

## 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/>



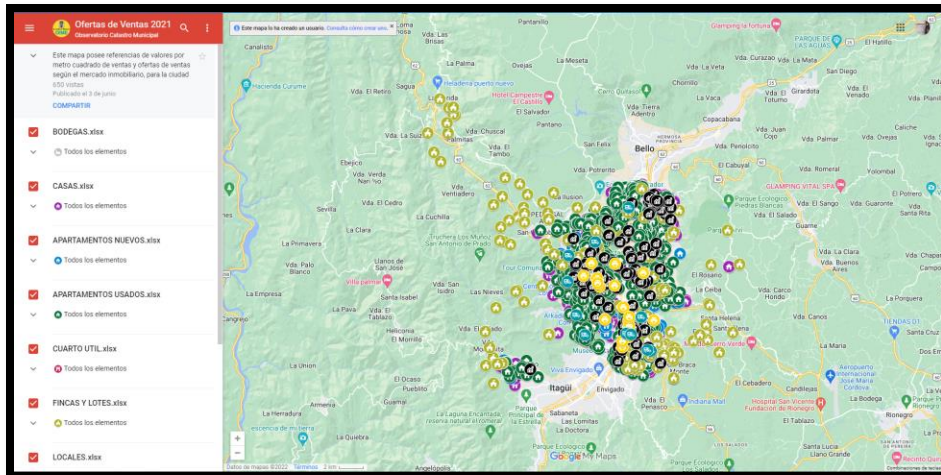
Clic en el botón Mapas de Ofertas de Ventas

Mapas de Ofertas de Ventas

Clic en el botón Información Capturada 2021

Información Capturada 2021

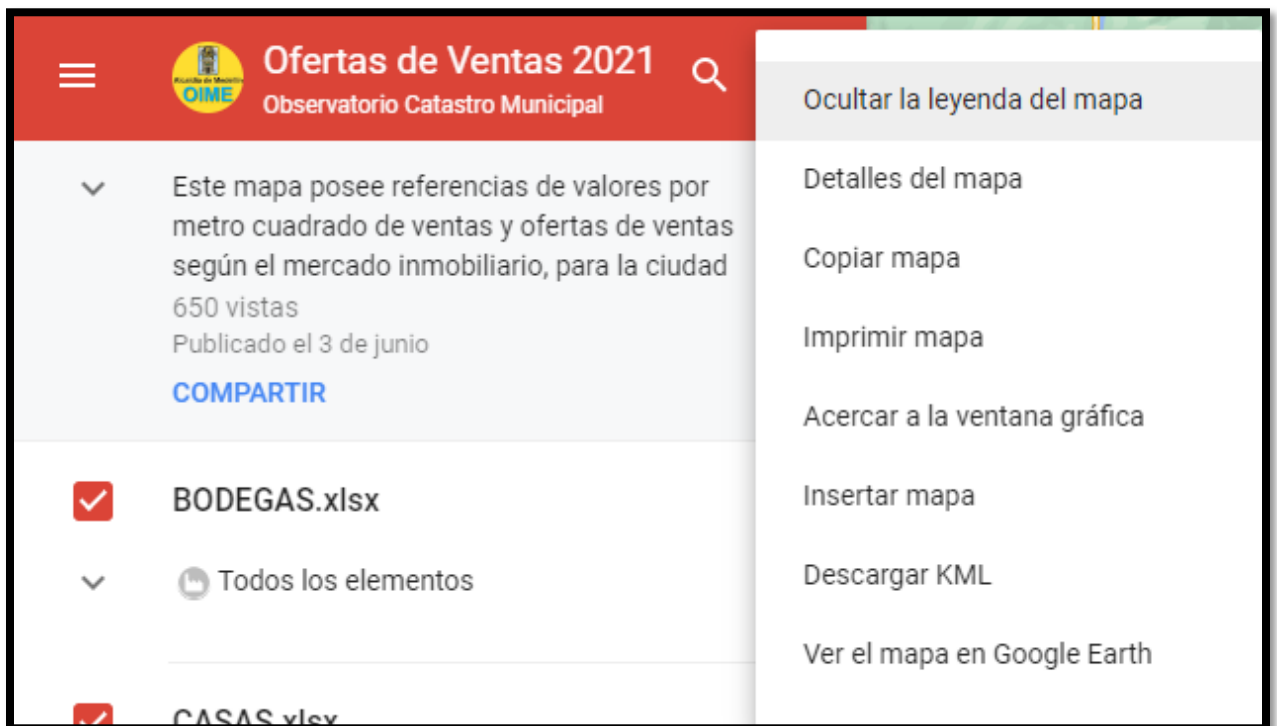
Este evento clic nos lleva a la siguiente pagina web:



