

Ejercicio Obtención de datos públicos Observatorio Inmobiliario De Medellín (OIME)

1 Obtención de datos públicos

Para las distribuciones de frecuencia necesitamos datos. Ahora, busquemos datos que nos puedan servir para nuestro proceso de formación.

Ingresemos a la siguiente página web:

- <http://catastrooime.blogspot.com/>



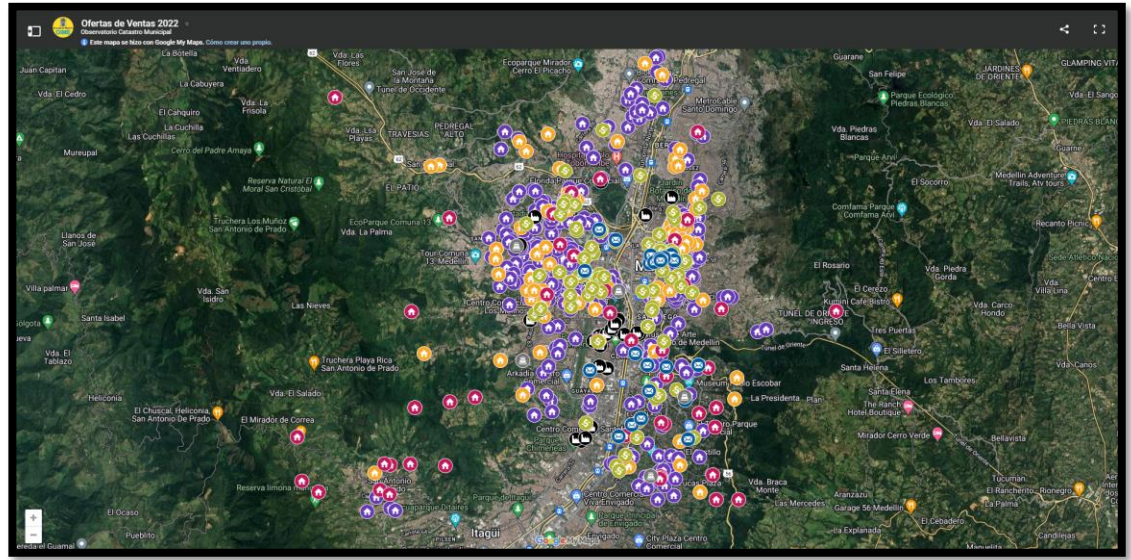
Clic en el botón Mapas de Ofertas de Ventas


Mapas de Ofertas de Ventas

Clic en el botón Información Capturada 2022

Información Capturada 2022

Este evento clic nos lleva a la siguiente página web:



-
- Clic en expandir 

Ahora daremos clic sobre los tres puntos en la parte superior izquierda del mapa

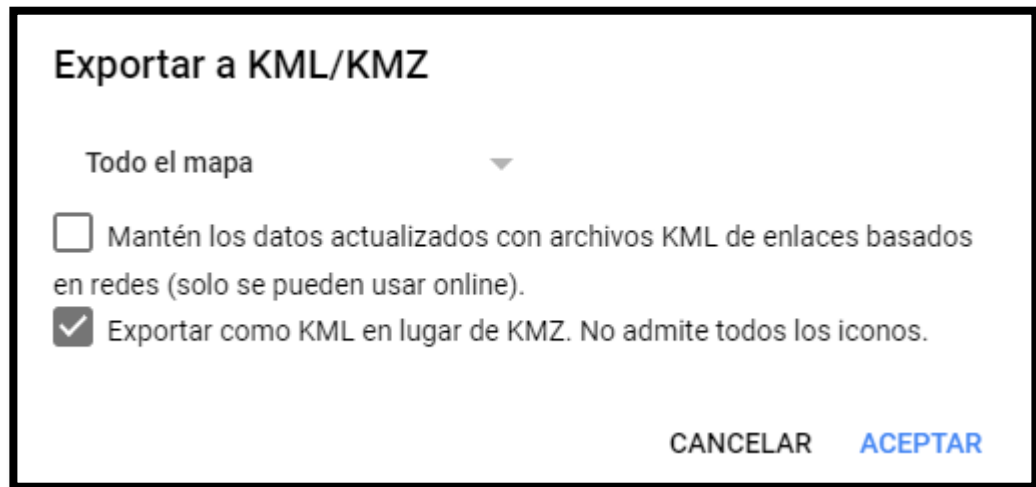


Clic en Descargar KML

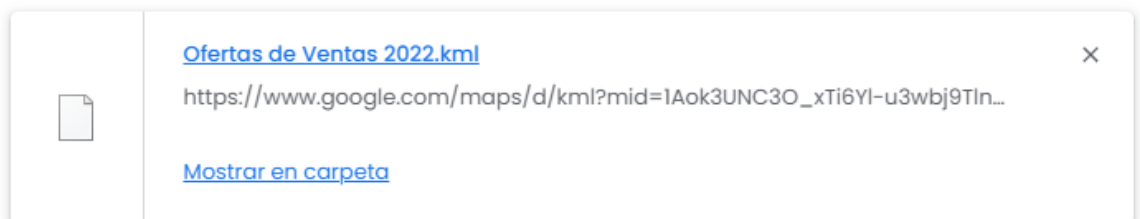



Marcamos la segunda opción y aceptar

3



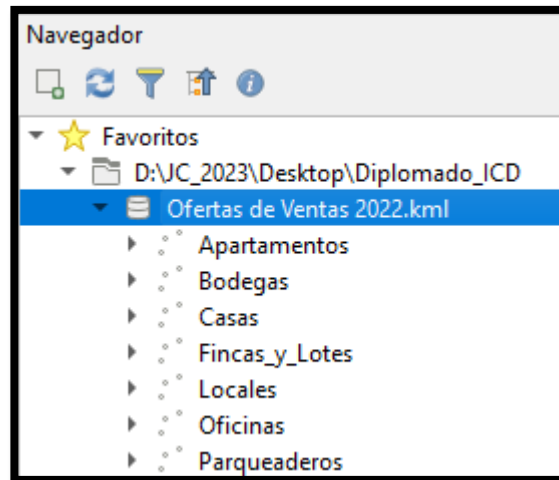
Verificamos el documento descargado en nuestra carpeta habitual de descargas



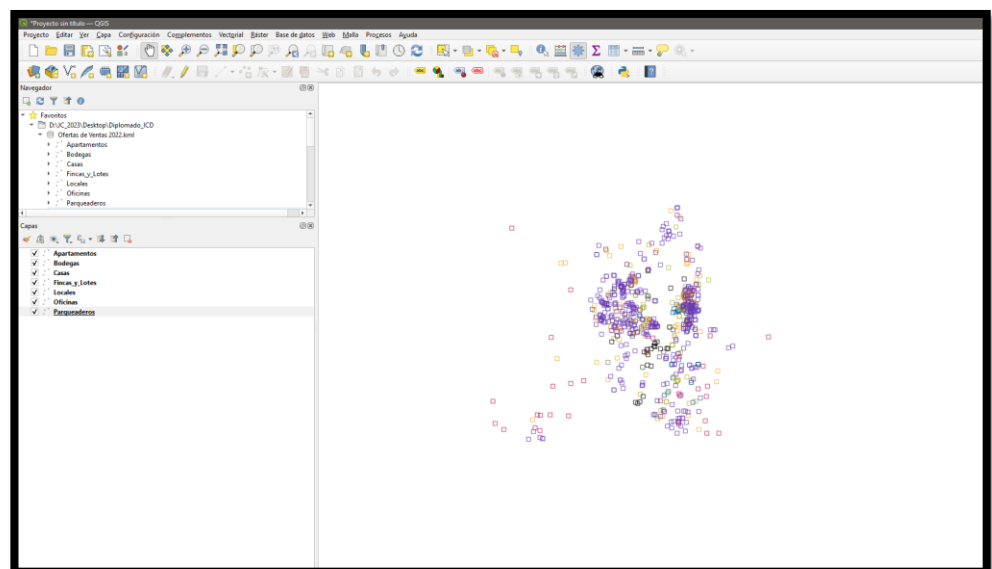
Nombre	Fecha de modificación	Tipo	Tamaño
 Ofertas de Ventas 2022.kml	5/07/2023 4:33 p. m.	Archivo KML	1.265 KB

