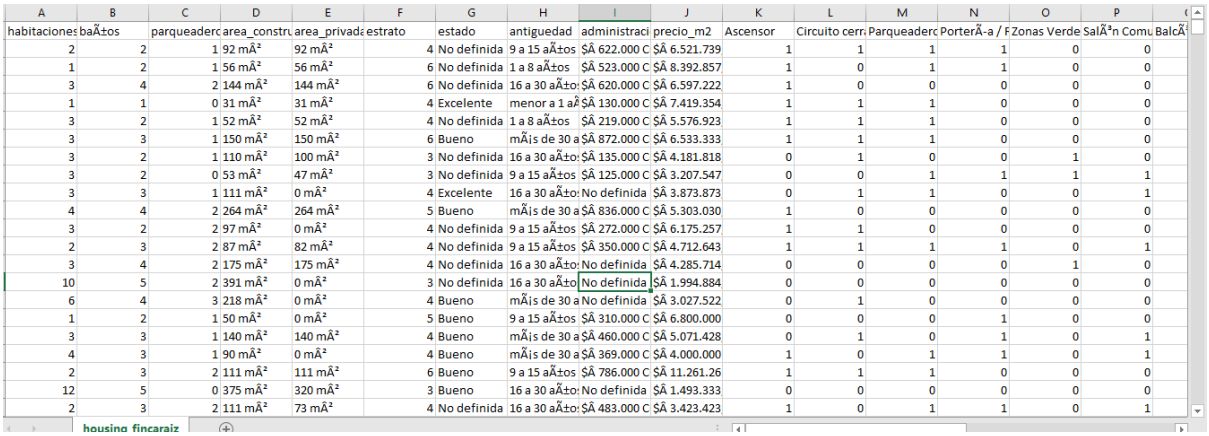
**Integrantes :**
Paula Rios - 91105
Oscar Velasquez - 91273