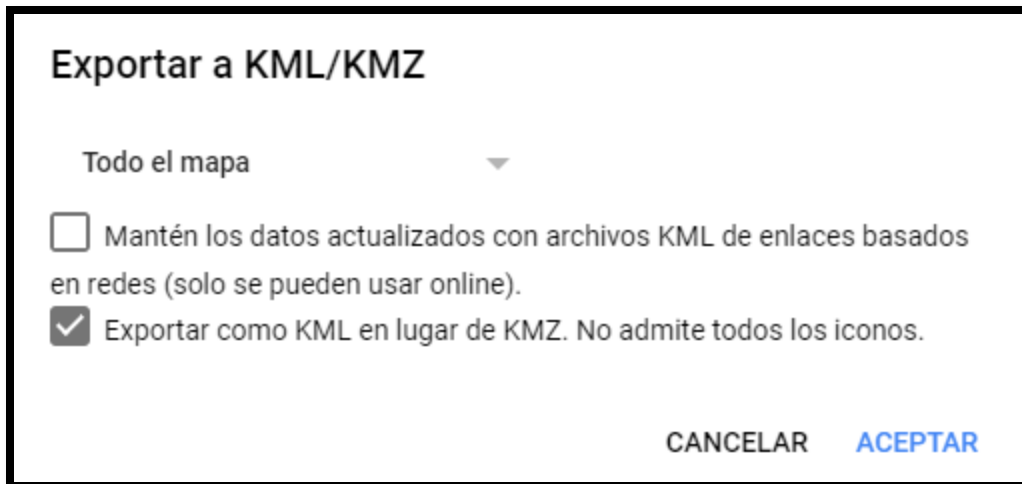
Ahora daremos clic sobre los tres puntos en la parte superior izquierda del mapa



