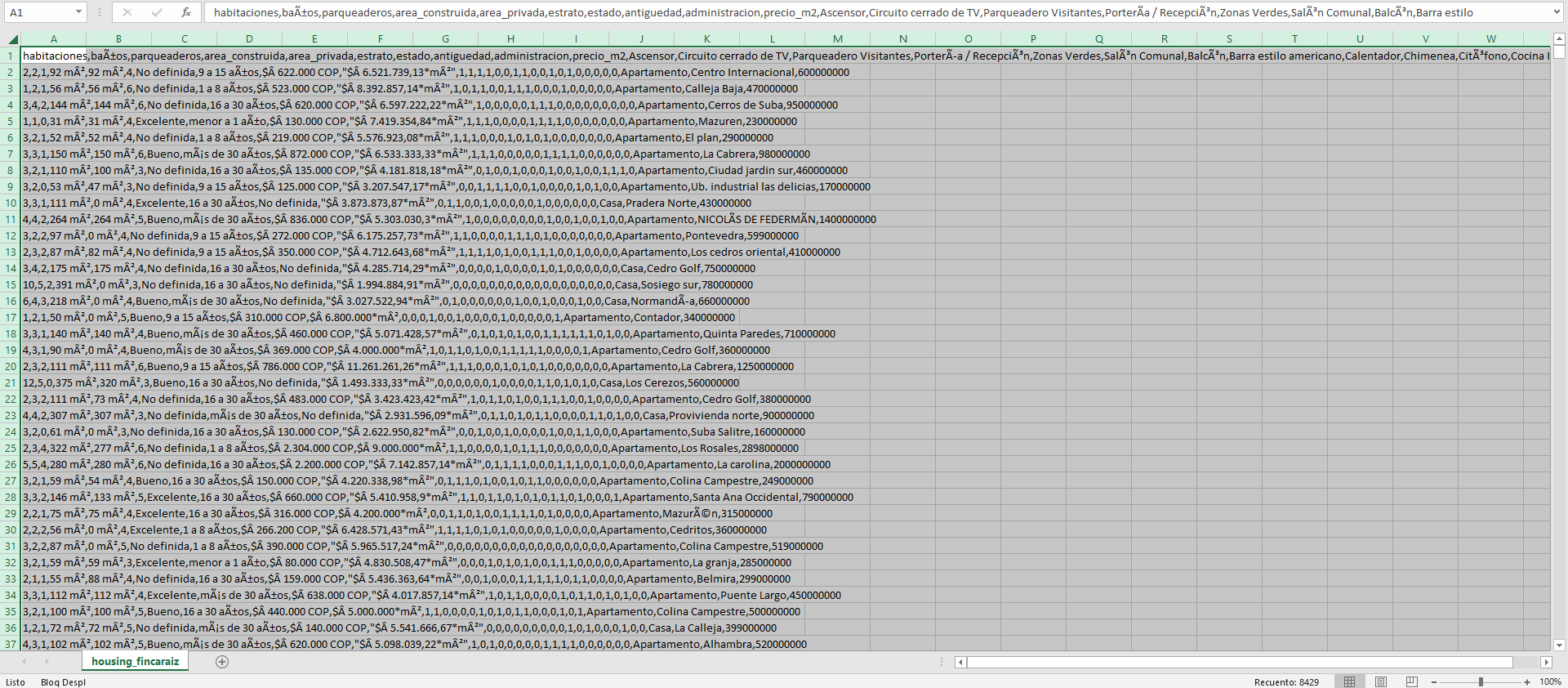
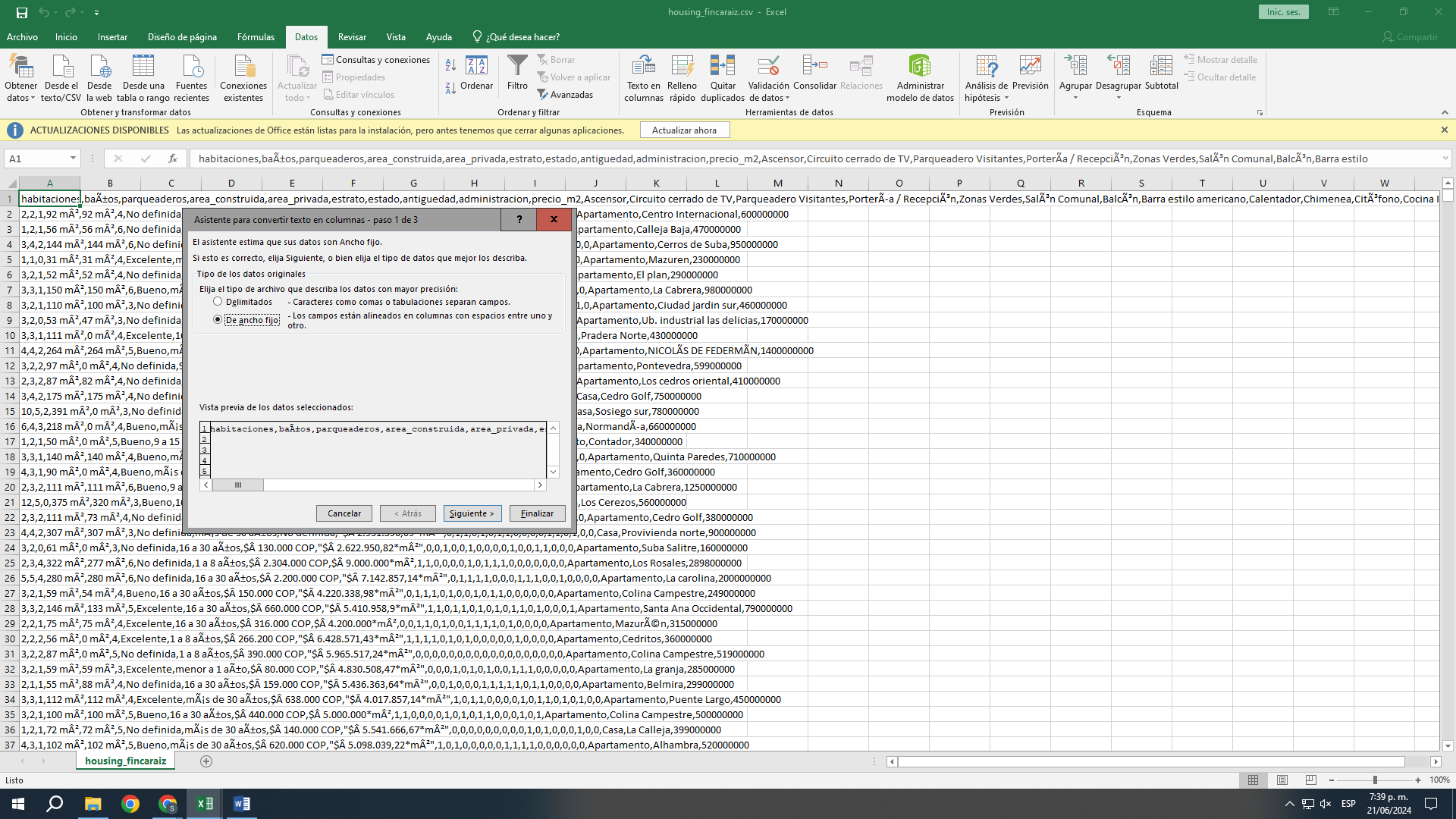
**Documentación laboratorio 3.**

