**Metodología CRIPS-DM**

**Fase I. Business Understanding.**

**Pagar es Fácil**

Es un eje de negocios digitales que permite a pequeños y medianos empresarios exhibir y vender productos o servicios, así como transaccionar con tarjetas de crédito y billetera virtual, transferir saldos a cuentas bancarias y pagar servicios básicos.

Pagar es Fácil ofrece múltiples servicios, pero la que se utilizó en esta investigación fue el servicio de MarketPlace. Pagar es Fácil aloja muchos comercios los cuales día a día suben sus productos a la plataforma e igualmente se transacciona con las compras realizadas por las personas.

Regularmente la plataforma recibe ingresos económicos por comisiones de cada venta, siempre y cuando el vendedor entregue correctamente el producto.

En base a lo mencionado anteriormente, Pagar es Fácil requiere conocer una predicción de ganancias de las ventas que se podrían generar en futuros años, conocer cuales son sus compradores más recurrentes, las categorías y productos más solicitados, y sus vendedores más confiables.

**Fase II. Data Understanding.**

EL dataset fue obtenido de las diferentes bases de datos proporcionadas por la plataforma, las cuales son las ventas conformadas por los clientes, vendedores, productos y categorías, a continuación, se detalla cada una ellas:

Cliente: Nombres y apellidos de los clientes.

Nombre\_producto: el nombre del producto.

Name\_comercio: el nombre del vendedor o mejor conocido como el comercio.

Precio\_producto: precio del producto proporcionado por el vendedor.

Precio\_producto\_venta: precio del producto sumado con la comisión de Pagar es Fácil.

Ganancias: es la resta de precio\_producto\_venta con precio\_producto obteniendo así las ganancias por producto por venta.

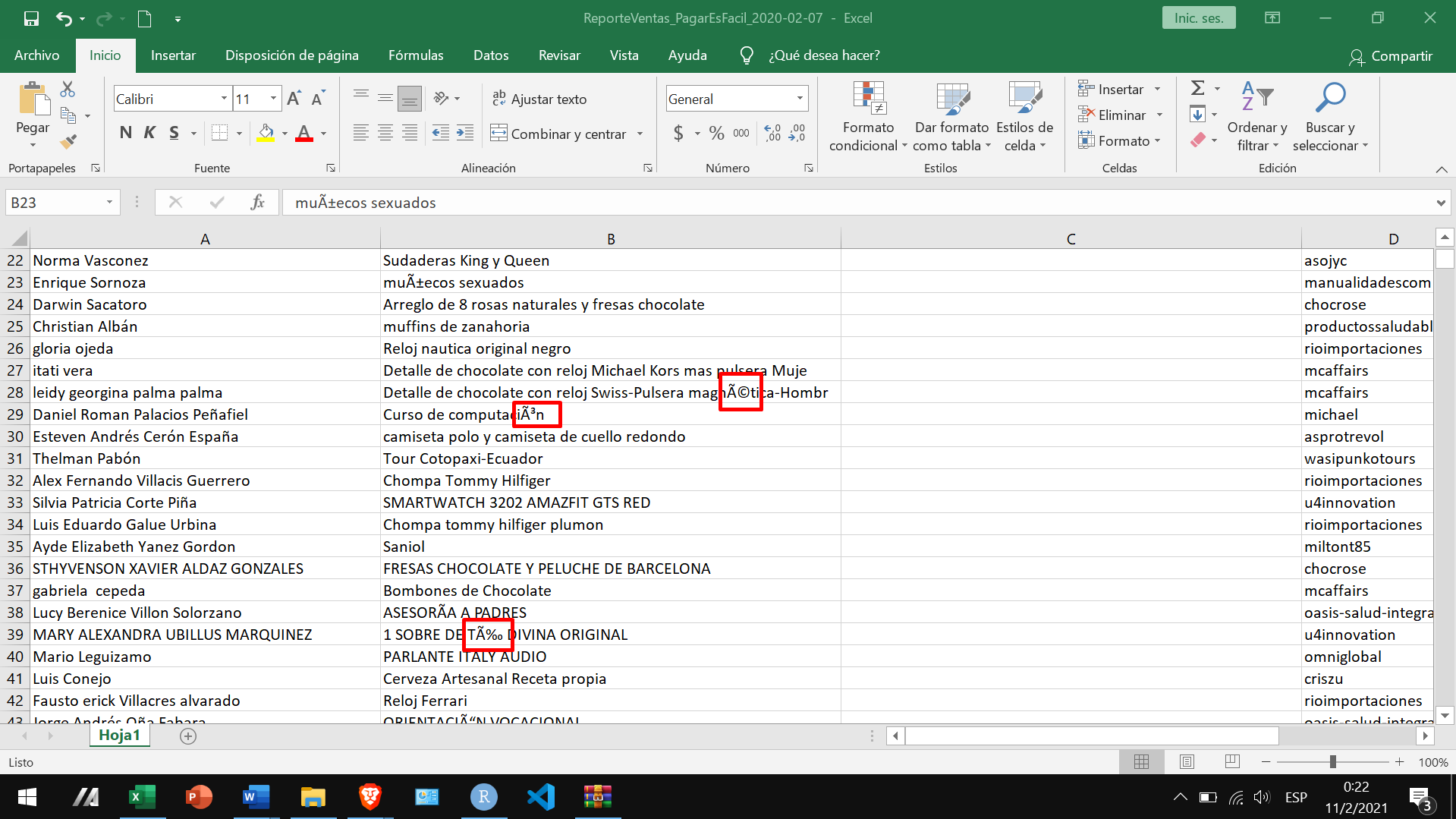
Category: las categorías de los productos.