Clic en Descargar KML



Marcamos la segunda opción y aceptar

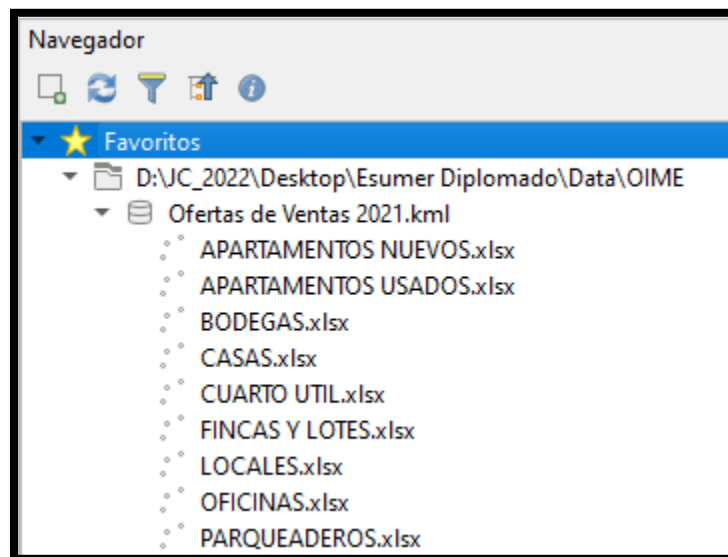


Verificamos el documento descargado en nuestra carpeta habitual de descargas

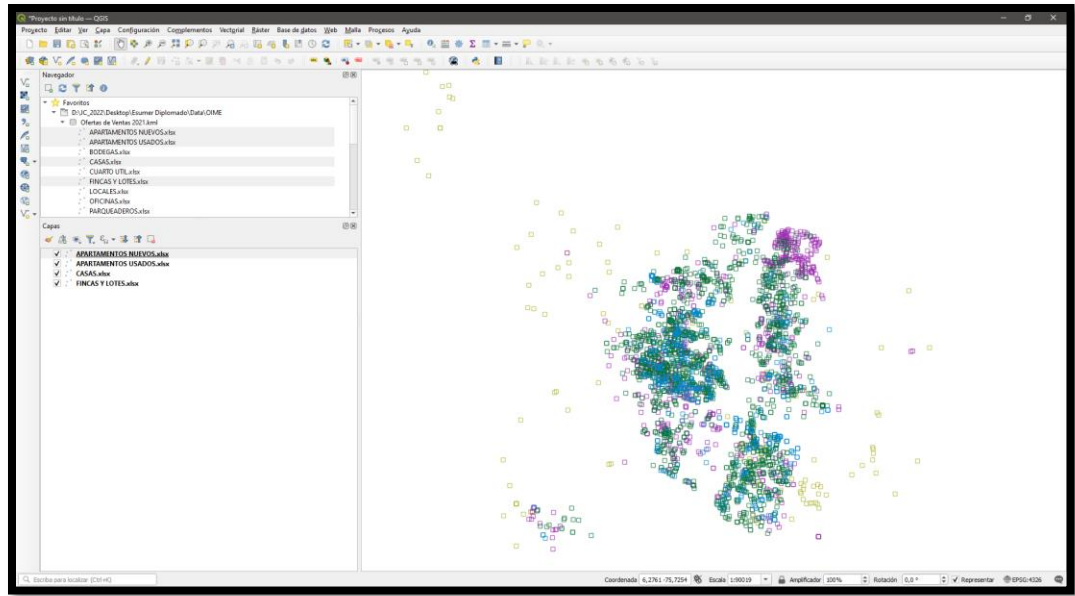
Name	Date modified	Type	Size
▼ Today (2)			
Ofertas de Ventas 2021.kml	22/12/2022 5:39 p. m.	KML File	7.067 KB

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.

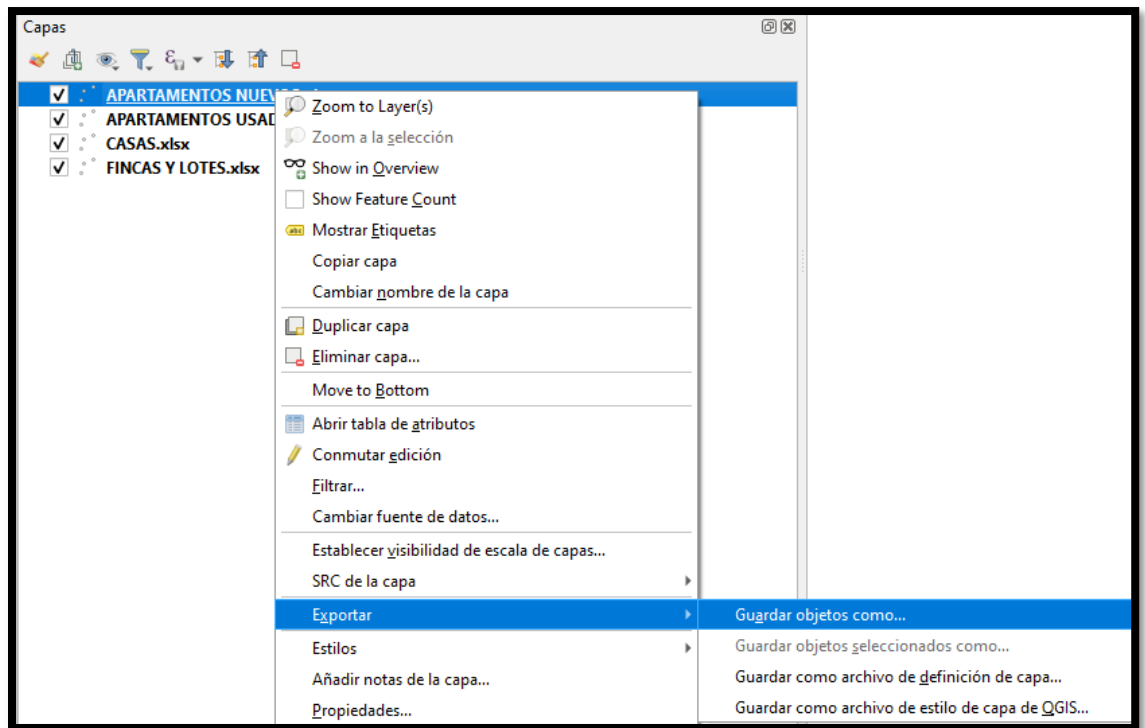
- Abrimos QGIS y anexamos una carpeta de datos en donde almacenaremos la información descargada, por ejemplo, la carpeta que muestro a continuación:



- Podemos cargar los datos que necesitemos simplemente con arrastras los puntos al panel de capas, para el presente caso agregaremos los valores de apartamentos nuevos, usados, casas y fincas y lotes.