Ahora, con la aplicación QGIS o ArcGIS procedemos a exportar estos datos a Excel, los pasos son los siguientes en QGIS, en ArcGIS son similares.

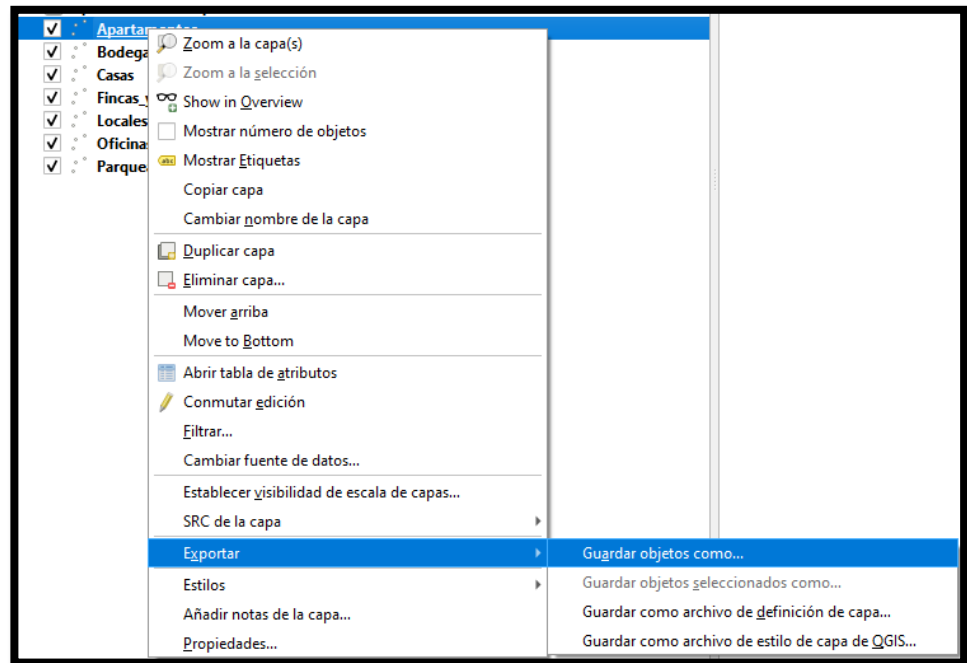
- Abrimos QGIS y anexamos una carpeta de datos en donde almacenaremos la información descargada, por ejemplo, la carpeta que muestro a continuación:



- Podemos cargar los datos que necesitemos simplemente con arrastras los puntos al panel de capas, para el presente caso agregaremos todos los valores.