# Laboratorio 3

- **Recolectar datos iniciales**



| habitaciones | baÃ±os | parqueadero | area_construi | area_privada | estrato | estado | antiguedad | administraci | precio_m2 | Ascensor | Circuito cerr | Parqueadero | PorterÃ-a / R | Zonas Verde | SalÃ³n Comu | BalcÃ³ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2 | 2 | 1 | 92 mÂ² | 92 mÂ² | 4 | No definida | 9 a 15 aÃ±os | $Â 622.000 C | $Â 6.521.739 | 1 | 1 | 1 | 1 | 0 | 0 | |
| 1 | 2 | 1 | 56 mÂ² | 56 mÂ² | 6 | No definida | 1 a 8 aÃ±os | $Â 523.000 C | $Â 8.392.857 | 1 | 0 | 1 | 1 | 0 | 0 | |
| 3 | 4 | 2 | 144 mÂ² | 144 mÂ² | 6 | No definida | 16 a 30 aÃ±o | $Â 620.000 C | $Â 6.597.222 | 1 | 0 | 0 | 0 | 0 | 0 | |
| 1 | 1 | 0 | 31 mÂ² | 31 mÂ² | 4 | Excelente | menor a 1 aÃ | $Â 130.000 C | $Â 7.419.354 | 1 | 1 | 1 | 0 | 0 | 0 | |
| 3 | 2 | 1 | 52 mÂ² | 52 mÂ² | 4 | No definida | 1 a 8 aÃ±os | $Â 219.000 C | $Â 5.576.923 | 1 | 1 | 1 | 0 | 0 | 0 | |
| 3 | 3 | 1 | 150 mÂ² | 150 mÂ² | 6 | Bueno | mÃ¡s de 30 a | $Â 872.000 C | $Â 6.533.333 | 1 | 1 | 1 | 0 | 0 | 0 | |
| 3 | 2 | 1 | 110 mÂ² | 100 mÂ² | 3 | No definida | 16 a 30 aÃ±o | $Â 135.000 C | $Â 4.181.818 | 0 | 1 | 0 | 0 | 1 | 0 | |
| 3 | 2 | 0 | 53 mÂ² | 47 mÂ² | 3 | No definida | 9 a 15 aÃ±os | $Â 125.000 C | $Â 3.207.547 | 0 | 0 | 1 | 1 | 1 | 1 | |
| 3 | 3 | 1 | 111 mÂ² | 0 mÂ² | 4 | Excelente | 16 a 30 aÃ±o | No definida | $Â 3.873.873 | 0 | 1 | 1 | 0 | 0 | 1 | |
| 4 | 4 | 2 | 264 mÂ² | 264 mÂ² | 5 | Bueno | mÃ¡s de 30 a | $Â 836.000 C | $Â 5.303.030 | 1 | 0 | 0 | 0 | 0 | 0 | |
| 3 | 2 | 2 | 97 mÂ² | 0 mÂ² | 4 | No definida | 9 a 15 aÃ±os | $Â 272.000 C | $Â 6.175.257 | 1 | 1 | 0 | 0 | 0 | 0 | |
| 2 | 3 | 2 | 87 mÂ² | 82 mÂ² | 4 | No definida | 9 a 15 aÃ±os | $Â 350.000 C | $Â 4.712.643 | 1 | 1 | 1 | 1 | 0 | 1 | |
| 3 | 4 | 2 | 175 mÂ² | 175 mÂ² | 4 | No definida | 16 a 30 aÃ±o | No definida | $Â 4.285.714 | 0 | 0 | 0 | 0 | 1 | 0 | |
| 10 | 5 | 2 | 391 mÂ² | 0 mÂ² | 3 | No definida | 16 a 30 aÃ±o | No definida | $Â 1.994.884 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 6 | 4 | 3 | 218 mÂ² | 0 mÂ² | 4 | Bueno | mÃ¡s de 30 a | No definida | $Â 3.027.522 | 0 | 1 | 0 | 0 | 0 | 0 | |
| 1 | 2 | 1 | 50 mÂ² | 0 mÂ² | 5 | Bueno | 9 a 15 aÃ±os | $Â 310.000 C | $Â 6.800.000 | 0 | 0 | 0 | 1 | 0 | 0 | |
| 3 | 3 | 1 | 140 mÂ² | 140 mÂ² | 4 | Bueno | mÃ¡s de 30 a | $Â 460.000 C | $Â 5.071.428 | 0 | 1 | 0 | 1 | 0 | 1 | |
| 4 | 3 | 1 | 90 mÂ² | 0 mÂ² | 4 | Bueno | mÃ¡s de 30 a | $Â 369.000 C | $Â 4.000.000 | 1 | 0 | 1 | 1 | 0 | 1 | |
| 2 | 3 | 2 | 111 mÂ² | 111 mÂ² | 6 | Bueno | 9 a 15 aÃ±os | $Â 786.000 C | $Â 11.261.26 | 1 | 1 | 1 | 0 | 0 | 0 | |
| 12 | 5 | 0 | 375 mÂ² | 320 mÂ² | 3 | Bueno | 16 a 30 aÃ±o | No definida | $Â 1.493.333 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 2 | 3 | 2 | 111 mÂ² | 73 mÂ² | 4 | No definida | 16 a 30 aÃ±o | $Â 483.000 C | $Â 3.423.423 | 1 | 0 | 1 | 1 | 0 | 1 | |

  - Las variables más prometedoras son: precio, habitacion, baños, nombre, area_construida y area_privada
  - Fuente única de datos, no se está fusionando con otra fuente
  - No hay datos faltantes, los datos están completos

- **Describir los datos**

- El archivo es en formato .csv
- Los datos se obtuvieron mediante web scraping
- El archivo tiene 31 columnas(atributos) y 8428 filas (registros)
- Contiene datos numéricos, alfanuméricos y de texto

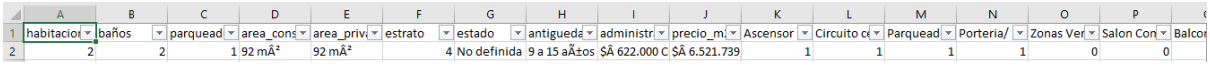
- **Explorar los datos**

  Las variables que son prometedoras para un análisis más profundo son: antigüedad, area_construida, area_privada, nombre y ubicación.