- 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.

Guardar capa vectorial como...

Formato: Hoja de cálculo de MS Office Open XML [XLSX]

Nombre de archivo: D:\JC\_2022\Desktop\Esumer Diplomado\Data\OIME\AptosNuevos.xlsx

Nombre de la capa: AptosNuevos

SRC: EPSG:4326 - WGS 84

Codificación: UTF-8

☐ Guardar sólo los objetos espaciales seleccionados

► **Seleccione campos a exportar y sus opciones de exportación**

☒ Persist layer metadata

▼ **Geometría**

Tipo de geometría: Automático

☐ Forzar multi tipo

☒ Incluir dimensión Z

► ☐ **Extensión (actual: ninguno)**

▼ **Opciones de capa**

OGR\_XLSX\_FIELD\_TYPES: AUTO

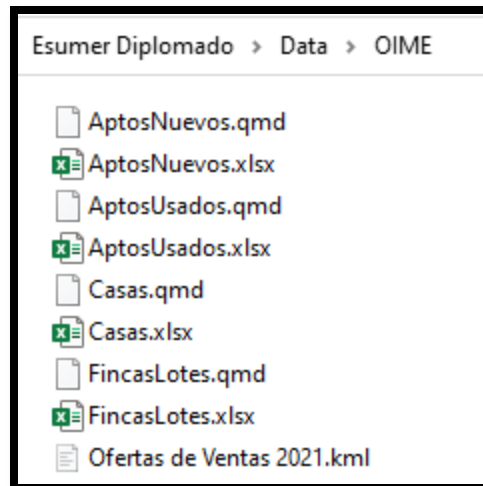
OGR\_XLSX\_HEADERS: AUTO

► **Opciones personalizadas**

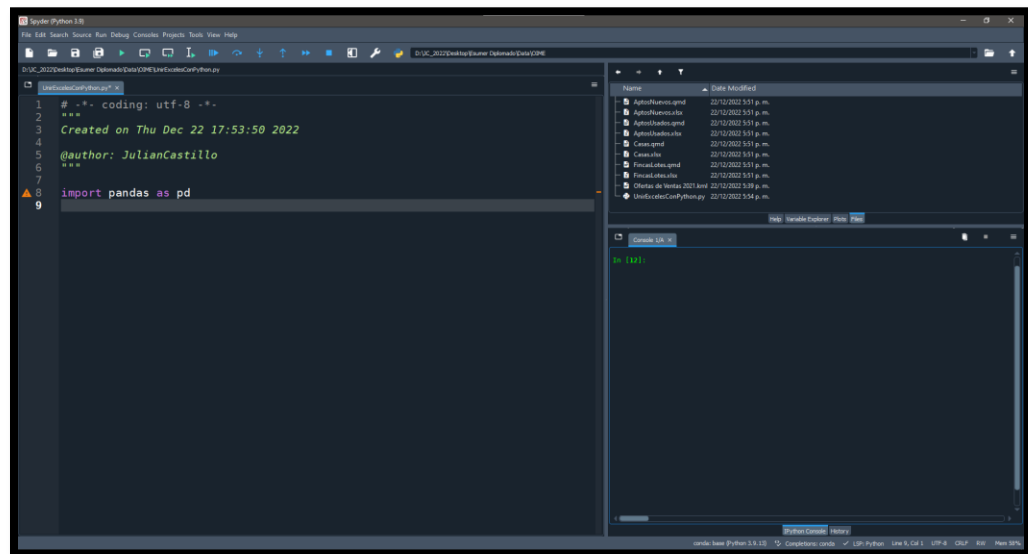
☒ Añadir archivo guardado al mapa

Aceptar Cancelar Ayuda

- 
- 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.