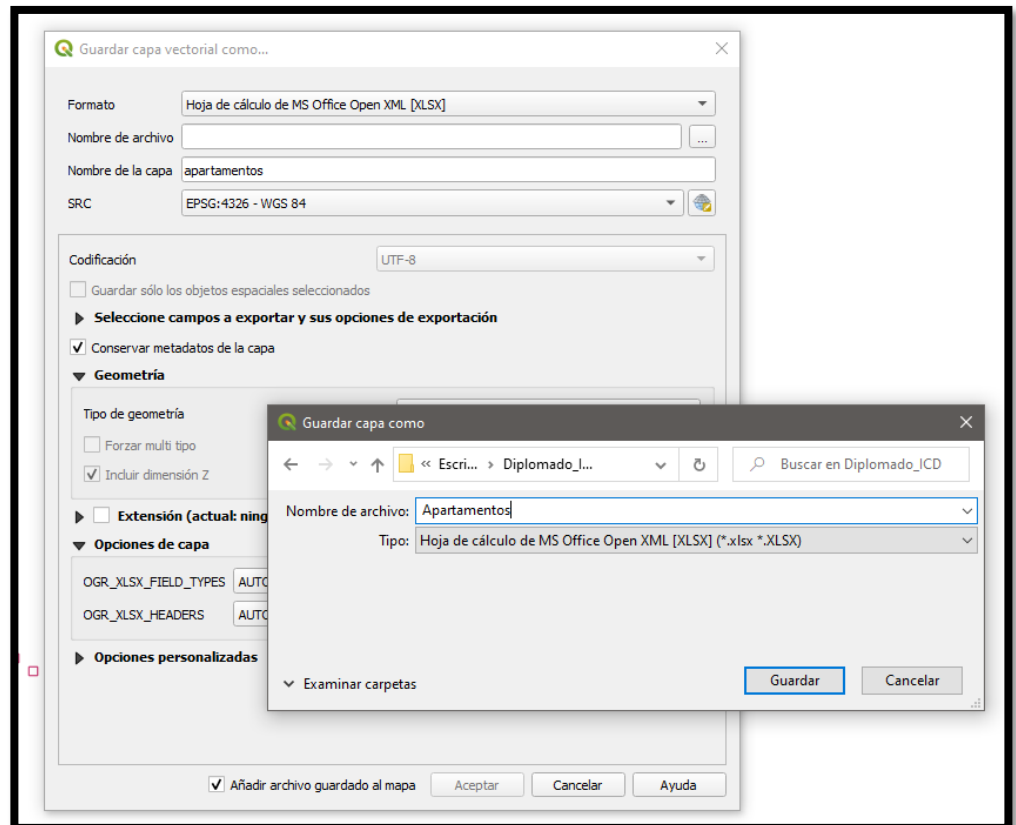


- Ahora en el panel de capas, con el clic derecho exportaremos los datos a archivos de Excel.



- Seleccionamos la opción MS Office y le asignamos el nombre a cada archivo y procedemos a guardarlos uno a uno.

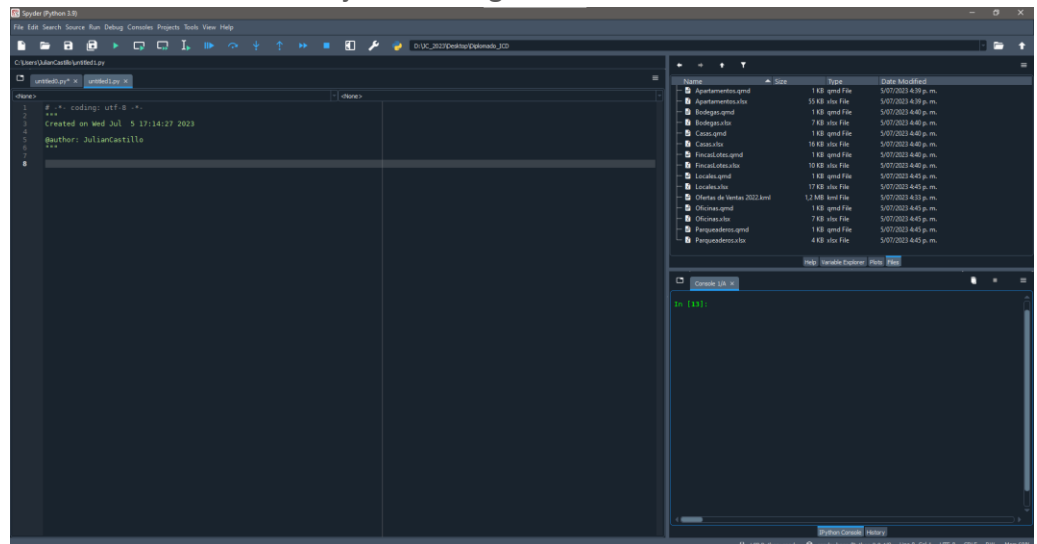
5



- Ahora con los documentos guardados debemos unificarlos en un solo documento. Esto podemos hacerlo de manera programática o de manera manual copiando y pegando datos. En el presente caso lo explicaré de manera programática.