- **Verificar los datos**

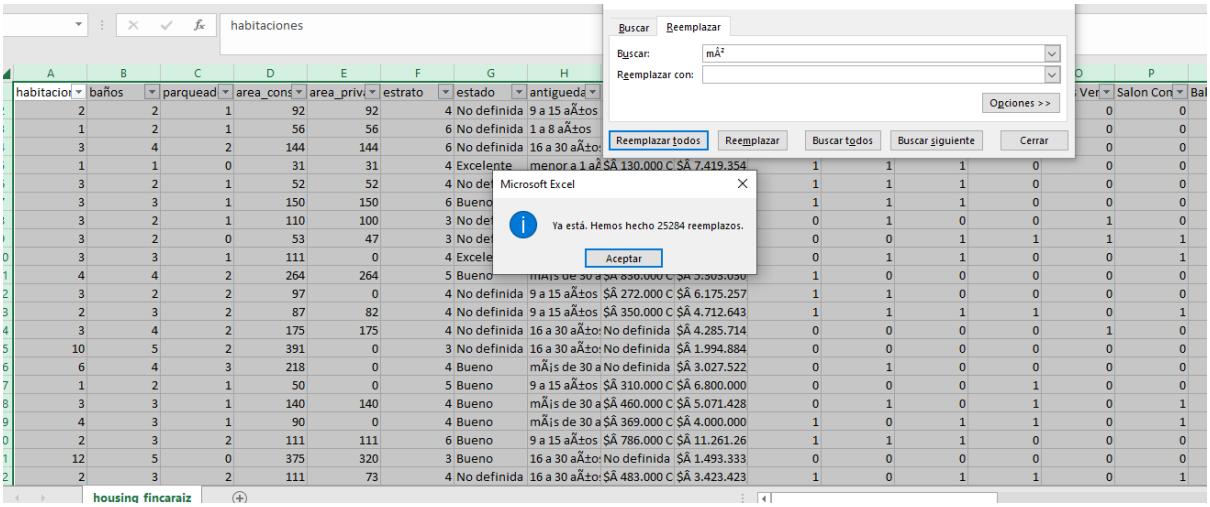
  Errores tipográficos en los datos:

- Corrección nombre de atributos

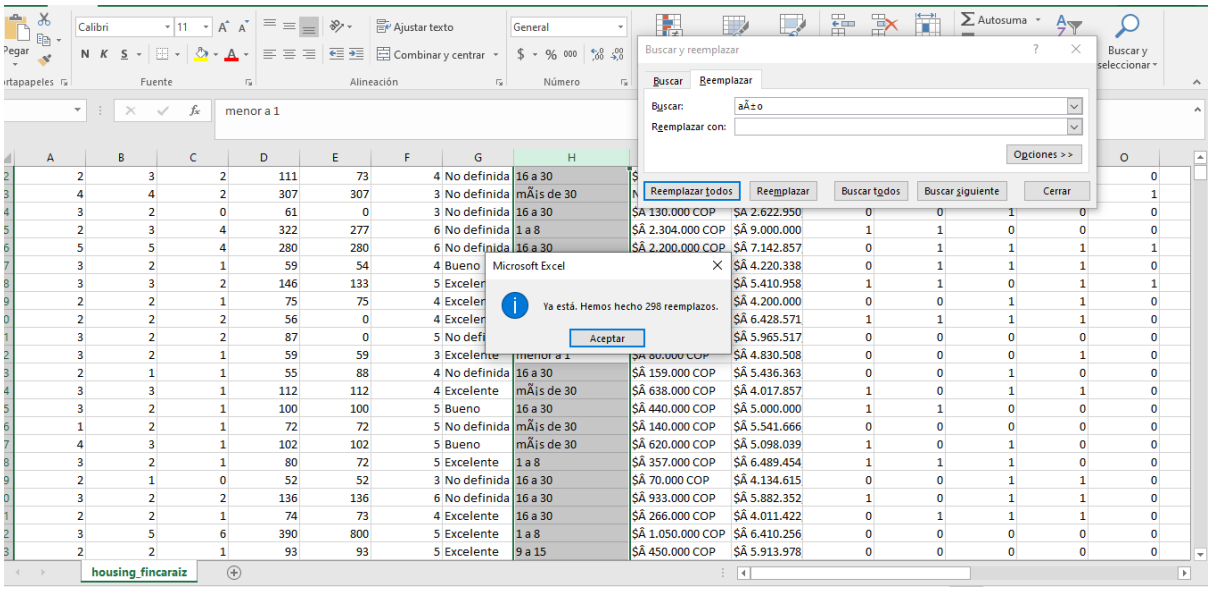


| habitacion | baños | parquead | area_cons | area_priv | estrato | estado | antigueda | administr | precio_m | Ascensor | Circuito c | Parquead | Porteria/ | Zonas Ver | Salon Con | Balcor |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2 | 2 | 1 | 92 mÂ² | 92 mÂ² | 4 | No definida | 9 a 15 aÃ±os | $Â 622.000 C | $Â 6.521.739 | 1 | 1 | 1 | 1 | 0 | 0 | |

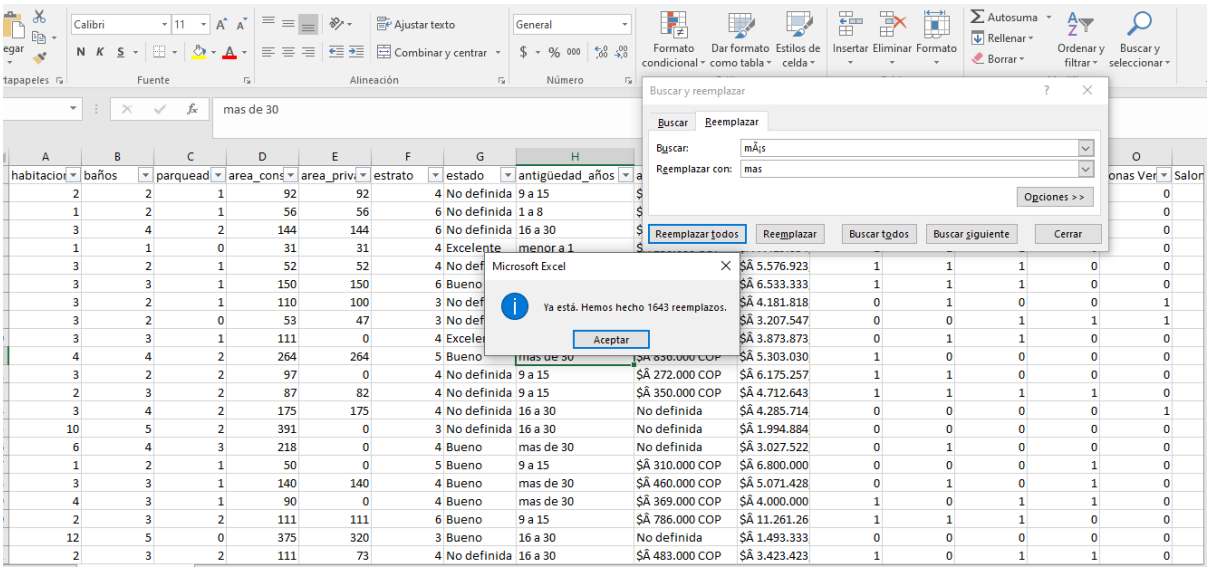
- Se cambian nombres de  variables area_construida y area_privada, por area_construida_m2 y area_privada_m2, se eliminan las unidades(mÂ² ) de los registros de las variables precio_m2, area_construida_m2 y area_privada_m2