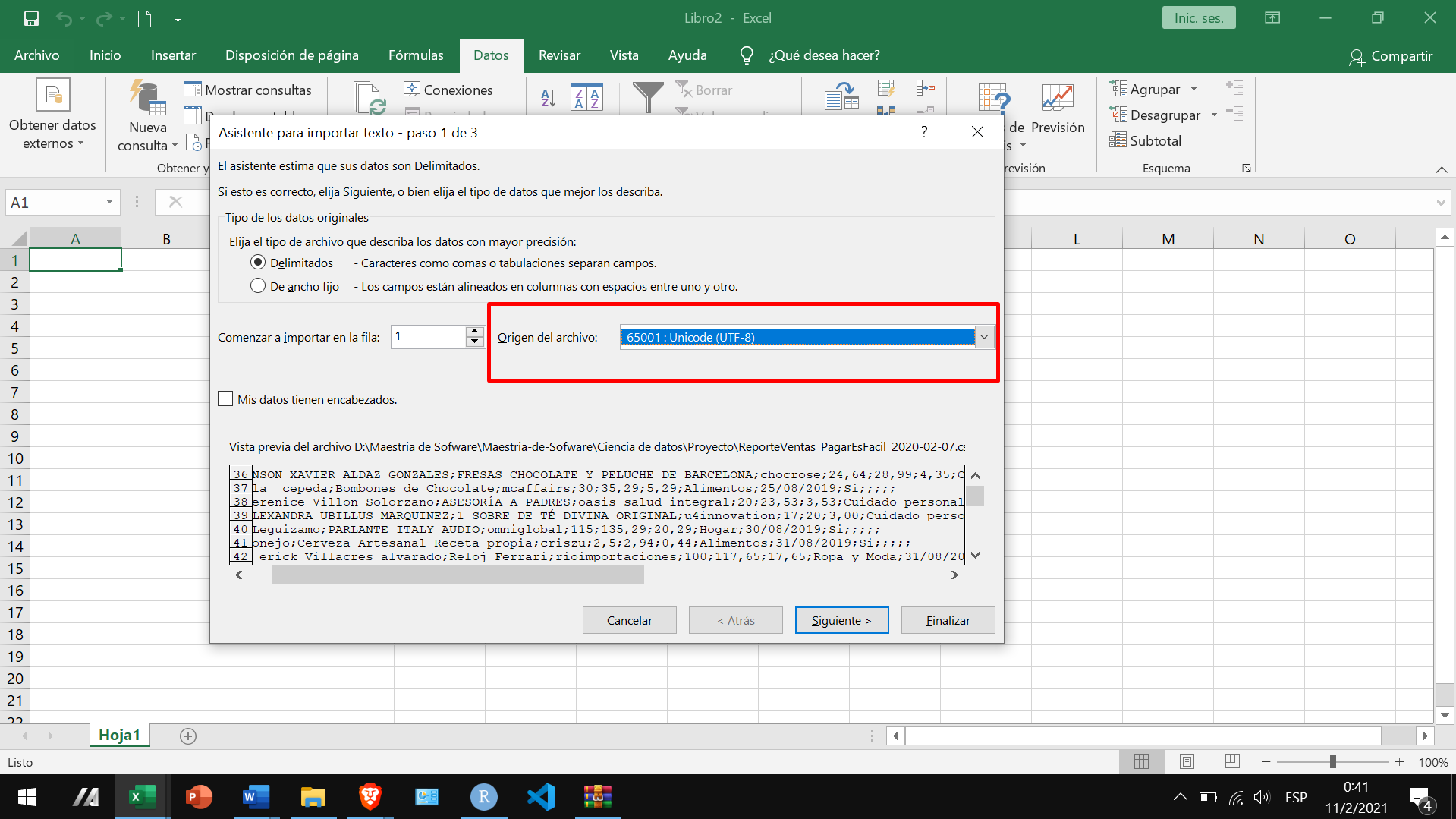
Fecha: fecha de venta.