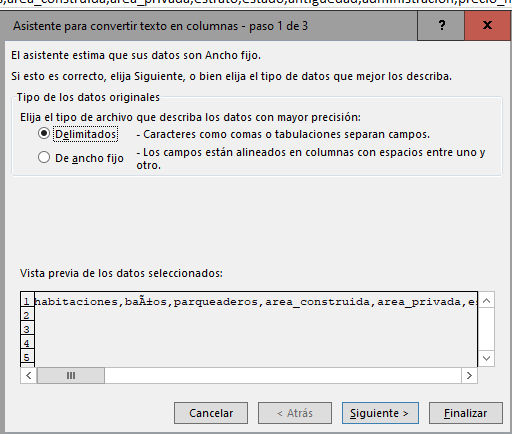
Primer paso:



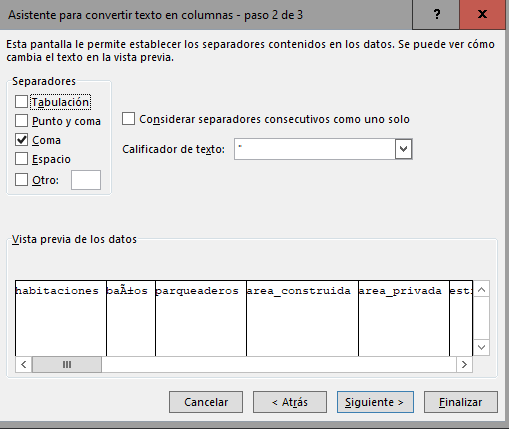
Se realizo la respectiva acomodación de los datos seleccionándolos todos.