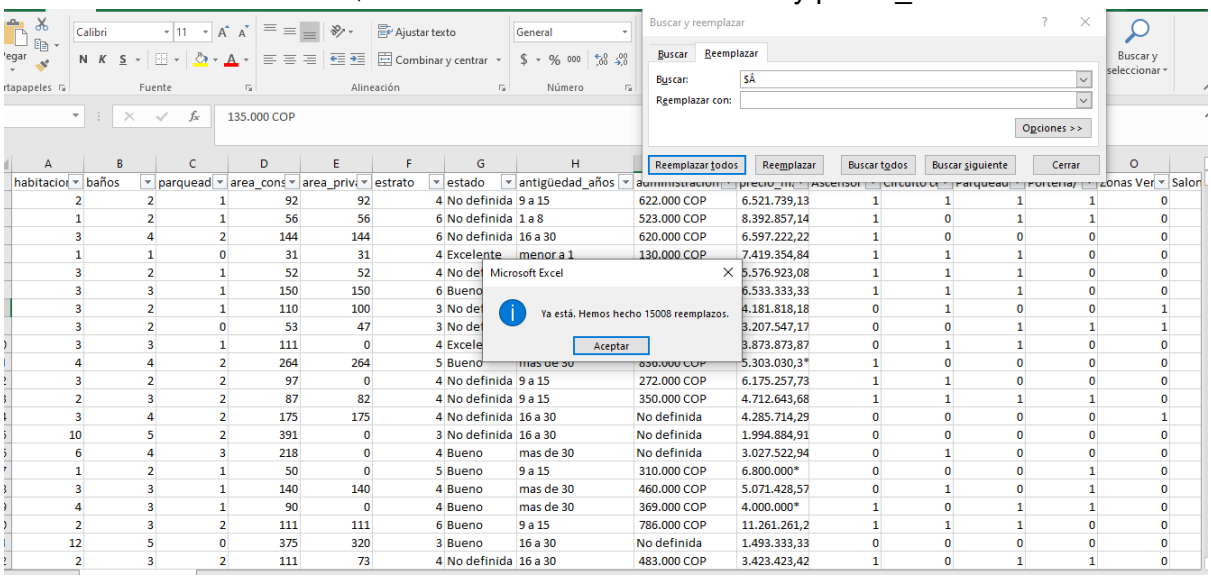

- Se cambia nombre de la variable antigüedad por antiguedad_años, se elimina la unidad de la antiguedad (aÃ±os, aÃ±o)




- Se cambian los registros de la variable antiguedad_años que contengan la palabra “mÃ¡s” por la palabra “mas”




- Se elimina “$Â” de la variable administracion y precio_m2




- Se elimina el “*” de la variable precio_m2, insertando una nueva columna en la que estaran unicamente valores numericos, y se elimina la columna que contiene el “ * “

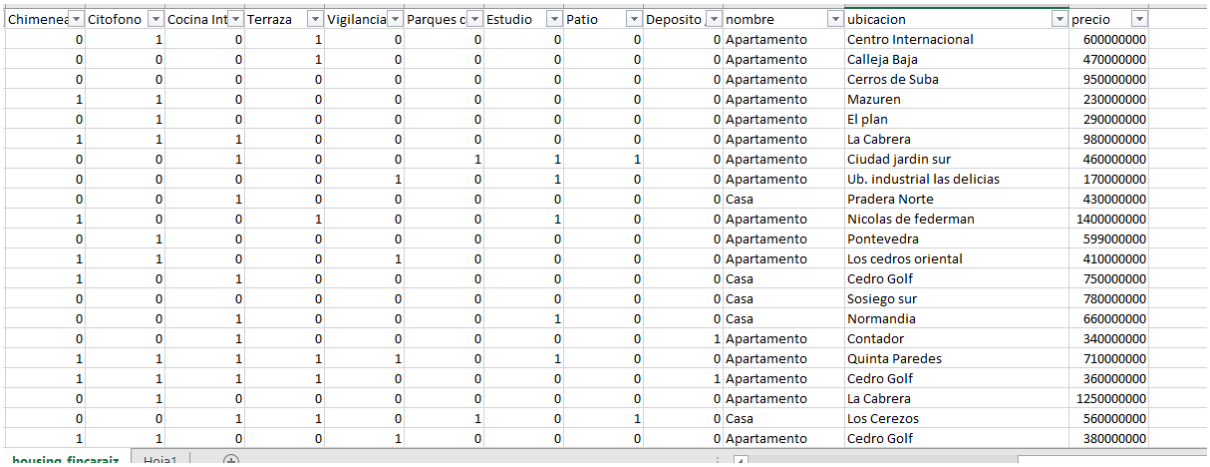
=SUSTITUIR(K2;"*";" ")