Venta\_concretada: variable utilizada para conocer si un vendedor entrego el producto al cliente.

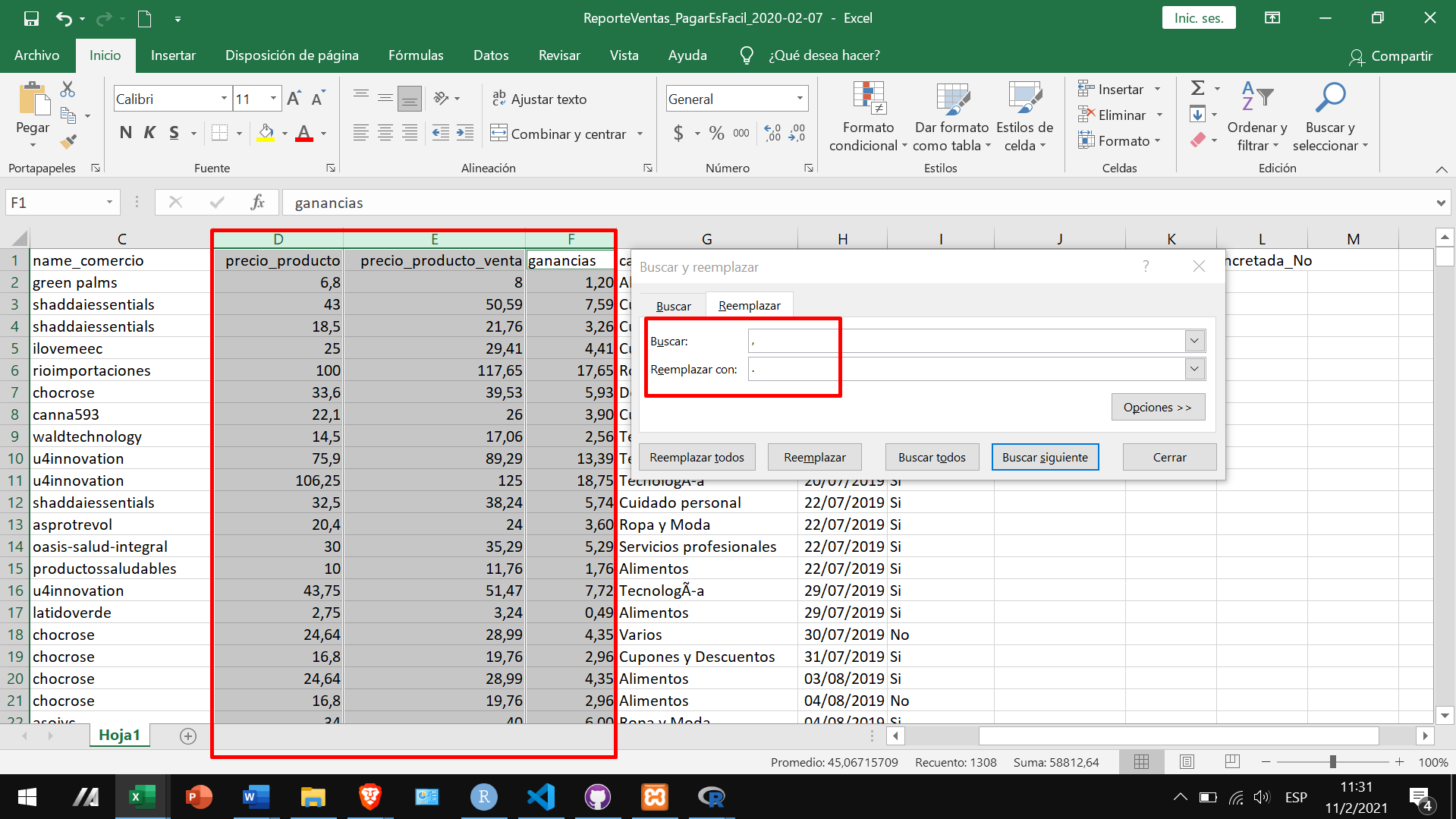
**Fase III. Data Preparation.**

Lo primero que se hizo con el dataset fue corregir aquellos campos con caracteres que están mal codificados.