Nombre	Fecha de modificación	Tipo	Tamaño
Apartamentos.qmd	5/07/2023 4:39 p. m.	Archivo QMD	2 KB
Apartamentos.xlsx	5/07/2023 4:39 p. m.	Hoja de cálculo d...	56 KB
Bodegas.qmd	5/07/2023 4:40 p. m.	Archivo QMD	2 KB
Bodegas.xlsx	5/07/2023 4:40 p. m.	Hoja de cálculo d...	8 KB
Casas.qmd	5/07/2023 4:40 p. m.	Archivo QMD	2 KB
Casas.xlsx	5/07/2023 4:40 p. m.	Hoja de cálculo d...	17 KB
FincasLotes.qmd	5/07/2023 4:40 p. m.	Archivo QMD	2 KB
FincasLotes.xlsx	5/07/2023 4:40 p. m.	Hoja de cálculo d...	11 KB
Locales.qmd	5/07/2023 4:45 p. m.	Archivo QMD	2 KB
Locales.xlsx	5/07/2023 4:45 p. m.	Hoja de cálculo d...	18 KB
Ofertas de Ventas 2022.kml	5/07/2023 4:33 p. m.	Archivo KML	1.265 KB
Oficinas.qmd	5/07/2023 4:45 p. m.	Archivo QMD	2 KB
Oficinas.xlsx	5/07/2023 4:45 p. m.	Hoja de cálculo d...	8 KB
Parqueaderos.qmd	5/07/2023 4:45 p. m.	Archivo QMD	2 KB
Parqueaderos.xlsx	5/07/2023 4:45 p. m.	Hoja de cálculo d...	5 KB

- Usando Spyder procederemos a crear un documento “.py” en el cual unificaremos los documentos.
 - Nuestro entorno de trabajo será el siguiente:



- Nuestro Script será el siguiente usando la función concat de pandas.

```
# -*- coding: utf-8 -*-
"""
Created on Wed Jul 5 16:47:50 2023

@author: JulianCastillo
"""
```

```
# -*- coding: utf-8 -*-
"""
Created on Thu Dec 22 17:53:50 2022
@author: JulianCastillo
"""

import pandas as pd
"""
Estas rutas se deben actualizar con los valores de tu
equipo de cómputo,
recuerda que Windows utiliza como separador de carpetas el
backslash \
y Python utiliza el slash como separador de carpetas /.
Si deseas convertir las rutas de Microsoft a rutas de
Python puedes utilizar
la librería "os" de Python, puedes consultar la
documentación en
https://docs.python.org/3/library/os.path.html
"""

Apartamentos = 'https://github.com/juliancastillo-
iudigital/DiplomadoICD/raw/main/Data/OIME/Apartamentos.xlsx'
Bodegas = 'https://github.com/juliancastillo-
iudigital/DiplomadoICD/raw/main/Data/OIME/Bodegas.xlsx'
Casas = 'https://github.com/juliancastillo-
iudigital/DiplomadoICD/raw/main/Data/OIME/Casas.xlsx'
FincasLotes = 'https://github.com/juliancastillo-
iudigital/DiplomadoICD/raw/main/Data/OIME/FincasLotes.xlsx'
Locales = 'https://github.com/juliancastillo-
iudigital/DiplomadoICD/raw/main/Data/OIME/Locales.xlsx'
Oficinas = 'https://github.com/juliancastillo-
iudigital/DiplomadoICD/raw/main/Data/OIME/Oficinas.xlsx'
Parqueadero = 'https://github.com/juliancastillo-
iudigital/DiplomadoICD/raw/main/Data/OIME/Parqueaderos.xlsx'

dfApartamentos = pd.read_excel(Apartamentos)
dfBodegas = pd.read_excel(Bodegas)
dfCasas = pd.read_excel(Casas)
dfFincasLotes = pd.read_excel(FincasLotes)
dfLocales = pd.read_excel(Locales)
dfOficinas = pd.read_excel(Oficinas) #Oficinas presenta
error, no cargaremos el dato
dfParqueadero = pd.read_excel(Parqueadero)

df = pd.concat([dfApartamentos, dfBodegas, dfCasas,
dfFincasLotes, dfLocales, dfParqueadero], axis=0)
df.rename(columns={'Age': 'Years'}, inplace=True)
cols = ['VALOR_COMERCIAL', 'VALOR_M']
```



```
df[cols] = df[cols].replace({'\$: ': ' ', ': ': ' '},
                             regex=True)

for idx, column in enumerate(df.columns):
    print(f"Column Name: {column}, Index: {idx}")
df.drop(df.columns[[1,2,3,4,5,6,7,8,9,10]], axis=1,
        inplace=True)

df.rename(columns={'FECHA_DE_INGRESO': 'Fecha'},
          inplace=True)
df.rename(columns={'TIPO_INVESTIGACION':
'TipoInvestigacion'}, inplace=True)
df.rename(columns={'TIPO_PREDIO': 'TipoPredio'},
          inplace=True)
df.rename(columns={'ESTADO': 'Estado'}, inplace=True)
df.rename(columns={'BARRIO': 'Barrio'}, inplace=True)
df.rename(columns={'ESTRATO': 'Estrato'}, inplace=True)
df.rename(columns={'AREA_PRIVADA': 'AreaPrivada'},
          inplace=True)
df.rename(columns={'AREA_LOTE': 'AreaLote'}, inplace=True)
df.rename(columns={'VALOR_COMERCIAL': 'ValorComercial'},
          inplace=True)
df.rename(columns={'VALOR_M__': 'ValorM2'}, inplace=True)
df.rename(columns={'LONGITUD': 'Longitud'}, inplace=True)
df.rename(columns={'LATITUD': 'Latitud'}, inplace=True)

df['ValorComercial'] = df['ValorComercial'].astype(float)
df['ValorM2'] = df['ValorM2'].astype(float)

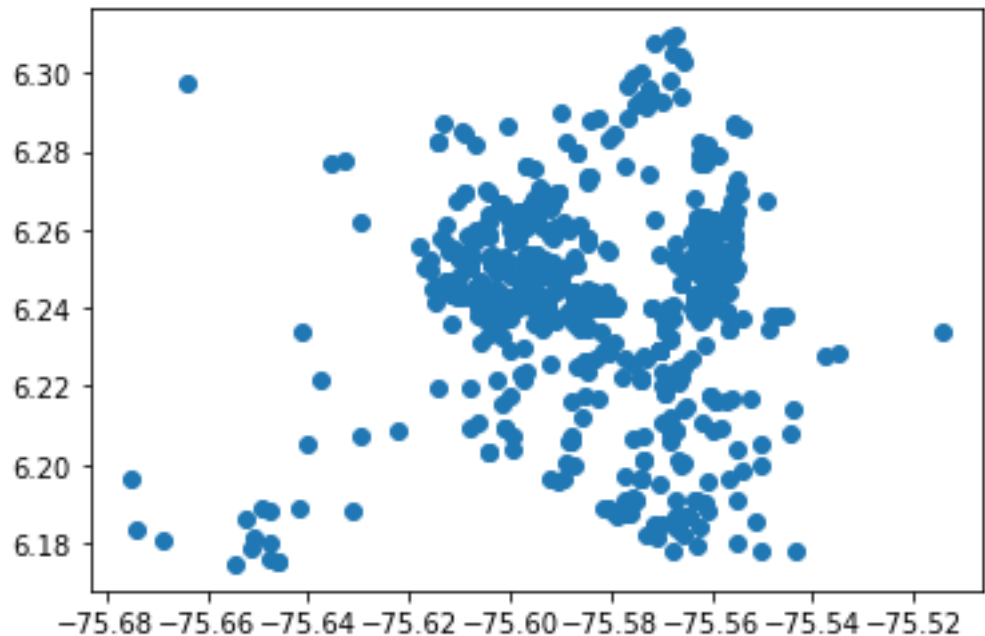
df['AreaPrivada'] = df['AreaPrivada'].replace({'0.0 M²':
' '}, regex=True)
df['AreaLote'] = df['AreaLote'].replace({'0.0 M²': ' '},
                                         regex=True)

df['AreaPrivada'] = df['AreaPrivada'].astype(int)
df['AreaLote'] = df['AreaLote'].astype(int)

df['ANIO'] = df.Fecha.str[6:]
df['ANIO'] = df['ANIO'].astype(int)
df['MES'] = df.Fecha.str[3:5]
df['MES'] = df['MES'].astype(int)
df['DIA'] = df.Fecha.str[:2]
df['DIA'] = df['DIA'].astype(int)

import matplotlib.pyplot as plt
plt.scatter(df['Longitud'], df['Latitud'])
```