| | D | E | F | G | H | I | J | K |
|---|---|---|---|---|---|---|---|---|
| | area_cons | area_priv | estrato | estado | antigüedad_años | administracion | precio_m2 | precio_m2 |
| 1 | 92 | 92 | 4 | No definida | 9 a 15 | 622.000 COP | 6.521.739,13 | 6.521.739,13* |
| 1 | 56 | 56 | 6 | No definida | 1 a 8 | 523.000 COP | 8.392.857,14 | 8.392.857,14* |
| 2 | 144 | 144 | 6 | No definida | 16 a 30 | 620.000 COP | 6.597.222,22 | 6.597.222,22* |
| 0 | 31 | 31 | 4 | Excelente | menor a 1 | 130.000 COP | 7.419.354,84 | 7.419.354,84* |
| 1 | 52 | 52 | 4 | No definida | 1 a 8 | 219.000 COP | 5.576.923,08 | 5.576.923,08* |
| 1 | 150 | 150 | 6 | Bueno | mas de 30 | 872.000 COP | 6.533.333,33 | 6.533.333,33* |
| 1 | 110 | 100 | 3 | No definida | 16 a 30 | 135.000 COP | 4.181.818,18 | 4.181.818,18* |
| 0 | 53 | 47 | 3 | No definida | 9 a 15 | 125.000 COP | 3.207.547,17 | 3.207.547,17* |
| 1 | 111 | 0 | 4 | Excelente | 16 a 30 | No definida | 3.873.873,87 | 3.873.873,87* |
| 2 | 264 | 264 | 5 | Bueno | mas de 30 | 836.000 COP | 5.303.030,3 | 5.303.030,3* |
| 2 | 97 | 0 | 4 | No definida | 9 a 15 | 272.000 COP | 6.175.257,73 | 6.175.257,73* |
| 2 | 87 | 82 | 4 | No definida | 9 a 15 | 350.000 COP | 4.712.643,68 | 4.712.643,68* |
| 2 | 175 | 175 | 4 | No definida | 16 a 30 | No definida | 4.285.714,29 | 4.285.714,29* |
| 2 | 391 | 0 | 3 | No definida | 16 a 30 | No definida | 1.994.884,91 | 1.994.884,91* |
| 3 | 218 | 0 | 4 | Bueno | mas de 30 | No definida | 3.027.522,94 | 3.027.522,94* |
| 1 | 50 | 0 | 5 | Bueno | 9 a 15 | 310.000 COP | 6.800.000 | 6.800.000* |
| 1 | 140 | 140 | 4 | Bueno | mas de 30 | 460.000 COP | 5.071.428,57 | 5.071.428,57* |
| 1 | 90 | 0 | 4 | Bueno | mas de 30 | 369.000 COP | 4.000.000 | 4.000.000* |
| 2 | 111 | 111 | 6 | Bueno | 9 a 15 | 786.000 COP | 11.261.261,26 | 11.261.261,26* |
| 0 | 375 | 320 | 3 | Bueno | 16 a 30 | No definida | 1.493.333,33 | 1.493.333,33* |
| 2 | 111 | 73 | 4 | No definida | 16 a 30 | 483.000 COP | 3.423.423,42 | 3.423.423,42* |

