As': 'Buenos Aires',  
    'Bs.As. ': 'Buenos Aires',  
    'B. Aires': 'Buenos Aires',  
    'B.Aires': 'Buenos Aires',  
    'Provincia de Buenos Aires': 'Buenos Aires',  
    'Santa Fe': 'Santa Fe',  
    'Córdoba': 'Córdoba',  
    'Cordoba': 'Córdoba',  
    'Tucumán': 'Tucumán',  
    'Mendoza': 'Mendoza',  
    'Río Negro': 'Río Negro'}
```