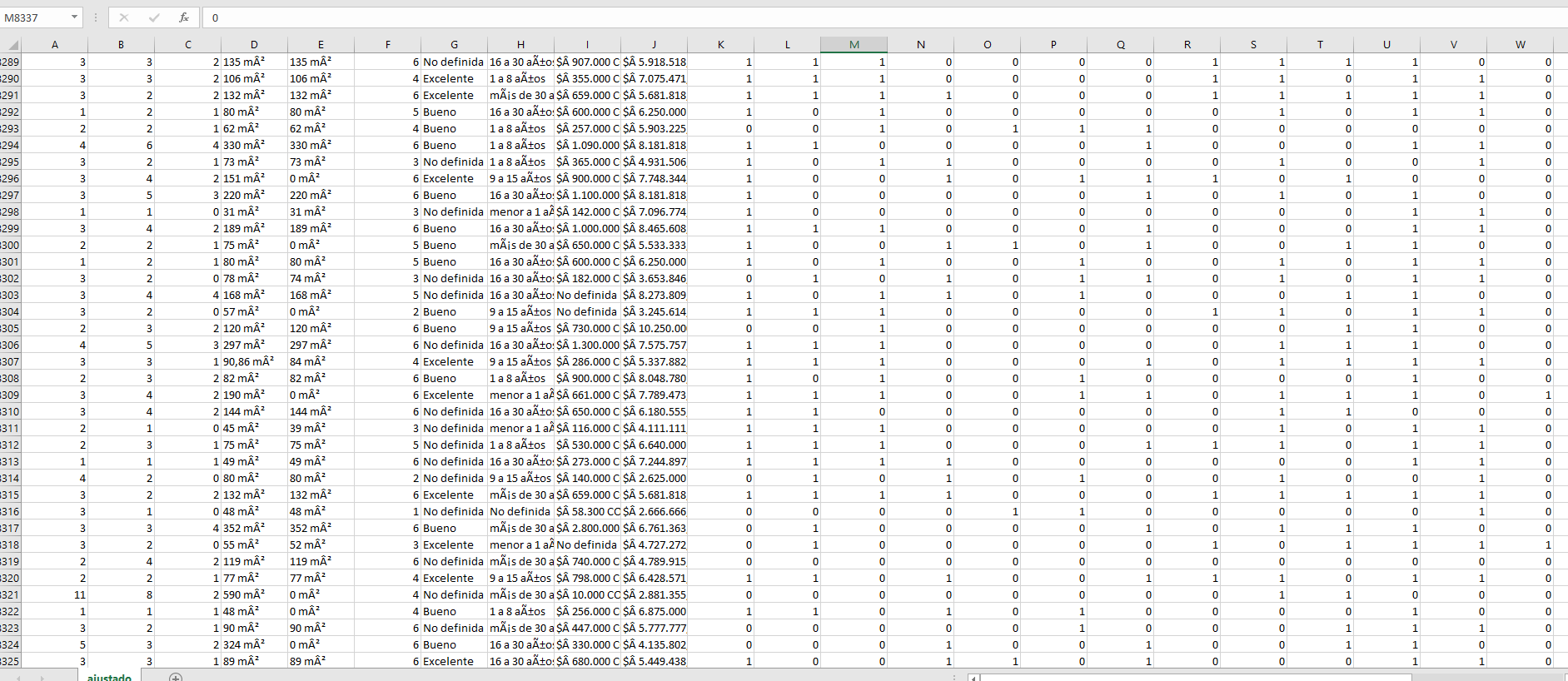
Ingresamos al apartado de datos y se selecciono texto en columnas.

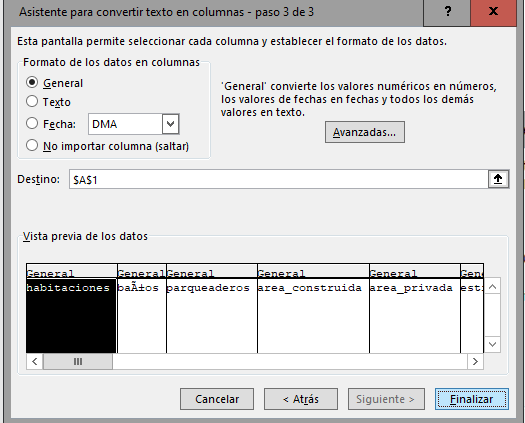


Se selecciono el campo de delimitados.

