

Perfilamiento de datos

Recipe (Caso de Orders parametrized):

- Tiene dos columnas, con un total de 73.595 filas.
- Todas las columnas (Date, Year, Month y Day) tienen los datos correctos, no hay ni datos en formato erróneos ni nulos.
- Los duplicados han sido eliminados.
- No se tienen outliers por el tipo de dato, pues son fechas de los eventos.
- El año más común es el 2015, el día más común es el 03 y el mes más común es mayo.

Recipe (Caso StockItems):

- Tiene 5 columnas, con un total de 227 filas.
- La columna ColorID tiene un total de 99 valores faltantes.
- La columna SizeLast tiene un total de 64 valores faltantes.
- La columna PriceLast tiene 2 valores en formato erróneo.
- Los ID son efectivamente únicos.
- Hay dos outliers (longitud del nombre del objeto en stock fuera de lo normal).
- La columna PriceLast tiene un total de 23 outliers.
- En total, hay 227 categorías de ítems en stock.
- Hay 33 colores diferentes (3-36) para los diferentes objetos en stock.

Recipe (Filter OrderLines Columns):

- En total hay nueve columnas, con 8430 filas.
- La totalidad de las columnas presentan valores correctos.
- Hay 5 columnas que funcionan como llave foránea para acceder a OrderID, OrderLineID, StockItemID, DateID y PackageTypeID.
- La columna para la cantidad presenta 470 outliers, un máximo de 100 unidades y un mínimo de una sola unidad.
- El precio por unidad tiene 2405 outliers, un precio máximo de 240 y un precio mínimo de 18.50.
- El TaxRate maneja un valor único de 15 para todas las filas.
- El precio total (cantidad*precio por unidad), maneja 889 outliers, un total máximo de 2400 y un total mínimo de 25.