- 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 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
"""

aptosnuevos = pd.read_excel('d:/JC_2022/Desktop/Esumer
Diplomado/Data/OIME/AptosNuevos.xlsx')
```

```

aptosusados = pd.read_excel('d:/JC_2022/Desktop/Esuner
Diplomado/Data/OIME/AptosUsados.xlsx')
casas = pd.read_excel('d:/JC_2022/Desktop/Esuner
Diplomado/Data/OIME/Casas.xlsx')
fincaslot = pd.read_excel('d:/JC_2022/Desktop/Esuner
Diplomado/Data/OIME/FincasLotes.xlsx')

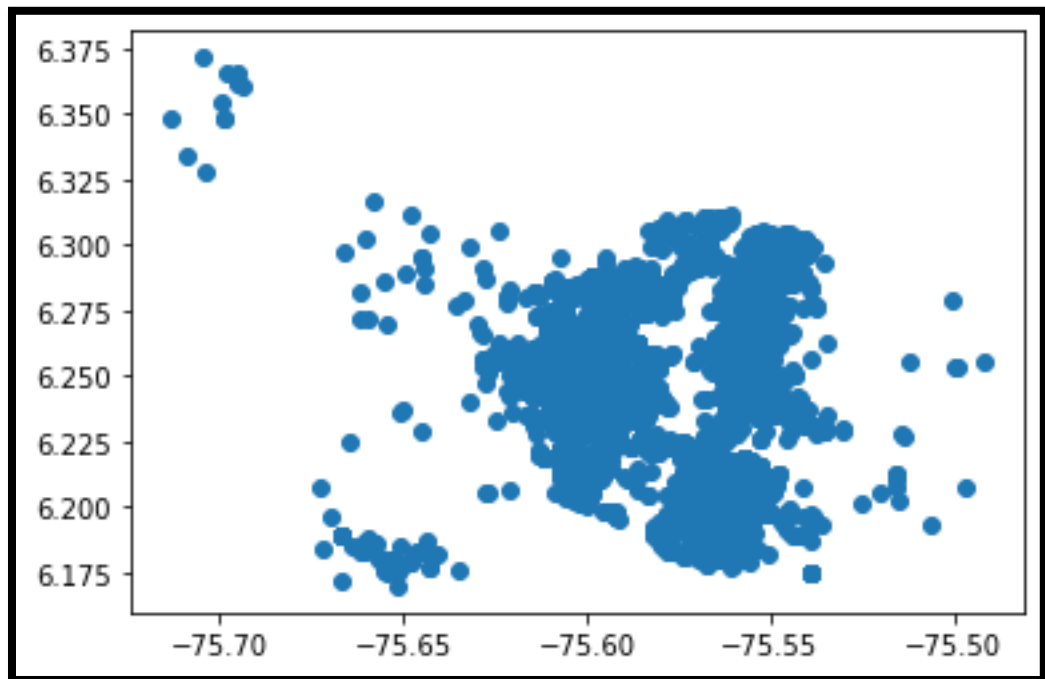
df = pd.concat([aptosnuevos, aptosusados, casas, fincaslot], axis=0)
cols = ['VALORCOMERCIAL', 'VALORM2']
df[cols] = df[cols].replace({'\': ''}, regex=True)

df['VALORCOMERCIAL'] = df['VALORCOMERCIAL'].astype(float)
df['VALORM2'] = df['VALORM2'].astype(float)
df.drop(df.columns[[1,2,3,4,5,6,7,8,9,10,21,22]], axis=1, inplace=True)
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 gráfico 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\_excel” de Pandas.

```

df.to_excel('d:/JC_2022/Desktop/Esuner
Diplomado/Data/OIME/DocumentoFinal.xlsx')