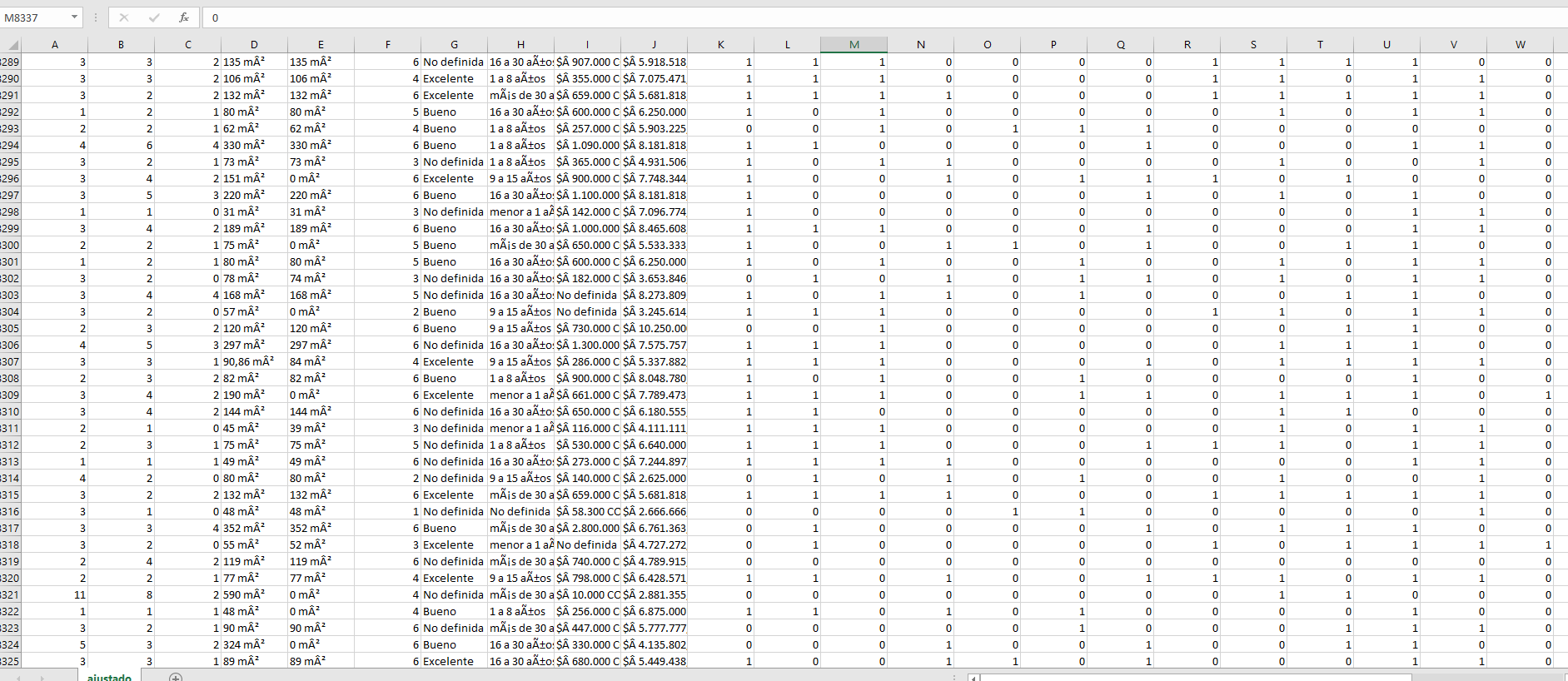


Se selección el campo “coma” para separar las columnas



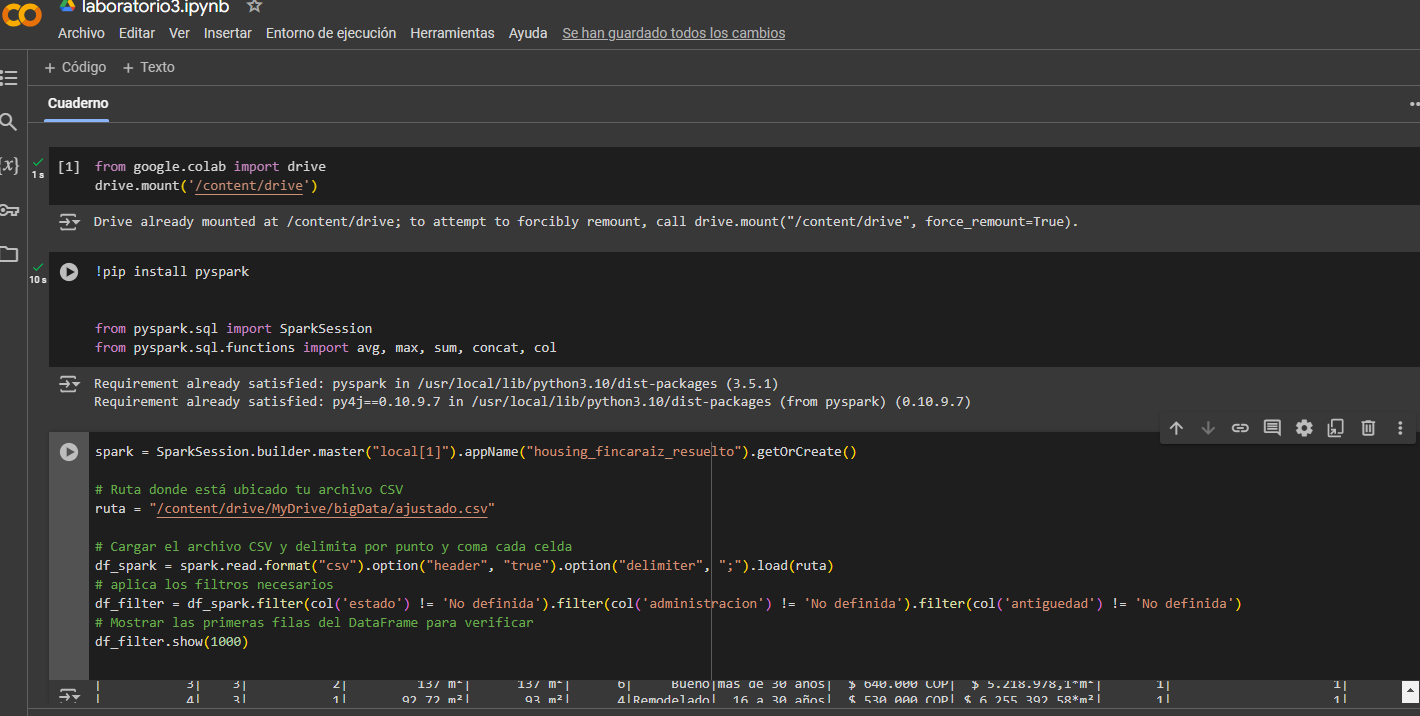


Se deja todos lo campos en general, y se presiona finalizar.



Se ajusto el documento respectivamente con los pasos previos.

**Nota:** solo se ajusta, las filas y las columnas. La limpieza de los datos se hace en pyton con la librería pyspark



**¿Qué variables (columnas, atributos) de la(s) tabla(s) o base(s) de datos parecen más prometedores?**

RTA: los campos más relevantes son:

Habitaciones, Baños, Parqueaderos, Área construida, Área privada, Estrato, Estado, Antigüedad, Administración, Precio, Precio\_m2, Ubicación.

**¿Qué variables parecen irrelevantes y pueden ser excluidos?**

RTA: Los campos irrelevantes son:

Ascensor, Circuito cerrado de TV, Parqueadero Visitantes, Portería / Recepción, Zonas Verdes, Salón Comunal, Balcón, Barra estilo americano, Calentador, Chimenea, Citófono, Cocina Integral, Terraza, Vigilancia, Parques cercanos, Estudio, Patio, Depósito / Bodega, nombre, ubicación

**¿Hay suficientes datos para sacar conclusiones generalizables o hacer predicciones**

**precisas?**

RTA: si hay suficiente información

**¿Hay demasiadas variables para el método de modelado de su elección?**

RTA: si hay suficientes variables para aplicar el método

**¿Está fusionando varias fuentes de datos? Si es así, ¿hay áreas que podrían plantear un**

**problema al fusionar?**

RTA: no, se utiliza únicamente una fuente de datos que contiene toda la información necesaria.

**¿Ha considerado cómo se manejan los valores que faltan en cada uno de sus orígenes de**

**datos?**

RTA: se omiten principalmente