6.521.739,13

| | D | E | F | G | H | I | J | K | L |
|---|---|---|---|---|---|---|---|---|---|
| | area_cons | area_priv | estrato | estado | antigüedad_años | administracion | precio_m2 | Ascensor | Circuito ce |
| 1 | 92 | 92 | 4 | No definida | 9 a 15 | 622.000 COP | 6.521.739,13 | 1 | 1 |
| 1 | 56 | 56 | 6 | No definida | 1 a 8 | 523.000 COP | 8.392.857,14 | 1 | 0 |
| 2 | 144 | 144 | 6 | No definida | 16 a 30 | 620.000 COP | 6.597.222,22 | 1 | 0 |
| 0 | 31 | 31 | 4 | Excelente | menor a 1 | 130.000 COP | 7.419.354,84 | 1 | 1 |
| 1 | 52 | 52 | 4 | No definida | 1 a 8 | 219.000 COP | 5.576.923,08 | 1 | 1 |
| 1 | 150 | 150 | 6 | Bueno | mas de 30 | 872.000 COP | 6.533.333,33 | 1 | 1 |
| 1 | 110 | 100 | 3 | No definida | 16 a 30 | 135.000 COP | 4.181.818,18 | 0 | 1 |
| 0 | 53 | 47 | 3 | No definida | 9 a 15 | 125.000 COP | 3.207.547,17 | 0 | 0 |
| 1 | 111 | 0 | 4 | Excelente | 16 a 30 | No definida | 3.873.873,87 | 0 | 1 |
| 2 | 264 | 264 | 5 | Bueno | mas de 30 | 836.000 COP | 5.303.030,3 | 1 | 0 |
| 2 | 97 | 0 | 4 | No definida | 9 a 15 | 272.000 COP | 6.175.257,73 | 1 | 1 |
| 2 | 87 | 82 | 4 | No definida | 9 a 15 | 350.000 COP | 4.712.643,68 | 1 | 1 |
| 2 | 175 | 175 | 4 | No definida | 16 a 30 | No definida | 4.285.714,29 | 0 | 0 |
| 2 | 391 | 0 | 3 | No definida | 16 a 30 | No definida | 1.994.884,91 | 0 | 0 |
| 3 | 218 | 0 | 4 | Bueno | mas de 30 | No definida | 3.027.522,94 | 0 | 1 |
| 1 | 50 | 0 | 5 | Bueno | 9 a 15 | 310.000 COP | 6.800.000 | 0 | 0 |
| 1 | 140 | 140 | 4 | Bueno | mas de 30 | 460.000 COP | 5.071.428,57 | 0 | 1 |
| 1 | 90 | 0 | 4 | Bueno | mas de 30 | 369.000 COP | 4.000.000 | 1 | 0 |
| 2 | 111 | 111 | 6 | Bueno | 9 a 15 | 786.000 COP | 11.261.261,26 | 1 | 1 |
| 0 | 375 | 320 | 3 | Bueno | 16 a 30 | No definida | 1.493.333,33 | 0 | 0 |
| 2 | 111 | 73 | 4 | No definida | 16 a 30 | 483.000 COP | 3.423.423,42 | 1 | 0 |

- Se eliminan errores tipograficos en la variable ubicacion



| Chimenea | Citofono | Cocina Int | Terraza | Vigilancia | Parques c | Estudio | Patio | Deposito | nombre | ubicacion | precio |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | Apartamento | Centro Internacional | 600000000 |
| 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | Apartamento | Calleja Baja | 470000000 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Apartamento | Cerros de Suba | 950000000 |
| 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Apartamento | Mazuren | 230000000 |
| 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Apartamento | El plan | 290000000 |
| 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | Apartamento | La Cabrera | 980000000 |
| 0 | 0 | 1 | 0 | 0 | 1 | 1 | 1 | 0 | Apartamento | Ciudad jardin sur | 460000000 |
| 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | Apartamento | Ub. industrial las delicias | 170000000 |
| 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | Casa | Pradera Norte | 430000000 |
| 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | Apartamento | Nicolas de federman | 1400000000 |
| 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Apartamento | Pontevedra | 599000000 |
| 1 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | Apartamento | Los cedros oriental | 410000000 |
| 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | Casa | Cedro Golf | 750000000 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Casa | Sosiego sur | 780000000 |
| 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | Casa | Normandia | 660000000 |
| 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | Apartamento | Contador | 340000000 |
| 1 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 0 | Apartamento | Quinta Paredes | 710000000 |
| 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | Apartamento | Cedro Golf | 360000000 |
| 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Apartamento | La Cabrera | 1250000000 |
| 0 | 0 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | Casa | Los Cerezos | 560000000 |
| 1 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | Apartamento | Cedro Golf | 380000000 |

housing_fincaraiz Hoja1

A continuación se anexa el archivo .csv  luego de la limpieza de datos:
[housing_fincaraiz_Lim]