**Preprocesamiento de los datos.**

* **Descripción del modelo relacional.**



**Ventas\_Prod (66.576 x 6)**. Incluye las transacciones, abiertas por renglón (un renglón = un producto), fecha y cantidades. MinFecha: 05/03/2014. MaxFecha: 11/05/2015. Resumen de cantidad de ventas por mes/año para detectar sesgos en la aplicación de reglas con datos de entrenamiento 2014 en 2015. Calcular largo promedio de las transacciones. DISTINCT Ventas\_ID: 6660. MaxVentas\_ID: 11.896. Por qué se produce el discontinuo? Cantidades con valores negativos y decimales. Cuando un renglón de una transacción tiene cantidades negativas, todos los renglones son negativos (posiblemente sean compras canceladas). Cuando UM2 es distinto de 0, UM1 representa las cantidades en miles.

**Ventas (6660 x 4)**. Asocia la venta a un cliente. Incluye campo Venta\_Fecha, que tiene los mismos valores que el campo Fecha de Ventas\_Prod. DISCTINCT CLI\_ID: 672.

**Clientes (677 x 5)**. Asocia ID del cliente al nombre. Aporta datos geográficos.

**Productos (4283 x 8)**. Asocia ID del producto a distintos atributos descriptivos, entre ellos categoría y subcategoría. DescAdic comienza con DescGen.

**Precio\_Sugerido (4283 x 2)**. Asocia ID del producto a precio. 144 productos con Precio = 0.

**Categoria (11 x 2)**. Asocia ID categoría a descripción.

**Sub\_Categoria (136 x 2)**. Ventas. Asocia ID subcategoría a descripción.