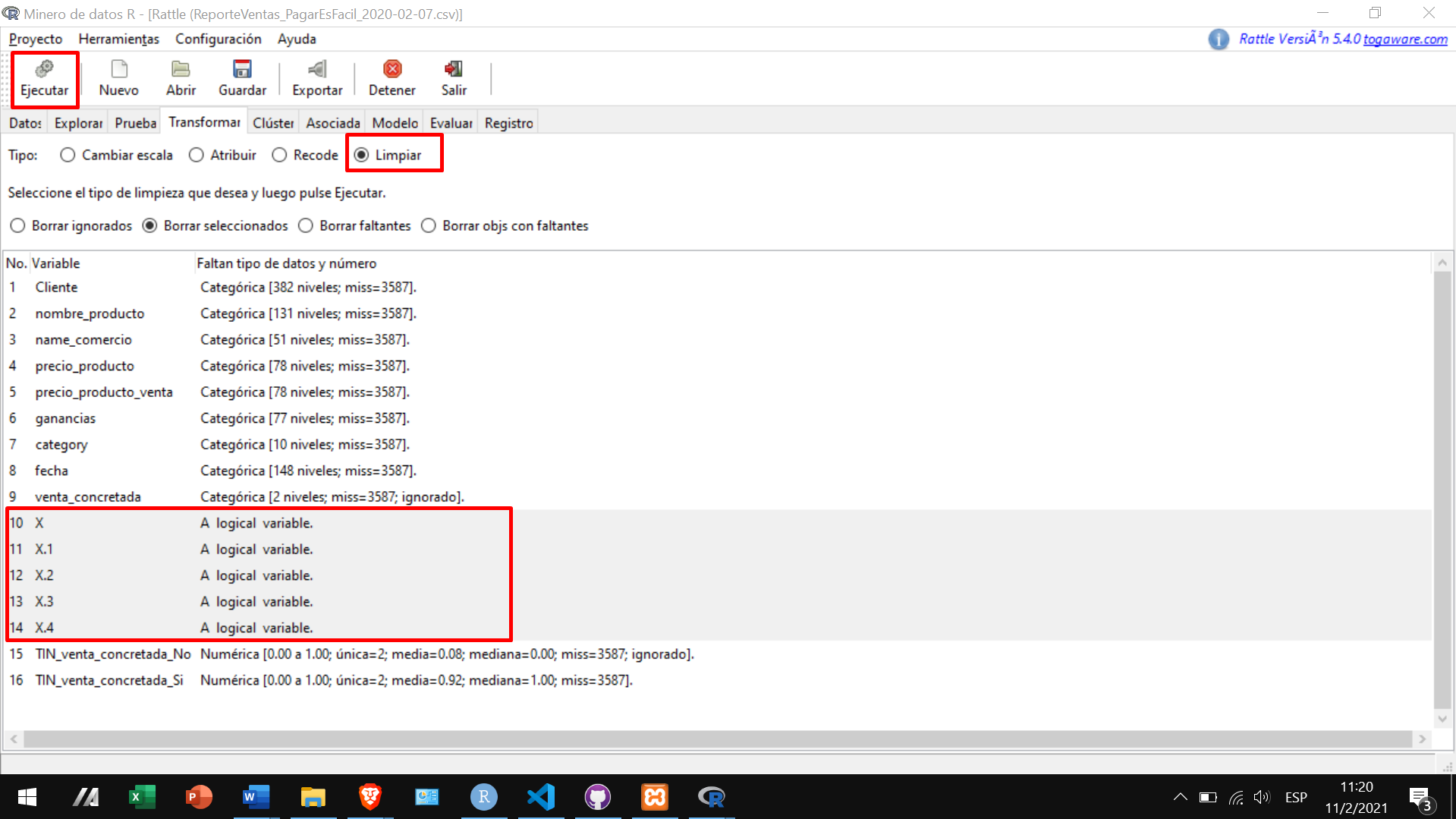




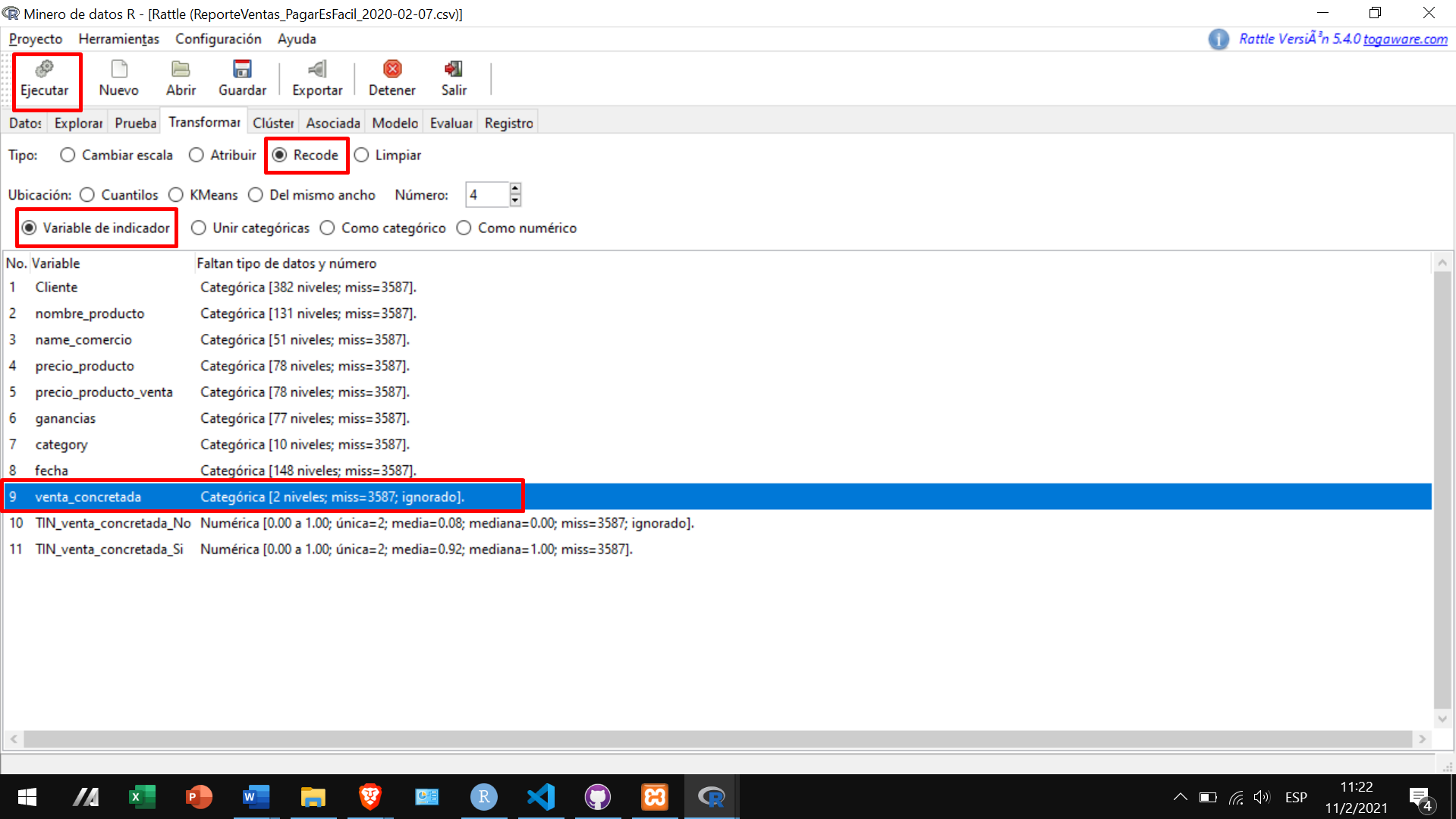
Cambiamos las comas que poseen los campos numéricos de precio\_producto, precio\_producto\_venta y ganancias por el punto, para que posteriormente la herramienta R nos los reconozca como variable numérica y no una cadena de texto.



Borrar aquellos campos adicionales que se crearon al transformar el archivo xlms a csv



Se utilizó la opción RECODE, para crear variables indicadoras a la columna venta\_concretada que poseen registros con valores de SI/NO y posteriormente utilizarlas para un posible análisis.



**Fase IV. Modeling.**

**Fase V. Evaluation.**

**Fase VI. Deployment.**