- Ya tenemos nuestro documento y el grafico de los puntos detallados como se muestra a continuación:



-
- Ahora subiremos nuestro documento final al GitHub del curso, exportaremos el documento usando la función “to_csv” de Pandas.

```
df.to_csv('D:\JC_2023\Desktop\Diplomado_ICD\Documento.csv',
index=False)
```

- Nuestro documento final es de la siguiente forma:

```
1 Name, Fecha, TipoInvestigacion, TipoPredio, Estado, Barrio, Estrato, AreaPrivada, AreaLote, ValorComercial, ValorM2, Longitud
2 181451.0, 14-01-2022, VENTA, APARTAMENTO, USADO, BOMBONA # 1, 3.0, 177, 41, 3700000000.0, 2090395.0, -75.56157993, 6.24319669, 2
3 181452.0, 14-01-2022, VENTA, APARTAMENTO, USADO, BOMBONA # 1, 4.0, 104, 15, 3900000000.0, 3750000.0, -75.56196249, 6.24351282, 2
4 181456.0, 14-01-2022, VENTA, APARTAMENTO, USADO, BOSTON, 4.0, 188, 32, 3500000000.0, 1861702.0, -75.56000993, 6.24931447, 2022, 1
5 181466.0, 17-01-2022, VENTA, APARTAMENTO, USADO, BOSTON, 4.0, 150, 92, 2300000000.0, 1533333.0, -75.55802651, 6.24946633, 2022, 1
6 181558.0, 19-01-2022, VENTA, APARTAMENTO, USADO, PRADO, 4.0, 167, 71, 3000000000.0, 1796407.0, -75.56188185, 6.26222026, 2022, 1
7 181570.0, 19-01-2022, VENTA, APARTAMENTO, USADO, PRADO, 4.0, 146, 58, 3500000000.0, 2397260.0, -75.56313921, 6.26112071, 2022, 1
8 181573.0, 19-01-2022, VENTA, APARTAMENTO, USADO, TEJELO, 3.0, 142, 30, 3500000000.0, 2464789.0, -75.56772297, 6.30498407, 2022, 1
```

-
- Estos documentos separados por comas son más sencillos de procesar y son más ligeros dándonos flexibilidad para procesar los datos.
- La URL de la ubicación del documento es la siguiente:
- <https://raw.githubusercontent.com/juliancastillo-iudigital/DiplomadoICD/main/Data/OIME/Documento.csv>

○

Ya tenemos nuestros datos organizados y procesados, ahora procedamos con nuestros análisis estadísticos.

Nota: Tenemos un total de 727 registros, consideras que deberíamos ampliar la cantidad de datos a 2021 para tener mejor significancia?

10

```
len(df)
727
```

Fin.

Por: Julián Andrés Castillo G.