```

- Nuestro documento final es de la siguiente forma:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1		Name	FECHA	INVESTIGACION	PREDIO	ESTADO	BARRIO	ESTRATO	AREA_PRIVADA	AREA_LOTE	VALORCOMERCIAL	VALORM2	LONGITUD	LATITUD	ANIO	MES	DIA
2	0	171123	05-02-2021	VENTA	APARTAMENTO	NUEVO	LA COLINA	3	126	76	300000000	2380952	-75,59396631	6,19830091	2021	2	5
3	1	171239	10-02-2021	VENTA	APARTAMENTO	NUEVO	ZONA EXPANSION PAJARITO	3	47	14	150000000	3191489	-75,60971395	6,28561574	2021	2	10
4	2	171240	10-02-2021	VENTA	APARTAMENTO	NUEVO	ZONA EXPANSION PAJARITO	3	47	14	160000000	3404255	-75,60971395	6,28561574	2021	2	10
5	3	171241	10-02-2021	VENTA	APARTAMENTO	NUEVO	ZONA EXPANSION PAJARITO	3	47	14	150000000	3191489	-75,60971395	6,28561574	2021	2	10
6	4	171247	11-02-2021	VENTA	APARTAMENTO	NUEVO	ZONA EXPANSION PAJARITO	3	50	18	175000000	3500000	-75,60927162	6,28522799	2021	2	11
7	5	171248	11-02-2021	VENTA	APARTAMENTO	NUEVO	ZONA EXPANSION PAJARITO	3	50	18	170000000	3400000	-75,60927162	6,28522799	2021	2	11
8	6	171270	12-02-2021	VENTA	APARTAMENTO	NUEVO	ZONA EXPANSION PAJARITO	3	52	18	150000000	2884615	-75,60927162	6,28522799	2021	2	12
9	7	171271	12-02-2021	VENTA	APARTAMENTO	NUEVO	ZONA EXPANSION PAJARITO	3	52	17	155000000	2980769	-75,60927162	6,28522799	2021	2	12
10	8	171364	17-02-2021	VENTA	APARTAMENTO	NUEVO	SAN GERMAN	4	71	16	340000000	4788732	-75,59139901	6,27051874	2021	2	17
11	9	171365	17-02-2021	VENTA	APARTAMENTO	NUEVO	SAN GERMAN	4	64	15	310000000	4843750	-75,59086143	6,26926496	2021	2	17
12	10	171366	17-02-2021	VENTA	APARTAMENTO	NUEVO	SAN GERMAN	4	67	15	320000000	4776119	-75,59139901	6,27051874	2021	2	17
13	11	171372	17-02-2021	VENTA	APARTAMENTO	NUEVO	SAN GERMAN	4	64	15	292000000	4562500	-75,59086143	6,26926496	2021	2	17
14	12	171374	17-02-2021	VENTA	APARTAMENTO	NUEVO	SAN GERMAN	4	86	15	416000000	4837209	-75,59007	6,27062346	2021	2	17
15	13	171378	17-02-2021	VENTA	APARTAMENTO	NUEVO	SAN GERMAN	4	75	13	365000000	4866667	-75,59007	6,27062346	2021	2	17
16	14	171384	17-02-2021	VENTA	APARTAMENTO	NUEVO	SAN GERMAN	4	70	20	320000000	4571429	-75,58930377	6,26874735	2021	2	17
17	15	171390	17-02-2021	VENTA	APARTAMENTO	NUEVO	SAN GERMAN	4	48	13	250000000	5208333	-75,59178686	6,27006898	2021	2	17
18	16	171393	17-02-2021	VENTA	APARTAMENTO	NUEVO	SAN GERMAN	4	63	19	330000000	5238095	-75,59178686	6,27006898	2021	2	17
19	17	171396	17-02-2021	VENTA	APARTAMENTO	NUEVO	SAN GERMAN	4	64	17	270000000	4218750	-75,59178686	6,27006898	2021	2	17
20	18	171399	17-02-2021	VENTA	APARTAMENTO	NUEVO	SAN GERMAN	4	89	15	430000000	4831461	-75,59043285	6,26973696	2021	2	17

- 
- Guardemos el documento como un archivo plano, en Excel podemos abrir el documento y con la opción guardar como, seleccionamos CSV (Comma separated values, valores separados por comas).

Nombre de archivo: DocumentoFinal.csv

Tipo: CSV (MS-DOS) (\*.csv)

- 
- Estos documentos separados por comas son mas 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-udea/EsumerDiplomado2022-1/main/Data/DocumentoFinal.csv>
- Es de anotar que nuestro documento al ser almacenado por Windows convierte los separadores de comas por punto y coma como se muestra a continuación:

https://raw.githubusercontent.com	x	+
←	→	↻
raw.githubusercontent.com/juliancastillo-udea/EsumerDiplomado2022-1/main/Data/DocumentoFinal.csv		
Name; FECHA; INVESTIGACION; PREDIO; ESTADO; BARRIO; ESTRATO; AREA_PRIVADA; AREA_LOTE; VALORCOMERCIAL; VALORM2; LONGITUD; LATITUD; ANIO; MES; DIA		
171123;05-02-2021;VENTA	; APARTAMENTO	; NUEVO ; LA COLINA;3;126;76;300000000;2380952;-75,59396631;6,19830091;2021;2;5
171239;10-02-2021;VENTA	; APARTAMENTO	; NUEVO ; ZONA EXPANSION PAJARITO;3;47;14;150000000;3191489;-75,60971395;6,28561574;2021;2;10
171240;10-02-2021;VENTA	; APARTAMENTO	; NUEVO ; ZONA EXPANSION PAJARITO;3;47;14;160000000;3404255;-75,60971395;6,28561574;2021;2;10
171241;10-02-2021;VENTA	; APARTAMENTO	; NUEVO ; ZONA EXPANSION PAJARITO;3;47;14;150000000;3191489;-75,60971395;6,28561574;2021;2;10
171247;11-02-2021;VENTA	; APARTAMENTO	; NUEVO ; ZONA EXPANSION PAJARITO;3;50;18;175000000;3500000;-75,60927162;6,28522799;2021;2;11
171248;11-02-2021;VENTA	; APARTAMENTO	; NUEVO ; ZONA EXPANSION PAJARITO;3;50;18;170000000;3400000;-75,60927162;6,28522799;2021;2;11
171270;12-02-2021;VENTA	; APARTAMENTO	; NUEVO ; ZONA EXPANSION PAJARITO;3;52;18;150000000;2884615;-75,60927162;6,28522799;2021;2;12
171271;12-02-2021;VENTA	; APARTAMENTO	; NUEVO ; ZONA EXPANSION PAJARITO;3;52;17;155000000;2980769;-75,60927162;6,28522799;2021;2;12
171364;17-02-2021;VENTA	; APARTAMENTO	; NUEVO ; SAN GERMAN;4;71;16;340000000;4788732;-75,59139901;6,27051874;2021;2;17
171365;17-02-2021;VENTA	; APARTAMENTO	; NUEVO ; SAN GERMAN;4;64;15;310000000;4843750;-75,59086143;6,26926496;2021;2;17
171366;17-02-2021;VENTA	; APARTAMENTO	; NUEVO ; SAN GERMAN;4;67;15;320000000;4776119;-75,59139901;6,27051874;2021;2;17
171372;17-02-2021;VENTA	; APARTAMENTO	; NUEVO ; SAN GERMAN;4;64;15;292000000;4562500;-75,59086143;6,26926496;2021;2;17
171374;17-02-2021;VENTA	; APARTAMENTO	; NUEVO ; SAN GERMAN;4;86;15;416000000;4837209;-75,59007;6,27062346;2021;2;17
171378;17-02-2021;VENTA	; APARTAMENTO	; NUEVO ; SAN GERMAN;4;75;13;365000000;4866667;-75,59007;6,27062346;2021;2;17
171384;17-02-2021;VENTA	; APARTAMENTO	; NUEVO ; SAN GERMAN;4;70;20;320000000;4571429;-75,58930377;6,26874735;2021;2;17
171390;17-02-2021;VENTA	; APARTAMENTO	; NUEVO ; SAN GERMAN;4;48;13;250000000;5208333;-75,59178686;6,27006898;2021;2;17
171393;17-02-2021;VENTA	; APARTAMENTO	; NUEVO ; SAN GERMAN;4;63;19;330000000;5238095;-75,59178686;6,27006898;2021;2;17
171396;17-02-2021;VENTA	; APARTAMENTO	; NUEVO ; SAN GERMAN;4;64;17;270000000;4218750;-75,59178686;6,27006898;2021;2;17
171399;17-02-2021;VENTA	; APARTAMENTO	; NUEVO ; SAN GERMAN;4;89;15;430000000;4831461;-75,59043285;6,26973696;2021;2;17
171464;18-02-2021;VENTA	; APARTAMENTO	; NUEVO ; PALENQUE;4;38;9;120000000;3157895;-75,59637123;6,27927074;2021;2;18
171465;18-02-2021;VENTA	; APARTAMENTO	; NUEVO ; PALENQUE;4;37;9;115000000;3108108;-75,59637123;6,27927074;2021;2;18
171846;13-04-2021;VENTA	; APARTAMENTO	; NUEVO ; EL ESTADIO;5;67;25;380000000;5671642;-75,5920807;6,25877456;2021;4;13

- 
- Este cambio debemos ajustarlo en nuestro documento cambiando el separador de los datos con el argumento "sep=';'".

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

Fin.

Por: Julián Andrés Castillo G.